CS 267 Assignment 0

Alex Albaugh
February 7, 2013

Biography

I am a first year PhD student in chemical engineering and a member of Teresa Head-Gordon's lab, which is broadly focused on computational studies of biomolecular systems. My area of focus is in combined quantum mechanical/molecular mechanical (QM/MM) simulations, particularly in developing novel ways for these two systems to interact. From this class I am hoping to learn how to develop efficient code for use in a computer cluster, particularly code that can interface well with quantum mechanical and molecular mechanical simulation routines.

Parallel Application: Folding@home

Folding@home is a distributed computing project where volunteers download software that allows their personal computers to run molecular dynamics simulations during idle processor time. The project has about 175,000 active cores and 167,000 donors.

Background

Proteins have a hierarchy of structure. The primary structure consists of the particular sequence of amino acids in the protein. The secondary structure consists of common, local shapes within the protein such as alpha-helices and beta-sheets. The tertiary and quaternary structures are the overall 3-D structure of entire proteins and multi-protein complexes, respectively. Folding@home uses the distributed computational resources of its thousands of donors to run molecular dynamics simulations that determine how a given protein structure folds into its final 3-D form (tertiary, quaternary structure).

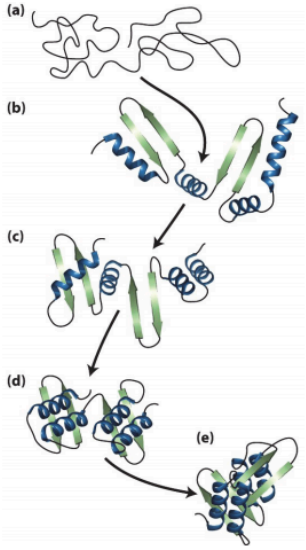

Figure 1: The path of a protein primary amino acid sequence (a) as it folds into its final 3-D structure (e). Alpha-helices are in blue and beta-sheets are in green.

Work on protein folding is important because diseases such as Alzheimer's, cystic fibrosis, mad cow, and even some cancers are thought to be caused by protein misfolding. Additionally, protein folding can help with drug development and protein design/engineering.

Parallelization

Folding@home does not have a static architecture as all of the computational resources are distributed across the systems of thousands of donors. The project takes advantage of a wide range of possible donor hardware including multi-core processors, GPUs, and even PlayStation 3s (from March 2007 to November 2012). Not only is the overall project parallel in the sense that the whole project is distributed, the software also takes advantage of parallel architectures on individual donor machines. As you might imagine, one of the challenges is designing different versions of the software to run on different platforms.

Given the nature of distributed computing minimal communication is desired since the components can physically be on the other side of the world, so the overall process works in the following manner. A donor downloads a "work unit" from the Folding@home server, their machine processes and completes the work unit, and then the user uploads the completed information back to the server. These work units consist of pieces of molecular dynamics simulations and are pieced together with Folding@home's internal machines. Each work unit is also given a deadline so that in the event of a client error, deletion, or other failure the unit is reassigned to a different donor so that work can proceed at a fast pace.

Performace

Folding@home has been extremely successful. Folding simulations with Folding@home can run into the millisecond (10^{-3}) timescale, which is impressive when considering that timesteps in these simulations can be in the picosecond (10^{-12}) or even femtosecond (10^{-15}) timescales. In terms of speed Folding@home operates in the petaflop range and in September 2007 became the first "computer" to crack the petaflop barrier. Although Folding@home does not qualify for the Top500 list, it would at least be within the top 5 as of the November 2012 list since its typical operating speed is around 5 petaflops.

In terms of the scientific value of Folding@home currently 100 papers have been published using data from the program. In addition most data from the Folding@home simulations agrees well with experimental findings. Finally, the work done through Folding@home has contributed to the development of experimental treatments for Alzeihmer's and Huntington's diseases.

References

1. "Folding@home: Lessons from eight years of volunteer distributed computing", A.L. Beberg, D.L. Ensign, G. Jayachandran, S. Khaliq, V.S. Pande, *Proceedings of the 2009 IEEE International Parallel and Distributed Processing Symposium,* **2009.**
2. folding.standford.edu
3. H. Lodish, A. Berk, C.A. Kaiser, M. Krieger, M.P. Scott, A. Bretscher, H. Ploegh, P. Matsudaira. *Molecular Cell Biology*. 6th ed. W.H. Freeman and Company.
4. top500.org