Computational Biophysics

RLE Group
Computational Biophysics Group

Academic and Research Staff
Professor Collin Stultz

Postdoctoral Fellows
Sophie Walker
Sarah Bowman (joint with Prof. Catherine Drennan)

Graduate Students
Charles Fisher
Elaine Gee
Orly Slavin
Priya Parayanthal (MEng)

Visiting Graduate Students
Inara Colon (Summer '10)

Visiting Faculty Students
Prof. Gustavo Lopez (U. Puerto Rico)

Undergraduate Students
Joy Okapala
Nebiyat Tsegaye

Collaborators
Prof. Catherine Drennan (MIT Chemistry)
Prof. John Guttag (MIT, EECS)
Prof. D. E. Ingber (Harvard Medical School)
Prof. Barbara Imperiali (MIT Chemistry)
Prof. Charlie Sodini (MIT, EECS)
Prof. Anantha Chandrakasan (MIT, EECS)

Technical and Support Staff
Arlene Wint

Introduction/Group Overview
The RLE Computational Biophysics Group is focused on understanding conformational changes in biomolecules that play an important role in common human diseases. The group uses an interdisciplinary approach combining computational modeling with biochemical experiments to make connections between conformational changes in macromolecules and disease progression. By employing two types of modeling, molecular dynamics and probabilistic modeling, hypotheses can be developed and then tested experimentally.

1. The Structure of Collagen and Collagenolysis

Project Staff:
Ramon Salsas (PhD '10), Paul Nerenberg (PhD '10), Prof. Collin Stultz

1.1 The Mechanism of Collagenolysis: A Substrate-centric View

Collagenolysis (collagen degradation) is a critical process in the progression of cancer metastasis, atherosclerosis, and other diseases. Despite considerable efforts to understand the steps involved, the exact mechanism of collagenolysis remains unknown. We have proposed a degradation mechanism in which collagen exists in multiple states, some featuring structures that are unwound in the vicinity of the collagenase cleavage site, and that collagenases preferentially bind to and stabilize the unwound structures before degradation occurs.

The focus of this work is to investigate several aspects of this degradation mechanism using both experimental and computational methods. In particular, we have used molecular dynamics simulations to explore the structure of human type I collagen in the vicinity of the collagenase cleavage site. Simulations of type I collagen in this region suggest that partial unfolding of the $\alpha 2$ chain is energetically preferred relative to unfolding of the $\alpha 1$ chains. Localized unfolding of the $\alpha 2$ chain leads to a metastable structure that has disrupted hydrogen bonds N-terminal to the collagenase cleavage site. Our data suggest that this disruption in hydrogen bonding pattern leads to increased chain flexibility, thereby enabling the $\alpha 2$ chain to sample different partially unfolded states.

We have recently conducted experiments that are consistent with these observations. In this work we demonstrate for the first time that the degradation of type I collagen at room temperature, a temperature well below type I collagen's melting temperature, does not require the presence of theMMPhemopexin-like domain. Moreover, peptide bond hydrolysis with MMP mutants that contain only the catalytic domain occurs at the unique collagenase cleavage site and not at other potential cleavage sites. Thus both peptide bond hydrolysis and enzyme specificity are achieved with the catalytic domain alone at room temperature. As full-length enzyme is thought to be necessary for collagenase-mediated unwinding, our data imply that enzyme-mediated unwinding is not required for collagenolysis in vitro.We therefore analyze our data in light of a conformational selection model where thermal fluctuations at the cleavage site cause collagen to adopt unfolded conformations that are complementary to the collagenase active site. Overall, our findings suggest that type I collagen can adopt locally unfolded states at room temperature and that collagenolysis occurs when collagenases cleave these locally unfolded states (Figure 1).
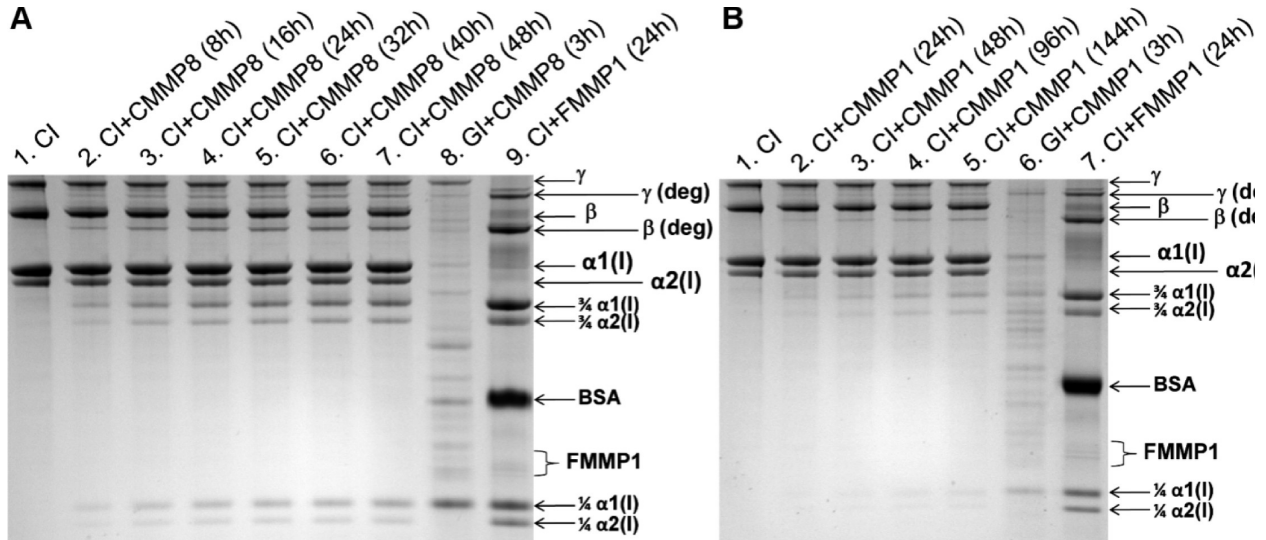
FIGURE 1: Degradation profiles of type I collagen at room temperature with CMMP8 and CMMP1. Intact type I collagen (CI) bands include monomeric R1(I) and R2(I) bands, dimeric β bands, and trimeric γ bands. The β and γ aggregates correspond to N-terminally cross-linked collagen molecules (41-44). Type I collagen degradation bands include R1(I) and R2(I) 3/4 and 1/4 fragments and degradation of cross-linked chains, βdeg and γdeg. (A) Type I collagen incubated with CMMP8. Lane 1: Type I collagen (150 µg/mL, 0.5 µM). Lanes 2-7: Type I collagen (150 µg/mL, 0.5 µM) incubated with the catalytic domain ofMMP8 (CMMP8) (25 µg/mL, 1.2 µM) for 8, 16, 24, 32, 40, and 48 h, respectively. Lane 8: Type I collagen gelatin (GI, 150 µg/mL, 1.5 µM) incubated with CMMP8 (13.8 µg/mL, 0.67 µM) for 3 h. Lane 9: Type I collagen (150 µg/mL, 0.5 µM) incubated with full-length MMP1 (FMMP1) (1.7 µg/mL, 30 nM) for 24 h. This lane contains a bovine serum albumin (BSA) band since FMMP1 is supplied in a buffer containing 1 mg/mL BSA. (B) Type I collagen incubated withCMMP1. Lane 1: Type I collagen (150 µg/mL, 0.5 µM). Lanes 2-5: Type I collagen (150 µg/mL, 0.5 µM) incubated with the catalytic domain of MMP1 (CMMP1) (40 µg/mL, 2.0 µM) for 24, 48, 96, and 144 h, respectively. Lane 6: Type I collagen gelatin (GI 150 µg/mL, 1.5 µM) incubated with CMMP1 (16.6 µg/mL, 0.83 µM) for 3 h. Lane 7: Type I collagen (150 µg/mL, 0.5 µM) incubated with full-lengthMMP1 (FMMP1) (1.7 µg/mL, 30nM) for 24 h.As with lane 9 in (A), this lane contains a BSA band since FMMP1 was supplied in a buffer containing 1 mg/mL BSA.
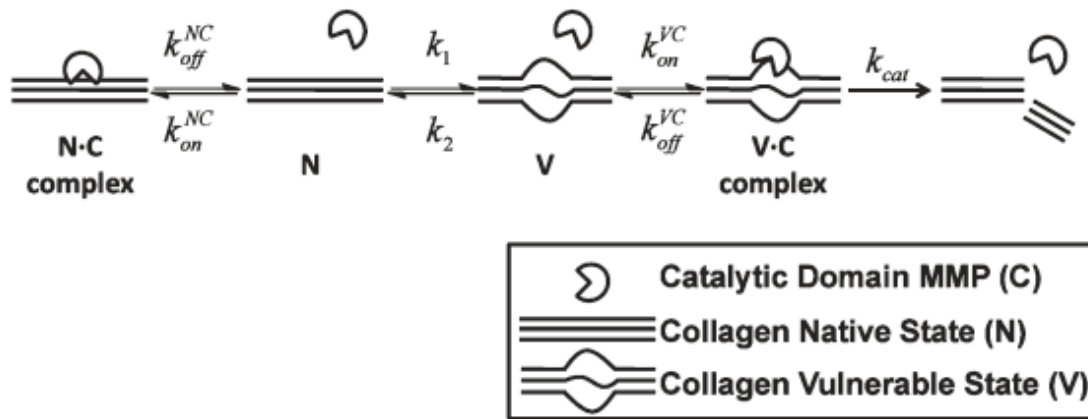

