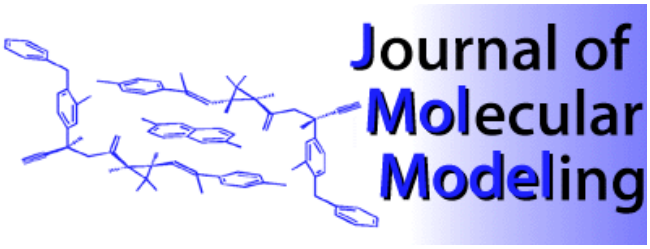
J. Mol. Model. **1999**, *5*, 17 – 36



FULL PAPER

A Civil Engineering Model of Protein Conformational Change

Wayne Lawton[1,2], Solaiyappan Meiyappan[1], Raghu Raghavan[1], Raju Viswanathan[1], and Yi Yu[1]

[1]Center for Information Enhanced Medicine; National University of Singapore; #03-14, 10 Science Park Road; The Alpha, Singapore Science Park II; Singapore 117684. Tel. (65) 8727624; Fax (65) 7756721.
E-mail: yuyi@ciemed.nus.edu.sg

[2]Department of Computational Science, National University of Singapore.

Received: 15 October 1998/ Accepted: 30 December 1998/ Published: 26 February 1999

Abstract We present a variational approach for the simulation of large conformational changes of proteins (including multiple protein chains/ligands) which takes advantage of their cross-linked one-dimensional nature, a structure which often occurs in civil engineering. Conformational changes are computed by incremental energy minimisation. We use an efficient finite element method for finding equilibria of complexes composed of inter-linked chains; this method is based on recent advances in the description of one-dimensional elasticity. Protein backbone elasticity, van der Waals repulsions, hydrogen bonds and salt bridges are taken into account, together with user-defined geometric distance constraints that may be imposed for purposes of simulating various binding processes based on chemical knowledge. These computational methods have been integrated into a system, *Proteinmorphosis*, which includes interactive visualisation. The conformational change of calmodulin upon peptide binding is examined as a first experiment. Allostery in hemoglobin, which consists of a cooperative oxygen binding mechanism, is a second, more sophisticated, numerical experiment. Different modelling strategies are designed to understand the allostery. The results for both molecules are consistent with existing hypotheses, and reproduce the known atomic positions after binding to within the experimental error. The modelling system is part of an on-going program to model structural biology, from protein structure to cell and tissue properties.

Keywords Protein conformational change, Incremental energy minimisation, Civil engineering model, Allostery

Introduction

The conformations and conformational changes of proteins are vital to their function. While the underlying processes responsible for such changes are ultimately quantum mechanical in nature, a complete modelling treatment of proteins at this level is beyond the reach of current computing technology. In this article, we describe a method, using a novel parametrisation, for simulating (possibly large) conformational changes within a classical mechanics framework, where quantum effects such as reactions at local sites provide a trigger for structural changes at the scale of the entire protein.

Correspondence to: Y. Yu

Using a macroscopic formulation of protein conformational changes, our modelling system *Proteinmorphosis* aims at fast and interactive computation with large proteins, in contrast to the very detailed, computationally intensive microscopic modelling. In many instances such a *civil engineering* [a] description, with its interactive flexibility, may be the most useful. The protein chain is viewed as a structural frame: a one-dimensional, elastic object with cross-linking interactions due to hydrogen bonds, salt bridges and disulphide bridges, in addition to the important 'excluded volume' van der Waals interactions. It is well known that the excluded volume effect is one of the most important determinants of protein structure. In this structural view, such intra-chain interactions act to determine conformation by acting on the protein backbone. The elastic properties of the backbone may be described in a manner analogous to the one-dimensional elasticity of thin rods [1, 2]. 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 engineering. We approximate a ligand binding process 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. Our current focus is on modelling conformational changes in the proteins calmodulin and hemoglobin. We also developed a convenient approach for the visualisation of large, complex proteins in a dynamic, interactive environment [3].

The model

Conformational change and allostery

Conformational changes occur when proteins interact with other molecules. These changes largely affect the biological functions of the proteins. Proteins may bind very tightly and specifically to other proteins, generating large complexes; and to nucleic acids, especially when controlling their replication and expression. In most cases, proteins interact with smaller molecules or ligands. In some cases, only small conformational changes occur near the local region of the binding sites. In other cases, substantial changes of protein conformations upon ligand binding are observed with both local movement and global rearrangement of subunits. Such conformational changes may help to maximise the interactions between the protein and the ligand and to minimise interactions with other components of the solvent.

[a] This term to describe the approach was coined by Professor George Rose of the Johns Hopkins University, who also suggested this problem to us.

The other role of protein conformational changes upon ligand binding is to produce functional alterations at other binding sites of the protein, as in allosteric proteins. Allosteric proteins control and coordinate chemical events in the cell. The theory as published in [4] was concerned mainly with cooperativity and feedback inhibition of enzymes. It says that cooperative ligand binding at multiple sites may arise in proteins with two or more conformations in equilibrium. It predicts that such proteins are likely to have several subunits symmetrically arranged, and that the conformations would differ by the arrangement of the subunits and number and/or energy of the bonds between them. In one conformation the subunits would be constrained by strong bonds that would resist the tertiary structure change needed for ligand binding. This state is generally called the T (for tense) state. In the other conformation, these constraints are relaxed, and the state is called the R (for relaxed) state for. In the transition between them, the symmetry of the protein molecule is conserved, so that the activity of all its subunits will be either equally low or equally high. This arrangement has the advantage that no direct interaction occurs between the ligands at distant binding sites and the regulatory metabolite which controls its activity. Control is entirely due to a change of protein conformation induced in the protein when it binds to the appropriate allostery-inducing ligand.

Incremental energy minimisation

Our method for determining the conformational change is one of incremental energy minimisation. We assume that, before binding, both protein and ligand are in their folded equilibrium conformations with the lowest energies. During the binding process, hydrogen bonds and salt bridges are breaking and forming, 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 proteins [4], this pathway becomes the stable transition between T state and R state with the alterations in quaternary structures.

This is the conceptual underpinning of incremental energy minimisation. 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 process, the system is never far from equilibrium, unless there are bifurcation points along the pathway at which the energy function is *catastrophic* [5]. Such points, if they exist, would show up as points where the Hessian

(matrix of second derivatives) of the energy function fails to be positive definite.

We first examine the possible degrees of freedom of protein molecules. A protein, to a first approximation, contains fixed bond length, bond angles, and planar segments. Figure 1 shows two sequential segments in a protein chain with vectors representing bonds between atoms and gray areas denoting planar regions. To this approximation, the only degree of freedom in the figure are rotations between planar segments, i.e. ϕ, ψ angles. 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 protein backbone, the sidechains exhibit less flexibility and therefore may be treated as rigid attachments to appropriate backbone atoms.

The total energy of the protein-ligand complex is the sum of their (backbone) elastic energy E_{el} and the interaction energy E_{int} (both within and across molecules). The elastic energy depends only on the shape (*i.e.*, equivalence up to rigid motion) of the backbone conformation (*i.e.*, uniquely specified by the backbone torsional ϕ, ψ angles). The interaction energies which arise are typically due to van der Waals repulsions, hydrogen bonds and salt bridges and depend only on the positions of all the atoms.

The protein-ligand complex phase manifold is $\mathcal{M} = T^d$, where T^d is a d-dimensional torus ($d = d_p + d_l + 6$ and d_p, d_l are numbers of torsional angles of the protein and the ligand, respectively) and the total potential energy E: $\mathcal{M} \to R$ is

$$E(\mathbf{x}) = E_{el}(\mathbf{x}) + E_{int}(\mathbf{x})$$
$$= E_{el}(\mathbf{x}_p) + E_{el}(\mathbf{x}_l) + E_{int}(\mathbf{x}_p, \mathbf{x}_l, \mathbf{x}_r)$$

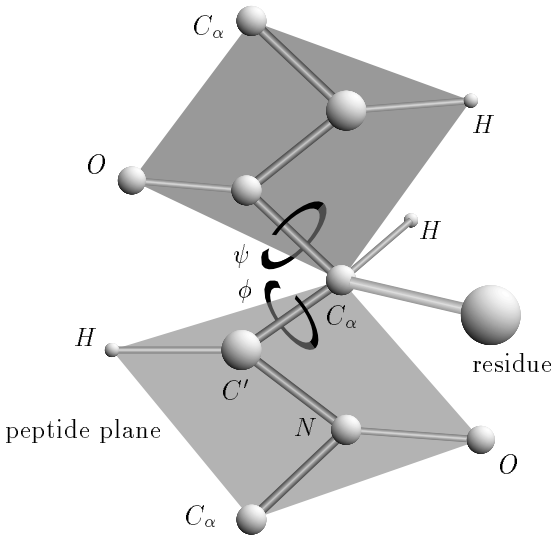


Figure 1 *Illustration of backbone torsion angles*

where $\mathbf{x} = [x_p\ \mathbf{x}_l\ \mathbf{x}_r] \in \mathcal{M}$. The elastic energy can be decoupled into two parts $E_{el}(\mathbf{x}_p)$ and $E_{el}(\mathbf{x}_l)$ with $\mathbf{x}_p \in T^p$ and $\mathbf{x}_l \in T^d$ being free variables, which uniquely determine the conformations of the protein and the ligand respectively. In contrast, the interaction energy E_{int} is fully coupled, depending also on the relative position and orientation of the protein and the ligand. The six degrees of freedom are captured by $\mathbf{x}_r \in T^6$.

In an allosteric ligand binding process, there are only a small number of equilibrium conformations. Without loss of generality, we assume there are only two equilibrium, in fact minimal energy, states: T and R states (the associated conformations are $\mathbf{x}^T$ and $\mathbf{x}^R$). That is, we have

$$\mathcal{J}(\mathbf{x}^T) = \mathcal{J}(\mathbf{x}^R) = 0 \tag{1}$$

and

$\mathcal{H}(\mathbf{x}^T)$ and $\mathcal{H}(\mathbf{x}^R)$ are positive definite, where $\mathcal{J}(\bullet)$ denotes the Jacobian and $\mathcal{H}(\bullet)$ denotes the Hessian of the total potential energy. To study the conformational changes from T state to R state, we need to find a trajectory from $\mathbf{x}^T$ to $\mathbf{x}^R$.

When the ligand binds to the protein, there is a dynamical process of breaking and forming of different sets of hydrogen bonds and salt bridges. It is also observed that there are substantial charge redistributions around the binding sites. The resulting equations are too difficult to solve on the computer in a short time with the dynamical form of the interaction energy. The only easy part of the interaction energy is the van der Waals interaction which depends solely on the position of all the atoms.

However, breaking and forming of hydrogen bonds and salt bridges as well as charge redistributions mainly occur in the transition period, especially at the bifurcation point when the protein-ligand complex overwhelms the potential energy barriers. These interactions do not change very much in the local region of both T and R equilibrium conformation states. van der Waals interactions play an important role in seeking minimum energy states around those local regions.

