The Structure and Internal Dynamics of Polypeptides Probed with Pyrene Excimer Fluorescence

by

Remi Julian Ronald Casier

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Chemistry

Waterloo, Ontario, Canada, 2021

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Ken A. Dill Professor Chemistry, Stony Brook University
Supervisor	Jean Duhamel Professor Chemistry, University of Waterloo
Internal Member	Mario Gauthier Professor Chemistry, University of Waterloo
Internal Member	Xiaosong Wang Professor Chemistry, University of Waterloo
Internal-External Member	Michael Tam Professor Chemical Engineering, University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

I completed all work and writing presented in this thesis with regular consultation with my supervisor, Jean Duhamel. This thesis consists in part of six manuscripts written for publication.

Chapter 2: Casier, R. and Duhamel, J. Pyrene Excimer Fluorescence as a Direct and Easy Experimental Means to Characterize the Length Scale and Internal Dynamics of Polypeptide Foldons. *Macromolecules* **2018**, *51*, 3450–3457. DOI: 10.1021/acs.macromol.8b00459

Chapter 3: Casier, R. and Duhamel J. Effect of Structure on Polypeptide Blobs: A Model Study Using Poly(L-lysine). *Langmuir* **2020**, *36*, 7980–7990. DOI: 10.1021/acs.langmuir.0c01347

Chapter 4: Casier, R. and Duhamel J. Effect of Like Charges on the Conformation and Internal Dynamics of Polypeptides Probed by Pyrene Excimer Fluorescence. *Macromolecules* **2020**, *53*, 5147–5157. DOI: 10.1021/acs.macromol.0c00836

Chapter 5: Casier, R. and Duhamel, J. The Effect of Amino Acid Size on the Internal Dynamics and Conformational Freedom of Polypeptides. *Macromolecules* **2020**, *53*, 9811 – 9822. DOI: 10.1021/acs.macromol.0c02153.

Chapter 6: Casier, R. and Duhamel, J. *Blob*-Based Approach to Estimate the Folding Time of Proteins Supported by Pyrene Excimer Fluorescence Experiments. *Macromolecules* **2020**, *53*, 9823 – 9853. DOI: 10.1021/acs.macromol.0c02201

Chapter 7: Casier, R. and Duhamel, J. *Blob*-Based Predictions of Protein Folding Times from the Amino Acid-Dependent Conformation of Polypeptides in Solution. *Macromolecules* **2021**, *ASAP*. DOI: 10.1021/acs.macromol.0c02617

Abstract

A combination of pyrene excimer fluorescence, fluorescence *blob* model (FBM), and molecular mechanics optimizations (MMOs) was applied to characterize a series of pyrene-labeled polypeptides. Analysis of the fluorescence decays with the FBM yielded information on the number (N_{blob}^{exp}) of amino acids (*aa*'s) contained within the volume of a *blob* and the rate constant (k_{blob}) at which two *aa*'s labeled with pyrene encountered one another inside a *blob*. The N_{blob}^{exp} value was used as a structural parameter which was compared to the N_{blob}^{theo} value obtained from MMOs conducted on constructs of pyrene-labeled polypeptides to determine the conformation adopted by the rigid polypeptides in solution.

The relationship that exists between N_{blob}^{exp} and macromolecular conformation was established by using a series of pyrene-labeled homopolypeptides, namely poly(*L*-glutamic acid) (P*L*Glu), poly(*D*,*L*-glutamic acid) (P*DL*Glu), and poly(*L*-lysine) (P*L*L). The coiled conformations of P*DL*Glu and P*L*L yielded the smallest *blob* sizes, and thus the smallest N_{blob}^{exp} values, due to their elongated conformation on the length-scale of a *blob*. N_{blob}^{exp} was found to be significantly larger, when P*L*Glu and P*L*L adopted an α-helical conformation, indicating that the *blob* size was directly related to conformational density. This information was then used to identify the unknown conformation of P*L*Glu in DMSO and determine the interhelical distance in bundles of P*L*L α-helices. To study the effect of ionic interactions, the protonated and deprotonated forms of P*DL*Glu and P*L*Glu were compared. N_{blob}^{exp} was found to remain constant, indicating that the temporal window provided by an excited pyrene was

sufficient to study the rigid polypeptides even when their dynamics were slowed by the repulsion of like charges.

Next, the effect that the *aa* sequence had on N_{blob}^{exp} and the internal dynamics ($k_{blob} \times N_{blob}^{exp}$) of a coiled polypeptide was investigated. To do this, a series of pyrene-labeled copolypeptides were synthesized through the ring opening polymerization of the *N*-carboxyanhydride monomers obtained from racemic mixtures of glutamic acid and either glycine, alanine, or carbobenzyloxylysine (Lys(Z)) to yield PGlyGlu, PAlaGlu, and PLys(Z)Glu, respectively. Both N_{blob}^{exp} and $k_{blob} \times N_{blob}^{exp}$ increased with decreasing *aa*'s side chain size (SCS). This indicated that the reduced steric hindrance generated by the smaller SCSs increased the internal dynamics and conformational freedom of the polypeptides. The effect, that the degree of comonomer incorporation had on the chain dynamics and internal density of polypeptides, was determined with a series of PAlaGlu samples prepared with alanine contents between 24 and 58 mol%. The constant N_{blob}^{exp} value obtained with these samples suggested that the *blob* size was independent of the comonomer composition and instead depended only on the presence of a few small *aa*'s. Lastly, the relationship between N_{blob}^{exp} and the side chain length (SCL) linking pyrene to the polypeptide was determined using a series of P*DL*Glu and P*L*L samples adopting a coiled conformation. This allowed for the quantitative determination of the contribution of the pyrenyl label to N_{blob}^{exp}. After correcting for its contribution and taking into account the effect of SCS and SCL on N_{blob}^{exp}, a program was implemented to estimate the *blob* size of each *aa* depending on the *aa* sequence in a heterogeneous polypeptide, such as a protein. This program was employed to determine the folding time of 145 proteins segmented into *blob*s. Comparison of the calculated and

experimental folding times yielded a correlation coefficient of 0.73, demonstrating the ability of *blob*-based approaches, such as the ones described in this thesis, to predict the folding time of proteins.

In summary, this thesis has introduced a novel *blob*-based approach to predict the folding time of proteins based on parameters that can be determined experimentally from the use of pyrene excimer fluorescence and racemic polypeptides. It opens an alternate path toward developing a better understanding for proteins on *how to fold graciously* (C. Levinthal).

Acknowledgements

I would like to extend my deepest gratitude to my supervisor Prof. Jean Duhamel for his invaluable contributions, support, and guidance throughout my degree. Without his continued mentorship, I would not be where I am today.

I would like to thank my advisory committee members Prof. Mario Gauthier, Prof. Xiaosong Wang, and Prof. Michael Tam for their valuable advice and discussions.

I would also like to thank my current and former peers in the Duhamel and Gauthier groups for all their contributions and great memories.

I would also like to thank my family for their unconditional support during my time at Waterloo.

Last, but certainly not least, I would like to thank WIN, IPR, OGS, NSERC, GWC^2, and the University of Waterloo for financial support.

Table of Contents

List of Figures

List of Tables

List of Schemes

List of Abbreviations

2D	two-dimensional
3D	three-dimensional
aa	amino acid
ACN	acetonitrile
AFM	atomic force microscopy
Ala	*D,L*-alanine
CD	circular dichroism
DMF	*N,N*-dimethylformamide
DMSO	dimethyl sulfoxide
DP	degree of polymerization
EEC	end-to-end cyclization
FBM	fluorescence *blob* model
FRET	Förster resonance energy transfer
Glu	glutamic acid
Gly	glycine
GPC	gel permeation chromatography
H-bond	hydrogen bond
HCP	hexagonal close packed
HX	hydrogen exchange
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
IRF	instrument response function
Lys(Z)	*D,L-N-ε*-carbobenzyloxylysine
MALS	multi angle light scattering
MMO	molecular mechanics optimization
MS	mass spectroscopy

NCA	amino acid *N*-carboxy anhydride
NMR	nuclear magnetic resonance
Np	Naphthalene
PAlaGlu	poly(*D*,*L*-alanine-*co*-*D*,*L*-glutamic acid)
PAlaGlu(OtBu)	poly(*D*,*L*-alanine-*co*-*D*,*L*-glutamic acid 5-*tert*-butyl ester)
PAlaGNa	sodium salt of poly(*D*,*L*-alanine-*co*-*D*,*L*-glutamic acid)
$PAla_xGlu_y$	poly(*D*,*L*-alanine-*co*-*D*,*L*-glutamic acid) with *x* and *y* mol% alanine and glutamic acid, respectively
PAMA	poly(*n*-alkyl methacrylate)
P*DL*Glu	poly(*D*,*L*-glutamic acid)
P*DL*GNa	sodium salt of poly(*D*,*L*-glutamic acid)
PEF	pyrene excimer formation/fluorescence
PGlu	poly(glutamic acid)
PGlyGlu	poly(glycine-*co*-*D*,*L*-glutamic acid)
PGlyGlu(OtBu)	poly(glycine-*co*-*D*,*L*-glutamic acid 5-*tert*-butyl ester)
PGNa	sodium salt of poly(glutamic acid)
P*L*Glu	poly(*L*-glutamic acid)
P*L*GNa	sodium salt of poly(*L*-glutamic acid)
P*L*L	poly(*L*-lysine)
PLys(Z)Glu	poly(*D*,*L*-*N*-ε-carbobenzyloxylysine-*co*-*D*,*L*-glutamic acid)
PLys(Z)Glu(OtBu)	poly(*D*,*L*-*N*-ε-carbobenzyloxylysine-*co*-*D*,*L*-glutamic acid 5-*tert*-butyl ester)
PPII	poly(proline) type II helix
Pro	proline
Py	pyrene
Py-	pyrene-labeled
PyAc-	pyreneacetic acid-labeled

PyBu-	pyrenebutyric acid-labeled
PyLM	pyrene-labeled macromolecule
PyMe-	pyrenemethylamine-labeled
SCL	side chain length
SCS	side chain size
SPC	single photon counting
SSF	steady-state fluorescence
SU	structural unit
THF	tetrahydrofuran
TRF	time-resolved fluorescence
UV	ultraviolet

List of Symbols

$[\eta]$	intrinsic viscosity
$[Py]_{loc}$	local pyrene concentration
$[\theta]$	molar ellipticity
$<n>$	average number of pyrenes contained within each *blob*
1^o	primary structure
2^o	secondary structure
3^o	tertiary structure
4^o	quaternary structure
C_∞	characteristic ratio
C_α	alpha-carbon
$Đ$	dispersity
D^*	poorly stacked pyrene excimer which emits with the lifetime τ_D
$d_{h\text{-}h}$	interhelix distance in bundled PLLs
dn/dc	refractive index increment
$E0^*$	properly stacked pyrene excimer which emits with the lifetime tE0
f_{Ala}	molar fraction of alanine
f_b(SCS)	bending function
f_{Mfree}	molar fraction of pyrenes which emit with the lifetime τ_M in the monomer decay
I_E	steady-state excimer fluorescence intensity
I_E/I_M	ratio of fluorescence intensities of excimer and monomer
I_M	steady-state monomer fluorescence intensity
k_2	rapid excimer formation rate constant
k_{blob}	pseudo-unimolecular rate constant of excimer formation in a *blob* containing a single excited and ground-state pyrene

$k_{blob} \times N_{blob}$	product of k_{blob} and N_{blob}, equivalent to the number of encounters per unit time between the structural units in a *blob*
k_{cy}	end-to-end cyclization rate constant
k_{diff}	bimolecular rate constant for diffusive encounters
M_n	number average molecular weight
N(SCL)	intrinsic N_{blob} value of an *aa* based on its side chain length
N_0	maximum number of structural units separating two pyrenyl labels while still maintaining good overlap
N_0^l	N_0 to the left side of the reference amino acid inside a *blob*
N_0^r	N_0 to the right side of the reference amino acid inside a *blob*
n_b	number of *blobs*
N_{blob}	number of structural units of a macromolecule contained in a *blob*
N_{blob}^{exp}	experimentally obtained N_{blob} value
N_{blob}^{theo}	N_{blob} value determined from MMOs
N_{fold}	number of amino acid residues which cooperatively fold together in a *foldon*
p	probability for an excited and ground-state pyrene to form excimer upon encounter
Py^*	excited pyrene species
Py_{diff}^*	excited pyrene species which form excimer by diffusion
Py_{free}^*	excited pyrene species which do not form excimer
Py_{k2}^*	excited pyrene species which form excimer with the rate constant k_2
r	correlation coefficient
τ_D	lifetime of poorly stacked excimer
τ_{E0}	lifetime of properly stacked excimer
τ_f	protein folding time
τ_M	lifetime of unquenched pyrene monomer
τ_p	conformational search time for one *aa*

V_{blob}	volume of a blob
x	molar fraction of structural units labeled with pyrene
x_{Np}	molar fraction of structural units labeled with naphthalene
$\Delta(\Delta G)$	difference in Gibbs free energy
ΔG	Gibbs free energy
ΔN_{blob}	change in N_{blob} due to intermolecular interactions
ΔN_{blob}^{theo}	theoretical change in N_{blob} due to intermolecular interactions determined with MMOs
ε	molar absorptivity coefficient
η	viscosity
θ, φ, ω	amino acid torsion angles
λ_{Py}	moles of pyrene per mass of macromolecule
ρ_p	protein hydrodynamic density
Ω	number of conformations
Ω_B	number of conformations in a *blob*
Ω_M	number of conformations among *blobs*
Ω_{Total}	total number of conformations in a protein segmented into *blobs*
Ω_u	number of unique conformations

Chapter 1

Introduction

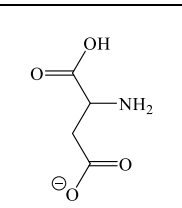
1.1 The Structure of Proteins

Proteins are biological macromolecules composed of one or more linear chains of amino acids (*aa*'s) referred to as polypeptides. The polypeptide chains are often folded into a specific three-dimensional arrangement, that confers a biological function to the macromolecular assembly.[1] The majority of proteins are composed of twenty different *aa*'s, which are referred to as the most common *aa*'s, but over 240 naturally occurring *aa*'s are known.[2] Of the 20 most common *aa*'s, all but one share the same general structure with the only difference arising from the identity of their side chain. Only proline and its cyclic side chain connected to the terminal amine breaks this pattern. Table 1.1 provides the structures of the 20 most common *aa*'s organized according to their side chain functionality.[3] The carbon atom bearing the side chain located between the amine and the carboxylic acid end groups of *aa*'s is referred to as the alpha-carbon (C_α) and each carbon extending along the side chain is indexed by sequential Greek letters. All the common *aa*'s but glycine, which has two hydrogens on its C_α, contain a chiral C_α and adopt an *L*-configuration in their natural form.

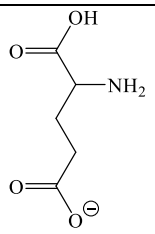
The structural arrangements of the *aa*'s constituting proteins are often divided into primary, secondary, tertiary, and quaternary structures.[1,3] The primary (1º) structure of a protein refers to the linear sequence of *aa*'s defining the sequence of a protein, and is also described as a polypeptide. Figure 1.1 illustrates a backbone fragment containing four generic *aa*'s covalently linked together through an amide bond, where R represents the *aa* side chain. The ends of the chain are flanked by an amine and a carboxylic acid on the so-called *N*- and *C*-termini, respectively. By convention, the *N*-terminus is used as the starting point of an *aa* sequence.

Table 1.1. Chemical structures of the common amino acids.

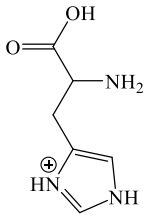
Ionic



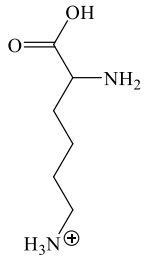
Aspartic acid



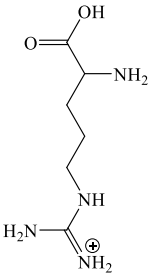
Glutamic acid



Histidine

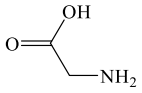


Lysine

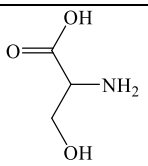


Arginine

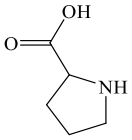
Hydrophilic



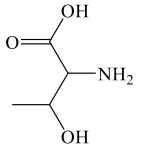
Glycine



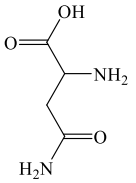
Serine



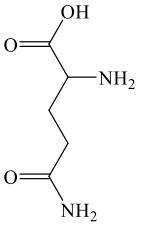
Proline



Threonine

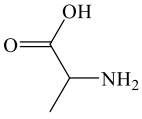


Asparagine

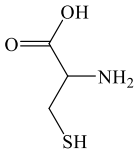


Glutamine

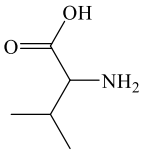
Hydrophobic



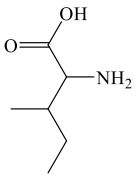
Alanine



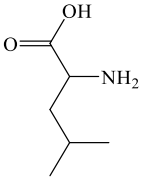
Cysteine



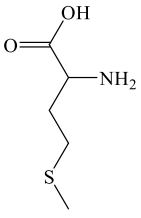
Valine



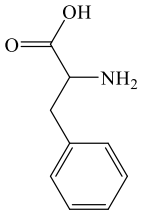
Isoleucine



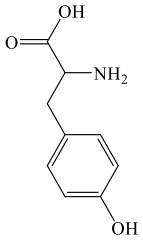
Leucine



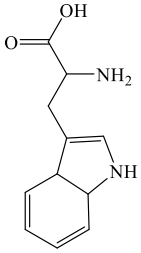
Methionine



Phenylalanine



Tyrosine



Tryptophan

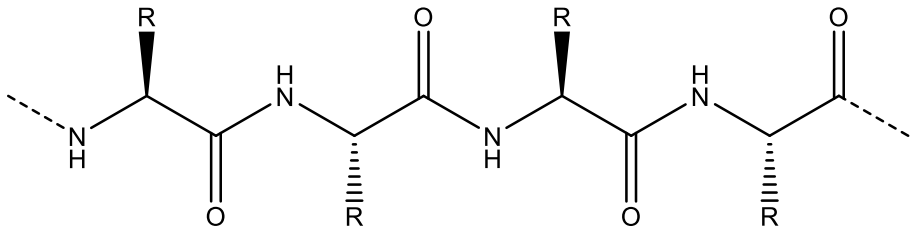


Figure 1.1. Fragment of a polypeptide chain containing four *L*-amino acids.

X-ray crystallography has shown that the amide linkages adopt rigid-planar conformations extending from the carbonyl oxygen to the amide proton.[4] All *aa*'s in a protein exist almost exclusively in the trans conformation (i.e. have a torsion angle (ω) equal to 180^{o}), except when they are followed by a proline, in which case steric hinderance between the side chains partially stabilizes the cis conformation.[5] This is explained by the ability of the lone pair of electrons on the nitrogen to resonate with the carbonyl of the adjoining *aa* (Figure 1.2) and it has been estimated that the C-N bond has ~30 to 40% double-bond characteristic.[6]

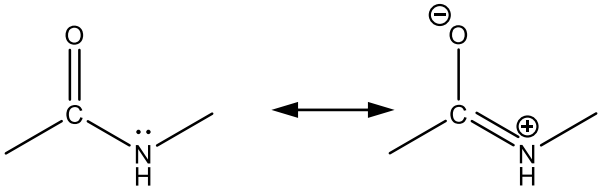


Figure 1.2. Resonance structures of the amide bond linking two *aa*'s in a polypeptide restricting the rotation of the amide bond.

The inability of *aa*'s to rotate about their amide bond restricts their orientation in a polypeptide backbone to the two torsional angles θ and φ depicted in Figure 1.3, where both angles equal 180º when the polypeptide is in an all trans conformation. The rotational freedom about θ and φ depends on the side chain (R) of the *aa*. The first carbon in the side chain (C_β) has the greatest effect on the rotational freedom due to its close proximity to the amide hydrogen and the carbonyl, which restricts some rotational angles due to steric hindrance. The conformationally available angles are typically represented by a Ramachandran diagram which is a 2D plot of allowed θ and φ values.[7] Interestingly, the Ramachandran diagrams of all *aa*'s but glycine and proline are quite similar to one another, which indicates that they share a similar degree of conformational freedom.[8] Since glycine has no C_β in its structure, glycine has significantly more rotational freedom resulting in many more allowed θ and φ values. On the other hand, the cyclic structure of proline significantly reduces its rotational freedom, resulting in a rather limited number of available conformations.

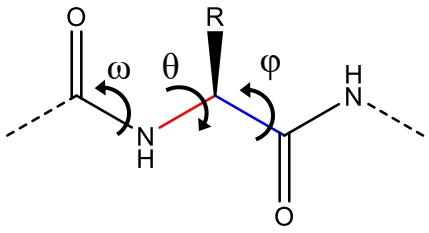


Figure 1.3. Torsional angles for each *aa* in a polypeptide chain. The amide torsional angle (ω) is typically constrained to 180º, while the angles θ (C_α-N bond in red) and φ (C_α-C bond in blue) are free to rotate. The arrows indicate the direction of increasing angle.

The secondary (2º) structures of a protein are defined by the regular patterns generated by a sequence of *aa*'s adopting a local conformation. The two most predominant secondary structures are α-helices and β-pleated sheets.[1] An α-helix consists of a regularly repeating twist of the amino acids defined by the torsion angles of $\theta = -57º$ and $\varphi = -47º$. The result is a right-handed helix containing 3.6 *aa*'s per turn and a pitch (rise per turn) of 5.4 Å.[1,3] The side chains of the *aa*'s extend away from the backbone, which allows for tight packing of the helical core. Additionally, every backbone amide proton is capable of hydrogen bonding with the carbonyl of the *aa* located 4 residues away. The tight packing of the core, along with the many hydrogen bonds generated between amide protons and carbonyl oxygens, makes the α-helix a very stable and rigid structure. The typical length of an α-helix is about 14 *aa's* (~3 turns) in globular proteins.[9] In addition to α-helices, other types of helical secondary structures including 3_{10}-helices, π-helices, and poly(proline) type II (PPII) helices are also encountered, but they are not as common due to their smaller number of favourable interactions as compared to α-helices.[1] The 6 Å pitch and 3 *aa*'s per turn of a 3_{10}-helix result in a much thinner helix with fewer hydrogen bonds than for an α-helix, while the 1.1 Å rise and 4.4 *aa*'s per turn of π-helices results in a wider helix with fewer van der Waals interactions due the presence of a small cavity in the helical core resulting from the larger diameter of the helix.[1] PPII helices, as their name implies, are often observed with oligoproline peptides and proline-rich segments in proteins.

β-sheets are also very common secondary structures. A β-sheet contains two or more chains whose alignment with one another results in the formation of hydrogen bonds between the backbone amides in the adjacent chains. The chain segments can run in the same direction

(i.e. the N to C alignment of the *aa*'s in each chain points to the same direction) or in opposite directions resulting in parallel or antiparallel sheets, respectively. Like α-helices, β-sheets are also stabilized by hydrogen bonding between the backbone amide protons and carbonyl oxygens and the van der Waals interactions between the tightly aligned chains.[3] Often the chain segments, or strands, constituting a β-sheet form loops near the sheet edges to facilitate better packing in the structure of proteins. These loops, or turns, typically involve 3 to 5 *aa*'s locked in specific conformations to allow the peptide chain to reverse its direction.[1]

The high-order tertiary (3º) and quaternary (4º) structures of proteins refer to the overall three-dimensional arrangement of the structural components. The shape of proteins consisting of a single peptide chain, or one segment in a multimeric protein, defines their tertiary structure, while the quaternary structure refers to the arrangement of folded monomeric subunits in proteins consisting of more than one peptide chain.

1.2 Driving Forces in Protein Folding

In the 1960's, Christian B. Anfinsen demonstrated that protein folding was reversible, implying that folding was thermodynamically driven and that the native state was the most thermodynamically stable conformation.[10,11] Since folding is a spontaneous process, the combination of all inter- and intra-molecular interactions must result in an overall negative free energy. Denaturation studies on globular proteins have shown that the free energy holding their native structure together amounts to 20 to 60 kJ/mol.[12] For comparison purpose, an isolated hydrogen bond has a free energy of ~20 kJ/mol,[13] nearly as much as an entire protein. Since

proteins have relatively small free energies, all interactions in a protein must contribute substantially toward the stability of the protein. The most common forces contributing to protein stability are discussed below.

Five of the 20 most common *aa*'s bear an ionic charge at physiological pH (Table 1.1), namely aspartic acid, glutamic acid, arginine, histidine, and lysine, the two former and the three latter *aa*'s having a negatively and a positively charged side chain, respectively. Since many of these ionic *aa*'s coexist in the sequence of proteins, essentially all proteins generate ionic interactions in the form of electrostatic attraction between two oppositely charged *aa*'s in close proximity to one another, often referred to as an ion pair or a salt bridge.[14] The large increase in electric potential predicted by Coulomb's law as two ions approach one another as a protein collapses into its dense native state was once believed to be the main force contributing to protein stability.[15] However this initial view has now shifted and the general consensus is that although ionic interactions are important for defining the 3D structure of a protein, their overall contribution toward the stability of a protein is rather small.[1,15–17] The strength of an ionic interaction depends on the local dielectric constant associated with the polarizability of their surrounding medium. Since each protein is unique, determining the dielectric constant near an ion pair is challenging, and so is the calculation of the contribution of ionic interactions to protein stability.[17] Experimentally, these contributions can be determined by pH induced denaturation experiments. These studies have shown that each surface salt bridge contributes ~3 – 5 kJ/mol to the stability of a protein, whereas a buried salt bridge can contribute over 10 kJ/mol.[18–20] One suggested reason for these rather low

contributions to stability is that favourable ionic interactions persist even when a protein is denatured, thereby minimizing changes in electrostatic potential between the folded and denatured states and limiting their stabilizing ability.[21] Since the frequency of ion pairs in many proteins is rather low (~ 1 ion pair per 30 residues), and buried salt bridges even less common (~1 in 5 salt bridges),[17] it is often estimated that ionic interactions contribute only ~ 40 kJ/mol[15,16] to the overall stability of a protein.

Hydrogen bonds (H-bonds) are one of the defining features of secondary structures. They therefore play a central part in the structure of a protein. They occur when a hydrogen atom is shared between two electronegative atoms such as O or N. Each amide bond contains a carbonyl oxygen capable of receiving, and a protonated nitrogen capable of donating, a proton to generate a H-bond. Furthermore, many of the *aa* side chains (Table 1.1) can also form H-bonds. The abundance of potential hydrogen bonding groups leads to their high prevalence in protein structures, with globular proteins containing ~1.1 H-bond per *aa* on average.[22] Of these H-bonds, ~65% occur between backbone amide groups (nearly all of which occur in 2^{o} structures), 23% between side chains and backbone amides, and the remainder 12% between side chains. Up until the 1990's, it was clear that H-bonding was important to the structure of a protein, but there was uncertainty about how much (if at all) H-bonds contributed to protein stability. Since then, advances in experimental techniques and increase in computational power have provided an answer to this question. Experimentally, the energies associated with H-bonds in proteins can be determined by point mutation studies whereby an *aa* capable of H-bonding is replaced with an *aa* which cannot (ex. serine to alanine).

Subtraction of the free energy (ΔG) of folding between a native and its variant protein yields the difference in free energy (Δ(ΔG)) due to the mutation. Since replacing one *aa* with another often results in changes of multiple non-specific interactions (ex. van der Waals, hydrophobicity, …) and not just H-bonding, these point mutation studies often compare the Δ(ΔG) values between several mutations through *double mutant cycles*[23] or between a same mutation at two locations in a protein, one where the native *aa* is involved in a H-bond and the other where it is not.[16] Since H-bonds are directional, their strength is related to the relative arrangements of the atoms involved and their contribution to stability can be rather varied depending on their location, but the majority of H-bonds between side chains have been found to favorably contribute ~1 – 4 kJ/mol towards protein stability.[16,23] Hydrogen bonds involving the backbone amides have also been shown to contribute ~ 4 kJ/mol to stability, leading to the conclusion that all H-bonds are important. However, since the majority of H-bonds occur between backbone amides, they contribute more to stability than H-bonds involving side chains.[24] To support these experimental results, similar values of ~2 – 8 kJ/mol for stabilization energy have been predicted through computer simulations when taking into account the aqueous environment of proteins.[13] The agreement between experimental and computational studies has provided strong support to the notion that H-bonds are very important to stability, contributing ~ 4 kJ/mol per H-bond. These results have been used to predict that up to ~40% of the overall stability of a protein may arise from H-bonds.[25]

On average, ~86% of the hydrophobic side chains are buried in the apolar core of a protein,[26] leading to tightly packed interiors. Using the atomic volumes of *aa*'s, Klapper

demonstrated that 75% of the space in the interior of proteins is occupied by atoms. This density was not only much higher than that of water, where just 36% of the water volume is occupied by atoms, but even larger than that of an array of close-packed spheres (71%), leading to the conclusion that the interior of proteins is closer to a solid than a liquid.[27] The tight packing leads to many molecular contacts resulting in strong van der Waals interactions between neighboring *aa*'s. To highlight the importance of these interactions in the dense interior, Van der Waals interactions of a methylene (-CH_2-) have been calculated in different environments. The interaction energy of methylene in the interior of a protein was calculated to equal ~13 kJ/mol, much larger than the ~8 kJ/mol interaction energies calculated for methylene in water and cyclohexane,[28] indicating that folding into a tightly packed structure increases van der Waals interactions, aiding in protein stability. Since van der Waals interactions are nonspecific, they can be difficult to measure directly as they are typically convoluted with one of the other major contributors to protein stability represented by the hydrophobic effect.[25]

The hydrophobic effect describes the tendency for hydrophobic groups to aggregate with one another in a polar solvent, such as water, in order to minimize the surface area in contact between the opposing phases.[1] Since water cannot H-bond with a hydrophobe, an ordered *shell* of water molecules forms at the interface, resulting in a significant entropy loss for the involved water molecules. Furthermore, many more ordered water molecules are present at the interface generated by the large surface area representing many small hydrophobic domains than at the interface of a smaller surface area generated by a much larger

domain combining many smaller hydrophobic domains. This difference in surface area implies that the aggregation of hydrophobes into large hydrophobic domains and the associated release of water molecules from the ordered surrounding shell to the disordered bulk increases the net entropy of the system, which increases its free energy. Therefore, the hydrophobic effect is non-specific and is driven by an entropic gain induced by water molecules. The contribution of the hydrophobic effect to protein stability is, like H-bonding, studied through point mutation experiments. Replacing one hydrophobic *aa* with another *aa* having a smaller side chain (ex. isoleucine to valine) results in a $\Delta(\Delta G)$ value representing the loss in stability due to the smaller hydrophobic side chain. As would be expected, a larger decrease in hydrophobic side chain size results in more negative (less stable) $\Delta(\Delta G)$ values, by an average of ~5 kJ/mol per methylene ($-CH_2-$) group removed.[16,25] The combination of the hydrophobic effect and van der Waals interactions has been predicted to contribute ~60% to the stability of a protein,[25] making it the main driving force for folding.

Besides the main contributors driving protein folding discussed above, other forces are known to destabilize proteins. The main destabilizing force is the conformational entropy of the polypeptide chain constituting the protein. An unfolded chain has a high degree of freedom, is able to adopt many different conformations, and has a high entropy. However, as the chain folds into its confined native state, it loses its ability to populate many of its previously available conformational states, resulting in a significant loss in entropy. Several techniques have estimated that this entropic loss ranges between 7 and 15 kJ/mol per *aa* at room temperature, but is more likely of being closer to the higher end of this range.[25,29] Another

contributor to protein instability is the number of isolated ionic groups buried in the apolar core of a protein, such as an isolated ionic side chain. This occurs when the folding of a protein forces an ionic side chain into a hydrophobic domain, preventing interactions with water and resulting in a net loss of free energy. For example, replacing a buried aspartic acid residue (anionic) in RNase Sa with another non-ionic *aa* increased its stability by ~12 kJ/mol.[30] In contrast, if the buried ionic side chain is capable of H-bonding, the side chain can contribute favorably to stability as discussed above.

1.3 Energy Landscape of Folding

Protein folding is often described in terms of an energy landscape, represented by a funnel-like diagram which is broad at the top and narrows to a single point at the bottom, such as the one depicted in Figure 1.4.[31] These folding funnels represent the free energy of a protein chain as a function of their conformational space. The internal free energy of the chain is represented by the vertical axis, while each horizontal axis represents a degree of freedom in the chain. Since every chain has many degrees of freedom (ex. every *aa* will have its own φ and θ values defining its conformation), energy funnels have many conformational dimensions, but for simplicity sake, only two conformational dimensions are often shown.

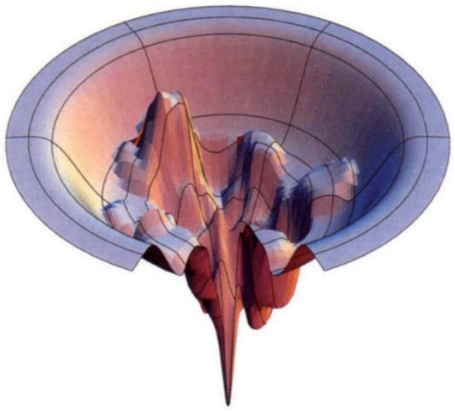


Figure 1.4. Illustration of an energy funnel with a rugged energy landscape containing energy barriers and kinetic traps. Reprinted by permission from Springer Nature: Nature, Nature Structural Biology, From Levinthal to Pathways to Funnels, Ken A. Dill and Hue S. Chan 1997.[31]

For each conformation, the internal energy is the sum of all the forces described above other than the conformational entropy of the chain, so that every conformation corresponds to a unique point on the energy surface. The entropy of the chain is represented by the width of the energy funnel in that a broader width at a specific energy represents greater conformational freedom and larger entropy. Folding is often described as being akin to a ball rolling from the top of the funnel, where an unfolded chain has a large internal energy and entropy, to the bottom, where the protein exists in its native state with low internal energy and entropy. As a protein seeks its lowest energy state, its conformational search tends to follow a path of lowest resistance (i.e. avoiding large peaks, or high energy conformations) to the bottom while being constantly nudged in any direction by Brownian motions. Since many conformations exist at

similar energies, no single conformational pathway is followed by the protein and there is no specific conformation at a given funnel depth. As such, folding must occur through an ensemble of pathways, each of which starting somewhere on the energy surface and leading to the native state. The surface of the energy funnel will be unique for each protein. A two-state protein (one which folds with a single rate constant) is often described with a macroscopically smooth funnel defined by a volcano-shaped energy funnel, such that there is one energy barrier with no stable intermediates.[32] On the other hand, multistate folders can have much more complex landscapes. They can exhibit plateaus, where the protein may have to search a great number of conformations before it can continue folding, and even have much more rugged landscapes, with many peaks and valleys. As the protein folds, it can get caught in a trap for a brief period of time before it is nudged out by Brownian motions that allow it to continue its downhill search. Such intermediates are sometimes referred to as 'off-pathway' intermediates, since they must go uphill (increase their energy) before the protein can continue to fold. However, there are no conformations which are on- or off-pathway in an energy landscape since folding does not follow a predefined path, as folding occurs simultaneously through an ensemble of pathways. On these rugged landscapes, the starting conformation can have a large effect on the folding kinetics of proteins. For example, some pathways may lead directly to the free energy minimum, meandering though many peaks and valleys while avoiding traps. However, other pathways cannot avoid all traps, which leads to slower and more complex folding kinetics.

1.4 Structural Characterization

Three-Dimensional Structure with Atomic Resolution: The structure of a protein is not only important for defining its biological function but is also used to help understand how proteins fold. There are three primary techniques which are applied to determine the overall three-dimensional structure of proteins; x-ray crystallography,[33] nuclear magnetic resonance (NMR),[34] and cryogenic-electron microscopy (cryo-EM).[35] The earliest method was x-ray crystallography. As the name suggests, this technique exposes a crystalized protein to a beam of x-rays in order to elicit its structure. The first and often most difficult step in this procedure is growing a protein crystal.[36] This is done by slowly changing the conditions (concentration, pH, buffers, salts, etc…) of a concentrated protein solution until the protein slowly grows into crystals. In general, proteins which are more flexible and contain less constrained domains are more difficult, or sometimes even impossible, to crystalize, limiting which proteins can be studied by x-ray crystallography. To make this step even more difficult, a crystal size of about 0.1 mm in the longest direction is needed, which often requires careful control of the crystallization process.[33] Once a suitable crystal is obtained, it is exposed to an x-ray beam which generates a diffraction pattern as the x-rays are scattered off the crystal structure. The diffraction pattern is then transformed into a map of electron density using specialized computer programs, which is used to determine the position of each *aa* in a unit cell (and thus also in the structure of a protein). The resolution of the electron density maps depends heavily on the ordering of the proteins in the crystal, but resolutions as high as ~ 1 – 3 Å are obtained for high quality crystals.[37] Since x-ray scattering requires uniform crystals, disorganized proteins or flexible segments of proteins, which often crystalize in different orientations, make

them invisible to this technique. One way to determine the structure of a difficult-to-crystalize protein is through NMR.

One of the main advantages of NMR experiments is that they yield the structure of a protein in solution.[34] This is done by placing the protein in a deuterated solvent (D_2O) and monitoring the relaxations of the nuclei spins as they are perturbed by a strong magnetic field.[38] Since proteins have large numbers of atomic nuclei in relatively similar environments, multi-dimensional NMR techniques are conducted to separate the signals of the nuclei in two or more frequency domains. These interactions can be observed either through bonds (ex. COSY, HSQC), through space (ex. NOESY, ROESY), or combinations of both (ex. ^{15}N-NOESY-HSQC). In particular, the cross relaxation of two nearby spin-active nuclei, known as the nuclear Overhauser effect (NOE), is often taken advantage of to determine the relative positioning of the atomic nuclei in a protein.[34] Each NMR signal is then used to build a list of restraints of the atoms in the protein, which is then employed to build a model of the location of each atom in the protein. Although the separation of NMR signals into multiple frequency dimensions reduces the chance of observing overlapping signals, it does not exclude this possibility, which makes NMR difficult or impossible to apply for the structure determination of larger proteins (historically up to ~ 25 kg/mol, but proteins up to 100's of kg/mol can now be studied by NMR).[39] Furthermore, only nuclei with an unpaired nuclear spin (i.e. having an odd atomic mass) interact with the external magnetic field, so that the response of each nucleus in a protein depends on their isotopic abundance.[38] Since the natural abundance of ^{13}C and ^{15}N nuclei is ~ 1% and 0.4%, respectively,[40] the vast majority of carbon and nitrogen atoms would

typically be difficult to detect by NMR, resulting in experiments which would take a prohibitively long time. To circumvent this problem, proteins studied by NMR are often prepared with isotope-enriched atoms.[41] However, this process increases the cost.

Another technique worth mentioning is cryo-EM, a relatively new technique which is gaining popularity.[42] It involves depositing a solution containing a protein on a screen[35] before vitrifying the sample by flash freezing, which traps the protein in its native conformation and prevents water from crystalizing. While maintaining the cryogenic temperature, the vitrified sample is then transferred into an electron microscope. Hundreds of thousands of images of the protein are taken at random orientations before being compiled to reconstruct the 3D structure. The limit of resolution of cryo-EM is presently unknown, but resolutions of 2.2 Å have been reported.[43] Therefore this technique can provide structures with near atomic-resolution without requiring protein crystallization. One of the current problems associated with cryo-EM is to overcome the conformational heterogeneity of proteins which complicates the imaging compilation.

Lower-Resolution Structural Information: Obtaining a structure with atomic resolution is time intensive, costly, and often unnecessary, since many experiments do not require such level of detail. The ability to detect or monitor a structural change undergone by a protein is usually enough. The techniques most commonly used for this purpose are circular dichroism (CD),[44] Fourier transform infrared spectroscopy (FTIR),[45] and fluorescence.[46] CD measures the difference in absorption of left-handed and right-handed polarized light. When an asymmetrical molecule is exposed to circularly polarized light, it may preferentially absorb

either the left or right-handed polarized light. For proteins, the CD signal of their amide bonds is typically measured as they absorb in the 180 – 230 nm range.[44] If a protein is folded, the regular repeating conformation of amide bonds involved in its secondary structural motives results in a difference in absorbance between the left and right-handed polarized light. This difference is quantified by the molar ellipticity (*[θ]*) and is a function of wavelength.[47] Since each secondary structure results in a different conformation for the amide bonds, each secondary structure has a unique *[θ]* spectrum.[44] This allows CD to not only determine if a structure is present, but also identify the nature of the secondary structures present in a protein and their abundance. CD is also often employed to monitor the conformational changes undergone by a protein after application of an external stimulus (denaturant, temperature, etc…).

FTIR measures the absorbance of molecules in solution in the IR range, which reports on the vibrational transitions of bonds.[45] The vibrational frequency and probability of absorption depend on the strength and polarity of the bond, and are therefore affected by inter- and intramolecular interactions. The infrared spectrum is typically plotted as a function of wavenumber ν expressed in cm^{-1}, which is proportional to the transition energy. The vibrational modes of amides and *aa* side chains have been characterized and each have their own absorption band, corresponding to a peak in the IR spectrum.[45] The most commonly used peak for determining the structure of a protein is the amide I band near 1650 cm^{-1}, corresponding to the CO stretching vibration of the amides. The amide I band is nearly independent of the *aa* side chain, but changes with backbone conformation. Consequently, it

is ideally suited to determine the secondary structure(s) of a protein. Since water absorbs near 1645 cm^{-1}, care must be taken to ensure that the water peak does not interfere with the absorption measurements. This can be done by using short pathlengths and high protein concentrations, or by using deuterated water which absorbs at the much lower frequency of ~1210 cm^{-1}.[48]

Another common tool used to measure conformational changes in a protein is fluorescence.[46] When a fluorophore is excited by the absorption of a photon, it will emit light at a longer wavelength than the excitation wavelength as it returns to the ground state. The fluorescence characteristics, such as wavelength and lifetime, depend on the polarity of the local environment. This implies that the emission of a fluorophore is different in an apolar environment (such as inside the core of a protein) versus a polar environment (such as when exposed to water). Fluorescence experiments on proteins can be done using the intrinsic fluorescence of phenylalanine, tyrosine, or tryptophan, or by labeling a dye onto a protein. In the simplest case, the fluorescence of the intrinsic fluorophores of a protein will change as the protein structure is perturbed. Since only the fluorescence of fluorophores experiencing an environmental change is affected, the change in fluorescence can be used to determine the extent and location of denaturation happening to the structure of a protein. Fluorescence is also commonly applied to probe the internal dynamics and structural intermediates of proteins as they fold, as will be discussed later.

Conformational Stability: The stability of proteins is typically determined through equilibrium unfolding experiments.[49] The free energy difference between a folded and unfolded protein

(ΔG) is defined by the equilibrium constant (K_f) between the folded and unfolded states as shown in Equation 1.1, where R is the ideal gas constant and T is the absolute temperature in Kelvin.

$$\Delta G = RT \ln K_f \qquad (1.1)$$

K_f is determined by the ratio of the molar fraction of proteins in their folded (x_f) and unfolded (x_u) state ($K_f = x_f/x_u$). Since many proteins exist predominantly in their folded state under physiological conditions, the population of unfolded proteins is very small making it difficult to determine K_f. However, the equilibrium between the folded and unfolded state can be perturbed by changing the experimental conditions, such as upon addition of a chemical denaturant like urea or guanidine hydrochloride, thereby increasing the population of unfolded proteins. A spectroscopic technique, such as FTIR, fluorescence, or CD which were just discussed, can be used to monitor the change in signal as a function of denaturant concentration. The exact handling of the experimental data depends on the equilibrium folding mechanism. In the simplest two-state case (the protein is either folded or unfolded with no intermediates), the spectroscopic signal at each denaturant concentration is compared to that of the completely folded and unfolded protein to determine x_f and x_u. The molar fractions are then used to generate a plot of ΔG as a function of denaturant concentration. ΔG is then extrapolated to zero concentration, which provides the conformational stability of the protein

without denaturant. In the case where the protein has a detectable folding intermediate, more complex models are required which typically involve the global analysis of multiple data sets.[50]

1.5 Folding Kinetics

The determination of the folding rate requires monitoring the formation of structure as a function of time. Absorbance, fluorescence, and CD are commonly used to monitor structure,[51] but other techniques such as NMR can also be employed.[52] In order to measure folding rates, a method must be used to quickly perturb the protein from an equilibrium state so that its spectroscopic signal can be monitored as the protein folds/unfolds. Stopped-flow techniques are often used for this purpose, but many other techniques exist. Ultimately, the choice of a specific technique depends on the time-scale of the experiment. Stopped-flow methods are limited by the mixing rate of two solutions. They measure the relaxation of proteins on the micro to millisecond timescale.[53] Although there are other techniques such as temperature jump experiments that can probe protein folding on the nanosecond timescale,[54] they typically do not fully denature a protein. This means that folding can only be probed from a partially unfolded state which can miss important steps in the folding kinetics. Therefore stopped-flow experiments are used more predominantly.[53] Stopped-flow experiments are conducted by quickly mixing two solutions. The refolding of a protein is measured by preparing a solution containing a protein and a denaturant. The solution is then rapidly diluted in a stopped-flow cell, lowering the denaturant concentration which allows the protein to refold into its native conformation. Similarly, unfolding is measured by quickly mixing a solution containing the

folded protein with a solution containing a denaturant. The detector response is then plotted as a function of time and fitted with an exponential function. A two-state protein can be fitted with a single exponential, whose analysis yields the folding rate constant. Multi-state proteins require sums of exponentials, where each exponential is attributed to a rate constant. In order to determine the rate constant of folding under native conditions without denaturant, the observed rate constant must be determined for many different denaturant concentrations. A Chevron plot is then constructed by plotting the log of the observed rate constant as a function of denaturant concentration. The characteristic v-shape of the plot is used to extrapolate the rate constants of folding and unfolding in the absence of denaturant.

Many small single-domain proteins fold according to two-state kinetics with no kinetically observable intermediates.[32] In this case, folding proceeds smoothly from the unfolded state to the folded state with no energy traps. On the other hand, many other proteins are multistate, implying that they have one or more kinetically observable intermediates. Consequently, these proteins can adopt a metastable state. Classically, folding intermediates can be produced in an on- or off-pathway, depending on the pathway for the formation of the folding intermediates and the folded state. On-pathway refers to an intermediate which forms in-between the folded and unfolded states, while an off-pathway intermediate forms an equilibrium with the unfolded state, but cannot directly transition into the folded state. The distinction between on- or off-pathway is important since it suggests that proteins may differ in their folding mechanisms, where on-pathway folders fold differently from off-pathway folders. However, it has been suggested that the observation of these different kinetics depends

on the relative stability of the folding intermediates rather than having to be on a specific pathway.[55]

Many globular proteins experience a fast folding step that takes place in less than a few tens of microseconds and is commonly referred to as the burst phase.[56] This rapid folding often occurs more rapidly than the mixing of the two solutions in most stopped-flow experiments, but can still be detected by comparing the amplitude of the exponentials as a function of denaturant concentration. This burst phase is characterized by the rapid collapse of the protein into a globular state.[53,57] The formation of the globular state is believed to be driven by non-specific hydrophobic interactions in the protein chain, leading to a significant decrease in the chain's radius of gyration. The rapid increase in chain compactness is also generally accompanied by the formation of local secondary structures which form on a similar timescale.[57,58] Theoretical models have suggested that the formation of secondary structures is driven by the compactness of the chain, since secondary structures are the only regular way to efficiently pack a linear chain into a confined space.[59,60] A review on the topic[61] suggests that the hydrophobic collapse and the initial formation of secondary structure may not be mutually exclusive and must occur simultaneously. The result is a molten globule of secondary structures in their non-native conformations, which then must continue their conformational search toward the natively folded state with a slower rate constant.

As a protein transitions from an unfolded to a folded state, the *aa*'s must diffusively encounter one another before they can form favorable interactions according to a process which ultimately leads to structure formation. Therefore, understanding the dynamics, or the rate at

which *aa*'s in a polypeptide chain interact with one another, is important to understand the folding process. Experimentally, diffusion-limited loop formation of unfolded polypeptide chains has been measured by fluorescence using end-to-end cyclization (EEC) experiments taking advantage of triplet-triplet energy transfer[62] or quenching.[63] EEC experiments involve the end-labeling of an oligopeptide with a fluorophore at one end and a quencher at the other end. Upon contact with one another, the excited fluorophore is quenched, resulting in a shortened lifetime. Time-resolved fluorescence measurements are applied to measure the rate of contact between the two dyes to provide a quantitative measure of the cyclization rate constant (k_{cy}) of the oligopeptide, which in turn provides information on how quickly the oligopeptide ends can diffusively encounter one another. k_{cy} depends both on the *aa* sequence of the oligopeptide and the chain length but takes values that are typically in the 10^8 s^{-1} range. Since EEC experiments can only be conducted on small (typically < 20 *aa*'s) oligopeptides, the observed rates are representative of dynamics experienced by either small proteins or during loop formation. Some studies have tried to quantify how the rate of loop formation depends on the *aa* sequence[62] and location[64] in a polypeptide chain. Unfortunately, the synthetic difficulties associated with the preparation of monodisperse polypeptides with specific labeling positions make these studies rather challenging, which explains why they are still limited to oligopeptides no greater than ~ 30 *aa*'s. The dynamics measured by these quenching studies provide information on the limiting rate constant for intrachain diffusive encounters of proteins near their unfolded state, and therefore are typically considered to be an upper limit for the rate at which proteins can explore their conformations.

1.6 Heterogeneity of Folding

The globular states of many proteins[65] (including barstar[66] and ribonuclease A[67]) have been shown to be heterogeneous. For ribonuclease A, at least 5 distinct intermediates have been discovered.[68] Binding studies of these intermediates have demonstrated that each intermediate follows an independent pathway with its own folding rate. The fact that each pathway was unique and different from one another suggested that folding did not follow one predefined pathway, as predicted by the energy landscape model. Further evidence in favor of an energy landscape comes from structural information on the globular intermediates determined using hydrogen exchange (HX) experiments. These experiments use either NMR or mass spectroscopy (MS) to monitor the rate of exchange of protons (typically the amide protons) in a protein with the deuterated solvent.[69] The rate of exchange depends on the accessibility of the amide protons to the solvent and their involvement in hydrogen bonds. Residues which are buried in the interior of a protein or are involved in hydrogen bonded networks, such as those generated by secondary structures, will exchange more slowly than those which are not. Therefore, HX experiments provide an experimental means to monitor the formation of structure with single *aa* resolution. Comparison of the structural intermediates of ribonuclease A showed that the *aa*'s in the intermediates had substantially different exchange rates.[67] This meant that the intermediates were conformationally heterogeneous, involving different folded and unfolded chain segments of the protein.

The conformational heterogeneity of proteins has also been studied by time-resolved Förster resonance energy transfer (*tr*-FRET) in barstar.[65] *tr*-FRET is similar to fluorescence quenching, but the quenching (or energy transfer) occurs through space. The efficiency of

energy transfer depends on the distance separating the donor and acceptor molecules to the power of six. Since FRET is extremely sensitive to distance, the FRET efficiency is often used to back calculate the distance between the donor and acceptor. In barstar,[65] a native tryptophan served as the donor and a single acceptor was attached onto a cysteine. The FRET efficiency in the labeled barstar was then measured in its folded state and in solutions containing increasing concentrations of denaturant. This allowed the distance between the FRET donor and acceptor (D-A) to be measured at various stages of folding. When folded, the D-A distances measured by FRET corresponded to the distances expected from the x-ray crystal structure. When denatured, the FRET efficiency decreased, which reflected an increase in the D-A distance. Careful analysis of the fluorescence decays revealed that there was in fact a distribution of D-A distances, indicative of proteins in different conformations. This process was repeated by systematically labeling four different cystines with the acceptor in order to measure the intramolecular distance, and therefore conformation, in different segments located throughout the protein as a function of denaturant concentration. The distribution of distances was most narrow at low denaturant concentrations, indicating that barstar adopted similar native-like folded conformations, as would be expected for proteins near a natively folded state. However, as the denaturant concentration was increased, there was significant broadening of the distribution of D-A distances, which indicated that the denatured state of barastar was heterogenous and was not occupied by similar conformational states. This clearly indicated that barsar must adopt many different conformational states when unfolded. The fact that the heterogeneity decreased as the protein approached its native state was taken as direct

experimental evidence of the energy landscape model of folding, where the conformational space of a protein becomes more limited as stability increases.

1.7 The Hierarchy of Folding

Despite evidence that folding does not follow a unique pathway, several trends in the way proteins fold have been established. The folding rate of proteins tends to correlate to their native topology, which is defined by the contacts between the *aa*'s in a folded protein. One of the common topological parameters used to quantify the structure of proteins is the relative contact order (RCO).[70] The RCO is determined from the average sequence distance separating all pairs of contacting *aa*'s in the high-resolution structures of proteins. Proteins with contacts between *aa*'s that are predominantly close together in the sequence have a low RCO, while those with less local contacts have a larger RCO. Therefore, the RCO provides a measure of the relative importance of local and non-local structures, where a local structure is generated by neighboring *aa*'s along a protein sequence. The negative correlation between RCO and folding rate in single domain proteins demonstrated that proteins with more local structures, such as α-helices and turns, folded more quickly than those with more non-local structures, such as β-sheets.[70] Since then, many other topological parameters have been used to correlate folding rates, but they all rely on the locality of contacts. These correlations can now be used to predict the folding rates of some proteins from their primary sequence.[71] The conclusion that short segments of a protein sequence involving short-range contacts fold more rapidly than segments involving long-range contacts demonstrates that the formation of local contacts must

be responsible for guiding the folding of proteins along their energy landscapes. In fact, contact order has been combined with a statistical mechanical model to predict not only the folding rate of proteins but also their folding pathways.[72] The model predicts that the dominant folding paths are those which contain only small loop closures (i.e. form local structures), again supporting the idea that folding occurs though local interactions of the *aa*'s.

Perhaps the strongest support for the existence of local folding domains comes from HX experiments. These experiments are done either under equilibrium conditions with a denaturant as described earlier, or with HX pulsed experiments. HX pulse experiments begin with an unfolded protein in D_2O, which allows all the amides to exchange with the solvent. A stopped flow apparatus is then used to initiate folding. By briefly exposing (or pulsing) the folding protein with a protonated buffer, a snapshot of the population of folding intermediates is determined as a function of folding time.[73] These experiments revealed that proteins contained groups of *aa*'s which were able to exchange on a similar time scale, which meant that the *aa*'s in each group must cooperatively fold together, since folding protects them from the exchange process. This observation led to the term *foldon*, which defines the groups of *aa*'s which cooperatively interact with one another to fold. *Foldons* have been found in over a dozen proteins, even those without kinetically observable intermediates.[74] Some proteins, such as apomyoglobin,[75,76] have multiple folding domains which simultaneously fold in parallel with one another. Other proteins, such as cytochrome c[77,78] show *foldons* appearing in a stepwise manner, indicating that the folding of each domain occurs consecutively after one another. However, other studies have demonstrated that parallel folding pathways can generate a similar

step-wise creation of *foldons*.[79] The size of *foldons* varies for each folding domain and protein, but typically is ~ 20 *aa*'s in size.[80] Again, this observation supports the fact that folding is driven by local interactions, but the underlying physical phenomena governing their size is not yet fully understood.

Circular permutation studies have also been used to study the folding domains in ribosomal protein S6.[79,81,82] Circular permutation experiments can be conducted on any protein whose *C* and *N* termini are in close spatial proximity to one another in the native state. The ends of the protein are connected to each other either directly or with a short linker, before the protein is being cut in another location to yield a protein with a similar native structure but a different *aa* sequence. Depending on the location of the chain incision, the circular permutants are found to differ in their structural stability, folding rate, and dominant folding pathway. However, despite all these differences the folding domains in S6 remained the same, even when the order of their folding was reversed.[79] Using the idea that folding was governed by domains constituted of *aa*'s that fold cooperatively, an entire β-strand in S6 (corresponding to the *aa*'s located in the center of the sequence of the wild-type protein) could be removed without perturbing the remainder of the folding domains and structure, directly supporting the idea that folding takes place locally.[82]

The mounting evidence that protein folding is a result of interactions between *aa*'s situated in a local volume led to the proposal that the structure of a protein could be predicted by considering individual small segments of the sequence of a protein rather than that of the entire protein. Based on this general principle, the zipping and assembly model (ZAM) was

developed to predict the three-dimensional structures of proteins solely from molecular force field simulations.[83,84] ZAM divides the sequence of a protein into a series of overlapping segments of 8 to 12 *aa*'s. Molecular dynamic simulations are then used to determine favorable interactions (such as H-bonds or hydrophobic interactions). The metastable segments are then incrementally combined together until the structure of the full protein is revealed. This method has been used to accurately predict structures for many different proteins, even when the protein sequence is as large as 112 *aa*'s.[84]

1.8 The Fluorescence Blob Model and Protein Folding

The fluorescence *blob* model (FBM) was introduced in the 1990s as a model to analyze the time-resolved fluorescence decays of a macromolecule randomly labeled with pyrene.[85,86] In principle, any dye and quencher pair can be used, but the ability of pyrene to self-quench to form an excimer simplifies the labeling process to a single dye. The FBM is built on the principle that a pyrene covalently bound onto a macromolecule will probe a finite volume, referred to as a *blob*, while it remains in the excited state. By doing so, a macromolecule is divided into a series of *blobs* among which the pyrenes will distribute themselves according to a Poisson distribution.[85–89] A depiction of a linear pyrene-labeled macromolecule compartmentalized into *blobs* is given in Figure 1.5. *Blobs* which contain only a single pyrene will be unable to form excimer, and therefore the pyrene will return to the ground-state with its unquenched monomer lifetime τ_M. If a *blob* contains more than one pyrene, the pyrenes can diffusively encounter one another to from an excimer, which results in the simultaneous quenching of the monomer fluorescence and the appearance of the excimer fluorescence.

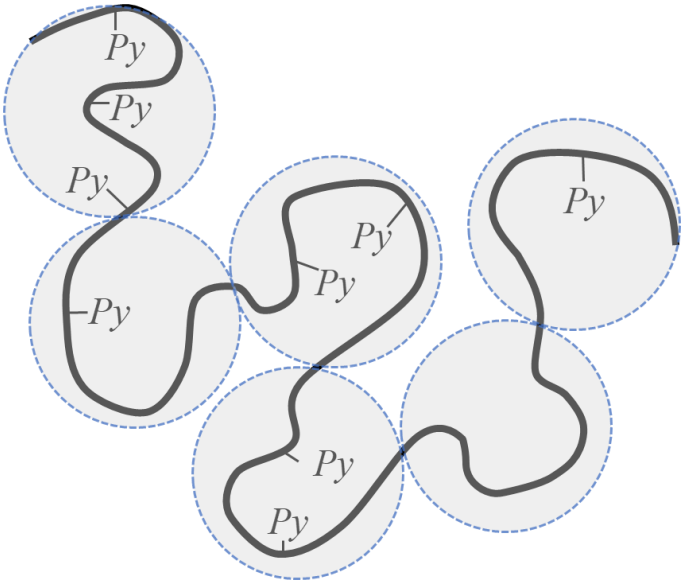


Figure 1.5. Compartmentalization of a pyrene-labeled polymer into *blobs* according to the FBM.

The FBM mathematical handling of the fluorescence decays has been described in detail in many publications.[87–89] Within the framework of the FBM, global analysis of the monomer and excimer decays of the pyrene-labeled macromolecule provides information not only on the internal dynamics of a polymer, but also on its conformation in solution. The FBM does so by returning two fundamental parameters describing the macromolecule. The first parameter is k_{blob}, the pseudo-unimolecular rate constant of excimer formation in a *blob* containing a single excited and ground-state pyrene. The pseudo-unimolecular k_{blob} equals the ratio of the bimolecular rate constant for diffusive encounters k_{diff} divided by the *blob* volume ($k_{blob} = k_{diff} \times (1/V_{blob})$) since $1/V_{blob}$ equals the concentration of one ground-state pyrene inside a *blob*. Since pyrene is covalently attached onto the macromolecule, k_{blob} reflects the rate

constant at which the structural units diffusively encounter one another in a *blob*, thus providing a direct measure of the mobility of the macromolecule on the length scale of a *blob*. The second parameter retrieved from the FBM is <*n*>, the average number of pyrenes contained within each *blob*. <*n*> reports on the local pyrene concentration in a *blob* which depends on two factors, the amount of pyrene attached onto the macromolecule and the conformation of the macromolecule. However, when <*n*> is combined with the molar fraction *x* of structural units, which are labeled with pyrene, and the molar fraction f_{Mfree} of pyrenes which emit with the lifetime τ_{M} in the monomer decay, the average number N_{blob} of structural units contained within a *blob* can be determined with Equation 1.2.

$$N_{\mathrm{blob}} = \frac{< n >}{x}\left(1 - f_{Mfree}\right) \qquad (1.2)$$

Since Equation 1.2 accounts for the pyrene content, N_{blob} is independent of *x* and therefore depends only on the local conformation and dynamics of the macromolecule. For example, the rapid movement of the structural units in a highly dynamic macromolecule allows pyrene to probe a larger volume, therefore increasing V_{blob} and N_{blob} as compared to a less dynamic macromolecule. In less dynamic macromolecules, where the backbone motions are limited and V_{blob} is constant, a tightly coiled macromolecule will have more structural units per unit volume than its loosely coiled counterpart, implying that the tightly coiled macromolecule will contain more structural units per *blob*, resulting in a larger N_{blob} value. Therefore, N_{blob} responds directly to the conformational freedom of macromolecules, with more tightly

constrained macromolecules having smaller *blob* sizes than those with more freedom. Accordingly, the dynamics of macromolecules, defined as the number of contact events in a *blob* per unit time, can be probed with the FBM by determining the product $k_{\text{blob}} \times N_{blob,}$ since it takes into account both the rate constant of encounters/interactions between structural units and the number of structural units involved. The FBM has been applied to retrieve reliable information on the internal dynamics of many macromolecules, and even used to reveal the solution conformation of several rigid macromolecules including amylose[90] and poly(*L*-glutamic acid)[91,92] with the aid of molecular mechanics optimizations (MMOs). The ability of the FBM to provide quantitative information on rigid macromolecules makes it an ideal experimental tool to probe the internal dynamics and conformations of polypeptides.

Protein folding is a very complex process, but there is increasing support in the notion that the solution to this problem may be found through the identification of the folding domains of a protein. While the existence and size of folding domains in proteins can be probed readily by HX experiments,[74,80] there still remain many unanswered questions about the nature of their existence. For example, the size of *foldons* is commonly reported to be ~ 15 to 35 *aa*'s,[74,80] but there has yet to be any experimental evidence explaining why folding domains exist over such a length scale or what effect, if any, the *aa* sequence has. Such fundamental questions are difficult to answer by studying proteins due to their complex nature, but as a review on pyrene-excimer formation points out,[89] the FBM may provide insight to answer such questions by studying simpler polypeptides. Conceptually, the division of a pyrene-labeled polypeptide into *blobs* is similar to the compartmentalization of a protein into *foldons*, as they both segment a

polypeptide chain into a series of subvolumes, which come together to define the whole. However, unlike *foldons*, the size of the subdomains (*blobs*) in the FBM have a clear physical rational for their existence, namely that the *blob* volume is defined by the volume that an excited pyrenyl label can probe through diffusive movements. The covalent attachment of pyrene onto a polypeptide ensures that the volume of a *blob* is directly related to the volume that can be explored by the pyrene-labeled *aa*'s. Therefore, N_{blob} provides a measure of the number of *aa*'s which can diffusively encounter one another on the ~1 μs temporal window provided by pyrene. Furthermore, the FBM also provides k_{blob}, a measure of the rate constant of encounter between the *aa*'s. Therefore, the FBM simultaneously determines not only the number of *aa*'s capable of diffusively encountering one another, but also the rate at which they do so. At a fundamental level, a protein is simply a polypeptide with a heterogeneous incorporation of 20 different structural units, or *aa*'s, which means that the solution behavior of a protein can be expected to be decomposed into the behavior of each individual element. Therefore, the FBM study of polypeptides with simple *aa* sequences could be used to better understand the complex behavior of proteins as they fold.

1.9 Research Goals and Thesis Outline

The overarching goal of this research was to apply the FBM to describe quantitatively the solution conformation and internal dynamics of polypeptides. In doing so, the hope is to convince the reader that the folding domains in proteins are created in the same manner as the *blobs* employed in the FBM analysis of pyrene-labeled polypeptides, and that the

characterization of polypeptide *blobs* will therefore provide a deeper understanding of protein folding. To achieve this goal, this thesis aimed to characterize first, how the formation of structure in polypeptides affected the *blob* size, by monitoring the N_{blob} value as a function of the conformation of a polypeptide, and second, how the *aa* sequence of an unstructured polypeptide affected its internal dynamics and conformational freedom. The following research chapters describe the experiments that were conducted in this thesis to address these two objectives.

A proof of principle that the FBM could report on the conformation and internal dynamics of polypeptides is presented in Chapter 2. The FBM was applied to analyze the fluorescence decays acquired with pyrene-labeled poly(*L*-glutamic acid) (P*L*Glu) and poly(*D*,*L*-glutamic acid) (P*DL*Glu), which adopted α-helical and random coil conformations in DMF, respectively. In combination with MMOs, N_{blob} was shown to report directly on the polypeptide conformation, with the more densely packed α-helical conformation affording a larger N_{blob} value than the unstructured coil, which demonstrated that N_{blob} was sensitive to the local chain density. At the same time, k_{blob} was found to report on the timescale expected for the dynamics of polypeptides. The fact that the N_{blob} range of ~ 10 to 20 *aa*'s matched closely the sizes of *foldons*,[74,80] and that k_{blob} reported on the expected time scale of polypeptide motions, was used as evidence that the FBM was appropriately suited to probe the domains of polypeptides as defined by the diffusive encounters of their *aa* constituents. This idea was further explored in Chapter 3 by studying the effect that charges had on the *blob* size and internal dynamics for P*L*Glu, P*LD*Glu, and another homopolypeptide, poly(*L*-lysine) (P*L*L).

Like P*L*Glu and P*DL*Glu, a combination of FBM and MMOs was able to successfully characterize both the conformation and dynamics of P*L*L in solution.

Chapter 4 aimed to solidify the first goal of this thesis, which aimed to understand how the formation of structure in a polypeptide affected N_{blob}. To do this, conformational changes in P*L*L were induced by changing the solvency conditions in order to map changes in the *blob* size for each of the induced conformations ranging from an unstructured coil to an organized bundle of α-helices. In agreement with Chapters 2 and 3, N_{blob} was shown to increase as the conformational density of P*L*L increased. MMOs were used to provide molecular-level details of the P*L*L conformation, including the interhelical distance separating the P*L*L α-helices inside a bundle.

Moving on to the second objective of this thesis work, a series of copolypeptides containing *aa*'s of increasing side chain size (SCS) were prepared in Chapter 5 to investigate how the structure of an *aa* affected the behavior of the copolypeptides in solution. Since the formation of structure in the copolypeptides was suppressed by the incorporation of a racemic mixture of *aa*'s, the increase in N_{blob} and $k_{blob}\times N_{blob}$ with decreasing SCS was attributed to the increased conformational freedom provided to the copolypeptides from the decreased steric hindrance experienced by the smaller SCSs. Chapter 6 built on these results by investigating the effect of comonomer content on the copolypeptides. Surprisingly, the N_{blob} value of the alanine-glutamic acid copolymers were independent of the alanine content, which suggested that the *blob* size was most affected by *aa*'s with small SCSs. With this information in mind,

the aptly called *blob*-based model was introduced to estimate the folding times of proteins based on the averaged *blob* size expected for pyrene-labeled polypeptides.

The final research Chapter (Chapter 7) refined this model in several significant ways. First the contribution of the pyrenyl label to N_{blob} was determined quantitatively by studying a series of polypeptides where pyrene was separated from the polypeptide backbone by increasing side chain lengths (SCLs). After correcting for the pyrene contribution and considering the effect of the SCS and SCL of different *aa*'s on N_{blob}, the exact *aa* sequence could then be taken into account to determine specific *blob* sizes for each protein. These *blob* sizes were then used to predict the folding times of 145 proteins which showed a strong correlation with the experimental folding time of these proteins, resulting in a correlation coefficient of 0.73. Along with Chapter 6, Chapter 7 suggested that *blob*-based approaches could become good predictors of the folding times of proteins in solution.

Lastly, Chapter 8 summarizes the results of the entire thesis and suggests future experiments to strengthen the conclusions that were drawn from this thesis, and to improve the estimation of protein folding times.

Chapter 2

Pyrene Excimer Fluorescence as a Direct and Easy Experimental Means to Characterize the Length Scale and Internal Dynamics of Polypeptide *Foldons*

Adapted with permission from Casier, R. Duhamel, J. Pyrene Excimer Fluorescence as a Direct and Easy Experimental Means to Characterize the Length Scale and Internal Dynamics of Polypeptide Foldons. *Macromolecules* **2018**, *51*, 3450 – 3457.

2.1 Abstract

This report provides evidence that the *blobs* characterized through the application of the fluorescence *blob* model (FBM) to the analysis of the fluorescence decays acquired with pyrene-labeled polypeptides are equivalent to the *foldons* used in the study of protein folding. The FBM was applied to characterize the length and time scale over which pyrene excimer formation (PEF) took place between pyrene labels covalently attached onto α-helical and partially helical poly(*L*-glutamic acid) (Py-P*L*Glu) in DMF and DMSO, respectively, and unfolded poly(*D*,*L*-glutamic acid) (Py-P*DL*Glu) in both DMF and DMSO. The *blob* size obtained for α-helical Py-P*L*Glu in DMF and the characteristic time determined for the backbone dynamics of unfolded Py-P*DL*Glu matched very closely the expected size of *foldons* and their characteristic folding times, respectively. In particular, the *blob* size was confirmed by conducting Molecular Mechanics Optimizations (MMOs) with HyperChem. Furthermore, the level of pyrene clustering along the polypeptides correlated nicely with their expected conformation, either coiled or helical for the Py-P*DL*Glu or Py-P*L*Glu constructs in DMF, respectively, Consequently, these results suggest that PEF experiments conducted on pyrene-labeled polypeptides provide valuable information on the time and length scales experienced by the amino acids located inside a polypeptide *blob*, and that if polypeptide *blobs* and *foldons* present similar length and time scales, both entities must be equivalent. Since dynamic or spatial information on *foldons* is usually retrieved by conducting NMR or hydrogen exchange mass spectrometry experiments, PEF might thus provide an alternative, possibly simpler, route toward the characterization of polypeptide *foldons* in solution.

2.2 Introduction

All advances made toward understanding the folding pathway of proteins must solve Levinthal's paradox which states that in order for proteins to fold on experimentally observed time-scales, protein folding cannot follow a random conformational search that would exhaust all possible arrangements of the conformational space, since such a search would be prohibitively time-consuming.[1,2] Consequently, any proposed protein folding mechanism must first ensure that reasonable folding times are achieved, i.e. the folding time of a protein must be considerably smaller than the lifetime of the organism generating the protein. Within this context, protein folding is typically described in terms of an energy funnel that guides the spontaneous folding of a high-energy unfolded chain towards partially folded intermediates of continuously decreasing free energy, until the lowest-energy native conformation is reached.[3–5] As part of this continued search for accommodating Levinthal's paradox, *foldons* were recently introduced to represent the subset (N_{Fold}) of amino acid residues which cooperatively fold together in different "sections" of the polypeptide chain. NMR or mass spectrometry (MS) hydrogen exchange (HX) experiments on partially folded proteins have found that short sequences of about 20 amino acids (N_{Fold}=20) exchange their amide hydrogens in D_2O with a similarly slow rate constant suggesting that they all belonged to a same *foldon*.[6-10] The obvious advantage of dividing the protein into a string of *foldons* is that the time required for a number (N/N_{Fold}) of *foldons* made of N_{fold} residues to fold is infinitely shorter than that for an N residue-long protein ($N >> N_{\mathrm{Fold}}$). Assuming that each residue takes about $\tau_p \sim 1$ ps to adopt one of three possible conformations and that N_{Fold} equals 20, the folding time of a 100 residue-long

protein (N=100) would drop from an impossible $3^{N}\times\tau_{p}$ = 1.6×10^{28} yr-long folding time down to $(N/N_{Fold})\times3^{NFold}\times\tau_{p}$ = 17 ms, a reasonable folding time for a living organism.[11] Consequently the *foldon* hypothesis has been reported as being the only viable choice toward resolving Levinthal's paradox.[12] As a result, *foldons* have been the object of considerable scientific scrutiny.

While the size of *foldons* in partially folded proteins has now been characterized,[6,8,12] their existence still raises several unanswered fundamental questions. For instance, having established the existence of *foldons* does not explain their *raison d'être* in the first place. Furthermore N_{Fold} determined by NMR and MS HX experiments for partially folded polypeptides should also be characterized for unfolded polypeptides, an important length scale since it corresponds to the onset of folding. Yet this information remains unknown since all amide hydrogens of a fully unfolded polypeptide would be equivalently exchangeable in an NMR or MS HX experiment. Finally, since the time scale of protein folding is critical to accommodate Levinthal's paradox, the time scale over which *foldons* fold is another important factor that should be included in the characterization of *foldons*.

A recent review suggested that these questions could be answered by conducting pyrene excimer fluorescence/formation (PEF) experiments on pyrene-labeled polypeptides.[13] This report represents a first step toward this goal, by characterizing PEF in two series of pyrene-labeled poly(*L*-glutamic acid) (Py-P*L*Glu) and poly(*D*,*L*-glutamic acid) (Py-P*DL*Glu) in *N*,*N*-dimethylformamide (DMF) and dimethylsulfoxide (DMSO). Due to the absence of stereoregularity, P*DL*Glu is a perfect example of polypeptide adopting a random coil

conformation in DMF and DMSO. P*DL*Glu would thus be representative of a polypeptide at the initial stage of folding. By contrast, P*L*Glu in DMSO and DMF adopts a partially[14] and fully[15] helical conformation, respectively, thus representing states of the polypeptide along and at the end of the folding pathway. These states of the polypeptides can be probed by monitoring excimer formation between every two pyrene labels covalently attached onto the polypeptides and applying the fluorescence *blob* model (FBM) to determine the time and length scale over which PEF occurs.[13]

The FBM recognizes that an excited fluorophore covalently attached onto a polymer cannot probe the entire polymer volume during its finite lifetime, but that it rather probes a subvolume of the polymer coil referred to as a *blob*. Within the framework of the FBM, a fluorescently labeled polymer is represented by a cluster of *blobs*, among which the randomly attached pyrene labels distribute themselves according to Poisson statistics. The mobility range of the excited pyrene is restricted to the diffusion of the polymer structural unit that pyrene is bound to. As such, the volume of a *blob* is limited by the mobility of the labeled structural unit, and therefore the *blob* volume reflects the structural unit mobility on the time-scale of backbone motion. As discussed in a recent review,[13] a *blob* should be to the FBM what a *foldon* is to protein folding, so that the characterization of a polypeptide *blob* is bound to provide structural and dynamic information about a *foldon*. These ideas were put to the test herein by applying the FBM to fit the pyrene monomer and excimer fluorescence decays of Py-P*L*Glu and Py-P*DL*Glu in DMF and DMSO.

2.3 Experimental

Materials. Acetic acid (glacial, Sigma), *N*,*N*′-diisopropylcarbodiimide (DIC, Sigma, 99 %), *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide hydrochloride (EDC·HCl, Sigma, ≥ 99.0 %), *N*,*N*-dimethylformamide (DMF, Sigma, ≥ 99.8 %), dimethyl sulfoxide (DMSO, Sigma, ≥ 99.9 %), hydrochloric acid (HCl, Fisher, 34 wt% in water), *N*-hydroxysuccinimide (NHS, Sigma, 98 %), 1-pyrenemethylamine hydrochloride (PyMA·HCl) and sodium hydroxide (NaOH, Sigma, ≥ 97.0 %) were used as received. Deionized water (DIW) was obtained from a Biopure Series 4400 Single Pass Reverse Osmosis system.

The poly(*L*-glutamic acid sodium salt) (P*L*GNa, Alamanda Polymers, M_n = 121300 g·mol^{-1}, *DP* = 803, *PDI* = 1.02) and poly(*D*,*L*-glutamic acid sodium salt) (P*DL*GNa, Alamanda Polymers, M_n = 118400 g·mol^{-1}, *DP* = 784, *PDI* = 1.06) were received with an unprotected *N*-terminus. To prevent possible complications arising from the reaction of the amine end with one of the glutamate side chains during the pyrene-labeling step, the *N*-terminal was capped with an acetyl group. To this end, succinimidyl acetate was first prepared by adding NHS (0.90 g, 7.8 mmol) to an ice-bath cooled round-bottom flask containing glacial acetic acid (5 mL, 87.4 mmol). With vigorous stirring, DIC (1.4 mL, 8.9 mmol) was added dropwise. After allowing to react overnight, the acetic acid was removed via rotary evaporation at 40 °C, and the resulting white paste was dissolved in a minimal amount of methanol. The methanol solution was then cooled in an ice-bath and the white crystal precipitate was collected via suction filtration yielding succinimidyl acetate (0.8 g, 5.1 mmol, 65 % yield). Acetylation of the *N*-terminus of PGlu was carried out by first dissolving PGNa (400 mg, 3.3 μmol) in a

solution of DMF (6 mL) and DIW (6 mL) in a 20 mL vial with vigorous stirring. After complete dissolution, succinimidyl acetate (31 mg, 200 μmol) was added and the solution was left to stir overnight. The acetylated PGNa was then purified by dialysis against a successive series of solutions starting with a 4:1 DMF:water mixture, followed by a 1:1 DMF:water mixture, water, and lastly an aqueous NaOH solution (pH 9). The solution was then freeze-dried to remove the water and stored in a −20 °C freezer.

Pyrene Labeling. Adapting previously published procedures,[16,17] both the acetylated P*L*GNa and P*DL*GNa were labeled with PyMA yielding the pyrene-labeled PGlu constructs Py-P*L*Glu and Py-P*DL*Glu, respectively. As an example of one synthesis, Py-P*L*Glu labeled with 6.9 mol% pyrene was prepared as follows: P*L*GNa (39.9 mg, 0.26 mmol eq. of glutamic acid (Glu)) was added to a 20 mL vial along with a small magnetic stir bar, 8 mL DMF and 8 mL DIW. After dissolving the P*L*GNa with rapid stirring, PyMA·HCl (7.7 mg, 29 μmol) was added in the solution and left to stir for 30 min. to ensure complete dissolution. EDC·HCl (5.6 mg, 29 μmol) was then added and the 20 mL vial was capped, covered with aluminum foil to prevent excess external light irradiation, and left to react for 36 h. The reaction mixture was then transferred to a dialysis tube (14 kDa molecular weight cut-off, Regenerated Cellulose) and dialysed against a 20 vol% DIW solution in DMF for 4 h in a 1 L beaker, after which time the dialysate solution was replaced with a 50 vol% DIW in DMF solution and dialysed for another 6 h. After this time, the solution was dialysed against pure DIW overnight, followed by dialysis against pH 6 DIW and pH 9 DIW for 4 h each, respectively. Before the last dialysate solution of pH 9 DIW was discarded, a ca. 20 mL aliquot of the dialysate solution was freeze-dried to

determine the mass ratio (λ_{DS}) of dialysate solids per g of Py-P*L*GNa in the dialysis solution. The purified Py-P*L*GNa solution was then accurately weighed into vials and freeze-dried to remove the water.

Pyrene Content. The fraction (x) of glutamate residues that were labeled with pyrene was determined using Equation 2.1, where λ_{Py} represents the moles of pyrene per gram of sample, and M_{GNa} and M_{PyGlu} are the molecular weights of a glutamic acid residue (151.1 g·mol^{-1}) and the pyrene-labeled residue (342.4 g·mol^{-1}), respectively. In order to determine λ_{Py}, a mass ($m_{Py\text{-}PGNa/DS}$) of at least 5.0 mg of freeze-dried pyrene-labeled PGNa was accurately weighed out. To ensure the accuracy of $m_{Py\text{-}PGNa/DS}$, the mass of dialysate solids (m_{DS}) in the Py-PGNa sample needed to be determined. This was done by multiplying $m_{Py\text{-}PGNa/DS}$ by λ_{DS} to yield m_{DS}. Subtracting m_{DS} from $m_{Py\text{-}PGNa/DS}$ yielded the actual mass of pure Py-PGNa ($m_{Py\text{-}PGNa}$). The Py-PGNa was then dissolved in 2 – 3 mL of DIW before 1 M HCl was added dropwise until the Py-PGNa precipitated from the solution as it was converted into Py-PGlu; two additional drops of HCl were then added to ensure the complete protonation of the PGNa backbone to PGlu. The suspension was then vortexed for ca. 30 s before it was frozen for freeze-drying. The freeze-dried Py-PGlu was then dissolved into a known volume of DMF (V_{DMF}). The UV absorbance of the solution was then measured (see Figure S2.1 in Supporting Information (SI)), and the molar concentration of pyrene [*Py*] was determined using a molar extinction coefficient of 40,000 M^{-1}·cm^{-1} at the 344 nm absorbance peak.[16] The pyrene content (λ_{Py}) expressed in mol of pyrene per gram of Py-PGNa was obtained by taking the ratio [Py]/($m_{Py\text{-}PGNa}/V_{DMF}$). The molar fraction ($f_{Py}$) could then be calculated by applying Equation 2.1.

$$x = \frac{M_{GNa}}{\lambda_{Py}^{-1} + M_{GNa} - M_{PyGlu}} \tag{2.1}$$

Steady-State Fluorescence. All steady-state fluorescence (SSF) spectra were acquired with a PTI spectrofluorometer equipped with a xenon arc lamp using an excitation wavelength of 344 nm and scanning the emission from 350 to 600 nm over 50 s with excitation and emission slit-widths of 2 and 1 nm, respectively. After preparing solutions containing ca. 2.5×10^{-6} M pyrene, oxygen was removed by out-gassing the solutions with a gentle flow of N_2 gas (high purity, 99.99 %) for ca. 25 min before acquiring the SSF spectrum. Oxygen is an efficient and well-known quencher of long-lived excited dyes like pyrene.[18] The fluorescence intensity of the monomer (I_M) and excimer (I_E) were calculated using the area under the SSF spectrum from 372 to 378 nm and from 500 to 530 nm, respectively. I_M and I_E were employed to determine the I_E/I_M ratio of the Py-PGlu solutions which represents the PEF efficiency for a given construct.

Time-Resolved Fluorescence. Samples were prepared in a similar manner as for the steady-state fluorescence measurements. The fluorescence decays of pyrene were acquired using an IBH TC-SPC spectrofluorometer in conjunction with a NanoLED-340 laser. All decays were collected with an excitation wavelength of 344 nm over 1024 channels. Using a time-per-channel of 2.04 $ns \cdot ch^{-1}$, the monomer fluorescence was collected until a total of at least 40,000 counts was obtained for the fluorescence decay maximum at an emission wavelength of 375 nm using a 370 nm cut-off filter in front of the emission lens to minimize light scattering. The excimer fluorescence decays were acquired with a decay maximum of at least 20,000 counts

at 510 nm with a 495 nm cut-off filter using a time-per-channel of 1.02 ns·ch^{-1}. The instrument response function was collected using a LUDOX dispersion in water with a peak maximum of at least 20,000 counts at 344 nm.

Circular Dichroism. Using a Jasco J-715 spectropolarimeter, circular dichroism (CD) spectra were averaged over ten scans from 400 to 250 nm using a 0.10 mm pathlength quartz cell containing a solution with a pyrene concentration of ca. 2.5×10^{-3} M. A solvent background correction was applied to each spectrum. The molar ellipticity spectra acquired with the Py-PGlu solutions are provided in Figure S2.2.

Fluorescence decay analysis: The fluorescence *blob* model (FBM) was used to fit the fluorescence decays of Py-P*L*Glu and Py-P*DL*Glu in solution. The FBM has been described in numerous reviews and the reader is directed toward these documents for more information (see reference #13). In brief, the FBM divides a macromolecule into segments of equal volume, where each unit volume is referred to as a *blob*. Since pyrene is covalently and randomly bound along the macromolecular backbone, pyrene will distribute itself among the *blobs* according to Poisson statistics.[19] Within each *blob*, the FBM considers four distinct species of pyrene. Py_{free} represents the excited pyrenes which do not form excimer during their lifetime and emit with their unquenched lifetime τ_M. Py_{k2} represents the excited pyrenes that are very close to a ground-state pyrene and can form excimer through a rapid rearrangement process with a rate constant k_2. Py_{diff} represents the excited pyrenes that diffuse to a ground-state pyrene with a rate constant k_{blob} to generate the Py_{k2} species. The species Py_{agg} represents the preassociated pyrenes which directly form an excimer upon excitation. Two excimer species were needed in

the analysis. This situation is often encountered when dealing with pyrene excimer generated in more rigid environments as with polypeptides where pyrene is bound to the polymer via an amide bond.[20] The excimers obtained with the strongest and smallest contributions were referred to as *E0** and *D** and were assumed to be the result of good and poor stacking between two pyrene labels, respectively. Equations S2.2 and S2.3 were used to fit the monomer and excimer fluorescence decays.

To find the optimal k_2 value for each PGlu construct, the monomer and excimer fluorescence decays were first fitted globally according to Equations S2.1 and S2.2 in SI using the *globmis90lbg* program that allowed for k_2 to be optimized in the analysis. After analysing all the pyrene contents for a given PGlu construct and solvent, the k_2 value was averaged and the decays were reanalysed with the program *globmis90obg* using a k_2 value fixed to the average. Using the framework of the FBM, the parameters retrieved from the fit of the monomer and excimer fluorescence decays yield structural and dynamical information on the macromolecule of interest. Among those, the parameters $<n>$, the average number of ground-state pyrenes within a *blob*, and k_{blob}, which represents the rate constant for diffusive encounters between two structural units inside a *blob* bearing a pyrene label, are of particular interest. Since pyrene is distributed randomly along the polymer, $<n>$ can be introduced into Equation S2.4 to calculate N_{blob}, which is the average number of structural units encompassed within a *blob*. In Equation S2.4, f_{Mfree} is the molar fraction of pyrene labels Py_{free} that emit with their unquenched lifetime τ_M in the monomer decay, and f_{Py} has been defined in Equation 2.1.

The monomer and excimer decays of Py-P*L*Glu and Py-P*DL*Glu in DMF and DMSO were globally fitted using Equations S2.1-2.2. The goodness-of-fit of the global analysis of the monomer and excimer decays was based on three criteria: Firstly, the fit had to lie between the experimental data points, which is quantified by the random distribution of the residuals about zero. Secondly, the autocorrelation function needed be randomly distributed around zero. Lastly the χ^2 value had to take a value smaller than 1.30, where a value of unity would be a perfect fit. An example of fit of the monomer and excimer fluorescence decays globally analyzed using Equations S2.1 and S2.2 is shown in Figure S2.3. The parameters retrieved from the analysis of the fluorescence decays are listed in Tables S2.1 – S2.6.

2.4 Results and Discussion

Circular dichroism: The conformation of the Py-P*L*Glu and Py-P*DL*Glu constructs was assessed by conducting circular dichroism (CD) experiments in both DMF and DMSO. Due to the strong absorption of DMF and DMSO around 200 nm where CD is typically applied to infer the existence of secondary structure in polypeptides, the molar ellipticity [θ] of the PGlu backbone could not be measured in this wavelength range. Fortunately, [θ] of pyrene has been shown to reliably report on the backbone conformation of Py-PGlu.[16,17,21] The CD spectra acquired with the Py-PGlu samples are shown in Figure S2.2 in SI. To quantify [θ] for each sample, the area under the 1Bb band of pyrene was plotted as a function of pyrene content. The resulting [θ] values are presented in Figure 2.1.

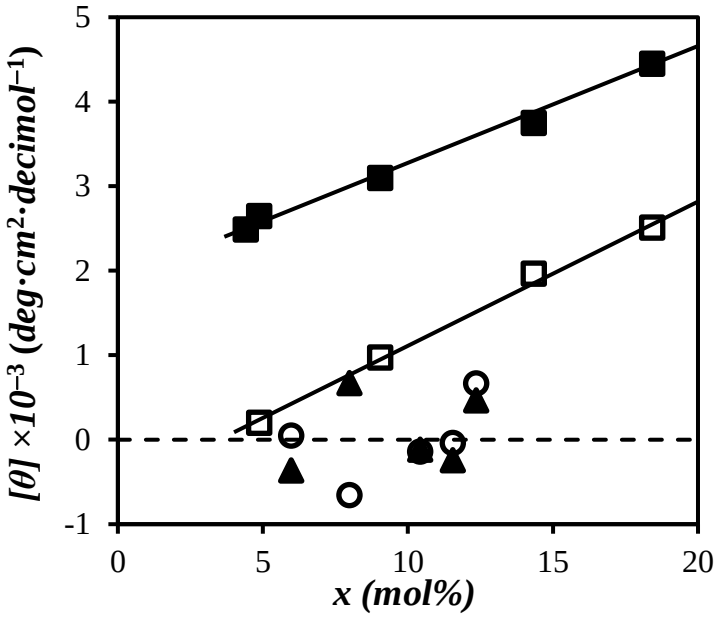


Figure 2.1. Average molar ellipticity values of the ^1B_b band of pyrene for Py-P*L*Glu as a function of pyrene content (*x*) in DMF (■) and DMSO (□) and Py-P*DL*Glu in DMF (▲) and DMSO (○). The solid and dashed lines are added to guide the eyes for the molar ellipticity of the Py-P*L*Glu and Py-P*DL*Glu constructs, respectively.

As shown in Figure 2.1, the Py-P*DL*Glu constructs exhibited very low [θ] values scattered around zero indicating that Py-P*DL*Glu adopts a coiled conformation in both DMF and DMSO. In DMF, Py-P*L*Glu exhibited a relatively high [θ] that increased with increasing pyrene content. These [θ] values matched very closely those reported earlier[16,17] and reflect the ordering of the pyrene labels along α-helical P*L*Glu in DMF. Lastly, Py-P*L*Glu showed significantly lower [θ] values as a function of pyrene content in DMSO than in DMF. However contrary to Py-P*DL*Glu where [θ] equaled zero within experimental error in both DMF and

DMSO, Py-P*L*Glu in DMSO yielded [θ] values that were clearly greater than zero. This suggested that Py-P*L*Glu retained some secondary structure in DMSO, probably a partially helical conformation. In summary, the CD results showed that P*DL*Glu adopted a fully coiled conformation in DMF and DMSO and that P*L*Glu adopted a partially and fully helical conformation in DMSO and DMF, respectively. Consequently, they represent important stages along the folding pathway of a polypeptide, at the beginning of folding when the chain is coiled, at intermediate stages when the polypeptide is partially folded, and at the end of folding when the polypeptide is fully folded.

Steady-state fluorescence: The normalized steady-state fluorescence spectra of the Py-P*L*Glu constructs in DMF are shown in Figure 2.2A. They exhibit the typical spectral features of a pyrene-labeled macromolecule with the pyrene monomer emitting sharp bands between 370 and 420 nm and the excimer showing a broad structureless fluorescence centered at 480 nm. As expected, the fluorescence intensity of the excimer relative to that of the monomer increased with increasing pyrene content, a consequence of the increased local pyrene concentration that promotes pyrene-pyrene encounters, and thus PEF. The same behavior was observed for the Py-P*DL*Glu sample, but Py-P*L*Glu formed substantially more excimer than Py-P*DL*Glu for a same pyrene content as illustrated in Figure 2.2B by the higher excimer-to-monomer fluorescence intensity (I_E/I_M) ratios. This result was attributed to the compact α-helix structure of Py-P*L*Glu that brought the pyrene labels closer to each other than the Py-P*DL*Glu random coil since the helical structure has much less excluded volume than a coiled chain.

Figure 2.2C displays the I_E/I_M ratios obtained for Py-P*L*Glu and Py-P*DL*Glu in both DMF and DMSO. The absence of stereoregularity for Py-P*DL*Glu leads it to adopt a coiled conformation in both DMF and DMSO, as confirmed by CD measurements (Figure 2.1), and the only difference in the I_E/I_M ratios should arise from the change in solvent viscosity between the two solvents. Since DMSO (η = 1.99 cP) is about twice as viscous as DMF (η = 0.802 cP) at 25 °C and since I_E/I_M is inversely proportional to solvent viscosity,[13] the I_E/I_M ratio of Py-P*DL*Glu in DMSO was expected to be about half of that obtained in DMF. However as seen in Figure 2.2C, the I_E/I_M ratios for Py-P*DL*Glu in DMSO and DMF overlapped. As it turns out, the efficiency of excimer formation between an excited and a ground-state pyrene has been found to be about twice as great in DMSO than in DMF.[22] As a result, the competing effects of solvent viscosity and PEF efficiency cancelled one another out, resulting in an unchanged I_E/I_M ratio between the two solvents. Knowing that the I_E/I_M ratios in DMF and DMSO were not affected by the difference in solvent viscosity, the differences observed in Figure 2.2C for Py-P*L*Glu in DMF and DMSO were indicative of a conformational change experienced by P*L*Glu between the two solvents. The I_E/I_M ratios of Py-P*L*Glu in DMSO tended to follow the same trend of the helical Py-P*L*Glu in DMF, but with a slightly lower I_E/I_M ratio, thus confirming the conclusions drawn from CD measurements that Py-P*L*Glu retained some helical conformation in DMSO. Consequently Py-P*DL*Glu in DMSO provided the opportunity to probe the dynamics of a partially folded polypeptide.

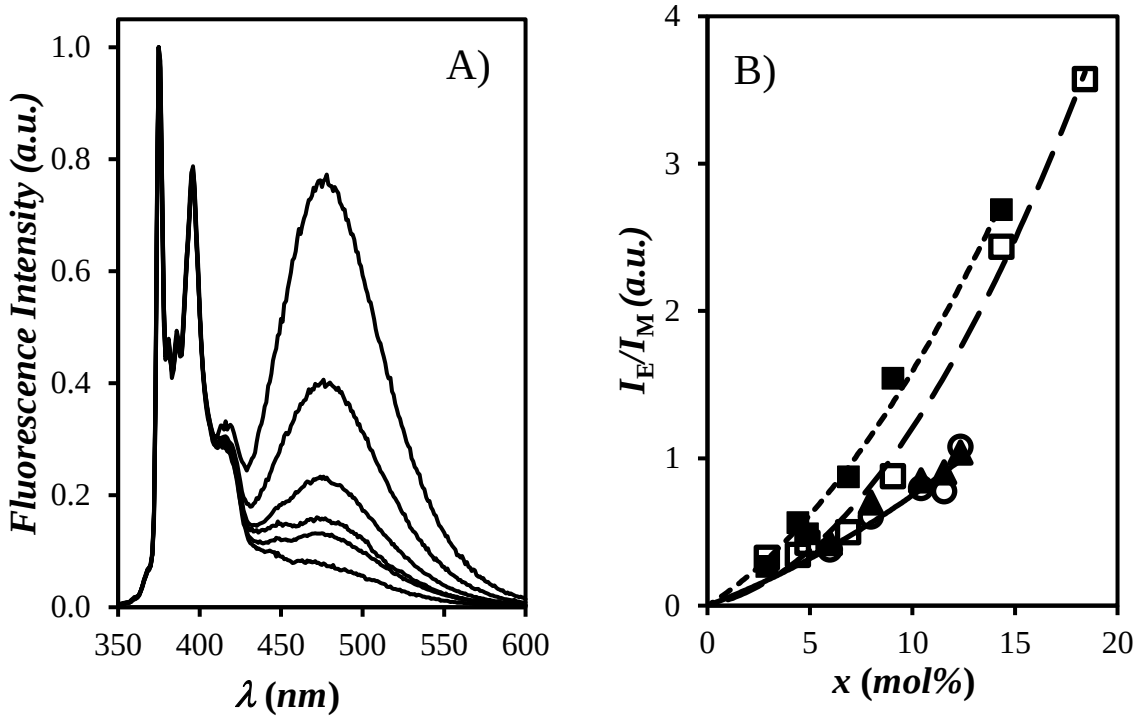


Figure 2.2. A) Steady-state fluorescence spectra of Py-P*L*Glu in DMF, B) the I_E/I_M ratios of Py-P*L*Glu (■) and Py-P*DL*Glu (▲) in DMF, and C) the I_E/I_M ratios of Py-P*L*Glu in DMF (■) and DMSO (□) and Py-P*DL*Glu in DMF (▲) and DMSO (○). The lines for DMF (solid) and DMSO (dashed) were added to guide the eyes.

Size of a blob: The monomer and excimer fluorescence decays of Py-P*L*Glu and Py-P*DL*Glu were globally analyzed with the FBM to yield the *blob* size (N_{blob}) representing the average number of structural units encompassed within a single *blob* and the rate constant k_{blob} of excimer formation within a *blob* that contains one excited and one ground-state pyrene. Since a *blob* represents the volume that one glutamic acid (Glu) bearing a pyrene label can probe while pyrene remains excited, it also represents that volume of the polypeptide in solution

where Glu's can interact with one another to generate preferential interactions. As such, a *blob* is not only analogous to a *foldon*, but also provides a rationale as to the nature of a *foldon*. N_{blob} was calculated according to the procedure described in SI and it was plotted in Figures 2.3A and B as a function of pyrene content for Py-P*L*Glu and Py-P*DL*Glu in DMF and DMSO. Within error, N_{blob} remained constant as a function of pyrene content for each PGlu chain for a given solvent, taking values of 21 ± 1 and 18 ± 1 for Py-P*L*Glu and 10 ± 1 and 11 ± 2 for Py-P*DL*Glu in DMF and DMSO, respectively. We note that the N_{blob} value of 21 ± 1 for Py-P*L*Glu in DMF matches that of 22 ± 2 found in earlier studies.[16,17]

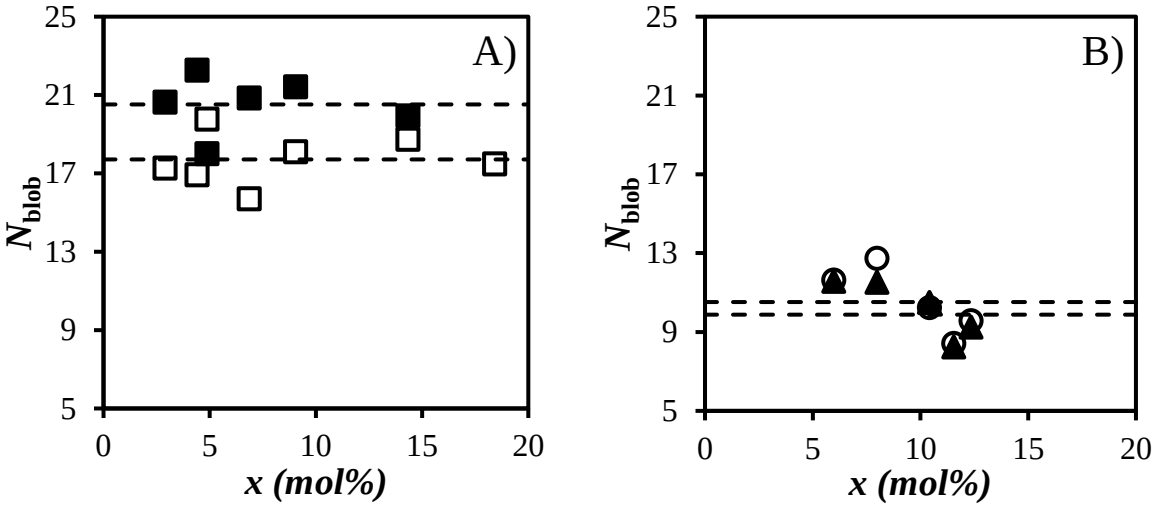


Figure 2.3. N_{blob} as a function of pyrene content for A) Py-P*L*Glu and B) Py-P*DL*Glu in DMF (solid) and DMSO (hollow). The dashed lines represent the averages.

Py-P*L*Glu was expected to adopt a helical conformation in DMF,[16,17,21] while Py-P*DL*Glu would adopt a coiled conformation in both solvents. Since a helical PGlu chain

generates less free volume than its coiled counterpart, the helical PGlu chain is expected to generate a denser environment. Consequently, the observed larger *blob* size of 21 ± 1 for Py-P*L*Glu in DMF than that of 10 ± 3 for Py-P*DL*Glu in both DMF and DMSO was consistent with the expected backbone conformations of the polypeptides. The similar N_{blob} values for Py-P*DL*Glu in DMF and DMSO also supported the conclusions from the I_E/I_M ratios and CD measurements that Py-P*DL*Glu exhibits a coiled structure regardless of the solvent. Similarly, the relatively large N_{blob} value of 18 ± 1 found for Py-P*L*Glu in DMSO reflected a partially folded conformation. Based on these N_{blob} values, it would appear that a *blob* would involve 10 Glu residues for an unfolded polypeptide, but that N_{blob} increases progressively to reach 18 Glu residues for a partially helical PGlu and 21 Glu's for a fully α-helical PGlu. We note with interest that partially or fully folded P*L*Glu yields N_{blob} values of about 20 which are comparable to reported N_{Fold} values.[6,8,12]

Molecular Mechanics Optimizations: The N_{blob} values of 10 and 21 obtained for the fully unfolded and helical P*L*Glu could be further validated by conducting Molecular Mechanics Optimizations (MMOs) with HyperChem where pyrene labels are attached *in silico* onto P*L*Glu and their ability to overlap is monitored as a function of the number of Glu's separating two pyrene labels according to an already published procedure.[7,8] A description of the procedure implemented for the MMOs is provided as SI. Figures 2.4A – D depict several of the HyperChem renderings used to determine the pyrene carbon overlap.

As seen in Figure 2.4E, a good overlap between two pyrene labels, corresponding to a greater than 7 number of carbon atoms of one pyrene overlapping the frame of the second

pyrene, was obtained for Glu's located up to 11 and 5 Glu's away from the reference pyrene for the helical and coiled chains, respectively. A good overlap observed by MMOs was expected to result in PEF and should be recorded by our fluorescence experiments. Since excimer can be formed in either direction from the reference residue, the *blob* size is expected to equal twice the number of residues for which good overlap can occur, plus one for the reference residue. This led to expected N_{blob} values of (11×2+1=) 23 and (5×2+1=) 11 for the helical and coiled conformations, respectively.

Since the coiled chain is not symmetrical, the *blob* size was determined three times by moving the reference pyrene to another starting position. On average, N_{blob} obtained by MMOs was found to equal 10 ± 1 for P*L*Glu in a random coil conformation. The MMOs conducted with the α-helical P*L*Glu was only determined once since the backbone had a set repeating structure, and the overlap could not change depending on the location of the reference pyrene. The expected N_{blob} values of 23 and 10 ± 1 determined by MMOs were found to match remarkably well the N_{blob} values determined by the FBM found to equal 21 ± 2 and 10 ± 3 for the helical and coiled chains, respectively, thus confirming that the experimentally measured N_{blob} values were within expectations.

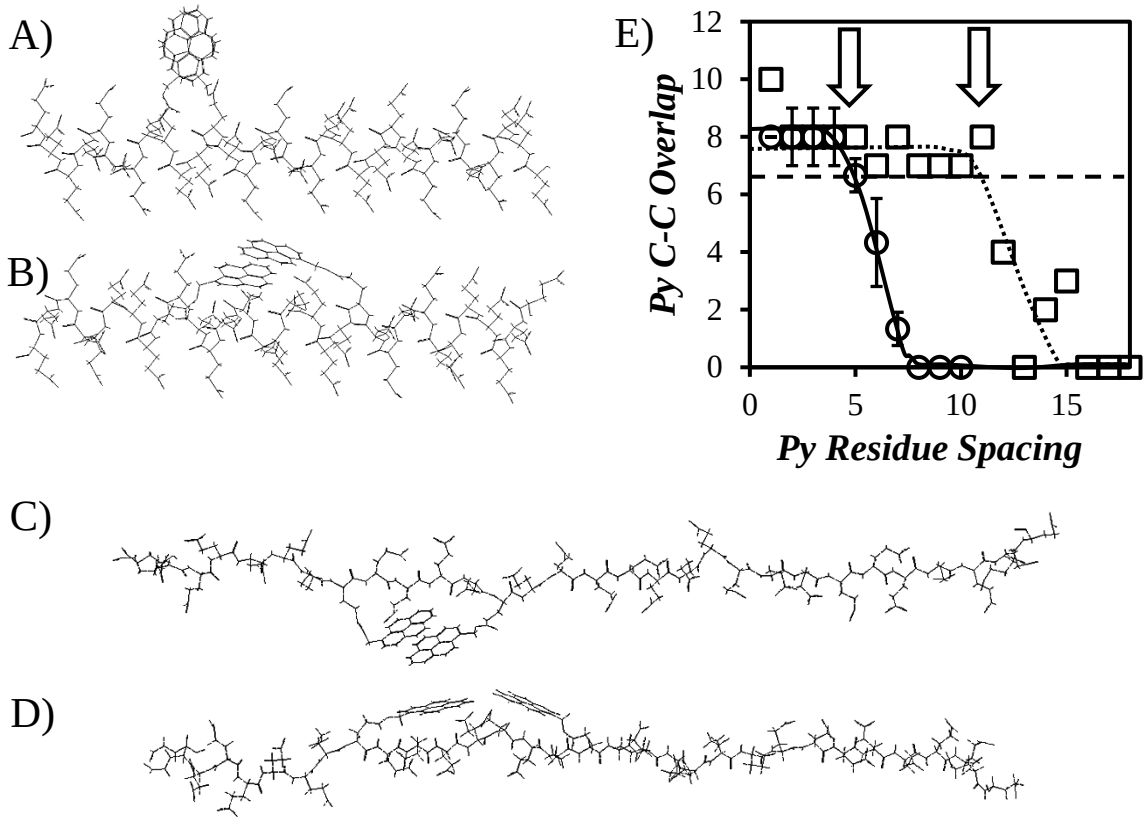


Figure 2.4. Illustrations of the quality of pyrene carbon overlap for Py-PGlu made of 32 glutamic acids and labeled with a reference pyrene on the 9th residue from the left and having the second pyrene labeled on the A) 12th (good overlap) and B) 21st (poor overlap) residue when PGlu adopts a helical conformation and, the C) 14th (good overlap) and, D) 17th (no overlap) residue when PGlu adopts a random coil conformation, respectively. E) Overlap between two pyrene pendants as a function of the number of Glu's between pyrene labels for a PGlu chain in a helical (□) and a random coil (○) conformation. The dashed line corresponding to a carbon-carbon overlap of 7 marks the boundary between a poor and good overlap.[16,17] The solid and dotted lines are used to guide the eye. The arrows indicate the onset of poor overlap.

Molar fractions of pyrene: One other set of parameters retrieved from the FBM was the molar fractions f_{diff}, f_{k2}, f_{agg}, and f_{free} of the different pyrene species. A plot of these molar fractions as a function of pyrene content is given in Figure 2.5. Several trends can be observed for the molar fractions as a function of pyrene content. In all cases, f_{free} decreased as the pyrene content increased. At low pyrene contents, the pyrene pendants were sufficiently far apart that each *blob* contained at most one pyrene so that PEF could not occur and the pyrene labels emitted as if they were free in solution. As the pyrene content increased, the number of pyrenes per *blob* increased, resulting in a decrease in f_{free}. The decrease was most notable in the Py-P*L*Glu samples, likely due to their helical conformation in both DMF and DMSO. Again, since the helical segments had less free volume than the coiled segments, the local pyrene concentration was expected to grow more rapidly in the helical rather than the coiled segments with increasing pyrene content.

Another well-defined trend in Figure 2.5 was the continuous increase in both f_{k2} and f_{agg} with increasing pyrene contents. The significant proportion of aggregated pyrenes reported by the FBM was supported by the UV absorption spectra of pyrene (Figure S2.1) having P_A values of 2.72 ± 0.03 and 2.72 ± 0.03 for Py-P*L*Glu and Py-P*DL*Glu in DMF, respectively. These P_A values are smaller than 3.0 which reflects pyrene aggregation.[23] As pyrene content increased, the probability of having pyrenes sufficiently close to one another to rapidly form excimer with a rate constant k_2 or instantaneously through direct excitation of a pyrene aggregate was expected to increase as well. This effect appeared to be more pronounced for the helical Py-P*L*Glu samples in DMF and DMSO than for the coiled Py-P*DL*Glu samples.

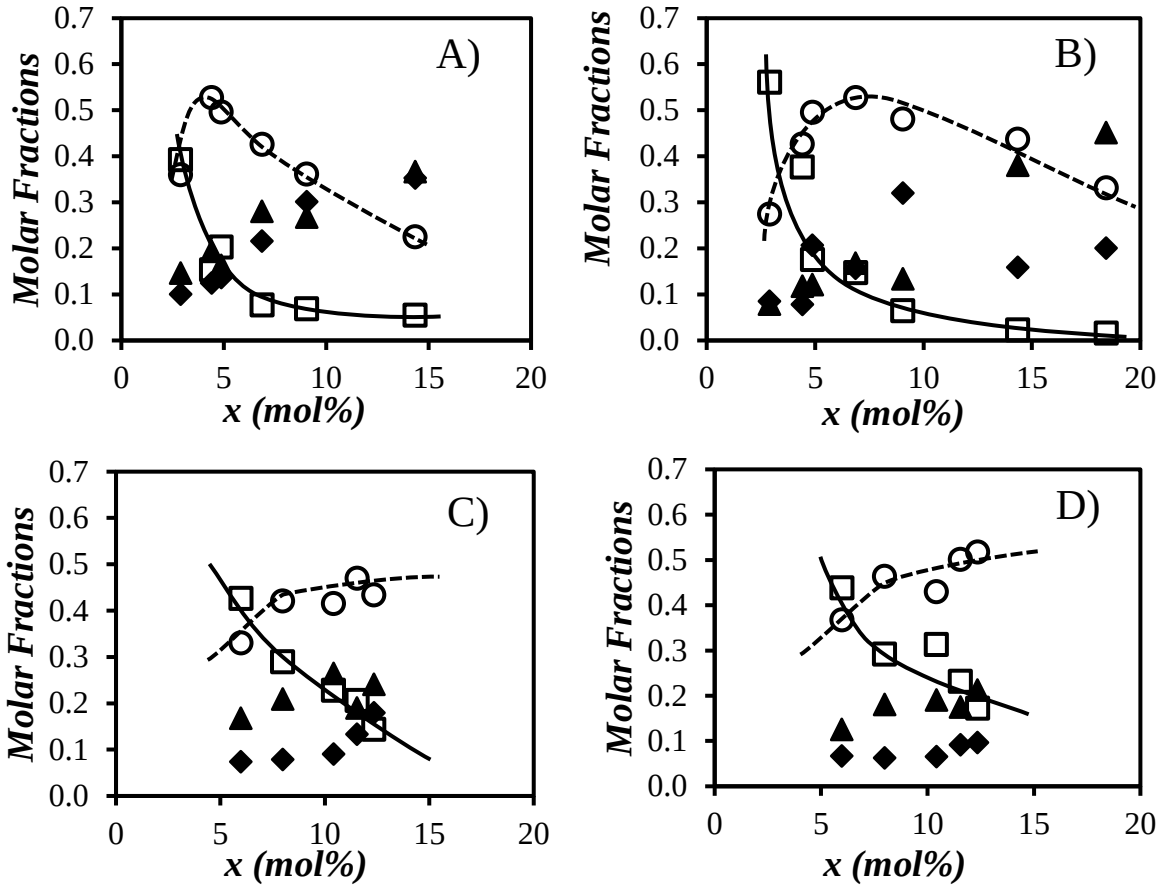


Figure 2.5. Molar fractions of pyrene species: f_{diff} (○), f_{free} (□), f_{k2} (▲), and f_{agg} (◆), for Py-P*L*Glu (A,B) and Py-P*DL*Glu (C,D) in DMF (A,C) and DMSO (B,D). To help guide the eyes, the pyrenes which rapidly form excimer (f_{k2} and f_{agg}) were shaded and lines were added for f_{free} (—) and f_{diff} (---).

Lastly, the trends of f_{diff} were slightly different between Py-P*L*Glu and Py-P*DL*Glu over similar pyrene contents. For Py-P*DL*Glu, f_{diff} increased slightly with pyrene content. Initially

at relatively low pyrene content, the pyrenes were spaced sufficiently far apart that the molar fraction (f_{free}) of isolated pyrenes was relatively high. As the pyrene content increased, the average distance between pyrene labels decreased resulting in an increase in the fraction of pyrenes that formed excimer by diffusion. This trend continued until the majority of *blobs* contained at least two pyrenes resulting in an f_{free} value that approached zero. A similar pattern of increasing f_{diff} was observed for the Py-P*L*Glu samples at pyrene contents below 5 and 7 mol% in DMF and DMSO, respectively. Since f_{free} tends to zero above these labeling levels, all the pyrene pendants were able to form excimer, and f_{diff} reached its maximum value. As more pyrene labels were attached onto the polypeptide, more pyrenes were close enough to rapidly form excimer without diffusion, resulting in a decrease in f_{diff} combined with the observed increase in f_{agg} and f_{k2}. As a quick aside, the f_{diff} values of Py-P*L*Glu in DMF were quite low at higher pyrene content, with $f_{diff} = 0.23$ at 14.3 mol% pyrene. Since the FBM relies on f_{diff} to retrieve information on the blob size and dynamics, it became more challenging to extract meaningful parameters from the analysis as f_{diff} decreased. For this reason, the highest pyrene content of Py-P*L*Glu with 18.4 mol% pyrene could not be meaningfully analysed in DMF by the FBM due to a prohibitively low f_{diff} value.

As a more in-depth study of the molar fractions, a simple simulation of the molar fractions of pyrene species that rapidly formed excimer without long-range diffusion (f_{k2} and f_{agg}) was conducted. Given that the fractions of pyrene that rapidly formed excimer did not require long-range dynamics, the fraction of these species should be equal to the probability of having pyrene labels close enough to one another to form excimer without backbone motion.

To determine this probability, a string of 800 boxes, which was close to the degree of polymerization for both the P*L*Glu and P*DL*Glu used in this study, were randomly filled with pyrene, where each box had a probability to be populated equal to *x*. For P*DL*Glu, excimer could form rapidly if the pyrenes were directly adjacent to one another (aggregated pyrenes) or if they were up to one residue away (pyrenes forming excimer through k_2). In addition to these restraints, the pyrenes which were directly adjacent to one another in space along the helix of P*L*Glu were also considered. Using HyperChem, this was found to include pyrenes two or three units away. So, for a pyrene in position *i* along the backbone, the pyrenes in position [*i*–2, *i*–1, *i*+1, *i*+2] or [*i*–4, *i*–3, *i*–2, *i*–1, *i*+1, *i*+2, *i*+3, *i*+4] were considered to be able to rapidly form excimer for the coiled P*DL*Glu and the helical P*L*Glu constructs, respectively. The sum of the molar fractions [$f_{agg} + f_{k2}$] of the simulated chains was determined by averaging over three independent simulations composed of 1000 chains in each simulation. The calculated probabilities were compared to the measured molar fractions in Figure 2.6 for Py-P*L*Glu and Py-P*DL*Glu in DMF, the more well-behaved solvent.

The overlapping trends for both the coiled and helical chains show very good agreement between the fractions obtained by the FBM and the simulated PGlu chains. In both cases, the sum [$f_{agg} + f_{k2}$] increased with increasing pyrene content, with the helical PGlu yielding higher fractions compared to the coiled PGlu for a same pyrene content. This expected trend can be rationalized by the higher local pyrene concentration resulting from the helical PGlu. In addition, the overlapping trends between the simulated and measured fractions support the claim that the PGlu chains adopt their expected conformations in solution and that

pyrene was indeed randomly distributed along the PGlu backbone during the labeling step. If labeling was to occur in a clustered manner, the sum [f_{agg} + f_{k2}] obtained from the molar fractions of pyrene measured from the FBM would be significantly higher than those of the simulation where the occupancy of pyrene in the boxes was selected randomly. The absence of such a deviation confirms that the pyrenes were distributed randomly along the PGlu backbone both in the helical and coiled conformation.

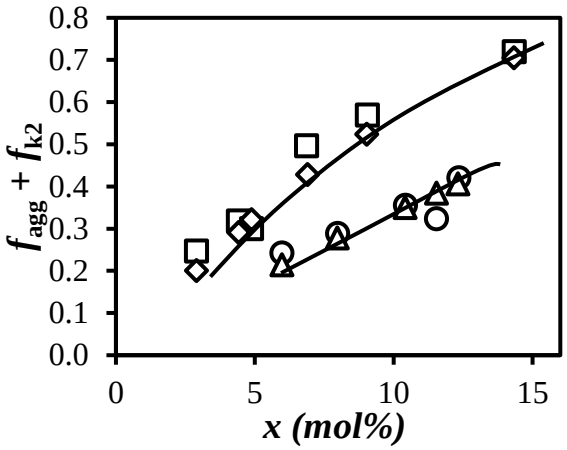


Figure 2.6. Comparison of the molar fraction of pyrenes in close proximity to one another along the PGlu backbone found for the helical Py-P*L*Glu experimentally (f_{agg} + f_{k2}, ◻) and simulated (◇) and for the coiled Py-P*DL*Glu experimentally (f_{agg} + f_{k2}, ○) and simulated (△) in DMF.

Blob Dynamics: The last FBM parameter to discuss was k_{blob} shown in Figures 2.7A and B, which took similar values for P*L*Glu and P*DL*Glu in both DMF and DMSO. However the

meaning of k_{blob} differs from P*L*Glu to P*DL*Glu. Since P*L*Glu is fully and partially helical in DMF and DMSO, k_{blob} for P*L*Glu in Figure 2.7A describes side chain motion since the P*L*Glu backbone is locked in an α-helical conformation. Consequently only k_{blob} for unfolded P*DL*Glu truly reflects the polypeptide backbone dynamics. In fact, $(k_{blob})^{-1}$ represents the characteristic time scale over which a 10-Glu sequence in P*DL*Glu undergoes backbone motion. Assuming that each Glu needs a characteristic time τ_p ($=10^z$ s) to probe one of three possible conformations, an estimate of τ_p is given by $(k_{blob}\times 3^{N_{blob}})^{-1}$ yielding a z value ($z = log_{10}(\tau_p)$) of -12.0 ± 0.7 depending on the k_{blob} and N_{blob} values obtained for a given Py-P*DL*Glu construct in Figure 2.7B. Consequently, our PEF experiments predict that τ_p is in the picosecond range, in perfect agreement with the expected dynamics of polypeptides.[1-5] Thus $(k_{blob})^{-1}$ is a representative folding time for a 10 Glu sequence along the P*DL*Glu backbone.