FIGURE 2: A conformational selection mechanism for collagenolysis with MMP catalytic domains. Collagen exists in an equilibrium between native (N) and vulnerable (V) states with the equilibrium determined by$K_{eq}=k_1/k_2$. The catalytic domain ofMMPs (C) interacts in a nonspecific manner with the native state with binding constant Kbind NC $=k_{on}$ NC/k_{off} NC, yielding theN3C complex. C binds to the vulnerable state, V, with binding constant Kbind VC = k_{on} VC/k_{off} VC, forming the V3C complex. The V3C complex is then degraded with catalytic rate kcat.

From these data we can calculate degradation rates and then use these data to determine thermodynamic and kinetic parameters for collagen degradation using the reaction scheme shown in Figure 2.

To obtain an estimate for Keq at room temperature, we quantified the extent of collagen degradation over time using the data shown in Figure 2A.We then fit the reaction scheme shown in Figure 4 to these data to obtain an estimate for Keq. We note that the previously discussed studies on a heterotrimeric type I collagen-like peptide, which contains the collagenase scissile bond, found that $(K_{bind}^{VC})_{CMMP8} = (0.7-1.1) \times 10^6$ M-1. As this peptide is a model for the collagenase cleavage site, we used this range of $(K_{bind}^{VC})_{CMMP8}$ in our numerical calculations of the model for type I collagen. We again computed the RMSE between the degradation time courses obtained with the model using many different values of Keq and compared these results to the experimentally determined degradation time course. Varying kcat of CMMP8 over 2 orders of magnitude, as we did for the aforementioned model peptide data, did not change the best fit values for Keq and caused the minimum RMSE to again vary by only 2% (Table S2 and Figure S2, Supporting Information). The best fits between the model and experiment are achieved when Keq=(1.7-2.1) $\times$ 10-3 (Figure 6A), and the corresponding degradation plot obtained from the model agrees well with experiment (Figure 6B). This value for Keq suggests that the folded triple-helical native state is more favorable at room temperature.

Despite the fact that CMMP1 has a higher specific activity than CMMP8 under our experimental conditions (as measured against a thiopeptolide substrate; see Materials and Methods), it degrades collagen with considerably lower efficiency than CMMP8 (Figure 2). To decipher the physical basis for this reduced activity, we applied the reaction scheme shown in Figure 4 to the CMMP1 degradation experiments. Given that Keq is an inherent property of type I collagen at a given temperature and both CMMP1 and CMMP8 collagen degradation experiments were performed at room temperature, Keq is the same for both reactions. Thus, we used our degradation model to determine the range of $(K_{bind}^{VC})_{CMMP1}$. As kcat corresponds to the catalytic rate once the protein has bound the vulnerable state (that is unfolded in the region of the scissile bond), it is estimated using previously determined kcat values for MMP1 degrading gelatin from type I collagen (Table S3, Supporting Information) (46). With Keq= (1.7-2.1) $\times$ 10-3, the best fits are found for $(K_{bind}^{VC})_{CMMP1}=(0.9-1.3) \times 10^4$ M-1 (Figure 7). These data suggest that CMMP1's binding affinity for the vulnerable state of collagen is nearly 2 orders of magnitude lower than that of CMMP8. Expressed as a difference of binding free energies, $\Delta\Delta G_{bind} = (\Delta G_{bind})_{CMMP1} - (\Delta G_{bind})_{CMMP8}$, the binding of CMMP1 to type I collagen is approximately 2.5 kcal/mol less favorable than the binding of CMMP8 to type I collagen.

In light of these considerations we expand our previous reaction scheme for the catalytic domain alone to consider the effects of the hemopexin-like domain on collagenolysis (Figure 3). In this reaction scheme, collagen exists in an equilibrium between native (N) and vulnerable (V) states, and either state can be bound by collagenases via the hemopexin-like domain, which contains binding sites for collagen (13, 32, 33, 68, 69). When the native state is bound, an N3H complex is formed that cannot be cleaved since the scissile bond in this structure is not accessible to the MMP active site. Since the catalytic domain of the enzyme is not bound to the scissile bond in the N3H complex and the putative binding site for the hemopexin-like domain in type I collagen is removed from the cleavage site (33, 62), the N3H complex can transition to a vulnerable state (V3H complex) via a conformational change similar to the one experienced by unbound collagen. Once the V3H complex is formed, the catalytic domain can then bind the accessible scissile bond, yielding the V3F complex, which then goes on to form degraded protein.
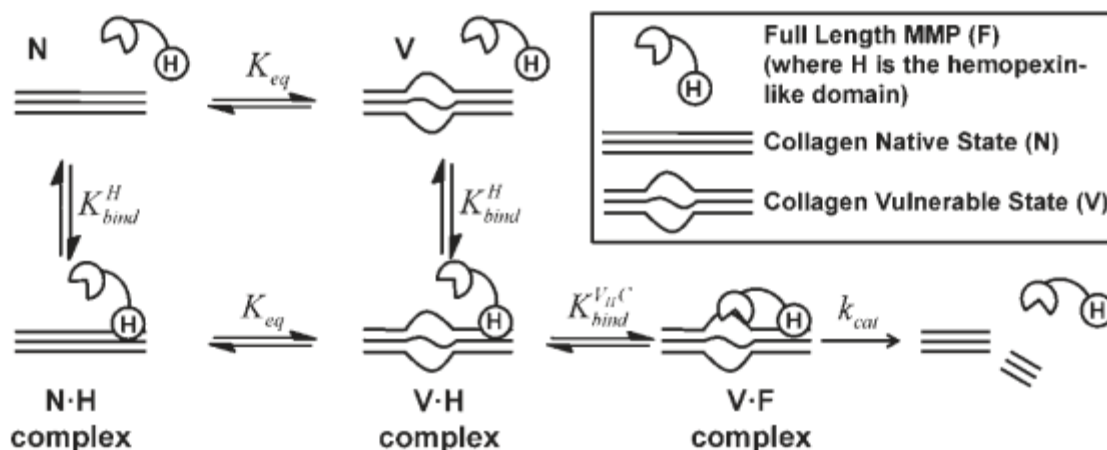
FIGURE 3: A conformational selection mechanism for collagenolysis with full-length collagenases. Collagen exists in an equilibrium between native (N) and vulnerable (V) states determined by the equilibriumconstantKeq. Full-lengthMMP (F) can bind to the native state of collagen via the hemopexin-like domain (H) with binding constant Kbind H , forming theN3Hcomplex.TheN3Hcomplex can transition to theV3Hcomplex via the equilibrium between native and vulnerable states,Keq.The full-length enzyme, F, can bind directly to the vulnerable state, V, via the hemopexin domain (H) forming the V3H complex, again determined by the binding constant Kbind H . Once aV3H complex is formed, the catalytic domain of the full-lengthMMP can bind to the vulnerable state with a binding constant Kbind VHC, forming theV3Fcomplex.TheV3F complex is then degraded with catalytic rate kcat.

