

A. Background

I have research experience in; percolation theory, diffusion and electrical conduction in disordered media, simulation of diffusion limited aggregates, spinodal decomposition in quenched binary phase fluids, fluctuation driven long-time (material) domain growth dynamics, dynamical chaos, cellular automata models for fluids at low Reynolds numbers, mechanical stability in polymeric and amorphous chalcogenide glasses, generic rigidity, graph-rigidity algorithms, protein conformational flexibility and protein stability.

B. Research Interests

General Directions: My research is best described as a mix of statistical and computational physics with algorithm development. My broad interests attract me to the interface of interdisciplinary research in the area of model building that optimizes for tractability, accuracy and computational efficiency. Although I do not carry out experiments in the laboratory, I believe it is essential to work directly with experimental data (and work closely with the people actually doing the experiments!) to truly develop the best models and assess how well the calculations and/or simulation results compare with observations in the real world.

Current Research: I am currently developing novel computational methods and statistical mechanical models to accurately predict in practical computing times, thermodynamic stability, conformational flexibility, kinetics and long-time dynamics of proteins and polypeptides. The data that is produced is overwhelming, and all these attributes are important to biological function, but not as independent entities. Therefore, we go one step further, and associate these physical properties together in a coherent way by developing quantified stability/flexibility relationships (QSFR). The information contained within the QSFR descriptors account for a wide range of thermodynamic and solvent conditions. We then take this bio-chemical-physical information and construct a QSFR database for proteins in aqueous solution. This process is a bioinformatics project in itself. However, we go further than this, and through collaborations, we wish to apply bioinformatics investigations using the QSFR database, and apply this knowledge to better understand protein function. In order to accomplish all these objectives, we are developing new software that will become user-friendly tools for making QSFR predictions. We have working-prototype models currently, but are undertaking a major improvement of the model, and are on the ground floor of building the QSFR database. The expected target date to have the software tools and database nearly bug free and available for free academic use is set for 2010 following a 5-year research plan. However, through collaborative efforts in beta testing, etc, we will have limited accessibility starting NOW.

Future Objectives: The computational methods my group has/is developing are general but optimized for the problem at hand. There are always new issues that arise when shifting to different problems. I wish to work on extending our biophysical models to accurately predict physical-chemical properties of small molecules, nucleic acids, supramolecules and condensed phases of water. Develop especially fast algorithms to simulate long-time dynamics in biopolymers that will help better understand their biological function. From applications point of view in our post genomic era, the problems related to rational design of proteins and polypeptides, and other biopolymers will require high-throughput bioinformatics research. To successfully complete these tasks will require non-traditional, innovative approaches in modeling.

C. Applications

1. Rational protein design, predict protein-substrate association and identify functional allosteric effects
2. Build accurate homology models, sequence and structural alignment tools
3. Develop faster ensemble-based electrostatic calculation methods for pKa and stability predictions
4. Protein folding predictions that include thermodynamic and mechanical properties of the transition state
5. Prediction of protein oligomerization states that accurately describe protein-protein interactions

