UNIVERSITY OF CALIFORNIA, MERCED

Protein Folding upon Binding and Conformational Rheostats: Novel Methods for their Analysis and Engineering Applications in Biosensing

A Dissertation Submitted in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in Chemistry and Chemical Biology

by

Thinh Doan Nguyen Luong

2021

Advisory Committee:

Professor Andy LiWang (Chair)

Professor Victor Muñoz (PI)

Associate Professor Eva De Alba

Assistant Professor Arvind Gopinath

Copyright

Chapter 3 is reproduced with permission from Suhani Nagpal, Thinh Doan Nguyen Luong, Mourad Sadqi, Victor Muñoz. Downhill (Un)Folding Coupled to Binding as a Mechanism for Engineering Broadband Protein Conformational Transducers. ACS Synth. Biol. 2020, 9, 9, 2427–2439

University of California, Merced

Graduate Division

The Dissertation of Thinh Doan Nguyen Luong is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

Professor Andy LiWang

_____

Professor Victor Muñoz

_____

Associate Professor Eva De Alba

_____

Assistant Professor Arvind Gopinath

Dedicated

To my beloved and deeply missed mom

TABLE OF CONTENTS

Acknowledgment

I would like to thank my advisor, Dr. Victor Muñoz, for his support and guidance through this research. I sincerely admire his keen intellect and brilliant approach to science. Dr. Muñoz has shown me how and what a true scientist should be; he has inspired me tremendously to overcome many challenges and continue my Ph.D. program. I will always look up to him for the rest of my professional career, striving relentlessly to become a well-rounded scientist.

I would also like to thank my committee members, Dr. Andy LiWang, Dr. Eva De Alba, and Dr. Arvind Gopinath, for their mentorships and invaluable feedback.

I would like to thank Dr. Mourad Sadqi for his support and guidance from my first day at UC Merced until today. I have learned a great deal owing to his scientific knowledge, his patience, and his advice.

I also like to express my gratitude toward all the former and current lab members for their kind-hearted support. Their scientific inputs have often enlightened my mind and aided me with my research. Thanks to Rama, Suhani, Nivin, and Ameed for always being there to help me out and make my journey so memorable, especially in the difficult setting of the unprecedented pandemic.

Thanks to all my family members and relatives for supporting me spiritually, believing in me, and lifting me up.

A portion of this thesis/dissertation are reprints of the materials as it appears in ACS-Synthetic Biology. The co-author Suhani Nagpal contributed equally to Chapters 2 and 3, and we share equal authorship on the publication(s). All other co-authors listed in the publication directed and supervised research.

List of Abbreviations

3D: three-dimensional

APBS: adaptive Poisson− Boltzmann solver

CD: Circular Dichroism

CR(s): Conformational Rheostat(s)

DSC: Differential Scanning Calorimetry

FRET: fluorescence resonance energy transfer

GpW: W protein of bacteriophage lambda

IDP(s): Intrinsically disordered protein(s)

IPDP(s): Partially disordered protein(s)

IPTG: isopropyl-β-D-thio-galactopyranoside

MD: Molecular dynamics

NMR: Nuclear magnetic resonance

PET: photo-induced electron transfer

PME: Particle Mesh Ewald

Q: the fraction of native backbone contacts

R_g: the radius of gyration

RMSD: root-mean-square deviation

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

SM-FRET: single-molecule- fluorescence resonance energy transfer

TCEP: tris(2-carboxyethyl) phosphine

TSCC: Triton Shared Computing Cluster

TFE: 2,2,2-Trifluoroethanol

List of Figures

List of Tables

Vita

Thinh Doan Nguyen Luong

Education

2021 University of California, Merced — Merced, CA
Ph.D. in Chemistry and Chemical Biology

2015 California State University, East Bay — Hayward, CA
Master of Science in Chemistry

2011 Université de Pierre et Marie Curie — Paris, France
Bachelor of Sciences and Technologies, Specialization in Chemistry

Research Experience

University of California, Merced — Merced, CA
Graduate Student Researcher with Dr. Victor Muñoz — 2015-present

- Designed and demonstrated an approach to dissect the energy landscape of intrinsically disordered proteins by combining recombinant protein expression and protein purification technique, and Circular Dichroism Spectroscopy to study the conformational change as a function of trifluoroethanol and to quantify the degree of its folding cooperativity.
- Employed an integrated approach of computational and experimental methods, including computational design, mutagenesis technique, and chemical labeling of purified protein with organic fluorophores to engineer pH transducer mechanism for biosensors that can produce an analogical signal, broad dynamic range, and fast response time.

California State University, East Bay — Hayward, CA
Master's thesis research conducted with Dr. Chul-Hyun Kim 2013-2015

- Performed the entire process of purifying T7 RNA polymerase enzyme from *E. Coli* and isolation RNA from *in vitro* RNA transcription and employed the NMR technique to investigate the interaction between the RNA and the coat protein of Brome Mosaic Virus and the role of RNA in regulating the replication process of Brome Mosaic Virus

Teaching Experience

- Teaching Assistant in 3 undergraduate courses, including General Chemistry I, General Chemistry II, and Organic Chemistry I at UC Merced (2015-2021)

- Teaching Assistant in Instrumental Analysis at CSU East Bay (2014)

Fellowships and Travel Awards

CCBM Scholar Travel Awards for the platform presentation at the 63rd Annual Meeting of Biophysical Society

CCBM SEED Project Award

Publication

1. Thinh D. N. Luong*, Suhani Nagpal*, Mourad Sadqi & Victor Muñoz**.** Molecular LEGO: A Modular Approach to Dissect the Folding Landscapes of Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* (*submitted to JACS*)
2. Suhani Nagpal*, Thinh D. N. Luong*, Mourad Sadqi & Victor Muñoz. Downhill (Un)Folding Coupled to Binding as Mechanism for Engineering Broadband Protein-Based Biosensors. *ACS Synth. Biol.* 2020, 9,9,2427-2439
3. Thinh D.N. Luong, Mourad Sadqi & Victor Muñoz. Fluorescence-based strategies to implement an optical signal reporting the conformational rheostat of pH biosensor. ACS Synth.Biol. (in preparation)

Abstract

Title: Protein Folding upon Binding and Conformational Rheostats: Novel Methods for their Analysis and Engineering Applications in Biosensing

Proteins are the work horses of the cell that perform the vast majority of functions essential for life. The mechanism by which proteins fold to their functional native state has been a subject of extensive research for more than 50 years now. Downhill folders are the class of proteins whose folding reaction is heterogeneous, non-cooperative, and happens without encountering a significant free energy barrier, resulting in ultrafast kinetics. The single ensemble of conformations of a global downhill folding protein moves gradually from highly disordered to the unique native structure when thermodynamic parameters that affect the protein's stability are changed (one-state folding). The gradual morphing of a one-state downhill folding protein structure in response to thermodynamic bias is referred to as a conformational rheostat. When such a conformational rheostat is coupled to binding an analyte, it can result in an ultrafast, broad band, and single-molecule analog biosensors. This thesis explores conformational rheostats as the mechanism behind the folding upon binding behavior of intrinsically disordered proteins and as broadband transducers towards engineering high-performance biosensors.

The second chapter of this thesis describes a new methodology that we have developed to study the conformational landscape of intrinsically partially disordered proteins (IPDP). This methodology is inspired by the LEGO game, where the sequence of an IPDP is deconstructed into its local structural elements and their possible combinations based on the 3D structure the IPDP acquires upon binding its partners. The local structural elements are hence analogous to LEGO building blocks, and their combinations report on the interactions among them, like the complementary indentations of LEGO pieces. In particular, we chose the IPDP NCBD as model IPDP to develop the proof of concept for the method. Our results showed that even though the NCBD is highly flexible and apparently disordered, there are strong local signals and different sets of long-range transient interactions. These sets of interactions stabilize the overall fold and compete with one another hence resulting in a dynamic ensemble. The methodology developed in Chapter 2 is expected to be extremely useful in characterizing the incipient cooperativity of virtually any IPDP in their unbound form, a capability that is currently unavailable.

Chapters 3 and 4 of this thesis deal with the design of a pH biosensor using the downhill folding protein gpW as a scaffold and unfolding coupled to ionization as a transducer. In chapter 3, a methodology for engineering conformational pH transducing into pH insensitive proteins using a histidine grafting approach was developed. The methodology was applied to the protein gpW to demonstrate an engineered, tunable broadband pH transducer based on the conformational rheostat mechanism. Chapter 4 explores general strategies for introducing fluorescence readouts capable of converting the gradual conformational changes of the rheostatic pH transducer into broadband

fluorescence-based pH biosensors. Strategies that exploit the Förster Resonance Energy Transfer (FRET) and Photo Induced Electron Transfer (PET) mechanisms were explored as potential means to convert changes in conformation into suitable fluorescence signals were explored and characterized. We discovered that FRET signals using fluorophores in the visible (required for high-sensitivity biosensing) are insensitive to the localized conformational changes associated with conformational rheostats in native-like conditions. In contrast, the very short-range distance dependence of PET (< 1 nm) enabled efficient, high signal-to-noise broadband pH sensing, thus emerging as an extremely useful strategy for implementing rheostatic fluorescence biosensing.

Thinh Doan Nguyen Luong

Chemistry & Chemical Biology, Ph.D.

University of California, Merced

2021

Committee Chair: Professor Andy LiWang

CHAPTER 1

1 Introduction

1.1 Protein folding problem

Proteins are macromolecules that play essential roles in diverse aspects of biological processes. Some of their many functions include enzymes that catalyze different chemical processes in living cell[1], a transport protein that carries molecules in and out of the cell, or antibodies that help to protect our body against infections.

Their functions are tightly associated with their three-dimensional (3D) structures which are in turn related to their amino acid sequences. Indeed, proteins must acquire their active native 3D states for proper functioning[2]. Numerous studies have been conducted in the last few decades to understand the unifying mechanism in structural dynamics and protein folding. In the 1960s, Christian Anfinsen conducted a groundbreaking experiment in which the folding/refolding process of the bovine pancreatic ribonuclease was studied using denaturing agents. The result showed that proteins can fold by themselves and hence that the information needed to determine their 3D structures was stored in their amino acid sequences[3].

Therefore, the questions of how proteins fold or why they fold into specific structures have been the fundamental problems in protein folding studies[2]. For instance, scientists have tried to determine how amino acids decide the unique 3D structure, or what is the mechanism that allows proteins to reach their native structure; how we can predict the 3D structure of a protein solely based on their given sequences; and whether we can engineer proteins for any function of interest.

Additional questions have arisen since the discovery of intrinsically disordered proteins. Studies in the last decades have found that about 40% of the human proteome comprises protein domains or regions that are intrinsically disordered (IDPs or IDRs)[4]. Misfolding of a protein becomes an underlying cause of many neurodegenerative diseases such as Alzheimer's or Parkinson's [5].

Such studies not just simply provide insights about protein dynamics or protein structure prediction but ideally could lead to enhanced understanding and perhaps better solutions to treat these diseases. Moreover, in parallel, scientists strive to transform this advanced knowledge into beneficial applications for humans and the environment in terms of designing novel proteins.

1.2 Energy landscape theory and barriers in protein folding reaction

The amino acid sequences of natural proteins have the ability to fold into their native structure, whether on their own or by interacting with partners. Levinthal proposed that these sequences must follow a specific folding pathway rather than randomly searching through an astronomical number of possible pathways to find their natives structures[6]. Indeed, it would take an astronomical amount of time for a protein molecule to randomly explore all of its possible conformations in order to find the native conformation, whereas, in reality, folding rates are much faster than that. This concept has been known as the "Levinthal paradox."

Later, the energy landscape theory emerged as a statistical method to describe the protein folding mechanism. Proteins exist as an ensemble of different conformations, including the native and unfolded state, in which the folding process is driven by entropic and enthalpic forces. The energy landscape is described as a funnel with high energy at the broad top, corresponding to the myriads of unfolded conformations, and with energy and entropy decreasing as conformations become more native-like (going down the funnel). At the early stages of folding, the funnel provides the gradient to guide the protein through the search until it reaches the low-energy native structure at the bottom[7]. In the native state, the protein is stabilized by enthalpy due to various non-covalent interactions, compensating for the loss in conformational entropy associated with the ordering of the protein[8].

The overall funnel shape does not mean it is necessarily smooth. Indeed, proteins can have a certain degree of roughness in the landscape that can transiently trap the protein as it searches for the native structure —in kinetic terms, the time it gets trapped ultimately affects its folding rate. The decompensation between enthalpy and entropy also generates a global free energy barrier, which is essentially an entropic bottleneck. Crossing the global free energy barrier determines the folding rate, particularly proteins with smooth landscapes: the higher the barrier, the longer the protein takes to get to the native conformation. This theory also predicts different folding scenarios in which proteins fold by crossing a free barrier (two-state folding) or without a barrier (downhill folding)[7].

1.3 Effect of marginal stability on protein folding mechanism

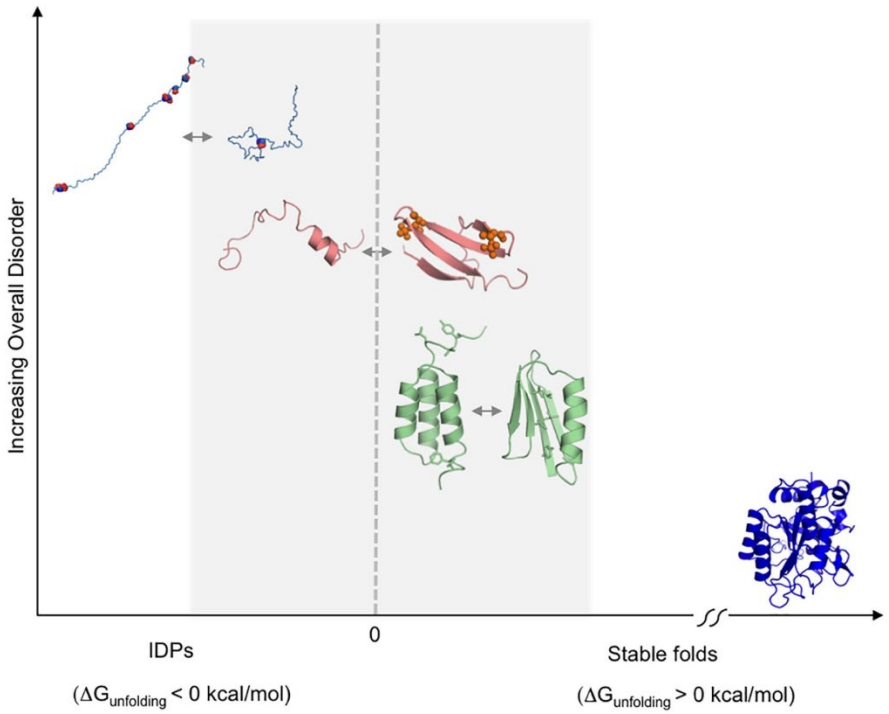

Figure 1.1 Examples of transitions from top left to the bottom right of different proteins with marginal stability: prostate associated Gene 4 (PAGE4), 4E-BP2; order-to-order fold switching between GA98 and GB98. In contrast, stable proteins such as subtilisin (shown in dark blue) do not undergo such changes. Adapted from Kulkarni, P.; Solomon, T. L.; He, Y.; Chen, Y.; Bryan, P. N.; Orban, J.(2018)[9]. Figure reprinted with permission.

Not all proteins can be classified as structures or disordered there is a spectrum of order-disorder for natural proteins[9]. Indeed, there are stable proteins with well-defined structures, whereas there are IDPs on the other extreme. Highly stable proteins such as proteases and structural proteins have relatively higher free energy barriers separating their native and unfolded states. As a result, they fold and unfold cooperatively in a two-state manner. The other extreme types of order-disorder transitions correspond to intrinsically disordered proteins whose sequences have a preponderance for polar, charged amino acids and paucity of bulky hydrophobic residues[10-12]. Their sequence is interspersed with structure-breaking residues such as glycine and proline. Beyond that limit, there are protein sequences known to remain disordered even when they bind their partner[13]. In an intermediate range, certain protein sequences are at the brink of stability, and hence their structure is extremely sensitive to solvent conditions and other environmental/biological conditions. These marginally stable sequences fold to their native state when bound to their partner (folding upon binding) or upon post-translational modifications such as phosphorylation or

glycosylation. The marginal stability of certain proteins can also result in a phenomenon called protein metamorphosis, where a particular sequence exists in two different folds. Nature utilizes protein metamorphosis for several functions, including entry of viruses into host cells and controlling circadian rhythms in cyanobacteria[14]. Downhill folding proteins are also, by definition, marginally stable because they are stabilized mostly by local interactions[15]. The lack of a significant free energy barrier results in gradual, highly heterogeneous, non-cooperative (un)folding. In summary, downhill or one state folding, IDP folding upon binding, and protein metamorphosis are all related phenomena that emerge from the marginal stability of certain protein sequences.

1.4 Downhill protein folding

Protein folding is conventionally described as an all-or-none process where the protein is toggling between two conformations (unfolded and folded) like a conformational switch[16]. Such a two-state folding scenario holds true if the protein folding process involves crossing a high free energy barrier (several *RT*) that separates the native and unfolded ensembles. Two-state folding is slow (requires crossing a high barrier) and results in highly cooperative transitions in which the protein does not significantly populate any partially folded conformations[17].

However, over the last two decades, biophysical work has shown that not all proteins fold in a two-state fashion and that there is an alternative type of protein folding mechanism known as downhill folding, which was predicted from the energy landscape theory[7]. These proteins fold ultra-fast (in the microsecond timescale) and thus fold by crossing marginal barriers (< 3RT) or even no barrier at all (less than 1 *RT*)[18, 19].

In the extreme case where the barrier height is insignificant relative to thermal energy (< 1 RT), the folding free-energy landscape exhibits a single well (instead of two) that shifts gradually from the native state to the unfolded state ensemble as denaturation stress increases[20]. In this case, when the protein is in mildly denaturing conditions, it populates a continuum of partly folded conformations rather than a linear combination of native and fully unfolded conformations[21].

These proteins are generally classified as downhill folders. Their behavior can be experimentally confirmed using various biophysical instrumental techniques, such as NMR, far UV-CD, DSC, SM-FRET, and thermodynamic and kinetic analyses to probe the transition and measure the rates[22]. A few proteins have been identified as downhill folders by employing these approaches, such as gpW[23], PDD[8], and BBL[22]. The common structural features of these proteins are their relatively small size and basic fold topology with a simple secondary structure type arrangement. In terms of folding cooperativity, these proteins exhibit broad transitions, with broadness that is directly proportional to the fraction of local interactions stabilizing their native structure (local versus total interactions)—for instance, the higher the local fraction, the lower the cooperativity.

1.5 Intrinsically disordered proteins (IDPs) and folding upon binding mechanism

IDPs do not fold into three-dimensional structures but exist as broad, non-random, conformational ensembles. Recent studies estimate that IDPs play essential roles in numerous fundamental biological processes such as transcription, cell cycle control, and signaling[4]. The key to their functioning is the ability to fold when bound to ligand partner(s) folding coupled to binding)[24-26], thus operating as morphing proteins. Some IDPs are devoid of any structure, even after binding to partners[13], but many are partially disordered (IPDP) and morph to accommodate the structural patterns of partners[27]. Such behavior enables sophisticated binding modes[17], including binding to multiple partners[28, 29], allostery[30], and structural adaptation[24]. The inability of IDPs to adopt a three-dimensional structure is attributed to their special amino acid compositions: high net charge and enriched proline content[11, 12]. IDP sequences also have low contents of hydrophobic residues[10].

Over the last two decades, research efforts have focused on their folding upon binding process. These studies have shown that IPDPs bind partners following two possible mechanisms. One of them is termed conformational selection, where the proteins will adopt a conformation similar to their bound state on their own and prior to binding their partners. Another possibility is that the binding to a partner occurs while the protein is still unfolded, and then folding occurs in the encounter complex[27]. However, what remains a mystery is the role (if any) that the folding mechanism of the IPDP plays in defining its binding/functional properties. Despite booming interest in IDP behavior, investigating their folding and energetics remains challenging because these methods rely on analyzing the data as all-or-none transitions between a unique native structure and the unfolded ensemble, and any observable will be averaged over a dynamic ensemble of structures.

1.6 Conformational rheostats (CRs)

As more studies have been conducted, it has been noted that certain features are shared by downhill folding proteins and IDPs: an abundance of local interactions, high net charge, and marginal stability. This connection between downhill folding and IDPs (particularly those that are partially disordered) suggests that IDPs may fold upon binding as conformational rheostats, that is, gradually or uncooperatively. This conformational behavior could be a functional mechanism adopted by IDPs to be able to specifically select partners, alternate between conformational selection and induced-fit binding, and exhibit allostery while being apparently disordered.

1.6.1 CR as a mechanism for transducing signals

As downhill folding protein gradually changes its conformational properties during (un)folding, such protein can act as a conformational rheostat rather than a conformational switch. Consequently, it has been proposed that the folding of a downhill folder coupled to

binding to an analyte or partner could result in gradual changes in conformation in response to various levels of binding (i.e., different analyte concentrations) as a function of the concentration of analytes or partners. The key here is that protein conformations with different degrees of native structure will also have a varying affinity to the analyte (maximal for the native state and minimal for the fully unfolded state) (Figure 1.1), thus producing a gradual (analog) signal[31].

In this regard, conformational rheostats can be used to design protein-based biosensors. The downhill folding protein will be coupled to the binding of an analyte of interest to produce an analog signal.

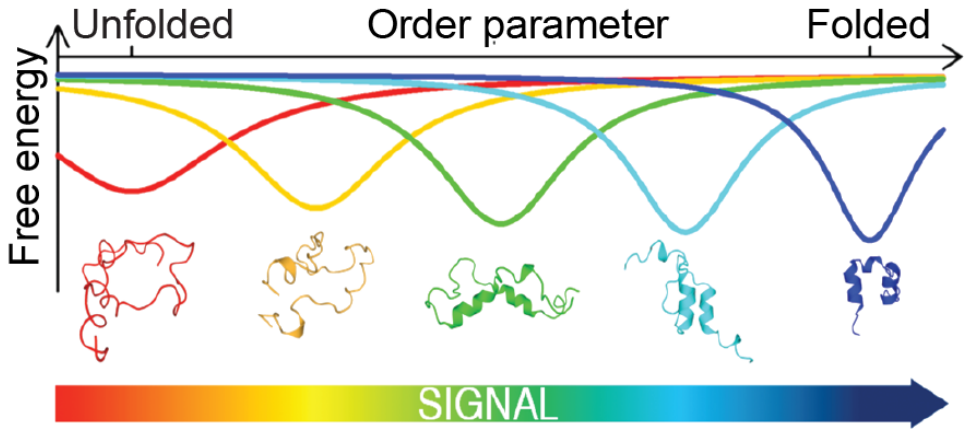

Figure 1.2 Sketch of a conformational rheostat based on the coupling of a signal to the folding ensemble of a one-state downhill folding protein module. Figure adapted from Cerminara, M.; Desai, T. M.; Sadqi, M.; Muñoz, V. (2012)[18]. Figure reprinted with permission.

1.6.2 CR as a mechanism behind IDP function

The structural disorder is generally considered necessary and sufficient to enable IPDPs to morph into any required shape on cue. However, the question is then: how does an IPDP manage to bind specifically, select among partners, and exhibit allostery? A possible solution to these puzzles is that IPDPs fold upon binding as conformational rheostats (CR)[17, 32], which is like a downhill folding protein. The critical connection here is that downhill folding domains are known to have IDP-like sequences and are stabilized mainly through local interactions, making them fast folding and marginally stable and hence partially disordered[33]. The distinctive functional feature of a CR is that the ensuing conformational ensembles are not simply disordered but contain built-in energetic biases towards specific (potentially multiple) structures. Such biases can provide an active driving force for selecting partners and allostery. CRs also undergo gradual order-disorder transitions, which would explain how IPDPs adapt or morph their ensembles in response

to different partners and integrate conformational selection and induced-fit binding in the same protein[33].

1.7 Research aims

1.7.1 Proposing a molecular LEGO approach to investigate CR as a mechanism that governs the folding upon binding behavior of IDPs

Investigating the role that the folding mechanism plays in how IPDPs bind and function requires approaches that resolve their conformational landscapes and energetics in the absence of partners. The classical approach to thermodynamically characterize folding behavior relies on determining the folding free energy (ΔG^{UN}) from analysis of equilibrium denaturation experiments of canonical proteins[34, 35]. Herein, the native signatures (e.g., via fluorescence or circular dichroism) are measured as a function of a denaturing agent, and the resulting sigmoidal curve is interpreted assuming a two-state folding equilibrium. When performed on many point mutations (designed based on the native structure), this analysis informs the energetic contributions of the structural perturbation caused by each mutation[36, 37]. In this experiment, the pre-transition signal defines the native state and the post-transition the unfolded state, whereas anything in between arises from a linear combination of the two states. The problem with IDPs is that their native states are broad ensembles that change properties dynamically, and so are their unfolded ensembles. Furthermore, denaturation experiments on IDPs lack defined baselines, so it is challenging to assess the degree of native structure at any given condition. Likewise, the inherently low cooperativity of IDP transitions results in nearly linear unfolding behavior that is not amenable to two-state thermodynamic analysis.

Chapter 2 focuses on addressing this challenge. For that purpose, a modular approach called molecular LEGO is introduced and demonstrated experimentally and computationally. The approach starts by decomposing an IPDP into its basic secondary structural elements or building blocks (defined from the 3D structures they form in the complex), as well as all of their feasible combinations (sets of consecutive elements). The combined LEGO elements recapitulate specific subsets of tertiary interactions in analogy to the complementary indentations between LEGO building blocks. The conformational analysis of LEGO elements allows dissecting contributions from local interactions and pairwise tertiary interactions. Moreover, the behavior of the simpler elements provides the correct reference ensembles to use for interpreting the results on higher-order elements. Such reference ensembles are essential to detect any subtle effects (additive or cooperative) that might occur in each combined element and the entire protein and convert them into energetic contributions using simple statistical thermodynamic analysis.

1.7.2 Designing a broadband biosensor based on the CR mechanism

Since the biosensor concept was first introduced in 1962[38], biosensor research has been growing at an increasing rate with various designs and applications from healthcare to

chemical and biological analysis and environmental monitoring[39]. A biosensor is an analytical device that comprises a biological sensing element such as biological material or biologically derived material. The biological element can connect or integrate with a physicochemical transducer to deliver a complex bioanalytical measurement[40]. Nowadays, there are multiple types and examples of successful biosensor designs[41], but biosensors capable of monitoring the levels of an analyte in real-time, in-situ, and inside living cells are still unavailable. For instance, the monitoring of key molecules for cellular processes such as ATP/ADP, protons, and metal ions (Copper, Calcium) can be invaluable to unravel cellular processes as well as the cellular mechanisms of disease. The requirements for such sensors are to be highly specific and sensitive in the biologically appropriate concentration range of the molecule of interest.

Protein-based biosensors are particularly interesting as their reliance on biomolecular recognition can potentially meet the requirements in specificity and selectivity[42]. Building a biosensor based on the biomolecular recognition event of a protein also requires a transducing mechanism that converts the binding into a traceable signal. It has been proposed that proteins can be implemented with a transducer based on a conformational change by coupling protein folding to the binding to the molecule of interest[43]. The functioning of such conformational protein-based biosensors relies on engineering the protein to be unfolded in the absence of analyte so that the binding event triggers its folding[44]. Accordingly, binding of the target molecule to the biosensor protein only occurs when the protein is in its native structure. Binding will provide free energy to stabilize the native state and thus switch the protein from the unfolded to the native state resulting in a significant conformational change. Such conformational change becomes the signal transducer[44]. With this design, the protein sensor acts like a conformational switch that produces a binary signal (bound or free). The signal, in turn, can be read out by introducing a spectroscopic signal that detects the conformational change, such as introducing extrinsic fluorophores to measure the distance between the two ends of the protein by Förster resonance energy transfer[45].

Inspired by the general principle of folding coupled to binding for protein-based biosensor design, we propose to apply this approach to proteins that fold via the downhill folding mechanism. The idea is that biosensors based on downhill folding modules will have much faster response time (downhill folding proteins fold in microseconds while typical single domain two-state proteins fold in the tens of milliseconds to minutes[8]), much broader dynamic range[16], and the potential to produce analog signals at the single-molecule level that could lead to nanoscale biosensors. Chapters 3 and 4 of this dissertation report our efforts to engineer broadband pH transducers on a pH-insensitive downhill folding protein using the CR concept (Chapter 3) and introduce fluorescence readouts that can effectively transform the conformational change into optical signals (Chapter 4).