Once in the local region of the equilibrium R state, the elastic energy of the backbone, the potential energy of stable cross-linking bonds and the van der Waals interaction energy should be sufficient to locate the desired equilibrium. One way to introduce an approximate interaction energy that provides a stable *pathway* between (local) equilibrium states is to introduce an extra energy term based on a certain set of distance constraints. This energy is applied incrementally as

$$E_{dis}(\mathbf{x}, \mathbf{l}) = \sum_{i=1}^{n_d} \frac{1}{2} \lambda \left(\left| \mathbf{x}_1^i - \mathbf{x}_2^i \right| - l_i \right)^2 \tag{2}$$

$$= \frac{1}{2} \left(|P\mathbf{x}| - \mathbf{l} \right)^T \lambda \left(|P\mathbf{x}| - \mathbf{l} \right) \tag{3}$$

where n_d is the number of distance constraints, $\mathbf{l} = [l_1\ l_2 \ldots l_{n_d}]$ is the corresponding ideal distance set, $\boldsymbol{\lambda} = \mathrm{diag}(\lambda_1\ \lambda_2 \ldots \lambda_{n_d})$ is the stiffness constant set, and P is a difference operator such that $\mathbf{x} = [\mathbf{x}_1^1 - \mathbf{x}_2^1\ \ \mathbf{x}_1^2 - \mathbf{x}_2^2 \ldots \mathbf{x}_1^{n_d} - \mathbf{x}_2^{n_d}]$. Thus, the total potential energy becomes

$$\hat{E}(\mathbf{x},\boldsymbol{\lambda}) = E(\mathbf{x}) + E_{dis}(\mathbf{x},\boldsymbol{\lambda}) \quad (4)$$

Here $\boldsymbol{\lambda}$ is a vector of control parameters. The whole binding process is divided into steps k = 1, 2, 3,…. The control parameter is also a monotonically increasing function of the step k: $0 < \lambda_i(k) < \lambda_i(k+1)$, and $\lambda_i(k) \to \infty$ if $k \to \infty$, i=1, 2,…,n_d.

The incremental energy minimisation starts from the initial equilibrium state T. Given a control trajectory $\boldsymbol{\lambda}_k$, k=1,2,…, we compute a phase trajectory $m_k \in \mathcal{M}$ such that m_k is a local minimum of $\hat{E}\lambda_k$. Quadratic minimisation method is used to calculate the local minimum at each step. Assuming $\mathbf{x}_k$ is a local minimum conformation at step k and the control parameter at step k+1 is $\boldsymbol{\lambda}_{k+1}=\boldsymbol{\lambda}_k + \delta\boldsymbol{\lambda}_k$, we have

$$\hat{E}_k(\mathbf{x}_k) = E(\mathbf{x}_k) + E_{dis}(\mathbf{x}_k,\lambda_k) \quad (5)$$

$$\hat{E}_{k+1}(\mathbf{x}_k) = E(\mathbf{x}_k) + E_{dis}(\mathbf{x}_k,\lambda_{k+1}) \quad (6)$$

The change of the equilibrium conformation $\delta\mathbf{x}_k=\mathbf{x}_{k+1}-\mathbf{x}_k$ is computed as

where $\mathbf{n} = \dfrac{P\mathbf{x}_k}{|P\mathbf{x}_k|}$

Global parametrisation

The torsional angles provide a set of local coordinates for the protein conformational change. A global set of coordinates, in the sense of the translational and rotational displacement of each atom, however, is more convenient for our purposes.

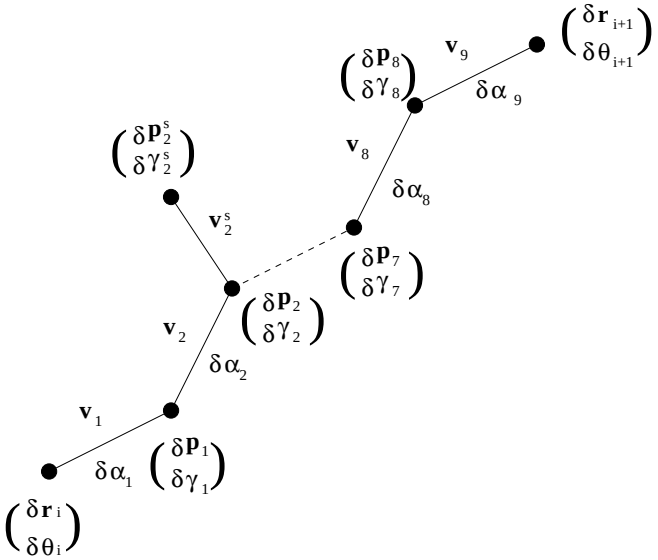


Figure 2 *Parametrisation of backbone degrees of freedom, in local and global coordinates*

Therefore, it is useful to parameterise changes in the degrees of freedom of protein chains using global coordinates and determine the transfer matrix between local and global coordinates.

As we mentioned before, there are only six torsional degrees of freedom for every three consecutive residues. If the i-th backbone atom is fixed, the (small) translational and rotational displacement of the i+10-th backbone atom is related to the increments of the intervening six torsion angles by a transfer matrix.

Figure 2 shows three consecutive residues and the first and the last backbone atoms are chosen as two nodal points. Their positions are denoted as $\mathbf{r}_i$ and $\mathbf{r}_{i+1}$ and their translational and rotational displacements are denoted as ($\delta\mathbf{r}_i^T$; $\delta\boldsymbol{\theta}_i^T$) and ($\delta\mathbf{r}_{i+1}^T$; $\delta\boldsymbol{\theta}_{i+1}^T$). Similarly, position and orientation displacements of the j-th backbone atom are written as ($\delta\mathbf{p}_j^T$; $\delta\beta_j^T$). Denoting the nine torsion angles and their changes as α_j and $\delta\alpha_j$, j=1,2,…,9, we have

$$\begin{pmatrix}\delta\mathbf{r}_{i+1}\\ \delta\theta_{i+1}\end{pmatrix} = \begin{pmatrix}\delta\mathbf{r}_i\\ \delta\theta_i\end{pmatrix} + \begin{pmatrix}\delta\theta_i\times\mathbf{v}_{1,10} + \sum_{k=1}^{8}\delta\alpha_k\mathbf{u}_k\times\mathbf{v}_{10k}\\ \sum_{k=1}^{9}\delta\alpha_k\mathbf{u}_k\end{pmatrix}$$

where $\mathbf{v}_{j,k}$ is the vector from backbone atom j to k. Writing $\mathbf{u}_j = \mathbf{v}_j/|\mathbf{v}_j|$ and $\mathbf{d}_j = \mathbf{u}_j \times \mathbf{v}_{j,10}$, we have

$$A_i\begin{bmatrix}\delta\alpha_1 & \delta\alpha_2 & \delta\alpha_4 & \delta\alpha_5 & \delta\alpha_7 & \delta\alpha_8\end{bmatrix}^T = D_i\begin{pmatrix}\delta\mathbf{r}_i\\ \delta\theta_i\\ \delta\mathbf{r}_{i+1}\\ \delta\theta_{i+1}\end{pmatrix}$$

where

$$A_i = \begin{pmatrix}\mathbf{d}_1^T & \mathbf{d}_2^T & \mathbf{d}_4^T & \mathbf{d}_5^T & \mathbf{d}_7^T & \mathbf{d}_8^T\\ \mathbf{u}_1^T & \mathbf{u}_2^T & \mathbf{u}_4^T & \mathbf{u}_5^T & \mathbf{u}_7^T & \mathbf{u}_8^T\end{pmatrix}$$

and

$$D_i\begin{pmatrix}-I & V_i & I & 0\\ 0 & -I & 0 & -I\end{pmatrix}$$

Here, the skew symmetric matrix V_i corresponds to the vector $\mathbf{v}_{1,10}$ – to every vector $\mathbf{a}$, we may associate a skew symmetric matrix written in terms of components of $\mathbf{a}$ as

$$A = \begin{pmatrix}0 & -a_3 & a_2\\ a_3 & 0 & -a_1\\ -a_2 & a_1 & 0\end{pmatrix}$$

For general values of $\mathbf{v}_{j,k}$, each A_i is stably invertible. If it is not, we must replace each singular or nearly singular A_i with a geometrically regularised version

$$A_i^r = U(D + \sigma I)V$$

where $\sigma > 0$ is small and $A_i = UDV$ is the singular value decomposition of A_i so that $A^r{}_i$ has 6 non-zero eigenvalues. Thus

$$\begin{bmatrix}\delta\alpha_1 & \delta\alpha_2 & \delta\alpha_4 & \delta\alpha_5 & \delta\alpha_7 & \delta\alpha_8\end{bmatrix}^T = \tilde{T}_i \begin{pmatrix}\delta\mathbf{r}_i \\ \delta\theta_i \\ \delta\mathbf{r}_{i+1} \\ \delta\theta_{i+1}\end{pmatrix}$$

with $\tilde{T}_i = \left(A_i^r\right)^{-1} D_i$. Once we get the element transfer matrix $\tilde{T}_i$, three rows of zeros are inserted and we have the corresponding transfer matrix T_i.

For each backbone atom $\mathbf{p}_j$, its position and orientation displacements are given in terms of changes in the nodal point degrees of freedom by a transfer matrix $T^b{}_j$:

$$\begin{pmatrix}\delta\mathbf{p}_j \\ \delta\beta_j\end{pmatrix} = T_j^b \begin{pmatrix}\delta\mathbf{r}_i \\ \delta\theta_i \\ \delta\mathbf{r}_{i+1} \\ \delta\theta_{i+1}\end{pmatrix} \tag{8}$$

where

$$T_j^b = \begin{pmatrix}\mathbf{d}_{1,j}^T & \mathbf{d}_{2,j}^T \ldots \mathbf{d}_{j-1,j}^T & 0 \ldots 0 \\ \mathbf{u}_1^T & \mathbf{u}_2^T \ldots \mathbf{u}_j^T & 0 \ldots 0\end{pmatrix} T_i + \begin{pmatrix}I & -V_{i,j} & 0 & 0 \\ 0 & I & 0 & 0\end{pmatrix}$$

Here, $\mathbf{d}_{k,j} = \mathbf{u}_k \times \mathbf{v}_{k,j+1}$ and the skew symmetric matrix $V_{i,j}$ is associated with the vector $\mathbf{v}_{1,j+1}$. Assuming each sidechain is rigidly attached to the corresponding backbone alpha carbon, we have

$$\begin{pmatrix}\delta\mathbf{p}_j^s \\ \delta\beta_j^s\end{pmatrix} = \begin{pmatrix}\delta\mathbf{p}_j \\ \delta\beta_j\end{pmatrix} + \begin{pmatrix}\delta\beta_j \times \mathbf{v}_j^s \\ 0\end{pmatrix} = T_j^s \begin{pmatrix}\delta\mathbf{r}_i \\ \delta\theta_i \\ \delta\mathbf{r}_{i+1} \\ \delta\theta_{i+1}\end{pmatrix}$$

where $\mathbf{v}_j^s$ is the vector from the backbone atom $\mathbf{p}_j$ to the sidechain atom $\mathbf{p}^s{}_j$

With the above derivation of the transfer matrices for both backbone atoms and sidechain atoms with respect to the nodal points along the backbone chain, it is easy to assemble them together to form one total transfer matrix. Supposing the total number of atoms to be n and the number of nodal points to be m, we have

$$\Delta P = T\Delta R \tag{9}$$

where the $6n \times 1$ vector ΔP contains the translational and rotational (small) displacements of all the atoms, the $6m \times 1$ vector ΔR is the displacements of the nodal points, and T is a $6n \times 6m$ global transfer matrix. The atoms and nodal points are arranged in the same order as in the protein polypeptide chain. Within each residue, the order of atoms is as follows: nitrogen, alpha carbon, carbon, oxygen, and sidechain atoms. It is obvious that the global transfer matrix T is block diagonal with each block of size $6m_i \times 12$, where m_i is the number of atoms in the i-th residue.

