doi:10.1016/j.jmb.2003.08.027 J. Mol. Biol. (2003) 333, 153–168

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Allosteric Changes in Protein Structure Computed by a Simple Mechanical Model: Hemoglobin T ↔ R2 Transition

Chunyan Xu, Dror Tobi and I. Bahar*

Center for Computational Biology and Bioinformatics and Department of Molecular Genetics and Biochemistry School of Medicine, University of Pittsburgh, Pittsburgh, PA 15213, USA

*Corresponding author

Information on protein dynamics has been usually inferred from spectroscopic studies of parts of the proteins, or indirectly from the comparison of the conformations assumed in the presence of different substrates or ligands. While molecular simulations also provide information on protein dynamics, they usually suffer from incomplete sampling of conformational space, and become prohibitively expensive when exploring the collective dynamics of large macromolecular structures. Here, we explore the dynamics of a well-studied allosteric protein, hemoglobin (Hb), to show that a simple mechanical model based on Gaussian fluctuations of residues can efficiently predict the transition between the tense (T, unliganded) and relaxed (R or R2, O_2 or CO-bound) forms of Hb. The passage from T into R2 is shown to be favored by the global mode of motion, which, in turn is driven by entropic effects. The major difference between the dynamics of the T and R2 forms is the loss of the hinge-bending role of $\alpha_1-\beta_2$ (or $\alpha_2-\beta_1$) interfacial residues at αPhe36-His45 and βThr87-Asn102 in the R2 form, which implies a decreased cooperativity in the higher affinity (R2) form of Hb, consistent with many experimental studies. The involvement of the proximal histidine βHis92 in this hinge region suggests that the allosteric propagation of the local structural changes (induced upon O_2 binding) into global ones occur *via* hinge regions. This is the first demonstration that there is an intrinsic tendency of Hb to undergo T → R2 transition, induced by purely elastic forces of entropic origin that are uniquely defined for the particular contact topology of the T form.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: hemoglobin; Gaussian network model; anisotropic network model; allostery

Introduction

Studies on hemoglobin (Hb), the oxygen carrier protein in red blood cells, have contributed significantly to our understanding of the molecular mechanism of allostery in general,[1] and Hb continues to serve as a model for understanding the role of intramolecular interactions in functional cooperativity. Human normal adult hemoglobin (HbA) is composed of four subunits having similar three-dimensional (3D) structures, two identical α-chains of 141 amino acid residues each referred to as α_1 and α_2 subunits, and two identical β-chains, β_1 and β_2, of 146 amino acid residues each. Each subunit has a heme-binding pocket that binds oxygen (Figure 1(a)). The oxygenation process of Hb is cooperative, i.e. binding of a first O_2 molecule enhances O_2 binding affinity, as evidenced by the decrease in free energy penalty observed with successive O_2 bound. This increased affinity underlies the uptake/release of O_2 over a relatively narrow range of pressure in the lungs/tissues.

There are two classical models proposed for describing the allosteric mechanism of Hb: the Monod, Wyman and Changeux (MWC) concerted

Abbreviations used: RMSD, root-mean-square deviation; GNM, Gaussian network model; ANM, anisotropic network model; NMA, normal mode analysis; EN, elastic network; HbA, adult hemoglobin; MWC, Monod, Wyman and Changeux; KNF, Koshland, Nemethy and Filmer; MD, molecular dynamics.

E-mail address of the corresponding author: bahar@pitt.edu

0022-2836/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.

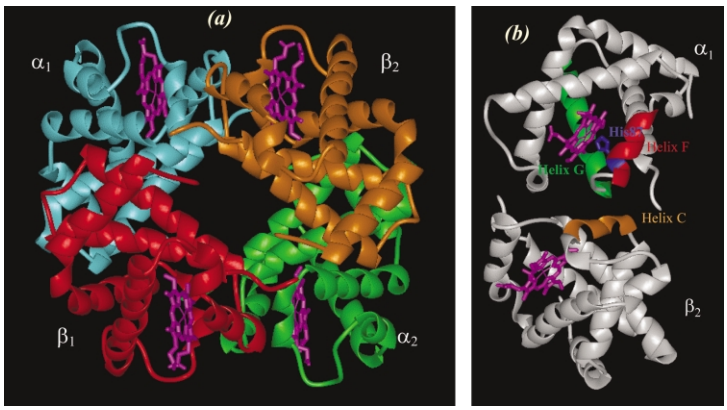

Figure 1. (a) Ribbon diagram of the Hb tetramer. The four monomers α_1, α_2, β_1 and β_2 are shown in cyan, green, red and orange, respectively, and the four heme groups are explicitly displayed in magenta. The RCSB PDB entry 1a3n (unliganded Hb) is used in this Figure. (b) The α_1 and β_2 subunits, and the secondary structural elements that will be shown to be reconfigured during the T–R2 transition are displayed. The proximal αHis87, helix F (81–89) and helix G (95–112) are shown in blue, red and green, respectively, for α_1. The helix C (36–41) is shown in orange for the β_2 subunit.

model,[2] and the Koshland, Nemethy and Filmer (KNF)[3] sequential model. The MWC model assumes that the low-affinity (tense, T) and high-affinity (relaxed, R) forms of the tetramers co-exist and can both bind O_2. In the absence of O_2 the T form dominates the ensemble of conformers. Upon O_2 binding, on the other hand, the R form is stabilized, and a shift in the dominant population from T to R is induced. All four molecules in a given tetramer have the same conformation at any time, i.e. there is a concerted (all-or-none) change between the T and R forms in all four subunits. This model was later supported by the oxy and deoxy-structures of Hb determined by Perutz.[4] The concerted mechanism of transition was the rotation of the $\alpha_2\beta_2$ dimer (subunits colored green and orange in Figure 1(a)) by a torsional angle of 15° relative to the $\alpha_1\beta_1$ dimer about an axis that crosses the α_1 and α_2 (cyan and green) subunits. The KNF model, on the other hand, allows for the co-existence of two different forms, R and T, in a given tetramer. Accordingly, each subunit can individually undergo a change from T to R form upon ligand binding, and the affinity of the neighboring subunits is changed through interactions at the subunit interfaces. Both the MWC and KNF models can fit the sigmoidal O_2 binding curves by suitable choice of parameters, and in particular the MWC model seems to be supported by spectroscopic, thermodynamic and kinetics studies.[5]

Yet, on a broader perspective, these classical models can be viewed as two extreme cases of a more complex scheme that involves a combination of these simplified schemes.[6,7] In particular, an intradimer (tertiary) cooperativity (within $\alpha_1\beta_1$ or $\alpha_2\beta_2$) (symmetry rule model) has been inferred, in addition to the interdimer (quaternary) cooperativity, from the free energy database compiled for Hb tetramers with hemesite analogs,[8] supported by recent experiments.[9] The role of the interactions at the $\alpha_1-\beta_1$ (or $\alpha_2-\beta_2$) subunit interface(s) and their coupling to those at the $\alpha_1-\beta_2$ (or $\alpha_2-\beta_1$) interface(s) are still being explored (see e.g. Chang *et al.*[10])

A third Hb conformation, designated as R2, has been observed in the carbon monoxy (CO)-bound form at pH 5.8.[11] This conformation is closer to the oxygenated state of Hb because the original R structure is formed under a high salt condition. R shows intermediate characteristics between T and R2.[12] Understanding the molecular mechanisms or intersubunit interactions underlying the transition between the two most distinctive forms, T and R2, is the task presently undertaken.

Our knowledge of structure–function relations of Hb is by and large provided by structural and spectroscopic studies. Efforts for understanding the mechanism of $T \leftrightarrow R$ transition have usually focused on the comparison of the available polymorphic structures of the multimer. While these studies help in visualizing the types of domain movements, they have limitations: (i) they require two or more structures for comparison; (ii) they are not predictive but simply compare already known structures; and (iii) the inferred mechanisms depend on the observed structures, which do not necessarily represent a complete or unbiased set, except for a few thoroughly studied cases (e.g. the study of HIV-1 protease flexibility by Karplus and co-workers.[13])

Molecular dynamics (MD) simulations can usually provide a more complete description of conformational motions. Yet, the MD simulations of Hb have generally been limited to its monomers or substructures due to the large size (574 residues or ~64.5 kDa) of the tetramer, except for a few recent studies.[14] Apart from the large size of the protein, it is also a challenge to sample the transition state in traditional MD simulations.[15] A remarkable study in this respect is the recent MD simulations of $T \rightarrow R$ transition by Mouawad *et al.*,[14] which explores the transition by the "Path Exploration with Distance Constraints" technique introduced by Guilbert and co-workers.[16] In this technique the MD trajectory is not free to explore the most probable pathways, but instead it explores a subset of pathways that fulfill the constraints. The initial and final conformations are

predefined and a potential that forces the molecule to approach the known final conformation is imposed. Similarly Zaloj & Elber[17] used the stochastic path approach for calculating trajectories between the T and the R states of Hb. As such, these methods do not predict the final conformation to be reached, but instead allows for the examination of the most probable transition pathways, or conformational kinetics, between two conformers. This is a useful approach in that it gives a detailed description of atomic interactions (hydrogen bonds, salt bridges, interatomic distances) that are affected by the transition. Our approach is different in that it predicts the most probable direction of conformational change given only the initial state, but at the cost of a loss in resolution (no atomic structure or specificity).

