Statement of my research plans and teaching interests

Genome sequencing projects provide the linear sequence information on proteins, but the understanding of the biological roles of these proteins requires knowledge of their structures. My current research involves deducing and predicting three-dimensional structures of proteins from the linear sequences based on high-throughput computational methods. Recently, we have developed a new Threading-ASSEmble-Refinement (TASSER) methodology [1], which, for the first time, can consistently draw the initial template structure identified by threading techniques [2] closer to its native structure. In the large-scale benchmark test of 2,234 representative proteins deposited in Brookhaven Protein Data Bank (PDB), TASSER can generate correct folds for 2/3 of cases without using homology information [1,3]. In recent applications, the TASSER has successfully generated all structures in the genomes of *M. genitalium*, *S. cerevisiae*, *E. coli*, as well as all G protein-coupled receptor proteins in the human genome. These results demonstrate not only that TASSER can provide more accurate structures than existing methods, but also that the algorithm can be exploited on a large scale for structural genomic studies. In recent 6th CASP experiment, a community-wide blind test/assessment of protein structure prediction methodologies, the models predicted by TASSER are consistently better than the best prediction by automated servers in all 26 targets/domains released as of October 10, 2004.

De novo protein design. As the reverse process to protein structure prediction, designing a new sequence to give a novel protein of desired structure and function is one of the most crucial challenges to contemporary biological and computational approaches on protein folding problems. It also provides potential opportunities for developing new drugs and therapeutics. Although considerable progress has recently occurred, most previous protein design methods, in general, have not been used to create new protein structures but rather redesign naturally occurring proteins so that they have enhanced stability or new functionality. The bottleneck of the de novo protein designs is the low accuracy of the free energy potential for describing specific protein sequence-structure pairs. One of my further research plans will be to exploit the TASSER force field to guide de novo computer-based protein design. As an important advantage of the TASSER approach in protein structure prediction, the TASSER force field is a combination of a variety of knowledge-based resources/threading-based restraints and has been systematically optimized by a large mount of protein structure decoys [4]. Our large-scale PDB benchmark test shows that TASSER can generate correct folds from scratch for about 70% of small proteins (40~100 residues) without using threading restraints [1]. This demonstrates the unique ability of TASSER force field that makes it possible to successfully redesign entirely new proteins at least up to 100 residues long. Rather than spending laborious and tedious human checking and inspections, with the help of quick Monte Carlo searching techniques [5], this approach can be automated and used for large-scale screen and design of novel proteins.

High-throughput atomic structure modeling. In principle, the atomic-level simulation could provide high-resolution description of the protein structure and the interaction potential. It is, however, computationally too expensive to be used in genome-scale structure calculations. In contrast, the reduced models, like TASSER, are readily

used for high-throughput structure modeling. However, the modest resolution partially limits its utilization in biotechnological applications, e.g. enzyme active size identifications, and ligand-protein binding recognitions. As the second part of my further research plans, which will be implemented in collaboration with Jeffrey Skolnick's group, we will incorporate the full-atomic refinement procedure into the TASSER modeling approach. The idea is to use the combined atomic potential and reduced force field to improve the accuracy of TASSER modeling. The goal is to generate high-resolution atomic models directly from the fast and high-throughput TASSER modeling, which could be used for recognizing enzyme active sites and for predicting protein-DNA interaction specificity.

Spider capture silk simulation. The third part of my future research plans will involve the mechanical investigations of biological macromolecules in living cells. In previous works, we constructed mechanical models for the successful understanding of entropic elasticity of the DNA/RNA and protein molecules under external force and torque [6-8]. Inspired by the recent observation of the striking exponential force-extension characteristics in spider-silk stretching experiments [9], I will work on constructing new dynamic models for the understanding of the mechanics of spider capture silk. Spider silk is a natural material produced by orb-web weaving spiders that has a high tensile strength comparable to steel; but unlike steel, it is also extremely resilient with the ability to be stretched to 10 times without breaking. One of the appealing dreams of material scientists is to understand the enigmatical mechanism of spider silk and produce analogs of synthetic silks in yeast and bacteria for all kind of construction purposes. Our preliminary analytical calculations [10] show a simple hierarchical chain model can naturally reproduce the exponential force-extension experimental data. In the next step work, we will simulate in real time the dynamic process and the elastic response of the spider silk molecules under the external stretching based on the hierarchical chain model. This effort will provide a route to solving the puzzle of what mechanism leads to the high tenacity and extensibility of the spider capture silk, which will also lend insights to the genetic design of special proteins like spider silk molecules as mentioned in the first project.

Though my career interests lie mainly in performing high quality research in physics, biophysics, and bioinformatics, I also genuinely enjoy teaching and understand the fundamental necessity of building solid undergraduate and graduate programs. I have myself benefited and learned from the guidance and advice by others over the years, and I would like to give the same kind of support to my own students. In the effort to train aspiring scientists and to educate students, I hope that I will be able to incorporate teaching into my research projects in a way that is advantageous to the department, my research, and my students' education.

Refereces:

1. Zhang Y, Skolnick J. Automated structure prediction of weakly homologous proteins on a genomic scale. Proc Natl Acad Sci U S A 2004;101:7594-7599.
2. Skolnick J, Kihara D, Zhang Y. Development and large scale benchmark testing of the PROSPECTOR 3.0 threading algorithm. Protein 2004;56:502-518.

3. Zhang Y, Skolnick J. Tertiary structure predictions on a comprehensive benchmark of medium to large size proteins. Biophys J 2004;87:2647-2655.
4. Zhang Y, Kolinski A, Skolnick J. TOUCHTONE II: A new approach to ab initio protein structure prediction. Biophys J 2003;85:1145-1164.
5. Zhang Y, Kihara D, Skolnick J. Local energy landscape flattening: Parallel hyperbolic Monte Carlo sampling of protein folding. Proteins 2002;48:192-201.
6. Zhou H, Zhang Y, Ou-Yang Z. Bending and base-stacking interactions in double-stranded DNA. Phys Rev Lett 1999;82:4560-4563.
7. Zhang Y. Approach to Monte Carlo calculation of the buckling of supercoiled DNA loops. Phys Rev E Stat Phys Plasmas Fluids 2000;62:R5923-5926.
8. Zhang Y, Zhou H, Ou-Yang Z. Stretching single-stranded DNA: interplay of electrostatic, base-pairing, and base-pair stacking interactions. Biophys J 2001;81(2):1133-1143.
9. Becker N, Oroudjev E, Mutz S, Cleveland JP, Hansma PK, Hayashi CY, Makarov DE, Hansma HG. Molecular nanosprings in spider capture-silk threads. Nat Mater 2003;2(4):278-283.
10. Zhou H, Zhang Y. A hierarchical chain model of spider capture silk elasticity. Phys Rev Lett 2004:Submitted.