Potential energy functions

The incremental energy minimisation method used in our civil engineering model is a quasistatic approach, i.e., the total energy in each step only contains potential energy but not kinetic energy. The total potential energy is assumed to be a function of the number and type of chemical species within the protein molecule and the distance between all pairs of atoms. To calculate the potential energy, it is divided into a number of terms which correspond to physical effects, which is the *force field* [6, 7]. We use standard forms of the various potential energy functions [6], except for the torsion potential energy, for which we use a quadratic form rather than the more commonly used cosine form (see equation(11) and the description following it); there is no special parameter adjustment in the simulations. We also use explicit potentials for hydrogen bonds and salt bridges [8], which some force fields do not employ.

In this paper, we assume the bond lengths and bond angles are fixed and the backbone torsional potential E_{tor} is the only contribution to the backbone elastic energy E_{el}. The interaction energy E_{int} contains van der Waals interaction energies E_{vdW} and hydrogen bond or salt bridge potentials E_H. Electrostatic potentials are not treated in this paper.

Thus, the total potential energy has the following form

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

Below, we will more specifically discuss each of these potential energy terms. We also need to evaluate the Hessian $\mathcal{H}$ and the Jacobian $\mathcal{J}$ of the total potential energy $\hat{E}$ in order to do the quadratic energy minimisation (7) at each step. The Hessian is also referred to as the *stiffness*, denoted by K, because it reveals the force-displacement relationship

$$\begin{pmatrix}\delta\mathbf{F} \\ \delta\tau\end{pmatrix} = K \begin{pmatrix}\delta\mathbf{r} \\ \delta\theta\end{pmatrix}$$

Similarly, the term *load* is often used to denote the negative Jacobian L^o–$\mathcal{J}(\mathbf{x})$. Rotations about bonds are described as torsion or dihedral angles, which are as usual taken to lie in the range –180° to +180°. Rotations about the N–C_α bond and about the C_α–C' bond of the peptide backbone are denoted by the usual ϕ and ψ angles. As mentioned earlier, due to the partial double bond character of the protein peptide chain, the torsion angle ω is almost rigid. The most com-

monly used form of the torsion angles potential is the Pitzer potential:

$$E_{tor} = \sum_i \frac{V_i}{2}\left[1+\cos\left(n_i\phi_i - \gamma_i\right)\right] \tag{10}$$

where V_i gives the energy barrier to rotation of the i-th torsion angle, γ_i is the reference angle where the torsional potential energy is a maximum and n_i is the multiplicity, that is, the number of potential minima in one full rotation.

However, it is found that the Pitzer potential is insufficient to give a full representation of the energy barriers of torsion angle change. The nature of torsional potentials is not fully understood and there is even some dispute about the periodicity and location of the energy minima of these functions [9]. High-level quantum mechanical data are now available on the conformational energies of the glycyl and alanyl dipeptides [6] which help to develop ϕ and ψ torsion parameters for the peptide backbone.

We use a simple harmonic form for ϕ and ψ torsional potentials in this paper; these angles lie in the –180° to 180° range from their respective references:

$$E_{tor} = \sum_{torsions} \frac{1}{2} k_{tor}\left(\phi_i - \gamma_i\right)^2 \tag{11}$$

where k_tor is the torsion constant. We use a value of k_{tor}=14 kcal/mole and use for γ_i the value taken by the corresponding torsion angle in the original conformation of the protein. We find computationally that the original conformation for the proteins we model here is in a state very close to equilibrium with the (full set of) potential choices we have made.

The multiplicity in the torsional potential (10) is at least in part handled in our model by the inclusion of non-bonded van der Waals interactions with all of the surrounding non-bonded atoms. The torsion and non-bonded interactions are highly coupled (see, for instance, [10]) and we find that the existence of the energy barrier present in the standard periodic form of the potential is reproduced in our approach as well.

With the above harmonic form of the torsional potential, it is easy to derive the corresponding stiffness matrix. Since

$$\delta E^i_{tor} = \frac{1}{2}\delta\phi_i^T \hat{K}^i_{tor}\delta\phi_i \tag{12}$$

and $\delta\phi_i = \tilde{T}_i\begin{pmatrix}\delta\mathbf{x}_i\\ \delta\mathbf{x}_{i+1}\end{pmatrix}$, the stiffness matrix with respect to the nodal atom i and i+1 is a 12 × 12 matrix

$$K^i_{tor} = \tilde{T}_i^T \hat{K}^i_{tor}\tilde{T}_i \tag{13}$$

where $\hat{K}^i_{tor} = k_{tor} I$.

Van der Waals interactions, which represent the excluded volume effect, as well as weak inter-molecular attractions, are described by a Lennard-Jones potential of the form:

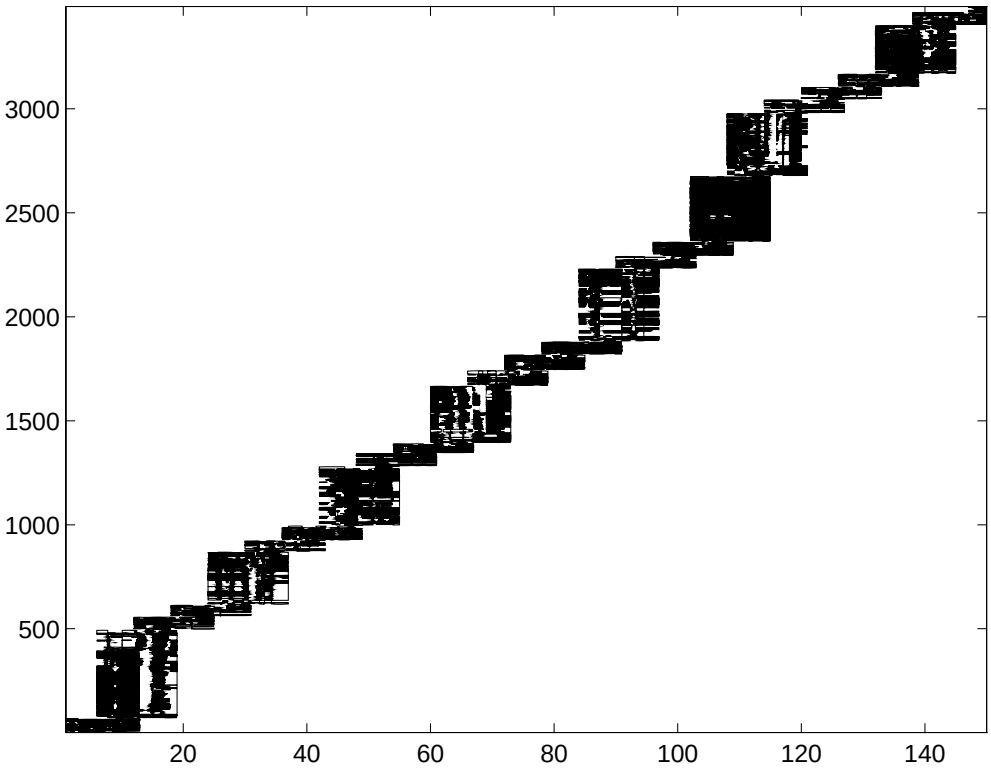


Figure 3 *Contour plot of the transformation matrix T used in the calmodulin peptide binding example*

$$E_{vdW}^{ij} = \frac{R^{12}}{r^{12}} - 2\frac{R^6}{r^6} \tag{14}$$

where $r=|\mathbf{x}_i - \mathbf{x}_j|$ and R is the summation of the van der Waals radii of the two atoms. Then

$$K_{vdW}^{ii} = \frac{\partial^2 E_{vdW}^{ij}}{\partial \mathbf{x}_i^2} = K_{vdW}^{jj} \tag{15}$$

and

$$K_{vdW}^{ij} = K_{vdW}^{ji} = -K_{vdW}^{ii} \tag{16}$$

The total van der Waals stiffness becomes

$$K_{vdW} = T^T \begin{pmatrix} K_{vdW}^{ii} & K_{vdW}^{ij} \\ K_{vdW}^{ji} & K_{vdW}^{jj} \end{pmatrix} T \tag{17}$$

where T is the transformation matrix discussed in the previous subsection.

It is computationally far too expensive to calculate the van der Waals interactions between all pairs of the atoms as the increase in computing time is of $O(n^2)$ where n is the total number of atoms. To reduce computing time, a cutoff radius R_{cut} is used, outside of which the van der Waals interactions are considered to be negligible and therefore not included. For most cases, we are more interested in the excluded volume effect, i.e. the van der Waals repulsion, and thus the cutoff radius is chosen to be the sum of the van der Waals radii of the two interacting atoms. The calculation of the van der Waals interactions then becomes a collision de-

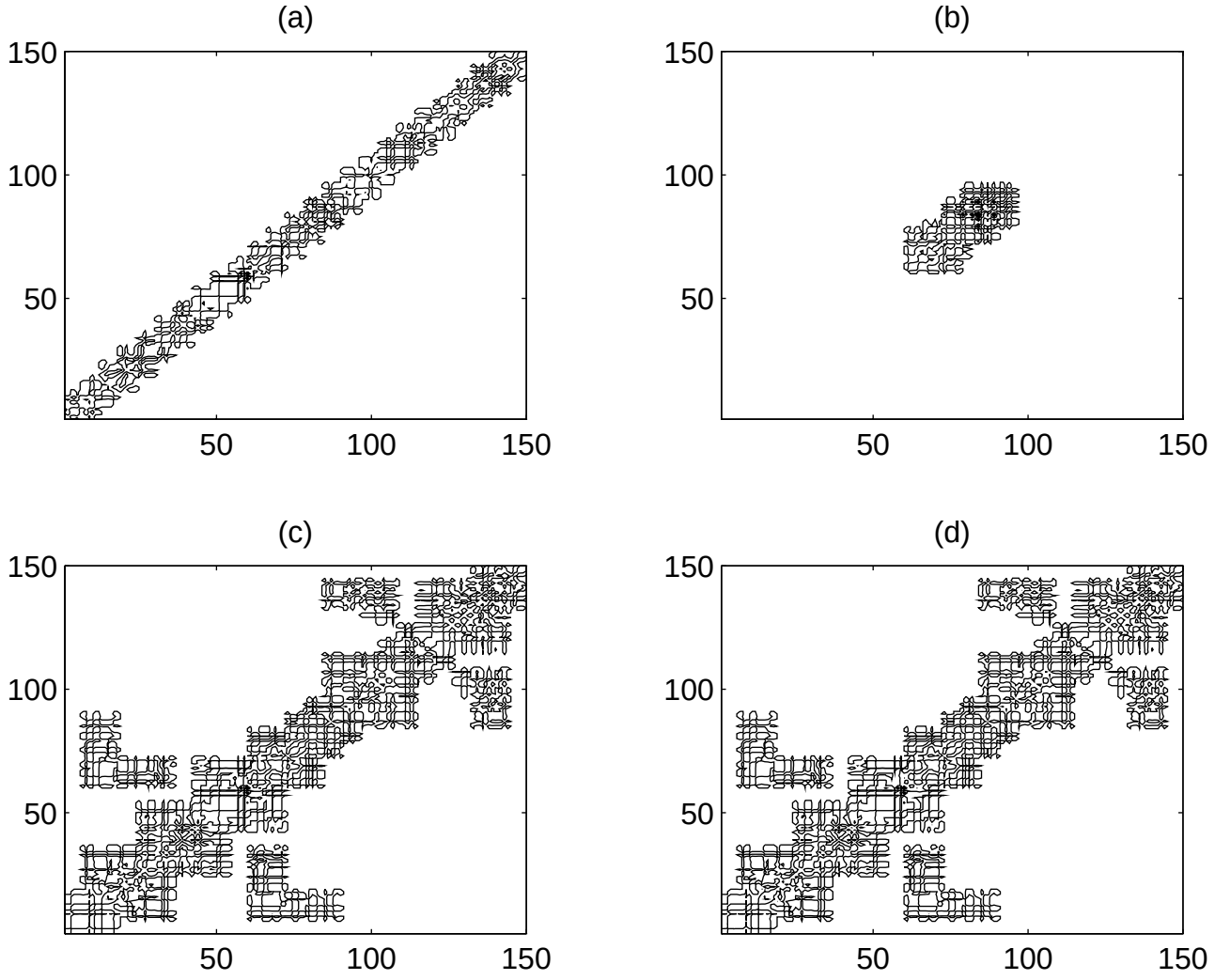


Figure 4 *Contour plot of the stiffness matrices used in the calmodulin peptide binding example: (a) K_{tor}; (b) K_H; (c) K_{vdW}; (d) K*

tection problem. In the appendix, we provide an efficient (linear in time and space) algorithm to detect collisions when each atom in the protein molecule can be approximated as a sphere with radius equal to its van der Waals radius. Similar algorithms are used in the molecular dynamics literature.