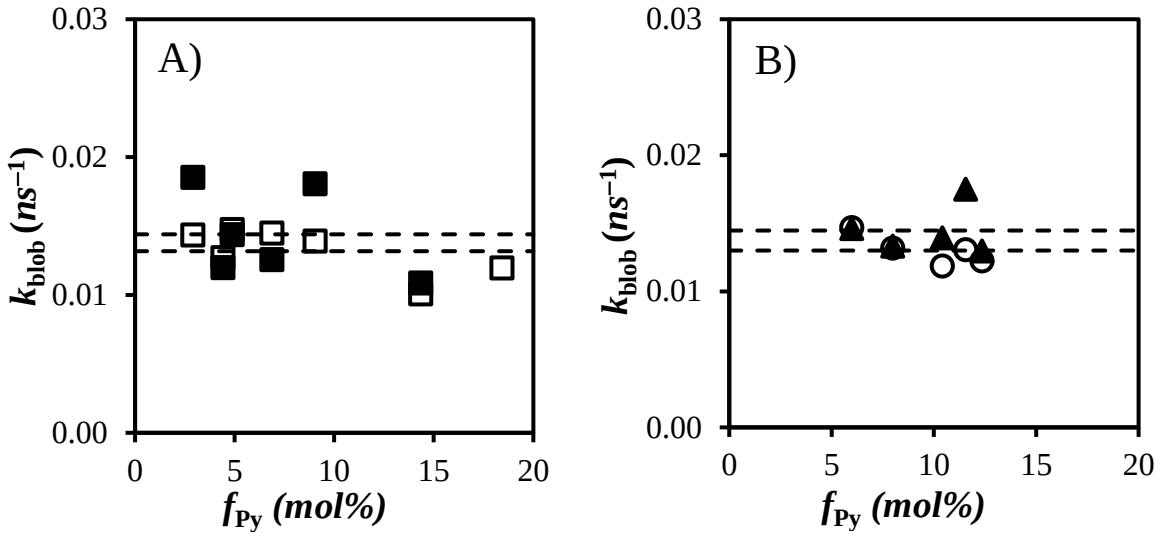


Figure 2.7. k_{blob} as a function of pyrene content for A) Py-PLGlu and B) Py-PDLGlu in DMF (solid) and DMSO (hollow). The dashed lines represent the averages.

Implications and limitations of the FBM analysis: Twenty years of FBM studies have established that an excited dye covalently attached onto a macromolecule probes a volume while it is excited that is referred to as a *blob*.[19] A *blob* is much smaller than the volume defined by the entire macromolecule such as the polymer coil for a linear chain. The FBM takes advantage of this insight by viewing the macromolecular volume as a cluster of *blobs* where dyes can distribute themselves randomly according to a Poisson distribution. The outcome of the present study is that polypeptides can be divided into *blobs* in the same manner as any other linear polymer chain such as poly(alkyl methacrylate)s, polystyrene, poly(*N*-isopropylacrylamide), poly(*N*,*N*-dimethylacrylamide), or polyisoprene, to name but a few.[13,20] Interestingly, the *blobs* found in this study for two series of Py-PGlu constructs have similar size and dynamics as the *foldons* used to study protein folding.

Whether a FBM study could be applied to actual proteins instead of polypeptides brings to the fore a number of experimental challenges. First, the FBM aims to characterize chain dynamics with the implication that the protein be unfolded, a key requirement to study the process of protein folding. Consequently, the study of protein chains dynamics would require that the protein not have secondary and tertiary structure or that a denaturing agent be introduced to avoid the existence of structural motives. Second, the compartmentalization of the polymer coil into *blobs* lends itself nicely to the random labeling of the structural units constituting the chain, the amino acids in the case of polypeptides or proteins. But whereas glutamic acids could be introduced randomly during the synthesis of polypeptides for random attachment of the pyrene labels, this requirement might be more complicated to achieve for an

actual protein whose sequence is well-defined. It must also be pointed out however that the requirement of having a well-defined sequence matters most to establish protein functionality or structure but is not expected to be as demanding to probe the chain dynamics involved in the folding of unfolded proteins. Third, the FBM has been mostly applied to pyrene-labeled macromolecules.[13,20] The use of a same dye (i.e. pyrene) has the advantage of reducing the massive variations encountered in the comparison of the quenching rate constants obtained in the study of end-to-end cyclization of oligopeptides with a variety of dyes having each their own set of photophysical properties.[24] Fourth, the hydrophobicity of pyrene prevents its use in aqueous solution for the study of polymer chain dynamics, an important limitation for the study of phenomena such as protein folding that are meant to take place in water. Consequently, the study of polypeptides with pyrene must be conducted in organic solvents that can solubilize pyrene but which are polar enough to also solubilize the polypeptides. In this context, DMF and DMSO are well-suited for the study of pyrene-labeled polypeptides.

Last, fluorescence quenching studies conducted on short oligopeptides focus on end-to-end cyclization.[13] By contrast, FBM studies deal with actual long chain polymers and probe the main chain dynamics.[13,20] As it turns out, the rate constants (k_{cy}) for end-to-end cyclization of fluorescently labeled oligopeptides are strongly influenced by the nature of the dyes used[24] and the location of the dyes whether at the chain ends or at internal position inside the oligopeptide.[25] Thus comparison of k_{cy} values obtained from different studies is difficult since they can span more than two orders of magnitude for a same oligopeptide depending on the set of dyes used in a given study[24] and k_{cy} has been found to decrease with increasing length of an

overhang added to the oligopeptide.[25] While some recommendable effort has been recently applied to reconcile the widely different k_{cy} values reported in the scientific literature for different dyes and a same oligopeptide,[24] pyrene was not included in this normalization exercise. It is however known that for synthetic polymers, end-labeled chains with a degree of polymerization N yield $k_{cy}\times N$ products that are 5-to-15 times smaller than the product $k_{blob}\times N_{blob}$ for a same polymer.[26,27] Variations in the normalization factor depend on the spacer used to covalently and randomly attach pyrene to the polymer, a longer and more flexible spacer resulting in larger $k_{blob}\times N_{blob}$ products.[28] It is thus presently difficult to compare the $k_{cy}\times N$ and $k_{blob}\times N_{blob}$ products obtained for oligo- and polypeptides using pyrene and other fluorescent dyes. This is an important limitation that will need to be addressed in the future.

2.5 Conclusions

In summary, this study provides strong support to the notion that PEF experiments yield N_{blob} values for folded and partially folded polypeptides that are similar to those of N_{Fold} obtained by hydrogen exchange experiments. Furthermore, k_{blob} was found to provide reliable information about the polypeptide backbone dynamics for P*DL*Glu in DMF and DMSO. Since polypeptide *blobs* and *foldons* appear to share similar length and time scales, we postulate that a *blob* and a *foldon* might be two equivalent quantities. But whereas hydrogen exchange experiments are limited to partially folded proteins, PEF experiments can characterize a *blob* along the entire path of the folding process, from a fully unfolded polypeptide as for Py-P*DL*Glu in DMF and DMSO to partially folded Py-P*L*Glu in DMSO and fully folded Py-P*L*Glu in DMF. Thus PEF, in combination with FBM analysis and MMOs appears to represent a

powerful means to characterize *foldon* size and dynamics for polypeptides and should nicely complement studies conducted with proteins by hydrogen exchange experiments.

Chapter 3

The Effect of Like-Charges on the Conformation and Internal Dynamics of Polypeptides Probed by Pyrene Excimer Fluorescence

Adapted with permission from Casier, R.; Duhamel, J. Effect of Like Charges on the Conformation and Internal Dynamics of Polypeptides Probed by Pyrene Excimer Fluorescence. *Macromolecules* **2020**, *53*, 5147 – 5157.

3.1 Abstract

Three series of pyrene-labeled polypeptides, namely poly(*L*-lysine) (Py-P*L*L), poly(*L*-glutamic acid) (Py-P*L*Glu), and poly(*D,L*-glutamic acid) (Py-P*DL*Glu), were studied in DMSO by monitoring their ability to form excimer between an excited and a ground-state pyrene. The effect that the charges of protonated Py-P*L*L (Py-P*L*L·HCl) and deprotonated Py-P*L*Glu (Py-P*L*GNa) and Py-P*DL*Glu (Py-P*DL*GNa) had on their conformation and dynamics was assessed by monitoring their fluorescence. The fluorescence decays were analyzed according to the fluorescence *blob* model (FBM) to determine N_{blob}, which is the number of structural units in a *blob*, and k_{blob}, which is the rate constant for diffusive encounters between structural units and their side chains inside a *blob*. FBM analysis indicated that the *blob* size for Py-P*L*Glu and Py-P*DL*Glu was unaffected by the presence of anionic charges, yielding N_{blob} values of 10.3 ± 1.7 and 18.2 ± 1.1 glutamic acid units, respectively. These N_{blob} values matched the values found for their uncharged counterparts. Molecular mechanics optimizations (MMOs) were then applied to determine the theoretical N_{blob}^{theo} value that could be obtained if Py-P*L*Glu adopted the conformation of a random coil, a polyproline type II helix, a 3_{10}-helix, or an α-helix. The agreement found between the N_{blob} value of 17.9 ± 1.1 for protonated P*L*Glu and deprotonated P*L*GNa and the N_{blob}^{theo} of 19 found for a 3_{10}-helical conformation suggested that this was the conformation adopted by P*L*Glu in DMSO. Py-P*L*L·HCl was studied in a similar manner, yielding an N_{blob} value of 14.3 ± 1.3 lysines which suggested a coiled conformation according to MMOs. Comparison of k_{blob} between charged and neutral polypeptides demonstrated that the presence of charges slowed the dynamics experienced by the amino

acids. Since the polypeptide *blobs* appeared to have features in terms of their size and dynamics that were similar to those of *foldons*, this study further supports the notion that *blobs* and *foldons* might be identical objects.

3.2 Introduction

Out of the 20 common amino acids, five are typically considered to be charged at physiological pH.[1] They all bear a side chain that is terminated by either a cation for arginine (Arg), histidine (His), and lysine (Lys), or an anion for aspartic acid (Asp) and glutamic acid (Glu). The prevalence of these amino acids in proteins results in the formation of salt bridges between the cations and the anions in the three dimensional structure of nearly every known functional protein.[2] These ionic interactions within proteins have been found to critically affect their biological functions,[3-5] structure and stability,[4,6,7] and ability to fold into their native conformation.[1,8,9] Despite the importance of ionic interactions in proteins, their study remains challenging and clear trends of the effect of charges on the process of protein folding are difficult to find. For example, the presence of a charged glutamic acid is thought to prevent hemoglobin from polymerizing, a process known to otherwise induce red blood cells to adopt a sickle shape which leads to anemia. Moreover, when the negatively charged glutamic acid is replaced by an apolar valine in sickly hemoglobin, this abnormal version of hemoglobin can homopolymerize into filaments since there is no energetic penalty to bury the apolar valine in the apolar interface between two sickle hemoglobin tetramers.[1] However, the importance of charges for proteins folding into their tertiary or quaternary structure should not be overstated

as some proteins such as ribosomal protein S6[9] and ubiquitin[10] can fold into their native conformation even after their charged residues have been removed, with the ubiquitin variant without charges adopting the same native state as wild type ubiquitin but with even increased stability. In addition, some recent computational studies have demonstrated that the inclusion of electrostatic interactions within a protein did not improve structure prediction, though it affected its long-range interactions with charged objects such as other proteins or oligonucleotides.[11]

Certainly, the inclusion of electrostatic forces between charged amino acids to the already complex task of following their displacement in solution as a polypeptide folds according to backbone dynamics creates another layer of complexity to the already complicated problem of protein folding. One approach that has been widely applied to simplify the study of protein folding has been to reduce the complexity associated with having to handle the entire polypeptide through its compartmentalization into smaller subdomains whose much smaller size makes their individual study manageable. The idea of folding subdomains has been conceptualized many times over the years in the form of *foldons*,[12,13] hierarchical segments,[14] and *blobs*.[15,16] The size of these subdomains has been probed in a number of manners. Hydrogen exchange (HX) experiments carried out on partially folded proteins has suggested that *foldons* are about 20 amino acids in size.[12,13] However a review on ionic interactions in proteins supported the notion that proteins fold in a hierarchical manner by determining that a large amount of ionic interactions (salt bridges) in proteins occur among residues less than 11 amino acids away from one another in the primary sequence,[7] thus

suggesting that *foldons* involving charged amino acids might be smaller in size. Whether considering that protein folding occurs through intermediates that take the form of *foldons* or hierarchical segments, these earlier studies looked at proteins in their native or near native conformation. By determining the number of amino acids that can undergo HX with the solvent by NMR or MS, information on the size of these folding subdomains can be obtained, since those amino acids that do not undergo HX are most likely inaccessible to the solvent because they are already involved in a folded motive of the protein. However, what these HX experiments have not uncovered is the underlying phenomena that lead to the specific *foldon* size of about 20 amino acids reported in the scientific literature,[12] which makes it difficult to determine what effect, if any, charges play in the size and dynamics of *foldons*.

The ideal technique for probing the effect that charged amino acids might have on the chain dynamics of polypeptides would have to do so on a length scale of a few tens of amino acids that matches the accepted *foldon* size.[12] As it turns out, the fluorescence *blob* model (FBM) is ideally suited for this purpose.[15-17] The FBM acknowledges that when a fluorophore is covalently attached to a polymer, it probes a volume referred to as a *blob* that is defined by its lifetime and the dynamics of the main chain. Random labeling of a polymer with a dye and its quencher results in a Poisson distribution of the quenchers among the *blobs* which can be taken advantage of to determine the average number $<n>$ of quenchers per *blob*. In turn, combining the known quencher content of the polymer with $<n>$ yields the number N_{blob} of structural units a *blob* is made of. In essence, N_{blob} is a measure of the size of a *blob*. To date, all values found for N_{blob} have been within a few tens of structural units[15–20] which is

comparable to the stated size of a *foldon*.[12,13] Most FBM experiments have been conducted with the dye pyrene due to its exceptional photophysical properties which have been reviewed several times.[16,21,22]

Preliminary FBM experiments on two series of pyrene-labeled poly(*L*-glutamic acid)s (P*L*Glu) and poly(*D,L*-glutamic acid)s (P*DL*Glu) yielded N_{blob} values of 20 and 11 in DMF, where P*L*Glu adopted an α-helical conformation and P*DL*Glu was a random coil, respectively.[23] These experiments also provided the rate constant k_{blob} describing the diffusive motions of the amino acids inside a *blob*, which reflected the dynamics of the polypeptide backbone inside the *blob* subdomains. The number of amino acids constituting the *blobs* of α-helical P*L*Glu were found to match closely the number of amino acids found in partially folded proteins obtained from HX experiments[12,13,24,25] which supported the view that the *blobs* identified within the framework of the FBM were similar, and possibly the same, entities as *foldons* that are considered to study protein folding.[16,23] This earlier study suggested that the FBM could be applied to study the effect that charges have on the size and dynamics of polypeptide *blobs*. To this end, this report describes the FBM study of a series of pyrene-labeled polypeptides in their charged state, namely deprotonated P*L*Glu (Py-P*L*GNa) and P*DL*Glu (Py-P*DL*GNa) and protonated poly(*L*-lysine) (Py-P*L*L·HCl). The effect that charges had on the *blob* size and dynamics as probed by the pyrene labels were determined by comparing the k_{blob} and N_{blob} values obtained from the FBM analysis of the fluorescence decays of the Py-P*L*Glu and Py-P*DL*Glu samples in DMSO under acidic and alkaline conditions. The size and dynamics of the *blobs* of the cationic polypeptide Py-P*L*L·HCl were also characterized. The N_{blob} values

obtained from this study were then used in conjunction with molecular mechanics optimizations to infer the backbone structure of the polypeptides in solution and the effect that charges had on the polypeptide structure and dynamics in DMSO. Consequently, this study characterizes quantitatively the effects that charges have on the conformation and dynamics of charged polypeptide model systems in DMSO and offers new insights on the behavior of these complex macromolecules in solution.

3.3 Experimental

Materials: Acetic acid (glacial, Sigma), butylamine (Aldrich, 99.5 %), *N*,*N*'-dicyclohexylcarbodiimide (DCC, Fluka Analytical, ≥ 99.0 %), *N*,*N*′-diisopropylcarbodiimide (DIC, Sigma, 99 %), *N*,*N*-diisopropylethylamine (DIPEA, Sigma, 99.5 %), *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide hydrochloride (EDC·HCl, Sigma, ≥ 99.0 %), *N*,*N*-dimethylformamide (DMF, Sigma, ≥ 99.8 %), dimethyl sulfoxide (DMSO, Sigma, ≥ 99.9 %), hydrochloric acid (HCl, Fisher, 34 wt% in water), *N*-hydroxysuccinimide (HOSu, Sigma, 98 %), 1-pyreneacetic acid (PyAcOH, Aldrich, 97 %), 1-pyrenemethylamine hydrochloride (PyMA·HCl), and sodium hydroxide (NaOH, Sigma, ≥ 97.0 %) were used as received. Deionized water (DIW) was obtained from a Biopure Series 4400 Single Pass Reverse Osmosis system. The synthesis of all small molecules not listed in the materials section are reported in the Supporting Information (SI).

Pyrene-labeled poly(glutamic acid)s: Pyrene-labeled poly(*L*-glutamic acid) (Py-P*L*Glu, DP = 803, Đ = 1.02) and poly(*D*,*L*-glutamic acid) (Py-P*DL*Glu, DP = 784, Đ = 1.06) were prepared

as reported in a previous study.[23] Both pyrene labeled poly(glutamic acid) samples were lyophilized and stored in their salt form in a – 20 °C freezer before being freshly lyophilized before use.

Pyreneacetic acid N-hydroxysuccinimide ester (2,5-dioxopyrrolidin-1-yl 2-(pyren-1-yl)acetate, PyAcOSu): 1-Pyreneacetic acid (PyAcOH) was activated via the *N*-hydroxysuccinimide ester to facilitate the labeling of poly(*L*-lysine) (P*L*L). In a 250 mL round-bottom flask, PyAcOH (1.00 g, 3.84 mmol, 1.0 eq.) and HOSu (0.71g, 6.2 mmol, 1.6 eq.) were dispersed in 100 mL of dicholoromethane (DCM) and left to stir for 15 minutes to ensure an even dispersion. DCC (1.17 g, 5.7 mmol, 1.5 eq.) was then added to the stirring dispersion, which made the dispersion quickly turn into a clear solution. Over the next hour, a fine white precipitate of dicyclohexylurea (DCU) appeared. The reaction was left to stir overnight. The next day, the reaction mixture was suction filtered (Whatman #1) to remove the precipitated DCU before condensing the solution mixture to ca. 10 mL of DCM with a rotary evaporator. Upon condensing the solution, more DCU precipitate formed which was subsequently removed via suction filtration. The PyAcOSu was then isolated by precipitation into ethanol (ca. 100 mL) to yield a yellow-brown solid (1.34 g, 3.75 mmol, 98% yield). The solid was dried under vacuum overnight before storage at – 20 °C. ^{1}H NMR (300 MHz, d_6-DMSO): δ = 8.35 – 8.07 (m, 9H), 4.90 (s, 2H), 2.79 (s, 4H) ppm.

Pyrene-labeled poly(L-lysine) (Py-PLL): A series of Py-P*L*L samples were prepared from the hydrobromide salt of poly(*L*-lysine) (P*L*L·HBr, Alamanda Polymers, M_n = 165,100 g·mol^{-1}, DP = 790, Đ = 1.06). The preparation of a Py-P*L*L sample where 8 mol% of the lysines were

labeled with pyrene is described in more detail hereafter. P*L*L·HBr (113 mg, 0.54 mmol lysine equivalents) was dissolved in DMSO (15 mL) with stirring over 30 minutes. DIPEA (108 μL, 0.62 mmol, 1.1 eq. per lysine) was then added to the solution which was left to stir until the solution became clear (ca. 1 hr). PyAcOSu (16 mg, 45 μmol, 0.08 eq. per lysine) was added to the clear homogeneous solution which was left to stir overnight. The next day, the crude reaction mixture was transferred to a dialysis tube (Spectrum Labs Spectra/Por 7 dialysis membrane RC tubing, MWCO = 8 kDa). The Py-P*L*L sample was then dialysed against DMF, followed by 70:30 vol% and 30:70 vol% DMF:water mixtures, and finally against water acidified with HCl to pH 6. Each dialysis step used ca. 800 mL of solvent and the dialysis set up was left to equilibrate for 4 hours. HCl was added to ensure that the unlabeled lysine residues were protonated and to remove any remaining HBr. A final dialysis was then conducted against acidified water overnight before the Py-P*L*L sample labeled with a molar fraction *x* of pyrenyl groups was freeze-dried to remove the water and stored in a –20 °C freezer. A similar protocol was followed with the additions of 0.01 to 0.12 eq. of PyAcOSu to prepare a total of six Py-P*L*L samples. The protonated Py-P*L*Ls (Py-P*L*L·HCl) were used in all fluorescence experiments.

UV-Vis absorption: UV-Vis absorption spectra were acquired with 1 nm steps using a Varian Cary 100 Bio spectrophotometer and a 1 cm pathlength UV-Vis cell. A background correction for the solvent was applied to each spectrum.

Dye content of Py-PLL·HCl: The molar fraction (*x*) of lysine residues which were labeled with 1-pyreneacetic acid was calculated using Equation 3.1. In Equation 3.1, λ_{Py} represents the

moles of pyrene per gram of sample and M_{Py} and M_L are the molar masses of a lysine residue bearing a pyrene pendant (M_{Py} = 370.45 g·mol^{-1}) and a lysine residue protonated with HCl (M_L = 164.63 g·mol^{-1}), respectively. λ_{Py} was determined by weighing freshly lyophilized Py-P*L*L·HCl and dissolving it in a known amount of water with sonication. After complete dissolution, a small aliquot (ca. 0.1 mL) of the aqueous Py-P*L*L·HCl solution was diluted with DMF (ca. 5 mL) and the absorbance of the solution, or a dilution thereof, was measured. Beer-Lambert law was applied to determine the pyrene concentration for the Py-P*L*L·HCl solution in DMF based on the molar absorption coefficient (ε_{Py}) of 37,300 M^{-1}·cm^{-1} at 345 nm in DMF for *N*-butyl-1-pyreneacetamide (PyAcNBu) used as a model compound.[26]

$$x = \frac{M_L}{\lambda_{Py}^{-1} + M_L - M_{Py}} \tag{3.1}$$

Steady-state fluorescence (SSF): Steady-state fluorescence spectra were acquired on a Horiba QM-400 spectrofluorometer equipped with a xenon arc lamp. Solutions in DMSO were prepared with ca. 2.7×10^{-6} M pyrene. A 10 μL aliquot of 1 M HCl was added to the Py-P*L*L solutions to ensure that the polypeptides would be in their protonated form (Py-P*L*L·HCl). Py-P*L*Glu and Py-P*DL*Glu were studied in their deprotonated form, namely Py-P*L*GNa and Py-P*DL*GNa, respectively, by adding 100 μL of pH 12 NaOH solution to 4 mL of the polypeptide in a DMSO solution containing ca. 2.7 × 10^{-6} M pyrene. After sonicating the samples for ca. 10 minutes, oxygen was removed by outgassing the samples with a gentle flow of N_2 gas

(Praxair, N4.0 or greater) for ca. 30 minutes before sealing the fluorescence cell and acquiring the SSF spectrum. Using excitation and emission slit widths of 1 nm, the emission of pyrene was monitored by exciting the sample at 344 nm and scanning the emission wavelength from 350 to 600 nm in 1 nm steps using a scanning rate of 10 $nm \cdot s^{-1}$. The fluorescence intensity of the monomer (I_M) and excimer (I_E) were calculated using the area under the SSF spectrum from 375 to 381 nm and from 500 to 530 nm, respectively.

Time-resolved fluorescence (TRF): Samples were prepared in a similar manner as for the SSF measurements. The fluorescence decays of pyrene were acquired using an IBH TC-SPC fluorometer equipped with a NanoLED-340 laser. All decays were collected using an excitation wavelength of 344 nm over 1024 channels. The monomer fluorescence was collected with a time-per-channel of 2.04 $ns \cdot ch^{-1}$ until a total of at least 40,000 counts was obtained at the fluorescence decay maximum at an emission wavelength of 378 nm, using a 370 nm longpass filter to minimize any potential contribution from light scattering. The excimer fluorescence decays were acquired with a decay maximum of at least 20,000 counts at 510 nm with a 470 or 495 nm longpass filter using a time-per-channel of 1.02 $ns \cdot ch^{-1}$. The instrument response function was collected using a LUDOX dispersion in water with a peak maximum of at least 20,000 counts at 344 nm.

Fluorescence decay analysis: The fluorescence decays of the pyrene-labeled polypeptides were analyzed using the fluorescence *blob* model (FBM) which has been the object of a number of reviews.[21,27,28] Within the framework of the FBM, the polypeptide is divided into segments of equal volume, where each unit volume is referred to as a *blob*. Within each *blob*, four distinct

pyrene species are expected to contribute to the overall fluorescence signal. Py_{free} represents the excited pyrenes which do not form excimer during their lifetime and emit with their unquenched lifetime τ_M. Py_{k2} represents the excited pyrenes that are close to a ground-state pyrene and can form excimer through a rapid rearrangement process with the rate constant k_2. Py_{diff} represents the excited pyrenes that first diffuse to a ground-state pyrene with a rate constant k_{blob} before generating the Py_{k2} species. The final species, Py_{agg} represents the pre-associated pyrenes which form excimer directly upon excitation. Py_{agg} includes two excimer species referred to as *E0** and *D** which are assumed to be the result of good and poor stacking between two pyrenyl labels, respectively. Using the framework of the FBM, the parameters retrieved from the global analysis of the monomer and excimer fluorescence decays yielded structural and dynamic information on the pyrene-labeled polypeptides.[20-23] The full expressions of the monomer and excimer decays are given in Equations S3.1 – S3.3. The monomer and excimer fluorescence decays were fitted globally according to Equations S3.1 and S3.2 and the parameters were optimized with the Marquardt-Levenberg algorithm.[29] The decay fits were deemed satisfactory when the χ^2 was smaller than 1.3 and the residuals and autocorrelation of the residuals were randomly distributed around zero. An example of the quality of the fits is provided as Figure S3.4 in SI. A list of all parameters retrieved from the FBM analysis of the fluorescence decays are listed in Tables S3.2 – S3.7. Among these parameters, $<n>$, the average number of ground-state pyrenes inside a *blob*, and k_{blob} are of particular interest. Since the pyrenyl labels are randomly distributed along the macromolecular backbone, $<n>$ can be introduced into Equation 3.2 to calculate N_{blob}, which is the average

number of structural units encompassed within a *blob*. In Equation 3.2, f_{Mfree} is the molar fraction of pyrenyl labels Py_{free} that emit in the monomer decay. k_{blob} provides information about the time scale over which pyrene excimer is being produced, and since the pyrenyl derivatives are covalently attached onto the macromolecule, k_{blob} describes the internal dynamics of the macromolecule of interest.

$$N_{blob} = \frac{< n >}{x}\left(1 - f_{Mfree}\right) \qquad (3.2)$$

3.4 Results

The lysines of P*L*L were labeled with 0.9 to 12.1 mol% of 1-pyreneacetic acid to yield a series of Py-P*L*L constructs. The Py-P*L*Glu and Py-P*DL*Glu series employed in this study had been prepared earlier by reacting 1-pyrenemethylamine with 4.4 to 14.3 mol% of the glutamic acids of P*L*Glu and 6.0 to 12.3 mol% of the glutamic acids of P*DL*Glu.[23] The short linker lengths of 1-pyreneacetic acid and 1-pyrenemethylamine ensured that the results obtained with the pyrene-labeled polypeptides reflected, as closely as possible, those of the unlabeled P*L*L and poly(glutamic acid)s, respectively. The structures of the three pyrene-labeled polypeptides studied are given in Figure 3.1.

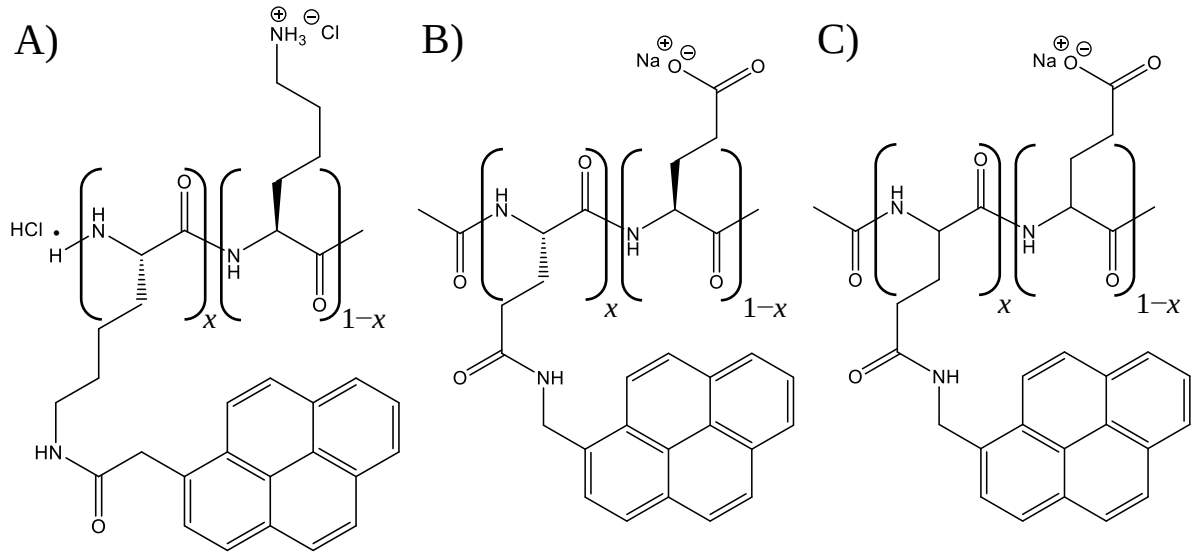


Figure 3.1. Chemical structures of A) Py-P*L*L·HCl, B) Py-P*L*GNa, and C) Py-P*DL*GNa, where *x* represents the molar fraction of residues labeled with pyrene.

Fluorescence spectra of the pyrene-labeled polypeptides: The hydrochloride salts of the Py-P*L*L samples (referred to as Py-P*L*L·HCl from here on) were studied in DMSO, which readily solubilizes both pyrene and the protonated form of P*L*L.[30] To ensure that all the lysine side chains bore an ammonium cation (Figure 3.1A), a small aliquot of $HCl_{(aq.)}$ was added to the DMSO solution as described in the Experimental section. The steady-state fluorescence (SSF) spectra of Py-P*L*L·HCl in DMSO are shown in Figure 3.2A for different pyrene contents. As expected for pyrene-labeled macromolecules, an increase in pyrene content resulted in an increase in fluorescence intensity of the excimer at 480 nm relative to that of the monomer at 378 nm as the local concentration of the pyrenyl pendants along the polypeptide backbone increased. The relative increase in the excimer fluorescence intensity (I_E) to that of the

monomer (I_M) was reflected by the increase in the I_E/I_M ratio with increasing pyrene content as shown in Figure 3.2D.

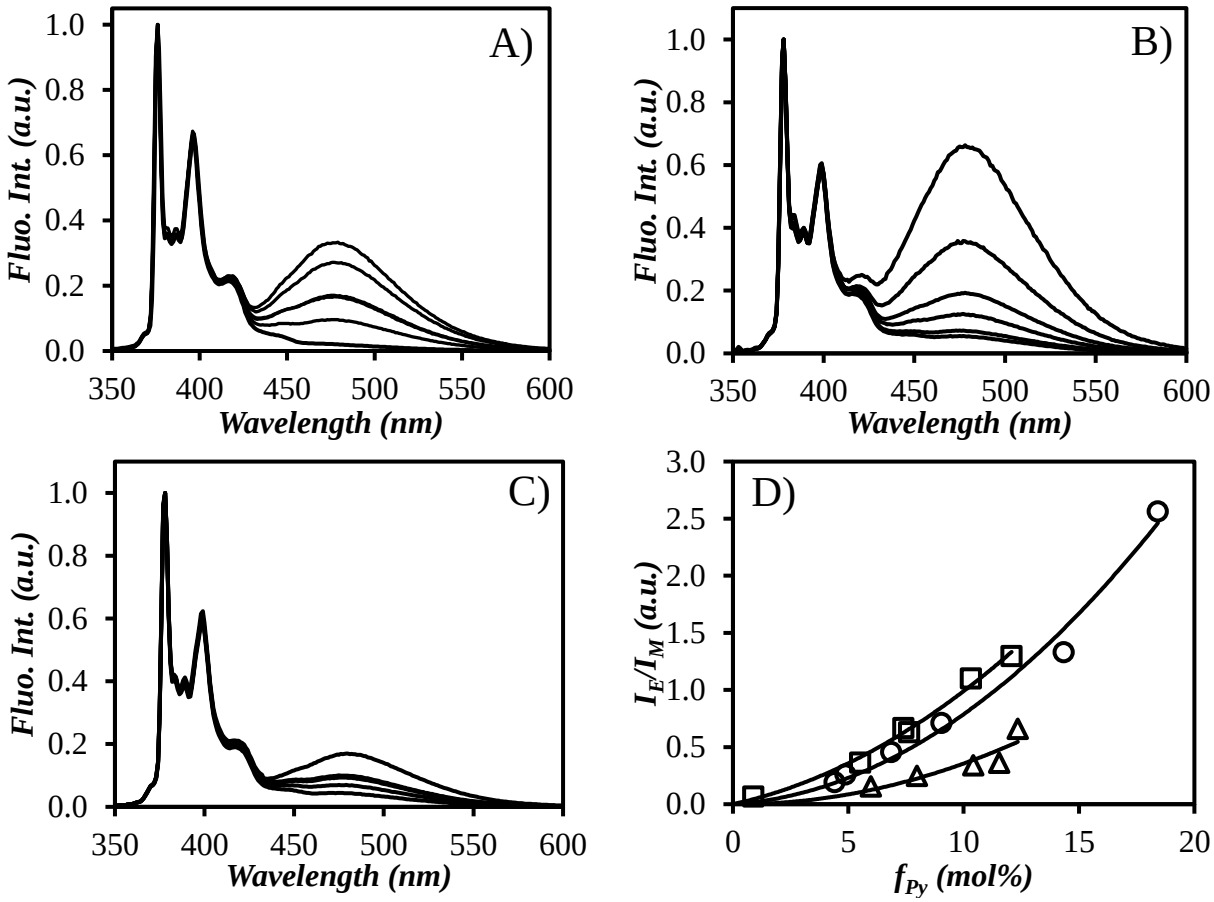


Figure 3.2. Steady-state fluorescence spectra of A) Py-P*L*L·HCl, B) Py-P*L*GNa, and C) Py-P*DL*GNa in DMSO, with D) their resulting I_E/I_M ratios: (□) Py-P*L*L·HCl, (○) Py-P*L*GNa, and (△) Py-P*DL*GNa. The solid lines in D) were added to guide the eye.

The pyrene-labeled poly(glutamic acid)s were previously studied in DMSO in their protonated form.[23] Herein the Py-P*L*GNa and Py-P*DL*GNa samples (the sodium salts of Py-

P*L*Glu and Py-P*DL*Glu, respectively) were also studied in DMSO by adding a small aliquot of $NaOH_{(aq.)}$ to the poly(glutamic acid)s solutions in DMSO. The Py-P*L*GNa and Py-P*DL*GNa constructs offered two examples of charged polypeptides whose behavior could be compared to that of the Py-P*L*L·HCl samples. However, since the unprotonated lysines quenched the fluorescence of pyrene, Py-P*L*L could not be studied by fluorescence in DMSO thus preventing the comparison of the solution behavior of Py-P*L*L·HCl with that of Py-P*L*L. Fortunately, the poly(glutamic acid)s did not quench the fluorescence of the pyrene labels regardless of whether they were protonated or not, so that the behavior of the poly(glutamic acid)s could be compared in their protonated or deprotonated form.

Trends similar to that of Py-P*L*L·HCl were also observed with the SSF spectra of both Py-P*L*GNa and Py-P*DL*GNa in, respectively, Figures 3.2B and C, where the excimer intensity increased with respect to that of the monomer with increasing pyrene content. In Figure 3.2D, Py-P*L*GNa and Py-P*L*L·HCl were found to have nearly overlapping I_E/I_M ratios for any given pyrene content, but their I_E/I_M ratios were significantly larger than the I_E/I_M ratio of Py-P*DL*GNa. The higher I_E/I_M ratios of Py-P*L*GNa compared to that of Py-P*DL*GNa was unexpected. The electrostatic repulsion induced by the carboxylate anions along the backbone was expected to lead to a coiled conformation for both polyglutamate salts in DMSO regardless of the side chain stereochemistry, as is typically observed in alkaline aqueous solutions.[31-33] If Py-P*L*GNa and Py-P*DL*GNa had a similar random coil conformation in DMSO, they should exhibit similar I_E/I_M ratios. This was not observed in Figure 3.2D. The increase in I_E/I_M from

Py-P*DL*GNa to Py-P*L*GNa reflected a denser conformation of the latter polypeptide in DMSO suggesting that it might retain some structure even in its deprotonated state.

The I_E/I_M ratios obtained for Py-P*L*L·HCl and the polyglutamate salts were a bit more difficult to compare because they involved different side chain lengths and pyrene derivatives which are known to affect the I_E/I_M ratios differently. The I_E/I_M ratio is proportional to the product $k_{diff}\times[Py]_{loc}$ where k_{diff} is the rate constant for pyrene excimer formation by diffusive encounters and $[Py]_{loc}$ is the local concentration of ground-state pyrenes experienced by an excited pyrene.[34,35] k_{diff} depends on the dynamics of the polymer backbone and the side chain to which pyrene is attached[16,21,27,28] and the probability (p) for an excited and ground-state pyrene to form excimer upon encounter.[36] For a same pyrene derivative, a longer linker between pyrene and the main chain has been shown to enhance pyrene mobility resulting in a larger k_{diff} and enhanced excimer formation.[34] The probability p depends on the nature of the pyrene substituent used to link pyrene to the polymer. For pyrene-labeled polymers, $[Py]_{loc}$ depends on the pyrene density along the polymer backbone which is related to the length of the linker connecting pyrene to the polymer, a longer linker resulting in a lower $[Py]_{loc}$, and the conformation of the polymer, a more compact conformation resulting in a higher $[Py]_{loc}$. Since k_{diff}, p, and $[Py]_{loc}$ are impossible to assess independently by SSF, time-resolved fluorescence measurements were conducted as they separate the effect of chain dynamics from $[Py]_{loc}$. To this end, the FBM was applied to the analysis of the fluorescence decays of the pyrene-labeled polypeptides in DMSO.

Characterization of polypeptide blob size: The size of the subdomains or *blobs*, where the pyrene-labeled side chains of the polypeptides interacted with one another, was characterized in DMSO by globally analyzing their monomer and excimer fluorescence decays with the FBM to yield N_{blob}, the average number of amino acids encompassed within a *blob*, from Equation 3.2. Figure 3.3 is a plot of N_{blob} as a function of pyrene content for Py-P*L*L·HCl, Py-P*L*GNa, and Py-P*DL*GNa in DMSO. Within experimental error, N_{blob} remained constant for all pyrene contents with respect to each polypeptide. The N_{blob} value was lowest for Py-P*DL*GNa, with an average value of 10.3 ± 1.7, and highest for Py-P*L*GNa with an N_{blob} value of 18.2 ± 1.1, with Py-P*L*L·HCl taking an intermediate N_{blob} value of 14.3 ± 1.3. The higher N_{blob} value of Py-P*L*GNa compared to that of Py-P*DL*GNa is consistent with the higher I_E/I_M ratios obtained by SSF, which confirms that the backbone of Py-P*L*GNa is more compact than that of Py-P*DL*GNa. This is typically observed when the backbone adopts a structured conformation, such as that of an α-helix.[23]

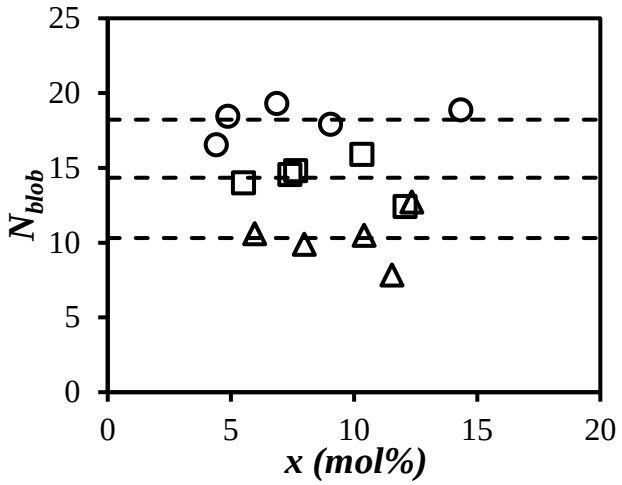


Figure 3.3. Plot of N_{blob} as a function of pyrene content for (□) Py-P*L*L, (○) Py-P*L*GNa, and (△) Py-P*DL*GNa. The dashed lines represent the averages of the experimental results.

Conformation of poly(L-lysine): The structure of Py-P*L*L·HCl in DMSO was examined first by conducting molecular mechanics optimizations (MMOs) to predict what N_{blob} should be if the Py-P*L*L·HCl constructs adopted a helical or coiled conformation. To this end, the maximal overlap experienced between two pyrenyl labels attached onto a P*L*L backbone was monitored as a function of the number of lysine residues separating one reference pyrenyl group from a second one which was allowed to move along the P*L*L backbone one lysine at a time. The extent of overlap between two pyrenyl labels was quantified by counting the number of carbon atoms of one pyrene overlapping the frame of the other pyrene. A carbon-carbon overlap between two pyrenes greater than or equal to seven was considered to result in a good pyrene-pyrene overlap conducive of excimer formation and was included in the calculation of the theoretical N_{blob} value (N_{blob}^{theo}). MMOs were conducted for both an α-helical and a randomly

coiled *PL*L. Since the α-helical P*L*L had set backbone positions, MMOs were only conducted for one position of a reference pyrenyl label. In contrast, the number of lysines yielding a good pyrene-pyrene overlap from the MMOs of the coiled P*L*L were averaged over three different positions of the reference pyrene along the P*L*L backbone to assess the effect of the reference pyrene position on the N_{blob}^{theo} value. A more detailed description of the protocol applied to conduct the MMOs is given in Supporting Information (SI).

Figure 3.4A displays the results of the MMOs for both the α-helical and coiled P*L*Ls and Figures 3.4B and C display polypeptide constructs illustrating good and poor overlap between pyrene residues. Figure 3.4A shows that a pyrenyl group is expected to only form excimer with another pyrenyl attached to residues located up to 6.7 ± 0.6 lysines away from the reference lysine when P*L*L is in a coiled conformation. This number of residues, located on one side of the reference residue, for which pyrene excimer formation (PEF) occurs is referred to as N_o (see Figure 3.4D). N_o for the coiled P*L*L was significantly smaller than for the α-helical P*L*L for which MMOs predict that PEF can occur between a pyrene attached onto a lysine located up to 15 residues away from the reference pyrene. However, the helical twist of the P*L*L backbone results in a dip in the C-C overlap for residues #13 and 14 in Figure 3.4A when the lysine side chains are placed on opposite sides of the helix. Since the overlap is less than seven for these 2 residues, they were excluded from the N_o calculations for the helical P*L*L so that N_o equaled 13 for the α-helical P*L*L.

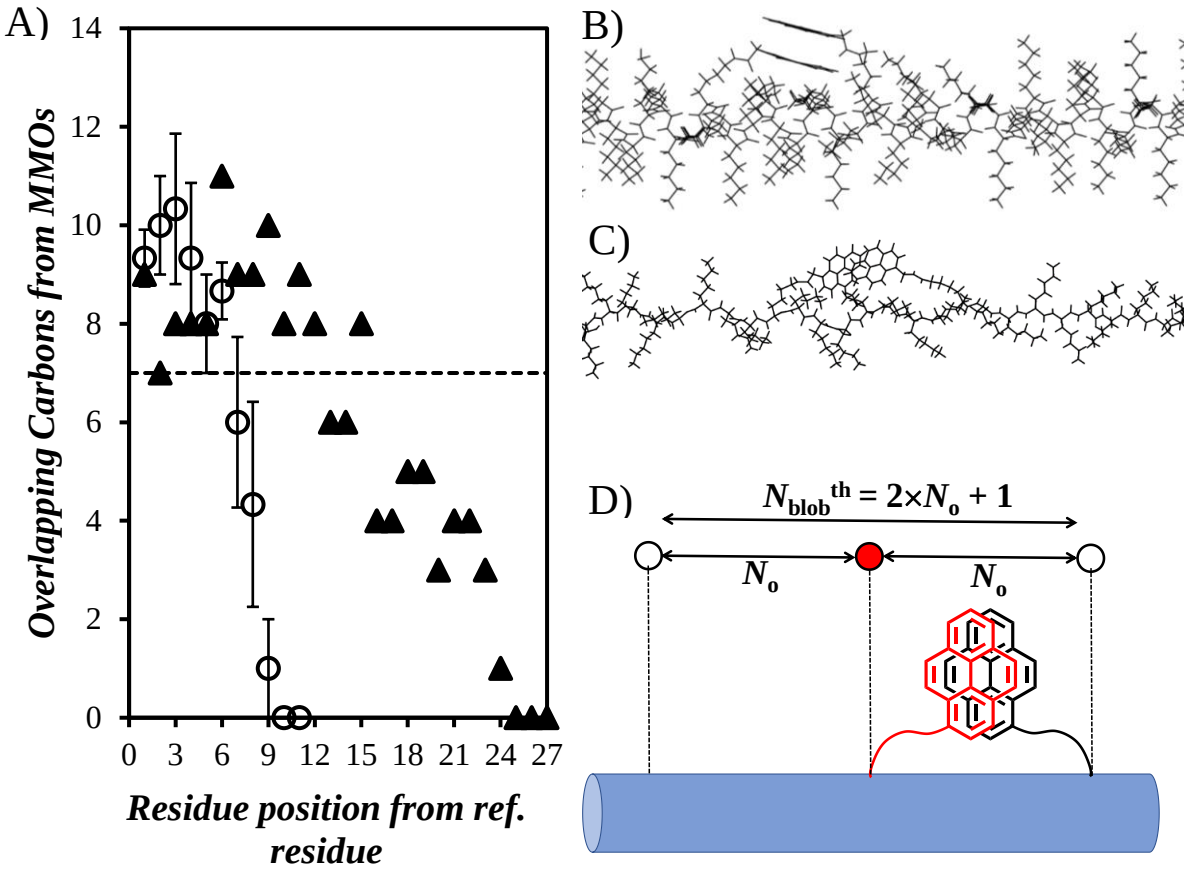


Figure 3.4. A) Plot of the maximum carbon-carbon overlap between two pyrene pendants as a function of the number of lysines between two pyrene labels for a P*L*L chain in an α-helical (▲) and coiled (O) conformation from MMOs. The dashed lines correspond to a carbon-carbon overlap of 7 marking the boundary between poor and good overlap between pyrenes. Example illustrations of B) good overlap between two pyrenes separated by 11 residues when P*L*L adopts a helical conformation and C) poor overlap between two pyrenes separated by 8 residues when P*L*L adopts a coiled conformation. D) Schematic representation of the N_{blob}^{theo} calculation based on the N_o value determined from MMOs with the reference pyrene in red and the secondary pyrene in black.

Given that the P*L*L side chains can form excimer on either side of the reference residue, the expected N_{blob}^{theo} value was calculated as $2\times N_o+1$ (see Figure 3.4D), where one amino acid was added to the product $2\times N_o$ to account for the reference residue. Since the MMOs yielded N_o values of 6.7 ± 0.6 and 13 for the coiled and α-helical Py-P*L*L·HCl, N_{blob}^{theo} equaled 14.3 ± 1.2 and 27 for the coiled and α-helical P*L*L, respectively. As it turns out, the experimental N_{blob} value of 14.3 ± 1.3 for Py-P*L*L·HCl in DMSO matched precisely the N_{blob}^{theo} value expected for Py-P*L*L·HCl in a coiled conformation, indicating that this was the conformation that P*L*L adopted in DMSO. This result agrees with previous studies that concluded that P*L*L adopted a coiled conformation in water under acidic conditions.[37]

Conformation of the poly(glutamic acid)s: It was determined in a previous study that the protonated Py-P*L*Glu and Py-P*DL*Glu constructs in DMSO yielded experimental N_{blob} values of 17.7 ± 1.3 and 10.5 ± 1.7, respectively.[23] These experimental N_{blob} values were attributed to a partially unfolded helix and a random coil, respectively, based on the N_{blob}^{theo} values found by MMOs to equal 23 for an α-helical P*L*Glu and 10 ± 1 for a randomly coiled P*L*Glu.[23] Figure 3.5A compares the results obtained in this study for Py-P*L*GNa and Py-P*DL*GNa with those of the uncharged Py-P*L*Glu and Py-P*DL*Glu. The experimental N_{blob} value of 10.3 ± 1.7 found in this study for the Py-P*DL*GNa samples in DMSO matched both the experimental N_{blob} value of 10.5 ± 1.7 for Py-P*DL*Glu in DMSO and N_{blob}^{theo} of 10 ± 1 obtained by MMOs for a random Py-P*L*Glu coil, indicating that both the protonated P*DL*Glu and deprotonated P*DL*GNa samples adopted a coiled conformation in DMSO. Consequently, the similar N_{blob} values of

the P*DL*Glu samples under either acidic or alkaline conditions implied that the side chains were able to probe the same volume whether or not the side chains were charged.

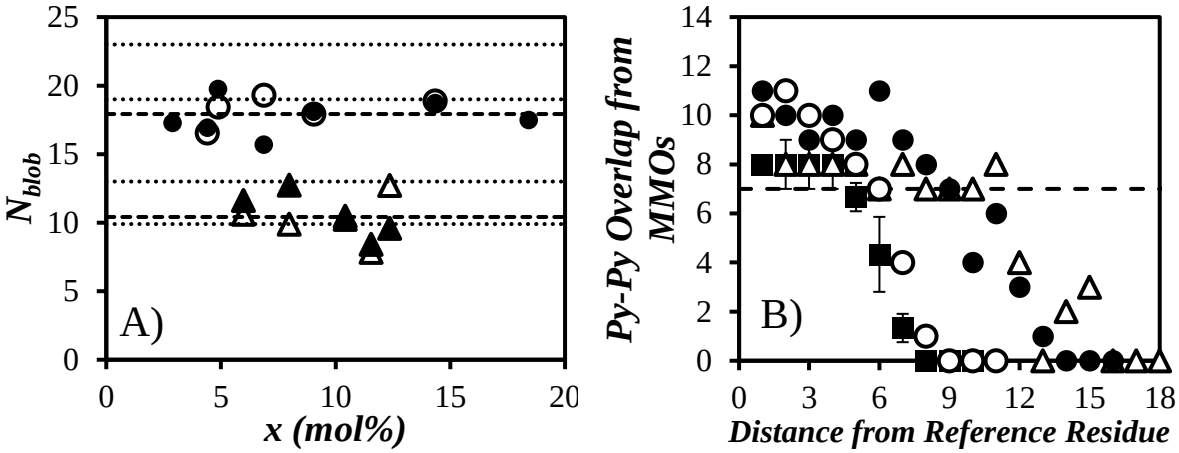


Figure 3.5. A) Plot of N_{blob} as a function of pyrene content for (circles) Py-P*L*Glu and (triangles) Py-P*DL*Glu in the (hollow) sodium salt and (filled) neutral form in DMSO. The dashed lines represent the averages of the experimental results. The dotted lines represent the N_{blob}^{theo} values predicted from MMOs for a PGlu with from top to bottom: (N_{blob} = 23) α-helix, (N_{blob} = 19) 3_{10}-helix, (N_{blob} = 13) PPII-helix, and (N_{blob} = 10) random coil. B) The overlap between two pyrene pendants as a function of distance between pyrene labels for a P*L*Glu chain in a (△) α-helix, (●) 3_{10}-helix, (○) PPII-helix, and (■) random coil conformation from MMOs. The dashed line corresponds to a pyrene-pyrene overlap of 7 carbons marking the boundary between poor and good overlap between pyrenes, yielding an N_o value of 9 and an $N_{blob}^{theo} = 2 \times N_o + 1 = 19$.

The smaller N_{blob} value of 10.3 ± 1.7 for the coiled Py-P*DL*GNa compared to the N_{blob} value of 14.3 ± 1.3 for the coiled Py-P*L*L·HCl in DMSO was due to the longer reach of the 1-pyreneacetyl derivative attached to lysine, which provided a larger local volume where the pyrene-labeled residues could interact with one another. This preliminary result indicates that the *blob* subvolume within a randomly coiled polypeptide increases with increasing amino acid side chain length as would be expected.[34] The increase in the *blob* size as a function of the length of the linker connecting pyrene to the polypeptide backbone will be studied in more depth in the future to further explore this concept.

The fact that Py-P*L*GNa in DMSO had a relatively large N_{blob} value of 18.2 ± 1.1, which was similar to the N_{blob} value of 17.7 ± 1.3 found for Py-P*L*Glu in DMSO,[23] was surprising for three reasons. First, DMSO is traditionally viewed as a solvent that denatures structural segments,[38-42] so that N_{blob} should be closer to 10, contrary to the N_{blob} values of 17.9 ± 1.1 found for the Py-P*L*Glu and Py-P*L*GNa samples in DMSO. Second, the similar N_{blob} values of about 18 found for both Py-P*L*Glu and Py-P*L*GNa samples suggested that P*L*Glu adopted a secondary structure in DMSO and that this structure remained unchanged regardless of whether the side chains were protonated or not. Third, the N_{blob} value of 17.9 ± 1.1 found for Py-P*L*Glu and Py-P*L*GNa in DMSO was larger than the N_{blob}^{theo} value of 10 for a P*DL*Glu coil, but smaller than that of 23 for an α-helix.[23] Consequently, if Py-P*L*Glu or P*L*GNa adopted a secondary structure in DMSO, the N_{blob} value corresponding to that structure was intermediate between that of a random coil and an α-helical conformation. As it turns out, these results agree with the conclusions drawn from a recent study on the structure of P*L*Glu in DMSO.[43]

The study concluded that P*L*Glu adopted an uninterrupted, rigid, rod-like structure, such as that of an α-helix or similar helical structure in DMSO. Based on the experimental N_{blob} values, the structure of Py-P*L*Glu in DMSO had to be less dense than that of an α-helix with a backbone having fewer residues-per-turn. It suggested that a 3_{10}-helix or poly(proline) type II (PPII) helix would be much likelier candidates for the structure of P*L*Glu in DMSO, since both have only 3 residues per turn compared to 3.6 for an α-helix. Furthermore, previous studies have indicated that the 3_{10}-helix is a conformational intermediate between an α-helix and a random coil,[44] and that DMSO can induce the formation of stable 3_{10}-helices[45] and PPII-helices[40,42] in some peptides.

Based on this insight, MMOs were repeated on Py-P*L*Glu constructs where the P*L*Glu backbone adopted a 3_{10} or a PPII-helical conformation. The numbers of carbon atoms between two overlapping pyrenes were plotted in Figure 3.5B as a function of the number of residues separating the reference pyrene from the secondary pyrene as it was attached sequentially to neighboring glutamic acids. The plot in Figure 3.5B yielded an N_o value of 9 for the 3_{10}-helix conformation which implied an N_{blob}^{theo} value of 19. The PPII-helix had a significantly lower N_o of 6 leading to an N_{blob}^{theo} value of 13. Figure 3.5A shows that the experimental N_{blob} value of 17.9 ± 1.1 for Py-P*L*Glu and Py-P*L*GNa agrees well with the N_{blob}^{theo} value of 19 found by MMOs for a 3_{10}-helix conformation, indicating that this is the likeliest structure of P*L*Glu and P*L*GNa in DMSO. It is important to note at this point that this study represents the first example where the combination of PEF, FBM, and MMOs is applied to predict the as yet unknown structure of a macromolecule in solution.

The existence of structured Py-P*L*GNa in DMSO while Py-P*L*L·HCl was structureless was interesting, but somewhat surprising given the chemical similarities between the two polypeptides. Possible reasons for these differences will be presented later in the Discussion section.

Dynamics of oligopeptide segments inside blobs: In addition to backbone structure, the FBM also provides quantitative information about the dynamics experienced by the pyrene labels inside the *blobs*. This information was obtained through comparison of the k_{blob} values where k_{blob} is the rate constant for the diffusive encounters between structural units bearing a pyrene derivative within a *blob* that contains only one excited and one ground-state pyrene label. In other words, k_{blob} describes the rate of encounter between amino acids within the *blob* sub-volume. To quantify the effect of the side chain charges on polypeptide chain dynamics, Figure 3.6A compares the k_{blob} values obtained between the poly(glutamic acid)s and their corresponding salts. The k_{blob} values of Py-P*L*Glu and Py-P*DL*Glu overlapped with an average value of 13.1 ± 1.4 μs^{-1}, while the k_{blob} values of Py-P*L*GNa and Py-P*DL*GNa overlapped at a significantly lower averaged k_{blob} value of 8.3 ± 1.0 μs^{-1}. The overlapping k_{blob} values of the *L* and *DL* configurations of the poly(glutamic acid)s in either their protonated or deprotonated state indicated that the dynamics within a *blob* remained constant irrespective of the conformation of the polypeptide backbone and suggested that the rigid backbone limited the volume of a *blob*.

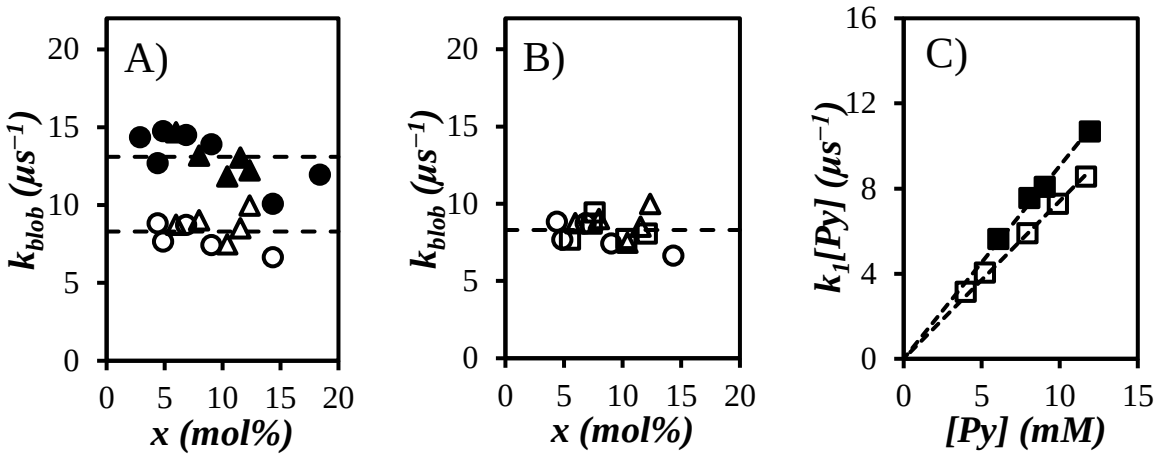


Figure 3.6. Plots of k_{blob} of A) (circles) Py-P*L*Glu and (triangles) Py-P*DL*Glu in the (hollow) sodium salt and (filled) neutral form and B) (□) Py-P*L*L·HCl, (○) Py-P*L*GNa, and (△) Py-P*DL*GNa in DMSO. The dashed lines represent the averages. C) Plot of $k_1[Py]$ as a function of concentration for (■) PyMAAc and (□) PyAcNBu in DMSO. The dashed lines represent the linear fits with slopes of 0.91 ± 0.01 and 0.74 ± 0.01 ns^{-1} for PyMAAc and PyAcNBu, respectively.

Since the only difference between the PGlu and PGNa samples was the solution pH, the decrease in k_{blob} from the protonated to the deprotonated PGlu constructs was expected to be solely due to the presence of charges on the glutamate side chains. Under acidic conditions, the side chains of the protonated poly(glutamic acid)s were free to interact with one another inside the *blobs,* and the dynamics were determined by the inherent mobility of the side chains. An increase in pH deprotonated the glutamic acid side chains thus resulting in charged side chains. The presence of these same charges along the P*L*GNa backbone now added an additional electrostatic repulsive force between the side chains which appeared to slow their

dynamics to the same extent for both conformations of the deprotonated poly(glutamic acid)s. It is important to note that although the repulsion between the side chains was enough to slow their dynamics, it did not reduce the volume that they could probe, since N_{blob} remained unchanged for a same conformation of the protonated or deprotonated poly(glutamic acid). The reduction in the dynamics of the side chains induced by the presence of charges agrees with the experimentally observed acceleration of the folding rate of the highly charged protein S6 upon removal of its charges.[9]

The k_{blob} values of Py-P*L*L·HCl were then compared to those of the Py-P*L*GNa and Py-P*DL*GNa constructs in Figure 3.6B. The presence of charges on both the Py-P*L*L·HCl and Py-PGNa constructs should normally allow for a more direct comparison between the two polypeptides. As a matter of fact, the average k_{blob} value of $8.3 \pm 0.7\ \mu s^{-1}$ found for Py-P*L*L·HCl in DMSO matched perfectly the average k_{blob} value of $8.3 \pm 1.0\ \mu s^{-1}$ determined for Py-P*L*GNa and Py-P*DL*GNa. Unfortunately, and despite the fact that both polypeptides bore charged side chains, the k_{blob} values could not be compared directly due to the difference in pyrene derivative used to label the side chains and the side chain length separating the pyrenyl label from the backbone.[34]

The first factor that may have influenced the dynamics perceived through the k_{blob} values shown in Figure 3.6B was the difference in pyrene derivatives which were used to label the polypeptides. It is known that different pyrene derivatives can have different photophysical constants[34,46] and in the case of dynamics, the probability of quenching (p) upon encounter between an excited and a ground-state pyrene becomes important.[36] To this end, p was

evaluated for two model compounds, namely *N*-butyl-1-pyreneacetamide (PyAcNBu) and 1-pyrenemethyl acetamide (PyMAAc), that closely resembled the pyrene derivatives used to prepare the Py-P*L*L·HCl and Py-PGNa constructs, respectively. The full description and discussion of this study is given in the SI, with the final results plotted in Figure 3.6C. In Figure 3.6C, k_1 is the rate constant of excimer formation derived from the Birks scheme[46] for PyAcNBu and PyMAAc, and [*Py*] is the concentration of the compounds in solution. Plotting the product k_1[*Py*] obtained from the global Birks scheme analysis of the monomer and excimer fluorescence decays against [*Py*] in Figure 3.6C yields straight lines whose slopes depend on *p* and the hydrodynamic radius of the molecules. Since both model compounds have similar sizes, the slopes of 0.91 ± 0.01 and 0.74 ± 0.01 ns^{-1} for, respectively, PyMAAc and PyAcNBu in Figure 3.6C suggest that *p* would be about 20% smaller for Py-P*L*L·HCl than for the Py-PGNa constructs. But since k_{blob} would be expected to increase by about 20% upon increasing the length of the spacer linking pyrene to the backbone from 5 atoms for the Py-PGlu constructs to 7 atoms for the Py-P*L*L constructs,[34] both effects cancel each other resulting in a same k_{blob} value for the Py-P*L*L·HCl and the Py-PGNa constructs in Figure 3.6B.

Thus, k_{blob} for Py-P*L*L·HCl would be expected to be about 20% larger or equal to 10 μs^{-1} if Py-P*L*L·HCl had been prepared with a pyrene derivative having a structure similar to 1-pyrenemethylamine used to label the poly(glutamic acid)s. Similarly, if k_{blob} for neutral Py-P*L*L could have been measured by preventing quenching of pyrene by the P*L*L amines, it would have been equal to 15.7 μs^{-1} after accounting for a 20% increase due to the longer linker between pyrene and the P*L*L backbone compared to k_{blob} for Py-PGlu. Here again, protonation

of the P*L*L side chains would result in a 60% decrease in side chain mobility. In turn, the 60% slowing of side chain mobility observed in this study for the poly(glutamic acid)s and P*L*L when dealing with charged residues in polypeptides would be expected to contribute to a lengthening of protein folding time as has been reported when charged residues are incorporated into the sequence of a protein.[9,10]

3.5 Discussion

The selection of DMSO as a solvent to study polypeptide chain dynamics was dictated by the hydrophobicity of pyrene which prevents the use of water, where pyrenyl labels would aggregate and lose their ability to probe polymer chain dynamics.[22] As DMSO is commonly used as a denaturing agent for proteins,[38-42] it was expected to solvate and denature the polypeptides investigated in this study, thus enabling the study of their chain dynamics. Yet, while DMSO solvated all polypeptides considered in this report and denatured the Py-P*L*L·HCl constructs, it did not denature the Py-P*L*Glu and Py-P*L*GNa samples. Furthermore, not only did the negatively charged Py-P*L*GNa yield the same N_{blob} value of 18 residues as Py-P*L*Glu, suggesting that P*L*GNa displayed the same type of structure in DMSO as P*L*Glu, but negatively charged Py-P*L*GNa retained its structure in DMSO while positively charged Py-P*L*L·HCl did not.

As it turned out, these apparently contradicting observations are not isolated and the view that DMSO always denatures polypeptides is being challenged.[42,43,45,47] In fact, it has been suggested that the α-helices found in proteins or polypeptides dissolved in DMSO might

adopt a 3_{10} or a poly(proline) type II (PPII) helical conformation.[42,43,45] A combination of PEF, FBM, and MMOs enabled to predict that the conformation of P*L*Glu and P*L*GNa in DMSO can only be that of a 3_{10}-helix and not that of a random coil, a PPII-helix, or an α-helix. Current techniques capable of structure predictions for macromolecules in solution are mainly spectroscopic techniques such as NMR,[39,42,43,45] circular dichroism,[39,43] or Fourier transform infrared (FTIR).[43,47] These techniques typically probe the structural units of macromolecules over short sub-nanometer distances. Changes in the spectral signature obtained from the structural units lead to conclusions about their local environment and enables one to predict the conformation of a given macromolecule in solution. In contrast, PEF probes distances along a macromolecule that can be as short as 3.4 Å (for two stacked and overlapping pyrenes leading to PEF) and as long as several nanometers depending on the length of spacer used to link a pyrene derivative to the polymer backbone.[34] It thus provides an alternative method that probes macromolecules over a more extended length scale and which is expected to nicely complement the spectroscopic techniques currently used to probe the conformation of macromolecules in solution.

The difference in conformation between the randomly coiled P*L*L·HCl and the 3_{10}-helical PGNa was attributed to the dual properties of DMSO. First, the dipole moment of 3.96 D for DMSO is much larger than that of 1.85 D for water,[48] which explains why DMSO forms strong hydrogen bonds with amine and amide protons.[49] The ability of DMSO at denaturing structured polypeptides has been attributed to the latter interactions, as they efficiently disrupt the H-bond network usually involved in the stabilization of secondary structures. In the case

of P*L*L·HCl, both interactions are important with the former and the latter leading to H-bonding between DMSO and the protons of the side chain amines and the backbone amides of P*L*L·HCl, respectively. Second, the dielectric constant of 47 for DMSO is about half that of 80 for water,[48] making DMSO a much less polar solvent. DMSO is believed to present its two methyl groups to solvate the butyl side chain of P*L*L,[42] further facilitating the solvation of P*L*L·HCl and its access to the amide protons of the P*L*L·HCl backbone, thus enabling the unfolding of P*L*L·HCl into a random coil.

Compared to P*L*L·HCl, the chemical structure of PGNa is much less favorable towards interactions with DMSO. First, the carboxylate anions terminating the side chains of P*L*GNa cannot interact with DMSO. Second, the ethylene portion of the P*L*GNa side chains is much shorter than the butyl portion of the P*L*L·HCl side chains, thus reducing the level of interactions with the methyl groups of DMSO. Together, the weaker interactions between DMSO and P*L*GNa must be hindering the access of DMSO to the backbone amide protons, resulting in the P*L*GNa retaining its structure in DMSO. The ability of side chains to sterically block DMSO from interacting with the amide backbone of peptides is a mechanism which has been proposed for the observed stability of some β–structures in DMSO.[32]

The apparent lack of effect found in this study induced by the deprotonation of P*L*Glu into P*L*GNa on the secondary structure of this polypeptide in DMSO can also be related to a similar absence of effect induced by the deprotonation of a hexa(*L*-aspartic acid) (H*L*AA) and a hexa(*L*-glutamic acid) (H*L*Glu) on the end-to-end distance of these end-tagged oligopeptides probed by fluorescence resonance energy transfer (FRET).[50] Contrary to similar hexapeptides

prepared with basic amino acids such as lysine, arginine, and histidine whose end-to-end distance increased upon decreasing the solution pH, the end-to-end distance of H*L*AA and H*L*Glu remained constant when the solution pH was increased from 1 to 12.[50] This study illustrating the different behaviors experienced by positively and negatively charged oligopeptides represents another example of the poorly understood effect that charges have on the conformation of polypeptides in solution.

The stability of the P*L*Glu 3_{10}-helix in DMSO regardless of whether the carboxylic acids were protonated or not offered a unique opportunity for probing the effect of charges on side chain mobility. It is well-established that the deprotonation of P*L*Glu in water induces a conformational change from an α-helix to a random coil.[31-33] A similar α-helix-to-random coil conformational change takes place upon protonation in water of the side chain amines in P*L*L.[51] These conformational changes complicate the assessment of the effect that charged side chains have on their mobility since charging the polypeptide is accompanied by a conformational change. Our PEF measurements demonstrate that such complications are eliminated for P*L*Glu and P*DL*Glu in DMSO, where they retain, respectively, their 3_{10}-helical and random coil conformation regardless of their charged state. Furthermore, PEF established that the sole origin of the 60% slower dynamics experienced by the charged side chains of deprotonated P*L*GNa and P*DL*GNa compared to protonated P*L*Glu and P*DL*Glu, was the presence of charges generated along the backbone. This conclusion was also supported by the results obtained with Py-P*L*L·HCl by accounting quantitatively for the difference with the Py-P*DL*Glu constructs in terms of spacer length of 7 versus 5 atoms separating the pyrene derivative from the

polypeptide backbone (see Figure 3.1) and the 20% difference in PEF efficiency between the 1-pyrenemethyl amine and 1-pyreneacetic acid derivatives (see Figure 3.6C). The reduction in side chain mobility observed for these model polypeptides upon adding charges to the polymers was also in agreement with literature reports that have found a reduction in folding time when changing from neutral to charged variants of a same protein.[9,10] The existence of charges along the polypeptide backbones appeared to affect side chain dynamics in a same manner, since the same reduction in k_{blob} values was found upon charging the neutral P*L*Glu, P*DL*Glu, and P*L*L samples into the charged P*L*GNa, P*DL*GNa, and P*L*L·HCl samples.

This study added the N_{blob} value of 14.3 ± 1.3 for the P*L*L·HCl random coil and 17.9 ± 1.1 for 3_{10}-helical Py-P*L*Glu and Py-P*L*GNa in DMSO to that of 21 ± 1 for α-helical Py-P*L*Glu in DMF and 10 ± 3 for the Py-P*DL*Glu random coil in both DMF and DMSO.[23] Figure 3.7 compares all the experimental N_{blob} values reported thus far in the literature for polypeptides with the N_{blob}^{theo} values corresponding to specific polypeptide conformations. Figure 3.7 clearly illustrates how the experimental N_{blob} values agglomerate along the diagonal and generate small clusters corresponding to specific polypeptide conformations associated with a distinct N_{blob}^{theo} value. In fact, Figure 3.7 suggests that the combination of PEF, FBM, and MMOs provides a new experimental means to probe the conformation of macromolecules in solution. Whereas it is well known that P*L*Glu is an α-helix in DMF and P*L*L·HCl in DMSO, P*DL*Glu in DMF and DMSO, and P*DL*GNa in DMSO are random coils, the conformation of P*L*Glu and P*L*GNa in DMSO remained unknown at the time of this study. The ability of Figure

3.7 to help predict the conformation of a polypeptide in solution was taken advantage of to propose that P*L*Glu and P*L*GNa in DMSO adopt a 3_{10}-helical conformation.

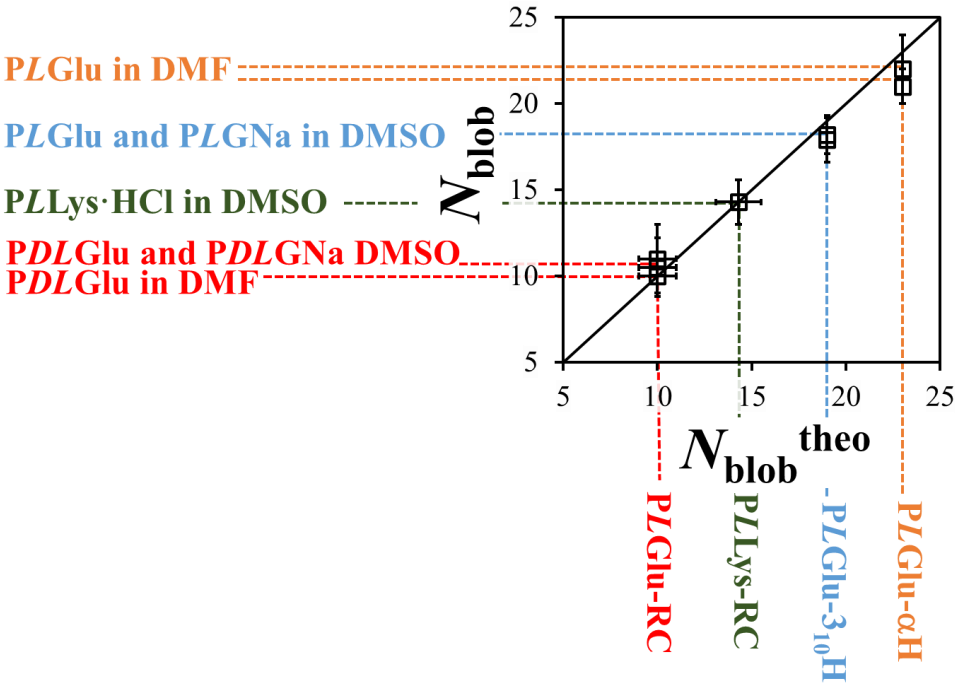


Figure 3.7. Comparison of the experimental N_{blob} and N_{blob}^{theo} values obtained with different polypeptides under different conditions. RC = random coil, 3_{10}H = 3_{10}-helix, αH = α-helix. P*L*Glu in DMF,[23,52] P*L*Glu in DMSO,[23] P*DL*Glu in DMF and DMSO,[23] and P*L*GNa, P*L*L·HCl, and P*DL*GNa in DMSO (this study).

In terms of the comparison between *blobs* and *foldons*, the N_{blob} values for the structured 3_{10} and α-helices of the Py-P*L*Glu samples in DMSO and DMF are around 20 residues. The similarity between the size of *foldons*, estimated to be around 20 amino acids as

determined by hydrogen exchange (HX) experiments conducted by NMR or mass spectrometry on structured or partially structured proteins,[12] and the size of *blobs* for structured polypeptides determined by PEF suggests that *blobs* and *foldons* might represent a same entity as was suggested earlier.[16,23] However, whereas the *foldon* size of unfolded polypeptides cannot be determined by HX experiments, since all amide hydrogens of an unfolded protein are fully exposed to the solvent and undergo HX at a similar rate preventing the determination of *foldon* size, the *blob* size can be determined for both structured and unstructured polypeptides. The FBM determined an N_{blob} value of 10.3 ± 1.7 and 14.3 ± 1.3 for the fully unfolded P*DL*Glu or P*L*L samples in DMSO, respectively. The larger *blob* size found for P*L*L compared to P*L*Glu could be attributed to the longer reach enabled by the lysine side chain and the 1-pyreneacetyl derivative. The N_{blob} values were lower for the unstructured P*DL*Glu and P*DL*GNa polypeptides compared to the structured P*L*Glu and P*L*GNa, reflecting the lower density experienced by the pyrenyl labels in the unfolded polypeptide coil. Finally, charged residues did not seem to affect the *blob* size for structured or unstructured polypeptides, suggesting that charged residues would not affect *foldon* size also if *blobs* and *foldons* were a same entity.

3.6 Conclusions

This study represents the first example where a combination of PEF, FBM, and MMOs were used to probe the effects that charged amino acids had on polypeptide conformation and dynamics in DMSO. The use of a series of pyrene-labeled homopolypeptides such as

poly(glutamic acid)s and P*L*L enabled the alteration of some specific structural parameters in a controlled manner such as the polypeptide configuration (structured P*L*Glu versus randomly coiled P*DL*Glu), the ionic state of the polypeptide (charged P*L*L·HCl, P*L*GNa, and P*DL*GNa versus neutral P*L*Glu and P*DL*Glu), or the nature and spacing between the side chain charges and the polypeptide backbone (five side chain atoms for the positively charged P*L*L·HCl versus four side chain atoms for the negatively charged P*L*GNa and P*DL*GNa). The combination of PEF, FBM, and MMOs enabled the study of the effects induced by changes in the aforementioned structural parameters onto polypeptide conformation and dynamics in DMSO. In particular, these measurements established that charged amino acids reduced side chain mobility by 60% for the polypeptides studied. They also demonstrated that while P*L*L·HCl and P*DL*GNa were random coils in DMSO, P*L*Glu and P*L*GNa remained structured, leading to the conclusion that they adopted a 3_{10}-helical conformation. This represented the first example where the combination of PEF, FBM, and MMOs was used to predict the unknown structure of a macromolecule in solution. Finally, the most important conclusion from this study might be that the *blobs* characterized through N_{blob} and k_{blob} display features in terms of size and dynamics that are similar to those expected of *foldons*, suggesting that *blobs* and *foldons* might actually represent a same entity.

Chapter 4

The Effect of Structure on Polypeptide Blobs: A Model Study Using Poly(L-Lysine)

4.1 Abstract

The conformation of a series of pyrene-labeled poly(*L*-lysine)s (Py-P*L*Ls) in 60:40 and 90:10 (v/v) acetonitrile:water mixtures was determined by comparing the results obtained from the fluorescence *blob* model (FBM) analysis of their fluorescence decays with those obtained from molecular mechanics optimizations (MMOs). P*L*L aggregates formed in both solutions, as demonstrated by FRET experiments between naphthalene- and pyrene-labeled P*L*Ls. The addition of an excess of unlabeled P*L*L allowed the conformational study of isolated Py-P*L*L embedded in a matrix of unlabeled P*L*Ls. By varying the acetonitrile (ACN) content of the solution from 60 to 90 vol% ACN, Py-P*L*L was found to undergo a conformational change from a random coil to an α-helix. The conformational change induced an increase in the maximum number of lysines (N_{blob}) separating two pyrene-labeled lysines that could still form an excimer between an excited and a ground-state pyrene. N_{blob} obtained from the FBM analysis increased from 15.2 ± 2.1 to 25.2 ± 1.2 lysines as P*L*L changed its conformation from a random coil to an α-helix. AFM revealed that the α-helical P*L*Ls organized themselves into structured bundles ~22 nm in diameter. FBM analysis of the decays acquired with a solution of aggregated Py-P*L*Ls in a 90:10 ACN:water mixture yielded a larger N_{blob} value of 36.6 ± 3.4. The increase in N_{blob} indicated that the Py-P*L*L constructs could now interact with one another in the helical bundles. This increase in N_{blob} was then used in conjunction with MMOs to determine an interhelical spacing of 2.9 ± 0.1 nm for Py-P*L*Ls in a bundle. This interhelical spacing resulted in a local density of 0.25 ± 0.01 g·cm^{-3} for the bundles of P*L*L α-helices, which was a reasonable density for a protein in solution. This study describes an experimental

means to probe the number of amino acids that interact with each other as the conformation of a polypeptide evolves from that of a random coil to that of an α-helix, to finally that of a bundle of α-helices.

4.2 Introduction

Uncovering the underlying physical phenomena that convert the one-dimensional sequence of a protein into its three-dimensional structure has been and continues to be the object of intense research,[1,2] as the three-dimensional structure of a protein defines its biological function.[3,4] However, proteins are such complex macromolecules that this task remains challenging. One approach that has been actively pursued to simplify the study of protein folding is to assume that a protein can be compartmentalized into subdomains where folding takes place cooperatively.[5–8] This assumption significantly simplifies the study of the folding process by reducing the total number of conformations that needs to be considered from being exceedingly large for the whole protein to manageable for an ensemble of smaller subdomains. As of late, the existence of these subdomains has been supported through the notion of *foldons*[5,9,10] and *blobs*.[7,8,11]

The existence of *foldons* has primarily been substantiated by hydrogen exchange (HX) experiments.[5,9,10] By monitoring the exchange rate of a protein amide protons by NMR or mass spectrometry (MS), groups of amino acids were found to have amide protons that exchanged with the solvent at a similar rate, indicating that they were part of a same subdomain.[5] The study of some proteins in their native or near native state by HX experiments have shown that

foldons contain between ~15 and 35 amino acids, depending on the amino acid sequence and location of the *foldon* within a protein.[5,9,12−14] Although HX experiments yield the *foldon* size, the origin of the specific size taken by *foldons* is not fully understood. Moreover, since HX experiments are conducted on proteins near their native state, the measured *foldon* size is only representative of the size of a folding subdomain near the end of the folding pathway and cannot provide information on if/how the *foldon* size changes as a protein folds. Such considerations matter as the number of conformations that needs to be considered to predict the folding time of a protein, an area of considerable scientific interest,[15-18] depends critically on the *foldon* size.[5] As it turns out, a fluorescence-based methodology was recently introduced to determine the number of structural units in a macromolecule that interact with one another on a time and length scale defined by the fluorescence of the dye pyrene.[8,19−21] It provides an alternative to the *foldon*-based approach to investigate how the number of interacting amino acids varies as the conformation of a polypeptide evolves along its folding pathway from a random coil to the organized assembly of secondary structures in the 3D structure of a protein, and it was applied in this study.

The methodology takes advantage of the ability of pyrene to form an excimer upon encounter between an excited and a ground-state pyrene. When covalently attached onto a macromolecule, reduced mobility and steric hindrance experienced by an excited pyrene imply that during the time pyrene remains excited, it can only form excimer by encountering a ground-state pyrene inside a small subvolume in the macromolecule referred to as a *blob*. A macromolecule can then be divided into a string of *blobs*.[21] These concepts led to the

formulation of the fluorescence blob model (FBM), which describes the kinetics of pyrene excimer formation (PEF) in a macromolecule viewed as a cluster of *blobs* where the pyrenyl labels are randomly distributed according to a Poisson distribution. FBM analysis of the time-resolved fluorescence decays of a macromolecule randomly labeled with pyrene yields $<n>$, the average number of pyrenes distributed randomly inside each *blob* according to a Poisson distribution. In turn, $<n>$ is used to calculate N_{blob}, which represents the maximum number of structural units separating two pyrene-labeled structural units inside a *blob* that allow excimer formation between an excited and a ground-state pyrene. Because the dye pyrene can self-quench to form an excimer,[22,23] pyrene acts as both a dye and a quencher, simplifying the labeling requirements of a macromolecule to be studied by the FBM since it only requires the attachment of a single dye. In addition, the long unquenched lifetime of pyrene provides a sufficient timespan to study rigid macromolecules, making pyrene an ideal dye for probing polypeptides.[24–27]

To date, the FBM has been applied to several pyrene-labeled polypeptides including poly(*L*-glutamic acid)s (P*L*Glu),[8,25,26,29] poly(*D*,*L*-glutamic acid)s (P*DL*Glu),[8,29] arborescent P*L*Glu's,[28] and protonated P*L*L.[29] A combination of circular dichroism and FBM indicated that Py-P*DL*Glu adopted an unstructured conformation in DMF with a corresponding N_{blob} value of 11, while Py-P*L*Glu adopted an α-helical structure with a significantly larger N_{blob} value of 20.[8] The similarity of N_{blob} when P*L*Glu adopted a structured conformation to that of the expected *foldon* size of ~20 suggested that *blobs* and *foldons* might represent similar entities.[5] Furthermore, arborescent Py-P*L*Glu's have been used as mimics of a protein in its molten

globular state near the end of the folding pathway. The increased crowding of the P*L*Glu helices with increased generation number resulted in a further enhancement of N_{blob} to a maximum value of 28 for the third generation (G3) arborescent Py-P*L*Glu.[28]

This application of the FBM to probe the folding subdomains of a polypeptide provided several distinct advantages. Since the size of a *blob* is defined by the diffusive encounters between an excited and ground-state pyrene attached onto a polypeptide, the N_{blob} value is a direct reflection of the number of amino acids included inside a *blob*, where the timescale over which the backbone and the side chains undergo diffusive motion controls pyrene-pyrene encounters. Thus, the FBM suggests that the size of the folding subdomains within a polypeptide is defined by the number of amino acids whose combined mobility is reflected by the ability of two pyrenyl labels located inside this subdomain or *blob* to form an excimer. The FBM also provides quantitative information on the dynamics experienced by amino acids inside a *blob* through the term k_{blob}, which represents the rate of encounter between amino acids inside a *blob*. Additionally, since N_{blob} is sensitive to the local concentration of amino acids within a *blob*, the FBM provides a direct tool to probe how the number of amino acids involved in a *blob* evolves along the entire folding pathway of a given polypeptide. This is contrary to the *foldons* characterized through HX experiments, which describe folded subdomains near the end of a protein folding pathway. The preliminary results obtained with the poly(glutamic acid) constructs support the notion that N_{blob} increases as the polypeptide becomes more structured.[8] However, this conclusion was reached by studying different poly(glutamic acid) constructs,

namely P*L*Glu and P*DL*Glu. To study how polypeptide structure affects N_{blob} in a more controlled manner, poly(*L*-lysine) (P*L*L) was employed in the present study.

P*L*L is an excellent model polypeptide to study how N_{blob} changes with structure due to its inherent malleability that enables P*L*L to adopt varied conformations.[30–35] By controlling the solvent composition, the solution pH and temperature, P*L*L is able to undergo conformational changes from an unfolded coil[30–33] to secondary structures such as α-helices,[30–34] β-sheets,[30–34] and PPII-helices,[33] and higher order aggregates of secondary structures including amyloid-like fibrils.[34,35] In particular, P*L*L adopts structures that are similar in nature to β–amyloid peptide aggregates and lithostathine helical filaments linked to Alzheimer's[36] and Creutzfeldt-Jakob disease,[37] respectively. Ordered helical domains, such as those generated in bundles of P*L*L helices,[34] are also encountered in transmembrane proteins.[38,39] The large range of P*L*L conformations provides an opportunity to use P*L*L as a mimic for a protein as it passes through the different stages along the folding pathway, from a coiled conformation[30–33] representative of an unfolded protein to a structured α-helix[30–34] reflecting the formation of secondary structural motives, all the way to α-helical P*L*L aggregates[34] which display similarities with the tertiary structure of some fully folded proteins.

To this end, a series of P*L*Ls were randomly labeled with the dye pyrene (Py-P*L*Ls). By monitoring the fluorescence of the Py-P*L*L constructs in acetonitrile:water mixtures, the FBM was applied to characterize the size of a P*L*L *blob* and its dynamics through the parameters N_{blob} and k_{blob}, respectively. The change in N_{blob}, as the solvent composition was varied, was used in conjunction with MMOs to characterize the conformational change from a

random coil to an α-helix undergone by P*L*L as the solvent composition was changed. The presence of higher-ordered Py-P*L*L aggregates was demonstrated using a combination of FRET and AFM. The FBM was then applied to characterize the conformation and the internal spatial arrangement of the secondary structures found in P*L*L aggregates. The determination of N_{blob} as P*L*L morphed from a random coil to an α-helix, and finally to an α-helix embedded in a bundle of P*L*L α-helices, similar to the arrangement of structural motives encountered in the tertiary structure of a protein, provided a first glimpse at how the size of a *blob*, and possibly the size of *foldons*, evolves along the folding pathway of a polypeptide.

4.3 Experimental

Materials: Acetic acid (glacial, Sigma), acetonitrile (ACN, Sigma, ≥ 99.9%), butylamine (Aldrich, 99.5 %), *N*,*N*'-dicyclohexylcarbodiimide (DCC, Fluka Analytical, ≥ 99.0 %), *N*,*N*'-diisopropylcarbodiimide (DIC, Sigma, 99 %), *N*,*N*-diisopropylethylamine (DIPEA, Sigma, 99.5 %), *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide hydrochloride (EDC·HCl, Sigma, ≥ 99.0 %), *N*,*N*-dimethylformamide (DMF, Sigma, ≥ 99.8 %), dimethyl sulfoxide (DMSO, Sigma, ≥ 99.9 %), hydrochloric acid (HCl, Fisher, 34 wt% in water), *N*-hydroxysuccinimide (HOSu, Sigma, 98 %), 1-naphthaleneacetic acid (NpAcOH, Aldrich, technical grade), 1-pyreneacetic acid (PyAcOH, Aldrich, 97 %), and sodium hydroxide (NaOH, Sigma, ≥ 97.0 %) were used as received. Deionized water (DIW) was obtained from a Biopure Series 4400 Single Pass Reverse Osmosis system. The synthesis of all small molecules not listed in the materials section are reported in the Supporting Information (SI).

Pyrene-labeled poly(L-lysine)s (Py-PLL)s: The hydrobromide salt of poly(*L*-lysine) (Alamanda Polymers, M_n = 165,100 g·mol^{-1}, DP = 790, $Đ$ = 1.06) was labeled with PyAcOSu according to the method described earlier[29] to yield Py-P*L*L whose chemical structure is shown in Figure 4.1A. The hydrochloride salt of Py-P*L*L was used in all experiments but was still referred to as Py-P*L*L for the sake of simplicity.

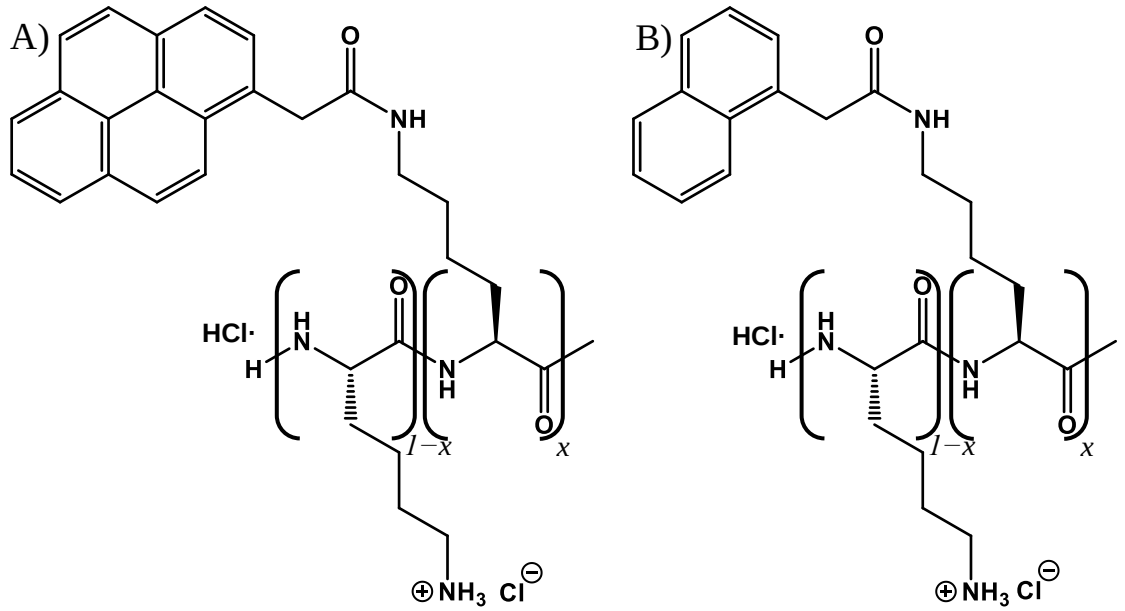


Figure 4.1. Chemical structures of A) Py-P*L*L and B) Np-P*L*L. *x* equals the molar fraction of lysines bearing a dye molecule.

Naphthalene-labeled poly(L-lysine) (Np-PLL): A procedure similar to that used earlier to prepare the Py-P*L*L constructs[29] was applied to label P*L*L with naphthalene by simply replacing PyAcOSu in the reaction with NpAcOSu (0.10 eq. per lysine). All other aspects of the synthesis

remained the same. As for Py-P*L*L, the hydrochloride salt of Np-P*L*L shown in Figure 4.1B was used in all experiments but was referred to as Np-P*L*L.

UV-Vis Absorption: The UV-Vis absorption spectrum was acquired using a Varian Cary 100 Bio spectrophotometer using a 1 cm pathlength cell. A background correction of the solvent was applied to each spectrum.

Dye content of Py- and Np-PLLs: The molar fraction (x) of lysine residues labeled with pyrene and naphthalene for the Py-P*L*L and Np-P*L*L constructs, respectively, was calculated using Equation 4.1. In Equation 4.1, λ_x represents the moles of dye per gram of sample, M_x equals the molar mass of a lysine residue bearing a dye pendant (M_{Py} = 370.45 g·mol^{-1}, M_{Np} = 296.37 g·mol^{-1}), and M_L is the molar mass of a lysine residue protonated with HCl (M_L = 164.63 g·mol^{-1}). λ_x was determined by accurately weighing freshly lyophilized dye-labeled P*L*L·HCl and dissolving it in a known amount of water with sonication. A small aliquot (ca. 0.1 mL) was diluted with DMF (ca. 5 mL) and the UV absorbance of the solution, or a dilution thereof, was measured to determine the molar concentration of the dye according to the Beer-Lambert law. A molar absorptivity coefficient of the model compound *N*-butyl-1-pyreneacetamide (PyAcNBu, ε_{Py} = 37,300 M^{-1}·cm^{-1} at 345 nm in DMF)[40] was used to determine λ_{Py} for the Py-P*L*Ls. *N*-Butyl-1-naphthalenacetamide (NpAcNBu), with a molar absorptivity coefficient (ε_{Np}) of 6,350 ± 21 M^{-1}·cm^{-1} at 284 nm in DMF (See Figure S4.1), was used to determine λ_{Np} for Np-P*L*L.

$$x = \frac{M_L}{\lambda_x^{-1} + M_L - M_x} \tag{4.1}$$

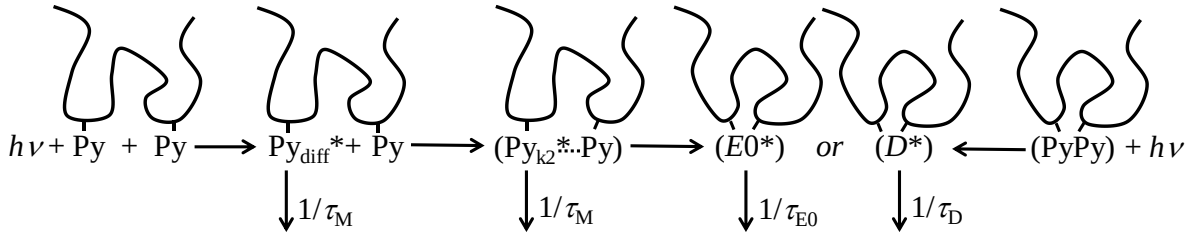
Steady-state fluorescence (SSF): Py-P*L*L was dissolved in a small volume of water acidified with HCl. After dissolution with sonication, an aliquot was diluted in DMF, where the absorption spectrum of the pyrenyl labels would not be distorted by the presence of pyrene aggregates, and the absorption was measured. The absorbance of the water stock solution was back-calculated and an appropriate volume of Py-P*L*L aqueous solution was added to a vial such that the pyrene concentration in a 4 mL solution would equal 2.7×10^{-6} M. The Py-P*L*L aqueous solution in the vial was acidified with 10 μL of 1 M HCl to ensure that all lysines of Py-P*L*L were protonated. Water was then added to the vial to bring the volume of the Py-P*L*L aqueous solution to 0.4 or 2.4 mL in preparation for the 90:10 or 60:40 ACN:water solutions, respectively. Further addition of 3.6 or 1.6 mL of ACN to the vial brought the solution volume to ca. 4 mL. The solutions were then sonicated for 20 minutes. Oxygen was removed from the solutions by outgassing with a gentle flow of N_2 (Praxair, N4.0) for ca. 25 minutes and the fluorescence cell was sealed to maintain an oxygen-free atmosphere. The SSF spectra were acquired with a Horiba QM-400 spectrofluorometer equipped with a xenon arc lamp. Pyrene was excited at 344 nm using an excitation slit width of 1 nm. The fluorescence spectra were recorded from 350 to 600 nm with a 1 nm resolution. The fluorescence intensity of the monomer (I_M) and excimer (I_E) of pyrene were calculated using the area under the SSF spectrum from 375 to 381 nm and from 500 to 530 nm, respectively, to determine the I_E/I_M ratio.

Förster resonance energy transfer (FRET): Steady-state FRET experiments were conducted on solutions that contained Np-P*L*L as the energy donor and Py-P*L*L as the energy acceptor. For each FRET experiment, three solutions were prepared according to the method outlined in the SSF section. The first solution with Py-P*L*L had a pyrene concentration of 2.7×10^{-6} M, the second solution with Np-P*L*L had a naphthalene concentration of 15.7×10^{-6} M, and the third solution contained both Py-P*L*L and Np-P*L*L at the same concentrations as those used to prepare the first two solutions. These three solutions were then transferred to fluorescence cells where they were outgassed with a gentle flow of N_2 (Praxair, N4.0) for ca. 25 minutes before acquiring their SSF spectrum. The SSF spectra were acquired with excitation and emission slit widths of 2 and 1 nm, respectively. The fluorescence was monitored by exciting the samples at 295 nm and scanning the emission from 300 to 570 nm.

Time-resolved fluorescence (TRF): Samples were prepared in a similar manner as those in the SSF measurements. The fluorescence decay of pyrene was acquired using an IBH TC-SPC spectrofluorometer equipped with a NanoLED-340 laser. All decays were collected using an excitation wavelength of 344 nm over 1,024 channels. The monomer fluorescence was collected with a time-per-channel of 2.04 $ns\cdot ch^{-1}$ until a total of at least 40,000 counts was obtained at the fluorescence decay maximum at an emission wavelength of 378 nm using a 370 nm longpass filter. The excimer fluorescence decays were acquired with a decay maximum of at least 20,000 counts at 510 nm with a 470 or 495 nm longpass filter using a time-per-channel of 1.02 $ns\cdot ch^{-1}$. The longpass filters were used to minimize potential light scattered by the solutions from reaching the detector. The instrument response function was collected using

a LUDOX dispersion whose signal was acquired at 344 nm with a peak maximum of at least 20,000 counts.

Fluorescence decay analysis: The time-resolved fluorescence decays of the monomer and excimer of Py-P*L*Ls were globally analyzed with the fluorescence *blob* model (FBM) using Equations S4.1 – S4.3 given in SI. The FBM acknowledges that a pyrenyl label covalently attached to a macromolecule probes a finite volume, referred to as a *blob*, while it remains excited. The macromolecule randomly labeled with pyrene can be compartmentalized into a cluster of *blobs* among which the pyrenyl labels distribute themselves randomly according to a Poisson distribution. The average number $<n>$ of pyrenes per *blob* retrieved from the FBM analysis is used in conjunction with the pyrene-content (λ_{Py}) to determine N_{blob} in Equation S4.4. The FBM assumes the existence of four pyrene species which have been illustrated in Scheme 4.1. Within each *blob*, the excited Py_{diff}* diffuses toward a ground-state pyrene with a rate constant k_{blob} to form the species Py_{k2}* corresponding to an excited pyrene located close to a ground-state pyrene. The two pyrenes then rearrange rapidly to form one of the two excimers *E*0* or *D** with the rate constant k_2. The excimer species *E0** and *D** represent excited pyrene dimers that are either well or poorly stacked and decay to the ground-state through their lifetime τ_{E0} and τ_D, respectively. The excimer can also be formed by the direct excitation of the pre-associated *E0* and *D* pyrene dimers. Excited pyrenes located inside *blobs* without ground-state pyrenes are referred to as the species Py_{free}, which emit with their unquenched lifetime τ_M. More information about the FBM can be found in earlier reviews.[20,21,41]



Scheme 4.1. Reaction scheme illustrating the different pyrene species leading to PEF.

Atomic Force Microscopy (AFM): A 90:10 ACN:water mixture containing 120 μg/mL P*L*L was prepared in the following manner. P*L*L (5 mg) was dissolved in water (5.95 mL) acidified with 50 μL of 1 M HCl. After dissolution, the P*L*L solution was filtered through a 0.2 μm PTFE filter to remove any dust. The filtered P*L*L solution (0.2 mL) was then added to filtered ACN (1.8 mL). The mixture was sonicated for 15 minutes and was deposited (ca. 2 drops) on a freshly cleaved mica surface that had been rinsed with filtered ACN. The P*L*L solution was left on the mica surface for 30 minutes before being rinsed with filtered ACN. After drying overnight in a dust-free environment, the AFM images were collected on a DI Dimension 3100 instrument operated with a silica cantilever in tapping mode with a 1 Hz scan rate.

4.4 Results

P*L*L is known to adopt a broad variety of conformations in solution, taking the form of α-helices,[30–34] PPII-helices,[33] β-sheets,[30–34] and fibrillar aggregates[34,35] depending on the solvency conditions and temperature. The most commonly used factor to induce a conformational change for P*L*L is pH. Unfortunately, fluorescence quenching by the primary amine ends of lysine prevents the fluorescence study of the pyrenyl labels bound to P*L*L at the

high pH used to induce P*L*L to adopt an α-helical conformation.[29] Besides pH, adjusting the solvent composition is another common approach to change the conformation of P*L*L.[30,32] A previous study reported that the addition of acetonitrile (ACN) to an aqueous solution of protonated P*L*L could induce P*L*L to undergo a coil-to-helix transition.[32] Circular dichroism (CD) measurements indicated that P*L*L was a random coil in ACN:water mixtures containing less than 70 vol% ACN, but that P*L*L adopted an α-helical conformation for ACN contents above 80 vol%. Since the fluorescence of Py-P*L*L could only be monitored when Py-P*L*L was protonated, this study used a 90:10 and 60:40 (v/v) ACN:water mixture to characterize the protonated Py-P*L*L in a helical and coiled conformation, respectively. Due to the hydrophobic nature of pyrene, the 60:40 ACN:water solution was selected over pure water in order to minimize pyrene-induced aggregation of Py-P*L*L.[42]

Solvent Induced Aggregation of PLL: Preliminary fluorescence measurements conducted with the Py-P*L*L solutions showed a stronger than expected excimer fluorescence with respect to the monomer emission. This observation suggested the formation of Py-P*L*L aggregates in solution. Since the presence of Py-P*L*L aggregates in solution would affect the interpretation of the results obtained by pyrene excimer fluorescence/formation (PEF), their existence was investigated by using Förster resonance energy transfer (FRET) with naphthalene and pyrene labels, since they constitute a well-known pair of FRET donor and acceptor,[43–45] respectively. A P*L*L sample was labeled with 1-naphthaleneacetic acid to yield a Np-P*L*L sample with a molar fraction of naphthalene-labeled lysine (x_{Np}) of 0.09. The absorbance and fluorescence spectra of the Np-P*L*L and Py-P*L*L samples are given in Figure S4.2. Since the Förster radius

(R_o) of the naphthalene-pyrene pair equals 2.9 nm[46] and the width of a single P*L*L helix equals ca. 1.6 nm when measured from one ε-ammonium on a lysine residue to another located on the opposite side of a P*L*L helix constructed with HyperChem, the presence of FRET would indicate that Np-P*L*L and Py-P*L*L would have to be within $2\times R_o$ (= 5.8 nm) of one another, and thus form an intermolecular aggregate.

The existence of P*L*L aggregates was first investigated in the 60:40 ACN:water mixture. The SSF spectra of solutions containing Py-P*L*L alone, Np-P*L*L alone, and a Py-P*L*L and Np-P*L*L mixture are shown in Figure 4.2A. The solution containing only Np-P*L*L exhibited a strong fluorescence characteristic of naphthalene, and the Py-P*L*L solution exhibited the weak emission of pyrene due to the low absorbance of Py-P*L*L at the 295 nm excitation wavelength (see absorption spectrum of Py-P*L*L in Figure S4.2A). The solution containing the mixture of Py-P*L*L and Np-P*L*L displayed the fluorescence bands characteristic of both the Py- and Np-P*L*L samples. By adding the SSF spectra of the Np-P*L*L solution to that of the Py-P*L*L solution, a theoretical spectrum was produced of a solution containing both Py-P*L*L and Np-P*L*L in the absence of FRET. Figure 4.2B compares this theoretical spectrum with the experimental spectrum obtained for the solution of the Np-P*L*L and Py-P*L*L mixture. The fluorescence spectrum of the Np-P*L*L and Py-P*L*L mixture exhibited a slight suppression of the naphthalene emission accompanied with an increase in pyrene emission. Although these changes are relatively small, this deviation from the theoretical spectrum indicated that FRET occurred between naphthalene and pyrene and that the P*L*Ls must be associated in the 60:40 ACN:water mixture.

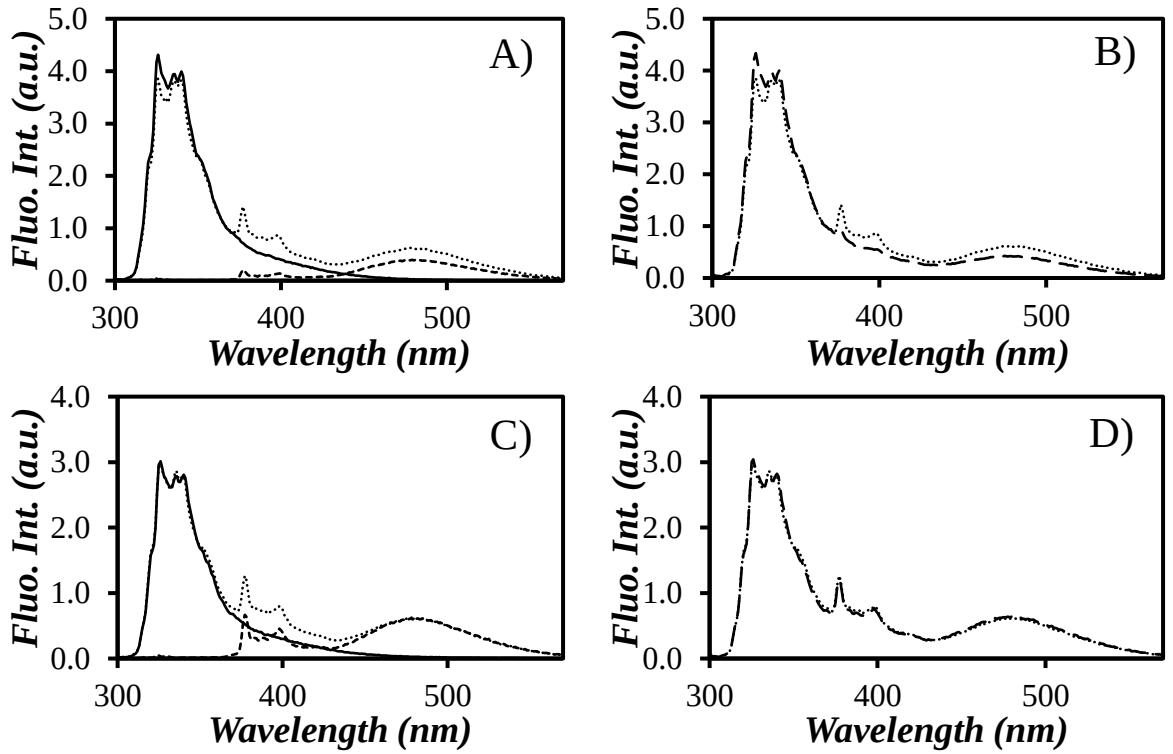


Figure 4.2. Steady-state fluorescence spectra in a 60:40 ACN:water mixture of (A,C) (- - -) Py-P*L*L alone, (−) Np-P*L*L alone, and (⋯) the Py-P*L*L and Np-P*L*L mixture. (B,D) Comparison of (⋯) the experimental fluorescence spectrum of the mixture of Py-P*L*L and Np-P*L*L and (− −) the theoretical fluorescence spectrum resulting from the addition of the individual Py-P*L*L and Np-P*L*L spectra. The fluorescence spectra were acquired (A,B) without and (C,D) with a 50-fold excess of unlabeled P*L*L. λ_{ex}= 295 nm, [Np] = 15.7 μM, x_{Np} = 0.09, [Py] = 2.7 μM, x = 0.12.

Isolation of the fluorescently labeled P*L*Ls could be achieved by adding a 50-fold excess of unlabeled P*L*L to the fluorescently labeled P*L*Ls. The addition of unlabeled P*L*L by no means prevented the aggregation of the P*L*Ls, but rather ensured that the fluorescently

labeled P*L*Ls were surrounded by unlabeled P*L*Ls. The successful isolation of the fluorescently-labeled P*L*Ls was demonstrated through the comparison of Figures 4.2C and D. Figure 4.2D shows that the theoretical fluorescence spectrum, obtained by adding the individual Py-P*L*L and Np-P*L*L fluorescence spectra, overlapped perfectly with the experimental spectrum of the Py-P*L*L and Np-P*L*L mixture. The perfect overlap of the fluorescence spectra in Figure 4.2D demonstrated the absence of FRET between Np-P*L*L and Py-P*L*L and that the addition of a 50-fold excess of unlabeled P*L*L was sufficient to isolate the fluorescently labeled P*L*Ls from one another. A similar procedure was applied to Py-P*L*L and Np-P*L*L in the 90:10 ACN:water mixture and the corresponding SSF spectra are given in Figure S4.3. The SSF spectra demonstrated aggregation of the fluorescently labeled P*L*Ls, but that addition of unlabeled P*L*L isolated the labeled P*L*Ls from one another. Therefore, these experiments demonstrated that addition of unlabeled P*L*L prevented intermolecular excimer formation in both the 60:40 and 90:10 (v/v) ACN:water mixtures.

Steady-State Fluorescence of Isolated Py-PLLs: Following the conclusions drawn from Figure 4.2D, the isolated Py-P*L*L constructs were first studied in the presence of a 50-fold excess of unlabeled P*L*L. The SSF spectra of the Py-P*L*Ls in 60:40 and 90:10 ACN:water mixtures are given in Figures 4.3A and B, respectively. For a given pyrene content *x*, significantly more excimer fluorescence was generated with respect to the monomer by the 90:10 ACN:water mixture compared to the 60:40 ACN:water mixture. Although the increase in pyrene excimer formation observed in Figure 4.3 agrees with an increased density induced by a coil-to-helix conformational change,[8] a conformational change is not the only factor that influences the I_E/I_M

ratio. The I_E/I_M ratio is proportional to the product $k_{diff}\times[Py]_{loc}$, where k_{diff} is the rate constant for pyrene excimer formation by diffusive encounters and $[Py]_{loc}$ is the concentration of ground-state pyrene experienced locally by an excited pyrene bound to a macromolecule.[47] Therefore, an increase in I_E/I_M can be due not only to an increase in $[Py]_{loc}$ that might have been induced by the conformational change of Py-P*L*L, but also to an increase in k_{diff} due to a decrease in solvent viscosity. As it turns out, an increase in the ACN content from 60 to 90 vol% was accompanied by a significant decrease in solvent viscosity η from 0.75 to 0.39 mPa·s.[48] The inverse relationship between k_{diff} and solvent viscosity leads to an increase in I_E/I_M with decreasing solvent viscosity. Changes in solvent viscosity can be accounted for by considering the product $I_E/I_M\times\eta$ which allows for a more direct comparison between the I_E/I_M ratios obtained in different solvents. The similar $I_E/I_M\times\eta$ products obtained in Figure 4.3C suggest that after correcting for viscosity, PEF only increased a little as the ACN content increased from 60 to 90 vol%. A conformational change in these two solvent mixtures was expected to be accompanied by a much larger increase in I_E/I_M than that reported in Figure 4.3C, since the backbone is much more condensed in a helical conformation which should result in a higher $[Py]_{loc}$.[8] Conclusions drawn about the I_E/I_M ratios obtained from different solvents, as is the case in Figure 4.3C, are further complicated by the solvent-dependent probability of forming a pyrene excimer upon encounter between an excited and a ground-state pyrene.[49] Fortunately, the dynamic and structural factors described by k_{diff} and $[Py]_{loc}$ that affect the I_E/I_M ratios of Py-P*L*L can be deconvoluted from one another through FBM analysis of the time-resolved fluorescence (TRF) decays.

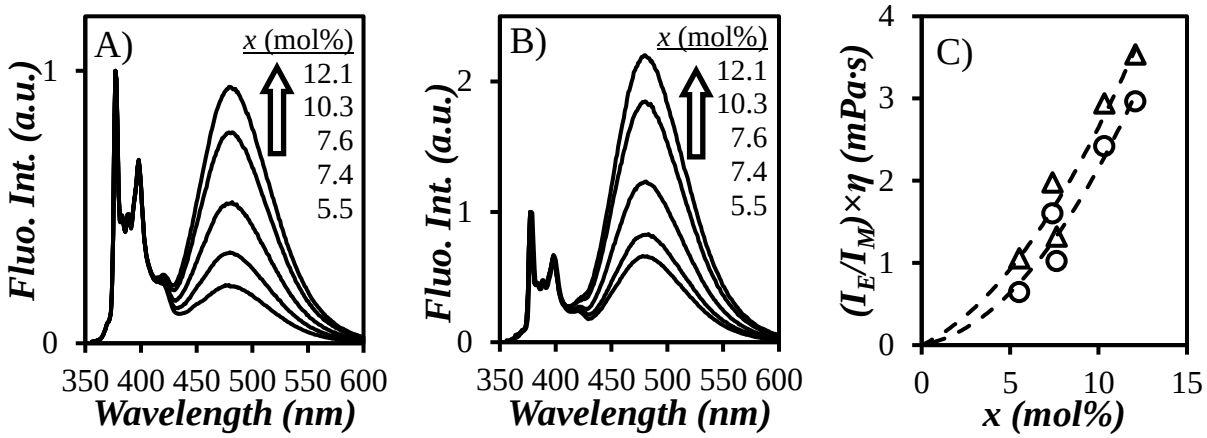


Figure 4.3. Steady-state fluorescence spectra of the Py-P*L*L constructs in aqueous solutions containing A) 60 and B) 90 vol% ACN with an excess of unlabeled P*L*L. C) Plot of $(I_E/I_M)\times\eta$ as a function of pyrene content for the Py-P*L*L constructs in aqueous solutions containing (O) 60 and (Δ) 90 vol% ACN with an excess of unlabeled P*L*L. The dashed lines were added to guide the eye.

Characterization of Individual PLLs by TRF: The fluorescence decays of Py-P*L*L in the ACN:water mixtures were analysed in accordance to the FBM. The FBM analysis yields two key parameters, namely k_{blob}, the rate constant of diffusive encounters between two structural units bearing a pyrenyl label inside a *blob*, and N_{blob}, the maximum number of lysines separating two pyrene-labeled lysines within a *blob* that still allow efficient pyrene excimer formation. Figure 4.4A displays N_{blob} as a function of pyrene content for the ACN:water mixtures containing Py-P*L*L with a 50-fold excess of unlabeled P*L*L. The addition of unlabeled P*L*L isolated the Py-P*L*Ls from one another, ensuring that N_{blob} represented isolated Py-P*L*Ls inside a matrix of aggregated P*L*Ls. For both the 60:40 and 90:10 ACN:water mixtures, N_{blob}

remained relatively constant as a function of pyrene content. The average N_{blob} value equalled 15.2 ± 2.1 and 25.2 ± 1.2 for Py-P*L*L in the 60 and 90 vol% ACN mixtures, respectively. The change in N_{blob} values between the two mixtures indicated that $[Py]_{loc}$ had changed, which is typically the case when a conformational change occurs. The experimental N_{blob} values could be compared to the theoretical N_{blob}^{theo} values obtained by molecular mechanics optimizations (MMOs) for a given Py-P*L*L conformation.

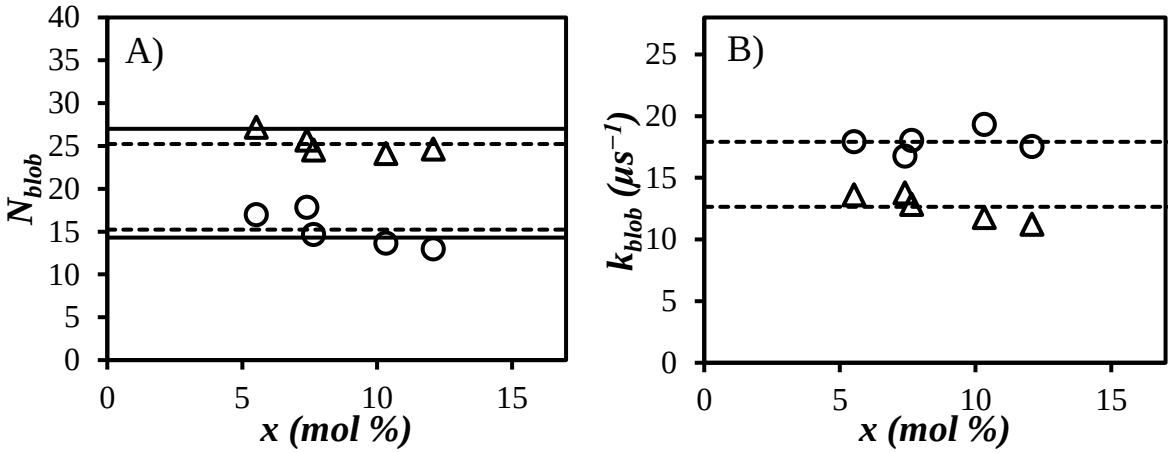


Figure 4.4. Plots of A) N_{blob} and B) k_{blob} as a function of the pyrene content of Py-P*L*L in aqueous solutions containing (○) 60 and (△) 90 vol% ACN with a 50-fold excess of unlabeled P*L*L. The dashed lines represent the average values and the solid lines represent the N_{blob}^{theo} values obtained from MMOs conducted on a Py-P*L*L construct simulated with HyperChem adopting a (N_{blob}^{theo} = 14.3 ± 1.2) random coil and (N_{blob}^{theo} = 27) α-helical conformation.

Such MMOs were performed in an earlier publication[29] and the main results for N_{blob}^{theo} are summarized hereafter. A P*L*L macromolecule was constructed in silico with the program

HyperChem. The P*L*L backbone was then extended before it was allowed to relax to yield a conformation representative of a randomly coiled P*L*L. The number of residues separating a reference pyrene-labeled lysine from a secondary pyrenyl label attached onto another lysine along the P*L*L backbone while still allowing pyrene excimer fluorescence (PEF) was then determined.[8,24,28,29] These MMOs yielded an N_{blob}^{theo} for a Py-P*L*L random coil of 14.4 ± 1.2 lysines.[29] The MMOs were then repeated for an α-helical Py-P*L*L construct. Since the backbone was now structured, the lysines were more densely packed in space. This increased lysine density resulted in an increased N_{blob}^{theo} of 27 for an α-helical Py-P*L*L construct.[29]

The experimental N_{blob} value of 15.2 ± 2.1 when Py-P*L*L was in the 60:40 ACN:water mixture matched closely the N_{blob}^{theo} value of 14.4 ± 1.2 for Py-P*L*L in a coiled conformation. It also agreed with the experimental N_{blob} value of 14.3 ± 1.3 obtained earlier for Py-P*L*L adopting a random coil conformation in DMSO.[29] The matching N_{blob} values confirmed that Py-P*L*L adopted a coiled conformation in the 60:40 ACN:water mixture, in agreement with the conformation determined by circular dichroism (CD) for protonated P*L*L in the same solvent mixture.[32] As the ACN content was increased to 90%, N_{blob} almost doubled to 25.2 ± 1.2 which matched the N_{blob}^{theo} value of 27 lysines found by MMOs for an α-helical Py-P*L*L construct. The good agreement found between the experimental N_{blob} and N_{blob}^{theo} confirmed that Py-P*L*L adopted an α-helical conformation in the 90:10 ACN:water mixture, again agreeing with published CD measurements of P*L*L under similar conditions.[32] The similarity between the experimental N_{blob} and N_{blob}^{theo} obtained from MMOs indicated that a combination of PEF,

FBM, and MMOs was able to successfully determine the conformation of isolated Py-P*L*L inside P*L*L aggregates.

