Intelligent Engineering Systems Through Artificial Neural Networks, Vol. 4, Eds. C.H. Dagli, B.R. Fernández, J. Ghosh, and R.T.S. Kumara, 1994, Am. Soc. Mechanical Engr.; NY

NEURAL NETWORKS TO COMPUTE MOLECULE DYNAMICS

LARRY S. LIEBOVITCH
Center for Complex Systems, Florida Atlantic University
Boca Raton, FL 33431 USA

NIKITA D. ARNOLD
Angewandte Physik, J. Kepler University
A-4040 Linz, Austria

LEV Y. SELECTOR
Department of Ophthalmology, Columbia University
New York, NY 10032 USA

ABSTRACT
In order to understand how large biomolecules, such as proteins, function as miniature machines, we need to compute the motions within these molecules. These molecular dynamics computations have been performed by evaluating the force on an atom, determining how far that force moves the atom in a time step, and repeating this procedure many times. However, because the forces change when the structure changes, these time steps must be kept very small, typically 10^{-15} s. Thus, it has not been possible to compute motions over the millisecond time scale that is biologically important. Neural networks share many properties with large biomolecules such as multiple energy minima, frustration, and ultrametricity. Thus, neural networks may provide a more natural and thus a more efficient method to compute molecular dynamics. We show how to construct a neural network that is analogous to a given molecule so that the dynamics of the neural network can be used to compute the dynamics of the molecule. The spatial structure of a molecule is encoded in the values of the nodes of the neural network and the energy structure of the molecule is encoded in the connection strengths between the nodes. As the network evolves in time it therefore computes the changing structure of the molecule. We illustrate these ideas by using a Hopfield network to compute the structure of cyclohexane switching from a chair to a twisted boat conformation.

INTRODUCTION

Motions in large biomolecules such as proteins are important in how these molecules function as structural units, catalyze chemical reactions, and bind ligands (Karplus and Petsko, 1990; McCammon and Harvey, 1987; Welch, 1986). To compute the spatial positions of the atoms of a protein as a function of time the standard method has been: 1) evaluate the force on an atom due to all the other atoms, 2) move that atom appropriately, and 3) repeat steps 1 - 2 many times (Karplus and Petsko, 1990; McCammon and Harvey, 1987). The limitation of this method is that the time steps must be kept very small, typically 10^{-15} s, so that the forces do not change during the time step. Thus, it is not possible to compute the motions over milliseconds or seconds that are of interest, for example, in our patch clamp experiments where we measure the durations of the open and closed states of ion channels in the cell membrane (Liebovitch et al., 1987; Liebovitch and Sullivan, 1987; Liebovitch, 1994).

By transforming one problem into another mathematically equivalent problem, we can sometimes find a much more efficient computational algorithm. For example, the usual algorithm for multiplication is to multiple the first number by each digit of the second number and sum these results multiplied by the place value of the digit of the second number. People typically use this algorithm to the base 10 and computers typically use this algorithm to the base 2. For numbers with N digits this algorithm requires approximately N^2 arithmetic operations. Another algorithm for multiplication is to compute the inverse fast Fourier transform of the convolution of the fast Fourier transforms of the bit patterns of the two numbers. This algorithm requires approximately N*log(N) operations. When multiplying two numbers with a million digits, and this is done for example when computing π to a million decimal places, this second algorithm is approximately 100,000 times faster than this first algorithm.

Computing molecular dynamics is slow since the time steps must be kept very small because the protein is always changing too fast and thus escaping from the computation. Our approach is to transform this problem into another computational structure that is a more natural fit to the protein, so that we can develop a more efficient computational algorithm. Proteins share many properties in common with neural networks. Hence, neural networks may provide a more efficient method to compute molecular dynamics.

NEURAL NETWORKS AND MOLECULAR DYNAMICS

A neural network consists of nodes and connections between them (Amit, 1989; McClelland and Rumelhart, 1989; Domany et al., 1991; Freeman and Skapura, 1991). Each node has a value. At each time step, the new value of a node depends on the values of the other nodes and the strengths of the connections between them.