2. Novel Methods for Modeling Intrinsically Disordered Proteins

Project Staff:
Charles Fisher, Prof. Collin Stultz

Modeling Intrinsically Disordered Proteins with Bayesian Statistics

The characterization of intrinsically disordered proteins is challenging because accurate models of these systems require a description of both their thermally accessible conformers and the associated relative stabilities or weights. These structures and weights are typically chosen such that calculated ensemble averages agree with some set of pre-specified experimental measurements; however, the large number of degrees of freedom in these systems typically leads to multiple conformational ensembles that are degenerate with respect to any given set of experimental observables. In this work we demonstrate that estimates of the relative stabilities of conformers within an ensemble are often incorrect when one does not account for the underlying uncertainty in the estimates themselves. Therefore, we present a method for modeling the conformational properties of disordered proteins that estimates the uncertainty in the weights of each conformer. The Bayesian Weighting (BW) formalism incorporates information from both experimental data and theoretical predictions to calculate a probability density over all possible ways of weighting the conformers in the ensemble. This probability density is then used to estimate the values of the weights. A unique and powerful feature of the approach is that it provides a built-in error measure that allows one to assess the accuracy of the ensemble. We validate the approach using reference ensembles constructed from the 5-residue peptide met-enkephalin and then apply the BW method to construct an ensemble of the K18 isoform of Tau protein. The K18 ensemble consists of a diverse set of conformations and agrees with all previously reported experimental data (Figure 4). Using this ensemble, we indentify a specific pattern of long-range contacts in K18 that correlates with the known aggregation properties of the sequence (Figure 5).
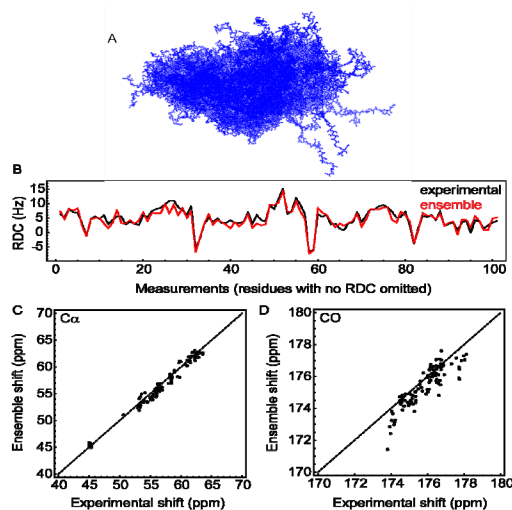
Figure 4. Application of the BW method to the K18 isoform of Tau. (A) A diverse set of 300 structures. (B) An overlay of the RDCs predicted from the ensemble and obtained from experiment shows good agreement (R = 0.94 and RMS = 1.31 Hz). The predicted RDCs are uniformly scaled to account for uncertainty in predicting the magnitude of alignment. (C) and (D) The Cα (R = 0.99 and RMS = 0.74 ppm) and CO (R = 0.88 and RMS = 0.72 ppm) chemical shifts obtained from the ensemble show good agreement with experiment.
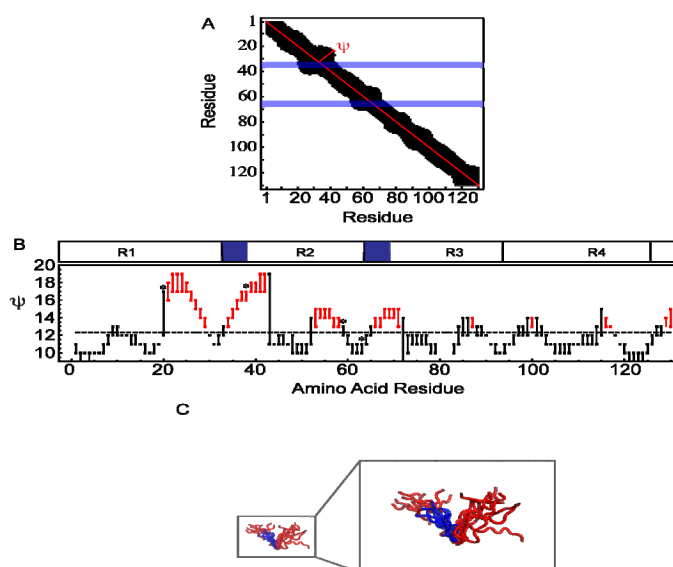

Figure 5. Analysis of long-range contacts in the K18 ensemble. (A) The contact map for K18 calculated from the Bayesian estimate for the weights. A black square indicates that the residues are within 25 Å on average. $\psi(i)$ is the length along the sequence to the farthest residue that makes a contact with residue i, and is shown in red as the distance from the diagonal. PHF6* (residues 33-38) and PHF6 (residues 64-69) are highlighted in blue. (B) 95% confidence intervals for ψ for each Cα in K18. The average ψ of all the residues is shown as a dashed line and residues with confidence intervals that lie above this line are shown in red. The PHF initiating hexapeptides are shaded blue in the map of the sequence and sites of mutation or phosphorylation known to alter the aggregation propensity of K18 *in vitro* are marked with a star (*) in the graph. (C) An overlay of the 10 most probable structures in the ensemble aligned via residues 20-44 (red), with the PHF6* colored blue, illustrates a turn motif.

2. Modeling the Unfolded State of alpha-Synuclein Protein

Sponsors:
HST, RLE

Project Staff:
Orly Ullman, Sophie Walker, Prof. Collin Stultz

α-Synuclein is an intrinsically disordered protein (IDP) that has been implicated in the pathogenesis of several neurodegenerative disorders such as Parkinsons disease (PD). A number of studies suggest that aggregates of α-synuclein are responsible, in part, for neuronal death and dysfunction in PD patients; therefore, obtaining insights into the mechanism of α-synuclein self-association is of immense importance. Since α-synuclein does not have a well-defined three-dimensional structure in solution, obtaining precise insights into the structural determinants of α-synuclein aggregation is problematic. This fact highlights the need for methods that can accurately model the unfolded state of IDPs. Once an adequate representation of the unfolded state of α-synuclein is obtained, and aggregation-prone conformers within this ensemble are identified, then selective inhibitors that prevent the self-association of aggregation-prone conformers can be constructed.

The objective of this proposal is to identify and characterize aggregation-prone conformers within the unfolded ensemble of α-synuclein. This is accomplished by building ensembles for both the wild-type protein and for mutants that are associated with increased aggregation potential. By comparing wild-type and mutant ensembles, we can identify potential aggregation-prone conformers.

The hypothesis that guides the construction of these ensembles is that IDPs, like α-synuclein, can be modeled as a set of energetically favorable conformers, where each conformer corresponds to a local energy minimum on a complex potential energy surface.

The Specific Aims of this work are:

1. Generate structural ensembles that adequately represent accessible conformations of wild-type α-synuclein using a series of computational methods developed in our laboratory. We have developed a novel method for modeling intrinsically disordered systems that uses Bayesian statistics to quantify the uncertainty in the underlying structural ensemble. A powerful feature of our approach is that it provides a built-in error measure that allows one to assess the accuracy of the constructed ensemble. Experimental data, such as chemical shifts and residual dipolar couplings are used to guide the generation of the appropriate ensembles.

2. Generate similar structural ensembles for α-synuclein mutants (A53T, A30P and E46K) that are associated with increased aggregation *in vitro* and identify potential aggregation-prone conformers by comparing ensembles corresponding to both wild-type and mutant proteins. As a part of this strategy we will also identify structures that are not associated with increased aggregation.