A wealth of theoretical[18,19] and experimental[20–22] studies provide evidence for the close link between dynamics and function.[23–25] Two computational techniques most widely used for characterizing collective dynamics, and in particular elucidating the dominant (low frequency) modes of motion that usually relate to functional mechanisms, are normal mode analysis (NMA)[26–28] and essential dynamics analysis (EDA).[29] NMA approximates the atomic force field by a harmonic form, and yields, *via* the eigenvalue decomposition of the Hessian matrix, the shapes and frequencies of the $3N - 6$ normal modes underlying the collective dynamics. Gerstein and co-workers recently made an extensive NMA study of protein motions and built a database of 3814 putative motions.[30] EDA, on the other hand, decomposes the covariance (or cross-correlations) matrix **C** extracted from MD simulations to reconstruct the motions along the principal coordinates. A third approach considers a series of time-evolved snapshots or conformations visited at fixed intervals, which are projected into a space of collective coordinates by a singular value decomposition technique.[31,32]

Here, we use a purely mechanical model significantly simpler than the approaches described above, called the Gaussian network model (GNM)[33–37] and its recent extension, the anisotropic network model (ANM).[38–41] GNM views the structure as an elastic network (EN), the nodes of which are usually selected as the α-carbon atoms. Identical springs connect the "interacting residues", i.e. the pairs of nodes located closer than a certain cutoff distance. In parallel with the statistical mechanical theories developed for polymer gels,[42] the topology of internal contacts is represented by a connectivity or Kirchhoff matrix Γ. Γ fully defines the equilibrium dynamics of the structure. The GNM motions are induced by $N - 1$ modes. ANM, the extension of the GNM to 3D, yields $3N - 6$ modes and provides us with the directionality. ANM may be viewed as a coarse-grained NMA with uniform spring constants, in parallel with the NMA with uniform potentials originally proposed by Tirion at the atomic level.[43]

Four major advantages of the ANM compared to traditional NMA are (i) the possibility of estimating the motions in the absence of energy functions and parameters, (ii) the mathematical simplicity (or computational efficiency), thus permitting serial analysis for sets of proteins, (iii) the possibility of performing the analysis with low resolution structures or contact data using a hierarchical renormalization method,[44] and (iv) the efficient applicability to large structures, complexes or assemblies.[44,45,41] The slowest mode, also referred as the global mode, is usually insensitive to atomic details,[38,46–49] which lends support to the use of such a simple approach. The major disadvantages are the approximation of the global energy minimum by a quadratic well, and the ensuing prediction of motions that occur within this global well only. Transitions over an energy barrier cannot therefore be predicted, unless these involve relatively low barriers that are smoothed out in the coarse-grained description of the structure and energetics.[38]

Here, we show that the passage from the T to the R2 form is driven by the global mode of motion of the tetramer modeled as a GNM, and can be predicted accurately by the ANM. This is the first demonstration that this functional transition is in accord with the most cooperative, purely mechanical motion inherently driven (or favored) by the tetrameric architecture, *per se*. This gives insights as to why a local structural change induced by oxygen binding can trigger an allosteric response: the molecule is mechanically predisposed to undergo this particular conformational change. Oxygen binding perturbs the position of the proximal βHis92, which coincides exactly with the global hinge region and effectively propagates a cooperative conformational change. From another standpoint, inasmuch as the equilibrium dynamics of such ENs composed harmonic oscillators is fully controlled by entropic driving forces[50] that are uniquely defined by network topology, the present study invites attention to the importance of entropy, and the underlying interresidue contact topology, in facilitating the functional passage between the two equilibrium forms of Hb. Further analysis by full-atomic energy minimization reveals the specific interactions that are redistributed by the structural changes originally driven by entropic effects.

Results and discussion

The results are organized in four sections: (i) verification of the quantitative agreement between theory and experiments by comparing X-ray crystallographic *B*-factors of individual residues to their predicted amplitudes of motions; (ii) analysis of the global mode of motion of the tetramer, which reveals the hinge sites essential for the concerted rotation of the $\alpha_1\beta_1$ dimer relative to the $\alpha_2\beta_2$ dimer, and demonstration of the radical change in the dynamics of the so-called switch

region at the $\alpha_1-\beta_2$ and $\alpha_2-\beta_1$ interfaces (Figure 1(b)) upon ligand-binding; (iii) analysis of the second slowest mode which draws attention to domain movements coordinated the $\alpha_1-\beta_1$ and $\alpha_2-\beta_2$ interfacial interactions; and (iv) demonstration that the transition T → R2 is induced by the global mode of motion, and may be viewed as a reconfiguration facilitated by entropic effects and stabilized by specific interactions redistributed after global reconfiguration. Therefore, the "tense" form has an entropic tendency to assume the "relaxed" form. The opposite transition, R2 → T, on the other hand, cannot be explained by purely elastic forces but necessitates energetic contributions.

Agreement between theoretically predicted and experimental *B*-factors

The RCSB Protein Data Bank (PDB)[51] entries 1a3n (T), 1hho(R) and 1bbb(R2)[11,52,53] were used in the present computations. The X-ray crystallographic temperature factors, also called Debye–Waller factors or *B*-factors (B_i), are compared with the GNM-predicted mean-square (ms) fluctuations $\langle \Delta \boldsymbol{R}_i \cdot \Delta \boldsymbol{R}_i \rangle$ of residues $1 \le i \le N$ in the tetramer ($N = 574$) in Figure 2. The theoretical *B*-factors are found by inserting the GNM-predicted ms fluctuations in the equality $B_i = 8\pi^2 \langle \Delta \boldsymbol{R}_i \cdot \Delta \boldsymbol{R}_i \rangle / \mathbf{3}$ (see Materials and Methods for details). The Figure illustrates the result for one of the αβ-dimers in the T form, the results for the other dimer being almost indistinguishable. The serial indices 1–141 and 142–287 along the abscissa (here and in the other Figures) correspond to the respective α and β-chains. We see that the α-subunit fluctuations profile is very similar to that of the β-subunit. The close agreement between theoretical results and X-ray crystallographic temperature factors lends support to the utility and applicability of the GNM as an approximate but physically meaningful model.

The GNM spring constant γ for the harmonic potentials between near neighboring (interacting) residues is taken here as 0.88 kcal/(mol Å^2) (for the T form) to match the absolute size of experimental factors. This parameter rescales the absolute size of the ms fluctuations in general, but the distribution (or relative magnitudes) of the fluctuations among residues remains unaffected and uniquely defined for a given architecture. The value of the spring constant is coupled to the choice of the cutoff distance r_c for defining the "connected" residues of the network. A cutoff distance of 7.0 Å is usually adopted in GNM studies, consistent with the first coordination shell radius of amino acids deduced from the knowledge-based examination of PDB structures at the coarse-grained level of single site (α-carbon) per residue.[54] These values are in accord with the results from the GNM analysis of 113 high resolution crystal structures by Phillips and co-workers[55] where the best correlation between theory and experiments indicated an optimal value of $k_B T/\gamma = 0.87(\pm 0.46)$ Å2 with $r_c = 7.3$ Å. In the ANM approach, on the other hand, detailed examination of the effective ranges of interactions as a function of the strength of interactions indicated a cutoff distance of 12–15 Å[39] for maintaining the same spring constant, and we used a cutoff distance of 13 Å in ANM computations.