We use the following (4,8) Lennard-Jones potential for the hydrogen bond:

$$E_H^{ij} = \frac{A_0}{r^8} - \frac{B_0}{r^4} \tag{18}$$

where $r=|\mathbf{x}_i-\mathbf{x}_j|$.

For salt bridges, we use the same form of the potential functions except for different selections of the two parameters A_0 and B_0. Currently all hydrogen bonds and salt bridges are specified in the beginning of the modelling in order to preserve desired secondary structures as well as the interactions between different subunits (in, say, a tetramer as in hemoglobin).

In a full computational scheme for ligand binding, we would very carefully choose a force field that has accurate interaction potentials. In this more modest treatment, we focus on allostery, i.e. we assume ligand binding is known, and examine its effects on protein conformation. We do this by introducing distance constraint potentials between pairs of atoms at $\mathbf{x}_i$ and $\mathbf{x}_j$ of the form

$$E_{dis}^{ij} = \frac{1}{2}\lambda_{ij}\left(\left|\mathbf{x}_i - \mathbf{x}_j\right| - \mathbf{l}_{ij}\right)^2 \tag{19}$$

where $l_{ij} > 0$ is the ideal distance and λ_{ij} is the control parameter. The total stiffness is just the summation of the individual stiffness matrices:

$$K = K_{tor} + K_{vdW} + K_H + K_{dis} \tag{20}$$

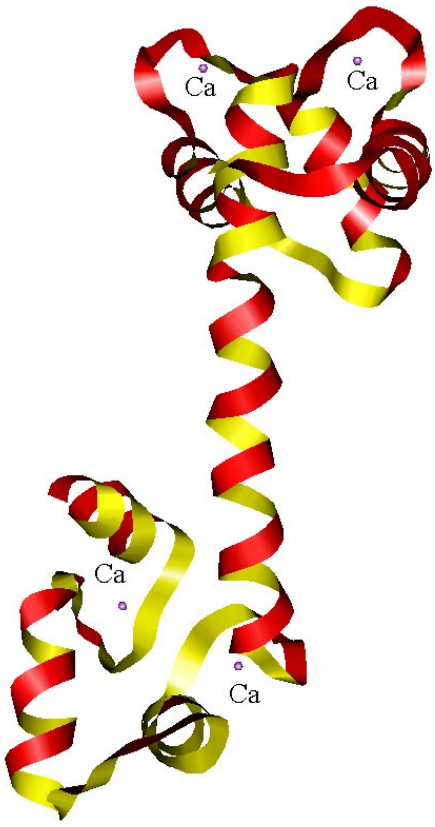


Figure 5 *Calcium binding protein—calmodulin (from PDB)*

Variational methods

The energy minimisation problem formulated in the last section is to compute a phase trajectory given a control trajectory. In addition to the distance constraints we introduced in order to approximate the dynamical interaction potentials, it is also possible to have other (linear) constraints. For example, we may like to fix some atoms in the protein-ligand complex, or make some sub-structures rigid. Such constraints are usually incrementally linear and can be written as $B\mathbf{x} = c$. Thus, under general assumptions, the energy minimisation problem reduces to a sequence of linearly constrained quadratic minimisation (LCQM) problems. We use k as the index of the LCQM steps and k=1,2,…. At each step *k*, assume $\mathbf{x}_k$ is an equilibrium conformation so that we have (5), and $B\mathbf{x}_k = c$. Applying an increment of the distance constraint energy, $\delta E_{\text{dis}}(\mathbf{x}_k, \delta\boldsymbol{\lambda}_k)$, the total energy becomes (6) and $\mathbf{x}_k$ is no longer an equilibrium conformation because the Jacobian of the total energy at $\mathbf{x}_k$, is not zero. The new equilibrium conformation $\mathbf{x}_{k+1} = \mathbf{x}_k\ \delta\mathbf{x}_k$ can be computed using Newton's method:

$$\text{minimize } \delta E_k = \frac{1}{2}\delta\mathbf{x}_k^T K_{\mathbf{x}_k}\delta\mathbf{x}_k - \delta L_{\mathbf{x}_k}^T\delta\mathbf{x}_k \tag{21}$$

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

The constrained minimum can be computed by solving the associated variational equation directly or by applying a relaxation method such as Uzawa's algorithm to a suitably augmented version [11]. The variational equation we get is linear: [12, 13]:

$$\begin{pmatrix} K_{\mathbf{x}_k} & B^T \\ B & 0 \end{pmatrix}\begin{pmatrix} \delta\mathbf{x}_k \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \delta L_{\mathbf{x}_k} \\ 0 \end{pmatrix} \tag{22}$$

Since the quadratic expansion of the total potential energy $\hat{E}_{k+1}$ in general ignores higher order terms in the local region of $\mathbf{x}_k$, several Newton iterations are needed to converge to the true equilibrium conformation $\mathbf{x}_{k+1}$.

Modeling experiments

Numerical analysis

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

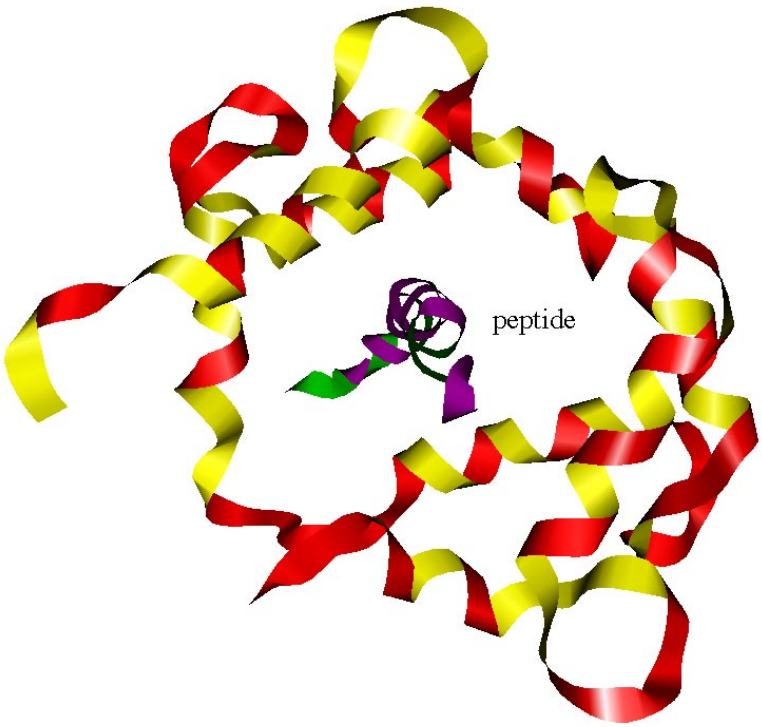
First, it decouples the movements of the atoms (in terms of the degrees of freedom) along the chain of the protein backbone. To characterise the movement of any atom along the polypeptide in terms of torsional angles requires all the torsional angles prior to that atom. In our parametrisation, however, the independent degrees of freedom are those of the nodal backbone atoms. A pair of successive nodal atoms along the protein backbone (separated by nine backbone bonds) define a (directed) *computational element*. 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, as shown in Figure 3, is banded with column block width 12.

Specifically, what this means is that we have reduced the degrees of freedom (d.o.f.) of the entire protein to that of the nodal backbone atoms *only*. Decoupling means that every atom is 'tied to' (*i.e.*, its motion is completely determined by) exactly one pair of nodal atoms which define a computational element, and is independent of the motion of any other nodal atoms.

Second, this decoupling has the important consequence that the resulting stiffness matrix of the d.o.f. of the set of nodal atoms is relatively sparse, since non-zero entries arise from the nodal d.o.f. of pairwise interacting computational elements only: each atom motion can be easily pulled back to that of the nodal atoms of its computational element 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 (see Figure 4). The torsional stiffness matrix 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

Figure 6 *Calmodulin-peptide complex (from PDB)*



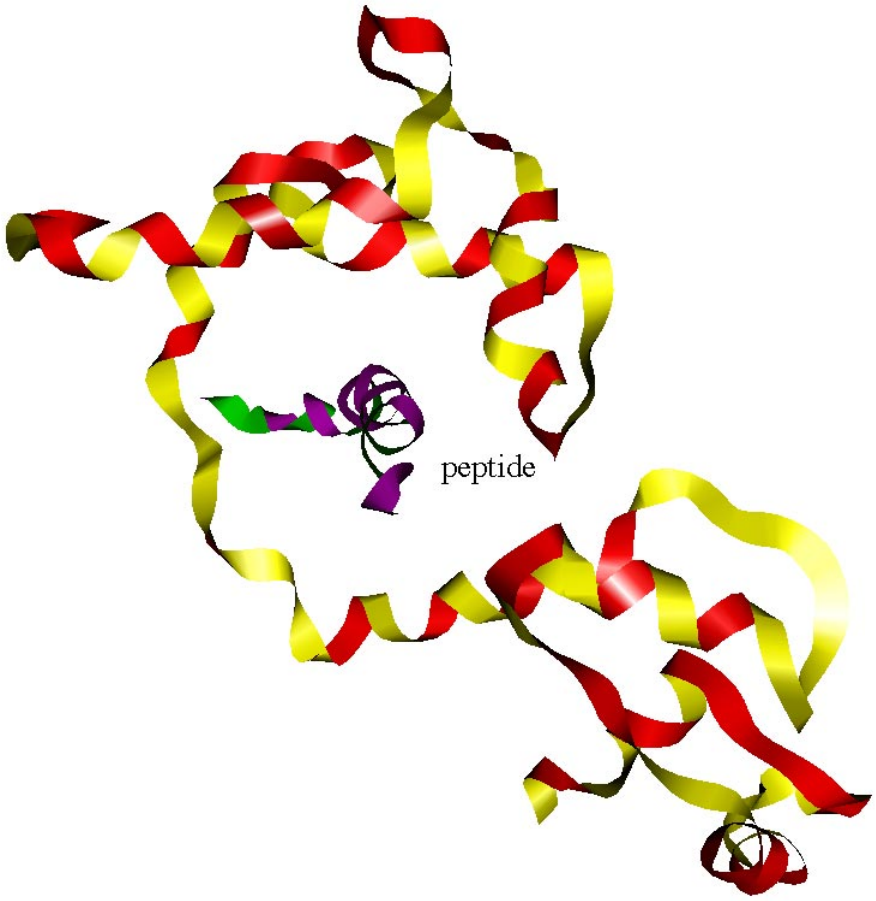


Figure 7 *Calmodulin-peptide complex with one distance constraint (simulation result)*

with some off-diagonal blocks and it is thus quite sparse. The sparsity leads to fast computation in solving the linearly constrained quadratic minimisation problem and results in a major improvement in computational efficiency. Although there have been many papers which aim to reduce the set of d.o.f. (see [14], for example), we have not seen elsewhere the decoupling technique we have described here, and we believe that this is an important contribution.