In Figure 4.4B, the k_{blob} value of 17.9 ± 0.9 μs^{-1} in the 60:40 ACN:water mixture was found to be significantly higher than the k_{blob} value of 12.6 ± 1.1 μs^{-1} obtained for the 90:10 ACN:water mixture. Since N_{blob} and k_{blob} are related to, respectively, $[Py]_{loc}$ and k_{diff} used in the analysis of the SSF spectra, the larger N_{blob} and smaller k_{blob} values obtained in the 90:10 ACN:water mixture suggested that the change in solvent composition had an opposite effect on the dynamics (k_{blob} and k_{diff}) and structure (N_{blob} and $[Py]_{loc}$) of the Py-P*L*L constructs. This observation rationalized the similar I_E/I_M ratios found in Figure 4.3C for the 60:40 and 90:10 ACN:water mixtures, since changes in solvent composition seemed to affect dynamics and structure in opposite manners.

The difference in k_{blob} observed in the two solvent mixtures was unlikely to be due to a change in polypeptide conformation, as a previous study showed that k_{blob} remained constant for a given polypeptide regardless of its backbone conformation.[8] One possible cause for the difference in k_{blob} might be a different probability (p) of forming excimer upon the encounter of an excited and ground-state pyrene, since it is known that p depends on the solvent composition.[28,49] The lower k_{blob} value in the less viscous 90:10 ACN:water mixture would suggest that the probability of forming excimer might be lower in this solvent mixture than in the more viscous 60:40 ACN:water mixture. Another factor which likely contributed to the differing k_{blob} values was the intermolecular association experienced by P*L*L in these solvent mixtures. Since the Py-P*L*L constructs were embedded inside a matrix of P*L*L aggregates, it

stands to reason that the dynamics experienced by the side chains would be affected by these interactions due to the numerous contacts between the P*L*L chains. The difference in k_{blob} may also be a result of differences in packing efficiencies between the P*L*Ls in the helical and coiled conformation, a denser packing of the P*L*L α-helices resulting in a lower k_{blob} value. In either case, it is apparent that the difference in k_{blob} values obtained in different solvent mixtures can be attributed to numerous factors that compete with each other and cannot be unambiguously assigned. Therefore, the focus of the study was placed on the structural characterization of the Py-P*L*L constructs through the N_{blob} value.

AFM Measurements of PLL: AFM was applied to characterize the P*L*L aggregates identified by FRET in the 90:10 ACN:water mixture. The PEF experiments confirmed that under these conditions, P*L*L was in an α-helical conformation, as had been established earlier by CD measurements.[32] A P*L*L solution in the 90:10 ACN:water mixture was deposited on a mica plate for imaging. A full description of the solution preparation is given in the Experimental section. By far, the majority of the P*L*Ls were incorporated into large P*L*L aggregates about 1 to 2 μm in diameter, as shown in Figures 4.5A and B. These aggregates appeared to be composed of rather disorganized building blocks, making it difficult to draw any conclusions from these images. Fortunately, the areas on the mica surface between aggregates showed better isolated P*L*L structures in Figure 4.5C, whose aggregation must have resulted in the large P*L*L aggregates observed in Figures 4.5A and B.

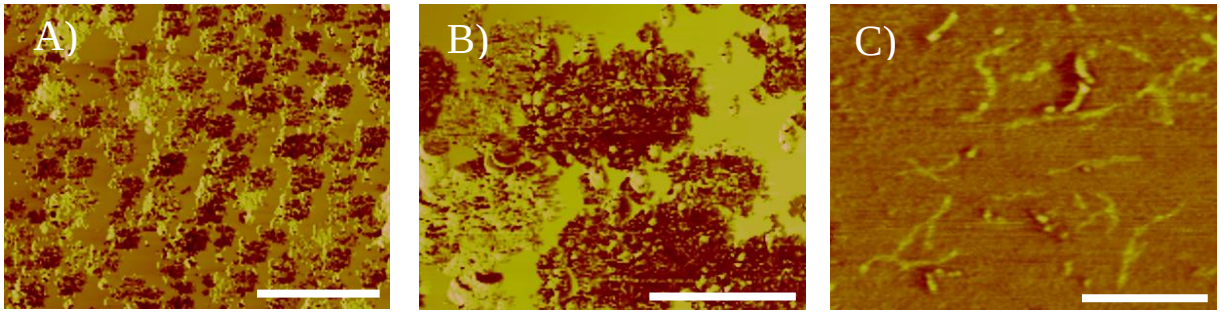


Figure 4.5. AFM phase images of P*L*L deposited on a mica surface from a 90:10 ACN:water mixture. A,B) Aggregates of P*L*L bundles and C) the bundles of P*L*L helices. The white bars indicate a length scale of A) 5 μm, B) 1.5 μm and C) 250 nm.

The structures in Figure 4.5C were identified as bundles of aligned P*L*L α-helices, as suggested by their fairly straight orientation. The DI Nanoscope software was used to determine that the bundles of α-helical P*L*Ls were 22.4 ± 3.7 nm wide (Figure S4.4). The average 22.4 nm width was significantly larger than the width of 1.6 nm expected for an individual P*L*L α-helix built with HyperChem. The bundle width was also much shorter than 110 nm, which would have been the end-to-end distance of a fully extended α-helical P*L*L with a DP of 790, assuming a helical rise of 0.14 nm per lysine. Thus Figure 4.5C suggested that the bundles were likely composed of an array of P*L*L α-helices aligned parallel to the main axis of the bundle, similarly to the bundles of P*L*L α-helices obtained in 90:10 methanol:water mixtures.[34] The images shown in Figures 4.5A – C suggest that P*L*L aggregation occurs over two length scales. The first length scale corresponds to the association of P*L*L helices into the organized bundles pictured in Figure 4.5C. These bundles then aggregate together in the large

macroscopically disorganized masses observed in Figures 4.5A and B. However, one key structural parameter missing from this analysis of P*L*L aggregation is the interhelical distance separating the P*L*L helices in the bundles shown in Figure 4.5C. How such information could be obtained is described in the next sections.

PEF in Bundles of PLL α-Helices: Since PEF in Py-P*L*L can only occur between an excited and a ground-state pyrene covalently attached onto the polypeptide, PEF reports on length scales no longer than a few nanometers and therefore provides a direct means to probe macromolecular structures on the nanometer scale. Consequently, a PEF-based study of a solution of aggregated Py-P*L*Ls would provide an experimental means to characterize the spatial arrangement of P*L*L α-helices inside P*L*L bundles on this length scale. To this end, the fluorescence of Py-P*L*L in a 90:10 ACN:water mixture was studied without unlabeled P*L*L. Figure 4.6A compares the I_E/I_M ratios of the Py-P*L*L solutions with and without unlabelled P*L*L. Since the solvent was the same in these experiments, direct comparison can be made from the pyrene excimer trends obtained with the Py-P*L*L samples. For a given pyrene content, the I_E/I_M ratio was higher for the Py-P*L*L solution without unlabeled P*L*L. The increased I_E/I_M ratios indicated that the Py-P*L*Ls were no longer isolated from one another and could now form excimer intermolecularly. To quantify this enhancement in PEF due to intermolecular interactions within the Py-P*L*L bundles, the fluorescence decays of the Py-P*L*L solutions in 90:10 ACN:water mixtures were analyzed according to the FBM.

Figures 4.6B and C compare the change in, respectively, the k_{blob} and N_{blob} values obtained for Py-P*L*L α-helices embedded in a P*L*L bundle and Py-P*L*L α-helices bundled

together. The k_{blob} value of 21.1 ± 2.5 μs^{-1} in Figure 4.6B for the Py-P*L*Ls bundles was significantly larger than the k_{blob} value of 12.6 ± 1.1 μs^{-1} for isolated Py-P*L*Ls inside unlabeled P*L*L bundles. Since the k_{blob} values were obtained in a same solvent mixture, the difference in k_{blob} could be discussed. The pseudo-unimolecular rate constant k_{blob} is equal to the product $k_{diff} \times (1/V_{blob})$ where V_{blob} is the *blob* volume and $1/V_{blob}$ is the concentration equivalent to one ground-state pyrene per *blob*. Therefore, the differing k_{blob} values indicated that either k_{diff} or V_{blob} had changed. Whether the Py-P*L*Ls were bundled with unlabeled or pyrene-labeled P*L*Ls, the Py-P*L*Ls experienced similar environments, implying that the volume V_{blob} probed by an excited pyrene should not change. This only left the term k_{diff} as the origin for the change in k_{blob}. The change in k_{diff} was attributed to changes in the probability (*p*) of two pyrenyl moities having the proper alignment for PEF, depending on whether the two pyrenyl labels where attached on a same helix (small *p* value) or on two different helices (high *p* value). When the Py-P*L*Ls were bundled with the unlabeled P*L*Ls, excimer formation could only occur from pyrenes originating from the same Py-P*L*L helix. This meant that a ground-state pyrene could only approach an excited pyrene located inside a *blob* from the two sides of the *blob* along the backbone. However, when the Py-P*L*Ls were bundled together, pyrenes from nearby helices could also contribute to the number of pyrenes in the *blob*. Since the *blob* could now be accessed by intermolecular pyrenes that had different orientations with respect to the intramolecular pyrenes, the increase in the number of possible orientations between two pyrenyl labels must have led to an increase in the probability *p* of having a proper alignment

for PEF. The increase in p is proposed as explanation for the increase in k_{diff} and thus k_{blob} observed between the experiments conducted with isolated and aggregated Py-P*L*L helices.

The increase in k_{blob} was also accompanied by an increase in N_{blob} from 25.2 ± 1.2 determined for an isolated Py-P*L*L α-helix to 36.6 ± 3.4 when the Py-P*L*Ls were bundled together. The increase in N_{blob} also indicated that the Py-P*L*Ls in the bundles of α-helices experienced local intermolecular interactions with one another. Since N_{blob} is a measure of the average number of lysine residues inside a *blob*, the increase in N_{blob} (ΔN_{blob}) observed for the Py-P*L*L bundles meant that the side chains of neighbouring Py-P*L*Ls contributed to the *blob* volume of a reference Py-P*L*L helix. Therefore, the ΔN_{blob} of 11.4 ± 3.7 was a result of an increase in $[Py]_{loc}$ experienced by the Py-P*L*L helices.

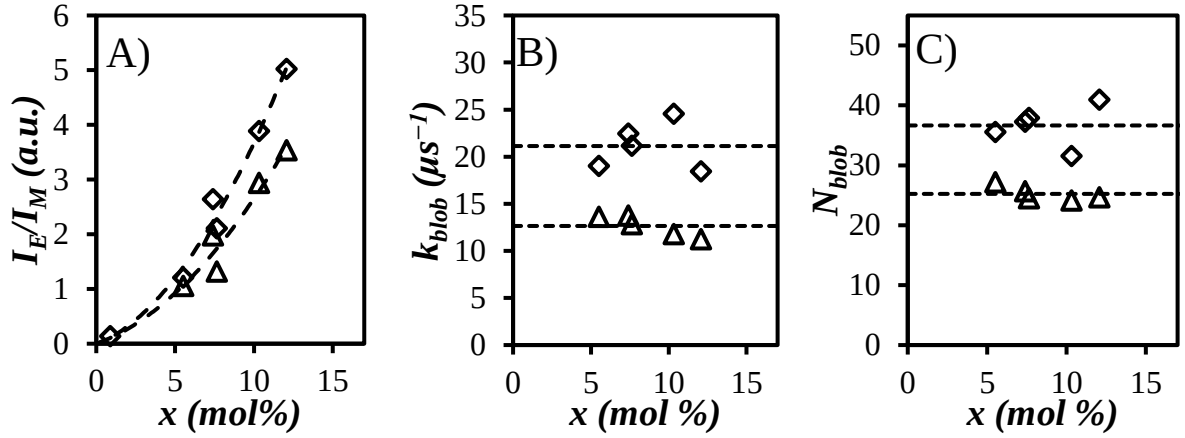


Figure 4.6. Plots of A) the I_E/I_M ratio, B) k_{blob}, and C) N_{blob} as a function of pyrene content for Py-P*L*L in a 90:10 ACN:water mixture (△) with and (◇) without a 50-fold excess of unlabeled P*L*L. The dashed lines in A) were added to guide the eye, and represent the averaged N_{blob} and k_{blob} values in B) and C), respectively.

Characterization of PLL bundles: The AFM image in Figure 4.5C indicated that the Py-P*L*Ls must be aligned parallel with one another inside the bundles. In addition to this, the presence of FRET between the Np- and Py-P*L*Ls indicated that the P*L*L α-helices must be packed within each bundle where they must have been separated by less than 5.8 nm from each other. Based on this information, it was assumed that the Py-P*L*Ls packed inside the P*L*L bundles according to a hexagonal-close-packing (HCP) arrangement. The experimental ΔN_{blob} value was used in conjunction with MMOs to calculate the density of Py-P*L*Ls in the bundles. Following a procedure published earlier,[28] MMOs were conducted to determine the increase in the theoretical N_{blob} value (ΔN_{blob}^{theo}) as a function of the distance ($d_{h\text{-}h}$) separating the axis of P*L*L α-helices constructed with HyperChem in a HCP arrangement. The central helix, referred to as the primary helix, was then labeled with pyrene halfway down the helix. A secondary pyrenyl label was then attached onto one of the residues on one of the six neighbouring helices. The pyrenes were then brought within 3.4 Å from each other to determine if excimer could form between the two pyrenes. If the frame of one pyrene could overlap with 7 or more carbons of the second pyrene when the two were 3.4 Å apart, the overlap between the two pyrenyl labels was considered sufficient to form an excimer. An example of two Py-P*L*Ls which were sufficiently close to one another to form excimer intermolecularly is shown in Figure 4.7A.

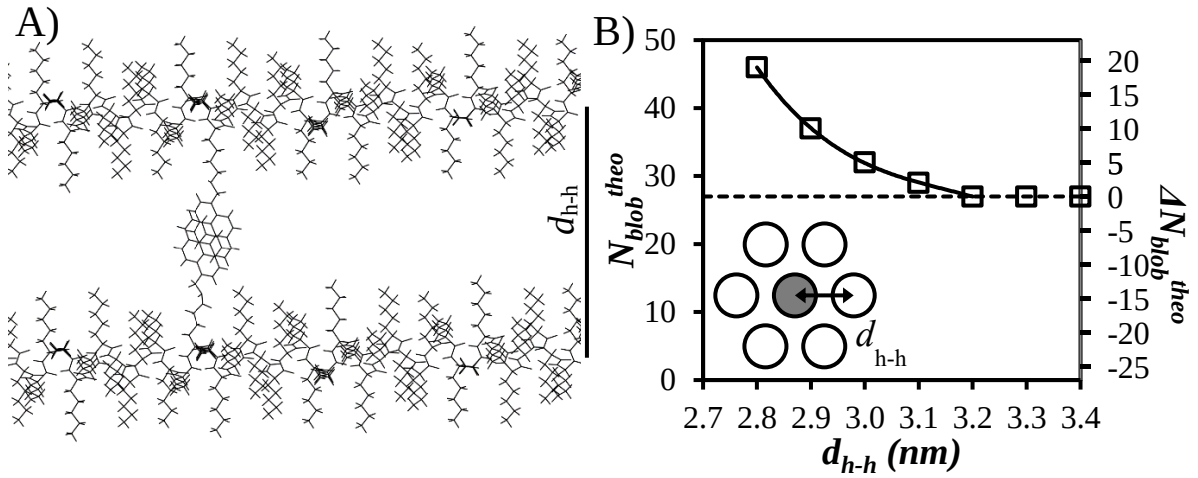


Figure 4.7. A) An example side-view of a MMO with good overlap between two pyrenes on adjacent P*L*L helices corresponding to an interhelical distance $d_{h\text{-}h}$ of 2.9 nm. B) N_{blob}^{theo} as a function of interhelical distance between Py-P*L*L helices in an HCP arrangement. ΔN_{blob}^{theo} represents the increase in N_{blob}^{theo} with respect to $N_{blob}^{theo} = 27$ for an isolated Py-P*L*L helix (- - -). The inlay depicts a simple head-on representation of the primary helix (grey) surrounded by six neighbouring helices (empty).

This process was repeated for all lysines on each of the neighbouring helices for a given $d_{h\text{-}h}$. A schematic of helices separated by $d_{h\text{-}h}$ in a HCP lattice is given as an inlay in Figure 4.7B. The number of pyrenes on the six neighbouring residues that could form excimer were then counted yielding ΔN_{blob}^{theo}. The determination of ΔN_{blob}^{theo} was repeated for $d_{h\text{-}h}$ values decreasing from 3.4 to 2.8 nm in 0.1 nm increments. Figure 4.7B displays ΔN_{blob}^{theo} and the total N_{blob}^{theo} as a function of $d_{h\text{-}h}$. The N_{blob}^{theo} value in Figure 4.7B is equal to the sum of

N_{blob}^{theo} = 27 for an isolated Py-P*L*L helix and ΔN_{blob}^{theo}. A detailed explanation of the MMO calculations is given in the SI.

Figure 4.7B shows that when the Py-P*L*L helices are more than 3.2 nm apart from one another in the HCP lattice, the Py-P*L*Ls are isolated. They are too far from one another to form excimer intermolecularly and can only form excimer intramolecularly resulting in the N_{blob}^{theo} value of 27. As $d_{h\text{-}h}$ decreased below 3.2 nm, the pyrene-labeled lysine side chains of neighbouring helices were close enough to one another to form excimer, resulting in the increase of N_{blob}^{theo} with decreasing $d_{h\text{-}h}$ shown in Figure 4.7B. To quantify the increase in N_{blob}^{theo} with respect to the interhelical distance, ΔN_{blob}^{theo} was fit with the empirical function given in Equation 4.3. The validity of Equation 4.3 holds true only within the bounds of $2.8 \leq d_{h\text{-}h} \leq 3.2$ nm.

$$\Delta N_{blob}^{theo} = -248.8 \times \left(d_{h-h} - 3.2\right)^3 - 34.8 \times \left(d_{h-h} - 3.2\right)^2 - 21.6 \times \left(d_{h-h} - 3.2\right) \quad (4.3)$$

The FBM analysis of the fluorescence decays acquired with the Py-P*L*L constructs yielded an N_{blob} value that increased from 25.2 ± 1.2 for an isolated Py-P*L*L α-helix to 36.6 ± 3.4 when the Py-P*L*Ls were bundled together, resulting in an experimental ΔN_{blob} of 11.4 ± 3.7. Equating the experimentally measured ΔN_{blob} with ΔN_{blob}^{theo} defined in Equation 4.3 yielded the interhelical distance of 2.9 ± 0.1 nm separating the bundled Py-P*L*Ls helices. This interhelical distance was consistent with the FRET studies which indicated that the helices must be within 5.8 nm of each other. It was also consistent with the interhelical distance of 2.5

± 0.1 nm found for bundles of α-helices of P*L*L obtained at pH 11.5 from 90:10 ethanol:water mixtures.[34] The slightly larger d_{h-h} value of 2.9 ± 0.1 nm found in this work might be a result of electrostatic repulsion between the P*L*L helices. Since the P*L*L bundles were found to have a diameter of 22.4 ± 3.7 nm by AFM, the calculated d_{h-h} value was used to determine that each bundle contained 55 ± 18 P*L*L helices across their cross section. Lastly, now that d_{h-h} was known, the P*L*L density (ρ) inside the bundles could be determined by applying Equation 4.4 representing the mass of P*L*L per HCP unit cell volume. In Equation 4.4, M_L = 164.63 g·mol^{-1} is the molar mass of a lysine residue protonated with HCl, N_A is Avogadro's number and Δh_L = 0.14 nm is the raise per lysine of a P*L*L α-helix. The derivation of Equation 4.4 is given in the SI.

$$\rho = \frac{2M_L}{\sqrt{3}N_A \Delta h_L d_{h-h}^{\ 2}} \tag{4.4}$$

With a d_{h-h} value of 2.9 nm obtained through the combination of PEF and MMOs, Equation 4.4 yielded a density of 0.25 ± 0.01 g·cm^{-3}. To put this density in perspective, it was compared to the density (ρ_P) of 41 proteins (see SI) which were calculated from the quantity 2.5/[η] where [η] is the intrinsic viscosity of the proteins.[50] While the density of the proteins listed in Table S4 took an average value of 0.6 ± 0.2 g/mL, the density of the P*L*L aggregates was found to match the lower ρ_P values found to equal 0.22 and 0.25 g/mL for α_{S2}-casein and clathrin, respectively. The similarity between the density of P*L*L aggregates and that of some

proteins suggests that the PEF experiments conducted with Py-P*L*L probe an environment similar to that encountered in some proteins. The density of the bundled P*L*L matched closely the 0.25 and 0.26 g·cm^{-3} internal densities of, respectively, second- and third-generation arborescent P*L*Glu's that were studied in DMF and were both viewed as macromolecules with a very dense interior.[28] It must be pointed out however that the protein densities measured by intrinsic viscosity appear to be half those measured by densimetry.[51] The discrepancy must be accounted for by the hydrodynamic diameter used in intrinsic viscosity measurements which is typically larger, due to solvation of the outer surface of the protein, than the volume of the protein core.

4.5 Discussion

The ACN:water mixtures employed in this study provided an experimental means to probe Py-P*L*L in different conformations, namely in the form of isolated random coils in a 60:40 ACN:water mixture with a 50-fold excess of unlabeled P*L*L, isolated α-helices in a 90:10 ACN:water mixture with a 50-fold excess of unlabeled P*L*L, and bundled α-helices in a 90:10 ACN:water mixture without unlabeled P*L*L. N_{blob} determined from the PEF experiments equaled 15.2 ± 2.1 for an isolated Py-P*L*L random coil, 25.2 ± 1.2 for an isolated Py-P*L*L α-helix, and 36.6 ± 3.4 for bundled Py-P*L*L α-helices. This steady increase in N_{blob} mirrored the increase in density expected for a polypeptide evolving along its folding pathway from a random coil to isolated secondary structures before ending as a structured protein with secondary motives spatially arranged in a well-defined 3D structure. The increase in density

experienced by a folding polypeptide results in an increased number of encounters between the amino acids diffusing inside the macromolecular volume defined by the polypeptide in solution, which is reflected by the increase in N_{blob}. Furthermore, the knowledge from CD measurements that P*L*L adopts a random coil and an α-helical conformation in, respectively, the 60:40 and 90:10 ACN:water mixtures,[31] enabled the validation of the experimental N_{blob} values by comparing them to the N_{blob}^{theo} values obtained by MMOs conducted on Py-P*L*L constructs with a same conformation. The good agreement between the experimental N_{blob} and theoretical N_{blob}^{theo} values obtained for a Py-P*L*L construct in a same conformation demonstrated the relevance of this approach based on a combination of PEF, FBM, and MMOs to probe the conformation of polypeptides in solution. It further supported similar conclusions which were drawn in earlier reports.[8,24-26,28,29]

Having used MMO to validate the experimental N_{blob} values obtained by PEF for known conformations of isolated Py-P*L*L constructs, N_{blob} determined for the Py-P*L*L bundles was then compared to that obtained by MMO conducted on an array of α-helical Py-P*L*Ls packed in a hexagonal arrangement as a function of the interhelix distance ($d_{h\text{-}h}$). N_{blob}^{theo} decreased from 46 to 27 when d_{h-h} increased from 2.8 to 3.2 nm, above which N_{blob}^{theo} remained constant and equal to 27, which is the N_{blob} value for an isolated Py-P*L*L α-helix. From the N_{blob}^{theo}-*vs*-$d_{h\text{-}h}$ trend obtained in Figure 4.7B, the N_{blob} of 36.6 ± 3.4 for bundled Py-P*L*L α-helices resulted in a $d_{h\text{-}h}$ value of 2.9 ± 0.1 nm. This $d_{h\text{-}h}$ value represented the missing length scale for the organization of the P*L*L α-helices at the molecular level in the P*L*L aggregates illustrated by the AFM pictures in Figure 4.5. Consequently, this study led to the conclusion

that α-helical P*L*Ls must be coming within 2.9 nm of each other to form 22 nm-wide bundles of helices in the 90:10 ACN:water mixture, before aggregating into much larger objects 1 – 2 μm in size. The 0.25 g/mL density of the P*L*L aggregates was also found to be representative of some proteins in solution.[50] The different conformations adopted by P*L*L as demonstrated by a combination of PEF experiments, FBM analysis, and MMOs are presented in Figure 4.8.

Isolated Coils

60:40 ACN:water with excess P*L*L

Isolated Helices

90:10 ACN:water with excess P*L*L

Bundles of helices

90:10 ACN:water without P*L*L

Figure 4.8. Depiction of the different conformations adopted by P*L*L under different experimental conditions as demonstrated by a combination of PEF, FBM, and MMOs.

The *blobs* used to compartmentalize the polypeptide volume based on the ability of pyrene labels to encounter in solution provide a means to characterize a finite volume within the macromolecular volume defined by the polypeptide describing how far the residues of a given polypeptide can reach in solution. As the polypeptide collapses onto itself to generate secondary structures that re-arrange themselves into the tertiary structure of a protein, the level of interactions between amino acids increases due to the shorter distances separating them,

resulting in the observed increase in N_{blob}. As was pointed out earlier,[7,8,29] the N_{blob} values that are recovered through these PEF experiments happen to be surprisingly similar to those expected from *foldons* used to characterize the cooperative folding of segments of a protein, and whose size has been reported to range between ~ 15 and 35 residues.[9] In fact, the N_{blob} values determined for Py-P*L*L found as an isolated random coil, an isolated α-helix, and an α-helix embedded in a bundle of Py-P*L*L α-helices equal 15.2 ± 2.1, 25.2 ± 1.2, and 36.6 ± 3.4, respectively, a range that matches remarkably well the expected range of *foldon* sizes. Of course, the N_{blob} value depends on the linker separating the pyrenyl label from the main chain, a longer linker allowing a larger reach and resulting in a larger N_{blob} value. A proper comparison between *blobs* and *foldons* requires accounting for the pyrene label and its linker, a correction which has not been applied yet for lack of sufficient data. It remains somewhat interesting, however, that the *blob*- and *foldon*-based procedures, which both aim at compartmentalizing a macromolecule into clusters of subvolumes, yield similar numbers of interacting structural units in these subvolumes when dealing with polypeptides. Nevertheless, the comparison between *blobs* and *foldons*, although interesting, should not distract from the fact that the concept of a *blob* is more general than that of the *foldon* since *blobs* can be defined for any macromolecule including structureless polypeptides such as intrinsically disordered proteins (IDPs),[52–54] which would be incapable of forming *foldons*. It must also be pointed out that compared to the HX experiments conducted on real proteins, the application of PEF experiments to aqueous solutions of pyrene-labeled proteins would certainly be challenging due to fluorescence quenching by some amino acids and pyrene aggregation in water.[42]

However, the PEF experiments described herein can nevertheless be applied to a wide range of macromolecular constructs that can be used as model systems to mimic the behaviour of more complex proteins. Such studies will no doubt be most helpful to sort out the underlying physical phenomena that control the folding of proteins and that might be impossible to clearly isolate with more complex proteins.

4.6 Conclusions

The conformations of P*L*L studied herein provided a snapshot of how the size of the subdomains within a protein where folding takes place may evolve at three stages along the folding pathway of a polypeptide. Py-P*L*L in its coiled conformation represented a protein in its fully denatured state and it was found to have a relatively small N_{blob}. The effect that the formation of structural motives, generated during the folding of a polypeptide, has on the size of a *blob* were investigated with the α-helical Py-P*L*L. The near doubling of the experimental N_{blob} value from 15.2 ± 2.1 for Py-P*L*L in a random coil conformation to 25.2 ± 1.2 as Py-P*L*L adopted an α-helical conformation suggested that the formation of structural motifs within a protein is accompanied by a significant increase in *foldon* size. Lastly, the bundled Py-P*L*Ls were used to mimic the packing of structural motifs in proteins in their native or near-native state. The interactions between the neighbouring Py-P*L*L helices resulted in a further increase in N_{blob} from 25.2 ± 1.2 to 36.6 ± 3.4. This increase in N_{blob} could be quantified to yield the interhelical distance between P*L*L α-helices bundled together in a filament which was found to equal 2.9 ± 0.1 nm, resulting in a density of 0.25 g/mL. This density was similar to that of

some proteins as determined by intrinsic viscosity measurements, leading to the conclusion that the PEF experiments with Py-P*L*L were capable of probing polypeptide environments similar to the interior of proteins.

Chapter 5

The Effect of Amino Acid Size on the Internal Dynamics of Polypeptides

5.1 Abstract

The fluorescence *blob* model (FBM) was applied to analyse the fluorescence decays of a series of pyrene-labeled polypeptides to better understand how the amino acid (*aa*) composition of a polypeptide affected its dynamics. Three pyrene-labeled polypeptides were prepared by copolymerizing racemic (*D,L*) mixtures of different *aa*'s, namely glycine (Gly), alanine (Ala), and carbobenzyloxylysine (Lys(Z)) with glutamic acid (Glu) to yield Py-PGlyGlu, Py-PAlaGlu, and Py-PLys(Z)Glu, respectively. All polypeptides contained 44 (± 3) mol% of Glu for fluorescence labeling with 1-pyrenemethylamine. The behavior of these three polypeptides was characterized in *N,N*-dimethylformamide (DMF) and dimethylsulfoxide (DMSO) by fluorescence. It was then compared with the behavior of pyrene-labeled poly(*D,L*-glutamic acid) (Py-PGlu) to assess the effect that each comonomer had on the backbone dynamics of the polypeptides. The FBM analysis of the fluorescence decays yielded the maximum number (N_{blob}) of residues separating two Glu's bearing a pyrenyl group while still allowing excimer formation between an excited and a ground-state pyrene. Py-PLys(Z)Glu yielded the same *blob* size (N_{blob} = 11) as for the Py-PGlu samples. In contrast, the incorporation of ~56 mol% Ala and Gly resulted in an increase in N_{blob} from 11 for Py-PGlu and Py-PLys(Z)Glu to 16 for Py-PAlaGlu and 23 for Py-PGlyGlu in DMSO. Considering that the internal dynamics of a polymer depend strongly on the size of its structural units and keeping proline aside, whose cyclic structure prevents backbone motion, this result implied that N_{blob} for the 17 largest *aa*'s of the 20 most common *aa*'s with 2 or more atoms in their side chain must take a value between 11 for PGlu and 16 for PAlaGlu. This rather narrow range of N_{blob} values suggests that the internal dynamics of polypeptides should be much simpler to predict than their structure since

only two *aa*'s, namely Gly and Ala, out of the 20 most common *aa*'s, appear to contribute differently to the internal dynamics of polypeptides.

5.2 Introduction

The function of many proteins is defined by their three-dimensional (3D) structure.[1,2] Therefore, the characterization of the structure and folding of proteins has been a main goal of protein science.[3] In this context, intense research has focused on better understanding the polypeptide dynamics involved in the folding leading to the 3D structure of proteins.[4-8] Polypeptide dynamics are also important in understanding the role of the unstructured segments, also referred to as intrinsically disordered regions (IDRs), that have been discovered in many proteins over the past few decades.[9–15] In fact, some functional proteins, referred to as intrinsically disordered proteins (IDPs), exhibit no regular structure under physiological conditions.[12–21] The rapidly increasing number of correlations uncovered between the function and dynamics of proteins indicates that the internal dynamics of proteins are important not only for controlling the folding pathway toward their final 3D-structure, but also in defining their function.[22–27] In the same manner that a protein sequence defines its 3D-structure, its sequence must also contribute to the internal dynamics of the protein.[27–30] This study thus investigates the effect that the *aa* composition of a polypeptide has on its internal dynamics.

The folding rate of a protein is often determined by spectroscopic techniques such as fluorescence,[31] UV/IR circular dichroism (CD),[32] NMR,[33] and electron paramagnetic resonance (EPR).[34] Although these techniques provide invaluable information on dynamics,

they are often conducted on real proteins. Consequently, such studies make it difficult to draw general conclusions on the effec, that a particular *aa* sequence might have on the internal dynamics of a protein since every protein is unique. To date, this effect has been studied primarily by monitoring the end-to-end cyclization (EEC) rate constant (k_{cy}) of fluorescently end-labeled oligopeptides as a function of sequence composition.[35-41] However, these EEC experiments are typically applied to short oligopeptides with a degree of polymerization under 25, due to the short distance range over which most fluorescence quenching mechanisms operate. Consequently, EEC experiments are not suited to study the longer polypeptides more representative of large proteins, whose two ends would be too far apart for quenching to occur.[42]

Interestingly, the short length scale, over which most quenching mechanisms occur, is taken advantage of by the fluorescence *blob* model (FBM), where *blobs* are used to compartmentalize a macromolecule randomly labeled with the dye pyrene (PyLM). The *blobs* are the loci for encounters between excited and ground-state pyrenes leading to pyrene excimer formation (PEF). According to the FBM, the PyLM can be viewed as a cluster of identical *blobs* among which the pyrenyl labels distribute themselves according to a Poisson distribution.[42–48] FBM analysis of the fluorescence decays of the PyLMs yields $\langle n \rangle$, the average number of ground-state pyrenes in the *blobs*, and k_{blob}, the rate constant describing the encounters between two pyrene-labeled structural units (SUs) inside a *blob*. In turn, $\langle n \rangle$ can be used to calculate N_{blob}, which represents the number of SUs contained inside one *blob*.[42,45,46] To date, the FBM has been applied to characterize the internal dynamics of numerous synthetic

polymers including randomly coiled pyrene-labeled polypeptides, namely poly(*D*,*L*-glutamic acid) (Py-P*DL*Glu) and poly(*L*-lysine) (Py-P*L*Lys).[48-50]

The FBM is applied in the present study to assess the effect that the *aa* composition of a polypeptide has on the dynamics of the polypeptide backbone. Since the internal dynamics of a linear polymer are expected to be strongly influenced by the molar mass of its constituting structural units in the absence of any other effect,[51,52] a series of *aa*'s were selected such that their side chain size (SCS) covered the range expected for most common *aa*'s. Glycine (Gly), with only hydrogens on its α-carbon, had the smallest SCS. The *aa* with the next smallest side chain was alanine (Ala), with a methyl group. The two other *aa*'s selected in this study were glutamic acid (Glu), since its side chain is comparable in size to many other *aa*'s, and *N*-ε-carbobenzyloxylysine (Lys(Z)) to represent the larger *aa*'s such as lysine or arginine, and less common *aa*'s such as pyrrolysine. Copolypeptides were prepared from racemic mixtures containing ~44 mol% of Glu and ~56 mol% of either Gly, Ala, or Lys(Z) to yield PGlyGlu, PAlaGlu, and PLys(Z)Glu, respectively. The racemic composition of the copolypeptides prevented secondary structure formation in solution, while the Glu residues provided a means for fluorescence labeling with 1-pyrenemethylamine. Maintaining the comonomer content close to 56 mol% enabled the characterization of the effects, that Gly, Ala, and Lys(Z) had on the chain dynamics of the polypeptides with respect to PGlu. This study illustrates how the FBM can be applied to assess the effect that different *aa*'s might have on the internal dynamics of polypeptides.

5.3 Experimental

Chemicals: Activated carbon (Sigma), *D*,*L*-alanine (Ala, Sigma, ≥ 99%), ammonium sulfate (Sigma, ≥ 99.0%), bis(trichloromethyl) carbonate (triphosgene, Oakwood Chemical, ≥ 99%), calcium hydride (CaH_2, Sigma, ≥ 95%), *N*-ε-carbobenzyloxy-*D*-lysine (H-*D*-Lys(Z)-OH, Bachem, 99.7%), *N*-ε-carbobenzyloxy-*L*-lysine (H-*L*-Lys(Z)-OH, Bachem, 99.5%), *N*,*N*′-diisopropylcarbodiimide (DIC, Sigma, 99 %), *N*,*N*-diisopropylethylamine (DIPEA, Sigma, 99.5%), dimethyl sulfoxide (DMSO, Sigma, ≥ 99.9%), *D*,*L*-glutamic acid 5-*tert*-butyl ester (H-*D*,*L*-Glu(OtBu)-OH, Chem-Impex Int'l Inc., > 98%), glycine (Gly, Sigma, ≥99%), hydrochloric acid (HCl, Fisher, 34 wt% in water), 1-hydroxybenzotriazole monohydrate (HOBt, Advanced ChemTech), (±)-α-pinene (Oakwood Chemical, 98%), 1-pyrenemethylamine hydrochloride (PyMA·HCl, Sigma, 95%), sodium bicarbonate (Sigma, ≥ 99.7%), and sodium hydroxide (NaOH, Sigma, ≥ 97.0%) were used as received. *n*-Butylamine (Sigma, 99.5%), *N*,*N*-dimethylformamide (DMF, Sigma, ≥ 99.8%), dioxane (≥ 99.5%), ethyl acetate (Sigma, ≥ 99.7%), and tetrahydrofuran (THF, Sigma, ≥ 99.0%) were dried according to procedures given in the SI before their use. Deionized water (DIW) was obtained from a Biopure Series 4400 Single Pass Reverse Osmosis system. *N*-Succinimidyl acetate (AcOSu) was synthesized according to a previously reported procedure.[48] N_2 gas (Praxair, N4.0) was used to keep the reaction mixtures under nitrogen atmosphere.

Synthesis of Glycine N-carboxyanhydride (GlyNCA): Glycine (8 g, 107 mmol, 1 eq) was dispersed in freshly dried THF (300 mL) with stirring in a three-necked round-bottom flask (500 mL), located in a fume hood, equipped with a condenser and an addition funnel, and maintained under nitrogen atmosphere. As the dispersion was heated to reflux, triphosgene (18

g, 61 mmol, 2/3 eq.), which had been in a sealed ampule due to its known toxicity, was weighed beforehand with a balance outside the fume hood before being dissolved in THF (60 mL) in an addition funnel. After the complete dissolution of triphosgene, α-pinene (20 mL, 125 mmol, 2.2 eq.) was added to the refluxing alanine suspension. The triphosgene solution was then added dropwise via the addition funnel over the course of ca. 2 hours and left to react for an additional hour. After the three-hour reaction, nitrogen was bubbled through the reaction mixture to evacuate any remaining phosgene. The yellow solution was filtered, condensed, and precipitated into heptane with vigorous stirring. After recovering the solid via suction filtration, the heterogeneous yellow and white solid was dispersed in ethyl acetate (50 mL) which dissolved the yellow masses, leaving only an off-white solid. The solid was collected and this process repeated which further reduced the yellow colour. Finally, the solid was dissolved in acetone and treated with activated carbon (ca. 2 g) before it was filtered and precipitated in petroleum ethers. After drying, GlyNCA was recovered as a white solid (5.98 g, 59 mmol, 55% yield) and used immediately in the polymerization described below. The ^{1}H NMR spectrum of GlyNCA is given in Figure S5.1 of the Supporting Information (SI).

Synthesis of D,L-Alanine N-carboxyanhydride (AlaNCA): *D,L*-Alanine (5 g, 56 mmol, 1 eq.) was dispersed in freshly dried ethyl acetate (200 mL) in an apparatus located in a fume hood and similar to the one used in the GlyNCA synthesis. α-Pinene (20 mL, 125 mmol, 2.2 eq.) was added into the refluxing alanine suspension. Triphosgene (11.1 g, 37 mmol, 2 eq. of phosgene) was dissolved in ethyl acetate (80 mL) taking the same precautions as for the GlyNCA synthesis before it was added dropwise over 2 hrs to the refluxing alanine dispersion.

After letting the mixture react for an additional hour, the transparent solution was removed from the heat and cooled before it was filtered to remove any unreacted alanine. The organic layer was then washed with an ice-cold bicarbonate solution (100 mL, 0.5 wt% $NaHCO_3$) and dried with sodium sulfate. The solution was then concentrated (ca. 30 mL) before it was added into heptane (ca. 300 mL) where AlaNCA was insoluble. The resulting heterogeneous mixture was vigorously stirred for 10 minutes which led to the formation of a white precipitate. The organic phase was then chilled overnight at –20 °C. The off-white solid was collected via suction filtration and dissolved in ethyl acetate (ca. 50 mL). Activated carbon (ca. 1 g) was added and left to stir for 15 minutes. The carbon was then removed by filtration and the clear solution was added into heptane (ca. 300 mL), stirred, and then chilled for 1 hour at –20 °C. The solid was collected and treated with activated carbon before precipitation into petroleum ethers. The activated carbon treatment was repeated until a bleach-white solid was obtained (typically 2 – 3 treatments in total). The solid was then dried under vacuum yielding AlaNCA (5.14 g, 45 mmol, 80% yield), which was used immediately in the copolymerization with Glu(OtBu)NCA. The ^{1}H NMR spectrum of AlaNCA is given in Figure S5.2 of SI.

Synthesis of D,L-Glutamic acid 5-tert-butyl ester N-carboxyanhydride (Glu(OtBu)NCA): The *tert*-butyl ester of *D,L*-glutamic acid (5 g, 25 mmol, 1 eq.) was converted into the NCA in a procedure similar to that of alanine using triphosgene (2.9 g, 10 mmol, 1.2 eq. of phosgene) and α-pinene (9 mL, 57 mmol, 2.3 eq.), while observing the same precautions when dealing with triphosgene. After the activated carbon treatment, the Glu(OtBu)NCA was precipitated into heptane rather than petroleum ether. The dried white solid (4.8 g, 21 mmol, 85% yield)

was stored at –20 °C under nitrogen atmosphere until needed. The ^{1}H NMR spectrum of Glu(OtBu)NCA is given in Figure S5.3 of SI.

Synthesis of N-ε-Carbobenzyloxy-L-lysine N-carboxyanhydride (L-Lys(Z)NCA): H-*L*-Lys(Z)-OH (5.1 g, 18 mmol, 1 eq.) was converted into the NCA using triphosgene (2.4 g, 8.1 mmol, 1.33 eq. of phosgene) and α-pinene (6.5 mL, 41 mmol, 2.3 eq.) following the procedure used for Glu(OtBu)NCA. A white solid (4.36 g, 14 mmol, 79% yield) was obtained. The dried white solid was stored at –20 °C under nitrogen atmosphere until needed. The ^{1}H NMR spectrum of *L*-Lys(Z)NCA is given in Figure S5.4 of SI.

Synthesis of N-ε-Carbobenzyloxy-D-lysine N-carboxyanhydride (D-Lys(Z)NCA): It was obtained as a white solid (4.94 g, 16 mmol, 92% yield) by applying the same procedure as for *L*-Lys(Z)NCA. The ^{1}H NMR spectrum of *D*-Lys(Z)NCA is shown in Figure S5.5 of SI.

Synthesis of Poly(glycine-co-D,L-glutamic acid 5-tert-butyl ester) (PGlyGlu(OtBu)): A 100 mL pear-shaped flask equipped with a spin vane was flamed under a flow of N_2 gas (50 $mL \cdot min^{-1}$) to remove any residual moisture. After cooling to room temperature, freshly dried and distilled DMF (50 mL) and dioxane (50 mL) were added and chilled with an ice bath. The freshly prepared GlyNCA (1.3 g, 13.0 mmol) and Glu(OtBu)NCA (2.3 g, 9.9 mmol) were dissolved in the DMF/dioxane mixture. The solution was cooled in an ice bath before freshly distilled *n*-butylamine (1.6 μL, 16.4 μmol, 1400:1 monomer:initiator) was added to initiate the polymerization. The same nitrogen flow of 50 $mL \cdot min^{-1}$ was maintained over the course of the polymerization to help eliminate the CO_2 released from the reaction. The conversion (*p*) of the polymerization was monitored by NMR. After 6 hours the reaction was stopped with the

addition of a drop of 1 M HCl solution at a conversion of ca. 0.14. PGlyGlu(OtBu) could not be precipitated in ether from the DMF/dioxane mixture, so instead the solution was added to water (200 mL) and the copolypeptide was salted-out with the addition of ammonium sulfate. The precipitated polypeptide was then collected by centrifugation (−20 °C, 10 min., relative centrifugal force (RCF) = 15k×*g*). After collection, PGlyGlu(OtBu) was precipitated from DMF (ca. 7 mL) into chilled ether (50 mL) three times to remove any remaining small-molecule impurities. After drying under vacuum, PGlyGlu(OtBu) was collected as a white power (0.4 g). Its ^{1}H NMR spectrum is given in Figure S5.6 in the SI.

Poly(D,L-alanine-co-D,L-glutamic acid 5-tert-butyl ester) (PAlaGlu(OtBu)): AlaNCA (1.3 g, 10.9 mmol) and Glu(OtBu)NCA (2.1 g, 9.2 mmol) were dissolved in a mixture of DMF (50 mL) and dioxane (20 mL) kept in an ice bath. The polymerization was initiated with *n*-butylamine (0.66 μL, 6.7 μmol, 3000:1 monomer:initiator). After 4 hours, the polymerization was stopped at ~0.11 conversion. The polypeptide was precipitated into chilled ether (4:1 ether:DMF), yielding a suspension of white solid in the ether solution. The suspension was centrifuged (−20 °C, 10 min., RCF = 15k×*g*) to collect the solid. The solid was then dissolved in DMF (10 mL) and passed through a PTFE filter (0.45 μm) to remove any insoluble matter. The filtered solution was added to chilled ether to reprecipitate the polypeptide. To ensure the complete removal of the unreacted monomer, the collected solid was dissolved into DMF (10 mL) and reprecipitated into chilled ether twice more. The precipitate was dried under vacuum overnight, yielding 0.3 g of PAlaGlu(OtBu). The NMR spectrum is given in Figure S5.7 in the SI.

Synthesis of poly(D,L-N_6-carbobenzyloxylysine-co-D,L-glutamic acid 5-tert-butyl ester) (PLys(Z)Glu(OtBu)): The polymerization was completed in a manner similar to that of PGlyGlu(OtBu). Glu(OtBu)NCA (2.1 g, 8.9 mmol), *D*-Lys(Z)NCA (2.2 g, 7.2 mmol), and *L*-Lys(Z)NCA (2.2 g, 7.2 mmol) was added with stirring to freshly dried and distilled DMF (90 mL) and the mixture was chilled with an ice bath. After dissolution, freshly distilled *n*-butylamine (1.2 μL, 12 μmol, 2000:1 monomer:initiator) was added to initiate the polymerization under a nitrogen atmosphere. After 12 hours, the polymerization had reached ca. 0.10 conversion, at which point the reaction was terminated. The polypeptide was then worked-up in a manner similar to PAlaGlu(OtBu). The resulting precipitate was dried under vacuum overnight yielding 0.55 g of PLys(Z)Glu(OtBu). The NMR spectrum is given in Figure S5.8 in the SI.

Polypeptide end-capping and deprotection: Before the Glu(OtBu) residues in the polypeptides were deprotected, the *N*-terminus was capped with an acetic group to prevent any reaction between the amine-end and the acid groups of the deprotected Glu. All the polypeptides were capped and deprotected in a similar manner. The NMR spectra of the polypeptides after deprotection are given in Figures 5.9 – 5.11 in the SI. An example procedure is given for PLys(Z)Glu(OtBu). PLys(Z)Glu(OtBu) (41 $kg \cdot mol^{-1}$, 0.55 g, 13 μmol) was dissolved in DMF (15 mL). After adding AcOSu (0.2 g, 1.3 mmol) and DIPEA (0.2 mL, 1.2 mmol), the solution was left to stir overnight. The next day, the capped PLys(Z)Glu(OtBu) was purified by dialysis using an 8 kDa regenerated cellulose dialysis tube. The polypeptide was dialyzed three times against DMF (400 mL) for 4 hours before being recovered by ether precipitation and

centrifugation. Once the amine end was capped, the white solid was dried under vacuum for three hours before it was dissolved in a mixture of trifluoroacetic acid (TFA, 7 mL) and water (0.7 mL) to deprotect the glutamic acid units. After two hours, PLys(Z)Glu was precipitated into ether (40 mL) and collected by centrifugation. The solid was then dried under vacuum for two hours to remove the majority of the remaining TFA. The solid was then purified by three precipitations from DMF (10 mL) into diethyl ether (40 mL). Each time PLys(Z)NCA was dissolved in DMF, 1 drop of 1 M HCl was added to ensure the complete removal of TFA. The white solid was again dried under vacuum to yield PLys(Z)Glu (37 $kg \cdot mol^{-1}$, 0.5 g, 13 μmol). The ^{1}H NMR spectrum of PLys(Z)Glu is provided in Figure S5.11 in SI.

Pyrene-labeling of the polypeptides: Pyrene-labeling of poly(glycine-*co*-*D*,*L*-glutamic acid), poly(*D*,*L*-alanine-*co*-*D*,*L*-glutamic acid), and poly(*D*,*L*-*N*-*ε*-carbobenzyloxylysine-*co*-*D*,*L*-glutamic acid) to yield Py-PGlyGlu, Py-PAlaGlu, and Py-PLys(Z)Glu, respectively, was conducted in a manner similar as in an earlier publication.[48] Some glutamic acids, representing a molar fraction *x* of the structural units (SUs) constituting the polypeptides, were labeled with 1-pyrenemethyl amine (PyMA). An example of the synthesis of Py-PGlyGlu (see chemical structure in Figure 5.1A) with a final pyrene content of 5 mol% is provided hereafter. PGlyGlu (0.1 g, 48 mol% Glu, equivalent to 1.1 mmol of structural units) was dissolved in DMSO (10 mL) with stirring. HOBt (12 mg, 76 μmol, 0.07 eq. to structural units), PyMA (17.5 mg, 65 μmol, 0.06 eq.), and DIPEA (distilled from CaH_2, 15.2 μL, 87 μmol, 0.08 eq.) were then added to the solution. After complete dissolution of the reagents, DIC (10.3 μL, 65 μmol, 0.06 eq.) was added and the solution was left to stir overnight. The next day, the resulting Py-PGlyGlu

was purified by dialysis (regenerated cellulose, 8 kDa cutoff). The solution was first dialyzed twice against DMF (600 mL) for four hours to remove the small molecules. The solution was then dialyzed against 2:1, 1:1, and 1:2 DMF:water solutions at pH 9. Finally, the solution was dialyzed twice against water (600 mL, pH 9). Once the dialysis was complete, the water was removed by lyophilization yielding the sodium salt of Py-PGlyGlu (Py-PGlyGNa) as a light-yellow solid. Each polypeptide was labeled with different pyrene contents according to a similar procedure where the reagent equivalents were adjusted for each of the target pyrene contents. The pyrene-labeled polypeptides were stored as their sodium salts in a –20 °C freezer until needed. They were converted into their protonated form before use with the addition of an excess of HCl. Their chemical structure is presented in Figure 5.1.

Pyrene-labeled poly(D,L-glutamic acid)s: Pyrene-labeled poly(*D,L*-glutamic acid)s (Py-PGlu's, DP = 784, $Đ$ = 1.06) were prepared as reported in a previous study.[48] The Py-PGlu's were stored in a –20 °C freezer and freshly lyophilized before use. The structure of Py-PGlu is given in Figure 5.1C.

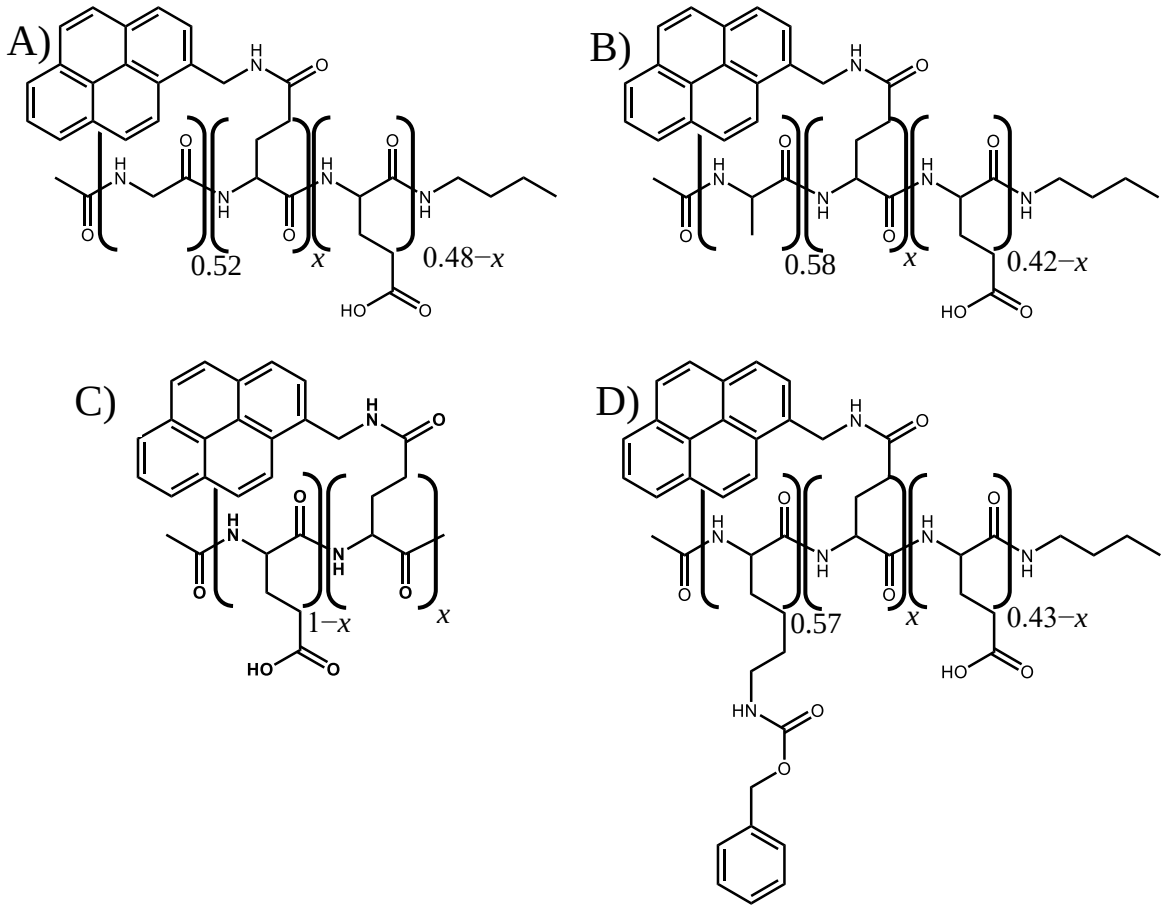


Figure 5.1. The chemical structures of A) Py-PGlyGlu, B) Py-PAlaGlu, C) Py-PGlu, and D) Py-PLys(Z)Glu. All stereocenters are racemic.

^{1}H NMR analysis: NMR spectra were acquired on a Bruker 300 MHz NMR spectrometer for polypeptide solutions in d_6-DMSO. The glutamic acid (Glu) content of the copolymers prepared with Ala and Lys(Z) was determined by comparing the integration of the *β*-protons of the two comonomers. Since Gly had no β-proton and its α-protons were shifted upfield, the α-protons of both Gly and Glu were used to determine the Glu content of PGlyGlu. The signal

of the *n*-butylamine initiator was observed for PGlyGlu and PLys(Z)Glu. The integration of the terminal methyl group was used to calculate the degree of polymerization (DP) of these polypeptides. Table 5.1 summarizes the comonomer content and DP of the polypeptides determined by NMR.

GPC analysis: The molecular weight distribution of the synthesized polypeptides was determined with a TOSOH 8321 GPC-WS GPC Workstation in conjunction with a Wyatt DAWN HELEOS multiangle light scattering (MALS) detector using two TSKgel Alpha-M 13 μm mixed bed columns in DMSO at 70 °C with a flow rate of 0.5 $mL \cdot min^{-1}$. Unfortunately, the d*n*/d*c* value of PGlyGlu(OtBu) was too low in DMSO to retrieve reliable results from the GPC analysis. Therefore, GPC analysis was only possible for the PAlaGlu(OtBu) and PLys(Z)Glu(OtBu) samples. The d*n*/d*c* values of PAlaGlu(OtBu) and PLys(Z)Glu(OtBu) were calculated through the ASTRA software package using the built-in 100% mass recovery method which yielded values of 0.0198 and 0.0438 $mL \cdot g^{-1}$, respectively. The GPC characterization was conducted on the protected polypeptides, since they provided better signal in the GPC detectors as compared to their deprotected counterparts.

The MALS detector array was used to calculate the absolute molecular weight of PAlaGlu(OtBu) (M_n = 29 $kg \cdot mol^{-1}$ (DP = 240), $Đ$ = 1.05) and PLys(Z)Glu(OtBu) (M_n = 41 $kg \cdot mol^{-1}$ (DP = 180), $Đ$ = 1.04). Although the GPC traces indicated that the polymers contained some low-molecular weight species (See Figure S5.12 in the SI), the analysis of the GPC traces described in SI indicated that they only contributed a small 0.13 and 0.05 weight fraction of the overall polymers for PAlaGlu(OtBu) and PLys(Z)Glu(OtBu), respectively. Due

to their relatively small contributions, these lower molecular weight species did not have much effect on the molecular weight or dispersity of the polymer, and as such the M_n and $Đ$ reported in Table 5.1 are characteristic of the main peak. It is worth pointing out that the polydispersity of polymer samples is irrelevant for the study at hand which is conducted with the FBM, specifically designed to handle polydisperse polymer samples.[44-46] Table 5.1 summarizes the molecular weight of the two polypeptides characterized by GPC analysis. The M_n's in Table 5.1 are the molecular weights of the polymers in their deprotected form.

Table 5.1. Glu content, number-average molecular weight (M_n), degree of polymerization (DP), and dispersity ($Đ$) of the deprotected polypeptides.

Polypeptide	Monomer Feed (mol% Glu)	Polypeptide Composition (mol% Glu)	M_n ($kg \cdot mol^{-1}$)	*DP*	*Đ*
PGlyGlu	43	48	14[a]	150[a]	-
PAlaGlu	46	42	23	240	1.05
PGlu[b]	-	100	100	784	1.07
PLys(Z)Glu	39	43	37	180, 176[a]	1.04

[a] Determined by NMR.

[b] Information provided by the certificate of analysis from Alamanda Polymers.

UV-Vis absorption measurements: The UV-Vis absorption spectra were acquired with a Varian Cary 100 Bio spectrophotometer using a UV-Vis absorption cell with a 1.0 cm pathlength. A background correction for the solvent was applied to each spectrum.

Pyrene content of the polypeptides: The molar fraction (x) of the structural units labeled with pyrene was determined with Equation 5.1. Where, y is the molar fraction of glutamic acid incorporated in the polypeptide determined by NMR in Table 5.1. M_Z is the molar mass of the comonomer (M_Z: M_{Gly} = 57.05, M_{Ala} = 71.08, $M_{Lys(Z)}$ = 262.31 g·mol^{-1}), M_{GNa} (= 144.15 g·mol^{-1}) is the molar mass of the sodium salt of glutamic acid, and M_{Py} (= 342.4 g·mol^{-1}) is the molar mass of a glutamic acid-bearing pyrene. Lastly, λ_{Py} equaled the moles of pyrene per gram of sample. To determine λ_{Py}, at least 5 mg of the lyophilized pyrene-labeled polypeptide was added to 5 mL of DMSO. A drop (0.1 mL) of 1 M HCl in water was added to the solution to protonate the glutamic acid residues. After sonicating for 15 minutes, the solution or dilution thereof was transferred to a 1.0 cm pathlength cell for UV absorption measurements. Using the Beer-Lambert law, the molar concentration of pyrene was determined using the absorbance peak maximum of pyrene at 346.5 nm and the molar absorptivity coefficient (ε) of the model compound *N*-(1-pyrenylmethyl)acetamide (PyMAAc) in DMSO (ε = 39,300 M^{-1}·cm^{-1}). The determination of ε at 346.5 nm is described in the SI.

$$x = \frac{(1-y)M_Z + yM_{GNa}}{\lambda_{Py}^{-1} + M_{GNa} - M_{Py}} \qquad (5.1)$$

Steady-state fluorescence (SSF): Steady-state fluorescence spectra were acquired on a Horiba QM-400 spectrofluorometer equipped with a xenon arc lamp. A 10 μL aliquot of 1 M HCl was added to the solutions containing ca. 2.7×10^{-6} M pyrene to ensure that the pyrene-labeled polypeptides were studied in their protonated form. After sonicating the samples for ca. 10

minutes, oxygen was removed by outgassing the samples with a gentle flow of N_2 gas (Praxair, N4.0 or greater) for ca. 30 minutes before sealing the fluorescence cell and acquiring the SSF spectrum. Using excitation and emission slit widths of 1 nm, the emission of pyrene was monitored by exciting the sample at 344 nm and scanning the emission wavelength from 350 to 600 nm in 1 nm steps using a scanning rate of 10 $nm \cdot s^{-1}$. The fluorescence intensity of the monomer (I_M) and excimer (I_E) were calculated using the area under the SSF spectrum from 375 to 381 nm and from 500 to 530 nm, respectively.

Time-resolved fluorescence (TRF): The pyrene-labeled polypeptide solutions were prepared in a similar manner as for the SSF measurements. The fluorescence decays of pyrene were acquired using an IBH time-correlated single photon counting (TC-SPC) fluorometer equipped with a NanoLED-340 laser. All decays were collected using an excitation wavelength of 344 nm selected with the excitation monochromator over 1,024 channels. To minimize light scattering from reaching the detector, a 370 nm longpass filter was employed to collect the monomer fluorescence at 378 nm with a time-per-channel of 2.04 $ns \cdot ch^{-1}$ until a minimum of 40,000 counts was obtained at the fluorescence decay maximum. Similarly, a 470 or 495 nm longpass filter was applied to the collection at 510 nm of the excimer fluorescence decays with at least 20,000 counts at the decay maximum using a time-per-channel of 1.02 $ns \cdot ch^{-1}$. The instrument response function (IRF) was determined by collecting the light scattered at 344 nm by a LUDOX dispersion in water with a peak maximum of at least 20,000 counts. The IRF was used in the deconvolution of the fluorescence decays for analysis.

Fluorescence decay analysis: The monomer and excimer decays of the pyrene-labeled polypeptides were fit globally according to the FBM using Equations S5.2 – 5.5. Within the framework of the FBM, a pyrene-labeled macromolecule is divided into a string of subvolumes referred to as *blobs*.[43] The volume of a *blob* is defined by the volume that an excited pyrene bound to the macromolecule can probe while it remains excited. Since pyrene is randomly attached to the macromolecule, the pyrenes distribute themselves among the *blobs* according to a Poisson distribution. Within a *blob*, an excited and a ground-state pyrene form excimer through a sequential pathway. First, an excited pyrene *Py**, referred to as Py_{diff}*, undergoes slow diffusive motions that are controlled by the diffusion of the pyrene-labeled SU. This displacement brings Py_{diff}* close to a ground-state pyrene (*Py*) with a rate constant k_{blob}. When Py_{diff}* and *Py* are within reach of each other, Py_{diff}* turns into the species Py_{k2}* which rapidly rearranges with a rate constant k_2 to form with the ground-state *Py* one of two excimers, *E0** or *D**, which emit with a lifetime τ_{E0} or τ_D, respectively. In contrast, the pyrene monomers Py_{diff}* and Py_{k2}* emit with their natural lifetime τ_M. The global fit of the monomer and excimer fluorescence decays with the FBM equations yields <*n*>, the average number of ground-state pyrenes within a *blob*. <*n*> is then used along with the molar fraction (*x*) of pyrene-labeled SUs in the macromolecule to determine N_{blob} according to Equation 5.2, which is the average number of SUs inside a *blob*. In Equation 5.2, f_{Mfree} represents the molar fraction of pyrenes Py_{free}* in the monomer decay that do not form excimer and emit with their lifetime τ_M.

$$N_{blob} = \frac{< n >}{x}\left(1 - f_{Mfree}\right) \qquad (5.2)$$

The monomer and excimer decays were fitted globally with Equations S5.2 – S5.4 in the SI, whose parameters were optimized according to the Marquardt-Levenberg algorithm.[53] The fits were deemed satisfactory when the χ^2 parameter was smaller than 1.3 and the residuals and autocorrelation of the residuals were randomly distributed around zero. An example of the fits is given as Figure S5.16 in the SI. The parameters retrieved from the FBM analysis of all the pyrene-labeled polypeptides are given in Tables S5.2 – 5.7. More information on the FBM can be found in earlier reviews.[44-46]

5.4 Results

Polypeptide Synthesis and Characterization: All copolymers were prepared through the ring-opening polymerization of the corresponding *aa* NCAs using *n*-butylamine as the initiator. The copolymerizations were stopped at conversions under 15% to minimize monomer composition drift. Although *aa* NCAs are known to have quite different reactivity ratios,[54,55] maintaining a low conversion in *co*- and *ter*-polymers prepared from *aa* NCAs has been shown to yield random compositions.[56] Therefore, both the low conversion and the similarity between the feed and copolymer compositions seen in Table 5.1 indicated that the monomers were incorporated randomly into the polypeptides studied herein.

Steady-State Fluorescence: The steady-state fluorescence (SSF) spectra of the pyrene-labeled polypeptides were collected in DMSO and DMF. An example of the SSF spectra is shown for the series of Py-PLys(Z)Glu samples in Figure 5.2A, where the pyrene content was varied from 4.5 to 13.5 mol%. The SSF spectra of the synthesized pyrene-labeled polypeptides in both

DMF and DMSO are provided in Figures S5.17 – S5.21 in SI. Figure 5.2A illustrates how PEF increased with increasing pyrene content (x) as a result of more pyrenyl encounters. The increase in excimer emission was quantified with the I_E/I_M ratio which was plotted as a function of pyrene content in Figures 5.2B and C for the solutions in DMSO and DMF, respectively.

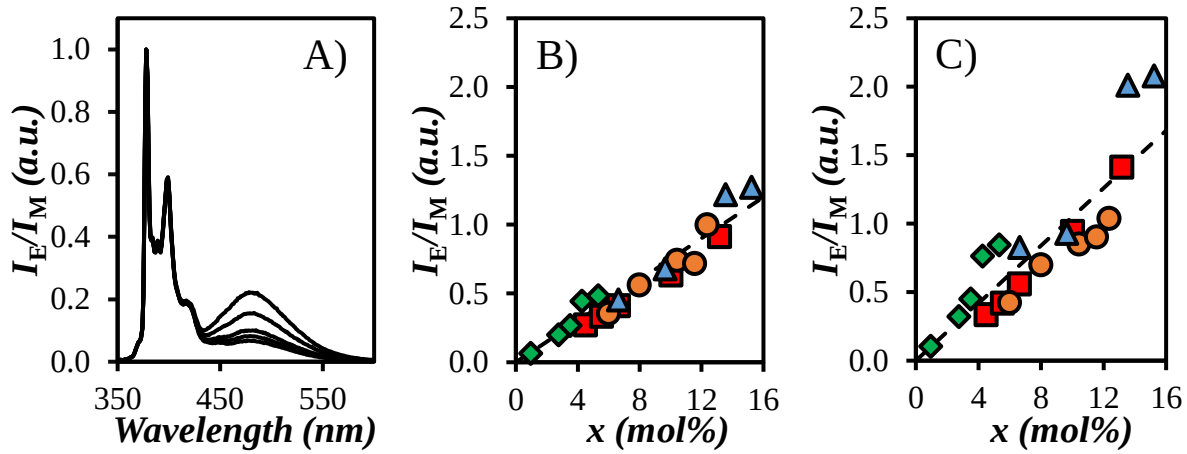


Figure 5.2. A) Steady-state spectra of the Py-PLys(Z)Glu samples in DMSO and the I_E/I_M ratios of for (◆) Py-PGlyGlu, (▲) Py-PAlaGlu, (●) Py-PGlu, and (■) Py-Lys(Z)Glu in B) DMSO and C) DMF. The dashed lines were added to guide the eyes. Error on data points in Figures 5.2B and 5.2C are smaller than the symbols.

Figures 5.2B and C show that most I_E/I_M ratios were not much larger in DMF (η = 0.784 mPa.s at 25 °C) than in DMSO (η = 1.987 mPa.s at 25 °C) despite the 2.5-fold viscosity difference between the two solvents. The similarity in the I_E/I_M ratios most certainly originated from the difference in the efficiency of excimer formation between the two solvents. A previous study showed that excimer formation for 1-pyrenemethylacetamide is about twice as

efficient in DMSO compared to DMF.[48] Since the I_E/I_M ratio is proportional to the efficiency of excimer formation[57] and inversely proportional to solvent viscosity,[58] the decrease in viscosity between DMSO and DMF was offset by the decrease in the efficiency of excimer formation, resulting in the similar I_E/I_M ratios between the two solvents.

Another interesting feature of the I_E/I_M ratios was that they did not reveal much difference between different polypeptides in a given solvent. This observation was unlike that made with the I_E/I_M ratios obtained with a series of pyrene-labeled poly(*n*-alkyl methacrylate)s (Py-PAMAs) where side chain size (SCS) was found to have a major effect on the I_E/I_M ratio, with smaller side chains resulting in more efficient excimer formation and higher I_E/I_M ratios.[52] This effect was primarily attributed to the progressive stiffening of the polymethacrylate backbone as SCS increased due to an increase in steric interactions. However, Py-PAMAs with side chains longer than 12 carbons yielded overlapping I_E/I_M ratios, suggesting that the I_E/I_M ratio did not change much once the backbone became too stiff. The similar I_E/I_M ratios obtained with the pyrene-labeled polypeptides suggest that the SCS does not influence the polypeptide dynamics to the same extent as for PAMAs, likely due to the inherent stiffness of the polypeptide backbone and the 44 mol% content of the bulkier Glu in the polypeptides, which dampened the effect that the smaller *aa*'s might have had.

Effect of Amino Acid Side Chain Size on the Blob Size: One of the most important parameters retrieved from the FBM is N_{blob} (see Equation 5.2), which is the number of SUs inside a *blob*. Since pyrene is covalently bound to the polypeptide backbone through a glutamic acid residue, N_{blob} represents the number of *aa*'s each glutamic acid can interact with on the timescale of the

polypeptide dynamics. Figures 5.3A and B show that N_{blob} remained constant as a function of pyrene content for each polypeptide in DMSO and DMF, respectively. Since the increased addition of bulkier pyrenyl labels to the polypeptides did not affect N_{blob}, the low level of pyrene labeling used in these experiments was not expected to perturb the internal dynamics of the polypeptides.

The effect induced by a given side chain was monitored as a function of its side chain size (SCS) defined in terms of the number of atoms contained within the side chain structure, excluding hydrogens. The SCSs were equal to 0 for glycine (Gly), 1 for alanine (Ala), 5 for glutamic acid (Glu), and 15 for carboxybenzyl protected lysine (Lys(Z)). Quantification of SCS in this manner allowed for a simple comparison between the comonomers incorporated into the polypeptides in Figure 5.3C. Figure 5.3C shows that within error limits, N_{blob} for Py-PGlu and Py-PLys(Z)Glu took a similar average value of 11.0 ± 1.4 and 11.5 ± 1.7 in DMSO and DMF, respectively. These experimental N_{blob} (N_{blob}^{exp}) values were similar to the theoretical N_{blob} (N_{blob}^{theo}) value of 10.3 ± 1.2 obtained by molecular mechanics optimizations (MMOs) for an extended Py-PGlu construct representative of a PGlu backbone in a random coil conformation.[48] These similar N_{blob} values suggested that Py-PGlu and Py-PLys(Z)Glu were fairly extended on the length scale of a *blob*.

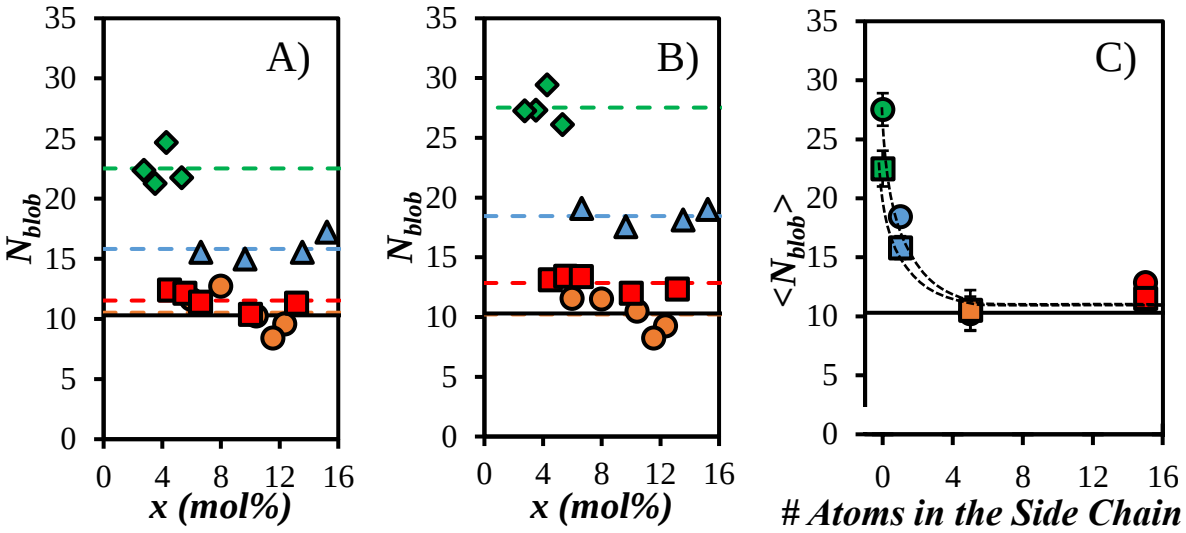


Figure 5.3. Plot of N_{blob} as a function of pyrene content for (◆) Py-PGlyGlu, (▲) Py-PAlaGlu, (●) Py-PGlu, and (■) Py-Lys(Z)Glu in A) DMSO and B) DMF. C) Comparison of the average N_{blob} values as a function of the number of non-hydrogen side chain atoms in (□) DMSO and (○) DMF. The dashed lines in Figures A, B) represent the averaged N_{blob} value. The solid line indicates the N_{blob}^{theo} value of 10.3 ± 1.2 for a rigid and extended polypeptide assumed to adopt a coiled conformation.

The first implication of this result was that the relatively small side chain of Glu was sufficient to fully stiffen the polypeptide backbone on the length scale of a *blob*, thereby making further increases in SCS irrelevant to the *blob* size. The second implication was that the eight *aa*'s out of the 20 most common *aa*'s with SCS equal to or larger than that of Glu should not alter the number of *aa*'s, which can locally interact with one another via diffusive encounters only, without specific interactions such as H-bonding or electrostatics. Under such

conditions, N_{blob} for an oligopeptide segment in a protein is expected to equal 11 if this segment is composed of *aa*'s bearing a side chain of size equal to or larger than the SCS of 5 for Glu.

Next, the effect that the smaller side chains of Ala and Gly had on N_{blob} was explored. Figure 5.3C shows that the N_{blob} values obtained for Py-PAlaGlu and Py-GlyGlu were both significantly larger than those obtained for Py-PGlu and Py-PLys(Z)Glu. The increase in N_{blob} from 11.3 ± 1.2 for Py-PGlu and Py-PLys(Z)Glu to 15.8 ± 1.0 for Py-PAlaGlu and finally 22.5 ± 1.5 for Py-PGlyGlu in DMSO indicated that *blobs* with Ala and Gly residues having smaller side chains were constituted of a larger number of *aa*'s. To ensure that this increase in N_{blob} for Py-PAlaGlu and Py-PGlyGlu was not due to any inherent effect induced by Gly and Ala, MMOs were conducted for PGlyGlu. Following a procedure which has been described numerous times,[47–50] N_{blob}^{theo} was found to equal 10.3 ± 1.2 for PGlyGlu, which matched the N_{blob}^{theo} value of 10.3 ± 1.2, that had been obtained for PGlu.[48] A full description of the MMO procedure is provided in the SI. The matching N_{blob}^{theo} values ruled out the possibility that the increase in N_{blob}^{exp} could be due to differences in the monomer/backbone structures that could have been induced by the introduction of Gly or Ala. It should also be noted that N_{blob}^{theo} is representative only of a backbone which is extended on the length scale of a *blob* (ca. a few nm)[48,49] and should only be used as a lower limiting value, since the backbone in solution may not be as extended on this length scale.

Since the increase in N_{blob}^{exp} for Py-PAlaGlu and Py-PGlyGlu could not be simply explained through MMOs, it must have been due to the increased freedom brought to the polypeptide backbone by Gly and Ala. As for the previous study on Py-PAMAs,[52] the decrease

in SCS resulted in greater freedom of the backbone due to the reduction of sterics imparted by the side chains. Two explanations can be suggested to rationalize this increase in N_{blob}^{exp}. The first explanation is that the low energy barrier for bond rotation enabled by the small side chains of Gly and Ala enhanced backbone mobility, which would in turn allow the excited pyrenyl pendants to probe a larger volume, as was proposed for the relatively flexible PAMAs.[52] The increase in the *blob* volume would thus result in a larger N_{blob}^{exp} value than N_{blob}^{theo} determined by MMOs. The second explanation would be that the incorporation of SUs with a smaller side chain would bring significant conformational freedom to the polypeptide backbone. The reduction in steric hinderance generated by the smaller side chains, particularly in the case of Gly which has only hydrogens on its α-carbon, provided greater rotational freedom between the *aa*'s in the backbone. This increased freedom may then allow the polypeptide backbone to adopt more conformations within a *blob*, resulting in the increased number of *aa*'s which can locally interact with one another as reported by N_{blob}^{exp}.

This second possibility is likely the major contributor to the increased N_{blob}^{exp} value for the polypeptides. Unlike PAMAs, the backbone of polypeptides is fairly rigid, which makes it unlikely that the increased backbone mobility induced by a shorter side chain could justify the 2-fold increase in N_{blob} from 11.3 for PGlu and PLys(Z)Glu to 22.5 for PGlyGlu. In contrast, a 3.6-fold increase in N_{blob} from 16.3 for poly(octadecyl methacrylate) to 58.6 for poly(methyl methacrylate) was reported for the much more flexible PAMAs in THF.[52] In addition, previous studies have indicated that the inclusion of small amino acids, in particular Gly, results in an increase in the local backbone density based on the end-to-end distance[59] or the C_∞ coefficient

of Gly-containing oligopeptides.[60−65] This again suggests that the increase in N_{blob}^{exp} in Figure 5.3 is the result of an increase in backbone conformational freedom.

The N_{blob}^{exp} values of Py-PAlaGlu and Py-PGlyGlu were also determined in DMF were they were found to equal 18.4 ± 0.7 and 27.5 ± 1.4, respectively. While the N_{blob}^{exp} value of 18.4 ± 0.7 for Py-PAlaGlu in DMF was comparable to that of 15.8 ± 1.0 for Py-PAlaGlu in DMSO, the N_{blob}^{exp} value of 27.5 ± 1.4 for Py-GlyGlu in DMF was substantially larger than that of 22.5 ± 1.5 for Py-PGlyGlu in DMSO. The larger N_{blob}^{exp} value obtained for the Py-PGlyGlu series in DMF was attributed to DMSO being a much better solvent toward PGlyGlu than DMF. Indeed, PGlyGlu could be dissolved in DMSO up to a concentration of 7.1 mg/mL, which enabled the measurement of its intrinsic viscosity (see Figure S5.22 in SI). In contrast, PGlyGlu was insoluble in DMF at a concentration as low as 1.5 mg/mL. To assess whether the larger N_{blob} value of 27.5 ± 1.4 for Py-GlyGlu in DMF was solely a result of poor solvent quality and not polymer aggregation, the I_E/I_M ratios of PGlyGlu were measured as a function of polymer concentration in DMF and DMSO. Figure S5.23 in the SI showed that the I_E/I_M ratios of PGlyGlu remained constant as a function of concentration. The constant I_E/I_M ratios demonstrated that the larger N_{blob} for PGlyGlu in DMF could not be a result of polymer aggregation and was solely a consequence of poor solvent quality.

Encounter Rate Constant within a Blob: The FBM analysis also retrieves the rate constant (k_{blob}) describing the diffusive encounters between two *aa*'s bearing one excited and one ground-state pyrene located inside a polypeptide *blob*. The k_{blob} values retrieved from the FBM analysis for the pyrene-labeled polypeptides in DMSO and DMF are given in Figures 5.4A and

B, respectively. Figures 5.4A and B show that for each pyrene-labeled polypeptide, k_{blob} remained constant within experimental error as a function of pyrene content.

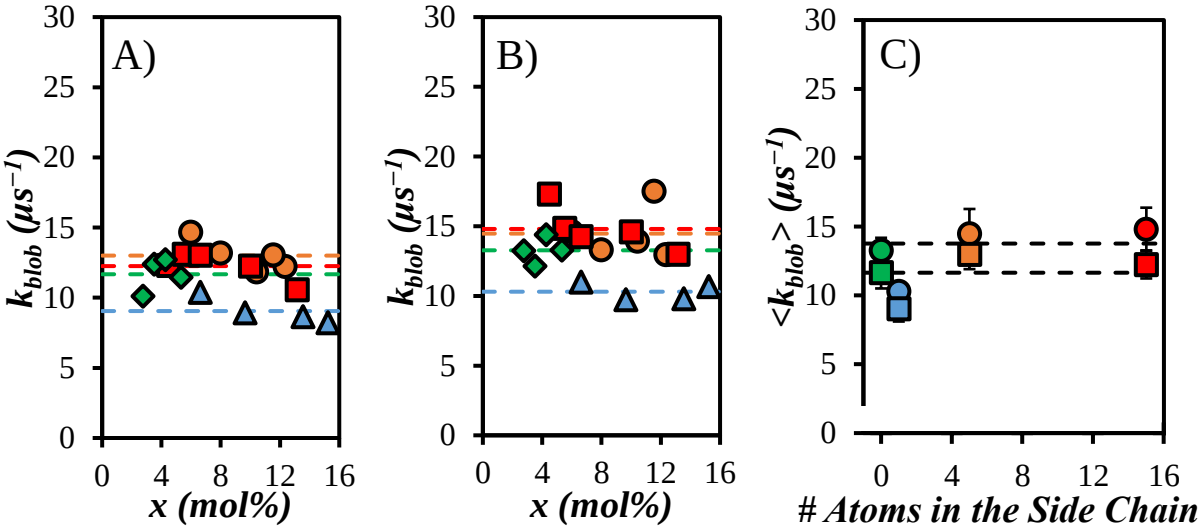


Figure 5.4. Plot of k_{blob} as a function of pyrene content for (◆) Py-PGlyGlu, (▲) Py-PAlaGlu, (●) Py-PGlu, and (■) Py-Lys(Z)Glu in A) DMSO and B) DMF. C) Comparison of the average k_{blob} values as a function of the copolymer side chain size in (□) DMSO and (○) DMF. The dashed lines represent the averages of the k_{blob} values obtained for a same series of polypeptide in Figures 5.4A and 4B and the averages of the k_{blob} values obtained in (bottom) DMSO and (top) DMF in Figure 5.4C.

Unlike N_{blob}^{exp} however, k_{blob} showed relatively little change as a function of SCS for all the pyrene-labeled polypeptides in Figure 5.4C, taking an average value of 11.6 ± 1.8 and 13.4 ± 2.2 μs^{-1} in DMSO and DMF, respectively. The fact that k_{blob} remained constant despite the increase in N_{blob}^{exp} for the smaller *aa*'s was an interesting result. k_{blob} is equal to the product

$k_{diff}\times(1/V_{blob})$ where k_{diff} is the bimolecular rate constant for diffusive encounters within a *blob* and $1/V_{blob}$ is the concentration equivalent to one pyrene-labeled *aa* inside a *blob*.[44] Therefore, a constant k_{blob} value implies that if the increase in N_{blob}^{exp} led to an increase in V_{blob} due to the enhancement in backbone mobility, k_{diff} must have increased in the exact same proportion to yield the fairly constant k_{blob} values observed in Figure 5.4C. Such a coincidence would be highly unlikely. A much likelier possibility is that both k_{diff} and V_{blob} are constant and that the increase in N_{blob}^{exp} inside the constant V_{blob} is solely due to increased conformational freedom of the backbone. In turn, k_{diff} would be independent of the *aa* composition, suggesting that the diffusive rate constant of *aa*'s within a *blob* is unaffected by changes in the *aa* composition of a *blob*, probably because the polypeptide backbone is too rigid.

Polypeptide Internal Dynamics: Good agreement between N_{blob}^{exp} and N_{blob}^{theo} values indicates that the segments of a polymer backbone within a *blob* remain immobile over the time scale of PEF (~1 μs). This condition has been observed when polypeptides form secondary structures such as P*L*Glu in DMF and DMSO, which adopts α- and 3_{10}-helical conformations,[48,49] respectively, α-helical P*L*Lys.HCl in a 90:10 acetonitrile:water mixture,[50] or when the polypeptides are extended such as randomly coiled PLys(Z)Glu in a 60:40 acetonitrile:water mixture[50] and P*DL*Glu[48,49] in DMF and DMSO. For these rigid backbones, V_{blob} can be assumed to remain constant and therefore k_{blob} directly reflects the dynamics of the SUs within a *blob*. However, in cases where the constancy of V_{blob} is not certain, the term $k_{blob}\times N_{blob}$ has been shown to provide a more accurate representation of the overall polymer dynamics since it includes both the information on *blob* size (N_{blob}) and the rate of encounters within a *blob*

(k_{blob}).[52,66] In this respect, the $k_{blob}\times N_{blob}$ product represents the number of contact events between SUs located inside a *blob* per unit time. A plot of $\langle k_{blob}\times N_{blob}\rangle$ averaged over all pyrene contents for a given series of pyrene-labeled polypeptides is given in Figure 5.5 as a function of SCS.

The trend of $\langle k_{blob}\times N_{blob}\rangle$-*vs*-SCS shown in Figure 5.5 indicates that the polypeptide segments inside the *blobs* experience slower internal dynamics as the SCS increases, with $\langle k_{blob}\times N_{blob}\rangle$ quickly reaching the limiting values of 0.14 ± 0.02 and 0.18 ± 0.03 ns^{-1} in DMSO and DMF, respectively. One reason for the slightly larger $\langle k_{blob}\times N_{blob}\rangle$ values obtained in DMF compared to DMSO is most certainly a consequence of the interplay between the 2.5-fold decrease in k_{blob} with increased viscosity from DMF to DMSO and the 1.5-fold increase in k_{blob}, with the higher probability of forming an excimer upon encounter between two pyrenyl labels in DMSO than in DMF.[48]

The trends shown in Figure 5.5 are also in agreement with end-to-end cyclization (EEC) studies of oligopeptides.[38,39,59,67] These studies found that k_{cy}, the rate constant for EEC, remained relatively constant for most amino acids, except when Gly or Pro were incorporated into an oligopeptide sequence. The presence of Pro decreased k_{cy}, while incorporating Gly increased k_{cy} by about two-fold for an oligopeptide of the same length.[38,59] This result agreed remarkably well with the $\langle k_{blob}\times N_{blob}\rangle$ values obtained, which demonstrated that there was an about two-fold increase in $\langle k_{blob}\times N_{blob}\rangle$ for PGlyGlu as compared to the other copolypeptides.

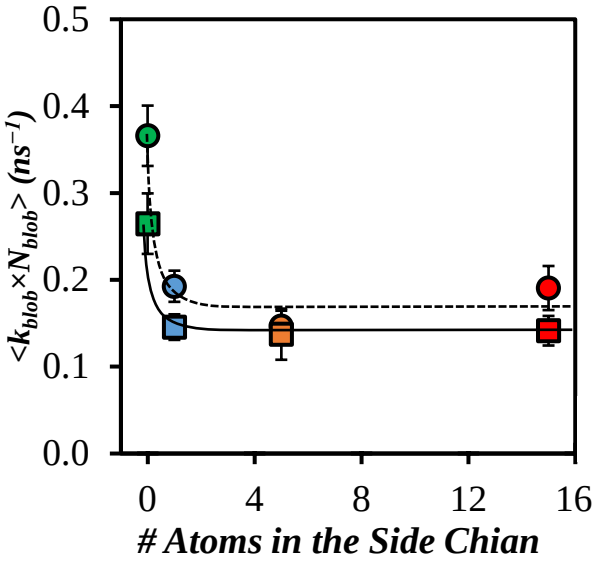


Figure 5.5. Plot of $<k_{blob}\times N_{blob}>$ as a function of the comonomer side chain size in (□, solid line) DMSO and (○, dashed line) DMF.

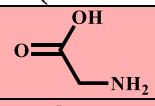
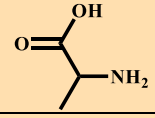
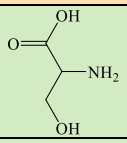
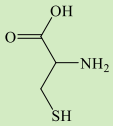
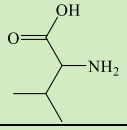
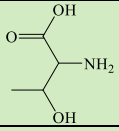
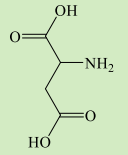
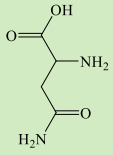
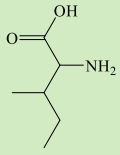
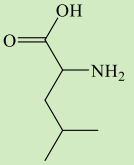
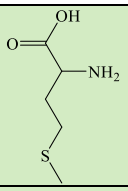
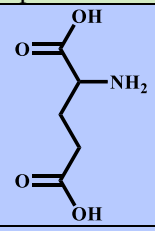
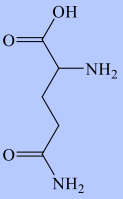
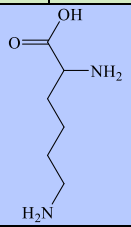
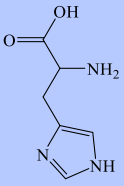
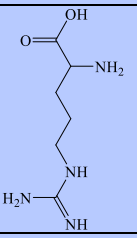
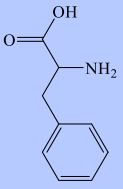
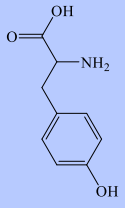
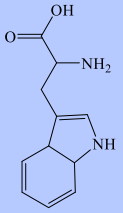
Interestingly, the plot of $<k_{blob}\times N_{blob}>$ as a function of SCS in Figure 5.5 was somewhat similar to that obtained for the PAMA series, for which $<k_{blob}\times N_{blob}>$ decreased with increasing SCS, until it reached a plateau for SCSs greater than 15 atoms corresponding to a dodecyl side chain.[52] In both cases, the plateau in $<k_{blob}\times N_{blob}>$ indicated that increasing the SCS stiffened the backbone, but that a limit exists to the effect that steric hindrance has on backbone dynamics. One main difference between the polypeptides and PAMAs was the SCS required to stiffen the backbone. The PAMAs required a 15 atom SCS to stiffen the backbone, whereas the copolypeptides were relatively unaffected by SCS other than the limiting case when the side chain contained a single hydrogen atom, indicating that the inclusion of even a single methyl group in the side chain was sufficient to stiffen the polypeptide backbone. This result

was most certainly a reflection of inherent stiffness of the polypeptides as compared to the PAMAs for the two following reasons. First, rotation about a peptide bond in the polypeptides was much more hindered than about a C-C bond in the PAMA samples and second, all polypeptides contained ~44 mol% of the somewhat bulky Glu, both effects contributing toward making the polypeptides much stiffer than the PAMA samples.

5.5 Discussion

Perhaps the most surprising result from this study was the relatively narrow range of N_{blob}^{exp} values retrieved for different *aa*'s in Figure 5.3. If one compares in Table 5.2 the different *aa*'s according to their SCS, 40% (8 out of 20) of the most common *aa*'s have side chains which contain 5 or more atoms. The constancy of N_{blob} between PGlu and PLys(Z)Glu with SCS of 5 and 15 atoms, respectively, and the excellent agreement between the N_{blob}^{exp} and N_{blob}^{theo} values obtained for an extended polypeptide suggest that PGlu and PLys(Z)Glu are fully extended over the length scale of a *blob*. Consequently, homopolypeptides prepared with *aa*'s having a SCS of 5 or more atoms would be expected to adopt a same extended conformation and have a same N_{blob} value of 11 ± 1. In fact, these experiments suggest that any of these 8 *aa*'s, or 40% of the 20 most common *aa*'s, should have the same effect on the conformation and dynamics of an unstructured polypeptide and can be used interchangeably.

Table 5.2. Experimental and predicted N_{blob} values for polypeptides incorporating ~44 mol% of amino acids with different side chain sizes (number of non-hydrogen atoms in the side chain).

Size	Amino acid chemical structure (Amino acids in bold have been investigated in this report)					N_{blob} DMF	N_{blob} DMSO
0						28 ± 2	23 ± 2
	Glycine						
1						18 ± 1	16 ± 1
	Alanine						
2						14 ± 3	14 ± 2
	Serine	Cysteine					
3							
	Valine	Threonine					
4							
	Aspartic acid	Asparagine	Isoleucine	Leucine	Methionine		
5+						11 ± 1	11 ± 1
	Glutamic acid	Glutamine	Lysine	Histidine			
							
	Arginine	Phenylalanine	Tyrosine	Tryptophan			

The copolypeptides PGlyGlu, PAlaGlu, and PLys(Z)Glu contained 44 (± 3) mol% glutamic acid on average, where Glu seems to be a good representative of 40% of the 20 most common *aa*'s. In fact, these copolypeptides were excellent model polymers to probe the effect that the 60% other *aa*'s with a smaller side chains would have on the conformation of unstructured proteins, whose sequence contains on average 40% of *aa*'s with a SCS of 5 or more atoms. Incorporation of the next 9 smaller *aa*'s with a side chain of 2 to 4 atoms into a polypeptide containing ~44 mol% Glu should yield an N_{blob} value intermediate between that of alanine and glutamic acid and should thus be equal to ~14 ± 2. This statement implies that 17 out of the 20 most common *aa*'s should have N_{blob} values between 11 and 16, a surprisingly narrow range. Only when alanine or glycine are present in the sequence would a more significant increase in N_{blob} be expected.

The assumption that N_{blob} for an *aa* is only a function of its SCS may appear like a large leap at first glance, as this does not take into account any structural features of a side chain such as branching. As it turns out, branching in the side chain does indeed cause a decrease in N_{blob} compared to its linear counterpart due to increased sterics.[52] This effect was quantified for the Py-PAMAs, for which N_{blob}^{exp} obtained with the PAMAs with *n*- and *t*-butyl side chains were compared. In both cases, the SCS equaled 6 atoms, but the branched *t*-butyl side chains yielded an N_{blob}^{exp} value that was 40% smaller than for the *n*-butyl groups. Although this is a rather significant decrease, there are several important differences between the PAMAs and polypeptides.

As discussed before, the inherent rigidity of the polypeptides appears to dampen the effect that SCS has on the backbone conformation, which led to significant changes in N_{blob} only for a SCS of 0 for glycine and 1 for alanine. This means that side chain branching should not be as impactful for the rigid polypeptides. The second and perhaps even more important difference is the range of N_{blob}^{exp} values. The PAMAs saw an increase in N_{blob}^{exp} from 16 to 59 as the SCS decreased from 20 to 3. However, Figure 5.3 demonstrates that 18 out of the 19 *aa*'s in Table 5.2 must take an N_{blob}^{exp} value between 11 and 16 in DMSO and between 11 and 18 in DMF. Therefore, any effect that branching (or any other structural features of an *aa*) must result in relatively small changes in N_{blob}^{exp} since the N_{blob}^{exp} values cannot be lower than the limiting N_{blob}^{exp} value of 11 ± 1 for an extended segment of a polypeptide, or higher than 16 or 18 since alanine has a smaller side chain. Consequently, all the *aa*'s highlighted in green in Table 5.2 with 2 – 4 non-hydrogen atoms in their side chains must have N_{blob}^{exp} values around 14 despite the changes in side chain structure and size. This represents a narrow range of N_{blob}^{exp} values.

Furthermore, the 8 *aa*'s in Table 5.2 with a SCS of 5 or more atoms will generate oligopeptide segments that are fully extended over the length scale of a *blob*. The fact that N_{blob}^{exp} is expected to be ~14 ± 2 for the 9 additional *aa*'s with a SCS of 2 or 3 atoms implies that for 17 out of 20 *aa*'s, peptide segments of 11 – 14 *aa*'s will be fairly extended in solution, unless they can generate a secondary structure. Of course, this possibility was avoided in the present study through the use of racemic mixtures of *aa*'s to prepare all polypeptides, which were thus unable to form a secondary structure. The more substantial changes in N_{blob} observed

when alanine or glycine are present in the backbone are due to the enhanced conformational freedom that they provide to the polypeptides.

The trend in Figure 5.3C for the data in DMSO, where the polypeptides are well solvated, was reproduced in Figure 5.6 to better illustrate how the N_{blob} values are related to the increased folding of the polypeptide chain due to the presence of smaller *aa*'s. This trend suggests a hierarchy of motions for the polypeptide backbone. The main chain of a polypeptide constituted of *aa*'s with a SCS of 5 or more non-hydrogen atoms is rigid and extended due to the stiff nature of the constituting peptide bonds and the steric hindrance of the bulky side chains. The addition of smaller *aa*'s such as alanine or glycine provides sufficient conformational freedom to enable the local bending of the chain, thus enhancing the encounter between two pyrenyl labels located in a same *blob*. As discussed in this report, these local motions are reflected by N_{blob}, k_{blob}, and the product $k_{blob}\times N_{blob}$. Longer range motions that would bring pyrenyl labels from neighboring *blobs* into the *blob* of reference are handled with the exchange term given by the product $k_e\times$[*blob*], where k_e is the rate constant describing the exchange of ground-state pyrene among *blobs* and [*blob*] is the local concentration of *blobs* inside the polymer coil. In terms of time scales, the ground-state pyrenyls diffuse in and out of the *blobs* with a $k_e\times$[*blob*] rate constant of $\sim 5\times10^6$ s^{-1}, encounters between two SU bearing a pyrenyl label inside a *blob* occur with a k_{blob} rate constant of $\sim 10\times10^6$ s^{-1}, and rapid re-arrangement between two pyrenyl labels to form an excimer takes place with a k_2 rate constant of $\sim 150\times10^6$ s^{-1} (see Tables S5.2 and S5.5).

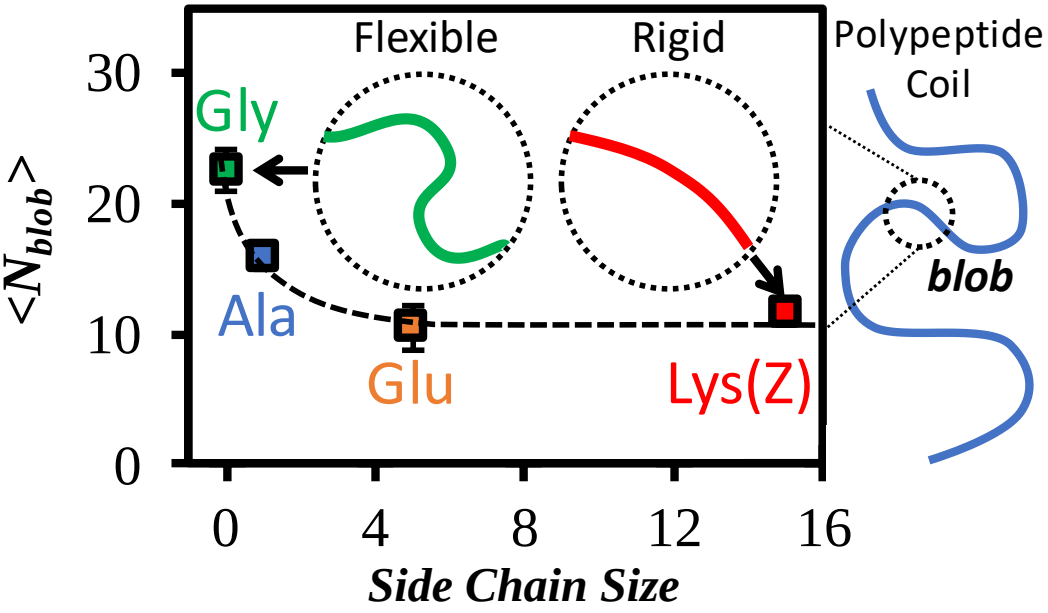


Figure 5.6. Illustration of the enhanced flexibility of the polypeptide backbone upon incorporation of smaller *aa*'s. (Data from Figure 5.3C obtained in DMSO)

One factor that must be kept in mind when considering the N_{blob}, k_{blob}, and $k_e \times [blob]$ parameters is that these parameters pertain to a glutamic acid modified with 1-pyrenemethylamine. As has been reported in the literature, these parameters depend critically on the length of the spacer linking pyrene to the polymer backbone.[49,68,69] For instance, N_{blob}^{theo} obtained by MMOs for an extended PGlu or PLys labeled with 1-pyrenemethylamine or 1-pyreneacetic acid equals 10.3 ± 1.2 Glu and 14.3 ± 1.3 Lys, respectively.[49] The difference is related to the length of the linker made of 5 and 7 atoms connecting pyrene to the polypeptide backbone of PGlu and PLys, respectively, whereby the longer reach of PLys labeled with 1-pyreneacetic acid results in a larger N_{blob}^{theo} than for PGlu labeled with 1-pyrenemethylamine.

Consequently, the FBM parameters obtained for polypeptides prepared with different spacers are internally consistent with each other, as long as the same spacer is used in a series of experiments.

The two organic solvents DMSO and DMF were used in these experiments. Since this study focused on the characterization of the internal dynamics of polypeptides, DMSO was deemed to be a better solvent for this purpose as the polypeptides were better dissolved as compared to DMF. The dynamic encounters between the pyrene-labeled structural units of the polypeptides monitored in DMSO would likely be similar to those experienced by proteins in water at the early stages of folding, before the hydrophobic collapse and the formation of secondary structures have had time to take place. Working with racemic polypeptides in DMSO eliminated hydrophobic interactions and prevented structure formation, which would otherwise reduce, or most likely halt, dynamic interactions between some *aa*'s in solution.

Perhaps the most important consequence of this work is that the rate of encounters $<k_{blob} \times N_{blob}>$ between *aa*'s, which dictates the probability of encounters between specific *aa*'s whose interactions result in the formation of structural elements as proteins fold, remains constant in Figure 5.5 for *aa*'s having a SCS equal to 1 (Py-PAlaGlu), 5 (Py-PGlu), and 15 (Py-PLys(Z)Glu. This is again a surprisingly simple outcome, as except for glycine (SCS=0) and proline, all other 18 *aa*'s in Table 5.2 are expected to encounter each other at a same frequency. It suggests that other than for Gly (and proline), the *aa* sequence appears to have a limited effect on the internal dynamics of a polypeptide.

Interestingly, the N_{blob} values reported in Figure 5.3 (and Table 5.2) were similar to the size of many of the known[70] and predicted[71] intrinsically disordered regions (IDRs) of proteins, which are typically constituted of a few 10's of amino acids. This observation suggests that the FBM yields information on a length scale that is relevant to IDRs in proteins. Since the size of a *blob* is defined by the number of *aa*'s which can encounter one another on the time scale of backbone and side chain motions, the agreement between the size of *blobs* and IDRs suggests that the size of IDRs in proteins might be defined by those *aa*'s which are capable of diffusional interactions with one another.

5.6 Conclusions

The FBM was applied to a series of pyrene-labeled polypeptides to probe the effect that different *aa*'s have on the conformation and internal dynamics of polypeptides. Considering the quasi-infinite number of possible structural arrangements resulting from the folding of proteins constituted of the 20 most common *aa*'s, this study suggests that the internal dynamics and conformation of polypeptides prepared from racemic mixtures of *aa*'s are infinitely simpler to understand. The plots of N_{blob} and the product $k_{blob} \times N_{blob}$ as a function of SCS in Figures 5.3 and 5.5, respectively, suggest the existence of a gradient in conformational freedom which, like the synthetic PAMAs studied earlier,[52] depends on the size of the *aa*'s side chain. However, the plots in Figures 5.3 and 5.5 suggest that this gradient only involves the smaller *aa*'s in Table 5.2. In fact, the results presented in this report imply that 8 + 9 = 17 *aa*'s out of the 20 most common *aa*'s generate an extended conformation for an unfolded polypeptide and

will adopt a conformation in solution with a similarly low density. Only the incorporation of alanine and glycine, with SCS of 1 and 0, respectively, in the polypeptide sequence seems to generate substantial conformational freedom resulting in a more compact polypeptide that would promote encounters between *aa*'s. This implies that since most *aa*'s have a same effect on polypeptide chain dynamics, Ala and Gly distributed along a polypeptide sequence are most efficient in promoting encounters between *aa*'s and thus must be key contributors to protein folding.

Chapter 6

Blob-Based Approach to Estimate the Folding Time of Proteins Supported by Pyrene Excimer Fluorescence Experiments

Adapted with permission from Casier, R.; Duhamel, J. *Blob*-Based Approach to Estimate the Folding Time of Proteins Supported by Pyrene Excimer Fluorescence Experiments. *Macromolecules* **2020**, *53*, 9823 – 9835.

6.1 Abstract

The response of the internal dynamics of a polypeptide to changes in its amino acid (*aa*) composition was investigated by applying the florescence *blob* model (FBM) to a series of pyrene-labeled poly(*D,L*-alanine-*co*-*D,L*-glutamic acid)s (PAlaGlu's) and poly(*D,L*-glutamic acid) (PGlu). The alanine content of the PAlaGlu samples was varied between 24 and 58 mol%, using PGlu (containing 0 mol% alanine) for comparison purposes. The FBM yielded the number N_{blob} of *aa*'s which could diffusively encounter one another in a *blob*, which is the volume probed by an excited pyrenyl label covalently attached to the polypeptide. The incorporation of 24 mol% of alanine was found to significantly increase N_{blob} from 11 (±1) for PGlu to 16 (±1) for the $Pala_{24}Glu_{76}$ sample in DMSO due to the enhanced conformational freedom provided by alanine to the polypeptide. Interestingly, further increases in the alanine content of the PAlaGlu samples from 24 up to 58 mol% did not change the N_{blob} value, implying that only a few flexible *aa*'s in the backbone of a polypeptide are required to disrupt its rigid conformation. The N_{blob} values of 16 for PAlaGlu and 23 found earlier for poly(glycine-*co*-*D,L*-glutamic acid) (PGlyGlu) were then used to estimate an average N_{blob} value of 19 *aa*'s for proteins based on their average *aa* composition, which should include at least one glycine or one alanine residue for a *blob* of this size. The N_{blob} value of 19 *aa*'s was then used to estimate the folding times of 145 different proteins by applying a simple hierarchical conformational search method to determine the number of conformations that could be adopted by the oligopeptide segment of a protein located inside a *blob*. Surprisingly for such a crude and simple approach, the method provided folding time (τ_F) estimates that were in good agreement with experimentally measured values, resulting in a correlation coefficient of 0.73. The good agreement found between the calculated and experimentally

determined τ_F's supports the notion that the folding of proteins occurs in and among localized subdomains which happen to be well represented by FBM analysis of the fluorescence decays of pyrene-labeled polypeptides.

6.2 Introduction

How a linear chain of amino acids (*aa*'s) folds spontaneously and rapidly into the specific 3D structure of a protein remains one of the most fundamental questions in protein science.[1] The wide range of protein compositions enabled by the near-infinite number of combinations of the constituting 20 most common *aa*'s has made this a rather challenging question to answer. Such an answer must first provide a rationale as to how polypeptides fold over such short times, considering that it would take an infinitely long folding time to probe the vastness of the entire conformational space theoretically available to the polypeptide backbone. Therefore, predicting the folding time of a protein based on its constitutional elements has become a means to assess which parameters govern the rate and mechanism for the folding of a protein. Among the physical parameters that have been investigated to date are the contact topography of the *aa*'s based on the local contact order,[2,3] the long-range order,[4] or total contact distance,[5,6] the protein sequence,[7,8] the *aa* properties[9] which include polarity, hydrophobicity, or solvent accessibility, the chain size,[10,11] the secondary structure content,[12] and the number of native contacts.[13] Bioinformatics and machine learning have also enabled the prediction of the folding time and in some cases even the folding mechanism.[14] However, since these techniques simply predict folding using correlations between the structure/sequence of a protein to those of other

known proteins, these studies lack the ability to provide an underlying physical rationale as to how proteins fold. Furthermore, the dependency of the physical parameters on the different types of proteins makes it difficult to conclude which parameters are most important to predict the folding time. For example, the parameters that seem to have the most effect on the folding rate of two-state proteins (those that fold in a single step) and three-state proteins are the contact order[2] and the chain length,[11] respectively. In addition, good correlations between the aforementioned physical parameters and the folding rate are not general and are usually found for specific structural classes of proteins based on their contents of α-helices or β-strands.[4,7,15]

Beyond the physical parameters discussed above, the intermolecular interactions experienced by *aa*'s in a protein define the mechanism and rate for the folding of a protein.[12,16,17] Over the past decades, it has become clear that folding predominantly occurs through *aa*'s interacting with each other in localized subdomains, often referred to as *foldons*.[18–21] Therefore, it stands to reason that the folding of proteins is also governed by the folding of the oligopeptide segments constituting the *foldons*. *Foldons* are typically probed by hydrogen exchange (HX) experiments, which have demonstrated that *aa*'s generate domains in proteins where they interact cooperatively to form secondary structures.[18,20] Although HX experiments probe *foldons* directly, the inherent complexity of proteins and the fact that these experiments determine the number of *aa*'s constituting an already folded domain make it difficult to assess how the *aa* composition of a *foldon* affects its size and the rate of *aa* interactions, since it is pre-set by the protein sequence. Consequently, the results of these experiments cannot be easily applied to predict folding times since the effect that individual

aa's have on *foldons* cannot be easily sorted out. In contrast, the fluorescence *blob* model (FBM) provides a means to determine quantitatively the number of *aa*'s in an unfolded synthetic polypeptide of known composition, where the *aa*'s interact locally with one another, as well as the frequency at which they do so.[22–25]

Conceptually, the FBM is ideally suited to study the subdomains of macromolecules, as it was derived by dividing a macromolecule into a series of subvolumes referred to as *blobs*.[26–28] Within the framework of the FBM, a *blob* is defined as the volume an excited dye (typically pyrene) covalently attached to the macromolecule can probe while it remains excited. Since pyrene is covalently linked to the macromolecule, which for the purpose of this report will be a polypeptide, the *blob* represents the volume where a set number of *aa*'s can locally diffuse and interact with one another. The random labeling of a polypeptide with a pyrene derivative ensures that the pyrenyl labels distribute themselves randomly among the *blobs* according to a Poisson distribution defined by the average number $<n>$ of pyrenes per *blob*. FBM analysis of the fluorescence decays acquired with polypeptides randomly labeled with pyrene yields $<n>$ which is used to calculate N_{blob}, the number of *aa*'s contained inside a *blob*. Consequently, N_{blob} represents the number of *aa*'s which interact locally with one another. The FBM analysis also yields the rate constant (k_{blob}) at which an excited and a ground-state pyrene encounter to form an excimer inside a *blob*. Again, since the pyrenyl labels are covalently attached onto the polypeptide, k_{blob} represents the time scale over which the *aa*'s encounter one another inside a *blob*.

To date, the FBM has been applied to several model polypeptides to provide information on how the *blob* size and internal dynamics respond to changes in the structure,[22,23] internal density,[23,29] and *aa* composition[24,25] of a homopolypeptide. If the folding time of a polypeptide depends on the size of the subdomains where *aa*'s interact with one another and since proteins are composed of 20 different *aa*'s, the effect that the *aa* composition of a copolypeptide might have on the *blob* size (N_{blob}) and internal dynamics (k_{blob}) would need to be clearly understood. The first step toward developing such an understanding was carried out by investigating a series of copolypeptides of known composition.[25] Three pyrene-labeled random copolypeptides referred to as Py-PGlyGlu, Py-PAlaGlu, and Py-PLys(Z)Glu were prepared that contained 44 (±3) mol% of glutamic acid (Glu) and 56 (±3) mol% of either glycine (Gly), alanine (Ala), or *N*-ε-carbobenzyloxylysine (Lys(Z)), respectively. The FBM revealed that compared to poly(glutamic acid) (PGlu), the incorporation of the larger Lys(Z) had no significant impact on N_{blob} or k_{blob} which were the same for PGlu and Py-PLys(Z)Glu, suggesting that all *aa*'s with large side chains resulted in similar backbone dynamics. However, the incorporation of the smaller *aa*'s (Ala and Gly) resulted in a significant increase in N_{blob} from 11 (±1) for PGlu to 15.8 (±1.0) for Py-PAlaGlu and 22.5 (±1.5) for Py-PGlyGlu in DMSO. The increase in N_{blob} demonstrated that the smaller *aa*'s provided the conformational freedom necessary for the polypeptide backbone to increase the number of *aa*'s which could interact with one another inside a *blob*. These preliminary results indicated that the size of a polypeptide *blob* was governed by the local conformational freedom of a polypeptide sequence, and that the small *aa*'s (Ala and Gly) would enhance the *blob* size while *aa*'s with

a size similar to or larger than Glu would have a same *blob* constituted of 11 (±1) *aa*'s. However, since all the copolypeptides studied had a constant ~44:56 molar ratio of Glu-to-different *aa*'s, the effect that the molar ratio of Glu-to-a given *aa* had on N_{blob} for a copolypeptide remained unknown. For instance, if an Ala content of 58 mol% raised N_{blob} from 11 for PGlu to 16 for PAlaGlu, the question remained as to what the increase in N_{blob} would be for a different Glu:Ala molar ratio such as 70:30? Answering this question would help establish the relationship between N_{blob} and the *aa* composition of a polypeptide, that would enable the determination of the different N_{blob} values for the *blobs* corresponding to different segments of the sequence of any polypeptide. In turn, these N_{blob} values could then be converted into theoretical folding times to be compared with the folding time experimentally determined for a given polypeptide.

The current study addresses this point by determining N_{blob} for a series of Py-PAlaGlu samples prepared with varying Ala:Glu molar ratios. Surprisingly, N_{blob} remained constant and equal to 15.8 (±1.0) regardless of the alanine content for PAlaGlu samples containing between 24 and 58 mol% Ala, leading to the tantalizing proposal that the size of a polypeptide *blob* might simply be determined by the smallest *aa*'s present in the oligopeptide sequence constituting a *blob*. Since PAlaGlu and PGlyGlu have an N_{blob} value of 15.8 (±1.0) and 22.5 (±1.5), respectively, a protein could be viewed as a string of *blobs* having an N_{blob} value of 19 *aa*'s, taken as the average of the N_{blob} values of PAlaGlu and PGlyGlu ($N_{blob} \sim (15.8 + 22.5)/2$), as long as each *blob* contains at least one Ala or Gly residue. As it turns out, Ala and Gly have, respectively, a 7.3 and 7.6% probability to appear in a protein sequence,[30] so that a hypothetical

blob constituted of 19 *aa*'s would be expected to contain 3 *aa*'s that could be either Ala or Gly. This would ensure that each 19 *aa*'s *blob* would have a greater than 95% probability to contain at least one Ala or Gly, thus validating the assumption made earlier. The 19 *aa*'s estimate for N_{blob} could then be applied to assess the viability of such a *blob*-based view for handling polymer dynamics and conformations. To this end, a procedure was implemented to predict the theoretical folding time of 145 proteins, assuming that the folding of a polypeptide is governed by the local interactions of *aa*'s confined inside *blobs* made of 19 *aa*'s. This crude and simple approach yielded a 1:1 relationship between the experimentally determined and mathematically predicted folding times for the 145 proteins considered with a correlation coefficient of 0.73. Such a strong correlation indicates that while the mathematical procedure can be much refined by mapping out in a more rigorous and detailed manner the relationship between N_{blob} and the polypeptide composition, the proposed *blob*-based methodology appears to be an appealing mathematical tool to estimate the folding times of proteins. Furthermore, this study provides further support to the notion[18] that folding is dictated by the interactions existing between *aa*'s localized in the subvolumes generated inside a protein.

6.3 Materials

Chemicals: *D*,*L*-Alanine (Ala, Sigma, ≥ 99%), *D*,*L*-glutamic acid 5-*tert*-butyl ester (H-*D*,*L*-Glu(OtBu)-OH, Chem-Impex Int'l Inc., > 98%), bis(trichloromethyl) carbonate (triphosgene, Oakwood Chemical, ≥ 99%), calcium hydride (CaH_2, Sigma, ≥ 95%), *N*,*N*′-diisopropylcarbodiimide (DIC, Sigma, 99 %), *N*,*N*-diisopropylethylamine (DIPEA, Sigma,

99.5%), dimethyl sulfoxide (DMSO, Sigma, ≥ 99.9%), hydrochloric acid (HCl, Fisher, 34 wt% in water), 1-hydroxybenzotriazole monohydrate (HOBt, Advanced ChemTech), (±)-α-pinene (Oakwood Chemical, 98%), potassium sulfate (Sigma, ≥ 99.0%), 1-pyrenemethylamine hydrochloride (PyMA·HCl, Sigma, 95%), and sodium hydroxide (NaOH, Sigma, ≥ 97.0%) were used as received. *n*-Butylamine (Sigma, 99.5%), *N*,*N*-dimethylformamide (DMF, Sigma, ≥ 99.8%), and dioxane (≥ 99.5%) were dried following procedures provided in the SI. Deionized water was obtained from a Biopure Series 4400 Single Pass Reverse Osmosis system. *N*-Succinimidyl acetate (AcOSu) was synthesized according to a previously reported procedure.[22] N_2 gas (Praxair, N4.0) was used to keep the reaction mixtures under nitrogen atmosphere.

Synthesis of D,L-alanine N-carboxyanhydride (AlaNCA) and D,L-glutamic acid 5-tert-butyl ester N-carboxyanhydride (Glu(OtBu)NCA): *D,L*-Alanine and the *tert*-butyl ester of *D,L*-glutamic acid were converted into AlaNCA and Glu(OtBu)NCA, respectively, according to a previously published procedure using triphosgene.[25]

Poly(D,L-alanine-co-D,L-glutamic acid)s (PAlaGlu's): Three PAlaGlu's were synthesized with contents ranging from 24 to 58 mol% alanine (f_{Ala} = 0.24 – 0.58). The synthesis of PAlaGlu with f_{Ala} = 0.58 ($PAla_{58}Glu_{42}$) has been previously described.[25] Each PAlaGlu was synthesized in a similar manner using an amine-initiated ring opening polymerization of the NCAs. An example synthesis is provided for the $PAla_{41}Glu_{59}$ sample with an f_{Ala} of 0.41. Freshly prepared AlaNCA (0.34 g, 3.0 mmol) and Glu(OtBu)NCA (1.00 g, 4.4 mmol) were dissolved in an ice-chilled mixture of dried DMF (45 mL) and dioxane (8 mL) with stirring.