CHAPTER 2

2 Molecular LEGO: a modular approach to dissect the folding landscape of intrinsically disordered proteins.

2.1 Introduction

Due to the heterogenicity of their conformational ensembles, understanding how IDPs fold or search through conformational space to fold into a structure that corresponds to its bound state would be helpful to understand their folding coupled to binding mechanism. The characteristics of the energy surface or precisely the quantitative measurement of the free energy barrier would be the key to understand their folding pathway. However, it is still challenging to study their transition state despite our advances in sophisticated instrumental methods because of their structural flexibility.

In the ’90s, extensive studies on various well-structured proteins were carried out to decipher their potential folding initiation sites by fragmentation of the polypeptide chain into a series of peptides that spans the length of the entire or part of the protein[46]. To characterize the early events in a protein folding pathway, conformational preferences of these short peptide fragments comprising of different secondary structural elements were examined in aqueous solution or with co-solvent such as 2,2,2-Trifluoroethanol (TFE). In this regard, the work of Dyson and Wright showcased that different structural motifs (alpha or beta) exhibit varied conformational propensities for early formation of local secondary structural elements and, conformations adopted by these peptide fragments are in general lowly populated and may reflect the actual secondary structure present in the folded state[47-49]. Itzhaki et al. investigated a series of five peptide fragments by dissecting the N-terminal chymotrypsin inhibitor 2, a prime example of a two-state folding protein searching for the nucleation sites during its folding process[50]. Those studies performed on two-state folding domains revealed weak native-like biases in the fragments[49]and the need for almost the entire protein to elicit any detectable folding[51]. Recently, a similar modular approach has been used to investigate the folding landscape of IDPs via molecular simulations, in which the much faster dynamics of the small protein fragments greatly enhance conformational sampling[52]. Another key advantage of a modular approach is that it facilitates the direct quantitative comparison between experiments and simulations.

Inspiring by those works and the LEGO toy in which building blocks associate with one another by complementary indentation, we came up with the molecular LEGO approach

as a tool to dissect the energy landscape of IDP. By combining computational and experimental techniques, we can characterize the folding behavior of the peptides that make up the whole protein and the protein itself. We reason that the high contributions from local interactions and minimal folding cooperativity expected for IPDPs[17] make them more suited for these types of studies.

Precisely, we performed the analysis integrating experiments and simulations. Experimentally, we studied the conformational behavior of NCBD and all of its LEGO elements (building blocks and combined elements) using circular dichroism and the non-specific structure-promoting cosolvent (TFE) as the thermodynamic variable. Computationally, we performed all-atom Molecular Dynamics (MD) simulations in explicit solvent, taking advantage of the shorter timescales (μs) involved in protein morphing phenomena and the current availability of IDP-improved force fields[53, 54]. Both experiments and simulations were interpreted and compared using an elementary statistical thermodynamic treatment of the helix/coil transition. This method allows us to probe the free energy contributing to the energy landscape and provide insight into the conformational bias, suggesting that the IDP might employ conformational rheostat to folding upon binding mechanism.

2.2 NCBD as a model for demonstrating the LEGO approach.

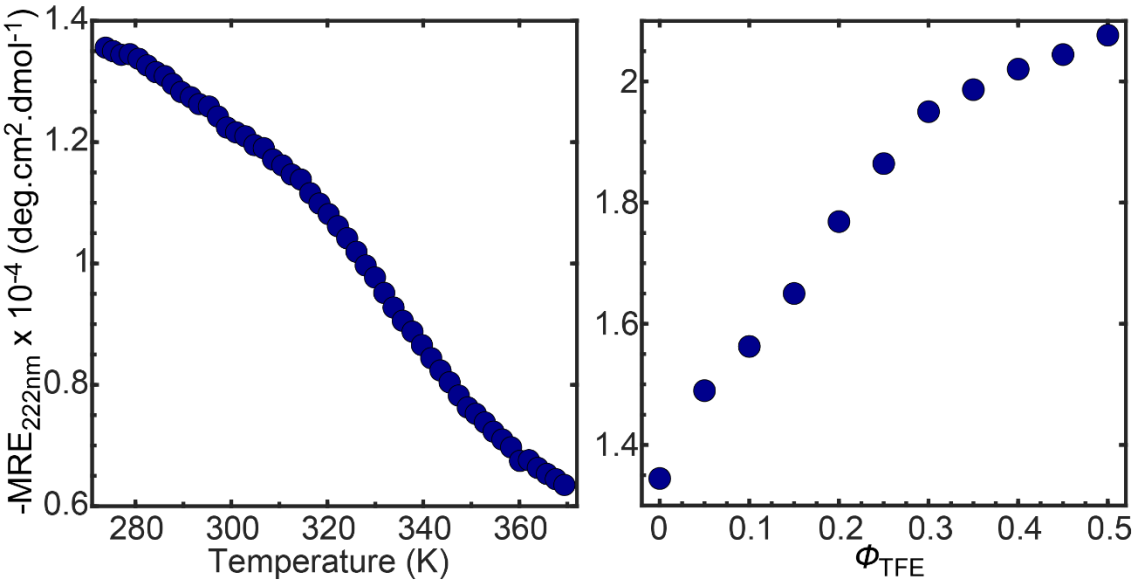

Figure 2.1 Characterization of (un)folding behavior of NCBD using Far-UV CD technique. Left. CD signal at 222 nm as a function of temperature. Right. CD signal at 222 nm as a function of TFE.

Here we demonstrate the molecular LEGO approach by dissecting the folding landscape of the protein NCBD (nuclear coactivator domain), categorized as an IPDP that still adopts some residual structure in its unbound ensemble. Most importantly, folding coupled to binding behavior of NCBD has been extensively studies by different methods, including NMR[55], molecular simulations[56-58], and SM-FRET[59-62], which give us fundamental references to evaluate our approach. NCBD binds to multiple structurally diverse partners such as the transactivation domain of p53 (p53-TAD)[63] and ACTR[55] or well-folded proteins like IRF-3[64], and structures of NCBD in complex with these partners reveal large topological variations. NCBD exhibits high α-helical content by CD without defined tertiary structure in its free form but adopts a three-helix bundle topology defined by a few mid-range contacts[65]. However, the (dis)ordering transitions of NCBD are broad and featureless, including its unfolding by temperature and stabilization via the cosolvent trifluoroethanol (Figure 2.1). These properties make NCBD an ideal model for a proof of concept of the molecular LEGO.

2.3 Design building blocks on NCBD

The first step in the molecular LEGO approach is the design of the building blocks and their combinations based on their secondary structures. IPDPs such as NCBD still have some residual structures that would allow us to design their fragments. We used the NMR ensemble as the basis for dividing the 59-residues sequence of NCBD into four specific blocks that represent the local (secondary) structural segments which are Helix 1 (H1), Helix 2 (H2), Helix 3 (H3), and the C-terminal tail (T). We further redefined the limits of the α-helices based on the predictions of helical propensity from AGADIR[66] to ensure that we did not break the sequence in the middle of the helix-forming regions. This step is straightforward because the NCBD sequence has a delimited helical propensity profile, including strong N- and C-helix capping signals and multiple proline residues in the connecting regions (Figure 2.2).

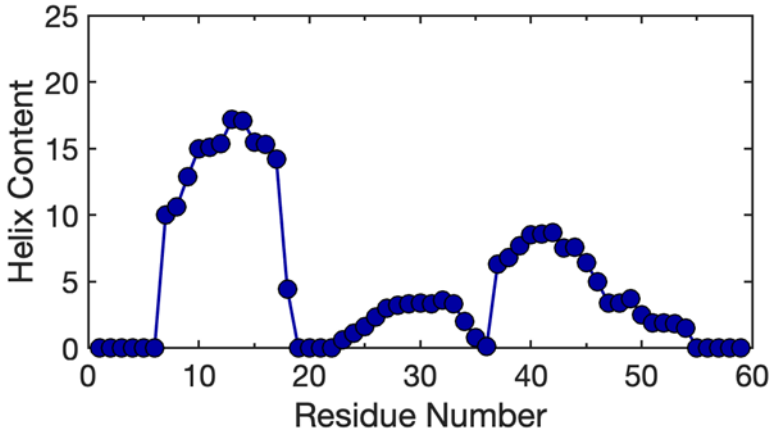

Figure 2.2 Predicted NCBD helical content from AGADIR shows three α-helices at the precise locations determined by NMR experiments[65]

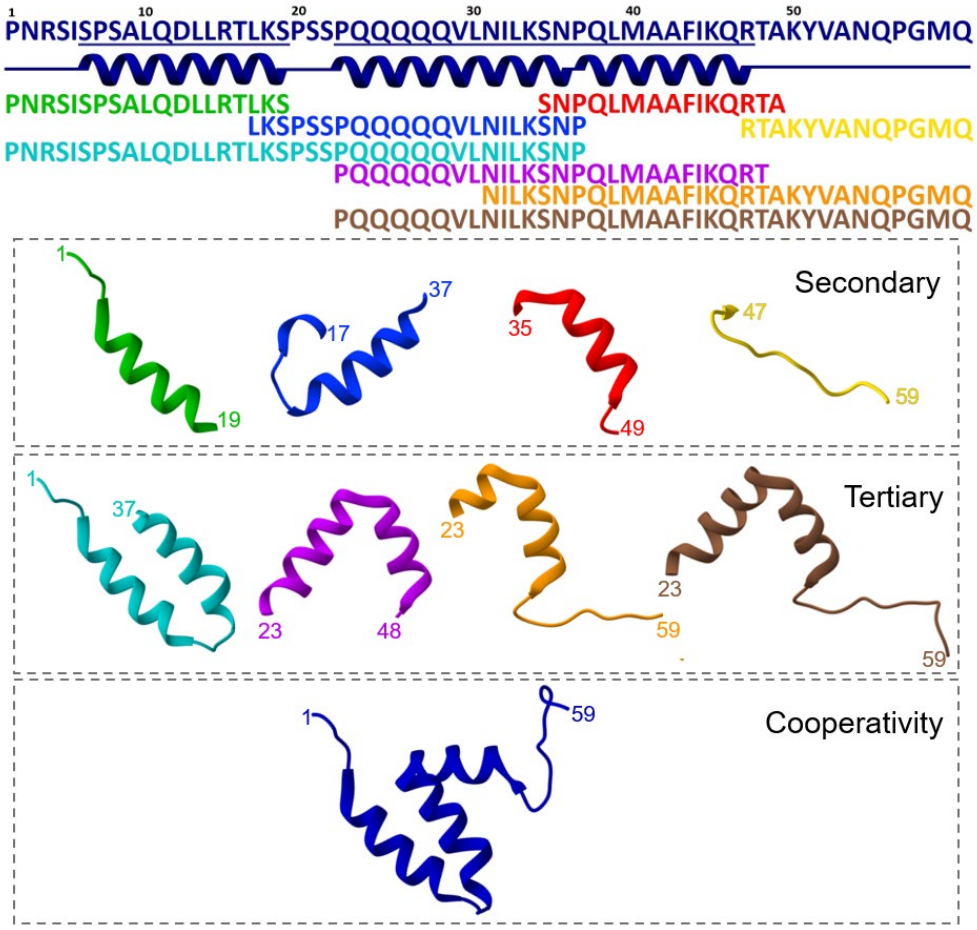

Figure 2.3 Molecular LEGO design. (From top to bottom) The complete NCBD sequence (ID: 2KKJ) and a diagram showing the 3 α-helices found in the NMR ensemble are shown in navy blue. The sequences of the eight fragments, designed according to the sequence and structural patterns of NCBD, are shown color-coded: building blocks in primary colors (H1 green, H2 blue, H3 red, T yellow), and the combined elements in corresponding secondary colors (H1–H2 cyan, H2–H3 magenta, H3–T orange, and H2–H3–T brown). Diagram showing the structure of each fragment and full-length protein in the NCBD NMR structure in cartoon representation. The color coding is maintained. The building blocks report on secondary structure propensities. Building block combinations report on pairwise (element to element) tertiary interactions: e.g., H1–H2 reports on the tertiary interactions between helices 1 and 2. Comparison of the fragments with the behavior of the entire protein reports on the degree of folding cooperativity.

We then defined four combinations of consecutive building blocks (H12, H23, H3T, H23T) that will report on the contributions to the NCBD folding landscape of pairwise tertiary interactions. Finally, the comparison of LEGO elements with the behavior of the entire protein will enable the inference of the overall contribution from global cooperativity to

the NCBD ensemble. Each fragment has few extra residues at the two ends to ensure the entire length of the helix. The complete molecular LEGO design for NCBD is shown in Figure 2.3.

2.4 Experimental Strategy to dissect conformational ensembles

To investigate whether the NCBD conformational ensemble has built-in energetic biases, we analyzed the conformational properties of all of its molecular LEGO elements by experiment and simulation. The MD simulation was done by Dr. Suhani Nagpal.

2.4.1 TFE titration experiment by far UV-CD spectroscopy.

The far-UV CD spectroscopy was employed in this study, which reports the average peptide bond conformation of the peptide and the protein. This technique is particularly sensitive to the α-helical structures, which is suitable in our case as NCBD and many IPDPs becoming α-helical upon binding. We use the cosolvent TFE as the thermodynamic variable to enhance the inherent conformational propensities of all the LEGO elements, including all the fragments and the NCBD protein.

Precisely, the TFE titration from 0 to 0.5 ϕ_{TFE} was performed. TFE is a polar/organic cosolvent that induces structure in peptides and proteins by strengthening the backbone intramolecular hydrogen bonds relative to the hydrogen bonds they make with water[67]. TFE has been widely used as a helix-promoting agent[68] but is also known to stabilize β-hairpin structures[69, 70] and promote hydrophobic interactions by its effects on the hydration shell[71]. The structure-promoting effect of TFE is sequence-independent, and hence we assume that its strength is constant for all the molecules in this study (see Methods). The TFE titration of the building block H1 monitored by far-UV CD is given in Figure 2.4 (left) as an example. In the absence of TFE, the CD spectrum of H1 indicates a population of ~25% α-helix with the remainder being random coil. The addition of TFE steadily increases the α-helical content of H1 until being plateau (from 0.3 to 0.5 ϕ_{TFE}). The increase in helical content is evident in the rise of the negative signal at 222 nm, the settling of the second minimum at ~208 nm, and the increase (up to 2.5-fold) in the ratio of the signal intensities at 193 nm and 222 nm[72]. The CD spectra as a function of TFE exhibit an isodichroic point at ~203 nm, consistent with a structuring transition in which each peptide bond is either coil or helix. Although different in quantitative terms, the TFE titrations of all the other LEGO elements and full NCBD are similar (all data shown in Figure 2.5).

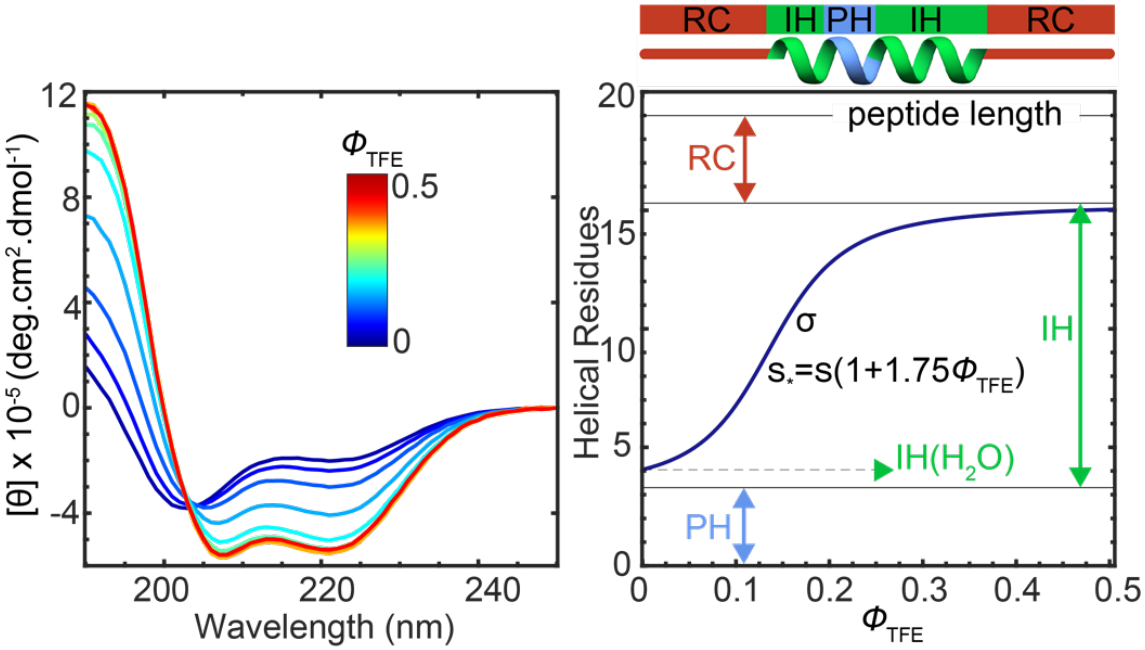

Figure 2.4 Experimental conformational analysis of NCBD and LEGO elements. To probe the energetic biases in the conformational ensemble of each molecule, we use TFE as a structure-promoting agent and monitor the changes in conformation by far-UV CD. The left panel shows the CD spectra (in molar ellipticity units) of H1 as a function of TFE concentration (i.e., volume fraction: ϕ_{TFE}) as an example. The data for all of the other molecules are given in Figure 2.5. The right panel summarizes the tripartite helix-coil analysis of the TFE titration for each molecule: preformed helical residues (PH) in blue, TFE-inducible helical residues (IH) in green, and TFE-insensitive random coil residues (RC) in red. The average number of helical residues (dark blue) is obtained from the CD spectra (see methods). The structure-promoting effect of TFE is equal to $\mathbf{1 + 1.75}\emptyset_{TFE}$ for all the molecules (sequence-independent), and each experimental curve is fit to equation (7) to extract the relevant parameters (nucleation σ, elongation s, IH and PH; RC is simply obtained from the molecule's total length as $\boldsymbol{RC = L_{mol} - IH - PH}$).

2.4.2 Helix-coil transition model to describe the formation of α-helices.

The results from the TFE titration experiments indicate that the interplay between TFE and the folding/structural propensities of the molecules in this study can be analyzed in terms of the statistical thermodynamics of the helix-coil transition[73, 74]. The helix-coil transition describes the formation of α-helices at the residue level as a nucleation (σ) and elongation (s) process[75]. The effect of TFE can be simply described as an increase in the elongation parameter (stronger hydrogen bonds), which promotes a cooperative (sigmoidal) transition to α-helix structure (Figure 2.4 right). However, we should not use a conventional homopolymer model in this case because the amino acid sequences of NCBD and LEGO

elements are highly heterogeneous. On the other hand, the CD signal is averaged over all the peptide bonds, and hence we needed a treatment that is realistic yet simple enough (few parameters) to enable the quantitative analysis of the experimental data. For this purpose, we implemented a tripartite helix-coil model based on the original Zimm-Bragg homopolymer treatment[74]. The tripartite model divides any polypeptide chain into three different types of units (peptide bonds): PH, which is already α-helical without TFE; RC, which are random coil regardless of TFE; and IH, which have residual α-helix population that can be significantly enhanced by TFE (Figure 2.4 right). The model can be used to calculate the average number of helical peptide bonds on any peptide/protein with only four parameters: the number of PH units, and σ, *s*, and the number of IH units. We also assumed the effect of TFE is sequence-independent, i.e., here it was set to $s_* = s(1 + 1.75\emptyset_{\mathrm{TFE}})$. In this treatment, the total helix length at maximal TFE is simply PH+IH.

2.5 Materials and Methods

Recombinant protein expression and purification

The gene encoding for the full NCBD protein was cloned as a His-tag fusion in the bacterial expression vector pBAT4, which is designed to facilitate the expression of unstable and disordered proteins. The plasmid was transformed into *E. coli* BL21(DE3) competent cells. Transformed cells were grown in LB broth at 310 K until the optical density at 600 nm reached a value of 1.2-1.4, followed by induction with IPTG at 291 K overnight. The cells were then harvested by centrifugation at 8000 rpm for 30 min. The pellet was resuspended until homogeneous in 20 mM phosphate buffer at pH 7.5. Cell lysis was carried out using the freeze-thaw method (6 cycles) followed by an ultracentrifugation step at 35,000 rpm for 30 min. After ultracentrifugation, the supernatant was collected and loaded onto a Nickel-column (His-Trap) as the first step in its purification by affinity chromatography. The loaded column was washed with binding buffer (20 mM Tris 150 mM NaCl, 10 mM Imidazole pH 7.5) followed by elution with a gradient from 0% to 100% of elution buffer (20 mM Tris 150 mM NaCl, 500 mM Imidazole pH 7.5). Fractions containing the NCBD protein were pooled and subjected to a second round of HPLC purification on a reverse-phase (RP) column using a 0%-95% acetonitrile gradient with 0.1% trifluoroacetic acid (TFA). All fractions containing pure NCBD were pooled, lyophilized, and stored at 253 K. Protein purity was assessed by SDS-PAGE and verified by electrospray mass spectrometry.

Peptide synthesis

Peptide molecules corresponding to the four building blocks and four combinations (sequences given in Figure 2.3) were chemically synthesized by Bio-Synthesis Inc. (Texas). Peptide purity was determined to be higher than 95% by RP-HPLC and mass spectrometry. Peptides were synthesized with acetylation and/or amidation at the N- and C-terminus, respectively, when the end of peptide corresponded to a non-terminal position in NCBD: H1 and H12 (free, amidated); H2, H3, and H23 (acetylated, amidated); T, H3T and H2H3T (acetylated, free).

Far-UV circular dichroism (CD) spectroscopy

Far UV-CD spectra were recorded from 190 nm to 250 nm with 1nm resolution and 2 nm bandwidth at 278 K on a Chirascan CD spectrometer from Applied-PhotoPhysics ltd. (UK) equipped with a temperature controller system. A rectangular cuvette with a pathlength of 1 mm was used to hold the peptide/protein samples. NCBD samples were prepared at 30 μM, and all the peptides were prepared at 50 μM, except H2 that was prepared at 25 μM due to its lower solubility in 20 mM Tris-HCl. Actual samples were prepared by dilution from a 500 μM stock solution for peptides (300 μM for NCBD) previously prepared in the same buffer. TFE titrations were performed as follows: 200 μL of the initial sample was added to the cuvette, temperature equilibrated for 10 minutes, and the CD spectrum was acquired. Then the volume of TFE required to reach a 0.05 TFE (volume fraction) solution was added. After mixing, the sample was inserted into the cuvette holder, temperature equilibrated for 10 min, and then recording of the spectrum. The same process was repeated iteratively to reach a final TFE volume fraction of 0.5 in 0.05 increments. The protein concentration was corrected for each step to account for the volume changes resulting from the TFE addition. The CD spectra were baseline subtracted (spectrum of a sample of the same buffer). All experiments were done in duplicate. The final protein concentrations of all the samples were calculated relative to an internal reference to minimize errors (due to the low or inexistent molar extinction coefficient at 280 nm of NCBD and peptides) using a three-step procedure: 1) the initial concentration of each sample was estimated from the concentration of the stock solution, which was determined either by absorbance at 280 nm for NCBD (two tyrosine residues, $\varepsilon(280nm) = 2 \times 1{,}280\ M^{-1}cm^{-1}$) or by weight using an analytical scale for all of the peptides; 2) a correction factor based on the ratio of the absorbance at 195 nm (very strong signal: $\varepsilon(195nm) \approx N \times 5{,}000 M^{-1}cm^{-1}$, where N is the number of peptide bonds in the molecule) of each sample relative to that of H1 as internal reference (corrected by the ratio of peptide bonds:N_x/N_{H1}) was applied to minimize pipetting errors, and 3) correction of the changes in total volume due to the TFE added to the sample.

Singular value decomposition (SVD)

The set of experimental CD spectra as a function of TFE were organized as an M x N matrix, with M being the set of wavelength values and N the TFE volume fractions. Each element in the matrix represents the molar ellipticity value of the molecule at a given wavelength and TFE volume fraction. The matrix of molar ellipticities was decomposed by SVD:

$$A = USV^T \qquad (1)$$

where U contains the CD spectral components, S is the diagonal matrix of singular values in decreasing ranking order, and V is a matrix with the amplitude of the U components as a function of TFE volume fraction. The first component corresponded to a pure α-helix CD spectrum, and hence its amplitude (first row of V) represents the changes in helical

content as a function of TFE, which can be converted to the average number of helical residues ($\langle k \rangle$) using the equation:

$$\langle k(\Phi_{TFE}) \rangle = \mathrm{V}_1 \cdot \mathrm{U}_1(222\ nm) \cdot \mathrm{S}_1/(-39{,}500\ deg.cm^2.dmol^{-1}) \qquad (2)$$

where the denominator is the molar ellipticity of one helical peptide bond within a formed α-helix.

Helix-Coil treatment

We describe the formation of helical structure using the Zimm-Bragg helix-coil theory. In the Zimm-Bragg model, each peptide bond can be in either helical conformation (h) or coil (c), and helix formation occurs by process of nucleation (cost of forming the first helical hydrogen bond, defined by the parameter σ) and elongation (defined by the parameter s). With this definition and using the coil as a reference state, the statistical weight matrix is defined as:

$$\mathbf{M} = \begin{pmatrix} 1 & \sigma s \\ 1 & s \end{pmatrix} \qquad (3)$$

for which the partition function is

$$q = (1,0)\mathbf{M}^n \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad (4)$$

where n is the number of peptide bonds in the molecule. The average number of helical residues is simply calculated as:

$$\langle k \rangle = n(s/\lambda_1)(1/2)\{1 + [(s-1) + 2\sigma\,]/[\,(1-s)^2 + 4\sigma s]^{1/2}\} \quad (5)$$

where λ_1 is the largest eigenvalue of the statistical weight matrix:

$$\lambda_1 = \{(1+s) + [(1-s)^2 + 4\sigma s]^{1/2}\}/2 \qquad (6)$$

This treatment is for a homopolymer. To adapt it to heteropolymers and introduce the effect of TFE (TFE titration), we use the tripartite helix-coil model shown in Figure 2.3, in which the heteropolymer sequence is divided into three types of units according to their average nucleation and elongation parameters: peptide bonds that are fully helical in water (PH), peptide bonds that are TFE-sensitive (i.e., inducible helix, IH) and peptide bonds that remain coil at all concentrations of TFE (RC). Using these definitions, we can calculate the average number of helical residues of a given peptide/protein as a function of TFE with the equation:

$$\langle k(\mathrm{TFE}) \rangle = \mathrm{PH} + \mathrm{IH}(s_*/\lambda_{1,*})(1/2)\{1 + [(s_* - 1) + 2\sigma\,]/[\,(1-s_*)^2 + 4\sigma s_*]^{1/2}\} \qquad (7)$$

where $s_* = s + 1.75\emptyset_{TFE}$ and $\lambda_{1,*}$ is the largest eigenvalue of the statistical weight matrix at each TFE volume fraction. The effect of TFE is generic, that is, independent of the protein sequence. Equation 7 provides the basis for the fits to the experimental data presented in Figures 2.6, 2.7, and 2.9.

