Dartmouth **Department of Computer Science**

Center for Structural Biology and Computational Chemistry

Bruce Randall Donald
Joan P. and Edward J. Foley Jr 1933 Professor
of Computer Science
Adjunct Professor of Chemistry and Biological Sciences

6211 Sudikoff Laboratory
Hanover, NH 03755-3510
voice: (603) 646-3173
fax: (603) 646-1672
email: brd@cs.dartmouth.edu
www.cs.dartmouth.edu/~brd/

November 26, 2003

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

Dear Dr. Steveninck,

I am happy to write to recommend Dr. Lincong Wang for a position in your department. Lincong has been a postdoctoral research associate in my lab for over two years, so I have collaborated with him closely and know him very well. Lincong is a very good research scientist in the field of Nuclear Magnetic Resonance (NMR) structural biology. Overall, his work involves creation of sophisticated computer models to assist researchers in analyzing three-dimensional structures of proteins, and designing drugs targeting them. Lincong's doctoral work focused on several important proteins and how they do their job in the human body. Vitamin A taken from food becomes retinoic acid in the body – the retinoic acid then is in a form that the body can use. In particular, retinoic acid has profound effects on the development of a human embryo, as well as cell growth in general. Analogs of retinoic acid have been used to treat cancers, especially leukemia and several skin diseases such as acne.

Lincong developed sophisticated new biochemical and biophysical techniques to understand the interactions of these proteins with retinoic acid. He also succeeded in the difficult task of resolving the structures of these proteins in solution. These functional-structural studies have provided the basic information for drug designs targeting these proteins, and have been used by others to design analogs of retinoic acid used to treat cancers and skin disorders. He has published his work as several first-author papers in very good journals, including *Biochemistry, Journal of Biological Chemistry,* and *Biochimica et Biophysica Acta.* The work had also been presented at the Annual Symposium of Protein Society and other good scientific conferences.

Lincong is a very productive researcher. In addition to his work on retinoic acid structure, he has also developed computational techniques (computational geometry in particular) for automated NMR assignment and structure determination, and rational drug design. He has pioneered the use of computational techniques to automate this process, and has created motion models for the proteins based on statistical mechanics and molecular dynamics to best interpret the information from NMR studies.

Before joining our team at Dartmouth, Lincong conducted his research at the University of Michigan on the developments of nuclear magnetic resonance (NMR) and computational methods to study protein dynamics: the internal motions of proteins in solution. Such dynamical information is critical for rational drug design since it complements the structural data obtained from x-ray diffraction and solution NMR, which are static in nature. His original contribution there was the development

of novel techniques to extract dynamical information by NMR relaxation measurements and by residual dipolar couplings. He has presented these discoveries at key meetings in the field, including the Frontiers of NMR in Molecular Biology (Keystone Symposia), etc. This work was published, also as first-author papers, in two very prestigious journals, *Proc. National Academy of Sciences* and *Journal of Magnetic Resonance.* Lincong has written several software packages to process the data obtained by his novel method. The name of the software is RDC-Dynamics and is freely available to the research community.

During his previous works at Geo-centers Inc., Lincong made significant contributions in unifying the data in the RCSB Protein Data Base (`http://www.rcsb.org/pdb`) by combining his skills and knowledge in both computer science and structural biology. The Protein Data Base is the only data base storing 3D structural data obtained by experimental methods, and is used by researchers worldwide.

Since joining our group, Lincong has made important contributions to our work on computational approaches to reprogramming enzyme specificity, towards combinatorial biosynthesis for small-molecule diversity. These take the form of a computational scoring function for evaluating protein design, rigorously based on statistical mechanics. Lincong has also made fundamental contributions by leading the way in the integration of residual dipolar coupling data into our structural genomics research.

Hence, Lincong is well-qualified and very creative in all aspects of his research, including computational methods for biomolecular NMR, NMR structural genomics, and computational modeling of protein motion from NMR dynamics.

At Dartmouth, Lincong done excellent work on novel algorithms for NMR structure determination. These are represented in a new paper, 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. L. Wang and B. R. Donald, under review at the *Journal of Biomolecular NMR.* Lincong's paper makes four major contributions to the method of determining protein structures by solution NMR spectroscopy using residual dipolar couplings as the main restraints. These contributions, I believe, will be valuable not only to the NMR community in particular and structural genomics in general, but also to structural biologists more broadly. This is because in both experimental and computational structural biology, exact computational methods have been, for the most part, elusive to date. Second, rigorous comparisons of structures derived from NMR vs. X-ray crystallography are made possible by Lincong's techniques, and these comparisons should be of general interest.

Lincong's is the first NMR structure determination algorithm that simultaneously uses exact solutions, systematic search and only 2 RDCs per residue. (A systematic search is a search over all possible conformations (solutions) that employs a provable pruning strategy that guarantees pruned conformations need not be considered further). Lincong's first contribution is the derivation of low-degree polynomial equations for computing, *exactly* and *in constant time*, dihedral (ϕ, ψ) angles from residual dipolar couplings (RDCs) measured on a single internuclear vector $\mathbf{v}$ in two different aligning media. The easily computable exact solutions eliminate the need for one-dimensional grid-search previously employed to compute the directions of $\mathbf{v}$ or two-dimensional grid-search to compute (ϕ, ψ) angles. Furthermore, these equations are very general and can easily be extended to compute both the backbone and sidechain dihedral angles from RDC data measured on any single vector in two aligning media. Finally, Lincong's method can be applied *mutatis mutandis* to derive similar equations for computing dihedral angles from RDCs in nucleic acids.