After all the NCA dissolved, the polymerization was initiated by the addition of 400 μL of a 1 μL/mL *n*-butylamine/DMF solution (4.0 μmol *n*-butylamine). The head-space of the flask was purged with a flow of nitrogen (50 mL/min) to displace the carbon dioxide released from the polymerization. The conversion was monitored by comparing the amide and α-proton signals of the polypeptide at 8.0 and 4.2 ppm to those of the NCA monomers at 8.9 – 9.0 and 4.4 ppm, respectively, by taking the ^{1}H NMR spectrum of the reaction mixture at different times. After 12 hours, the polymerization was stopped at a conversion of ca. 14 mol% by the addition of 1 drop of 1 M HCl. The solution was then transferred to a dialysis tube (8 kDa cutoff) and dialyzed against DMF (500 mL). After 4 hours, the dialysate was replaced with fresh DMF to ensure the complete removal of low molecular weight species. At 4 hour intervals, the dialysate was replaced with a succession of 2:1, 1:1, and 1:2 DMF:water mixtures. A final dialysis step against pure water overnight was used to precipitate the $PAla_{41}Glu(OtBu)_{59}$ which was then recovered by centrifugation. To prevent complications after the Glu(OtBu) deprotection, the *N*-terminus of $PAla_{41}Glu(OtBu)_{59}$ was capped with an acetic group following the protocol described hereafter. The collected $PAla_{41}Glu(OtBu)_{59}$ (0.13 g, 6 μmol) was dissolved in DMF (50 mL) and AcOSu (0.2 g, 1.3 mmol)[22] was added. The solution was left to stir for 2 days to ensure that the *N*-terminus would have enough time to react with AcOSu. Water (450 mL) was then added to the mixture, followed by the addition of potassium sulfate until the capped $PAla_{41}Glu(OtBu)_{59}$ precipitated from the solution. The precipitate was then collected by centrifugation. The ^{1}H NMR spectra of the PAlaGlu(OtBu) samples with f_{Ala} equal to 0.24 and 0.41 are given in Figures S6.1 – 6.2 in the SI, respectively. The glutamic acids were deprotected

by stirring $PAla_{41}Glu(OtBu)_{59}$ (0.13 g, 6 μmol) in a mixture of TFA (7 mL) and water (0.7 mL). After 2 hours, the deprotected $PAla_{41}Glu_{59}$ was precipitated into cold ether and collected by centrifugation. The $PAla_{41}Glu_{59}$ was then dissolved in DMF (5 mL) with 1 drop of 1 M HCl and filtered through a 0.22 μm PTFE filter to remove small traces of insoluble material. The $PAla_{41}Glu_{59}$ was then recovered again by precipitation into ether and centrifugation. Two additional DMF to ether precipitations were conducted to ensure the complete removal of low molecular weight species. $PAla_{58}Glu_{42}$ was then dried under vacuum overnight to yield a white powder (0.1 g). The ^{1}H NMR spectra of PAlaGlu's with f_{Ala} equal to 0.24 and 0.41 are given in Figures S6.3 – 6.4 in the SI, respectively.

Pyrene-labeled poly(D,L-alanine-co-D,L-glutamic acid)s (Py-PAlaGlu's): The PAlaGlu samples were all labeled with pyrene in a similar manner. Each PAlaGlu was labeled with different amounts of pyrene. An example of $PAla_{41}Glu_{59}$ with a target pyrene content of 6 mol% is provided. $PAla_{41}Glu_{59}$ (22 mg, f_{Ala} = 0.41), 1-pyrenemethylamine (3.7 mg, 14 μmol), HOBt (9.1 mg, 59 μmol), and DIPEA (3 μL, 17 μmol) were added to DMSO (17 mL). Once dissolved, DIC (2.4 μL, 15 μmol) was added and the solution was left to stir overnight in the dark. The pyrene-labeled $PAla_{41}Glu_{59}$ sample was then passed through a 0.22 μm PTFE filter into a dialysis tube (8 kDa cutoff). The solution was then dialyzed twice against DMF (300 mL) for four hours, before being dialyzed against 2:1, 1:1, and 1:2 DMF:water mixtures for 4 hours each. Finally, the solution was dialyzed twice against water (pH 9 – 10). Once the dialysis was completed, the water was removed by lyophilization yielding the sodium salt of Py-$PAla_{41}Glu_{59}$ (Py-$PAla_{41}GNa_{59}$). To minimize CO_2 in the air from changing the solutions

pH, the dialysis beaker was sealed with aluminum foil and an elastic band. The chemical structure of the Py-PAlaGlu samples are given in Figure 6.1A.

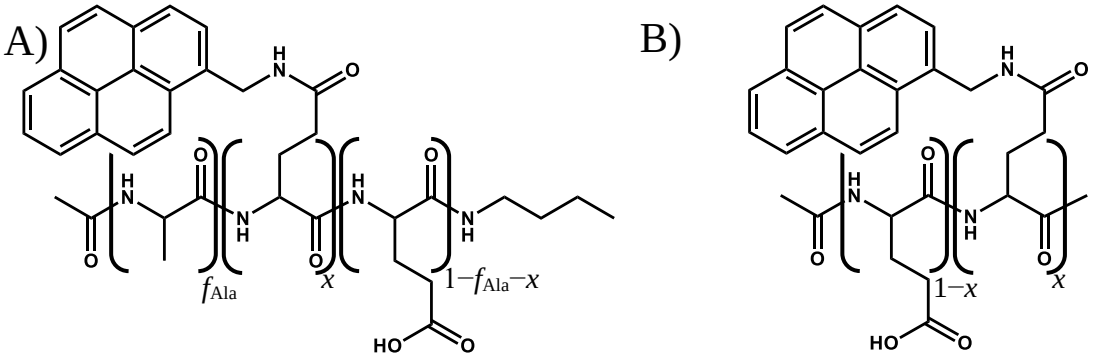


Figure 6.1. Chemical structure of A) the random copolypeptides Py-PAlaGlu and B) the homopolypeptide Py-PGlu. All stereocenters are racemic. The molar fractions of alanine and pyrene-labeled glutamic acids in the copolypeptide are given by f_{Ala} and x, respectively.

Pyrene-labeled poly(D,L-glutamic acid) (Py-PGlu): The Py-PGlu samples were synthesized as previously reported.[22,24] The chemical structure of Py-PGlu is given in Figure 6.1B.

^{1}H NMR Analysis: Proton NMR spectra were acquired on a Bruker 300 MHz NMR in d_6-DMSO. The alanine content (f_{Ala}) of the copolymers was determined by comparing the integration of the β-protons of the two comonomers. The ^{1}H NMR spectra used for the calculations of f_{Ala} are given in the SI and a summary of the corresponding f_{Ala} values is provided in Table 6.1.

Table 6.1. Characterization of poly(glutamic acid) (PGlu) and poly(alanine-*co*-glutamic acid) (PAlaGlu).

Polypeptide	Monomer Feed (mol% Ala)	Polypeptide Composition		M_n (kg·mol^{-1})	*DP*	*Đ*
		mol% Ala (f_{Ala})	mol % Glu			
PGlu[a]	-	0	100	100	784	1.06
$PAla_{24}Glu_{76}$	20	24	76	10	90	1.10
$PAla_{41}Glu_{59}$	41	41	59	16	150	1.21
$PAla_{58}Glu_{42}$	54	58	42	23	240	1.05

[a] Information provided by certificate of analysis from Alamanda Polymers.

GPC Characterization: The molecular weight distribution of the synthesized polypeptides was determined with a TOSOH 8321 GPC-WS GPC Workstation in conjunction with a Wyatt DAWN HELEOS multiangle light scattering (MALS) detector using two TSKgel Alpha-M 13 μm mixed bed columns, in DMSO at 70 °C with a flowrate of 0.6 mL·min^{-1}. The *dn*/*dc* values were found to be too low when the polypeptides were deprotected, therefore the GPC analysis was conducted on the PAlaGlu(OtBu) samples. The d*n*/d*c* values of PAlaGlu(OtBu) with f_{Ala} = 0.24 and 0.41 were calculated to equal 0.0251 and 0.0224 mL·g^{-1}, respectively, using the ASTRA software package with the built-in 100% mass recovery method. The GPC traces are given in the SI as Figures S6.4A and B. The MALS detector array was used to calculate the absolute molecular weights. PAlaGlu(OtBu) with f_{Ala} = 0.24 and 0.41 were found to have an M_n of 14 kg·mol^{-1} (DP = 90, *Đ* = 1.10) and 21 kg·mol^{-1} (DP = 150, *Đ* = 1.21), respectively. It must be noted that polydispersity of the copolypeptides had no bearing on the FBM analysis

of the decays acquired with the pyrene-labeled polypeptides, as the FBM is designed to handle polydisperse polymer samples. A summary of the molecular weight and dispersity of all the deprotected PAlaGlu samples is given in Table 6.1.

UV-Vis Absorbance Measurements: The UV-Vis absorption spectra were acquired with a Varian Cary 100 Bio spectrophotometer using a UV-Vis absorption cell with a 1.0 cm pathlength. A background correction for the solvent was applied to each spectrum.

Pyrene Content Determination: The molar fraction x of the structural units labeled with pyrene in Figure 6.1 was determined with Equation 6.1. In Equation 6.1, f_{Ala} is the molar fraction of alanine incorporated in the polypeptide, M_{Ala} (= 71.08 g·mol^{-1}) is the molar mass of alanine, M_{GNa} (= 144.15 g·mol^{-1}) is the molar mass of the sodium-salt of glutamic acid, and M_{Py} (= 342.4 g·mol^{-1}) is the molar mass of a glutamic acid labeled with 1-pyrenemethylamine. Lastly, the number of moles of pyrene per gram of sample (λ_{Py}) was determined by UV absorbance measurements. Since the Py-PAlaGNa sample was dialyzed against a NaOH solution, the lyophilized solid contained both Py-PAlaGNa and salts. The mass of salt (m_S) in the mixture was determined by lyophilizing ca. 20 mL of the last dialysate. The mass of salt per unit volume (λ_S) was then determined by weighing the remaining solid after lyophilization. Since the volume of lyophilized Py-PAlaGNa solution was known, m_S in the Py-PAlaGNa solid mixture (m_{S+Py}) was determined by multiplying λ_S by the volume of lyophilized Py-PAlaGNa solution. The mass fraction f of pure Py-PAlaGNa was then calculated by $(m_{S+Py} - m_s)/m_{S+Py}$. At least 5 mg of lyophilized Py-PAlaGNa solid was dissolved in ca. 5 mL of DMSO. The mass of pure Py-PAlaGNa was determined by multiplying the weighed mass by the mass fraction f. A drop

of 1 M HCl in water was added to the solution to protonate the glutamic acid residues. After sonicating for 15 minutes, the solution was diluted with DMF and an aliquot was transferred to a 1.0 cm pathlength cell for UV-Vis absorption measurements. The molar concentration of pyrene was determined using the Beer-Lambert law at the absorbance peak maximum of pyrene at 346.5 nm with the molar absorptivity coefficient of the model compound *N*-(1-pyrenylmethyl)acetamide (PyMAAc) in DMSO (ε = 39,300 $M^{-1}\cdot cm^{-1}$).[25] Dividing the pyrene molar concentration by the polymer mass concentration yielded the pyrene content λ_{Py}.

$$x = \frac{f_{Ala} M_{Ala} + (1 - f_{Ala}) M_{GNa}}{\lambda_{Py}^{-1} + M_{GNa} - M_{Py}} \tag{6.1}$$

Steady-State Fluorescence (SSF) Measurements: Steady-state fluorescence spectra were acquired on a Horiba QM-400 spectrofluorometer equipped with a xenon arc lamp. A ~5 mL solution of pyrene-labeled polypeptide was prepared such that it contained ca. 2.7×10^{-6} M pyrene. A 10 μL aliquot of 1 M HCl was added to ensure that all the Glu residues were protonated. The samples were sonicated for ca. 10 minutes and outgassed with a gentle flow of N_2 gas (Praxair, N4.0 or greater) for ca. 30 minutes to remove oxygen which quenches the fluorescence of pyrene. The fluorescence cell was then sealed to preserve the nitrogen atmosphere before acquiring the SSF spectrum. The sample was then excited at 346 nm using an excitation and emission slit widths of 1 nm. The SSF spectrum was collected in 1 nm steps from 350 to 600 nm at a scanning rate of 10 $nm\cdot s^{-1}$. The fluorescence intensity of the monomer (I_M) and excimer (I_E) were calculated using the area under the SSF spectrum from 375 to 381

nm and from 500 to 530 nm, respectively. The relative intensity of the excimer to that of the monomer was quantified using the ratio I_E/I_M.

Time-Resolved Fluorescence Measurements: The Py-PAlaGlu solutions were prepared in a similar manner as for the SSF measurements. The fluorescence decays were acquired with an IBH TC-SPC fluorometer equipped with a NanoLED-340 laser. The monomer and excimer decays were collected over 1,024 channels. The monomer decay was collected at 378 nm with a time-per-channel of 2.04 $ns \cdot ch^{-1}$ until a total of at least 40,000 counts was obtained at the decay maximum. A 370 nm longpass filter was used to minimize any potential light scattering. The excimer fluorescence decays were acquired with a time-per-channel of 1.02 $ns \cdot ch^{-1}$ to at least 20,000 counts at 510 nm with a 470 or 495 nm longpass filter. The instrument response function was collected at 344 nm with a LUDOX dispersion in water and was used in the deconvolution of the fluorescence decays in the decay analysis.

Fluorescence Decay Analysis: The monomer and excimer fluorescence decays were globally analysed according to the fluorescence *blob* model (FBM) with Equations S6.1 – S6.3 in the SI. The FBM requires a macromolecule which has been labeled with a dye and quencher. Since pyrene has the ability to self-quench as it forms an excimer, it is often used since it acts as both the dye and quencher, thereby simplifying the labeling procedure. Within the framework of the FBM, a pyrene-labeled macromolecule is divided into subvolumes, referred to as *blobs*, where a *blob* is defined by the volume probed by an excited pyrenyl label covalently attached to the macromolecule (Figure 6.2).

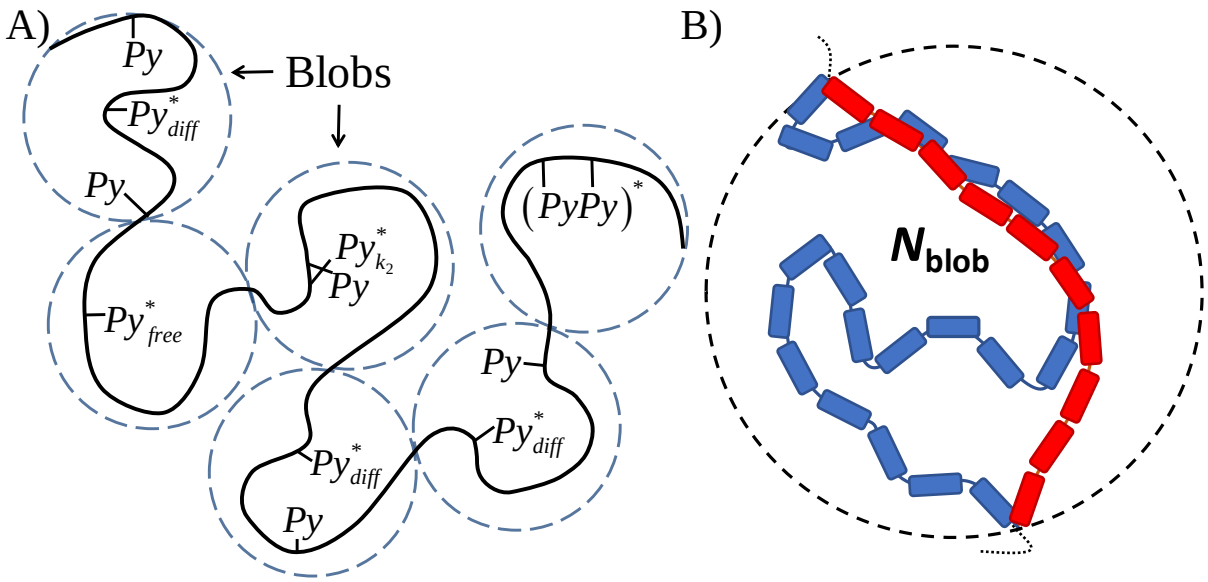


Figure 6.2. A) Illustration of the compartmentalization of a linear polymer chain into *blobs*. The chain is randomly labeled with pyrenyl groups which (Py_{diff}*) are attached on structural units (SUs) undergoing diffusive motions in solution described by k_{blob}, (Py_{k2}*) are in close proximity and form excimer via a rapid re-arrangement with a rate constant k_2, (Py_{free}*) are free in solution and emit with their natural lifetime τ_M, and (*PyPy*)* are aggregated and form excimer instantaneously. B) Illustration of a *blob* for a (red, 10 SUs) rigid and (blue, 21 SUs) flexible polymer.

The FBM utilizes a sequential process for excimer formation. An excited pyrene, referred to as Py_{diff}*, and a ground-state pyrene (*Py*) diffuse inside the *blob* with a rate constant k_{blob}. Since the pyrenyl labels are bound to the macromolecule, k_{blob} reflects the diffusive motion of the structural unit to which pyrene is bound, and therefore provides information on the internal dynamics of the macromolecule. Once Py_{diff}* and *Py* are in close proximity, Py_{diff}*

morphs into the species Py_{k2}* which then undergoes a rapid rearrangement with *Py* to form an excimer with a rate constant k_2. The process forms one of two excimer species referred to as *E** and *D** which emit with their natural lifetimes τ_E and τ_D, respectively. Provided the pyrenes are distributed randomly along the macromolecule, the number of ground-state pyrenes within a *blob* follows a Poisson distribution. This is taken advantage of by the FBM which yields the average number $<n>$ of ground-state pyrenes per *blob*, which is then used in conjunction with the pyrene content *x* calculated with Equation 6.1 to determine N_{blob}, the average number of structural units within the volume of a *blob*, whose expression is given in Equation 6.2. In Equation 6.2, f_{Mfree} is the molar fraction of pyrenyl labels which emit with their unquenched lifetime τ_M in the monomer decay.

$$N_{blob} = \frac{< n >}{x}\left(1 - f_{Mfree}\right) \qquad (6.2)$$

6.4 Results

Three series of pyrene-labeled poly(*D*,*L*-alanine-*co*-*D*,*L*-glutamic acid)s (PAlaGlu's) were prepared to determine the effect of copolypeptide composition on the size (N_{blob}) and dynamics (k_{blob}) of a *blob*. The glutamic acid (Glu) and alanine (Ala) comonomers were selected for the following reasons. First, Glu provides a carboxylic acid for the covalent attachment of 1-pyrenemethylamine. Second, since pyrene-labeled poly(*D*,*L*-glutamic acid) (Py-PGlu) and several other copolypeptides containing Glu had been previously characterized, these studies

provided a benchmark against which the behavior of the Py-PAlaGlu samples could be compared.[22,25] Third, Ala was chosen as the second comonomer since previous research had found that a Py-PAlaGlu sample prepared with 58 mol% Ala generated *blobs* with an N_{blob} value of 16 (±1) *aa*'s, substantially larger than that of 11 (±1) *aa*'s for PGlu. While the increase in N_{blob} could be clearly attributed to the enhanced backbone conformational freedom brought by the Ala residues,[25] the effect that the content of Ala residues constituting a PAlaGlu sample might have on N_{blob} remained to be assessed. Therefore, a series of Py-PAlaGlu samples were prepared with alanine contents (f_{Ala}) of 0.24, 0.41, and 0.58 to determine how N_{blob} varied with the Ala content of the PAlaGlu samples. Finally, since the folding of the polypeptide into a structured conformation would interfere with the measurement of N_{blob} and k_{blob} for a purely coiled polypeptide according to the FBM, racemic mixtures of Glu and Ala were used to prepare the copolypeptides and ensure that they were studied in a purely unstructured coiled conformation.

Steady-State Fluorescence: The fluorescence spectra of the Py-PAlaGlu and Py-PGlu samples were acquired in DMSO. The fluorescence spectra of the Py-PAla$_{58}$Glu$_{42}$ series, with a molar fraction of alanine (f_{Ala}) of 0.58, are shown in Figure 6.3A for different pyrene contents (x). The fluorescence spectra of the other Py-PAlaGlu's and Py-PGlu series are given in Figure S6.6 in the SI. As expected, Figure 6.3A shows that the excimer intensity increases with increasing x due to the increased local concentration of pyrenyl labels. The increase in excimer fluorescence was quantified by calculating the I_E/I_M ratio as described in the experimental section and plotting it as a function of pyrene content in Figure 6.3B for all the Py-PAlaGlu

and Py-PGlu samples. Figure 6.3B showed that the I_E/I_M ratios clustered around a single line, indicating that the I_E/I_M ratios were independent of the alanine content of the copolypeptides. At first glance, this would suggest that the incorporation of alanine had no effect on the internal dynamics or conformation of the copolypeptides. In fact, this result could also be a consequence of two effects cancelling each other. Within the FBM framework, PEF for a rigid backbone would occur between two pyrenyl labels close to each other, resulting in a small N_{blob} but a large k_{blob}. Similarly, a more flexible backbone would allow two pyrenyl labels to form excimer inside a larger *blob* with a larger N_{blob} associated with a smaller k_{blob}. This discussion highlights the importance of distinguishing between the local density and the internal dynamics of macromolecules which are described by N_{blob} and k_{blob}, respectively. An analysis of fluorescence spectra based on the I_E/I_M ratios cannot distinguish between the two effects and yields an average response. A previous study on pyrene-labeled polypeptides also found that their I_E/I_M ratio did not show much difference with the polypeptide composition.[25] Differences in behavior can be assessed more easily through the fluorescence decay analysis according to the FBM. Consequently, the FBM analysis was applied to the fluorescence decays acquired with the Py-PAlaGlu samples.

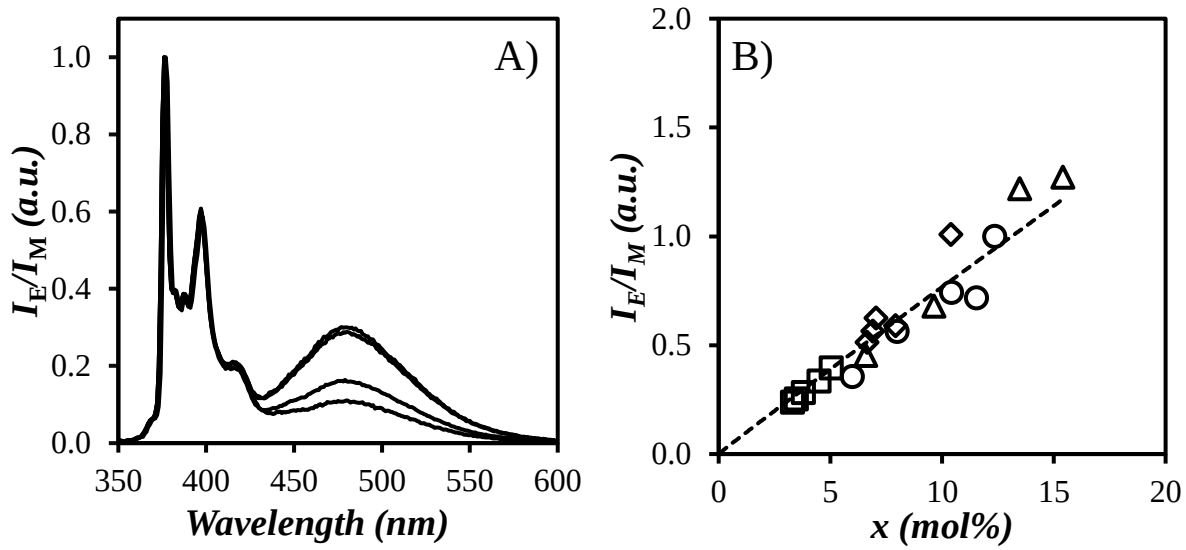


Figure 6.3. A) Steady-state fluorescence spectra of Py-PAlaGlus (f_{Ala} = 0.58) (from bottom to top: x = 0.07, 0.10, 0.14, 0.15) and B) the I_E/I_M ratios for the Py-PGlu and Py-PAlaGlu samples in DMSO as a function of pyrene content x. f_{Ala} = (○) 0, (□) 0.24, (◇) 0.41, and (△) 0.58. The dashed line was added to guide the eyes.

Blob Size: The size of a *blob* is characterized by N_{blob}, the number of *aa*'s contained within the volume of a *blob*, which was calculated from Equation 6.2. Since the same linker was used to connect the pyrenyl label to the backbone of all the copolypeptides, namely 1-pyrenemethylamine covalently attached onto Glu, and all the polypeptides were in a coiled conformation due to their racemic composition, N_{blob} provided a direct measure of the conformational freedom of the copolypeptide backbone resulting from the inclusion of alanine. Figure 6.4A shows that for the Py-PAlaGlu and Py-PGlu samples in DMSO, N_{blob} remained

constant as a function of pyrene content, indicating that the *blob* size was independent of the pyrene content.

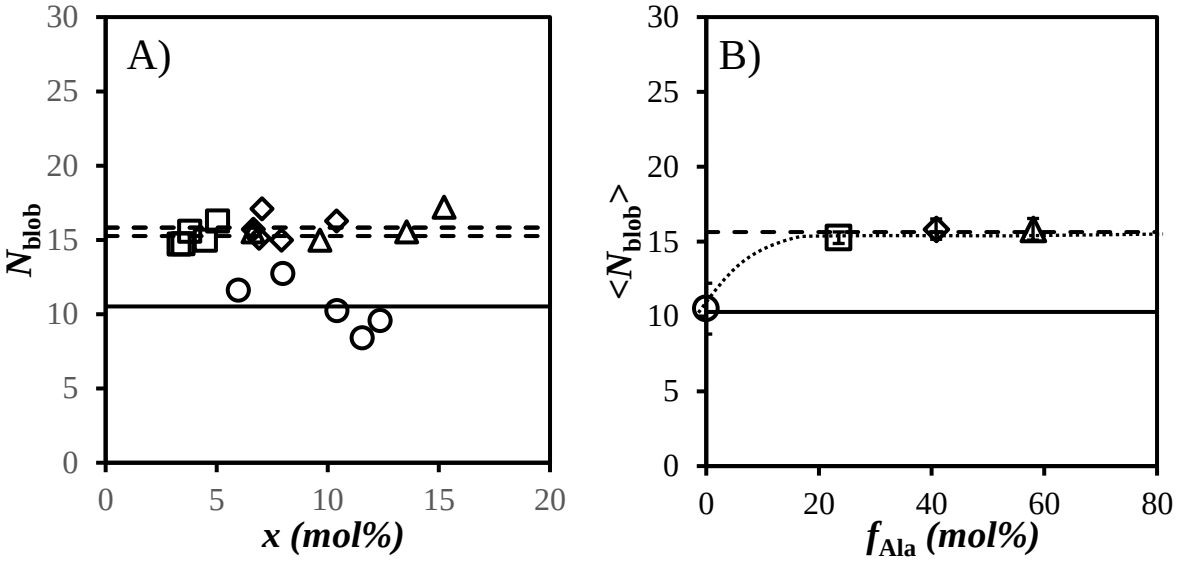


Figure 6.4. Plot of A) N_{blob} as a function of pyrene content and B) the average N_{blob} value as a function of polypeptide alanine content for the Py-PGlu and Py-PAlaGlu samples in DMSO. f_{Ala} = (○) 0, (□) 0.24, (◇) 0.41, and (△) 0.58. The dashed lines represent A) the N_{blob} values for a given copolypeptide averaged over all pyrene contents and B) the average N_{blob} value (15.6 (±0.8)) of the Py-PAlaGlu samples. The solid line indicates the N_{blob}^{theo} value of 10.3 (±1.2) for a rigid polypeptide in a coiled conformation.

Figure 6.4A shows that N_{blob} was the lowest for Py-PGlu, taking an averaged value of 10.5 (±1.7) matching the theoretical N_{blob}^{theo} value of 10.3 (±1.2) determined from molecular mechanics optimizations (MMOs).[23] In the MMOs, the N_{blob}^{theo} value is representative of a

polypeptide whose backbone is elongated over a length scale of several nanometers and is independent of the *aa* sequence.[25] Figure 6.4A shows that the Py-PAlaGlu samples had significantly larger N_{blob} values than N_{blob}^{theo}, indicating that the incorporation of alanine into the Glu backbone enhanced the backbone conformational freedom of the copolypeptides. In order to characterize how the alanine content affected the *blob* size of the copolypeptides, the average N_{blob} value of the Py-PGlu and Py-PAlaGlu samples was plotted as a function of the alanine content (f_{Ala}) in Figure 6.4B. The trend in Figure 6.4B was rather interesting as it showed that all the Py-PAlaGlu samples had similar N_{blob} values equal to 15.6 ± 0.8 independent of alanine content. This indicated that the *blob* size, and therefore the conformational freedom, of the copolypeptides was independent of the alanine content above 24 mol%.

This result suggested that relatively low amounts of alanine incorporated into the Glu backbone were able to fully disrupt the elongated (on the length scale of a *blob*) PGlu coil. Alanine, with its small side chain, provided enhanced rotational freedom to the sections of the backbone where it was incorporated, which likely allowed the polypeptide backbone to adopt a more condensed conformation. The constant N_{blob} value in Figure 6.4B indicated that only a few alanines needed to be incorporated for the polypeptide to locally adopt this denser conformation. The ability of the conformational density of the polypeptide chain to be controlled by simply adding a few points of enhanced conformational freedom at random locations in the backbone suggests that the number of *aa*'s which can interact with one another

in a *blob* is controlled by the *aa*'s with the smallest side chains (and greatest conformational freedom) found within a *blob*.

Encounter Rate within a Blob: The diffusive encounters between two pyrenyl labels located inside a same *blob* are described by the rate constant k_{blob}. Since the pyrenyl labels are bound to the Glu residues, k_{blob} represents the frequency at which two Glu's interact with each other inside a *blob*. A plot of k_{blob} as a function of pyrene content for the pyrene labeled polypeptides is given in Figure 6.5A. Figure 6.5A shows that k_{blob} remained constant within experimental error as a function of pyrene content for each pyrene-labeled polypeptide, again indicating that pyrene content did not affect the results obtained by the FBM.

Figure 6.5B shows that the average k_{blob} value decreased continuously with increasing f_{Ala} from 13.0 (±1.1) μs^{-1} for Py-PGlu to 9.0 (±0.9) μs^{-1} for Py-P$Ala_{58}Glu_{42}$. The decrease in $<k_{blob}>$ was reasonable when considering the larger N_{blob} values obtained upon incorporating alanine into the PGlu backbone. Larger N_{blob} values reflected *blobs* that were either larger or denser, both effects contributing to the decrease in k_{blob} observed in Figure 6.5B. Nevertheless, the ~30% decrease in $<k_{blob}>$ observed with f_{Ala} increasing from 0 to 0.58 was similarly modest as the ~50% increase in N_{blob} shown in Figure 6.4B.

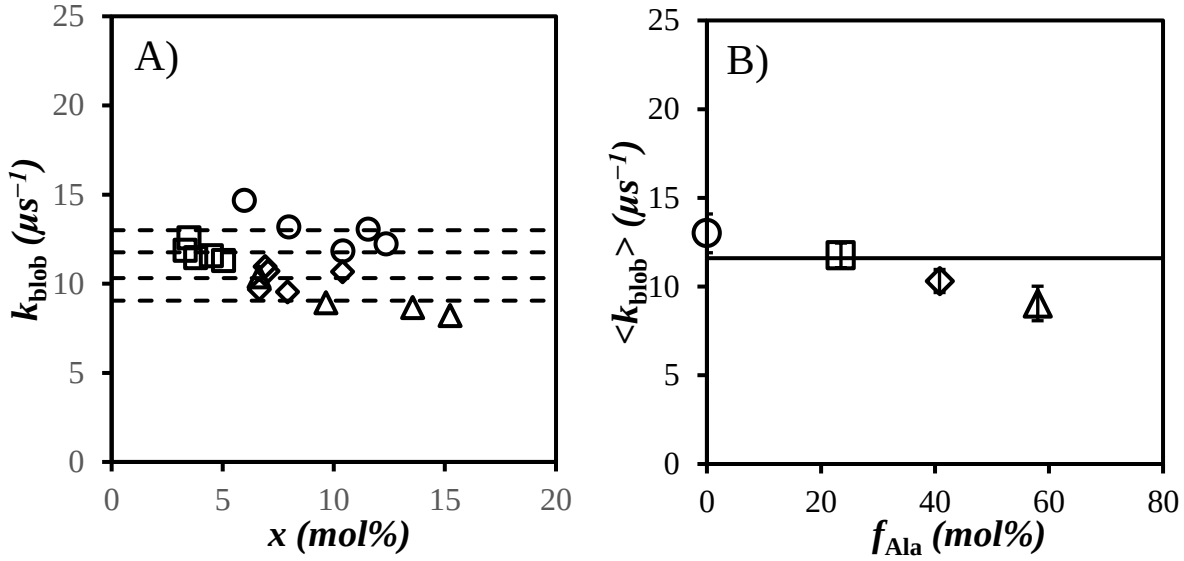


Figure 6.5. Plot of A) k_{blob} as a function of pyrene content of the Py-PGlu and Py-PAlaGlu samples in DMSO and B) the average k_{blob} value as a function of polypeptide alanine content. f_{Ala} = (○) 0, (□) 0.24, (◇) 0.41, and (△) 0.58. The dashed lines represent the k_{blob} values for a given copolypeptide averaged over all pyrene contents and the solid line is the average k_{blob} value of copolypeptides composed of ca. 44% Glu determined in a previous study.[24]

Blob Dynamics: The product $k_{blob}\times N_{blob}$ reflects the frequency of contacts between all *aa*'s within a *blob*, and therefore provides a direct measure of the internal dynamics of a polypeptide. A plot of $k_{blob}\times N_{blob}$ as a function of pyrene content is given in Figure 6.6A for the Py-PGlu and Py-PAlaGlu samples.

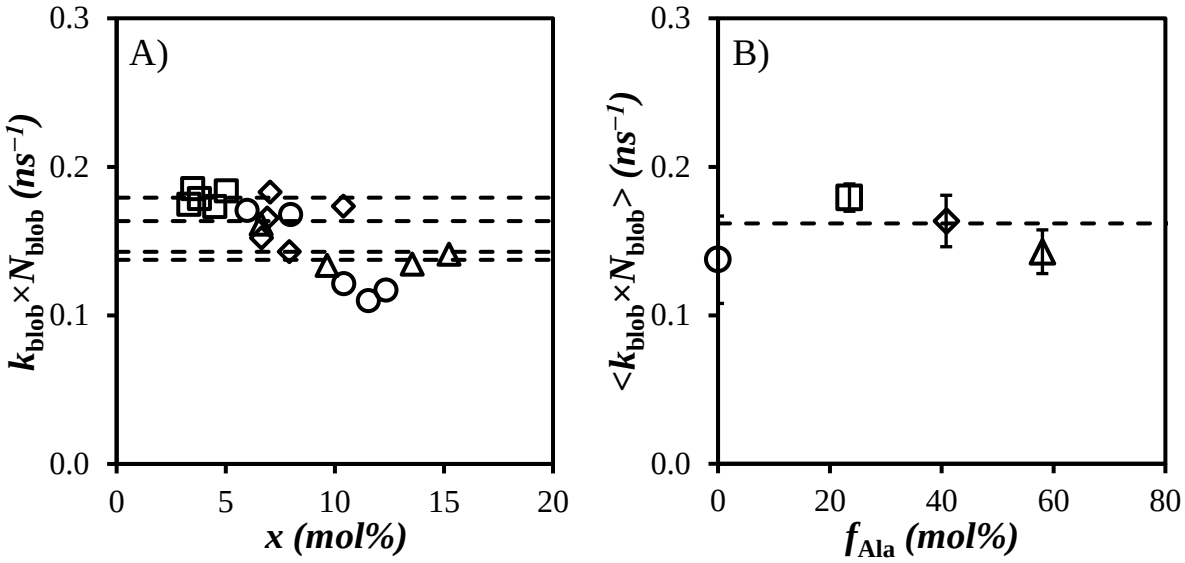


Figure 6.6. Plot of A) $k_{blob} \times N_{blob}$ as a function of pyrene content and B) the average $k_{blob} \times N_{blob}$ product as a function of copolypeptide alanine content for the Py-PGlu and Py-PAlaGlu samples in DMSO. f_{Ala} = (○) 0, (□) 0.24, (◇) 0.41, and (△) 0.58. The dashed lines represent the averages.

Figure 6.6A shows that although scattered, the product $k_{blob} \times N_{blob}$ did not change much as a function of pyrene content. To compare the frequency of contacts experienced by the *aa*'s constituting a *blob* between the different copolypeptides, the $k_{blob} \times N_{blob}$ values were averaged and plotted as a function of alanine content in Figure 6.6B. $<k_{blob} \times N_{blob}>$ equaled 0.14 (±0.03) ns^{-1} for Py-PGlu, while it took an average value of 0.16 (±0.02) ns^{-1} for the Py-PAlaGlu samples. Within experimental error, the $<k_{blob} \times N_{blob}>$ values remained relatively unchanged with alanine content. The constancy of the frequency of *aa*'s encounters deduced from $<k_{blob} \times N_{blob}>$ combined with the increase in N_{blob} in the presence of 24 mol% of Ala indicates that frequency of encounters between *aa*'s does not change much with f_{Ala}.

6.5 Discussion

The two main results obtained from this and an earlier study regarding the effect of polypeptide composition on N_{blob} are illustrated in Figure 6.7. The trend of N_{blob} as a function of *aa* side chain size (SCS) is shown in Figure 6.7A. It was obtained with a series of copolypeptides prepared from racemic mixtures of Glu and a selected few other *aa*'s, namely glycine (Gly), alanine (Ala), and N_{ε}-carbobenzyloxy lysine (Lys(Z)). Figure 6.7A led to the conclusion that the side chain size of an *aa* was one of the primary factors which affected the *blob* size for a polypeptide. Since there was only a significant increase in N_{blob} when alanine and glycine were incorporated, and since they have the smallest side chains of all *aa*'s, it was concluded that the majority of *aa*'s would have similar N_{blob} values in the range of 11 to 14. Therefore, a simplistic view would then be that N_{blob} for a polypeptide may be expected to be about 12 *aa*'s regardless of its *aa* sequence unless alanine or glycine is present, which would then result in a larger N_{blob} value. The question then became, by how much would the *blob* size increase when a set amount of alanine or glycine would be incorporated in a copolypeptide? The current study suggests, that the *blob* size is independent of the alanine content in Figure 6.7B (which is a reproduction of Figure 6.4B) as long as the alanine content remains above ca. 24 mol%. This conclusion suggested that the mere presence of some alanines within a *blob* was enough to enlarge the *blob* size to a value that remained constant with alanine content. Based on these results, it may be expected that N_{blob} for polypeptides would range between 11 and 14 *aa*'s in size unless Ala or Gly were present in a *blob* at which point the *blob* size would increase to 16 or 23 *aa*'s,

respectively. In fact, these conclusions suggested that a *blob*-based view of the internal dynamics and conformations of polymers in solution, such as those derived from applying the FBM, might result in relatively simple rules that could be taken advantage of to implement a mathematical procedure to predict the folding time of proteins. The following sections describe how such a procedure could be designed.

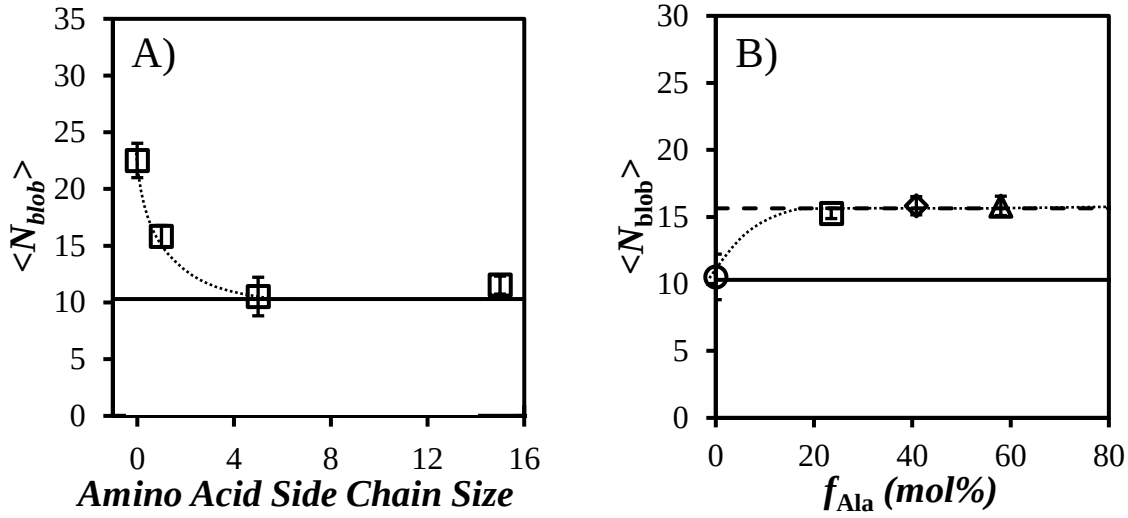


Figure 6.7. Plots of the average N_{blob} value as a function of A) the *aa* side chain size (SCS) and B) the alanine content of the PAlaGlu copolypeptide in DMSO. The copolypeptides in A) were prepared with a ~44:56 molar ratio of *D*,*L*-glutamic acid-to-another racemic comonomer.[25] The SCS of 0, 1, 5, and 15 corresponds to glycine, *D*,*L*-alanine, *D*,*L*-glutamic acid, and *D*,*L*-N_{ε}-carbobenzyloxy lysine used as a comonomer. The solid line represents the $N_{blob}^{theo} = 10.3 \pm 1.2$ value for a rigid polypeptide in a coiled conformation.

Blob Model and Polypeptide Folding: Levinthal's paradox states that the number of conformations available in the conformational space of a protein far exceeds the number of conformations that can be searched by a polypeptide backbone during the time a protein folds.[31] In its simplest form, the estimated folding time of a protein may be expected to scale as 3^N, where *N* is the number of *aa*'s in the sequence of a protein and 3 represents the degrees of freedom of each *aa*.[31] This approach quickly fails since it predicts folding times far greater than those observed experimentally. Despite this breakdown, the size of a protein is still an important parameter, as demonstrated by the rather good correlation between the square root of the sequence length of a protein and its folding time.[10] To circumvent this second paradox, the conformational search of the polypeptide backbone has been suggested to take place within subvolumes of the polypeptide coil that encompass smaller segments of the protein. These subvolumes have been identified as *foldons* along the folding pathway of a protein[18] or *blobs*[32] to define the number of *aa*'s capable of diffusively interacting with one another and therefore fold together on a similar time scale. From a *blob*-based perspective, the *aa*'s within a *blob* will diffusively interact with one another resulting in preferential interactions, beginning the folding process. This process occurs simultaneously throughout all the *blobs*. As the *aa*'s in an individual *blob* begin to fold, the *blobs* must also reorient with respect to each other in order for the protein to reach its native conformation. Based on this model of folding, the number of conformations that a polypeptide must search through as it folds into its native conformation may be significantly reduced as discussed below.

Conformations of a Polypeptide: In order to estimate the relevant number of conformations a protein can adopt based on the above description of folding, two parameters must be known, namely, the *blob* size (N_{blob}) and the number of *blobs* (n_b) contained in a protein. If these values are known, they can be used to estimate the number of conformations that a number N_{blob} of *aa*'s can adopt within a *blob* and that a number n_b of *blobs* can adopt among themselves. To do this, the renormalization group (RG) theory was employed. RG theory provides a simple equation for estimating the number of conformations that a self-avoiding polymer with N segments can adopt as the segments arrange themselves on a lattice.[33–35] In Equation 6.3, C depends on the coordination number of the lattice, γ depends on the dimensionality, and μ is the effective coordination number.

$$\Omega(N) = CN^{\gamma-1}\mu^N \tag{6.3}$$

On a 3-dimentional simple cubic lattice, the constants C and γ are equal to 1.17 and 7/6, respectively.[34,36] Since μ is equal to the degree of freedom of each chain segment, it depends on the solvent quality and takes a value as high as 4.68 in an isothermal solvent[33,34] to as low as 3.19 in a poor solvent.[37] Although Equation 6.3 was derived for polymers with many chain segments,[34] it turns out that Equation 6.3 still provides a very good estimate even if the polymer contains few chain segments (see SI). It should thus remain valid to predict the number of conformations adopted by the small number of segments encompassed inside a polypeptide *blob*.

To calculate the number of conformations a protein can adopt, Equation 6.3 must be modified to estimate the number of conformations among a number N_{blob} of *aa*'s in a *blob* or among the n_b *blobs* in a polypeptide. To do this, Equation 6.3 is rewritten in terms of the generic value *n*, which is equal to the number of objects (whether they be *aa*'s or *blobs*) in a chain. For simplicity sake, the modification of Equation 6.3 is discussed in terms of the number of *aa*'s within a *blob*, but a similar argument can be made in terms of the number of *blobs* within a polypeptide. A *blob* containing *n aa*'s is considered first, where each *aa* occupies a single cell in the lattice. For each self-avoiding walk of the *aa*'s in a *blob*, there will be *n* filled sites. Since the number of segments (*N* in Equation 6.3) is equal to the number of filled sites minus 1, a *blob* of size *n* will generate $n - 1$ segments. Therefore, Equation 6.4 was defined in terms of the number of *aa*'s (or filled sites) by the change in variable of *N* in Equation 6.3 to *n* by using the identity $N = n - 1$.

$$\Omega(n) = C(n-1)^{\gamma-1}\,\mu^{n-1} \qquad (6.4)$$

Equation 6.4 is built from a lattice model, which means that the number of conformations predicted by Equation 6.4 includes the conformations of *n aa*'s which are similar but differ in the direction in which the chain initially propagates in the 3D lattice (i.e. there are conformations which are superimposable but rotated about the origin). This can be seen by considering the population of conformations which are similar but differ in the lattice-positioning of the second *aa* in the *blob*. The first *aa* occupies the first site and defines the

starting point. Since the second *aa* must be placed in an adjacent site, there are μ sites available (μ is the coordination number). This means that there are μ possible directions in which the second *aa* can be placed, each defining the direction the chain will initially propagate. If the remainder of the $n - 2$ *aa*'s share the same conformational arrangement with respect to themselves, this results in μ conformations which differ only by a rotation about the starting point. Since it is only the number of conformations with respect to the *aa*'s in a *blob* and not their absolute orientation in space which matters (i.e. the orientation of a folding chain in space does not affect the number of conformations it must search to fold), the rotational isomers for each conformation projected along the initial μ directions should not be counted towards the conformational search of the folding *aa*'s. Therefore, the number of unique (or relevant) conformations (Ω_u) a *blob* containing *n aa*'s can adopt (or a chain containing *n* objects) is defined in Equation 6.5 as $\Omega(n)/\mu$.

$$\Omega_u(n) = \frac{C(n-1)^{\gamma-1}\,\mu^{n-1}}{\mu} = C(n-1)^{\gamma-1}\,\mu^{n-2} \qquad (6.5)$$

Since proteins typically collapse into compact structures, the μ term in Equation 6.5 is set equal to 3.19 since a polymer in a poor solvent may be expected to experience a similar conformational freedom to that of the dense globular conformations experienced by a folding protein. All the constants in Equation 6.5 are known (C, γ, and μ). Consequently, the only variable that remained to be determined to calculate the number of conformations in a *blob* is the number ($n = N_{blob}$) of *aa*'s contained in a *blob*. Connecting Equation 6.5 back to the FBM,

the number (Ω_B) of conformations that a *blob* containing N_{blob} *aa*'s can take is equal to $\Omega_u(N_{blob})$, where $n = N_{blob}$ in Equation 6.5. As the *aa*'s within individual *blobs* simultaneously interact/fold with one another, the *blobs* must also orientate themselves with respect to each other. Therefore, the *blobs* must also undergo a conformational search among themselves for a polypeptide chain to completely fold. Following a similar logic, Equation 6.5 can also be used to define the number of conformations *n blobs* can adopt among themselves. Therefore, if a polypeptide were to contain n_b *blobs*, the number of conformations (Ω_M) among those *blobs* would be equal to $\Omega_u(n_b)$ (where $n = n_b$ in Equation 6.5). By viewing a polypeptide as a string of *blobs* and applying the so-called *blob* approach to folding, each *blob* would generate Ω_B possible conformations while there would be Ω_M conformations among the *blobs*. Since the *aa*'s within each *blob* simultaneously interact with one another, the total number of conformations a polypeptide may be expected to adopt could be approximated by the product $\Omega_B \times \Omega_M$.

Size of a Polypeptide Blob: The *blob* size of a coiled macromolecule is affected by the distance separating the pyrenyl label from the polypeptide backbone[24,38] as well as the structure and composition of the structural units constituting a *blob*.[25,39] For example, the 7 atom-long linker connecting pyrene to a randomly coiled poly(*L*-lysine) resulted in a *blob* size of 14 *aa*'s while the 5 atom-long linker connecting pyrene to the randomly coiled PGlu backbone resulted in a smaller *blob* size equal to 10 *aa*'s.[24] The use of 1-pyrenemethylamine to label the polypeptides extends the glutamic acid side chains by only one methylene group and the pyrenyl derivative. This represents one of the shortest possible side chain extensions for labeling with a dye

derivative, which should minimize the effect that 1-pyrenemethylamine has on the *blob* size. In addition to the linker length, it has been demonstrated that the *blob* size of 10 expected for a pyrene-labeled glutamic acid in PGlu depends also on the nature of the neighbouring residues, as was illustrated in Figure 6.7A. Since most *aa*'s have side chain sizes similar to that of Glu,[25] the N_{blob} value of a pyrene-labeled Glu provides a good starting point for these calculations. Therefore, a polypeptide *blob* is expected to be about 10 *aa*'s in size, but if Ala or Gly is present, the *blob* size increases to a value of 16 or 23, respectively. The current study demonstrates that the increase in N_{blob} due to the presence of these small *aa*'s is independent of the amount of Ala, and probably Gly, which is present in the *blob* (above 24 mol% for Ala) and therefore the size of a polypeptide *blob* can be expected to be controlled by the presence of a few Ala and Gly residues.

As a first approximation, the *blob* size of a polypeptide was taken to equal 19 *aa*'s as the average of the two N_{blob} values obtained for PAlaGlu (N_{blob} = 15.8) and PGlyGlu (N_{blob} = 22.5). Implicit with this approximation is that each *blob* should contain at least one Gly or Ala to generate the necessary conformational freedom resulting in a larger N_{blob} value. As it turns out, 7.8 and 7.3 mol% of the *aa*'s constituting a protein are expected to be alanines and glycines, respectively.[30] This means that a *blob* constituted of 19 *aa*'s should have a greater than 95% probability of having 3 *aa*'s being either Ala or Gly, thus validating the assumption.

Based on these considerations, a polypeptide was approximated as a string of *blobs*, each containing 19 *aa*'s. Since many proteins contain a large number of *aa*'s, their sequence length is much larger than a *blob* and the chain ends represent a small component of the overall

chain and therefore contribute very little to the *blob* size. However, when the chain length is relatively short (on the order of 1 to 3 times the *blob* size),[29] the *blob* size of a polypeptide is expected to decrease since the *aa*'s near the chain ends can only interact with *aa*'s in the chain. To account for the change in N_{blob} due to the chain length, Equation 6.6 was previously derived for polymers with a degree of polymerization (DP) $\geq N_{blob}$.[29] In Equation 6.6, N_0 is the largest number of *aa*'s separating two pyrene-labeled *aa*'s from one another while still allowing pyrene-pyrene encounters for pyrene excimer formation.

$$< N_{blob} >= \frac{1}{DP}\left(-N_0^{\,2} - N_0 + DP\left(2N_0 + 1\right)\right) \qquad (6.6)$$

Using a pyrene-labeled *aa* as a reference point, the pyrenyl labels can form excimer with other pyrene-labeled *aa*'s up to N_0 structural units away. Since the pyrenyl pendant used as a reference can form an excimer with another pyrenyl label located on either side of the reference point, N_{blob} is defined as $2N_0 + 1$, where the *'+ 1'* accounts for the reference label.[29] Using the estimated N_{blob} value of 19, an average *blob* in a polypeptide can be expected to have an N_0 value of 9 *aa*'s. Using an N_0 value of 9, Equation 6.6 provides an estimate of the average *blob* size of any polypeptide based on its sequence length or *DP*.

Estimating a Protein Folding Time: Based on the above discussion, the number of conformations a polypeptide can adopt is calculated as $\Omega_{Total} = \Omega_B \times \Omega_M$, which depends only on the *blob* size and the number of *blobs* contained within a polypeptide. If folding is a result of random diffusive encounters between local *aa*'s, then the average folding time of a protein

(τ_F) would be the sum over all possible conformations (Ω_{Total}) of the time (τ_p) it takes to probe a given conformation. If it is assumed that τ_p is similar for each conformational state, then the folding time is simply equal to $\Omega_{Total} \times \tau_p$, where τ_p is the effective conformational search time. Therefore, τ_F may be estimated with Equation 6.7.

$$\tau_F = \Omega_B(<N_{blob}>) \times \Omega_M(DP/<N_{blob}>) \times \tau_p \qquad (6.7)$$

Equation 6.7 requires two factors; the length or *DP* of a protein and the size of its *blobs*. Using the estimated N_0 value of 9 and the sequence length of a protein (*DP*), the $<N_{blob}>$ value of different proteins was calculated by applying Equation 6.6. Using the calculated $<N_{blob}>$ value, the number of conformations, that the *aa*'s within a *blob* can adopt, is equal to $\Omega_B = \Omega_u(<N_{blob}>)$ from Equation 6.5. Since each *blob* contains $<N_{blob}>$ *aa*'s, the protein will contain $n_b = DP/<N_{blob}>$ *blobs* and therefore, the number of conformations among *blobs* is equal to $\Omega_M = \Omega_u(n_b)$. Using a $\tau_p = 1 \times 10^{-12}$ s which is typically assumed for the conformational search of one *aa*,[40-43] Figure 6.8 compares the calculated τ_F's to the experimentally measured τ_F's ($= 1/k_f$) of 145 different proteins.[44,45] A list of the calculated $<N_{blob}>$ and τ_F values obtained for these proteins are given in the SI.

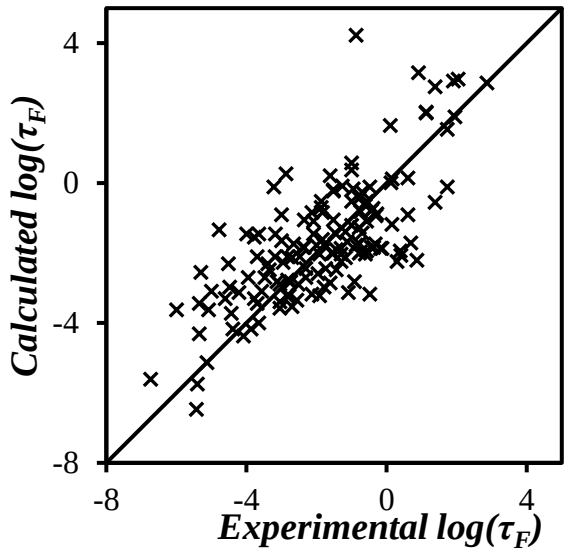


Figure 6.8. Log-log plot of protein folding times calculated by applying a *blob* model approach as a function of experimentally determined protein folding times. The solid line represents the line of equality. The correlation coefficient $r = 0.73$.

Strikingly, Figure 6.8 shows a rather good agreement between the experimental τ_F's and those estimated using a *blob*-based approach over 8 orders of magnitude. This good agreement is all the more surprising when considering that no parameter needed to be adjusted for the τ_F calculations based of Equation 6.7. The total number of configurations used in Equation 6.7 is defined by $<N_{blob}>$, whose expression is based on N_0, which was set to equal 9 *aa*'s in Equation 6.6, and the number of *blobs* (n_b), which is given by the ratio $DP/<N_{blob}>$ and is thus defined by the length of the protein sequence. The last parameter for the calculation is τ_P, the time taken by an *aa* to probe one conformation, which was set to equal 1 ps based on an assumption made in several reports.[40-43] Yet, if τ_P is optimized to find a better match

between the calculated and experimental τ_P values shown in Figure 6.8, the optimization returns a τ_P value of 0.98 ps, which is essentially 1 ps and justified its use in the τ_P calculations.

Figure 6.8 shows that the calculated τ_F values were correlated quite strongly to the experimentally measured values with a correlation coefficient r of 0.73. Additionally, the datapoints were scattered about the equality line, demonstrating that the folding times predicted by the FBM were not only correlated but also scaled directly with the experimentally determined values. On average, the calculated folding times were able to predict the experimentally observed τ_F's within 1.05 ($\pm$0.84) orders of magnitude. Furthermore, the proteins used in Figure 6.8 include a wide range of protein types with sequence length ranging from 16 to 307 *aa*'s, including two- and three-state proteins,[4,7] as well as α, β, and mixed proteins,[7,15] whose secondary structure is composed primarily of α-helices, β-strands, and mixtures of the two, respectively. The ability of the proposed *blob*-based approach to predict the folding times of such a wide range of protein types supports the idea that the folding times of proteins are defined by the encounters between *aa*'s taking place within local subdomains.

Comparison with Published Trends: Numerous studies can be found in the literature that successfully correlate the folding time (or folding rates) of proteins with physical and empirical parameters. Correlation coefficients range from 0.6 to nearly unity,[1,7,46] the latter value indicating that the model can perfectly predict folding rates. Typically the highest correlations (0.75 to 0.99)[7,44–48] are obtained using machine learning algorithms, which optimize correlations between calculated and experimental parameters. Despite these very high correlations, the generated models have no physical basis and therefore can only provide

information on the parameters that might be important to folding, but not as to why they are important.[1] Furthermore, the statistical relevance of these methods was questioned in several reviews due to model overfitting, because of the relatively small data sets and overtraining, which leads to great correlations for the included set, but often fails when applied to larger data sets.[1,49]

The predictive power of physical models yields slightly lower correlation coefficients in the range of 0.6 to 0.9, but typically less than 0.8.[1,15,46,49,50] Often these models are only applied on an individual class of proteins, which requires sorting proteins into groups based on their folding kinetics (two- vs multi-state) and structural makeup (α, β, and mixed). This leads to rather small datasets of ca. 20 – 60 proteins, resulting in few correlations that can be generalized to larger data sets such as the one used in Figure 6.8. In larger datasets (< 80 proteins), the correlations between physical parameters and folding rates are typically represented by correlation coefficients smaller than 0.75. Larger correlations with higher correlation coefficients (> 0.8) are often reported, but these are typically only obtained by adding empirical scaling factors to the physical data, which are then optimized to obtain the best correlation.[8,51] For example, the local topography of a folded protein is often parametrized using the absolute contact order (ACO), which is determined by the native contacts in a proteins structure. Correlation coefficients of ACO and folding rate have been reported to be as high as 0.87 for 107 proteins when additional empirical factors were used.[51] However, without such factors the correlations are in the 0.69 to 0.73 range[1,3] depending on the dataset. Another example used a similar topological parameter related to contact yielded a correlation

coefficient of 0.69 across 95 proteins.[46] The 0.73 coefficient determined in Figure 6.8 may not be the largest correlation coefficient in the literature, but it demonstrates that a *blob*-based approach is capable of predicting folding times as effectively as many other methods.

Despite their similar predictive power, the *blob*-based method described in this report approaches the prediction of protein folding time from a perspective that is completely different from all other methods. In the present work, folding times are predicted based on the size of the folding domains/*blobs* of proteins estimated solely from the experimentally determined local conformations and dynamics of polypeptides, without using any structural information known about the native proteins. Contrary to any other methods which assess the correlations between folding times and different parameters such as protein size or ACO, the experimentally determined folding times were compared directly with the predicted folding times and a 1:1 correspondence was obtained. Considering the crudeness of the proposed approach, the surprisingly good agreement obtained in Figure 6.8 between the predicted and experimentally determined folding times suggests that *blob*-based approaches might be ideally suited to determine the folding times of macromolecules.

Limitations and Future Refinements: The already good agreement found in Figure 6.8 between calculated and experimental τ_F's might be further improved by reducing the number of assumptions currently needed in the τ_F calculations. These improvements would involve assessing how N_{blob} is affected by 1) the side chain length of all *aa*'s, 2) the presence of the 1-pyrenemethylamine label that extends the reach of the Glu side chains, and 3) the heterogeneity encountered in the sequence of proteins. As discussed above, the N_{blob} values used for the τ_F

calculations in Figure 6.8 were determined by labeling the Glu with 1-pyrenemethylamine in a series of copolypeptides. Since N_{blob} is known to depend on the length of the linker separating the pyrenyl label from the polypeptide backbone, these N_{blob} values are only representative of Glu residues labeled with this specific pyrene derivative. Therefore, a future study must be conducted to quantify the effect that the pyrene derivative used to label the polypeptide has on the N_{blob} value so that it can be accounted for. However, since the 1-pyrenemethylamine label extends the Glu side chain by only one methylene unit (and the pyrene molecule) the results obtained in this report with 1-pyrenemethylamine are expected to closely reflect the behaviour of the Glu side chain itself and therefore in this particular case, the N_{blob} value would not be expected to change much.

Since N_{blob} depends on the length of the linker connecting the pyrenyl label to the polypeptide backbone,[38] the N_{blob} values would also be expected to change with each *aa* since each *aa* has a different side chain onto which a pyrene derivative would have to be attached. This means that each *aa* would be expected to have its own characteristic N_{blob} value depending on its side chain length. Consequently, a *blob*-based approach would benefit greatly if the dependency of N_{blob} on the *aa* side chain length was determined. The current study used Glu to determine the N_{blob} values of the copolypeptides. As mentioned earlier, Glu happens to have a side chain whose size is comparable to that of many other *aa*'s. This means that although the N_{blob} values used in the τ_F calculations were not exact, these N_{blob} values should be representative of many other *aa*'s.

Perhaps the greatest limitation in the current calculation is the assumption that each *blob* in a protein is composed of a same number N_{blob} of *aa*'s. Since the size of a *blob* depends on the local *aa* sequence, a protein must be composed of *blobs* of varying size. The current study demonstrated that the *aa*'s with a small side chain resulted in a larger N_{blob} value that did not change with the number of small *aa*'s being incorporated inside a *blob*. Furthermore, the incorporation of several flexible *aa*'s (~ 4 out of 16 *aa*'s or ~25 % of 20 *aa*'s) was sufficient to fully disrupt the rigid polypeptide backbone. Assessing in a more systematic manner how these smaller *aa*'s affect N_{blob} would most certainly help refine the predictions made with a *blob*-based approach. The results obtained from the aforementioned improvements could then be incorporated in the *blob*-based approach to provide a detailed map of the different *blob* sizes associated with each *aa* of the sequence of a protein, which in turn would then be used to refine the calculations and obtain more precise folding times.

6.6 Conclusions

The size and dynamics of different polypeptide *blobs* were characterized as a function of the Ala content of different series of Py-PAlaGlu samples. Aside from an initial increase in the *blob* size from 11 (±1) for PGlu to 16 (±1) for $PAla_{24}Glu_{76}$ with a relatively small Ala content, N_{blob} of the copolypeptides did not change as a function of alanine content up to an f_{Ala} value of 0.58. The dynamics of the polypeptide *blobs*, monitored through the term $k_{blob} \times N_{blob}$, showed no change between 0.14 (±0.03) ns^{-1} for PGlu and 0.16 (±0.02) ns^{-1} when averaged for all PAlaGlu samples. These results suggest that the flexibility, and therefore the *blob* size, of

polypeptides is controlled primarily by the smallest and most flexible *aa*'s, even if those *aa*'s are present in relatively low amounts.

Since alanine and glycine are the smallest *aa*'s, they are expected to be the two most important *aa*'s for determining the *blob* size in proteins. The N_{blob} values of 16 (±1) and 23 (±2) for PAlaGlu and PGlyGlu, respectively, were averaged to give an estimated *blob* size of 19 *aa*'s. Considering that proteins fold through localized domains, the sequence length of a protein and its *blob* size estimated to equal 19 *aa*'s were used to estimate the folding time of 145 different proteins and peptides. The calculated and experimentally determined folding times were found to agree quite well with each other as demonstrated by the 0.73 correlation coefficient between the folding times of the 145 proteins spanning 8 orders of magnitude in Figure 6.8. This procedure, which approaches the prediction of the folding times of protein from estimates of experimentally determined folding subdomains, represent a complete departure from the types of parameters employed in all other procedures currently applied to estimate τ_F's. Yet and considering its rather crude implementation, the trend shown in Figure 6.8 suggests a strong and direct correlation between the predicted and experimentally determined τ_F's. The ability to predict the folding time of proteins based simply on their sequence length and *blob* size appears to be a worthy path toward predicting the τ_F's of proteins and provides further support that the *blobs* determined through the FBM could actually represent the domains of *aa*'s where cooperative folding occurs in proteins.

Chapter 7

Blob-Based Predictions of Protein Folding Times from the Amino Acid Dependent Conformation of Polypeptides in Solution

Adapted with permission from Casier, R.; Duhamel, J. *Blob*-Based Predictions of Protein Folding Times from the Amino Acid-Dependent Conformation of Polypeptides in Solution. *Macromolecules* **2021**, *ASAP*, DOI: 10.1021/acs.macromol.0c02617.

7.1 Abstract

A combination of fluorescence *blob* model (FBM) and molecular mechanics optimizations was applied to determine the number (N_{blob}) of amino acids (*aa*'s) located in the volume probed by an excited pyrene for poly(*L*-lysine) labeled with 1-pyrenebutyric acid (PyBu-P*L*L), as it adopted an extended conformation in DMSO. The N_{blob} value of PyBu-P*L*L was then compared to those of other pyrene-labeled polypeptides adopting a similarly extended conformation over the length scale of a *blob* to quantify the contribution of the pyrenyl label to N_{blob}. After subtracting the pyrene contribution, the plot of the corrected N_{blob} as a function of *aa* side chain reach (SCR) led to the conclusion that, for a polypeptide in an extended conformation, increasing the SCR of an *aa* by one bond resulted in a ~1.8 *aa* increase in N_{blob}. This information could then be applied to predict the intrinsic N_{blob} value (N(SCR)) of any *aa* based on its SCR as long as this *aa* was part of an *aa* sequence in an extended conformation. The increase in N_{blob} resulting from the conformational freedom imparted by smaller *aa*'s to a polypeptide backbone was accounted for by multiplying N(SCR) with the bending function f_b(SCS), which was determined experimentally for the different side chain sizes (SCS) of *aa*'s. By scanning the sequence of a protein, an N_{blob} value was calculated for each *aa* from its N(SCR) and f_b(SCS) values. The N_{blob} values were then sorted from largest to lowest and averaged to yield $<N_{blob}>$. Renormalization group theory was applied to determine the number (Ω) of conformations, that would be available to a protein based on its $<N_{blob}>$ value, and the average number of *blobs* found in the protein sequence. Multiplying Ω by the time required by an *aa* to adopt its possible conformations yielded the protein folding time (τ_F). A correlation

coefficient of 0.73 was obtained from the comparison of the calculated and experimentally determined τ_F values, demonstrating the ability of this model to predict the protein folding times within ± 1.4 orders of magnitude from the expected value.

7.2 Introduction

Recent improvements in our ability to predict the folding times of proteins have led to a greater understanding of the protein folding phenomenon.[1,2] These predictions have benefited from the identification of some key parameters closely associated with characteristic features of proteins, that best correlate with their folding times. In turn, these parameters have enabled an understanding of the many underlying physical phenomena, that control the folding kinetics and mechanisms of proteins.[1,2] Accounting for these physical phenomena has led to the implementation of a wide number of approaches based on theoretical,[3,4] thermodynamic,[5–8] and topological[9] considerations. Using classic topological correlations, Plaxico demonstrated in 1998 that the concept of relative contact order, which defines the average number of amino acids (*aa*'s) separating all pairs of contacting residues in the native structure of a protein normalized to its length, was correlated to the folding rate of two-state proteins.[9] Since then, there has been continuing developments in the procedures using the contact order and related contact topological parameters to predict the folding rates of proteins. Some of these descriptors include the average topological information,[10] cliquishness,[11,12] maximum intrachain contact entanglement,[13] absolute contact order,[14] geometric contact,[15] helical content,[16] long-range order,[17] and total contact distance.[18] These methods yielded strong

correlations between the different parameters of interest and experimental folding rates, resulting in correlation coefficients that were greater than 0.7. However, these methods are often applied to a single structural or folding kinetic class of proteins and they all require explicit information on the native state of proteins.[19,20] Another approach is to calculate these, and other, descriptors from databases of protein structures and combine the results to obtain an average value of each descriptor for a specific amino acid.[5,15,21–24] The agglomerated results are then used to predict the folding times of a specific protein based on its sequence without requiring the knowledge of its structure. However, this approach is still built on the correlations that are generated between the specific 3D structures of proteins and the selected defining parameter(s).

Besides the use of these structural parameters, the localized interactions and dynamics experienced by the *aa*'s constituting a protein also affect the ability of a protein to fold.[25] Therefore, a method capable of quantifying the dependency of the internal dynamics of a protein on its constituting *aa*'s would provide valuable insights into the protein folding process. One such method is based on the application of the fluorescence *blob* model (FBM) to probe the internal dynamics of polypeptides labeled with the dye pyrene. The FBM has been applied to predict the folding time of proteins based solely on parameters that were determined experimentally to describe the local dynamics and conformational freedom of polypeptides in solution.[26] Analysis of fluorescence decays of pyrene-labeled polypeptides with the FBM yields the number of *aa*'s, which interact locally with one another on the time-scale of molecular motions undergone by the polypeptide backbone in solution.[26–32] To do this, the

polypeptide backbone is compartmentalized into domains referred to as *blobs*, where each *blob* is defined by the volume that a pyrenyl pendant can probe while remaining excited.[33–35] Since pyrene is covalently bound to the side chain of an *aa*, a *blob* represents the subvolume within a polypeptide coil, that an *aa* bearing a pyrenyl label can probe. Therefore, the number (N_{blob}) of *aa*'s encompassed within each *blob* is equal to the number of residues which locally interact with one another inside such a domain.

To date, the FBM has been applied to probe the internal dynamics and local conformations of pyrene-labeled homopolypeptides, including poly(*L*-glutamic acid) (P*L*Glu),[27–29] poly(*D*,*L*-glutamic acid) (P*DL*Glu),[29,31] and poly(*L*-lysine) (P*L*L),[30,31] as well as copolypeptides of *D*,*L*-glutamic acid and glycine, *D*,*L*-alanine, or carboxybenzyl protected *D*,*L*-lysine (Lys(Z)), referred to as PGlyGlu, PAlaGlu, and PLyz(Z)Glu, respectively.[26,32] Since P*DL*Glu and the copolypeptides were all labeled with pyrene in the same manner, namely with a 1-pyrenemethylene (PyMe) derivative covalently attached onto a Glu via an amide bond, it provided the opportunity to explore how the *aa* composition of the copolypeptides affected the N_{blob} values of the pyrene-labeled Glu's. The increase in N_{blob} from 11.0 (±1.4) for PyMe-PLys(Z)Glu and PyMe-PDLGlu to 15.8 (±1.0) and 22.5 (±1.5) for, respectively, PyMe-PAlaGlu and PyMe-PGlyGlu demonstrated that in dimethylsulfoxide, the inclusion of alanine and glycine increased the flexibility and conformational freedom of the copolypeptide backbone compared to that of PGlu, resulting in an increase in the number of *aa*'s interacting with one another inside a *blob*.[32] Interestingly, N_{blob} was found to remain constant for a series of PyMe-PAlaGlu samples with alanine contents ranging from 24 to 58 mol%, which

suggested that the *blob* size was controlled by the presence of small *aa*'s in its sequence and not their level of incorporation in a *blob*.[26] Since proteins contain, on average, about 7.3 and 7.6% alanine and glycine, respectively, the combined ~15% probability of finding an alanine or glycine in any protein sequence led to the assumption that a protein *blob* was composed of 19 *aa*'s (N_{blob} ~ (15.8 + 22.5)/2 based on the N_{blob} values of 15.8 and 22.5 obtained for PyMe-PAlaGlu and PyMe-GlyGlu, respectively), as a *blob* made of 19 *aa*'s would have a greater than 95% probability of containing at least one glycine or one alanine. Using the idea that a protein folds through localized interactions between its constituting structural units, a procedure was developed to estimate the folding time of a protein based on the average N_{blob} value of 19 *aa*'s and the sequence length of a protein. The procedure was applied to 145 proteins resulting in a 0.73 correlation coefficient between the logarithms of the experimental and calculated folding times.[26] Although a strong correlation was obtained, the first iteration of the model was rather crude since it relied on a number of assumptions, which are described hereafter.

The first assumption was that the N_{blob} values used in these calculations were representative of all *aa*'s, although they were truly based on the response of Glu residues labeled with 1-pyrenemethylamine (PyMA). In fact, since the *blob* size depends on the linker length separating the pyrenyl label from the polymeric backbone,[28,36] each *aa* would be expected to have its own distinctive N_{blob} value with respect to its specific side chain reach, thus resulting in a distribution of N_{blob} values. Since PyMA was a pyrene derivative, that would minimally extend the side chain of the Glu residues used for pyrene-labeling, the second assumption was that PyMA would provide the most accurate description of the dynamics

experienced by a Glu residue. Of course, while PyMA would affect the motions of the Glu side chain in the least intrusive manner, the determined N_{blob} values still reflected the contributions from both the reach of the Glu side chain as well as the pyrenyl label itself. The third assumption was that every protein would have a similar *blob* size of ~ 19 *aa*'s regardless of its sequence based on the probability for a 19 *aa* sequence to contain at least one glycine or one alanine. However, since the size of a *blob* depends on the nature of the *aa*'s in the sequence of the oligopeptide located within the volume of a *blob*, the *blob* size is expected to fluctuate throughout the protein sequence.

The present study addresses these issues by proposing an improved *blob*-based procedure, that predicts the folding time of proteins by calculating N_{blob} for each *aa* in any given protein sequence, in a manner that accounts and corrects for the side chain reach (SCR) of each *aa*, the contribution of a pyrenyl label to the size of a *blob*, and the exact sequence of a protein. The improved *blob*-based procedure is implemented by designing polypeptides with well-defined features in terms of chemical composition, conformation, and length of spacer connecting the pyrenyl label to the polypeptide backbone. The fluorescence decays of these pyrene-labeled polypeptides were analyzed according to the FBM to retrieve the parameters required for the proposed *blob*-based procedure to calculate the size and number of *blobs* constituting a polypeptide. In turn, *blob* size and number of *blobs* were fed into a program to predict the folding times of 145 proteins, which were compared to those obtained experimentally. Consequently, this study illustrates how experiments can be designed to

determine the parameters needed in the implementation of *blob*-based approaches to better predict the folding times of proteins.

7.3 Experimental

Chemicals: Acetonitrile (ACN, Sigma, ≥ 99.9%), dichloromethane (DCM, Sigma, ≥ 99.8%), *N'*,*N'*-dicyclohexylcarbodiimide (DCC, Sigma, 99%), dimethyl sulfoxide (DMSO, Sigma, ≥99.9%), *N*,*N*-dimethylformamide (DMF, Sigma, ≥ 99.9%), 2-propanol (isopropanol, Sigma, 99%), *N*-hydroxysuccinimide (HOSu, Sigma, 98%), *N*,*N*-diisopropylethylamine (DIPEA, Sigma, ≥ 99%), poly(*L*-lysine) (P*L*L·HBr, Alamanda Polymers, M_n = 165,100 g·mol^{-1}, *DP* = 790, *Đ* = 1.06), and 1-pyrenebutyric acid (Sigma, 97%) were used as received. Deionized water was obtained from a Biopure Series 4400 Single Pass Reverse Osmosis system.

1-Pyrenebutyric acid N-hydroxysuccinimide ester (PyBuOSu): 1-Pyrenebutyric acid (1.0 g, 3.5 mmol, 1 eq.) and HOSu (0.48 g, 4.2 mmol, 1.2 eq.) were dissolved in a solution of DCM (60 mL) and ACN (20 mL). DCC (0.83 g, 4.0 mmol, 1.2 eq.) was added and the solution was left to stir overnight in the dark. The reaction was filtered to remove the urea side product and condensed to ~ 25 mL on a rotary evaporator. The solution was then filtered again to remove additional urea that had precipitated. The solution was condensed to ~ 10 mL and isopropanol (60 mL) was added to precipitate PyBuOSu. The PyBuOSu product was then collected by suction filtration. The brown solid was dissolved in DCM (10 mL) and precipitated in isopropanol (60 mL) twice more. The solid was dried under vacuum to yield PyBuOSu (0.79

g, 2.0 mmol, 59% yield). The ^{1}H NMR spectrum of PyBuOSu is provided as Figure S7.1 in the Supporting Information (SI).

1-Pyrenebutyric acid-labeled poly(L-lysine) (PyBu-PLL): The PyBu-P*L*L samples were prepared from the hydrobromide salt of poly(*L*-lysine) (P*L*L·HBr). An example preparation of a PyBu-P*L*L sample, where 4.4 mol% of the lysines were labeled with pyrene, is described. P*L*L·HBr (34.2 mg, 0.16 mmol Lys) was added to a stirring solution of DIPEA (32 μL, 1.1 eq. per Lys) and DMSO (5 mL). Once P*L*L·HBr had dissolved, a solution of PyBuOSu in DMSO (0.34 mL, 10.2 mg·mL^{-1}, 0.05 eq. per Lys) was added. The stirring solution was left to stir overnight in the dark. PyBu-P*L*L was purified by dialysis against successive solutions of DMF, a series of 2:1, 1:1, 1:3 DMF:water mixtures, and lastly water acidified with HCl (pH ~ 5). The hydrochloride salt of PyBu-P*L*L was then freeze dried and stored at –20 ºC until it was used for fluorescence measurements.

UV-Vis absorbance measurements: The absorption spectra were acquired with a Varian Cary 100 Bio spectrophotometer using a 1.0 cm quartz UV-Vis absorption cell. A background correction for the solvent was applied to each spectrum.

Pyrene content determination: The pyrene content of the PyBu-P*L*L samples was determined with Equation 7.1. In Equation 7.1, x is the molar fraction of lysine residues bearing pyrene, M_{Lys} is the molar mass of the hydrochloride salt of a lysine residue (= 164.63 g·mol^{-1}), M_{Py} is the molar mass of a lysine bearing a pyrenyl label (= 370.45 g·mol^{-1}), and λ_{Py} is the number of moles of pyrene per mass of PyBu-P*L*L. λ_{Py} was determined by measuring the absorbance of a PyBu-P*L*L solution in DMSO with a known mass concentration. To this end, 10 mg of

lyophilized PyBu-P*L*L was dissolved in 5 mL of DMSO. After sonicating for 30 minutes, an aliquot of the solution was diluted and transferred to a 1.0 cm pathlength cell for UV-Vis absorption measurements. The molar concentration of pyrene was determined using the Beer-Lambert law at the absorbance peak maximum of pyrene at 346 nm with the molar absorptivity coefficient of 1-pyrenebutyric acid used as a model compound in DMSO (ε = 41,400 $M^{-1}\cdot cm^{-1}$).[37]

$$x = \frac{M_{Lys}}{\lambda_{Py}^{-1} + M_{Lys} - M_{Py}} \tag{7.1}$$

Steady-state fluorescence (SSF): The steady-state fluorescence spectra were acquired on a Horiba QM-400 spectrofluorometer equipped with a xenon arc lamp. Samples were prepared by diluting a solution of PyBu-P*L*L in DMSO such that the concentration of pyrene was equal to 2.4 μM. A small aliquot (10 μL) of 1 M HCl was added to a 4 mL solution of PyBu-P*L*L to ensure that the lysines were protonated. The solution was then sonicated for 15 minutes and outgassed with nitrogen (Praxair, N4.0) to remove oxygen. The fluorescence cell was then sealed and the SSF spectra acquired using an excitation wavelength of 346 nm. The excitation and emission slit widths were set to 1 nm. The emission was monitored from 350 to 600 nm in 1 nm increments with a scanning rate of 10 $nm\cdot s^{-1}$. The fluorescence intensity of the pyrene excimer (I_E) and monomer (I_M) were determined by integrating the intensity of the fluorescence spectra from 500 to 530 nm and from 376 to 382 nm, respectively. Dividing I_E by I_M yielded the I_E/I_M ratio which was used to quantify the efficiency of pyrene excimer formation (PEF).

Time-resolved fluorescence (TRF): The fluorescence decays of PyBu-P*L*L were acquired with an IBH TC-SPC fluorometer equipped with a NanoLED-340 laser using the same samples as for the SSF measurements. The monomer and excimer fluorescence decays were acquired over 1,024 channels using a time-per-channel of 1.02 ns·ch^{-1}. The monomer decay was acquired at 378 nm to a decay maximum of at least 40,000 counts. The excimer decay was acquired at 510 nm to a decay maximum of at least 20,000 counts. A 370 or 490 nm long pass filter was placed at the entrance of the emission monochromator to minimize stray light and scattering photons from reaching the detector during the acquisition of the monomer and excimer TRF decays, respectively. The instrument response function was acquired using a LUDOX dispersion in water for deconvolution of the fluorescence decays.

Fluorescence decay analysis: The monomer and excimer decays of PyBu-P*L*L were globally analyzed using the fluorescence *blob* model (FBM). The FBM has been extensively described in previous publications[33-35,38] and the details for the current study are given in the SI. Briefly, within the framework of the FBM, the polypeptide is divided into *blobs* which represent the volume that an excited pyrenyl label attached onto P*L*L can probe. Taking advantage of the random labeling of pyrene, the FBM determines the average number (<*n*>) of pyrenes per *blob*, which is then used along with the pyrene content to determine N_{blob}, the average number of structural units inside a *blob*, according to Equation 7.2. In Equation 7.2, *x* is the molar fraction of pyrene-labeled *aa*'s as defined in Equation 7.1 and f_{Mfree} represents the molar fraction of isolated pyrenyl labels that do not form excimer and are detected in the TRF decays of the pyrene monomer.

$$N_{blob} = \frac{< n >}{x} \times \left(1 - f_{Mfree}\right) \qquad (7.2)$$

7.4 Results and Discussion

The effects that the side chain reach of an *aa* and the composition of a polypeptide have on the number of *aa*'s, which locally interact with one another, were studied using a series of pyrene-labeled polypeptides. In these fluorescence experiments, N_{blob} defines the number of *aa*'s, which are contained within the volume of a *blob*. Therefore, N_{blob} depends both on the size of a *blob*, defined by the length of the spacer linking the pyrenyl label to the polypeptide backbone,[28,36] and the local density of the polypeptide backbone, defined by its conformation.[26,32] First, the effect that the side chain reach of an *aa* had on N_{blob} was determined by monitoring the change in N_{blob} as the length of the spacer connecting the pyrenyl label to the polypeptide backbone was increased by using 1-pyrenemethylamine, 1-pyreneacetic acid, and 1-pyrenebutyric acid to label poly(*D*,*L*-glutamic acid) (PyMe-P*DL*Glu), poly(*L*-lysine) (PyAc-P*L*L), and poly(*L*-lysine) (PyBu-P*L*L), respectively. The chemical structures of PyMe-P*DL*Glu, PyAc-P*L*L, and PyBu-P*L*L are provided in Figures 7.1A – C.

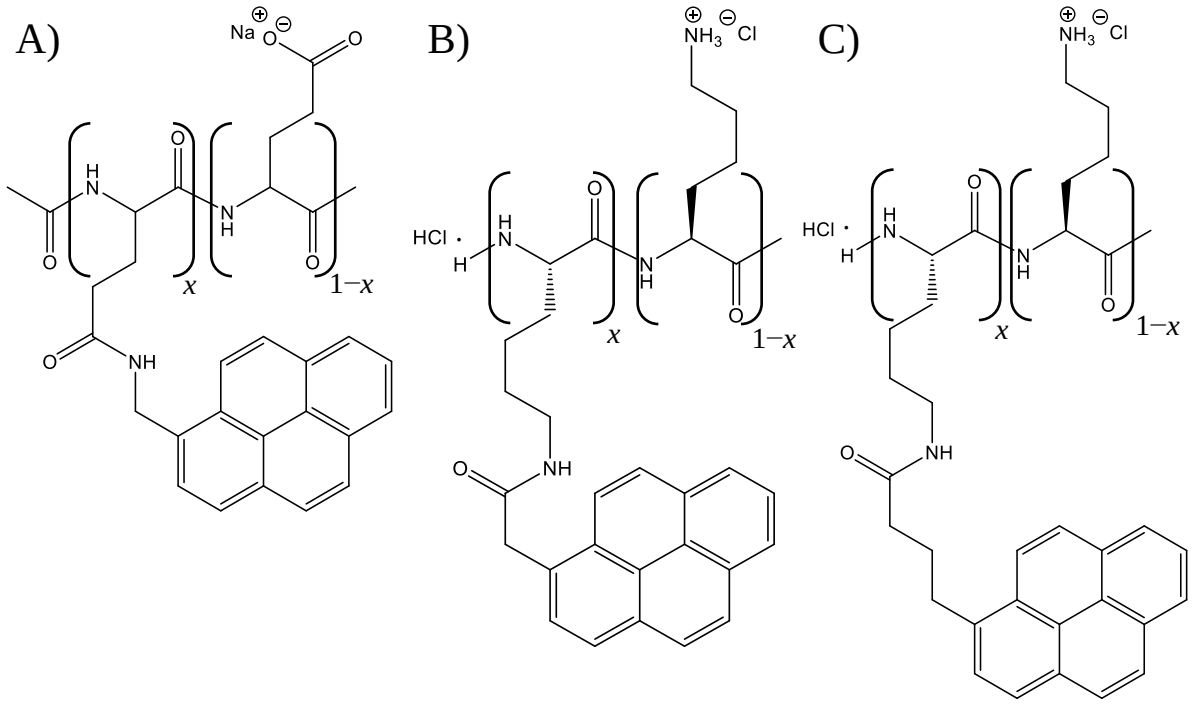


Figure 7.1. Chemical structure of A) PyMe-P*DL*Glu (5), B) PyAc-P*L*L (7), and C) PyBu-P*L*L (9). The length of the linker separating the pyrenyl label from the polypeptide backbone, expressed in terms of spacer atoms, is noted in the parenthesis.

PyMe-P*DL*Glu and PyAc-P*L*L have been previously characterized using the FBM.[26,29-31] The N_{blob} value of 14.3 (±1.3) *aa*'s obtained for PyAc-P*L*L in DMSO was found to be larger than that of 10.5 (±1.7) for PyMe-P*DL*Glu.[29–31] Molecular mechanic optimizations (MMOs) were conducted on the PGlu and P*L*L backbones to demonstrate that in DMSO, both PyMe-P*DL*Glu and PyAc-P*L*L adopted coiled conformations, where the rigid backbone was extended on the length scale of a *blob*. Since both polypeptides adopted a similar conformation, it was concluded that the 7-atom long linker connecting the pyrenyl label to the polypeptide backbone

of PyAc-P*L*L allowed the pyrenyl label to probe a larger volume compared to that afforded by the 5-atom linker of PyMe-P*DL*Glu and therefore included more amino acids within its larger *blob* volume. To quantify the dependence of the *blob* size on the side chain length, a third pyrene-labeled polypeptide, namely PyBu-P*L*L, was prepared using 1-pyrenebutyric acid to generate a 9 atom-long linker. However, since PyBu-P*L*L had not yet been characterized, its fluorescence was first studied to determine its *blob* size and ensure that it adopted a conformation similar to that of PyMe-P*DL*Glu and PyAc-P*L*L.

PyBu-PLL characterization: The steady-state fluorescence spectra of PyBu-PLL in DMSO are provided in Figure S7.2 in the SI. Since the fluorescence spectra looked typical for that of a 1-pyrenebutyric acid labeled polymer, the TRF decays of the pyrene monomer and excimer were acquired and analyzed with the FBM to determine N_{blob} and provide information on the polypeptide conformation. The parameters retrieved from the analysis were listed in Tables S7.1 – S7.2 in SI and were used to calculate N_{blob} based on Equation 7.2. Figure 7.2A shows, that N_{blob} plotted as a function of pyrene content remained relatively constant for PyBu-P*L*L in DMSO, taking an average value of 17.7 (±1.9). Since N_{blob} responds to both the length of the linker connecting the pyrenyl label to the polypeptide backbone[28,36] and the local conformation of the polypeptide backbone,[26,32] it is important to isolate the two effects from one another so that each effect can be individually quantified. Therefore, the conformation of PyBu-P*L*L in DMSO was assessed by using molecular mechanics optimizations (MMOs) to determine the theoretical *blob* size (N_{blob}^{theo}) of a rigid P*L*L backbone in a coiled conformation following previously reported procedures.[29,30]

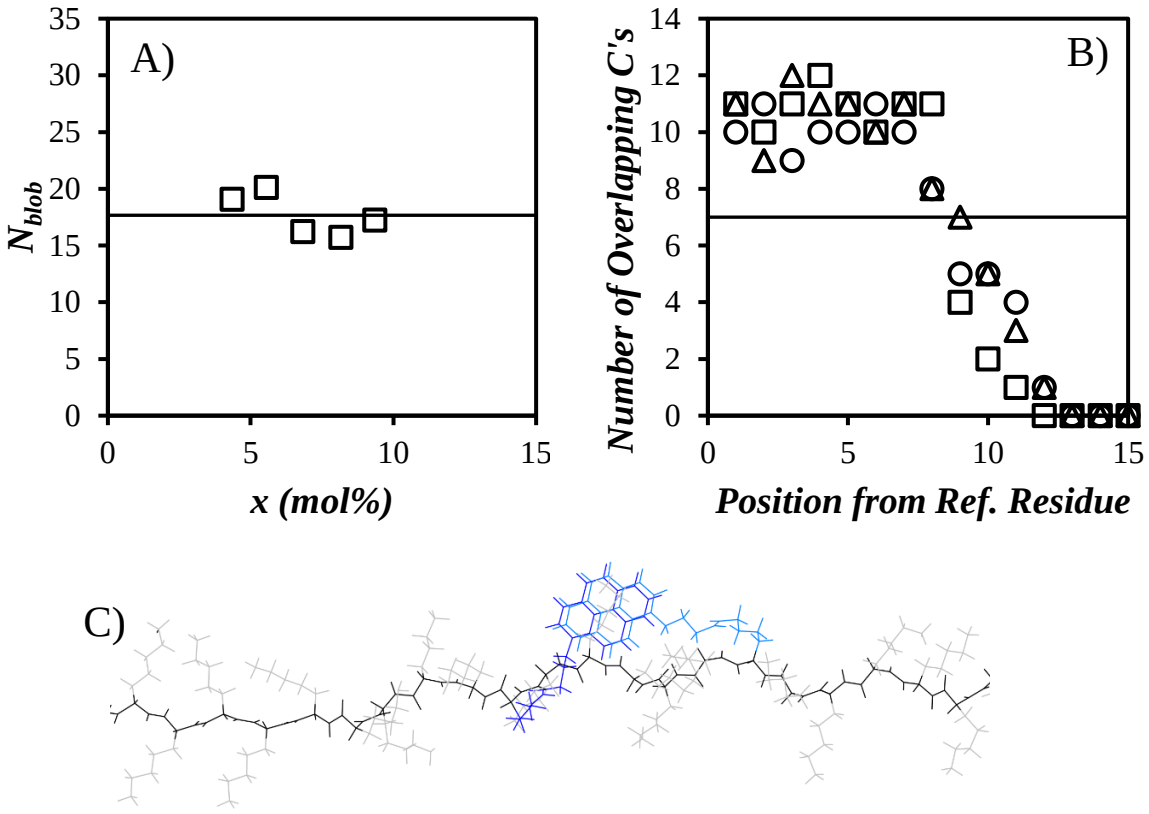


Figure 7.2. A) Plot of N_{blob} as a function of pyrene content for PyBu-P*L*L in DMSO. The line represent the average experimental $\langle N_{blob} \rangle$ and calculated N_{blob}^{theo}, which take similar values of 17.7 (±1.9) and 17.7 (±1.2), respectively. B) Number of carbon atoms determined by MMOs between two overlapping pyrenyl pendants as a function of the number of lysines separating two pyrene-labeled lysines for a P*L*L chain in a coiled conformation. The line represents the breakpoint between good and poor pyrene-pyrene overlap. C) Example rendering of the MMO for PyBu-P*L*L in an extended backbone conformation (black) with two pyrenyl labels showing good overlap (light and dark blue) located 6 lysines apart from one another. The unlaleled side chains of lysine are in shaded grey.

To this end, a 45 unit-long P*L*L was constructed in the program HyperChem. The chain ends were stretched apart, resulting in an elongated conformation. The chain end constraints were then removed so that the P*L*L backbone would adopt a relaxed conformation. Once the root-mean-squared energy gradient was less than 0.1 kcal·Å^{-1}·mol^{-1}, the conformational optimization was considered complete. After fixing the backbone in place, a reference lysine residue was labeled with 1-pyrenebutyric acid (PyBA). A second residue was also labeled with PyBA. To determine if the pyrenyl labels were close enough to form an excimer, the polypeptide backbone was fixed in the MMO and the two pyrenes were pulled towards one another, such that they could stack parallel to each other with a 3.4 Å separation gap. The extent of overlap between the two pyrenyl labels was quantified by counting the number of carbons of the pyrene on the secondary lysine residue, that were observable within the frame of the pyrene on the primary lysine residue. The pyrene-pyrene pairs resulting in 7 or more overlapping carbons were considered to generate enough overlap to form an excimer. The procedure was repeated for each residue up to 15 residues away from the reference pyrene. Figure 7.2B plots the number of overlapping carbons as a function of the distance separating the two pyrene-labeled residues using three different reference positions. An example rendering of two pyrenyl labels attached onto the P*L*L backbone and showing good overlap is given in Figure 7.2C. The maximum number (N_0) of structural units separating the two pyrenyl labels while still resulting in good overlap is defined as the reach of a pyrene-labeled side chain. Figure 7.2B indicates that of the three reference positions, two positions resulted in sufficient overlap to form an excimer up to 8 units away, while the other showed sufficient

overlap up to 9 units away. Therefore, a lysine labeled with PyBA would have on average an N_0 value of 8.3 (±0.6) lysines, assuming that the backbone adopted the extended conformation expected of a random coil. Since the reference pyrenyl label can interact with a secondary pyrenyl label located on either side of the reference pyrene along the P*L*L chain, the total number of structural units capable of interacting with one another is equal to $2\times N_0 + 1$, where the '+ 1' accounts for the reference residue. The N_0 value of 8.3 (±0.6) thus resulted in an N_{blob}^{theo} value of 17.7 (±1.2), which precisely matched the experimentally obtained N_{blob} value of 17.7 (±1.9) in Figure 7.2A and demonstrated that the PyBu-P*L*L must adopt an extended coiled conformation in DMSO. This result was in agreement with the conclusions previously drawn for a series of P*L*Ls labeled with 1-pyreneacetic acid acidified in DMSO.[30,31] This provides another demonstration of the ability of the FBM in conjunction with MMOs to provide quantitative information on the conformation of polypeptides in solution independently of the nature of the pyrenyl label.

N_{blob} *and amino acid side chain reach:* Since PyMe-P*DL*Glu, PyAc-P*L*L, and PyBu-P*L*L all adopted the same extended coil conformation in DMSO, their different N_{blob} values arose solely from differences in the length of the spacer connecting the pyrenyl label to the polypeptide backbone. A plot of N_{blob} as a function of the number of atoms in the spacer linking pyrene to the backbone of a polypeptide adopting an extended conformation is given in Figure 7.3. Since N_{blob} represents the number of *aa*'s which locally interact with one another, Figure 7.3 represents a calibration curve for predicting the *blob* size as a function of linker length. Such a curve can then be used to predict the N_{blob} value for any *aa* based on its side chain reach (SCR),

where SCR is defined by the fewest number of continuously connected atoms spanning the length of the side chain. Since the N_{blob} values reported in Figure 7.3 are representative of side chains bearing a pyrenyl label, the contribution to N_{blob} from the pyrene molecules must be accounted for.

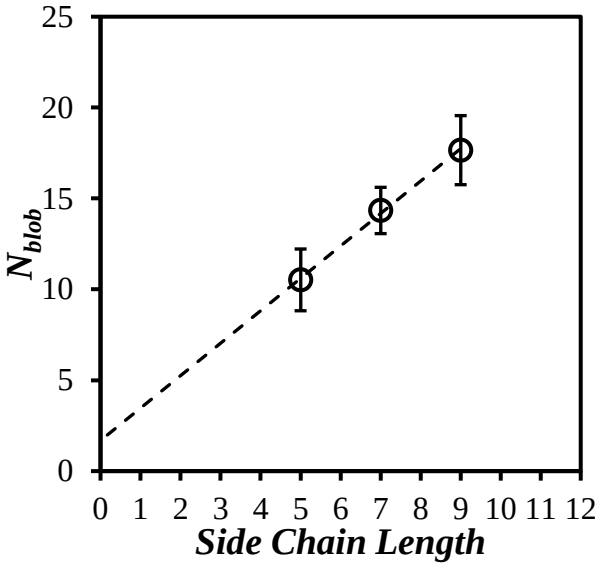


Figure 7.3. Plot of N_{blob} as a function of the length of the linker (referred to as side chain length) connecting pyrene to the backbone of a polypeptide in an extended conformation. The dashed line represents the line of best fit: $y = 1.78\ (\pm 0.07)\ x + 1.68\ (\pm\ 0.52)$, $R^2 = 1.00$.

When the trend in Figure 7.3 was extrapolated to a 0-atom SCR, N_{blob} took a value of 1.68 (± 0.52). Since it would be impossible for an *aa* with a zero-length side chain to interact with any *aa*'s other than itself, this *aa* would be the only *aa* contained within its volume and the *blob* size should take a value of 1. Therefore, the contribution of pyrene to N_{blob} must equal the intercept minus 1. Since the pyrene molecule is constant among all the pyrene-labeled

polypeptides, it is reasonable to assume that its contribution remains constant with changing side chain reach. It was therefore concluded that a pyrene molecule contributed 0.68 *aa* to the *blob* sizes in Figure 7.4. Consequently, if pyrene were removed from the side chain, the new relationship between side chain reach and N_{blob} would be 1.78 times the side chain reach (SCR) expressed in number of atoms plus 1 as described in Equation 7.3. This result is reasonable since adding one atom to all side chains should increase the reach of each side chain toward another side chain to its left and to its right by 4 atoms in total or 1.33 *aa*'s if the polypeptides was a fully extended string of atoms. Accounting for residual bending in the conformation of a polypeptide adopting an extended conformation explains the 1.78 increase given in Equation 7.3. Although this relationship would be impossible to check by fluorescence, the fact that N_{blob} represents the number of *aa*'s which locally interact with one another dictates that these local interactions must still be present even when pyrene is removed from the side chain.

$$N_{blob} = 1.78 \times SCL + 1 \qquad (7.3)$$

N_{blob} and polypeptide composition: The second parameter affecting N_{blob} is the local backbone conformation. This section aims to provide first, a brief summary of how the N_{blob} value of coiled polypeptides responds to local conformational changes induced by the local *aa* composition and second, the implications of the resulting plots which are later used to calculate the *blob* sizes in proteins. The effect of local *aa* composition was first studied with a series of copolypeptides prepared from a mixture of 44 (±3) mol% *D,L*-glutamic acid and 56 (±3) mol%

of either glycine (Gly), *D*,*L*-alanine (Ala), or *D*,*L*-*N*-ε-carbobenzyloxylysine (Lys(Z)) and referred to as PyMe-PGlyGlu, PyMe-PAlaGlu, and PyMe-PLys(Z)Glu, respectively.[32] The chemical structures of the copolypeptides are provided in Figures 7.4A – D.

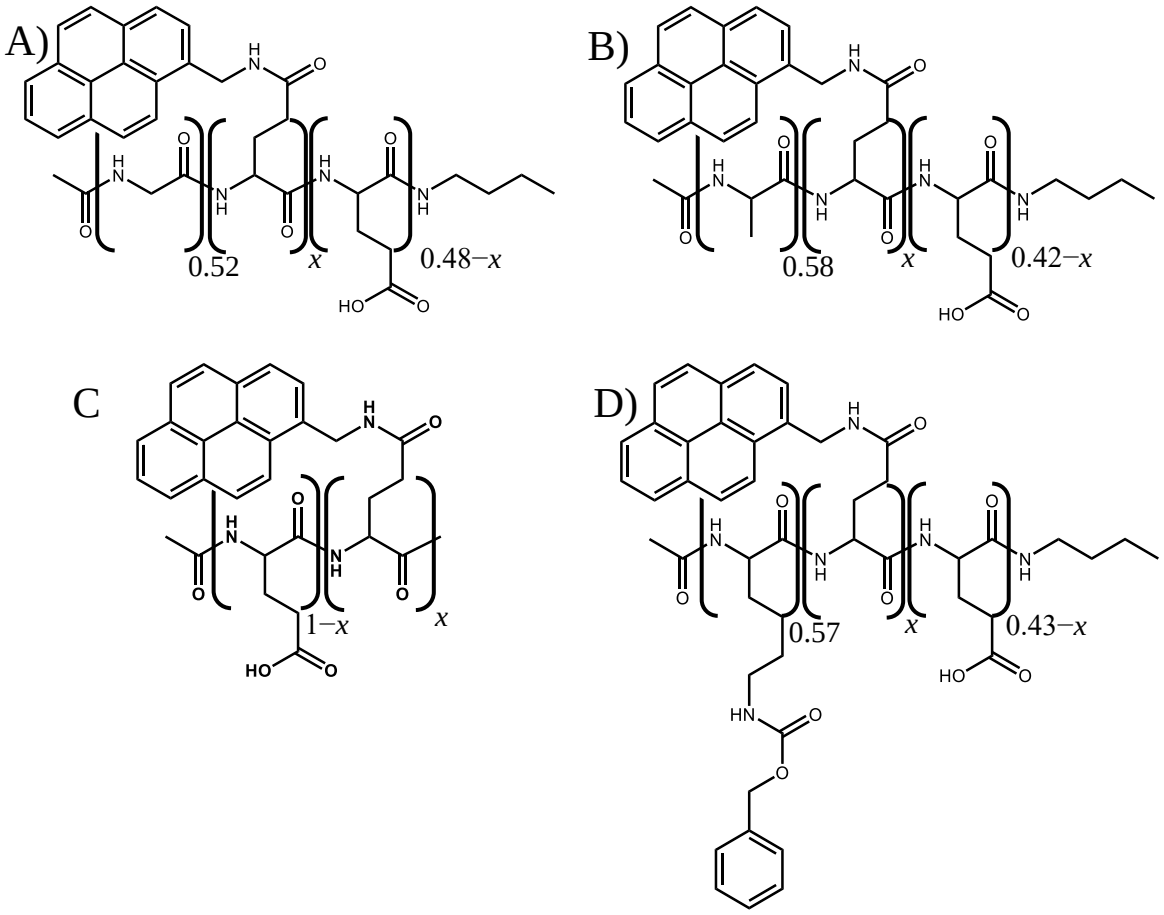


Figure 7.4. Chemical structures of A) PyMe-PGlyGlu, B) PyMe-PAlaGlu, C) PyMe-P*DL*Glu, and D) PyMe-PLys(Z)Glu. All stereocenters are racemic.

The N_{blob} values of the copolypeptides were plotted in Figure 7.5A as a function of the comonomer side chain size (SCS, in terms of non-hydrogen atoms). Interestingly, the N_{blob} values of 22.5 (±1.5) for PyMe-PGlyGlu and 15.8 (±1.0) for PyMe-PAlaGlu were larger than those of the plateau value of 11.0 (±1.4) for PyMe-PLys(Z)Glu and PyMe-P*DL*Glu. Since each polypeptide used 1-pyrenemethylamine to label a fraction of the glutamic acid residues, the 5-atom linker length remained constant and therefore, the differences in N_{blob} must have been due to a change in backbone conformation. The racemic mixtures of *aa*'s used to prepare the copolypeptides ensured that they could not form secondary structure and adopted a coiled conformation in solution. Therefore, the increase in N_{blob} with decreasing comonomer SCS was not due to the formation of secondary structures, but rather to an enhancement in the backbone conformational freedom due to reduced steric hindrance experienced by the *aa*'s with small side chains. The reduced steric hindrance provided points of greater conformational freedom throughout the backbone and therefore allowed more *aa*'s to interact with one another in a *blob*. This led to the conclusion that the N_{blob} value of a polypeptide depended on the local *aa* sequence contained within a *blob*, such that with all else being equal, *blobs* containing small *aa*'s would have larger N_{blob} values than those that exclusively contained larger *aa*'s.

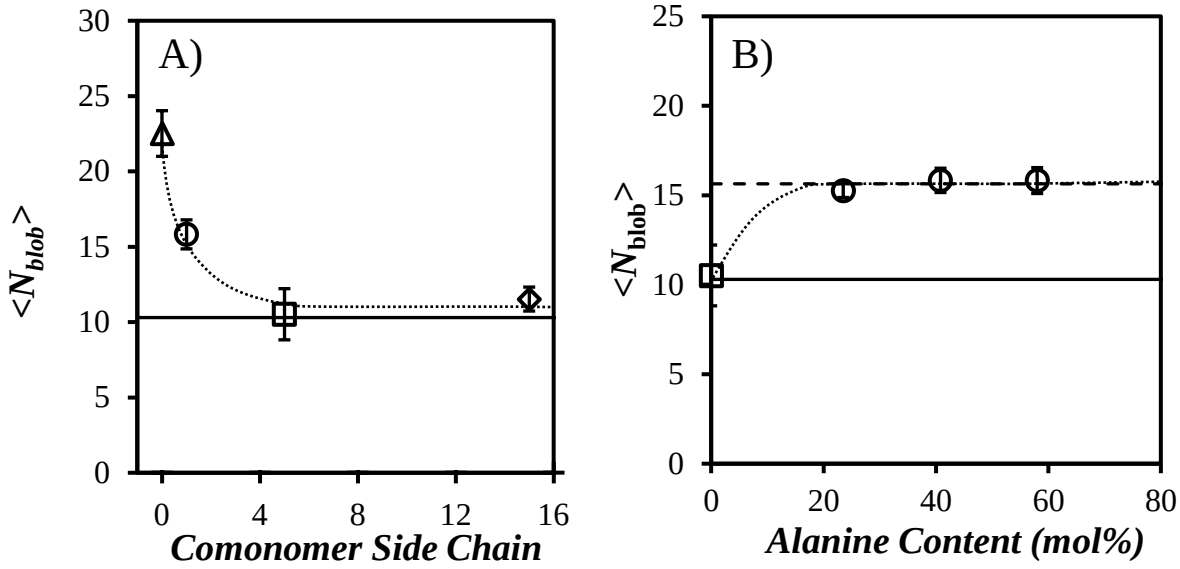


Figure 7.5. Plots of the average N_{blob} values of (△) PyMe-PGlyGlu, (○) PyMe-PAlaGlu, (□) PyMe-P*DL*Glu, and (◇) PyMe-PLys(Z)Glu in DMSO as a function of A) comonomer side chain size (Gly = 0, Ala = 1, Glu = 5, and Lys(Z) = 15) and B) alanine content. The solid lines represent the N_{blob}^{theo} value of 10.3 (±1.2) for a rigid polypeptide coil having a 7 atom-long linker connecting pyrene to the polypeptide backbone. The dashed line represents the average N_{blob} value of 15.6 (±0.8) for the PyMe-PAlaGlu's and the dotted lines were added to guide the eyes.

Since the copolypeptides in Figure 7.4 contained 44 (±3) mol% Glu and 56 (±3) mol% comonomer, the dependence of N_{blob} on the level of comonomer incorporated into the backbone needed to be studied next. This was accomplished by using a series of PyMe-PAlaGlu samples with alanine contents ranging from 24 to 58 mol%.[26] Figure 7.5B shows that compared to PyMe-P*DL*Glu (which contained no alanine), the incorporation of 24 mol%

alanine into the Glu backbone resulted in a sharp increase in N_{blob} from 10.5 (±1.7) to 15.6 (±0.8). Surprisingly, N_{blob} then remained constant as a function of alanine content indicating that the backbone conformational freedom of the copolypeptides remained unchanged. This result demonstrated that the incorporation of low levels of small *aa*'s were able to control the backbone flexibility of a polypeptide on the length scale of a *blob*.

Estimating the blob size of an aa in a protein: At the beginning of this discussion, it is important to discuss the definition of side chain reach (SCR) and side chain size (SCS), which are used as two different concepts in this study. SCR represents the number of atoms, including H-atoms, that extend the side chain of an *aa* into the solvent, thus reflecting the reach of an *aa* side chain. In contrast, SCS reflects the bulkiness of the side chain, which most affects bending, and as such is given by the total number of non-hydrogen atoms in the side chain. As a result, lysine with its ε-aminobutyl side chain has a SCR and SCS equal to 6 and 5, respectively. The SCR and SCS values of the 20 most common *aa*'s have been listed in Table S7.4 in Supporting Information (SI).

All the polypeptides studied above were either homopolypeptides or random copolypeptides. Consequently, the distribution of *aa*'s was similar along the entire backbone and the *blobs* were therefore symmetrical. In other words, the number of *aa*'s (N_0), that a reference pyrenyl label could interact with along the backbone on each of its two flanks, was the same in both directions such that N_{blob} equaled $2\times N_0+1$. However, since the *aa* sequence of proteins is heterogeneous, the N_0 value for a given *aa* may be different from one side of the reference pyrenyl label relative to the other. As a result, the N_{blob} value of an *aa* in a

heterogenous sequence is defined by the modified equation $N_0^{l} + N_0^{r} + 1$, where N_0^{l} and N_0^{r} represent the N_0 values to each of the right (*r*) or left (*l*) side of the reference *aa* inside a *blob*. For this reason, the N_{blob} determination of the *aa*'s in a protein will be described with respect to N_0 and not N_{blob}. A simplistic diagram showing the relationship between N_0 and N_{blob} in a homogeneous polypeptide with an elongated backbone is given in Figure 7.6 where the polypeptide backbone is first assumed to adopt an extended conformation on the length scale of a *blob*.

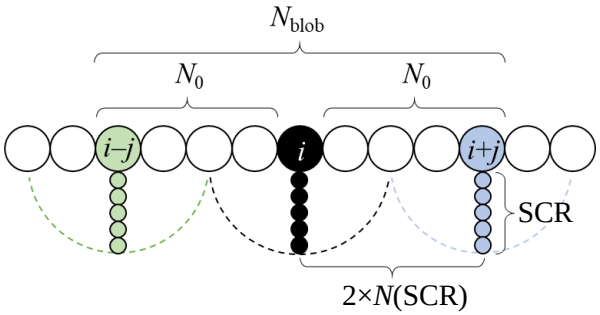


Figure 7.6. Schematic representation of the components defining the N_{blob} value for the reference *aa* (black) in a polypeptide containing homogeneous *blobs* and adopting an extended conformation. The N_0 values depend on the side chain reach (SCR) of the reference residue and the *aa*'s to the left (green) and right (blue). The number (*N*(SCR)) of *aa*'s which are included in the *intrinsic reach* of a single side chain is a function of SCR.

As discussed above, the results obtained from the FBM suggest that the number of *aa*'s, that any particular *aa* interacts with in a polypeptide, is a function of its SCR and its surrounding sequence. Therefore, determining the *blob* size of an *aa* in a protein must take

both these considerations into account. The effect of SCR is discussed first. As depicted in Figure 7.6, the N_0 value of an *aa* depends not only on the SCR of the reference *aa*, but also on the SCRs of the surrounding *aa*'s. Therefore, the term N(SCR) (referred to as the intrinsic reach) was introduced to represent the number of *aa*'s located on one side of the reference *aa*, whose side chain tip could interact with the side chain tip of the reference *aa* (Figure 7.6). The first step in estimating the *blob* size of a protein was to determine N(SCR) as a function of SCR. This was accomplished by using the N_{blob} values in Figure 7.3. Equation 7.3 defines N_{blob} as a function of the SCR of an *aa* inside a pyrene-free *blob*. Since all the polypeptides used to generate Equation 7.3 had homogeneous *blobs* obtained with homopolypeptides, the corresponding reach (N_0) of each side chain was determined with the relationship existing between N_0 and N_{blob} ($N_0 = (N_{blob} - 1)/2 = SCR \times 1.78/2$ based on Equation 7.3). Lastly, one still needed to consider that N_0 represents the reach of two *aa*'s touching each other, and that it is defined by the combined SCRs of these two *aa*'s. In this case, since each *blob* studied so far was defined by *aa*'s having the same SCRs, the N_0 values of these polypeptides represented twice the intrinsic reach of each *aa* (i.e. $N_0 = 2 \times N$(SCR) for symmetrical *blobs*). These relationships were then used to generate a plot of N(SCR) as a function of SCR for pyrene-free *aa*'s in Figure 7.7A. The linear relationship in Figure 7.7A could now be used to determine the N(SCR) value for any *aa* based on its SCR according to Equation 7.4. For the purpose of determining N(SCR), the SCR of an *aa* was determined as the furthest possible distance separating the tip of its side chain to its α-carbon in terms of the number of atoms/bonds

(including hydrogen atoms). Table S7.4 in SI provides the SCR and the *N*(SCR) values of the 20 most common *aa*'s.

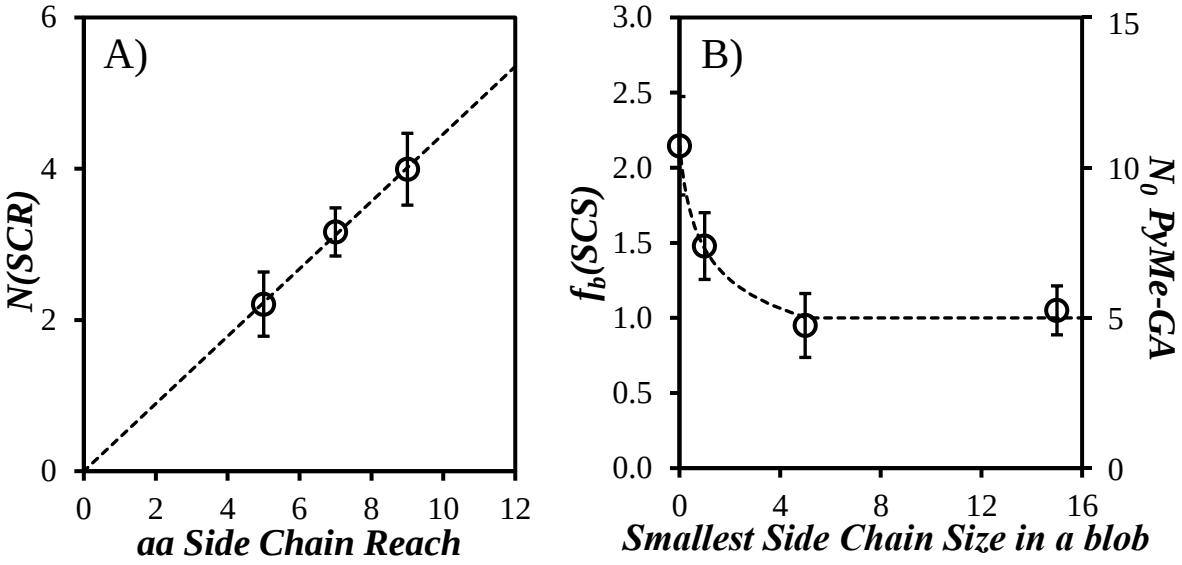


Figure 7.7. Plot of A) *N*(SCR) as a function of *aa* side chain reach (SCR) and B) N_0 of the pyrene-labeled polypeptides in Figure 7.5A and the bending function (f_b(SCS)) by which N_0 increased from its plateau value of 5.0 (±0.7) as a function of side chain size (SCS). The dashed lines represent the lines of best fit equal to A) *N*(SCR) = 0.45 (±0.02) × SCR and B) defined by Equation 7.4.

$$N(SCR) = \frac{N_0}{2} = \frac{N_{blob} - 1}{4} = \frac{1.78 \times SCR}{4} = 0.45 \times SCR \qquad (7.4)$$

The *N*(SCR) value determined from Figure 7.7A represents the number of structural units an *aa* can interact with in one direction along a polypeptide backbone if and only if the backbone adopts a rigid extended conformation (like the one shown in Figure 7.2B). However,

Figure 7.5A demonstrated that a polypeptide backbone was only rigid if the *blob* contained *aa*'s with side chain sizes (SCSs) of 5 non-hydrogen atoms or greater. Since the sequence of a protein often contains many different *aa*'s, it is very likely that the *blob* of a protein will contain a mixture of *aa*'s with side chains both larger and smaller than 5 atoms which, respectively, do not and do affect the size of a *blob*. Therefore, the effect that the local *aa* sequence inside a *blob* has on the N(SCR) value must be determined. In particular, it will affect the term N_0 which is used to determine N_{blob} in Equation 7.4 and which represents the number of *aa*'s that a reference residue can interact with along one of its flanks, after taking into consideration the local polypeptide (or protein) conformation. Thus, the N_0 value for a symmetrical *blob* is expected to equal $[2\times N(\text{SCR})]\times f_b(\text{SCS})$, where f_b(SCS) is the bending function accounting for the change in the N(SCR) values due to the local backbone conformational freedom of a polypeptide resulting from the SCS of an *aa*. f_b(SCS) was estimated from the increase in N_{blob} from 11.0 (±1.4) to 22.5 (±1.5) in Figure 7.5A as the SCS decreased from 5 to 0 atoms. Since this behavior was only representative of the 5 atom-long side chain of the 1-pyrenemethylamine-labeled Glu (SCR = 5), it needed to be generalized. This was accomplished by first determining the corresponding N_0 ($=(N_{blob} - 1)/2$) values which were plotted as the secondary axis in Figure 7.7B as a function of SCS. All N_0 values in Figure 7.7B were then normalized by dividing them by the plateau value of 5.0 (±0.7) found for PyMe-P*DL*Glu and PyMe-PLys(Z)Glu to yield the bending function f_b(SCS). The bending function represented the factor by which N(SCR) increased due to the additional conformational freedom provided by the introduction of *aa*'s with smaller SCS in the local *aa* composition

inside a *blob*. f_b(SCS) was plotted as a function of SCS on the primary axis in Figure 7.7B. SCS values of 5 or greater resulted in an f_b(SCS) value equal to unity. An f_b(SCS) value of 1.0 indicates that an oligopeptide segment occupying a *blob* containing only large *aa*'s is in an extended conformation and that the N_0 value is simply equal to the *N*(SCR) value determined from Figure 7.7A. However, if a protein *blob* contains a smaller amino acid, the *N*(SCR) value would be increased by the factor f_b(SCS) to account for its increased conformational freedom. To determine which SCS within a *blob* should be used to calculate f_b(SCS) in Figure 7.7B, the PyMe-PAlaGlu copolymers were used. Figure 7.5B showed that without alanine, a P*DL*Glu backbone had a *blob* size of 10.5 (±1.7) *aa*'s. However, if 2 – 3 of those Glu's, corresponding to an alanine content per *blob* of 0.24×10.5 = 2.54 for the $PAla_{0.24}Glu_{0.76}$ copolypeptide, were replaced by an alanine, the *blob* size increased by ~48%. Again, since a *blob* is defined by the number of *aa*'s on both sides of the reference *aa* bearing a pyrenyl label, this meant that on average there only needed to be ~1 (=2.54/2) alanine on each side of the reference residue to increase its *blob* size. In other words, the *N*(SCR) value of the PyMe-Glu residue increased by ~ 48% (Figure 7.7B) due to the inclusion of a single alanine inside the *blob*. This indicated that the N_0 value of an *aa* responds to the smallest *aa* in its range, even at very low levels of incorporation. Consequently, the *aa* with the smallest SCS among the N_0 residues within reach of a reference *aa* defines f_b(SCS) for the reference *aa*. The f_b(SCS) values in Figure 7.7B were fit empirically with the piecewise function given as Equation 7.5 allowing for the calculation of f_b(SCS) for any SCS. Equation 7.5 uses the SCS of the smallest *aa* located within N_0 units

of the reference residue. Figure 7.7B could then be used as a calibration curve to determine f_b(SCS) for the local composition experienced by a reference *aa* inside a *blob*.

$$f_b(SCS) = \begin{cases} 1.57 \times (SCS + 0.3)^{-0.266} & SCS < 5 \\ 1 & SCS \geq 5 \end{cases} \qquad (7.5)$$

Since N(SCR) and f_b(SCS) could be determined as a function of the SCR and SCS for all *aa*'s constituting the local sequence surrounding a reference *aa*, the N_{blob} value could now be estimated for any *aa* in a heterogeneous polypeptide, such as a protein. The process used to calculate N_{blob} for a reference *aa* represented as the i^{th} *aa* in a heterogeneous sequence is illustrated in Figure 7.8. The process is based on the fact that for two *aa*'s to interact, their side chains must be able to touch. Therefore, N_0 was calculated by determining the *aa* that was the furthest away from the i^{th} *aa* used as reference in Figure 7.8, such that the side chains of the i^{th} *aa* and a second *aa* located at position $i - j$ in a polypeptide (i.e. j *aa*'s to the left of the reference *aa*) could still touch one another. Since the SCR of an *aa* defines its N(SCR) value, the furthest the *aa*'s at position i and $i - j$ could be separated from one another and still interact, while the polypeptide retained an elongated conformation, was the sum of their N(SCR) values (= N(SCR(i)) + N(SCR(i-j))). However, since the local conformation depends on the SCS of the *aa*'s located between (and including) the *aa*'s at positions i and $i - j$ (SCS($i \rightarrow i$-j)), the sum of their N(SCR) values needed to be multiplied by f_b(SCS($i \rightarrow i$-j)) to account for the backbone conformation. As explained in the discussion pertaining to the bending function, f_b(SCS($i \rightarrow i$-

j)) takes the value $f_b(SCS(l))$ of the smallest *aa* located at position *l* such that $i - j \leq l \leq i$. Therefore, the largest number of *aa*'s in a heterogeneous polypeptide separating the *aa*'s at positions *i* and $i - j$, while still allowing them to interact, would be $[N(SCR(i)) + N(SCR(i-j))]\times f_b(SCS(l))$. If the number *j* of *aa*'s was less than the product $[N(SCR(i)) + N(SCR(i-j))]\times f_b(SCS(l))$, then the two *aa*'s were considered close enough to interact. A similar reasoning would apply for a secondary *aa* located *k* *aa*'s to the right of the reference *aa* at position *i*. Since a *blob* is defined by the largest number of *aa*'s separating the reference *aa* from a secondary *aa*, the N_0^l and N_0^r values for the N_0 value left and right of the reference *aa* at position *i*, were set equal to the largest number of *aa*'s separating the reference *aa* at position *i* and a secondary *aa* at position $i - j$ or $i + k$, respectively, while still allowing them to touch via their side chain. Therefore, as depicted in Figure 7.8, N_0^l and N_0^r equaled the largest integers *j* and *k* satisfying the conditions described in Equation 7.6,

$$j \leq [N(SCR(i-j)) + N(SCR(i))]\times f_b(SCS(l)) \quad (7.6a)$$

$$k \leq [N(SCR(i)) + N(SCR(i+k))]\times f_b(SCS(l)) \quad (7.6b)$$

where *l* represents the position of the *aa* with the smallest SCR such that $i - j \leq l \leq i$ or $i \leq l \leq i+k$ depending on whether the sequence to the left or the right of the reference *aa* is being considered, respectively. Finally, the N_{blob} value was calculated as $N_0^l + N_0^r + 1$.