There are three major time consuming parts in the computation: collision detection, formation of the transformation matrix and the total stiffness matrix, and the solving of the linear system for the linear constraint quadratic minimisation. Among these three, the latter two parts are relatively computationally more expensive. We use a fast collision detection algorithm with linear complexity O(n) where n is the number of atoms. The collision detection is only performed

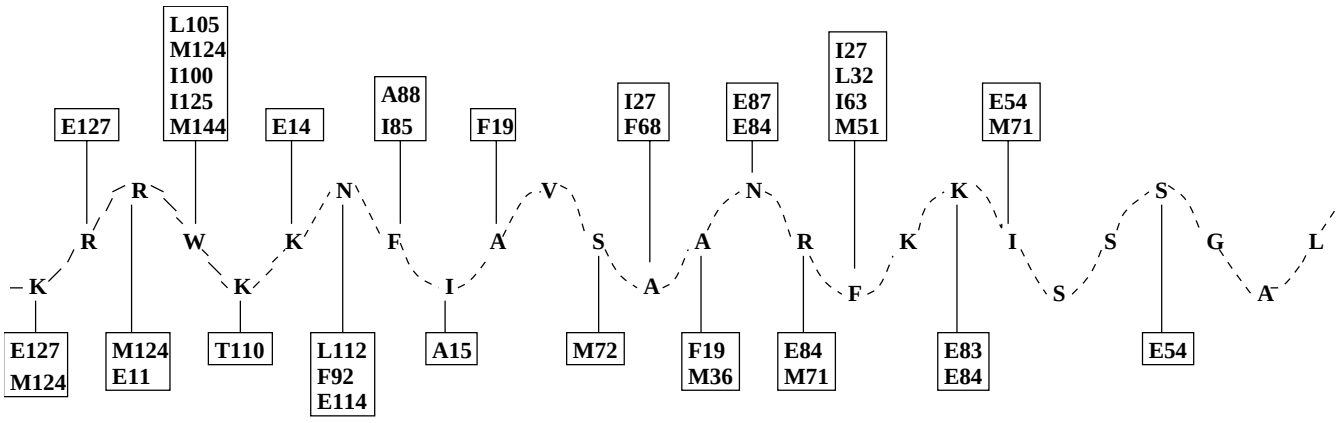


Figure 8 *Contacts between calmoduin and the peptide*

in the first iteration of every distance constraint increment. It further reduces the overall computation. The complexity of formation of the transformation matrix and the total stiffness matrix is also linear in the number of atoms. As they are fully decoupled with the global parametrisation, however, parallel and concurrent processing techniques can be used. To solve the linear system, we use least squares solvers from standard numerical libraries. Taking advantage of the nature of incremental energy minimisation and the specific sparse pattern of the stiffness matrix, optimal iterative solvers have been selected.

Calmodulin peptide binding

Calmodulin is a monomeric signal transduction protein consisting of a chain of 148 amino acids that is capable of binding to up to 4 Ca^{2+} ions. The molecule is dumbbell-shaped [15] with an eight-turn solvent-exposed central alpha helix connecting the two pairs of EF-hand domains (see Figure 5). The letters E and F denote alpha helices and each EF hand consists of an alpha helix of about ten residues, a ten-amino-acid loop, and a second alpha helix. While the presence of such a long central helix in solution has been questioned, several lines of evidence suggest that some amino acids within this region, as well as in both hydrophobic pockets formed within the EF-hand domains, are critical for target interaction.

The calmodulin-binding regions have been studied in several cases and consist of approximately 20 amino acids. Several examples of peptides that bind quite tightly to calmodulin have been revealed. Studies using fluorescence quenching and photoaffinity labeling indicate that the hydrophobic portions of the peptide are partially buried upon interaction with calmodulin [16-18]. Experiments with doubly labelled peptides, furthermore, show that both ends of such peptides are capable of interacting with the two EF-hand domains of calmodulin simultaneously. This means that the central alpha helix has to be structurally flexible for the two lobes of calmodulin to bind to the target amphipathic peptide.

In our simulation experiment, we study the conformational change of calmodulin when it binds to a synthesised 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. The unbound calmodulin is from *drosophila melanogaster* expressed in *E coli*. The calmodulin in the bound calmodulin peptide complex is also from *drosophila* and the peptide is synthesised from rabbit skeletal myosin light-chain kinase. The 3-dimensional structure (2bbm) determined by multidimensional NMR is shown in Figure 6 [19].

The 3-dimensional structure of the calmodulin peptide complex is roughly a compact ellipsoid. All eight helices of calmodulin (A to D in the N-lobe and E to H in the C-lobe) are in close proximity to the peptide and wrap around it.

To begin, we first apply just one distance constraint between a pair of atoms (near the centres) in the N-lobe and in the C-lobe respectively. In other words, we try to pull the two lobes of the calmodulin closer to each other. The simulation

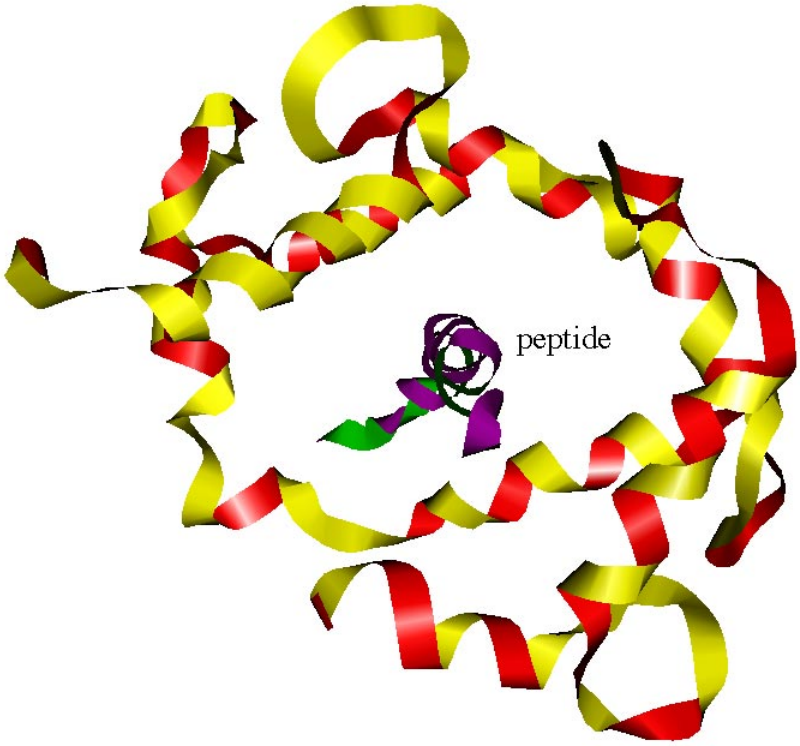


Figure 9 *Calmodulin-peptide complex (simulation result)*

result is shown in Figure 7. It has features roughly similar to the calmodulin-peptide complex (Figure 6) and shows how the long central helix serves as an *expansion joint*. It is also observed that the two lobes are not directly approaching each other. There are significant relative rotations (around 90 degrees) between the two lobes. This is exactly a major feature in the three dimensional structure of the bound calmodulin-peptide from the Protein Data Bank.

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. Figure 8 shows the close contacts between calmodulin and the peptide. We have chosen 24 such distance constraints based on the contacts. This small set of 24 distance constraints (compared to the approximately 1,000,000 possible distances or atom pairs that can be chosen between calmodulin and the peptide from the PDB data for the calmodulin-peptide complex) is applied incrementally as discussed earlier to 'drive' the binding process, so that the corresponding distances eventually reach their final values. The values for these 24 distances have been taken from the bound state conformation in the PDB. The final configuration (of the entire protein) reached by the simulation is determined just by these distance constraints and the interactions we have specified. The corresponding simulation result is shown in Figure 9. Figure 10 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 Protein Data Bank. There are only minor differences in the lobes, and the large conformational change in calmodulin upon peptide binding is manifested almost completely in the 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 Å) between the peptide and calmodulin. These qualitative observations are the same as what we get from the analysis of the three dimensional structure of the bound calmodulin-peptide from Protein Data Bank. We have also made a quantitative comparsion between the simulation result and the known data. We compute the differences of the distances between every corresponding pair of

Figure 10 *The motion-blur picture of calmodulin peptide-binding (simulation result)*

alpha carbon atoms in our simulation result **x** and in the experimentally determined configuration in the Protein Data Bank, $\hat{\mathbf{x}}$:

$$e_{ij} = r_{ij} - \hat{r}_{ij} = \left|\mathbf{x}\left(C_\alpha^i\right) - \mathbf{x}\left(C_\alpha^j\right)\right| - \left|\hat{\mathbf{x}}\left(C_\alpha^i\right) - \hat{\mathbf{x}}\left(C_\alpha^j\right)\right| \quad (23)$$

The maximum e_{ij} is 1.6Å which is less than the resolution of the structure obained from NMR. The RMS deviation

$$R_e = \sqrt{\frac{\sum_{i,j} e_{ij}^2}{n_{C_\alpha}\left(n_{C_\alpha} - 1\right)/2}} \quad (24)$$

is 0.155Å. Here $n_{C\alpha}$ is the number of alpha carbon atoms.

Cooperative hemoglobin oxygen binding

Hemoglobin, the oxygen-carrying pigment in red blood cells, has a molecular weight of 64,458 [20, 21]. It has a protein part, the globin, and four nonprotein heme groups. 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 a five-coordinated iron. The sixth coordinated position of iron can be occupied by a molecule of oxygen, O_2.