Intrasubunit and intersubunit couplings

The cross-correlation function $C_{ij} = \langle \Delta \boldsymbol{R}_i \cdot \Delta \boldsymbol{R}_j \rangle / (\langle \Delta \boldsymbol{R}_i \cdot \Delta \boldsymbol{R}_i \rangle \langle \Delta \boldsymbol{R}_j \cdot \Delta \boldsymbol{R}_j \rangle)^{\mathbf{1/2}}$ measures how/if the

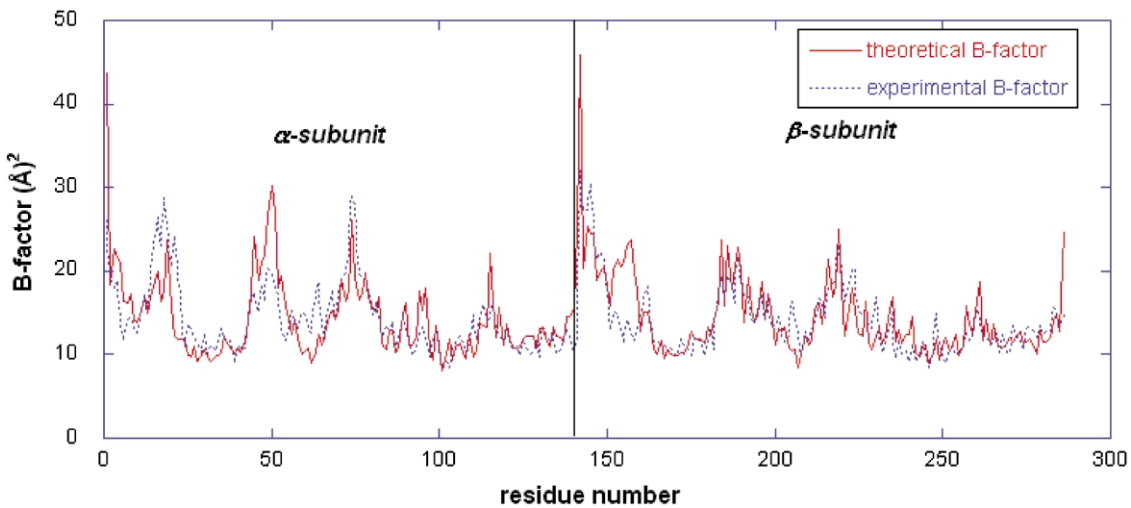

Figure 2. Comparison of theoretical (continuous curve) and experimental (broken curve) temperature factors for the α-carbon atoms in the chains α_1 and β_1 for HbA $\alpha_2\beta_2$ tetramer. Results for subunits α_2 and β_2 (not shown), are almost identical with those found for α_1 and β_1. Experimental data are taken from the PDB file 1a3n.[53] Theoretical results are calculated using the ms fluctuations $\langle \Delta \boldsymbol{R}_i \cdot \Delta \boldsymbol{R}_i \rangle$ computed by the GNM in $B_i = 8\pi^2 \langle \Delta \boldsymbol{R}_i \cdot \Delta \boldsymbol{R}_i \rangle / 3$. The correlation coefficient between the theory and experiments is found to be 0.71. Calculations repeated for the form R2 (not shown) yielded the same level of agreement with experiments (correlation coefficient 0.67).

movement of a given residue *i* correlates with that of residue *j*. C_{ij} varies between −1 and 1. It is close to 1 when the two residues are highly correlated, i.e. they concertedly move in the same direction in space. C_{ij} approaches −1, when the two residues are anti-correlated, i.e. they are coupled but move in opposite directions. C_{ij} tends to be zero when the two residues are uncorrelated. In this latter case they either do not move at all, or their motions are perpendicular to each other. Cross-correlations are conveniently displayed by correlation maps, the axes of which are the residue serial numbers.

The correlation map for the T form is shown in Figure 3 (see the color code in the legend). Similar maps were found for the forms T, R and R2. Residues within each subunit are found to be highly (auto)correlated, as evidenced by the square blocks along the diagonal. In other words, each of the individual subunits α_1, α_2, β_1 and β_2 undergo coherent, almost rigid-body movements *en bloc*. In addition to intrasubunit couplings, we observe distinctive intersubunit couplings between α_1 and β_1, and between α_2 and β_2 subunits. This indicates that there is a hierarchical coupling, firstly within monomers, and secondly within the dimers $\alpha_1\beta_1$ and $\alpha_2\beta_2$, consistent with the view of the tetramer as a dimer of cooperative αβ dimers.[9] The

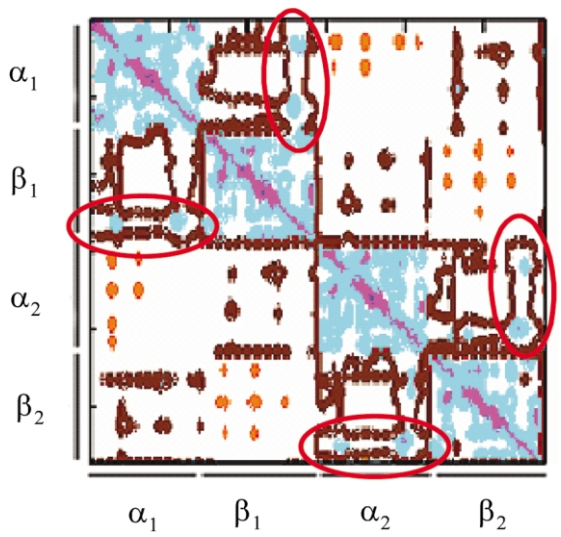

Figure 3. Correlation map for the cross-correlations C_{ij} between the motions of the residues. The two axes refer to residue indices. The regions colored magenta, blue, brown and orange refer to strongly correlated, correlated ($C_{ij} > 0$), uncorrelated ($C_{ij} \approx 0$) and anticorrelated ($C_{ij} < 0$) pairs, respectively. The subunits are indicated by the bars parallel with the axes. Blue blocks indicate positively correlated units (intrasubunit correlations for α_1, β_1, α_2 and β_2 domains) and the off-diagonal encircled (by ellipses) blue regions indicate the interdomain correlations between subunits α_1 (or α_2) and β_1 (or β_2). Anticorrelated (concerted but opposite direction) motions are observed for the subunit pair α_1–α_2, as well as the pair β_1–β_2. Results presented for 1a3n. 1bbb and 1hho showed similar behavior.

regions encircled by the red ellipses indicate the correlated pairs ($\alpha_1$27–37 and $\beta_1$115–132), and ($\alpha_1$100–117 and $\beta_1$107–127) that establish the coupling between the monomers that form the dimer $\alpha_1\beta_1$. We note that β131Gln, recently shown by site-directed mutations to play an important role in coordinating the communication between the $\alpha_1\beta_1$ and $\alpha_1\beta_2$ interfaces,[10] belongs to this group of residues. On the other hand, subunits α_1 and α_2 are anti-correlated ($C_{ij} < 0$), as well as subunits β_1 and β_2. The anti-correlated motions of these subunits are consistent with the counter-rotations of the $\alpha_1\beta_1$ dimer and $\alpha_2\beta_2$ dimer, which will be shown below to be the dominant global mode of motion of the tetramer.

The global motion is coordinated by a hinge site at the α_1–β_2 (or α_2–β_1) interface

Physical significance of the slowest (global) mode

This mode of motion involves the overall molecule, and this high cooperativity is usually relevant to function. It is the mode surviving (or dominating) at relatively long times, although it is likely to be influenced by friction and random force collisions with the solvent. The peaks in the slowest mode shape indicate the regions that enjoy the highest conformational flexibility when viewed from a relatively long time-window, whereas minima represent the hinge centers in these global modes. There is rotational flexibility at the hinge centers, but no translational mobility, hence their low ms fluctuations in the global mode.

Global dynamics of the T form

The two α-chains exhibit similar dynamics, as well as the two β-chains. We show the representative slow mode shape for one of the α and β-subunits in Figure 4(a). Three minima regions are distinguished in each case, at residues 33, 36–45, 85–94, and 138–141 in α-chains, and 35–43, 87–102, and 142–146 for β-chains. These are the global hinge-bending sites of the tetramer. Examination of the tetramer structure shows that these residues are all located at the α_1–β_2 (or α_2–β_1) interface (Figure 4(b)). The cross-correlations in the global mode (not shown) further show that subunits α_1 and β_1 move along the same principal direction; and subunits α_2 and β_2 undergo concerted, opposite direction, motions with respect to the dimer $\alpha_1\beta_1$. All these observations are consistent with the *en bloc* motion of the dimer $\alpha_1\beta_1$ relative to $\alpha_2\beta_2$ coordinated by the hinges (minima in the global mode shape) at the α_1–β_2 and β_1–α_2 interfaces; the particular mechanism of this motion, a global torsion, will be analyzed further below. We note that βTrp37 was recently shown by Ho and coworkers to be engaged in a slow conformational exchange in the CO-bound form of Hb in the presence of IHP, while βTrp14 and βTrp15 were not,[56] which is consistent with the participation of