3. Generate solutions of α-synuclein that are enriched with potential aggregation-prone conformers and solutions enriched with conformers that are not associated with increased aggregation. This will be accomplished with a disulfide scrambling assay. The aggregation rate of these proteins will be tested *in vitro*.

Our long term objective is to develop a library of aggregation-prone conformers that will serve as the starting point for a ligand discovery strategy that specifically targets aggregation-prone conformers.

Results:

We divided α-synuclein into smaller segments each of 8 residue length for which we performed a near exhaustive search. We then combine these segments to form the full length protein.

In order to perform a wide and efficient search of the energy landscape for each of the segments, we use replica exchange molecular dynamics simulations (REMD). Molecular Dynamics simulations were done using the CHARMM force-field and an implicit solvent model [34, 42]. In this method, a number of simulations (replicas) are run in parallel over a specified temperature range. In order to sample a variety of conformations without being constrained by energy barriers, replicas at adjacent temperatures are allowed to exchange periodically, with an acceptance probability based on the Metropolis criterion. We obtained 5000 possible conformations for each of the segments. Once equilibrated, 5000 structures are collected at temperature 300K. We combined the segments into one protein using a clustering algorithm to cluster segments for which the overlapping residues share similar conformations. The wide combinatorial space allows us to attach the segments such that a variety of full conformations will be achieved. The final structural library contained ~200,000 structures.

For these preliminary results, we restricted the size of the structural library to limit the computational costs associated with building an ensemble. We performed a pruning routine to ensure that structures within the set are sufficiently different from one another so that the weights can be considered independent from one another. In the pruning process two structures are chosen randomly from the large structural library, if the pair-wise RMSD of the Cα is lower than a chosen cutoff,

Figure 6: Comparison of ensemble derived observables and experiment (x-axis). The y-axis in (A) is to the Cα chemical shifts and in (B) is the RDCs.

Figure 7: Inter-residue contacts with 95% confidence intervals. x-axis=residue number; y-axis=number of residue contacts. The vertical dashed lines delineate the NAC region.

we discard of the structure with the higher energy. We chose the RMSD cutoff to be 19.2Å in order to obtain 304 structures. Application of the BW algorithm to these structures yielded an ensemble that gave excellent agreement with the experimental Cα chemical shifts and the N-H RDCs (Figure 6).
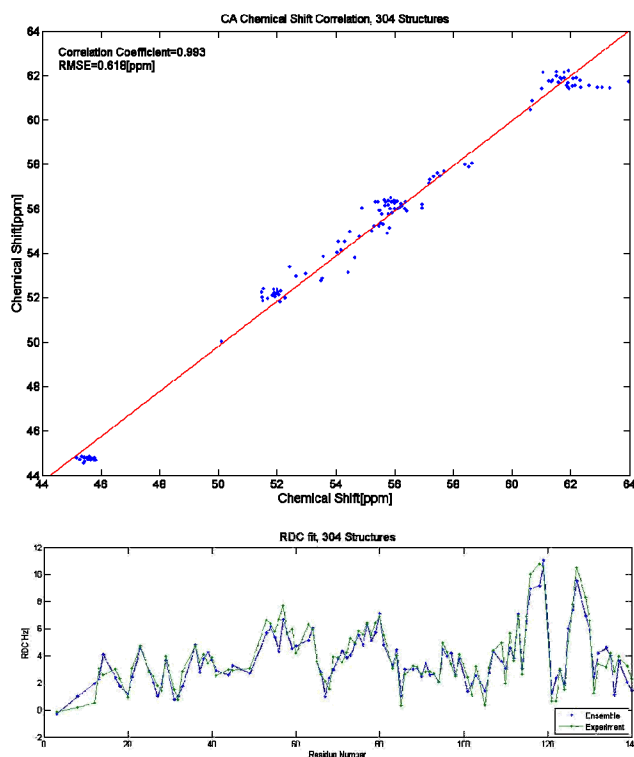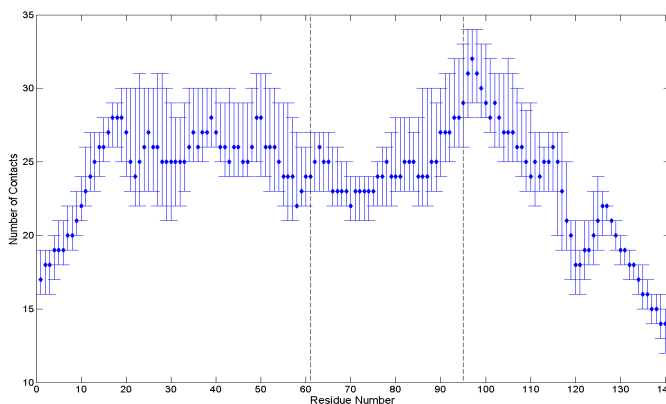
However, since agreement with experiment does not guarantee that the model is accurate, we computed the uncertainty estimate for this ensemble. Since the uncertainty parameter $\sigma_{\bar{W}} = 0.5$, we express calculated parameters from the model along with their associated confidence intervals. For example, figure 7 shows the average number of intramolecular contacts for each residue in the α-synuclein sequence. A contact corresponds to an inter-residue distance within 25 angstroms; i.e., interatomic distances that can be detected with paramagnetic residue experiments [43].

Several things are apparent from Figure 7. First our uncertainty in the number of contacts is less for residues near the N-terminus (1-20) and at the C-terminus (120-140). Furthermore, the number of contacts is greatest in the acidic C-terminal region and in the amphipathic N-terminal region. Moreover, the NAC(8-18) segment (residues 68-78), which plays a role in the initiation of α-synuclein aggregation [15, 16], has fewer contacts relative to the N and C-termini of the molecule. These data are in qualitative agreement with previous results from PRE experiments, which suggest that there are residual long range contacts between the C-terminal region and residues that are N-terminal to the NAC segment [43]. In addition, these results are consistent with the notion that the central NAC(8-18) segment is relatively free of long range contacts and therefore may be more available to form inter-molecular contacts that promote fibril formation.

Figure 8: SDS PAGE of α-synuclein, A30P, E46K and A53T following purification.

Protein Purification and Aggregation Assays:
We have successfully purified α-synuclein and the point mutants A30P, E46K and A53T using a technique refined in our laboratory. Briefly, bacterial cells are harvested by centrifugation and cell pellets are resuspended in resuspension buffer. Resuspended cells are then disrupted by sonication under ice-cold conditions and the resulting solution is boiled for 20 minutes. Given α-synuclein's solubility and extreme heat stability, it remains in solution after boiling whereas most other proteins aggregate. Following centrifugation, the supernatant was loaded onto a weak ion exchange column and eluted in a high salt buffer. Following desalting, the sample was run through a strong anion exchange column. After a subsequent desalting step, the purity of the purified protein was assessed via SDS-PAGE (Figure 8).
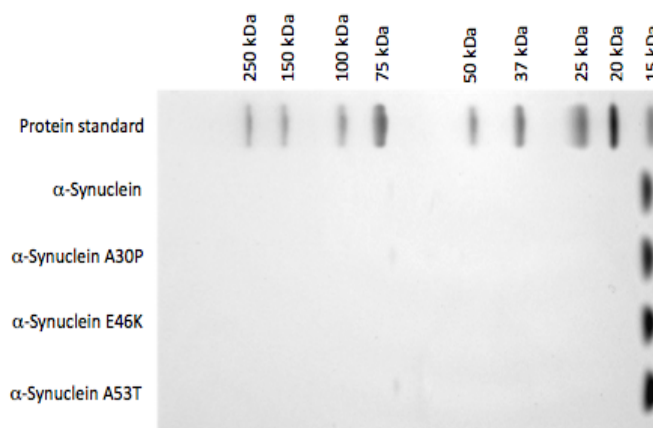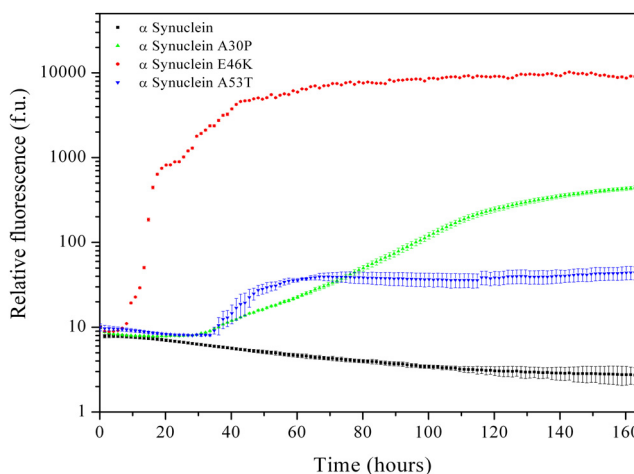

Figure 9: Thioflavin T fluorescence of α-synuclein, A30P, E46K and A53T demonstrating varying rates of amyloid fibril formation.