Estimation of pairwise tertiary interactions and cooperativity

In the Zimm-Bragg model, the statistical weight (*w)* of a given helical conformation is given by $w = \sigma s^i$, where *i* is the number of helical peptide bonds. We can calculate the statistical weight expected for a fully folded molecule containing two helical elements (molecular LEGO's building blocks) that are not interacting with one another as the product of the statistical weights of the fully formed helical elements. Hence, the contributions from tertiary interactions between the two elements can be obtained from the ratio between the statistical weights of the entire molecule divided by the product of the weights of its separated elements as:

$$\Delta G_{mn} = -RT\ln(\frac{w_{mn}}{w_m w_n}) \qquad (8)$$

where $w_m = \sigma_m s_m^{k_m}$ and $w_n = \sigma_n s_n^{k_n}$ are the statistical weights of the fully-induced helical conformation of building blocks *m* and *n*, and $w_{mn} = {\sigma_{mn}}^2 s_{mn}^{k_{mn}}$ is the statistical weight of a molecule containing building blocks *m* and *n* in a full helical conformation. In these expressions, *k* is the number of residues that need to become helical to form the helix(es) defined in the NCBD NMR ensemble (dashed lines in Figure 2.6, 2.7 and 2.9), that is, $k = H_{NMR} - \text{PH}$. We used this procedure to calculate pairwise interactions between H1 and H2 and between H2 and H3. For the tail (T), we considered that its effect on a combined molecule is to extend H3 rather than nucleating a new one (w_{3T} only includes 1 nucleation and w_{23T} includes 2). After the pairwise tertiary interactions have been estimated, the same calculation can be carried out for the entire protein to estimate the overall folding cooperativity. In this case, the fully formed conformation includes three helices, and hence $w_{NCBD} = {\sigma_{NCBD}}^3 s_{NCBD}^{k_{NCBD}}$, relative to the product of the statistical weights of the four elements. The overall folding cooperativity is finally obtained as:

$$\Delta G_{coop} = \Delta G_{NCBD} - (\Delta G_{H12} + \Delta G_{H23T}) \qquad (9)$$

We performed these calculations for the experimental data using the helix-coil parameters given in Table 2.1 and for the MD simulations using nucleation and elongation parameters obtained from the analysis of the trajectories.

All-atom MD simulations.

Computational Design Method and Analysis were performed by Dr. Suhani Nagpal with the computation methods describes below.

We carried MD simulations in explicit solvent using the GROMACS package[76-78] and the Charmm22* force field[79]. Water molecules were described using the TIP3P model. Periodic boundary conditions were used, and long-range electrostatic interactions were treated with the Particle Mesh Ewald (PME)[80] summation using a grid spacing of 0.16 nm combined with a fourth-order cubic interpolation to derive the potential and forces in-between grid points. The real space cutoff distance was set to 1.2 nm, and the van der Waals cutoff to 1.2 nm. The bond lengths were fixed[81] , and a time step of 2 fs was used for the numerical integration of the equations of motion. Coordinates were recorded every 10 ps.

For NCBD, we performed two separate 12 μs trajectories starting from the lowest energy structure of the NCBD NMR ensemble (PDB ID: 2KKJ). The protein was placed in a dodecahedral water box (volume = 262.38 nm^3) large enough to contain the protein and at least a 1.0 nm layer of solvent on all sides. The structure was solvated with 8,216 water molecules, and six Cl^- ions were added to neutralize the system. The starting coordinates for the 8 NCBD fragments (as defined in Figure 2.3) were extricated from the protein's PDB file. The fragments were acetylated and/or amidated as needed to replicate the chemically synthesized peptides (H1, H12 free and amidated; H2, H3, H23 acetylated and amidated; T, H3T, H23T acetylated and free). The CHARMM22* force field was then adjusted to include the parameters for N-acetylation and C-amidation. Box dimensions were kept sufficiently large to account for the high flexibility and large-scale motions expected on these peptides. Two 2-μs trajectories were performed for each fragment (three 2 μs trajectories for the larger fragments H12 and H23T.

In all cases, the starting structure was subjected to energy minimization using the steepest descent method. All systems were equilibrated at a constant temperature of 310 K utilizing the two-step ensemble procedure (NVT and NPT). First, the system was subjected to NVT (constant number of particles, volume, and temperature) equilibration for 100 ps with the position of the protein restrained, followed by NPT (constant number of particles, pressure, and temperature) equilibration for 2 ns each. The simulations were subjected to the modified Berendsen thermostat with a 0.1 ps relaxation time[82] to maintain the temperature. The structures were then subjected to Parrinello-Rahman with 0.2 ps relaxation time for pressure coupling[83] at 1 bar before the production run was started. All the simulations were run on the Triton Shared Computing Cluster (TSCC) at the San Diego Supercomputing center (SDSC).

Analysis of MD simulations.

The number of native contacts per residue was calculated from each MD trajectory with a 1-ns time step and using the NMR structure as the reference of native contacts. Contacts were defined using a 0.5 nm cutoff between any two pairs of heavy atoms that are at least 3 residues apart in the sequence. The number of native contacts trajectory was then converted into the fraction of native contacts (Q).

We used the peptide bonds as a basic conformational unit to compare with experimental data analyzed with the Zimm-Bragg model. Each trajectory was then analyzed to assign each peptide bond of the simulated molecule to either helix or coil state at each time frame. The helical state (*H*) was defined according to the local conformation (dihedral angles) and backbone hydrogen bonding status. These processed trajectories were finally used to calculate the number of helical residues per time frame and the average fraction helix per residue for each molecule.

Analysis of dihedral angles. We classified the conformation of a peptide bond unit based on its flanking ψ and ϕ angles. Particularly, we defined a helical peptide bond (*h*) when its dihedral angles are -50° < ψ < -17° and -80° < ϕ < -50°, and coil peptide bond (*c*) as everything else.

Analysis of hydrogen bonds. A hydrogen bond between residues *i and i*+4 was considered formed when the donor-acceptor distance was < 0.35 nm and the donor-hydrogen-acceptor angle >160°. We computed every hydrogen bond formed at each time frame using the MD Analysis python toolkit: we first evaluated all possible hydrogen bonds per time frame, and then every time a *i,i*+4 hydrogen bond was formed according to our criteria, we assigned a hydrogen-bonded state (h_{HB}) to peptide bonds *i*+1, *i*+2, and *i*+3.

Computing helix nucleation and elongation. We define the elongation parameter (*s*) as the equilibrium constant between the helix and coil states of the central peptide bond in a triplet. For a given time frame, the helix state of the central peptide bond is any of the following: [c h h], [h h h] or [h h c]; and the coil state is either [h c c], [h c h] or [c c h]. For a predefined helical segment, s is simply the average of the elongation for all the peptide bonds within it.

The nucleation parameter (σ) is defined as the equilibrium constant for the formation of the first helical (*i, i*+4) hydrogen bond (flanked by coil peptide bonds). To calculate σ, we used a rolling window of 7 peptide bonds and defined nucleation on the third peptide bond ($\tilde{h}$) as:

$$\sigma = \frac{1}{t}\sum (cc\tilde{h}hhcc)$$

where *t* is the number of time frames in the trajectory. The final parameters for one molecule were determined as the average over all the available MD trajectories.

*Time-averaged native contact map***.** We calculated the time-averaged probability of finding each native contact (native contacts for NCBD and all the fragments defined from the coordinates of the NMR structure) between two residues that are at least 3 apart in the sequence. For each time frame, a contact was considered formed when at least one heavy atom of the first residue was within 0.5 nm distance of at least one heavy atom of the second residue.

2.6 Results and Discussion

2.6.1 2.6.1 Conformational Propensities of LEGO building blocks

TFE titrations of all the LEGO elements and the NCBD protein are shown in Figure 2.5. All the spectra exhibit a helix-coil transition features with an isodichroic point at ~203 nm, two negative bands at 193 nm and 222 nm, in which their ratio of intensities are increased as a function of TFE.

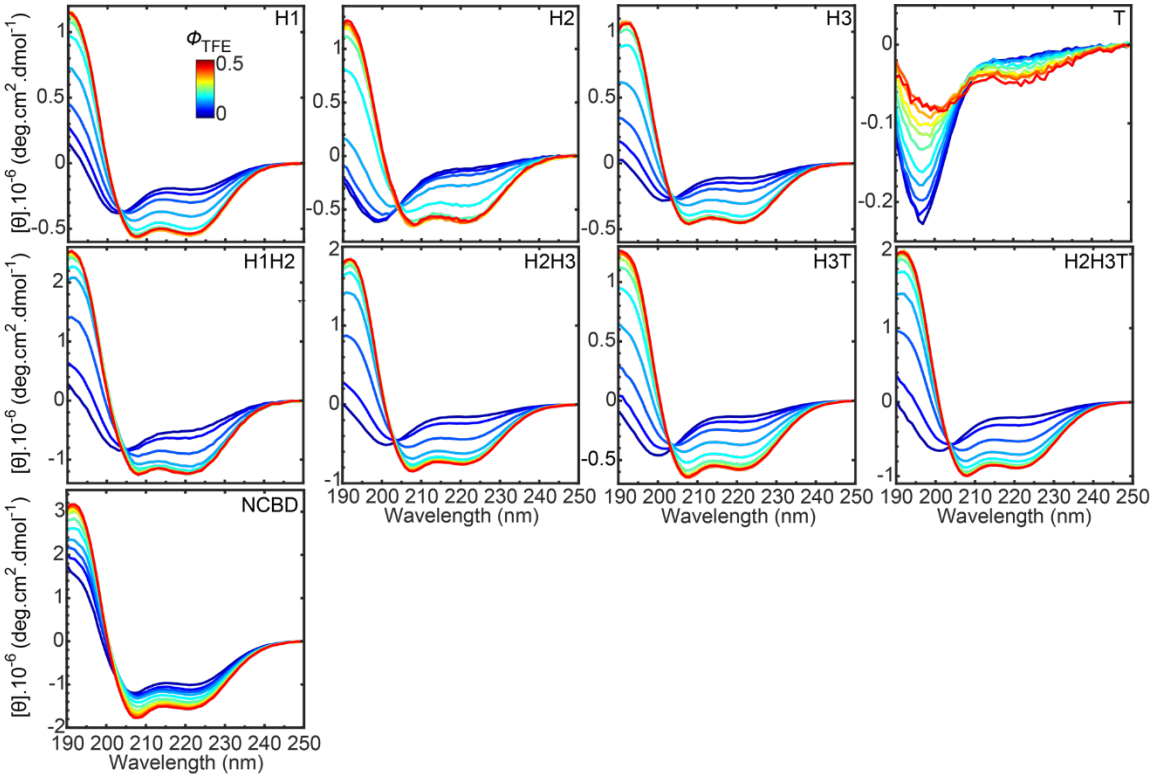

Figure 2.5 Far UV-CD spectra (in molar ellipticity units) of all molecular LEGO elements and full NCBD as a function of TFE volume fraction (ϕ_{TFE} from 0 to 0.5) at 278 K.

We then analyzed the CD spectra as a function of TFE for each peptide/protein to determine the number of helical peptide bonds using singular value decomposition (SVD) analysis. Each dataset was expressed in molar ellipticity units; and a value of –39,500 deg.cm^2.dmol^{-1} for the molar ellipticity at 222 nm of one helical peptide bond[84, 85] was used to convert the data into the number of helical peptide bonds as a function of TFE (see Methods). These data were fitted to the tripartite helix-coil model as described in the Methods section.

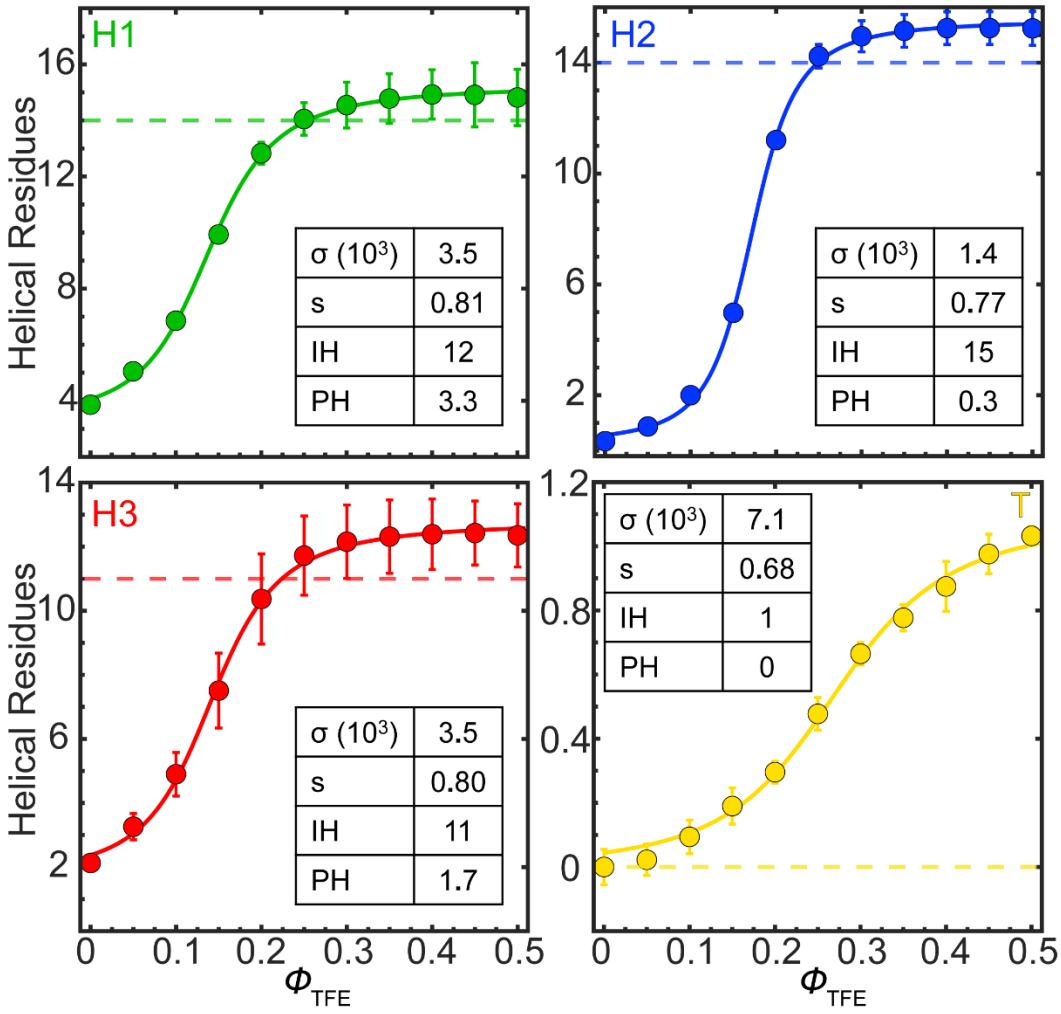

Figure 2.6 LEGO building blocks: secondary structure propensities and local interactions. The experimental conformational analysis of the 4 NCBD building blocks color-coded as in Figure 2.3. The panels show the average number of helical residues (circles) and experimental error, obtained from two independent measurements, as a function of the TFE volume fraction for H1, H2, H3, and T. The colored curves represent the fit to equation (7), and the parameters from the fit are given in the inset. Dash lines (--) indicate the number of helical residues determined from the NMR structure.

The results of the conformational analysis for the four building blocks (H1, H2, H3, and T) are provided in Figure 2.6. In general, these results indicate that the three regions containing α-helices in the NCBD NMR structure have a residual helical structure on their own and are very sensitive to TFE. Of all the elements, H1 has the highest residual helical structure. The maximal helix length of H1, H2, and H3 (i.e., at the highest TFE) is only slightly longer (about one residue) than the helices in the NMR ensemble, which suggests that local signals tightly control the location and extent of the NCBD helices. It is also

apparent that the tail (T) does not have a detectable helical structure, but a minimal helix of about 1 residue (i.e., 1 hydrogen-bonded peptide bond, or 1 helix turn) forms at the highest TFE.

The helical transitions induced by TFE are sigmoidal and can be accurately reproduced by the tripartite helix-coil model using four floating parameters per molecule and a standard parameter for the TFE stabilization (Figure 2.4). The helix-coil parameters for each building block are given as an inset in each panel of Figure 2.6. This experimental analysis shows that the cost of nucleation (σ) for H1, H2, and H3 is within the range of the values found in idealized model peptides used to investigate helix stability[86]. H1 and H3 are slightly easier to nucleate and hence less cooperative than H2. Elongation is somewhat lower than 1 for all the sequences, which explains both their residual helix content (on an infinitely long helix $s = 1$ results in 50% helix content) but also their high sensitivity to TFE (minor tweaks raise *s* above 1). T is interesting because even though it has a minuscule helical propensity overall, it contains a one-turn region that seems primed to become helical by stabilizing factors.

These results demonstrate that the sequence of NCBD contains very specific local signals. Such signals prime certain regions on NCBD to form α-helices upon mild stabilization by other factors (i.e., TFE, tertiary interactions, partner binding) and also seem to define their limits. The consistency between the local conformational biases in the isolated building blocks and the structural ensemble of the full protein suggests that local interactions play a major role in determining the folding landscape of NCBD. These biases are also likely to modulate NCBD's binding properties consistently with a conformational rheostat mechanism.

2.6.2 Conformational biases through pairwise tertiary interactions

The results for the combined LEGO elements are given in Figure 2.7. The behaviors of these molecules should highlight any contributions from pairwise tertiary interactions to the NCBD conformational ensemble. Qualitatively, the experimental results are similar to those of the building blocks: i) residual helical structure in native conditions, ii) strong response to TFE, iii) sigmoidal TFE transitions, and iv) helix populations within the helix lengths of the NCBD NMR ensemble. However, the comparison between the combined LEGO elements and the compounded effects of their separate building blocks (obtained from Figure 2.6, shown as grey curves) reveal significant differences that demonstrate the presence of transient interactions between elements.

We can identify some general effects. All of the combined elements exhibit enhanced sensitivity to TFE, as manifested by the curves with higher slopes and plateauing at lower ϕ_{TFE}, as well as their slightly higher σ and *s* helix-coil parameters. Notably, the experiments do not detect significant increases in residual helical structure in the absence of TFE. Hence, the energetic biases introduced by pairwise tertiary interactions are insufficient on their own to increase the helical content by experimentally detectable levels.

The other general effect is that the thermodynamic coupling between consecutive LEGO building blocks seems to have a significant impact on redefining the maximal helix lengths, most notably for H3.

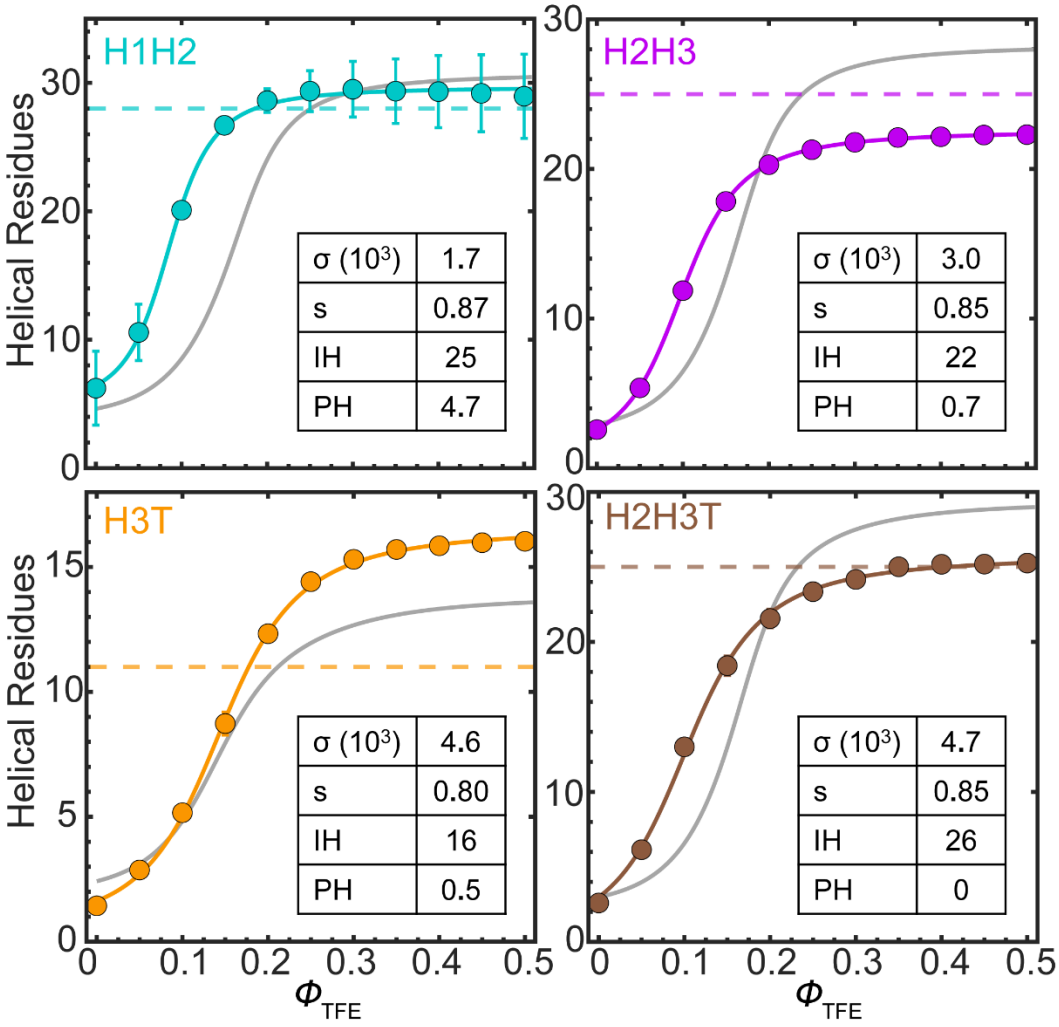

Figure 2.7 Combinations of building blocks: mapping pairwise tertiary interactions. The experimental conformational analysis of the 4 combinations of building blocks that are color-coded in Figure 2.3. The panels show the average number of helical residues (circles) and experimental error, obtained from two independent measurements, as a function of the TFE volume fraction for H12, H23, H3T, and H23T. The colored curves represent the fit to equation (7), and the parameters are given in the inset. The grey curves show the compounded curves of the relevant building blocks for each combination (e.g., H1 and H2 for H12) and represent the reference behavior expected for the combined fragment if the effect is

additive (no tertiary interactions). Dashed lines (--) indicate the number of helical residues determined from the NMR structure.

On an individual basis, we find that the interactions between H1 and H2 are stronger than between 2 and 3. H1H2 does, in fact, exhibit some increase in residual helical content. The effects on H2H3 were detected from the response to TFE in experiments. The impact of the tail on H3 is interesting, as the added C-terminal sequence seems to stimulate the extension of the helix beyond what is observed in the NMR ensemble. The helix extension is evident in the experiments (3 residues longer maximal helix length). In other words, whereas the tail does not nucleate much of a helix on its own, it effectively elongates a helix formed in its preceding sequence.

The experiments on H2H3 indicate a maximal helix of ~23 residues, whereas, in the NCBD NMR ensemble, this region extends over 25, and the H2 and H3 building blocks sum up to 28. At least part of this difference seems to arise from local capping effects of the region connecting helices 1 and 2, which are absent in H2H3 and H2H3T (Figure 2.3).

Furthermore, the presence of helix 2 seems to impede the elongation of H3 into T. This is readily apparent in experiments, which show that H2H3T has a maximum helix of 26 in perfect agreement with the NCBD NMR ensemble. In contrast, the maximal helix lengths of H2 and H3T add up to 30. Strikingly, there also seem to be *"non-native"* interactions (not found in the NCBD NMR ensemble) between H2 and T with an increase in elongation (s) for H2H3T relative to H3T, jointly with a reduced maximal helix length.

2.6.3 Comparison between experimental and computational analysis

As previously mentioned, the same type of analysis was computational performed on the single fragments, all the combinations of the single fragments and full-length proteins. These experimental and computational study results will give us a complete picture as the simulations might provide details at the atomic levels that are not tribute in the CD experiments.

Figure 2.8 shows the simulation data of all the molecular building blocks and their super secondary fragments, as represented in Figure 2.3. Regarding the conformational propensities of the single fragments, we have seen a good agreement between the experiments and the simulation, including the presence of residual helical structures, the average helix population per molecule, especially in the H1 and H3, and the detection of some marginal helical propensity in the T. In addition, the extension of the helical regions in the simulations is in excellent agreement with those of the NCBD NMR ensemble (Figure 2.9A). This result further confirms that the helical regions in the NCBD ensemble are defined by strong local signals.

With respect to the double fragments, we also observed consistency between the experiments and MD simulation, which reinforce the effect of pairwise interactions between the helices (Figure 2.8B), including the interaction between H1 and H2, H2 and

H2 and H3, and H3 and T. Most importantly, the simulation indicated that the helical extension that we saw in the experimental data when H3 connected to the T was driven by the local interactions (helix-coil cooperativity). This extension was also predicted from the AGADIR (Figure 2.2), which further supports the local origin of this effect.

The effects of pairwise interactions on the three-helix lengths are more individualized. For instance, simulations of H1H2 show that pairwise tertiary interactions between H1 and H2 2 increase the intrinsic helical population (mostly at the end of helix 2), but these interactions do not seem to change the maximal length of the helix.

The main discrepancy between experiments and simulations is quantitative: the interplay between helices 2 and 3 with the tail results in a strong stabilization of the two helices in the simulations. The effect is, however, more subtle in experiments. Hence, the simulations overestimate the helical population of the relevant molecules relative to experiments, most particularly H3T and H2H3T, and to a lesser extent, H2H3.

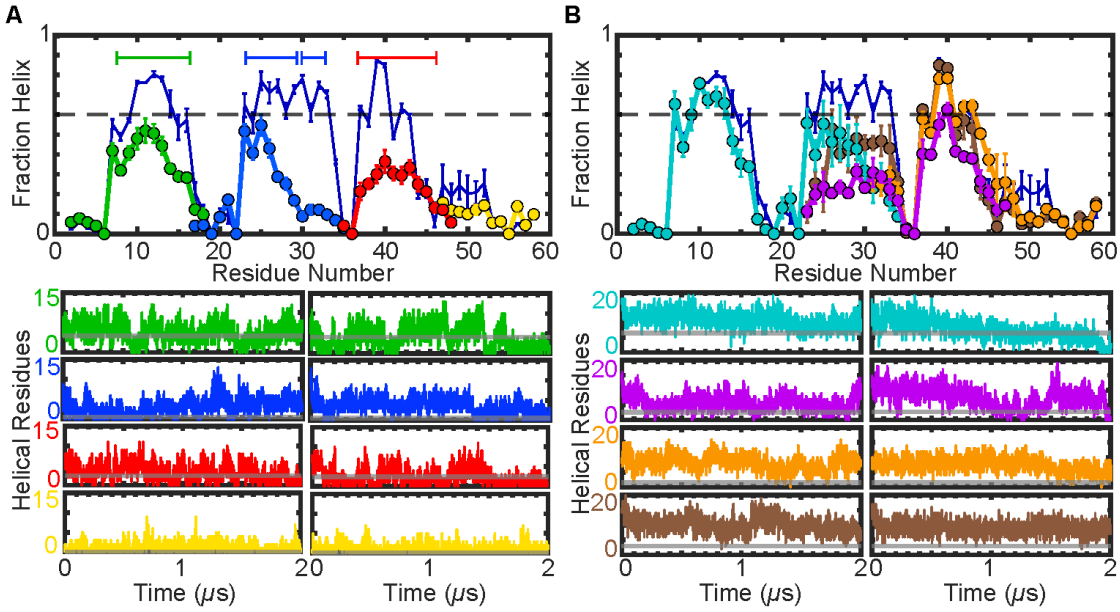

Figure 2.8 MD simulations analysis of all the LEGO building blocks and their combinations. Panel A (top) shows the helical propensity per residue for the 4 building blocks determined from two 2 μs MD simulations. Panel B (Top) shows the helical propensity per residue for the 4 combined LEFO elements obtained from two 2 μs MD simulations. The helical propensity profile for the full-length protein (see Figure 2.9) is shown with a thin navy line for reference. The horizontal lines signal the helix length (consecutive residues with at least 10% fraction helix) emerging from these simulations. The grey dashed line indicates a 60% helicity threshold. Error bars indicate the standard error of two trajectories. The bottom panels show the

time evolution of the number of helical residues for each molecule in two separate 2 μs MD trajectories. The horizontal grey lines indicate the average number of helical residues determined from experiments at $\phi_{TFE}=0$ (ordinate intercept in Figures 2.6 and 2.7), shown for comparison.