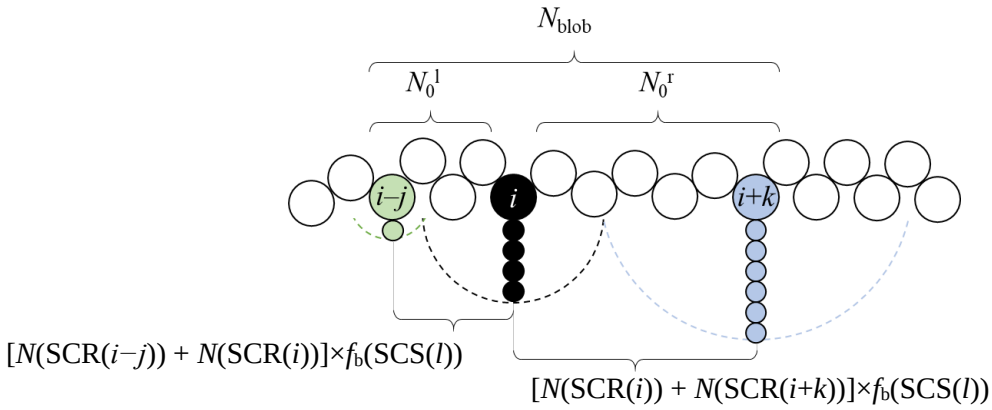


Figure 7.8. Schematic representation of the components defining the N_{blob} value for the reference *aa* (*i*, black) in a heterogeneous polypeptide. The N_0 values depend on the side chain reaches (SCRs) and side chain sizes (SCSs) of the reference residue (*i*) and its neighboring *aa*'s to the left ((*i*-*j*)th *aa*, green) and right ((*i*+*k*)th *aa*, blue). The superscripts 'l' and 'r' denote whether the N_0 values are to the left or to the right of the reference *aa* at position *i*, respectively. The numbers *i*, *j*, and *k* denote the position of each *aa* in the sequence. The SCR of each *aa* defines its *N*(SCR) value according to Figure 7.7A. The SCS of the smallest *aa* with an index *l* between *i*–*j* and *i* (SCS(*i*-*j*→*i*)) and between *i* and *i*+*k* (SCS(*i*→*i*+*k*) defines the f_b(SCS(*l*)) values to the left and right, respectively. The values of $j = N_0^l$ and $k = N_0^r$ are determined by taking the largest integers satisfying the condition [*N*(SCR(*i*–*j*)) + *N*(SCR(*i*))]×f_b(SCS(*i*-*j*→*i*)) ≥ *j* and [*N*(SCR(*i*)) + *N*(SCR(*i*+*k*))]×f_b(SCS(*i*→*i*+*k*)) ≥ *k*, respectively, as described in Equation 7.6. $N_{blob} = N_0^l + N_0^r + 1$.

Determining the blobs in a protein: Now that a procedure had been implemented to determine the N_{blob} values of each *aa* in the sequence of a protein, the next step was to use these N_{blob}

values to determine how many *blob*s a protein might contain and what their average size would be. Since every *aa* was assigned an N_{blob} value, a procedure was implemented to determine which of these N_{blob} values best defined the protein. The largest N_{blob} value was identified first, since it would be expected to generate the greatest number of interactions and therefore be central to the folding process. Next, the N_{blob} values for all other *aa*'s contained within the *blob* were discarded since these *aa*'s (which by definition have smaller N_{blob} values) were considered to be contained within the larger *blob*. This marked the boundaries of the first *blob* of the protein. The next largest N_{blob} value in the remaining sequence was then found and the process repeated, marking the boundaries of the second *blob*. Continuing this iterative process, the entire sequence was compartmentalized into *blobs*, at which point the N_{blob} values were averaged to yield the proteins average N_{blob} value, and tallied to give n_b, the number of *blobs*. An example illustrating this procedure is given in the SI. In some cases, the search for the largest remaining N_{blob} value returned two or more *aa*'s with the same N_{blob} values. In this situation, one of the *aa*'s was selected at random since it was equally likely for the *blob* to be centered about any of the *aa*'s with the same N_{blob}. For this reason, the calculation of the proteins n_b and N_{blob} value was averaged over 1000 iterations of this process to obtain representative values. The resulting $<N_{blob}>$ and $<n_b>$ values were therefore equal to average *blob* size and number of *blobs*, respectively, contained within the protein. Since the boundaries of the *blobs* may intersect with one another, the product $<N_{blob}>\times<n_b>$ is often greater than the protein length since this product 'double counts' the *aa*'s contained within the overlapping regions of the *blobs*. While this double counting might appear troubling at first glance, it is

clear that the conformational search of one *aa* whose side chain can interact with the side chains of other *aa*'s in a neighboring *blob* does not stop at the boundary of the neighboring *blob* and will involve these neighboring *aa*'s. A discussion on the implications of overlapping *blobs* is given in the SI.

Estimating the folding time of a protein: The folding of proteins takes place in compartmentalized domains, as demonstrated by hydrogen exchange experiments.[39] These domains consist often of continuous segments in the sequence of a protein, demonstrating that folding involves the rearrangement of a linear oligopeptide segment contained within a subvolume of the protein. The number of *aa*'s contained within each of these folding subvolumes must therefore depend on the local chain conformation and dynamics. Since N_{blob} represents the number of *aa*'s which locally interact with one another in the subvolume of a polypeptide, it provides a method to estimate the number of *aa*'s that encounter diffusively to generate the folding domains within a protein. By viewing a protein as a string of *blobs*, the number of conformations and therefore the folding time of a protein can be calculated. Starting from an unfolded protein in a coiled conformation, the *aa*'s contained within each *blob* begin to simultaneously interact with one another, starting the folding process. As folding continues, the *blobs* must also rearrange among themselves so that the folding domains can assemble into their native conformation. Based on this model, the number of conformations that a protein can adopt is divided into two components, namely the number (Ω_{B}) of conformations resulting from the re-arrangement of *aa*'s inside a *blob* and the number (Ω_{M}) of re-arrangements among the *blobs* constituting the protein. Since each *blob* contains N_{blob} *aa*'s, Ω_{B} is defined by the

number of conformations enabled by N_{blob} structural units. Similarly, Ω_M depends on the number (n_b) of *blob*s contained within the protein. To calculate the number of conformations among a number N_{blob} of *aa*'s and n_b *blobs*, Equation 7.7 was used.

$$\Omega(n) = C(n-1)^{\gamma-1}\,\mu^{n-2} \tag{7.7}$$

As discussed in an earlier publication,[26] Equation 7.7 defines the number of unique conformations (Ω) that a polymer with *n* structural units can adopt. It is based on the results of the renormalization group theory of self-avoiding walks. On a 3D cubic lattice, the constants *C* and γ are equal to 1.17 and 7/6, respectively,[40,41] while the constant μ is the effective coordination number, which depends on the solvent quality. Since the dense globular structure of a folding protein is expected to be similar to that of a polymer in a poor solvent, μ was set equal to 3.19, which is the calculated degree of freedom of a polymer in such conditions.[42]

Based on the above description of a folding protein, each *blob* folds simultaneously so that the conformational searches of the *blobs* parallel each another. As the *aa*'s within each *blob* undergo their conformational search, the *blobs* are also probing their conformations, which compounds the number of conformations within a *blob* (Ω_B) with the number of conformations among the *blob*s (Ω_M). Assuming that an *aa* requires some particular time τ_p to probe a single conformation, the folding time (τ_F) needed for this conformational search to reach the native conformation of a protein is defined in Equation 7.8, using the average *blob* size $\langle N_{blob} \rangle$ and number of *blobs* $\langle n_b \rangle$ of a protein. In Equation 7.8, τ_p is set equal to 1×10^{-12}

s, which corresponds to the time-scale of local molecular motions and is commonly used as the conformational search time of an *aa* in a protein.[43–46]

$$\tau_F = \Omega\left(< N_{blob} >\right) \times \Omega\left(< n_b >\right) \times \tau_p \qquad (7.8)$$

Correlation of Calculated Folding Times: Following the above procedure, the primary sequence of 145 proteins were converted into strings of N_{blob} values. The N_{blob} values were then used to divide the sequence of the proteins into strings of *blobs*. The average number of *blobs* ($<n_{\text{b}}>$) constituting a protein and their average N_{blob} values ($<N_{\text{blob}}>$) were then used to estimate its folding time using Equation 7.8. The calculated folding times were then compared to experimentally determined τ_{F} values in Figure 7.9.[47,48] A list of the proteins and their experimental and calculated τ_{F}, $<N_{blob}>$, and $<n_{\text{b}}>$ values is provided in Table S7.6 in the SI. The SI also contains more considerations about the procedure applied for predicting the τ_{F} values.

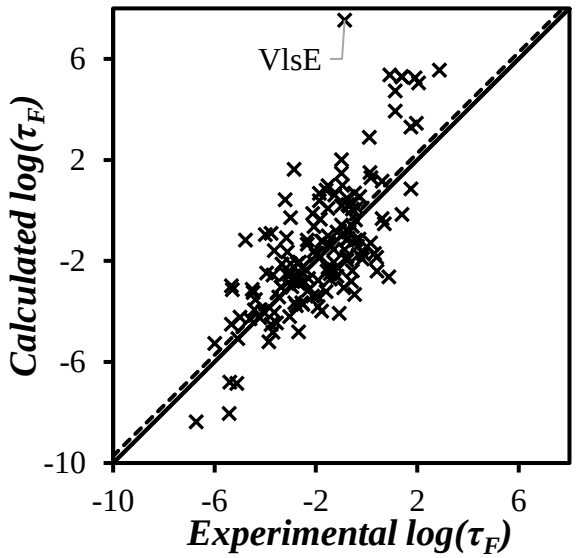


Figure 7.9. Plot of the logarithm of the folding time (τ_F) of proteins calculated from the fluorescence *blob* model as a function of the logarithm of experimentally determined τ_F values. The solid line represents the equality line. The correlation coefficient r of all the proteins is equal to 0.73. Excluding VlsE, r = 0.75 and the line-of-best-fit (dashed): y = 1.00 (±0.07) x + 0.3 (±0.2).

Figure 7.9 shows that overall, the calculated folding times match the experimentally determined ones relatively well with a correlation coefficient of 0.73. The one significant outlier appears to be the V1sE protein (PDB: 1L8W). VlsE is often excluded from correlation trends due to its abnormal folding rate.[49–52] The use of topological descriptors (ex. contact order) suggests that VlsE should fold ~ 4 orders of magnitude faster due to its primarily α-helical structure.[49] Interestingly, polypeptide dynamics used in this study predicts that VlsE should fold much more slowly due to its large size (and many *blobs*). In either case, the folding

time of VlsE is often an outlier in correlation trends. If the VlsE protein is excluded, the correlation coefficient in Figure 7.9 increases to 0.75 demonstrating that polypeptide dynamics are a good predictor for the folding times of proteins, regardless of their size, structural class, or folding kinetics. Furthermore, the 1.00 (±0.07) slope of the line-of-best-fit indicated that the predicted folding times showed a 1:1 relationship with the experimentally determined values and the 0.3 (±0.2) intercept showed that the values were similar. In fact, the average difference between the calculated and observed folding times was used to show that the *blob*-based method described herein was able to predict folding times within ±1.4 orders of magnitude of the experimentally determined values.

The 0.75 correlation coefficient obtained using a *blob*-based approach to predict the folding time of proteins is rather comparable to correlation coefficients obtained using topological parameters. However, what makes the *blob* approach different is that the method explicitly predicts folding times and does so using only information pertaining to the internal dynamics of polypeptides in solution as a function of *aa* composition. In contrast, most other experiment-based models correlate the topology (typically contact topology) of the folded proteins which requires information from the high-resolution structures of the natively folded proteins. In this respect, the *blob*-based approach is the first example in the literature which solely relies on experimentally obtained information on the dynamics of polypeptides to predict the folding time of proteins.

A similar *blob*-based methodology has been applied previously to predict protein folding times, but it used many approximations.[26] The main assumption was that all proteins

had a *blob* size of ~19 *aa*'s which decreased for smaller proteins due to chain end effects. The N_{blob} value of 19 *aa*'s was taken as the average of the N_{blob} values of 22.5 (±1.5) for PyMe-PGlyGlu and 15.8 (±1.0) for PyMe-PAlaGlu. Although this assumption led to a similar correlation coefficient of 0.73 for the same 145 proteins considered, the current study is most important because it takes into consideration the molecular details of a protein sequence and eliminates any possible contribution from the pyrenyl labels used to determine the *blob* sizes. The precise *aa* sequence of each protein was taken into account while determining the size and number of *blobs* of each protein. These corrections were important, as the current model now demonstrates that the folding time of proteins can be calculated based on their individual internal dynamics and local conformation, even when pyrene is accounted for and removed from the process.

7.5 Conclusions

The dynamic and structural parameters retrieved from applying the FBM to specifically designed pyrene-labeled polypeptides were employed to propose a theoretical framework enabling the compartmentalization of a protein sequence into *blobs*, which were specific to each *aa* constituting a protein sequence. This led to the implementation of a program which was used to predict the folding times of proteins. The excellent agreement found between the predicted and experimental folding times suggests that the short folding time of proteins is a mere consequence of polymer chain dynamics, which promote the local interactions in subvolumes of the polypeptide coil, and that these interactions are well described by a *blob*-

based approach like the FBM. This study opens the path to numerous experiments that should be aimed at further refining the form of the f_b(SCS) function defined in Equation 7.5 and establishing a broader library of N_{blob} values for more *aa*'s and different lengths of spacers separating the pyrenyl label from the polypeptide backbone. It is certainly the main advantage of the proposed methodology that experiments can now be designed to probe the specific effects that selected *aa*'s might have on the folding times of proteins.

Chapter 8

Concluding Remarks and Future Work

8.1 Thesis Summary

The primary goal of this thesis was to assess whether a combination of pyrene excimer fluorescence (PEF), fluorescence *blob* model (FBM) analysis, and molecular mechanics optimization (MMO) could characterize the conformation and internal dynamics of polypeptides in solution. The proof of principle for this primary goal was established in Chapter 2 by using poly(*L*-glutamic acid) (P*L*Glu) and poly(*D*,*L*-glutamic acid) (P*DL*Glu) as model polypeptides. P*L*Glu and P*DL*Glu were labeled with 1-pyrenemethylamine to yield Py-P*L*Glu and Py-P*DL*Glu, respectively. Circular dichroism (CD) was used to assess the conformation of Py-P*L*Glu and Py-P*DL*Glu in DMF and DMSO. The absence of molar ellipticity (*[θ]*) of the ^1B_b band of pyrene for Py-P*DL*Glu demonstrated that Py-P*DL*Glu adopted a coiled conformation in both solvents. An increase in [θ] with increasing pyrene content of Py-P*L*Glu indicated that Py-P*L*Glu adopted an α-helical conformation in DMF. While DMSO was expected to denature the Py-P*L*Glu helix into a random coil, the non-zero *[θ]* values suggested that Py-P*L*Glu adopted a structure intermediate between an α-helix and a random coil in DMSO. The fluorescence of the monomer and excimer decays of Py-P*L*Glu and Py-P*DL*Glu were analyzed within the framework of the FBM to yield N_{blob}, the number of structural units (Glu residues) contained within a *blob*, which is the volume probed by an excited pyrene bound to the polypeptide, and k_{blob}, the quenching rate constant in a *blob* containing one exited and one ground-state pyrene. The N_{blob} value of Py-P*DL*Glu was found to be similar in DMF and DMSO, taking an average value of 10 (±1), while the N_{blob} value of Py-P*L*Glu was significantly larger in DMF and DMSO, taking values of 21 (±1) and 18 (±1),

respectively. To help with the assessment of N_{blob}, molecular mechanics optimizations (MMOs) were utilized. α-Helical and coiled conformations of PGlu were constructed *in silico* and were used to determine the largest number of Glu's, referred to as the theoretical N_{blob} value (N_{blob}^{theo}), that could separate two pyrene-labeled Glu residues and still allow for PEF. N_{blob}^{theo} was found to equal 23 and 10 (±1) when PGlu adopted an α-helical and a coiled conformation, respectively. Comparison of the theoretical N_{blob}^{theo} and experimental N_{blob} values led to the conclusion that Py-P*L*Glu adopted an α-helical conformation in DMF and Py-P*DL*Glu adopted a coiled conformation in both DMF and DMSO, in agreement with the results obtained by CD measurements. Interestingly, the N_{blob} values obtained from the FBM were similar to the folding domains found in many proteins that are made of ~ 20 amino acids (*aa*'s), which suggested that the *blobs* characterized by the FBM had a similar length scale as the domains generated during protein folding. The internal dynamics of the polypeptides were investigated next with the term k_{blob}, which represents the rate constant at which the *aa*'s encompassed inside a *blob* encounter each other. The Py-P*L*Glu α-helices and Py-P*DL*Glu random coils yielded similar k_{blob} values, which suggested that interactions between the side chains of the Glu residues occurred at a similar frequency regardless of the backbone conformation. Whereas k_{blob} describes the encounter frequency between Glu residues, the inverse of k_{blob} (k_{blob}^{-1}) reports on the time scale over which these encounters occur. For Py-P*DL*Glu, it was found that the 10 Glu's constituting a *blob* would interact with one another on a ~ 73 ns time scale, which is the same time scale expected for the internal dynamics experienced by a hypothetical protein made of N_{blob} *aa*'s (3^{Nblob}×1 ps ~ 60 ns). The ability for

the FBM to not only provide information on the structure and dynamics of a polypeptide, but to do so on the same time- and length-scale at which proteins fold, suggested that the diffusive encounters between the *aa*'s, which define the *blob* of a polypeptide, are also responsible for defining the folding domains in proteins. This insight provided the opportunity to use the FBM to explore in a controlled manner how changes in the conformation and amino acid (*aa*) composition of a polypeptide affected the number of *aa*'s which can locally interact with one another in a *blob* and the rate at which they do so.

In Chapter 3, the FBM analysis of polypeptides was pushed further by investigating the effect that the charges borne by *aa*'s had on the *blob* size and dynamics. This was accomplished by comparing the N_{blob} and k_{blob} values, obtained for Py-P*L*Glu and Py-P*DL*Glu in their protonated form in Chapter 2 to those obtained in their deprotonated form, referred to as Py-P*L*GNa and Py-P*DL*GNa, respectively. The FBM revealed that the *blob* size (N_{blob}) remained constant in DMSO, whether or not the Glu side chains were charged. Since Py-P*DL*Glu and Py-P*DL*GNa both adopted a coiled conformation, the similar N_{blob} values of ~10 indicated that the *blob* volume was independent of whether the side chains were charged. Interestingly, the relatively large N_{blob} value of ~18 for Py-P*L*Glu in DMSO persisted even when it was deprotonated, which indicated that P*L*Glu and P*L*GNa must retain some structured conformation in DMSO. Since the N_{blob} value of ~18 *aa*'s did not match either of the N_{blob}^{theo} values of 23 or 10 obtained by MMOs for P*L*Glu in either an α-helical or random coil conformation, respectively, it implied that Py-P*L*Glu and Py-P*L*GNa in DMSO did not adopt either conformation. Consequently, the MMOs were extended to include Py-P*L*Glu constructs

adopting the polyproline type-II (PPII) and 3_{10}-helical conformations. The 3_{10}-helical conformation of PGlu yielded an N_{blob}^{theo} value of 19, matching the N_{blob} value of ~ 18 *aa*'s obtained for Py-P*L*Glu in DMSO and identifying the 3_{10}-helix as the probable conformation adopted by P*L*Glu in DMSO. This marked the first time that the FBM had been used to identify the unknown conformation of a macromolecule in solution. Comparison of the k_{blob} values obtained for the protonated and deprotonated Py-PGlu samples showed that the presence of the charged glutamate side chains led to a ~ 40% decrease in the side chain mobility due to their ionic repulsions. A pyrene-labeled poly(*L*-lysine) (Py-P*L*L) was also studied in its protonated (charged state). Using a combination of MMOs and the FBM, Py-P*L*L was found to adopt a coiled conformation in DMSO with an N_{blob} value of 14 (±1). Compared to the coiled Py-P*DL*Glu, the longer side chain of lysine connecting the pyrenyl label onto the P*L*L backbone allowed the excited pyrene to probe a larger volume, resulting in a larger *blob* size despite the similarity in their conformations. This highlighted the important effect that *aa* side chain length had on the *blob* size. The similar N_{blob} values found for different polypeptides regardless of the side chain charges and the dependency of N_{blob} on the length of the spacer separating the pyrenyl label from the polypeptide backbone suggested that the size of the folding domains in proteins are also independent of the presence of charges, and that the domain size depends only on the conformation and side chain length of the *aa*'s.

The increase in N_{blob} of Py-P*DL*Glu from 10 in a coiled conformation to 21 for Py-P*L*Glu in an α-helical conformation showed that the *blob* size depended on the local conformation adopted by the polypeptide. To further investigate this effect, Chapter 4 aimed

to characterize how the *blob* size of Py-P*L*L changed as P*L*L folded into more complex architectures. The smallest *blob* size of Py-P*L*L was found to be 14 (±1), when P*L*L adopted a coiled conformation in either DMSO or a 60:40 acetonitrile:water mixture. When the acetonitrile content of the acetonitrile:water mixture was increased to 90%, the Py-P*L*Ls aggregated into large structures. The addition of a 50-fold excess of unlabeled P*L*L isolated the Py-P*L*Ls from one another, which allowed for their individual characterization. The N_{blob} value of the isolated Py-P*L*Ls equaled 25 (±1), corresponding to the N_{blob}^{theo} value expected for an α-helical conformation using MMOs. AFM images of the Py-P*L*Ls revealed that the large aggregates were composed of organized bundles of Py-P*L*Ls aligned parallel to one another. With this information, the FBM was then applied to analyze the fluorescence decays of a solution of the aggregated Py-P*L*Ls. The presence of intermolecular interactions between the Py-P*L*L α-helices led to a further increase in N_{blob} to 37 (±3). The increase in N_{blob} by ~12 lysines upon incorporation into a P*L*L bundle was used to determine that the bundled P*L*L helices must have been separated by an interhelical distance of 2.9 nm, resulting in a density of 0.25 g/mL. The evolution of N_{blob} as Py-P*L*L underwent a conformational change from a random coil, to an α-helix, and finally to bundles of α-helices provided a snapshot of how the number of *aa*'s capable of diffusively encountering one another increased with increasing structural order. This insight led to the conclusion that the folding subdomains generated in a protein are expected to increase in size along its folding pathway due to the increased number of interactions experienced by the *aa*'s.

The firm conclusions drawn in Chapters 2 – 4, that the FBM could successfully characterize the ordered conformations of polypeptides in solution, led to the investigation in Chapter 5 of the effect that the *aa* sequence of a polypeptide has on its conformational freedom and dynamics, when the polypeptide is in an unstructured state. This study was conducted by preparing a series of copolypeptides containing racemic mixtures of *D,L*-Glu and either glycine (Gly), *D,L*-alanine (Ala), or carbobenzyloxy protected *D,L*-lysine (Lys(Z)) to produce PGlyGlu, PAlaGlu, and PLys(Z)Glu, respectively. The copolypeptides were all prepared with a similar Glu content of ~ 44 mol% to ensure that the only difference in the copolypeptide composition was the choice of the *aa* comonomer. Furthermore, the racemic mixtures of *aa*'s used in the preparation of the copolypeptides ensured that they could not form secondary structures in solution. The FBM results of the pyrene-labeled copolypeptides were then compared to those obtained with Py-P*DL*Glu to determine what effect the side chain size (SCS) of an *aa* had on the polypeptide *blobs*. The N_{blob} values of Py-PLys(Z)Glu was similar to that of Py-P*DL*Glu, which suggested that all copolypeptides prepared with *aa*'s having a SCS similar to, or larger than, Glu were expected to adopt an elongated conformation on the length-scale of a *blob*. Compared to the average 11 (±1) *blob* size of Py-P*DL*Glu and Py-PLys(Z)Glu, N_{blob} was found to increase to 16 (±1) for Py-PAlaGlu and 23 (±2) for Py-PGlyGlu in DMSO due to a reduction in steric hinderance between the small side chains of Ala and Gly and their neighboring residues. The reduction in steric hindrance meant that the polypeptide backbone was now much more flexible, allowing many more *aa*'s to interact with one another within a *blob*. The surprisingly narrow range of N_{blob} values obtained in this study suggested that 17 of

the 20 most common *aa*'s were expected to have only very minor effects on the backbone conformation of a polypeptide, resulting in estimated *blob* sizes of 11 to 14 *aa*'s. Significant changes in a polypeptide conformational freedom required that either Ala or Gly be present in the sequence. The internal dynamics experienced by the copolypeptides were also compared with the product $k_{blob} \times N_{blob}$, which represents the number of encounters between the *aa*'s in a *blob* per unit time. $k_{blob} \times N_{blob}$ was found to remain constant for all the pyrene-labeled polypeptides except for Py-PGlyGlu, which resulted in ~twice the frequency of encounters between the *aa*'s. This suggested that the *aa* sequence of a polypeptide has a very limited effect on its internal dynamics unless Gly (and very likely proline, which was not included in the study) is present.

The preliminary results of Chapter 5 indicated that the incorporation of *aa*'s containing small side chains defined the number of *aa*'s capable of interacting with one another in a polypeptide sequence and the frequency at which they did so. However, these conclusions were based on copolypeptides containing a large (~60 mol%) fraction of comonomer. Chapter 6 compared a series of Py-PAlaGlu samples with alanine contents ranging from 24 to 58 mol% to Py-P*DL*Glu in DMSO to investigate how the *blob* size and internal dynamics of a polypeptide respond to the level of incorporation of small *aa*'s. $k_{blob} \times N_{blob}$ was found to decrease with increasing Ala content due to either the increase in *blob* volume or *blob* density. More interestingly, N_{blob} increased from 11 (±2) for Py-P*DL*Glu (which contained 0 mol% Ala) to 16 (±1) for Py-PAlaGlu containing 24 mol% of Ala, at which point N_{blob} remained constant as a function of Ala content. The constancy of N_{blob} with Ala content was rather surprising and

indicated that the *blob* size was apparently controlled by the smallest *aa* present in a *blob*. This insight led to an attempt of estimating the folding time of 145 proteins by viewing a protein as a string of folding *blobs*. Assuming that each *blob* in a protein contained either an Ala or Gly (the smallest *aa*'s), the *blob* size of a protein was estimated to be ~ 19 *aa*'s, the average of the N_{blob} values obtained for Py-PAlaGlu and Py-PGlyGlu. Although this approximation was rather crude, a 0.73 correlation coefficient and a near 1:1 scaling was achieved between the calculated and experimentally determined folding times of 145 proteins. The ability of the proposed *blob*-based method to predict the folding times of proteins supported the earlier conclusions drawn in Chapters 2 – 4 that protein folding is governed by the local interactions of the *aa*'s contained inside *blobs*. This study also suggested that *blob*-based approaches such as those presented in this thesis might be ideal mathematical tools to represent the diffusive encounters between the *aa*'s constituting *foldons*, which in turn could provide an effective means to better understand protein folding.

The final experimental chapter refined the method for estimating the *blob* size of a protein by taking into account the side chain length (SCL) of each *aa*, the contribution of pyrene towards a *blob* size, and the exact *aa* sequence of proteins. The effect of SCL and the contribution of pyrene was accounted for by comparing the *blob* sizes of a series of pyrene-labeled polypeptides with pyrene attached at incrementally longer SCLs from the polypeptide backbone. A plot of N_{blob} against SCL showed that a 1 atom increase in the SCL of a reference *aa* resulted in a 0.89 increase in the number of *aa*'s that the reference *aa* could interact with on each of its two sides along the oligopeptide segment constituting its *blob*. Extrapolating the

SCL to 0 atoms revealed that the pyrene molecule increased the *blob* size of the pyrene-labeled polypeptides by 0.7 *aa*'s. This information was then used to calculate the inherent *blob* size (N(SCL)) for each *aa* when the polypeptide backbone was in an extended conformation. The results of Chapters 5 and 6 were then generalized to account for changes in the N_{blob} values due to the increase in conformational freedom imparted by *aa*'s having a small SCS by defining the bending function f_b(SCS), which was defined by the local sequence surrounding an *aa*. The *blob* size for each *aa* in the sequence of a protein was determined from the product N(SCL)×f_b(SCS). The *blobs* were then sorted based on their size to determine the average size and number of *blobs* in a protein. These numbers were then used to estimate the protein folding time based on the number of configurations available to a protein according to the procedure that had been described in Chapter 6. The calculated folding times of 144 proteins were within ~ ±1.4 order of magnitude of the experimentally determined values, yielding a correlation coefficient of 0.75. The ability of the proposed *blob*-based approach to predict the folding time of proteins suggests that the complexity of protein folding arises from the chain dynamics and interactions experienced by the oligopeptide segments that define the subvolumes, where the folding of a protein actually takes place.

8.2 Future Work

The main conclusion of the studies presented in this thesis, namely that the polypeptide *blobs* can be viewed as the domains where folding occurs in proteins leaves this area open to many avenues of research. The early chapters focused on the effect that the polypeptide structure has

on *blob* size. This effect could be investigated further by studying the *blobs* generated by other secondary structures found in proteins, such as β-sheets. Preparing macromolecules with well-defined densities of polypeptide chains, such as a bottle-brush type architectures with polypeptide branches, could be used to investigate in a controlled manner changes in the *blob* size and internal dynamics of the polypeptide chains as a function of the density of the polypeptide branches. These types of studies could then be applied to map the changes in a polypeptide behavior, as the side chains interact with one another in an increasingly crowded environment.

The latter chapters focused on how the *aa* sequence affected the behavior of a polypeptide in solution. The assumption that the *blob* size of a polypeptide is defined by the inclusion of a single small *aa* was based on the constant N_{blob} values obtained for the Py-PAlaGlu samples as a function of alanine content in Chapter 6. To this end, two additional experiments should be conducted to further support this conclusion. First, the lowest alanine content of the PAlaGlu samples used in Chapter 6 was 24 mol%. Consequently, a PAlaGlu with a lower (~5 – 10 mol%) alanine content should be prepared to assess whether or not a constant N_{blob} value continues to apply for very low alanine contents. Second, it would be beneficial to confirm that any copolypeptide prepared with small *aa*'s exhibit the same pattern as the PAlaGlu series, namely that their *blob* size is independent of the small *aa* content. To test this assumption, a series of Py-PGlyGlu's with a glycine content ranging from ~ 10 to 40 mol% could be prepared to evaluate how N_{blob} varies as a function of glycine content.

The experiments described in this thesis led to the broader question of how the heterogeneous sequence of a polypeptide affects the *blob* size and chain dynamics. Chapter 7 handled this question by introducing the bending function f_b(SCS), which defines the increase in the *blob* size of an *aa* due to the flexibility of its surrounding oligopeptide sequence. The f_b(SCS) term was built from the results obtained with the copolypeptides prepared in Chapters 5 – 7. In addition to the suggested experiments mentioned above, additional experiments could be designed to improve the accuracy of f_b(SCS). Some simple experiments would involve the preparation of Py-P*X*Glu copolypeptides, where *X* is any *aa* of choice, to map the response of the N_{blob} value to each *aa*, while keeping the Glu residue for pyrene labeling. One noteworthy *aa* would be proline, whose cyclic structure could not be accounted for in the current studies. The comparison of leucine and isoleucine copolypeptides would also be of interest due to their relatively small side chains and their difference in branching. Preparation of these two copolypeptides would be used to investigate the impact that branching has on the dynamics of a copolypeptide. Furthermore, f_b(SCS) relies on the relative change in N_{blob} for a 1-pyrenemethyl amine attached onto glutamic acid resulting in a 5 atom-linker as a function of comonomer side chain size. To ensure that the *blob* size defined by the length of the side chain of an *aa* responds to chain flexibility according to the same f_b(SCS) function, the effect of the comonomer side chain size (SCS) should be investigated for copolypeptides containing pyrene attached onto *aa*'s other than glutamic acid. Lysine-based copolypeptides would be ideal candidates for these studies since 1-pyreneacetic acid could easily be attached onto the ε-amine of lysine, resulting in a 7-atom linker separating pyrene from the polypeptide backbone. Lastly,

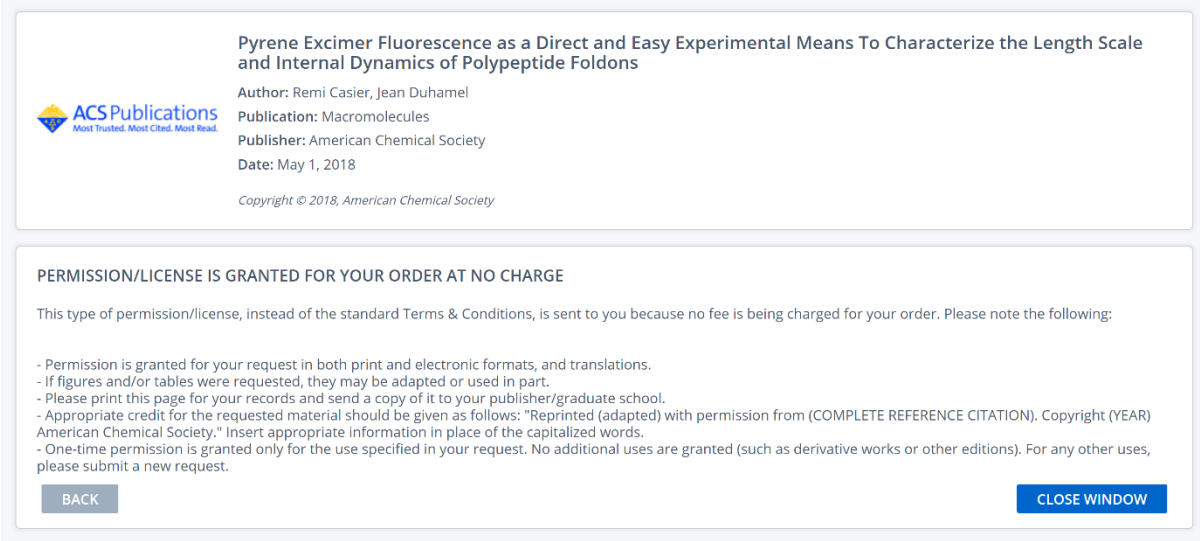
f_b(SCS) assumes that polypeptides containing more than 2 different *aa*'s behaved in a similar manner to the studied copolypeptides (i.e. only the smallest *aa* in a *blob* is important). This assumption should be verified by preparing polypeptides from 3 or more different *aa*'s. Out of the *aa*'s studied in this thesis, the terpolypeptide of glutamic acid, alanine, and carbobenzyloxy-lysine (Lys(Z)) could be used to determine if the very large side chain of Lyz(Z) would dampen the increased flexibility imparted to the polypeptide backbone by alanine, or if alanine does indeed define the *blob* size regardless of any other larger *aa*'s. The results obtained by applying the FBM to polypeptides of increasingly complex sequences would be used to refine f_b(SCS), to more precisely predict the *blob* sizes in the pseudo-infinite combination of *aa*'s defining a protein. To verify the ability of the procedure introduced in Chapter 7 to predict the *blob* size in proteins, the FBM could be applied to a denatured protein. The protein would need to have a pseudo-random distribution of *aa*'s to which pyrene could be labeled, such as glutamic acid or lysine. After denaturing the protein with guanidinium chloride in DMSO, it could be labeled with pyrene and studied with the FBM. The results obtained from the FBM could then be used to ensure that f_b(SCS) can accurately predict the *blob* size of proteins, and to refine the selection criteria for defining the *blobs* constituting a protein.

Letters of Copyright Permissions

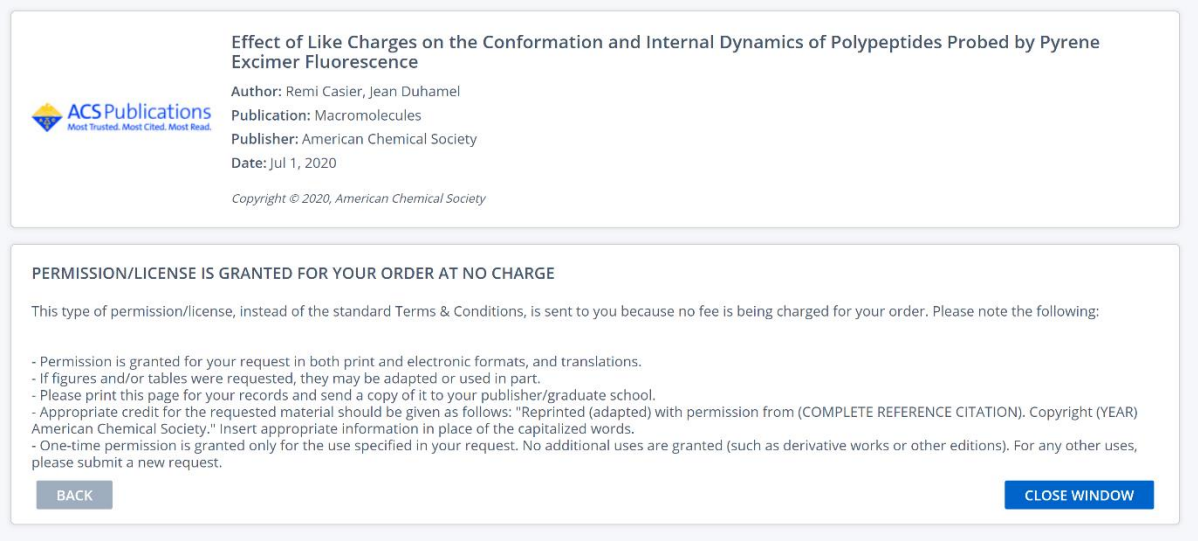
Permissions for reproduction of Figure 1.4.

License Number	4927240776812
License date	Oct 13, 2020
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Structural Biology
Licensed Content Title	From Levinthal to pathways to funnels
Licensed Content Author	Ken A. Dill et al
Licensed Content Date	Jan 1, 1997
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	The Structure and Internal Dynamics of Polypeptides Probed with Pyrene Excimer Fluorescence
Institution name	University of Waterloo
Expected presentation date	Jan 2021
Portions	Figure 4

Permissions for reproduction of Chapter 2.



Permissions for reproduction of Chapter 3.



Effect of Like Charges on the Conformation and Internal Dynamics of Polypeptides Probed by Pyrene Excimer Fluorescence

Author: Remi Casier, Jean Duhamel
Publication: Macromolecules
Publisher: American Chemical Society
Date: Jul 1, 2020

Copyright © 2020, American Chemical Society

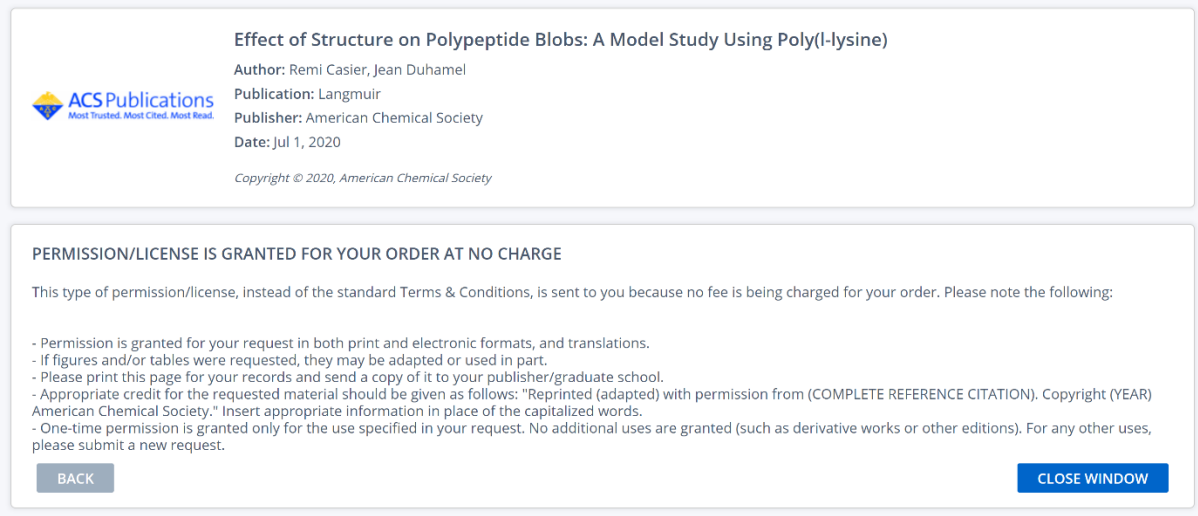
PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK CLOSE WINDOW

Permissions for reproduction of Chapter 4.



Effect of Structure on Polypeptide Blobs: A Model Study Using Poly(l-lysine)

Author: Remi Casier, Jean Duhamel
Publication: Langmuir
Publisher: American Chemical Society
Date: Jul 1, 2020

Copyright © 2020, American Chemical Society

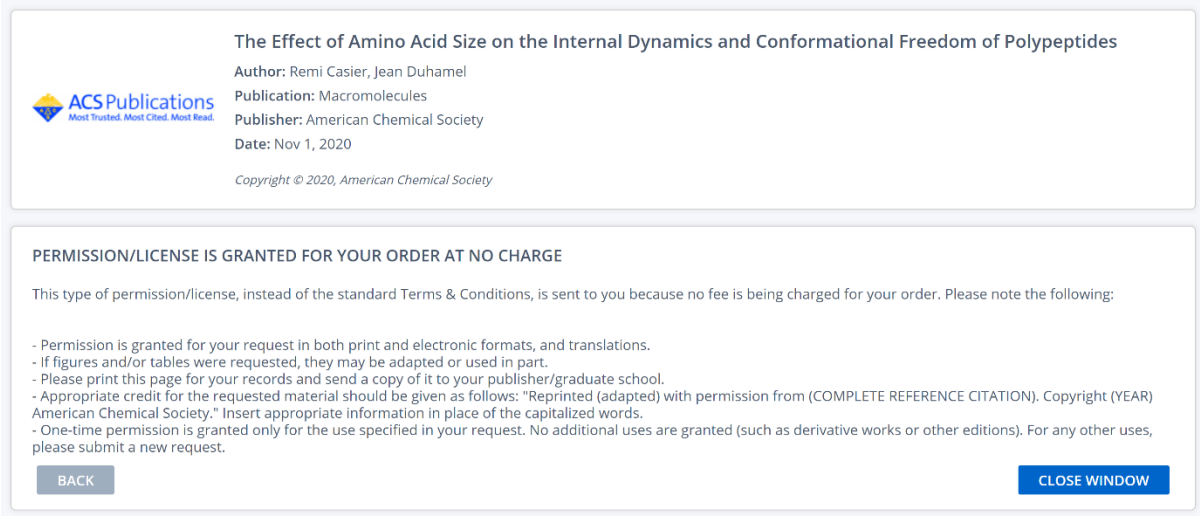
PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK CLOSE WINDOW

Permissions for reproduction of Chapter 5.



The Effect of Amino Acid Size on the Internal Dynamics and Conformational Freedom of Polypeptides

Author: Remi Casier, Jean Duhamel
Publication: Macromolecules
Publisher: American Chemical Society
Date: Nov 1, 2020

Copyright © 2020, American Chemical Society

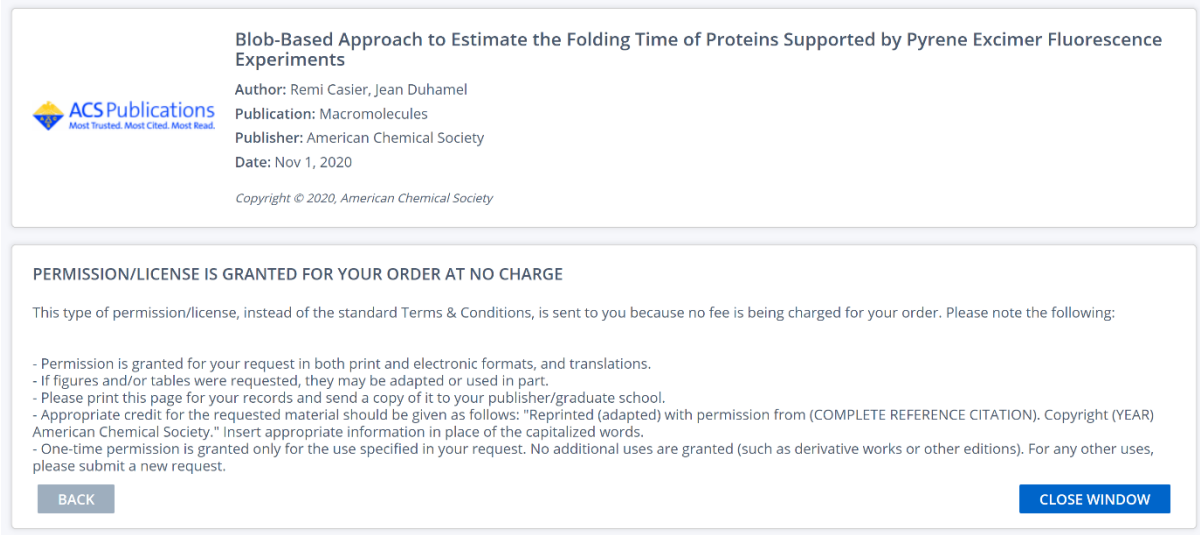
PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK CLOSE WINDOW

Permissions for reproduction of Chapter 6.



Blob-Based Approach to Estimate the Folding Time of Proteins Supported by Pyrene Excimer Fluorescence Experiments

Author: Remi Casier, Jean Duhamel
Publication: Macromolecules
Publisher: American Chemical Society
Date: Nov 1, 2020

Copyright © 2020, American Chemical Society

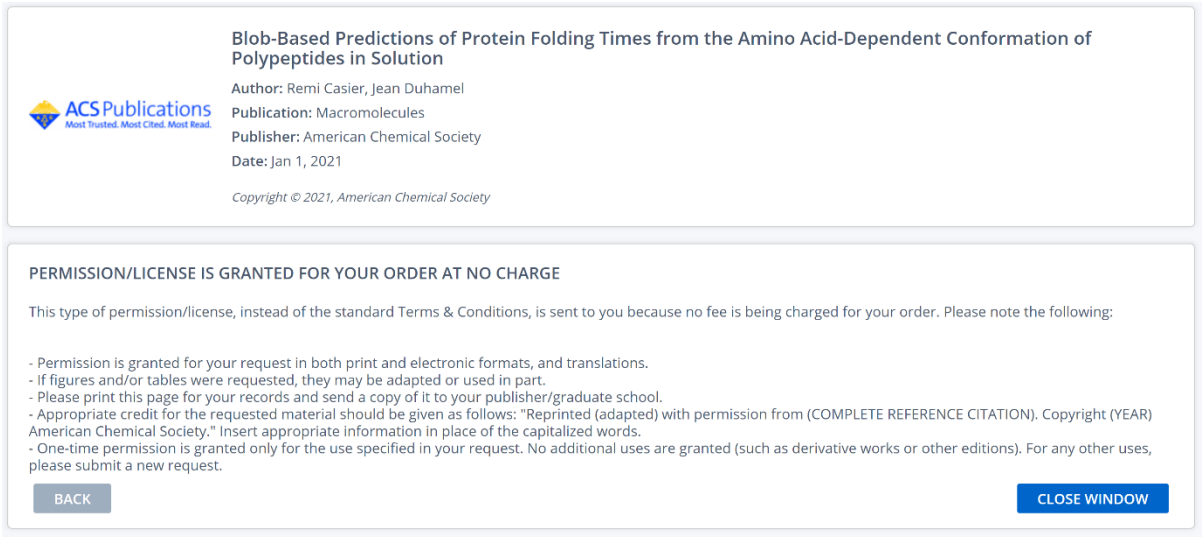
PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK CLOSE WINDOW

Permissions for reproduction of Chapter 7.



Blob-Based Predictions of Protein Folding Times from the Amino Acid-Dependent Conformation of Polypeptides in Solution

Author: Remi Casier, Jean Duhamel

Publication: Macromolecules

Publisher: American Chemical Society

Date: Jan 1, 2021

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK CLOSE WINDOW

References

Chapter 1

1. Voet, D.; Voet, J. G. Biochemistry, 4th ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2011, pp 221 – 290.
2. Wagner, I.; Musso, H. New Naturally Occurring Amino Acids *Angew. Chem. Int. Ed. Engl.* **1983**, *22*, 816 – 828
3. Bahar, I.; Jernigan, R. L.; Dill, K. A. Protein Actions: Principles & Modeling; Garland Science, Taylor & Francis Group, LLC: New York, NY, 2017, pp 1 – 28.
4. Pauling, L. The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry, 3ed ed.; Cornell University Press: Ithaca, NY, 1960, pp 498.
5. Alderson, T. R.; Lee, J .H.; Charlier, C. C.; Ying, J.; Bax, A. Propensity for Cis-Proline Formation in Unfolded Proteins. *ChemBioChem* **2018**, *19*, 37 – 42.
6. Kemnitz, C. R.; Loewen, M. J. "Amide Resonance" Correlates with a Breadth of C-N Rotation Barriers. *J. Am. Chem. Soc.* **2007**, *129*, 2521 – 2528.
7. Ramachandran, G. N.; Ramakrishnan, C. Sasisekharan, V. Stereochemistry of Polypeptide Chain Configurations. *J. Mol. Biol.* **1963**, *7*, 95 – 99.
8. Chakrabarti, P.; Pal, D. The Interrelationships of Side-Chain and Main-Chain Conformations in Proteins. *Prog. Biophys. Mol. Biol.* **2001**, *76*, 1 – 102.
9. Kumar, S.; Bansal, M. Geometrical Characteristics of α-Helices in Globular Proteins. *Biophys. J.* **1998**, *75*, 1935 – 1944.

10. Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. H. Jr. The Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide Chain. *Proc. Natl. Acad. Sci. U. S. A.* **1961**, *47*, 1309 – 1314.
11. Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science* **1973**, *181*, 223 – 230.
12. Pace, C. N. Conformational Stability of Globular Proteins. *Trends Biochem. Sci.* **1990**, *15*, 14 – 17.
13. Sheu, S.-Y., Yang, D.-Y.; Selzle, H. L.; Schlag, E. W. Energetics of Hydrogen Bonds in Peptides. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12683 – 12687.
14. Zhou, H.-X.; Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem. Rev.* **2018**, *118*, 1691 – 1741.
15. Dill, K. A. Dominant Forces in Protein Folding. *Biochemistry* **1990**, *29*, 7133 – 7155.
16. Pace, C. N.; Scholtz, J. M.; Grimsley, G. R. Forces Stabilizing Proteins. *FEBS Lett.* **2014**, *558*, 2177 – 2184.
17. Barlow, D. J.; Thornton, J. M. Ion-Pairs in Proteins. *J. Mol. Biol.* **1983**, *168*, 867 – 885.
18. Fersht, A. R. Conformational Equilibria in α- and δ- chymotrypsin: The Energetics and Importance of the Salt Bridge. *J. Mol. Biol.* **1972**, *64*, 497 – 509.
19. Perutz, M. F.; Raidt, H. Stereochemical Basis of Heat Stability in Bacterial Ferredoxins and in Haemoglobin A2. *Nature* **1975**, *225*, 256 – 259.

20. Anderson, D. E.; Becktel, W. J.; Dahlquist, F. W. pH-Induced Denaturation of Proteins: A single Salt Bridge Contributes 3 – 5 kcal/mol to the Free Energy of Folding of T4 Lysozyme. *Biochemistry* **1990**, *29*, 2403 – 2408.
21. Pace, C. N.; Alston, R. W.; Shaw, K. L. Charge-Charge Interactions Influence the Denatured Sate Ensemble and Contribute to Protein Stability. *Protein Sci.* **2000**, *9*, 1359 – 1398.
22. Stickle, D. F.; Presta, L. G.; Dill, K. D.; Rose, G. D. Hydrogen Bonding in Globular Proteins. *J. Mol. Biol.* **1992**, 226, 1143 – 1159.
23. Bowie, J. Membrane Protein Folding: How Important are Hydrogen Bonds? *Curr. Opin. Struct. Biol.* **2011**, *21*, 42 – 49.
24. Pace, C. N.; Fu, H.; Fryar, L. K.; Landua, J.; Trevino, S. R.; Schell, D.; Thurlkill, R. L.; Imura, S.; Scholtz, J. M.; Gajiwala, K.; Sevcik, J.; Urbanikova, L.; Myers, J. K.; Takano, K.; Hebert, E. J.; Shirley, B. A.; Grimsley, G. R. Contribution of Hydrogen Bonds to Protein Stability. *Protein Sci.* **2014**, *23*, 652 – 661.
25. Pace, C. N.; Fu, H.; Fryar, K. L.; Landua, J.; Trevino, S. R.; Shirley, B. A.; Hendricks, M. M.; Iimura, S.; Gajiwala, K.; Scholtz, J. M.; Grimsley, G. R.; Contribution of Hydrophobic Interactions to Protein Stability. *J. Mol. Biol.* **2011**, 408, 514 – 528.
26. Lesser, G. J.; Rose, G. D. Hydrophobicity of Amino Acid Subgroups in Proteins. *Proteins* **1990**, *8*, 6 – 13.

27. Klapper, M. H. On the Nature of the Protein Interior. *Biochim. Biophys. Acta.* **1971**, *229*, 557 – 566.
28. Lazaridis, T.; Archontis, G.; Karplus, M. Enthalpic Contribution to Protein Stability: Insights from Atom-Based Calculations and Statistical Mechanics. *Adv. Protein Chem.* **1995**, *47*, 231 – 306.
29. Brady, G. D.; Sharp, K. A. Entropy in Protein Folding and in Protein-Protein Interactions. *Curr. Opin. Struct. Biol.* **1997**, *7*, 215 – 221.
30. Trevino, S. R.; Gokulan, K.; Newsom, S.; Thurlkill, R. L.; Shaw, K. L.; Mitkevich, V. A.; Makarov, A. A.; Sacchettini, J. C.; Scholtz, J. M.; Pace, C. N. Asp79 Makes a Large, Unfavorable Contribution to the Stability of RNase Sa. *J. Mol. Biol.* **2005**, *354*, 967 – 978.
31. Dill, K. A.; Chan, H. S. From Levinthal to Pathways to Funnels. *Nat. Struct. Biol.* **1997**, *4*, 10 – 19.
32. Rollins, G. C.; Dill, K. A. General Mechanism of Two-State Protein Folding Kinetics. *J. Am. Chem. Soc.* **2014**, *136*, 1142 – 11427.
33. Smyth, M. S.; Martin, J. H. J. X Ray Crystallography. *Mol. Pathol.* **2000**, *53*, 8 – 14.
34. Markwick, P. R. L.; Malliavin, T.; Nilges, M. Structural Biology by NMR: Structure, Dynamics, and Interactions. *PLoS Comput. Biol.* **2008**, *4*, e1000168.
35. Nwanochie, E.; Uversky, V. N. Structure Determination by Single-Particle Cryo-Electron Microscopy: Only the Sky (and Intrinsic Disorder) is the Limit. *Int. J. Mol. Sci.* [Online] **2019**, *20*, 4186; https://doi.org/10.3390/ijms20174186.

36. Carter, C. W.; Baldwin, E. T.; Frick, L. Statistical Design of Experiments for Protein Crystal Growth and the Use of a Precrystallisation Assay. *J. Cryst. Growth* **1988**, 90, 60 – 73.

37. Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M. Protein Crystallography for Non-Crystallographers, or How to Get the Best (But Not More) from Published Macromolecular Structures. *FEBS J.* **2008**, *275*, 1 – 21.

38. Silverstein, R. M.; Webster, F. W.; Kimble, D. J. Spectrometric Identification of Organic Compounds, 7th ed; John Wiley & Sons, Inc. Hoboken, NJ, 2005, pp 127 -143.

39. Jiang, Y.; Kalodimos, C. G. NMR Studies of Large Proteins. *J. Mol. Biol*. **2017**, *429*, 2667 – 2676.

40. Holden, N. E. "Table of the Isotopes", in Lide, D. R., Ed., CRC Handbook of Chemistry and Physics, 86th ed., CRC Press, Boca Raton FL, 2005.

41. Acton, T. B.; Xiao, R.; Anderson, S.; Aramini, J.; Buchwald, W. A.; Ciccosanti, C.; Conover, K.; Everett, J.; Hamilton, K.; Huang, Y. J.; Janjua, H.; Kornhaber, G.; Lau, J.; Lee, D. Y.; Liu, G.; Maglaqui, M.; Ma, L.; Mao, L.; Patel, D.; Rossi, P.; Sahdev, S.; Shastry, R.; Swapna, G. V. T.; Tang, Y.; Tong, S.; Wang, D.; Wang, H.; Zhao, L.; Montelione, G. T. Chapter Two - Preparation of Protein Samples for NMR Structure, Function, and Small-Molecule Screening Studies. *Methods Enzymol*. **2011**, *493*, 21 – 60.

42. Edgelman, E. H. The Current Revolution in Cryo-EM. *Biophys. J.* **2016**, *110*, 1008 - 1012.

43. Bartesaghi A., Merk A., Subramaniam S. 2.2 Å Resolution Cryo-EM Structure of β-Galactosidase in Complex with a Cell-Permeant Inhibitor. *Science* **2015**, *348*, 1147 – 1151.
44. Greenfield, N. J. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. *Nat. Protoc.* **2006**, *1*, 2876 – 2890.
45. Barth, A.; Infrared Spectroscopy of Proteins. *Biochim. Biophys. Acta. Biomembr.* **2007**, *1767*, 1073 – 1101.
46. Lakowicz, J. R. Principles of Fluorescence Spectroscopy, 3rd Ed. Springer: Singapore, 2006, pp 529 - 567.
47. Beychok, S. Circular Dichroism of Biological Macromolecules. *Science* **1966**, *154*, 1288 – 1299.
48. Venyaminov, S. Y.; Prendergast, F. G. Water (H2O and D2O) Molar Absorptivity in the 1000 – 4000 cm−1 Range and Quantitative Infrared Spectroscopy of Aqueous Solutions. *Anal. Biochem.* **1997**, *248*, 234 – 245.
49. Sancho, J. The Stability of 2-State, 3-State and More-State Proteins from Simple Spectroscopic Techniques…Plus the Structure of the Equilibrium Intermediates at the Same Time. *Arch. Biochem. Biophys.* **2013**, *531*, 4 – 13.
50. Walters, J.; Milam, S. L.; Clark, A. C. Chapter 1 Practical Approaches to Protein Folding and Assembly: Spectroscopic Strategies in Thermodynamics and Kinetics. *Methods Enzymol.* **2009**, *455*, 1 – 39.

51. Wallace, L. A.; Matthews, C. R. Sequential vs. Parallel Protein-Folding Mechanisms: Experimental Tests for Complex Folding Reactions. *Biophys. Chem.* **2002**, *101 – 102*, 113 – 131.

52. Frieden, C.; Hoeltzli, S. D.; Ropson, I. J. NMR and Protein Folding: Equilibrium and Stopped-Flow Studies. *Protein Sci.* **1993**, *2*, 2007 – 2014.

53. Lapidus, L. L.; Yao, S.; McGarrity, K. S.; Hertzog, D. E.; Tubman, E. Protein Hydrophobic Collapse and Early Folding Steps Observed in a Microfluidic Mixer. *Biophys. J.* **2007**, *93*, 218 – 224.

54. Chen, Y.; Ding, F.; Nie, H.; Serohijis, A. W.; Sharma, S.; Wilcox, K. C.; Yin, S.; Dokholyan, N. V. Protein Folding: Then and Now. *Arch. Biochem. Biophys.* **2008**, *469*, 4 – 19.

55. Baldwin, R. L. On-Pathway versus Off-Pathway Folding Intermediates. *Fold. Des.* **1996**, *1*, R1 – R8.

56. Chen, E.; Goldbeck, R.A.; Kliger, D.S. Probing Early Events in Ferrous Cytochrome C Folding with Time-Resolved Natural and Magnetic Circular Dichroism Spectroscopies. *Curr Protein Pept. Sci*. **2009**, *10*, 464 – 475.

57. Akiyama, S.; Takahashi, S.; Ishimori, K.; Morishima, I. Stepwise Formation of α-Helices During Cytochrome C Folding. *Nat. Struct. Biol.* **2000**, *7*, 514 – 520.

58. Kubelka, J. Hofrichter, J.; Eaton, W. A. The Protein Folding 'Speed Limit'. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76 – 88.

59. Banavar, J.R.; Maritan, A.; Micheletti, C.; Trovato, A. Geometry and Physics of Proteins. *Proteins* **2002**, *47*, 315 – 322.

60. Chan, H. S.; Dill, K. A. Origins of Structure in Globular Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 6388 – 6392.

61. Uversky, V. N.; Fink, A. L. The Chicken-Egg Scenario of Protein Folding Revisited. *FEBS Lett.* **2002**, *515*, 79 – 83.

62. Krieger, F.; Fierz, b.; Bieri, O.; Drewello, M.; Kiefhaber, T. Dynamics of Unfolded Polypeptide Chains as Model for the Earliest Steps in Protein Folding. *J. Mol. Biol.* **2003**, *332*, 265 – 274.

63. Huang, F.; Nau, W. M. A Conformational Flexibility Scale for Amino Acids in Peptides. *Angew. Chem. Int. Ed.* 2003, 42, 2269 – 2272.

64. Fierz, B.; Kiefhaber, T. End-to-End vs Interior Loop Formation Kinetics in Unfolded Polypeptide Chains. *J. Am. Chem. Soc.* **2007**, *129*, 672 – 679.

65. Sridevi, K.; Lakshmikanth, G. S.; Krishnamoorthy, G.; Udgaonkar, J. B. Increasing Stability Reduces Conformational Heterogeneity in a Protein Folding Intermediate Ensemble. *J. Mol. Biol* **2004**, *337*, 699 – 711.

66. Shastry, M. C. R.; Udgaonkar, J. B. The Folding Mechanism of Barstar: Evidence for Multiple Pathways and Multiple Intermediates. *J. Mol. Biol.* **1995**, *247*, 1013 – 1027.

67. Houry, W. A.; Scheraga, H. A. Structure of a Hydrophobically Collapsed Intermediate on the Conformational Folding Pathway of Ribonuclease A Probed by Hydrogen-Deuterium Exchange. *Biochemistry* **1996**, *35*, 11734 – 11746.

68. Houry, W. A.; Scheraga, H. A. Nature of the Unfolded State of Ribonuclease A: Effect of Cis-Trans X-Pro Peptide Bond Isomerization. *Biochemistry* **1996**, *35*, 11719 – 11733.

69. Marcsisn, S. R.; Engen, J. R.; Hydrogen Exchange Mass Spectrometry: What is it and What Can it Tell Us? *Anal. BioAnal. Chem.* **2010**, *397*, 967 – 972.

70. Plaxco, K. W.; Simons, K. T.; Baker, D. Contact Order, Transition State Placement and the Refolding Rates of Single Domain Proteins. *J. Mol. Biol.* **1998**, *277*, 985 – 994.

71. Gromiha, M. M; Thangakani, A. M.; Selvaraj, S. FOLD-RATE: Prediction of Protein Folding Rates from Amino Acid Sequence. *Nucleic Acids Res.* **2006**, *34*, W70 – W74.

72. Weikl, T. R.; Palassini, M.; Dill, K. A. Cooperativity in Two-State Protein Folding Kinetics. *Protein Sci.* **2004**, *13*, 822 – 829.

73. Krishna, M. M. G.; Hoang, L.; Lin, Y.; Englander, S. W. Hydrogen Exchange Methods to Study Protein Folding. *Methods* **2004**, *34*, 51 – 64.

74. Englander, S. W.; Mayne, L. The Case for Defined Protein Folding Pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 8253 – 8258.

75. Sanctis, G. D.; Ascoli, F.; Brunori, M. Folding of Apominimyoglobin, *Proc. Natl. Sci. U. S. A.* **1994**, *91*, 11507 – 11511.

76. Baldwin, R. L. The Search for Folding Intermediates and the Mechanism of Protein Folding. *Annu. Rev. Biophys.* **2008**, *37*, 1 – 21.

77. Bai, Y.; Sosnick, T. R.; Mayne, L.; Englander, S. W. Protein Folding Intermediates: Native-State Hydrogen Exchange. *Science* **1995**, *269*, 192 – 197.

78. Englander, S. W.; Mayne, L.; Krishna, M. M. Protein Folding and Misfolding: Mechanism and Principles. *Q. Rev. Biophys*. **2007**, *40*, 287 – 326.

79. Haglund, E.; Lind, J.; Oman, T.; Ohman, A.; Mäler, L.; Oliveberg, M.; The HD-Exchange Motions of Ribosomal Protein S6 are Insensitive to Reversal of the Protein-Folding Pathway. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 21619 – 21624.

80. Englander, S. W.; Mayne, L. The Nature of Protein Folding Pathways. *Proc. Natl, Acad. Sci. U. S. A.* **2014**, *111*, 15873 – 15880.

81. Haglund, E.; Lindberg, M. O.; Oliveberg, M. Changes of Protein Folding Pathways by Circular Permutation. *J. Biol. Chem.* **2008**, *283*, 27904 – 27915.

82. Haglund, E.; Danielsson, J.; Kadhirvel, S.; Lindberg, M. O.; Logan, D. K.; Oliveberg, M. Trimming Down a Protein Structure to its Bare Foldons. *J. Biol. Chem.* **2012**, *287*, 2731 – 2738.

83. Ozkan, S. B.; Wu, G. A.; Chodera, J. D.; Dill, K. D. Protein Folding by Zipping and Assembly. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 11987 – 11992.

84. Shell, M. S.; Ozkan, S. B.; Voelz, V.; Wu, G. A.; Dill, K. A. Blind Test of Physics-Based Prediction of Protein Structures. *Biophys. J.* **2009**, *96*, 917 – 924.

85. Duhamel, J.; Yekta, A.; Winnik, M. A.; Jao, T. C.; Mishra, M. K.; Rubin, I. D. A Blob Model to Study Polymer Chain Dynamics in Solution. *J. Phys. Chem.* **1993**, *97*, 13708–13712.

86. Mathew, A.; Siu, H.; Duhamel, J. A blob Model to Study Chain Folding by Fluorescence. *Macromolecules* **1999**, *32*, 7100 − 7108.

87. Duhamel, J. Polymer Chain Dynamics in Solution Probed with a Fluorescence Blob Model. *Acc. Chem. Res.* **2006**, *39*, 953 – 960.

88. Duhamel, J. Global Analysis of Fluorescence Decays to Probe the Internal Dynamics of Fluorescently Labeled Macromolecules. *Langmuir* **2014**, *30*, 2307 – 2324.

89. Farhangi, S.; Duhamel, J. Long Range Polymer Chain Dynamics Studied by Fluorescence Quenching. *Macromolecules* **2016**, *49*, 6149-6162.

90. Li, L.; Duhamel, J. Conformation of Pyrene-Labeled Amylose in DMSO Characterized with the Fluorescence Blob Model. *Macromolecules* **2016**, *49*, 7965–7974.

91. Duhamel, J.; Kanagalingam, S.; O'Brien, T. J.; Ingratta, M. W. Side-Chain Dynamics of an α-Helical Polypeptide Monitored by Fluorescence. *J. Am. Chem. Soc.* **2003**, *125*, 12810–12822.

92. Ingratta, M.; Duhamel, J. Effect of Side-Chain Length on the Side-Chain Dynamics of α–Helical Poly(*L*-Glutamic Acid) as Probed by a Fluorescence Blob Model. *J. Phys. Chem. B* **2008**, *112*, 9209–9218.

Chapter 2

1. Levinthal, C. Are there Pathways for Protein Folding. *J. Chim. Phys.* **1968**, *65*, 44-45.
2. Levinthal, C. How to Fold Graciously. *Spectroscopy in Biological Systems. Proceedings University of Illinois Bulletin* 1969, University of Illinois Press, Urbana, IL, pp 22-24.
3. Pande, V. S.; Grosberg, A. Y.; Tanaka, T. Heteropolymer Freezing and Design: Towards Physical Model of Protein Folding. *Rev. Mod. Phys.* **2000**, *72*, 259-314.
4. Dill, K. A.; MacCallum, J. L. The Protein-Folding Problem, 50 years on. *Science* **2012**, *338*, 1042-1046.
5. Eaton, W. A.; Wolynes, P. G. Theory, Simulation, and Experiments Show that Proteins Fold by Multiple Pathways. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E9761-E9762.
6. Panchenko, A. R.; Luthey-Schulten, Z.; Cole, R. The Foldon Universe: A Survey of Structural Similarity and Self-Recognition of Independent Folding Units. *J. Mol. Biol.* **1997**, *272*, 95-105.
7. Bai, Y.; Sosnick, T. R.; Mayne, L.; Englander, W. E. Protein Folding Intermediates: Native-State Hydrogen Exchange. *Science* **1995**, *269*, 192-197.
8. Walters, B. T.; Mayne, L.; Hinshaw, J. R.; Sosnick, T. R.; Englander, S. W. Folding of Large Protein at High Structural Resolution. *Proc. Natl. Acad. Sci.* **2013**, *110*, 18898-18903.
9. Hingorani, K. S.; Gierasch, L. M. Comparing Protein Folding *in Vitro* and *in Vivo*: Foldability Meets the Fitness Challenge. *Curr. Opin. Struct. Biol.* **2014**, *24*, 81-90.

10. Hu, W.; Walters, B. T.; Kan, Z. Y.; Mayne, L.; Rosen, L. E.; Marguse, S.; Englander, S. W. Stepwise Protein Folding at Near Amino Acid Resolution by Hydrogen Exchange and Mass Spectrometry. *Proc. Natl. Acad. Sci.* **2013**, *110*, 7684-7689.
11. Naganathan, A. N.; Muñoz, V. Scaling of Folding Times with Protein Size. *J. Am. Chem. Soc.* **2005**, *127*, 480-481.
12. Englander, S. W.; Mayne, L. The Nature of Protein Folding Pathways. *Proc. Natl. Acad. Sci.* **2014**, *111*, 15873-15880.
13. Farhangi, S.; Duhamel, J. Long Range Polymer Chain Dynamics Studied by Fluorescence Quenching. *Macromolecules* **2016**, *49*, 6149-6162.
14. Matsumoto, M.; Watanabe, H.; Yoshioka, K. Electric and hydrodynamic properties of polypeptides in solution. II. Conformation of Poly(*L*-glutamic acid) in Various Organic Solvents. *Biopolymers* **1970**, *9*, 1307-1317.
15. Yamaoka, K.; Ueda, K. Reversing-Pulse Electric Birefringence of (Glu)n. Part 2. Reversing-Pulse Electric Birefringence Study of Helical Poly(alpha-L-glutamic acid) in *N*,*N*-Dimethylformamide with Emphasis on a New Data Analysis for Polydisperse System. *J. Phys. Chem.* **1982**, *86*, 406-413.
16. Duhamel, J.; Kanagalingam, S.; O'Brien, T.; Ingratta, M. Side-Chain Dynamics of an α–Helical Polypeptide Monitored by Fluorescence. *J. Am. Chem. Soc.* **2003**, *125*, 12810-12822.

17. Ingratta, M.; Duhamel, J. Effect of Side-Chain Length on the Side-Chain Dynamics of α–Helical Poly(L-glutamic acid) as Probed by a Fluorescence Blob Model. *J. Phys. Chem. B* **2008**, *112*, 9209-9218.
18. Lakowicz, J. R. Principles of Fluorescence Spectroscopy, 3rd Ed. Springer: Singapore, 2006, p 278.
19. Mathew, A.; Siu, H.; Duhamel, J. A *Blob* Model to Study Chain Folding by Fluorescence. *Macromolecules* **1999**, *32*, 7100-7108.
20. Duhamel, J. New Insights in the Study of Pyrene Excimer Fluorescence to Characterize Macromolecules and their Supramolecular Assemblies in Solution. *Langmuir* **2012**, *28*, 6527-6538
21. Shoji, O.; Okumura, M.; Kuwata, H.; Sumida. T.; Kato, R.; Annaka, M.; Yoshikuni, M.; Nakahira, T. Secondary Structure and Side Chain Chromophore Orientation in Poly(L-glutamines) Having Pyrene Chromophores in the Side Chains. *Macromolecules* **2001**, *34*, 4270–4276.
22. Hall, T. Study of Arborescent Poly(L-glutamic acid) by Pyrene Excimer Formation. Unpublished, MSc. Thesis, University of Waterloo, Waterloo, ON, Canada, 2012.
23. Winnik, F. M. (1993) Photophysics of Preassociated Pyrenes in Aqueous Polymer Solutions and in Other Organized Media. *Chem. Rev. 93*, 587–614.

24. Jacob, M. H.; D'Souza, R. N.; Schwarzlose, T.; Wang, X.; Huang, F.; Haas, E.; Nau, W. M. Method-Unifying View of Loop-Formation Kinetics in Peptide and Protein Folding. *ASAP J. Phys. Chem. B.* **2018**.
25. Fierz, B.; Kiefhaber, T. End-to-End vs Interior Loop Formation Kinetics in Unfolded Polypeptide Chains. *J. Am. Chem. Soc.* **2007**, *129*, 672-679.
26. Ingratta, M.; Hollinger, J.; Duhamel, J. A Case for Using Randomly Labeled Polymers to Study Long Range Polymer Chain Dynamics by Fluorescence. *J. Am. Chem. Soc.* **2008,** *130*, 9420-9428.
27. Yip, J.; Duhamel, J.; Qiu, X. P.; Winnik, F. M. Long-Range Polymer Chain Dynamics of Pyrene-Labelled Poly(*N*-isopropylacrylamide)s Studied by Fluorescence. *Macromolecules* **2011**, *44*, 5363-5372.
28. Probing Side Chain Dynamics of Branched Macromolecules by Pyrene Excimer Fluorescence. Farhangi, S.; Duhamel, J. *Macromolecules* **2016**, *49*, 353-361.

Chapter 3

1. Zhou, H.-X.; Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem. Rev.* **2018**, *118*, 1691 – 1741.
2. Højgaard, C.; Kofoed, C.; Espersen, R.; Johansson, K. E.; Villa, M.; Willemoës, M.; Lindorff-Larsen, K.; Teilum, K.; Winther, J. R. A Soluble, Folded Protein without Charged Amino Acid Residues. *Biochemistry* **2016**, *55*, 3949 – 3956.

3. Feng, L.; Campbell, E. B.; Hsiung, Y.; MacKinnon, R. Structure of a Eukaryotic CLC Transporter Defines an Intermediate State in the Transport Cycle. *Science* **2010**, *330*, 635 – 641.
4. Maniccia, A. W.; Yang, W.; Johnson, J. A.; Li, S.; Tjong, H.; Zhou, H. X.; Shaket, L. A.; Yang, J. J. Inverse Tuning of Metal Binding Affinity and Protein Stability by Altering Charged Coordination Residues in Designed Calcium Binding Proteins. *PMC Biophys.* **2009**, *2*, 11.
5. Kalodimos, C. G.; Biris, N.; Bonvin, A. M. J. J.; Levandoski, M. M.; Guennuegues, M.; Boelens, R.; Kaptein, R. Structure and Flexibility Adaptation in Nonspecific and Specific Protein-DNA Complexes. *Science* **2004**, *5682*, 386 – 389.
6. Sheinerman, F. B.; Norel, R.; Honig, B. Electrostatic Aspects of Protein-Protein Interactions. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153 – 159.
7. Kumar, S.; Nussinov, R. Close-Range Electrostatic Interactions in Proteins. *ChemBioChem.* **2002**, *3*, 604 – 617.
8. Tanford, C. Protein Denaturation. C. Theoretical Models for the Mechanism of Denaturation. *Adv. Protein Chem.* **1970**, *24*, 1 – 95.
9. Kurnik, M.; Hedberg, L.; Danielsson, J.; Oliveberg, M. Folding Without Charges. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 5705 – 5710.
10. Loladze, V. V.; Makhatadze, G. I. Removal of Surface Charge-Charge Interactions From Ubiquitin Leaves the Protein Folded and Very Stable. *Protein Sci.* **2002**, *11*, 174 – 177.

11. Tsai, M.-Y. Zheng, W.; Balamurugan, D.; Schafer, N. P.; Kim, B. L.; Cheung, M. S.; Wolynes, P. G. Electrostatics, Structure Prediction, and the Energy Landscapes for Protein Folding and Binding. *Protein Sci.* **2016**, *25*, 255 – 269.
12. Englander, S. W.; Mayne, L. The Nature of Protein Folding Pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 15873 – 15880.
13. Englander, W.; Mayne, L. The Case for Defined Protein Folding Pathways. *Proc. Natl. Acad. Sci.* **2017**, *114*, 8253 – 8258.
14. Baldwin R. L.; Rose, G. D. Is Protein Folding Hierarchic? I. Local Structure and Peptide Folding. *Trends Biochem. Sci.* **1999**, *24*, 26 – 33.
15. Irondi, K.; Zhang, M.; Duhamel, J. Study of the Semidilute Solutions of Poly(*N*,*N*-dimethylacrylamide) by Fluorescence and its Implications to the Kinetics of Coil-to-Globule Transitions. *J. Phys. Chem. B* **2006**, *110*, 2628-2637.
16. Farhangi, S.; Duhamel, J. Long Range Polymer Chain Dynamics Studied by Fluorescence Quenching. *Macromolecules* **2016**, *49*, 6149-6162.
17. Mathew, A.; Siu, H.; Duhamel, J. A *Blob* Model to Study Chain Folding by Fluorescence. *Macromolecules* **1999**, *32*, 7100-7108.
18. Li, L.; Duhamel, J. Conformation of Pyrene-Labeled Amylose in DMSO Characterized with the Fluorescence Blob Model. *Macromolecules* **2016**, *49*, 7965 – 7974.

19. Kanagalingam, S.; Spartalis, J.; Cao, T.-M.; Duhamel, J. Scaling Relations Related to the Kinetics of Excimer Formation between Pyrene Groups Attached onto Poly(N,N-dimethylacrylamide)s. *Macromolecules* **2002**, *35*, 8571 – 8577.
20. Farhangi, S.; Weiss, H.; Duhamel, J. Effect of Side-Chain Length on the Polymer Chain Dynamics of Poly(alkyl methacrylate)s in Solution. *Macromolecules* **2013**, *46*, 9738 – 9747.
21. Duhamel, J. New Insights in the Study of Pyrene Excimer Fluorescence to Characterize Macromolecules and their Supramolecular Assemblies in Solution. *Langmuir* **2012**, *28*, 6527-6538.
22. Winnik, F. M. Photophysics of Preassociated Pyrenes in Aqueous Polymer Solutions and in Other Organized Media. *Chem. Rev.* **1993**, *93*, 587–614.
23. Casier, R.; Duhamel, J. Pyrene Excimer Fluorescence as a Direct and Easy Experimental Means to Characterize the Length Scale and Internal Dynamics of Polypeptide Foldons. *Macromolecules* **2018**, *51*, 3450 – 3457.
24. Konermann, L.; Pan, J.; Liu, H.-Y. Hydrogen Exchange Mass Spectrometry for Studying Protein Structure and Dynamics. *Chem. Soc. Rev.* **2011**, *40*, 1224-1234.
25. Bai, Y.; Sosnick, T. R.; Mayne, L.; Englander, S. W. Protein Folding Intermediates: Native-State Hydrogen Exchange. *Science* **1995**, *269*, 192-197.

26. Thoma, J. L.; Duhamel, J.; Bertocchi, M. J.; Weiss, R. G. Long Range Polymer Chain Dynamics of Highly Flexible Polysiloxane in Solution Probed by Pyrene Excimer Fluorescence. *Polymers* **2018**, *10*, 345.
27. Duhamel, J. Global Analysis of Fluorescence Decays to Probe the Internal Dynamics of Fluorescently Labeled Macromolecules. *Langmuir* **2014**, *30*, 2307 – 2324.
28. Duhamel, J. Polymer Chain Dynamics in Solution Probed with a Fluorescence Blob Model. *Acc. Chem. Res.* **2006**, *39*, 953 – 960.
29. Press, W. H.; Flanery, B. P.; Tenkolsky, S. A.; Vetterling, W. T. Numerical Receipes in Fortran: The Art of Scientific Computing; Cambridge University Press: Cambridge and New York, 1992; pp 523−528.
30. Zhao, D.; Li, Q.; Duan, E.; Li, H.; Shen, X. Solubility of l-Lysine Hydrochloride in Dimethyl Sulfoxide, Methanol, Ethanol, Water, and Glycol between (283 and 323) K. *J. Chem. Eng. Data* **2009**, *54*, 2126 – 2127.
31. Imahori, K.; Tanaka, J. Ultraviolet Absorption Spectra of Poly(*L*-glutamic acid). *J. Mol. Biol.* **1959**, *1*, 359-364.
32. Bychkova, V. E.; Ptitsyn, O. B. ; Barskaya, T. V. Thermodynamic Parameters of Helix-Coil Transition in Polypeptide Chains I. Poly(L-glutamic acid). *Biopolymers* **1971**, *10*, 2161-2179.

33. Nitta, K.; Yoneyama, M.; Ohno, N. Polymer Concentration Dependence of the Helix to Random Coil Transition of a Charged Polypeptide in Aqueous Salt Solution. *Biophys. Chem.* **1975**, *3*, 323-329.
34. Farhangi, S.; Duhamel, J. Probing Side Chain Dynamics of Branched Macromolecules by Pyrene Excimer Fluorescence. *Macromolecules* **2016**, *49*, 353 – 361.
35. Farhangi, S. Duhamel, J. Pyrenyl Derivative with a Four-Atom Linker That Can Probe the Local Polarity of Pyrene-Labeled Macromolecules. *J. Phys. Chem. B* **2016**, *120*, 834 – 842.
36. Lakowicz, J. R. Principles of Fluorescence Spectroscopy, 3rd Ed. Springer: Singapore, 2006, p 281.
37. Myer, Y. P. The pH-Induced Helix-Coil Transition of Poly-L-lysine and Poly-L-glutamic Acid and the 238-mμ Dichroic Band. *Macromolecules* **1969**, 2, 624 – 628.
38. Jackson, M. Mantsch, H. H. Beware of Proteins in DMSO. *Biochim. Biophys. Acta.* **1991**, *1078*, 231 – 235.
39. Kotik, M.; Radford, S. E.; Dobson, C. M. Comparison of the Refolding of Hen Lysozyme from Dimethyl Sulfoxide and Guanadinium Chloride. *Biochemistry* **1995**, *34*, 1714-1724.
40. Iwase, H.; Hirai, M.; Arai, S.; Mitsuya, S.; Shimizu, S.; Otomo, T.; Furusaka, M. Comparison of DMSO-Induced Denaturation of Hen Egg-White Lysozyme and Bovine α-Lactalbulmin. *J. Phys. Chem. Solids* **1999**, *60*, 1379-1381.

41. Arakawa, T.; Kita, Y.; Timasheff, S. N. Protein Precipitation and Denaturation by Dimethyl Sulfoxide. *Biophys. Chem.* **2007**, *131*, 62 – 70.

42. Chakraborty, S.; Hosur, R. V. NMR Insight into the Core of GED Assembly by H/D Exchange Coupled with DMSO Dissociation and Analysis of the Denatured State. *J. Mol. Biol.* **2011**, *405*, 1202-1214.

43. Berbeć, S.; Dec, R.; Molodenskiy, D.; Wielgus-Kutrowska, B; Johannessen, C.; Hernik-Magoń, A.; Tobias, F.; Bzowska, A; Ścibisz, G.; Keiderling, T. A.; Svergun, D.; Dzwolak, W. β2-Type Amyloidlike Fibrils of Poly-l-glutamic Acid Convert into Long, Highly Ordered Helices upon Dissolution in Dimethyl Sulfoxide. *J. Phys. Chem. B* **2018**, *122*, 11895 – 11905.

44. Ogasawara, N.; Kasahara, K.; Iwai, R.; Takahashi T. Unfolding of α-Helical 20-Residue Poly-Glutamic Acid Analyzed by Multiple Runs of Canonical Molecular Dynamics Simulations. *PeerJ* **2018**, DOI: 10.7717/peerj.4769

45. Bellanda, M.; Peggion, E.; Mammi, S.; Bürgi, R.; vanGunsteren, W. Conformational Study of an Aib-Rich Peptide in DMSO by NMR. *J. Pept. Res.* **2001**, *57*, 97 − 106.

46. Birks, J. B.; Dyson, D. J.; Munro, I. H. 'Excimer' Fluorescence. II. Lifetime Studies of Pyrene Solutions *Proc. R. Soc. A* **1963**, *275*, 575 – 588.

47. Mirtic, A.; Grdadolnik, J. The Structure of Poly(L-lysine) in Different Solvents. *Biophys. Chem.* **2013**, *175-176*, 47-53.

48. *CRC Handbook of Chemistry and Physics* ,77th Ed.; Boca Raton, FL, 1996.

49. Zheng, Y.-J.; Ornstein, R. L. A Molecular Dynamics and Quantum Mechanics Analysis of the Effect of DMSO on Enzyme Structure and Dynamics: Subtilisin. *J. Am. Chem. Soc.* **1996**, *118*, 4175-4180.

50. Norouzy, A.; Assaf, K. I.; Zhang, S.; Jacob, M. H.; Nau, W. M. Coulomb Repulsion in Short Polypeptides. *J. Chem. Phys. B* **2015**, *119*, 33-43.

51. Chou, P. Y.; Scheraga, H. A. Calorimetric Measurements of Enthalpy Change in the Isothermal Helix-Coil Transition of Poly(*L*-lysine) in Aqueous Solution. *Biopolymers* **1971**, *10*, 657-680.

52. Ingratta, M.; Duhamel, J. Effect of Side-chain Length on the Side-chain Dynamics of α-Helical Poly(L-glutamic acid) as Probed by a Fluorescence Blob Model. *J. Phys. Chem. B* **2008**, *112*, 9209-9218.

Chapter 4

1. Dill, K. A.; Chan, H. S. From Levinthal to Pathways to Funnels. *Nat. Struct. Mol. Biol.* **1997**, *4*, 10 – 19.

2. Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. Navigating the Folding Routes. *Science* **1995**, *267*, 1619 – 1620.

3. Petsko, G. A.; Ringe, Dagmar. *Protein Structure and Function*, New Science Press Ltd: London, 2004; pp 50.

4. Dobson, C. M. Protein Folding and Misfolding. *Nature* **2003**, *426*, 884 – 890.

5. Englander, S. W.; Mayne, L. The Nature of Protein Folding Pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 15873 – 15880.

6. Baldwin R. L.; Rose, G. D. Is Protein Folding Hierarchic? I. Local Structure and Peptide Folding. *Trends Biochem. Sci.* **1999**, *24*, 26 – 33.

7. Farhangi, S.; Duhamel, J. Long Range Polymer Chain Dynamics Studied by Fluorescence Quenching. *Macromolecules* **2016**, *49*, 6149-6162.

8. Casier, R.; Duhamel, J. Pyrene Excimer Fluorescence as a Direct and Easy Experimental Means to Characterize the Length Scale and Internal Dynamics of Polypeptide Foldons. *Macromolecules* **2018**, *51*, 3450 – 3457.

9. Englander, W.; Mayne, L. The Case for Defined Protein Folding Pathways. *Proc. Natl. Acad. Sci.* **2017**, *114*, 8253 – 8258.

10. Maity, H.; Maity, M.; Krishna, M. M. G.; Mayne, L.; Englander, S. W. Protein Folding: The Stepwise Assembly of Foldon Units. *Proc. Natl. Acad. Sci.* **2005**, *102*, 4741 – 4746.

11. Irondi, K.; Zhang, M.; Duhamel, J. Study of the Semidilute Solutions of Poly(*N*,*N*-dimethylacrylamide) by Fluorescence and its Implications to the Kinetics of Coil-to-Globule Transitions. *J. Phys. Chem. B* **2006**, *110*, 2628-2637.

12. Walters, B.T.; Mayne, L.; Hinshaw, J. R.; Sosnick, T. R.; Englander, S. W. Folding of a large protein at high structural resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 18898 – 18903.

13. Hu, W.; Walters, B. T.; Kan, Z. Y., Mayne, L.; Rosen, L. E.; Marqusee, S.; Englander, S. W. Stepwise Protein Folding at Near Amino Acid Resolution by Hydrogen Exchange and Mass Spectometry. *Proc. Natl. Acad. Sci.* **2013**, *110*, 7684 – 7689.

14. Bai, Y.; Sosnick, T. R.; Mayne, L.; Englander, S. W. Protein folding intermediates: Native-state hydrogen exchange. *Science* **1995**, *269*, 192 – 197.

15. Gelman, H.; Gruebele, M. Fast Protein Folding Kinetics. *Q. Rev. Biophys.* **2014**, *47*, 95-142.

16. Sturzenegger, F.; Zosel, F.; Holmstrom, E. D.; Buholzer, K. J.; Makarov, D. E.; Nettels, D.; Schuler, B. Transition Path Times of Coupled Folding and Binding Reveal the Formation of an Encounter Complex. *Nature Comm.* **2018**, *9*, 1-11.

17. Jeon, J.; Thurber, K. R.; Ghirlando, R.; Yau, W.-M.; Tycko, R. Application of Millisecond Time-Resolved Solid State NMR to the Kinetics and Mechanism of Melitin Self-Assembly. *Proc. Natl. Acad. Sci.* **2019**, *116*, 16717-16722.

18. Ivankov, D. N.; Finkelstein, A. V. Solution of Levinthal's Paradox and a Physical Theory of Protein Folding Times. *Biomolecules* **2020**, *10*, 250.

19. Farhangi, S.; Weiss, H.; Duhamel, J. Effect of Side-Chain Length on the Polymer Chain Dynamics of Poly(alkyl methacrylate)s in Solution. *Macromolecules* **2013**, *46*, 9738 – 9747.

20. Duhamel, J. New Insights in the Study of Pyrene Excimer Fluorescence to Characterize Macromolecules and their Supramolecular Assemblies in Solution. *Langmuir* **2012**, *28*, 6527-6538.

21. Duhamel, J. Polymer Chain Dynamics in Solution Probed with a Fluorescence Blob Model. *Acc. Chem. Res.* **2006**, *39*, 953 – 960.

22. Förster, Th.; Kasper, K. Ein Konzentrationsumchlag der Fluoreszenz des Pyrens. *Z. Elektrochem.* **1955**, *59*, 976 – 980.

23. Birks, J. B.; Dyson, D. J.; Munro, I. H. 'Excimer' Fluorescence. II. Lifetime Studies of Pyrene Solutions *Proc. R. Soc. A* **1963**, *275*, 575 – 588.

24. Li, L.; Duhamel, J. Conformation of Pyrene-Labeled Amylose in DMSO Characterized with the Fluorescence Blob Model. *Macromolecules* **2016**, *49*, 7965 – 7974.

25. Ingratta, M.; Duhamel, J. Effect of Side-Chain Length on the Side-Chain Dynamics of α-Helical Poly(L-glutamic acid) as Probed by a Fluorescence Blob Model *J. Phys. Chem. B* **2008**, *112*, 9209– 9218

26. Duhamel, J.; Kanagalingam, S.; O'Brien, T.; Ingratta, M. Side-Chain Dynamics of an α-Helical Polypeptide Monitored by Fluorescence *J. Am. Chem. Soc.* **2003**, *125*, 12810–12822.

27. Thoma, J. L.; Duhamel, J.; Li, M.-J.; Bertocchi, M. J.; Weiss, R. G. Long-Range, Polymer Chain Dynamics of a "Stiff" Polymer. Fluorescence from Poly (isobutylene-alt-

maleic anhydride) with N-(1-Pyrenylmethyl) succinimide Groups. *Macromolecules* **2017**, *50*, 3396 – 3403.

28. Hall, T.; Whitton, G.; Casier, R.; Gauthier, M.; Duhamel, J. Arborescent Poly(l-glutamic acid)s as Standards To Study the Dense Interior of Polypeptide Mesoglobules by Pyrene Excimer Fluorescence. *Macromolecules* **2018**, *51*, 7914-7923

29. Casier, R. Duhamel, J. The Effect of Like-Charges on the Conformation and Internal Dynamics of Polypeptides Probed by Pyrene Excimer Fluorescence. *Macromolecules* just accepted **2020**.

30. Yasui, S. C.; Keiderling, T. A. Vibrational Circular Dichroism of Polypeptides. 8. Poly(lysine) Conformations as a Function of pH in Aqueous Solution. *J. Am. Chem. Soc.* **1986**, *108*, 5576 – 5581.

31. Paterlini, M. G.; Freedman, T. B.; Nafie, L. A. Vibrational Circular Dichroism Spectra of Three Conformationally Distinct States and an Unordered State of Poly(L-lysine) in Deuterated Aqueous Solution. *Biopolymers* **1986**, *25*, 1751 – 1765.

32. Arunkumar, A. I.; Kumar, T. K. S.; Sivaraman, T.; Yu, C. Acetonitrile-Induced Conformational Transitions in Poly-L-Lysine. *Int. J. Biol. Macromol.* **1997**, *21*, 299 – 305.

33. Mirtič, A.; Grdadolnik, J. The Structure of Poly-L-Lysine in different Solvents. *Biophys. Chem.* **2013**, *175 – 176*, 47 – 53.

34. Cieślik-Boczula, K. Alpha-Helix to Beta-Sheet Transition in Long-Chain Poly-L-Lysine: Formation of Alpha-Helical Fibrils by Poly-L-Lysine. *Biochimie* **2017**, *137*, 106 – 114.

35. Dzwolak, W.; Ravindra, R.; Nicolini, C.; Jansen, R.; Winter, R. The Diastereomeric Assembly of Polylysine Is the Low-Volume Pathway for Preferential Formation of β-Sheet Aggregates. *J. Am. Chem. Soc.* **2004**, *126*, 3762 – 3768.

36. Selkoe, D. J. The Molecular Pathology of Alzheimer's Disease. *Neuron* **1991**, *6*, 487 – 498.

37. Laurine, E.; Grégoire, C.; Fändrich, M.; Engemann, S.; Marchal, S.; Thion, L.; Mohr, M.; Monsarrat, B.; Michel, B.; Dobson, C. M.; Wanker, E.; Erard, M.; Verdier, J. M. Lithostathine Quadruple-Helical Filaments Form Proteinase K-Resistant Deposits in Creutzfeldt-Jakob Disease. *J. Biol. Chem.* **2003**, *278*, 51770 – 51778.

38. Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. Membrane Proteins. *Molecular Biology of the Cell*, 4th edition. New York: Garland Science; 2002.

39. Bockaert, J.; Pin, J. P.; Molecular Tinkering of G Protein-Coupled Receptors: An Evolutionary Success. *EMBO J* **1999**, *18*, 1723 – 1729.

40. Thoma, J. L.; Duhamel, J.; Bertocchi, M. J.; Weiss, R. G. Long Range Polymer Chain Dynamics of Highly Flexible Polysiloxane in Solution Probed by Pyrene Excimer Fluorescence. *Polymers* **2018**, *10*, 345.

41. Duhamel, J. Global Analysis of Fluorescence Decays to Probe the Internal Dynamics of Fluorescently Labeled Macromolecules. *Langmuir* **2014**, *30*, 2307 – 2324.

42. Winnik, F. M. Photophysics of Preassociated Pyrenes in Aqueous Polymer Solutions and in Other Organized Media. *Chem. Rev.* **1993**, *93*, 587–614.

43. Winnik, F. M.; Regismond, S. T. A. Fluorescence Methods in the Study of the Interactions of Surfactants with Polymers. *Colloids Surf. A.* **1996**, *118*, 1 – 39.

44. Zhang, M.; Duhamel, J. Study of the Microcrystallization of Ethylene−Propylene Random Copolymers in Solution by Fluorescence. *Macromolecules* **2007**, *40*, 661 – 669.

45. Zhang, Q.; Kim, D.; Li, L.; Patel, S.; Duhamel, J. Surfactant Structure-Dependent Interactions with Modified Starch Nanoparticles Probed by Fluorescence Spectroscopy. *Langmuir* **2019**, *35*, 3432 – 3444.

46. Berlman, I. B. *Energy Transfer Parameters of Aromatic Compounds*; Academic Press: New York, 1973; pp 309.

47. Farhangi, S.; Duhamel, J. Probing Side Chain Dynamics of Branched Macromolecules by Pyrene Excimer Fluorescence. *Macromolecules* **2016**, *49*, 353 – 361.

48. Thompson, J. W.; Kaiser, T. J.; Jorgenson, J. W. Viscosity Measurements of Methanol–Water and Acetonitrile−Water Mixtures at Pressures up to 3500 Bar Using a Novel Capillary Time-of-Flight Viscometer. *J. Chromatogr. A.* **2006**, *1134*, 201 – 209.

49. Lakowicz, J. R. Principles of Fluorescence Spectroscopy, 3rd Ed. Springer: Singapore, 2006, p 281.

50. Harding, S. E. The Intrinsic Viscosity of Biological Macromolecules. Progress in Measurements, Interpretation, and Application to Structure in Dilute Solution. *Prog. Biophys. Molec. Biol.* **1997**, *68*, 207-262.
51. Zaccai, N. R.; Serdyuk, I. N. ; Zaccai, J. Methods in Molecular Biophysics, 2nd Ed., Cambridge University Press, Cambridge, UK, 2017, p 199.
52. Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic´ , Z. Intrinsic Disorder and Protein Function. *Biochemistry* **2002**, *41*, 6573 − 6582.
53. Dyson, H. J.; Wright, P. E. Coupling of Folding and Binding for Unstructured Proteins. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54 − 60.
54. Dyson, H. J.; Wright, P. E. Intrinsically Unstructured Proteins Their Functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197 − 208.

Chapter 5

1. Petsko, G. A.; Ringe, Dagmar. *Protein Structure and Function*, New Science Press Ltd: London, 2004; pp 50.
2. Dobson, C. M. Protein Folding and Misfolding. *Nature* **2003**, *426*, 884 – 890.
3. Ivankov, D. N.; Finkelstein, A. V. Solution of Levinthal's Paradox and a Physical Theory of Protein Folding Times. *Biomolecules* **2020**, *10*, 250.

4. Doose, S.; Neuweiller, H.; Sauer, M. Fluorescence Quenching by Photoinduced Electron Transfer: A Reporter for Conformational Dynamics of Macromolecules. *ChemPhysChem* **2009**, *10*, 1389 – 1398.

5. Ries, J.; Schwarze, S.; Johnson, C. M.; Neuweiler, H. Microsecond Folding and Domain Motions of a Spider Silk Protein Structural Switch. *J. Am. Chem. Soc.* **2014**, *136*, 17136 – 14144.

6. Nenov, A.; A Becarra, S.; Rivalta, I.; Cerullo, G.; Mukamel, S.; Garavelli, M. Tracking Conformational Dynamics of Polypeptides by Nonlinear Electronic Spectroscopy of Aromatic Residues: A First-Principles Simulation Study. *ChemPhysChem* **2014**, *15*, 3282 – 3290.

7. Petersson, E. J.; Goldberg, J. M.; Wissner, R. F. On the Use of Thioamides as Fluorescence Quenching Probes for Tracking Protein Folding and Stability. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6827 – 6837.

8. Holtkamp, W.; Kokic, G.; Jäger, M.; Mittelstaet, J.; Komar, A. A.; Rodnina, M. V. Cotranslational Protein Folding on the Ribosome Monitored in Real Time. *Science* **2015**, *350*, 1104 – 1107.

9. Dunker, A. K.; Obradovic, Z.; Romero, P.; Garner, E. C. Intrinsic Protein Disorder in Complete Genomes. *Genome Inform.* **2000**, *11*, 161 – 171.

10. Obradovic, Z.; Peng, K.; Vucetic, S.; Radivojac, P.; Brown, C. J.; Dunker, A. K. Predicting Intrinsic Disorder from Amino Acid Sequence. *Proteins* **2003**, *53*, 566 – 572.

11. DeForte, S.; Uversky, V. N. Resolving the Ambiguity: Making Sense of Intrinsic Disorder when PDB Structures Disagree. *Protein Sci.* **2016**, *25*, 676 – 688.
12. Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradović, Z. Intrinsic Disorder and Protein Function. *Biochemistry* **2002**, *41*, 6573 – 6582.
13. Dyson, H. J.; Wright, P. E. Coupling of Folding and Binding for Unstructured Proteins. *Curr. Opin. Struc.* Biol. **2002**, *12*, 54 – 60.
14. Dyson, H. J.; Wright, P. E. Intrinsically Unstructured Proteins and Their Functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197 – 208.
15. Berlow, R. B.; Dyson, H. J.; Wright, P. E. Functional Advantages of Dynamic Protein Disorder. *FEBS Lett.* **2015**, *589*, 2433 – 2440.
16. Uversky, V. N. Natively Unfolded Proteins: A Point Where Biology Waits for Physics. *Protein Sci.* **2002**, *11*, 739 – 756.
17. Ganguly, D.; Zhang, W.; Chen, J. Synergistic Folding of Two Intrinsically Disordered Proteins: Searching for Conformational Selection. *Mol. BioSyst.* **2012**, *8*, 198 – 209.
18. Demarest, S. J.; Martinez-Yamout, M.; Chung, J.; Chen, H.; Xu, W.; Dyson, H. J.; Evans, R. M.; Wright, P. E. Mutual Synergistic Folding in Recruitment of CBP/p300 by p160 Nuclear Receptor Coactivators. *Nature* **2002**, *415*, 549 – 553.
19. Schuler, B.; Borgia, A.; Borgia, M. B.; Heidarsson, P. O.; Holmstorm, E. D.; Nettels, D.; Sottini, A. Binding Without Folding – The Bimolecular Function of Disordered Polyelectrolyte Complexes. *Curr. Opin. Struc. Biol.* **2020**, *60*, 66 – 76.

20. Cato, L.; Stott, K.; Watson, M.; Thomas, J. O. The Interaction of HMGB1 and Linker Histones Occurs Through their Acidic and Basic Tails. *J. Mol. Biol.* **2008**, *384*, 1264 – 1272.

21. Sigalov, A.; Aivazian, D.; Stern, L. Homooligomerization of the Cytoplasmic Domain of the T Cell Receptor ζ Chain and of Other Proteins Containing the Immunoreceptor Tyrosine Based Activation Motif. *Biochemistry* **2004**, *43*, 2049 – 2061.

22. Wright, P. E.; Dyson, H. J. Intrinsically Unstructured Proteins: Re-Assessing the Protein Structure-Function Paradigm. *J. Mol. Biol.* **1999**, *293*, 321 – 331.

23. Kohem, A. Role of Dynamics in Enzyme Catalysis: Substantial versus Semantic Controversies. *Acc. Chem. Res.* **2015**, *48*, 466 – 473.

24. Frauenfelder, H,; Chen, G.; Berendzen, J.; Fenimore, P. W.; Jansson, H.; McMahon, B. H.; Stroe, I. R.; Swenson, J.; Young, R. A Unified Model of Protein Dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 5129 – 5134.

25. Eisenmesser, E. Z.; Millet, O.; Labeikovsky, W.; Korzhnev, D. M.; Wolf-Watz, M.; bosco, D. A.; Skalicky, J. J.; Kay, L. E.; Kern, D. Intrinsic Dynamics of an Enzyme Underlies Catalysis. *Nature* **2005**, *438*, 117 – 121.

26. Agarwal, P. L. Role of Protein Dynamics in Reaction Rate Enhancement by Enzymes. *J. Am. Chem. Soc.* **2005**, *127*, 15248 – 15256.

27. Uversky, V. N. Intrinsically Disordered Proteins and Their "Mysterious" (Meta)Physics. *Front. Phys.* **2019**, *7*, doi: 10.3389/fphy.2019.00010.

28. Yang, J.; Zng, Y.; Liu, Y.; Gao, M.; Liu, S.; Su, Y.; Huang, Y. Electrostatic Interactions in Molecular Recognition of Intrinsically Disordered Proteins. *J. Biomol. Sruct. Dyn.* **2019**, doi: 10.1080/07391102.2019.1692073.

29. Ivankov, D. N.; Finkelstein, A. V. Prediction of Protein Folding Rates From The Amino Acid Sequence-Predicted Secondary Structure. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 8942 – 8944.

30. Schwalbe, H. Fiebig, K. M.; Jones, J. A.; Grimshaw, S. B.; Spencer, A. Glaser, S. J.; Smith, L. J.; Dobson, C. M. Structural and Dynamicl Properties of a Denatured Protein. Heteronuclear 3D NMR Experiments and Theoretical Simulations of Lysozyme in 8 M Urea. *Biochemistry* **1997**, *36*, 8977 – 8991.

31. Muñoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. Folding Dynamics and Mechanism of β-Hairpin Formation. *Nature* **1997**, *390*, 196 – 199.

32. Jones, K.; Wittung-Stafshede, P. The Largest Protein Observed To Fold by Two-State Kinetic Mechanism Does Not Obey Contact-Order Correlation. *J. Am. Chem. Soc.* **2003**, *125*, 9606 – 9607.

33. Grantcharova, V. P.; Baker, D. Folding Dynamics of the src SH3 Domain. *Biochemistry* **1997**, *36*, 15685 – 15692.

34. García-R. EPR of Site-Directed Spin-Labeled Proteins: A Powerful Tool to Study Structural Flexibility. *Arch. Biochem. Biophys.* **2020**, *684*, 108323.

35. Haas, E.; Katchalski-Katzir, E.; Steinberg, I. Brownian Motion of the Ends of Oligopeptide Chains in Solution as Estimated by Energy Transfer Between the Chain Ends. *Biopolymers* **1978**, *17*, 11 – 31.

36. Hudgins, R.R.; Huang, F.; Gramlich, G.; Nau, W.M. A Fluorescence-Based Method for Direct Measurement of Submicrosecond Intramolecular Contact Formation in Biopolymers: An Exploratory Study with Polypeptides. *J. Am. Chem. Soc.* **2002**, *124*, 556–564.

37. Norouzy, A.; Assaf, K. I.; Zhang, S.; Jacob, M. H.; Nau, W. M. Coulomb Repulsion in Short Polypeptides. *J. Chem. Phys. B* **2015**, *119*, 33 – 43.

38. Krieger, F.; Fierz, B.; Bieri, O.; Drewello, M.; Kiefhaber, T. Dynamics of Unfolded Polypeptide Chains as Model for the Earliest Steps in Protein Folding. *J. Mol. Biol.* **2003**, *332*, 265 – 274.

39. Krieger, F.; Möglich, A.; Kiefhabler, T. Effect of Proline and Glycine Residues on Dynamics and Barriers of Loop Formation in Polypeptide Chains. *J. Am. Chem. Soc*. **2005**, *127*, 3346 – 3352.

40. Fierz, B.; Kiefhaber, T. End-to-End vs Interior Loop Formation Kinetics in Unfolded Polypeptide Chains. *J. Am. Chem. Soc.* **2007**, *129*, 672 – 679.

41. Shahabi, M.; Hajihosseini, R.; Nau, W. M.; Noghabi, K. A.; Norouzy, A. Augmenting Peptide Flexibility by Inserting Gamma-Aminobutyric Acid (GABA) in Their Sequence. *Int. J. Pept. Res. Ther.* **2020**, https://doi.org/10.1007/s10989-020-10054-2.

42. Farhangi, S.; Duhamel, J. Long Range Polymer Chain Dynamics Studied by Fluorescence Quenching. *Macromolecules* **2016**, *49*, 6149 – 6162.

43. Mathew, A.; Siu, H.; Duhamel, J. A *Blob* Model to Study Chain Folding by Fluorescence. *Macromolecules* **1999**, *32*, 7100 – 7108.

44. Duhamel, J. Polymer Chain Dynamics in Solution Probed with a Fluorescence Blob Model. *Acc. Chem. Res.* **2006**, *39*, 953 – 960.

45. Duhamel, J. New Insights in the Study of Pyrene Excimer Fluorescence to Characterize Macromolecules and their Supramolecular Assemblies in Solution. *Langmuir* **2012**, *28*, 6527 – 6538.

46. Duhamel, J. Global Analysis of Fluorescence Decays to Probe the Internal Dynamics of Fluorescently Labeled Macromolecules. *Langmuir* **2014**, *30*, 2307 – 2324.

47. Li, L.; Duhamel, J. Conformation of Pyrene-Labeled Amylose in DMSO Characterized with the Fluorescence Blob Model. *Macromolecules* **2016**, *49*, 7965 – 7974.

48. Casier, R.; Duhamel, J. Pyrene Excimer Fluorescence as a Direct and Easy Experimental Means to Characterize the Length Scale and Internal Dynamics of Polypeptide Foldons. *Macromolecules* **2018**, *51*, 3450 – 3457.

49. Casier, R.; Duhamel, J. The Effect of Like-Charges on the Conformation and Internal Dynamics of Polypeptides Probed by Pyrene Excimer Fluorescence. *Macromolecules* **2020**, *53*, 5147 – 5157.

50. Casier, R.; Duhamel, J. The Effect of Structure on Polypeptide Blobs: A Model Study Using Poly(L-Lysine). *Langmuir* **2020**, *36*, 7980 – 7990.

51. Thoma, J. L.; Duhamel, J.; Bertocchi, M. J.; Weiss, R. G. Long Range Polymer Chain Dynamics of Highly Flexible Polysiloxane in Solution Probed by Pyrene Excimer Fluorescence. *Polymers* **2018**, *10*, 345/1 – 345/15.

52. Farhangi, S.; Weiss, H.; Duhamel, J. Effect of Side-Chain Length on the Polymer Chain Dynamics of Poly(alkyl methacrylate)s in Solution. *Macromolecules* **2003**, *46*, 9738 – 9747.

53. Press, W. H.; Flanery, B. P.; Tenkolsky, S. A.; Vetterling, W. T. Numerical Recipes in Fortran: The Art of Scientific Computing; Cambridge University Press: Cambridge and New York, 1992; pp 523 − 528.

54. Oya, M.; Uno, K.; Iwakura, Y.; Polymerization of α-Amino Acid N-Carboxy Anhydride. IX. Copolymerization of α-Amino Acid N-Carboxy Anhydride Heterogeneous System. *J. Polym. Sci. A-1* **1972**, *10*, 613 – 623.

55. Hadjichristidis, N.; Iatrou, H.; Pitsikalis, M.; Sakellariou, G. Synthesis of Well-Defined Polypeptide-Based Materials via the Ring-Opening Polymerization of α-Amino Acid *N*-Carboxyanhydrides. *Chem. Rev*. **2009**, *109*, 5528 – 5578.

56. Wamsley, A.; Jasti, B.; Phiasivongsa, P.; Li, X., Synthesis of random terpolymers and determination of reactivity ratios of N-carboxyanhydrides of leucine, β-benzyl aspartate, and valine. J. *Polym. Sci. A Polym. Chem.* **2004**, *42*, 317 – 325.

57. Lakowicz, J. R. Principles of Fluorescence Spectroscopy, 3rd Ed. Springer: Singapore, 2006, pp 281.

58. Ingratta, M.; Duhamel, J. Effect of Viscosity on Long-Range Polymer Chain Dynamics in Solution with a Fluorescence Blob Model. *Macromolecules* **2009**, *42*, 1244–1251.

59. Soranno, A.; Longhi, R.; Bellini, T.; Buscaglia, M. Kinetics of Contact Formation and End-to-End Distance Distributions of Swollen Disordered Peptides. *Biophys. J.* **2009**, *96*, 1515 – 1528.

60. Oka, M.; Baba, Y.; Kagemoto, A.; Nakajima, A. Random-Coiled Conformation of Polypeptide Chains. *Polym. Bull.* **1989**, *21*, 385 – 392.

61. Flory, P. J. Statistical Mechanics of Chain Molecules, Interscience Publishers, New York, 1969.

62. Oka, M.; Hayashi, T.; Nakajima, A. Random-Coiled Conformation of Polypeptide Chains II. Experimentally Evaluated Characteristic Ratio of Poly(*N*5-2-Hydroxyethyl-L-Glutamine) and Theoretical Conformational Analysis of Poly(L-Glutamine) and Poly(L-Glutamic Acid). *Polym. J.* **1985**, *17*, 621 – 631.

63. Brant, D. A.; Miller, W. G.; Flory, P. J. Conformational Energy Estimates for Statically Coiling Polypeptide Chains. *J. Mol. Biol.* **1967**, *23*, 47 – 65.

64. Tanaka, S.; Nakajima, A. Conformational Energy and Unperturbed Chain Dimension of Polypeptide Copolymers. *Polym. J.* **1971**, *2*, 725 – 739.

65. Oka, M.; Nakajima, A. Random-Coiled Conformation of Polypeptide Chains I. Theoretical Conformational Analysis of Polypeptide Chains with Aromatic Side Chains. *Polym. J.* **1984**, *16*, 693 – 709.

66. Ingratta, M.; Mathew, M.; Duhamel, J. How Switching the Substituent of a Pyrene Derivative From a Methyl to a Butyl Affects the Fluorescence Response of Polystyrene Randomly Labeled with Pyrene. *Can. J. Chem.* **2010**, *88*, 217 – 227.

67. Krieger, F.; Fierz, B.; Axthelm, F.; Joder, K.; Meyer, D.; Kiefhaber, T. Intrachain Diffusion in a Protein Loop Fragment from Carp Parvalbumin. *Chem. Phys.* **2004**, *307*, 209 – 215.

68. Ingratta, M.; Duhamel, J. Effect of Side-chain Length on the Side-chain Dynamics of α-Helical Poly(L-glutamic acid) as Probed by a Fluorescence Blob Model. *J. Phys. Chem. B* **2008**, *112*, 9209 – 9218.

69. Farhangi, S.; Duhamel, J. Probing Side Chain Dynamics of Branched Macromolecules by Pyrene Excimer Fluorescence. *Macromolecules* **2016**, *49*, 353 – 361.

70. Dunker, A. K.; Lawson, . D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, . S.; Oldfield, C. .; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, .; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. Intrinsically Disordered Proteins. *J. Mol. Graph* **2001**, *19*, 26 – 59.

71. Dunker, A. K.; Oldfield, C. J.; Meng, .; Romero, P.; Yang, J. Y.; Chen, . W.; Vacic, V.; Obradovic, Z.; Uversky, V. N. The Unfoldomics Decade: An Update on Intrinsically Disordered Proteins. *BMC Genomics* **2008**, 9, S1, doi: 10.1186/1471-2164-9-S2-S1.

Chapter 6

1. Ivankov, D. N.; Finkelstein, A. V. Solution of Levinthal's Paradox and a Physical Theory of Protein Folding Times. *Biomolecules* **2020**, *10*, 250.
2. Plaxco, K. W.; Simons, K. T.; Baker, D. Contact Order, Transition State Placement and the Refolding Rates of Single Domain Proteins. *J. Mol. Biol.* **1998**, *277*, 985 – 994.
3. Ivankov, D. N.; Garbuzynskiy, S. O.; Alm, E.; Plaxco, K.W.; Baker, D.; Finkelstein, A. V. Contact Order Revisited: Influence of Protein Size on the Folding Rate. *Protein Sci.* **2003**, *12*, 2057 – 2062.
4. Gromiha, M. M.; Selvaraj, S. Comparison Between Long-Range Interactions and Contact Order in Determining the Folding Rate of Two-State Proteins: Application of Long-Range Order to Folding Rate Prediction. *J. Mol. Biol.* **2001**, *310*, 27 – 32.
5. Zhou, H.; Zhou, Y. Folding Rate Prediction Using Total Contact Distance. *Biophys. J.* **2002**, *82*, 458 – 463.
6. Zhang, L.; Li, J.; Jiang, Z.; Xia, A. Folding Rate Prediction Based on Neural Network Model. *Polymer* **2003**, *44*, 1751 – 1756.