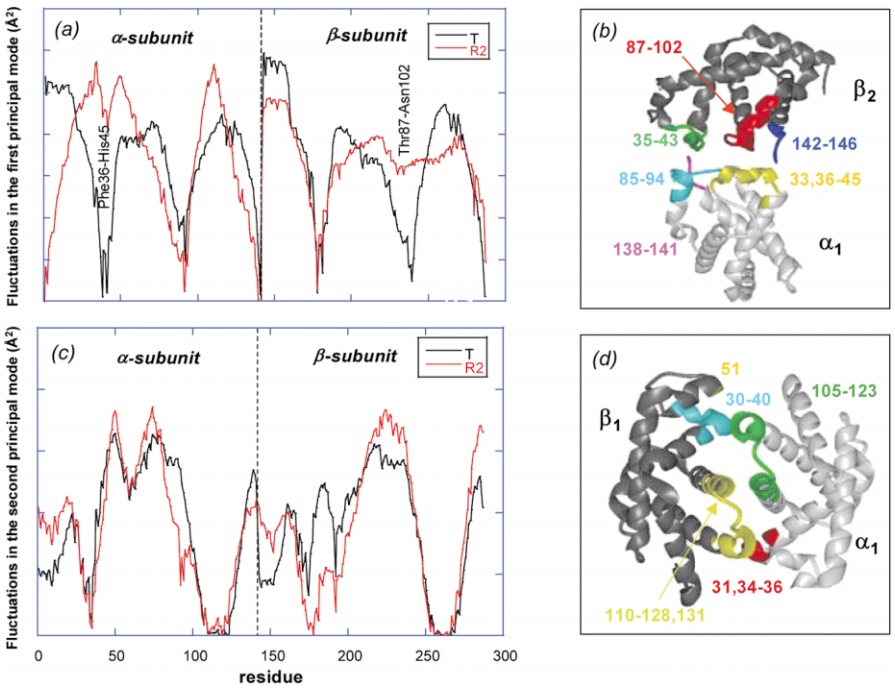

Figure 4. (a) Distribution of ms fluctuations driven by the global mode of motion. Results are presented for unliganded HbA T(1a3n) and carbon-monoxygenated HbA R2 (1bbb). The R curve falls between the T and R2 curves in accord with its intermediate structure (data not shown). Maxima refer to flexible regions, and minima to hinge sites. Note that compared with T, two minima at α36–45 and β87–102 disappear in the R form, suggesting a loosening or relaxation, and thereby an impediment in the mechanical function of the hinge sites. The dynamics of α and β-subunits are very similar in the T form, while the dynamics are different in the R2 form. (b) The hinge regions of the slowest mode correspond to the $\alpha_1\beta_2$ interface. Only the $\alpha_1\beta_2$ subunits for the T form are shown for clarity. The hinge regions are colored, as well as their labels (by the same color, for clarity). (c) Second slowest mode shape. Minima correspond to hinge sites at the $\alpha_1-\beta_1$ and $\alpha_2-\beta_2$ interfaces. (d) Hinge sites operative in the second slowest mode correspond to hinge sites at the $\alpha_1-\beta_1$ and $\alpha_2-\beta_2$ interfaces, shown for the R2 form.

βTrp37 (and not βTrp14 and βTrp15) in the global hinge site at the $\alpha_1-\beta_2$ interface.

Comparison of the dynamics of the T and R2 forms

The most drastic difference between the global mode shapes of the two forms is the change in the mobility of the R2 residues αPhe36-His45 and βThr87-Asn102 (Figure 4(a)). Interestingly, these segments precisely include the so-called switch region at the $\alpha_1-\beta_2$ interface (Figure 4(b)). The first mode of the tetramer thus directly interferes with the hinge-bending role of the switch regions at the $\alpha_1-\beta_2$ and $\alpha_2-\beta_1$ interfaces. Another difference lies in the N termini of the α-chains that are flexible in T and restricted in the R2 form. We are not yet sure of the functional implication of this difference in mobility at this region. Calculations repeated for the R form (not shown) confirmed the intermediate behavior of R between the T and R2 forms.

Allosteric changes mediated by the hinge region and the implication of the distinctive dynamics of the switch residues in the two forms

How does O_2 binding to a given subunit enhance the O_2 affinity of the other subunits in the same tetramer? Direct heme–heme interactions are out of the question because of the physical distance between them. Therefore the oxygen-binding signal at the heme pocket must propagate through the α–β interface to affect the binding affinity of the neighboring subunits. The present study draws attention to two stretches of residues, α36–45 and β87–102, at the $\alpha_1-\beta_2$ and $\alpha_2-\beta_1$ interfaces. These undergo a strongly cooperative change in dynamics from highly constrained (T form) to a flexible (or loose) (R2 form) behavior. The fact that β87–102, which is being affected directly upon O_2 binding, assumes a hinge role in the global dynamics suggest that local structural changes propagate into a global ones *via* hinge regions. Furthermore the fact that these residues loose their hinge role in the R state allows for efficient obstruction of the large-scale propagation of motions across subunits α_1 and β_2 (or α_2 and β_1) and explains the reduced cooperativity of the R form.

The presently observed decrease in communication efficiency and range (due to the loss of hinge function of residues at the α β interface), or reduced cooperativity, is a feature of the higher affinity (R) form of Hb emphasized in several experimental studies. It is well known that in the deoxy-Hb, the heme porphyrins are domed such

that the iron is displaced out of the heme plane towards the proximal histidine residues (α87 or β92) at the eighth position of helix F (F8).[4,57] Upon oxygenation, the iron moves into the heme plane, dragging the F helix that moves laterally ~1 Å across the heme. These changes propagate to the FG segment and to the neighboring subunit C-helix, and break several hydrogen bonds and salt bridges in the $\alpha_1\beta_2$ and $\alpha_2\beta_1$ interfaces (Figure 1(b)). Our analysis indicates that among the proximal histidine residues, those on the β-chain (βHis92) assume a critical hinge role rather than those on the α-chains (αHis87), which is consistent with the observation that the rHb (βH92G) cannot retain the T-form with distal ligand bound, while rHb (αH87G) does.[58]

Using genetically engineered mutants, Ho and colleagues have shed light on intersubunit hydrogen bonds that impart cooperativity and allostery in HbA. The interactions in which βAsp99 is engaged emerge as important determinants of the stability of the T form. For example, the mutant αV96W in the R form converts to the deoxy-quaternary structure (T) upon addition of an allosteric effector, inositol hexaphosphate,[59] which forms a hydrogen bond between αTrp96 and βAsp99 at the $\alpha_1\beta_2$ interface. Likewise, the mutation βD99N shows greatly reduced cooperativity and increased oxygen affinity, due to the disruption of the hydrogen bonds between hydrogen bonds βAsp99 and αTyr42 at the subunit interface.[60] The double mutant (βD99N and αY42D), on the other hand, restores the allosteric effect lost in the β D99N mutant,[60] because the second (compensatory) mutation αY42D restores the hydrogen bonds between the subunit interfaces. These observations are consistent with the presently revealed critical role of residues α36–45 and β87–102.

Tetramer global dynamics controlled by $\alpha_1\beta_2$ dimer dynamics

In the T form, the dynamics of the α and β-subunits exhibit common features (black curve in Figure 4(a)). In the R2 form, on the other hand, they differ (Figure 4(a), red curve) due to the relaxation of α36–45 and β87–102 at the $\alpha_1\beta_2$ (or $\alpha_2\beta_1$) interface. The change in dynamics at this particular interaction site seems to control the quaternary structure dynamics. Rivetti and co-workers showed that the binding curve was non-cooperative for a single Hb crystal that remained in the T conformation upon oxygen binding.[61] Deoxyhemoglobin encapsulated in silica gel, trapped in the T form, was also observed to bind oxygen non-cooperatively.[62] Heme binding module substitution experiments also indicate that the heme binding structure of tetrameric Hb is closely coupled to subunit interactions.[63] These observations raise the question: to what extent do the interactions at the $\alpha_1-\beta_2$ (or $\beta_2-\alpha_1$) interface control the dynamics of the tetramer?

To address this issue, we repeated the GNM analysis for individual α and β subunits, and for the dimers $\alpha_1\beta_1$ and $\alpha_1\beta_2$ (or $\alpha_2\beta_2$ and $\alpha_2\beta_1$). The dynamics of subunits were found to be almost indistinguishable when examined as individual subunits. The dimer $\alpha_1\beta_1$ (Figure 5(a)), considered separately, also yielded almost indistinguishable features for the α and β-subunits' global modes. The distinctive dynamics of the α and β-subunits in the tetramer R2 form could be captured only when the $\alpha_1\beta_2$ (or $\alpha_2\beta_1$) dimer was considered (compare Figure 5(b) and (c)). This confirms that the interaction between $\alpha_1$and β_2 *is* the major determinant of the dynamics of the tetramer, and implies that cooperative oxygen binding is conveyed by $\alpha_1-\beta_2$ (and $\alpha_2-\beta_1$) interaction(s).

