

PROTEINMORPHOSIS: A MECHANICAL MODEL FOR PROTEIN CONFORMATIONAL CHANGES

S. Meiyappan, R. Raghavan, R. Viswanathan, Y. Yu
Centre for Information Enhanced Medicine (CIEMED)
National University of Singapore
#03 – 14, *The Alpha, Science Park II*
Singapore 117684

W. Lawton
Centre for Information Enhanced Medicine (CIEMED) &
Computational Science Department
National University of Singapore
10 *Kent Ridge Crescent*
Singapore 119260

Proteinmorphosis is a physically-based interactive modeling system for simulating large or small conformational changes of proteins and protein complexes. It takes advantage of the cross-linked one-dimensional nature of protein chains. The user can, based on her chemical knowledge, pull pairs of points (lying either on a single protein or on different molecules) together by specifying geometric distance constraints. The resulting conformation(s) of the molecule(s) of interest is computed by an efficient finite element formalism taking into account elasticity of the protein backbone, van der Waals repulsions, hydrogen bonds, salt bridges and the imposed distance constraints. The conformational change is computed incrementally and the result can be visualized as an animation; complete interactivity is provided to position and view the proteins as desired by the user. Physical properties of regions on the protein can also be chosen interactively. The conformational change of calmodulin upon peptide binding is examined as a first experiment. It is found that the result is satisfactory in reproducing the conformational change that follows on peptide binding. We use *Proteinmorphosis* to study the cooperative hemoglobin oxygen binding mechanism in a second, more sophisticated, experiment. Different modeling strategies are designed to understand the allosteric (cooperative) binding process in this system and the results are found to be consistent with existing hypotheses.

1 Introduction

The change of conformation of a protein, as it binds to ligands and other proteins, is crucial to its function. Our goal is to create fast and interactive computation with large proteins, to complement the very detailed, computationally intensive microscopic modeling familiar from molecular dynamics. We achieve

this by using a macroscopic, *civil engineering*^a approach to the problem. The protein chain is viewed as a structural frame: mostly one-dimensional, elastic objects with cross-links due to hydrogen bonds, salt bridges and the like. The elastic properties of its backbone may be described by a formalism analogous to that of thin rods¹, although the protein backbone usually has a geometrically constrained set of degrees of freedom. While torsional angles can typically undergo large changes, bond angles and bond lengths along the backbone change very little, and some bonds are very rigid. Including breakable cross-links is equivalent to a nonlinear frame structure in civil engineering². We approximate protein conformational changes adiabatically by *pulling* a small number of points to prescribed interatomic distances, as a passage through successive equilibria until the final conformation is reached, under the action of appropriate intermolecular forces. We also developed a convenient approach for the visualization of large, complex proteins in a dynamic, interactive environment.

2 Mechanical Model for Protein Complexes

2.1 Incremental energy minimization

Our method for determining the conformational change is one of incremental energy minimization. We assume that, before binding, both protein and ligand reside in their folded equilibrium conformations with the lowest energies. During the binding process, hydrogen bonds and salt bridges break and form, and other interaction energies are introduced between the protein and the ligand. In the process, the energy landscape in terms of the total set of conformational degrees of freedom of the protein-ligand complex changes. If the time scale for equilibration is fast enough, the overall conformation may be taken to be instantaneously an equilibrium, or local minimum energy state, of this moving energy landscape.

Thus, it is essential to find a stable pathway along which the protein and the ligand change their conformations from one low energy equilibrium state to another. For allosteric proteins³, this pathway becomes a stable transition between initial and final states with corresponding alterations in quaternary structures.

This is the conceptual underpinning of incremental energy minimization. The binding process is modeled as the incremental application of a set of distance constraints, with the system correspondingly reaching a local, nearby energy minimum. Throughout the process, due to the incremental nature of the

^aThis term to describe the approach was coined by Professor George Rose of Johns Hopkins University, who also suggested this problem to us.