7. Gromiha, M. M. Thangakani, A. M.; Selvaraj; S. FOLD-RATE: Prediction of Protein Folding Rates from Amino Acid Sequence. *Nucleic Acid Res.* **2006**, *34*, W70 – W74.
8. Ouyang Z.; Liang, J. Predicting Protein Folding Rates from Geometric Contact and Amino Acid Sequence. *Protein Sci.* **2008**, *17*, 1256 – 1263.
9. Gromiha, M. M. Importance of Native-State Topology for Determining the Folding Rate of Two-State Proteins. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1481 – 1485.
10. Naganathan, A. N.; Muñoz, V. Scaling of Folding Times with Protein Size. *J. Am. Chem. Soc.* **2005**, *127*, 480 – 481.
11. Galzitskaya, O. V.; Garbuzynskiy, S. O.; Ivankov, D. N.; Finkelstein, A. V. Chain Length is the Main Determinant of the Folding Rate for Proteins with Three-State Folding Kinetics. *Proteins* **2003**, *51*, 162 – 166.
12. Ivankov, D. N.; Finkelstein, A. V. Prediction of Protein Folding Rates from the Amino Acid Sequence-Predicted Secondary Structure. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 8942 – 8944.
13. Micheletti, C. Prediction of Folding Rates and Transition-State Placement from Native-State Geometry. *Proteins* **2003**, *51*, 74 – 84.
14. Aumpuchin, P.; Kikuchi, T. Prediction of Folding Mechanisms for Ig-Like Beta Sandwich Protein Based on Inter-Residue Average Distance Statistics Methods. *Proteins* **2019**, *87*, 120 – 135.

15. Sancho, D. D.; Muñoz, V. Integrated Prediction of Protein Folding and Unfolding Rates from Only Size and Structural Class. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17030 – 17043.
16. Gromiha, M. M.; Selvaraj, S. Inter-Residue Interactions in Protein Folding and Stability. *Prog. Biophys. Mol. Biol.* **2004**, *86*, 235 – 277.
17. Uversky, V. N. Intrinsically Disordered Proteins and Their "Mysterious" (Meta)Physics. *Front. Phys.* **2019**, *7*, doi: 10.3389/fphy.2019.00010
18. Englander, S. W.; Mayne, L. The Nature of Protein Folding Pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 15873 – 15880.
19. Baldwin R. L.; Rose, G. D. Is Protein Folding Hierarchic? I. Local Structure and Peptide Folding. *Trends Biochem. Sci.* **1999**, *24*, 26 – 33.
20. Maity, H.; Maity, M.; Krishna, M. M. G.; Mayne, L.; Englander, S. W. Protein Folding: The Stepwise Assembly of Foldon Units. *Proc. Natl. Acad. Sci.* **2005**, *102*, 4741 – 4746.
21. Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. The Protein Folding Problem. *Annu. Rev. Biophys.* **2008**, *37*, 289 – 316.
22. Casier, R.; Duhamel, J. Pyrene Excimer Fluorescence as a direct and Easy Experimental Means to Characterize the Length Scale and Internal Dynamics of Polypeptide Foldons. *Macromolecules* **2018**, *51*, 3450 – 3457.
23. Casier, R.; Duhamel, J. Effect of Structure on Polypeptide Blobs: A Model Study Using Poly(L-lysine). *Langmuir* **2020**, *36*, 7980 – 7990.

24. Casier, R. Duhamel, J. Effect of Like Charges on the Conformation and Internal Dynamics of Polypeptides Probed by Pyrene Excimer Fluorescence. *Macromolecules* **2020**, *53*, 5147 – 5157.
25. Casier, R. Duhamel, J. The Effect of Amino Acid Size on the Internal Dynamics and Conformational Freedom of Polypeptides. *Macromolecules* just accepted **2020**.
26. Mathew, A.; Siu, H.; Duhamel, J. A *Blob* Model to Study Chain Folding by Fluorescence. *Macromolecules* **1999**, *32*, 7100 – 7108.
27. Duhamel, J. Polymer Chain Dynamics in Solution Probed with a Fluorescence Blob Model. *Acc. Chem. Res.* **2006**, *39*, 953 – 960.
28. Duhamel, J. New Insights in the Study of Pyrene Excimer Fluorescence to Characterize Macromolecules and their Supramolecular Assemblies in Solution. *Langmuir* **2012**, *28*, 65276538.
29. Hall, T.; Whitton, G.; Casier, R.; Gauthier, M.; Duhamel, J. Arborescent Poly(L-glutamic acid)s as Standards To Study the Dense Interior of Polypeptide Mesoglobules by Pyrene Excimer Fluorescence. *Macromolecules* **2018**, *51*, 7914 – 7923.
30. Brooks, D. J.; Fresco, J. R.; Lesk, A. M.; Singh, M. Evolution of Amino Acid Frequencies in Proteins Over Deep Time: Inferred Order of Introduction of Amino Acids into the Genetic Code. *Mol. Biol. Evol.* **2002**, *19*, 1645 – 1655.
31. Levinthal, C. How to Fold Graciously. *Spectroscopy in Biological Systems. Proceedings University of Illinois Bulletin* **1969**, University of Illinois Press, Urbana, IL, pp 22 – 24.

32. Farhangi, S.; Duhamel, J. Long Range Polymer Chain Dynamics Studied by Fluorescence Quenching. *Macromolecules* **2016**, *49*, 6149 – 6162.
33. Sykes, M. F.; Guttmann, A. J.; Watts, M. G.; Roberts, P. D. The Asymptotic Behaviour of Self-Avoiding Walks and Returns on a Lattice. *J. Phys. A: Gen. Phys.* **1972**, *5*, 653 – 660.
34. Watts, M. G. Application of the Method of Padé Approximants to the Excluded Volume Problem. *J. Phys. A: Math. Gen.* **1975**, *8*, 61 – 66.
35. Cifra, P.; Romanov, A. On the Entropy of Single Flexible Polymer Chains. *Makromol. Chem.* **1986**, *187*, 2289 – 2297.
36. Öttinger, H. G. Computer Simulations of Three-Dimensional Multiple-Chain Systems: Scaling Laws and Virial Coefficients. *Macromolecules* **1985**, *18*, 93 – 98.
37. Mańka, A.; Nowicki, W.; Nowicka, G. Monte Carlo Simulations of a Polymer Chain Conformation. The Effectiveness of Local Moves Algorithms and Estimation of Entropy. *J. Mol. Model* **2013**, *19*, 3659 – 3670.
38. Farhangi, S.; Duhamel, J. Probing Side Chain Dynamics of Branched Macromolecules by Pyrene Excimer Fluorescence. *Macromolecules* **2016**, *49*, 353 – 361.
39. Farhangi, S.; Weiss, H.; Duhamel, J. Effect of Side-Chain Length on the Polymer Chain Dynamics of Poly(alkyl methacrylate)s in Solution. *Macromolecules* **2013**, *46*, 9738 – 9747.
40. Plotkin, S. S.; Onuchic, J. N. Understanding Protein Folding with Energy Landscape Theory Part I: Basic Concepts. *Q. Rev. Biophys*. **2002**, *35*, 111 – 167.

41. Berezovsky, I. N.; Trifonov, E. N. Loop Fold Structure of Proteins: Resolution of Levinthal's Paradox. *J. Biomol. Struct. Dyn.* **2002**, *20*, 5 – 6.
42. Finkelstein, A. V. 50+ Years of Protein Folding. *Biochem. (Mosc.)* **2018**, *83*, S3 – S18.
43. Feng, H.; Zhou, Z.; Bai, Y. A Protein Folding Pathway with Multiple Folding Intermediates at Atomic Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 5026 – 5031.
44. Manavalan, B., Kuwajima, K.; Lee, J. PFDB: A standardized Protein Folding Database with Temperature Correction. *Sci. Rep.* **2019**, *9*, 1588.
45. Garbuzynskiy, S. O.; Ivankov, D. N.; Bogatyreva, N. S.; Finkelstein, A. V. Golden Triangle for Folding Rates of Globular Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 147 – 150.
46. Censoni, L.; Martínez, L. Prediction of Kinetics of Protein Folding with Non-Redundant Contact Information. *Bioinformatics* **2018**, *34*, 4034 – 4038.
47. Capriotti, E.; Casadino, R. K-Fold: A Tool for the Prediction of the Protein Folding Kinetic Order and Rate. *Bioinformatics* **2007**, *23*, 385 – 386.
48. Gromiha, M. M. A Stastical Model for Predicting Protein Folding Rates from Amino Acid Sequence with Structural Class Information. *J. Chem. Inf. Model*. **2005**, *45*, 494 – 501.
49. Corrales, M.; Cuscó, P.; Usmanova, D. R.; Chen, H.-C.; Bogatyreva, N. S.; Filion, G. J.; Ivankov, D. N. Machine Learning: How Much Does It Tell About Protein Folding Rates? *PLoS ONE* **2015**, *10*, e0143166.

50. Baiesi, M.; Orlando, E.; Seno, F.; Trovato, A. Exploring the Correlation Between the Folding Rates of Proteins and the Entanglement of Their Native States. *J. Phys. A: Math. Theor*. **2017**, *50*, 504001.
51. Finkelstein, A. V.; Bogatyreva, N. S.; Garbuzynskiy, S. O. Restrictions to Protein Folding Determined by the Protein Size. *FEBS Lett.* **2013**, *587*, 1884 – 1890.

Chapter 7

1. Ivankov, D. N.; Finkelstein, A. V. Solution of Levinthal's Paradox and a Physical Theory of Protein Folding Times. *Biomolecules* **2020**, *10*, 250.
2. Aumpuchin, P.; Kikuchi, T. Prediction of Folding Mechanisms for Ig-Like Beta Sandwich Protein Based on Inter-Residue Average Distance Statistics Methods. *Proteins* **2019**, *87*, 120 – 135.
3. Finkelstein, A. V.; Bogatyreva, N. S.; Garbuzynskiy, S. O. Restrictions to Protein Folding Determined by the Protein Size. *FEBS Lett.* **2013**, *587*, 1884 – 1890.
4. Bogatyreva, N. S.; Finkelstein, A. V. Cunning Simplicity of Protein Folding Landscapes. *Protein Eng. Des. Sel.* **2001**, *14*, 521 – 523.
5. Muñoz, V.; Eaton, W. A. A Simple Model for Calculating the Kinetics of Protein Folding from Three-Dimensional Structures. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11311 – 11316.

6. Naganathan, A. N.; Muñoz, V. Scaling of Folding Times with Protein Size. *J. Am. Chem. Soc.* **2005**, *127*, 480 – 481.
7. Sancho, D. D.; Doshi, U.; Muñoz, V. Protein Folding Rates and Stability: How Much Is There Beyond Size? *J. Am. Chem. Soc.* **2009**, *131*, 2074 -2075.
8. Sancho, D. D.; Muñoz, V. Integrated Prediction of Protein Folding and Unfolding Rates from Only Size and Structural Class. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17030 – 17043.
9. Plaxco, K. W.; Simons, K. T.; Baker, D. Contact Order, Transition State Placement and the Refolding Rates of Single Domain Proteins. *J. Mol. Biol.* **1998**, *277*, 985 – 994.
10. Censoni, L.; Martínez, L. Prediction of Kinetics of Protein Folding with Non-Redundant Contact Information. *Bioinformatics* **2018**, *34*, 4034 – 4038.
11. Dill, K. A.; Fiebig, K. M.; Chan, H. S. Cooperativity in Protein-Folding Kinetics. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 1942-1946.
12. Micheletti, C. Prediction of Folding Rates and Transition-State Placement from Native-State Geometry. *Proteins* **2003**, *51*, 74 – 84.
13. Baiesi, M.; Orlando, E.; Seno, F.; Trovato, A. Exploring the Correlation Between the Folding Rates of Proteins and the Entanglement of Their Native States. *J. Phys. A: Math. Theor*. **2017**, *50*, 504001.
14. Ivankov, D. N.; Garbuzynskiy, S. O.; Alm, E.; Plaxco, K.W.; Baker, D.; Finkelstein, A. V. Contact Order Revisited: Influence of Protein Size on the Folding Rate. *Protein Sci.* **2003**, *12*, 2057 – 2062.

15. Ouyang, Z.; Liang, J. Predicting Protein Folding Rates from Geometric Contact and Amino Acid Sequence. *Protein Sci.* **2008**, *17*, 1256 – 1263.
16. Ivankov, D. N.; Finkelstein, A. V. Prediction of Protein Folding Rates from the Amino Acid Sequence-Predicted Secondary Structure. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 8942 – 8944.
17. Gromiha, M. M.; Selvaraj, S. Comparison Between Long-Range Interactions and Contact Order in Determining the Folding Rate of Two-State Proteins: Application of Long-Range Order to Folding Rate Prediction. *J. Mol. Biol.* **2001**, *310*, 27 – 32.
18. Zhou, H.; Zhou, Y. Folding Rate Prediction Using Total Contact Distance. *Biophys. J.* **2002**, *82*, 458 – 463.
19. Galzitskaya, O. V.; Garbuzynskiy, S. O.; Ivankov, D. N.; Finkelstein, A. V. Chain Length is the Main Determinant of the Folding Rate for Proteins with Three-State Folding Kinetics. *Proteins* **2003**, *51*, 162 – 166.
20. Gromiha, M. M.; Selvaraj, S. Inter-Residue Interactions in Protein Folding and Stability. *Prog. Biophys. Mol. Biol.* **2004**, *86*, 235 – 277.
21. Capriotti, E.; Casadino, R. K-Fold: A Tool for the Prediction of the Protein Folding Kinetic Order and Rate. *Bioinformatics* **2007**, *23*, 385 – 386.
22. Gromiha, M. M. Importance of Native-State Topology for Determining the Folding Rate of Two-State Proteins. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1481 – 1485.

23. Gromiha, M. M. A Stastical Model for Predicting Protein Folding Rates from Amino Acid Sequence with Structural Class Information. *J. Chem. Inf. Model*. **2005**, *45*, 494 – 501.
24. Gromiha, M. M. Thangakani, A. M.; Selvaraj; S. FOLD-RATE: Prediction of Protein Folding Rates from Amino Acid Sequence. *Nucleic Acid Res.* **2006**, *34*, W70 – W74.
25. Uversky, V. N. Intrinsically Disordered Proteins and Their "Mysterious" (Meta)Physics. *Front. Phys.* **2019**, *7*, doi: 10.3389/fphy.2019.00010.
26. Casier, R.; Duhamel, J. Blob-Based Approach to Estimate the Folding Time of Proteins Supported by Pyrene Excimer Fluorescence. *Macromolecules* just accepted **2020**.
27. Duhamel, J.; Kanagalingam, S.; O'Brien, T. J.; Ingratta, M. W. Side-Chain Dynamics of an α-Helical Polypeptide Monitored by Fluorescence. *J. Am. Chem. Soc.* **2003**, *125*, 12810–12822.
28. Ingratta, M.; Duhamel, J. Effect of Side-Chain Length on the Side-Chain Dynamics of α-Helical Poly(*L*-Glutamic Acid) as Probed by a Fluorescence Blob Model. *J. Phys. Chem. B* **2008**, *112*, 9209–9218.
29. Casier, R.; Duhamel, J. Pyrene Excimer Fluorescence as a direct and Easy Experimental Means to Characterize the Length Scale and Internal Dynamics of Polypeptide Foldons. *Macromolecules* **2018**, *51*, 3450 – 3457.
30. Casier, R.; Duhamel, J. Effect of Structure on Polypeptide Blobs: A Model Study Using Poly(L-lysine). *Langmuir* **2020**, *36*, 7980 – 7990.

31. Casier, R. Duhamel, J. Effect of Like Charges on the Conformation and Internal Dynamics of Polypeptides Probed by Pyrene Excimer Fluorescence. *Macromolecules* **2020**, *53*, 5147 – 5157.
32. Casier, R. Duhamel, J. The Effect of Amino Acid Size on the Internal Dynamics and Conformational Freedom of Polypeptides. *Macromolecules* just accepted **2020**.
33. Duhamel, J.; Yekta, A.; Winnik, M. A.; Jao, T. C.; Mishra, M. K.; Rubin, I. D. A Blob Model to Study Polymer Chain Dynamics in Solution. *J. Phys. Chem.* **1993**, *97*, 13708 – 13712.
34. Mathew, A. K.; Siu, H.; Duhamel, J. A Blob Model to Study Chain Folding by Fluorescence. *Macromolecules* **1999**, *32*, 7100 – 7108.
35. Duhamel, J. Global Analysis of Fluorescence Decays to Probe the Internal Dynamics of Fluorescently Labeled Macromolecules. *Langmuir* **2014**, *30*, 2307 – 2324.
36. Farhangi, S.; Duhamel, J. Probing Side Chain Dynamics of Branched Macromolecules by Pyrene Excimer Fluorescence. *Macromolecules* **2016**, *49*, 353 – 361.
37. Li, L.; Duhamel, J. Conformation of Pyrene-Labeled Amylose in DMSO Characterized with the Fluorescence Blob Model. *Macromolecules* **2016**, *49*, 7965 – 7974.
38. Farhangi, S.; Duhamel, J. Long Range Polymer Chain Dynamics Studied by Fluorescence Quenching. *Macromolecules* **2016**, *49*, 6149 – 6162.
39. Englander, S. W.; Mayne, L. The Nature of Protein Folding Pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 15873 – 15880.

40. Sykes, M. F.; Guttmann, A. J.; Watts, M. G.; Roberts, P. D. The Asymptotic Behaviour of Selfavoiding Walks and Returns on a Lattice. *J. Phys. A: Gen. Phys.* **1972**, *5*, 653 – 660.
41. Öttinger, H. G. Computer Simulations of Three-Dimensional Multiple-Chain Systems: Scaling Laws and Virial Coefficients. *Macromolecules* **1985**, *18*, 93 – 98.
42. Mańka, A.; Nowicki, W.; Nowicka, G. Monte Carlo Simulations of a Polymer Chain Conformation. The Effectiveness of Local Moves Algorithms and Estimation of Entropy. *J. Mol. Model* **2013**, *19*, 3659 – 3670.
43. Plotkin, S. S.; Onuchic, J. N. Understanding Protein Folding with Energy Landscape Theory Part I: Basic Concepts. *Q. Rev. Biophys*. **2002**, *35*, 111 – 167.
44. Berezovsky, I. N.; Trifonov, E. N. Loop Fold Structure of Proteins: Resolution of Levinthal's Paradox. *J. Biomol. Struct. Dyn.* **2002**, *20*, 5 – 6.
45. Finkelstein, A. V. 50+ Years of Protein Folding. *Biochem. (Mosc.)* **2018**, *83*, S3 – S18.
46. Feng, H.; Zhou, Z.; Bai, Y. A Protein Folding Pathway with Multiple Folding Intermediates at Atomic Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 5026 – 5031.
47. Manavalan, B., Kuwajima, K.; Lee, J. PFDB: A standardized Protein Folding Database With Temperature Correction. *Sci. Rep.* **2019**, *9*, 1588.
48. Garbuzynskiy, S. O.; Ivankov, D. N.; Bogatyreva, N. S.; Finkelstein, A. V. Golden Triangle for Folding Rates of Globular Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 147 – 150.

49. Jones, K.; Witting-Stafshede, P. The Largest Protein Observed To Fold by Two-State Kinetic Mechanism Does Not Obey Contact Order Correlation. *J. Am. Chem. Soc.* **2003**, *125*, 9606 – 9607.

50. Punta, M.; Rost, B. Protein Folding Rates Estimated from Contact Predictions. *J. Mol. Biol.* **2005**, *348*, 507 – 512.

51. Istomin, A. Y.; Jacobs, D. J.; Livesay, D. R. On the Role of Structural Class of Protein with Two-State Folding Kinetics in Determining Correlations Between Its Size, Topology, and Folding Rate. *Protein Sci.* **2007**, *16*, 2564 – 2569.

52. Samiotakis, A., Witting-Safshede, P.; Cheung, M. S. Folding, Stability and Shape of Proteins in Crowded Environments: Experimental and Computational Approaches. *Int. J. Mol. Sci.* **2009**, *10*, 572 – 588.

References: Appendix A

S3

1. Casier, R.; Duhamel, J. Pyrene Excimer Fluorescence as a Direct and Easy Experimental Means To Characterize the Length Scale and Internal Dynamics of Polypeptide Foldons. *Macromolecules* **2018**, *51*, 3450 – 3457.
2. Adzhubei, A. A.; Sternberg, M. J. E. Left-Handed Polyproline II Helices Commonly Occur in Globular Proteins. *J. Mol. Biol.* **1993**, *229*, 472 – 493.
3. Farhangi, S. Duhamel, J. Pyrenyl Derivative with a Four-Atom Linker That Can Probe the Local Polarity of Pyrene-Labeled Macromolecules. *J. Phys. Chem. B* **2016**, *120*, 834 – 842.
4. Lakowicz, J. R. *Principles of Fluorescence 2nd Ed.*; Kluwer Academic: NY, 1999.
5. Birks, J. B.; Dyson, D. J.; Munro, I. H. 'Excimer' Fluorescence. II. Lifetime Studies of Pyrene Solutions *Proc. R. Soc. A* **1963**, *275*, 575 – 588.
6. Hall, T. Study of Arborescent Poly(L-Glutamic Acid) by Pyrene Excimer Formation. MSc. Thesis, University of Waterloo, Waterloo, ON, Canada, **2012**, UWSpace: http://hdl.handle.net/10012/6972.

S4

1. Hall, T.; Whitton, G.; Casier, R.; Gauthier, M.; Duhamel, J. Arborescent Poly(l-glutamic acid)s as Standards To Study the Dense Interior of Polypeptide Mesoglobules by Pyrene Excimer Fluorescence. *Macromolecules* **2018**, *51*, 7914-7923

2. Harding, S. E. The Intrinsic Viscosity of Biological Macromolecules. Progress in Measurements, Interpretation, and Application to Structure in Dilute Solution. *Prog. Biophys. Mol. Biol.* **1997**, *68*, 207-262.

S6

1. Sykes, M. F.; Guttmann, A. J.; Watts, M. G.; Roberts, P. D. The Asymptotic Behaviour of Selfavoiding Walks and Returns on a Lattice. *J. Phys. A: Gen. Phys.* **1972**, *5*, 653 – 660.
2. Manavalan, B., Kuwajima, K.; Lee, J. PFDB: A standardized Protein Folding Database With Temperature Correction. *Sci. Rep.* **2019**, *9*, 1588.
3. Garbuzynskiy, S. O.; Ivankov, D. N.; Bogatyreva, N. S.; Finkelstein, A. V. Golden Triangle for Folding Rates of Globular Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 147 – 150.

S7

1. Muñoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. Folding Dynamics and Mechanism of β-hairpin Formation. *Nature* **1997**, *390*, 196 – 199.
2. Manavalan, B., Kuwajima, K.; Lee, J. PFDB: A standardized Protein Folding Database with Temperature Correction. *Sci. Rep.* **2019**, *9*, 1588.
3. Garbuzynskiy, S. O.; Ivankov, D. N.; Bogatyreva, N. S.; Finkelstein, A. V. Golden Triangle for Folding Rates of Globular Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 147 – 150.

Appendix A:

Supporting Information (SI)

S2 – SI for Chapter 2

UV-Vis absorption spectra.

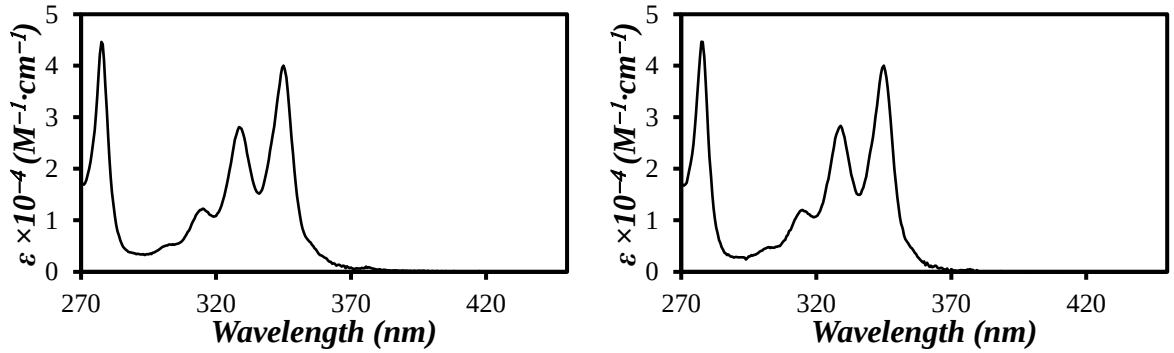


Figure S2.1: Example UV absorption spectra of Py-P*L*Glu (left) and Py-P*DL*Glu (right) in DMF.

Molar ellipticity spectra obtained from circular dichroism.

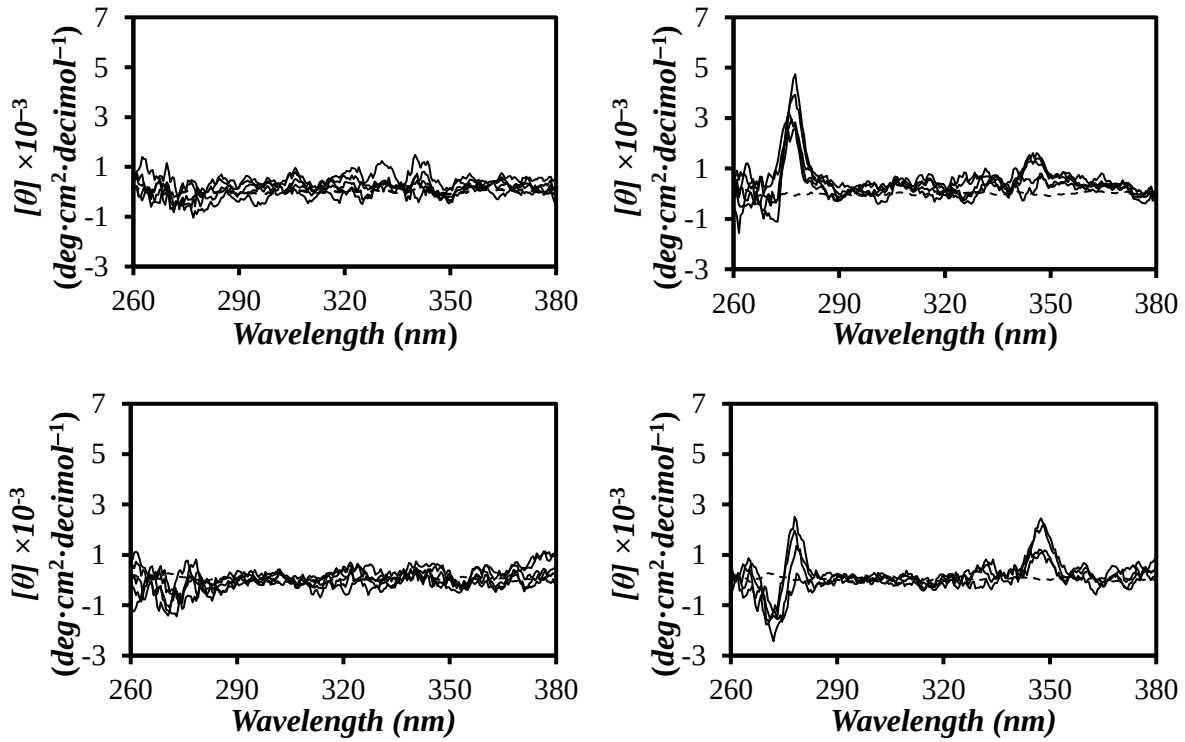


Figure S2.2: Circular dichroism of Py-P*L*Glu (right) and Py-P*DL*Glu (left) in DMF (top) and DMSO (bottom). The solvent baseline is given as the dashed lines.

Equations for Fluorescence Decay Analysis

Equation for monomer fluorescence decays:

$$[Py^*]_{(t)} = [Py^*_{diff}]_{(t)} + [Py^*_{k2}]_{(t)} + [Py^*_{free}]_{(t)} = [Py_{diff}]_o \exp\left(-\left(A_2 + \frac{1}{\tau_M}\right)t - A_3\left(1-\exp(-A_4 t)\right)\right)$$

$$+\left([Py_{k2}]_o + [Py_{diff}]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\right)\exp\left(-\left(k_2+\frac{1}{\tau_M}\right)t\right)$$

$$-[Py_{diff}]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\exp\left(-\left(A_2+iA_4+\frac{1}{\tau_M}\right)t\right)$$

$$+[Py_{free}]_o \exp\left(-\frac{t}{\tau_M}\right)\Bigg] \qquad \text{(S2.1)}$$

Equation for excimer fluorescence decays:

$$[E^*]_{(t)} = [E0^*]_{(t)} + [D^*]_{(t)} = k_2\left(\left([Py_{k2}(E0)]_o + [Py_{diff}(E0)]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\right)\right.$$

$$\times\frac{\exp\left(-\frac{t}{\tau_{E0}}\right)-\exp\left(-\left(k_2+\frac{1}{\tau_M}\right)t\right)}{k_2+\frac{1}{\tau_M}-\frac{1}{\tau_{E0}}}$$

$$\left.+[Py_{diff}(E0)]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\frac{\exp\left(-\left(A_2+iA_4+\frac{1}{\tau_M}\right)t\right)-\exp\left(-\frac{t}{\tau_{E0}}\right)}{A_2+iA_4+\frac{1}{\tau_M}-\frac{1}{\tau_{E0}}}\right)$$

$$+k_2\left(\left([Py_{k2}(D)]_o+[Py_{diff}(D)]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\right)\times\frac{\exp\left(-\frac{t}{\tau_D}\right)-\exp\left(-\left(k_2+\frac{1}{\tau_M}\right)t\right)}{k_2+\frac{1}{\tau_M}-\frac{1}{\tau_D}}\right.$$

$$\left.+[Py_{diff}(D)]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\frac{\exp\left(-\left(A_2+iA_4+\frac{1}{\tau_M}\right)t\right)-\exp\left(-\frac{t}{\tau_D}\right)}{A_2+iA_4+\frac{1}{\tau_M}-\frac{1}{\tau_D}}\right)$$

$$\left.+[E0]_o\times\exp\left(-\frac{t}{\tau_{E0}}\right)+[D]_o\times\exp\left(-\frac{t}{\tau_D}\right)\right] \qquad \text{(S2.2)}$$

In Equations S2.1 and S2.2, the parameters A_2, A_3, and A_4 are given in Equations S2.3.a–c.

$$A_2 =< n >\times\frac{k_{blob}k_e[blob]}{k_{blob}+k_e[blob]} \qquad \text{(S2.3.a)}$$

$$A_3 =< n >\times\left(\frac{k_{blob}}{k_{blob}+k_e[blob]}\right)^2 \qquad \text{(S2.3.b)}$$

$$A_4 =k_{blob}+k_e[blob] \qquad \text{(S2.3.c)}$$

$$N_{blob}=\frac{< n >}{f_{Py}}\left(1-f_{Mfree}\right) \qquad \text{(S2.4)}$$

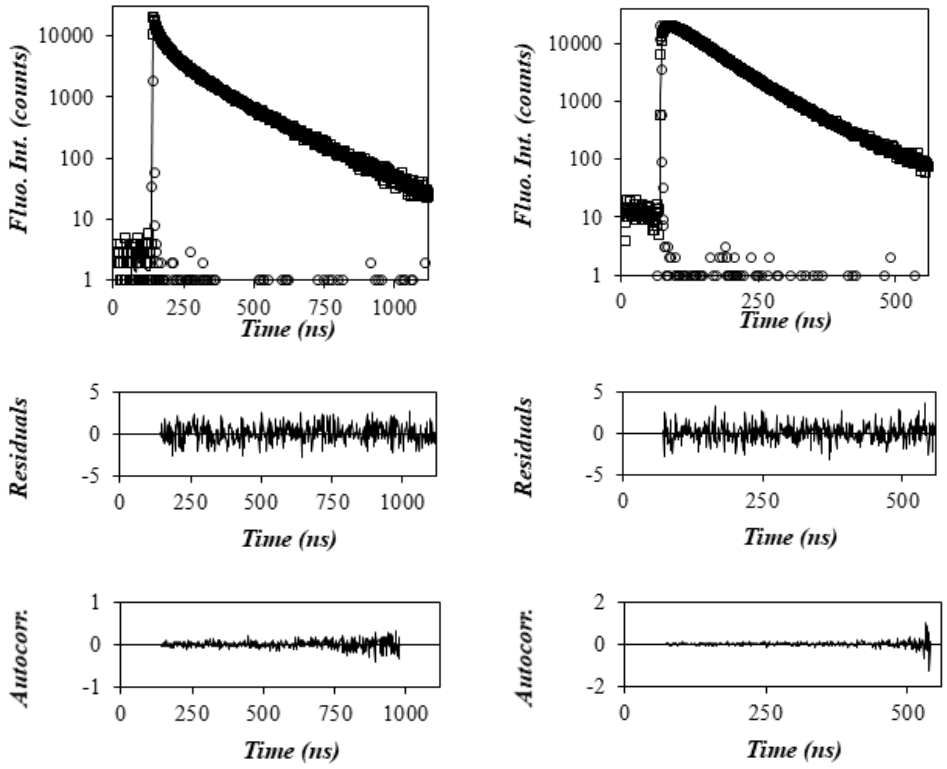


Figure S2.3: Monomer (left, λ_{em} = 375 nm) and excimer (right, λ_{em} = 510 nm) fluorescence decays (□) of Py-P*L*Glu (f_{Py} = 6.9 mol%) in DMF. The instrument response function (○) and the fit of the global analysis of the FBM (—) are overlaid with the fluorescence decays. λ_{ex} = 344 nm, [Py] = 2.5 × 10^{-6} M. χ^2 = 1.10.

Molecular Mechanics Optimizations

HyperChem Procedure: A P*L*Glu helix was constructed of 32 glutamic acid residues using the built-in alpha helix conformation parameters (Phi = –58 °, Psi = –47 °, and Omega = 180 °). The glutamate side-chain on the 9th Glu residue from the C-terminus was then modified to include a 1-pyrenemethyl amide unit. A second pyrene pendent was then added on the 10th residue. The helix backbone was prevented from moving during the optimization by constraining the atoms via the 'FIX ATOMS' constraint. Restraints were then placed on the pyrene rings such that the pyrene pendants could be brought together in a planar manner to achieve a 3.4 Å distance between the pyrenes. A typical placement of restraints were between the tops (C7 in a pyrene ring) and bottoms (C2 in a pyrene ring) of the pyrene molecules. The molecular mechanics optimization (MMO) used a Fletcher-Reeves Conjugate gradient in vacuo with a termination condition of a RMS gradient of 0.1 kcal·Å^{-1}·mol^{-1}. The optimization was considered successful when the planes of the pyrene pendants were parallel to one another with no bends present in the plane of the pyrene molecule. To ensure that the maximum overlap between the pendants was obtained, the optimization was conducted a second time as follows: The pyrene furthest away form the backbone (the 'red' pyrene in Figure S2.4A) was rotated 180 ° around the methyl bond attached to the 1-position of the pyrene. This resulted in flipping the pyrene relative to its starting position (Figure S2.4C). The MMO optimization was then completed again resulting in Figure S2.4E. (Note the difference in orientation of the 1-methyl groups of the pyrene pendants in Figures S2.4B and S2.4F.)

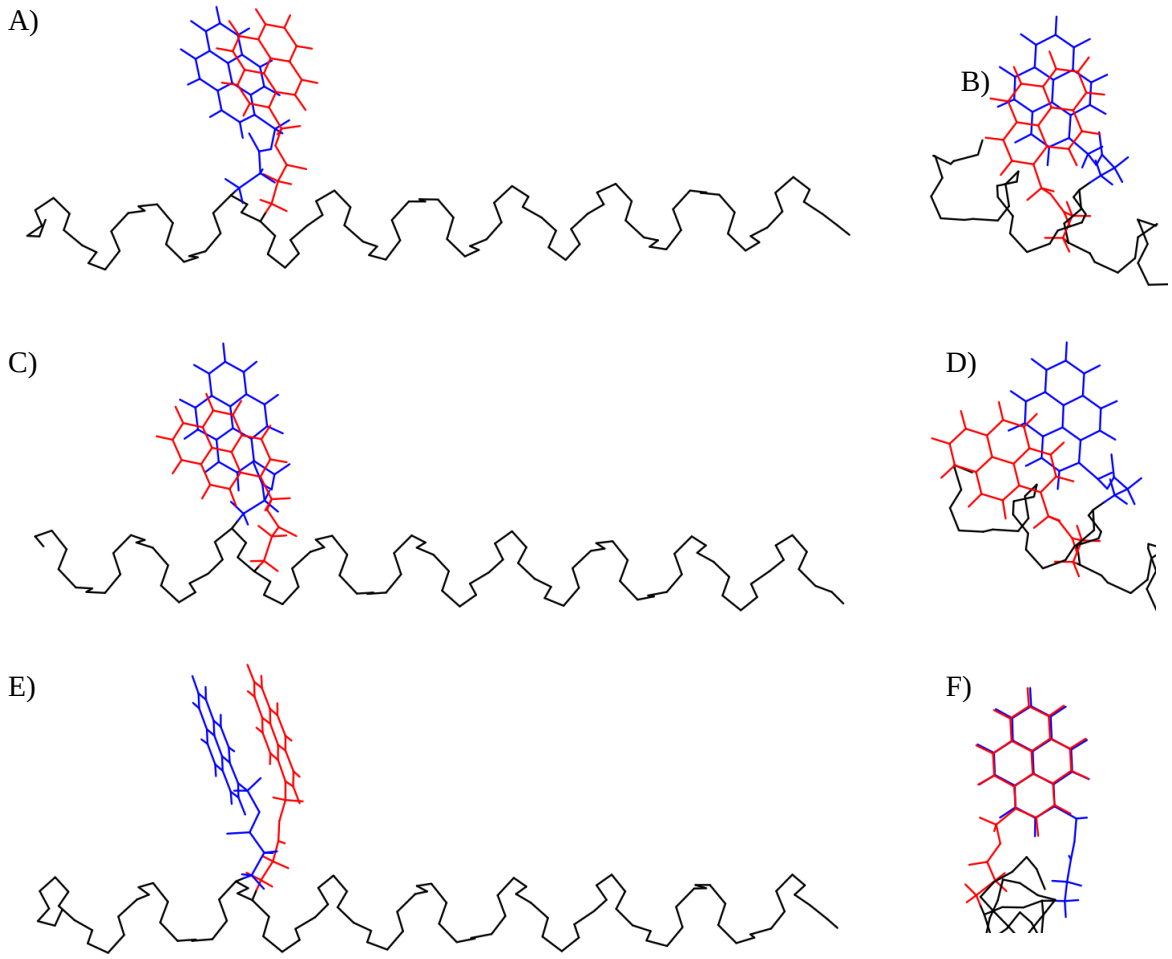


Figure S2.4: Hyperchem renderings of A,E) the side-view of the two pyrene conformations used for determining the maximum carbon overlap of the two pyrene pendants located on the 9th and 10th Glu residue from the C-Terminus with the corresponding B,F) pyrene-top-down views, respectively. C,D) displays the intermediate 180 ° rotation of the 'red' pyrene along its 1-methyl bond. Note: the sidechains along the PGlu backbone were hidden to highlight the pyrene-labeled residues.

Once the optimization was complete, and the above requirements met, the overlap of the pyrenes was determined by highlighting the pyrene on the 9th residue and orientating the screen such that it was orthogonal to the plane of the pyrene molecule. The number of carbons of the pyrene on the 10th residue that fell within the frame of the reference pyrene was then counted. The higher number of the two conformations was taken as the maximal overlap number. This procedure was then repeated via moving the second pyrene along the PLGlu backbone until no overlap was observed for several units in a row.

The pyrene carbon overlap for a P*L*Glu random coil was conducted in a similar manner as the helical P*L*Glu. To obtain a randomly coiled P*L*Glu segment, a P*L*Glu helix containing 32 Glu units was built and then stretched by setting a restraint of 999 Å between the 1st and 32^{ed} unit. After optimization, the restraint was then removed and the chain segment allowed to relax with no restraints present, yielding the conformation of the P*L*Glu random coil. The backbone was then constrained using the 'FIX ATOMS' constraint and the pyrene carbon overlap was determined in the same manner as above. Since the coil segment was not symmetrical, the pyrene labeling was repeated three times; with the first pyrene labeled on the 9th, 10th and 11th Glu residue from the C-terminus, respectively.

Parameters retrieved from the fluorescence decay analysis.

Table S2.1: Parameters retrieved using the FBM analysis of the pyrene monomer decays in DMF.

Construct	f_{Py} (mol%)	k_2 (ns^{-1})	f_{Mk2}	$k_e[blob]$ (μs^{-1})	f_{mdiff}	k_{blob} (μs^{-1})	<*n*>	τ_M (ns)	f_{Mfree}	χ^2
Py-P*DL*Glu	6.0	0.15	0.18	5.3	0.36	15	1.3	215	0.46	1.00
	8.0	0.15	0.23	2.9	0.46	13	1.3	215	0.32	1.05
	10.4	0.15	0.29	3.6	0.46	15	1.5	215	0.25	1.06
	11.5	0.15	0.22	3.1	0.54	12	1.2	215	0.24	1.07
	12.4	0.15	0.29	3.1	0.53	14	1.4	215	0.18	1.10
Py-P*L*Glu	2.9	0.15	0.16	4.0	0.40	19	1.1	215	0.44	1.15
	4.4	0.15	0.22	3.6	0.60	12	1.2	215	0.18	1.17
	4.9	0.15	0.19	3.7	0.57	14	1.1	215	0.24	1.15
	6.9	0.15	0.36	2.8	0.54	13	1.6	215	0.10	1.10
	9.0	0.15	0.38	2.9	0.52	18	2.1	215	0.10	1.21
	14.3	0.15	0.57	1.7	0.35	11	3.1	215	0.08	1.24

Table S2.2: Parameters retrieved using the FBM analysis of the pyrene excimer decays in DMF.

Construct	f_{Py} (mol%)	f_{Ek2}	τ_{E0} (ns)	$f_{EdiffE0}$	f_{EE0}	τ_{D} (ns)	f_{EdiffD}	f_{ED}	χ^2
Py-P*DL*Glu	6.0	0.29	50	0.53	0.13	107	0.05	0.00	1.00
	8.0	0.30	36	0.36	0.03	68	0.24	0.09	1.05
	10.4	0.34	41	0.38	0.00	70	0.16	0.12	1.06
	11.5	0.24	41	0.34	0.16	72	0.25	0.01	1.07
	12.4	0.28	46	0.34	0.20	75	0.17	0.01	1.10
Py-P*L*Glu	2.9	0.24	54	0.47	0.17	103	0.12	0.00	1.15
	4.4	0.23	43	0.43	0.11	83	0.19	0.04	1.17
	4.9	0.20	38	0.37	0.17	69	0.26	0.00	1.15
	6.9	0.30	40	0.43	0.02	83	0.03	0.21	1.10
	9.0	0.28	55	0.29	0.32	94	0.09	0.00	1.21
	14.3	0.39	39	0.13	0.37	70	0.11	0.00	1.24

Table S2.3: Fraction of pyrene species calculated using parameters retrieved from the FBM analysis of the pyrene monomer and excimer decays in DMF.

Construct	f_{Py} (mol%)	f_{k2}	f_{diffE0}	f_{diffD}	f_{diff}	f_{aggE0}	f_{aggD}	f_{agg}	f_{free}	χ^2
Py-P*DL*Glu	6.0	0.17	0.30	0.03	0.33	0.07	0.00	0.07	0.43	1.00
	8.0	0.21	0.25	0.17	0.42	0.02	0.06	0.08	0.29	1.05
	10.4	0.26	0.29	0.12	0.42	0.00	0.09	0.09	0.23	1.06
	11.5	0.19	0.27	0.20	0.47	0.13	0.00	0.13	0.21	1.07
	12.4	0.24	0.28	0.14	0.43	0.17	0.00	0.18	0.14	1.10
Py-P*L*Glu	2.9	0.15	0.29	0.07	0.36	0.10	0.00	0.10	0.39	1.15
	4.4	0.19	0.36	0.16	0.53	0.09	0.03	0.12	0.15	1.17
	4.9	0.16	0.21	0.29	0.50	0.14	0.00	0.14	0.20	1.15
	6.9	0.28	0.40	0.02	0.42	0.02	0.20	0.21	0.07	1.10
	9.0	0.27	0.28	0.09	0.36	0.30	0.00	0.30	0.07	1.21
	14.3	0.37	0.12	0.10	0.22	0.35	0.00	0.35	0.05	1.24

Table S2.4: Parameters retrieved using the FBM analysis of the pyrene monomer decays in DMSO.

Construct	f_{Py} (mol%)	k_2 (ns^{-1})	f_{Mk2}	$k_e[blob]$ (μs^{-1})	f_{mdiff}	k_{blob} (μs^{-1})	<*n*>	τ_M (ns)	f_{Mfree}	χ^2
Py-P*DL*Glu	6.0	0.15	0.14	8.6	0.39	15	1.3	170	0.47	1.12
	8.0	0.15	0.19	4.2	0.50	13	1.5	170	0.31	1.09
	10.4	0.15	0.20	5.7	0.46	12	1.6	170	0.33	1.12
	11.5	0.15	0.19	4.3	0.55	13	1.3	170	0.26	1.09
	12.4	0.15	0.24	3.0	0.56	12	1.5	170	0.19	1.13
Py-P*L*Glu	2.9	0.13	0.09	6.5	0.30	14	1.3	170	0.61	1.04
	4.4	0.13	0.13	3.9	0.46	13	1.3	170	0.41	1.19
	4.9	0.13	0.15	5.4	0.63	15	1.2	170	0.22	1.13
	6.9	0.13	0.20	4.6	0.63	15	1.3	170	0.17	1.03
	9.0	0.13	0.20	2.9	0.71	14	1.8	170	0.10	1.22
	14.3	0.13	0.45	3.2	0.52	10	2.8	170	0.03	1.14
	18.4	0.13	0.56	4.5	0.42	12	3.3	170	0.02	1.15

Table S2.5: Parameters retrieved using the FBM analysis of the pyrene excimer decays in DMSO.

Construct	f_{Py} (mol%)	f_{Ek2}	τ_{E0} (ns)	$f_{EdiffE0}$	f_{EE0}	τ_{D} (ns)	f_{EdiffD}	f_{ED}	χ^2
Py-P*DL*Glu	6.0	0.22	49	0.64	0.10	133	0.02	0.02	1.12
	8.0	0.26	35	0.24	0.08	58	0.41	0.00	1.09
	10.4	0.28	47	0.59	0.10	103	0.04	0.00	1.12
	11.5	0.23	46	0.57	0.11	82	0.08	0.02	1.09
	12.4	0.26	40	0.41	0.08	68	0.21	0.03	1.13
Py-P*L*Glu	2.9	0.18	54	0.60	0.17	145	0.03	0.03	1.04
	4.4	0.19	42	0.49	0.10	79	0.20	0.02	1.19
	4.9	0.15	48	0.49	0.25	84	0.12	0.00	1.13
	6.9	0.20	46	0.47	0.18	75	0.15	0.00	1.03
	9.0	0.14	43	0.22	0.33	67	0.30	0.01	1.22
	14.3	0.39	40	0.34	0.02	64	0.11	0.14	1.14
	18.4	0.46	42	0.28	0.01	63	0.06	0.19	1.15

Table S2.6: Fraction of pyrene species calculated using parameters retrieved from the FBM analysis of the pyrene monomer and excimer decays in DMSO.

Construct	f_{Py} (mol%)	f_{k2}	f_{diffE0}	f_{diffD}	f_{diff}	f_{aggE0}	f_{aggD}	f_{agg}	f_{free}	χ^2
Py-P*DL*Glu	6.0	0.13	0.36	0.01	0.37	0.06	0.01	0.07	0.44	1.12
	8.0	0.18	0.17	0.29	0.46	0.06	0.00	0.6	0.29	1.09
	10.4	0.19	0.40	0.03	0.43	0.07	0.00	0.07	0.31	1.12
	11.5	0.17	0.43	0.06	0.50	0.08	0.01	0.09	0.23	1.09
	12.4	0.21	0.34	0.17	0.51	0.07	0.03	0.10	0.17	1.13
Py-P*L*Glu	2.9	0.08	0.26	0.01	0.27	0.07	0.01	0.09	0.56	1.04
	4.4	0.12	0.30	0.12	0.43	0.06	0.02	0.08	0.38	1.19
	4.9	0.12	0.40	0.10	0.50	0.21	0.00	0.21	0.18	1.13
	6.9	0.17	0.40	0.13	0.53	0.16	0.00	0.16	0.15	1.03
	9.0	0.13	0.21	0.28	0.48	0.31	0.01	0.32	0.07	1.22
	14.3	0.38	0.33	0.11	0.44	0.02	0.14	0.16	0.02	1.14
	18.4	0.45	0.28	0.06	0.33	0.01	0.19	0.20	0.02	1.15

S3 – SI for Chapter 3

Synthetic Procedures

Pyreneacetic acid N-hydroxysuccinimide ester (2,5-dioxopyrrolidin-1-yl 2-(pyren-1-yl)acetate, PyAcOSu): 1-Pyreneacetic acid (PyAcOH) was first activated into the N-hydroxysuccinimide ester to facilitate the pyrene-labeling of poly(lysine). In a 250 mL round-bottom flask, PyAcOH (1.00 g, 3.84 mmol, 1.0 eq.) and *N*-hydroxysuccinimide (HOSu, 0.71g, 6.2 mmol, 1.6 eq.) was dispersed in 100 mL of dichloromethane (DCM) and left to stir for 15 minutes to ensure an even dispersion. *N*,*N*'-Dicyclohexylcarbodiimide (DCC, 1.17 g, 5.7 mmol, 1.5 eq.) was then added as the dispersion was stirred, which resulted in the solution quickly becoming transparent. Over the next hour the fine white precipitate of dicyclohexylurea (DCU) appeared. The reaction was left to stir overnight. The next day, the reaction mixture was suction filtered (Whatman #1) to remove the precipitated DCU before condensing the solution mixture to ca. 10 mL on a rotary evaporator. Upon condensing the solution, more DCU precipitate appeared which was subsequently removed via suction filtration. PyAcOSu was then isolated by precipitation into ethanol (ca. 100 mL) to yield a yellow-brown solid (1.34 g, 3.75 mmol, 98% yield). The solid was dried under vacuum overnight before storage at – 20 °C. ^{1}H NMR (300 MHz, d_6-DMSO): δ = 8.35 – 8.07 (m, 9H), 4.90 (s, 2H), 2.79 (s, 4H) ppm.

2,5-Dioxopyrrolidin-1-yl acetate (acetic acid N-hydroxysuccinimide ester, AcOSu): AcOSu was synthesized according to a previously reported procedure.[1]

N-(Pyren-1-ylmethyl)acetamide (PyMAAc): 1-Pyrenenethylamine hydrochloride (PyMA·HCl, 0.32 g, 1.2 mmol, 1 eq.) and AcOSu (0.24 g, 1.7 mmol, 1.4 eq.) were dispersed in 18 mL of

rapidly stirred DCM. *N*,*N*'-Diisopropylethylamine (DIPEA, 0.41 mL, 2.4 mmol, 2.0 eq.) was then added to the turbid mixture which slowly became clear over 2 hours. After leaving the clear solution to stir overnight, the solution turned turbid and the now turbid mixture was filtered (Whatman #1) to remove the white solid. The clear organic layer was then washed twice with 20 mL of 1 M HCl, twice with 20 mL of 2 M NaOH, and a final wash with 20 mL brine. The organic layer was then dried with sodium sulfate. The dried DCM was then decanted from the sodium sulfate before its removal via rotary evaporation. The resulting solid was dried under vacuum to yield PyMAAc (0.21 g, 0.8 mmol, 65 % yield) as a light-yellow solid. ^{1}H NMR (300 MHz, d_6-DMSO): δ = 8.55 (br. t, 1H), 8.39 – 8.02 (m, 9H), 4.99 (d, J = 5.7 Hz, 2H), 1.92 (s, 3H) ppm.

N-Butyl-2-(pyren-1-yl)acetamide (PyAcNBu): PyAcOSu (0.21 g, 0.58 mmol, 1 eq.) was dissolved in 18 mL of stirring DCM. To the clear brown solution, *N*-butylamine (0.17 mL, 1.7 mmol, 3 eq.) was added before leaving the solution to stir overnight. The next morning, the clear solution was washed twice with 20 mL aliquots of 1 M HCl, twice with 20 mL of 2 M NaOH, and once with brine. The DCM solution was then dried with sodium sulfate. The solution was decanted before DCM was rotary evaporated. The resulting brown solid was dried under vacuum to yield PyAcNBu (0.14 g, 0.45 mmol, 78 % yield) as a dark-brown solid. ^{1}H NMR (300 MHz, d_6-DMSO): δ = 8.41 – 7.99 (m, 10H), 4.18 (s, 2H), 3.08 (dt, J_d = 6.1 Hz, J_t = 6.6 Hz, 2H), 1.39 (m, 2H), 1.27 (m, 2H), 0.85 (t, J = 7.2Hz, 3H).

Equations of the monomer and excimer decays

The monomer and excimer decays were globally analyzed using Equations S3.1 and S3.2 for the monomer and excimer decays, respectively.

$$[Py^*]_{(t)} = [Py^*_{diff}]_{(t)} + [Py^*_{k2}]_{(t)} + [Py^*_{free}]_{(t)} = [Py_{diff}]_o \exp\left(-\left(A_2 + \frac{1}{\tau_M}\right)t - A_3\left(1-\exp(-A_4 t)\right)\right)$$

$$+\left([Py_{k2}]_o + [Py_{diff}]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\right)\exp\left(-\left(k_2+\frac{1}{\tau_M}\right)t\right)$$

$$-[Py_{diff}]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\exp\left(-\left(A_2+iA_4+\frac{1}{\tau_M}\right)t\right)$$

$$+[Py_{free}]_o \exp\left(-\frac{t}{\tau_M}\right)\Bigg] \qquad \text{(S3.1)}$$

$$[E^*]_{(t)} = [E0^*]_{(t)} + [D^*]_{(t)} = k_2\Bigg(\left([Py_{k2}(E0)]_o + [Py_{diff}(E0)]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\right)$$

$$\times\frac{\exp\left(-\frac{t}{\tau_{E0}}\right)-\exp\left(-\left(k_2+\frac{1}{\tau_M}\right)t\right)}{k_2+\frac{1}{\tau_M}-\frac{1}{\tau_{E0}}}$$

$$+[Py_{diff}(E0)]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\frac{\exp\left(-\left(A_2+iA_4+\frac{1}{\tau_M}\right)t\right)-\exp\left(-\frac{t}{\tau_{E0}}\right)}{A_2+iA_4+\frac{1}{\tau_M}-\frac{1}{\tau_{E0}}}\Bigg)$$

$$+k_2\left(\left([Py_{k2}(D)]_o+[Py_{diff}(D)]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\right)\times\frac{\exp\left(-\frac{t}{\tau_D}\right)-\exp\left(-\left(k_2+\frac{1}{\tau_M}\right)t\right)}{k_2+\frac{1}{\tau_M}-\frac{1}{\tau_D}}\right.$$

$$\left.+[Py_{diff}(D)]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\frac{\exp\left(-\left(A_2+iA_4+\frac{1}{\tau_M}\right)t\right)-\exp\left(-\frac{t}{\tau_D}\right)}{A_2+iA_4+\frac{1}{\tau_M}-\frac{1}{\tau_D}}\right)$$

$$\left.+[E0]_o\times\exp\left(-\frac{t}{\tau_{E0}}\right)+[D]_o\times\exp\left(-\frac{t}{\tau_D}\right)\right] \qquad \text{(S3.2)}$$

In Equations S3.1 and S3.2, the parameters A_2, A_3, and A_4 are given in Equations S3.3.a–c.

$$A_2=<n>\times\frac{k_{blob}k_e[blob]}{k_{blob}+k_e[blob]} \qquad \text{(S3.3.a)}$$

$$A_3=<n>\times\left(\frac{k_{blob}}{k_{blob}+k_e[blob]}\right)^2 \qquad \text{(S3.3.b)}$$

$$A_4=k_{blob}+k_e[blob] \qquad \text{(S3.3.c)}$$

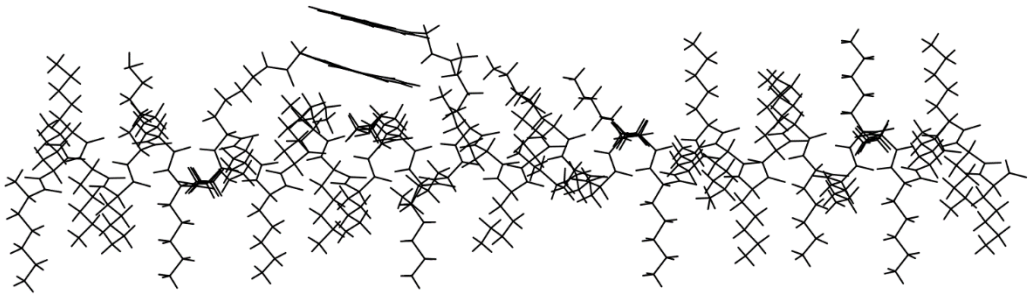
Molecular mechanics optimizations for Py-P*L*L·HCl and Py-P*L*Glu

The helical P*L*L was constructed in Hyperchem 8.0.7 for windows using the built-in α-helix database (Phi = −58 °, Psi = −47 °, Omega = 180 °) consisting of 45 *L*-lysine residues. The ammonium-end of the 10th lysine residue from the N-terminus was modified by attaching a 1-pyrenacetic acid unit, referred to as the reference residue. A second residue was then modified in a similar manner. The backbone of the P*L*L helix was locked in place using the 'FIX ATOMS' selection. Restraints were then added between the two pyrene pendants to force them to come together in such a way that the planes of the pyrene pendants were held parallel to one another and separated by 3.4 Å. The MMO to minimize the energy of the restrained molecule was then performed using a Polak-Ribiere (Conjugate gradient) algorithm in vacuo until a RMS gradient less than 0.1 kcal·Å^{-1}·mol^{-1} was achieved. The extent of overlap between the two parallel pyrene pendants was then assessed by superimposing the two pyrene rings and counting the number of carbon atoms in the second pyrene that overlapped the frame of the reference pyrene. The second pyrene was then moved to the next lysine side chain and the above procedure was repeated. This process was continued until the two pyrenes could no longer touch and the overlap between the two became zero. Figure S3.1A depicts the good overlap experienced between two pyrene pendants when the second pyrene was located 11 residues away from the reference pyrene.

To construct a P*L*L chains adopting a random coil conformation, a chain of 45 lysines was created and it was fully stretched by constraining the chain ends to be very far apart from one another. An end-to-end distance of 999 Å was used, but any number larger than the

extended chain would have sufficed. The stretched chain was then relaxed until the RMS gradient was less than 0.1 kcal·Å^{-1}·mol^{-1}, yielding the relaxed random coil conformation for the 45-mer oligolysine used for the MMOs. The structure of the relaxed chain with pyrene on the 10th and 18th residue from the N-terminus is shown in Figure S3.1B. Since the coiled segment is not symmetrical, the MMOs were repeated three times by changing the location of the refence pyrene from the 10th, 11th, and 12th and averaging the number of overlapping carbons obtained for the three overlapping pyrene pairs.

A)



B)

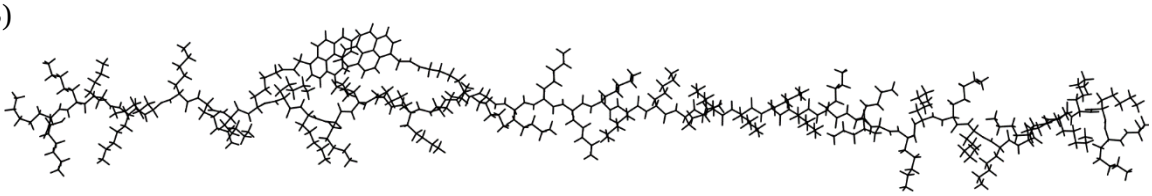


Figure S3.1. Illustrations of the quality of pyrene carbon overlap for Py-PLL·HCl made of 45 lysine and labeled with a reference pyrene on the 10th residue from the left and having the second pyrene labeled on the A) 21st residue (good overlap) when PLL adopted a helical conformation and B) 18th residue (poor overlap) when PLL adopted a coiled conformation.

The MMOs were repeated for Py-P*L*Glu in a 3_{10}-helix and PPII-helix conformation using a 30-mer oligo(*L*-glutamic acid) in a manner similar to that applied to generate the α-helical Py-P*L*L·HCl, constructs. The built-in angles for the 3_{10}-helix were Phi = −49 °, Psi = −26 °, and Omega = 180 °. The PPII-helix was constructed using Phi = −75 °, Psi = 145 °, and Omega = 180 °.[2] Due to the symmetry of the helix, the MMOs were repeated only once using the 4th glutamic acid residue from the *N*-terminus as reference.

Quenching efficiency of PyAcNBu and PyMAAc in DMSO

Although both poly(lysine) and poly(glutamic acid) were labeled with pyrene via an amide bond, the pyrenyl motive resided on opposite sides of the amide bond for the two polypeptides. For Py-P*L*L·HCl, the 1-pyrenemethyl group is bonded to the carbonyl side, while it is bonded to the nitrogen side of the amide for Py-P*L*Glu and Py-P*DL*Glu (Figure 3.1). Since this difference in configuration involves a single methylene group separating the pyrenyl from either atom of the amide bond, it might affect the photophysical properties of pyrene.[3] Some of these properties include the monomer lifetime (τ_M), the molar absorptivity coefficient (ε), quantum yield (ϕ), and the efficiency (p) of the pyrene derivative to form excimer upon encounter between two pyrenes.[4] τ_M is typically found by labeling a small amount of pyrene onto the polypeptide and fitting the fluorescence decay with a mono-exponential function, while ε is typically determined by measuring the absorbance of a model compound. The model compounds PyAcNBu and PyMAAc were used to determine ε for Py-P*L*L·HCl and Py-P*L*Glu (and Py-P*DL*Glu), respectively. Determining τ_M and ε is enough to fully characterize the pyrene-labeled polypeptides according to the FBM, however in order to fully compare the results between the two, p should be determined. To do this the Birks Scheme was applied to the analysis of the monomer and

excimer fluorescence decays of the model compounds PyAcNBu and PyMAAc. The kinetic scheme representing pyrene excimer formation according to the Birks scheme is described in Scheme S3.1.[5]

$$\begin{array}{ccccccc} Py + Py + h\nu & \longrightarrow & Py & + & Py^* & \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} & E^* \\ & & & & \downarrow \frac{1}{\tau_M} & & \downarrow \frac{1}{\tau_E} \end{array}$$

Scheme S3.1. Birks scheme for pyrene excimer formation.

In Scheme S3.1, an excited pyrene *Py**, with lifetime τ_M, encounters another ground-state pyrene *Py* which can result in the formation of excimer E* with a rate k_1. E* can then either dissociate to regenerate *Py* and *Py** with the rate constant k_{-1} or decay to the ground-state with the lifetime τ_E. The time-dependent concentrations of the pyrene monomer and excimer produced according to the kinetic scheme shown in Scheme S3.1 have been solved for and their equations are given in Equations S3.4–3.6, where $X = k_1 + \tau_M^{-1}$ and $Y = k_{-1} + \tau_E^{-1}$.[5]

$$[E^*] = \frac{k_1[Py^*]_o}{\sqrt{(X-Y)^2 + 4k_1k_{-1}}}\left[-\exp(-t/\tau_1) + \exp(-t/\tau_2)\right] \quad \text{(S3.4)}$$

$$[Py^*] = \frac{[Py^*]_o}{\sqrt{(X-Y)^2 + 4k_1k_{-1}}}\left[\left(X - \tau_2^{-1}\right)\times\exp(-t/\tau_1) - (X - \tau_1^{-1})\times\exp(-t/\tau_2)\right] \quad \text{(S3.5)}$$

$$\tau_1^{-1} = \frac{X + Y + \sqrt{(X-Y)^2 + 4k_1k_{-1}}}{2} \quad \text{(S3.6.a)}$$

$$\tau_2^{-1} = \frac{X + Y - \sqrt{(X-Y)^2 + 4k_1k_{-1}}}{2} \quad \text{(S3.6.b)}$$

Samples containing 4 to 12 mM of a given model compound in DMSO were prepared. After purging with nitrogen, the steady-state and time-resolved fluorescence of the samples were then acquired according to a procedure similar to the one outlined in the experimental section using a time-per-channel of 2.04 ns·ch^{-1} for the excimer decay. In Equation S3.7, κ is an instrument constant of the steady-state fluorometer, and ϕ_E^0 and ϕ_M^0 are the fluorescence quantum yields of the monomer and excimer, respectively.

$$\frac{I_E}{I_M} = \kappa \frac{\phi_E^0}{\phi_M^0} \frac{k_1[Py]}{k_{-1} + \tau_E^{-1}} \tag{S3.7}$$

Equation S3.7 implies that a plot of I_E/I_M should be linear with respect to concentration with a slope that depends on the instrument constant, the quantum yields, the excimer dissociation rate constant and lifetime, and the rate constant of excimer formation. A plot of I_E/I_M as a function of concentration is given in Figure S3.2 for PyAcNBu and PyMAAc.

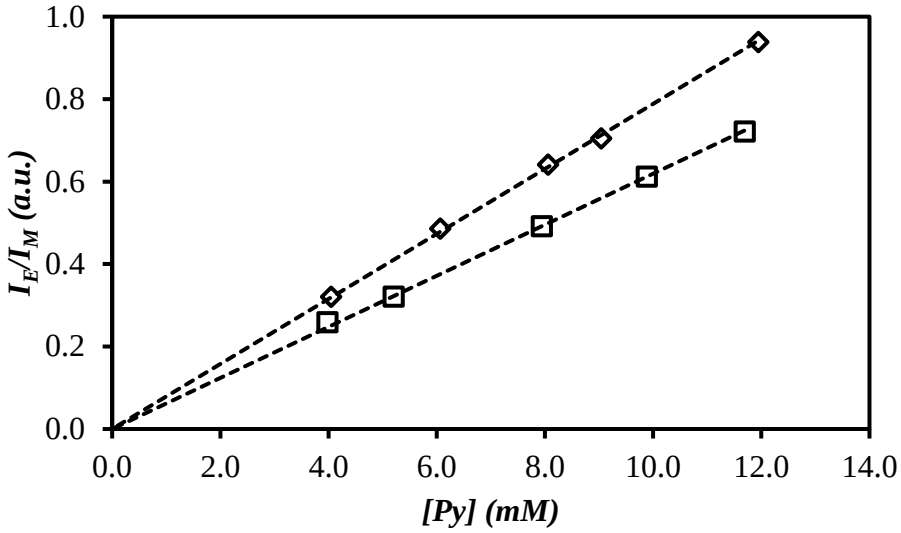


Figure S3.2. I_E/I_M as a function of concentration for (◇) PyMAAc and (□) PyAcNBu in DMSO. The dashed lines represent the linear fits with slope of 78.8 ± 0.4 and 61.9 ± 0.3 M^{-1} for PyMAAc and PyAcNBu, respectively.

As seen in Figure S3.2, PyMAAc has a slope a little over 25% larger than for PyAcNBu. Since all the measurements were acquired on the same instrument, κ is constant between the two series, and implies that the rate constants and/or quantum yields vary between the two derivatives. To remove the effects of quantum yield and isolate k_1, time-resolved fluorescence was utilized. The fluorescence decays were fitted according to Equations S3.4–3.6 using the in-house program *globirks32bg*. The parameters retrieved from the fits are listed in Table S1, including the product $k_1 \times [Py]$. A rate constant describing a diffusion-controlled reaction such as k_1 for excimer formation is expected to obey Equation S3.8, where R is the encounter radii of the two reactants, D is the combined diffusion coefficients of the two reactants, and p is the probability of having reaction occurring upon contact. The Stokes-

Einstein equation for D is given in Equation S3.9, where k_B is the Boltzmann constant, T is temperature, η is viscosity, and r is the hydrodynamic radius.

$$k_1 = 4\pi N_A RDp \tag{S3.8}$$

$$D = \frac{k_B T}{6\,\pi\,\eta\,r} \tag{S3.9}$$

Combining Equations S3.8 and S3.9 suggests that k_1 should be proportional to pR/r, so that a plot of $k_1 \times [Py]$ vs. $[Py]$ for PyAcNBu and PyMAAc in DMSO should be linear with a slope proportional to pR/r. This is indeed the case as shown in Figure S3.3.

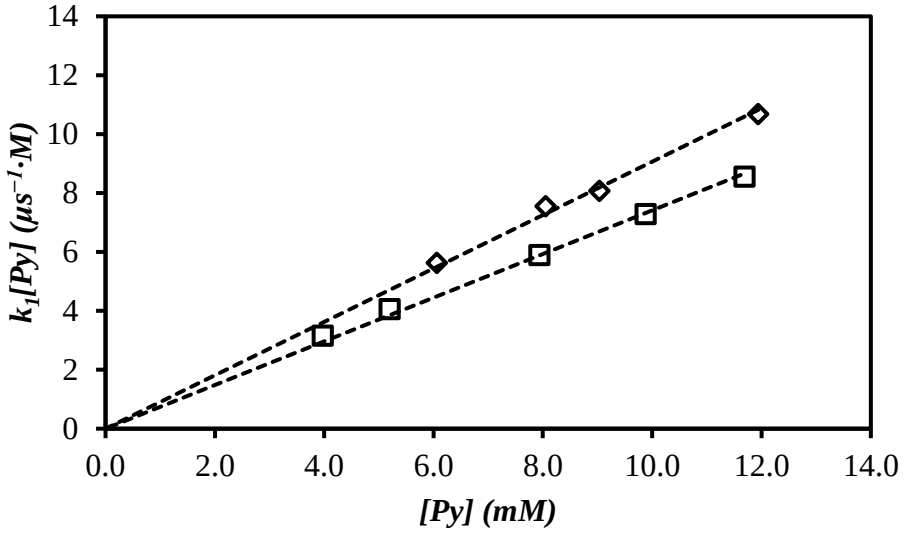


Figure S3.3. Plot of $k_1 \times [Py]$ as a function of concentration for (◇) PyMAAc and (□) PyAcNBu in DMSO. The dashed lines represent the linear fits with slopes of 0.91 ± 0.01 and 0.74 ± 0.01 ns^{-1} for PyMAAc and PyAcNBu, respectively.

In Figure S3.3, both model compounds exhibited a linear relationship between $k_1 \times [Py]$ and $[Py]$ with slopes of 0.91 and 0.74 ns^{-1} for PyMAAc and PyAcNBu, respectively. The slope of 0.91 ± 0.01 ns^{-1} for PyMMAc was found to be similar with the previously determined value of 0.99 ± 0.02 ns^{-1}.[6] Since both PyAcNBu and PyMAAc are similar in size, with PyAcNBu having an additional three methylene groups in its structure, they are expected to have similar hydrodynamic volumes and therefore similar *R* and *r* values. As a result, Figure S3.3 demonstrates that PyMAAc is about 20% more efficient at forming excimer than PyAcNBu in DMSO.

Parameters retrieved from the Birks Scheme analysis of PyAcNBu and PyMAAc in DMSO

Table S3.1. Parameters retrieved from the Birks scheme analysis of the monomer and excimer decays of PyAcNbu and PyMAAc in DMSO.

	Conc. (mM)	χ^2	τ_1 (ns)	a_{M1}	τ_2 (ns)	a_{M2}	a_{E1}	a_{E2}	k_1[Py] ($\mu s^{-1} \cdot M$)	k_{-1} (μs^{-1})	τ_E (ns)
PyAcNBu τ_M = 163 ns	4.0	1.29	47	0.09	123	0.91	-0.63	0.66	4.6	3.2	64
	5.2	1.30	46	0.10	112	0.90	-0.69	0.73	3.6	4.1	59
	7.9	1.27	45	0.11	93	0.89	-0.85	0.88	2.3	5.9	54
	9.9	1.26	44	0.13	83	0.87	-0.96	0.99	1.8	7.3	51
	11.7	1.13	44	0.17	76	0.83	-1.12	1.15	1.5	8.6	51
PyMAAc τ_M = 185 ns	6.1	1.28	44	0.11	105	0.89	-0.64	0.66	3.0	5.6	55
	8.1	1.29	44	0.15	89	0.85	-0.85	0.88	2.3	7.6	53
	9.0	1.10	42	0.13	83	0.87	-0.82	0.84	1.9	8.1	49
	11.9	1.29	42	0.19	70	0.81	-1.30	1.32	1.3	10.7	49

Example global-fit of the monomer and excimer decays of Py-P*DL*GNa according to the FBM

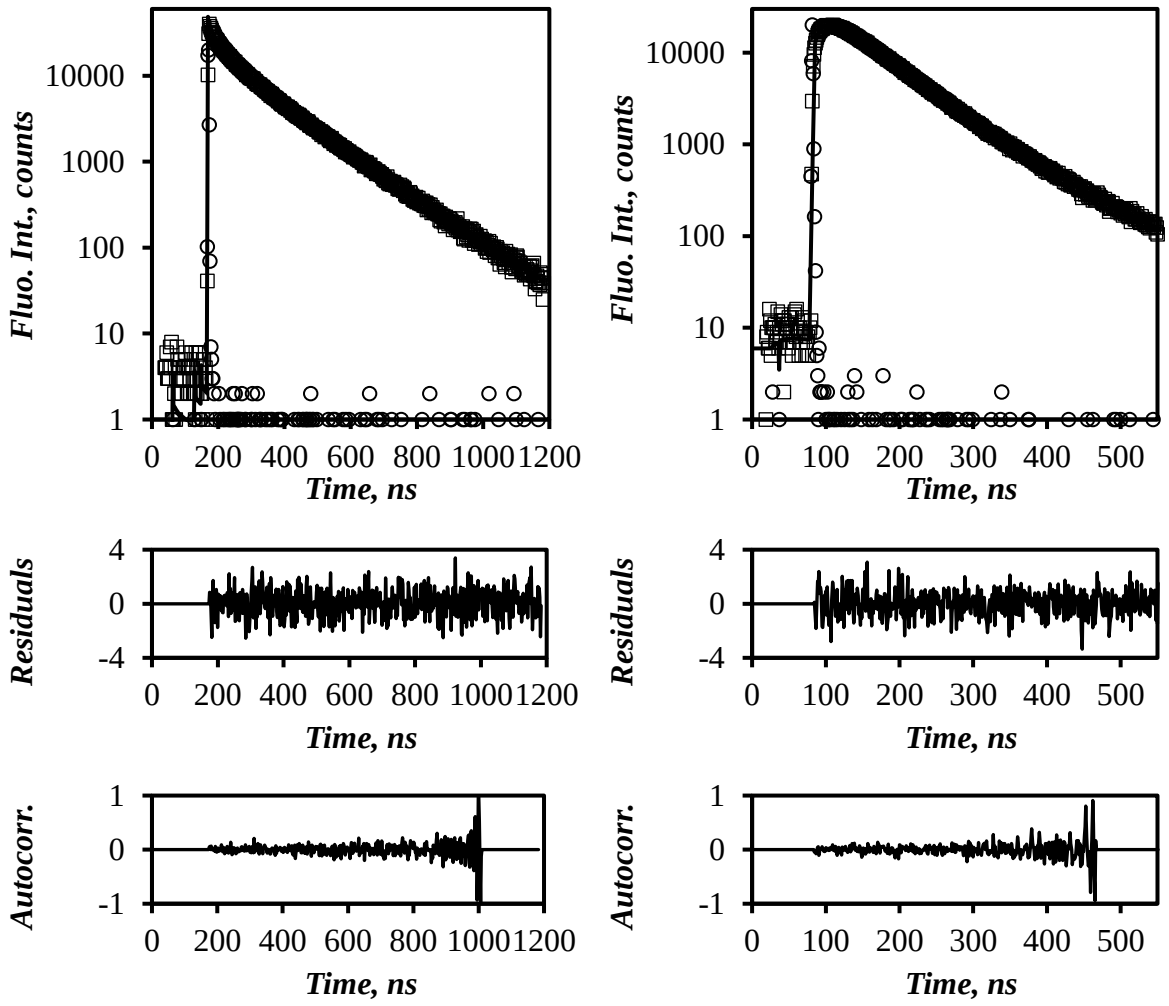


Figure S3.4. Monomer (left, λ_{em} = 375 nm) and excimer (right, λ_{em} = 510 nm) fluorescence decays (□) of Py-P*DL*GNa (x = 10.4 mol%) in DMSO. The instrument response function (○) and the fit of the global analysis of the FBM (—) are overlaid with the fluorescence decays. λ_{ex} = 344 nm, [Py] = 2.7×10^{-6} M. χ^2 = 1.03

Parameters retrieved from the FBM analysis of Py-P*LL*·HCl in DMSO solutions under acidic conditions

Table S3.2. Parameters retrieved using the FBM analysis of the pyrene monomer decays of Py-P*LL*·HCl under acidic conditions.

Conditions	x (mol%)	k_2 (ns^{-1})	f_{Mk2}	$k_e[blob]$ (μs^{-1})	f_{mdiff}	k_{blob} (μs^{-1})	<n>	f_{Mfree}	χ^2
DMSO τ_M = 157 ns	5.5	0.09	0.17	4.0	0.57	7.7	1.04	0.26	1.06
	7.6	0.09	0.30	5.4	0.59	9.4	1.28	0.12	1.23
	7.4	0.09	0.22	4.3	0.62	8.7	1.27	0.15	1.22
	10.3	0.09	0.36	3.1	0.60	7.7	1.70	0.04	1.10
	12.1	0.09	0.36	4.5	0.57	8.1	1.61	0.07	1.20

Table S3.3. Parameters retrieved using the FBM analysis of the pyrene excimer decays of Py-P*LL*·HCl under acidic conditions.

Conditions	x (mol%)	f_{Ek2}	τ_{E0} (ns)	$f_{EdiffE0}$	f_{EE0}	τ_D (ns)	f_{EdiffD}	f_{ED}	χ^2
DMSO	5.5	0.22	39.8	0.35	0.03	52.9	0.36	0.05	1.06
	7.6	0.30	44.4	0.35	0.05	45.4	0.24	0.07	1.23
	7.4	0.25	39.2	0.51	0.00	60.7	0.18	0.07	1.22
	10.3	0.34	41.6	0.37	0.00	51.1	0.20	0.10	1.10
	12.1	0.35	42.4	0.36	0.02	51.4	0.20	0.07	1.20

Table S3.4. Molar fraction of pyrene species calculated using parameters retrieved from the FBM analysis of the pyrene monomer and excimer decays of Py-P*L*L·HCl under acidic conditions.

Conditions	x (mol%)	f_{k2}	f_{diffE0}	f_{diffD}	f_{diff}	f_{aggE0}	f_{aggD}	f_{agg}	f_{free}	χ^2
DMSO	5.5	0.16	0.26	0.27	0.53	0.02	0.04	0.05	0.25	1.06
	7.6	0.27	0.32	0.21	0.53	0.04	0.06	0.10	0.10	1.23
	7.4	0.21	0.43	0.15	0.58	0.00	0.06	0.06	0.14	1.22
	10.3	0.32	0.35	0.19	0.55	0.00	0.10	0.10	0.03	1.10
	12.1	0.33	0.33	0.19	0.52	0.01	0.07	0.09	0.06	1.20

Parameters retrieved from the FBM analysis of Py-PLGlu and Py-PDLGlu in DMSO under alkaline conditions

Table S3.5. Parameters retrieved using the FBM analysis of the pyrene monomer decays of pyrene-labeled poly(glutamic acid) in DMSO under alkaline conditions.

Construct	x (mol%)	k_2 (ns^{-1})	f_{Mk2}	$k_e[blob]$ (μs^{-1})	f_{mdiff}	k_{blob} (μs^{-1})	$<n>$	f_{Mfree}	χ^2
Py-P*DL*GNa τ_M = 170 ns	6.0	0.084	0.15	7.1	0.26	8.8	1.56	0.59	1.04
	8.0	0.084	0.19	5.1	0.38	9.0	1.38	0.43	1.04
	10.4	0.084	0.26	3.3	0.42	7.5	1.61	0.32	1.03
	11.5	0.084	0.24	3.8	0.46	8.5	1.29	0.30	0.98
	12.4	0.084	0.35	6.5	0.50	10.0	1.85	0.15	1.10
Py-P*L*GNa τ_M = 170 ns	4.4	0.076	0.15	4.4	0.46	8.8	1.18	0.38	1.06
	4.9	0.076	0.22	3.6	0.52	7.7	1.23	0.27	1.05
	6.9	0.076	0.37	3.9	0.53	8.7	1.49	0.11	1.10
	9.0	0.076	0.41	2.5	0.51	7.4	1.76	0.08	1.09
	14.3	0.076	0.56	3.0	0.41	6.6	2.81	0.04	1.18

Table S3.6. Parameters retrieved using the FBM analysis of the pyrene excimer decays of pyrene-labeled poly(glutamic acid) in DMSO under alkaline conditions.

Construct	x (mol%)	f_{Ek2}	τ_{E0} (ns)	$f_{EdiffE0}$	f_{EE0}	τ_{D} (ns)	f_{EdiffD}	f_{ED}	χ^2
Py-P*DL*GNa	6.0	0.30	50.6	0.50	0.17	133	0.03	0.00	1.04
	8.0	0.28	44.9	0.48	0.15	84	0.09	0.00	1.04
	10.4	0.33	47.1	0.50	0.12	107	0.03	0.01	1.03
	11.5	0.30	45.8	0.47	0.13	79	0.10	0.00	0.98
	12.4	0.36	49.8	0.50	0.10	107	0.00	0.05	1.10
Py-P*L*GNa	4.4	0.20	45.4	0.54	0.16	99	0.07	0.03	1.06
	4.9	0.24	48.1	0.52	0.16	100	0.06	0.03	1.05
	6.9	0.28	49.1	0.33	0.32	88	0.06	0.00	1.10
	9.0	0.34	50.7	0.39	0.23	99	0.03	0.01	1.09
	14.3	0.43	38.1	0.21	0.12	65	0.11	0.13	1.18

Table S3.7: Fraction of pyrene species calculated using parameters retrieved from the FBM analysis of the pyrene monomer and excimer decays of pyrene-labeled poly(glutamic acid) in DMSO under alkaline conditions.

Construct	x (mol%)	f_{k2}	f_{diffE0}	f_{diffD}	f_{diff}	f_{aggE0}	f_{aggD}	f_{agg}	f_{free}	χ^2
Py-P*DL*GNa	6.0	0.13	0.23	0.01	0.24	0.08	0.00	0.08	0.55	1.04
	8.0	0.17	0.29	0.06	0.35	0.09	0.00	0.09	0.39	1.04
	10.4	0.24	0.36	0.02	0.38	0.09	0.01	0.10	0.29	1.03
	11.5	0.22	0.35	0.07	0.42	0.09	0.00	0.09	0.27	0.98
	12.4	0.31	0.43	0.00	0.43	0.09	0.04	0.13	0.13	1.10
Py-P*L*GNa	4.4	0.13	0.36	0.05	0.40	0.11	0.02	0.13	0.33	1.06
	4.9	0.19	0.40	0.04	0.44	0.12	0.02	0.14	0.23	1.05
	6.9	0.26	0.31	0.06	0.37	0.30	0.00	0.30	0.08	1.10
	9.0	0.32	0.36	0.03	0.39	0.21	0.01	0.23	0.06	1.09
	14.3	0.42	0.20	0.10	0.31	0.12	0.12	0.25	0.03	1.18

S4 – SI for Chapter 4

Synthesis of small molecules

Naphthaleneacetic acid N-hydroxysuccinimide ester (2,5-dioxopyrrolidin-1-yl 2-(naphthalen-1-yl)acetate, NpAcOSu): 1-Naphtheleneacetic acid (2.00 g, 10.7 mmol, 1.0 eq.) and *N*-hydrosuccinimide (HOSu) (1.85 g, 16.1 mmol, 1.5 eq.) was dissolved in acetonitrile (ACN, 150 mL) in a 250 mL round-bottom flask with stirring. To the stirring solution, *N*,*N*'-dicyclohexylcarbodiimide (DCC, 2.45 g, 11.9 mmol, 1.1 eq.) was added and left to stir overnight. The solution was then suction filtered (Whatman #1) to remove the dicyclohelyurea (DCU) precipitate that formed. The reaction solution was then condensed on a rotavap (ca. 20 mL) and refiltered to remove the additional DCU precipitate. The solution was then mixed with ethanol (ca. 100 mL) and chilled to – 20 °C, which resulted in the appearance of a light-yellow oily layer at the bottom of the flask. The ethanol layer was decanted and diethyl ether (100 mL) was added with vigorous stirring and then cooled to – 20 °C which resulted in the formation of an off-white waxy solid. The solvent was decanted and the solid was dried under vacuum overnight to yield NpAcOSu (2.34 g, 8.3 mmol, 77% yield) as an off-white waxy solid. ^{1}H NMR (300 MHz, d_6-DMSO): δ = 7.99 – 7.42 (m, 7H), 4.58 (s, 2H), 2.78 (s, 4H).

N-Butyl-2-(naphthalen-1-yl)acetamide (NpAcNBu): NpAcOSu (0.41 g, 1.4 mmol, 1.0 eq.) was dissolved in dichloromethane (DCM, 20 mL) with stirring. *n*-Butylamine (0.44 mL, 4.4 mmol, 3.2 eq.) was added and the mixture was left to stir for 72 hours. The white precipitate that had formed was removed via suction filtration (Whatman #1) and discarded. The clear organic filtrate was then washed sequentially with 50 mL aliquots of 1 M HCl, 2 M NaOH, and brine (water saturated with NaCl) before being dried with sodium sulfate. The dried organic layer

was then decanted and rotary evaporated to dryness. This yielded NpAcNBu (0.30 g, 1.2 mmol, 86% yield) as a white solid. ^{1}H NMR (300 MHz, d_6-DMSO): δ = 8.18 – 7.40 (m, 8H), 3.87 (s, 2H), 3.03 (dt, J_d = 6.0 Hz, J_t = 6.6 Hz, 2H), 1.35 (m, 2H), 1.27 (m, 2H), 0.85 (t, J = 7.2Hz, 3H).

NpAcNBu molar absorptivity coefficient (ε_{Np})

NpAcNBu was dissolved in known amounts of DMF to create a series of six solutions with differing NpAcNBu concentrations. The UV absorbance of the solutions was measured with a 1.00 cm pathlength cell, and the absorption maximum at 284 nm was recorded. A background correction of pure DMF was applied to each sample measurement. An example of the obtained absorption spectrum is given in Figure S4.1A. A plot of the corrected absorbance at the naphthalene absorption peak maximum as a function of concentration is given in Figure S4.1B. Using the Beer-Lambert law, ε_{Np} was found to equal 6350 ± 20 $M^{-1}\cdot cm^{-1}$ from the slope of the plot.

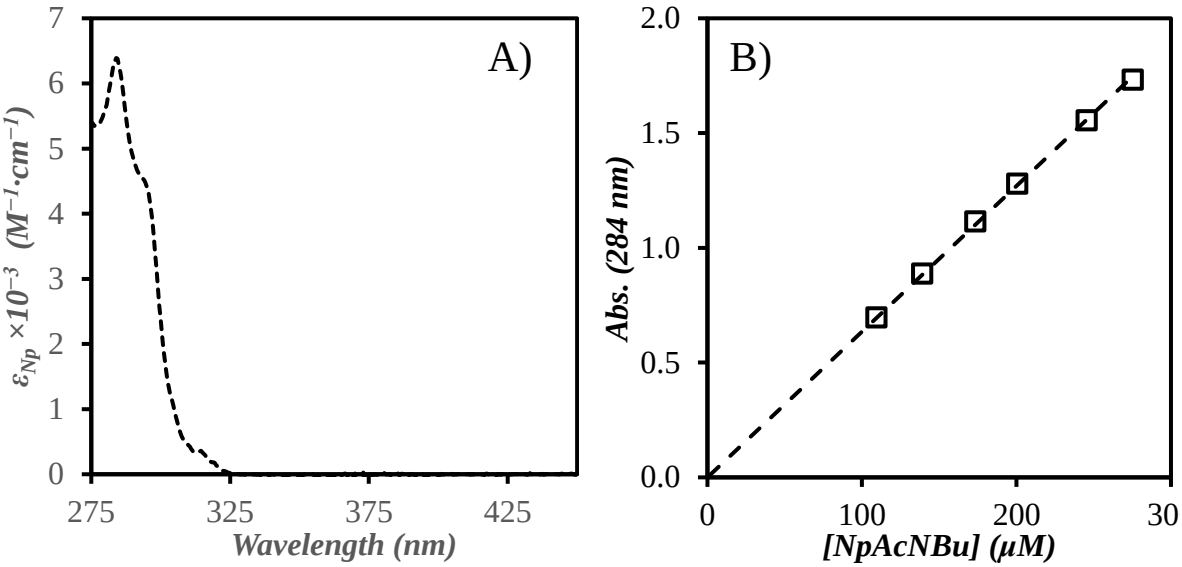


Figure S4.1. Plot of A) the absorbance spectrum of NpAcNBu and B) the absorption at 284 nm as a function of concentration. The dashed line in B represents the linear fit of the data with a slope equal to 6.35 ± 0.02 mM^{-1} and a forced intercept of zero, R^2 = 1.00. Solvent = DMF.

Absorbance and steady-state fluorescence spectra of Py-P*L*L and Np-P*L*L

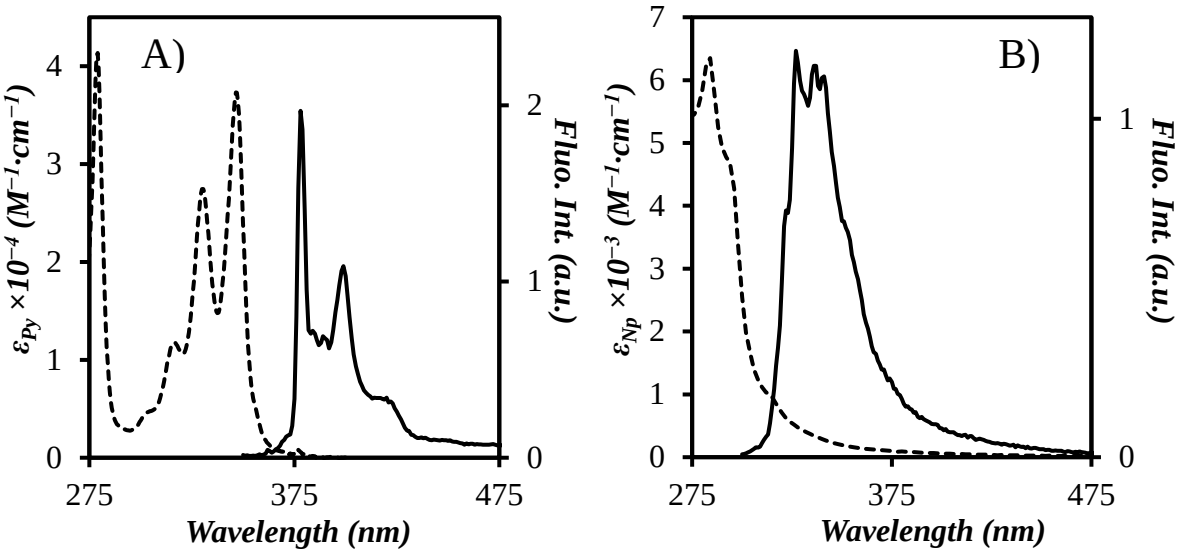


Figure S4.2. Plots of the (_ _) absorbance and (___) emission spectra of A) Py-P*L*L and B) Np-P*L*L in DMF. A) λ_{ex}= 344 nm, [Py] = 2.7 μM, x_{Py} = 0.01. B) λ_{ex}= 295 nm, [Np] = 15.7 μM, x_{Np} = 0.09.

Steady-state fluorescence spectra of Np- and Py-P*L*L in a 90:10 ACN:water mixture

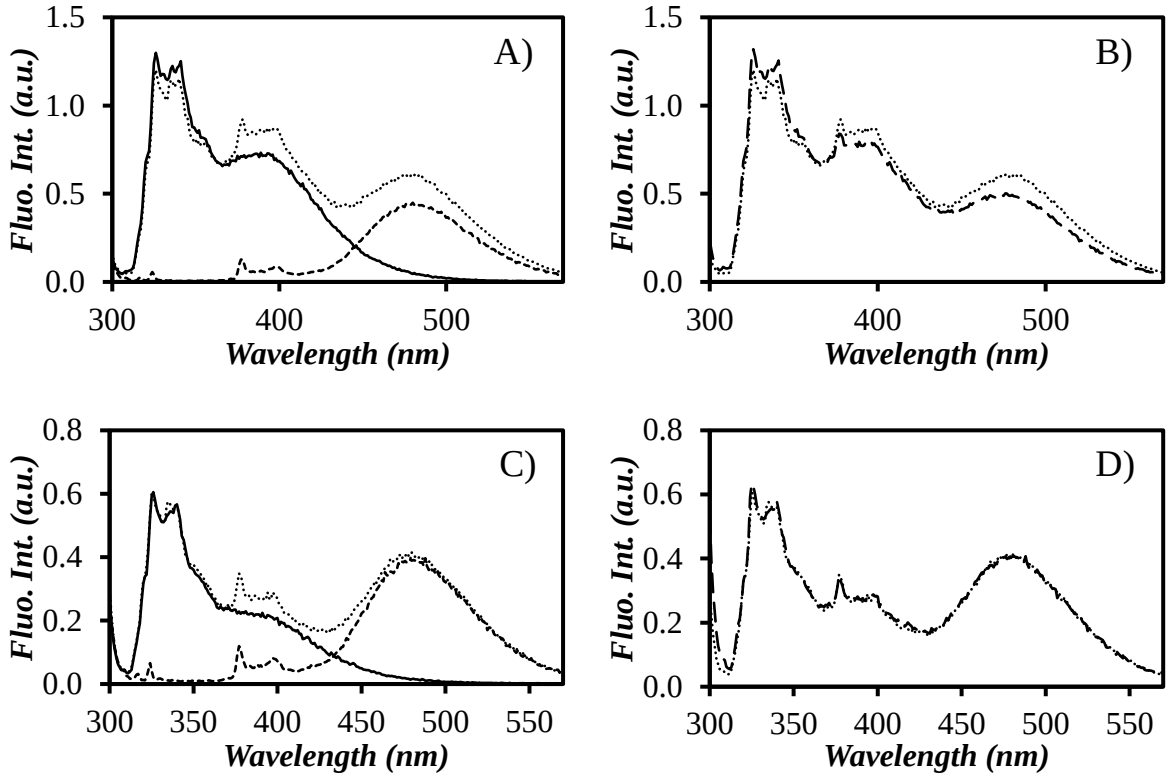


Figure S4.3: Steady-state fluorescence spectra of (A,C) (- - -) Py-P*L*L alone, (−) Np-P*L*L alone, and (⋯) a Py-P*L*L and Np-P*L*L mixture in a 90:10 ACN:water solution. (B,D) Comparison of (⋯) the experimental fluorescence spectrum of the mixture of Py-P*L*L and Np-P*L*L and (− −) the theoretical fluorescence spectrum resulting from the addition of the individual Py-P*L*L and Np-P*L*L spectra. The fluorescence spectra were acquired (A,B) without and (C,D) with a 50-fold excess of unlabeled P*L*L. λ_{ex}= 295 nm, [Np] = 15.7 μM, x_{Np} = 0.09, [Py] = 2.7 μM, x_{Py} = 0.12.

Steady-state fluorescence spectra of Py-P*LL*s in a 90:10 ACN:water mixture

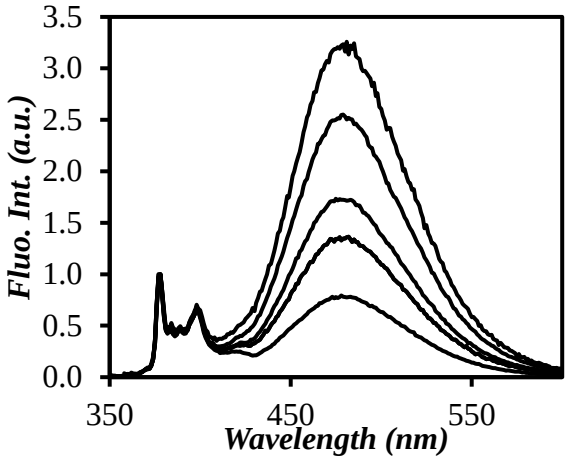


Figure S4.4. Steady-state fluorescence spectra of Py-P*LL* in a 90:10 ACN:water mixture. $[Py] = 2.7\times10^{-6}$ M.

Equations used to analyze the fluorescence decays of Py-P*LL*

The monomer and excimer decays were globally analyzed using Equations S4.1 and S4.2 for the monomer and excimer decays, respectively.

$$[Py^*]_{(t)} = [Py^*_{diff}]_{(t)} + [Py^*_{k2}]_{(t)} + [Py^*_{free}]_{(t)} = [Py_{diff}]_o \exp\left(-\left(A_2 + \frac{1}{\tau_M}\right)t - A_3\left(1-\exp(-A_4 t)\right)\right)$$

$$+\left([Py_{k2}]_o + [Py_{diff}]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\right)\exp\left(-\left(k_2+\frac{1}{\tau_M}\right)t\right)$$

$$-[Py_{diff}]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\exp\left(-\left(A_2+iA_4+\frac{1}{\tau_M}\right)t\right)$$

$$+[Py_{free}]_o\exp\left(-\frac{t}{\tau_M}\right)\Bigg] \qquad \text{(S4.1)}$$

$$[E^*]_{(t)}=[E0^*]_{(t)}+[D^*]_{(t)}=k_2\Bigg(\Bigg([Py_{k2}(E0)]_o+[Py_{diff}(E0)]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\Bigg)$$

$$\times\frac{\exp\left(-\frac{t}{\tau_{E0}}\right)-\exp\left(-\left(k_2+\frac{1}{\tau_M}\right)t\right)}{k_2+\frac{1}{\tau_M}-\frac{1}{\tau_{E0}}}$$

$$+[Py_{diff}(E0)]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\frac{\exp\left(-\left(A_2+iA_4+\frac{1}{\tau_M}\right)t\right)-\exp\left(-\frac{t}{\tau_{E0}}\right)}{A_2+iA_4+\frac{1}{\tau_M}-\frac{1}{\tau_{E0}}}\Bigg)$$

$$+k_2\Bigg(\Bigg([Py_{k2}(D)]_o+[Py_{diff}(D)]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\Bigg)\times\frac{\exp\left(-\frac{t}{\tau_D}\right)-\exp\left(-\left(k_2+\frac{1}{\tau_M}\right)t\right)}{k_2+\frac{1}{\tau_M}-\frac{1}{\tau_D}}$$

$$+[Py_{diff}(D)]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\frac{\exp\left(-\left(A_2+iA_4+\frac{1}{\tau_M}\right)t\right)-\exp\left(-\frac{t}{\tau_D}\right)}{A_2+iA_4+\frac{1}{\tau_M}-\frac{1}{\tau_D}}\Bigg)$$

$$+ [E0]_o \times \exp\left(-\frac{t}{\tau_{E0}}\right) + [D]_o \times \exp\left(-\frac{t}{\tau_D}\right)\Bigg] \quad \text{(S4.2)}$$

In Equations S4.1 and S4.2, the parameters A_2, A_3, and A_4 are given in Equations S4.3.a–c.

$$A_2 = < n > \times \frac{k_{blob} k_e [blob]}{k_{blob} + k_e [blob]} \quad \text{(S4.3.a)}$$

$$A_3 = < n > \times \left(\frac{k_{blob}}{k_{blob} + k_e [blob]}\right)^2 \quad \text{(S4.3.b)}$$

$$A_4 = k_{blob} + k_e [blob] \quad \text{(S4.3.c)}$$

The N_{blob} value was calculated using Equation S4.4, where $< n >$ is the average number of ground-state pyrenes per *blob* retrieved from the FBM analysis, x is molar fraction of lysines bearing a pyrene label, and f_{Mfree} is the fraction of pyrenyl labels that emit with its unquenched lifetime (τ_M) in the monomer decay.

$$N_{blob} = \frac{< n >}{x}\left(1 - f_{Mfree}\right) \quad \text{(S4.4)}$$

Parameters retrieved from the FBM analysis of Py-P*L*L in ACN:water mixtures

Table S4.1. Parameters retrieved using the FBM analysis of the pyrene monomer decays of pyrene-labeled poly(*L*-lysine) in acidified ACN:water mixtures.

Conditions	x (mol%)	k_2 (ns^{-1})	f_{Mk2}	$k_e[blob]$ (μs^{-1})	f_{mdiff}	k_{blob} (μs^{-1})	$<n>$	f_{Mfree}	χ^2
60:40 ACN:water 50-fold P*L*L τ_M = 215 ns	5.5	0.20	0.25	4.0	0.40	17.9	1.46	0.36	1.26
	7.6	0.20	0.31	4.4	0.44	18.0	1.49	0.25	1.07
	7.4	0.20	0.40	3.6	0.45	16.8	1.56	0.15	1.17
	10.3	0.20	0.45	4.1	0.45	19.3	1.56	0.10	1.12
	12.1	0.20	0.51	3.6	0.41	17.5	1.71	0.08	1.25
90:10 ACN:water 50-fold P*L*L τ_M = 203 ns	5.5	0.19	0.41	4.2	0.55	13.6	1.56	0.04	1.19
	7.6	0.19	0.52	4.7	0.47	12.8	1.90	0.01	1.21
	7.4	0.19	0.55	4.1	0.44	13.8	1.93	0.02	1.20
	10.3	0.19	0.64	3.4	0.36	11.8	2.50	0.00	1.28
	12.1	0.19	0.68	2.6	0.32	11.2	2.98	0.00	1.26
90:10 ACN:water τ_M = 203 ns	5.5	0.34	0.34	4.4	0.49	19.0	1.99	0.01	1.11
	7.6	0.34	0.34	6.9	0.46	21.2	2.91	0.01	1.11
	7.4	0.34	0.34	5.1	0.44	22.5	2.77	0.01	1.07
	10.3	0.34	0.34	6.0	0.31	24.6	3.27	0.01	1.16
	12.1	0.34	0.34	5.9	0.30	18.5	4.97	0.01	1.04

Table S4.2. Parameters retrieved using the FBM analysis of the pyrene excimer decays of pyrene-labeled poly(*L*-lysine) in acidified ACN:water mixtures.

Conditions	x (mol%)	f_{Ek2}	τ_{E0} (ns)	$f_{EdiffE0}$	f_{EE0}	τ_{D} (ns)	f_{EdiffD}	f_{ED}	χ^2
60:40 ACN:water 50-fold P*L*L	5.5	0.27	38.0	0.26	0.29	67.7	0.18	0.00	1.26
	7.6	0.27	38.7	0.26	0.33	66.0	0.14	0.00	1.07
	7.4	0.29	40.5	0.19	0.37	63.0	0.14	0.01	1.17
	10.3	0.29	40.5	0.13	0.42	58.4	0.16	0.00	1.12
	12.1	0.31	38.6	0.11	0.31	58.3	0.14	0.14	1.25
90:10 ACN:water 50-fold P*L*L	5.5	0.24	30.2	0.22	0.00	49.3	0.10	0.45	1.19
	7.6	0.28	30.3	0.05	0.30	50.6	0.20	0.17	1.21
	7.4	0.27	23.4	0.01	0.24	45.4	0.21	0.27	1.20
	10.3	0.26	26.0	0.06	0.15	50.7	0.09	0.44	1.28
	12.1	0.25	27.7	0.00	0.51	49.0	0.12	0.12	1.26
90:10 ACN:water	5.5	0.38	36.6	0.14	0.22	56.7	0.23	0.04	1.11
	7.6	0.20	33.0	0.03	0.57	58.3	0.14	0.05	1.11
	7.4	0.31	36.9	0.04	0.42	57.9	0.20	0.03	1.07
	10.3	0.28	30.1	0.01	0.49	55.2	0.12	0.10	1.16
	12.1	0.10	27.7	0.00	0.63	54.6	0.04	0.23	1.04

Table S4.3. Fraction of pyrene species calculated using parameters retrieved from the FBM analysis of the pyrene monomer and excimer decays of pyrene-labeled poly(*L*-lysine) in acidified ACN:water mixtures.

Conditions	x (mol%)	f_{k2}	f_{diffE0}	f_{diffD}	f_{diff}	f_{aggE0}	f_{aggD}	f_{agg}	f_{free}	χ^2
60:40 ACN:water 50-fold P*L*L	5.5	0.19	0.18	0.13	0.31	0.21	0.00	0.21	0.28	1.26
	7.6	0.22	0.21	0.11	0.32	0.27	0.00	0.27	0.18	1.07
	7.4	0.27	0.17	0.12	0.30	0.33	0.01	0.34	0.10	1.17
	10.3	0.27	0.12	0.15	0.27	0.40	0.00	0.40	0.06	1.12
	12.1	0.29	0.10	0.14	0.24	0.29	0.13	0.42	0.05	1.25
90:10 ACN:water 50-fold P*L*L	5.5	0.23	0.22	0.09	0.31	0.00	0.44	0.44	0.02	1.19
	7.6	0.28	0.05	0.20	0.25	0.30	0.16	0.46	0.01	1.21
	7.4	0.27	0.01	0.21	0.22	0.24	0.27	0.51	0.01	1.20
	10.3	0.26	0.06	0.09	0.15	0.15	0.44	0.59	0.00	1.28
	12.1	0.25	0.00	0.12	0.12	0.51	0.12	0.63	0.00	1.26
90:10 ACN:water	5.5	0.37	0.13	0.23	0.37	0.21	0.04	0.25	0.01	1.11
	7.6	0.20	0.03	0.14	0.18	0.56	0.05	0.62	0.00	1.11
	7.4	0.31	0.04	0.20	0.25	0.41	0.03	0.44	0.00	1.07
	10.3	0.28	0.01	0.12	0.13	0.49	0.10	0.59	0.00	1.16
	12.1	0.10	0.00	0.04	0.04	0.63	0.23	0.86	0.00	1.04

Width determination of P*LL* bundles

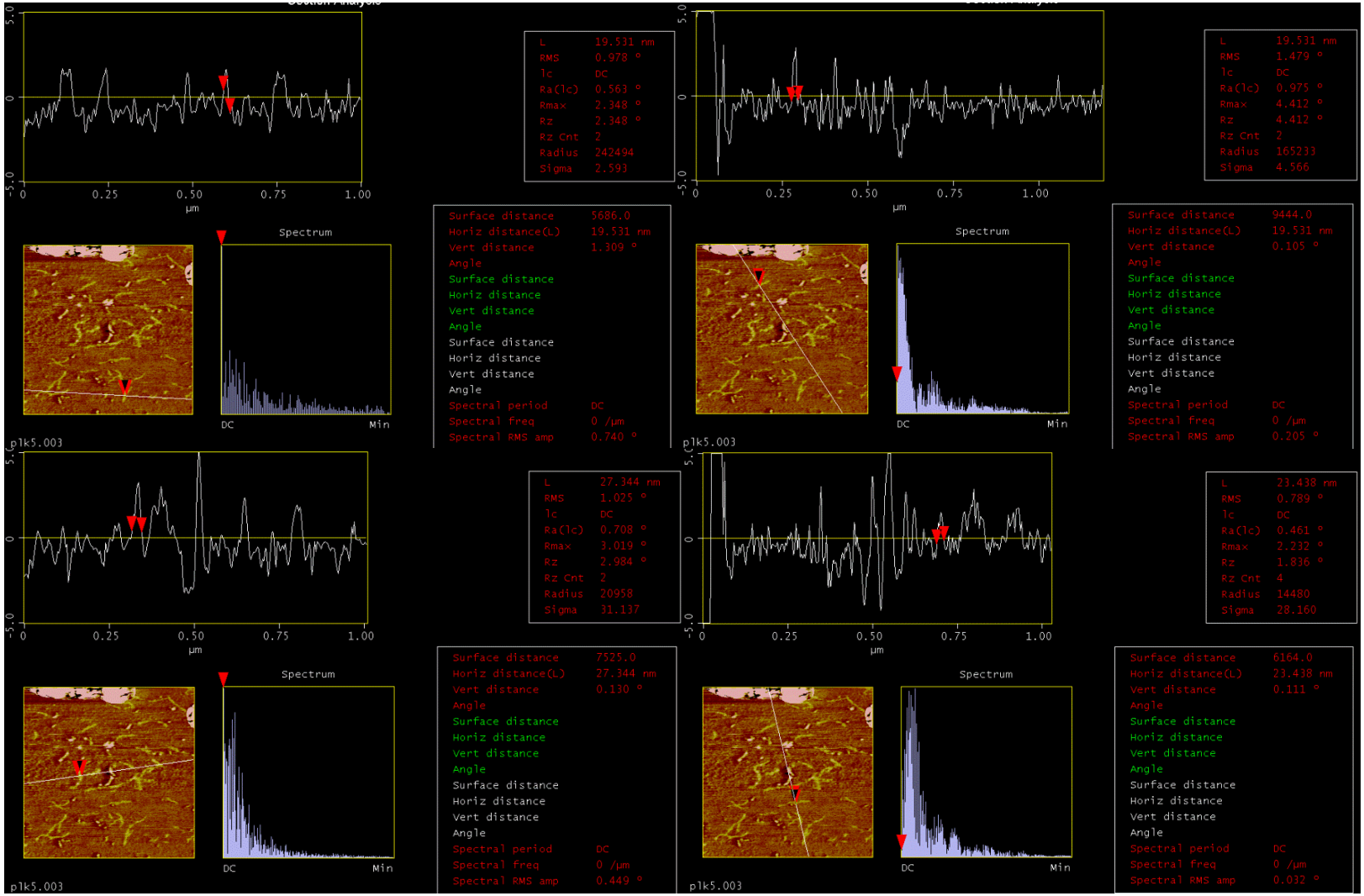


Figure S4.5. AFM phase images used to determine the width of P*LL* bundles

Molecular Mechanic Optimizations of Intermolecular PEF

The procedure described herein was adapted from a previously published procedure.[1] A P*L*L α-helix containing 55 lysines was constructed in HyperChem (phi = − 58 °, psi = − 47 °, omega = 180 °). The helix was then aligned with the z-axis. This helix, referred to as the primary helix, was used as a reference for the remaining procedure. Figure S4.6 outlines the orientation of the reference helix in a HCP lattice. Each neighboring helix was defined by two angles θ and ϕ and the interhelical distance $d_{\text{h-h}}$. The angle θ defines the position of the neighboring helix around the reference helix and the angle ϕ defines the relative orientation of the helix in a given position.

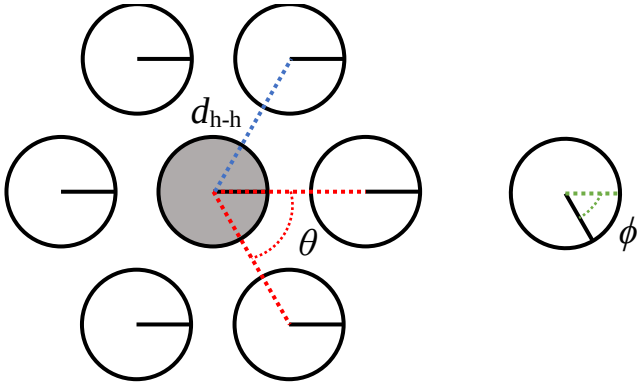


Figure S4.6. Schematic representation of the (grey) reference Py-P*L*L helix in a HCP lattice with an interhelical distance $d_{\text{h-h}}$. The angle θ defines the position of the (white) neighboring helices in the HCP lattice. The lines in the circles represent the relative orientation of each helix and the angle ϕ defines the deviation from this relative orientation.

Rather than constructing the entire lattice of P*L*L helices at once, the MMOs were individually conducted between the reference helix and each of the neighboring helices. To determine if the angle ϕ was an important factor in the MMOs, the $\Delta N_{\text{blob}}^{\text{theo}}$ was determined

for a neighboring helix located at $d_{h\text{-}h}$ = 3.0 nm and θ = 0 ° as a function of ϕ. To this end, a second helix was constructed and aligned with the *z*-axis in HyperChem, the helix was then rotated by an angle ϕ about the *z*-axis and translated 3.0 nm in the *y*-direction. Pyrene was then labeled onto the 27th residue from the *N*-terminus of the primary helix. The second helix was then also labeled with a pyrene on the 27th residue. The two pyrenes were then brought together such that their planes were parallel and 3.4 Å apart from each other. If seven or more carbon atoms of one pyrene overlapped with the frame of the other pyrene, the pyrenes were considered to have sufficient overlap to form excimer. The position of the pyrene on the second helix was moved sequentially along the backbone (both towards the *N*- and *C*-terminus) to determine if excimer could be formed between the reference and the second helix. The total number of pyrenes that could from excimer were then tallied, yielding ΔN_{blob}^{theo} for the given ϕ. Figure S4.7A displays the determined ΔN_{blob}^{theo} values for the second helix with $d_{h\text{-}h}$ = 3.0 nm and θ = 0 ° as a function of ϕ in 60 ° increments. Due to the symmetry of the helix, ΔN_{blob}^{theo} was only calculated up to a ϕ angle of 180 °. The values of ϕ = 180 – 360 ° were taken as the mirror of the ϕ = 0 – 180 ° values. Figure S4.7A shows that ΔN_{blob}^{theo} varied by at most 1 as a function of ϕ. This small variation indicated that ϕ had a relatively minor contribution to ΔN_{blob}^{theo}, and therefore only ϕ = 0 ° was considered for the remaining MMOs. The validity of this assumption was also supported by a previous publication,[1] which also demonstrated that ϕ had only minor effects on ΔN_{blob}^{theo}.

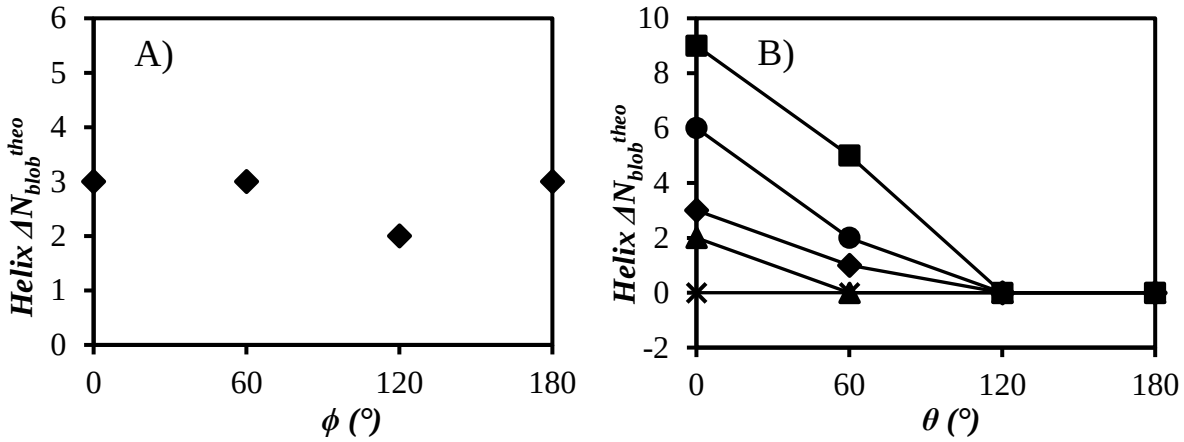


Figure S4.7. A) Plot of ΔN_{blob}^{theo} as a function of ϕ between the reference helix and a second helix located at an interhelical distance of 3.0 nm and a θ angle of 0 °. B) Plot of ΔN_{blob}^{theo} between the reference helix and a helix located at angles θ and ϕ = 0 ° with a $d_{h\text{-}h}$ of (■) 2.8, (●) 2.9, (◆) 3.0, (▲) 3.1 and (×) 3.2 nm. The lines were added to guide the eyes for the $d_{h\text{-}h}$ trends.

The ΔN_{blob}^{theo} between the remaining secondary helix locations, defined by the angle θ and the distance $d_{h\text{-}h}$, were determined as follows. The reference helix was constructed following the instructions outlined above. A second helix was then constructed in a similar manner. The second helix was rotated by θ and translated by a distance $d_{h\text{-}h}$ in the *y*-direction. The reference helix was then rotated by the angle θ, to yield the two helices with ϕ = 0 °. The secondary helices located at θ = 240 and 300 ° were symmetrical to the helices located at θ = 60 and 120 °, respectively. Therefore, the ΔN_{blob}^{theo} values for θ = 240 and 300 ° were set to equal the ΔN_{blob}^{theo} values of θ = 60 and 120 °, respectively. Figure S4.7B displays a plot of ΔN_{blob}^{theo} as a function of θ for $d_{h\text{-}h}$ values ranging from 2.8 to 3.2 nm. The ΔN_{blob}^{theo} for Py-

P*L*L in the helical bundle with interhelical distances of $d_{h\text{-}h}$ was calculated as the weighted average of the ΔN_{blob}^{theo} values between the individual secondary helices corresponding to a given interhelical distance. The ΔN_{blob}^{theo} values of the helices located at θ = 60 and 120 ° were counted twice to account for the helices at θ = 240 and 300 °. For example, ΔN_{blob}^{theo} of Py-P*L*L in a bundle with an interhelical distance of 2.8 nm was calculated as (9 + 2×5 + 2×0 + 0) = 19. Since the Py-P*L*L helix had an N_{blob}^{theo} value of 27 from intramolecular PEF, the total N_{blob}^{theo} expected for Py-P*L*L in a HCP bundle with a $d_{h\text{-}h}$ of 2.8 nm was 19 + 27 = 46.

Density of Py-P*L*L Bundles

The density of P*L*Ls within the lattice is defined as the mass of P*L*L per unit volume. The outline of a unit cell from helices packed in a HCP arrangement is given in Figure S4.8.

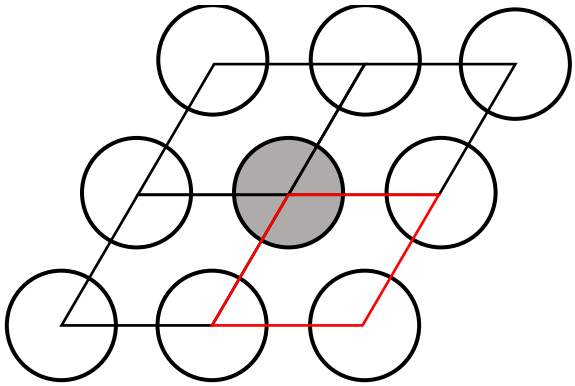


Figure S4.8. Schematic representation of the unit cells of a HCP arrangement. The red parallelogram highlights a single unit cell.

