November 15, 2003

Prof. Rob de Ruyter van Steveninck
Biocomplexity Faculty Search Committee
Biocomplexity Institute, Indiana University
Swain Hall West 117, Bloomington IN47405-7105

Dear Sir,

I am writing to apply for the position that you advertised in the November 5 issue of *Nature*. I am a research associate at the Computer Science Department of Dartmouth College. I am very interested in continuing my research on computational structural biology at your institute.

For the past ten years I have been conducting research in three distinct but related areas: protein biochemistry, biophysics, and computer science. My Ph.D. and postdoctoral research have centered on using and developing experimental methods, primarily NMR and biochemical techniques, to elucidate the function of proteins in terms of their structure and dynamics in solution. My current research focuses on design and implementation of algorithms for NMR structure determination using very sparse restraints, as well as algorithms for automatic resonance assignment.

As my C.V. shows, I have successfully applied multidisciplinary approaches to solve complex problems in biochemistry. I believe that I can make a unique contribution to your excellent research program in computational structural biology. In particular, I am interested in a number of interdisciplinary projects including the development of algorithms and NMR techniques for structure-based drug design and for probing protein-protein interaction. I feel that my broad research experience and strong background in physics and mathematics will continue to be a valuable asset for collaborating successfully with researchers from different areas such as molecular biology, biochemistry, computer science, and physical chemistry.

I am enclosing my C.V. and statement of research. Letters of recommendation are being mailed under separate cover. Thank you very much for your consideration.

Sincerely,

Lincong Wang

# Research Statement

**Lincong Wang**

For the past ten years I have been conducting research in three distinct but related areas: biochemistry, biophysics and computer science. A common theme of my research is using both experimental techniques and computational methods to elucidate the biochemical function of proteins.

## Previous Research Experience

**Dissertation research.** My Ph.D. project under Prof. Honggao Yan aimed to explain the function of proteins through biochemical measurement, structure determination and the characterization of dynamics in solution. Specifically, the study focused on human cellular retinoic acid binding proteins (CRABPs), primarily type II CRABP (CRABPII). CRABPs bind retinoic acid (RA), a metabolite of vitamin A. RA has profound effects on cell growth, differentiation and morphogenesis. Its analogs (retinoids) have been used to treat skin diseases and cancers, especially leukemia. To investigate the biochemical function of CRABPII, especially the residues important for RA binding, we developed a sensitive binding assay for measuring the relative dissociation constants of a series of site-directed CRABPII mutants [10, 15]. The binding data showed that two arginine residues (R111 and R132) contribute most to RA binding. In addition, we applied solution nuclear magnetic resonance spectroscopy (NMR) to examine the interaction between the bound RA and CRABPII, and to characterize the mechanism of RA exiting from the binding site [14]. The NMR data, namely, the nuclear Overhauser effect (NOE) distances between the bound RA and CRABPII, showed that the conformation of the bound RA and its interaction with CRABPII are, *qualitatively*, the same as those observed in the crystal structure of holo-CRABPII in complex with an *all-trans* RA [1]. However, a *quantitative* analysis of the NOE data indicated that, *in solution*, the $\beta$-ionone ring and its neighbors, the C-7 and C-8 positions, of the bound RA are more flexible than its carboxylate end. The NMR results correlate well with the biochemical studies which showed that the modifications at the $\beta$-ionone ring and the C-7 and C-8 positions of retinoids can be tolerated while the carboxylate group is critical for binding. In the crystal structure of holo-CRABPII the residues in contact with the $\beta$-ionone ring, the C-7 and C-8 positions of RA formed a narrow entrance leading to a deeply buried binding site. The high flexibility of these residues, as inferred from the observed NOE data, may make it possible for the buried RA to escape from the binding site through the entrance which, in the crystal structure, is too narrow to allow the entire RA to pass. To further investigate RA binding mechanism, especially to examine the roles of R111 and R132 in the binding, we again used NMR to solve the solution structures of both the wild-type apo-CRABPII [9] and a mutant CRABPII (R111M) [13], *and* to characterize the dynamics of the wild-type apo-CRABPII, a series of mutants (R111M in particular) [10, 15] and holo-CRABPII [14]. From the NMR structure and dynamics of the wild-type apo-CRAPBII we concluded that *in solution* the residues lining the RA entrance are highly flexible and the entrance itself is much more open than the entrance in holo-CRABPII crystal structure. However, compared with the wild-type apo-CRABPII the ligand entrance in the NMR structure of R111M mutant is much less open. Furthermore, the dynamical data of R111M mutant

indicate that the residues lining the entrance are more rigid. We thus proposed that R111 contributes not only to the binding of RA through the interaction of its guanidinium group with the carboxyl group of RA but also is critical for maintaining an open entrance and for enhancing the flexibility of the entrance.