Following purification we performed a fibril formation assay refined from Li et al. Wild-type α-synuclein and each of point mutants were assayed at a concentration of 2 mg/mL. The solution was incubated at 37°C with shaking in the presence of 25 μM Thioflavin T (ThT). ThT is a dye that fluoresces in the presence of amyloid fibrils. The fluorescence was measured using an excitation wavelength of 450nm with an emission wavelength of 482nm [50]. Using this technique, fibril formation is observed over approximately 1 week (see Figure 9).

α-Synuclein and its mutants can form soluble oligomers in solution that represent non-fibrillar aggregates. Therefore, in addition to the fibril-forming assay described above, we conducted dynamic light scattering (DLS) experiments to gain insights into the size distribution of the samples. An increased amount of larger soluble particles is consistent with increased formation of soluble non-fibrillar aggregates. From the results shown in Figure 10, it is apparent that the point mutants form oligomers at different rates, a result corroborating published data.

We have shown, thus far, that we can successfully purify and characterize wild-type α-synuclein and its variants using a range of techniques. These techniques will be applied to α-synuclein variants containing cysteine residues and consequent disulfide bonds.
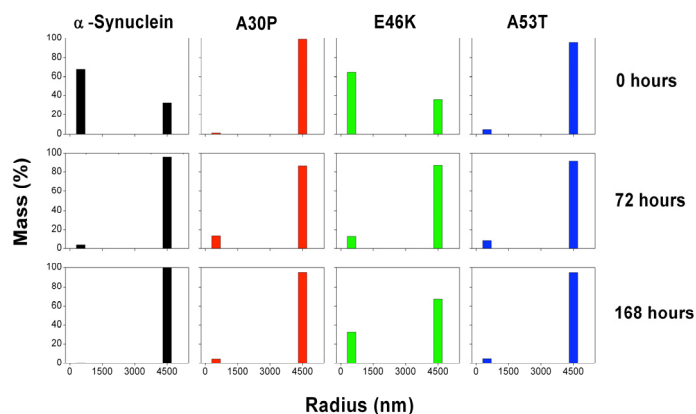

Figure 10: Dynamic light scattering data for α-synuclein, A30P, E46K and A53T exhibiting the different rates of oligomer formation.

3. Experimental analysis of Tau protein

Sponsors:
NIH 5R21NS063185-02

Project Staff:
Sophie Walker, Nebiyat Tsegaye, Joy Okpala, Prof. Collin Stultz

Our objective is to characterize the unfolded state of both wild-type Tau and its aggregation prone variants. By isolating conformers that are associated with increased aggregation, we are working to develop detailed structural models that provide insight into mechanism of Tau aggregation. To achieve this we have produced wild-type Tau, mutant Tau and solutions of Tau that are enriched with potential aggregation-prone conformers. The later has been accomplished through the introduction of disulfide bonds via cysteine residues. The aggregation rate of the proteins in these solutions will be tested *in vitro* and compared to that of wild type.

3.1 Protein Purification and Aggregation Assays.

We have successfully purified the longest naturally occurring isoform of Tau (hTau40) using a techniques refined in our laboratory. Additionally, we have purified a variant of Tau that is associated with increased risk of Alzheimer's disease (Tau ΔK280) and shorter variants of Tau (Tau K18 and Tau K18 ΔK38). The purity of each protein was estimated by densitometry, with the purity of each to be in excess of 90% (see Figure 11.)
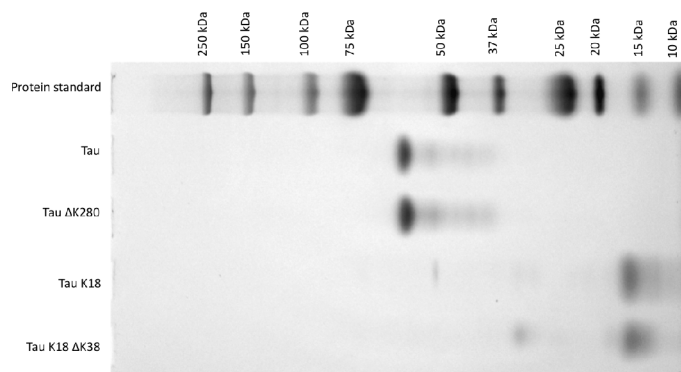
Figure 11: SDS PAGE of Tau, Tau ΔK280, Tau K18 and Tau K18 ΔK38 following purification.

Following purification we performed fibril formation assays refined from Barghorn et al (1). 50 μM solutions of Tau, Tau ΔK280, Tau K18 and Tau K18 ΔK38 were incubated at 37°C with shaking in the presence of 25 μM Thioflavin T (ThT). ThT is a dye that fluoresces in the presence of amyloid fibrils. The fluorescence was measured using an excitation wavelength of 450nm with an emission wavelength of 482nm. Using this technique, fibril formation is observed over 7 days (see Figure 12). The results indicate that the fibril forming capacity of each species is different, congruent with results in the literature.
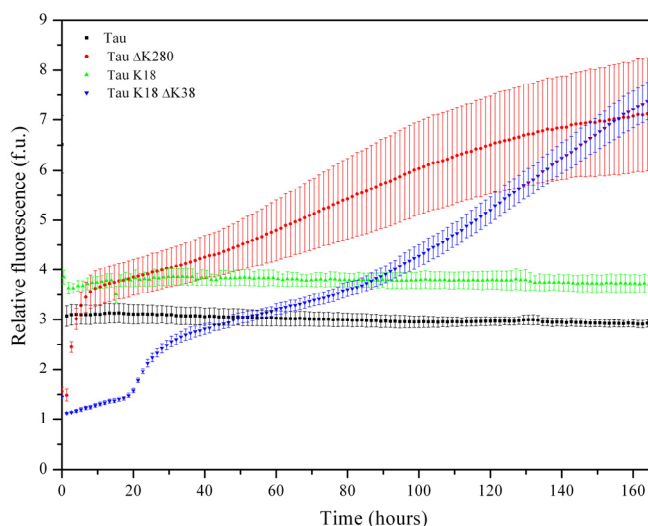

Figure 12: Thioflavin T fluorescence of Tau, Tau ΔK280, Tau K18 and Tau K18 ΔK38 demonstrating varying rates of amyloid fibril formation.