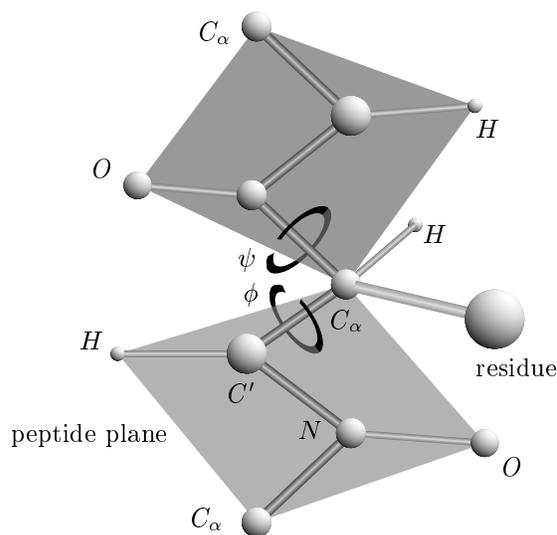


Figure 1: Illustration of backbone torsion angles.

process, the system is never far from equilibrium, unless there are bifurcation points along the pathway at which the energy function is *catastrophic*⁴. Such points, if they exist, would show up as points where the Hessian of the energy function has non-positive-definite eigenvalues.

2.2 Parametrization using global coordinates

We first examine possible degrees of freedoms of protein molecules. A protein, to a first approximation, contains fixed bond lengths, bond angles, and planar segments. The only degrees of freedom in this approximation are rotations between planar segments, i.e. ϕ , ψ angles⁵ (see Figure 1). Furthermore, every third bond along the backbone is effectively rigid due to its partial double bond character. The protein backbone thus possesses a set of rather constrained degrees of freedom. In comparison to the protein backbone, the sidechains exhibit less flexibility and thus can be treated as rigid attachments to backbone atoms.

The torsional angles provide a local set of coordinates for the protein conformational change. It is more convenient, however, to use a set of global coordinates for the translational and rotational displacement of each atom. Therefore, it is useful to parameterize the degrees of freedom of protein chains

using global coordinates and to use a transfer matrix between local and global coordinates.

As mentioned before, there are only 6 torsional degrees of freedom for every three consecutive residues. If the i -th backbone atom is fixed, the translational and rotational displacement of the $(i + 9)$ -th backbone atom may be related to the increments of the intervening 6 torsion angles by a transfer matrix as follows. In general, incremental changes of the 6 torsional degrees of freedom of such a segment may be related to those of the 6 degrees of freedom for relative translations and rotations between the endpoint backbone atoms i and $(i + 9)$ by a nonsingular transfer matrix. Such backbone segments of 9 bond-lengths each form the computational elements of our method, and endpoint atoms of these segments are referred to as *nodal* atoms. It is important to note that the incremental movement of any atom on a computational element may be related linearly to the incremental changes in the degrees of freedom of the nodal atoms only. Thus, the set of translations and rotations of the nodal atoms is our complete set of degrees of freedom.

2.3 Potential energy functions

The total energy in each step of the incremental energy minimization only contains potential energy but not kinetic energy. We use empirical potentials called *force field* detailed, for instance, in reference⁶.

The total potential energy has the following form

$$E(\mathbf{x}, \boldsymbol{\lambda}) = E_{tor}(\mathbf{x}) + E_{vdW}(\mathbf{x}) + E_H(\mathbf{x}) + E_{dis}(\mathbf{x}, \boldsymbol{\lambda}), \quad (2.1)$$

where E_{tor} is the torsional potential which is a quadratic approximation of the Pitzer potential, E_H is the hydrogen bond potential (we use a Lennard-Jones form), E_{vdW} is the Lennard-Jones van der Waals potential, and E_{dis} is the incrementally introduced distance restraint potentials of the form $E_{dis}^{ij} = \frac{1}{2}\lambda_{ij} (|\mathbf{x}_i - \mathbf{x}_j| - l_{ij})^2$, where $l_{ij} > 0$ is the ideal distance we desire and λ_{ij} is a control parameter.

The Hessian (matrix of second derivatives with respect to the degrees of freedom) of the energy function is the stiffness matrix used in the Newton method outlined below.

2.4 Variational methods

The energy minimization problem, under general assumptions, consists of a sequence of linearly constrained quadratic minimization (LCQM) problems.