A neural network is like a large biomolecule, such as a protein, in a number of different ways (Karplus and Petsko, 1990; McCammon and Harvey, 1987; Welch, 1986; Fraunfelder, 1986; Sasai and Wolynes, 1990; Frauenfelder, et al., 1991; Wolynes, 1991). A neural network has many nodes. The protein has many atoms. The nodes of a neural network can interact with each other at short range or long range. This is analogous to the short range atomic bonds and long range electrostatic forces in a protein. In both neural networks and proteins local interactions generate the global structure. The neural network is a "frustrated" system. That is, there are conflicting constraints on the values of the nodes so that there is no single energy minima. The protein is also a "frustrated" system. Sidechains, regions, and subunits can have multiple orientations in space that conflict with each other so that there is no single energy minima. Because of these conflicting constraints, neural networks have an energy function with many, local energy minima. Similarly, proteins also have a potential energy function with many, local energy minima which correspond to similar, but not identical, conformational structures. Because of this complex energy surface, the energy structure of a neural network is "ultrametric", that is, the energy of the network may need to increase a little before it can decrease. Similarly, a protein may need to increase in energy, by unfolding slightly, before it can decrease in energy by refolding into a different conformation.

To use a neural network to compute molecular dynamics we:
1) encode the spatial structure of the molecule in the values of the nodes of the neural network, 2) encode the energy structure (or force structure) of the molecule in the connection strengths, 3) update the values of the nodes, and 4) as the values of the nodes change they compute the changing structure of the molecule.

HOPFIELD NETWORK TO COMPUTE MOLECULAR DYNAMICS

A Hopfield network (Hopfield, 1982, 1984; Hopfield and Tank, 1986) is one type of neural network that can be used to compute molecular dynamics. In this network all the nodes are connected to all the other nodes. The connection strengths are symmetric ($J_{ij} = J_{ji}$) which implies

that a thermodynamically valid energy function can be defined.

To determine the connection strengths we choose a set of p memories that correspond to the stable conformational structures and/or the transition states of the molecule. Each such memory μ consists of a set of N values ξ of the N nodes, namely

$$\xi^{\mu} = (\xi_1{}^{\mu},...,\xi_N{}^{\mu}), \qquad \mu = 1...p \tag{1}$$

The connections strengths are then determined from these memories

$$J_{ij} = -\frac{1}{N^2} \sum_{\mu=1}^{p} \alpha_{\mu} \, \xi_i^{\mu} \, \xi_j^{\mu} \tag{2}$$

where α_{μ} is the energy associated with the conformational structure of memory μ.

CYCLOHEXANE

As a test of this method we used a neural network to compute the molecular dynamics of cyclohexane. Cyclohexane is a small molecule consisting of a ring of 6 carbon atoms and their attached hydrogen atoms. It has three basic conformational structures: a chair conformation which is the most stable state corresponding to the deepest local minima in the potential energy function, a twisted boat conformation which is a slightly less stable state corresponding to a more shallow local energy minima, and a boat conformation which is a metastable state corresponding to a saddle point in the energy function (Pickett and Strauss, 1970). Due to thermal fluctuations the molecule is constantly switching between its chair and twisted boat conformational states.

To compute the dynamics of this molecule we used a Hopfield network with 24 nodes.

To encode the spatial structure the nodes were distributed equally around the central plane of the molecule. If the bond connecting two carbon atoms was above the plane at the position of a node, then the value of that node was +1. If the bond connecting two carbon atoms was below the plane at the position of a node, then the value of that node was -1.

To determine the connection strengths we used the 4 memories that correspond to the 1 unique chair and 3 unique twisted boat conformational states. The energy of the least stable transition state was defined to be zero. With respect to this state, it is known that the energy of the chair is -14.5 kcal/mol and the energy of the twisted boat is -6.5 kcal/mol (Pickett and Strauss, 1970). The connection strengths were then determined from these 4 memories and their associated

energies.

The nodes were updated using Glauber dynamics.

The first result we found was that even though the network was constructed with memories representing only the chair and twisted boat states, nonetheless, the network also contains the boat state. The neural network even correctly predicts the approximate energy value of the boat conformation. The energy computed by the neural network is -4.8 kcal/mol compared to the measured energy of -6.0 kcal/mol. The boat arises out of the interactions of the chair and twisted boat memories. A property that arises from such interactions is called an emergent property. The few memories chosen to represent the molecule have captured the natural form of the molecule. The neural network reproduces additional properties of the molecule that may appear distinct to us, but are actually emergent properties of the stable conformational states of the molecule. Instead of thinking of a protein as a very complex potential function, we can think of it as constructed from a small set of "memory" structures. The interaction between these memory structures then produces the complex potential.