Lincong's second contribution is the development of a novel minimization algorithm based on depth-first search and biophysical insights. In contrast to (e.g.) simulated annealing approaches, his minimization is built upon the exact solutions for computing backbone (ϕ, ψ) angles from RDC data and exhaustive search. The minimization algorithm represents a very good way to compute structures as accurately as possible using very sparse restraints and will be of interest to the structural

biology community in general.

Lincong's third contribution is the computation of β-sheets from RDC data alone, which fundamentally extends a previous method of Prestegard and coworkers (which targeted only entirely helical proteins). Lincong's algorithm can be applied to determine the backbone structures of proteins consisting of either α-helices or β-sheets or both and thus has much wider applications since a majority of proteins have both α-helices and β-sheets.

Lincong's fourth contribution is the demonstration, for the first time, that the orientations of both α-helices and β-strands can be computed accurately and efficiently using only RDCs measured on a single bond vector type (NH) in two aligning media. With a minimum number of additional distance restraints a three dimensional structure can be computed consequently. Compared with other algorithms for computing backbone structures using RDCs, Lincong's algorithm achieves similar results but requires less data, relies less on statistics from the PDB and does not depend on molecular dynamics. Since RDCs can be acquired and assigned much more quickly than NOEs in general, his results show it is possible to compute structures very rapidly and inexpensively using mainly RDC restraints. For this reason, I believe believe that his algorithm will be of interest to the NMR, structural genomics, and structural biology communities.

Lincong has also collaborated with my other students on novel algorithms for automated NMR resonance assignment. This work (A Polynomial-Time Nuclear Vector Replacement Algorithm for Automated NMR Resonance Assignments. C. Langmead, A. Yan, R. Lilien, L. Wang, and B. R. Donald, RECOMB '03; and *Jour. of Computational Biology*, in press) should be of broad interest to the NMR and structural biology communities. It is often the case that a structural biologist needs NMR resonance assignments for a protein of known structure, or given an homologous structure. For example, the structure might have been solved by X-ray crystallography, but one wishes to use NMR to investigate protein-protein interactions (via chemical shift mapping), protein-ligand binding (via SAR by NMR or line-broadening analysis), or dynamics (via, e.g., nuclear spin relaxation analysis). It would be desirable to have a fast resonance assignment procedure that exploits the *a priori* structural knowledge. Our method employs the global properties of ^{1}H-^{15}N residual dipolar couplings (RDCs) to match geometric constraints on bond vectors to a known structural model. The algorithm achieves high accuracy on real NMR data and runs in only a few minutes. The required spectra can be recorded in about a day and only ^{15}N-labeling is required.

The method, called *Nuclear Vector Replacement* can also be used to assign a new protein, using an homologous structure. The RDC assignments can then be used for structure refinement. Our technique is named by analogy to molecular replacement (MR) in crystallography, and complements Ad Bax's molecular fragment replacement method in several ways (for example, the algorithm operates on unassigned spectra and unassigned RDCs). I believe our method fills an important gap in NMR structural biology. As noted in the paper, we hope the technique will be applied in structural genomics: Current efforts in structural genomics are expected to determine experimentally many more protein structures, thereby populating the "space of protein structures" more densely. This large number of new structures should make techniques such as X-ray crystallography MR and computational homology modeling more widely applicable for the determination of future structures. In the same way that MR attacks a critical informational bottleneck (phasing) in X-ray crystallography, an analogous technique for "MR by NMR" should address the NMR resonance assignment bottleneck. Hence, we developed NVR as a new RDC-based algorithm, which computes assignments that correlate experimentally-measured RDCs to a given *a priori* whole-protein 3D structural model. I believe this algorithm could be important for "MR by NMR".

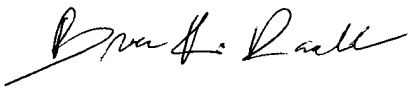
NVR represents a major advance over earlier work, and is much better suited to high-throughput application. A careful comparison is given to previous methods, revealing that NVR requires substantially less spectrometer time, fewer NMR experiments, cheaper isotopic labeling, and is computationally much more efficient. Finally, NVR allowed us to conduct a systematic analysis of some error characteristics of RDCs relative to 20 structural models. We discuss the information content

of RDCs relative to other NMR data. This paper should be of interest to both the NMR community and to other structural biologists. Lincong's major contribution on this paper was as our expert on the experimental and computational aspects of RDCs.

From a scientific point of view, Lincong's results are as strong as those of my former graduate student Chris Langmead (now an Assistant Professor in the Computer Science Department at Carnegie-Mellon), and somewhat stronger than those of my former postdoc Chris Bailey-Kellogg (now an Assistant Professor at Purdue). Both Langmead and Bailey-Kellogg work on computational aspects of biomolecular NMR. However, Lincong's mastery of physics, NMR, biophysics, and biochemistry is much stronger than either Chris. The mathematical talents of all three young scientists are comparable. While Langmead and Bailey-Kellogg have more grounding in standard computer science, Lincong is nevertheless quite impressive and creative computationally. Lincong may have the best computational background I have seen in all my experimental NMR co-workers. However, Langmead and Bailey-Kellogg are better at writing and giving technical talks. I think this is largely a matter of experience; since coming to Dartmouth, Lincong has worked hard on his skills at writing and giving talks.

Lincong has mentored and guided students and postdocs in our lab, and has served on the thesis committees of two graduate students. I feel Lincong has done very good work and is a very good scientist. While at Dartmouth, Lincong has produced two very good papers on new topics. He has excellent training and should continue to make significant contributions to structural and computational biology when he heads his own laboratory. Lincong is unique in having a rare, insightful, and integrated understanding the biophysical side as well as the computational side of NMR structural biology and computational biophysics. Please feel free to contact me if you have any further questions about Lincong.

Sincerely,



Bruce Randall Donald