D. Biophysical Modeling, Methodology and Algorithms

1. **Distance Constraint Model (DCM):** Since the late 1950s with the introduction of Ising-like models to explain the alpha-helix to coil transition, there has been a desire for *reduced protein models* to capture the essential physics of protein folding. Also important, are related issues pertaining to protein stability and molecular cooperativity that need to be accurately modeled so that improved functional efficiency can be engineered by computer aided design in fast turn-around times. To this end, it is desirable to have a simple model that can balance energy and entropy among molecular constituents, while representing solvation effects in terms of enthalpy and entropy component contributions. A common approach that can satisfy the *simple* and *fast* criteria has been to invoke free energy decomposition schemes. Unfortunately, models based on the additivity assumption of free energy components become inaccurate when component parts within a protein or associated with protein-solvent interactions strongly couple. These long-standing problems are addressed using a *Distance Constraint Model (DCM)* that is computationally tractable in practical computing times (D.J. Jacobs, et. al. 2003, Phys Rev E 68, 061109). The DCM is based on free energy decomposition, and invokes network rigidity as an underlying mechanical interaction to accurately account for non-additivity in component entropies. The DCM generalizes network rigidity calculations used in FIRST [see Jacobs, et. al., Proteins, 44 (2001)] to take into account temperature.
2. **Rigidity Propagation:** By studying the helix-coil transition in polypeptides the DCM was compared to the classic Zimm-Bragg and Lifson-Roig Ising-like models. The exact DCM calculations are difficult because rigidity at the beginning of a chain **non-locally depends** on the presence or absence of hydrogen bonds further down the chain. The approach used was based on the generalization of prior work on propagating rigidity down a Bethe Lattice (P.M. Duxbury, et al., 1999 Phys. Rev. E 59, 2084-2092). Here, interlacing loops formed by backbone hydrogen bonds creates much more difficulty. With A. Heckathorne, a Masters student, we developed recursion relations that propagate rigidity down a homogeneous polypeptide. This was a key step for constructing a novel transfer matrix method.
3. **Transfer Matrix Method:** Rigidity propagation was successfully formulated as a transfer matrix problem (D.J. Jacobs, et al., 2003 Phys. Rev. E 68, 061109) for homogeneous polypeptides. Accounting for the long-range interactions was accomplished in a similar way as in my early work with random walks in disordered media (D.J. Jacobs and H. Nakanishi, 1990 Phys. Rev. A 41, 706-719) where geometrical long-range correlations are modeled as a Markov chain problem by expanding the state-space. Subsequently, the transfer matrix method was generalized to heterogeneous polypeptides with M. Lee, a Masters student, and Greg Wood, a post doctoral researcher (M.S. Lee, et al., 2004 J. Phys. Cond. 16 S5035-S5046).
4. **Hybrid Monte Carlo - Mean Field Calculation:** The DCM requires calculating a partition function over all distinct constraint topologies. Because of the astronomical number of accessible states, the calculation is restricted to constraint topologies that are obtained by perturbing away from the native state. A mean-field approach was introduced to calculate free energy landscapes in a reduced 2-dimensional constraint space (D. Jacobs and S. Dallakyan, Biophysical Journal, 88, 1-13 (2005). Adoptive grid methods were employed to optimize calculations. A product probability measure is used, in similar way as used in my prior work with Lattice Boltzmann approximations while in the group of Matthieu Ernst (D.J. Jacobs and A. Masters, 1994 Phys. Rev. E 49, 2700-2710). Conformational entropy is determined within the DCM using generic rigidity calculations based on methodologies I developed in the group of Mike Thorpe and in collaboration with Bruce Hendrickson. (US patent # 6014449, 2000).
5. **Developing Transferable DCM-Parameters:** With Dennis Livesay (a physical chemist at Cal Poly Pomona) we are extending the DCM to explicitly account for several additional physical effects using a larger set of transferable parameters. Over the next five years, we plan to incorporate:
 - i) Explicit modeling of hydrophobic interactions
 - ii) Explicit modeling of Van der Waal or packing interactions
 - iii) Decomposing backbone and sidechain properties
 - iv) Distinguishing between residue types

Our more sophisticated parameterization will initially retain a small number of phenomenological adjustable parameters. Eventually, we wish to eliminate all fitting parameters by finding parameter dependencies on solvent conditions. This work requires careful consideration of the essential interactions governing protein stability. An important factor affecting protein stability is the long-range electrostatic interaction, which is sensitive to solvent conditions (i.e. pH and ionic strength).

6. **Hybrid DCM-Electrostatic Calculations:** We wish to account for shifts in protein stability due to changes in electrostatic energy depending on pH and ionic strength. At present, we perform Poisson-Boltzmann (PB) continuum calculations on static structure to obtain a better representation of the hydrogen bond network using the most probable protonation state. We have preliminary data suggesting a change in free energy between the native and unfolded states due to electrostatic contributions is additive. We interpolated this change as a linear function of a global flexibility order parameter, similar to how a magnetic field couples to the magnetization order parameter in Landau theory applied to a magnetic spin system. We initiated the combining of the PB-electrostatic and DCM calculations and this work requires to be funded, but is in progress currently. The current/future challenge is to devise a tractable, fast and accurate computation.
7. **Thermodynamic Properties of Hydrogen Bonds:** A model to calculate local thermodynamic properties of a hydrogen bond was constructed by using a single particle Schrodinger equation. An effective potential is used where electron charge transfer is accounted for using group inductive electronegativity plus a self-energy Born term to calculate partial charge on atoms (D. Karabay, et al., to be submitted to J. Amer. Chem. Soc.) The calculated local enthalpies and entropies will be incorporated into the DCM. Moreover, future work would entail developing better models for the "QSAR"-electronegativity hardness parameters.
8. **Hybrid DCM-GNM analysis:** During summer of 2004, my group has defined a hybrid model combining DCM with a new type of Elastic Network Model that is distinctly different in scope by those implemented by Bahar and Jernigan. Our new method tracks vibrational modes as a function of thermodynamic condition and the macrostate of a protein described by its global flexibility order parameter. Current emphasis is to identify differences in normal mode vibrations between the native and transition states, and to study biologically relevant dynamics in the native fold. Shelley Green (a MS student at CSUN is working on this project).
9. **Future work on Helix-coil Transition:** A C++ helix program was made to calculate α -helix to coil transitions. The transfer matrix will be extended to include 3-10 and π -helical states to investigate phase diagrams and make accurate prediction of stability curves. Further generalization of the matrix representation will allow study of the folding kinetics. A web-based tool will be created to provide predictions similar to AGADIR, but will take better account of solvent effects.
10. **Future work on Multiscale Modeling:** QSFR-results generated from the DCM will be used to produce novel methods to efficiently simulate long-time motions in proteins and other biopolymers. Preliminary ideas about constructing a hybrid method combining DCM with continuum elasticity theory to study motion in membrane bound proteins. This work will help improve understanding of the roles played by protein and membrane dynamics that together control opening/closing of ion-channels, and transport through those channels.

