AB INITIO PROTEIN TERTIARY STRUCTURE PREDICTION: COMPARATIVE-GENETIC ALGORITM WITH GRAPH THEORETICAL METHODS

Final Technical Report

April 2001

Susan K. Gregurick

PREPARED FOR THE UNITED STATES DEPARTMENT OF ENERGY/ALFRED P. SLOAN FOUNDATION

Work Preformed Under a DOE/Sloan Postdoctoral Fellowship

DISCLAIMER

During the period from September 1, 1998 until September 1, 2000 I was awarded a Sloan/DOE postdoctoral fellowship to work in collaboration with Professor John Moult at the Center for Advanced Research in Biotechnology (CARB). Our research project, *Ab Initio Protein Tertiary Structure Prediction and a Comparative Genetic algorithm,* yielded promising initial results. In short, the project is designed to predict the native fold, or native tertiary structure, of a given protein by inputting only the primary sequence of the protein (one or three letter code). The algorithm is based on a general learning, or evolutionary algorithm and is called Genetic Algorithm (GAs). In our particular application of GAs, we search for native folds, or lowest energy structures, using two different descriptions for the interactions of the atoms and residues in a given protein sequence. One potential energy function is based on a free energy description, [Pedersen97] while the other function is a threading potential derived by Moult and Samudrala.[Samudrala97] This modified genetic algorithm was loosely termed a Comparative Genetic Algorithm and was designed to search for native folded structures on both potential energy surfaces, simultaneously. We tested the algorithm on a series of peptides ranging from 11 to 15 residues in length, which are thought to be independent folding units and thereby will fold to native structures independent of the larger protein environment.[Pedersen97] Our initial results indicated a modest increase in accuracy, as compared to a standard Genetic Algorithm.[Pedersen97] We are now in the process of improving the algorithm to increase the sensitivity to other inputs, such as secondary structure requirements. The project did not involve additional students and as of yet, the work has not been published.

While working at CARB, I also had the opportunity to collaborate with a number of gifted scientists. I would like to take this opportunity to discuss each of the following collaborations, and the resulting publications.

In collaboration with Sherif A. Kafafi, we calculated the electronic and spectroscopic prosperities of eleven conformers of D-glucopyranose and D-mannopyranose using both an improved density functional method (K2-BVWN) as well as an empirical vibrational self-consistent field method (VSCF). One advantage of the presented density functional method was that the frequencies did not have to be rescaled, a particularly troublesome problem for many *ab initio* calculations. We were also able to show that the scaling of the calculation goes as N^3 where N is the number of basis functions. We found that we were not only able to reproduced the very difficult to calculate anomeric effect, but we were also able to reproduce the experimental vibrational frequencies for all eleven structures studied. This work illustrated the versatility of the K2-BVWN density functional method in calculating energetic and spectroscopic properties of carbohydrate systems. The work resulted in one invited publication in the Journal of Carbohydrate Chemistry, vol 18 pg 867 (1999). The project did not involve additional students, other than myself.

In collaboration with Fredrick P. Schwartz, Y. Shi, S. Wang, B. D. Wladkowski and S. Krueger at NIST and CARB, we studied the conformational changes in Cyclic AMP Receptor Protein when binding with DNA and the co-factor cyclic AMP, as measured in solution by small angle neutron scattering. We found, based on the experimental

scattering results, that the radius of gyration of the CRP protein increased significantly when binding with a 40 base-pair consensus sequence Con-DNA. My contribution to this collaborative effort was to model the conformational changes in the protein and calculate a scattering profile from our energy-minimized model, in order to compare to the experimentally measured scattering profile. The conformational studies of the CRP protein involved walking along eight well-defined conformational coordinates and at each sampling point calculate the radius of gyration and scattering curves to best approximate the experimental data. Using this method, we were able to show that the observed conformational change is due to large amplitude motions in the hinge region of the protein, between the C and B alpha-helix. The calculated scattering curves and distance distribution function of our model compare quite well with those obtain from the experimental data. This work is to be submitted presently to the Journal of Biological Chemistry.

In collaboration with Galina M. Chaban and R. Benny Gerber at the Hebrew University, and the University of California, Irvine, we also studied the vibrational spectroscopy of a small molecule, N-Methylacetamide (NMA). N-Methylacetamide was chosen as a good model for the peptide linkage in proteins and peptides, and thus we are able to initiate a study of the spectroscopy and mode-coupling effects in peptides. This project involved developing an improved empirical potential energy surface to describe the NMA molecule, as well as calculating a 33-by-33 pair-wise mode coupled vibrational potential energy surfaces using a second-order Moller-Plesset *ab initio* electronic structure calculation, as implemented in the GAMESS Electronic Structure Calculation Package.[Schmidt93] Using both potential energy surfaces and the anharmonic Correlated Corrected Vibrational Self-Consistent Field (CC-VSCF) method, we calculated the vibrational frequencies of NMA and compared to existing experimental data. This work illustrated the accuracy of the ab-initio CC-VSCF method in reproducing the NMA experimental vibrational frequencies without the need for rescaling. We also developed an expression to describe qualitatively, the degree of mode-to-mode coupling within a given molecule (termed coupling strength). Using both the CC-VSCF frequencies, and the expression for the coupling strength, it is possible to directly calculate both the probability of intramolecular vibrational energy transfer, as well as the two-dimensional infrared spectra of NMA. This work is to be submitted to the Journal of Physical Chemistry, and was presented at the 2001 International Conference on Computational Nanoscience held on Hilton Head Island, March 19-21. There were no additional students trained on this project.

My final collaboration was with Susan Krueger, James Zondlo, and Edward Eisenstein at the Center for Advanced Research in Biotechnology. In this collaboration, we studied the conformation of the GroEL/GroES complex, as measured by small angle neutron scattering. I developed a model for the asymmetric ring opening in the GroEL protein of the complex. I also calculated the neutron scattering profiles and radius of gyrations for these models. This work was to assist James Zondlo in the interpretation of his experimental results for incorporation into his PhD thesis. The work remains an unpublished thesis. James Zondlo, a graduate student, actively participated in this project.

J.T. Pedersen and J. Moult, JMB **269**, 240 (1997)
R. Samudrala and J. Moult, JMB **275**, 893 (1998)
M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, N. Matsunaga, K.A. Nguyet, S. Sa, T.L. Windus, M. Dupuis, and J. A. Montgomery, JCC **14**, 1347 (1993)