2.6.4 Global Stabilization Effects in the NCBD Ensemble

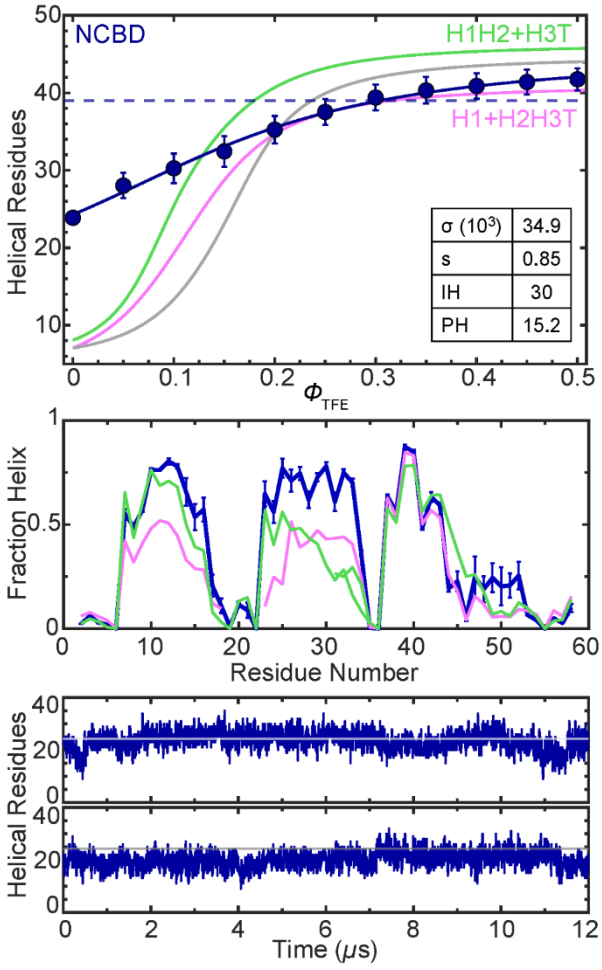

Figure 2.9 Cooperativity in the NCBD conformational landscape. Average number of helical residues (circles) and experimental error, obtained from two independent measurements, as a function of the TFE volume fraction for full-length NCBD. The navy-blue curve represents the fit to equation (7) with the parameters given in the inset. The grey curve shows the compounded curves of the 4 building blocks (H1, H2, H3, T). The pink and light green curves show the compounded curves of H12 with H3T and of H1 with H23T, respectively. The dashed line (--) represents the total number of helical residues from the NCBD NMR structure.

The molecular LEGO analysis allows us to interpret the NCBD TFE-induced transition and examine the role of cooperative effects and non-pairwise interactions (Figure 2.9). The experimental results highlight the seemingly uncooperative (non-sigmoidal) TFE-mediated transition of NCBD (Figure 2.9). This behavior is similar to its temperature denaturation (Figure 2.1) and consistent with previous studies by others[65]. Here it is important to note that the TFE dependence of NCBD is in stark contrast with that of its elements (Figures 2.6 and 2.7). The source for such difference becomes apparent when the NCBD results are looked at in relative terms. By compounding different LEGO elements, we can establish the behavior that would be expected if only local interactions contribute to the NCBD ensemble (grey profile), or if one adds the contributions from H1-H2 tertiary interactions (green profile), or from the interactions between H2, H3, and T (pink profile). These reference profiles reveal that NCBD has much higher helical content than expected from the sum of its parts: about 24 average helical residues in water, relative to 6-7 residues for the three compounded sets of LEGO elements (interactions between helices 1-2 and helices 2-3 do not significantly increase the residual helix, see Figure 2.7). The helix-coil analysis of the NCBD TFE transition indicates that about 15 residues, out of those 24, are fully helical (PH), whereas the remainder corresponds to many additional residues with a partial helical population (i.e., ~30%). Hence, all of the helix-inducible residues (IH) in NCBD already have high helical content in water, which enormously facilitates nucleation, resulting in at least 10-fold higher σ relative to nucleation in the LEGO elements, whereas the cost for elongation (s) is comparable (inset Figure 2.9). In other words, the low sensitivity to TFE of NCBD is not because its conformational ensemble is disordered, but because it is already highly primed towards forming α-helical structure due to interactions that are only present in the full-length protein.

The effect of TFE on well-folded proteins is complex, and switches from native stabilizing at lower volume fractions to denaturing as TFE becomes predominant[67]. We see in the IPDP NCBD that the native stabilizing effect seems to extend up to higher TFE concentrations. Indeed, at 0.5 $\phi_{TFE,}$ NCBD reaches about 41 helical residues, which is in excellent agreement with the NCBD NMR ensemble (dashed line in Figure 2.9). However, the helix-coil parameters indicate that above 0.5 ϕ_{TFE}, TFE will still induce additional helix (~4 residues more in total), hence starting to induce non-native conformations. The extended native stabilizing regime of TFE on NCBD could be due to the fact that this protein is natively α-helical and lacks a defined hydrophobic core[67].

As for the sources of the conformational biases that are unique to the full NCBD sequence, we can obtain key insights from the MD simulations. Notably, the simulations on NCBD (total of 24 μs) reproduce the main experimental results very closely, particularly its overall helical population in water (Figure 2.9 bottom). These simulations show that the increase in the native helical structure of NCBD concentrates mainly on H2 (Figure 2.9 center). This makes sense because, out of the three helical regions, H2 has the lowest intrinsic propensity (Figure 2.6). We also find that, along the simulated trajectories, H2 engages in interactions with the other two helices. Hence the effects of both sets of interactions on the stabilization of H2 are, at the very least, additive. The additive effect of pairwise

interactions between helices 1-2 and 2-3 on the stabilization of the second helix is evident in the comparison of the fraction helix profile of NCBD versus the H1H2+H3T (green) and H1+H2H3T (pink) compounded profiles (Figure 2.9 center). Indeed, this comparison demonstrates that H1 is primarily stabilized by 1-2 interactions, and H3 is stabilized or defined by its interplay with H2 and T. In contrast, the centrally positioned H2 becomes much more stable in the presence of both flanking helices.

Briefly, the combination of experimental and simulation results demonstrated a significant contribution of non-native or long-range interaction in stabilizing the helical structure in the NCBD ensemble.

2.6.5 Interaction network of NCBD

In addition to forming native contacts, the NCBD trajectories show the transient formation of many long-range non-native interactions not seen in the NMR ensemble, particularly between the T and H1 and between H1 and H3 (Figure 2.10). However, although these interactions are not present in the NMR ensemble, they still are consistent with an overall antiparallel helix bundle topology. Importantly, their contribution to the stabilization/formation of the helical structure in the NCBD ensemble seems highly significant. For instance, "non-native" interactions with H1 make T regain some of the helix structure that is suppressed by H2 (Figure 2.9 center). The interactions between H1 and H3, which were not resolved by NMR[87] but are formed in the simulations, could also contribute to the stabilization of the three-helix bundled ensemble.

The contribution of the non-native contacts was demonstrated in Figure 2.10 by MD simulation. NCBD populates a compact ensemble with a great deal of transient but frequent (significantly populated) long-range interactions. Panel A of Figure 2.10 shows the time-averaged "native" contacts observed in NCBD (bottom right) versus those on the LEGO elements (top left). These maps reveal that the H1H2 and H2H3 mostly recapitulate the patterns of native interactions present in the full NCBD, although, in these molecules, the contacts are somewhat less probable. However, in the NCBD ensemble, there seems also to be a significant number of non-native interactions, and which are longer range than the super-secondary structural patterns recapitulated in the LEGO elements (Figure 2.10A right). In the simulations, these long-range "non-native" interactions as the differential factor in cooperatively biasing the conformational landscape of NCBD.

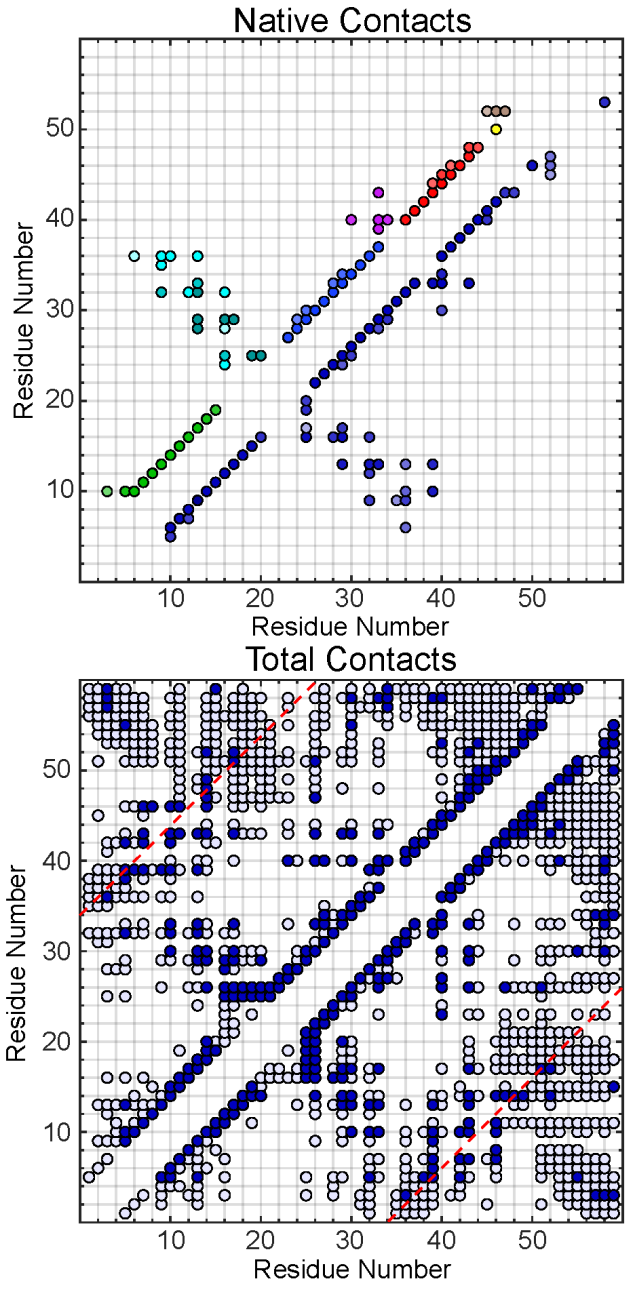

Figure 2.10 NCBD residue-residue interaction maps. Contact maps showing the residue-residue contacts that are formed during the simulations. A contact is considered formed when at least one atom of residue i is within a cutoff distance of 0.5 nm of at least one atom of residue j for a continuous-time of at least 70% of 10 ns. (A). Residue-residue native contacts (i.e., contacts present in the NMR ensemble structure). The top left triangle shows the native residue-residue contacts on all of the molecular LEGO elements (color code as in Figure 2.3), and the bottom right on

the full-length NCBD. The color intensity reflects the time-averaged probability of observing the contact in the log scale, with the lightest color corresponding to a probability between 10^{-4} and 10^{-3} and the strongest intensity for probabilities between 10^{-1} and 1. The LEGO building blocks only have local contacts. For the combined LEGO elements, we only show non-local contacts (> *i*, *i*+5) to avoid clutter. (B). Residue-residue total contacts (native and non-native) were observed in the simulations of full NCBD. Contacts (following the exact definition given above) have been parsed in two groups: dark navy blue for contacts present at least 10% (≥ 0.1 probability) and light navy blue for contacts present for at least 1% but less than 10%. The diagonal red dashed lines signal the maximum threshold for native interactions (≤ i, i+34) defined as per the long-range NOEs reported in the NMR structure.

2.6.6 The NCBD conformational landscape: estimation of energetic contributions

To estimate the energetic contributions that each set of interactions have on the NCBD ensemble, we resourced the helix-coil parameters from all the molecules (Table 2.1). Using these parameters and some simple assumptions, we calculated the statistical weight for forming each of the full α-helices on each molecule. From the ratio between the statistical weight of a complex molecule (e.g., H1H2, or NCBD) and the product of the weights of its building blocks, we estimated the difference in free energy for each set of interactions (see methods).

From Table 2.2, the results indicate that the free energy contributions from each set of pairwise tertiary interactions (H1-H2 and H2-H3) are ~5-6 kJ/mol, which is comparable to the average perturbation induced by single point mutations on folded proteins[88]. The interplay between H2 and H3 with the T contributes ~3 kJ/mol more, whereas the overall stabilization of the NCBD ensemble amounts to ~30 kJ/mol. The NCBD overall stabilization is hence comparable to the unfolding free energies measured by chemical denaturation on many two-state folding proteins[19], even though NCBD is an IPDP without a defined 3D structure. However, this comparison is somewhat misleading because the 30 kJ/mol of NCBD is in reference to a completely disordered ensemble (compounded building blocks). In contrast, chemically unfolded states, especially of the least stable and faster folding proteins, have significant local structure levels[32].

	H1	H2	H3	T	H1H2	H2H3	H3T	H2H3T	NCBD
σ (10^3)	3.5	1.4	3.5	7.1	1.7	3.0	4.6	4.7	34.9
s	0.81	0.77	0.8	0.68	0.87	0.85	0.8	0.85	0.85
IH	12	15	11	1	25	22	16	26	30
PH	3.3	0.3	1.7	0	4.7	0.7	0.5	0	15.2

Table 2.1 Helix-coil parameters from all the building block elements, their combinations, and the full-length NCBD.

These calculations also enable the estimation of cooperative (non-additive) contributions to the NCBD ensemble. Particularly, subtracting the pairwise interactions from the NCBD total stabilization leads to an estimate of ~17 kJ/mol for the global NCBD cooperativity. This value presumably includes the contribution from simultaneously forming the interactions between H1–H2 and H2–H3 in the ensemble, as well as the non-native interactions (H1 with H3 and T) that we see in the simulations (Figure 2.10). Importantly, the simulations also reveal that some of these sets of interactions do not form simultaneously but alternate with one another, which explains why they do not favor a unique structure but rather provide global stabilization to a broad, dynamic ensemble. Therefore, the NCBD ensemble that emerges from our results is one with strong conformational biases but conflicting tertiary interactions. This description is entirely consistent with a conformational rheostat ensemble providing the mechanism to drive the complex binding/functional behavior of NCBD.

	ΔG_{exp} (kJ/mol)
H1–H2	-5.1
H2–H3	-5.7
H3–T	-1.0

H2–H3-T	-8.6
NCBD	-30.8
Cooperativity	-16.6

Table 2.2 Energetic contributions from pairwise tertiary interactions to the NCBD ensemble. The change in free energy ($\Delta \boldsymbol{G}$) for given composite molecules (combinations or full protein) that is due to non-additive contributions (tertiary interactions) estimated from the σ and *s* parameters of the composite molecule relative to its building block elements from experiments and simulations (see methods). The cooperativity is obtained by subtracting the tertiary contributions for H1–H2 and H2–H3–T from the NCBD total change in free energy.

2.7 Conclusions

We have successfully introduced a molecular approach to measure the energetics that determine IDP's conformational landscapes in the absence of partners based on the hierarchical organization of protein structure. The disordered protein is dissected into its putative secondary structure elements and their super-secondary arrangements as the first step. The secondary structure elements are considered the LEGO building blocks, whereas the combined elements probe the role of pairwise tertiary interactions, in analogy to the complementary indentations between LEGO blocks that give rise to higher-order toy structures. The conformational biases of LEGO elements and the entire protein are then investigated via experiments, which provide the global behaviors and reality checks, and molecular simulations to gain structural insights at the atomistic level. Finally, the comparative hierarchical analysis of conformational behaviors using elementary statistical mechanical models render estimates of the free energy contributions from each set of relevant interactions to the folding landscape.

In principle, the approach should be easily generalizable to other IPDPs, and hence it adds a powerful tool to the IDP research toolbox. In this regard, we outline some basic rules for applying the molecular LEGO to other disordered proteins:

- A key element is the design of the LEGO elements. Ideally, one should use a structural ensemble of the unbound protein determined with one of the existing approaches for generating IDP ensembles from experimental structural restraints[89-91]. An alternative could be a structure of the IDP in a complex with a partner. In a worst-case scenario, the design could be based on secondary structure prediction profiles[92].

- Because these proteins are flexible/disordered, it is preferable to use a structure-promoting cosolvent as a thermodynamic variable, facilitating more direct comparison with their folding upon binding. TFE is a good option, particularly for IDPs that form α-helical structures (free or upon binding). Other alternatives are osmolytes, such as betaine and TMAO[93], and salts, given that IDPs have very high net charges[94].
- The conformational analysis should be carried out with techniques sensitive to the backbone conformation. Residue-averaged information is sufficient to address general mechanistic questions, as we do here using circular dichroism or with infrared spectroscopy. NMR provides residue-specific structural information, with the caveat of being much more labor-intensive (i.e., for studying all of the LEGO elements). To interpret the conformational biases of broad ensembles, it is essential to use a statistical thermodynamic treatment rather than assuming a two-state transition. The analysis could still be fairly simplified, but it should consider conformational entropy explicitly in terms of ensembles of microstates. In this regard, the molecular simulations allow the researcher to test the significance of the model used to analyze the experiments.

As we demonstrated the approach on the NCBD protein, we have revealed key information related to the energetic factors that control the conformational behavior of the NCBD or IDPs in general. Indeed, we find that the NCBD folding landscape has built-in energetic biases that compete to stabilize the different conformational sub-ensembles that NCBD forms in complex with various partners. This behavior highlights an active internal mechanism to select partners and modulate affinity, which is likely essential to its recruiting role as transcription coactivator[95], and it is possibly the first direct demonstration of a conformational rheostat. Thus, the molecular LEGO approach emerges as a powerful tool for dissecting the conformational ensembles of IPDPs and detect the presence of subtle, functionally relevant, energetic biases.

CHAPTER 3

3 Engineering a conformational rheostatic pH transducer by (un)folding coupled to binding mechanism

3.1 Introduction

The engineering of protein folding/unfolding equilibria coupled to binding to a suitable ligand offers a generalizable strategy for developing biosensors that exploit the unparalleled specificity and selectivity of protein-mediated biomolecular recognition[43]. The strategy entails engineering the protein to be intrinsically unstable in the absence of ligand and use the free energy provided by binding to the analyte in question (which only binds to the native structure) to trigger refolding transducing the binding event into a monitorable output[96]. These transducers toggle between the unfolded-free and the folded-bound states given that their folding mechanism is usually two-state, and the native structure is only formed upon binding the ligand, thus exemplifying the operation of conformational switches[97]. The results are binary signals and typically sigmoidal saturation curves that provide sensitivity to ligand concentrations within 20-fold below and above the apparent Kd or C50[98]. Another characteristic of such transducers is that their time response is ultimately determined by the rate of folding into their native state, which can take up to minutes for two-state folding proteins[88]. These characteristics make it challenging to produce protein transducers capable of broadband and/or real-time sensing, features that are often desired in biosensing applications. In that regard, one exciting possibility is to use downhill folding proteins as scaffolds for building transducers based on (un)folding coupled to binding.

Downhill proteins fold and unfold in very short times (microseconds) and change their structural properties gradually upon (de)stabilization, which results in broad, structurally heterogeneous (un)folding transitions. In fact, it has been proposed that the thermodynamic coupling between a biological signal and the gradual (un)folding of a downhill protein can result in conformational rheostats, a mechanism by which the protein produces analog responses to the input strength, e.g., ligand concentration. The noncanonical features of downhill protein folding present a unique opportunity for building conformational transducers with broadband sensitivity and real-time response.

Initially, we explored the merit of this idea on the BBL domain, a showcase of the most extreme, one-state downhill (un)folding behavior[22] and microsecond folding kinetics[99].

BBL folding is naturally pH-sensitive due to two histidine residues that are partially buried within the protein core. A detailed study of the pH response of BBL has shown that this protein changes its structure gradually over 4 orders of magnitude in proton concentration and can record changes in pH with response times of a few microseconds[18]. The BBL study did not address whether such remarkable broadband response is extensible to other downhill (un)folding coupled to binding processes, or rather a unique result of natural selection on BBL.

We investigate this hypothesis by engineering pH transducing into the naturally pH insensitive, downhill folding protein gpW. Notably, we engineered histidine grafts into its hydrophobic core to induce unfolding via histidine ionization. There are several reasons for using pH sensing for this study: (1) proton binding-release is a relatively straightforward process to engineer into proteins using histidine grafts; (2) histidine grafting allows for the introduction of multiple proton binding sites onto the protein scaffold as a strategy for amplification or modulation of the transducer response; (3) ionization–deionization processes are the fastest reactions in aqueous solution because proton transfer is orders of magnitude faster than conventional diffusion-controlled processes[100]. Such an ultrafast binding process makes the transducer response time to be solely determined by the conformational transition of the protein, which takes place in microseconds for downhill folders.

We then designed and tested ionization effects via computational modeling and studied experimentally the four most promising single grafts and two double grafts. All tested mutants become reversible pH transducers in the 4−9 range, and their response increases proportionally to how buried the histidine graft is. Importantly, the pH-dependent reversible (un)folding occurs in a rheostatic fashion, so the engineered transducers can detect up to 6 orders of magnitude in $[H^+]$ for single grafts and even more for double grafts. Our results demonstrate that downhill (un)folding coupled to binding produces the gradual, analog responses to the ligand (here H^+) that are expected of conformational rheostats, and which make them a powerful mechanism for engineering transducers with sensitivity over many orders of magnitude in ligand concentration (broadband).

3.2 GpW protein as a scaffold for pH conformational rheostatic transducer

To determine the extensibility of the broadband response of transducers based on the downhill (un)folding coupled to binding and rationalize its structural/energetic determinants, we decided to de novo engineer pH transducing into gpW (W protein of bacteriophage lambda). GpW is a protein that folds into an α+β topology[101] and is stable and unaffected by pH in the 4−9 range[102]. Moreover, gpW folds and unfolds in microseconds and exhibits the thermodynamic features of a downhill folding domain, including a minimally cooperative unfolding transition and multiprobe dependence[23], as well as anatomically heterogeneous thermal unfolding process as probed both by NMR and long-time-scale molecular dynamics (MD) simulations[102]. Relative to BBL, gpW is

considerably larger (62 vs. 45 residues), is a full gene product rather than an excised domain[23], and features an all-antiparallel fold with two distinct hydrophobic cores (Figure 3.1) that offers various structural loci for engineering pH transducing.

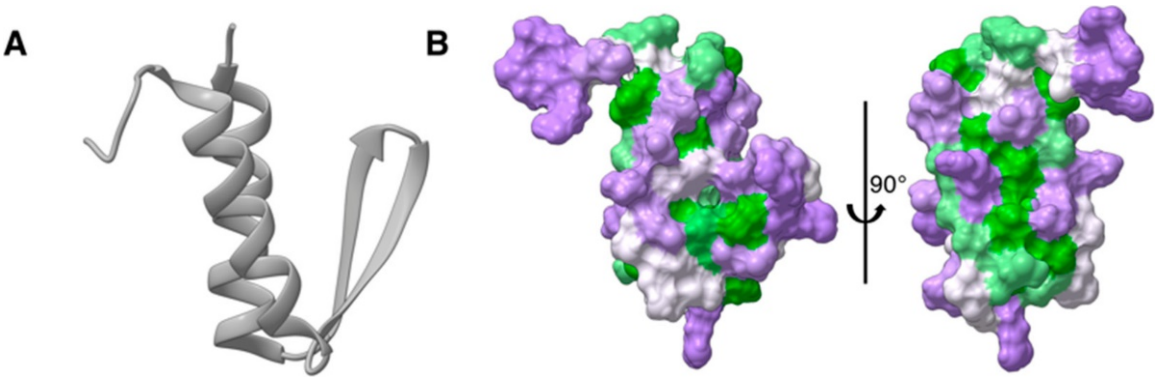

Figure 3.1 Structural features of gpW. (A) The protein is composed of 62 residues forming an all-antiparallel α + β topology consisting of one β-hairpin and two α-helices. (B) Molecular surface representation of the gpW native structure with color-coding signifying the degree of hydrophobicity from polar (purple), intermediate (white), to hydrophobic (green). The two projections highlight the cores between helix-2 and the β-hairpin (left) and between the two helices (right).

3.3 Design strategy for the conformational pH transducer based on histidine grafting.

To de novo engineer pH transducing into gpW, we resorted to a histidine grafting strategy by which we introduce histidine residues into structurally targeted protein locations. Several groups have reported that the dual aromatic/ionic character of histidine and its pKa value close to physiological pH (i.e., 6 − 6.5) can induce pH-dependent conformational changes in the native ensemble of a variety of proteins[103-108]. The role of histidine ionization as a trigger of conformational transitions on proteins is also exploited functionally by nature, like in the transmembrane protein OmpG, which opens and closes its central pore in response to the (de)ionization of two histidine residues[109]. The molecular mechanism behind these conformational changes hinges on the ability of the histidine residue to interact favorably with a surrounding hydrophobic environment in nonionic form and strongly destabilize the same environment when ionized (i.e., due to charge desolvation). Namely, a histidine that is buried into the core of a protein destabilizes the folded state upon ionization by an amount proportional to its pK_a dropping ($\Delta\Delta G_{+0}^{UN} = 2.3026 \times RT\Delta pK_a$). Mutational analyses have shown that the pK_a shifts of buried histidine residues vary widely, being ultimately determined by the local environment in the native structure, including the degree of burial and the interactions with surrounding residues[110].

Practically, we implemented a computational/experimental histidine grafting strategy in four steps: (i) identification of structural loci in gpW suitable for accommodating a partially buried histidine graft; (ii) mutation design in silico, followed by computational assessment of mutant stability through all-atom MD simulations in explicit solvent; (iii) selection of mutations that do not drastically perturb the stability of the native fold in nonionic form, and production in the lab; (iv) computational and experimental analysis of the pH response of each select mutation. Our results on four single and two double histidine grafts on gpW show that the approach reliably introduces a pH-dependent unfolding mechanism onto a naturally pH insensitive protein in the 4−9 range. Moreover, the structural changes of the mutants can extend for over >6 orders of magnitude in [H^+], demonstrating that downhill (un)folding coupled to binding is a powerful approach for engineering broadband conformational transducers.

3.4 Materials/Methods

3.4.1 Computational methods

Computational Design and Analysis were performed by Dr. Suhani Nagpal with the computation methods describes below.

Design strategy

To design mutations to histidine in core positions with varying degrees of solvent exposure of the protein gpW (PDB ID: 2L6Q), we used the Chimera tool and DUET algorithm[111]. A combination of structural analysis to identify target locations and stereochemical criteria was used to identify conservative replacements to histidine (e.g., enough room to accommodate the imidazole ring). Target mutation sites were ranked according to the predicted change in stability upon mutation calculated with DUET. The six single-point mutations to histidine were designed with the Chimera tool and refined via energy minimization. The fully solvent-accessible histidine[23] in gpW was also replaced with Ala to investigate the effects of ionization of the natural histidine. As a further computational test of the intrinsic native stability, all-atom MD trajectories in explicit solvent were run for each mutant in deprotonated form as well as for wild-type gpW.

All-atom MD simulations

MD simulations were performed in the GROMACS suite[112] using the OPLS all-atom force field[113]. Water molecules were modeled with the TIP4P representation[114]. Periodic boundary conditions were used, and long-range electrostatic interactions were treated with the Particle Mesh Ewald (PME) summation using a grid spacing of 0.16 nm combined with a fourth-order cubic interpolation to derive the potential and forces in-between grid points[80]. The real space cutoff distance was set to 1.0 nm, and the van der Waals cutoff to 1.0 nm. The bond lengths were fixed[81], and a time step of 2 fs was used for numerical integration of the equations of motion. Coordinates were recorded every 10 ps. The

simulations were performed at 310 K starting from the coordinates of the lowest energy conformer in the gpW NMR structural ensemble modified to carry the mutations to histidine and the H15A pseudo-wildtype. The protein was placed in a dodecahedral water box large enough to contain protein and at least 1.0 nm of solvent on all sides (volume ~233 nm^3). The structure was solvated with ~7300 water molecules, and 4−5 Cl^- ions were added to neutralize the system. The starting structure was subjected to energy minimization using the steepest descent method. All systems were equilibrated at a constant temperature of 310 K utilizing the two-step ensemble process (NVT and NPT). First, the system was subjected to NVT (constant number of particles, volume, and temperature) equilibration for 100 ps with the position of the protein restrained, followed by NPT (constant number of particles, pressure, and temperature) equilibration for another 100 ps. The simulations were subjected to the modified Berendsen thermostat with 0.1 ps relaxation time to maintain the exact temperature[115], followed by Parrinello− Rahman[116] with 0.2 ps relaxation time for pressure coupling at 1 bar before the production run was started. All the simulations were run on the Triton Shared Computing Cluster (TSCC) at the San Diego Supercomputing center (SDSC). The total simulation time per variant was 1 μs for the wild-type, 0.75 μs for the mutants L7H, M18H, F35H, and V40H, and 0.4 μs for A10H and A13H.