**Postdoctoral research.** My postdoctoral project supervised by Prof. Erik Zuiderweg centered on the development of NMR techniques for measuring protein dynamics (internal motion) and for elucidating protein function in terms of their dynamics and structure. We designed several NMR experiments to measure, respectively, the longitudinal cross-correlation in protonated proteins [8], transverse $^{13}$CO-$^{1}$HN (dipole-dipole) / $^{13}$CO chemical shift anisotropy (CSA) cross-correlation [3] and $^{15}$N and $^{13}$CO CSA / CSA cross-correlation [4]. In the work on longitudinal cross-correlation we first derived, from quantum and statistical mechanics, a system of differential equations describing the relaxation of a three spin system with both the dipole-dipole and CSA interactions. We then developed new NMR techniques for measuring various relaxation rates based on an approximation to the system of differential equations. These new experiments [8, 3, 4] make it possible to measure those internal motions which are inaccessible by the existing NMR techniques ($T_1/T_2$, heteronuclear NOE measurement, $T_{1\rho}$ and H-D exchange). In particular, we were able to observe local anisotropic motions through the measurement of the transverse $^{13}$CO-$^{1}$HN (dipole-dipole) / $^{13}$CO (CSA) cross-correlation [3]. We also developed a novel computational method for extracting slow dynamics from residual dipolar coupling (RDC) data [12]. The new experiments and computational method [12] as well as the existing NMR techniques were applied to examine the internal motion a ribonuclease (binase) over time scales ranging from pico-seconds to seconds. Through an in-depth analysis of all the dynamical data we correlated, for the first time, the catalysis of an enzyme with its internal motions: the relaxation rates of the residues involved in the catalysis of binase compare well with its $k_{cat}$ value [12].

**Current research.** My project at Dartmouth with Prof. Bruce Donald focuses on the development of algorithms for high throughput structure determination using very sparse NMR restraints [11, 7, 6, 5], primarily the orientational restraints from RDCs, and for automatic NMR resonance assignment [2]. Briefly, our algorithm for NMR structure determination first, divides an entire protein into secondary structure elements, then computes the orientation and local conformation of each of the elements by a systematic search method using only 2 RDCs per residue (the RDCs of backbone NH vectors in two aligning media [6] *or* the RDCs of both NH and CH vectors in a single medium [5]). Finally, a 3-dimensional backbone structure is constructed by translating these elements using very sparse NOE distances between them. The systematic search itself is made feasible and efficient by a set of new low-degree monomials derived from RDC equation and protein backbone geometry for computing, *exactly* and *in constant time*, the directions of an internuclear vector and the sines and cosines of backbone dihedral $(\phi, \psi)$ angles. Compared with other, heuristic NMR fold determination algorithms, our exact, systematic search-based algorithm uses fewer restraints but achieves similar or better accuracy. To our knowledge, our algorithm is the first NMR structure determination algorithm that simultaneously uses exact solutions, systematic search and only 2 RDCs per residue. With these new algorithms we demonstrated that solution NMR spectroscopy can play a major role in structural genomics by determining protein structures rapidly and inexpensively. Currently, we have also derived a set of new equations for computing analytically the best $(\phi, \psi)$ solutions for an individual residue using more than 3 RDCs per residue. An algorithm based on these equations is being developed.

# Future Research

For my future research I will take a multidisciplinary approach to solve challenging problems in biochemistry using experimental techniques, computational methods and theoretical models. I expect to collaborate with researchers from different areas such as molecular biology, enzymology, computer science and physical chemistry. Obviously, the research direction will be influenced by such collaborations afforded by my next position. At present, I am particularly interested in developing algorithms and experimental techniques, especially NMR, for *structure-based drug design, detection of protein-protein interaction* and *high throughput structure determination using sparse restraints.*

**Structure-based drug design.** Among the difficulties faced by structure-based drug design (or ligand-protein docking) are (a) conformational change and structure elasticity of active sites of proteins in solution, (b) uncertainty in the static structures solved by either X-ray diffraction or solution NMR, (c) inaccuracy of force fields and lack of accurate thermodynamical data. By an active site we mean either the substrate binding and catalysis site of an enzyme or the ligand binding site(s) of a binding protein. By structure elasticity we mean the capacity and extent of conformational changes. Based on my experience with CRABPs and binase I believe that, for a structure-based drug design algorithm to be robust, experimentally measured dynamical data must be incorporated *quantitatively*. I plan to develop a unified model built upon physical approximations to describe the conformational change, structure elasticity as well as the uncertainty in both structure and force field. Such a unified model will make it possible to incorporate data from diverse sources such as the experimental data from NMR and other spectroscopies, and statistical data extracted from the Protein Data Bank (PDB). I will explore new NMR techniques for measuring protein dynamics, especially the sidechain dynamics of residues important for ligand binding and enzyme catalysis, and new algorithms for structure-based drug design which will incorporate various dynamical data. The proposed research will be performed on proteins with pharmaceutical significance, possibly in collaboration with other researchers. The computational results will be further tested in a wet lab.