The new equilibrium conformation $\mathbf{x}_{k+1} = \mathbf{x}_k + \delta\mathbf{x}_k$ can be computed using Newton's method:

$$\text{minimize } \delta E_k = \frac{1}{2} \delta\mathbf{x}_k^T K_{\mathbf{x}_k} \delta\mathbf{x}_k - \delta L_{\mathbf{x}_k}^T \delta\mathbf{x}_k, \quad (2.2)$$

subject to $B\delta\mathbf{x}_k = 0$.

The corresponding constrained minimum can be computed by solving a linear set of equations. Since this procedure uses only a local quadratic expansion of the total potential energy E_{k+1} in the region of \mathbf{x}_k , several Newton iterations are needed to converge to the true equilibrium conformation.

3 Modeling Experiments

3.1 Calmodulin peptide binding

Calmodulin is a monomeric (single subunit) protein consisting of a chain of 148 amino acids that is capable of binding to up to 4 Ca^{2+} ions. The molecule is 'strikingly' dumbbell-shaped⁸ with an eight-turn solvent-exposed central alpha helix connecting the two pairs of EF-hand domains (see Figure 2 (a)).

Calmodulin exerts its role by activating more than 20 different enzymes in eukaryotic cells, and the calmodulin-binding domain (peptide) has been identified in several enzymes. Several experiments reveal that both ends of such peptides are simultaneously capable of interacting with the two EF-hand domains of calmodulin⁹. This result implies that the interconnecting central alpha helix has to be structurally flexible in order for the two lobes of calmodulin to bind to the ends of the target amphipathic peptide.

In this experiment, we study the conformational change of calmodulin when it binds to a synthesized 26-residue peptide from skeletal-muscle myosin light-chain kinase. The 3-dimensional structures of both the unbound calmodulin and the calmodulin-peptide complex are from the Protein Data Bank at Brookhaven National Laboratory. The unbound calmodulin is from *drosophila melanogaster* expressed in *E coli*. The 3-dimensional structure (4cln), by X-ray crystallography, is refined at 2.2 Å resolution using molecular dynamics. The calmodulin in the bound calmodulin-peptide complex is also from *drosophila* and the peptide is synthesized from rabbit skeletal myosin light-chain kinase. The 3-dimensional structure (2bbm) is determined by multidimensional NMR.

As shown in Figure 2 (b), the 3-dimensional structure of the calmodulin-peptide complex is roughly a compact ellipsoid. The calmodulin N-terminal and C-terminal lobes are wrapped around the peptide and bind to opposite ends of it.

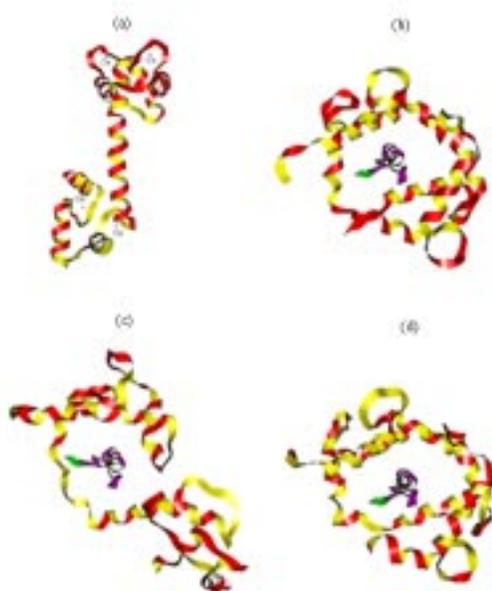


Figure 2: The three dimensional structure of (a) calmodulin (from PDB), (b) calmodulin-peptide complex (from PDB), (c) calmodulin-peptide complex (simulation result with 1 distance constraint), and (d) calmodulin-peptide complex (simulation result with 24 distance constraints).

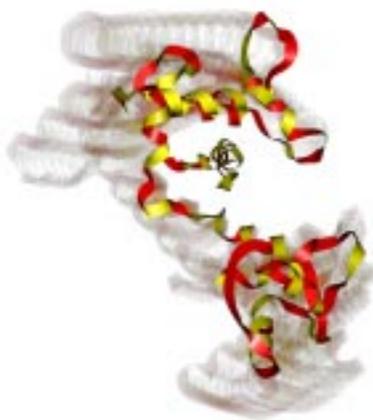


Figure 3: Motion-blur picture of calmodulin-peptide binding (simulation result).