Both heme and globin are essential for oxygen transport, which depends on partial pressure of O_2 in the blood. Where O_2 partial pressure is high, as in lungs, hemoglobin binds oxygen and becomes bright red (oxyhemoglobin — the three-dimensional structure is shown in Figure 12). Then when oxygenated hemoglobin reaches tissues where O_2 partial pres-

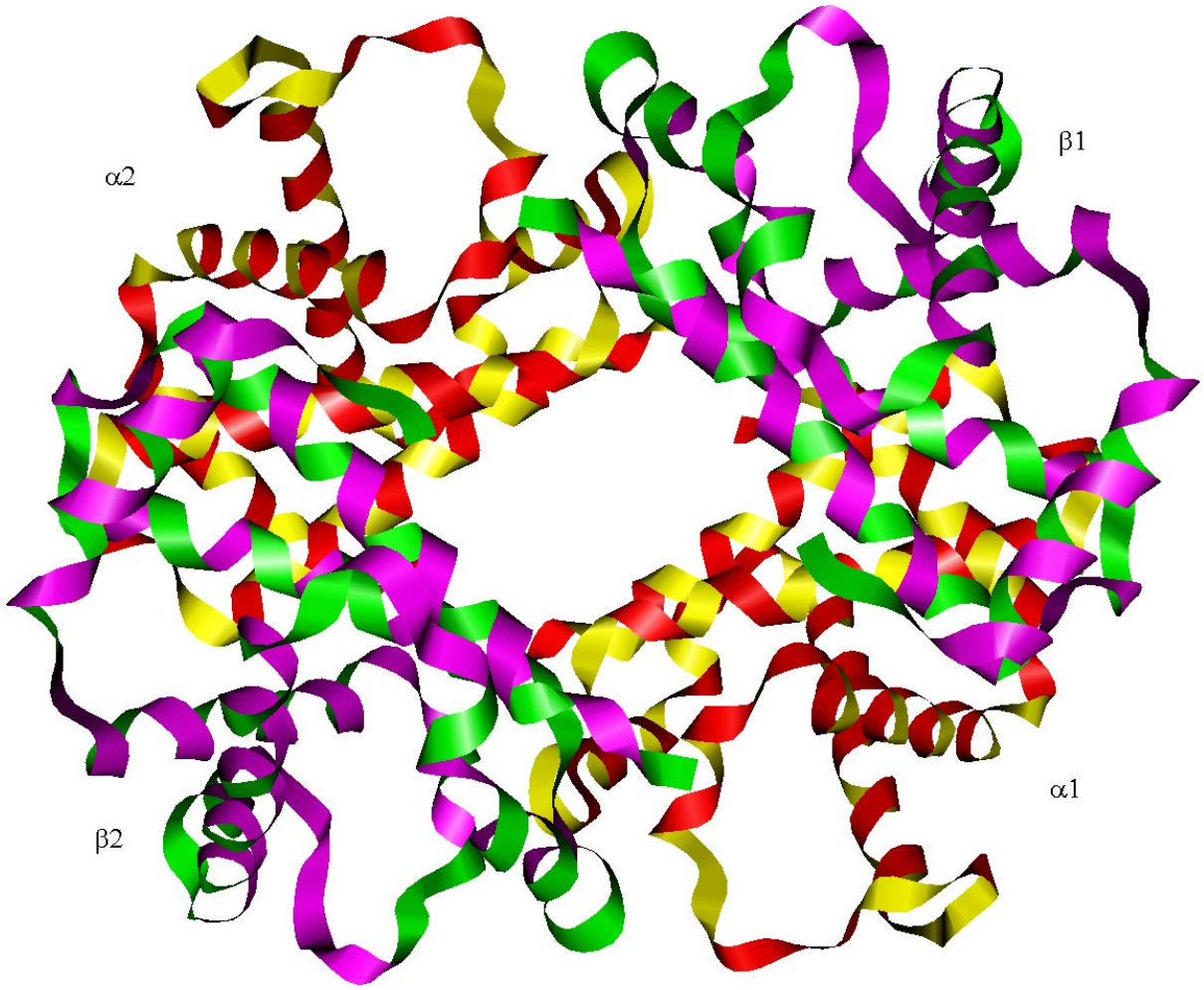


Figure 11 *Deoxygenated hemoglobin (from PDB)*

sure is low, it releases O_2 and turns purple (deoxyhemo-globin, the three-dimensional structure is shown in Figure 11).

The hemoglobin tetramer is formed with two dimers of α_1–β_1 and α_2–β_2. The structure of each subunit is similar and is also close to the structure of myoglobin. The secondary structures in hemoglobin are only helices (as shown in Figure 13) and the β subunit has one more helix (in standard notation, the helices are labelled A through H) than the α subunit.

In Perutz's original work [22], it was noticed that hemoglobin changes its quaternary structure upon oxygen binding. The α_1–β_1 and α_2–β_2 dimers are symmetrically rearranged and one approximately rotates about 12 to 15 degrees from the other.

This is an allosteric transition between oxygenated hemoglobin and deoxygenated hemoglobin and these two states are both equilibrium states. The T and R structures (deoxy and oxyhemoglobin respectively) differ in the arrangement of the four subunits (quaternary structure), and in the conformation of the subunits (tertiary structure).

Based on allostery theory [23], the oxygen binding process is cooperative. When oxygen binds to one subunit of a deoxygenated hemoglobin and transits from T state to R state, the oxygen binding at other subunits becomes easier. When one subunit of an oxygenated hemoglobin releases oxygen and transits from the R state to the T state, the oxygen release at other subunits also becomes easier. Such cooperative oxygen binding is vital for the function of hemoglobin as an oxygen carrier.

The three-dimensional structures of human deoxyhemoglobin (1hga) and oxyhemoglobin (1hho) that we use are both from Protein Data Bank using X-ray crystallography at a resolution of 2.4 Å. Before the discussion of the simulation steps, we first look at some important differences in structure between deoxyhemoglobin and oxyhemoglobin [24].

• Each α_1–β_1 and α_2–β_2 half of the molecule moves roughly as a rigid body, and the two halves slide over one another.

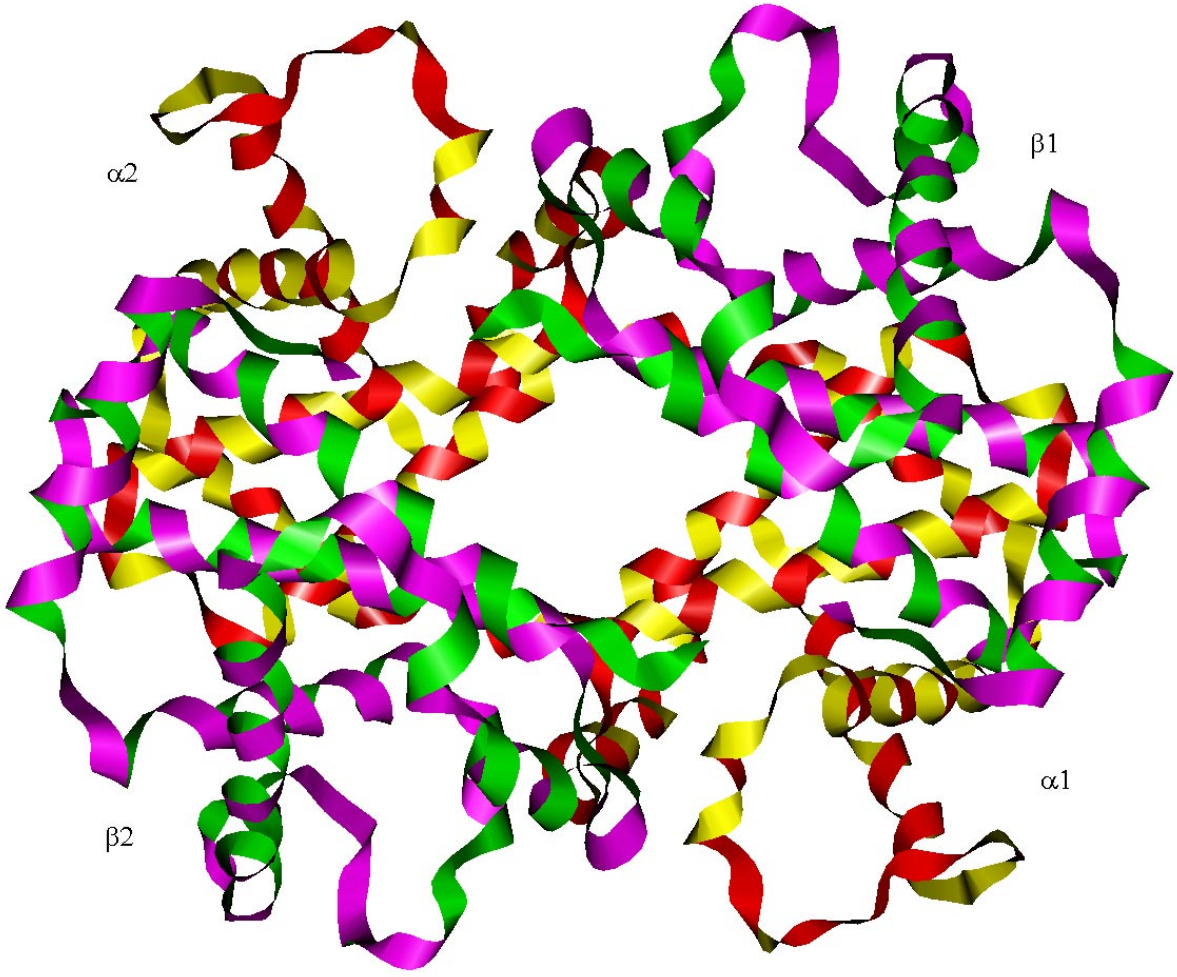


Figure 12 *Oxygenated hemoglobin (from PDB)*

• The structure of the interface between α_1 and β_1 remains the same. The α_1–β_1 interface therefore gives a frame of reference for the description of tertiary structure changes.

• The translation of the hemes and rotation of the β heme (relative to the E helix and the α_1–β_1 interface) removes the binding site from the vicinity of residue Val E11.

• There is a small movement at the contact between α_1 FG corner and β_2 C helix (flexible joint), while there is a large movement of about 6 Å at the contact between the α_1 C helix and the β_2 FG corner (switch region). Each position of the switch is stabilised by a different set of hydrogen bonds. Other hydrogen bonds in the two contacts of deoxyhemo-globin are broken in oxyhemoglobin. The contact region of deoxyhemoglobin is shown in Figure 14.

• 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 F helices and FG corners relative to α_1–β_1 interface. The enviroment around that area is shown in Figure 15.

The objective of our simulation is not only to reproduce the structure of oxyhemoglobin starting 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 for cooperative oxygen binding.

One possible hypothesis is that the movement of the iron atom itself acts as trigger [22]. 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 fail to observe the allosteric transition with some other hypotheses [20] for the trigger mechanism.

Perutz has emphasised the importance of the hydrogen bonds in the contacts between different dimers [22]. 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 introduced a set of distance constraints to break these hydrogen bonds. The simulation result reproduces the cooperative allosteric transition and is very close to (in the sense of pairwise distances) the configuration from the Protein Data Bank. It is shown in the upper-right of Figure 16 (upper-left of Figure 16 is the data from the Protein Data Bank).