The second slowest mode

Significance

While the first GNM mode usually defines the most cooperative and functional conformational changes, the second mode can play a complementary role in some cases and it may be of interest to examine the associated motion as another cooperative mechanism inherently favored by the tetrameric architecture. The contribution of the second mode to the dynamics of Hb is found to be weaker than that of the first mode's by a factor of 2 (the ratio of the corresponding eigenvalues). Yet, it is interesting to find that this mode controls the relative movements of the dimers $\alpha_1\beta_2$ and $\alpha_2\beta_1$ in the T form, and $\alpha_1\alpha_2$ and $\beta_1\beta_2$ in the R2 form, in contrast to mode 1, where the dimer $\alpha_1\beta_1$ motion relative to $\alpha_2\beta_2$ was controlled. Figure 4(c) displays the fluctuation behavior of the α and β-subunits in this mode, and Figure 4(d) illustrates the regions (residues α 9, 30–31, 34–35, 104–127 and β 3, 33–34, 110–132 for T (not shown); α 31, 34–35, 105–123 and β 30–40, 51, 110–128 for R2) emerging as the hinge-bending centers (minima in part c) in this mode. These are all located at the $\alpha_1-\beta_1$ (or $\alpha_2-\beta_2$) interface.

The role of $\alpha_1-\beta_1$ interface in Hb allostery is still a matter of debate. Ho and colleagues showed that the mutation $\beta_1$31Gln in the $\alpha_1-\beta_1$ interface can affect the reactivity of the SH group of β93Cys (located at the $\alpha_1-\beta_2$ interface) toward *p*-mercuribenzoate. This result indicates the communication between the $\alpha_1\beta_1$ and $\alpha_1\beta_2$-subunit interfaces. Ackers measured the free energy of dissociation of the tetramer to αβ dimers for all possible ligation states and found they agreed mostly with the predictions of the MWC model except those for a small intradimer cooperativity.[8] Further experiments by Ackers and co-workers support a Symmetry Rule model of allosteric regulation.[9]

In order to assess the most critical interactions and their functional implications at the $\alpha_1-\beta_1$ interface, we ranked the hydrogen bonds at the $\alpha_1-\beta_1$ (or $\alpha_2-\beta_2$) interface on the basis of their flexibility in the GNM mode 2. The hydrogen bond formed between αPro114 O and βHis116 NE2 is found to

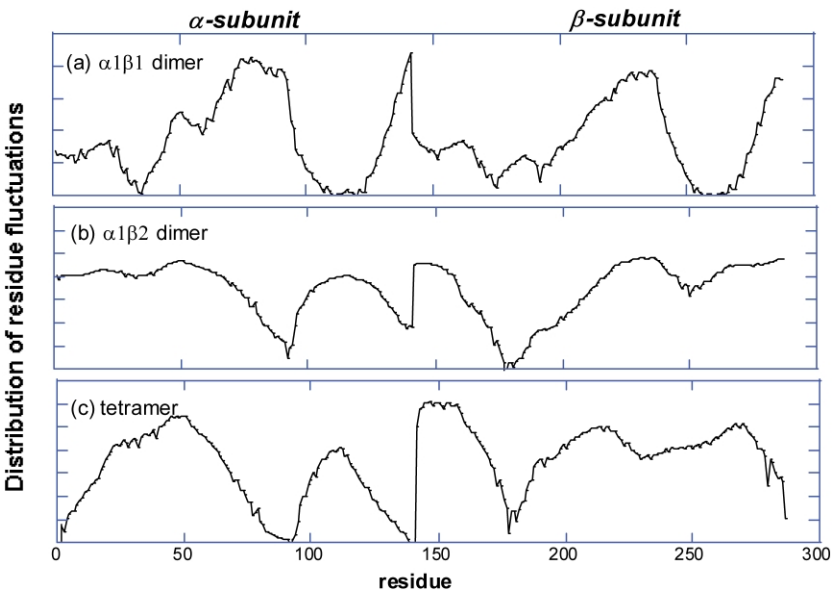

Figure 5. Comparison of the first principal mode of R2 tetramer (c) to that of $\alpha_1\beta_1$ (a) and $\alpha_1\beta_2$ (b) dimers. The coordinates for the R2 tetramer, the $\alpha_1\beta_1$ dimer or $\alpha_1\beta_2$ dimer of R2 are used to calculate the square fluctuation driven by the slowest mode. We can see that in the $\alpha_1\beta_1$ dimer, the α-subunit dynamics is similar to the β-subunit dynamics while in the $\alpha_1\beta_2$ dimer, the α-subunit dynamics is very different from the β-subunit dynamics, and the profiles $\alpha_1\beta_2$ dimer and R tetramer are very similar, indicating that the global mode shape of the tetramer is predominantly controlled by that of the $\alpha_1\beta_2$ (or $\alpha_2\beta_1$) dimer.

be the most constrained, which implies that this bond is highly stable for both the T and R2 forms. Other pairs of residues subject to stable interactions are αPhe117-βArg30, αArg31-βPhe122, αArg31-βGln127, αHis103-βGln131, αHis122-βArg30 and αAsp126-βTyr35.

T ↔ R2 transition is driven by the most coordinated (global) motion of Hb

The following approach is used to characterize the global structural changes favored by elastic modes. We start with the T form. Using ANM, we find the displacements (or deformations) driven by the slowest mode of motion, designated as $\{\Delta \boldsymbol{R}_i^{(1)}\}$. The superscript (1) in $\{\Delta \boldsymbol{R}_i^{(1)}\}$ indicates the first (slowest) mode. The fluctuating (or deformed) conformations visited by the action of the global mode are found by combining the calculated displacements with the known position vectors as $\boldsymbol{R}_i^{\text{new}} = \boldsymbol{R}_i \pm \alpha \Delta \boldsymbol{R}_i^{(1)}$, where $\alpha(\geq 0)$ is a scalar that determines the size of the displacement. Two sets of position vectors are obtained from this analysis, which refer to positive and negative α values. The corresponding reconfigured T forms, shortly designated as $T^+(\alpha)$ and $T^-(\alpha)$ respectively, are minimized using the package of programs MOIL[64] with a tolerance of 0.2 kcal/mol. The root-mean-square deviation (RMSD) between each energy-minimized conformation and the two stable forms T and R2 are examined. The two sets of conformations, $T^+(\alpha)$ and $T^-(\alpha)$, are equally distant from the original equilibrium state (T), the departure increases with increasing α, as expected. On the other hand, comparison with the R2 form reveals that one of the reconfigured T forms, $T^+(\alpha)$, gradually approaches the R2 state as α increases (while the other, $T^-(\alpha)$, moves farther away) (Figure 6, open circles). Thus, the purely elastic (or entropic) forces initiate deformations that are potentially conducive to the R2 state.

Figure 7 compares the T (top) and R2 (middle) forms, with the more favorable reconfigured T form ($T^+(\alpha)$, bottom) that we computed. The spatial orientations and conformations of the $\alpha_1\beta_1$ subunits (colored gray) are almost identical in the

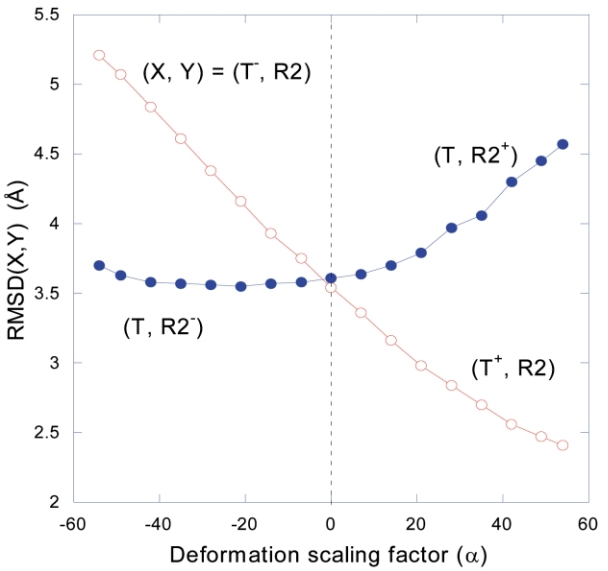

Figure 6. Comparison of the deformed structures generated by ANM with the different forms of Hb. The open circles describe the RMSD value (Å) between the R2 form and the deformed structures ($T^+(\alpha)$, right) and ($T^-(\alpha)$, left) obtained for the T form by varying the parameter α that scales the extent of departure from the original state as induced by mode 1 (see equation (15)). The filled circles refer to the RMSD values between the T form and reconfigured R2 forms ($R2^+(\alpha)$ and $R2^-(\alpha)$). We note that T^+ gradually approaches the R2 form (RMSD value of 2.41 Å reached at a $\alpha = 54$), while the reconfigured R2 form does not approach the T form. The RMSD values are calculated in each case by energy minimizing (with MOIL) the ANM-prediced coarse-grained deformed conformations.