The volume of a unit cell V_{cell} is equal to the product of the cell area in Figure S4.8 and its depth *l*. The cell area is defined by the product of its length (equal to $d_{h\text{-}h}$) and its height

$d_{h-h} \times \sin(60\,°)$. Figure S4.8 shows that each unit cell contains the cumulative cross-sectional area equal to one P*L*L helix. Therefore, the mass in a unit cell (M_{PLL}) is equal to the mass of lysine per unit length in an α-helix multiplied by the depth of a unit cell l. The mass of lysine per unit length in a helix is equivalent to the mass of one lysine unit (M_L/N_A) divided by the raise per lysine in the helix (Δh_L = 0.14 nm). Combining M_{PLL} and V_{cell}, the density ρ in Equation S4.5 yields the density of P*L*L in the HCP lattice separated with an interhelical distance of $d_{h\text{-}h}$.

$$\rho = \frac{M_{PLL}}{V_{cell}} = \frac{\left(\frac{1\ \ helix}{cell}\right) \times 1 \quad cell \times M_L \times \left(\frac{1}{N_A}\right) \times \left(\frac{1}{\Delta h_L}\right) \times l}{\left(d_{h-h}\right)^2 \times \sin\left(60°\right) \times l}$$

$$= \frac{\frac{M_L}{N_A \Delta h_L}}{\frac{\sqrt{3}}{2}\left(d_{h-h}\right)^2} = \frac{2M_L}{\sqrt{3} N_A \Delta h_L {d_{h-h}}^2} \qquad \text{(S4.5)}$$

Table S4.4. Intrinsic viscosity[2] and density of proteins

Protein	M_n (g/mol)	$[\eta]$ (mL/g)	ρ_P (g/mL)
Aldolase	150000	4.0	0.63
aS2-casein	23000	11.3	0.22
α-globulin	300000	3.5	0.71
Angiotensin converting enzyme	200000	4.4	0.57
Arachin (11S groundnut globulin)	300000	4.5	0.56
Aspartate transcarbamylase	307000	4.5	0.56
Brassin M (oilseed rape 11S globulin)	300000	3.8	0.66
Brassin R (mustard seed 11S (globulin)	300000	3.8	0.66
Carmin (11S safflower seed globulin)	30000	3.9	0.64
Catalase	250000	3.9	0.64
Chymotrypsinogen A	25660	2.5	1.00
Clathrin	180000	10.0	0.25

Conalbumin	75500	3.5	0.71
Conbrassin M (2S mustard seed globulin)	15000	5.4	0.46
Conbrassin R (2S rapeseed globulin)	15000	7.3	0.34
Concarmin (2S safflower seed globulin)	15000	6.5	0.38
Conhelianthin (2S sunflower seed globulin)	15000	7.3	0.34
Consesamin (2S sesame seed globulin)	15000	4.1	0.61
Cytochrome C		2.7	0.94
Fetuin	48000	7.0	0.36
Glycinin (11S soybean globulin)	300000	4.6	0.54
Gossypin (11S cottonseed globulin)	300000	4.0	0.63
Helianthin (11S sunflower seed globulin)	300000	3.8	0.67
Hemocyanin	3750000	5.6	0.45
Hemoglobin	68000	3.6	0.69
Lactate dehydeogenase	138000	3.9	0.64
Linin (11S linseed globulin)	300000	3.7	0.68
Lupin protein isolate	390000	6.8	0.37
Lysozyme	14320	2.7	0.93
Metamerythrin	107000	3.6	0.69
Mucin peptide		7.7	0.32
Myoglobin	17190	3.3	0.77
Myosin S1 head	110000	6.4	0.39
Neurophysin monomers	10000	5.5	0.45
Ovalbumin	45000	5.5	0.46
Plasminogen	143000	8.0	0.31
PRN60 glycoprotein of the retrovirus feline leukemia	6300	7.5	0.33
Prothrombin	18500	4.1	0.61
Ribonuclease A	13700	3.3	0.76
Serum Albumin	65000	3.7	0.68
Thrombin	33400	3.8	0.66

S5 – SI for Chapter 5

Solvent drying

Ethyl acetate: Ethyl acetate (500 mL) was refluxed with calcium hydride (2.5 g) for 3 hours, then distilled under a nitrogen atmosphere. The first ca. 15 mL of distilled ethyl acetate was discarded. The distilled ethyl acetate was stored under nitrogen until used.

THF: THF was dried in a similar procedure as ethyl acetate.

Dioxane: Dioxane (100 mL) was pre-dried by mixing with CaH_2 (0.2 g) for 2 hours. The dioxane was then refluxed with sodium (ca. 0.1 g) for 3 hours. The dioxane was then distilled under reduced pressure at 40 °C and was collected in a flame-dried Schlenk tube. The first 10 mL of distillate was collected in a separate ampule and discarded. The distilled dioxane was stored under nitrogen in the dark until used.

DMF: DMF (200 mL) was stirred with CaH_2 (0.4 g, 5 wt%) overnight in the dark under a nitrogen atmosphere. The calcium hydride was removed by filtration under a nitrogen atmosphere. The DMF was then distilled under reduced pressure at a temperature no greater than 35 °C to minimize degradation. The DMF was collected into a flame-dried Schlenk tube covered with aluminum foil to minimize light exposure. The first 15 mL of distillate was collected in a separate ampule and discarded. The distilled DMF was stored under nitrogen in the dark until used.

n-Butylamine: *n*-Butylamine (20 mL) was refluxed with CaH_2 (0.1 g) for 2 hours before it was distilled under vacuum at 40 °C into a Schlenk flask. The first 3 mL of distillate were discarded. The distilled *n*-butylamine was stored under nitrogen.

NMR: Amino acid NCAs

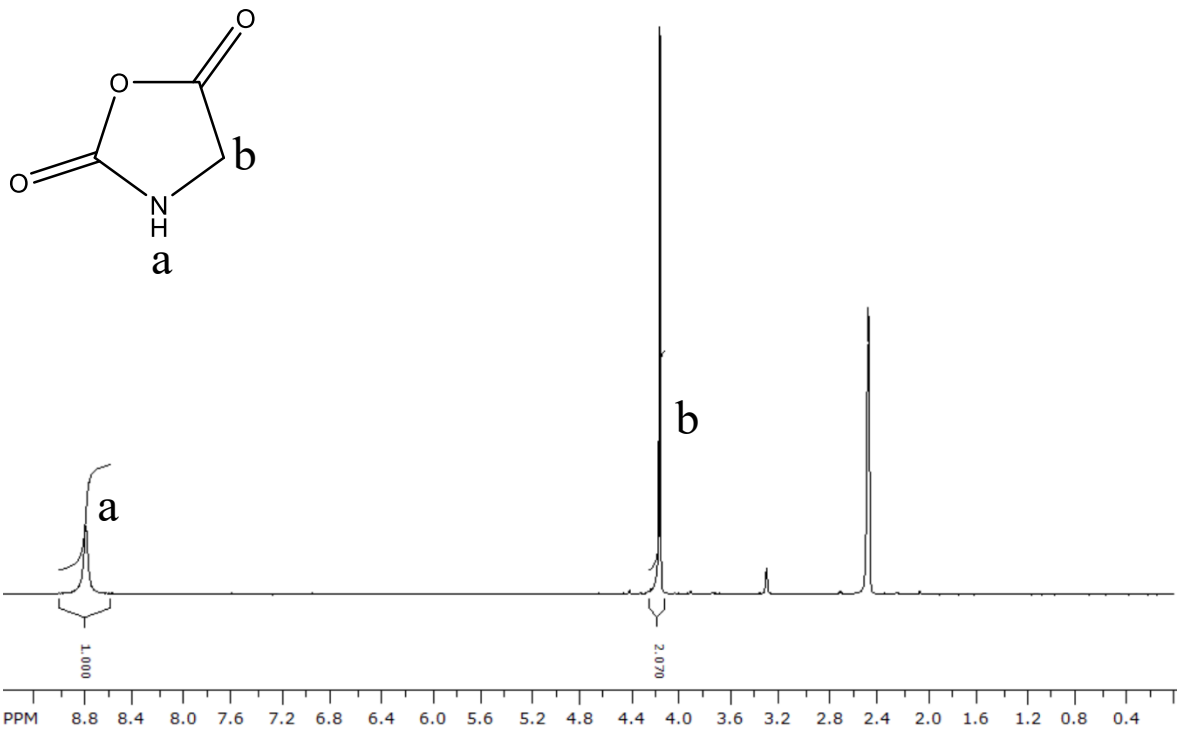


Figure S5.1. ^{1}H NMR spectrum of GlyNCA. 300 MHz, d_6-DMSO: δ = 8.80 (s, 1H), 4.16 (s, 2H) ppm. Residual water (3.29 ppm) and solvent (2.47 ppm) peaks are also present.

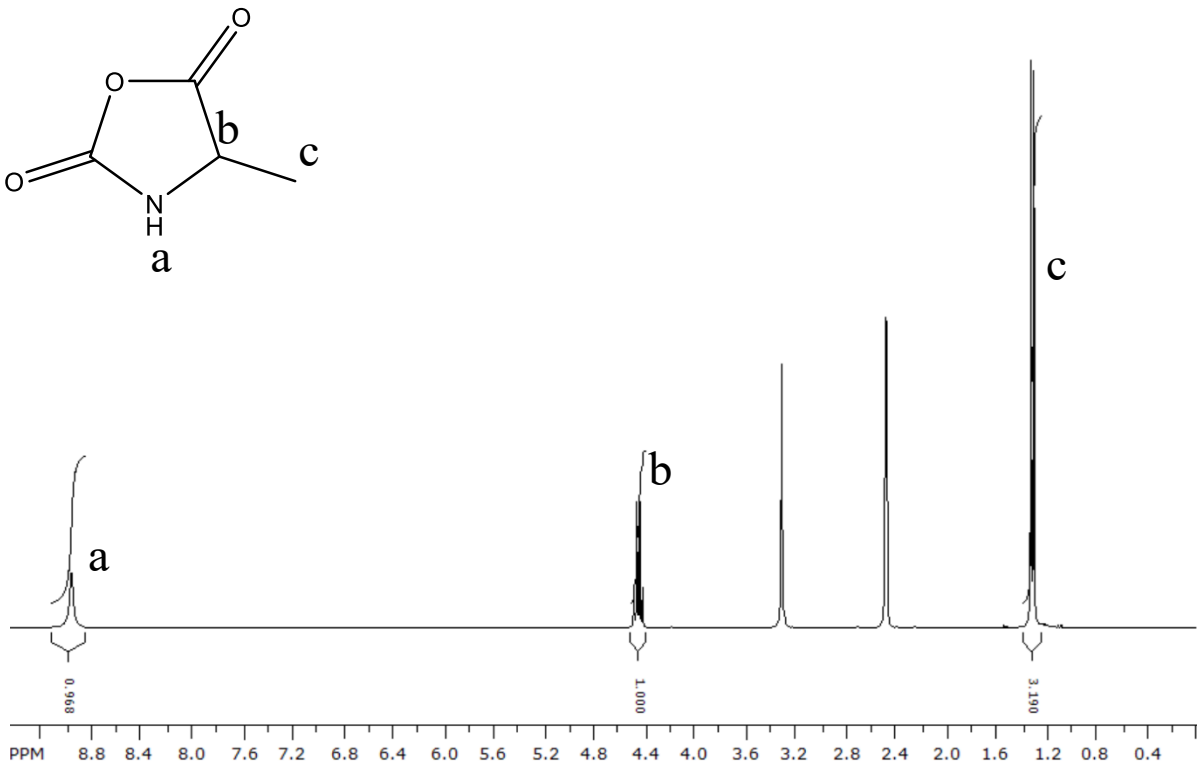


Figure S5.2. ^{1}H NMR spectrum of *D,L*-AlaNCA. 300 MHz, d_6-DMSO: δ = 8.98 (s, 1H), 4.45 (q, 1H), 1.30 (d, 3H) ppm. Residual water (3.31 ppm) and solvent (2.47 ppm) peaks are also present.

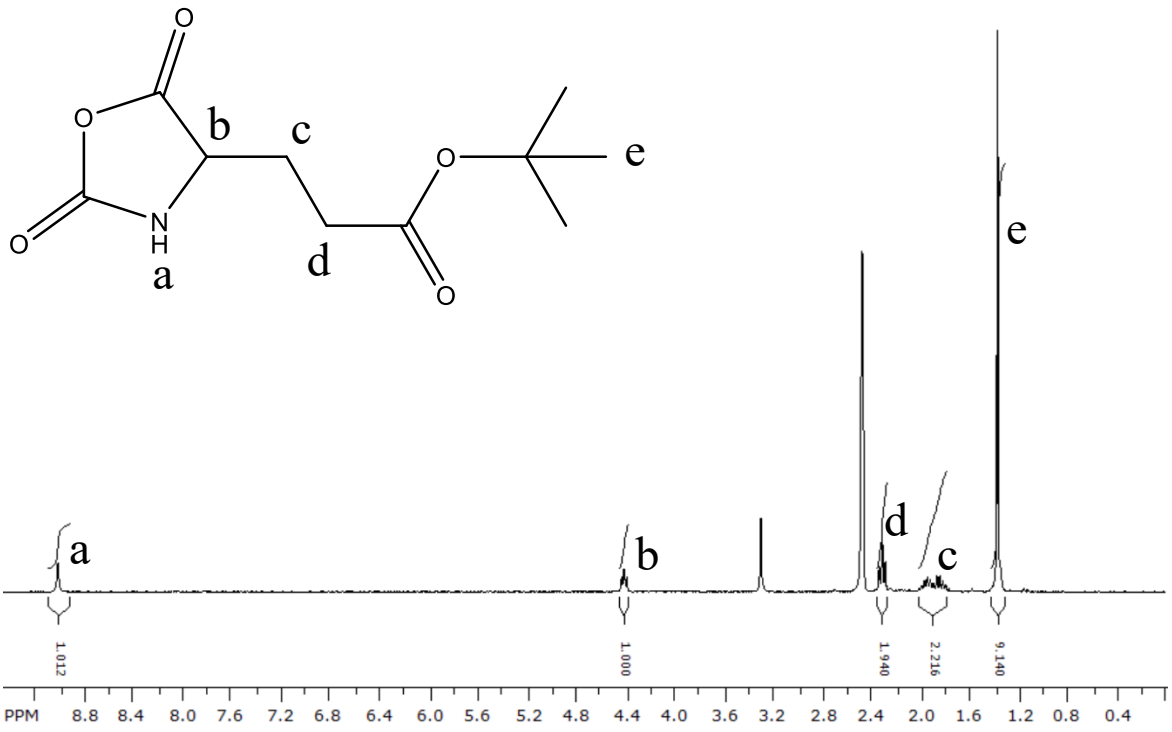


Figure S5.3. ^{1}H NMR spectrum of *D,L*-Glu(OtBu)NCA. 300 MHz, d_6-DMSO: δ = 9.03 (s, 1H), 4.41 (t, 1H), 2.31 (t, 2H), 2.00 – 1.79 (m, 2H), 1.37 (s, 9H) ppm. Residual water (3.29 ppm) and solvent (2.47 ppm) peaks are also present.

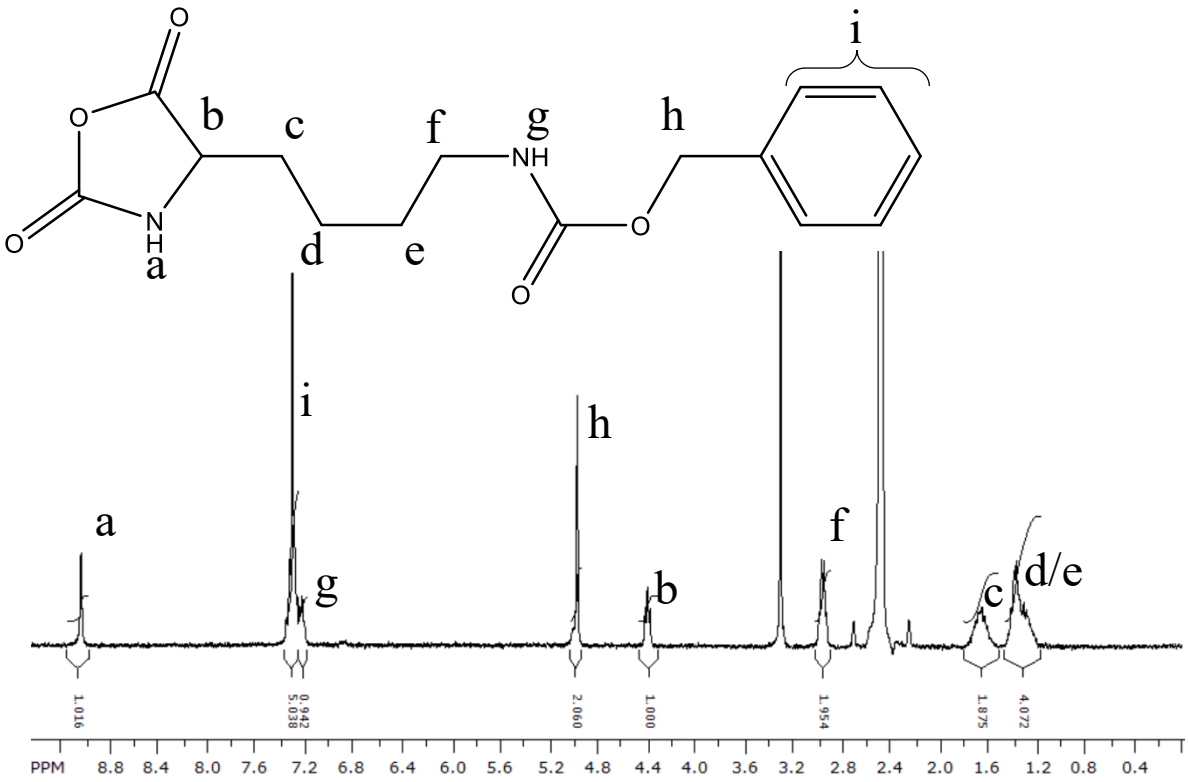


Figure S5.4. ^{1}H NMR spectrum of *D*-Lys(Z)NCA. 300 MHz, d_6-DMSO: δ = 9.05 (s, 1H), 7.31 (m, 5H), 7.23 (t, 1H), 4.97 (s, 2H), 4.39 (t, 1H), 2.95 (br q, 2H), 1.65 (m, 2H), 1.43 – 1.21 (m, 4H) ppm. Residual water (3.30 ppm) and solvent (2.47 ppm) peaks are also present.

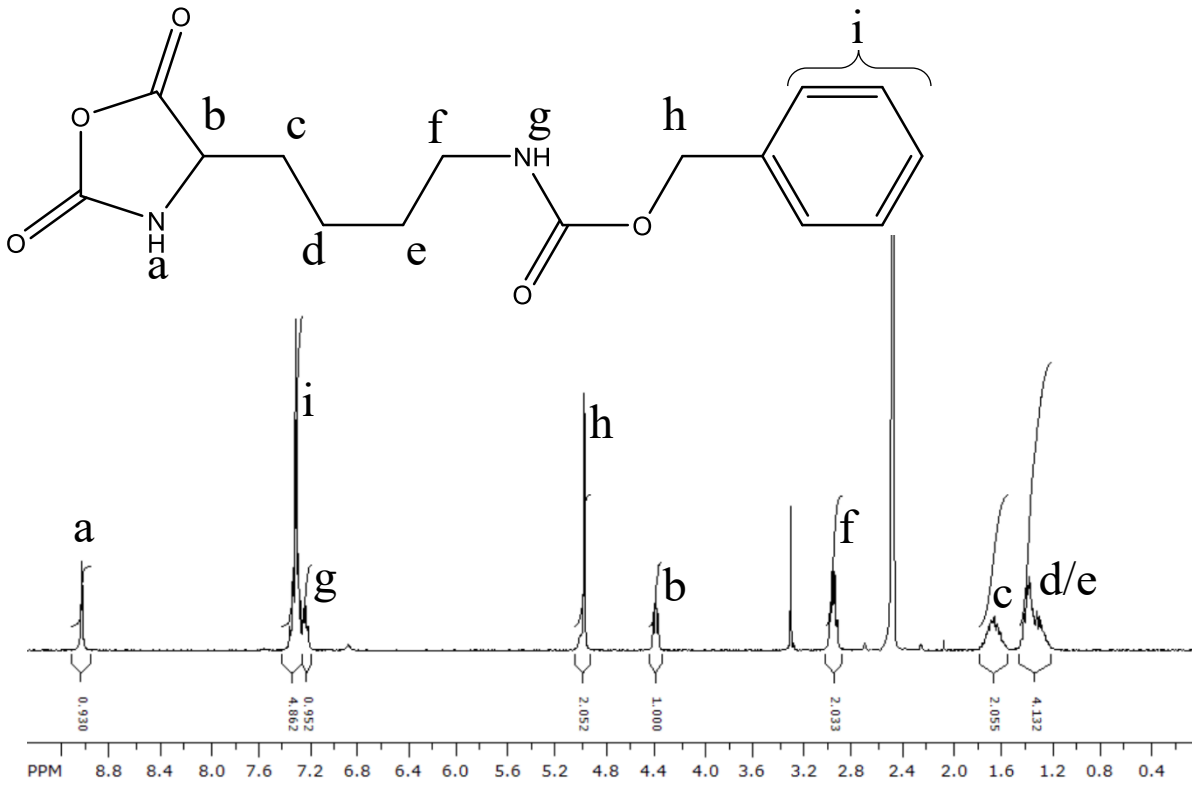


Figure S5.5. ^{1}H NMR spectrum of *L*-Lys(Z)NCA. 300 MHz, d_6-DMSO: δ = 9.05 (s, 1H), 7.31 (m, 5H), 7.23 (t, 1H), 4.97 (s, 2H), 4.39 (t, 1H), 2.95 (br q, 2H), 1.65 (m, 2H), 1.43 – 1.21 (m, 4H) ppm. Residual water (3.30 ppm) and solvent (2.47 ppm) peaks are also present.

NMR: Polypeptides

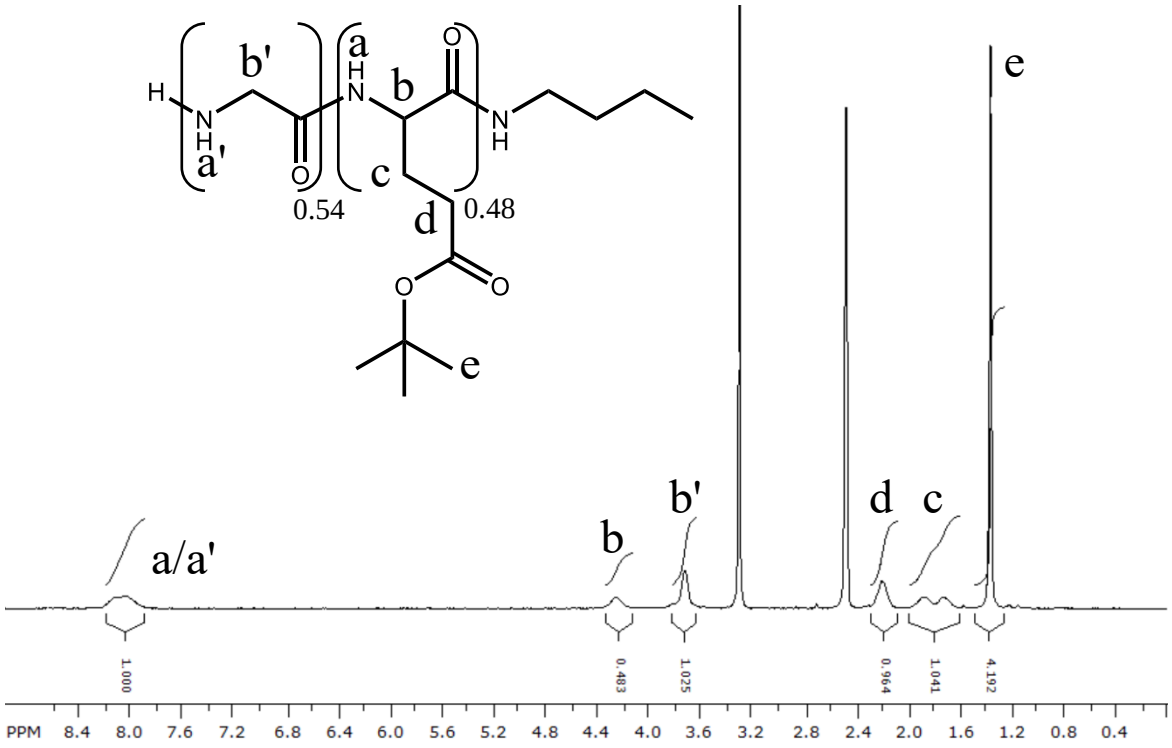


Figure S5.6. ^{1}H NMR spectrum of PGlyGlu(OtBu). 300 MHz, d_6-DMSO: δ = 8.09 (1H, amide), 4.25 (2H, Gly-α), 3.72 (1H, Glu-α), 2.23 (2H, γ-Glu), 1.88 – 1.73 (2H, β-Glu), 1.25 – 1.21 (m, 4H), 0.82 (t, 3H), and 1.36 (9H, tBu) ppm. Residual water (3.31 ppm) and solvent (2.47 ppm) peaks are also present.

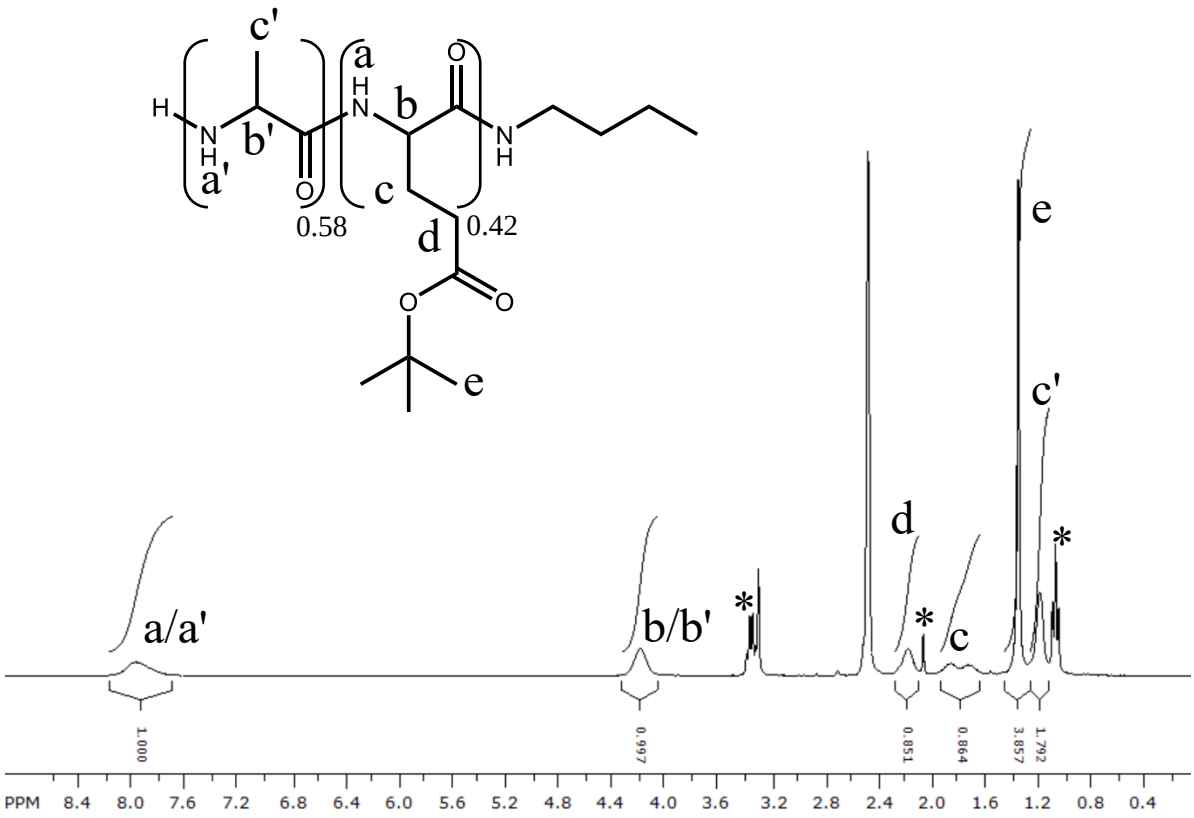


Figure S5.7. ^{1}H NMR spectrum of PAlaGlu(OtBu). 300 MHz, d_6-DMSO: δ = 8.00 (1H, amide), 4.20 (1H, α), 2.19 (2H, γ-Glu), 1.83 – 1.71 (2H, β-Glu), 1.33 (9H, tBu), 1.18 (3H, β-Ala). The peaks marked with an * are from residual diethyl ether (3.38 and 1.09 ppm) and acetone (2.09 ppm). Residual water (3.36 ppm) and solvent (2.47 ppm) peaks are also present.

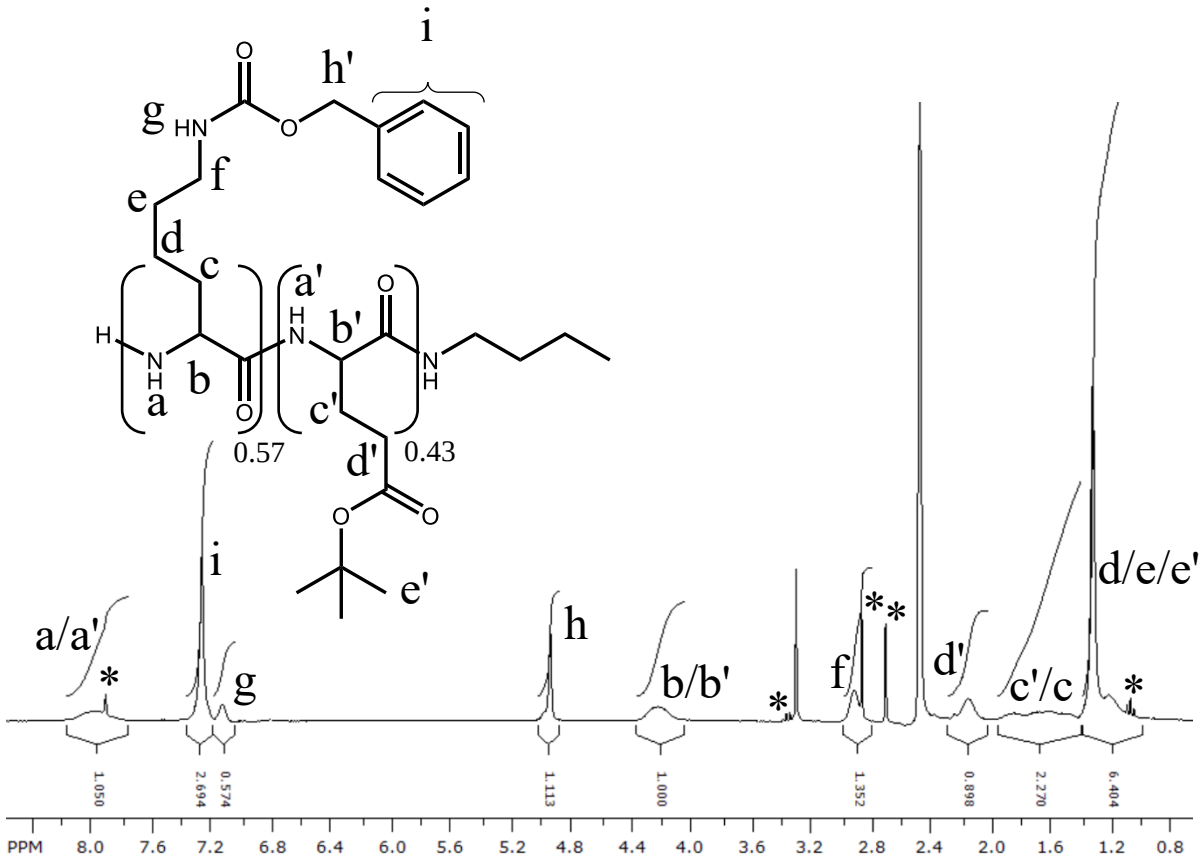


Figure S5.8. ^{1}H NMR spectrum of PLys(Z)Glu(OtBu). 300 MHz, d_6-DMSO: δ = 8.00 (1H, amide), 7.79 (5H, Bz), 7.14 (1H, ε-Lys amide), 4.95 (2H, Bz-CH_2), 4.23 (1H, α), 2.92 (2H, ε-Lys), 2.18 (2H, γ-Glu), 1.87 – 1.72 (2H, β-Glu), 1.58 – 1.48 (2H, β-Lys), 1.33 – 1.21 (4H, δ-Lys and γ-Lys), 1.31 (9H, tBu). The peaks marked with an * are from residual DMF (7.92, 2.86, and 2.70 ppm) and diethyl ether (3.38 and 1.09 ppm). Residual water (3.30 ppm) and solvent (2.47 ppm) peaks are also present.

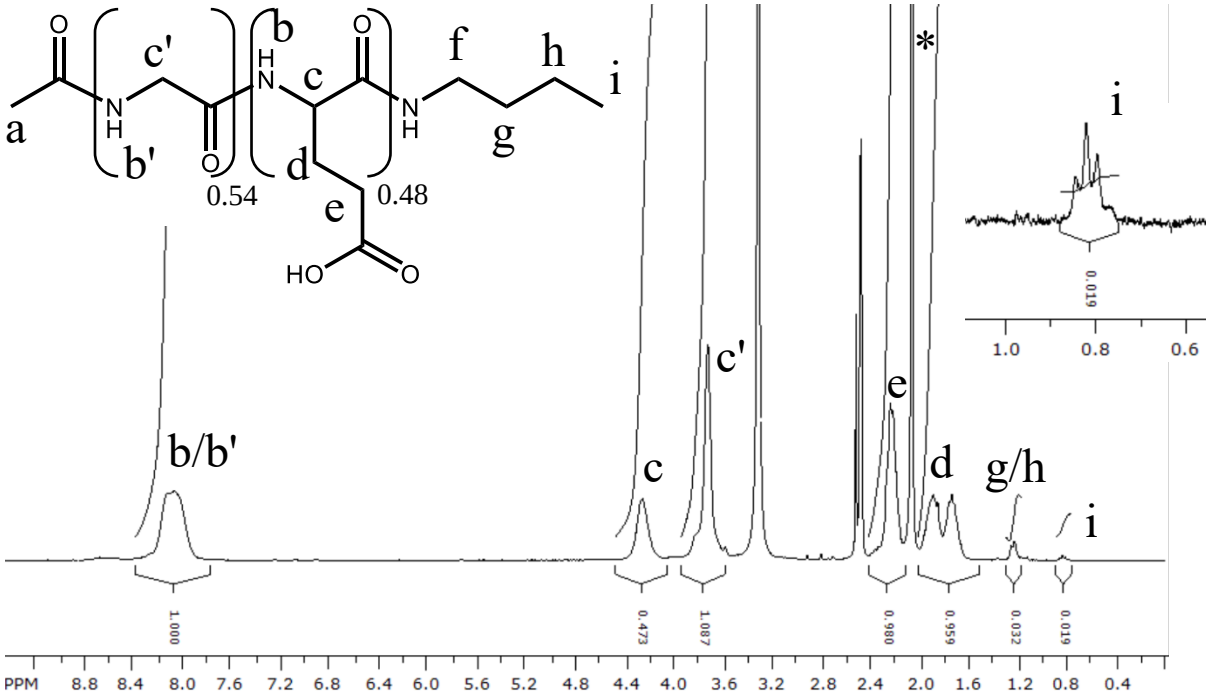


Figure S5.9. ^{1}H NMR spectrum of PGlyGlu. 300 MHz, d_6-DMSO: δ = 8.09 (1H, amide), 4.25 (2H, Gly-α), 3.72 (1H, Glu-α), 2.23 (2H, γ-Glu), 1.88 – 1.73 (2H, β-Glu), 1.25 – 1.21 (m, 4H), 0.82 (t, 3H) ppm. Protons ‘a’ (expected 1.73 ppm) coincide with ‘d’ while protons ‘f’ (expected 3.05 ppm) were not resolved. The peak marked with an * is residual acetone (2.05 ppm). Residual water (3.31 ppm) and solvent (2.47 ppm) peaks are also present.

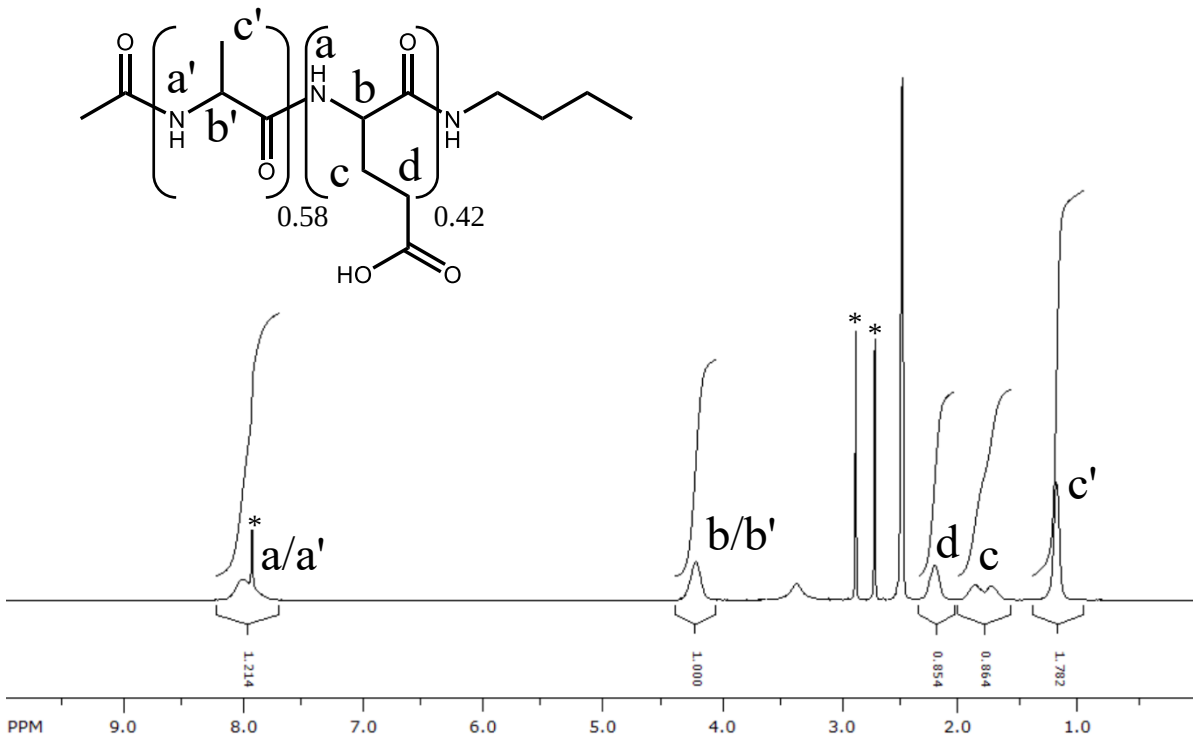


Figure S5.10. ^{1}H NMR spectrum of PAlaGlu. 300 MHz, d_6-DMSO: δ = 8.00 (1H, amide), 4.20 (1H, α), 2.19 (2H, γ-Glu), 1.83 – 1.71 (2H, β-Glu), 1.18 (3H, β-Ala). The peaks marked with an * are from residual DMF (7.92, 2.86, and 2.70 ppm). Residual water (3.36 ppm) and solvent (2.47 ppm) peaks are also present.

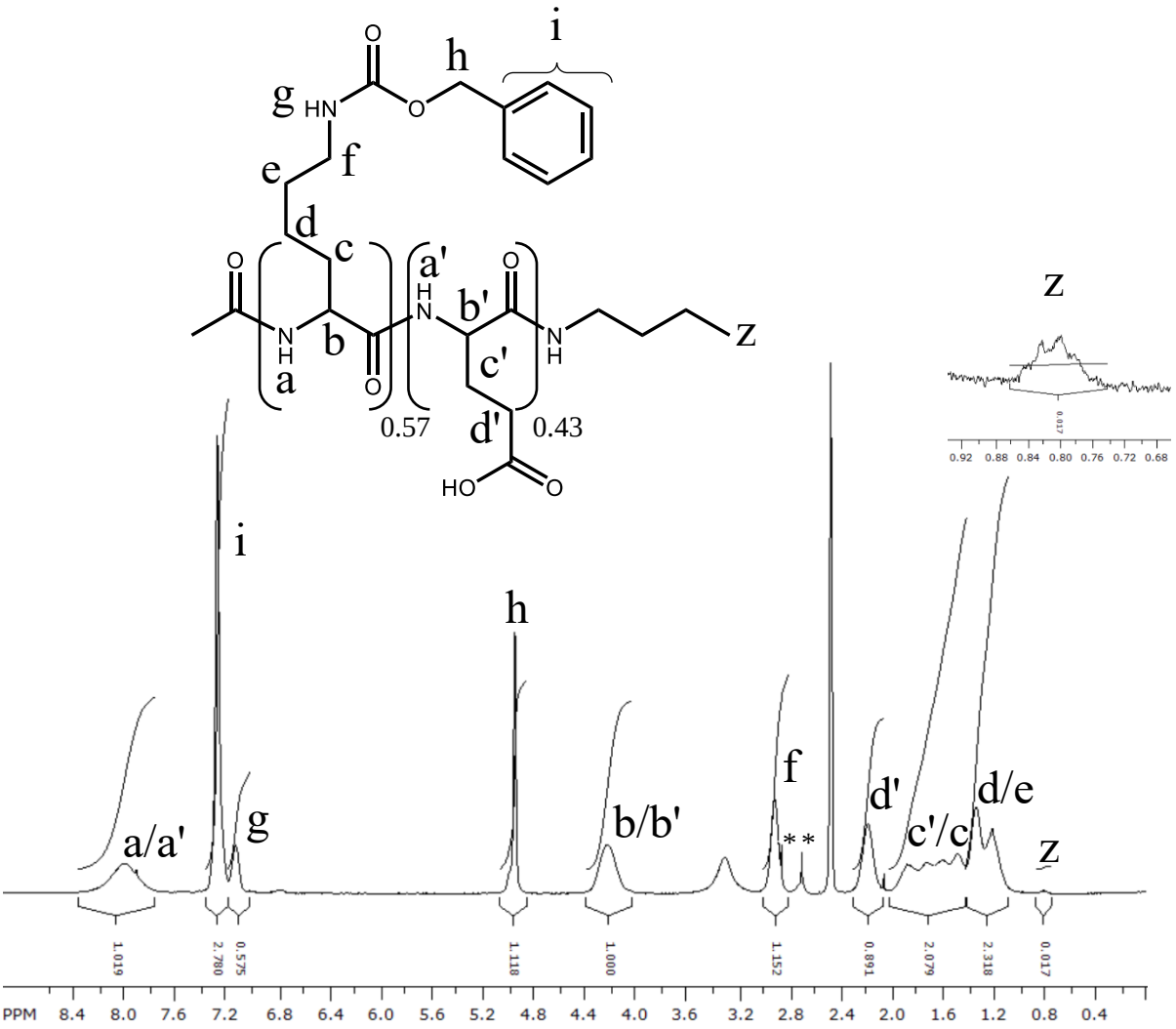


Figure S5.11. ^{1}H NMR spectrum of PLys(Z)Glu. 300 MHz, d_6-DMSO: δ = 8.00 (1H, amide), 7.79 (5H, Bz), 7.14 (1H, ε-Lys amide), 4.95 (2H, Bz-CH_2), 4.23 (1H, α), 2.92 (2H, ε-Lys), 2.18 (2H, γ-Glu), 1.87 – 1.72 (2H, β-Glu), 1.58 – 1.48 (2H, β-Lys), 1.33 (2H, δ-Lys), 1.21 (2H, γ-Lys), 0.8 (t, 3H). The peaks marked with an * are from residual DMF (7.92, 2.86, and 2.70 ppm). Residual water (3.30 ppm) and solvent (2.47 ppm) peaks are also present.

GPC traces and their analysis

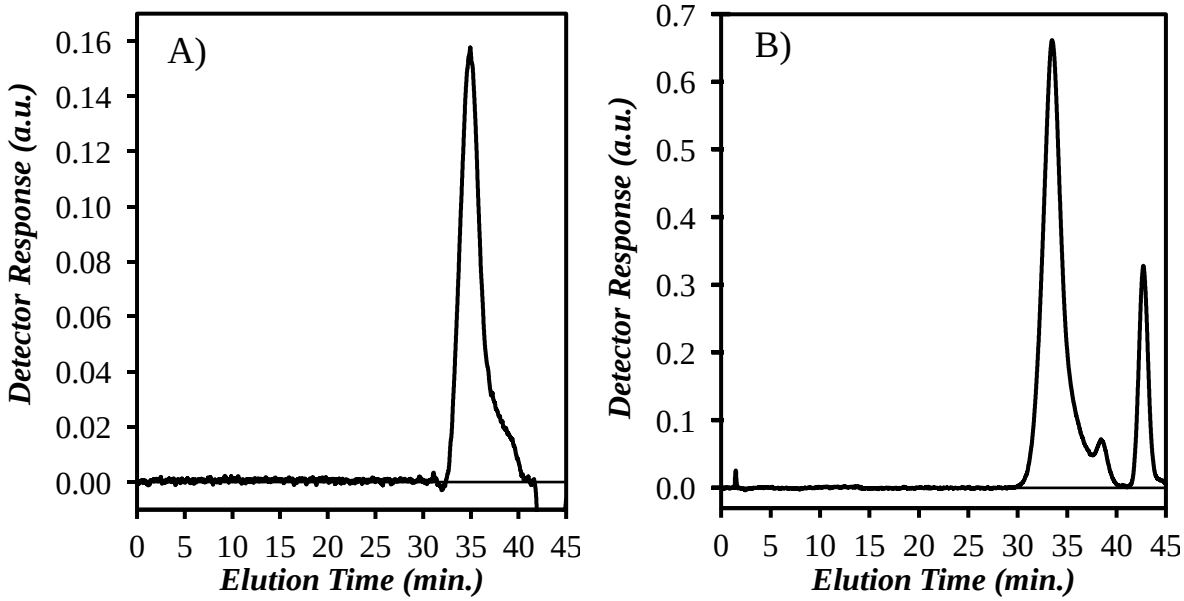


Figure S5.12. The GPC traces of A) PAlaGlu(OtBu) an B) PLys(Z)Glu(OtBu) in DMSO at 70 °C with a 0.5 mL·min^{-1} flow rate. The peak at ca. 43 min. is the solvent elution peak.

The raw GPC traces in Figures S5.12A and B showed small shoulders for elution volumes between 35 and 40 mL. To quantify the small shoulders in the GPC traces, the traces were fit with a sum of Gaussians according to Equation S5.1. In Equation S5.1 the elution time is denoted by x, a_i is the weight of the i^{th} Gaussian defined by its mean b_i and standard deviation c_i. The parameters retrieved from the fits are given below in Table S5.1. The GPC trace of PAlaGlu(OtBu) was fit with the sum of 5 gaussians, while that of PLys(Z)Glu(OtBu) required only 4. Figures S5.13A and B demonstrate that the parameters given in Table S5.1 accurately represented the experimental GPC traces.

$$f(x) = \sum_i \frac{a_i}{c_i\sqrt{2\pi}} \exp\left(-\frac{(x-b_i)^2}{2c_i^2}\right) \qquad \text{(S5.1)}$$

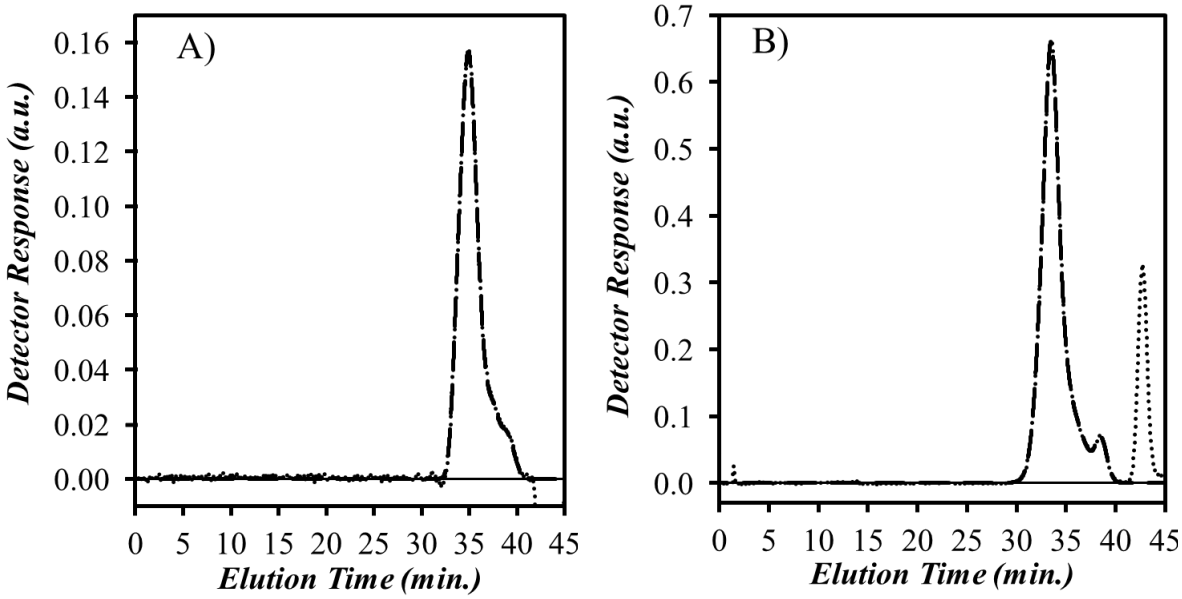


Figure S5.13. Comparison between the (⋯) experimental GPC traces and the (− −) sum-of-Gaussians fit for A) PAlaGlu(OtBu) and B) PLys(Z)Glu(OtBu). The solvent elution peak at ca. 43 min. was not included in the analysis.

Based on the b_i values given in Table S5.1, the Gaussians were assigned either to the main peak or peak shoulder. The b_i values representing Gaussians close to the peak maximum were considered to be part of the main peak, while those at higher elution volumes were considered to be part of the peak shoulder. Figures S5.14 shows that the main peak could be successfully isolated from the experimental GPC trace using this procedure. The weight fraction of the shoulder was then calculated as the normalized sum of the a_i values assigned to

the peak shoulder (see Table S5.1). This analysis showed that the shoulders contributed to rather small fractions of the GPC traces, taking values of 0.13 and 0.05 for PAlaGlu(OtBu) and PLys(Z)Glu(OtBu), respectively. To highlight the relative weights of each of these fractions, the inlays in Figures 5.14A and B compare main peak and shoulder contributions.

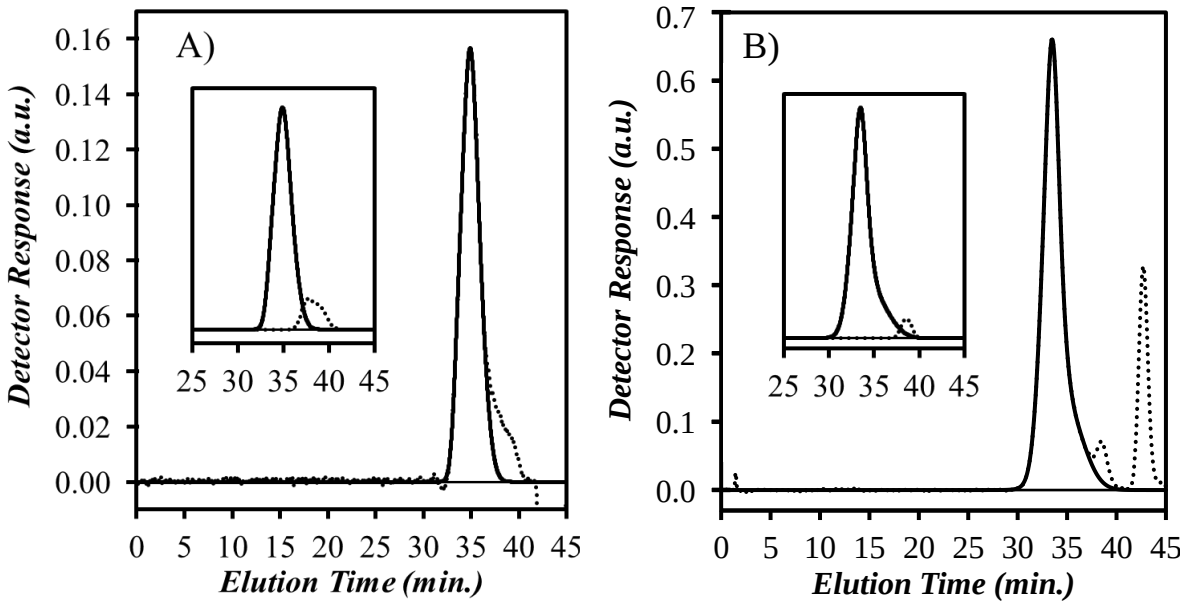


Figure S5.14. Plot of the (—) main peak compared to the (⋯) experimental GPC trace for A) PAlaGlu(OtBu) and B) PLys(Z)Glu(OtBu). The inlays compare the (—) main peak and (⋯) peak shoulder obtained from the fit of the GPC traces.

Table S5.1. Parameters retrieved from the sum-of-Gaussians fit of the GPC traces.

Polymer GPC Trace	i	a_i	b_i	c_i	Relative Area**
PAlaGlu(OtBu)	1	0.36	33.7	0.49	0.08
	2	2.13	34.7	0.73	0.48
	3	1.33	35.7	0.89	0.30
	4*	0.26	37.5	0.58	0.06
	5*	0.32	38.9	0.76	0.07
PLys(Z)Glu(OtBu)	1	10.21	33.3	0.92	0.57
	2	1.82	33.6	0.50	0.10
	3	5.02	34.9	1.72	0.28
	4*	0.81	38.6	0.57	0.05

* Used to represent the peak shoulder.

** Equal to $\frac{a_i}{\sum_i a_i}$

Determination of the molar absorptivity coefficient of PyMAAc in DMSO

The molar absorbance coefficient (ε_{Py}) of the model compound 1-pyrenemethyl acetamide (PyMAAc) was calculated by applying Beer-Lambert's law to the absorbance at 346.5 nm of a PyMAAc solution in DMSO using a 1 cm pathlength UV cell. ε_{Py} was determined to equal 39,300 ± 300 $M^{-1}\cdot cm^{-1}$ from the slope of the plot of absorbance vs. PyMAAC concentration shown in Figure S5.15.

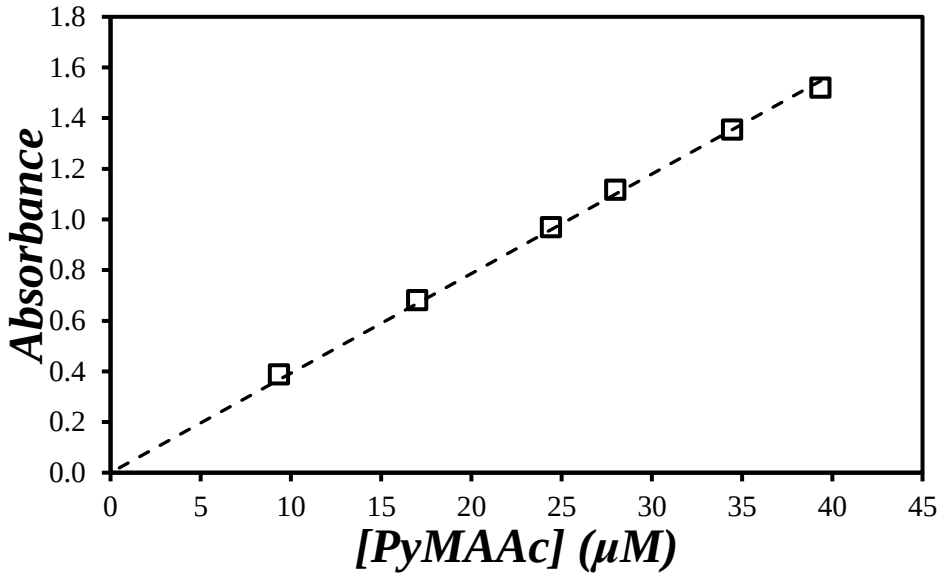


Figure S5.15. Absorbance of PyMAAc at 346.5 nm as a function of concentration. The straight line represents the linear fit of the data with a slope equal to 39,300 ± 300 M^{-1}. R^2 = 1.00.

Equations used to globally analyze the pyrene-labeled polypeptides

The monomer and excimer fluorescence decays of the pyrene-labeled polypeptides were globally analysed using Equations S5.2 and Equations S5.3 or S5.4 for the monomer and excimer decays, respectively. More precisely, all monomer fluorescence decays were fit with Equation S5.2. Equations S5.3 or S5.4 were used to fit the excimer decays depending on whether the longer-lived and poorly stacked pyrene dimers (*D**) were assumed to be produced through direct excitation of pyrene aggregates only, or a combination of direct excitation of pyrene aggregates and diffusive encounters between pyrenyl labels, respectively. Equation S5.3 was applied when too little excimer was generated and the contribution from *D** to the excimer fluorescence decays was small, typically less than 4% (see f_{ED} in Table S5.3). This

was the case for the more flexible Py-PGlyGlu and Py-PAlaGlu series. Application of the more general Equation S5.4 required a stronger contribution from the species *D** such as that obtained for the Py-PGlu and Py-PLys(Z)Glu series whose more rigid backbone favored the formation of more *D** species.

$$[Py^*]_{(t)} = [Py^*_{diff}]_{(t)} + [Py^*_{k2}]_{(t)} + [Py^*_{free}]_{(t)} = [Py^*_{diff}]_o \exp\left(-\left(A_2 + \frac{1}{\tau_M} \right) t - A_3 \left(1 - \exp(-A_4 t) \right) \right)$$

$$+ \left([Py^*_{k2}]_o + [Py^*_{diff}]_o e^{-A_3} \sum_{i=0}^{\infty} \frac{A_3^i}{i!} \frac{A_2 + iA_4}{A_2 + iA_4 - k_2} \right) \exp\left(-\left(k_2 + \frac{1}{\tau_M} \right) t \right)$$

$$- [Py^*_{diff}]_o e^{-A_3} \sum_{i=0}^{\infty} \frac{A_3^i}{i!} \frac{A_2 + iA_4}{A_2 + iA_4 - k_2} \exp\left(-\left(A_2 + iA_4 + \frac{1}{\tau_M} \right) t \right)$$

$$+ [Py^*_{free}]_o \exp\left(-\frac{t}{\tau_M} \right) \qquad \text{(S5.2)}$$

$$[E^*]_{(t)} = [E0^*]_{(t)} + [Py^*]_{(t)} = k_2 \left(\left([Py_{k2}(E0)]_o + [Py_{diff}(E0)]_o e^{-A_3} \sum_{i=0}^{\infty} \frac{A_3^i}{i!} \frac{A_2 + iA_4}{A_2 + iA_4 - k_2} \right) \right.$$

$$\times \frac{\exp\left(-\frac{t}{\tau_{E0}} \right) - \exp\left(-\left(k_2 + \frac{1}{\tau_M} \right) t \right)}{k_2 + \frac{1}{\tau_M} - \frac{1}{\tau_{E0}}}$$

$$\left. + [Py_{diff}(E0)]_o e^{-A_3} \sum_{i=0}^{\infty} \frac{A_3^i}{i!} \frac{A_2 + iA_4}{A_2 + iA_4 - k_2} \frac{\exp\left(-\left(A_2 + iA_4 + \frac{1}{\tau_M} \right) t \right) - \exp\left(-\frac{t}{\tau_{E0}} \right)}{A_2 + iA_4 + \frac{1}{\tau_M} - \frac{1}{\tau_{E0}}} \right)$$

$$+ [E0]_o \times \exp\left(-\frac{t}{\tau_{E0}}\right) + [D^*]_o \times \exp(-t/\tau_D) \tag{S5.3}$$

$$[E^*]_{(t)} = [E0^*]_{(t)} + [D^*]_{(t)} = k_2 \left(\left([Py_{k2}(E0)]_o + [Py_{diff}(E0)]_o e^{-A_3} \sum_{i=0}^{\infty} \frac{A_3^i}{i!} \frac{A_2 + iA_4}{A_2 + iA_4 - k_2} \right) \right.$$

$$\times \frac{\exp\left(-\frac{t}{\tau_{E0}}\right) - \exp\left(-\left(k_2 + \frac{1}{\tau_M}\right)t\right)}{k_2 + \frac{1}{\tau_M} - \frac{1}{\tau_{E0}}}$$

$$\left. + [Py_{diff}(E0)]_o e^{-A_3} \sum_{i=0}^{\infty} \frac{A_3^i}{i!} \frac{A_2 + iA_4}{A_2 + iA_4 - k_2} \frac{\exp\left(-\left(A_2 + iA_4 + \frac{1}{\tau_M}\right)t\right) - \exp\left(-\frac{t}{\tau_{E0}}\right)}{A_2 + iA_4 + \frac{1}{\tau_M} - \frac{1}{\tau_{E0}}} \right)$$

$$+ k_2 \left(\left([Py_{k2}(D)]_o + [Py_{diff}(D)]_o e^{-A_3} \sum_{i=0}^{\infty} \frac{A_3^i}{i!} \frac{A_2 + iA_4}{A_2 + iA_4 - k_2} \right) \times \frac{\exp\left(-\frac{t}{\tau_D}\right) - \exp\left(-\left(k_2 + \frac{1}{\tau_M}\right)t\right)}{k_2 + \frac{1}{\tau_M} - \frac{1}{\tau_D}} \right.$$

$$\left. + [Py_{diff}(D)]_o e^{-A_3} \sum_{i=0}^{\infty} \frac{A_3^i}{i!} \frac{A_2 + iA_4}{A_2 + iA_4 - k_2} \frac{\exp\left(-\left(A_2 + iA_4 + \frac{1}{\tau_M}\right)t\right) - \exp\left(-\frac{t}{\tau_D}\right)}{A_2 + iA_4 + \frac{1}{\tau_M} - \frac{1}{\tau_D}} \right)$$

$$+ [E0]_o \times \exp\left(-\frac{t}{\tau_{E0}}\right) + [D]_o \times \exp\left(-\frac{t}{\tau_D}\right) \tag{S5.4}$$

In Equations S5.2 and S5.3, the parameters A_2, A_3, and A_4 are given in Equations S5.5.a-c.

$$A_2 = < n > \times \frac{k_{blob} k_e[blob]}{k_{blob} + k_e[blob]} \quad \text{(S5.5.a)}$$

$$A_3 = < n > \times \left(\frac{k_{blob}}{k_{blob} + k_e[blob]} \right)^2 \quad \text{(S5.5.b)}$$

$$A_4 = k_{blob} + k_e[blob] \quad \text{(S5.5.c)}$$

Example global-fit of the monomer and excimer decays of PyLys(Z)Glu according to the FBM

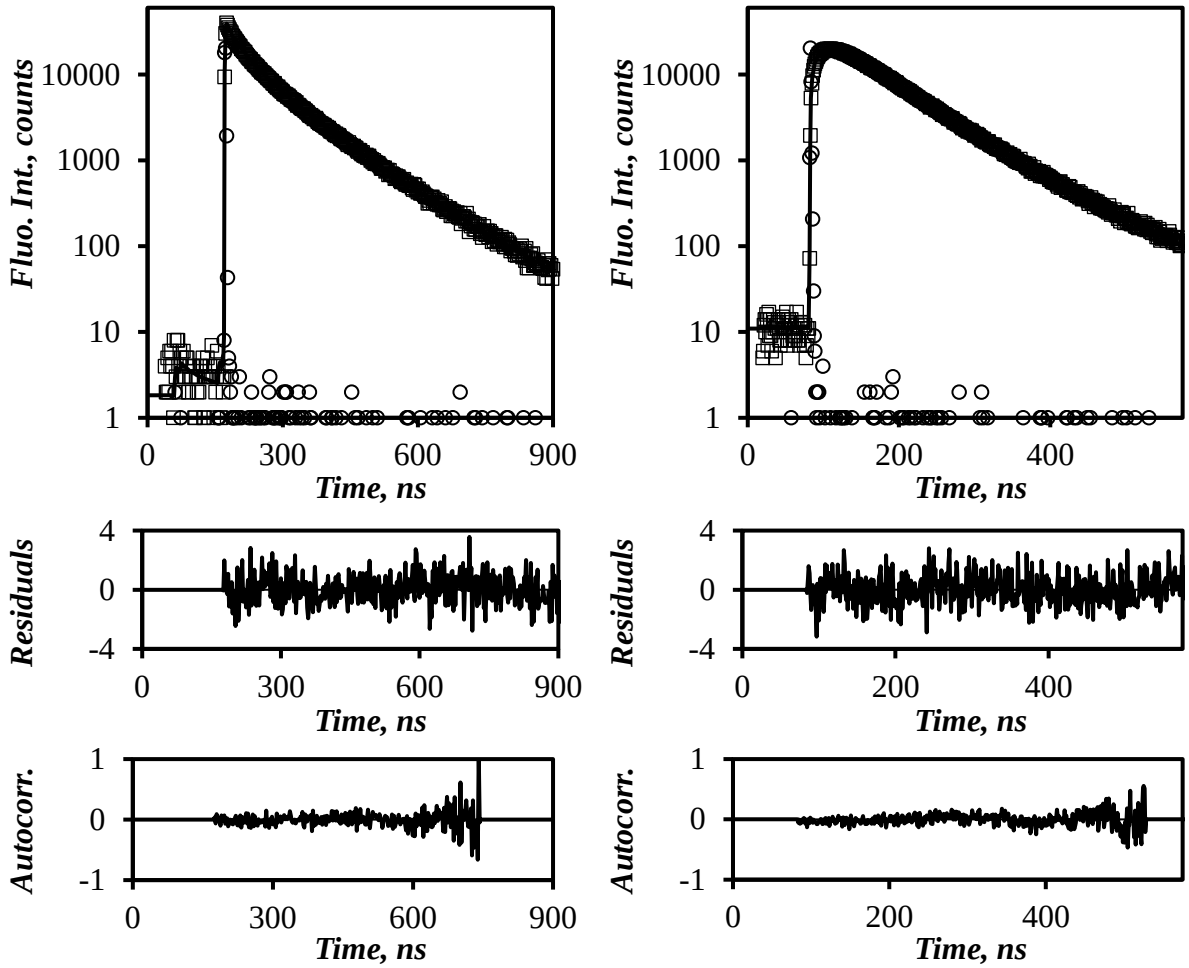


Figure S5.16. Monomer (left, λ_{em} = 375 nm) and excimer (right, λ_{em} = 510 nm) fluorescence decays (□) of Py-PLys(Z)Glu (x = 10.0 mol%) in DMSO. The instrument response function (○) and the fit of the global analysis of the FBM (—) are overlaid with the fluorescence decays. λ_{ex} = 344 nm, [Py] = 2.7×10^{-6} M. χ^2 = 1.09.

Parameters retrieved from the FBM of the pyrene-labeled polypeptides

Table S5.2. Parameters retrieved using the FBM analysis of the pyrene monomer decays of pyrene-labeled polypeptides in acetified DMSO.

Polypeptide	x (mol%)	k_2 (ns^{-1})	f_{Mk2}	$k_e[blob]$ (μs^{-1})	f_{Mdiff}	k_{blob} (μs^{-1})	$<n>$	f_{Mfree}	χ^2
Py-PGlyGlu τ_M = 170 ns	2.8	0.14	0.07	13.8	0.39	12.4	1.28	0.54	1.26
	3.5	0.14	0.09	9.3	0.48	10.1	1.39	0.43	1.08
	4.3	0.14	0.14	7.0	0.62	12.7	1.39	0.24	1.15
	5.3	0.14	0.16	7.3	0.60	11.5	1.52	0.24	1.27
Py-PAlaGlu τ_M = 172 ns	6.6	0.14	0.13	6.3	0.75	10.4	1.16	0.12	1.14
	9.6	0.14	0.17	5.0	0.77	8.9	1.53	0.05	1.14
	13.5	0.14	0.25	4.6	0.74	8.2	2.36	0.01	1.27
	15.2	0.14	0.25	4.6	0.73	8.7	2.40	0.02	1.15
Py-PGlu τ_M = 170 ns	6.0	0.15	0.14	8.6	0.39	14.7	1.31	0.47	1.12
	8.0	0.15	0.19	4.2	0.50	13.2	1.48	0.31	1.10
	10.4	0.15	0.20	5.7	0.46	11.9	1.60	0.34	1.12
	11.5	0.15	0.19	4.3	0.55	13.1	1.30	0.26	1.09
	12.3	0.15	0.24	3.0	0.57	12.2	1.46	0.19	1.13
Py-PLys(Z)Glu τ_M = 160 ns	4.5	0.15	0.09	8.3	0.49	12.3	0.96	0.42	1.01
	5.5	0.15	0.12	6.4	0.60	13.1	0.93	0.28	1.12
	6.6	0.15	0.16	5.3	0.65	13.0	0.93	0.19	1.24
	10.0	0.15	0.19	5.3	0.68	12.3	1.21	0.14	1.09
	13.2	0.15	0.25	3.6	0.69	10.6	1.58	0.06	1.12

Table S5.3. Parameters retrieved using the FBM analysis of the pyrene excimer decays of pyrene-labeled polypeptides in acidified DMSO.

Polypeptide	x (mol%)	f_{Ek2}	τ_{E0} (ns)	$f_{EdiffE0}$	f_{EE0}	τ_D (ns)	f_{EdiffD}	f_{ED}	χ^2
Py-PGlyGlu	2.8	0.13	49	0.79	0.04	141*	-	0.07	1.26
	3.5	0.15	48	0.81	0.03	145*	-	0.04	1.08
	4.3	0.17	52	0.75	0.00	156*	-	0.01	1.15
	5.3	0.20	51	0.75	0.05	153*	-	0.01	1.27
Py-PAlaGlu	6.6	0.14	52	0.80	0.04	119*	-	0.02	1.14
	9.6	0.18	52	0.78	0.03	141*	-	0.01	1.14
	13.5	0.24	47	0.71	0.00	88*	-	0.04	1.27
	15.2	0.24	49	0.70	0.05	94*	-	0.02	1.15
Py-PGlu	6.0	0.22	49	0.64	0.10	134	0.02	0.02	1.12
	8.0	0.26	35	0.24	0.08	58	0.41	0.01	1.10
	10.4	0.28	47	0.59	0.10	104	0.04	0.00	1.12
	11.5	0.23	46	0.57	0.11	82	0.08	0.01	1.09
	12.3	0.26	40	0.41	0.08	68	0.21	0.03	1.13
Py-PLys(Z)Glu	4.5	0.14	48	0.66	0.07	99	0.12	0.00	1.01
	5.5	0.16	50	0.63	0.05	87	0.16	0.00	1.12
	6.6	0.19	45	0.39	0.02	68	0.37	0.02	1.24
	10.0	0.21	42	0.48	0.05	66	0.26	0.00	1.09
	13.2	0.25	40	0.48	0.00	65	0.21	0.05	1.12

* Excimer decays were fit with Equation S5.3.

Table S5.4. Fraction of pyrene species calculated using parameters retrieved from the FBM analysis of the pyrene monomer and excimer decays of pyrene-labeled polypeptides in acidified DMSO.

Polypeptide	x (mol%)	f_{k2}	f_{diffE0}	f_{diffD}	f_{diff}	f_{E0}	f_{D}	f_{agg}	f_{free}	χ^2
Py-PGlyGlu	2.8	0.06	0.38	-	0.38	0.00	0.03	0.03	0.52	1.26
	3.5	0.09	0.46	-	0.46	0.00	0.02	0.02	0.42	1.08
	4.3	0.13	0.58	-	0.58	0.06	0.01	0.07	0.22	1.15
	5.3	0.15	0.58	-	0.58	0.03	0.01	0.04	0.23	1.27
Py-PAlaGlu	6.6	0.12	0.71	-	0.71	0.03	0.02	0.05	0.11	1.14
	9.6	0.17	0.74	-	0.74	0.03	0.01	0.04	0.05	1.14
	13.5	0.24	0.70	-	0.70	0.00	0.04	0.04	0.01	1.27
	15.2	0.24	0.69	-	0.69	0.05	0.02	0.07	0.01	1.15
Py-PGlu	6.0	0.13	0.36	0.01	0.37	0.06	0.01	0.07	0.44	1.12
	8.0	0.18	0.17	0.29	0.46	0.06	0.00	0.06	0.29	1.10
	10.4	0.19	0.40	0.03	0.43	0.07	0.00	0.07	0.31	1.12
	11.5	0.18	0.44	0.06	0.50	0.08	0.01	0.09	0.23	1.09
	12.3	0.21	0.34	0.18	0.52	0.07	0.03	0.10	0.17	1.13
Py-PLys(Z)Glu	4.5	0.09	0.39	0.07	0.47	0.04	0.00	0.04	0.40	1.01
	5.5	0.12	0.46	0.12	0.58	0.04	0.00	0.04	0.26	1.12
	6.6	0.16	0.32	0.30	0.62	0.02	0.01	0.03	0.19	1.24
	10.0	0.18	0.42	0.23	0.65	0.05	0.00	0.05	0.13	1.09
	13.2	0.24	0.46	0.20	0.66	0.00	0.05	0.05	0.05	1.12

Table S5.5. Parameters retrieved using the FBM analysis of the pyrene monomer decays of pyrene-labeled polypeptides in acidified DMF.

Polypeptide	x (mol%)	k_2 (ns^{-1})	f_{Mk2}	$k_e[blob]$ (μs^{-1})	f_{Mdiff}	k_{blob} (μs^{-1})	$<n>$	f_{Mfree}	χ^2
Py-PGlyGlu τ_M = 215 ns	2.8	0.13	0.12	10.4	0.43	12.4	1.42	0.44	1.21
	3.5	0.13	0.16	8.1	0.50	11.6	1.50	0.34	1.15
	4.3	0.13	0.23	6.7	0.58	14.4	1.58	0.18	1.18
	5.3	0.13	0.27	6.3	0.56	13.1	1.73	0.18	1.21
Py-PAlaGlu τ_M = 215 ns	6.6	0.13	0.30	4.2	0.62	11.0	1.39	0.09	1.08
	9.6	0.13	0.35	3.6	0.61	9.7	1.76	0.04	1.08
	13.5	0.13	0.47	3.1	0.53	10.7	2.60	0.01	1.24
	15.2	0.13	0.50	2.4	0.49	9.8	2.78	0.01	1.15
Py-PGlu τ_M = 215 ns	6.0	0.15	0.18	5.3	0.36	14.6	1.28	0.46	1.00
	8.0	0.15	0.23	2.9	0.46	13.3	1.34	0.32	1.05
	10.4	0.15	0.29	3.6	0.46	13.9	1.46	0.25	1.06
	11.5	0.15	0.22	3.1	0.54	17.5	1.25	0.24	1.07
	12.3	0.15	0.29	3.1	0.53	13.0	1.39	0.18	1.10
Py-PLys(Z)Glu τ_M = 200 ns	4.5	0.15	0.12	10.3	0.42	17.3	1.10	0.46	1.06
	5.5	0.15	0.16	7.4	0.48	14.8	1.15	0.35	1.08
	6.6	0.15	0.19	5.2	0.57	14.3	1.17	0.25	1.20
	10.0	0.15	0.29	4.5	0.60	14.6	1.35	0.11	1.20
	13.2	0.15	0.33	4.1	0.61	13.0	1.72	0.06	1.06

Table S5.6. Parameters retrieved using the FBM analysis of the pyrene excimer decays of pyrene-labeled polypeptides in acidified DMF.

Polypeptide	x (mol%)	f_{Ek2}	τ_{E0} (ns)	$f_{EdiffE0}$	f_{EE0}	τ_D (ns)	f_{EdiffD}	f_{ED}	χ^2
Py-PGlyGlu	2.8	0.21	53	0.73	0.04	179*	-	0.03	1.21
	3.5	0.23	52	0.72	0.04	181*	-	0.01	1.15
	4.3	0.25	55	0.64	0.10	179*	-	0.01	1.18
	5.3	0.30	54	0.63	0.06	173*	-	0.01	1.21
Py-PAlaGlu	6.6	0.30	53	0.62	0.04	125*	-	0.04	1.08
	9.6	0.34	48	0.59	0.00	94*	-	0.07	1.08
	13.5	0.39	47	0.43	0.00	75*	-	0.18	1.24
	15.2	0.42	46	0.41	0.00	75*	-	0.17	1.15
Py-PGlu	6.0	0.29	50	0.53	0.13	108	0.05	0.00	1.00
	8.0	0.30	36	0.36	0.03	68	0.24	0.09	1.05
	10.4	0.34	41	0.38	0.00	70	0.16	0.12	1.06
	11.5	0.24	41	0.34	0.16	72	0.25	0.01	1.07
	12.3	0.28	46	0.34	0.20	75	0.17	0.01	1.10
Py-PLys(Z)Glu	4.5	0.19	53	0.64	0.12	178	0.02	0.03	1.06
	5.5	0.23	52	0.63	0.08	128	0.05	0.00	1.08
	6.6	0.24	37	0.33	0.05	68	0.38	0.00	1.20
	10.0	0.29	39	0.13	0.10	61	0.48	0.00	1.20
	13.2	0.31	36	0.24	0.09	60	0.35	0.01	1.06

* Excimer decays were fit with Equation S5.3.

Table S5.7. Fraction of pyrene species calculated using parameters retrieved from the FBM analysis of the pyrene monomer and excimer decays of pyrene-labeled polypeptides in acidified DMF.

Polypeptide	x (mol%)	f_{k2}	f_{diffE0}	f_{diffD}	f_{diff}	f_{E0}	f_{D}	f_{agg}	f_{free}	χ^2
Py-PGlyGlu	2.8	0.12	0.42	-	0.42	0.02	0.02	0.04	0.43	1.21
	3.5	0.15	0.49	-	0.49	0.03	0.01	0.04	0.32	1.15
	4.3	0.21	0.53	-	0.53	0.08	0.01	0.09	0.17	1.18
	5.3	0.25	0.53	-	0.53	0.05	0.01	0.06	0.17	1.21
Py-PAlaGlu	6.6	0.12	0.71	-	0.71	0.03	0.02	0.05	0.11	1.08
	9.6	0.17	0.74	-	0.74	0.03	0.01	0.04	0.05	1.08
	13.5	0.24	0.70	-	0.70	0.00	0.04	0.04	0.01	1.24
	15.2	0.24	0.69	-	0.69	0.05	0.02	0.07	0.01	1.15
Py-PGlu	6.0	0.17	0.30	0.03	0.33	0.07	0.00	0.07	0.43	1.00
	8.0	0.21	0.25	0.17	0.42	0.02	0.06	0.08	0.29	1.05
	10.4	0.26	0.29	0.12	0.42	0.00	0.09	0.09	0.23	1.06
	11.5	0.19	0.27	0.20	0.47	0.13	0.00	0.13	0.21	1.07
	12.3	0.24	0.29	0.14	0.43	0.17	0.01	0.18	0.14	1.10
Py-PLys(Z)Glu	4.5	0.11	0.37	0.01	0.38	0.07	0.02	0.09	0.42	1.06
	5.5	0.16	0.42	0.04	0.46	0.05	0.00	0.05	0.33	1.08
	6.6	0.18	0.25	0.29	0.54	0.04	0.00	0.04	0.24	1.20
	10.0	0.26	0.12	0.43	0.55	0.09	0.00	0.09	0.10	1.20
	13.2	0.30	0.22	0.34	0.56	0.08	0.01	0.09	0.05	1.06

Steady-state fluorescence spectra of pyrene-labeled polypeptides

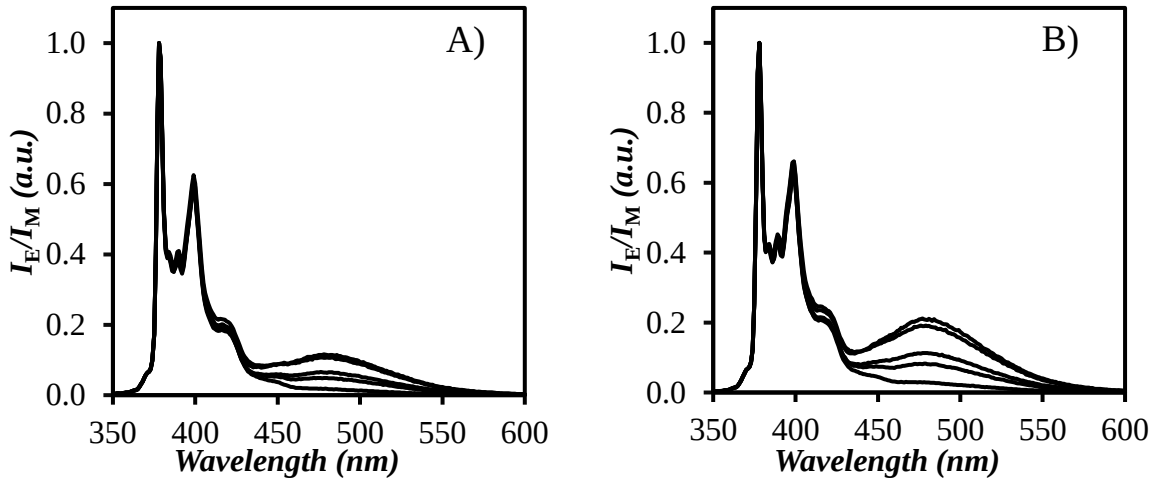


Figure S5.17. The steady-state fluorescence spectra of Py-PGlyGlu in A) DMSO and B) DMF.

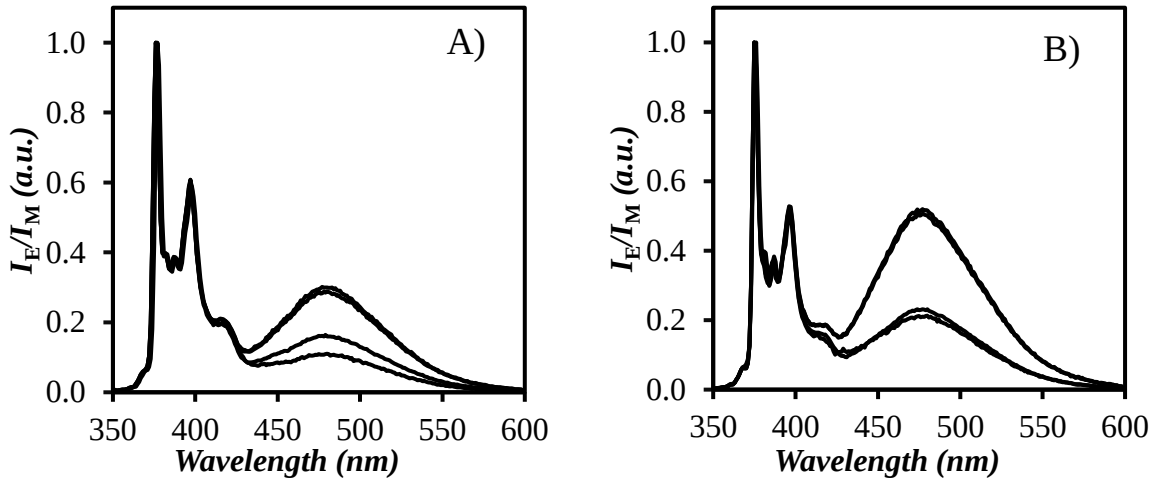


Figure S5.18. Steady-state fluorescence spectra of Py-PAlaGlu in A) DMSO and B) DMF.

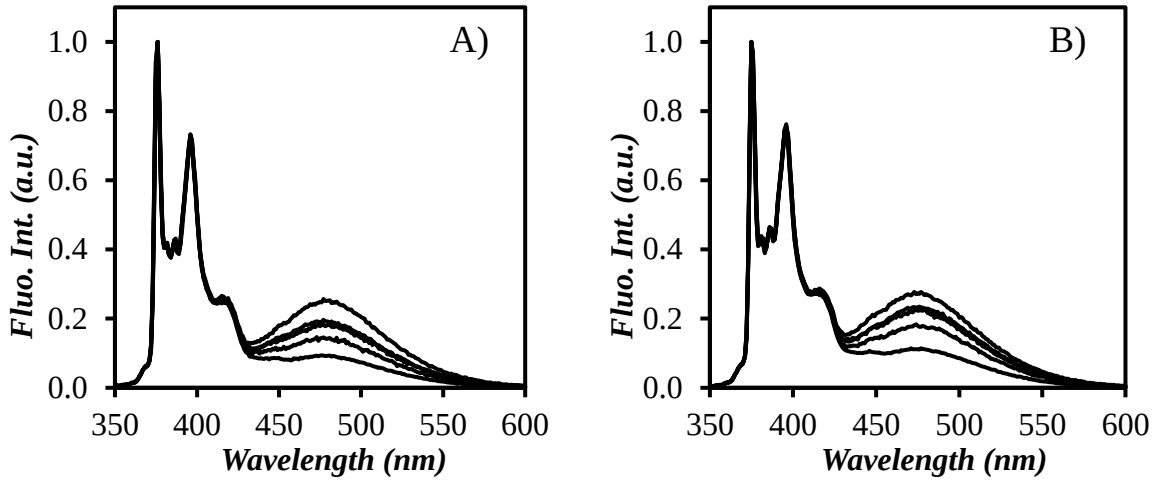


Figure S5.19. Steady-state fluorescence spectra Py-PGlu in A) DMSO and B) DMF.

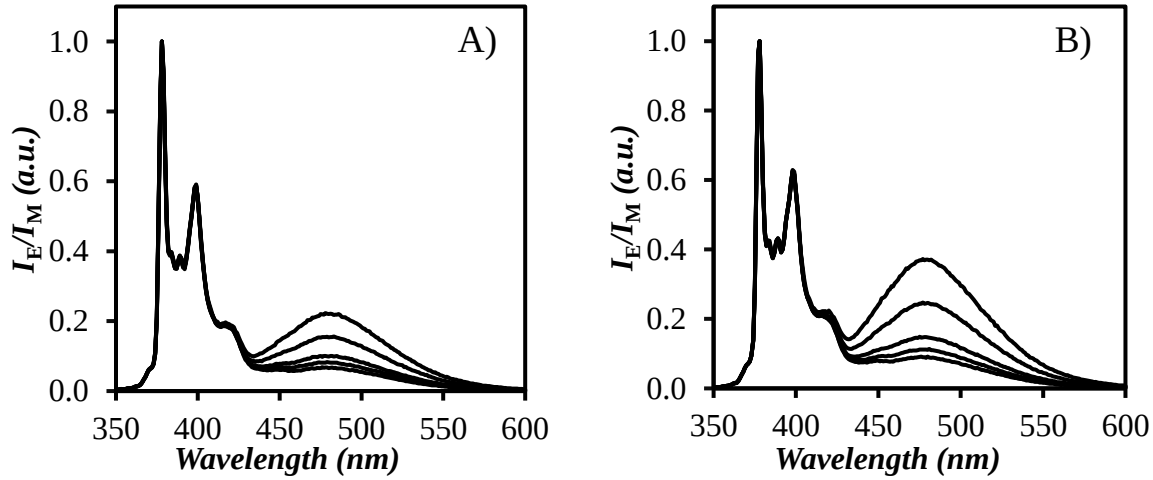


Figure S5.20. The steady-state fluorescence spectra of Py-PLys(Z)Glu in A) DMSO and B) DMF.

Molecular mechanics optimizations (MMOs) of PGlyGlu

The program HyperChem was used to construct an oligopeptide with 31 *aa*'s where 14 *aa*'s (i.e. 45%) were Glu and 17 *aa*'s (i.e. 55%) were Gly, resulting in a Glu content similar to that of Py-PGlyGlu. Using the random number generator provided by MS Excel, the Glu were distributed randomly throughout the backbone. After stretching and relaxing the backbone to produce a random-coil conformation of the polypeptide, the backbone atoms were locked in place to prevent them from moving during the optimizations. A pyrenyl label was then attached onto a reference Glu residue. Sequentially, a second pyrenyl label was placed onto neighbouring Glu residues. The pyrenes were then brought together and the overlap between the two pyrenyl planes was quantified by counting the number of carbons of one pyrene overlapping the frame of the other pyrene. Pyrene cannot be labeled onto Gly, so if a Gly was present in the residue position of interest the Gly side chain was replaced with that of the pyrene-labeled Glu. Since the backbone was locked in place, changing identity of the structural unit (by replacing the side chain) had no effect on the backbone conformation, but allowed the MMOs to be conducted for all the reference positions along the backbone. A plot of the number of overlapping carbons as a function of the number of *aa*'s separating the two pyrene-labeled *aa*'s is given in Figure S5.21. This plot indicated that two pyrenyls attached onto two Glu's separated by up to 5 or 6 residues could overlap satisfactorily and form an excimer. After averaging for three different reference positions, N_{blob}^{theo} was found to equal 10.3 ± 1.2 for PGlyGlu.

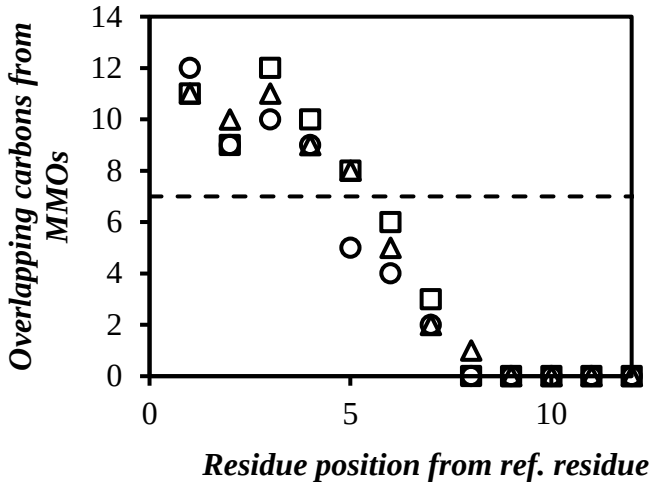


Figure S5.21. The number of overlapping carbons between two pyrenes attached onto poly(GlyGlu) (55% Gly) as a function of the number of residues separating the two structural units bearing a pyrenyl label. The dashed line equals a carbon overlap of 7 which was considered to provide sufficient overlap for the pyrenes to form excimer efficiently.

Intrinsic viscosity of PGlyGlu in DMSO

The intrinsic viscosity of PGlyGlu was determined using a universal-size 75 Cannon Ubbelohde viscometer in a temperature bath maintained at 25.0 ºC. The average of thrice measured flow times of pure DMSO and PGlyGlu solutions in DMSO with concentration ranging from 2.6 to 7.l mg/mL are given in Table S5.8.

Table S5.8. Flow times of PGlyGlu in DMSO.

Concentration (mg/mL)	Flow Time (s)
0 (pure DMSO)	229.26 ± 0.09
7.1	247.48 ± 0.08
5.3	243.26 ± 0.03
4.2	240.68 ± 0.14
3.5	238.80 ± 0.21
2.6	236.45 ± 0.08

The flow times t were then used to calculate the reduced ($\eta_{red} = (t\text{-}t_0)/(t \times c)$) and inherent ($\eta_{inh} = \ln(t/t_0)/c$) viscosities as a function of copolypeptide concentration c, where t_0 is the flow time of the pure solvent. Figure S5.22 shows that both the reduced and inherent viscosity increased linearly with decreasing copolypeptide concentration. Using the Huggins ($[\eta] = \lim_{c \to 0} \eta_{red}$) and Kraemer ($[\eta] = \lim_{c \to 0} \eta_{inh}$) relationships, the intrinsic viscosity was found to equal 12.81 (±0.04) mL/g for PGlyGlu in DMSO.

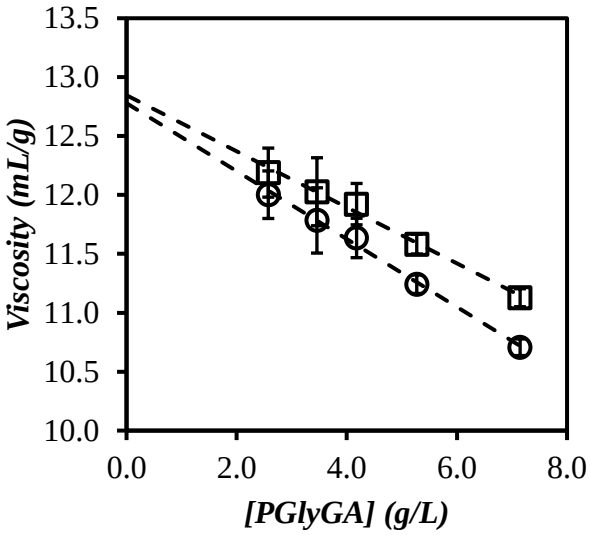


Figure S5.22. Plot of the (□) reduced and (○) inherent viscosities of PGlyGlu in DMSO as a function of concentration. The dashed lines represent the lines of best fit: $\eta_{red} = -0.238c + 12.81$, $R^2 = 0.99$. $\eta_{inh} = -0.288c + 12.78$, $R^2 = 1.00$.

I_E/I_M ratios of PGlyGlu as a function of polymer concentration

The I_E/I_M ratios of PGlyGlu were measured as a function of copolypeptide concentration to check for the presence of intermolecular interactions. Since the fluorescence of the Py-PGlyGlu samples are typically prepared such that the molar concentration of pyrene equals 2.7×10^{-6} M, the mass concentration of copolypeptide depends on the pyrene content of the sample, where a lower pyrene content is associated with a larger mass concentration of copolypeptide. For example, the pyrene content of Py-PGlyGlu ranged from 2.8 to 5.3 mol%, which corresponds to copolypeptide concentration of 8.9 and 4.9 mg/L, respectively. To ensure

that no polymer aggregation took place in this concentration range, the I_E/I_M ratios of the Py-PGlyGlu samples having the highest and lowest pyrene content were measured across the 4.9 – 8.9 mg/L concentration range. Figure S5.23 shows that the I_E/I_M ratios were independent of concentration demonstrating that excimer was only formed intramolecularly.

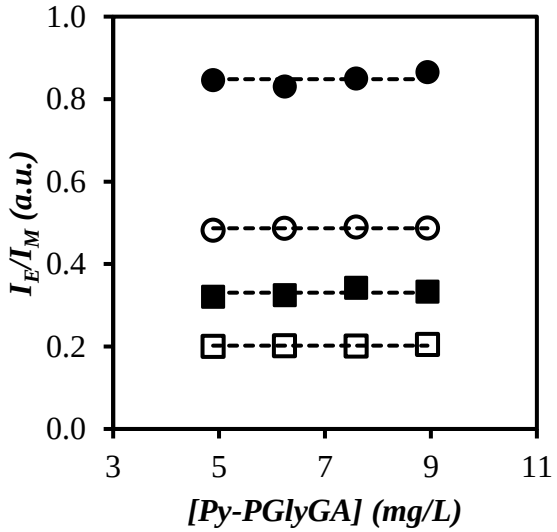


Figure S5.23. Plot of the I_E/I_M ratio of Py-PGlyGlu with pyrene contents equal to (squares) 2.8 and (circles) 5.3 mol% in (hollow) DMSO and (solid) DMF.

S6 – SI for Chapter 6

Solvent drying

Ethyl acetate: Ethyl acetate (500 mL) was refluxed with calcium hydride (2.5 g) for 3 hours, then distilled under a nitrogen atmosphere. The first ca. 15 mL of ethyl acetate distilled was discarded. The distilled ethyl acetate was stored under nitrogen until used.

THF: THF was dried in a similar procedure as ethyl acetate.

Dioxane: Dioxane (100 mL) was pre-dried by mixing with CaH_2 (0.2 g) for 2 hours. The dioxane was then refluxed with sodium (ca. 0.1 g) for 3 hours. The dioxane was then distilled under reduced pressure at 40 °C and was collected into a flame-dried Schlenk tube. The first 10 mL of distillate was collected in a separate ampule and discarded. The distilled dioxane was stored under nitrogen in the dark until used.

DMF: DMF (200 mL) was stirred with CaH_2 (0.4 g, 5 wt%) overnight in the dark under a nitrogen atmosphere. The calcium hydride was removed by filtration under a nitrogen atmosphere. The DMF was then distilled under reduced pressure at a temperature no greater than 35 °C to minimize degradation. The DMF was collected into a flame-dried Schlenk tube covered with aluminum foil to minimize light exposure. The first 15 mL of distillate was collected in a separate ampule and discarded. The distilled DMF was stored under nitrogen in the dark until used.

n-Butylamine: *n*-Butylamine (20 mL) was refluxed with CaH_2 (0.1 g) for 2 hours before it was distilled under vacuum at 40 °C into a Schlenk flask. The first 3 mL of distillate were discarded. The distilled *n*-butylamine was stored under nitrogen.

NMR spectra of the PAlaGlu(OtBu) samples

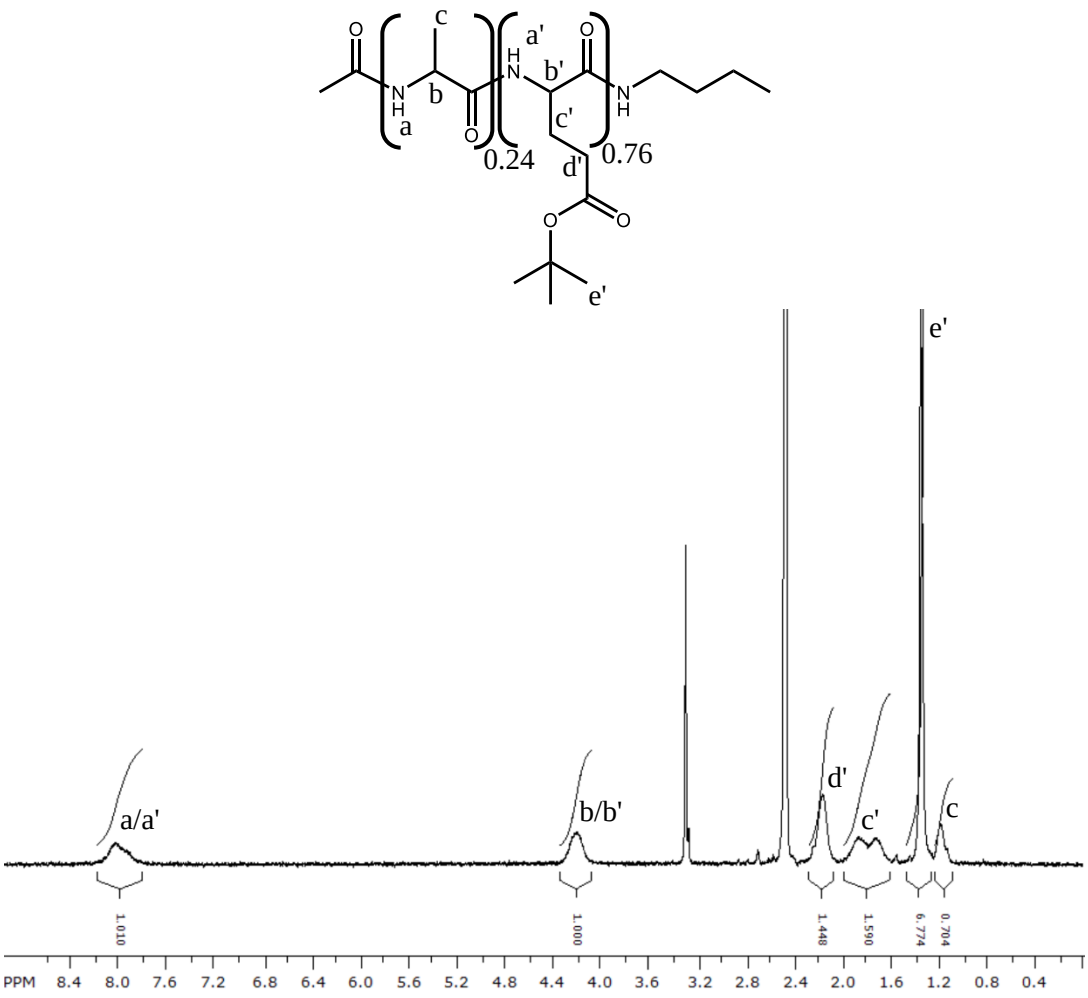


Figure S6.1. H NMR spectrum of $PAlaGlu_{24}(OtBu)_{76}$. 300 MHz, d_6-DMSO: δ = 8.00 (1H, amide), 4.20 (1H, α), 2.19 (2H, γ-Glu), 1.83 – 1.71 (2H, β-Glu), 1.34 (9H, tBu), 1.18 (3H, β-Ala). Residual water (3.36 ppm) and solvent (2.47 ppm) peaks are also present.

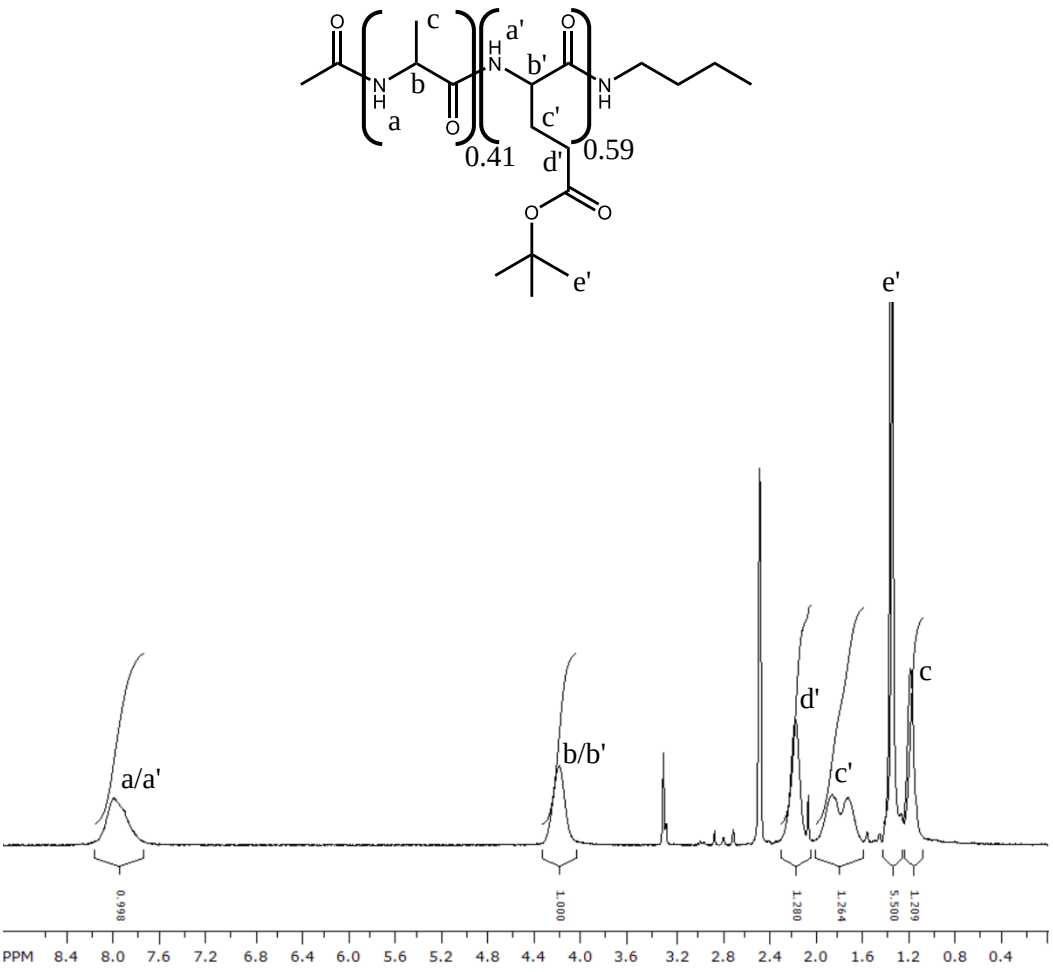


Figure S6.2. H NMR spectrum of $PAla_{41}Glu(OtBu)_{59}$. 300 MHz, d_6-DMSO: δ = 8.00 (1H, amide), 4.20 (1H, α), 2.19 (2H, γ-Glu), 1.83 – 1.71 (2H, β-Glu), 1.34 (9H, tBu), 1.18 (3H, β-Ala). Residual DMF (7.92, 2.86, and 2.70 ppm) and acetone (2.05). Residual water (3.36 ppm) and solvent (2.47 ppm) peaks are also present.

NMR Spectra of the PAlaGlu samples

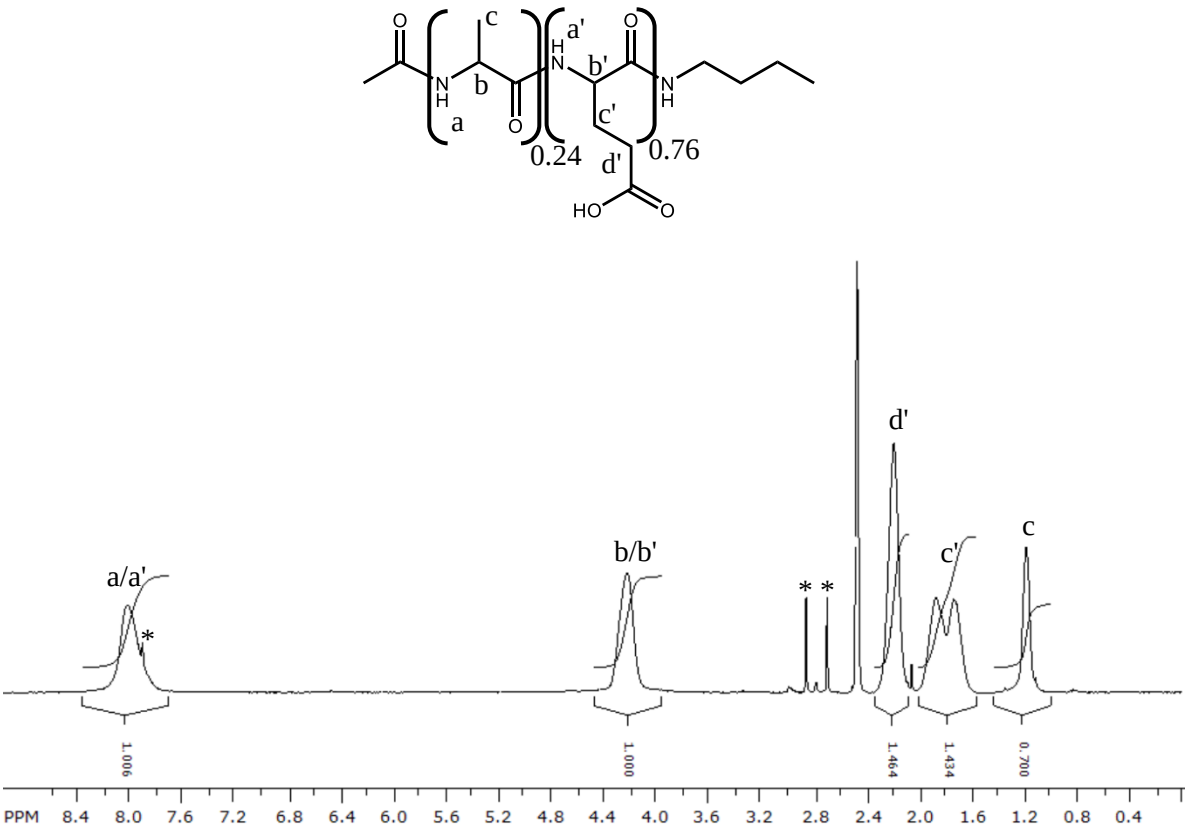


Figure S6.3. H NMR spectrum of $PAla_{24}Glu_{76}$. 300 MHz, d_6-DMSO: δ = 8.00 (1H, amide), 4.20 (1H, α), 2.19 (2H, γ-Glu), 1.83 – 1.71 (2H, β-Glu), 1.34 (9H, tBu), 1.18 (3H, β-Ala). The peaks marked with a * are from DMF. The residual solvent (2.47 ppm) peak is also present.

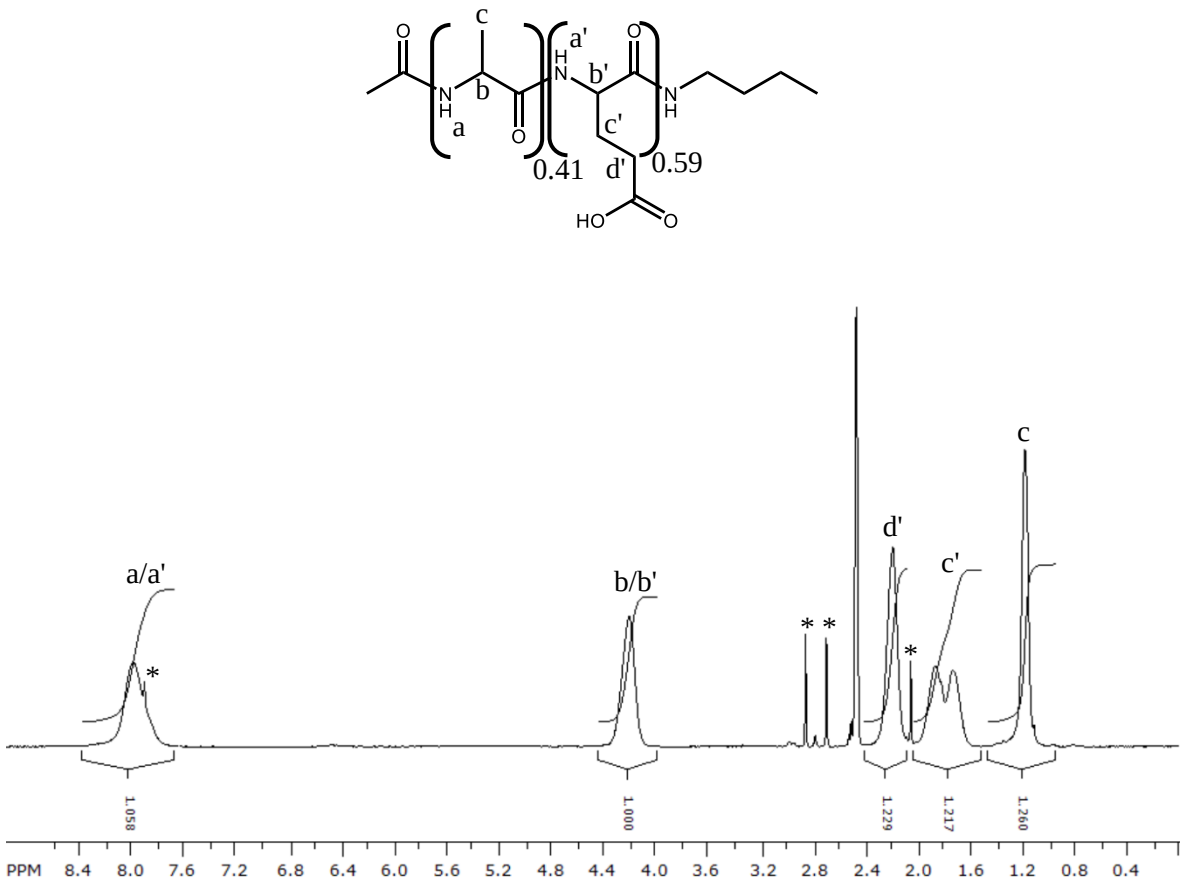


Figure S6.4. H NMR spectrum of $PAla_{41}Glu_{59}$. 300 MHz, d_6-DMSO: δ = 8.00 (1H, amide), 4.20 (1H, α), 2.19 (2H, γ-Glu), 1.83 – 1.71 (2H, β-Glu), 1.34 (9H, tBu), 1.18 (3H, β-Ala). The peaks marked with a * are from residual DMF and acetone. The residual solvent (2.47 ppm) peak is also present.

GPC traces of the PAlaGlu(OtBu) samples

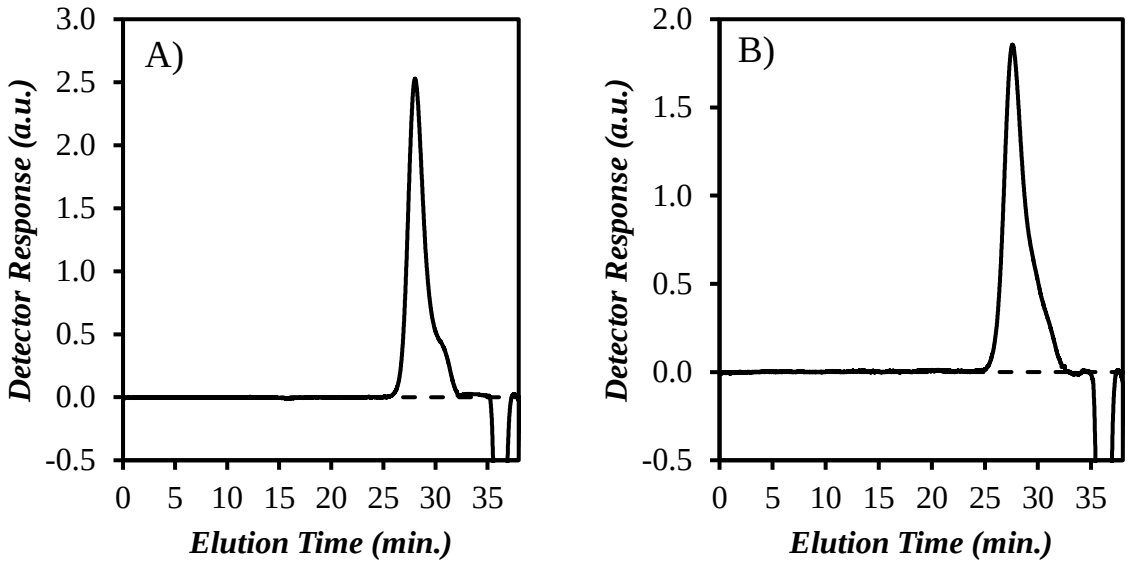


Figure S6.5. The DRI GPC traces of A) $PAlaGlu_{24}(OtBu)_{76}$ and B) $PAlaGlu_{41}(OtBu)_{59}$ in DMSO at 70 °C with a 0.6 $mL \cdot min^{-1}$ flow rate. The peak at ca. 36 min. is the solvent elution peak.

Steady-state fluorescence spectra of the Py-PAlaGlu samples in DMSO

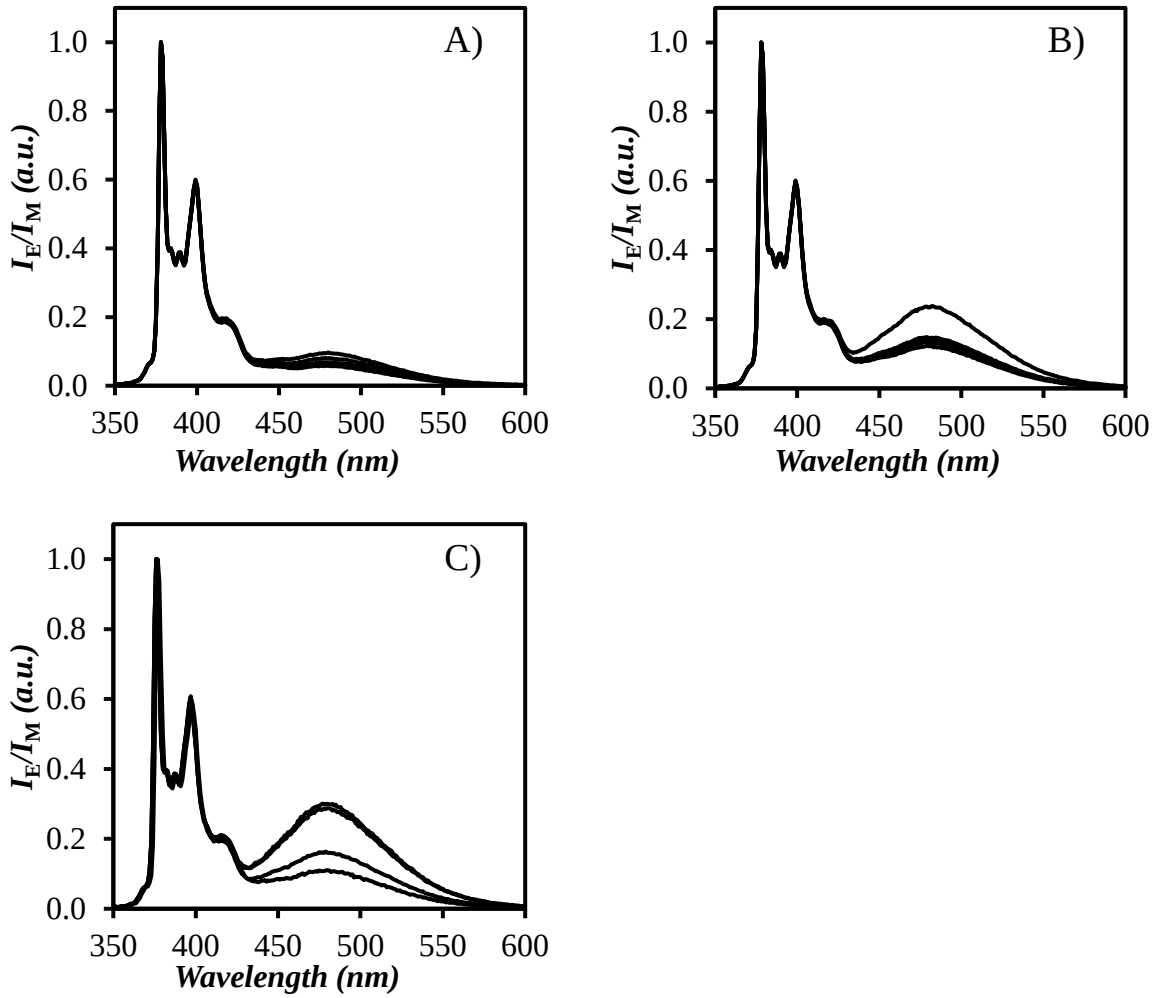


Figure S6.6. Steady-state spectra of Py-PAlaGlu's in DMSO; f_{Ala} = A) 0.24, B) 0.41, and C) 0.58.