3.4.2 Experimental methods.

Recombinant protein expression

All gpW variants, including the wild-type, single mutants, and double mutants, were produced using recombinant means. Wild-type gpW and the single His mutants were cloned as full genes in the bacterial expression vector pBAT4. Double His mutants were cloned as SUMO-His-tag fusions to improve protein expression and facilitate the expression of these unstable mutants. Plasmids containing the various gpW genes were transformed into E. coli BL21(DE3) competent cells. Cells were grown in LB broth at 310 K until the optical density at 600 nm reached a value of 1.0−1.2, followed by induction with isopropyl-β-D-thio-galactopyranoside (IPTG). After IPTG addition, cells were kept in growing conditions at 310 K for 4 h and then harvested and centrifuged at 8000 rpm for 30 min. The pellets were resuspended until homogeneous in the buffer required for protein purification. Cell lysis was performed using the freeze−thaw method (6 cycles). The lysates were subjected to ultracentrifugation at 35 000 rpm for 30 min.

Protein purification

The supernatant obtained after ultracentrifugation was loaded onto an HPLC HiTrap SP cation exchange column (GE Healthcare) and eluted with a gradient from 0 to 1 M NaCl in 20 mM phosphate buffer at pH 6.5. The fractions containing the gpW protein variant were pooled and subjected to a second round of HPLC purification on a reverse-phase (RP) column using a 0−95% Acetonitrile gradient with 0.1% trifluoroacetic acid (TFA) for elution. Double His mutants (SUMO-His-tag fusions) were purified with the first step of

affinity chromatography by loading the sample onto a Nickel Column (His-Trap) followed by washing with binding buffer (20 mM Tris pH 7.5, 150 mM NaCl, and 10 mM Imidazole), and elution with a gradient from 0% to 100% elution buffer (20 mM Tris pH 7.5, 150 mM NaCl, and 500 mM Imidazole). The protein tags were cleaved using ULP1 by incubation at 277 K overnight and subsequently removed by a second pass through the Nickel column (HisTrap) followed by Reverse Phase chromatography (as above). All fractions containing pure gpW variant were pooled, lyophilized, and stored at 253 K. Protein purity was assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), and mutant identity verified by electrospray mass spectrometry.

Far-UV circular dichroism (CD) spectroscopy

All the protein samples were prepared at 30 μM concentration in 20 mM buffer by dilution from 300 μM stocks prepared in the same buffer. Citrate buffer was used for experiments in the 3−6 pH range. Phosphate buffer was used for pH 7, and Tris-HCl buffer for the 8−9 pH range. HCl in water was used to prepare samples at pH 2. CD spectra were measured from 195 to 250 nm with 1 nm resolution and 2 nm effective bandwidth. Thermal denaturation curves monitored by CD were performed on a Chirascan CD spectrometer from Applied Photo Physics equipped with a temperature controller system. A cuvette with a path length of 1 mm was used. Thermal denaturation experiments were performed by recording CD spectra every 2 K from 273 to 373 K. CD spectra were baseline subtracted (spectrum of the same buffer). All experiments were done in duplicate. The relative final protein concentration of all samples for each gpW variant was calculated using a two-step procedure to minimize errors due to the small molar extinction coefficient of gpW at 280 nm: (1) concentrations were estimated from the stock solution concentration determined by absorbance at 280 nm (gpW has one tyrosine, $\varepsilon(280\ \text{nm}) = 1280\ M^{-1}cm^{-1}$) after applying the corresponding dilution factor; and (2) a correction factor based on the ratio of the absorbance at 195 nm ($\varepsilon(195\ \text{nm}) \approx 300\ 000\ M^{-1}cm^{-1}$) of the sample relative to that of an internal reference (wild-type gpW at pH 2) was applied to minimize pipetting errors.

Reversibility test of pH-induced protein (un)folding

The far UV-CD spectrum of the samples was recorded from 190 to 250 nm at 293 K. The initial samples were prepared by dissolving the lyophilized protein in water at a concentration of 30 μM in a microcentrifuge tube. The pH was adjusted by adding either NaOH (0.01 M or 0.1M) or HCl (0.01 M or 0.1M) in 1 μL steps until reaching the desired pH value (3 or 7). The sample was transferred to the CD cuvette, temperature equilibrated for 10 min, and the spectrum was acquired. The solution was then transferred back to the microcentrifuge tubes and adjusted to the other extreme pH (7 or 3), adding NaOH or HCl in 1 μL steps measuring the final pH with an ultrathin electrode. The sample was retransferred to the CD cuvette for spectrum measurement. After the acquisition, the entire process was repeated to adjust the pH back to the original value (3 or 7) by adding HCl or NaOH in 1 μL steps and recording the CD spectrum at the reversed pH. The added volumes

were recorded and used to correct for the actual changes in protein concentration. The entire procedure was done three times, and the triplicated spectra were averaged.

3.5 Results

3.5.1 Investigating pH sensitivity of the protein gpW wild-type

GpW does indeed fold and unfold in microseconds[23], but to be a suitable scaffold for this study, it also needs to be naturally insensitive to pH in the neutral to the mildly acidic range that is most biologically relevant (between 4 and 9). From an amino acid composition viewpoint, gpW has a sole histidine (H15) located on the exterior of helix-1 and fully solvent-exposed, and therefore unlikely to experience significant pK_a shifts.

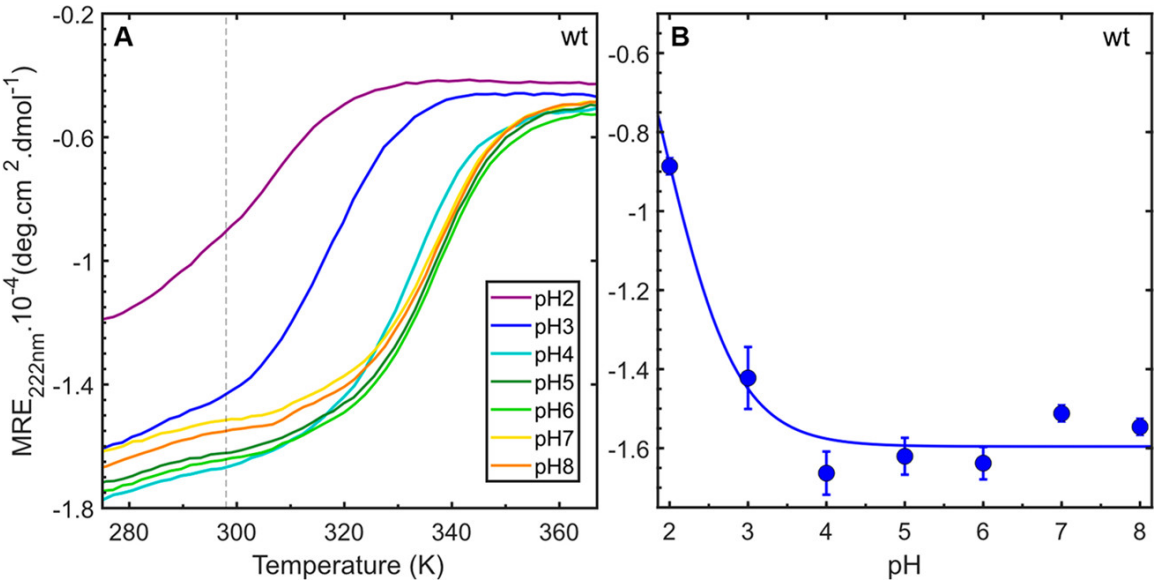

Figure 3.2 (A) Thermal unfolding curves of gpW at different pH monitoring the changes in the native α-helical content. Each curve is an average of two sets of data. Curves are colored according to the visible light spectrum, with the lowest pH corresponding to the highest energy (purple). (B) The blue circles represent the α-helical content of gpW as a function of pH at 298 K. The blue line is to guide the eye.

We determined the natural pH sensitivity of gpW by measuring the thermal unfolding of the wild-type (wt) protein as a function of pH. The results of these experiments are given in Figure 3.2. Panel A shows the thermal unfolding of gpW measured by CD at 222 nm (mean residue ellipticity, $MRE_{222\ nm}$, which reports on the strong native α-helical signal of gpW) as a function of pH. Protein stability appears completely unaffected down to pH 5. At pH 4 there are signs of slight destabilization that are sharply enhanced at lower pH. The behavior of gpW below pH 4 reflects the protonation of its glutamate and aspartate

residues, which results in a large positive net charge on the protein (pI > 10) that promotes its unfolding.

The destabilization at very low pH is also evident as a drop of the denaturation midpoint temperature (T_m) (Figure 3.3). However, from the viewpoint of the native signal at room temperature (298 K), the dependence on pH is flat down to pH 4; past this point, gpW loses native CD signal, but even at pH 2 there is still about 30% native signal left (Figure 3.2). As an additional test, we produced and studied the H15A gpW mutant, which confirmed that removing the imidazole at position 15 does not significantly affect gpW stability as measured by CD, nor its pH sensitivity in the 4–9 range. In other words, gpW emerges as a suitable protein scaffold for engineering conformational pH transducing.

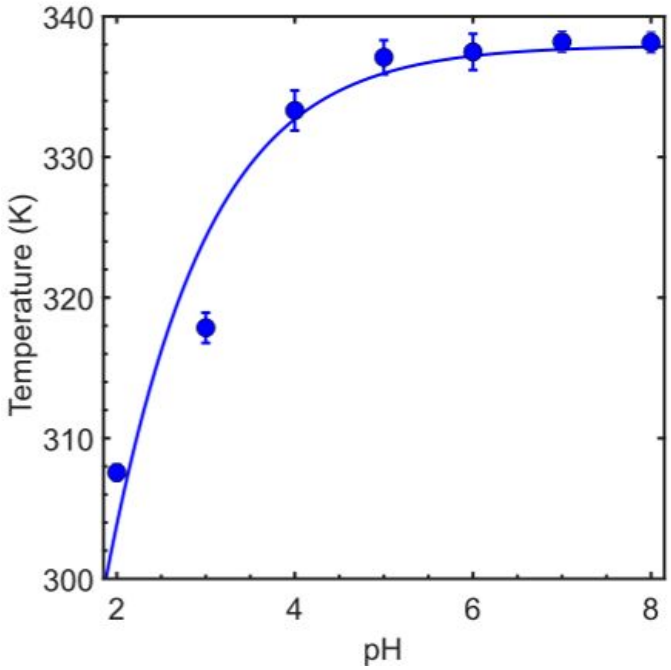

Figure 3.3 Midpoint temperature of wild-type gpW in the pH range 2 to 8 (circles) determined by calculating the derivative of the thermal unfolding curves. The blue line is an exponential curve shown to guide the eye.

3.5.2 Mutation Design Strategy in silico

All the computational analysis were performed by Dr. Suhani Nagpal

Histidine grafting principal

The imidazole ring of histidine is ionizable with a standard pK_a around 6.5. When the imidazole is deprotonated, its aromatic character predominates, and thus it can form stabilizing interactions with neighboring hydrophobic residues within the protein core. On

the other hand, a protonated histidine located in a buried position destabilizes the native structure due to the substantial energetic penalty involved in desolvating the charge. This net destabilization shifts the effective pK_a to lower values (*i.e.*, needing higher proton concentrations to become ionized). The larger the pK_a drop, the stronger the destabilization of the native state induced by histidine ionization, which can eventually drive protein unfolding once the destabilization is comparable to the intrinsic stability of the native state[117]. In this regard, as most downhill folding domains[17], gpW's native state has relatively low intrinsic stability, *i.e.*, about 14 kJ/mol[23]. Two important implications emerge from these considerations: (1) histidine grafts should be structurally conservative to avoid excessive destabilization of the gpW scaffold that could result in a protein that is unfolded over the entire pH range; (2) the grafts should still be sufficiently buried (and experience pK_a downshifts) to be able to trigger unfolding upon protonation at mildly acidic pH.

Single mutants design

We used structural analysis to select the locations on gpW for histidine grafting with these principles in mind. The recipient sites are residues that participate in one of the two gpW hydrophobic cores (Figure 3.1) and have enough space to accommodate the imidazole ring into the cavity without introducing significant steric clashes. We identified six such positions: A10, A13, L7, M18, F35, and V40. The six locations have varying degrees of solvent exposure, as seen in Figure 3.4. Hence, providing us with the flexibility to engineer different pH responses and explore how to maximize the transducer dynamic range.

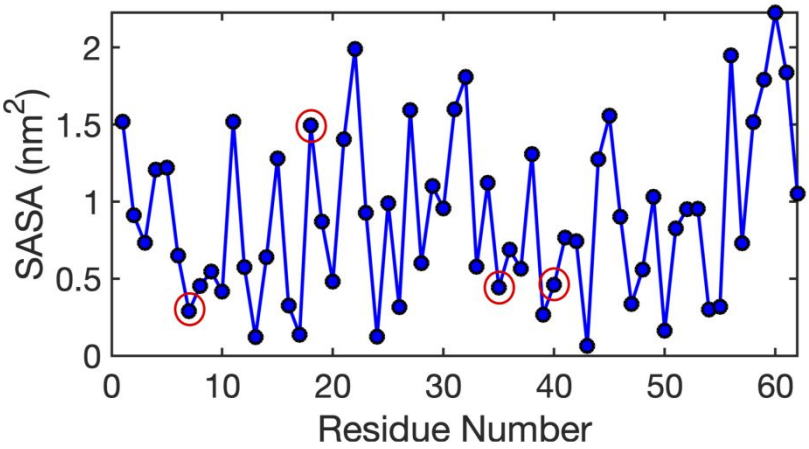

Figure 3.4 Solvent accessible surface area per residue of the gpW native structure calculated using UCSF Chimera tool with 0.14 nm solvent probe. Red circles indicate each of the four selected mutations.

We then designed the histidine mutations *in silico* and evaluated the effect on the stability of gpW using the DUET algorithm (Table 3.1). The calculations with DUET indicated that histidine substitutions into the L7, A10, and A13 positions could reduce the native stability

of gpW at room temperature by more than half, potentially placing these grafts at the brink of stability even at neutral pH.

Mutant	ΔΔG
L7H	-6.02
M18H	2.26
F35H	-0.17
V40H	-2.01
A10H	-6.65
A13H	-6.40

Table 3.1 Change in stability ($\Delta\Delta \boldsymbol{G}^{\boldsymbol{UN}}_{\boldsymbol{M-WT}} = \Delta \boldsymbol{G}^{\boldsymbol{UN}}_{\boldsymbol{M}} - \Delta \boldsymbol{G}^{\boldsymbol{UN}}_{\boldsymbol{wt}}$)in kJ/mol of gpW single histidine mutants as predicted by the DUET algorithm[111].

3.5.3 Molecular dynamics analysis of histidine graft stability

We then used atomistic MD simulations to investigate the intrinsic destabilization induced by the histidine grafts in deprotonated form. Particularly we simulated each of the six mutants and the wild-type for 400 ns. Figure 3.5 shows the time trajectories of the root-mean-square deviation (RMSD). Given the marginal stability and ultrafast folding of gpW, we expected these relatively short trajectories to display significant structural fluctuations and possibly even global unfolding. The control trajectory on wild-type gpW shows distinct structural transitions within the first 100 ns, followed by stabilization onto a relatively low RMSD ensemble. These fluctuations correspond mostly to the β-hairpin flapping in and out from its interaction with the two helices, which remain closely in contact throughout the simulation. The lower stability and enhanced structural dynamics of the gpW hairpin have been reported before from NMR analysis and long-timescale MD simulations[102]. Simulations of M18H, F35H, and V40H showed minimal structural fluctuations throughout the entire trajectory, with RMSD below 0.4 nm throughout (Figure 3.5). L7H displays larger structural fluctuations than the other three grafts and significantly higher mean RMSD (about 0.55 nm), consistent with the large destabilization predicted by DUET (Table 3.1) and its fully buried location gpW. The structural fluctuations of L7H are more

frequent, but they still are comparable in magnitude (≤0.65 nm) to those experienced by the wild-type.

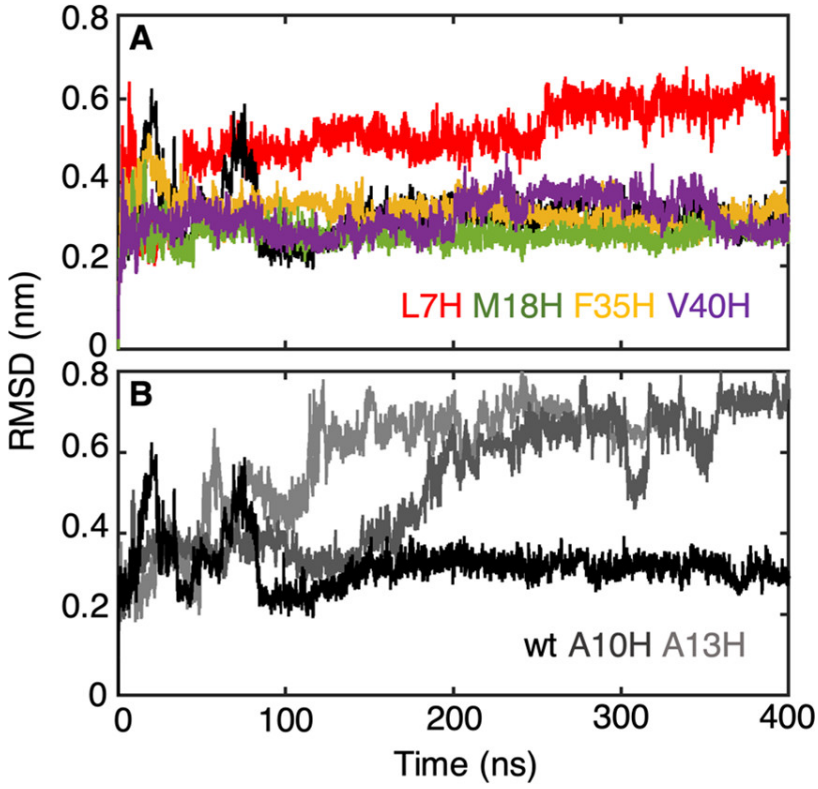

Figure 3.5 Time evolution of the root-mean-square deviation (RMSD) relative to the lowest energy conformer of the gpW NMR ensemble from MD simulations of gpW (black) and the six designed single (deprotonated) histidine substitutions. (A) Trajectories of the four mutants show structural fluctuations below the threshold (0.65 nm). (B) Trajectories of the mutants that exceeded the 0.65 nm RMSD threshold.

Moreover, as in the wild-type, the structural fluctuations of L7H are concentrated on the β-hairpin. These results suggest that L7H is probably a viable graft. In contrast, A10H and A13H show even larger structural fluctuations that get close to 0.8 nm and may not be fully equilibrated after 400 ns (Figure 3.5B). In addition, the conformational fluctuations of A10H and A13H involve the entire structure, which is again consistent with the large destabilization predicted by DUET (Table 3.1) and the more aggressive design of these mutations (introducing a bulky imidazole at a core location where there was only a methyl group). Therefore, we ruled out the A10H and A13H grafts for further study. We extended the simulation time up to 750 ns (Figure 3.6) for the other four grafts to ascertain whether the structural fluctuations (especially on L7H) would stabilize or continue evolving toward more unfolding. The longer trajectories showed small-scale reversible transitions with

signs of stabilization around their characteristic mean values. We decided to focus on L7H, M18H, F35H, and V40H as select grafts from these combined results.

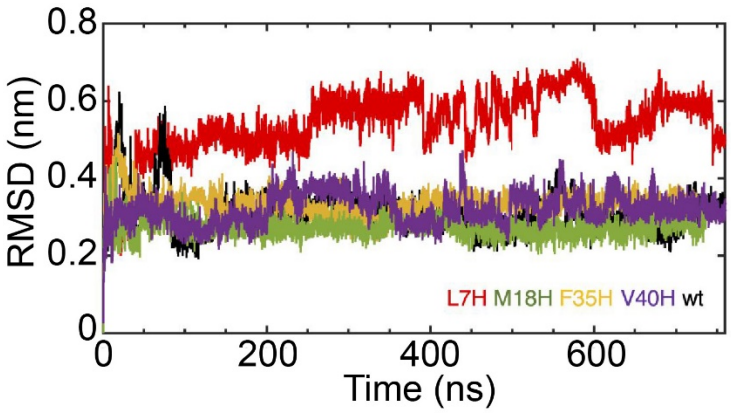

Figure 3.6 Time evolution of RMSD for the MD trajectories of the wild-type and four selected mutants.

3.5.4 Experimental analysis of conformational pH transducing

Modulating the transducer mechanism via single grafts

Figure 3.7 shows the unfolding of the four grafts as a function of pH and temperature measured by CD. From these data, we can determine the destabilization induced by the mutation on gpW (*i.e.*, from data at pH 7) and the pH response of each graft. All mutations have a destabilizing effect on the native ensemble relative to the wild-type. The changes in free energy upon mutation are given in Table 3.2. As for the pH response, visual inspection reveals the loss of α-helical structure as pH becomes mildly acidic. For most of them, the loss in structure is apparent even at pH 6. Overall, all grafts show enhanced disordering relative to the wild-type as pH drops from 8 to 4.

Protein	ΔG
Wild-type	13.5
L7H	4.5
M18H	8.1

F35H	5.9
V40H	5.5
M18H-F35H	0.75
F35H-V40H	-1.8

Table 3.2 Experimental unfolding free energy of the wild-type gpW, single and double histidine grafts at pH 7 in kJ/mol (using the native state as reference).

On the other hand, the overall magnitude and response range of the pH effect varies widely between grafts, as we anticipated in our design strategy. The pH response ranges from being minimally stronger than the wild-type's for M18H to the rather dramatic response of L7H, which at pH 5 and room temperature already exhibits less native signal than the wild-type at pH 2. F35H and V40H have intermediate responses. Therefore, the experiments confirm the coupling between graft protonation and gpW unfolding.

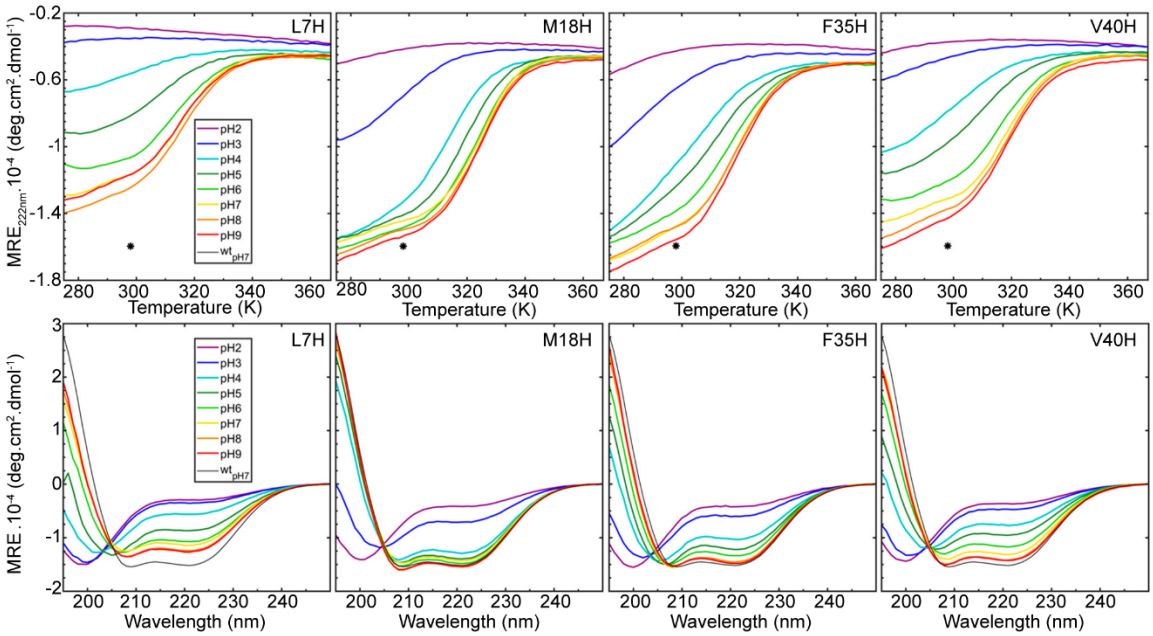

Figure 3.7 Temperature and pH unfolding of the four gpW histidine grafts. Each curve or spectrum is an average of 2 experiments. The curves and spectra are colored following the visible light spectrum, with the lowest pH corresponding to the highest energy (violet). The spectra at various pH values shown on the bottom row

correspond to 298 K. The black asterisk on the top row indicates the unfolding curve of wild-type gpW at room temperature, pH 7. The spectrum shown as a thinner black line on bottom row panels corresponds to the wild-type at pH 7.

In general, the pretransition baseline for temperature denaturation drops as the pH lowers (*e.g.*, F35H in Figure 3.7 top), with the possible exception of M18H, which is the graft with the weakest pH transducing properties. It is of note that in native-like conditions (neutral pH and room temperature), L7H displays a 25% decrease in α-helical CD signal relative to the wild-type, consistently with the strongest pH response for this graft. pH also changes the degree of the helical signal of the thermally unfolded state (*e.g.*, pH 2 unfolding curves in Figure 3.7 top), which becomes increasingly disordered (less negative CD at 222 nm) proportionally to the strength of the graft's pH response.

Notably, the gradual unfolding of gpW coupled to histidine ionization results in transducers with varying responses depending on the graft's location, which exhibit broadband sensitivity relative to a conformational switch mechanism. This point was our working hypothesis, and the results from the four histidine grafts confirm it. There are two parameters to evaluate the performance of these conformational pH transducers: (1) the maximum signal difference, which defines the transducer's sensitivity; (2) the dynamic range or response bandwidth, which defines the pH range of the transducer sensitivity. Here we use the CD signal at 222 nm at room temperature to indicate the transducer response. The data for the four grafts is given in Figure 3.8. This figure demonstrates that all single grafts are conformational pH transducers with sensitivity at pH > 4, whereas the wild-type is insensitive in that range (see Figure 3.2). Therefore, the histidine grafting approach works as a general strategy to engineer conformational pH transducers into proteins. Figure 3.8 also highlights the broadband behavior of these transducers. As it was hinted in Figure 3.7 (bottom), the mutants feature significant differences in both the native (*i.e.*, high pH) and unfolded (at pH 2) CD signals. This behavior is due to the gradual (dis)ordering of downhill folding proteins in contrast to the interconversion between well-defined end-states of conformational switches. Interestingly, because the changes in the native and unfolded baselines go in the same direction (less negative ellipticity), the maximum signal change of the transducer is roughly invariant (about 11 000–12 000 deg.cm^2.dmol^{-1}). In parallel, the dynamic range varies systematically as a function of the graft. M18H has a dynamic range of about 3 pH units that is minimally extended over that of a switch transducer. Moreover, its sensitivity only spans the 2–5 pH range, which indicates that the perturbation from H18 ionization is not capable by itself of inducing the unfolding of gpW. M18H is also the mutant with the highest stability in nonionized form (Table 3.1), which means it has a larger repository of folding free energy to compensate for the change in free energy upon histidine ionization. The next step up in the dynamic range is F35H, which extends the response up to pH 6.5 and has maximal sensitivity around pH 4. V40H starts to show a slight decrease in native signal at neutral pH and exhibits nearly linear CD signal changes over an impressively extended range of [H^+] (from 10 nM to 10 mM, *i.e.*, pH 8–2). In other words, V40H is a distinct example of an ultra-broadband pH transducer that could be used as a foundation for implementing high-performance pH

sensors. L7H also has a broadband pH response, but not as much because this mutant becomes fully unfolded at pH 3 and exhibits a fairly substantial loss in helical signal at neutral pH (see Figure 3.7 bottom), which penalizes its maximal signal difference (~9000 deg.cm^2.dmol^{-1}). The pH response of these mutants, particularly of L7H, highlights that the interplay between the intrinsic stability of the protein and the free energy perturbation upon ionization is key for optimizing the dynamic range and maximum signal of rheostatic conformational transducers.