To begin, we first apply just one distance constraint between a pair of atoms near the centers of N-lobe and C-lobe respectively. The simulation result is shown in Figure 2 (c). Compared to the three dimensional structure of the bound calmodulin-peptide complex, it has broadly similar features of binding. In particular, it reveals the fact that the long central helix serves as an *expansion joint*. It is also observed that the two lobes do not directly approach each other. There are significant relative rotations (a twist of around 90 degrees) between the two lobes.

With more careful analysis, it is found that there are extensive interactions between calmodulin and the peptide during the binding. All seven basic residues of the peptide make salt bridges with calmodulin. It is thus more reasonable to introduce distance constraints between calmodulin and the peptide. All together there are 24 distance constraints which have been chosen based on the contacts and the simulation result is shown in Figure 2 (d). Figure 3 is a motion-blur figure of one intermediate conformation.

The simulation result is very close to the three dimensional structure of the bound calmodulin-peptide from the Protein Data Bank shown in Figure 2 (b). There are only minor differences in the lobes, and the large conformational change in calmodulin upon peptide binding is manifested almost completely by changes of helix ϕ and ψ angles of residues 73 to 77. The two lobes bend about 100 degrees and twist about 120 degrees. There are about 185 contacts ($< 4\text{\AA}$) between the peptide and calmodulin.

Quantitatively, the differences (between simulated and observed bound states) of the distances between every two alpha carbon atoms are used to construct a distance deviation measure: $e_{ij} = |\mathbf{x}(C_\alpha^i) - \mathbf{x}(C_\alpha^j)| - |\hat{\mathbf{x}}(C_\alpha^i) - \hat{\mathbf{x}}(C_\alpha^j)|$. The maximum deviation is 1.6\AA which is less than the 2.2\AA resolution of the structure from NMR. The RMS deviation is 0.012\AA .

3.2 Cooperative hemoglobin oxygen binding

Hemoglobin, the oxygen-carrying pigment in red blood cells, has a molecular weight of 64,458¹⁰. It has a protein part, the globin, and a nonprotein part, the heme. Normal adult human hemoglobin (Hb A) consists of four polypeptide chains - two alpha chains, each containing 141 amino acids, and two beta chains, each containing 146 amino acids. Each of these chains is folded around a unit of heme and an atom of iron. Globin is bound to heme by a coordinated bond linking a histidine (F8 His) and the five-coordinated iron. The sixth-coordinated position of iron can be occupied by a molecule of oxygen, O₂.

Hemoglobin changes its quaternary structure upon oxygen binding¹¹. It is an allosteric transition between the oxygenated hemoglobin (R state) and

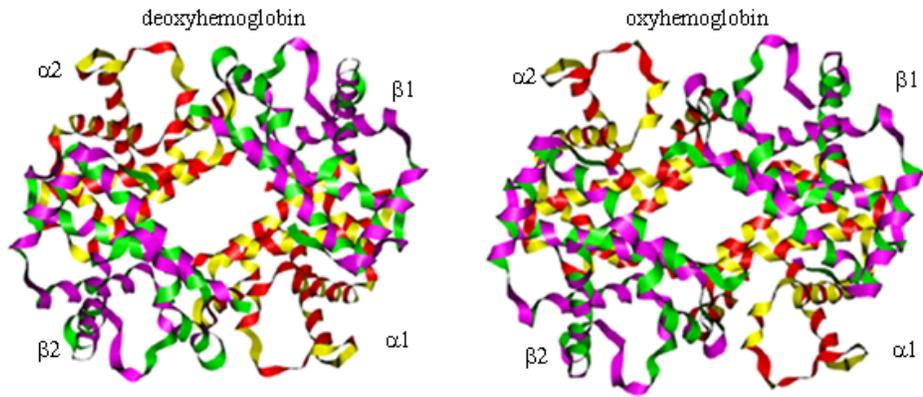


Figure 4: Structures of deoxyhemoglobin and oxyhemoglobin (from PDB).