Equations used to globally analyze the pyrene-labeled polypeptides

The monomer and excimer decays of the Py-PAlaGlus were fit with Equations S6.1 and S6.2.

$$[Py^*]_{(t)} = [Py^*_{diff}]_{(t)} + [Py^*_{k2}]_{(t)} + [Py^*_{free}]_{(t)} = [Py_{diff}]_o \exp\left(-\left(A_2 + \frac{1}{\tau_M}\right)t - A_3\left(1-\exp(-A_4 t)\right)\right)$$

$$+\left([Py_{k2}]_o + [Py_{diff}]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\right)\exp\left(-\left(k_2+\frac{1}{\tau_M}\right)t\right)$$

$$-[Py_{diff}]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\exp\left(-\left(A_2+iA_4+\frac{1}{\tau_M}\right)t\right)$$

$$+[Py_{free}]_o \exp\left(-\frac{t}{\tau_M}\right)\Bigg] \qquad \text{(S6.1)}$$

$$[E^*]_{(t)} = [E0^*]_{(t)} + [D^*]_{(t)} = k_2\Bigg(\left([Py_{k2}]_o + [Py_{diff}]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\right)$$

$$\times\frac{\exp\left(-\frac{t}{\tau_{E0}}\right)-\exp\left(-\left(k_2+\frac{1}{\tau_M}\right)t\right)}{k_2+\frac{1}{\tau_M}-\frac{1}{\tau_{E0}}}$$

$$+[Py_{diff}]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\frac{\exp\left(-\left(A_2+iA_4+\frac{1}{\tau_M}\right)t\right)-\exp\left(-\frac{t}{\tau_{E0}}\right)}{A_2+iA_4+\frac{1}{\tau_M}-\frac{1}{\tau_{E0}}}\Bigg)$$

$$+[E0]_o\times\exp\left(-\frac{t}{\tau_{E0}}\right)+[D]_o\times\exp\left(-\frac{t}{\tau_D}\right)\Bigg] \qquad \text{(S6.2)}$$

In Equations S6.1 and S6.2, the parameters A_2, A_3, and A_4 are given in Equations S6.3.a-c.

$$A_2 = < n > \times \frac{k_{blob} k_e [blob]}{k_{blob} + k_e [blob]} \quad \text{(S6.3.a)}$$

$$A_3 = < n > \times \left(\frac{k_{blob}}{k_{blob} + k_e [blob]} \right)^2 \quad \text{(S6.3.b)}$$

$$A_4 = k_{blob} + k_e [blob] \quad \text{(S6.3.c)}$$

The parameters k_{blob}, $k_e[blob]$, and $<n>$ represent the rate constant for the diffusive encounters between two structural units bearing a pyrenyl label, the product of the rate constant for exchange (k_e) of ground-state pyrene between *blobs* and the local *blob* concentration, and the average number of pyrenyl labels inside a *blob*, respectively. The pre-exponential factors corresponding to the pyrene species $[Py_{diff}^*]_M$, $[Py_{k2}^*]_M$, and $[Py_{free}^*]_M$ in Equation S6.1 and $[Py_{diff}^*]_E$, $[Py_{k2}^*]_E$, $[E0^*]_E$, and $[D^*]_E$ in Equation S6.2, where the indices *M* or *E* indicate that the pre-exponential factors were obtained from the monomer or the excimer fluorescence decays, were used to determine the molar fractions f_{diff}, f_{k2}, f_{free}, f_{E0}, and f_D whose expressions are given in Equations S6.4 – S6.8.

$$f_{diff} = \frac{1}{1 + \frac{[Py_{k2}^*]_M}{[Py_{diff}^*]_M} + \frac{[Py_{free}^*]_M}{[Py_{diff}^*]_M} + \frac{[E0^*]_E}{[Py_{diff}^*]_E} + \frac{[D^*]_E}{[Py_{diff}^*]_E}} \quad \text{(S6.4)}$$

$$f_{k2} = f_{diff} \times \frac{[Py_{k2}{}^*]_M}{[Py_{diff}{}^*]_M} = f_{diff} \times \frac{[Py_{k2}{}^*]_E}{[Py_{diff}{}^*]_E} \quad \text{(S6.5)}$$

$$f_{free} = f_{diff} \times \frac{[Py_{free}{}^*]_M}{[Py_{diff}{}^*]_M} \quad \text{(S6.6)}$$

$$f_{E0} = f_{diff} \times \frac{[E0^*]_E}{[Py_{diff}{}^*]_E} \quad \text{(S6.7)}$$

$$f_D = f_{diff} \times \frac{[D^*]_E}{[Py_{diff}{}^*]_E} \quad \text{(S6.8)}$$

The molar fraction for one of the pyrene species obtained from a given fluorescence decay, such as the molar fraction f_{Mfree} used in Equation 6.2 in the main text for the Py_{free}* species in the monomer decays that represent the pyrene species Py_{diff}*, Py_{k2}*, and Py_{free}* would be obtained as shown in Equation S6.9.

$$f_{Mfree} = \frac{[Py_{free}{}^*]}{[Py_{diff}{}^*] + [Py_{k2}{}^*] + [Py_{free}{}^*]} = \frac{f_{free}}{f_{diff} + f_{k2} + f_{free}} \quad \text{(S6.9)}$$

Example global-fit of the monomer and excimer decays of Py-PAlaGlu according to the FBM

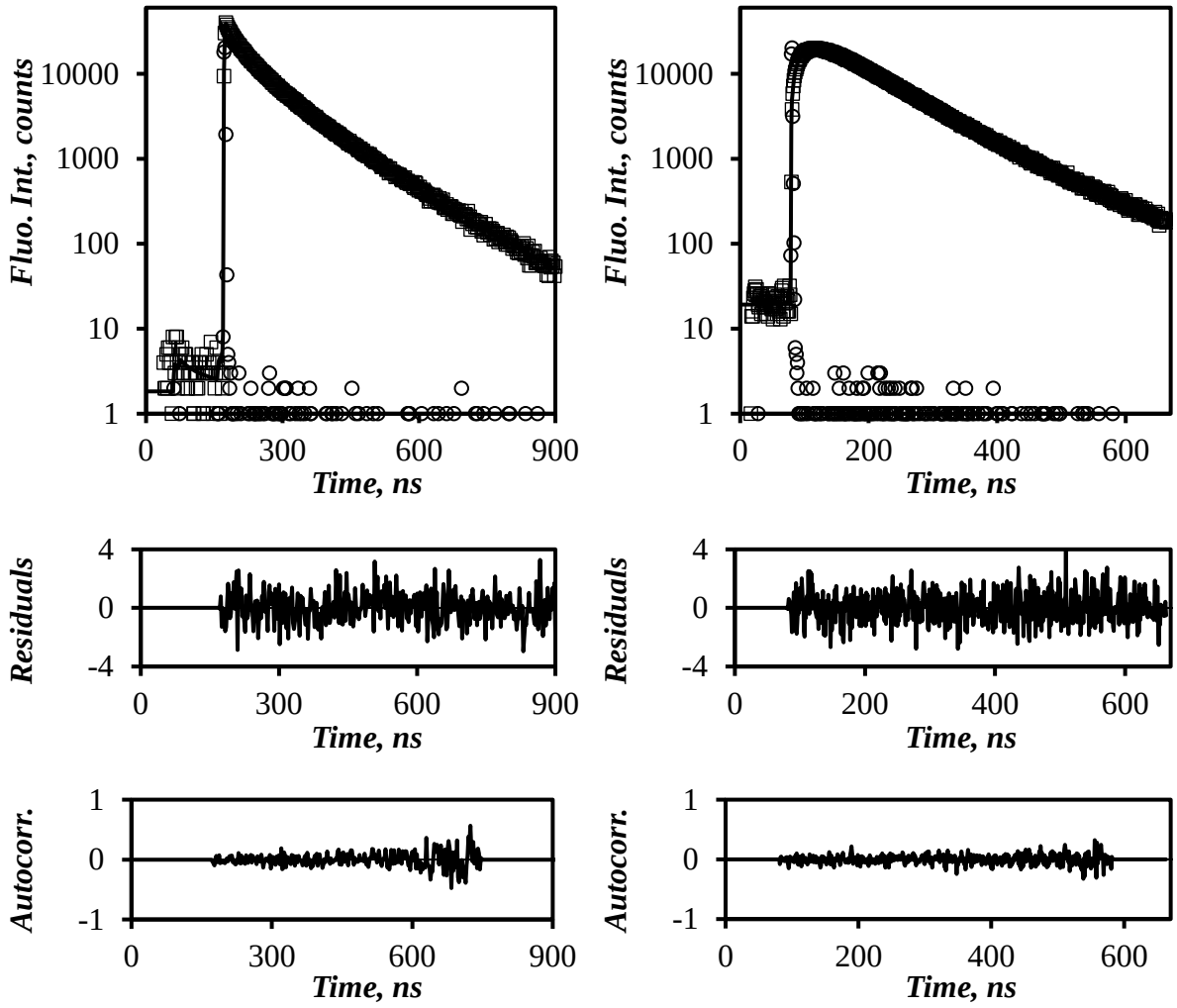


Figure S6.7. Monomer (left, λ_{em} = 375 nm) and excimer (right, λ_{em} = 510 nm) fluorescence decays (□) of Py-PAlaGlu (f_{Ala} = 0.24, x = 10.0 mol%) in DMSO. The instrument response function (○) and the fit of the global analysis of the FBM (—) are overlaid with the fluorescence decays. λ_{ex} = 344 nm, [Py] = 2.7×10^{-6} M. χ^2 = 1.16.

Parameters retrieved from the FBM analysis of the Py-PAlaGlu samples

Table S6.1. Parameters retrieved using the FBM analysis of the pyrene monomer decays of Py-PAlaGlu in acidified DMSO.

f_{Ala}	x (mol%)	k_2 (ns^{-1})	f_{Mk2}	$k_e[blob]$ (μs^{-1})	f_{Mdiff}	k_{blob} (μs^{-1})	$<n>$	f_{Mfree}	χ^2
0.24 τ_M = 172 ns	3.3	0.14	0.08	8.8	0.55	11.9	0.77	0.36	1.16
	3.8	0.14	0.09	8.9	0.58	12.6	0.76	0.32	1.16
	3.5	0.14	0.10	8.3	0.63	11.5	0.81	0.28	1.04
	4.5	0.14	0.12	8.0	0.67	11.6	0.86	0.22	1.24
	5.0	0.14	0.13	6.8	0.71	11.3	0.98	0.16	1.17
0.41 τ_M = 172 ns	6.6	0.13	0.18	5.4	0.72	9.7	1.16	0.10	1.15
	6.9	0.13	0.16	9.7	0.71	11.0	1.20	0.13	1.22
	7.9	0.13	0.16	6.4	0.75	9.5	1.30	0.09	1.19
	7.0	0.13	0.19	7.1	0.73	10.7	1.31	0.08	1.10
	10.4	0.13	0.26	6.4	0.70	10.7	1.76	0.04	1.06
0.58 τ_M = 172 ns	6.6	0.14	0.13	6.3	0.75	10.4	1.16	0.11	1.14
	9.6	0.14	0.17	5.0	0.77	8.9	1.53	0.05	1.14
	13.5	0.14	0.25	4.6	0.74	8.2	2.36	0.01	1.27
	15.2	0.14	0.25	4.6	0.73	8.7	2.40	0.01	1.15

Table S6.2. Parameters retrieved using the FBM analysis of the pyrene excimer decays of Py-PAlaGlu in acidified DMSO.

f_{Ala}	x (mol%)	f_{Ek2}	τ_{E0} (ns)	$f_{EdiffE0}$	f_{EE0}	τ_D (ns)	f_{ED}	χ^2
0.24	3.3	0.13	52	0.83	0.01	153	0.02	1.16
	3.8	0.13	54	0.83	0.04	103	0.00	1.16
	3.5	0.13	50	0.83	0.01	134	0.03	1.04
	4.5	0.14	52	0.82	0.04	80	0.00	1.24
	5.0	0.15	49	0.80	0.01	110	0.05	1.17
0.41	6.6	0.19	50	0.77	0.04	-	0.00	1.15
	6.9	0.17	47	0.78	0.04	-	0.00	1.22
	7.9	0.17	48	0.79	0.04	-	0.00	1.19
	7.0	0.20	49	0.77	0.03	-	0.00	1.10
	10.4	0.25	48	0.70	0.05	-	0.00	1.06
0.58	6.6	0.14	52	0.80	0.04	119	0.02	1.14
	9.6	0.18	52	0.78	0.03	141	0.01	1.14
	13.5	0.24	47	0.71	0.00	88	0.04	1.27
	15.2	0.24	49	0.70	0.05	94	0.02	1.15

Table S6.3. Fraction of pyrene species calculated using parameters retrieved from the FBM analysis of the pyrene monomer and excimer decays of Py-PAlaGlu in acidified DMSO.

f_{Ala}	x (mol%)	f_{k2}	f_{diffE0}	f_{diff}	f_{E0}	f_{D}	f_{agg}	f_{free}	χ^2
0.24	3.3	0.08	0.54	0.54	0.01	0.02	0.02	0.36	1.16
	3.8	0.09	0.57	0.57	0.03	0.00	0.03	0.31	1.16
	3.5	0.09	0.61	0.61	0.01	0.02	0.03	0.27	1.04
	4.5	0.11	0.64	0.64	0.03	0.00	0.03	0.21	1.24
	5.0	0.12	0.68	0.68	0.01	0.04	0.05	0.15	1.17
0.41	6.6	0.17	0.69	0.69	0.04	0.00	0.04	0.10	1.15
	6.9	0.15	0.68	0.68	0.03	0.00	0.03	0.13	1.22
	7.9	0.16	0.72	0.72	0.03	0.00	0.03	0.08	1.19
	7.0	0.18	0.70	0.70	0.03	0.00	0.03	0.08	1.10
	10.4	0.24	0.67	0.67	0.05	0.00	0.05	0.04	1.06
0.58	6.6	0.12	0.71	0.71	0.03	0.02	0.05	0.11	1.14
	9.6	0.17	0.74	0.74	0.03	0.01	0.04	0.05	1.14
	13.5	0.24	0.70	0.70	0.00	0.04	0.04	0.01	1.27
	15.2	0.24	0.69	0.69	0.05	0.02	0.07	0.01	1.15

Polymer Conformations and Degree of Polymerization

To ensure that the equation derived from renormalization group (RG) theory, which was developed for polymers with large chain lengths, would still hold true for a relatively small number of segments, Table S6.4 compares the number of conformations that a polymer can adopt for chain lengths of up to 20 structural units. The previously determined direct counting

method provides the exact number of conformations a chain of N segments can adopt in a 3D simple cubic lattice,[1] while the column labeled RG provides the result of Equation S6.4. In Equation S6.4, C = 1.17, γ = 7/6, and μ = 4.68.

$$\Omega(N) = CN^{\gamma-1}\mu^{N} \qquad \text{(S6.4)}$$

Unfortunately, the direct counting method which provides the exact number of conformations for shorter chains was only applied for isothermal solvents where μ = 4.68. Although μ = 3.19 for a poor solvent was used in the calculations described in the main text, the good agreement found in Table S4 between the exact number of conformations determined directly and with Equation S4 suggests that Equation S6.4 would also apply fairly well to shorter chains in poor solvents as it does in isothermal solvents.

Table S6.4. The number of conformations a polymer can adopt in a good solvent on a cubic lattice as a function of chain length calculated from renormalization group (RG) theory compared to a direct counting method.

Chain Length	*N*	Direct Counting[1]	RG	% Difference
2	1	6	5.5	8.7
3	2	30	29	4.0
4	3	150	144	3.7
5	4	726	709	2.3
6	5	3534	3449	2.4
7	6	16926	16652	1.6
8	7	81390	80025	1.7
9	8	387966	383255	1.2
10	9	1853886	1830675	1.3
11	10	8809878	8726417	0.9
12	11	41934150	41527241	1.0
13	12	1.9884×10^{8}	1.9735×10^{8}	0.8
14	13	9.4397×10^{8}	9.3675×10^{8}	0.8
15	14	4.4689×10^{9}	4.4421×10^{9}	0.6
16	15	2.1175×10^{10}	2.1046×10^{10}	0.6
17	16	1.0012×10^{11}	9.9643×10^{10}	0.5
18	17	4.7373×10^{11}	4.7145×10^{11}	0.5
19	18	2.2377×10^{12}	2.2293×10^{12}	0.4
20	19	1.0576×10^{13}	1.0536×10^{13}	0.4

List of proteins and their folding times

Table S6.5. List of peptides and proteins with their size, experimental[2,3] and calculated folding times, expected average *blob* size $\langle N_{blob} \rangle$, and expected number of *blobs*. $\langle N_{blob} \rangle$ was calculated using Equation 4 in the main text. The identity of each protein is provided as the code issued by the protein data bank (PDB) and if only part of the protein sequence was used in a folding experiment, it is indicated in the parentheses.

Identity	DP	Experimental $\log(\tau_F)$	Calculated $\log(\tau_F)$	$\langle N_{blob} \rangle$	# *blobs* $(=DP/\langle N_{blob} \rangle)$
1PGB (41 - 56)	16	-5.43	-6.46	13.40	1.19
1L2Y	20	-5.40	-5.75	14.53	1.38
α-Helix	21	-6.73	-5.61	14.74	1.42
1T8J	25	-5.12	-5.13	15.43	1.62
1PIN (6–39)	34	-4.08	-4.37	16.39	2.07
1VII (42–76)	35	-5.34	-4.31	16.46	2.13
1E0L	37	-4.39	-4.18	16.60	2.23
1E0M	37	-3.87	-4.18	16.60	2.23
1K9Q	40	-3.65	-4.00	16.79	2.38
1W4E	45	-4.43	-3.73	17.04	2.64
1PRB (7–53)	47	-5.99	-3.63	17.12	2.74
2WXC	47	-5.08	-3.63	17.12	2.74
1E0G	48	-3.04	-3.58	17.17	2.80
1OKS	49	-2.69	-3.53	17.20	2.85
1W4J	51	-5.34	-3.44	17.28	2.95
2WQG	51	-3.69	-3.44	17.28	2.95
2LLH (19–70)	52	-3.04	-3.39	17.31	3.00
1BA5	53	-2.56	-3.35	17.34	3.06
1IDY	54	-3.78	-3.30	17.37	3.11
1ENH	54	-4.60	-3.30	17.37	3.11
1DIV (1–56)	56	-2.84	-3.22	17.43	3.21
1SRL (9–64)	56	-1.91	-3.22	17.43	3.21
1PGB	56	-2.78	-3.22	17.43	3.21

1AVZ (Chain C: 85–141)	57	-2.13	-3.17	17.46	3.26
1SHG (6–62)	57	-0.48	-3.17	17.46	3.26
1JO8	58	-1.09	-3.13	17.49	3.32
1AU7 (103–160)	58	-4.21	-3.13	17.49	3.32
1FEX	59	-3.56	-3.09	17.52	3.37
1SS1 (2–60)	59	-4.99	-3.09	17.52	3.37
2PTL (18–77)	62	-1.78	-2.97	17.59	3.52
2L6R	62	-4.47	-2.97	17.59	3.52
3CI2	64	-2.52	-2.89	17.64	3.63
1C8C	64	-3.04	-2.89	17.64	3.63
2CRO (1–65)	65	-1.61	-2.85	17.66	3.68
1C9O	66	-3.13	-2.81	17.68	3.73
1G6P	66	-2.74	-2.81	17.68	3.73
2VKN (1–66)	66	-0.91	-2.81	17.68	3.73
1CSP	67	-2.82	-2.78	17.70	3.79
1RYK	69	-3.95	-2.70	17.74	3.89
1MJC	69	-2.30	-2.70	17.74	3.89
1UZC (3–71)	69	-3.47	-2.70	17.74	3.89
1DKT (5–76)	72	-1.95	-2.59	17.79	4.05
2A3D	73	-5.30	-2.55	17.81	4.10
1RFA (56–130)	75	-3.34	-2.48	17.84	4.20
2JMC	75	-1.43	-2.48	17.84	4.20
1M9S (391–466)	76	-1.74	-2.45	17.86	4.26
1UBQ	76	-2.30	-2.45	17.86	4.26
1LMB (6–85)	80	-4.52	-2.31	17.92	4.46
1KDX (586–666)	80	-3.39	-2.31	17.92	4.46
1O6X	81	-2.95	-2.27	17.93	4.52
1EHB (3–84)	82	-1.30	-2.24	17.95	4.57
1PNJ (3–84)	82	0.30	-2.24	17.95	4.57
1PRS (91–173)	83	0.87	-2.20	17.96	4.62
1HDN	85	-1.17	-2.14	17.99	4.73
2BKF (1–85)	85	-2.69	-2.14	17.99	4.73
1IMQ	86	-3.18	-2.10	18.00	4.78
1ST7	86	-3.69	-2.10	18.00	4.78
1AYI (1–86)	86	-2.48	-2.10	18.00	4.78
1NTI	86	-2.82	-2.10	18.00	4.78

1K8M (1–87)	87	0.41	-2.07	18.01	4.83
1PRS (1–88)	88	-1.30	-2.04	18.02	4.88
1K85	88	-0.61	-2.04	18.02	4.88
1GXT (4–91)	88	-1.91	-2.04	18.02	4.88
1TEN (803–891)	89	-0.78	-2.00	18.03	4.94
1BTA	89	-1.52	-2.00	18.03	4.94
1TIT	89	-1.56	-2.00	18.03	4.94
2BJD (12–101)	90	-0.74	-1.97	18.04	4.99
1FNF (1325–1415)	91	0.39	-1.94	18.06	5.04
3ZRT (3–93)	91	-0.56	-1.94	18.06	5.04
1DIV (58–149)	92	-1.43	-1.90	18.07	5.09
1TP3 (309–401)	93	-1.30	-1.87	18.08	5.14
1WIT	93	-0.18	-1.87	18.08	5.14
3ZRT (97–189)	93	-0.13	-1.87	18.08	5.14
2VH7 (5–98)	94	-0.36	-1.84	18.09	5.20
1GM1 (9–102)	94	-0.43	-1.84	18.09	5.20
1AUE (Chain B: 2022–2115)	94	-2.35	-1.84	18.09	5.20
1TTG	94	-2.39	-1.84	18.09	5.20
1FHT (2–97)	96	-2.00	-1.77	18.11	5.30
1N88	96	-0.87	-1.77	18.11	5.30
1FA3	96	-1.78	-1.77	18.11	5.30
1RIS (1–97)	97	-2.65	-1.74	18.12	5.35
2X7Z (311–407)	97	-0.32	-1.74	18.12	5.35
1HNG (2–98)	97	-0.78	-1.74	18.12	5.35
1APS	98	0.69	-1.71	18.13	5.41
2QJL	99	-1.13	-1.68	18.14	5.46
1E41_ (93–192)	100	-3.00	-1.65	18.15	5.51
1PUC (2–102)	101	-1.82	-1.61	18.15	5.56
1SPR (2–104)	103	-3.78	-1.55	18.17	5.67
1ARR (1–106)	106	-4.00	-1.46	18.20	5.83
1CUN (7–112)	106	-2.08	-1.46	18.20	5.83
1YYJ	106	-3.65	-1.46	18.20	5.83
2J5A (3–108)	106	-3.17	-1.46	18.20	5.83
1CUN (113–219)	107	-1.48	-1.43	18.20	5.88
1D6O	107	-0.69	-1.43	18.20	5.88

1BNI (3–110)	108	-1.17	-1.39	18.21	5.93
1U4Q (1662–1771)	110	-4.78	-1.33	18.23	6.03
1QAU (14–125)	112	-0.78	-1.27	18.24	6.14
2KDI	114	-1.00	-1.21	18.26	6.24
1QTU (1–115)	115	0.16	-1.18	18.26	6.30
2MYO	118	-2.08	-1.09	18.28	6.45
1HCD	118	-0.56	-1.09	18.28	6.45
2GA5	119	-2.35	-1.06	18.29	6.51
1EKG	119	-1.52	-1.06	18.29	6.51
1ADW	123	-0.30	-0.94	18.31	6.72
1J5U	124	-3.00	-0.91	18.32	6.77
1FGA (20–143)	124	0.61	-0.91	18.32	6.77
5L8I (3–127)	125	-0.28	-0.88	18.33	6.82
3O4D	126	-2.13	-0.85	18.33	6.87
2VIL	126	-1.82	-0.85	18.33	6.87
3O49	127	-0.74	-0.82	18.34	6.93
3CHY	128	-0.43	-0.79	18.34	6.98
1HRH (427–556)	130	-0.38	-0.73	18.35	7.08
1IFC	131	-1.87	-0.70	18.36	7.14
1OPA (1–133)	133	-0.61	-0.64	18.37	7.24
1CBI	136	1.39	-0.55	18.38	7.40
1RG8	137	-0.56	-0.52	18.39	7.45
3O4B (11–147)	137	-1.87	-0.52	18.39	7.45
2FS6	137	-1.00	-0.52	18.39	7.45
1NFI (67–206)	140	-0.78	-0.43	18.40	7.61
3H08	146	-0.69	-0.26	18.43	7.92
3F6R (2–148)	147	-1.52	-0.23	18.43	7.97
2A5E (9–156)	148	-1.52	-0.20	18.44	8.03
2PQE	149	-0.96	-0.17	18.44	8.08
1K0S	151	-3.21	-0.11	18.45	8.18
1A6N	151	-0.48	-0.11	18.45	8.18
1I1B (3–153)	151	1.74	-0.11	18.45	8.18
1DWR (1–152)	152	-1.26	-0.08	18.45	8.24
2RN2	155	0.13	0.00	18.47	8.39
1RA9	159	0.16	0.12	18.48	8.60
2KLL	160	0.61	0.15	18.48	8.66

1L63 (1–162)	162	-1.61	0.21	18.49	8.76
1LOP	164	-2.87	0.26	18.50	8.87
1FTG (2–169)	168	-1.00	0.38	18.51	9.08
1PHP (1–175)	175	-1.00	0.58	18.53	9.44
1PHP (186–394)	209	1.74	1.53	18.62	11.23
1IO2	213	0.11	1.64	18.63	11.44
1IGS (27–248)	222	1.95	1.89	18.64	11.91
1UCH (5–230)	226	1.13	2.00	18.65	12.12
1B9C (4–230)	227	1.13	2.03	18.65	12.17
1THF	253	1.39	2.75	18.69	13.53
3BLM	257	2.87	2.86	18.70	13.74
1V9E	259	1.91	2.91	18.70	13.85
4BLM (31–291)	261	2.04	2.97	18.70	13.95
1WQ5	268	0.91	3.16	18.71	14.32
1L8W (Chain B: 29–335)	307	-0.87	4.22	18.76	16.37

S7 – SI for Chapter 7

NMR of PyBuOSu

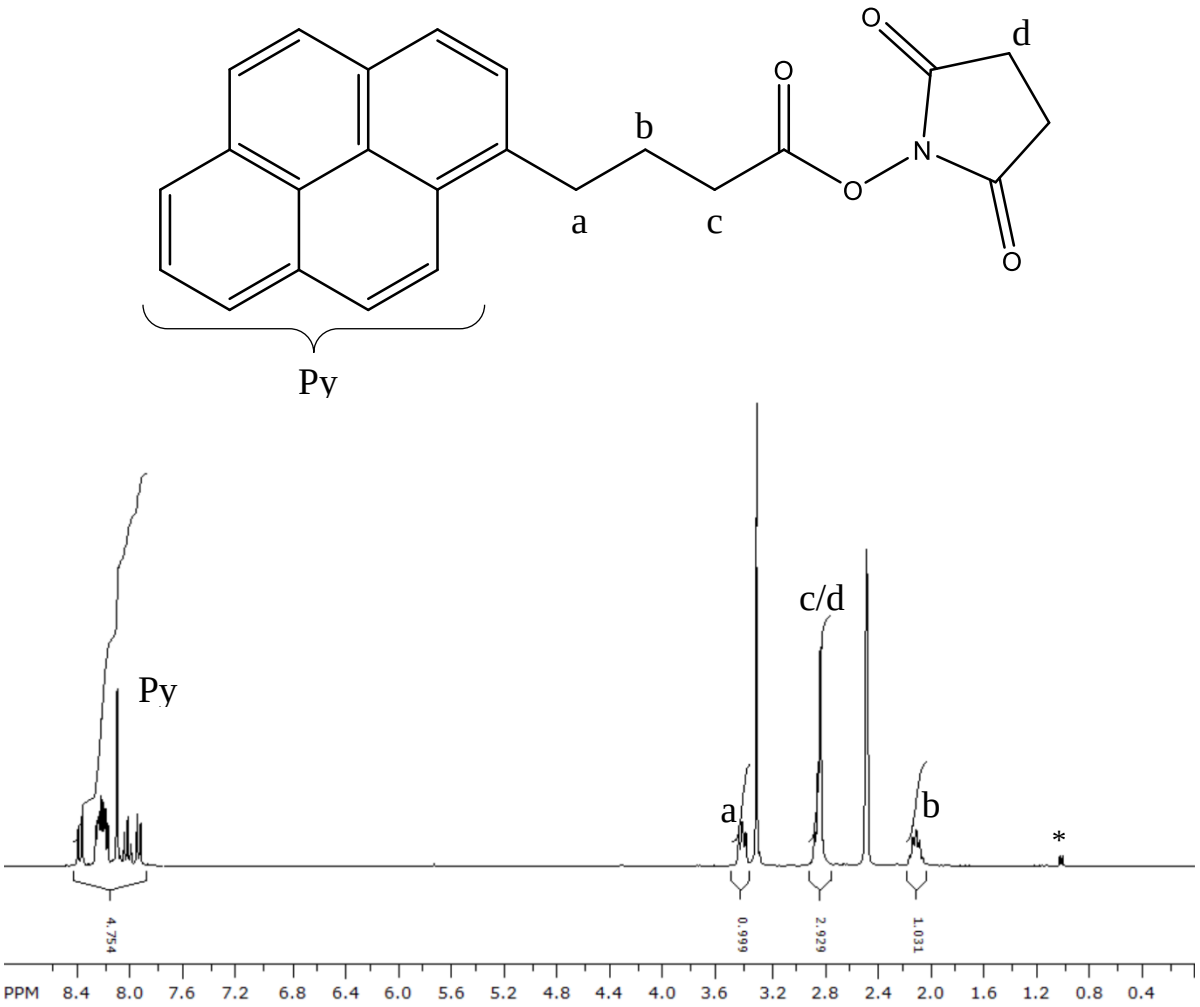


Figure S7.1. ^{1}H-NMR spectrum of PyBuOSu. 300 MHz, d_6-DMSO: δ = 8.4 – 7.9 (9H, m), 3.41 (2H, t), 2.9 – 2.8 (6H, m) and 2.10 (2H, p). The peak marked with an asterisk (*) is from residual isopropanol (1.04 ppm). Residual water (3.31 ppm) and solvent (2.47 ppm) peaks are also present.

Steady-state fluorescence spectra and their I_E/I_M ratios of PyBu-P*L*L in DMSO

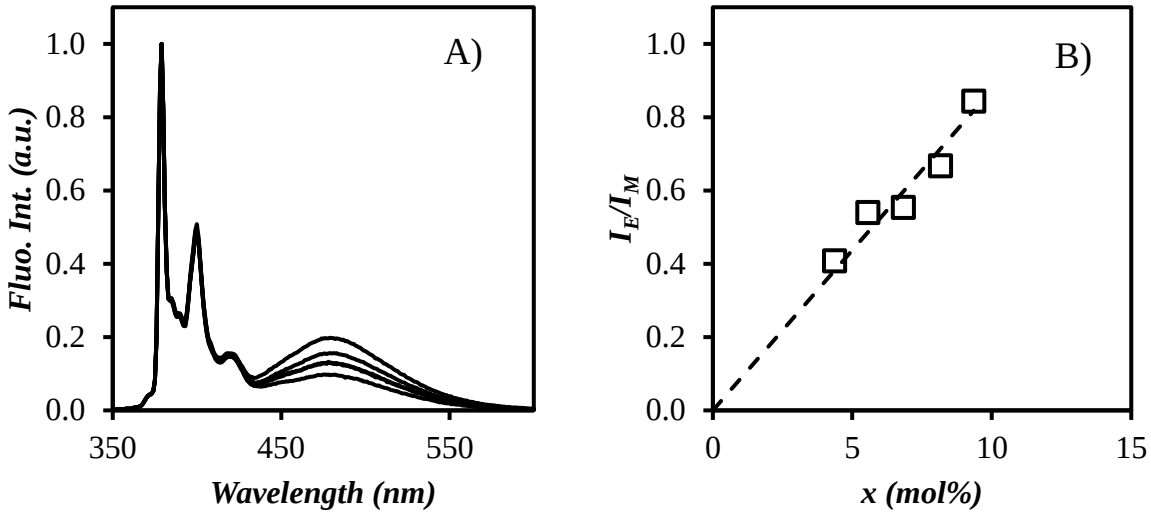


Figure S7.2. A) Steady state fluorescence spectra of PyBu-P*L*L in DMSO and B) I_E/I_M ratios of PyBu-P*L*L in DMSO as a function of pyrene content. λ_{ex} = 344 nm, [Py] = 2.4×10^{-6} M

Equations used to analyze the fluorescence decays of PyBu-P*L*L

The monomer and excimer decays of the PyBu-P*L*Ls were fit with Equations S7.1 and S7.2.

$$[Py^*]_{(t)} = [Py^*_{diff}]_{(t)} + [Py^*_{k2}]_{(t)} + [Py^*_{free}]_{(t)} = [Py^*_{diff}]_o \exp\left(-\left(A_2 + \frac{1}{\tau_M}\right)t - A_3\left(1-\exp(-A_4 t)\right)\right)$$

$$+\left([Py^*_{k2}]_o + [Py^*_{diff}]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\right)\exp\left(-\left(k_2+\frac{1}{\tau_M}\right)t\right)$$

$$-[Py^*_{diff}]_o e^{-A_3}\sum_{i=0}^{\infty}\frac{A_3^i}{i!}\frac{A_2+iA_4}{A_2+iA_4-k_2}\exp\left(-\left(A_2+iA_4+\frac{1}{\tau_M}\right)t\right)$$

$$+[Py^{*}_{free}]_o \exp\left(-\frac{t}{\tau_M}\right)\Bigg] \tag{S7.1}$$

$$[E^*]_{(t)} = [E0^*]_{(t)} + [D^*]_{(t)} + [ES^*]_{(t)} = k_2 \Bigg(\Bigg([Py^{*}_{k2}]_o + [Py^{*}_{diff}]_o e^{-A_3} \sum_{i=0}^{\infty} \frac{A_3^i}{i!} \frac{A_2 + iA_4}{A_2 + iA_4 - k_2} \Bigg)$$

$$\times \frac{\exp\left(-\frac{t}{\tau_{E0}}\right) - \exp\left(-\left(k_2 + \frac{1}{\tau_M}\right)t\right)}{k_2 + \frac{1}{\tau_M} - \frac{1}{\tau_{E0}}}$$

$$+[Py^{*}_{diff}]_o e^{-A_3} \sum_{i=0}^{\infty} \frac{A_3^i}{i!} \frac{A_2 + iA_4}{A_2 + iA_4 - k_2} \frac{\exp\left(-\left(A_2 + iA_4 + \frac{1}{\tau_M}\right)t\right) - \exp\left(-\frac{t}{\tau_{E0}}\right)}{A_2 + iA_4 + \frac{1}{\tau_M} - \frac{1}{\tau_{E0}}} \Bigg)$$

$$+ k_2 \Bigg(\Bigg([Py^{*}_{k2}]_o + [Py^{*}_{diff}]_o e^{-A_3} \sum_{i=0}^{\infty} \frac{A_3^i}{i!} \frac{A_2 + iA_4}{A_2 + iA_4 - k_2} \Bigg)$$

$$\times \frac{\exp\left(-\frac{t}{\tau_D}\right) - \exp\left(-\left(k_2 + \frac{1}{\tau_M}\right)t\right)}{k_2 + \frac{1}{\tau_M} - \frac{1}{\tau_D}}$$

$$+[Py^{*}_{diff}]_o e^{-A_3} \sum_{i=0}^{\infty} \frac{A_3^i}{i!} \frac{A_2 + iA_4}{A_2 + iA_4 - k_2} \frac{\exp\left(-\left(A_2 + iA_4 + \frac{1}{\tau_M}\right)t\right) - \exp\left(-\frac{t}{\tau_D}\right)}{A_2 + iA_4 + \frac{1}{\tau_M} - \frac{1}{\tau_D}} \Bigg)$$

$$+ [E0^*]_o \times \exp\left(-\frac{t}{\tau_{E0}}\right) + [D^*]_o \times \exp\left(-\frac{t}{\tau_D}\right) + [ES^*]_o \times \exp\left(-\frac{t}{\tau_S}\right) \tag{S7.2}$$

In Equations S7.1 and S7.2, the parameters A_2, A_3, and A_4 are given in Equations S7.3.a-c.

$$A_2 = < n > \times \frac{k_{blob} k_e [blob]}{k_{blob} + k_e [blob]} \quad \text{(S7.3.a)}$$

$$A_3 = < n > \times \left(\frac{k_{blob}}{k_{blob} + k_e [blob]} \right)^2 \quad \text{(S7.3.b)}$$

$$A_4 = k_{blob} + k_e [blob] \quad \text{(S7.3.c)}$$

In Equations S7.1 – S7.3, the parameters k_{blob}, $k_e[blob]$, and $<n>$ represent the rate constant for the diffusive encounters between two structural units bearing a pyrenyl label, the product of the rate constant for exchange (k_e) of ground-state pyrene between *blobs* and the local *blob* concentration, and the average number of pyrenyl labels inside a *blob*, respectively. Analysis of the fluorescence decays assumes that five pyrene species are present in solution, namely those pyrenes that are isolated and cannot form excimer (Py_{free}*), are subject to the diffusive motion of the structural units they are attached to (Py_{diffE0}* or Py_{diffD}* depending on whether they form a well (*E0**) or poorly (*D**) stacked dimer, respectively), or re-arrange quickly to form an excimer *E*0* or *D** with a rate constant k_2 (Py_{k2E0}* and Py_{k2E0}*, respectively). Finally, a short-lived decay is often observed in the excimer fluorescence decays and is attributed to a residual contribution from short-lived excimer (*ES**) or light scattering. It is typically not taken into account in the calculation of the molar fractions of the different

pyrene species. Analysis of the monomer fluorescence decays with Equation S7.1 yields the molar fractions f_{Mdiff}, f_{Mk2}, and f_{Mfree} of the pyrene species Py_{diff}*, Py_{k2}*, and Py_{free}*, where the M-subscript indicates that these pyrene species were detected in the monomer fluorescence decays. Analysis of the excimer fluorescence decays with Equation S7.2 yields the molar fractions f_{EdiffE0}, f_{EdiffD}, f_{Ek2E0}, f_{Ek2D}, f_{EE0}, f_{ED}, and f_{ES} of the pyrene species Py_{diffE0}*, Py_{diffD}*, Py_{k2E0}*, Py_{k2D}*, $E0$*, D*, and ES*, where the E-subscript indicates that these pyrene species were detected in the excimer fluorescence decays. The molar fractions f_{Mdiff}, f_{Mk2}, f_{Mfree}, f_{EdiffE0}, f_{EdiffD}, f_{Ek2E0}, f_{Ek2D}, f_{EE0}, f_{ED}, and f_{ES} obtained from the monomer or the excimer fluorescence decays were used to determine the molar fractions f_{diff}, f_{k2}, f_{free}, f_{E0}, and f_{D} whose expressions are given in Equations S7.4 – S7.8. The additional lifetime τ_{S} = 3.5 ns and its corresponding contribution to the excimer decay $[ES^*]_o$ was added to account for the presence of a short lifetime present in the excimer decay which was attributed to light scattering.

$$f_{diff} = \frac{1}{1 + \dfrac{f_{Mk2} + f_{Mfree}}{f_{Mdiff}} + \dfrac{f_{EE0} + f_{ED}}{f_{EdiffE0} + f_{EdiffD}}} \tag{S7.4}$$

$$f_{k2} = f_{diff} \times \frac{f_{Mk2}}{f_{Mdiff}} = f_{diff} \times \frac{f_{Ek2E0} + f_{Ek2D}}{f_{EdiffE0} + f_{EdiffD}} \tag{S7.5}$$

$$f_{free} = f_{diff} \times \frac{f_{Mfree}}{f_{Mdiff}} \tag{S7.6}$$

$$f_{E0} = f_{diff} \times \frac{f_{EE0}}{f_{EdiffE0} + f_{EdiffD}} \tag{S7.7}$$

$$f_D = f_{diff} \times \frac{f_{ED}}{f_{EdiffE0} + f_{EdiffD}} \quad \text{(S7.8)}$$

N_{blob} was calculated using Equation S7.9, where x is molar fraction of lysines bearing a pyrene label.

$$N_{blob} = \frac{< n >}{x}\left(1 - f_{Mfree}\right) \quad \text{(S7.9)}$$

Example of a global-fit of the monomer and excimer decays of PyBu-P*L*L according to the FBM

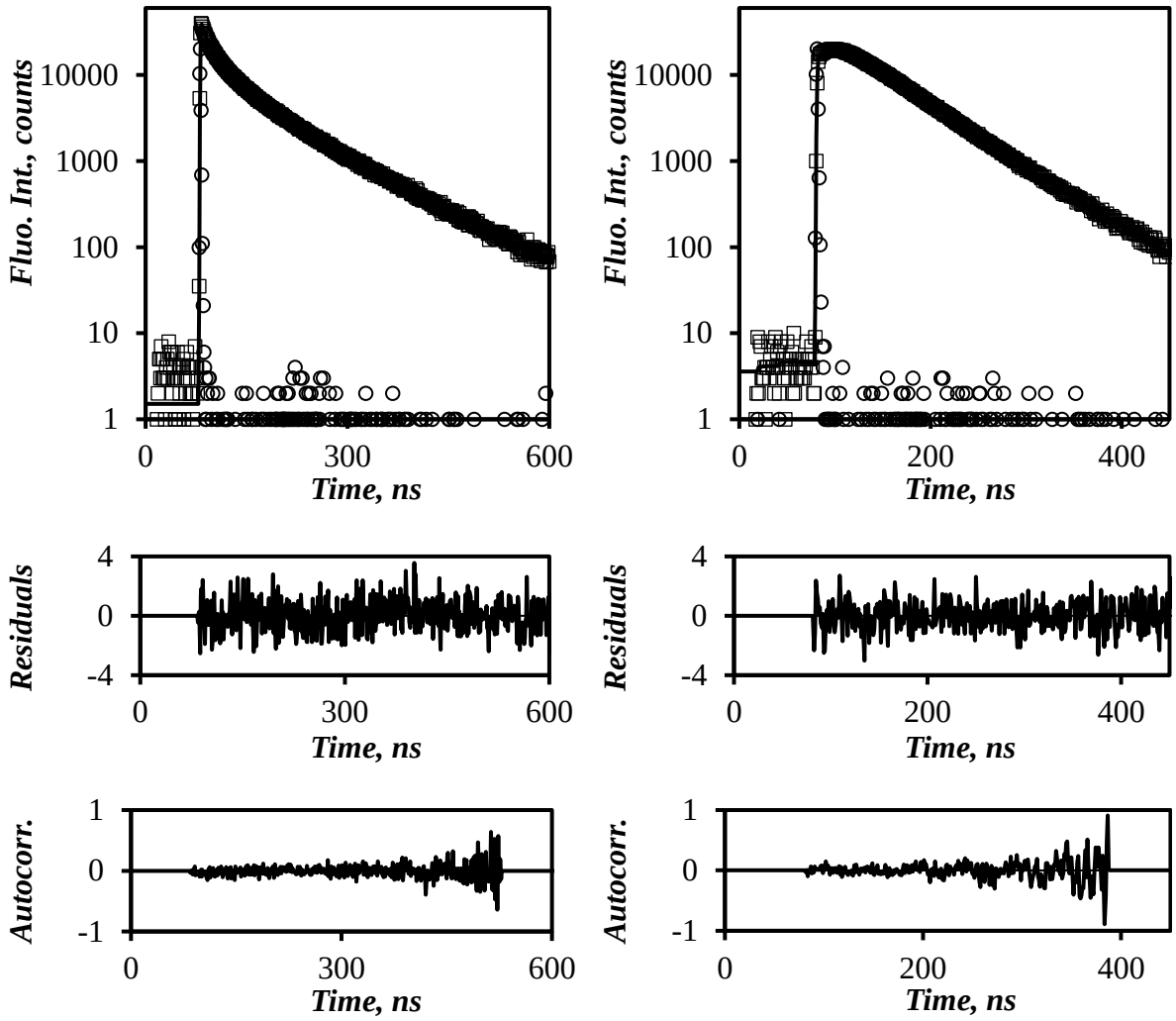


Figure S7.3. Monomer (left, λ_{em} = 378 nm) and excimer (right, λ_{em} = 510 nm) fluorescence decays (□) of PyBu-P*L*L (x = 9.4 mol%) in DMSO. The instrument response function (○) and the fit of the global analysis of the FBM (—) are overlaid with the fluorescence decays. λ_{ex} = 344 nm, [Py] = 2.4×10^{-6} M. χ^2 = 1.15.

Parameters retrieved from the FBM of PyBu-P*L*L

Table S7.1. Parameters retrieved using the FBM analysis of the pyrene monomer decays of PyBu-P*L*L in acidified DMSO.

Polypeptide	x (mol%)	k_2 (ns^{-1})	f_{Mk2}	$k_e[blob]$ (μs^{-1})	f_{Mdiff}	k_{blob} (μs^{-1})	$<n>$	f_{Mfree}	χ^2
PyBu-P*L*L τ_M = 135 ns	4.4	0.24	0.32	3.9	0.49	20.7	1.03	0.19	1.14
	5.5	0.24	0.33	7.9	0.61	24.3	1.18	0.05	1.15
	6.8	0.24	0.31	5.0	0.60	19.0	1.28	0.09	1.15
	8.2	0.24	0.31	6.9	0.66	19.9	1.32	0.03	1.16
	9.4	0.24	0.34	5.4	0.63	20.9	1.66	0.03	1.15

Table S7.2. Parameters retrieved using the FBM analysis of the pyrene excimer decays of PyBu-P*L*L in acidified DMSO.

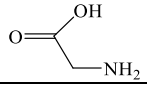
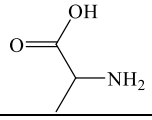
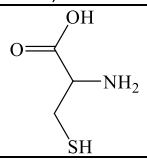
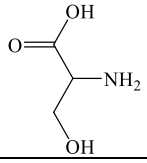
Polypeptide	x (mol%)	f_{Ek2}	τ_{E0} (ns)	$f_{EdiffE0}$	f_{EE0}	τ_D (ns)	f_{EdiffD}	f_{ED}	τ_S (ns)	f_S	χ^2
PyBu-P*L*L τ_M = 135 ns	4.4	0.27	33.4	0.16	0.03	45.2	0.25	0.00	3.5	0.29	1.14
	5.5	0.22	37.0	0.19	0.10	38.4	0.21	0.05	3.5	0.24	1.15
	6.8	0.26	37.8	0.15	0.03	54.4	0.30	0.00	3.5	0.26	1.15
	8.2	0.23	40.6	0.18	0.00	41.7	0.30	0.10	3.5	0.20	1.16
	9.4	0.24	36.9	0.10	0.00	47.1	0.34	0.13	3.5	0.19	1.15

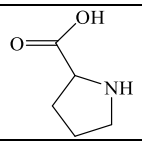
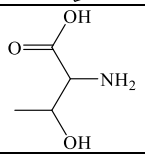
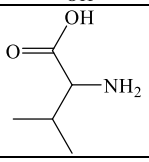
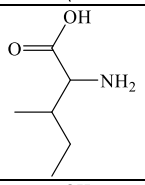
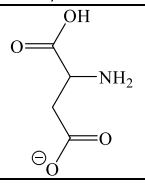
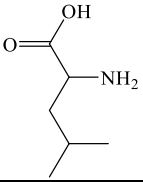
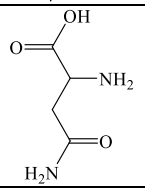
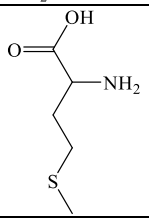
Table S7.3. Fraction of pyrene species calculated using parameters retrieved from the FBM analysis of PyBu-P*L*L in acidified DMSO.

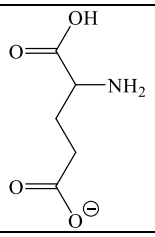
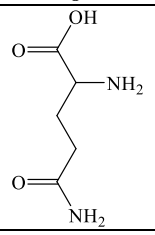
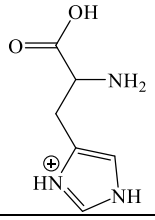
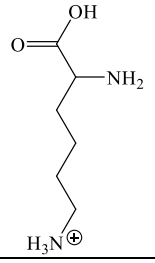
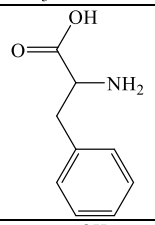
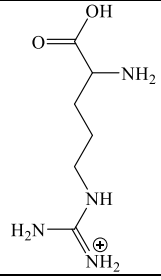
Polypeptide	x (mol%)	f_{k2}	f_{diffE0}	f_{diffD}	f_{diff}	f_{E0}	f_{D}	f_{agg}	f_{free}	χ^2
PyBu-P*L*L τ_M = 135 ns	4.4	0.31	0.18	0.29	0.47	0.04	0.00	0.04	0.19	1.14
	5.5	0.27	0.23	0.27	0.50	0.12	0.06	0.18	0.04	1.15
	6.8	0.30	0.19	0.39	0.58	0.03	0.00	0.04	0.08	1.15
	8.2	0.28	0.21	0.36	0.57	0.00	0.12	0.12	0.03	1.16
	9.4	0.28	0.12	0.41	0.53	0.00	0.16	0.16	0.03	1.15

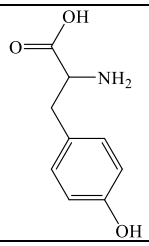
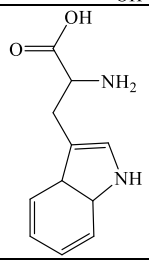
Amino acid side chain lengths and sizes

Table S7.4. List of the common 20 amino acids, their side chain reach and size, N(SCR), and f_b(SCS) values. N(SCR) and f_b(SCS) were calculated using Figures 7A and B, respectively, in the main text.

Amino Acid	Structure	Side Chain Reach (SCR)	N(SCR)	Side Chain Size (SCS)	f_b(SCS)
Glycine		1	0.45	0	2.16
Alanine		2	0.89	1	1.46
Cysteine		3	1.34	2	1.26
Serine		3	1.34	2	1.26

Proline		3	1.34	3	1.14
Threonine		3	1.34	3	1.14
Valine		3	1.34	3	1.14
Isoleucine		3	1.78	4	1.06
Aspartic acid		3	1.34	4	1.06
Leucine		4	1.78	4	1.06
Asparagine		4	1.78	4	1.06
Methionine		5	2.23	4	1.06

Glutamic acid		4	1.78	5	1.00
Glutamine		5	2.23	5	1.00
Histidine		5	2.23	6	1.00
Lysine		6	2.68	5	1.00
Phenylalanine		6	2.68	7	1.00
Arginine		7	3.12	7	1.00

Tyrosine		7	3.12	8	1.00
Tryptophan		7	3.12	10	1.00

Example procedure for the calculation of $\langle N_{blob} \rangle$ and $\langle n_b \rangle$

The N_0^l, N_0^r, and N_{blob} values for each *aa* in the sequence of the C-terminal β-hairpin of protein G (residues 41 – 56; protein data bank (PDB) code: 1PGB) is provided in Table S7.5. A depiction of N_0^l, N_0^r, and N_{blob} is given as Figure 8 in the main text. The values in Table S7.5 were then used to determine $\langle N_{blob} \rangle$ and $\langle n_b \rangle$ for the protein.

Table S7.5. Example procedure for determining a proteins N_{blob} and n_b values. The N_0^l, N_0^r, and N_{blob} values calculated for each *aa* in the sequence of the C-terminal β-hairpin of protein G (residues 41 – 56, PDB code 1PGB).[1] The protein is divided into *blobs* based on the *aa* N_{blob} values. In this example, the protein is divided into two *blobs*, one generated by tyrosine (Y) at position 5 and the other by valine (V) at position 14. The protein contains $n_b = 2$ *blobs* with an average N_{blob} of 9.5 (=(12+7)/2).

Sequence	**G**	**E**	**W**	**T**	**Y**	**D**	**D**	**A**	**T**	**K**	**T**	**F**	**T**	**V**	**T**	**E**
aa position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
N_0^l	0	1	2	3	4	3	4	5	6	7	6	7	3	4	3	4
N_0^r	4	3	7	3	7	4	5	4	3	4	3	4	3	2	1	0
aa N_{blob}	5	5	10	7	12	8	10	10	10	12	10	12	7	7	5	5
blob 1					**12**											
Remainder													7	7	5	5
blob 2														**7**		

The first step used to calculate $\langle N_{blob} \rangle$ and $\langle n_b \rangle$ was to determine the largest N_{blob} value in the protein. The largest N_{blob} value for the 1PGB fragment in Table S7.5 is 12, which was found for the *aa*'s at position 5, 10, and 12. Since there is more than one *aa* with the same N_{blob}

value, one was picked at random, and all *aa's* within N_0^l and N_0^r *aa*'s from the chosen *aa* were removed from the sequence. Assuming that tyrosine (Y) at position 5 was chosen, the 4 *aa*'s to the left of tyrosine were removed from the sequence since $N_0^l = 4$. Similarly, since the N_0^r value of tyrosine was 7, the 7 *aa's* to the right were also removed. The 12 (= 7+3+1) *aa's* removed from the sequence marked the first *blob* (*blob* 1 in Table S7.5, highlighted in green). This left behind the *aa*'s at positions 13 to 16. Of the remaining sequence, 7 was the largest N_{blob} value, which was found for the *aa*'s at position 13 and 14. Again, since more than one *aa* shared the same N_{blob} value, one was chosen at random. Assuming that valine (V) at position 14 was selected, the *aa's* located N_0^l and N_0^r *aa*'s to, respectively, the left and right of valine were removed to yield *blob* 2 in Table S7.5, highlighted in blue, which consumed the remainder of the sequence. As a side note, the N_0^l value of valine extended past the remaining sequence by 3 *aa*'s (positions 10 – 12 in Table S7.5). Consequently, these 3 *aa*'s are expected to be contained inside both *blobs* and therefore contribute to the N_{blob} values of both *blobs*. Now that the entire sequence had been segmented into *blobs*, the number of *blobs* n_b was determined by tallying the *blobs* and N_{blob} was calculated by averaging the N_{blob} values of the *blobs*. In this case, since the entire sequence was contained within two *blobs* $n_b = 2$, and $N_{blob} = (12+7)/2 = 9.5$.

The above procedure was repeated 1000 times, each time choosing a random *aa* among those sharing the largest remaining N_{blob} value. The 1000 iterations of N_{blob} and n_b calculations were than averaged to obtain $\langle N_{blob} \rangle$ and $\langle n_b \rangle$ for the protein. For the 1PGB fragment in Table S7.5, the 1000 iterations led to $\langle N_{blob} \rangle$ and $\langle n_b \rangle$ values of 9.19 and 2.36, respectively.

Additional discussion on the *blob* size calculations

As discussed in the main text, N_0 was defined as the number of *aa*'s separating a reference *aa* from another *aa* while still allowing the two *aa*'s to interact with each other. As a result, a *blob* can contain an *aa* which might not directly interact with the reference *aa*. An example of this effect would be a sequence containing a glutamic acid, alanine, and lysine in positions *i*, *i*+5, and *i*+6, respectively. If alanine has the smallest SCS in the sequence, f_b(SCS) equals 1.46 (Table S7.4). Since [*N*(glutamic acid) + *N*(alanine)]×f_b(SCS) = (1.78 + 0.89)×1.46 = 3.90 is less than 5 (the number of *aa*'s separating alanine from glutamic acid), alanine at position *i*+5 is predicted to not be able to interact with the glutamic acid at position *i*. However, [*N*(glutamic acid) + *N*(lysine)]×f_b(SCS) = (1.78 + 2.68)×1.46 = 6.51 is greater than 6, implying that lysine at position *i*+6 can interact with glutamic acid. This means that the N_0 value of glutamic acid will include alanine, although these two *aa*'s are not predicted to directly interact with one another. Although some smaller *aa*'s might not be able to touch the reference *aa* inside a *blob*, these small *aa*'s would still contribute to increasing the number of conformations taken by the oligopeptide segment constituting the *blob* and must be included in the calculations.

To determine the average *blob* size representative of a protein, the *aa*'s included in N_0^l and N_0^r of the *aa* with the largest N_{blob} value were considered to be contained within its *blob.* This had two implications. The first was that the *aa*'s near a *blob*'s periphery were expected to interact with other *aa*'s outside the volume of the *blob*. Second, the *blobs* in a protein will often overlap with one another, meaning that some *aa*'s will be contained in more than one *blob*.

This interpretation of the *blobs* suggests that a protein should not contain discrete segments of interacting *aa*'s but rather that the *aa*'s interact with one another in a continuum and that $<N_{blob}>$ simply represents the ensemble average number of *aa*'s capable of interacting with one another. Although the 'double counting' of some *aa*'s shared by neighboring *blobs* towards $<N_{\mathrm{blob}}>$ might first appear to artificially increase the total number of conformation Ω of the protein, the *aa*'s contained in more than one *blob* do contribute to the Ω value of each *blob*, since these *aa*'s must adopt their preferred conformation in all *blobs* involved, and as such must be counted toward the final Ω.

Lastly, since the predicted folding times depended on experimental results, there are bound to be errors in the predicted values. To ensure that the predicted τ_f's would not change too much if the constraints defining the contacts between two *aa*'s were relaxed, the calculation of folding times were repeated assuming that the *aa*'s could interact with one another if the reach between them was extended by Δl bond lengths as depicted in Figure S4. The number of *aa*'s contained within Δl bond lengths is equal to $N(\Delta l)\times f_b(\mathrm{SCS})$, thereby extending the N_0 values by $N(\Delta l)\times f_b(\mathrm{SCS})$. The new relationships defining the N_0's are provided in Figure S4. A negative Δl would illustrate a scenario where the $N(\mathrm{SCR})$ values overestimated the number of *aa*'s, while a positive Δl would represent either an underestimation of $N(\mathrm{SCR})$ or a situation where the side chains could interact with one another without directly touching each other, such as through hydrogen bonding with one another or with a nearby molecule (ex. water).

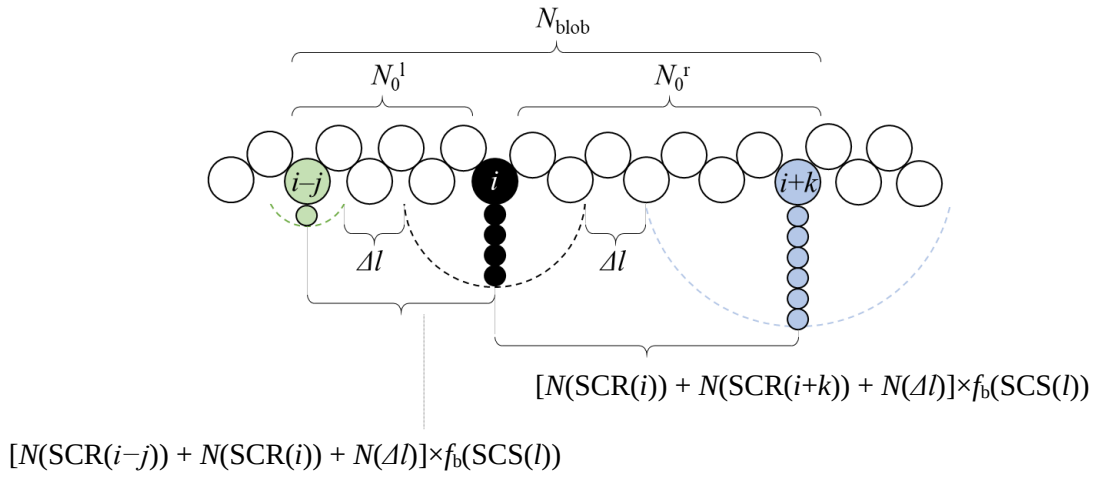


Figure S7.4. Schematic diagram of the N_{blob} calculation assuming that *aa*'s could interact with one another by a distance defined by their SCRs plus an additional Δl bond lengths.

To assess how changes in Δl affected the results, the correlation coefficient between the logarithms of calculated and experimental folding times and their average difference was plotted as a function of Δl in Figures S5A and B, respectively. Figure S5A shows that r was similar (~0.74) for all the Δl values ranging from −1 to 3, indicating that the correlation was little affected by this range of Δl values. The average difference between the calculated and experimental folding times in Figure S5B shows that the lowest difference was obtained when $\Delta l = 0$, and the further the deviation from 0, the larger the difference. However, there is almost no change in the predicted folding times for small deviations in Δl. Overall, small Δl values had a minimal effect on the parameters in Figures S5A and B, demonstrating that the method used to determine the τ_f's is stable to perturbations in *blob* sizes due to potential effects not included in the current program.

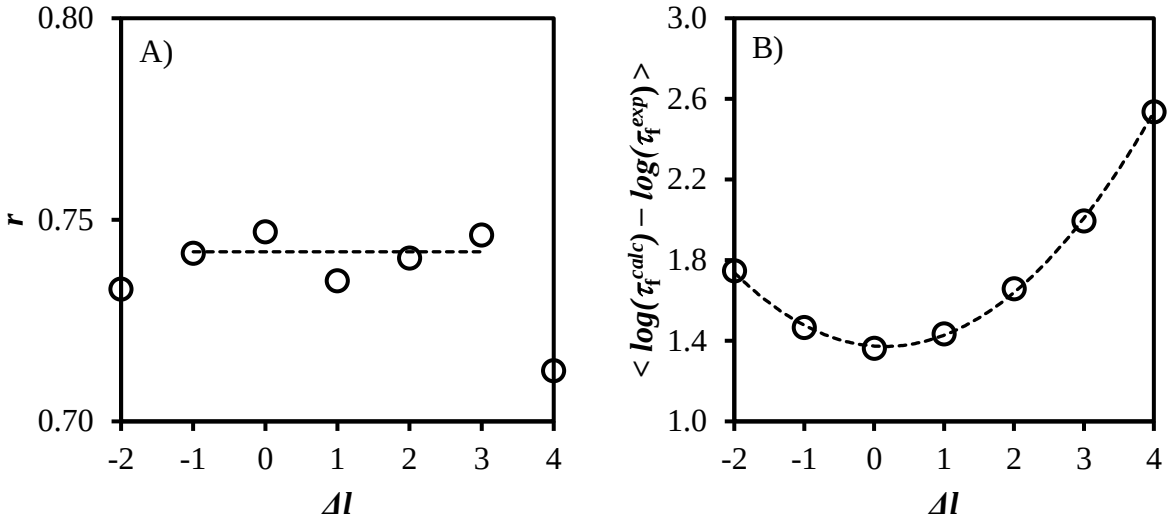


Figure S7.5. Plots of A) the correlation coefficient (*r*) and B) the average difference between the logarithms of the τ_f's calculated using Equation 6 in the main text (τ_f^{calc}) and the experimentally measured ones (τ_f^{exp}) as the constraints of the N_0 values defining the N_{blob} values changed by the length of a single atom (Δl). N_0^l and N_0^r were defined by the largest integer satisfying the conditions $j \leq [N(\mathrm{SCR}(i-j))+N(\mathrm{SCR}(j))+N(\Delta l)]\times f_b(\mathrm{SCS}(i\text{-}j\rightarrow i))$ and $k \leq [N(\mathrm{SCR}(i))+N(\mathrm{SCR}(i+k))+ N(\Delta l)]\times f_b(\mathrm{SCS}(i\rightarrow i+k))$, respectively.

Proteins and their folding times

Table S7.6. List of peptides and proteins with their size, experimental[2,3] and calculated folding times, the average *blob* size $\langle N_{blob} \rangle$, and average number of *blobs* $\langle n_b \rangle$. The identity lists the code from the protein data bank (PDB) and if only part of the sequence was used for the folding experiment, it is indicated in the parentheses.

Identity	DP	Experimental $\log(\tau_F)$	Calculated $\log(\tau_F)$	$\langle N_{blob} \rangle$	$\langle n_b \rangle$
1PGB (41 - 56)	16	-5.43	-8.2	9.19	2.36
1L2Y	20	-5.40	-7.3	12.00	2.00
α-Helix	21	-6.73	-7.1	4.75	6.06
1T8J	25	-5.12	-5.6	10.83	3.00
1PIN (6–39)	34	-4.08	-3.2	17.50	2.00
1VII (42–76)	35	-5.34	-5.3	19.50	2.00
1E0L	37	-4.39	-5.8	17.33	3.00
1E0M	37	-3.87	-3.3	13.00	4.00
1K9Q	40	-3.65	-4.6	16.33	3.00
1W4E	45	-4.43	-4.5	14.96	4.51
1PRB (7–53)	47	-5.99	-5.3	10.83	6.00
2WXC	47	-5.08	-5.1	13.25	4.00
1E0G	48	-3.04	-3.9	18.88	3.34
1OKS	49	-2.69	-5.0	11.47	6.26
1W4J	51	-5.34	-4.0	12.33	6.00
2WQG	51	-3.69	-4.5	13.75	4.00
2LLH (19–70)	52	-3.04	-4.2	14.47	4.47
1BA5	53	-2.56	-4.2	15.00	5.00
1IDY	54	-3.78	-3.6	13.31	5.00
1ENH	54	-4.60	-4.4	12.73	6.01
1DIV (1–56)	56	-2.84	-3.8	14.00	6.00
1SRL (9–64)	56	-1.91	-2.9	15.39	4.35
1PGB	56	-2.78	-4.1	14.02	5.77
1AVZ (Chain C: 85–141)	57	-2.13	-2.3	15.71	4.83
1SHG (6–62)	57	-0.48	-4.3	15.18	5.48

1JO8	58	-1.09	-3.6	12.15	7.00
1AU7 (103–160)	58	-4.21	-3.9	12.83	6.00
1FEX	59	-3.56	-3.9	12.42	6.00
1SS1 (2–60)	59	-4.99	-4.2	11.54	7.32
2PTL (18–77)	62	-1.78	-3.3	13.33	6.00
2L6R	62	-4.47	-2.7	15.80	5.00
3CI2	64	-2.52	-4.3	12.86	7.00
1C8C	64	-3.04	-2.5	17.39	5.00
2CRO (1–65)	65	-1.61	-3.7	13.85	7.00
1C9O	66	-3.13	-2.8	20.00	4.00
1G6P	66	-2.74	-2.4	17.76	5.32
2VKN (1–66)	66	-0.91	-2.9	15.70	5.48
1CSP	67	-2.82	-2.4	17.00	5.00
1RYK	69	-3.95	-2.8	16.55	5.77
1MJC	69	-2.30	-1.3	16.91	5.00
1UZC (3–71)	69	-3.47	-3.5	13.43	7.00
1DKT (5–76)	72	-1.95	-2.6	11.97	8.51
2A3D	73	-5.30	-2.9	15.00	6.00
1RFA (56–130)	75	-3.34	-2.9	17.00	6.00
2JMC	75	-1.43	-3.0	16.47	5.37
1M9S (391–466)	76	-1.74	-2.0	19.80	5.00
1UBQ	76	-2.30	-2.6	13.29	7.66
1LMB (6–85)	80	-4.52	-2.8	14.01	7.00
1KDX (586–666)	80	-3.39	-3.1	13.30	7.88
1O6X	81	-2.95	-2.6	12.89	8.33
1EHB (3–84)	82	-1.30	-2.4	15.57	7.00
1PNJ (3–84)	82	0.30	-1.9	17.83	6.00
1PRS (91–173)	83	0.87	-2.9	15.00	7.00
1HDN	85	-1.17	-2.4	14.53	7.51
2BKF (1–85)	85	-2.69	-2.6	12.22	9.22
1IMQ	86	-3.18	-3.1	12.71	9.48
1ST7	86	-3.69	-2.4	13.87	8.25
1AYI (1–86)	86	-2.48	-3.0	12.37	9.24
1NTI	86	-2.82	-2.0	13.47	8.27
1K8M (1–87)	87	0.41	-2.3	15.43	7.00
1PRS (1–88)	88	-1.30	-2.0	14.62	8.50
1K85	88	-0.61	-2.6	11.23	10.39

1GXT (4–91)	88	-1.91	-1.9	14.71	7.00
1TEN (803–891)	89	-0.78	-2.6	14.75	8.53
1BTA	89	-1.52	-2.4	15.71	7.00
1TIT	89	-1.56	-2.2	13.34	9.09
2BJD (12–101)	90	-0.74	-1.7	18.01	7.00
1FNF (1325–1415)	91	0.39	-1.5	15.50	8.00
3ZRT (3–93)	91	-0.56	-2.2	12.98	9.46
1DIV (58–149)	92	-1.43	-2.6	13.22	9.00
1TP3 (309–401)	93	-1.30	-0.6	16.13	8.00
1WIT	93	-0.18	-2.3	15.38	8.00
3ZRT (97–189)	93	-0.13	-2.1	15.06	8.67
2VH7 (5–98)	94	-0.36	-2.0	16.38	8.00
1GM1 (9–102)	94	-0.43	-1.6	17.86	7.00
1AUE (Chain B: 2022–2115)	94	-2.35	-1.3	17.86	7.00
1TTG	94	-2.39	-1.9	14.50	8.00
1FHT (2–97)	96	-2.00	-2.0	14.43	9.24
1N88	96	-0.87	-1.8	15.35	8.73
1FA3	96	-1.78	-0.4	17.27	7.48
1RIS (1–97)	97	-2.65	-1.3	14.11	9.00
2X7Z (311–407)	97	-0.32	-2.0	15.21	8.49
1HNG (2–98)	97	-0.78	-1.5	14.99	8.54
1APS	98	0.69	-1.6	18.13	8.00
2QJL	99	-1.13	-1.9	13.67	9.00
1E41 (93–192)	100	-3.00	-2.1	12.99	10.00
1PUC (2–102)	101	-1.82	-1.7	14.29	9.25
1SPR (2–104)	103	-3.78	-2.2	16.51	8.84
1ARR (1–106)	106	-4.00	-1.7	16.25	9.00
1CUN (7–112)	106	-2.08	-1.2	14.43	9.51
1YYJ	106	-3.65	-1.8	12.48	11.52
2J5A (3–108)	106	-3.17	-1.8	13.00	12.00
1CUN (113–219)	107	-1.48	-1.3	12.71	11.78
1D6O	107	-0.69	-0.3	16.98	8.50
1BNI (3–110)	108	-1.17	-1.3	18.12	7.35
1U4Q (1662–1771)	110	-4.78	-1.6	13.53	11.29
1QAU (14–125)	112	-0.78	-1.8	15.58	9.25

2KDI	114	-1.00	-0.8	14.92	11.06
1QTU (1–115)	115	0.16	-0.6	14.60	10.00
2MYO	118	-2.08	-0.7	16.22	9.66
1HCD	118	-0.56	-0.8	15.90	10.00
2GA5	119	-2.35	-0.8	12.68	11.83
1EKG	119	-1.52	-0.8	14.14	11.00
1ADW	123	-0.30	-0.7	13.90	11.00
1J5U	124	-3.00	-0.9	15.20	11.36
1FGA (20–143)	124	0.61	0.4	16.50	10.00
5L8I (3–127)	125	-0.28	-0.8	19.22	9.00
3O4D	126	-2.13	0.1	16.90	10.00
2VIL	126	-1.82	-0.2	17.44	9.00
3O49	127	-0.74	-0.2	17.60	10.00
3CHY	128	-0.43	-0.6	14.38	12.00
1HRH (427–556)	130	-0.38	-0.6	13.67	13.67
1IFC	131	-1.87	0.2	18.89	9.00
1OPA (1–133)	133	-0.61	0.3	17.80	9.93
1CBI	136	1.39	0.2	13.80	13.00
1RG8	137	-0.56	0.9	17.25	10.00
3O4B (11–147)	137	-1.87	0.7	17.45	11.00
2FS6	137	-1.00	0.7	14.32	13.15
1NFI (67–206)	140	-0.78	0.3	14.07	14.00
3H08	146	-0.69	0.9	14.46	13.00
3F6R (2–148)	147	-1.52	0.5	13.61	13.68
2A5E (9–156)	148	-1.52	1.5	18.12	10.95
2PQE	149	-0.96	0.5	18.09	11.00
1K0S	151	-3.21	0.2	13.59	14.35
1A6N	151	-0.48	0.7	15.13	13.34
1I1B (3–153)	151	1.74	0.5	13.80	15.00
1DWR (1–152)	152	-1.26	0.6	15.34	13.00
2RN2	155	0.13	1.2	15.56	14.51
1RA9	159	0.16	1.2	14.67	15.00
2KLL	160	0.61	0.4	12.18	17.22
1L63 (1–162)	162	-1.61	1.4	15.70	13.00
1LOP	164	-2.87	0.5	15.56	14.76
1FTG (2–169)	168	-1.00	1.6	17.07	14.00
1PHP (1–175)	175	-1.00	1.6	17.25	12.76

1PHP (186–394)	209	1.74	3.1	15.43	18.17
1IO2	213	0.11	3.4	15.32	17.50
1IGS (27–248)	222	1.95	3.7	14.27	19.64
1UCH (5–230)	226	1.13	4.8	13.91	22.52
1B9C (4–230)	227	1.13	4.2	16.62	18.22
1THF	253	1.39	4.8	15.14	22.39
3BLM	257	2.87	4.9	16.26	21.79
1V9E	259	1.91	5.2	16.44	21.00
4BLM (31–291)	261	2.04	5.7	15.57	21.46
1WQ5	268	0.91	5.3	14.98	22.69
1L8W (Chain B: 29–335)	307	-0.87	7.6	15.40	26.54

Python program used to determine <N_{blob}> and <n_b>

```
"""
@author: Remi Casier
Note: lines beginning with '#' indicate comments, and were added to aid in
understanding the code
"""
#imported libraries
import numpy as np
import random

#the main program
def mainprogram():
# number of atoms of additional separation - converted to length in terms
of number of aa's using SCL relationship
    additional_separation = 0
# number of times the program will calculate <nb>
    numberrepeats = 1000
# beginning of printout
    print('protein#, Tf, DP, avgNblob, Avg#blobs')
# repeat program for each protein listed below
    for pp in range(0,sequences.shape[0]):
# array of <nb> values for one protein
        numberofblobs = []
# array of <Nblob> values for one protein
        nblobvalue = []
#define the aa reaches
        ireach = aareachpoly(sequences[pp][1])
#left reaches, accounts for left end
        leftreach= aaleftreachpoly(ireach)
#right reaches, accounts for right end
        rightreach= aarightreachpoly(ireach)
```

```
  #calculate N0 to the right
          N0right = []
          for A1 in range(0,len(rightreach)):
              SCS = AAsize[sequences[pp][1][A1]]
  # counts aa's in a blob iff the aa's can touch
              count = 0
  # furthest distance two aa's can touch
              count2 = 0
              for A2 in range(A1+1, len(rightreach)):
                  SCS2 = AAsize[sequences[pp][1][A2]]
                  if SCS2 < SCS:
                      SCS = SCS2
                  bendingfactor = enhancement(SCS)
                  distance = A2-A1
                  delta = (rightreach[A1]+leftreach[A2])
                  deltaen = delta*bendingfactor
                  if   deltaen   +   atomtodistance(additional_separation)>=
distance:
                      count += 1
                      if distance > count2:
                          count2 = distance
              N0right.append(count2)

      #calculate N0 to the left
           N0left = []
           for A1 in range(0,len(leftreach)):
               SCS = AAsize[sequences[pp][1][A1]]
               count = 0
               count2 = 0
               for A2 in range(0, A1):
                   SCS2 = AAsize[sequences[pp][1][A1-A2-1]]
                   if SCS2 < SCS:
                       SCS = SCS2
                   bendingfactor = enhancement(SCS)
                   distance = A2+1
                   delta = (leftreach[A1]+rightreach[A1-A2-1])
                   deltaen = delta*bendingfactor
                   if  deltaen  +  atomtodistance(additional_separation)  >=
distance :
                       count += 1
                       if distance > count2:
                           count2 = distance
               N0left.append(count2)

      # Nblob values of each aa in the protein
           nb1 = []
           for L in range(0,len(N0right)):
               nblobsize=N0right[L]+N0left[L]+1
               nb1.append(nblobsize)

      # splitting program
```

```
    # maximum blob size in protein/fragment
         maxnblob = np.round(max(nb1),3)
         for R in range (0,numberrepeats):
             remainder = 1
             firstsplit = split(nb1,N0left,N0right)
    # array of fragments from the initial split
             isplits = [firstsplit]
    # define an array for each blob: [Nblob, location]
             blobs = []
             blobs.append([firstsplit[0],firstsplit[2][0]])
             iteration = 0
    # splits the arrays until the blob size of all new arrays is 0
             while remainder !=0:
                 iteration += 1
                 remainder = len(isplits)
                 remaindersplits = []
                 for r in range(0,remainder):
                     for leftright in range (0,2): # 0 = left, 1 = right
                         bv = isplits[r][2+leftright][1]
                         if len(bv)==0:
                             continue
                         rl = (isplits[r][2+leftright][2])
                         rr = (isplits[r][2+leftright][3])
                         splitresult = split(bv,rl,rr)
                         remaindersplits.append(splitresult)
                         blobs.append([splitresult[0],splitresult[2][0]])
                 isplits = remaindersplits
                 remainder = len(remaindersplits)
     # appends the list of nb values
             numberofblobs.append(len(blobs))


             blobvalue2 = []
             for b in range (0, len(blobs)):
                 blobvalue2.append(blobs[b][0])
     #appends the list of <Nblob> values
             nblobvalue.append(navg(blobvalue2))

#prints results
print(pp+1,',',sequences[pp][0],',',sequences[pp][3],',',len(sequences[pp][
1]),',',navg(nblobvalue),',',navg(numberofblobs))


  # tables of definitions and functions
  #_____________________________________________________________________
  # the number of atoms in the length of the side chain, with hydrogen
  AAlen = {'A' : 2,
           'R' : 7,
           'N' : 4,
           'D' : 3,
           'C' : 3,
           'Q' : 5,
```

```
            'E' : 4,
            'G' : 1,
            'H' : 5,
            'I' : 4,
            'L' : 4,
            'K' : 6,
            'M' : 5,
            'F' : 6,
            'P' : 3,
            'O' : 10,
            'S' : 3,
            'U' : 3,
            'T' : 3,
            'W' : 7,
            'Y' : 7,
            'V' : 3}
  #number of atoms in side chain, hydrogens are excluded
  AAsize = {'A' : 1,
            'R' : 7,
            'N' : 4,
            'D' : 4,
            'C' : 2,
            'Q' : 5,
            'E' : 5,
            'G' : 0,
            'H' : 6,
            'I' : 4,
            'L' : 4,
            'K' : 5,
            'M' : 4,
            'F' : 7,
            'P' : 3,
            'O' : 13,
            'S' : 2,
            'U' : 2,
            'T' : 3,
            'W' : 10,
            'Y' : 8,
            'V' : 3}
  ##Functions
  #calcualtes the number average
  def navg(x):
      n = 0
      d = 0
      for i in x:
          n += i
          d += 1
      return n/d

  #splits the protein at the largest Nblob value and subtracts the aa's
inside the blob from the remain sequences
  def split(nblobs,leftreaches,rightreaches):
```

```
  #returns the locations of the maximum nblobs
      maxnblobloc = np.where(nblobs == np.amax(nblobs))[0]
      lsplit = []
      rsplit = []
      nmaxes = len(maxnblobloc)
  #location of nblob used for split - if there are two or more locations for
the max(nblob) value, one is picked at random
      loc =  random.choice(maxnblobloc)
      size1  = loc-round(leftreaches[loc]) #checks for bounds
      if size1 < 0:
          size1 = 0
      lsplit.append(loc)
      lsplit.append(nblobs[:size1])
      lsplit.append(leftreaches[:size1])
      lsplit.append(rightreaches[:size1])
      size2 = loc+round(rightreaches[loc])+1 #checks for bounds
      if size2 > len(nblobs):
          size2 = len(nblobs)
      rsplit.append(loc)
      rsplit.append(nblobs[size2:])
      rsplit.append(leftreaches[size2:])
      rsplit.append(rightreaches[size2:])
      return max(nblobs),nmaxes,lsplit,rsplit

  # returns the reach of an amino acid based on the side chain length
  def aareach(aa):
      return 0.892074*(AAlen[aa])/2

  def atomtodistance(number):
      return 0.892074*(number)/2

  # returns an array containing the reaches of each amino acid in the
polypeptide given the sequence
  def aareachpoly(sequence):
      reacharray = []
      for i in sequence:
          reacharray.append(aareach(i))
      return reacharray

  # imposes the left bounds on the aa reaches in a polypeptide given reach
array
  def aaleftreachpoly(aareach):
      leftreacharray = []
      for i in range(0,len(aareach)):
          if aareach[i]>i:
              leftreacharray.append(i)
          else:
              leftreacharray.append(aareach[i])
      return leftreacharray

  # imposes the right bounds on the aa reaches in a polypeptide given reach
array
```

```
  def aarightreachpoly(aareach):
      rightreacharray = []
      for i in range(0,len(aareach)):
          if aareach[i]>len(aareach)-1-i:
              rightreacharray.append(len(aareach)-1-i)
          else:
              rightreacharray.append(aareach[i])
      return rightreacharray #imposes the right bounds on the aa reachs in
a polypeptide given reach array

  #returns the factor by which the reach should be enhanced
  def enhancement(size):
      if size >= 5.0:
          return 1.0
      elif size < 5:
          return 1.567121893*(size+0.3)**-0.265852605

  #protein information [PDB ID, sequence, null(not used), log(tf)]
  sequences = np.array([
              ['1PGB (41 - 56)','GEWTYDDATKTFTVTE',0,-5.431798276],
              ['1L2Y', 'NLYIQWLKDGGPSSGRPPPS',0,-5.397940009],
              ['Alpha Helix','AAAAAAAAAAAAAAAAAAAA',0,-6.73156447],
              ['1T8J','AYRVPSYDFSRSDELAKLLRQHAGA',0,-5.124674886],
              ['1PIN (6-39)','KLPPGWEKRMSRSSGRVYYFNHITNASQWERPSG',0,-
4.08236813],
              ['1VII (42-
76)','LSDEDFKAVFGMTRSAFANLPLWKQQNLKKEKGLF',4195.90,-5.341822127],
              ['1E0L','GATAVSEWTEYKTADGKTYYYNNRTLESTWEKPQELK',0,-
4.386374267],
              ['1E0M','SMGLPPGWDEYKTHNGKTYYYNHNTKTSTWTDPRMSS',0,-
3.865220889],
              ['1K9Q','FEIPDDVPLPAGWEMAKTSSGQRYFLNHIDQTTTWQDPRK',0,-
3.648073648],
              ['1W4E','NRRVIAMPSVRKYAREKGVDIRLVQGTGKNGRVLKEDIDAWLAGG',0,-
4.429803715],
              ['1PRB (7-
53)','LKNAKEDAIAELKKAGITSDFYFNAINKAKTVEEVNALKNEILKAHA',0,-5.99326385],
              ['2WXC','GSQNNDALSPAIRRLLAEWNLDASAIKGTGVGGRLTREDVEKHLAKA',0,-
5.081245438],

['1E0G','DSITYRVRKGDSLSSIAKRHGVNIKDVMRWNSDTANLQPGDKLTLFVK',0,-3.040061373],

['1OKS','ASRSVIRSIIKSSRLEEDRKRYLMTLLDDIKGANDLAKFHQMLMKIIMK',0,-
2.69262578780016],

['1W4J','GSREVAAMPAARRLAKELGIDASKVKGTGPGGVITVEDVKRWAEETAKATA',0,-
5.341822127],

['2WQG','GSADYSSLTVVQLKDLLTKRNLSVGGLKNEWVQRLIKDDEESKGESEVSPQ',0,-
3.691503096],
```

```
            ['2LLH (19-
70)','KGPSSVEDIKAKMQASIEKGGSLPKVEAKFINYVKNCFRMTDQEAIQDLWQW',0,-
3.040061373],

['1BA5','RKRQAWLWEEDKNLRSGVRKYGEGNWSKILLHYKFNNRTSVMLKDRWRTMKKL',0,-
2.562337443],

['1IDY','MEVKKTSWTEEEDRILYQAHKRLGNRWAEIAKLLPGRTDNAIKNHWNSTMRRKV',0,-
3.778361993],

['1ENH','RPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFQNKRAKI',0,-
4.603521508],
            ['1DIV (1-
56)','MKVIFLKDVKGKGKKGEIKNVADGYANNFLFKQGLAIEATPANLKALEAQKQKEQR',0,-
2.844628856],
            ['1SRL (9-
64)','TFVALYDYESRTETDLSFKKGERLQIVNNTEGDWWLAHSLTTGQTGYIPSNYVAPS',0,-
1.91089572],

['1PGB','MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTE',0,-
2.779484684],
            ['1AVZ (Chain C: 85-
141)','TLFVALYDYEARTEDDLSFHKGEKFQILNSSEGDWWEARSLTTGETGYIPSNYVAPV',0,-
2.128042961],
            ['1SHG (6-
62)','KELVLALYDYQEKSPREVTMKKGDILTLLNSTNKDWWKVEVNDRQGFVPAAYVKKLD',0,-
0.47772393],

['1JO8','PWATAEYDYDAAEDNELTFVENDKIINIEFVDDDWWLGELEKDGSKGLFPSNYVSLGN',0,-
1.085736205],
            ['1AU7 (103-
160)','KRRTTISIAAKDALERHFGEHSKPSSQEIMRMAEELNLEKEVVRVWFCNRRQREKRVK',0,-
4.212656474],

['1FEX','GRIAFTDADDVAILTYVKENARSPSSVTGNALWKAMEKSSLTQHSWQSLKDRYLKHLRG',0,-
3.561214752],
            ['1SS1 (2-
60)','ADNKFNKEQQNAFWEILHLPNLNEEQRNGFIQSLKDDPSQSANLLAEAKKLNDAQAPKA',0,-
4.994386542],
            ['2PTL (18-
77)','VTIKANLIFANGSTQTAEFKGTFEKATSEAYAYADTLKKDNGEYTVDVADKGYTLNIKFA',0,-
1.780607376],

['2L6R','MVRQEELAAARAALHDLMTGKRVATVQKDGRRVEFTATSVSDLKKYIAELEVQTGMTQRRRG',0,
-4.473233164],

['3CI2','LKTEWPELVGKSVEEAKKVILQDKPEAQIIVLPVGTIVTMEYRIDRVRLFVDKLDNIAQVPRVG',
0,-2.518907995],

['1C8C','MATVKFKYKGEEKQVDISKIKKVWRVGKMISFTYDEGGGKTGRGAVSEKDAPKELLQMLAKQKK',
0,-3.040061373],
```

['2CRO (1-
65)','MQTLSERLKKRRIALKMTQTELATKAGVKQQSIQLIEAGVTKRPRFLFEIAMALNCDPVWLQYGT',0,
-1.606889583],

['1C9O','MQRGKVKWFNNEKGYGFIEVEGGSDVFVHFTAIQGEGFKTLEEGQEVSFEIVQGNRGPQAANVVKL
',0,-3.12692027],

['1G6P','MRGKVKWFDSKKGYGFITKDEGGDVFVHWSAIEMEGFKTLKEGQVVEFEIQEGKKGPQAAHVKVVE
',0,-2.736055236],
['2VKN (1-
66)','DNFIYKAKALYPYDADDDDAYEISFEQNEILQVSDIEGRWWKARRANGETGIIPSNYVQLIDGPEE',0
,-0.912018412],

['1CSP','MLEGKVKWFNSEKGFGFIEVEGQDDVFVHFSAIQGEGFKTLEEGQAVSFEIVEGNRGPQAANVTKE
A',0,-2.822914132],

['1RYK','MNKDEAGGNWKQFKGKVKEQWGKLTDDDMTIIEGKRDQLVGKIQERYGYQKDQAEKEVVDWETRNE
YRW',0,-3.952079785],

['1MJC','SGKMTGIVKWFNADKGFGFITPDDGSKDVFVHFSAIQNDGYKSLDEGQKVSFTIESGAKGPAAGNV
TSL',0,-2.301760754],
['1UZC (3-
71)','QPAKKTYTWNTKEEAKQAFKELLKEKRVPSNASWEQAMKMIINDPRYSALAKLSEKKQAFNAYKVQTEK
',0,-3.474355855],
['1DKT (5-
76)','QIYYSDKYDDEEFEYRHVMLPKDIAKLVPKTHLMSESEWRNLGVQQSQGWVHYMIHEPEPHILLFRRPL
PKK',0,-1.954325169],

['2A3D','MGSWAEFKQRLAAIKTRLQALGGSEAELAAFEKEIAAFESELQAYKGKGNPEVEALRKEAAAIRDE
LQAYRHN',0,-5.298392679],
['1RFA (56-
130)','NTIRVFLPNKQRTVVNVRNGMSLHDCLMKALKVRGLQPECCAVFRLLHEHKGKKARLDWNTDAASLIG
EELQVDF',0,-3.344067511],
['2JMC (2-
77','MGPREVTMKKGDILTLLNSTNKDWWKVEVNDRQGFVPAAYVKKLDSGTGKELVLALYDYQESGDNAPSYS
PPPPP',0,-1.43317179],
['1M9S (391-
466)','QGTLASHRCKALTVDREARNGGKLWYRLKNIGWTKAENLSLDRYDKMEYDKGVTAYARVRNASGNSVW
TKPYNTAG',0,-1.737177928],

['1UBQ','MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKEST
LHLVLRLRGG',0,-2.301760754],
['1LMB (6-
85)','PLTQEQLEDARRLKAIYEKKKNELGLSQESVADKMGMGQSGVGALFNGINALNAYNAALLAKILKVSVE
EFSPSIAREIY',0,-4.516662612],
['1KDX (586-
666)','GVRKGWHEHVTQDLRSHLVHKLVQAIFPTPDPAALKDRRMENLVAYAKKVEGDMYESANSRDEYYHLL
AEKIYKIQKELE',0,-3.387496959],

['1O6X','MRSLETFVGDQVLEIVPSNEEQIKNLLQLEAQEHLQLDFWKSPTTPGETAHVRVPFVNVQAVKVFL
ESQGIAYSIMIEDVQ',0,-2.95320247694211],

['1EHB','AVKYYTLEEIQKHNNSKSTWLILHYKVYDLTKFLEEHPGGEEVLREQAGGDATENFEDVGHSTDAR
ELSKTFIIGELHPDDR',0,-1.302883446],
['1PNJ (3-
84)','MSAEGYQYRALYDYKKEREEDIDLHLGDILTVNKGSLVALGFSDGQEAKPEEIGWLNGYNETTGERGDF
PGTYVEYIGRKKI',0,0.299663193],
['1PRS (91-
173)','PRARFFYKEQFDGKEVDLPPGQYTQAELERYGIDNNTISSVKPQGLAVVLFKNDNFSGDTLPVNSDAP
TLGAMNNNTSSIRIS',0,0.868588964],

['1HDN','MFQQEVTITAPNGLHTRPAAQFVKEAKGFTSEITVTSNGKSASAKSLFKLQTLGLTQGTVVTISAE
GEDEQKAVEHLVKLMAELE',0,-1.17259510113878],
['2BKF (1-
85)','MEPQVTLNVTFKNEIQSFLVSDPENTTWADIEAMVKVSFDLNTIQIKYLDEENEEVSINSQGEYEEALK
MAVKQGNQLQMQVHEG',0,-2.692625788],

['1IMQ','MELKHSISDYTEAEFLQLVTTICNADTSSEEELVKLVTHFEEMTEHPSGSDLIYYPKEGDDDSPSG
IVNTVKQWRAANGKSGFKQG',0,-3.18337855235084],

['1ST7','VSQLFEEKAKAVNELPTKPSTDELLELYALYKQATVGDNDKEKPGIFNMKDRYKWEAWENLKGKSQ
EDAEKEYIALVDQLIAKYSS',0,-3.691503096],
['1AYI (1-
86)','MELKNSISDYTEAEFVQLLKEIEKENVAATDDVLDVLLEHFVKITEHPDGTDLIYYPSDNRDDSPEGIV
KEIKEWRAANGKPGFKQ',0,-2.475478547],

['1NTI','SQAEFDKAAEEVKHLKTKPADEEMLFIYSHYKQATVGDINTERPGMLDFKGKAKWDAWNELKGTSK
EDAMKAYIDKVEELKKKYGI',0,-2.822914132],
['1K8M (1-
87)','MGQVVQFKLSDIGEGIREVTVKEWYVKEGDTVSQFDSICEVQSDKASVTITSRYDGVIKKLYYNLDDIA
YVGKPLVDIETEALKDLE',0,0.408236813],
['1PRS (1-
88)','MANITVFYNEDFQGKQVDLPPGNYTRAQLAALGIENNTISSVKVPPGVKAILYQNDGFAGDQIEVVANA
EELGPLNNNVSSIRVISVP',0,-1.30288344570976],

['1K85','HMAPTAPTNLASTAQTTSSITLSWTASTDNVGVTGYDVYNGTALATTVTGTTATISGLAADTSYTF
TVKAKDAAGNVSAASNAVSVKT',0,-0.608012274664553],
['1GXT (4-
91)','NTSCGVQLRIRGKVQGVGFRPFVWQLAQQLNLHGDVCNDGDGVEVRLREDPEVFLVQLYQHCPPLARID
SVEREPFIWSQLPTEFTIR',0,-1.91089572037431],
['1TEN (803-
891)','LDAPSQIEVKDVTDTTALITWFKPLAEIDGIELTYGIKDVPGDRTTIDLTEDENQYSIGNLKPDTEYE
VSLISRRGDMSSNPAKETFTT',0,-0.781730067425853],

['1BTA','KKAVINGEQIRSISDLHQTLKKELALPEYYGENLDALWDCLTGWVEYPLVLEWRQFEQSKQLTENG
AESVLQVFREAKAEGCDITIILS',0,-1.52003068666138],

['1TIT','MHHHHHHSSLIEVEKPLYGVEVFVGETAHFEIELSEPDVHGQWKLKGQPLTASPDCEIIEDGKKHI
LILHNCQLGMTGEVSFQAANAKSAANLKVKEL',0,-1.56346013485171],
['2BJD (12-
101)','MLKRMYARVYGLVQGVGFRKFVQIHAIRLGIKGYAKNLPDGSVEVVAEGYEEALSKLLERIKQGPPAA
EVEKVDYSFSEYKGEFEDFETY',0,-0.738300619235528],

['1FNF (1325-
1415)','TGLDSPTGIDFSDITANSFTVHWIAPRATITGYRIRHHPEHFSGRPREDRVPHSRNSITLTNLTPGT
EYVVSIVALNGREESPLLIGQQST',0,0.390865033712927],
['3ZRT (3-
93)','MEYEEITLERGNSGLGFSIAGGTDNPHIGDDPSIFITKIIPGGAAAQDGRLRVNDSILFVNEVDVREVT
HSAAVEALKEAGSIVRLYVMRR',0,-0.564582826474227],
['1DIV (58-
149)','AAEELANAKKLKEQLEKLTVTIPAKAGEGGRLFGSITSKQIAESLQAQHGLKLDKRKIELADAIRALG
YTNVPVKLHPEVTATLKVHVTEQK',0,-1.43317179028073],
['1TP3 (309-
401)','REPRRIVIHRGSTGLGFNIVGGEDGEGIFISFILAGGPADLSGELRKGDQILSVNGVDLRNASHEQAA
IALKNAGQTVTIIAQYKPEEYSRFE',0,-1.30288344570976],

['1WIT','LKPKILTASRKIKIKAGFTHNLEVDFIGAPDPTATWTVGDSGAALAPELLVDAKSSTTSIFFPSAK
RADSGNYKLKVKNELGEDEAIFEVIVQ',0,-0.178060737580333],
['3ZRT (97-
189)','AEKVMEIKLIKGPKGLGFSIAGGVGNQHIPGDNSIYVTKIIEGGAAHKDGRLQIGDKILAVNSVGLED
VMHEDAVAALKNTYDVVYLKVAKPS',0,-0.130288344570976],
['2VH7 (5-
98)','TLISVDYEIFGKVQGVFFRKHTQAEGKKLGLVGWVQNTDRGTVQGQLQGPISKVRHMQEWLETRGSPKS
HIDKANFNNEKVILKLDYSDFQIVK',0,-0.364807364798731],
['1GM1 (9-
102)','KPGDTFEVELAKTDGSLGISVTGGVNTSVRHGGIYVKAIIPKGAAESDGRIHKGDRVLAVNGVSLEGA
THKQAVETLRNTGQVVHLLLEKGQVP',0,-0.434294481903252],
['1AUE (Chain B: 2022-
2115)','ILWHEMWHEGLEEASRLYFGERNVKGMFEVLEPLHAMMERGPQTLKETSFNQAYGRDLMEAQEWCRK
YMKSGNVKDLTQAWDLYYHVFRRISKQ',0,-2.34519020227756],

['1TTG','VSDVPRDLEVVAATPTSLLISWDAPAVTVRYYRITYGETGGNSPVQEFTVPGSKSTATISGLKPGV
DYTITVYAVTGRGDSPASSKPISINYRT',0,-2.38861965046789],
['1FHT (2-
97)','VPETRPNHTIYINNLNEKIKKDELKKSLYAIFSQFGQILDILVSRSLKMRGQAFVIFKEVSSATNALRS
MQGFPFYDKPMRIQYAKTDSDIIAKMK',0,-1.99775461675496],

['1N88','MKTAYDVILAPVLSEKAYAGFAEGKYTFWVHPKATKTEIKNAVETAFKVKVVKVNTLHVRGKKKRL
GRYLGKRPDRKKAIVQVAPGQKIEALEGLI',0,-0.868588963806504],

['1FA3','GEWEIIDIGPFTQNLGKFAVDEENKIGQYGRLTFNKVIRPCMKKTIYENEGFREIKGYEYQLYVYA
SDKLFRADISEDYKTRGRKLLRFNGPVPPP',0,-1.78060737580333],
['1RIS (1-
97)','MRRYEVNIVLNPNLDQSQLALEKEIIQRALENYGARVEKVEELGLRRLAYPIAKDPQGYFLWYQVEMPE
DRVNDLARELRIRDNVRRVMVVKSQEPF',0,-2.64919633960984],
['2X7Z (311-
407)','KPVSEKIMEIKLIKGPKGLGFSIAGGVGNQHWPGDNSIYVTKIIEGGAAHKDGKLQIGDKLLAVNNVA
LEEVTHEEAVTALKNTSDFVYLKVAKPTS',0,-0.321377916608406],
['1HNG (2-
98)','DSGTVWGALGHGINLNIPNFQMTDDIDEVRWERGSTLVAEFKRKMKPFLKSGAFEILANGDLKIKNLTR
DDSGTYNVTVYSTNGTRILNKALDLRIL',0,-0.781730067425853],

['1APS','STARPLKSVDYEVFGRVQGVCFRMYAEDEARKIGVVGWVKNTSKGTVTGQVQGPEEKVNSMKSWLS
KVGSPSSRIDRTNFSNEKTISKLEYSNFSVRY',0,0.694871171045203],

['2QJL','MVNVKVEFLGGLDAIFGKQRVHKIKMDKEDPVTVGDLIDHIVSTMINNPNDVSIFIEDDSIRPGII
TLINDTDWELEGEKDYILEDGDIISFTSTLHGG',0,-1.12916565294845],
['1E41 (93-
192)','GEEDLCAAFNVICDNVGKDWRRLARQLKVSDTKIDSIEDRYPRNLTERVRESLRIWKNTEKENATVAH
LVGALRSCQMNLVADLVQEVQQARDLQNRSGA',0,-2.99663192513244],
['1PUC (2-
102)','SKSGVPRLLTASERERLEPFIDQIHYSPRYADDEYEYRHVMLPKAMLKAIPTDYFNPETGTLRILQEE
EWRGLGITQSLGWEMYEVHVPEPHILLFKREKD',0,-1.82403682399366],
['1SPR (2-
104)','AEEWYFGKITRRESERLLLNPENPRGTFLVRESETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGF
YITSRTQFSSLQQLVAYYSKHADGLCHRLTNVCPT',0,-3.77836199255829],
['1ARR*dimer (1-
106','MKGMSKMPQFNLRWPREVLDLVRKVAEENGRSVNSEIYQRVMESFKKEGRIGAMKGMSKMPQFNLRWPR
EVLDLVRKVAEENGRSVNSEIYQRVMESFKKEGRIGA',0,-3.99550923350992],
['1CUN (7-
112)','MVHQFFRDMDDEESWIKEKKLLVSSEDYGRDLTGVQNLRKKHKRLEAELAAHEPAIQSVLDTGKKLSD
DNTIGKEEIQQRLAQFVDHWKELKQLAAARGQRLEESL',0,-2.084613513],

['1YYJ','ADLEDNWETLNDNLKVIEKADNAAQVKDALTKMRAAALDAQKATPPKLEDKSPDSPEMKDFRHGFD
ILVGQIDDALKLANEGKVKEAQAAAEQLKTTIRAYNQKYG',0,-3.648073648],
['2J5A (3-
108)','HYKTLRYYETVFAVKPTLSEEEMKKKFEQVKEFIKQKGGEILYEEDWGMRQLAYPIQKFNNARYFLVQ
FKTENPQLPNELDFQLKIDEDVIRWLNIQIKESEVKKN',0,-3.170349718],
['1CUN (113-
219)','EYQQFVANVEEEEAWINEKMTLVASEDYGDTLAAIQGLLKKHEAFETDFTVHKDRVNDVCANGEDLIK
KNNHHVENITAKMKGLKGKVSDLEKAAAQRKAKLDENSA',0,-1.476601238],

['1D6O','GVQVETISPGDGRTFPKRGQTCVVHYTGMLEDGKKFDSSRDRNKPFKFMLGKQEVIRGWEEGVAQM
SVGQRAKLTISPDYAYGATGHPGIIPPHATLVFDVELLKLE',0,-0.694871171],
['1BNI (3-
110)','VINTFDGVADYLQTYHKLPDNYITKSEAQALGWVASKGNLADVAPGKSIGGDIFSNREGKLPGKSGRT
WREADINYTSGFRNSDRILYSSDWLIYKTTDHYQTFTKIR',0,-1.172595101],
['1U4Q (1662-
1771)','ANKQQNFNTGIKDFDFWLSEVEALLASEDYGKDLASVNNLLKKHQLLEADISAHEDRLKDLNSQADS
LMTSSAFDTSQVKDKRETINGRFQRIKSMAAARRAKLNESHRL',0,-4.777239301],
['1QAU (14-
125)','NVISVRLFKRKVGGLGFLVKERVSKPPVIISDLIRGGAAEQSGLIQAGDIILAVNDRPLVDLSYDSAL
EVLRGIASETHVVLILRGPEGFTTHLETTFTGDGTPKTIRVTQP',0,-0.781730067],

['2KDI','MHHHHHHGEFQIFAKTLTGKTITLEVESSDTIDNVKSKIQDKEGIPPDQQRLIWAGKQLEDGRTLS
DYNIQRESTLHLVLRLRGGSMGGAADEEELIRKAIELSLKESRNSGGY',0,-0.998877308],
['1QTU (1-
115)','GSMAGEDVGAPPDHLWVHQEGIYRDEYQRTWVAVVEEETSFLRARVQQIQVPLGDAARPSHLLTSQLP
LMWQLYPEERYMDNNSRLWQIQHHLMVRGVQELLLKLLPDDRSPGIH',0,0.156346013],

['2MYO','MCDKEFMWALKNGDLDEVKDYVAKGEDVNRTLEGGRKPLHYAADCGQLEILEFLLLKGADINAPDK
HHITPLLSAVYEGHVSCVKLLLSKGADKTVKGPDGLTALEATDNQAIKALLQ',0,-2.084613513],

['1HCD','MGNRAFKSHHGHFLSAEGEAVKTHHGHHDHHTHFHVENHGGKVALKTHCGKYLSIGDHKQVYLSHH
LHGDHSLFHLEHHGGKVSIKGHHHHYISADHHGHVSTKEHHDHDTTFEEIII',0,-0.564582826],

['2GA5','TDGQVVPQEVLNLPLEKYHEEADDYLDHLLDSLEELSEAHPDCIPDVELSHGVMTLEIPAFGTYVI
NKQPPNKQIWLASPLSGPNRFDLLNGEWVSLRNGTKLTDILTEEVEKAISKSQ',0,-2.345190202],
['1EKG (90-
208)','LDETTYERLAEETLDSLAEFFEDLADKPYTFEDYDVSFGSGVLTVKLGGDLGTYVINKQTPNKQIWLS
SPSSGPKRYDWTGKNWVYSHDGVSLHELLAAELTKALKTKLDLSSLAYSGK',0,-1.520030687],

['1ADW','ATHEVHMLNKGESGAMVFEPAFVRAEPGDVINFVPTDKSHNVEAIKEILPEGVESFKSKINESYTL
TVTEPGLYGVKCTPHFGMGMVGLVQVGDAPENLDAAKTAKMPKKARERMDAELAQVN',0,-0.299663193],
['1J5U (7-
103)','MRKPIEHTADIAYEISGNSYEELLEEARNILLEEEGIVLDTEEKEKMYPLEETEDAFFDTVNDWILEI
SKGWAPWRIKREGNELKVTFRKIRKKEGTEIKALTYHLLKFERDGDVLKTKVVFDT',0,-2.996631925],
['1FGA (20-
143)','PKRLYCKNGGFFLRIHPDGRVDGVREKSDPHIKLQLQAEERGVVSIKGVCANRYLAMKEDGRLLASKC
VTDECFFFERLESNNYNTYRSRKYTSWYVALKRTGQYKLGSKTGPGQKAILFLPMS',0,0.608012275],
['5L8I (3-
127)','FTGKFEMESEKNYDEFMKLLGISSDVIEKARNFKIVTEVQQDGQDFTWSQHYSGGHTMTNKFTVGKES
NIQTMGGKTFKATVQMEGGKLVVNFPNYHQTSEIVGDKLVEVSTIGGVTYERVSKRL',0,-0.277948468],
['3O4D (11-
140)','PVLLKSTETGQYLRINPDGTVDGTRDRSDPHIQFQISPEGNGEVLLKSTETGQYLRINPDGTVDGTRD
RSDPHIQFQISPEGNGEVLLKSTETGQYLRINPDGTVDGTRDRSDPHIQFQISPEGNG',0,-
2.128042961],

['2VIL','VELSKKVTGKLDKTTPGIQIWRIENMEMVPVPTKSYGNFYEGDCYVLLSTRKTGSGFSYNIHYWLG
KNSSQDEQGAAAIYTTQMDEYLGSVAVQHREVQGHESETFRAYFKQGLIYKQGGVASGMK',0,-
1.824036824],

['3O49','KPVLLKSTETGQYLRINPDGTVDGTRDRSDQHIQFQVSPEGGGEVLLKSTETGQYLRINPDGTVDG
TRDRSDQHIQFQVSPEGGGEVLLKSTETGQYLRINPDGTVDGTRDRSDQHIQFQVSPEGGG',0,-
0.738300619],

['3CHY','ADKELKFLVVDDFSTMRRIVRNLLKELGFNNVEEAEDGVDALNKLQAGGYGFVISDWNMPNMDGLE
LLKTIRADGAMSALPVLMVTAEAKKENIIAAAQAGASGYVVKPFTAATLEEKLNKIFEKLGM',0,-
0.434294482],
['1HRH (427-
556)','YQLEKEPIVGAETFYVDGAANRETKLGKAGYVTNKGRQKVVPLTNTTNQKTELQAIYLALQDSGLEVN
IVTDSQYALGIIQAQPDKSESELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAGI',0,-
0.382179144],
['1IFC (1-
131)','AFDGTWKVDRNENYEKFMEKMGINVVKRKLGAHDNLKLTITQEGNKFTVKESSNFRNIDVVFELGVDF
AYSLADGTELTGTWTMEGNKLVGKFKRVDNGKELIAVREISGNELIQTYTYEGVEAKRIFKKE',0,-
1.867466272],
['1OPA (1-
133)','TKDQNGTWEMESNENFEGYMKALDIDFATRKIAVRLTQTKIIVQDGDNFKTKTNSTFRNYDLDFTVGV
EFDEHTKGLDGRNVKTLVTWEGNTLVCVQKGEKENRGWKQWVEGDKLYLELTCGDQVCRQVFKKK',0,-
0.608012275],

['1CBI','PNFAGTWKMRSSENFDELLKALGVNAMLRKVAVAAASKPHVEIRQDGDQFYIKTSTTVRTTEINFK
VGEGFEEETVDGRKCRSLPTWENENKIHCTQTLLEGDGPKTYWTRELANDELILTFGADDVVCTRIYVRE',0,1
.389742342],
['1RG8(1-
137)','FNLPPGNYKKPKLLYCSNGGHFLRILPDGTVDGTRDRSDQHIQLQLSAESVGEVYIKSTETGQYLAMD

TDGLLYGSQTPNEECLFLERLEENHYNTYISKKHAEKNWFVGLKKNGSCKRGPRTHYGQKAILFLPLPV',0,-
0.564582826],

['3O4B','HFNLPPGNYKKPVLLKSTETGQYLRINPDGTVDGTRDRSDTHIQFQISPEGNGEVLLKSTETGQYL
RINPDGTVDGTRDRSDTHIQFQISPEGNGEVLLKSTETGQYLRINPDGTVDGTRDRSDTHIQFQISPEGNG',0,
-1.867466272],

['2FS6','PNFSGNWKIIRSENFEELLKVLGVNVMLRKIAVAAASKPAVEIKQEGDTFYIKTSTTVRTTEINFK
VGEEFEEQTVDGRPCKSLVKWESENKMVCEQKLLKGEGPKTSWTRELTNDGELILTMTADDVVCTRVYVRE',0,
-0.998877308],
['1NFI (67-
206)','LTEDGDSFLHLAIIHEEKALTMEVIRQVKGDLAFLNFQNNLQQTPLHLAVITNQPEIAEALLGAGCDP
ELRDFRGNTPLHLACEQGCLASVGVLTQSCTTPHLHSILKATNYNGHTCLHLASIHGYLGIVELLVSLGADV',0
,-0.781730067],

['3H08','MEKTITIYTDGAASGNPGKGGWGALLMYGSSRKEISGYDPATTNNRMELMAAIKGLEALKEPARVQ
LYSDSAYLVNAMNEGWLKRWVKNGWKTAAKKPVENIDLWQEILKLTTLHRVTFHKVKGHSDNPYNSRADELARLA
IKENS',0,-0.694871171],
['3F6R (2-
148)','SKVLIVFGSSTGNTESIAQKLEELIAAGGHEVTLLNAADASAENLADGYDAVLFGCSAWGMEDLEMQD
DFLSLFEEFDRIGLAGRKVAAFASGDQEYEHFCGAVPAIEERAKELGATIIAEGLKMEGDASNDPEAVASFAEDV
LKQL',0,-1.520030687],
['2A5E (9-
156)','MEPSADWLATAAARGRVEEVRALLEAGALPNAPNSYGRRPIQVMMMGSARVAELLLLHGAEPNCADPA
TLTRPVHDAAREGFLDTLVVLHRAGARLDVRDAWGRLPVDLAEELGHRDVARYLRAAAGGTRGSNHARIDAAEGP
SDIPD',0,-1.520030687],

['2PQE','ATSTKKLHKEAATLIKAIDGDTVKLMYKGQAMTFRLLLVDTAETKHTKKGVEKYGAEASAFTKKMV
ENAKKIEVEFDKGQRTDKYGRGLAYIYADGKMVNEALVRQGLAKVAYVYKGNNTHEQLLRKSEAQAKKEKLNIWS
EDNADSGQ',0,-0.95544786],

['1K0S','MKTLADALKEFEVLSFEIDEQALAFDVDNIEMVIEKSDITPVPKSRHFVEGVINLRGRIIPVVNLA
KILGISFDEQKMKSIIVARTKDVEVGFLVDRVLGVLRITENQLDLTNVSDKFGKKSKGLVKTDGRLIIYLDIDKI
IEEITVKEGV',0,-3.213779166],

['1A6N','VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASEDLKKHGV
TVLTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGDFGADAQGAMNKALELFRKD
IAAKYKELGY',0,-0.47772393],
['1I1B (3-
153)','VRSLNCTLRDSQQKSLVMSGPYELKALHLQGQDMEQQVVFSMSFVQGEESNDKIPVALGLKEKNLYLS
CVLKDDKPTLQLESVDPKNYPKKKMEKRFVFNKIEINNKLEFESAQFPNWYISTSQAENMPVFLGGTKGGQDITD
FTMQFVSS',0,1.737177928],
['1DWR (1-
152)','GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVV
LTALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIA
AKYKELGFQ',0,-1.259453998],

['2RN2','MLKQVEIFTDGSCLGNPGPGGYGAILRYRGREKTFSAGYTRTTNNRMELMAAIVALEALKEHCEVI
LSTDSQYVRQGITQWIHNWKKRGWKTADKKPVKNVDLWQRLDAALGQHQIKWEWVKGHAGHPENERCDELARAAA
MNPTLEDTGYQVEV',0,0.130288345],

['1RA9','MISLIAALAVDRVIGMENAMPWNLPADLAWFKRNTLDKPVIMGRHTWESIGRPLPGRKNIILSSQP

```
GTDDRVTWVKSVDEAIAACGDVPEIMVIGGGRVYEQFLPKAQKLYLTHIDAEVEGDTHFPDYEPDDWESVFSEFH
DADAQNSHSYCFEILERR',0,0.160688958],

['2KLL','SSITGISPITEYLASLSTYNDQSITFALEDESYEIYVEDLKKDEKKDKVLLSYYESQHPSNESGDG
VDGKMLMVTLSPTKDFWLHANNKEHSVELHKCEKPLPDQAFFVLHNMHSNCVSFECKTDPGVFIGVKDNHLALIK
VDSSENLCTENILFKLSET',0,0.608012275],
             ['1L63 (1-
162)','MNIFEMLRIDEGLRLKIYKDTEGYYTIGIGHLLTKSPSLNAAKSELDKAIGRNTNGVITKDEAEKLFN
QDVDAAVRGILRNAKLKPVYDSLDAVRRAALINMVFQMGETGVAGFTNSLRMLQQKRWDEAAVNLAKSRWYNQTP
NRAKRVITTFRTGTWDAYK',0,-1.606889583],

['1LOP','MVTFHTNHGDIVIKTFDDKAPETVKNFLDYCREGFYNNTIFHRVINGFMIQGGGFEPGMKQKATKE
PIKNEANNGLKNTRGTLAMARTQAPHSATAQFFINVVDNDFLNFSGESLQGWGYCVFAEVVDGMDEVDKIKGVAT
GRSGMHQDVPKEDVIIESVTVSE',0,-2.866343581],
             ['1FTG (2-
169)','KKIGLFYGTQTGKTESVAEIIRDEFGNDVVTLHDVSQAEVTDLNDYQYLIIGCPTWNIGELQSDWEGL
YSELDDVDFNGKLVAYFGTGDQIGYADNFQDAIGILEEKISQRGGKTVGYWSTDGYDFNDSKALRNGKFVGLALD
EDNQSDLTDDRIKSWVAQLKSEFGL',0,-0.998877308],
             ['1PHP (1-
175)','MNKKTIRDVDVRGKRVFCRVDFNVPMEQGAITDDTRIRAALPTIRYLIEHGAKVILASHLGRPKGKVV
EELRLDAVAKRLGELLERPVAKTNEAVGDEVKAAVDRLNEGDVLLLENVRFYPGEEKNDPELAKAFAELADLYVN
DAFGAAHRAHASTEGIAHYLPAVAGFLMEKEL',0,-0.998877308],
             ['1PHP (186-
394)','DRPFTAIIGGAKVKDKIGVIDNLLEKVDNLIIGGGLAYTFVKALGHDVGKSLLEEDKIELAKSFMEKA
KEKGVRFYMPVDVVVADRFANDANTKVVPIDAIPADWSALDIGPKTRELYRDVIRESKLVVWNGPMGVFEMDAFA
HGTKAIAEALAEALDTYSVIGGGDSAAAVEKFGLADKMDHISTGGGASLEFMEGKQLPGVVALEDK',0,1.737
177928],

['1IO2','MKIAGIDEAGRGPVIGPMVIAAVVVDENSLPKLEELKVRDSKKLTPKRREKLFNEILGVLDDYVIL
ELPPDVIGSREGTLNEFEVENFAKALNSLKVKPDVIYADAADVDEERFARELGERLNFEAEVVAKHKADDIFPVV
SAASILAKVTRDRAVEKLKEEYGEIGSGYPSDPRTRAFLENYYREHGEFPPIVRKGWKTLKKIAEKVESEKK',0
,0.10857362],
             ['1IGS (27-
248)','QRPIISLNERILEFNKRNITAIIAEYKRKSPSGLDVERDPIEYSKFMERYAVGLSILTEEKYFNGSYE
TLRKIASSVSIPILMKDFIVKESQIDDAYNLGADTVLLIVKILTERELESLLEYARSYGMEPLIEINDENDLDIA
LRIGARFIGINSRDLETLEINKENQRKLISMIPSNVVKVAESGISERNEIEELRKLGVNAFLIGSSLMRNPEKIK
EFIL',0,1.954325169],
             ['1UCH (5-
230)','RWLPLEANPEVTNQFLKQLGLHPNWQFVDVYGMDPELLSMVPRPVCAVLLLFPITEKYEVFRTEEEEK
IKSQGQDVTSSVYFMKQTISNACGTIGLIHAIANNKDKMHFESGSTLKKFLEESVSMSPEERARYLENYDAIRVT
HETSAHEGQTEAPSIDEKVDLHFIALVHVDGHLYELDGRKPFPINHGETSDETLLEDAIEVCKKFMERDPDELRF
NAIALSAA',0,1.129165653],
             ['1B9C (4-
230)','GEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFTYGVQCF
SRYPDHMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYN
SHNVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLL
EFVTAAGIT',0,1.129165653],

['1THF','MLAKRIIACLDVKDGRVVKGSNFENLRDSGDPVELGKFYSEIGIDELVFLDITASVEKRKTMLELV
EKVAEQIDIPFTVGGGIHDFETASELILRGADKVSINTAAVENPSLITQIAQTFGSQAVVVAIDAKRVDGEFMVF
TYSGKKNTGILLRDWVVEVEKRGAGEILLTSIDRDGTKSGYDTEMIRFVRPLTTLPIIASGGAGKMEHFLEAFLA
GADAALAASVFHFREIDVRELKEYLKKHGVNVRLEGL',0,1.389742342],
```

```
['3BLM','KELNDLEKKYNAHIGVYALDTKSGKEVKFNSDKRFAYASTSKAINSAILLEQVPYNKLNKKVHINK
DDIVAYSPILEKYVGKDITLKALIEASMTYSDNTANNKIIKEIGGIKKVKQRLKELGDKVTNPVRYEIELNYYSP
KSKKDTSTPAAFGKTLNKLIANGKLSKENKKFLLDLMLNNKSGDTLIKDGVPKDYKVADKSGQAITYASRNDVAF
VYPKGQSEPIVLVIFTNKDNKSDKPNDKLISETAKSVMKEF',0,2.866343581],

['1V9E','SHHWGYGKHNGPEHWHKDFPIANGERQSPVDIDTKAVVQDPALKPLALVYGEATSRRMVNNGHSFN
VEYDDSQDKAVLKDGPLTGTYRLVQFHFHWGSSDDQGSEHTVDRKKYAAELHLVHWNTKYGDFGTAAQQPDGLAV
VGVFLKVGDANPALQKVLDALDSIKTKGKSTDFPNFDPGSLLPNVLDYWTYPGSLTTPPLLESVTWIVLKEPISV
SSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK',0,1.91089572],
              ['4BLM (31-
291)','KTEMKDDFAKLEEQFDAKLGIFALDTGTNRTVAYRPDERFAFASTIKALTVGVLLQQKSIEDLNQRIT
YTRDDLVNYNPITEKHVDTGMTLKELADASLRYSDNAAQNLILKQIGGPESLKKELRKIGDEVTNPERFEPELNE
VNPGETQDTSTARALVTSLRAFALEDKLPSEKRELLIDWMKRNTTGDALIRAGVPDGWEVADKTGAASYGTRNDI
AIIWPPKGDPVVLAVLSSRDKKDAKYDDKLIAEATKVVMKALN',0,2.041184065],

['1WQ5','MERYESLFAQLKERKEGAFVPFVTLGDPGIEQSLKIIDTLIEAGADALELGIPFSDPLADGPTIQN
ATLRAFAAGVTPAQCFEMLALIRQKHPTIPIGLLMYANLVFNKGIDEFYAQCEKVGVDSVLVADVPVEESAPFRQ
AALRHNVAPIFICPPNADDDLLRQIASYGRGYTYLLSRAGVTGAENRAALPLNHLVAKLKEYNAAPPLQGFGISA
PDQVKAAIDAGAAGAISGSAIVKIIEQHINEPEKMLAALKVFVQPMKAATRS',0,0.912018412],

['1L8W','NKFYQSVIQLGNGFLDVFTSFGGLVAEAFGFKSDPKKSDVKTYFTTVAAKLEKTKTDLNSLPKEKS
DISSTTGKPDSTGSVGTAVEGAIKEVSELLDKLVKAVKTAEGASSGTAAIGEVVADADAAKVADKASVKGIAKGI
KEIVEAAGGSEKLKAVAAAKGENNKGAGKLFGKAGAAAHGDSEAASKAAGAVSAVSGEQILSAIVTAADAAEQDG
KKPEEAKNPIAAAIGDKDGGAEFGQDEMKKDDQIAAAIALRGMAKDGKFAVKDGEKEKAEGAIKGAAESAVRKVL
GAITGLIGDAVSSGLRKVGDSV',0,-0.868588964],
  ])

  #runs the program: 'mainprogram'
  if __name__ == '__main__':
    mainprogram()
```