In addition to the fibril-forming assay described above, we conducted dynamic light scattering (DLS) experiments to gain insights into the rate of formation of soluble oligomers formed by Tau and its variants (see Figure 13). The results demonstrate that each variant forms large soluble aggregates at different rates; results corroborated by published data.
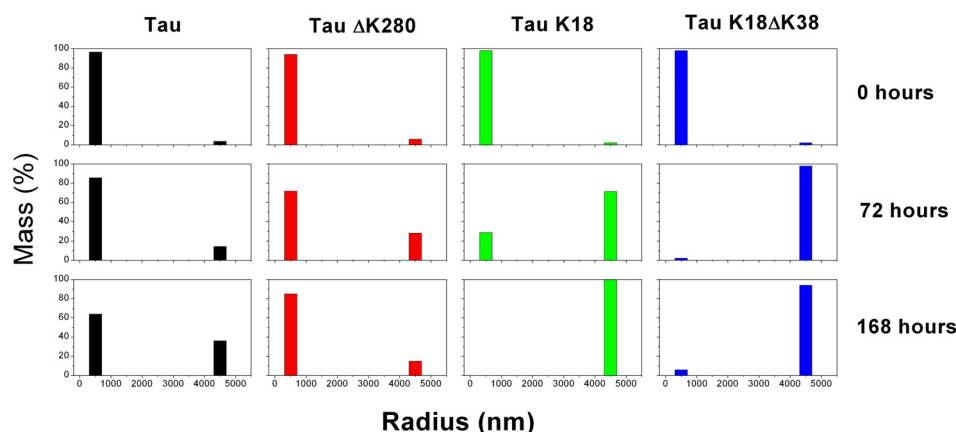
Figure 13: Dynamic light scattering data for Tau, Tau ΔK280, Tau K18 and Tau K18 ΔK38 exhibiting the variable rates of oligomer formation.

We have shown, thus far, that we can successfully purify and characterize Tau and its variants using a range of techniques. These techniques will be applied to Tau variants containing cysteine residues and consequent disulfide bonds described in the following section.

3.2 Generation of solutions of wild type and mutant Tau K18 enriched for conformers that are predicted to exhibit increased aggregation propensity.

Using a computational method called Energy-minima Mapping and Weighting (EMW) that was developed in our laboratory [2,3] we have developed structural ensembles that represent accessible conformations of Tau K18. We have identified aggregation prone conformers in the ensemble and through the introduction of cysteines and the subsequent formation of intramolecular disulfide bonds under oxidizing conditions; we have generated solutions that are enriched for these specific conformers.

We used the algorithm Disulfide by Design (DbD) [4] to identify sites within this structure that could be mutated to cysteines. The positions of these and the resulting disulfide bonds as dictated by DbD are shown in Figure 14.
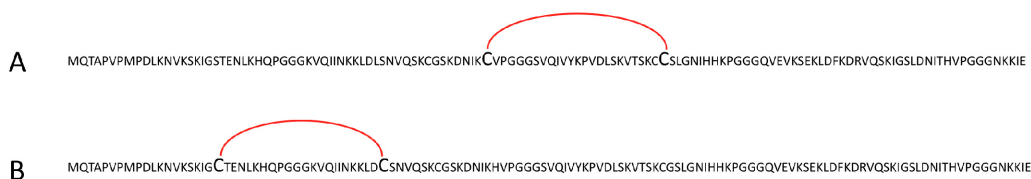

Figure 14: Disulfide bonding patterns as identified by DbD **(A)** Tau K18 H57C G81C and **(B)** Tau K18 S20C L42C.

The methods used to purify Tau were refined for the expression and purification of cysteine containing mutants. The purity of the mutants and their ability to form disulfide bonds was assessed by SDS PAGE (see Figure 15). The results confirmed that under oxidizing conditions both inter- and intramolecular disulfide bonds are formed. Purification of the conformers containing only intramolecular disulfide bond will be carried out prior to characterization of the aggregation propensity, the results of which will be compared to that of wild type Tau and the variants described in section 3.1.
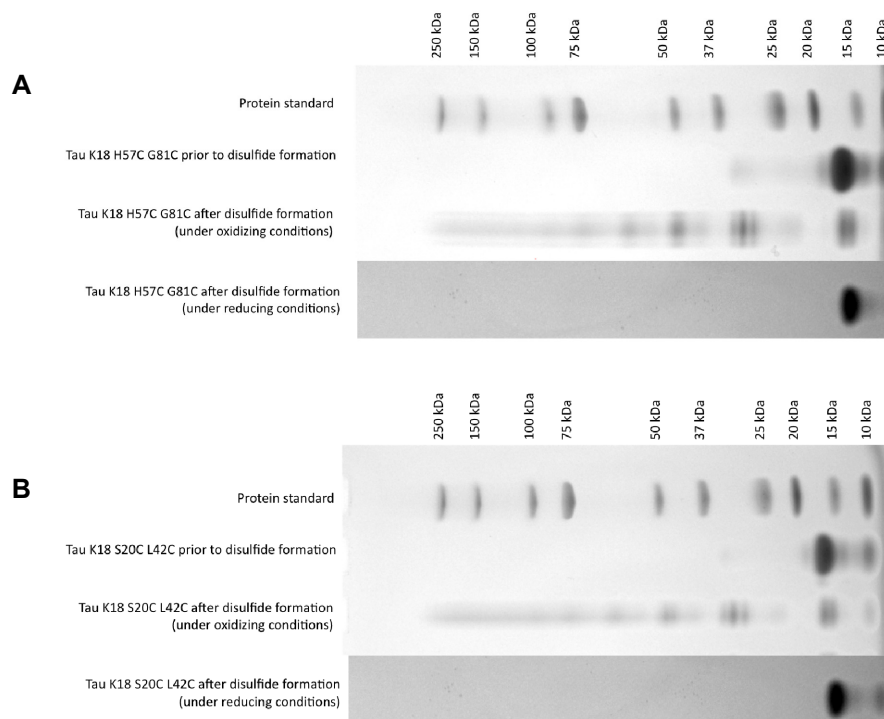
Figure 15: SDS PAGE of **(A)** Tau K18 H57C G81C and **(B)** Tau K18 S20C L42C, demonstrating the successful formation of disulfide-bonded species and the formation of a single band upon subsequent reduction.

If the predicted ensemble has accurately captured the unfolded state of these proteins, then some of these stabilized conformations should exhibit increased aggregation *in vitro*. These results will be used to validate and refine the results of the simulations. Discrepancies between the predicted findings and the experimental results will initiate a re-examination of the models and further improvements.

4. Computational and Theoretical Studies of the Nickel Regulatory Protein NikR

Sponsors:
HST, RLE,

Project Staff:
Sarah Bowman, Prof. Collin Stultz, Prof. Catherine Drennan,

4.1 NikR: A Brief Introduction to the Problem
Transcription factors regulate a wide variety of genes in the cell and play a crucial role in maintaining cellular homeostasis. A major unresolved issue is how transcription factors find their specific DNA binding sequence in the vast expanse of the cell and how they do so at rates that appear faster than the diffusion limit. Here, we relate an atomic-detail model that has been developed to describe the transcription factor NikR's mechanism of DNA binding to the broader theories of how transcription factors find their binding sites on DNA. NikR is the nickel regulatory transcription factor for many bacteria, and NikR from escherichia coli is one of the best studied ligand-mediated transcription factors. For the E. coli NikR protein, there is a wide variety of structural, biochemical, and computational studies that provide significant insight into the NikR-DNA binding mechanism. We find that the two models, the atomic-level

model for E. coli NikR and the cellular model for transcription factors in general, are in agreement, and the details laid out by the NikR system may lend additional credence to the current models for transcription factors searching for DNA.
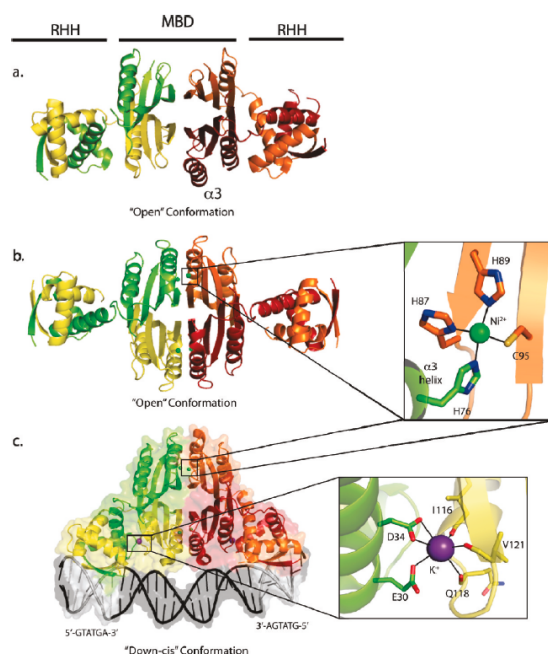