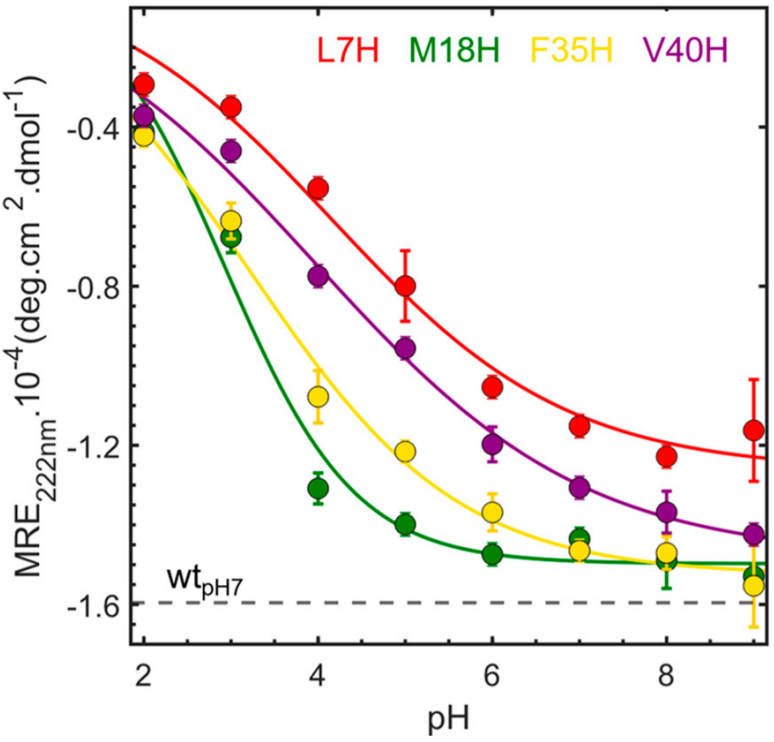

Figure 3.8 CD signal at 222 nm as a function of pH of all the single mutants at 298 K (circles) and their corresponding colored lines to guide the eye. The dashed line (- -) represents the CD signal of wild-type gpW in its folded state (pH 7) as reference.

Reversibility of pH transducing

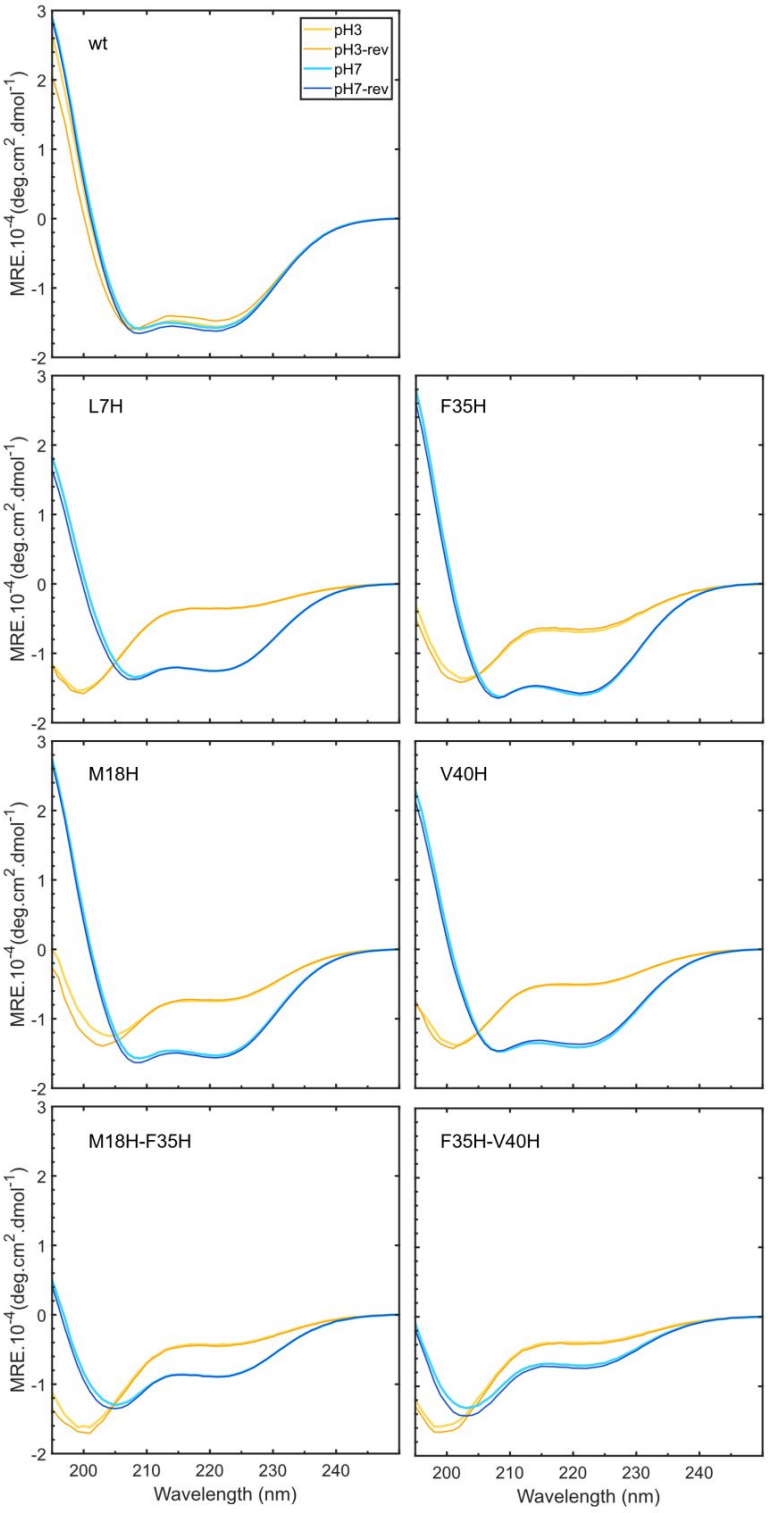
wt
pH3
pH3-rev
pH7
pH7-rev
L7H
F35H
M18H
V40H
M18H-F35H
F35H-V40H
MRE.10^{-4}(deg.cm^2.dmol^{-1})
Wavelength (nm)
200
210
220
230
240
3
2
1
0
-1
-2

Figure 3.9 Reversibility of thermal denaturation experiments on gpW and mutants. Each panel shows the far-UV CD spectrum of one gpW variant (wt, single and double grafts) measured at 293 K before (light shade) and after (dark shade) the temperature denaturation experiment. Blue shades are for experiments at pH 7, and orange shades for experiments at pH 3.

An effective pH transducer should respond to pH changes in both directions and in a fully reversible fashion. Routine control experiments in which we recorded the CD spectrum at room temperature of each graft before and after thermal denaturation (experiments on Figure 3.7 top) showed that the reversibility upon temperature denaturation was better than 95% for all grafts at all pH values tested (Figure 3.9). The reversibility after temperature denaturation is a very stringent test of the reversibility of protein unfolding transitions because high temperature enhances the aggregation of partially unfolded polypeptides[118].

As an additional control, we performed a simple test of reversibility of the pH transducing mechanism upon a drastic change in pH occurring at 293 K. The results from tests are shown in Figure 3.10 for the graft with the broadest dynamic range (V40H). The pH transducing mechanism exhibited excellent reversibility when pH was changed from 3 to 7 and reverted back to 3 or changed from 7 to 3 and reverted back to 7. These experiments demonstrate that the histidine grafts on gpW operate as reversible pH transducers.

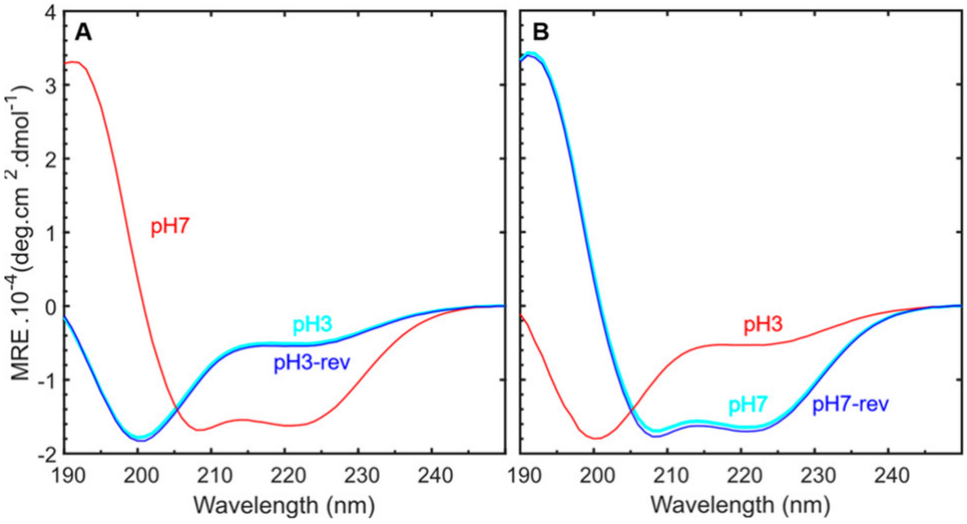

Figure 3.10 Reversible pH transducers. Far-UV CD spectrum of histidine graft V40H upon extreme changes in pH, both in neutralizing (A) and acidifying (B) directions at 293 K. The reversibility is shown on the V40H as an example. Cyan signals the initial condition, red the changed condition, and dark blue the reversion to the initial condition.

Modulating the transducer mechanism via multiple grafts

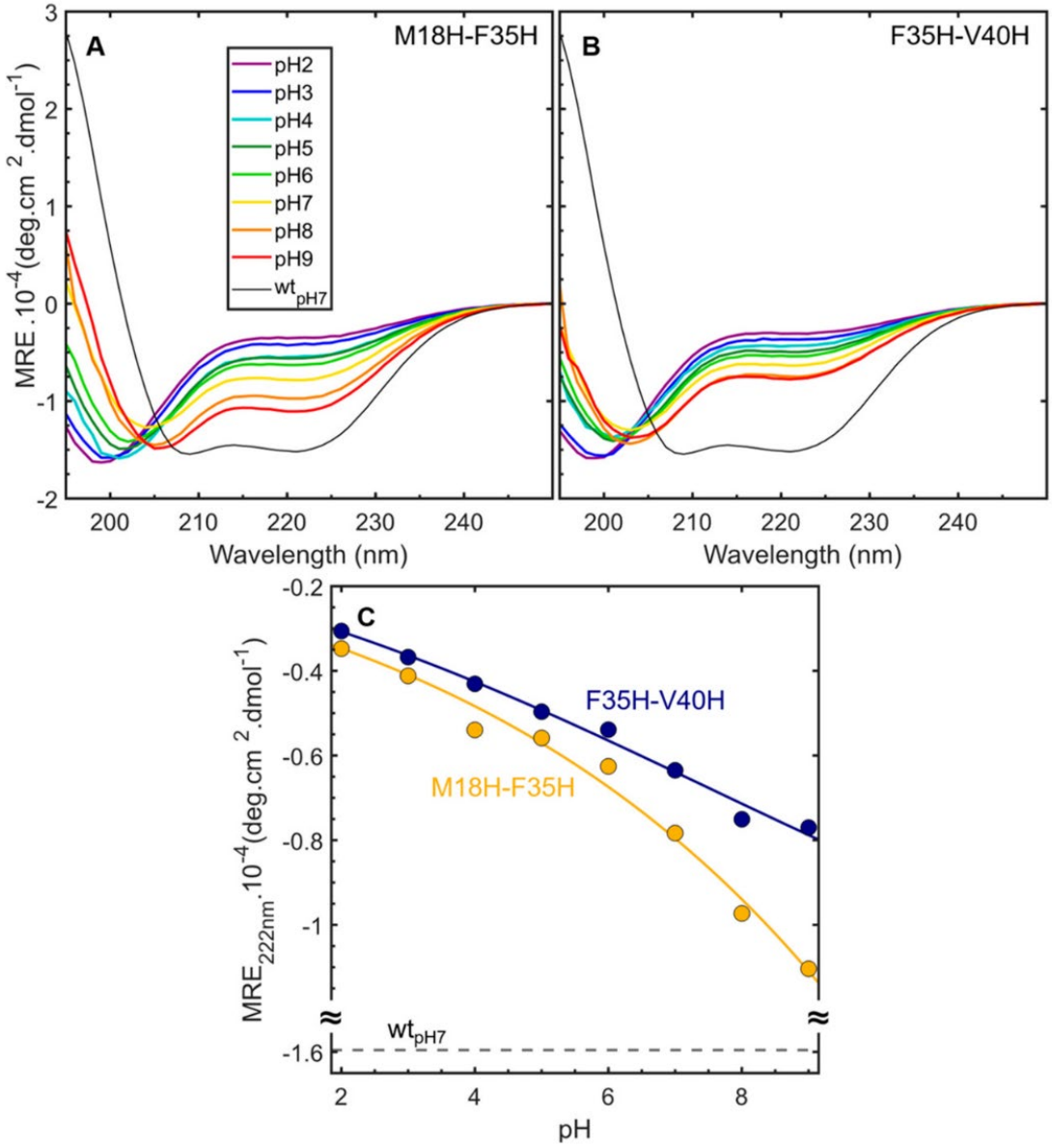

Figure 3.11 pH transducing on double histidine grafts. Far-UV CD spectra of the double histidine grafts M18H–F35H (A) and M18H–V40H (B) at 298 K as a function of pH. (C) CD signal at 222 nm as a function of pH for the double grafts (circles). The corresponding colored lines are to guide the eye. The dashed line (--) represents the CD signal at 222 nm of wild-type gpW at pH 7.

Our results indicate that the response of the pH transducer depends on the structural environment of the histidine graft: more buried positions lead to a broader dynamic range. The question remains whether the effects of more than one graft are additive or exhibit

positive or negative cooperativity. We decided to explore this issue by producing double grafts. However, we had to be careful as the single gpW grafts are already marginally stable (Table 3.2). We ruled out L7H since it is already at the brink of native stability and targeted M18H–F35H as a conservative double graft and F35H–V40H as a more aggressive one (based on Figure 3.8 and Table 3.2).

The experimental analysis of the double grafts highlights a slightly different pH response, which provides further insight into the coupling between downhill (un)folding and proton binding. In particular, the maximum α-helical signal (at the highest pH) is reduced, and it only reaches 70% of the wild-type native signal for M18H–F35H (Figure 3.11A) and 50% for F35H–V40H (Figure 3.11B). As a result, the CD signal at 222 nm has a convex dependence with pH (Figure 3.11C) rather than the sigmoidal-concave dependence of the single grafts. The reason behind these differences is that the double mutants are already partially unfolded at room temperature, even when the histidine residues are deionized. Therefore, histidine ionization can only tilt the already partially unfolded ensemble toward more disorder, resulting in the convex pH dependence shown in Figure 3.11C (half of a broad negative sigmoidal curve). The partially disordered native-like ensembles of the double grafts are most evident when compared with the CD spectrum of the wild-type gpW at neutral pH (Figure 3.11A, B). This result is consistent with the quantitative analysis of their thermal denaturation at neutral pH, which indicates that at room temperature, both of these proteins are close to or just past the denaturation midpoint (Table 3.2).

Regarding their properties as conformational pH transducers, the double mutants feature a reduction in maximum CD signal that is concomitant to their marginal intrinsic stability (7500 and 5000 $deg.cm^2.dmol^{-1}$, respectively). On the other hand, their response is fairly linear over the entire pH range (2–9), indicating that the double grafts in conjunction with a marginally stable nonionized downhill folding scaffold result in lower sensitivity and on ultra-broadband pH transducers.

3.6 Conclusions

The pH transducers we have engineered into gpW, whether single or double grafts, exhibit conformational changes that extend well beyond 2.5 orders of magnitude. Interestingly, the dynamic range in pH sensitivity increases proportionally to the magnitude of the perturbation produced by ionization (Figure 3.8). The dynamic range eventually extends over seven orders of magnitude in $[H^+]$, resulting in ultra-broadband pH transducers. Such transducers would offer impressive improvements over the dynamic range of existing pH sensors[119].

We have explored a strategy for engineering conformational transducers for biosensor applications based on thermodynamically coupling the (un)folding process of a downhill folding protein domain to binding an analyte of interest. The underlying hypothesis was that such coupling might give rise to rheostatic (analog) rather than switching (binary) conformational transducers[17]. As a protein scaffold, we chose the fast, downhill folder

gpW, pH as an analyte, and histidine grafting as an approach to engineer the conformational transducer.

Our results shed light on the interplay between downhill (un)folding and binding and how it can give rise to conformational transducers with unique properties. We can extract several practical lessons. First, the coupling between histidine ionization and downhill (un)folding converts the destabilization directly into structural changes. This feature is inherent to downhill (un)folding domains, which have flexible native ensembles, and gradual (continuous) unfolding transitions. Therefore, even relatively minor free energy perturbations result in structural changes that can be detected. In other words, the structural changes are not limited to the interconversion between a native and an unfolded state but also involve the gradual (dis)ordering of these ensembles.

Technical challenges notwithstanding, we can conclude that rheostatic conformational transducers add a new, exciting tool to the biosensor engineering toolbox. The sharp response of switching transducers will be preferable for applications where the range in ligand concentration is narrow and minor changes in ligand levels must be detected. An example of ultrasharp response is physiological temperature sensors, which need to detect changes of even a fraction of a degree. On the other hand, rheostatic transducers could provide improved performance to monitor signals that vary widely. For instance, pH changes between 8 and 4 inside living cells statically and dynamically, depending on the cellular compartment[120] and/or the cell's metabolic status. Currently, there are intracellular pH sensors (fluorophore-based and fluorescent protein-based) for either the neutral or the acidic (lysosomal) ranges[121]. However, there are no pH sensors available that can simultaneously operate in all intracellular locations and/or all metabolic stages, even though this capability is widely recognized as essential to understand the role of pH homeostasis in cell biology and physiology[122]. Another example of a broadband biological signal is the second messenger Ca^{2+}, which is found at only 50 nM in the cytosol, 500 μM in the ER, and at mM levels extracellularly, which changes drastically and quickly (sub-ms) in response to variate inputs[123]. Here again, there are chemical and recombinant (*e.g.*, calmodulin-based) Ca^{2+} sensors available[124-126], but none of them get close to covering the vast physiological [Ca^{2+}] range nor the fast, local concentration spikes arising from signaling pulses, such as those associated with cell contractility[127]. Conformational rheostat transducers would be, in this respect, tremendously useful to enable the broadband sensing required to effectively track widely varying biological signals/properties, such as Ca^{2+} and pH.

CHAPTER 4

4 Fluorescence-based strategies to implement an optical signal reporting the conformational rheostat behavior of pH biosensor.

4.1 Introduction

In the previous chapter, we reported an approach to build a pH conformational transducer that is sensitive to pH changes across >6 orders of magnitude (pH 3 to 9) by engineering histidine grafts onto a downhill folding protein. Circular dichroism (CD) spectroscopy was utilized to monitor the helical contents change of the proteins as a function of pH, taking advantage of the high helical content that gpW exhibits when is folded. Thus, our results illustrated that CR-based transducers are not only a result of natural selection but can also be engineered for interesting applications.

Although we successfully demonstrated that such pH transducer could produce a broadband response, the CD technique is not suitable to build an optical signal reporter for a complex environment such as in a living cell or blood sample because this technique would work well only in a sample that is completely transparent that does not have any other proteins or nucleic acids present (they all have CD signal), and for ligands that are also invisible to CD. Therefore, in this chapter, we aimed to introduce fluorescence signals that can be effectively coupled to the CR changes as a function of pH that we induced in gpW. The goal is to implement a signal reporter based on fluorescence that eventually circumvents these limitations and can be effectively used as a biosensor. Indeed, fluorescence-based biosensors are the most widely used optical sensing due to their high sensitivity and great potential in biomolecule imaging to study the organization and functions of living systems[128, 129]. These fluorescence-based biosensors will be of great interest to eventually be able to detect analytes at the single-molecule level in real-time.

In this regard, the working strategies to construct the fluorescence outputs onto this specific gpW pH transducer are being discussed based on fundamental fluorescence principles. Indeed, numerous fluorescence emission phenomena and suitable fluorophores can be explored to optimize the output[130]. Because we are interested in converting conformational changes in fluorescence signals, we targeted two phenomena that are proportional to the distance between the fluorophores/molecules: Förster resonance energy transfer (FRET)

and photo-induced electron transfer (PET). Methods based on these two phenomena are frequently used to study the conformational dynamics of biomolecules and their complexes. However, their general properties and dependence on the distance are entirely different, with FRET being sensitive to relatively long distances (3-10 nm) and PET being sensitive to very short distances (<1 nm). These two methods are hence complementary.

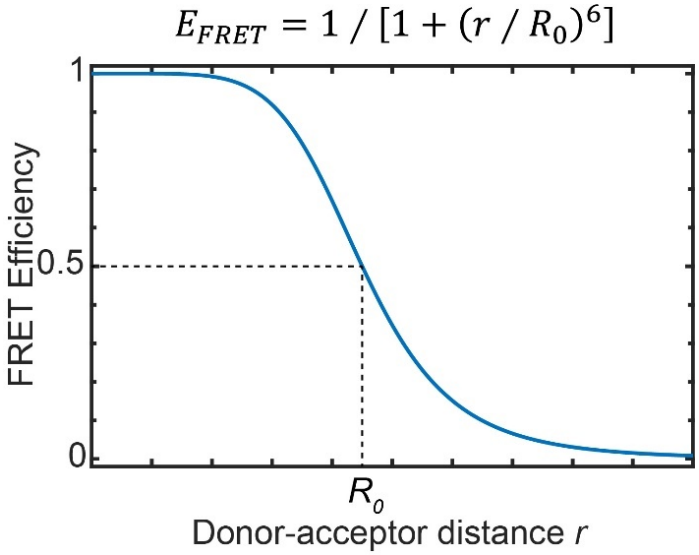

Figure 4.1 FRET acts as a spectroscopic ruler. FRET effect depends on the characteristic R_0 and inversely proportional to the six power of the distance between the donor and acceptor ($1/r^6$)

FRET-based biosensors rely on a pair of fluorophores being covalently attached to the protein. The donor's fluorophore emission overlaps with the excitation of the acceptor fluorophore, causing the deactivation of the donor and excitation of the acceptor (and hence a change in color) through Förster Resonance Energy Transfer, which as a dipole-dipole interaction is, to a first-order approximation, inversely proportional to the six power of the distance between the donor and acceptor ($1/r^6$). Thus, the FRET pair acts as a spectroscopic ruler that can measure the distance variations between two chosen label-attachment points on the protein as the protein's conformation changes upon binding to the analyte. FRET is a prevailing tool since it is a ratiometric (two-wavelength) measurement; therefore, it is independent of protein concentration (in contrast to fluorescence intensity measurements). The efficiency of the FRET effect depends on the characteristic R_0 (Förster distance at which the efficiency of transfer is 50%) that is unique for each donor-acceptor dye pairs (Figure 4.1). The signal is known to be roughly linear in a distance region within 0.5 to 1.5 times R_0. For common Alexa FRET pairs currently commercialized, which have an R_0 of ~5 nm, the linear distance range falls into 2.5 to 7.5 nm.

Despite all these advantages, FRET could pose limitations when the functional length scale does not match its distinctive sensitivity range. Consequently, PET has been used to fill in the gap. The PET phenomenon relies on the quenching of a fluorophore when in molecular

contact with an effective electron transfer acceptor. For biomolecules, one can use the fluorophore Atto655 and the aminoacid tryptophan or the nitrogenous base guanine as electron transfer donors[131]. PET requires both the fluorophore-quencher pair to be at van der Waals contact[132], which means the technique is sensitive to very short distances (below 1 nm) [131]. Also, PET is exponentially dependent on the distance between the fluorophore and quencher. Therefore, PET is a promising tool to monitor conformational changes that do not involve large differences in the end-to-end distance (Figure 4.2). However, PET reports the fluorescence intensity, which linearly depends on the sensor concentration and is not ratiometric like FRET.

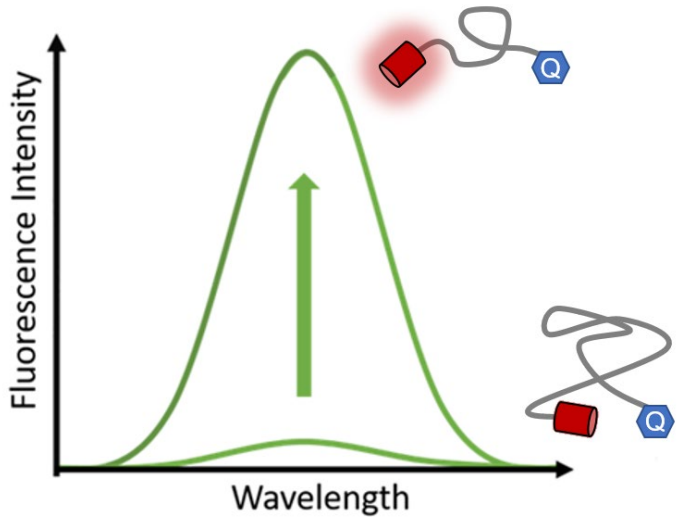

Figure 4.2 PET is a tool to monitor conformational changes. PET is exponentially dependent on the distance between the fluorophore and quencher

This chapter describes how both FRET and PET readouts were implemented on gpW pH transducers (engineered mutants L7H and M18HF35H) and characterized for their ability to convert the conformational change into an optical signal. In the FRET constructs, two cysteines were incorporated in the protein sequence to label specifically with Alexa 488 (Donor) and Alexa 594 (Acceptor). In the constructs that were designed for PET sensing, a single cysteine was incorporated in the sequence to be labeled with Atto 655, and a tryptophan residue was incorporated at a suitable location within the scaffold to enable PET. Our results showed that FRET readouts are insensitive to the local changes of the rheostatic pH transducer, whereas the PET fluorescence implementation could generate a high signal to noise (factor of 2) fluorescence readout that operates over a broad dynamic range in $[H^+]$ (pH 3 to pH 9).

4.2 Materials/Methods

Recombinant protein expression and purification

All gpW variants, including three gpW-L7H and the M18H-F35H variants that targeted different fluorescent probes, were cloned as entire genes in the bacterial expression vector pBAT4. The M18H-F35H variant was cloned as SUMO-His-Tag fusions to improve protein expression and purification of unstable protein. All the gpW variants used in the current study were expressed and purified as described in Chapter 2. In the case of the double mutants, which were expressed as fusion protein along with sumo tag, the tag was removed following the similar protocol described in Chapter 2. The proteins were lyophilized and stored at 253 K.

Fluorescent labeling gpW variants with Alexa 488, Alexa 594, and ATTO 655

ATTO 655 maleimide was used to label the PET constructs, and a pair of Alexa 488 maleimide and Alexa 594 maleimide was used for FRET constructs. All proteins were labeled by attaching the fluorescent dye to the inserted cysteine residues. Proteins were dissolved in Tris buffer at pH 6.5, 150 mM NaCl, and reducing agent, TCEP (tris(2-carboxyethyl)phosphine that is ten times higher than the protein concentration. After incubation for 30 minutes, the TCEP was removed by centrifugal filter (3 kDa cutoff). ATTO 655 was added to the protein solution with a 5 to 1 ratio for the PET constructs. The Alexa 488 and Alexa 594 were added to the protein with the following ratio 1.2:0.8:1 (Alexa 488: Alexa 594: protein) for the FRET constructs. The labeling was performed by incubating for 2 hours at room temperature or 277 K overnight with constant stirring. Then 10X BME (β-mercaptoethanol) was added to the protein mixtures. The labeled proteins were purified using RP-HPLC with a 0–95% Acetonitrile gradient with 0.1% trifluoroacetic acid (TFA) for elution. All fractions containing labeled proteins were pooled and aliquot into small vials and kept at 193 K. The purity and identity of the labeled proteins were verified by electrospray ionization mass spectrometry.

Fluorescence spectroscopy

Proteins samples were prepared at different pH by adding the labeled proteins 20 mM citrate buffer for experiments in the 3–6 pH range, phosphate buffer pH 7, and Tris-HCl buffer for the 8–9 pH range. A final concentration of 0.05 % v/v Tween 20 was added to all the buffer for the PET constructs. Fluorescence emission spectra were recorded between 630 nm and 740 nm by exciting the ATTO 655 at 620 nm or between 490 nm and 720 nm by exciting the Alexa 488 at 480 nm. The pH titration on the L7H-helix PET construct was performed three times, and pH titration of L7H-Cter was performed five times to calculate the standard deviation due to the concentration dependence of the PET fluorescence technique.