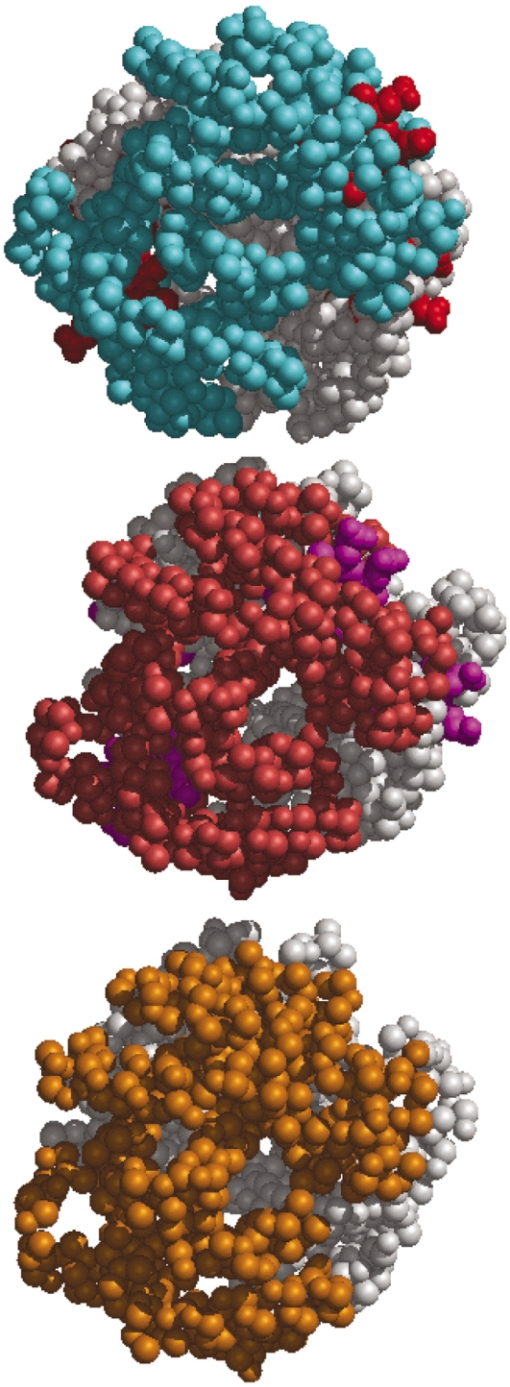

Figure 7. Comparison of the T (top), R2 (middle) and reconfigured T (bottom) forms of Hb. The T and R2 forms refer to the PDB structures deoxy-Hb[53] and CO-bound Hb.[11] The reconfigured T is predicted by equation (15). We started from the T form and compute its most probable global change using ANM. Note the similarity between the deformed T form predicted by ANM (bottom) and the experimental R2 form (middle; RMSD, 2.4 Å).

three cases. Their RMSD values after pairwise optimal superimposition yields <1.2 Å RMSD for all pairs. The top portions of the diagrams show the $\alpha_2\beta_2$ subunits in blue (T), pink (R2) or orange (reconfigured T). The heme groups are red and magenta in the upper two diagrams. Comparison of the upper two diagrams permits us to visualize the structural difference between the T and R2 forms: a counterclockwise rotation of the $\alpha_2\beta_2$ unit in the R2 form about the out-of-plane axis (by $>15°$) relative to its orientation in the T form. This torsional movement is facilitated by the weakening of a few severely constrained interactions at the $\alpha_1-\beta_2$ and $\alpha_2-\beta_1$ interfaces, which are predicted above, in agreement with the information deduced from experiments (see Figure 4(a) and (b)).

The striking feature in Figure 7 is the close similarity between the two lower diagrams, i.e. the R2 and the reconfigured T forms. The RMSD values between these two structures is 2.32 Å. The present theory thus predicts that the R2 form is a configuration that is directly stabilized by the global relaxation mode of the T form, i.e. the T form possesses an inherent potential derived from entropic effects for being reconfigured into the R2 form.

In order to gain an atomic description of the reconfiguration path from the T state toward the R2 state, the above-minimized reconfigured T forms were inspected. In the T state αThr41 and βArg40 do not form a close contact. The distance between the amine group of βArg40 side-chain and the carbonyl backbone of αThr41 is 5.65 Å. However, in the R2 state the two amino acid residues approach and the distance becomes 3.76 Å. We were able to depict this motion by observing the intermediate forms of reconfigured T state. A further analysis shows that in the minimized T state the above distance is 4.24 Å; upon reconfiguration of the T state the two amino acid residues approach and the distance becomes 2.8 Å (see Figure 8(a)). Another example for the change in close interactions at the $\alpha_1\beta_2$ interface is the triad βTrp37, βAsn102 and αAsp94. In the T state, βTrp37 forms a hydrogen bond between the NH group of its indole ring and the side-chain carboxyl group of αAsp94. βAsn102 does not form close contacts with these two amino acid residues. The distance between the βAsn102 side-chain carbonyl group and the βTrp37 indole NH is 5.45 Å, and the distance between the βAsn102 side-chain amine and the αAsp94 side-chain carboxyl is 5.55 Å. In the R2 state βAsn102 approaches βTrp37 and αAsp94 and the above distances become 2.98Å and 2.77 Å, respectively. The hydrogen bond between βTrp37 and αAsp94 is broken and the two amino acid residues move apart. Upon reconfiguration of the T state the distance between the βAsn102 carbonyl group and the βTrp37 indole NH shorten from 4.54 Å to 4.1 Å, and the distance between the βAsn102 side-chain amine and the αAsp94 side-chain carboxyl group shorten from 5.56 Å to 3.34 Å. Figure 8(b) shows how in the intermediate states of reconfigured T form βAsn102 approach βTrp37 and αAsp94 to form close contacts. These switches between interacting amino acid pairs are consistent with previous studies of Ho and co-workers.[8]

As a final test, we repeated the same analysis by

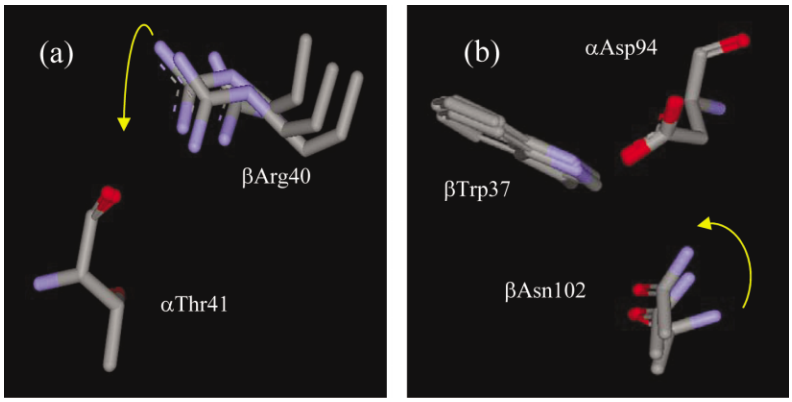

Figure 8. Changes in the contacts at the $\alpha_1-\beta_2$ interface of the intermediates visited upon the reconfiguration of the T forms. The Figure shows three states obtained by changing α the scaling factor ($\alpha = 0$, $\alpha = 28$ and $\alpha = 54$). (a) βArg40 approaches the carbonyl backbone of αThr41. (b) βAsn102 approaches βTrp37 and αAsp94 to form contact between its amine group and the αAsp94 carboxyl group, and between its carbonyl group and the βTrp37 indole NH.

taking the R2 form as the starting configuration and generating the deformed conformations $R^{+}(\alpha)$ and $R^{-}(\alpha)$. The results from this analysis are presented in Figure 6 (filled circles). A weak tendency to approach the T form is discernible in $R^{-}(\alpha)$ at around $\alpha = -21$, although further conformational change in the same direction does not progress towards the T form, but leaves almost unchanged the original RMSD value. Thus, in contrast to the T → R2 transition, the backward passage R2 → T does not appear to be a global change favored by purely entropic effects. This behavior can be rationalized by the fact that entropic effects might favor transitions in the direction of increasing disorder (or weakening internal interactions), as is the case for T → R2. The opposite change, on the other hand, is accompanied by an enhancement in cooperativity manifested by the constraints at the hinge sites (Figure 2), which cannot be favored by entropic effects, but by specific interactions that stabilize particular contacts.

It is worth noting that the dynamics of the tetramer results from the superposition of several modes, and other modes can also contribute to the transition between the two forms, in the order of decreasing importance as we consider higher frequency modes. As described above, the second slowest mode drives a cooperative motion that is totally different from the first mode. However, some of the higher frequency modes were observed to induce fluctuations similar to those of the first slowest mode. While the statistical contribution of these modes is much smaller, this observation invites attention to the occurrence of multiple pathways for the transition between T and R forms, as also pointed out in a classical NMA performed by Mouawad & Perahia for Hb tetramer.[65] Interestingly this study also showed the presence of a set of preferential modes that brings the T form close to the R form, while the opposite passage could not be observed, i.e. normal modes followed by energy minimization starting from the R state, did not allow the structure to assume a conformation significantly different from the R form. The R form was then concluded to be trapped in a deep and narrow potential energy minimum.