FIGURE 16: Crystal structures of full-length E. coli NikR. (a)Apo-NikRstructurewith disordered R3 helices in an "open" conformation. (b) Ni2þ - boundNikRin an open conformation with ordered R3 helices and nickel ions represented as green spheres and an inset of the high-affinity nickel site (green spheres). (c) NikR-DNA complex with NikR in a "down-cis" conformation with potassiumions (purple spheres) bound between the MBD and RHH domains and an inset view of the potassium site. The NikR-DNA complex is shown in surface representation to illustrate contacts between NikR and DNA, and the DNA motif responsible for specific recognition by NikR is indicated.
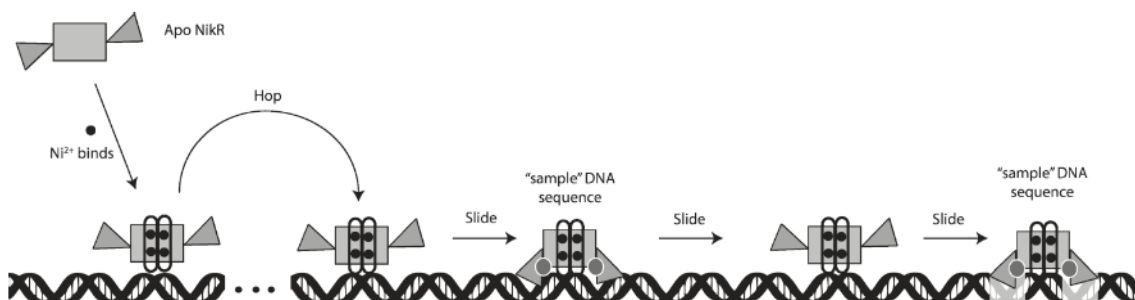

FIGURE 17: Proposed model for NikR's search for the nik operon on DNA. The RHH domains are represented as gray triangles. The MBD is represented as a rectangle. R3 helices are represented as ovals. Nickel ions are represented as black circles and potassiumions as gray circles.DNA is represented as a black double helix. The twoDNAsites highlighted in gray represent the two half-sites just upstreamof the nik operon to which NikR specifically binds in the cell to repress the transcription of nikABCDE, thus indicating the site whereNikRhas the highest affinity forDNA.

The wealth of data arising from crystallographic, biochemical, and computational studies on E. coli NikR supports a sliding diffusion mechanism that employs at least three distinct NikR conformers, with each conformer having a different affinity for the nik operon. When no metal is bound to NikR, the protein can

freely diffuse with no or little affinity [less than micromolar affinity (2)] forDNA.Once nickel ions bind toNikR, the R3 helices and loops preceding thembecome ordered, and this allows the protein to interact with DNA through nonspecific hydrogen bonds to backbone phosphate groups (Figure 16, black dashed lines). In this sense, the one-dimensional search along the DNA sequence does not begin until nickel ions bind at high affinity nickel sites on NikR. However, because nickel binding is not sufficient to induce the RHH domains to adopt a down-cis conformation, the presence of nickel does not ensure that the protein will recognize the specific nucleic acid sequence in the major groove of DNA. The nickel-bound structure of NikR with the RHH domains in an open conformation therefore represents a nonspecific binding conformation of NikR, whose localization on DNAis sequence-independent (Figure 17). While we do not have a good experimental number for the affinity of such a nonspecific interaction between Ni2þ-bound NikR and DNA, because it is below the detection level of the standard gel shift assay an estimate for the nonspecific DNA affinity can be obtained from base substitution experiments in which bases in the nik operon binding motif are individually substituted. Individual mutations can lead to significant reductions in affinity (>1000-fold).

5. Understanding Fibronectin Fibrillogenesis by Modeling Physiologic Unfolding of Fibronectin with Steered Molecular Dynamics Simulation

Project Staff:
Elaine Gee, Prof. D.E. Ingber, Prof. Collin Stultz

5.1 Cell-Mediated Fibronectin Unfolding

When a particular Fibronectin (FN) polymerization is essential for the development and repair of the these computational simulations.

The extracellular matrix (ECM) controls cell behavior by affecting adhesion, migration, gene expression, proliferation, and survival through biochemical and mechanical cues. The mature ECM develops from a provisional fibronectin (FN) matrix, which other ECM molecules, such as collagen, require for deposition. The process of FN assembly, or FN fibrillogenesis, plays an important role in ECM formation and function during the development and repair of the extracellular matrix in physiological processes such as embryogenesis, migration, angiogenesis, and wound healing as well as in pathological conditions as in cancer. The process of FN assembly requires both integrin binding to FN as well as cell contractility, which confers strain to the attached molecule. The 10th repeat of FN domain type III (10FNIII) contains the integrin-binding sequence RGD and has been shown to be mechanically weak. However, the role of this module in the molecular process of fibrillar assembly remains unclear. By studying the mechanism by which FN transduces cell-traction forces into the biochemical process of matrix assembly, one can understand the mechanical principles that define the cell microenvironment. The goal of this project is to understand how structural unfolding of 10FNIII by traction force affects FN mechanotransduction.

Data suggest that FN fibrillogenesis is facilitated by cell traction forces (on the order of nNs) transferred to FN via bound membrane integrin receptors. Forces applied to distinct regions of FN drive the formation of partially unfolded FN intermediates that then bind to additional FN molecules. Past studies have focused on FN unfolding by applying pulling force at the N- and C-termini of protein fragments rather than at the site of cell binding. However, since cell membrane integrins bind FN at a surface-exposed RGD loop on 10FNIII, we investigate the mechanism of cell-mediated FN unfolding by using steered molecular dynamics to apply tensile force at the RGD site. While pulling at the N- and C-termini of 10FNIII leads to unfolding trajectories that sample different unfolding pathways as previously published, we find that

pulling at the RGD site unfolds 10FNIII along one well-defined pathway. This pathway accesses a partially unfolded intermediate structure with solvent exposed hydrophobic N-terminal residues. Moreover, additional unfolding of this intermediate triggers an arginine residue, which is required for high affinity cell binding, to separate from the adjacent RGD loop. Based on these results, we propose a mechanism by which cell traction induces a conformational change in FN that possibly enables FN-FN associations and cell detachment, thereby promoting the binding of new FN to the cell surface receptor to promote subsequent cycles of adhesion, unfolding, and fibril formation that drive FN matrix assembly (Figure 18). Currently, investigations are underway to provide experimental support for the proposed mechanism obtained from these computational simulations using biophysical approaches including force spectroscopy techniques and cell-based fluorescence assays.
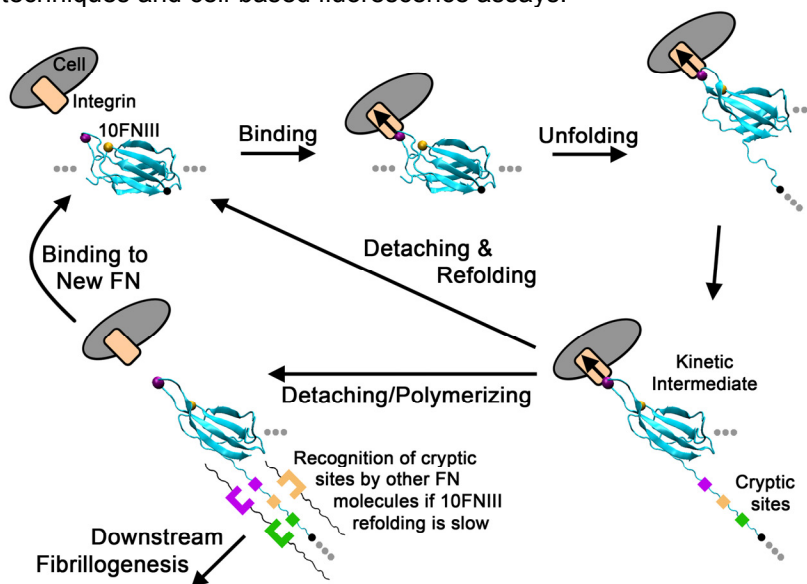

Figure 18. Cell-mediated Fibrillogenesis. The proposed model involves the formation of a partially unfolded intermediate during cell-traction mediated unfolding. Potential cryptic sites that promote FN aggregation and fibrillogenesis may exist in this partially unfolded structure. Subsequent unfolding past this intermediate induces the movement of a crucial arginine (gold bead) at a secondary binding site to a position far from the RGD loop (magenta bead labels the central glycine). This frees the cell from the RGD ligand and enables it to reattach to another FN molecule to propagate the process. If 10FNIII refolding is relatively slow, then unfolded portions of this domain are available to make contacts with other FN molecules to stimulate downstream fibrillogenesis.