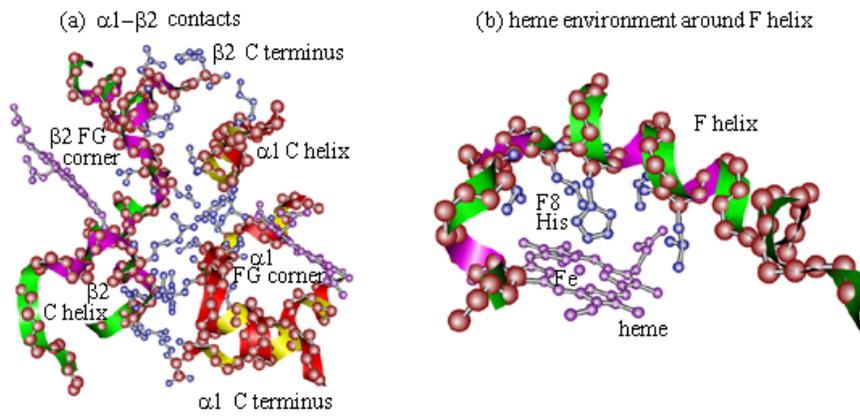


Figure 5: (a) Subunit contacts at $\alpha 1\beta 2$ interface; (b) Heme environment around F helix.

deoxygenated hemoglobin (T state) and these two states are both equilibrium states. The T and R structures differ in the relative arrangement and the conformations (quaternary and tertiary structures respectively) of the four subunits.

The three dimensional structures of human deoxyhemoglobin (1hga) and oxyhemoglobin (1hho) are both from Protein Data Bank using X-ray crystallography at resolution of 2.4\AA as shown in Figure 4. Before the discussion of the simulation steps, we first look at some important differences in structure between deoxyhemoglobin and oxyhemoglobin.

- Each $\alpha_1 - \beta_1$ and $\alpha_2 - \beta_2$ dimer of the molecule moves roughly as a rigid body, and the two dimers slide over one another.
- The structure of the interface between α_1 and β_1 remains the same. The $\alpha_1 - \beta_1$ interface therefore gives a frame of reference for the description of tertiary structure changes.
- There is a small movement at the contact between the α_1 FG corner and the β_2 C helix (flexible joint), while there is a large movement of about 6\AA at the contact between α_1 C helix and β_2 FG corner (switch region). Each position of this switch is stabilized by a different set of hydrogen bonds. Other hydrogen bonds in the two contacts of deoxyhemoglobin are broken in oxyhemoglobin. The contact region of deoxyhemoglobin is shown in Figure 5 (a).
- Tilting of the asymmetric proximal histidine F8 is associated with a movement of the iron atom towards the heme plane. It also results in the motion of the F helices and FG corners relative to the $\alpha_1 - \beta_1$ interface. The environment around that area is shown in Figure 5 (b).

The objective of our simulation is not only to reproduce the oxyhemoglobin from deoxyhemoglobin, but also to study the possible triggers for such an allosteric transition. In the literature, there are many hypotheses presented by different researchers about the mechanism of cooperative oxygen binding.

One such hypothesis is the movement of the iron atom. In deoxyhemoglobin, iron is out of the heme plane and it moves towards the heme plane when oxygen binds to it. The first simulation we did, therefore, was to introduce distance constraints between F8 His and Val E11 because they come closer together with the iron movement. But we did not observe any rearrangement of the quaternary structure. We also failed to observe the allosteric transition with some other hypotheses¹⁰ for the trigger mechanism.