To find the real trigger, we noticed that in the absence of oxygen, the four hemes are in a relaxed, essentially strain-

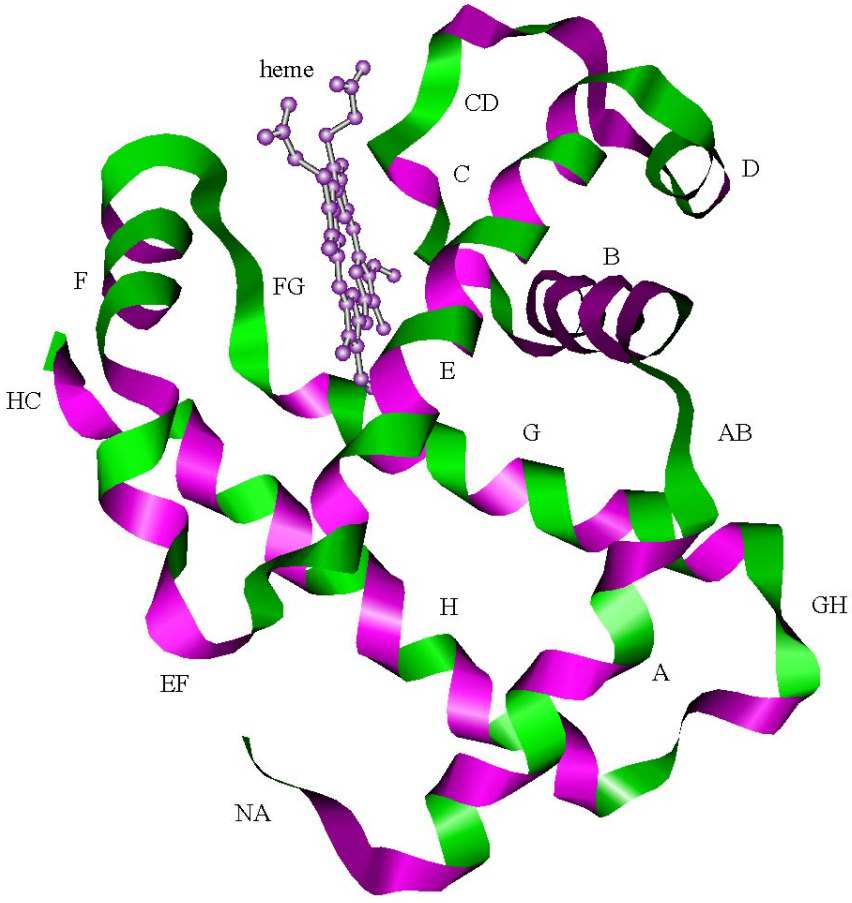


Figure 13 *Structure of hemoglobin β subunit*

free enviroment with F8 His tilted off-axis, the heme plane domed toward the F8 His, and an out-of-plane iron atom. Strain is produced when an oxygen binds on the other side of the heme, and this strain is transmitted through side chains and the F helix to the FG corners and to terminal salt bridges and hydrogen bonds. When the strain becomes great enough, the molecule shifts from T to R state. The shift relieves the strain on the already oxygenated heme and also allows unoxygenated hemes a potential freedom of motion that enhances their affinity for oxygen.

Therefore, it is possible that the straightening up of the tilted F8 helix and the corresponding movement of the F helix might be the trigger. Indeed, the movement of the F-helix and the FG-corner was found to be produced, in careful molecular mechanics computations of the heme environment [25], as a result of the strains arising in the allosteric core due to change in heme conformation upon ligand binding. Further, the change in the allosteric core conformation was found in [25] to produce a tertiary structural change such that tertiary-tertiary contact regions underwent alterations in structure in overall accord with experimental observations. The thermodynamic implications of tertiary structural change for hemoglobin cooperativity were discussed in [26] and found to be significant. Thus, tertiary structural change may be taken to be a trigger for quaternary structural change.

Accordingly, we introduce a set of distance constraints to simulate such a trigger. There are 16 of them between F-helix and α_1–β_1 reference interface in each subunit. These constraints, which constitute only a very small subset of the possible atom pairs one can choose from the atoms of the molecule, are applied incrementally exactly as described earlier for calmodulin. The result is shown in the lower-left of Figure 16.

The result is remarkably close to the oxygenated hemoglobin conformation from the Protein Data Bank: defining a pairwise (for every pair of alpha-Carbons) distance difference between the simulated and experimentally determined conformations as we did earlier in [23] gives an RMS distance difference of 0.215 Å.

It is worth emphasising that our results are robust in the sense that choosing different small sets of distance constraints between the relevant regions still yield results with small RMS distance differences from the PDB R-structure conformation. The important thing seems to be the choice of approximate regions of interactions in the molecule. Moreover, the simulation reproduces the *consequence* (in contrast to the previous simulation experiment, where this was used as an input)

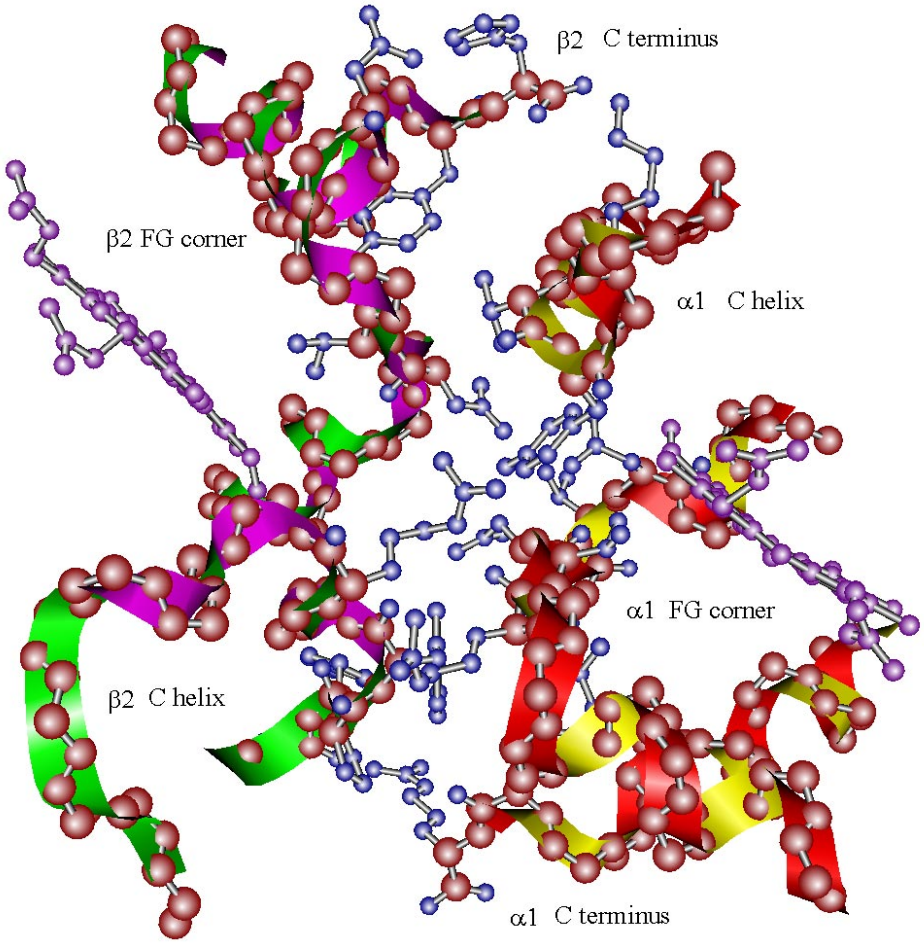


Figure 14 *Subunit contact at α_1–β_2 interface*

that the hydrogen bond network in deoxyhemo-globin is broken during the allosteric transition.

To further verify the correctness of our model, we have done another computational experiment. With the same helix-movement trigger, oxygen binding without the hydrogen bond network was simulated. This corresponds to a real experiment where the hydrogen bond network is dissolved using chemical methods [27]. Our result is the same as the real experiment result: there is no longer any cooperativity for the binding. As shown in the lower-right of Figure 16, the four subunits independently expand themselves.

Conclusions

The civil engineering model we have developed does not solve the protein folding problem. *Proteinmorphosis* starts with the given three-dimensional structure of the protein and studies its conformational changes. It is also important to emphasise that it is a fully mechanics-based approach to computing conformational changes, as opposed to computer animation approaches [28] that use interpolation between observed initial and final structures. As shown in this paper, the conformational changes from our simulation results are realistic and our approach has a variety of applications in drug design.

There are also many differences between our civil engineering approach and other physically-based molecular modelling systems such as SCULPT [29, 30]. First, the finite element parametrisation employed in *Proteinmorphosis* 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 minimisation 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

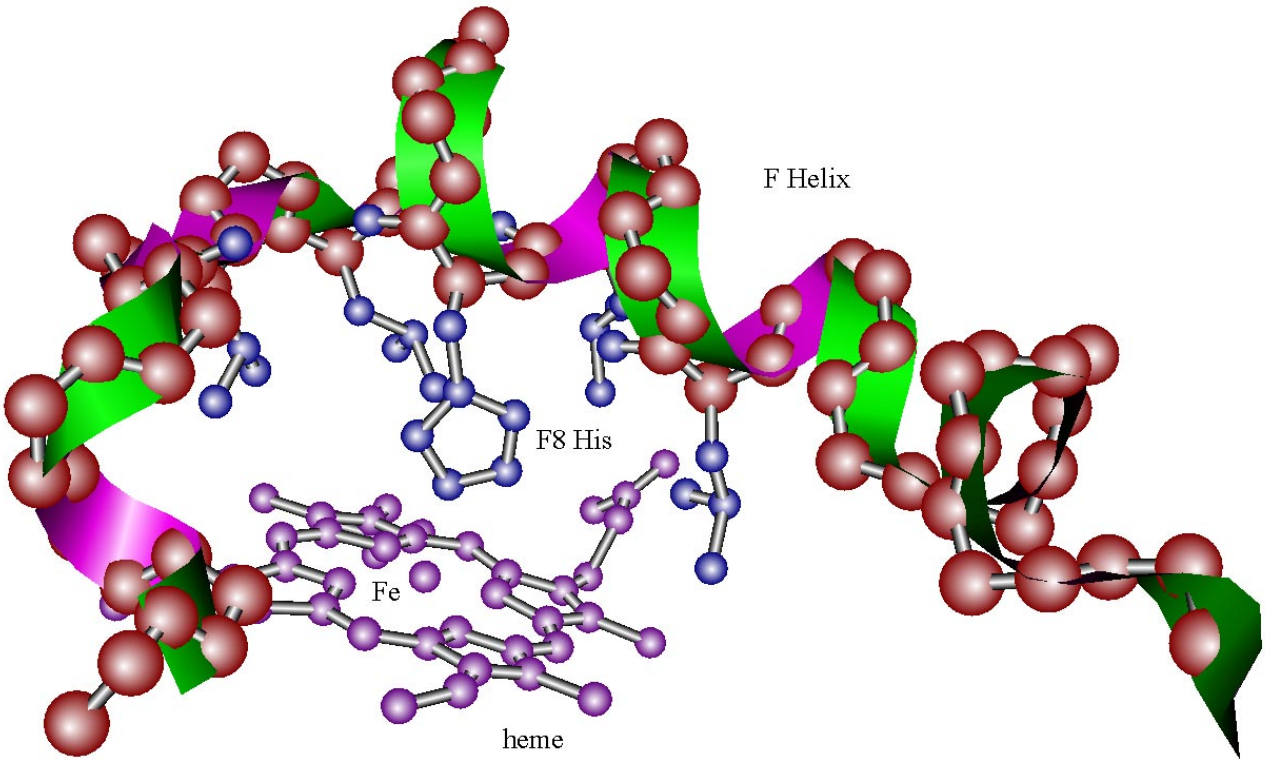


Figure 15 *Possible trigger for the hemoglobin T-R transition*

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.