4.3 Results

4.3.1 Strategy for engineering FRET-based signal readouts for the pH sensing conformational rheostat transducer

Our previous work showed that four single histidine grafts (L7H, M18H, F35H, V40H) and two double histidine grafts (M18H-F35H and F35H-V40H) on gpW exhibit pH-dependent structural changes that span across >6 orders of magnitude in [H^+] and can be characterized using CD spectroscopy. L7H showed the most extensive effect with linear CD signal changes in the range of 10 nM to 100 mM [H^+]. Therefore, we decided to use this variant as the basis for the design of the fluorescence-based signal output. Specifically, an extra cysteine was added at the N-terminal (N-term), e.g., position 1, and the second cysteine was introduced near the C-terminal (C-term) by replacing the methionine at position 56. These mutations allow us to chemically label the L7H scaffold with a FRET-pair to monitor the end-to-end distance (L7H-FRET) (Figure 4.3 left). Another FRET-construct was also designed to probe the distance between the β-hairpin and the helix-1. This second design was based on our MD simulation data, which showed that the β-hairpin experiences the largest structural fluctuations induced by acidification[102]. In this case, we used the double mutant M18H-F35H as a scaffold because the F35H mutation directly impacts the packing of the hairpin onto the two helices of gpW. We added one cysteine at the N-term (position 1) of M18-F35H, and replaced an aspartate in the beta-hairpin turn (position 29) with cysteine (Figure 4.3 right).

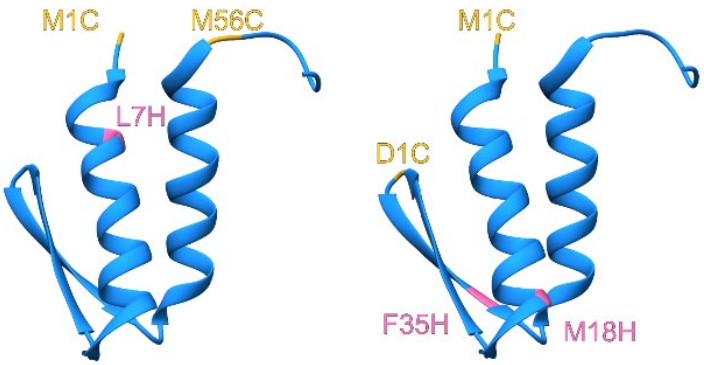

Figure 4.3 Designs of FRET constructs for gpW variants. L7H-FRET construct (left) and M18H-F35H-FRET construct (right). Locations of the Cysteine were shown in yellow.

The choice of this variant was based on the following reasons: (1) F35 is located on the beta-hairpin that forms the hydrophobic pocket with helix-2, and its mutation to histidine plus protonation results in the largest structural fluctuations of the beta-hairpin according to MD simulations; (2) F18-F35H also exhibits broadband pH-sensitivity that extends

across the pH 9-2 range; (3) the F18-F35H is partially unfolded at neutral pH and hence highly primed to unfold in the neutral pH range. Therefore, we expected that M18-F35H, combined with the strategy to probe the N-term to position 29 distance, would produce monitorable changes in end-to-end distance in the neutral to basic pH range and in the more acidic range. In both designs, Alexa-488 and Alexa-594 were used as the FRET pair of choice because they have many advantageous features for FRET measurements: high brightness, high photostability, quantum yield close to 1, and R_0 close to 5 nm. Moreover, these fluorophores are pH-insensitive at the relevant pH range (down to 3). Both fluorophores were labeled onto the cysteines using maleimide chemistry.

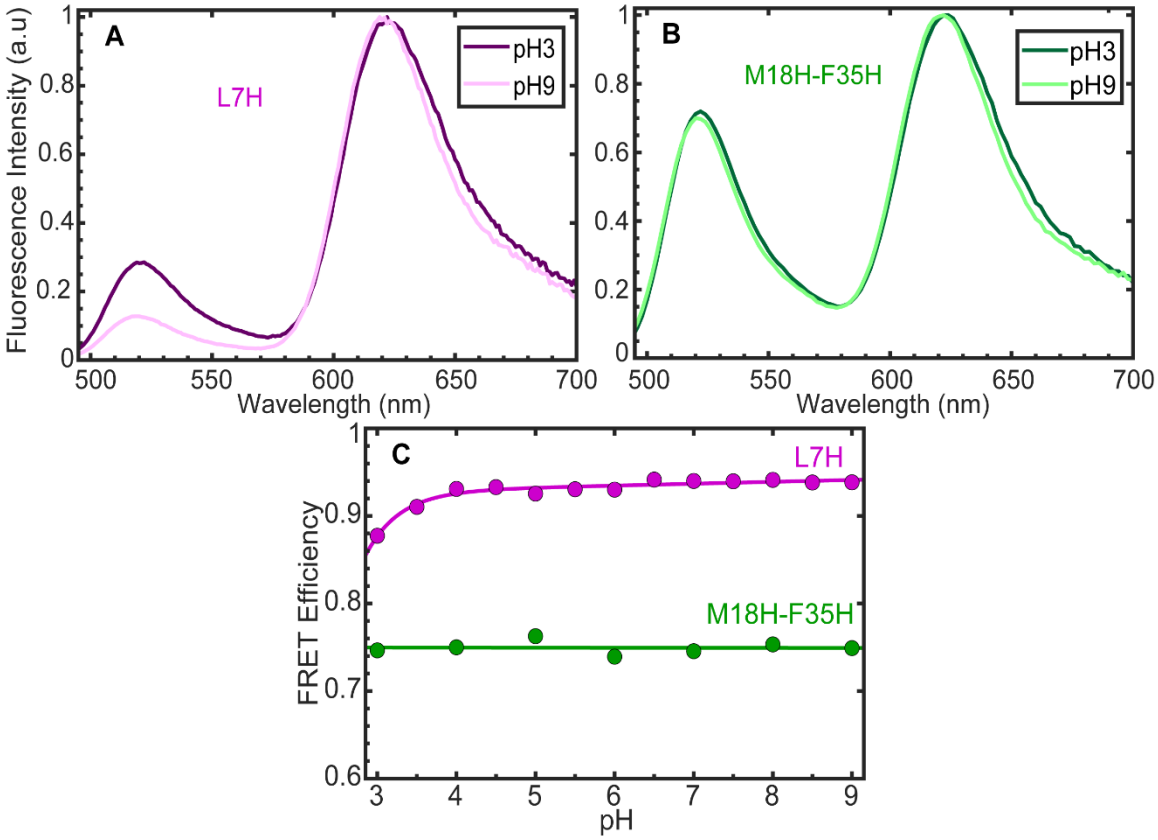

Figure 4.4 Fluorescence characterization of different FRET constructs at different pH conditions. Normalized fluorescence emission spectra of L7H (A) and M18H-F35H (B) at acidic (pH 3) and basic (pH 9) conditions. The spectra were recorded at 293 K and using excitation at 480 nm. FRET efficiency of L7H (magenta) and M18H-F35H (green) as a function of pH (C) with a polynomial fit curve shown to guide the eye.

Figure 4.4 shows the changes in fluorescence of donor and acceptor for the two FRET constructs at the highest and lowest pH values (Figure 4.4.A for L7H and 4.4.B for M18H-F35H), as well as the FRET efficiency as a function of pH (Figure 4.4.C). The FRET efficiency is inversely proportional to the sixth power of the distance, and hence it should

be highest at basic pH and lowest at acidic pH (Figure 4.1). In the L7H-FRET construct, we observed a very high FRET efficiency at pH 9 (above 0.9) and a lower FRET at pH 3, as judged by the intensity ratio of the donor and acceptor bands. These results indicate that FRET is sensitive to the conformational changes that take place on L7H upon histidine protonation, as detected by CD (Figure 3.8). However, the FRET efficiency as a function of pH is essentially constant from pH 9 to 4. Hence, L7H is only sensitive in the pH range (acidic) in which the protein is most unfolded. Given the high FRET in this region, a possible explanation for the lack of sensitivity throughout the pH 9 to 4 range is that the end-to-end distance of the protein stays within the flat region of the FRET efficiency dependence (e.g., distances <3 nm) (Figure 4.1). To test this hypothesis and confirm that the FRET readout is indeed responsive to larger changes in end-to-end distance, we performed an equilibrium chemical denaturation experiment on L7H using GdmCl (guanidinium chloride) as denaturant[133] and FRET efficiency as the probe. This experiment showed that upon chemical denaturation, which results in a highly expanded denatured state[134, 135], L7H experience a large decrease in FRET efficiency that goes from over 0.9 to about 0.55 (Figure 4.5). This control experiment demonstrates that FRET is responsive to changes in the unfolding of L7H but that these changes need to be large enough to bring the FRET curve to the linear range. However, from a practical standpoint, our results demonstrate that the FRET readout on L7H is inadequate for pH sensing in general and for broadband sensing in particular.

We then focused on the alternative M18H-F35H-FRET scaffold. In addition to the reasons outlined above, we also expected this variant to result in lower FRET efficiency when in the folded state because the end-to-end distance is, in this case, longer (Figure 4.3). This would be advantageous because it would bring the FRET changes within the linear response region of the FRET efficiency curve. The fluorescence spectrum of the donor and acceptor at pH 9 (when the folded structure is most stable) on M18H-F35H does show a significantly lower acceptor/donor ratio, which indicates lower FRET (Figure 4.4 B). However, the spectrum at pH 3 looked essentially identical (Figure 4.4 B), and the determination of the FRET efficiency as a function of pH is completely flat across the entire pH range (Figure 4.4 C). This result implies that, whereas CD and simulations show large structural fluctuations of the β-hairpin when M18H-F35H is protonated, the averaged distance between the hairpin turn and the N-terminus does not change. This scenario is very similar to what happens with the acid denaturation of the one-state downhill folder BBL, which is also caused by the protonation of two buried histidines (in this case, naturally occurring) [136].

The implication is that FRET readouts appear to be highly insensitive to the conformational changes induced on the gpW scaffold by pH. In other words, these large conformational changes result in very small differences in compaction of the protein, hence in FRET efficiency for fluorophores in the visible range.

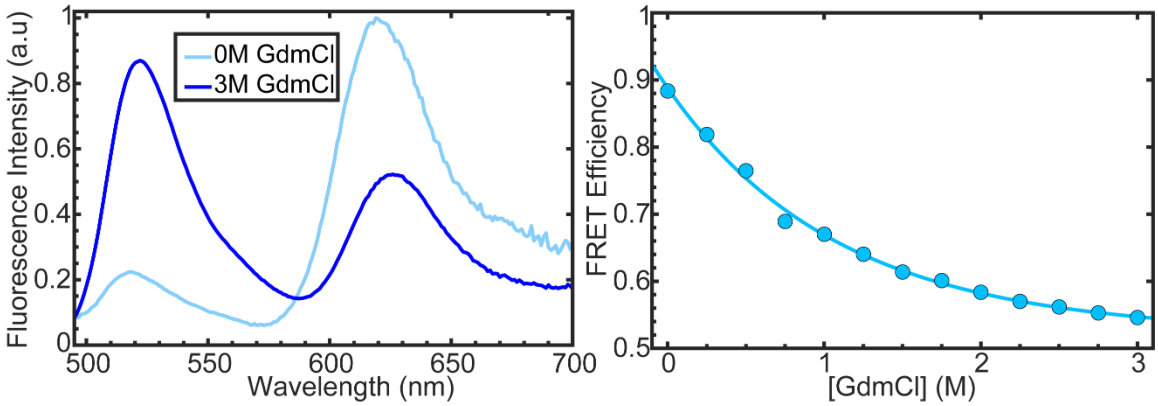

Figure 4.5 Equilibrium chemical unfolding experiment of L7H-FRET using GdmCl as a denaturant. The experiments were performed at pH7 and 293K. Normalized fluorescence emission spectra with excitation at 480 nm (left). FRET efficiency as a function of GdmCl with a polynomial fit curve shown to guide the eye (right).

4.3.2 PET readout strategy for converting the pH conformational rheostat transducer into a broadband pH biosensor

As the end-to-end distance changes of the L7H variant of gpW as a function of pH appeared to be less than 1 nm, the PET technique would, in principle, be more suitable to detect these local changes. Therefore, we designed two L7H constructs to implement the PET signal readouts into L7H. As fluorophore, we used ATTO 655, which has a high quantum yield and is well known to undergo high-efficiency PET using tryptophan or guanine as electron acceptor[137].

To implement the PET readout, we designed two constructs in which ATTO 655 was added to an engineered cysteine, and a tryptophan was added by protein engineering at a suitable position so that the electron donor and acceptor are within 1 nm in the folded structure. In the first one, we replaced the arginine residue at position 3 with a tryptophan and replaced the glutamine at position 53 with cysteine to fluorescently label the protein with ATTO655 (Figure 4.6 left). With this construct, the pair of quencher-fluorophore were placed right at the first residue of the helix-1 and the endpoint of helix-2. Such construct (L7H-helix) aimed to precisely monitor the conformational changes caused by the unwinding of the helix induced by pH. The second construct (L7H-Cterm) was engineered to add the tryptophan at position 1 and replacing methionine 56 with cysteine for ATTO 655 labeling (Figure 4.6 right). The rationale here was that by placing the fluorophores in the flexible tails, the conformational changes of the protein could be more effectively transduced into a change in average distance in the required range (unfolded with distance > 1nm).

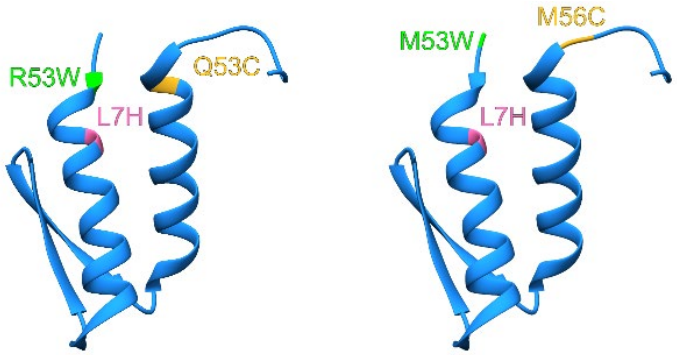

Figure 4.6 Design of PET constructs for gpW variants. L7H-helix construct (left), and L7H-Cterm construct (right). Locations of the Cysteine were shown in yellow and the Tryptophan in green.

Figure 4.7 shows the results we obtained for the L7H-helix construct. In this case, we observed that L7H-helix showed high fluorescence intensity at pH 3. The intensity then dropped abruptly at pH 4 and remained flat up to pH 9. The fluorescence intensity decreased by 2-fold between pH3 and pH 9. Our data revealed that the distance between the tryptophan and the dye did not significantly change until the protein was nearly fully protonated, at which point it changed drastically. This behavior could be useful for a high contrast/high-resolution monitoring of pH in the 3 to 4 range, but it is not suitable for broadband sensing, which was our purpose.

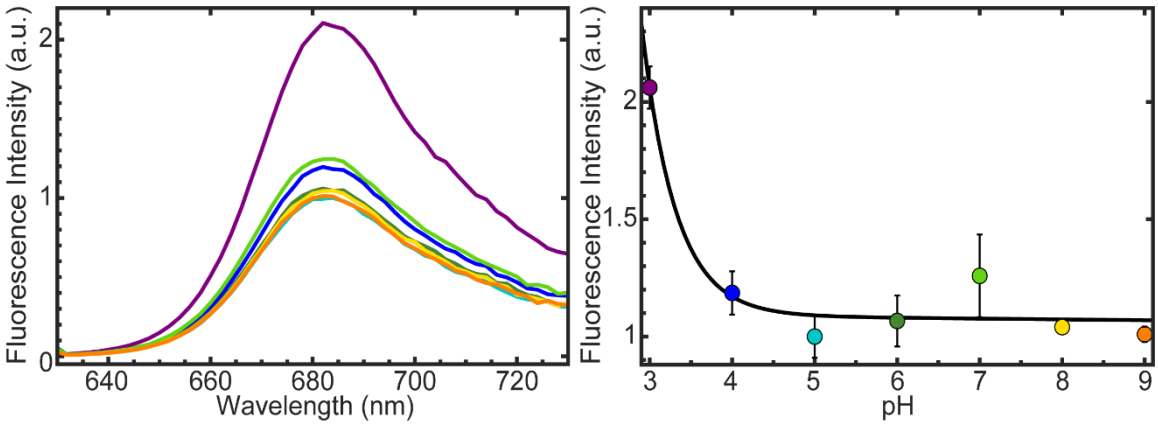

Figure 4.7 Fluorescence intensity change of L7H implemented with a PET readout at the end of the two helices. Normalized fluorescence emission spectra at various pH and 293 K with excitation at 620 nm (left). The spectra are colored following the visible light spectrum with the lowest pH corresponding to the highest energy (violet). Total fluorescence intensity was calculated from the area under each curve and plotted as function pH, and the data points are colored according to spectra

on the left (Right). The curve through the data points is a polynomial fit to guide the eye.

In contrast, the L7H-Cterm PET readout exhibited a gradual reduction in fluorescence intensity from its high value at pH 9 down to pH3 (Figure 4.8). The maximum change in fluorescence intensity is still of 2-fold, like in the L7H-helix variant, but in this case, it is spread out over the entire pH range in a nearly linear fashion. Therefore, placing the fluorophore and electron acceptor at the flexible tails seems to facilitate the effective transducing of the pH induced conformational change onto gradual changes in distance sensitive to PET (over the narrow range slightly below and above 1 nm). Interestingly, the curve that we observe by PET in this construct tracks very closely the conformational changes observed by CD (Figure 4.8 right inset), including the gradual tapering off the change in the pH 7 to 9 range. The implication is that the combination of PET and a careful optimization of the electron donor and acceptor placement results in a broadband fluorescence detection of pH over nearly 8 orders of magnitude in [H^+]. This is the first example available of a fluorescence-based biosensor that is solely capable of monitoring pH over the entire range that is physiologically significant and beyond.

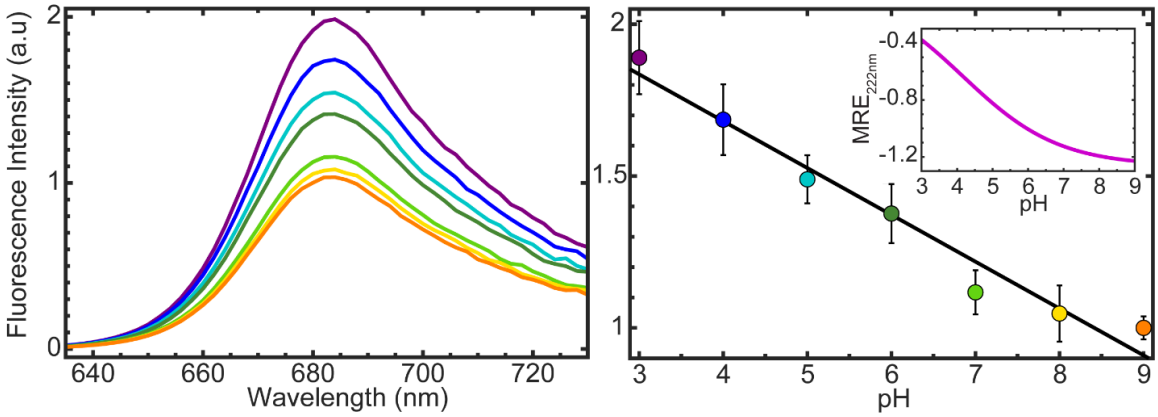

Figure 4.8 Fluorescence intensity change of L7H variant of gpW labeled with ATTO 655 at the C-terminal. Normalized fluorescence emission spectra at different pH and at 293 K with excitation at 620 nm (top). The spectra are colored following the visible light spectrum with the lowest pH corresponding to the highest energy (violet). Total fluorescence intensity was calculated from the area under each spectrum as a pH function and is plotted in the panel on the left with the symbols colored accordingly (left). The inset represents the conformational change of L7H as a function of pH using the Far-UV CD as a spectroscopic probe.

4.4 Conclusions

In the current study, we were able to engineer a fluorescence signal readout of the conformational changes that take place in our broadband pH conformational transducer. We tried several strategies, but the readout proved to be more effective in converting conformational changes within a compact globular ensemble onto a fluorescence signal using the PET effect. Notably, PET combined with the placement of fluorophore and quencher in the close distance in the folded structure, but in regions flexible enough to effectively transfer the average conformation onto a distance change produced the best output. The PET readouts are not ratiometric, but they still have high sensitivity (detection of fluorescence) and relatively high contrast (about 2-fold change) compared to other fluorescence-based signals. Therefore, its combination with the conformational rheostat pH transducer engineered on a downhill folding domain via histidine grafting provides a direct demonstration of a general principle for building broadband fluorescence biosensors.

We have successfully designed a PET-based biosensor to monitor pH with broadband capabilities. This system may be as well be capable of fast response times because it is based on the (un)folding coupled to protonation of the microsecond folding protein gpW[23]. By the same token, it would be exciting to determine whether this biosensor is capable of producing analog signals at the single-molecule level, which is another potential property of conformational rheostatic biosensors. Both the temporal resolution and the single-molecule analog response could be studies using single-molecule fluorescence spectroscopy. The PET effect results in reduced emission by static quenching (it does not change the fluorophore's emission lifetime), so to detect the PET signal as a function of pH would require measuring the photon output per unit of time from single-molecules rather than monitoring changes in a lifetime. However, these measurements are relatively straightforward with current single-molecule confocal fluorescence microscopes. Such experiments could demonstrate the potential for analog pH detection in real-time (sub-milliseconds) and with nanoscale (single-molecule) resolution, features that would be extremely useful for biomedical applications[138, 139].

For the purpose of developing entirely genetically encoded biosensors that could be used for in-vivo monitoring of intracellular pH, the PET readout strategy would require to use natural fluorophores-quencher pairs that can efficiently undergo electron transfer in the living cell. In this regard, a few fluorophores that have been recently derived from green fluorescence proteins or fluoresceine could be considered[140, 141].

Summary and Future Directions

This dissertation demonstrated that downhill folding proteins could be engineered to undergo (un)folding coupled to binding, resulting in rheostatic pH transducing with broadband capabilities. For this purpose, we implemented and characterized an engineering strategy based on grafting histidines on locations of the protein with different degrees of solvent exposure. We showed how the changes in solvent exposure and the number of grafts used could be used to tune properties of the transducer, such as its dynamic range and maximal signal. Moreover, we succeeded in developing suitable fluorescence readouts that effectively convert the localized conformational changes of conformational rheostats onto changes in fluorescence intensity that effectively track the broadband pH dependence. These results provide the basis for developing novel biosensor design tools that exploit rheostatic conformational transducer features. However, some more steps need to be worked out to develop a biosensor for practical uses in life imaging applications. A significant factor is the development of fully genetically encoded versions of the sensor based on PET readouts. For this purpose, one could use an extrinsic fluorophore that is able to penetrate the cells or to use recently developed PET-competent variants of the green fluorescent protein. The proof-of-concept work developed in this thesis should provide the basis for such future developments.

In parallel, the molecular LEGO approach revealed that conformational rheostat effects are not just an intriguing feature of downhill folding that could be exploited for technological applications, but rather seems to be the molecular mechanism controlling the complex folding upon binding that IDPs or, more precisely, IPDPs, used to exert their biological functions. With this approach, we have added a much-needed tool to the IDP research toolset. This new tool permits to dissect the folding energy landscape of an IDP in its unbound form, which is also a crucial factor to understand the relationships between the conformational properties of the IDP, their mechanisms of folding coupled to binding, and their biological functions. This approach preferentially targets IDPs that have some degree of residual structure. However, it is potentially extensible to much more disordered IDPs. In this case, we can use the 3D structures they form in complex with their partners as a starting point for the design of the LEGO components. In our proof of concept, we studied an IPDP using a structure-promoting cosolvent and CD as a spectroscopic technique to monitor the conformational changes on the protein, together with molecular dynamics simulations. Another key element of the approach is the analysis of the data (experiments and simulations) with statistical thermodynamic models that capture the gradual conformational changes observed in these proteins. In the future, the molecular LEGO

could be carried out tracking conformation by NMR, which provides residue-level structural resolution.

Lastly, a large fraction of the work discussed in this thesis has been performed in close collaboration with molecular simulation studies. The hand-in-hand combination of experiments and simulations is starting to become the new standard for molecular biophysics research. Here we have made some important strides in this direction. Moreover, the good agreement that we obtained between the experiments and computation proved the potential impact that such integration can bring to the field, with the simulations providing all the structural and mechanistic details, and key experiments and analysis providing the essential benchmarks to validate the results of the simulations or highlight their shortcomings and guide their refinement.

Bibliography

1. Protein's Functions in the Body. 2020.

2. Englander, S. W.; Mayne, L., The nature of protein folding pathways. *Proceedings of the National Academy of Sciences* **2014,** *111* (45), 15873-15880.

3. Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. H., The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences* **1961,** *47* (9), 1309-1314.

4. Babu, M. M., The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochemical Society Transactions* **2016,** *44* (5), 1185-1200.

5. Luheshi, L. M.; Crowther, D. C.; Dobson, C. M., Protein misfolding and disease: from the test tube to the organism. *Current Opinion in Chemical Biology* **2008,** *12* (1), 25-31.

6. Levinthal, C., Are there pathways for protein folding? *Journal de Chimie Physique* **1968,** *65*, 44-45.

7. Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G., Theory of protein folding: the energy landscape perspective. *Annual Review of Physical Chemistry* **1997,** *48*, 545-600.

8. Naganathan, A. N.; Muñoz, V., Thermodynamics of Downhill Folding: Multi-Probe Analysis of PDD, a Protein that Folds Over a Marginal Free Energy Barrier. *The Journal of Physical Chemistry B* **2014,** *118* (30), 8982-8994.

9. Kulkarni, P.; Solomon, T. L.; He, Y.; Chen, Y.; Bryan, P. N.; Orban, J., Structural metamorphism and polymorphism in proteins on the brink of thermodynamic stability. *Protein Science* **2018,** *27* (9), 1557-1567.

10. Dyson, H. J.; Wright, P. E., Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology* **2005,** *6* (3), 197-208.

11. Oldfield, C. J.; Dunker, A. K., Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annual Review of Biochemistry* **2014,** *83* (1), 553-584.

12. Uversky, V. N.; Dunker, A. K., Understanding protein non-folding. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2010,** *1804* (6), 1231-1264.

13. Borgia, A.; Borgia, M. B.; Bugge, K.; Kissling, V. M.; Heidarsson, P. O.; Fernandes, C. B.; Sottini, A.; Soranno, A.; Buholzer, K. J.; Nettels, D.; Kragelund, B.

B.; Best, R. B.; Schuler, B., Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **2018,** *555* (7694), 61-66.

14. Chang, Y. G.; Cohen, S. E.; Phong, C.; Myers, W. K.; Kim, Y. I.; Tseng, R.; Lin, J.; Zhang, L.; Boyd, J. S.; Lee, Y.; Kang, S.; Lee, D.; Li, S.; Britt, R. D.; Rust, M. J.; Golden, S. S.; LiWang, A., Circadian rhythms. A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. *Science* **2015,** *349* (6245), 324-8.

15. Muñoz, V.; Sadqi, M.; Naganathan, A. N.; de Sancho, D., Exploiting the downhill folding regime via experiment. *Frontiers in Life Science Journal* **2008,** *2* (6), 342-353.

16. Naganathan, A. N.; Li, P.; Perez-Jimenez, R.; Sanchez-Ruiz, J. M.; Muñoz, V., Navigating the Downhill Protein Folding Regime via Structural Homologues. *Journal of the American Chemical Society* **2010,** *132* (32), 11183-11190.

17. Munoz, V.; Campos, L. A.; Sadqi, M., Limited cooperativity in protein folding. *Current Opinion in Structural Biology* **2016,** *36*, 58-66.

18. Cerminara, M.; Desai, T. M.; Sadqi, M.; Muñoz, V., Downhill Protein Folding Modules as Scaffolds for Broad-Range Ultrafast Biosensors. *Journal of the American Chemical Society* **2012,** *134* (19), 8010-8013.

19. Jackson, S. E., How do small single-domain proteins fold? *Folding and Design* **1998,** *3* (4), R81-R91.

20. Liu, J.; Campos, L. A.; Cerminara, M.; Wang, X.; Ramanathan, R.; English, D. S.; Muñoz, V., Exploring one-state downhill protein folding in single molecules. *Proceedings of the National Academy of Sciences* **2012,** *109* (1), 179-184.