E. Applications to Physical Chemistry and Biochemistry

1. **Hypothesis that Network Rigidity Provides a Ubiquitous Nucleation Mechanism:** In publication D.J. Jacobs, et al., 2003 Phys Rev E 68, 061109, we show the nucleation mechanism is an outcome of the formation/breaking of rigid clusters. The nucleation process is problem specific --- dependent on system size, and the type and topology of cross-linking constraints. To date, all our work supports this hypothesis, but much research will be required to firmly demonstrate it. To this end, I plan to work with many experimentalists to obtain large sets of systematic experimental data. Initial collaborations

are currently underway with Maria Luisa Tasyaco (City College of New York), Nick Pace (Texas A&M University) and Frank Robb (Univ. Maryland Biotech Institute).

- Solvent Effects: Cold and Heat Denaturation:** The DCM accounts for hydration effects and other solvent effects, and thereby is able to accurately model both cold and heat denaturation (D. Jacobs and G. Wood, 2004 Biopolymers 75 1-31). In dilute solution the DCM-parameters are expressed as linear functions of solute-concentration, which is less restrictive than the “m-value” description in physical chemistry to predict thermodynamic stability as a function of mixed solvent conditions. This work is important to industrial protein engineering and to understand protein stability in the cell.
- Heat Capacity in Proteins:** The minimal DCM applied to proteins reproduces essential features of heat capacity measurements well (D.R. Livesay, et al., 2004 FEBS Letters 576, 468-476). No other computational method reproduced heat capacity curves in proteins to such a degree. After 3 free parameters are obtained by fitting, protein flexibility is quantified as a function of thermodynamic condition. Based on the suggestion of George Makhatadze, we will solve the inverse problem by optimizing a few free parameters so that multiple homologous proteins will have similar flexibility characteristics. If the common dogma that homologous mesophilic and thermophilic proteins exhibit similar flexibility characteristics at their respective operating temperatures is true, then we might be able to predict the transition temperatures and heat capacity curves using the minimal DCM. Some preliminary work in this direction can be found in: D.R. Livesay and D.J. Jacobs, *Conserved quantitative stability/flexibility relationships (QSFR) in an orthologous RNase H pair*. Proteins: Structure, Function, and Bioinformatics, **In Press**
- Free Energy Landscapes and Protein Folding:** After fitting to heat capacity, we directly obtain free energy landscapes in constraint space and in terms of a global flexibility order parameter. Many proteins show two-state behavior indicating a first order transition. However, we find proteins that exhibit behavior best described as a continuous phase transition and these cases match well with experimental evidence. Although we are effectively describing protein folding/unfolding, the model currently relies on perturbing away from the native state using a known 3D structure. In the future, the kinetics of protein folding/unfolding will be investigated in more detail. A critically important first step will be to develop simple but accurate models for describing the unfolded (disordered) state.
- Future Work on Small Molecules:** We have preliminary results describing observed thermodynamic properties of individual amino acids using the DCM. Since the DCM is based on a general premise, it will be applied to small flexible molecules important for catalytic chemical reactions.

F. Computational Biology

- Quantitative Stability and Flexibility Relationships (QSFR):** The DCM provides a means to calculate thermodynamic stability and mechanical characteristics in calculation times that are $\sim 10^{10}$ times faster than traditional molecular dynamics. The method would serve as a tool for high throughput applications for the identification of a set of site mutations that change the thermodynamic conditions required for function, while maintaining essential molecular cooperativity. Distributing software and constructing a web-based database are part of current proposals that are high priority elements necessary to move into protein design applications.
- Comprehensive Bioinformatic Studies:** Based on general physical reasoning, we demand the DCM to be valid over a wide spectrum of applications. Therefore, it is imperative to include comprehensive bioinformatic studies that directly apply DCM predictions, and have components that are independent of the DCM. In collaboration with Dennis Livesay on joint projects currently underway, we are testing the hypothesis: “Quantitative stability/flexibility relationships are conserved across protein families” using comprehensive bioinformatic protocols. This work will also lead our group to be involved in high performance computing to scale up our calculations and take advantage of massive parallel computing.

3. **Future Work on Hidden Markov Models Using QSFR:** Sequence and structural alignment tools will be developed based on QSFR. This future work will take place once the methods discussed above settle down and accuracy in predictions is benchmarked through comprehensive bioinformatic studies.

F. Summary

My planned research heavily involves the Distance Constraint Model recently developed in my lab. The focus problems listed above are diverse in breadth, yet they are all connected through my unique expertise in one core area that spans different disciplines. There are always risk factors involved in research, such as relying on one methodology. However, the degree of multiple successes recently made with the DCM indicates that this risk factor is reasonably low within the next 6-years. In this time span, my research will be focused on the DCM and its associated hybrid-parts. Beyond 5 years, if the DCM falls wayside to better methods, than I am readily able to adapt to new ideas (and approaches), and I will use the best computational methods available.