It is worth contrasting our model with molecular dynamics methods commonly applied to study protein dynamics. An all-atom approach, integrated over small time steps (see, for example, [31]), provides accurate results, albeit at the cost of enormous computation time for large molecules involving major conformational changes. In contrast, the less

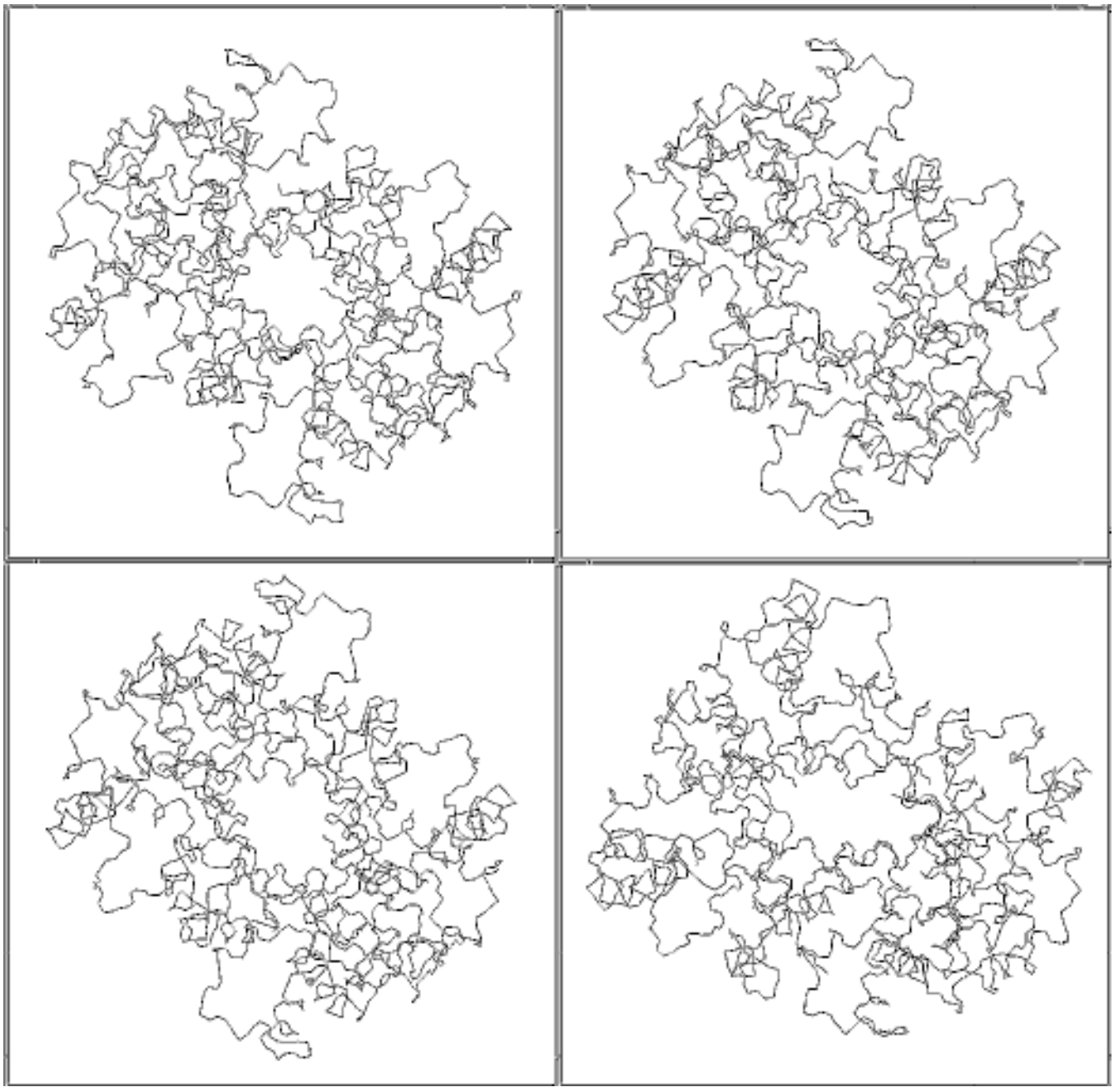


Figure 16 *Simulation result of hemoglobin oxygen binding*

detailed, 'macroscopic' model we use with its reduced set of degrees of freedom may seem crude, but has the advantage of significantly faster computation. Perhaps somewhat surprisingly, it does however produce 'correct' conformations, at least for the examples we have discussed. The key input for this seems to be the choice of 'reasonable' sets of interacting regions. In the examples we studied as test cases, the final structures (experimentally determined) were of course available to us for picking regions between which to apply distance constraints. We believe that our model, especially where cooperativity is involved, is a useful complement to more detailed and accurate modelling methods in the following sense.

First, where a large change in conformation is involved, local ligand binding effects may be computed effectively with molecular dynamics. These structural changes produce local distortions that may be used as an input to our civil engineering model in order to predict gross conformational changes, which would be difficult with more detailed models. Second, one's knowledge of chemistry may be invoked, perhaps together with a detailed molecular mechanics model, to define regions of likely interaction between proteins and other protein or drug molecules, especially where the presence of flexible regions can result in significant changes in conformation. Again, this region identification serves as an input to our model which may be usefully employed to predict final structure. Third, one can easily include dynamics in our macroscopic model in order to make an initial study of protein dynamics and to gain an understanding of low energy modes.

The main contributions of our mechanical model for protein conformational changes are its conceptual ideas using a macroscopic civil engineering approach and the efficient global parametrisation that it employs. The two examples of calmodulin and hemoglobin serve to validate our model. Potential biomedical applications of *Proteinmorphosis* in the near future include problems such as rational drug design.

Appendix

The algorithm for collision detection, referred to in Section 2, is developed based on three facts. First, a cube is a good approximation to a sphere. Checking the collision between two spherical atoms can be simplified by checking the collision between two cubes (voxels) whose widths equal the diameters of the associated atoms. This approach will not miss any collisions. Second, for each atom, only a small number of atoms fit within its cutoff distance because each atom's electron shell occupies a nonzero volume. A very conservative bound on the number of collisions to the atom is the division between $8R_{cut}^3$, the cutoff volume, and the volume of the smallest atom. The limited number of collisions per atom is also a consequence of the 1-dimensional nature of the protein polypeptide chain. Third, the sizes of the atoms do not differ by very much. In protein molecule, there are only five types of atoms, i,e, C(carbon), S(sulphur), H(hydrogen), N(nitrogen) and O(oxygen). Hydrogen has the smallest van der Waals radius of around 1 Å. The sulphur atom is somewhat larger but it does not occur very often and the rest have almost the same van der Waals radius.

The collision detection starts with dividing the space containing the whole protein molecule into voxels by imposing a uniform 3-dimensional grid on it. If an atom occupies part of a voxel, that voxel is marked as being occupied either partially or wholly by the atom. If the voxel has already been marked by another atom, an exact comparison is made between the two atoms. A voxel may be marked by more than one atom. An atom may mark more than one voxel.

The most critical issue here is to choose the proper size of grid. On one hand, the grid should be as small as possible so that the likelihood of a real collision is high when two atoms mark the same voxel. On the other hand, we want the grid to be as large as possible so that every atom has to mark very few voxels. In our case, we make a trade-off to choose the van der Waals radius of the oxygen atom as the grid size. It is easy to verify that the algorithm takes O(n) steps.

Acknowledgment We are grateful to Ramesh Subramonian for the collision detection algorithm described in the appendix. We would like to thank Professor James Anderson of the Department of Radiology of the Johns Hopkins University for providing computational resources. Most of our program has been run on the 6-processor SGI Infinite Reality in his lab.

Reference

1. Lawton, W.; Poston, T.; Raghavan, R.; Ranjan, S. R.; Viswanathan, R.; Wang, Y. P.; Yu, Y. Variational Methods in Biomedical Computing *the Proceedings, Conference on Computational Science for the 21st Century, Tours, France* John Wiley: Sussex, England, 1997.
2. Lawton, W.; Raghavan, R.; Viswanathan, R. *CIeMed report*, 1997.
3. Meiyappan, S.; Raphel, J.; Viswanathan, R.; Raghavan, R.; Yu, Y. *CIeMed report*, 1997.
4. Monod, J.; Wyman, J.; Changeux, J.-P. *J. Mol. Biol.* **1965**, *12*, 88.
5. Poston, T.; Stewart, I. *Catastrophe theory and its applications*, Dover Publishers: London, 1997.
6. Cornell, W. D.; Cieplak, P.; Bayly, C. I. *et. al., Journal of the American Chemical Society* **1995**, *117*, 5179.
7. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.
8. Berg, J., personal communication, 1996.
9. Lifson, S. In *NATO Advanced Study Institute/FEBS Advanced Course No. 78. Current Methods in Structural Molecular Biology*; 1981; p 1.
10. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta Jr., S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
11. Arrow, K. J.; Hurwicz, L.; Uzawa, H. *Studies in Linear and Nonlinear Programming*; Stanford University Press: Stanford, California, 1958.

12. Duff, I. S. "The solution of augmented systems", *Numerical Analysis 1993*, 1994, 40.
13. Fortin, M.; Glowinski, R. *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary Value Problems*; North Holland NY, 1983.
14. Turner, J.; Weiner, P.; Robson, B.; Venugopal, R.; Schubele III, H.; Singh, R. *J. Comput. Chem.* **1995**, *16*, 1271.
15. Lippard, S.; Berg, J. *Principles of Bioinorganic Chemistry*; University Science Books: Mill Valley, California, 1994.
16. Ikura, M.; Clore, G. M.; Gronenborn, A. M.; Zhu, G.; Klee, C. B.; Bax, A. *Science* **1992**, *256*, 632.
17. Meador, W. E.; Means, A. R.; Quiocho, F. A. *Science* **1993**, *262*, 1718.
18. Meador, W. E.; Means, A. R.; Quiocho, F. A. *Science* **1992**, *257*, 1251.
19. Taylor, D. A.; Sack, J. S.; Maune, J. F.; Beckingham, K.; Quiocho, F. A. *J. Biol. Chem.* **1991**, *266*, 21375.
20. Dickerson, R.E.; Geis, I. *Hemoglobin: Structure, Function, Evolution, and Pathology,*; The Benjamin/Cummings Publishing Company, Inc.: Menlo Park, California, 1983.
21. Fermi, G.; Perutz, M. *Atlas of Molecular Structures in Biology: 2. Haemoglobin and Myoglobin*; Clarendon Press: Oxford, 1981.
22. Perutz, M. *Nature* **1970**, *228*, 726.
23. Perutz, M. *Mechanisms of Cooperativity and Allosteric Regulation in Proteins*; Cambridge University Press: Cambridge, 1990.
24. Baldwin, J.; Chothia, C. *J. Mol. Biol.* **1979**, *129*, 175.
25. Gelin, B. R.; Lee, A. W.; Karplus, M. *J. Mol. Biol.* **1983**, *171*, 489.
26. Lee, A. W.; Karplus, M. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 7055.
27. Barrick, D.; Ho, N.; Simplaceanu, V.; Dahlquist, F. W.; Ho, C. *Molecular signaling in hemoglobin: a test of the Perutz model for cooperativity*; University of Oregon: preprint, 1996.
28. http://bioinfo.mbb.yale.edu/MolMovDB/
29. Surles, M. C. *Computer Graphics* **1992**, *26*, 221.
30. Surles, M. C.; Richardson, J. S.; Richardson, D. C.; Brooks Jr., F. P. *Protein Science* **1994**, *3*, 198.
31. McCammon, J. A.; Harvey, S. C. *Dynamics of proteins and nucleic acids*; Cambridge University Press: Cambridge, 1991.