21. Cho, S. S.; Weinkam, P.; Wolynes, P. G., Origins of barriers and barrierless folding in BBL. *Proceedings of the National Academy of Sciences* **2008,** *105* (1), 118-123.

22. Garcia-Mira, M. M.; Sadqi, M.; Fischer, N.; Sanchez-Ruiz, J. M.; Muñoz, V., Experimental Identification of Downhill Protein Folding. *Science* **2002,** *298* (5601), 2191-2195.

23. Fung, A.; Li, P.; Godoy-Ruiz, R.; Sanchez-Ruiz, J. M.; Munoz, V., Expanding the realm of ultrafast protein folding: gpW, a midsize natural single-domain with alpha+beta topology that folds downhill. *Journal of the American Chemical Society* **2008,** *130* (23), 7489-95.

24. Dyson, H. J.; Wright, P. E., Coupling of folding and binding for unstructured proteins. *Current Opinion in Structural Biology* **2002,** *12* (1), 54-60.

25. Sugase, K.; Dyson, H. J.; Wright, P. E., Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **2007,** *447* (7147), 1021-1025.

26. Wright, P. E.; Dyson, H. J., Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology* **1999,** *293* (2), 321-331.

27. Wright, P. E.; Dyson, H. J., Linking folding and binding. *Current Opinion in Structural Biology* **2009,** *19* (1), 31-38.

28. Waters, L.; Yue, B.; Veverka, V.; Renshaw, P.; Bramham, J.; Matsuda, S.; Frenkiel, T.; Kelly, G.; Muskett, F.; Carr, M.; Heery, D. M., Structural diversity in p160/CREB-binding protein coactivator complexes. *Journal of Biological Chemistry* **2006,** *281* (21), 14787-95.

29. Uversky, V. N., Unusual biophysics of intrinsically disordered proteins. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2013,** *1834* (5), 932-951.

30. Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J., The ensemble nature of allostery. *Nature* **2014,** *508* (7496), 331-339.

31. Naganathan, A. N.; Doshi, U.; Fung, A.; Sadqi, M.; Muñoz, V., Dynamics, Energetics, and Structure in Protein Folding. *Biochemistry* **2006,** *45* (28), 8466-8475.

32. Campos, L. A.; Sadqi, M.; Muñoz, V., Lessons about Protein Folding and Binding from Archetypal Folds. *Accounts of Chemical Research* **2020,** *53* (10), 2180-2188.

33. Muñoz, V.; Cerminara, M., When fast is better: protein folding fundamentals and mechanisms from ultrafast approaches. *Biochemical Journal* **2016,** *473* (17), 2545-2559.

34. Tanford, C., Protein denaturation. *Advances in Protein Chemistry* **1968,** *23*, 121-282.

35. Morris, E. R.; Searle, M. S., Overview of protein folding mechanisms: experimental and theoretical approaches to probing energy landscapes. *Current Protocols in Protein Science* **2012,** *Chapter 28*, Unit 28.2.1-22.

36. Fersht, A. R., Protein folding and stability: the pathway of folding of barnase. *FEBS Letters* **1993,** *325* (1), 5-16.

37. Naganathan, A. N.; Muñoz, V., Insights into protein folding mechanisms from large scale analysis of mutational effects. *Proceedings of the National Academy of Sciences* **2010,** *107* (19), 8611-8616.

38. Bhalla, N.; Jolly, P.; Formisano, N.; Estrela, P., Introduction to biosensors. *Experientia supplementum* **2016,** *60* (1), 1-8.

39. Mehrotra, P., Biosensors and their applications - A review. *Journal of oral biology and craniofacial research* **2016,** *6* (2), 153-159.

40. Turner, A. P. F., Biosensors: sense and sensibility. *Chemical Society Reviews* **2013,** *42* (8), 3184-3196.

41. Zhou, Q.; Son, K.; Liu, Y.; Revzin, A., Biosensors for Cell Analysis. *Annual Review of Biomedical Engineering* **2015,** *17* (1), 165-190.

42. Kunzelmann, S.; Solscheid, C.; Webb, M. R., Fluorescent biosensors: design and application to motor proteins. *Experientia supplementum* **2014,** *105*, 25-47.

43. Vallée-Bélisle, A.; Plaxco, K. W., Structure-switching biosensors: inspired by Nature. *Current Opinion in Structural Biology* **2010,** *20* (4), 518-26.

44. Oh, K. J.; Cash, K. J.; Plaxco, K. W., Beyond molecular beacons: optical sensors based on the binding-induced folding of proteins and polypeptides. *Chemistry* **2009,** *15* (10), 2244-2251.

45. Zadran, S.; Standley, S.; Wong, K.; Otiniano, E.; Amighi, A.; Baudry, M., Fluorescence resonance energy transfer (FRET)-based biosensors: visualizing cellular dynamics and bioenergetics. *Applied Microbiology and Biotechnology* **2012,** *96* (4), 895-902.

46. Fersht, A. R., Nucleation mechanisms in protein folding. *Current Opinion in Structural Biology* **1997,** *7* (1), 3-9.

47. Dyson, H. J.; Merutka, G.; Waltho, J. P.; Lerner, R. A.; Wright, P. E., Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. I. Myohemerythrin. *Journal of Molecular Biology* **1992,** *226* (3), 795-817.

48. Dyson, H. J.; Rance, M.; Houghten, R. A.; Wright, P. E.; Lerner, R. A., Folding of immunogenic peptide fragments of proteins in water solution: II. The nascent helix. *Journal of Molecular Biology* **1988,** *201* (1), 201-217.

49. Dyson, H. J.; Wright, P. E., Peptide conformation and protein folding. *Current Opinion in Structural Biology* **1993,** *3* (1), 60-65.

50. Itzhaki, L. S.; Neira, J. L.; Ruiz-Sanz, J.; de Prat Gay, G.; Fersht, A. R., Search for Nucleation Sites in Smaller Fragments of Chymotrypsin Inhibitor 2. *Journal of Molecular Biology* **1995,** *254* (2), 289-304.

51. Neira, J. L.; Itzhaki, L. S.; Otzen, D. E.; Davis, B.; Fersht, A. R., Hydrogen exchange in chymotrypsin inhibitor 2 probed by mutagenesis11Edited by J.Karn. *Journal of Molecular Biology* **1997,** *270* (1), 99-110.

52. Lindsay, R. J.; Mansbach, R. A.; Gnanakaran, S.; Shen, T., Effects of pH on an IDP conformational ensemble explored by molecular dynamics simulation. *Biophysical Chemistry* **2021,** *271*, 106552.

53. Best, R. B., Computational and theoretical advances in studies of intrinsically disordered proteins. *Current Opinion in Structural Biology* **2017,** *42*, 147-154.

54. Chong, S.-H.; Chatterjee, P.; Ham, S., Computer Simulations of Intrinsically Disordered Proteins. *Annual Review of Physical Chemistry* **2017,** *68* (1), 117-134.

55. 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* (6871), 549-553.

56. Knott, M.; Best, R. B., A Preformed Binding Interface in the Unbound Ensemble of an Intrinsically Disordered Protein: Evidence from Molecular Simulations. *PLOS Computational Biology* **2012,** *8* (7), e1002605.

57. Naganathan, A. N.; Orozco, M., The Native Ensemble and Folding of a Protein Molten-Globule: Functional Consequence of Downhill Folding. *Journal of the American Chemical Society* **2011,** *133* (31), 12154-12161.

58. Zhang, N.; Liu, F.-F.; Dong, X.-Y.; Sun, Y., Molecular Insight into the Counteraction of Trehalose on Urea-Induced Protein Denaturation Using Molecular Dynamics Simulation. *The Journal of Physical Chemistry B* **2012,** *116* (24), 7040-7047.

59. Ferreon, A. C.; Ferreon, J. C.; Wright, P. E.; Deniz, A. A., Modulation of allostery by protein intrinsic disorder. *Nature* **2013,** *498* (7454), 390-4.

60. Kim, J.-Y.; Meng, F.; Yoo, J.; Chung, H. S., Diffusion-limited association of disordered protein by non-native electrostatic interactions. *Nature Communications* **2018,** *9* (1), 4707.

61. 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 Communications* **2018,** *9* (1), 4708.

62. Zosel, F.; Mercadante, D.; Nettels, D.; Schuler, B., A proline switch explains kinetic heterogeneity in a coupled folding and binding reaction. *Nature Communications* **2018,** *9* (1), 3332.

63. Lee, C. W.; Martinez-Yamout, M. A.; Dyson, H. J.; Wright, P. E., Structure of the p53 transactivation domain in complex with the nuclear receptor coactivator binding domain of CREB binding protein. *Biochemistry* **2010,** *49* (46), 9964-9971.

64. Qin, B. Y.; Liu, C.; Srinath, H.; Lam, S. S.; Correia, J. J.; Derynck, R.; Lin, K., Crystal Structure of IRF-3 in Complex with CBP. *Structure* **2005,** *13* (9), 1269-1277.

65. Kjaergaard, M.; Teilum, K.; Poulsen, F. M., Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP. *Proceedings of the National Academy of Sciences of the United States of America* **2010,** *107* (28), 12535-40.

66. Muñoz, V.; Serrano, L., Elucidating the folding problem of helical peptides using empirical parameters. *Nature Structural Biology* **1994,** *1* (6), 399-409.

67. Buck, M., Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. *Quarterly Reviews of Biophysics* **1998,** *31* (3), 297-355.

68. Luo, P.; Baldwin, R. L., Mechanism of Helix Induction by Trifluoroethanol: A Framework for Extrapolating the Helix-Forming Properties of Peptides from Trifluoroethanol/Water Mixtures Back to Water. *Biochemistry* **1997,** *36* (27), 8413-8421.

69. Blanco, F. J.; Serrano, L., Folding of protein G B1 domain studied by the conformational characterization of fragments comprising its secondary structure elements. *European Journal of Biochemistry* **1995,** *230* (2), 634-49.

70. Searle, M. S.; Zerella, R.; Williams, D. H.; Packman, L. C., Native-like β-hairpin structure in an isolated fragment from ferredoxin: NMR and CD studies of solvent effects on the N-terminal 20 residues. *Protein Engineering, Design and Selection* **1996,** *9* (7), 559-565.

71. Reiersen, H.; Rees, A. R., Trifluoroethanol may form a solvent matrix for assisted hydrophobic interactions between peptide side chains. *Protein Engineering, Design and Selection* **2000,** *13* (11), 739-43.

72. Holzwarth, G.; Doty, P., THE ULTRAVIOLET CIRCULAR DICHROISM OF POLYPEPTIDES. *Journal of the American Chemical Society* **1965,** *87*, 218-28.

73. Lifson, S.; Roig, A., On the Theory of Helix—Coil Transition in Polypeptides. *The Journal of Chemical Physics* **1961,** *34* (6), 1963-1974.

74. Zimm, B. H.; Bragg, J. K., Theory of the Phase Transition between Helix and Random Coil in Polypeptide Chains. *The Journal of Chemical Physics* **1959,** *31* (2), 526-535.

75. Doshi, U.; Muñoz, V., Kinetics of α-helix formation as diffusion on a one-dimensional free energy surface. *Chemical Physics* **2004,** *307* (2), 129-136.

76. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19-25.

77. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **1995,** *91* (1), 43-56.

78. Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. In *Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS*, Solving Software Challenges for Exascale, Cham, 2015//; Markidis, S.; Laure, E., Eds. Springer International Publishing: Cham, 2015; pp 3-27.

79. Piana, S.; Lindorff-Larsen, K.; Shaw, D. E., How robust are protein folding simulations with respect to force field parameterization? *Biophysical Journal* **2011,** *100* (9), L47-L49.

80. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993,** *98* (12), 10089-10092.

81. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **1997,** *18* (12), 1463-1472.

82. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984,** *81* (8), 3684-3690.

83. Parrinello, M.; Rahman, A., Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Physical Review Letters* **1980,** *45* (14), 1196-1199.

84. Chen, Y.-H.; Yang, J. T.; Chau, K. H., Determination of the helix and β form of proteins in aqueous solution by circular dichroism. *Biochemistry* **1974,** *13* (16), 3350-3359.

85. Hennessey, J. P.; Johnson, W. C., Information content in the circular dichroism of proteins. *Biochemistry* **1981,** *20* (5), 1085-1094.

86. Scholtz, J. M.; Baldwin, R. L., The mechanism of alpha-helix formation by peptides. *Annual Review of Biophysics and Biomolecular Structure* **1992,** *21*, 95-118.

87. Kjaergaard, M.; Teilum, K.; Poulsen, F. M., Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP. *Proceedings of the National Academy of Sciences of the United States of America* **2010,** *107* (28), 12535-40.

88. De Sancho, D.; Muñoz, V., Integrated prediction of protein folding and unfolding rates from only size and structural class. *Physical Chemistry Chemical Physics* **2011,** *13* (38), 17030-17043.

89. Brookes, D. H.; Head-Gordon, T., Experimental Inferential Structure Determination of Ensembles for Intrinsically Disordered Proteins. *Journal of the American Chemical Society* **2016,** *138* (13), 4530-4538.

90. He, Y.; Nagpal, S.; Sadqi, M.; de Alba, E.; Muñoz, V., Glutton: a tool for generating structural ensembles of partly disordered proteins from chemical shifts. *Bioinformatics* **2019,** *35* (7), 1234-1236.

91. Jensen, M. R.; Ruigrok, R. W.; Blackledge, M., Describing intrinsically disordered proteins at atomic resolution by NMR. *Current Opinion in Structural Biology* **2013,** *23* (3), 426-35.

92. Chou, P. Y.; Fasman, G. D., Prediction of the secondary structure of proteins from their amino acid sequence. *Advances in Enzymology and Related Areas of Molecular Biology* **1978,** *47*, 45-148.

93. Bolen, D. W., Protein stabilization by naturally occurring osmolytes. *Methods in Molecular Biology* **2001,** *168*, 17-36.

94. Babu, M. M.; Kriwacki, R. W.; Pappu, R. V., Versatility from Protein Disorder. *Science* **2012,** *337* (6101), 1460.

95. Wang, Y.; Chu, X.; Longhi, S.; Roche, P.; Han, W.; Wang, E.; Wang, J., Multiscaled exploration of coupled folding and binding of an intrinsically disordered molecular recognition element in measles virus nucleoprotein. *Proceedings of the National Academy of Sciences* **2013,** *110* (40), E3743.

96. Plaxco, K. W.; Soh, H. T., Switch-based biosensors: a new approach towards real-time, in vivo molecular detection. *Trends in Biotechnology* **2011,** *29* (1), 1-5.

97. Ha, J.-H.; Loh, S. N., Protein conformational switches: from nature to design. *Chemistry* **2012,** *18* (26), 7984-7999.

98. Vallée-Bélisle, A.; Ricci, F.; Plaxco, K. W., Thermodynamic basis for the optimization of binding-induced biomolecular switches and structure-switching biosensors. *Proceedings of the National Academy of Sciences* **2009,** *106* (33), 13802-13807.

99. Li, P.; Oliva, F. Y.; Naganathan, A. N.; Muñoz, V., Dynamics of one-state downhill protein folding. *Proceedings of the National Academy of Sciences* **2009,** *106* (1), 103-108.

100. Crooks, J. E., Chapter 3 Proton Transfer to and from Atoms other than Carbon. In *Comprehensive Chemical Kinetics*, Bamford, C. H.; Tipper, C. F. H., Eds. Elsevier: 1977; Vol. 8, pp 197-250.

101. Sborgi, L.; Verma, A.; Munoz, V.; de Alba, E., Revisiting the NMR structure of the ultrafast downhill folding protein gpW from bacteriophage lambda. *PloS one* **2011,** *6* (11), e26409.

102. Sborgi , L.; Verma, A.; Piana, S.; Lindorff-Larsen, K.; Cerminara, M.; Santiveri, C. M.; Shaw, D. E.; de Alba, E.; Muñoz, V., Interaction Networks in Protein Folding via Atomic-Resolution Experiments and Long-Time-Scale Molecular Dynamics Simulations. *Journal of the American Chemical Society* **2015,** *137* (20), 6506-6516.

103. White, K. A.; Ruiz, D. G.; Szpiech, Z. A.; Strauli, N. B.; Hernandez, R. D.; Jacobson, M. P.; Barber, D. L., Cancer-associated arginine-to-histidine mutations confer a gain in pH sensing to mutant proteins. *Science Signaling* **2017,** *10* (495), eaam9931.

104. Deochand, D. K.; Perera, I. C.; Crochet, R. B.; Gilbert, N. C.; Newcomer, M. E.; Grove, A., Histidine switch controlling pH-dependent protein folding and DNA binding in a transcription factor at the core of synthetic network devices. *Molecular BioSystems* **2016,** *12* (8), 2417-2426.

105. Perier, A.; Chassaing, A.; Raffestin, S.; Pichard, S.; Masella, M.; Ménez, A.; Forge, V.; Chenal, A.; Gillet, D., Concerted Protonation of Key Histidines Triggers Membrane Interaction of the Diphtheria Toxin T Domain*. *Journal of Biological Chemistry* **2007,** *282* (33), 24239-24245.

106. Narayan, A.; Naganathan, A. N., Switching Protein Conformational Substates by Protonation and Mutation. *The Journal of Physical Chemistry B* **2018,** *122* (49), 11039-11047.

107. Malevanets, A.; Chong, P. A.; Hansen, D. F.; Rizk, P.; Sun, Y.; Lin, H.; Muhandiram, R.; Chakrabartty, A.; Kay, L. E.; Forman-Kay, J. D.; Wodak, S. J., Interplay of buried histidine protonation and protein stability in prion misfolding. *Scientific Reports* **2017,** *7* (1), 882.

108. Boyken, S. E.; Benhaim, M. A.; Busch, F.; Jia, M.; Bick, M. J.; Choi, H.; Klima, J. C.; Chen, Z.; Walkey, C.; Mileant, A.; Sahasrabuddhe, A.; Wei, K. Y.; Hodge, E. A.; Byron, S.; Quijano-Rubio, A.; Sankaran, B.; King, N. P.; Lippincott-Schwartz, J.; Wysocki, V. H.; Lee, K. K.; Baker, D., De novo design of tunable, pH-driven conformational changes. *Science* **2019,** *364* (6441), 658-664.

109. Yildiz, Ö.; Vinothkumar, K. R.; Goswami, P.; Kühlbrandt, W., Structure of the monomeric outer-membrane porin OmpG in the open and closed conformation. *The EMBO Journal* **2006,** *25* (15), 3702-3713.

110. Edgcomb, S. P.; Murphy, K. P., Variability in the pKa of histidine side-chains correlates with burial within proteins. *Proteins: Structure, Function, and Bioinformatics* **2002,** *49* (1), 1-6.

111. Pires, D. E. V.; Ascher, D. B.; Blundell, T. L., DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Research* **2014,** *42* (W1), W314-W319.

112. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005,** *26* (16), 1701-1718.

113. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society* **1996,** *118* (45), 11225-11236.

114. Abascal, J. L. F.; Vega, C., A general purpose model for the condensed phases of water: TIP4P/2005. *The Journal of Chemical Physics* **2005,** *123* (23), 234505.

115. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007,** *126* (1), 014101.

116. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981,** *52* (12), 7182-7190.

117. Chaudhury, S.; Ripoll, D. R.; Wallqvist, A., Structure-based pKa prediction provides a thermodynamic basis for the role of histidines in pH-induced conformational transitions in dengue virus. *Biochemistry and Biophysics Reports* **2015,** *4*, 375-385.

118. Ibarra-Molero, B.; Naganathan, A. N.; Sanchez-Ruiz, J. M.; Muñoz, V., Modern Analysis of Protein Folding by Differential Scanning Calorimetry. *Methods in Enzymology* **2016,** *567*, 281-318.

119. Richardson, D. S.; Gregor, C.; Winter, F. R.; Urban, N. T.; Sahl, S. J.; Willig, K. I.; Hell, S. W., SRpHi ratiometric pH biosensors for super-resolution microscopy. *Nature Communications* **2017,** *8* (1), 577.

120. Casey, J. R.; Grinstein, S.; Orlowski, J., Sensors and regulators of intracellular pH. *Nature Reviews Molecular Cell Biology* **2010,** *11* (1), 50-61.

121. Han, J.; Burgess, K., Fluorescent Indicators for Intracellular pH. *Chemical Reviews* **2010,** *110* (5), 2709-2728.

122. Srivastava, J.; Barber, D. L.; Jacobson, M. P., Intracellular pH sensors: design principles and functional significance. *Physiology (Bethesda)* **2007,** *22*, 30-9.

123. Bagur, R.; Hajnóczky, G., Intracellular Ca2+ Sensing: Its Role in Calcium Homeostasis and Signaling. *Molecular Cell* **2017,** *66* (6), 780-788.

124. Czarnik, A. W., Desperately seeking sensors. *Chemistry and Biology* **1995,** *2* (7), 423-8.

125. Miyawaki, A.; Griesbeck, O.; Heim, R.; Tsien, R. Y., Dynamic and quantitative Ca2+ measurements using improved cameleons. *Proceedings of the National Academy of Sciences of the United States of America* **1999,** *96* (5), 2135-40.

126. Mank, M.; Santos, A. F.; Direnberger, S.; Mrsic-Flogel, T. D.; Hofer, S. B.; Stein, V.; Hendel, T.; Reiff, D. F.; Levelt, C.; Borst, A.; Bonhoeffer, T.; Hübener, M.; Griesbeck, O., A genetically encoded calcium indicator for chronic in vivo two-photon imaging. *Nature Methods* **2008,** *5* (9), 805-11.

127. Clapham, D. E., Calcium signaling. *Cell* **2007,** *131* (6), 1047-58.

128. Tainaka, K.; Sakaguchi, R.; Hayashi, H.; Nakano, S.; Liew, F. F.; Morii, T., Design strategies of fluorescent biosensors based on biological macromolecular receptors. *Sensors (Basel)* **2010,** *10* (2), 1355-1376.

129. Tamura, T.; Hamachi, I., Recent Progress in Design of Protein-Based Fluorescent Biosensors and Their Cellular Applications. *ACS Chemical Biology* **2014,** *9* (12), 2708-2717.

130. Sharma, A.; Khan, R.; Catanante, G.; Sherazi, T. A.; Bhand, S.; Hayat, A.; Marty, J. L., Designed Strategies for Fluorescence-Based Biosensors for the Detection of Mycotoxins. *Toxins (Basel)* **2018,** *10* (5), 197.

131. Doose, S.; Neuweiler, H.; Sauer, M., A close look at fluorescence quenching of organic dyes by tryptophan. *Chemphyschem* **2005,** *6* (11), 2277-85.

132. Doose, S.; Neuweiler, H.; Sauer, M., Fluorescence quenching by photoinduced electron transfer: a reporter for conformational dynamics of macromolecules. *Chemphyschem* **2009,** *10* (9-10), 1389-98.

133. Arakawa, T.; Timasheff, S. N., Protein stabilization and destabilization by guanidinium salts. *Biochemistry* **1984,** *23* (25), 5924-5929.

134. Merchant, K. A.; Best, R. B.; Louis, J. M.; Gopich, I. V.; Eaton, W. A., Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proceedings of the National Academy of Sciences of the United States of America* **2007,** *104* (5), 1528-33.

135. Guo, L.; Chowdhury, P.; Glasscock, J. M.; Gai, F., Denaturant-induced expansion and compaction of a multi-domain protein: IgG. *Journal of molecular biology* **2008,** *384* (5), 1029-1036.

136. Sadqi, M.; Lapidus, L. J.; Muñoz, V., How fast is protein hydrophobic collapse? *Proceedings of the National Academy of Sciences* **2003,** *100* (21), 12117-12122.

137. Neuweiler, H.; Sauer, M., Using photoinduced charge transfer reactions to study conformational dynamics of biopolymers at the single-molecule level. *Current Pharmaceutical Biotechnology* **2004,** *5* (3), 285-98.

138. Lord, S. J.; Lee, H.-l. D.; Moerner, W. E., Single-molecule spectroscopy and imaging of biomolecules in living cells. *Analytical Chemistry* **2010,** *82* (6), 2192-2203.

139. Akkilic, N.; Geschwindner, S.; Höök, F., Single-molecule biosensors: Recent advances and applications. *Biosensors and Bioelectronics* **2020,** *151*, 111944.

140. Lv, X.; Yu, Y.; Zhou, M.; Hu, C.; Gao, F.; Li, J.; Liu, X.; Deng, K.; Zheng, P.; Gong, W.; Xia, A.; Wang, J., Ultrafast Photoinduced Electron Transfer in Green Fluorescent Protein Bearing a Genetically Encoded Electron Acceptor. *Journal of the American Chemical Society* **2015,** *137* (23), 7270-7273.

141. Adams, S. R.; Tsien, R. Y., Preparation of the membrane-permeant biarsenicals FlAsH-EDT2 and ReAsH-EDT2 for fluorescent labeling of tetracysteine-tagged proteins. *Nature Protocols* **2008,** *3* (9), 1527-1534.

Protein	**Amino Acid Sequence**
gpW-wt	MVRQE ELAAA RAALH DLMTG KRVAT VQKDG RRVEF TATSV SDLKK YIAEL EVQTG MTQRR RG
gpW-H15A	MVRQE ELAAA RAALA DLMTG KRVAT VQKDG RRVEF TATSV SDLKK YIAEL EVQTG MTQRR RG
gpW-L7H	MVRQE EHAAA RAALH DLMTG KRVAT VQKDG RRVEF TATSV SDLKK YIAEL EVQTG MTQRR RG
gpW-M18H	MVRQE ELAAA RAALH DLHTG KRVAT VQKDG RRVEF TATSV SDLKK YIAEL EVQTG MTQRR RG
gpW-F35H	MVRQE ELAAA RAALH DLMTG KRVAT VQKDG RRVEH TATSV SDLKK YIAEL EVQTG MTQRR RG

gpW-V40H	MVRQE ELAAA RAALH DLMTG KRVAT VQKDG RRVEF TATSH SDLKK YIAEL EVQTG MTQRR RG
gpW-sumo-M18H-F35H	MGHHH HHHGS DSEVN QEAKP EVKPE VKPET HINLK VSDGS SEIFF KIKKT TPLRR LMEAF AKRQG KEMDS LRFLY DGIRI QADQA PEDLD MEDND IIEAH REQIG GVRQE ELAAA RAALH DLHTG KRVAT VQKDG RRVEH TATSV SDLKK YIAEL EVQTG MTQRR RG
gpW-sumo-F35H-V40H	MGHHH HHHGS DSEVN QEAKP EVKPE VKPET HINLK VSDGS SEIFF KIKKT TPLRR LMEAF AKRQG KEMDS LRFLY DGIRI QADQA PEDLD MEDND IIEAH REQIG GVRQEE LAAAR AALHD LMTGK RVATV QKDGR RVEHT ATSHS DLKKY IAELEV QTGMT QRRRG
gpW-L7H-FRET	MCTRQ EEHAA ARAAL HDLMT GKRVA TVQKD GRRVE FTATS VSDLK KYIAE LEVQT GCTQR RRG
gpW-sumo-M18H-F35H-FRET	MGHHH HHHGS DSEVN QEAKP EVKPE VKPET HINLK VSDGS SEIFF KIKKT TPLRR LMEAF AKRQG KEMDS LRFLY DGIRI QADQA PEDLD MEDND IIEAH REQIG GCVR QEELA AARAA LHDLH TGKRV ATVQK CGRRV EHTAT SVSDL KKYIA ELEVQ TGMTQ RRRG
gpW-L7H-PET-helix	MGTWQ EEHAA ARAAL HDLMT GKRVA TVQKD GRRVE FTATS VSDLK KYIAE LEVCT GMTQR RRG
gpW-L7H-PET-Cter	MWTRQ EEHAA ARAAL HDLMT GKRVA TVQKD GRRVE FTATS VSDLK KYIAE LEVQT GCTQR RRG

NCBD5	MGHHH HHHHH ENLYF QGPPR SISPS ALQDL LRTLK SPSSP QQQQQ VLNIL KSNPQ LMAAF IKQRT AKYVA NQPGM Q

Appendix B

Investigating the pH dependence of the gpW-H15A

The figure below shows the α-helical content of gpW-H15A compared to the gpW-wt as a function of pH at 298 K. The blue line and magenta line guide the eyes. The result indicates that Histidine 15 does not have any effect on the pH sensitivity.