6. ECG Markers to Predict Cardiovascular Death

Sponsors:
HST, RLE, Texas Instruments

Project Staff:
Priya Parayanthal, Prof. Collin Stultz, Prof. John Guttag

6.1 ECG Measures to Predict Cardiovascular Outcomes

Patients who have had an acute coronary syndrome (ACS) are at higher risk for developing future adverse cardiovascular events such as cardiovascular death, congestive heart failure, or a repeat ACS. Several methods and metrics developed to assess the risk of ACS patients for a future adverse cardiovascular outcome, including heart rate variability, heart rate turbulence, deceleration capacity, T-wave Alternans, and Morphologic Variability. We aim to determine whether employing machine learning

methods can achieve even greater success in risk stratification of ACS patients for cardiovascular death into high risk and low risk groups

A maximum margin linear classifier is used to separate patients into high risk and low risk groups. The ECG signal is converted into the frequency domain to form the power spectrum and SVM's with a Linear Kernel is applied to this power spectrum. To mitigate the effects of outliers in the training data as well as address the fact that the data may not be linearly separable, slack is introduced into our optimization problem which allows for violations of the margin constraint and even misclassifications. A cost factor is introduced to account for how much training errors on positive examples outweigh errors on negative examples. In classifying cardiovascular death, classifying a patient who actually had cardiovascular death as low risk is much more costly than classifying a patient who did not have cardiovascular death as high risk.

The linear support vector machine relaxed quadratic programming problem that is solved is

$$\min \frac{1}{2}\|\theta\|^2 + C\sum_{i=1}^{n} \xi_i \quad \text{subject to}$$

$$y_i(\underline{\theta} \cdot \underline{x}_i + \theta_0) \geq 1 - \xi_i \quad \text{where } i = 1,...,n$$

$$\xi_i \geq 0 \text{ where } i = 1,...,n$$

and the final classification function is: $\quad f(\underline{x};\underline{\theta},\theta_0) = sign(\underline{\theta} \cdot \underline{x} + \theta_0)$

The decision boundary essentially divides the space into a positive and negative half and this classification function assigns an example to one of those halves depending on which side of the decision boundary it falls.

The results show that the Linear Kernel outperforms the existing morphologic variability metric based on odds ratio. The following graph shows the results of the Odds Ratios resulting from using the morphologic variability metric versus the Linear SVM across the 10 cross validation trials.
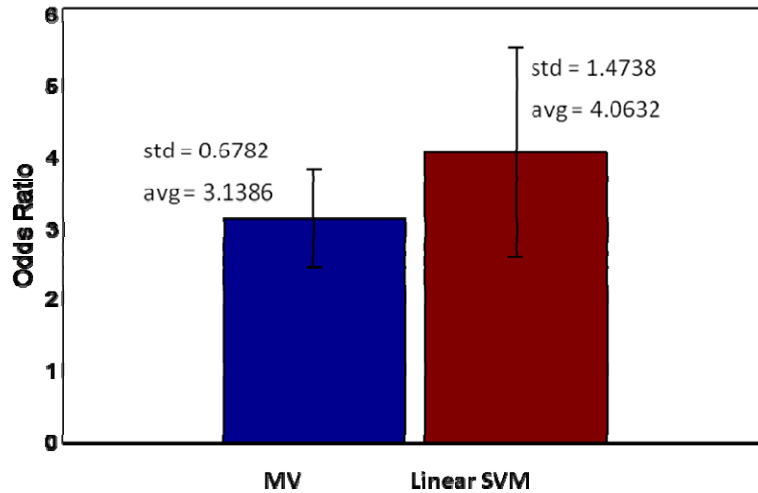

Figure 19: MV versus Linear SVM Odds Ratio averages and standard deviations across 10 test/train trials

Publications

Journal Articles, Published

1 Syed Z., **Stultz CM.** Kellis M., Indyk P. Guttag J. Motif Discovery in Physiological Datasets: A Methodology for Inferring Predictive Elements. *ACM Transactions on Knowledge Discovery from Data* 4(1), Article 2, 1-23, 2010.**

2 Salsas-Escat R., **Stultz CM.** Conformational Selection and Collagenolysis in Type III Collagen. *Proteins: Structure, Function, and Bioinformatics* 78, 325-335, 2010.**

3 Salsas-Escat R., Nerenberg P., **Stultz CM.** Cleavage site specificity and conformational selection in type I collagen degradation. *Biochemistry,* 49, 4147–4158, 2010.**

4 Gurry T., Nerenberg P., **Stultz CM.** The Contribution of Inter-chain Salt Bridges to Triple Helical Stability in Collagen. *Biophysical Journal* 98 (11), 2634-2643, 2010.**

5 Phillips C., Schreiter E., **Stultz CM.** Drennan C. Structural Basis of Low Affinity Nickel Binding to the Nickel-Responsive Transcription Factor NikR from Escherichia coli. *Biochemistry* 49, 7830-7838, 2010.**

6 Phillips C., **Stultz CM.**, Drennan C. Searching for the nik operon: how a ligand-responsive transcription factor hunts for its DNA binding site. *Biochemistry* 49, 7757-7763, 2010.**

7 Fisher CK., Huang A., **Stultz CM.** Modeling Intrinsically Disordered Proteins with Bayesian Statistics. *Journal of the American Chemical Society* 132, 14919-14927, 2010.**

8 Chen MM., Bartlett AI., Nerenberg PS., Friel CT., Hackenberger CPR., **Stultz CM.**, Radford SE., Imperiali B. Perturbing the folding energy landscape of the bacterial immunity protein Im7 by site-specific N-linked glycosylation. *Proceedings of the National Academy of Sciences*. 107 (52) 22528-22533, 2010.**

Journal Articles, Submitted

Judy S. Liu1, Christian R. Schubert1,2,4, David J. Tischfield1,2, Andrew E. Baltus1,2, Collin M. Stultz3,4 and Christopher A. Walsh1,2 , Kinesin1a (Kif1a) mediated vesicle transport requires doublecortin (DCX)-microtubule binding (under review).

Conference Abstracts:

1. Syed Z., Scirica BM., Morrow DA., **Stultz CM.**, Guttag JV. Association of Morphologic Variability with Ventricular Tachycardia and Ventricular Pause in Patients Treated with Ranolazine or Placebo Following non-ST-Elevation Acute Coronary Syndrome. American College of Cardiology, 1 page, March 2010.**

2. Syed Z., Scirica BM., Morrow DA., **Stultz CM.**, Guttag JV. Heart Rate Turbulence and Deceleration Capacity Are Associated with Myocardial Infarction Following Non-ST-Elevation Acute Coronary Syndrome. Heart Rhythm Society Scientific Sessions, 1 page, May 2010.[1]*

3. Syed Z., Scirica BM., Mohanavelu S., Morrow DA., **Stultz CM.**, Guttag JV. Association of Heart Rate Turbulence, Deceleration Capacity and Morphologic Variability with Sudden Cardiac Death Following Non-ST-Elevation Acute Coronary Syndrome: Results from the MERLIN-TIMI 36 Trial. AHA Scientific Sessions, 1 page, Nov. 2010.**

---

[1]* Outgrowth of Supervised Student Research