Based on free energy calculations, Eaton and co-workers[66] found that the T ↔ R transition state is closer to the R state than the T state. Another interesting docking experiment was performed by Janin & Wodak.[67] A reaction pathway between the T and R forms was constructed by requiring the intermediate conformations to retain the largest buried surface area between the $\alpha_1\beta_1$ and $\alpha_2\beta_2$ dimers. The passage from the T state to the R state was found to comply with this requirement, i.e. the T state αβ dimers can rotate to a near R state while maintaining much of the buried surface area originally present in the T state. The opposite passage, on the other hand, from R to T state could not be achieved due to significant loss in the buried surface area. This suggests a reaction pathway in which the αβ dimer in the T state rotates to a near R state intermediate in which the αβ dimers undergo a tertiary conformational rearrangement towards the R state. Our results agree qualitatively with the above studies. Notably, unlike the above NMA study, the present results are found without including any specific, atomic energy functions or parameters, but simply taking rigorous account of contact topology in the native state, demonstrating the dominant role of contact topology (or associated entropic effects) in determining the functional transition from T to R2 form.

Time-scale of the observed process

Gibson found that after photodissociation, carbon monoxide rebinds much faster than the initial rate of deoxyhemoglobin and carbon monoxide binding.[68] This suggested that this "fast-reacting" deoxyhemoglobin has a different affinity and possibly a different structure. Later Hopfield and colleagues found that the fast-reacting hemoglobin is the Hb that has not switched from R to T structure after photodissociation.[69] By nanosecond-resolved spectroscopy, Hofrichter and co-workers showed that the quaternary structure controls the overall rate of CO dissociation,[70] and these quaternary changes are seen at the scale of tens of microseconds.[71]

The time-scale of the present global mode is in the sub-nanoseconds regime ($0.62\ cm^{-1}$) as

estimated from the approximation $\omega = (\gamma\lambda_1/m)^{1/2}$ where λ_1 is the first mode eigenvalue (0.0545) and m is the Hb mass (64.5 kDa). This time-scale is comparable to that of the low frequency mode of aspartate transcarbamylase (0.52 cm^{-1}) obtained by NMA,[72,6] and differs from the one calculated by NMA[73] by a factor of three, approximately. Time-resolved resonance Raman and circular dichroism measurements indicate that the transition of Hb might occur in several steps, which is confirmed by time-resolved UV magnetic circular dichroism with βTrp37-α94Asp as marker.[74] Ho and co-workers observed a conformational exchange on the micro- to millisecond time-scale by studying the dynamics of βTrp37 in NMR.[22] Our calculations most likely refer to the initiation stage of the transition from the T state to a near R state intermediate, and indicate the mechanical tendency of the overall architecture to undergo this passage, rather than the completion of the transition. This tendency of the molecule is consistent with the studies of Janin & Wodak[67] and Eaton and co-workers[66] that suggest the occurrence of a transition state close to the R form.

Conclusion

Here, we demonstrate that a simple elastic network model, GNM/ANM, can predict the transition from the T form to the R2 form of Hb. Numerous mutational and spectroscopic studies have shown that the hydrogen bond and salt bridge formation at the $\alpha_1\beta_2$ interface are key to the allosteric activity of Hb. Without knowledge of any chemical and biological data on Hb function, using structural (purely geometric) information alone, we show that the hinge center for the most coordinated motion of Hb corresponds to the switch region at the $\alpha_1\beta_2$ interface. Numerous liganded and unliganded Hb revealed a number of structures that fall in between T and R2 states. Our studies tie the global motion of Hb directly to this functional conformational transition. Additionally, we identify a number of other interactions at the $\alpha_1-\beta_1$ interface, which can play a role in establishing the communication of the dimers.

How is it possible to predict conformational transitions with a simple, non-specific, linear model? Previous applications have shown that the global mode of proteins is insensitive to the details of structure; and the collective domain movements, hinge sites for these movements can be predicted without necessitating detailed knowledge of atomic structure and energy parameters. Yet, one would expect the movements predicted by EN models to occur in the close neighborhood of the original equilibrium state, i.e. these should be fluctuations within the same global energy minimum. In a recent comparison with MD simulations,[38] we observed that the motions that involve a passage over an energy barrier can be approximated by the ANM, provided that these occur in the close neighborhood of the native state. The coarse-grained description adopted in the ANM smoothes out the energy surface and allows for sampling nearby configurations that are otherwise separated by local energy barriers. The prediction of the T → R2 transition of Hb by this model also invites attention to the validity of this conjecture.

An important implication of the present analysis is that purely mechanical, or in the thermodynamics sense, purely entropic, effects play a significant role in inducing the T → R2 transition. The picture that emerges is that in the presence of a perturbation, or a destabilizing effect such as ligand binding, the most probable conformation to be visited by the T form is the R form, the passage being favored by entropic effects; and the reconfigured form is then stabilized by a redistribution of interresidue interactions (hydrogen bonds, salt bridges, etc.) at the $\alpha_1-\beta_2$ (and $\alpha_2-\beta_1$) interface(s). This is the first demonstration that entropic effects significantly contribute to directing the functional transition of Hb.

One may conceive that the methodology can be extended to even larger structural rearrangements of multimeric complexes or assemblies by further coarse-graining the structure. This can be an extremely efficient way of estimating cooperative structural changes, because GNM takes a few seconds, ANM takes about two hours of computing time on a SGI workstation with R12000 processor for predicting the size (GNM) and mechanism (ANM) of the most cooperative motions intrinsically favored by a given architecture of the size of Hb tetramer. It is sufficient to have one known structure, even at low resolution (of the order of electron microscopy images) for predicting the most likely cooperative movements. The next challenging application could be, for example, to be able to predict the EM image of a GroEL–GroES complex in the R form determined by Saibil and co-workers[75] starting from the well-known complex in the T form.[76]

Materials and Methods

GNM

In the GNM the structure is viewed as an EN. Each node is identified by a single residue (in our case Cα atom), and pairs of residues located within a cutoff distance, r_c, are assumed to be connected by elastic springs. The heme is modeled by taking a representative set of atoms that retain the same level of resolution as the protein (approximately 3.8 Å separation between adjacent atoms), using the atomic coordinates of the hemes in the PDB-deposited structures. Calculations are verified to be insensitive to the inclusion/exclusion of small molecules (like O_2 or CO). The difference in the dynamics of the ligand-bound forms essentially results from the distortion in the atomic coordinates of the tetramer, rather than the explicit inclusion of these small molecules. A uniform spring constant γ is adopted for all pairs. The dynamics of the network is fully defined by the topology

of interresidue contacts, or the Kirchhoff matrix $\mathbf{\Gamma}$, also called connectivity or adjacency matrix in graph theory.[77,78] For a network of N interacting sites, the elements of $\mathbf{\Gamma}$ are defined as:

$$\Gamma_{ij} = \begin{cases} -\mathbf{1} & \text{if } i \neq j \text{ and } \boldsymbol{R}_{ij} \leq r_c \\ \mathbf{0} & \text{if } i \neq j \text{ and } \boldsymbol{R}_{ij} > r_c \\ -\sum_{i \neq j} \Gamma_{ij} & \text{if } i = j \end{cases} \quad (1)$$

where $\boldsymbol{R}_{ij}$ is the distance between sites i and j (given by X-ray crystallography, NMR or EM). A cutoff distance of 7 Å is adopted in GNM calculation. $\mathbf{\Gamma}$ conveys the same type of information as conventional contact maps. Its non-zero off-diagonal terms indicate the interacting pairs, and its ith diagonal term is equal to the coordination number of the ith site. The probability distribution in $\Delta\boldsymbol{R}_i$ and the associated free energy of entropic origin are given by:

$$W(\Delta\boldsymbol{R}_i) = \exp\{-3(\Delta\boldsymbol{R}_i)^2/\mathbf{2}\langle(\Delta\boldsymbol{R}_i)^2\rangle\}$$
$$\Delta G_i = -T\Delta S_i = k_B T \ln W(\Delta\boldsymbol{R}_i) \quad (2)$$

where k_B is the Boltzmann constant, T is the absolute temperature and $\langle(\Delta\boldsymbol{R}_i)^2\rangle$ is the ms fluctuation of residue i. The correlations $\langle\Delta\boldsymbol{R}_i \cdot \Delta\boldsymbol{R}_j\rangle$ between the $\Delta\boldsymbol{R}_i$ and $\Delta\boldsymbol{R}_j$ are found from the statistical mechanical average:[33]

$$\langle\Delta\boldsymbol{R}_i \cdot \Delta\boldsymbol{R}_j\rangle = (\mathbf{1}/Z_N)\int(\Delta\boldsymbol{R}_i \cdot \Delta\boldsymbol{R}_j)e^{-V/k_B T}d\{\Delta R\}$$
$$= (3k_B T/\gamma)[\mathbf{\Gamma}^{-1}]_{ij} \quad (3)$$

where $[\boldsymbol{\Gamma}^{-1}]_{ij}$ denotes the ijth element of the inverse of $\mathbf{\Gamma}$, $\{\Delta\boldsymbol{R}\}$ is the $3N$-dimensional column vector $\{\Delta\boldsymbol{R}_1, \Delta\boldsymbol{R}_2, \ldots\Delta\boldsymbol{R}_N\}$ of the fluctuations of all nodes, and Z_N is the configurational partition function:

$$Z_N = \int \exp\left(-\frac{\gamma}{2k_B T}\{\Delta R^T\}\mathbf{\Gamma}\{\Delta\boldsymbol{R}\}\right)d\{\Delta\boldsymbol{R}\} \quad (4)$$