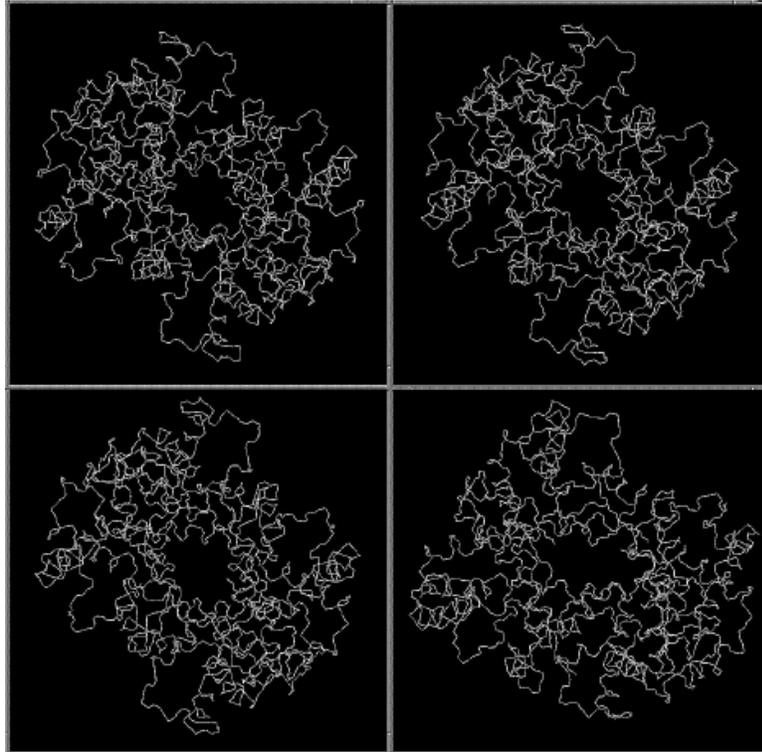


Figure 6: Hemoglobin-oxygen binding: clockwise from top left, (a) PDB data, (b) simulation based on breaking hydrogen bonds, (c) simulation showing lack of cooperativity without hydrogen bonds, and (d) simulation based on a trigger mechanism.

Perutz has emphasized the importance of the hydrogen bonds in the contacts between different dimers¹¹. Deoxyhemoglobin is held in its T state mainly by a network of hydrogen bonds and salt bridges that connect the amino- and carboxy-terminal regions of both chains. In oxyhemoglobin, this hydrogen bond network is broken. For consistency checking, we introduce a set of distance constraints to break those hydrogen bonds. The simulation result reproduces the cooperative allosteric transition and is very close to data from the Protein Data Bank. It is shown in the upper-right of Figure 6 (the upper-left of Figure 6 is the data from Protein Data Bank).

To find the real trigger, we notice that the straightening up of the tilted F8 helix and the corresponding movement of F helix is very essential for coop-

erative oxygen binding. We therefore introduce a set of distance constraints to simulate this and the result is shown in the lower-left of Figure 6. The result is very close to the data from the Protein Data Bank with an RMS deviation of 0.023. Moreover, the simulation reproduces the *consequence* (in contrast to the previous simulation experiment, where this was used as an input) that the hydrogen bond network in deoxyhemoglobin is broken during the allosteric transition.

To further verify the correctness of our model, we have done another experiment. With the same trigger, we simulate oxygen binding without the hydrogen bond network. This corresponds to a real experiment where the hydrogen bond network was broken using chemical methods². Our result is the same as the result of the real experiment: there is no longer any cooperativity due to oxygen binding. As shown in the lower-right of Figure 6, the four subunits just independently expand themselves.

4 Discussion and conclusions

4.1 Advantages of the approach

From the numerical computation point of view, the major significance of our approach is the novel parameterization of the degrees of freedom in terms of the global translations and rotations of nodal atoms (in a fixed inertial reference frame). Such a parameterization has significant advantages, especially for problems involving large conformational changes.

First, it decouples the movements of the atoms (in terms of the degrees of freedom) along the chain of the protein backbone. To characterize the movement of any atom along the polypeptide in terms of torsional angles requires all the torsional angles prior to that atom. In our parameterization, however, the movement of any atom can be uniquely determined by the translational and rotational displacements of the two nodal atoms of the specific element on which that atom lies. Thus, all the atom movements are fully decoupled and the transformation matrix is banded with column block width 12.

Second, all the potentials can be easily pulled back to the nodal atoms of elements through the transformation matrix. If we look at the form of the stiffness matrix K , it is obvious that the total stiffness matrix will be a full matrix in terms of torsional angle coordinates. In terms of our set of nodal degrees of freedom, however, it has a special form. The stiffness matrix in our parameterization due to backbone torsional energy K_{tor} is strictly banded with bandwidth 12. Some off-diagonal blocks are introduced due to hydrogen bond potentials and van der Waals potentials. Even so, the total stiffness matrix is

relatively banded with some off-diagonal blocks and it is thus quite sparse. The sparsity leads to fast computation in solving the linearly constrained quadratic minimization problem.