We evolved the network in time long enough to follow the changing structure of cyclohexane through 10,000 switches between the chair and twisted boat conformations. The computation time depended on the temperature assumed for the molecule. At room temperatures, the run through 10,000 switches required approximately 15 minutes on a Silicon Graphics IRIS workstation. From the dwell times in each state we then computed the distribution of dwell times (first passage times) for the chair and twisted boat states. These distributions are analogous to the open and closed dwell time distributions of ion channel proteins that we measure in our patch clamp experiments (Liebovitch et al., 1987; Liebovitch and Sullivan, 1987; Liebovitch, 1994).

We chose cyclohexane because there is good nuclear magnetic resonance experimental data (Anet and Bourn, 1967) and other molecular dynamics simulations (Elber and Karplus, 1987) of this molecule. Thus, we hoped to compare the results of our neural network method with these known results to determine if our new method worked. However, we could not do this because: 1) we could not determine the physical time in seconds that corresponds to each updating step in the neural network and 2) we could not compare the shape of our dwell time distributions to the previous results because our results extend over a much larger range of time scales than has ever been measured experimentally or computed by other methods. Details of these calculations as described in Liebovitch et al. (1994).

This work was supported in part by NIH grant EY06234.

REFERENCES

Amit, D. J. (1989) *Modeling Brain Function*. (Cambridge University Press, NY).

Anet, F.A.L., Bourn, A.J.R. (1967). Nuclear magnetic resonance studies of ring inversion in cyclohexane - d_{11}. *J. of Amer. Chem. Soc.* 89,760-768.

Domany, E., van Hemmen, J.L., Schulten, K. Eds. (1991). *Models of Neural Networks*. (Springer-Verlag, NY).

Elber, R., Karplus, M. (1987). A method for determining reaction rates in large molecules: application to myoglobin. *Chem. Phys. Letters* 139,375-380.

Freeman, J.A., Skapura, D.M. (1991). *Neural Networks: Algorithms, Applications, and Programming Techniques*. (Addison-Wesley, NY).

Fraunfelder, H. (1986). Proteins and glasses. In *Structure and Dynamics of Nucleic Acids, Proteins, and Membranes*. Clementi, E., Chin, S. Eds. (Plenum Press, NY), 169-177.

Fraunfelder, H., Silgar, S.G., Wolynes, P.G. (1991). The energy landscapes and motions of proteins. *Science* 254,1598-1603.

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA*. 79,2554-2558.

Hopfield, J.J. (1984) Neurons with graded responses have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA*. 81,3088-3092.

Hopfield, J.J.,Tank, D.W. (1986). Computing with neural circuits: a model. *Science* 233,625-633.

Karplus, M., Petsko G.A.(1990). Molecular dynamics simulation in biology. *Nature* 347,631-639.

Liebovitch, L.S., Fischbarg, J., Koniarek, J.P. (1987). Ion channel kinetics: a model based on fractal scaling rather than multistate Markov processes. *Math. Biosci.* 84,37-68.

Liebovitch, L.S., Sullivan, J.M. (1987). Fractal analysis of a voltage-dependent potassium channel from cultured mouse hippocampal neurons. *Biophys. J.* 52,979-988.

Liebovitch, L.S. (1994) Single shannels: from Markovian to fractal models, In *Cardiac Electrophysiology: From Cell to Bedside 2nd Ed.,* Zipes, D.P., Jalife, J., Eds. (W. B. Saunders, Philadelphia), in press.

Liebovitch, L.S., Arnold, N.D., Selector, L.Y. (1994). Neural networks to compute molecular dynamics. *J. Biological Systems,* in press.

McCammon, J.A., Harvery, S.C. (1987). *Dynamics of Proteins and Nucleic Acids*. (Cambridge University Press, NY).

McClelland, J.L., Rumelhart, D.E. (1989) *Explorations in Parallel Distributed Processing*. (MIT Press, Cambridge, MA).

Pickett, H.M., Strauss, H.L. (1970). Conformational structure, energy, and inversion rates of cyclohexane and some related oxanes. *J. Am. Chem. Soc.* 92,7281-7290.

Sasai, M., Wolynes, P.G. (1990). Molecular theory of associative memory Hamiltonian models of protein folding. *Phys. Rev. Lett.* 65,2740-2743.

Welch, R. R. (1986). *The Fluctuating Enzyme*. (John Wiley & Sons, NY).

Wolynes, P.G. (1991). Search and recognition: spin glass engineering as an approach to protein structure prediction. In *Biologically Inspired Physics*. Pelditi, P., Ed. (Plenum Press, NY), 15-37.