**Protein-protein interaction.** The detection of protein-protein interactions (protein-protein docking) based on the static structures of proteins by computational methods is a more formidable task than ligand-protein docking and faces many challenges. One of them is the high flexibility and large structural uncertainty of the solvent exposed sidechains on a protein surface. In a similar manner as described in the previous paragraph I will first develop a unified model for an efficient representation of a protein surface taking into account both the flexibility and position uncertainty of sidechains. Then, algorithms for protein-protein docking will be developed under the unified model using a minimum number of experimental restraints such as RDCs, chemical shifts, transfer NOEs from NMR and statistical restraints from the PDB. In addition, new NMR techniques for detecting protein-protein interactions will be developed. I will also explore simulation methods such as Brownian dynamics simulation to study the process of protein-protein interaction.

**High throughput structure structure determination using sparse restraints.** I would like to extend my current work on algorithm design for high throughput structure determination using very sparse restraints. In particular, I plan to develop NMR techniques and efficient algorithms for determining (a) an entire structure using only RDCs, (b) nucleic acid structures using RDCs and (c) sidechain conformations in an RDC-derived backbone structure using, for example, sequence-dependent rotamer library.

# References


[1] G. J. Kleywegt, T. Bergfors, H. Senn, P. Le Motte, B. Gsell, K. Shudo, and T. A. Jones. Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-trans-retinoic acid and a synthetic retinoid. *Structure*, 2:1241–1258, 1994.

[2] C. Langmead, A. Yan, R. Lilien, L. Wang, and B. R. Donald. A polynomial-time nuclear vector replacement algorithm for automated NMR resonance assignments. In *7th Annual Intl. Conf. on Comp. Mol. Biol.*, pages 176–187, Berlin, Germany, 2003.

[3] Y. Pang, M. Pellecchia, L. Wang, A. V. Kurochkin, and E. R. P. Zuiderweg. Evidence for extensive anisotropic local motions in a small enzyme using a new method to determine NMR cross-correlated relaxation rates in the absence of resolved scalar couplin. *J. Biomol. NMR*, 14:297–306, 1999.

[4] M. Pellecchia, Y. Pang, L. Wang, , A. V. Kurochkin, A. Kumar, and E. R. P. Zuiderweg. Quantitative measurement of cross-correlations between $^{15}$N and $^{13}$CO chemical shift anisotropy relaxation mechanisms by multiple quantum NMR. *J. Am. Chem. Soc.*, 121:9165–9170, 1999.

[5] L. Wang and B. R. Donald. Exact solutions for internuclear vectors and backbone dihedral angles from NH and CH residual dipolar couplings in *a single* medium, and their application in a systematic search algorithm for determining protein backbone structure. *(prepared)*.

[6] L. Wang and B. R. Donald. Exact solutions for internuclear vectors and backbone dihedral angles from NH residual dipolar couplings in *two media*, and their application in a systematic search algorithm for determining protein backbone structure. *J. Biomol. NMR (in review)*.

[7] L. Wang and B. R. Donald. An exact algorithm for determining protein backbone structure from a minimum number of residual dipolar couplings. In *8th Annual Intl. Conf. on Comp. Mol. Biol.*, San Diego, CA (*in review*), 2004.

[8] L. Wang, A. V. Kurochkin, and E. R. P. Zuiderweg. An iterative fitting procedure for the determination of longitudinal NMR cross–correlation rates. *J. Magn. Reson.*, 144:175–185, 2000.

[9] L. Wang, Y. Li, F. Abildgaard, J. L. Markley, and H. Yan. NMR solution structure of type II human cellular retinoic acid binding protein: Implications for ligand binding. *Biochemistry*, 37:12727–36, 1998.

[10] L. Wang, Y. Li, and H. Yan. Human cellular retinoic acid binding proteins: Quantitative analysis of the ligand binding properties of the wild-type and site-directed mutants. *J. Biol. Chem.*, 272:1541–1547, 1997.

[11] L. Wang, R. Mettu, R. Lilien, and B. R. Donald. An exact algorithm for determining protein backbone structure from NH residual dipolar couplings. In *IEEE Computer Society Bioinformatics Conference*, pages 611–612, Stanford University, CA, 2003.

[12] L. Wang, Y. Pang, T. Holder, J. Brender, A. V. Kurochkin, and E. R. P. Zuiderweg. Functional dynamics in the active site of the ribonuclease binase. *Proc. Natl. Acad. Sci. USA*, 98:7684–7689, 2001.

[13] L. Wang and H. Yan. NMR study suggests a major role for arg111 in maintaining the structure and dynamical properties of type II cellular retinoic acid binding protein. *Biochemistry*, 37:13021–32, 1998.

[14] L. Wang and H. Yan. NMR studies of type II human cellular retinoic acid binding protein. *Biochimica et Biophysica Acta*, 1433:240–252, 1999.

[15] H. Yan, L. Wang, and Y. Li. A novel method for measuring the binding properties of the site-directed mutants of the proteins that binding hydrophobic ligands: Application to cellular retinoic acid binding proteins. *Techniques in Protein Chemistry VIII (Marshak, D. R., Ed.)*, pages 449–456, 1997. Academic Press, San Diego.