4.2 Comparison with SCULPT

SCULPT is an interactive molecular modeling system¹³ which is similar in spirit to the *macroscopic* approach of *Proteinmorphosis*. It is therefore worth comparing the two systems. There are three major differences between the modeling approaches of the two systems which we will briefly discuss here.

First, the finite element parameterization employed in *Proteinmorphosis* is different. As discussed in the preceding subsection, our parameterization enables an effective kinematic decoupling of an atom's incremental movements from those atoms which do not belong to the corresponding computational element.

Second, the present system uses a (second order) Newton method to solve the incremental constrained quadratic minimization problem, in contrast to the first order gradient descent method which SCULPT uses. Second order methods are generally more stable, usually do not require step size adjustments and yield significantly faster convergence, and we consequently expect *Proteinmorphosis* to be a more effective tool for simulating large changes in conformation. For calmodulin (1374 atoms), our program achieves 2.5 updates per second on a single processor of the SGI Infinite Reality and 11.2 updates per second on two processors of the Cray T94. With about 1000 updates for the complete conformational change of calmodulin, the corresponding figures for the entire computation are about 7 minutes and 90 seconds respectively.

Finally, SCULPT uses springs attached to external fixed points in order to apply external forces. In contrast, *Proteinmorphosis* applies distance constraints between atom pairs or pairs of groups of atoms. This is a more natural mechanism for simulating binding processes; no *external* forces are involved. The constraints only serve to drive the system to a bound state equilibrium; the constraint forces are small when the final state is reached and only the intrinsic forces due to interatomic potentials are present at the end of the process.

References

1. W. Lawton, T. Poston, R. Raghavan, S. R. Ranjan, R. Viswanathan, Y. P. Wang and Y. Yu, "Variational Methods in Biomedical Computing", *the Proceedings, Conference on Computational Science for the 21st Century, Tours, France*, (John Wiley, Sussex, England, 1997).

2. S. S. Antman, *Nonlinear Problems in Elasticity*, (Springer-Verlag, Berlin, 1995).
3. J. Monod, J. Wyman and J-P. Changeaux, "On the nature of allosteric transitions: a plausible model", **12**, 88-118, (1965).
4. T. Poston and I. Stewart, *Catastrophe theory and its applications*, Dover Publishers, London, 1997.
5. T.E. Creighton, *Proteins: structures and molecular properties*, W.H. Freeman, New York, 1993.
6. W. D. Cornell, P. Cieplak, C. I. Bayly and *et. al.*, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules", *Journal of the American Chemical Society*, **117**, 5179-5197, (1995).
7. S. Lifson, "Potential energy functions for structural molecular biology", *NATO Advanced Study Institute/FEBS Advanced Course No. 78. Current Methods in Structural Molecular Biology*, 1-27, (1981).
8. S. Lippard and J. Berg, *Principles of Bioinorganic Chemistry*, (University Science Books, Mill Valley, California, 1994).
9. W. E. Meador, A. R. Means and F. A. Quiocho, "Modulation of calmodulin plasticity in molecular recognition on the basis of X-ray structures", *Science*, **262**, 1718-1722, (1993).
10. R. E. Dickerson and I. Geis, *Hemoglobin: Structure, Function, Evolution, and Pathology*, (The Benjamin/Cummings Publishing Company, Inc., Menlo Park, California, 1983).
11. M. Perutz, "Stereochemistry of cooperative effects in haemoglobin", *Nature*, **228**, 726-739, (1970).
12. D. Barrick, Nancy Ho, V. Simplaceanu, F.W. Dahlquist and Chien Ho, "Molecular signaling in hemoglobin: a test of the Perutz model for cooperativity", University of Oregon preprint, (1996).
13. M. C. Surles, J. S. Richardson, D. C. Richardson and F. P. Brooks, Jr., "Sculpting proteins interactively: continual energy minimization embedded in a graphical modeling system", *Protein Science*, **3**, 198-210, (1994).