The cross-correlations in Figure 3 are found from:

$$C_{ij} = \langle(\Delta\boldsymbol{R}_i \cdot \Delta\boldsymbol{R}_j)\rangle/[\langle(\Delta\boldsymbol{R}_i)^2\rangle\langle(\Delta\boldsymbol{R}_j)^2\rangle]^{1/2}$$
$$= [\mathbf{\Gamma}^{-1}]_{ij}/([\mathbf{\Gamma}^{-1}]_{ii}[\mathbf{\Gamma}^{-1}]_{jj})^{1/2} \quad (5)$$

The potential energy V in equation (3) may be written as:

$$V = (\gamma/2)\{\Delta R^T\}\mathbf{\Gamma}\{\Delta\boldsymbol{R}\}$$
$$= (\gamma/2)\sum_i \sum_j h(r_c - \boldsymbol{R}_{ij})(\Delta\boldsymbol{R}_j - \Delta\boldsymbol{R}_i)^2 \quad (6)$$

where the superscript T denotes the transpose, the summations are performed over all interaction sites, $h(x)$ is the Heavyside step function equal to 1 if $x \geq 0$, and zero otherwise, thus sorting all pairs located closer than the cutoff distance r_c. The ms fluctuations of residues and corresponding X-ray crystallographic B-factors $B_i = 8\pi^2\langle(\Delta\boldsymbol{R}_i)^2\rangle/\mathbf{3}$ are found using equation (3) with $i = j$, i.e.:

$$\langle(\Delta\boldsymbol{R}_i)^2\rangle = (\mathbf{3}k_B T/\gamma)[\mathbf{\Gamma}^{-1}]_{ii} \qquad B_i = \mathbf{8}\pi^2(k_B T/\gamma)[\mathbf{\Gamma}^{-1}]_{ii} \quad (7)$$

Equation (7) provides an extremely efficient way of estimating the ms fluctuations of individual residues. We determine $[\mathbf{\Gamma}^{-1}]$ by eigenvalue transformation of $\mathbf{\Gamma}$ (see below) and reconstruction of $\mathbf{\Gamma}^{-1}$ after removal of the zero eigenvalue. The GNM software is accessible on the internet†.

Finally we note that the theory yields a unique solution for the distribution of fluctuations for each particular architecture. However, the absolute sizes of fluctuations depend on the spring constant γ. We usually adjust γ to match the average B-factor (over all residues) from experiments. γ is thus found to be 0.88 kcal/(mol Å^2) using $r_c = 7.0$ Å. ANM calculations necessitate the adoption of a larger cutoff distance in order to maintain the same spring constant and to ensure the occurrence of no more than six zero eigenvalues. We used the cutoff distance of 13 Å in ANM calculations, consistent with our previous examination[39] of the change in r_c with γ. The ANM results were verified to be almost insensitive to the choice of r_c provided that $r_c > 10$ Å.

Mode analysis (MA)

A major advantage of the GNM is the possibility of determining the global modes of a quaternary structure with a minimal computational cost. To this aim, the first step is the eigenvalue decomposition of $\mathbf{\Gamma}$ as:

$$\mathbf{\Gamma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1} \quad (8)$$

where $\mathbf{U}$ is the orthogonal matrix whose columns $\boldsymbol{u}_k$, $1 \leq k \leq N$, are the eigenvectors of $\mathbf{\Gamma}$, and $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues (λ_k). The eigenvalues represent the frequencies of the $N-1$ non-zero GNM modes, usually organized in ascending order, i.e. $\lambda_1 \leq \lambda_i \leq \lambda_{N-1}$, and $\lambda_N = 0$. The ith element $(\boldsymbol{u}_k)_i$of the kth eigenvector describes the motion of residue i along the kth principal coordinate. The ms fluctuations of residues can be rewritten as a weighted sum over all modes as:[34]

$$\langle(\Delta\boldsymbol{R}_i)^2\rangle = \sum_k [(\Delta\boldsymbol{R}_i)^2]_k = \sum_k (\mathbf{3}k_B T/\gamma)[\lambda_k^{-1}(\boldsymbol{u}_k)i(\boldsymbol{u}_k)i] \quad (9)$$

and the contribution of the kth mode as:

$$[\Delta\boldsymbol{R}_i \cdot \Delta\boldsymbol{R}_j]_k = (\mathbf{3}k_B T/\gamma)[\gamma_k^{-1}\boldsymbol{u}_k\boldsymbol{u}_k^T]_{ij} \quad (10)$$

The shape of the kth mode, obtained by plotting $(\boldsymbol{u}_k)_i(\boldsymbol{u}_k)_i$ as a function of residue index i for $1 \leq i \leq N$, represents the distribution of residue mobilities in mode k. By definition this distribution is normalized, i.e $\sum_i \times (\boldsymbol{u}_k)i(\boldsymbol{u}_k)i = \mathbf{1}$. λ_k^{-1} serves as a statistical weight for mode k. The slowest modes thus make the largest contribution to $\langle(\Delta\boldsymbol{R}_i)^2\rangle$.

ANM

The fluctuations estimated by the GNM are by definition isotropic, i.e. $\langle(\Delta X_i)^2\rangle = \langle(\Delta Y_i)^2\rangle = \langle(\Delta Z_i)^2\rangle = \langle(\Delta\boldsymbol{R}_i)^2\rangle/\mathbf{3}$, and the system is described by $N-1$ internal modes, instead of $3N-6$, which provide information on the fluctuation sizes, only. No information on the fluctuation directions is provided. ANM addresses this deficiency.[39] ANM is in spirit similar to GNM, no residue specificity or non-linear effects are present. The major difference is that the potential for each pair (i, j) of interacting residues is $\{(\gamma/2)(|\boldsymbol{R}_{ij}| - |\boldsymbol{R}_{ij}^0|)^2\}$ in the ANM, as opposed to $\{(\gamma/2)(\Delta\boldsymbol{R}_{ij} \cdot \Delta\boldsymbol{R}_{ij})\}$ in the GNM. Here $\boldsymbol{R}_{ij}^0$ is the original distance vector between nodes i and j, $\boldsymbol{R}_{ij}$ is that after deformation, and $\Delta\boldsymbol{R}_{ij} = \boldsymbol{R}_{ij} - \boldsymbol{R}_{ij}^0$. In the

† http://www.ccbb.pitt.edu/CCBBResearchDynProGNM.htm

ANM, $\mathbf{\Gamma}$ is replaced by the $3N \times 3N$ connectivity matrix $\boldsymbol{H}$, the contacts along the X, Y and Z-directions (instead of a radial distance of R) being explicitly considered. $\boldsymbol{H}$ is proportional to the Hessian matrix of NMA with the two major simplifications: (i) all residues are uniform, coupled with the same force constant; and (ii) a coarse-grained representation (as opposed to full atomic) is adopted. Like GNM, ANM yields an analytical solution, thus avoiding simulations with sophisticated potentials or energy minimization. $\boldsymbol{H}$ is a $N \times N$ supermatrix, the off-diagonal super-elements of which are:

$$\mathbf{H}_{ij} = \begin{bmatrix} \partial^2 V/\partial X_i \partial X_j & \partial^2 V/\partial X_i \partial Y_j & \partial^2 V/\partial X_i \partial Z_j \\ \partial^2 V/\partial Y_i \partial X_j & \partial^2 V/\partial Y_i \partial Y_j & \partial^2 V/\partial Y_i \partial Z_j \\ \partial^2 V/\partial Z_i \partial X_j & \partial^2 V/\partial Z_i \partial Y_j & \partial^2 V/\partial Z_i \partial Z_j \end{bmatrix} \quad (11)$$

and the diagonal super elements are found from the negative sum of the off-diagonal terms, as $\mathbf{\Gamma}$. The second derivatives of V with respect to the components of the position vectors can be readily found from equation (6), e.g. $\partial^2 V/\partial X_i \partial Y_j = -\gamma(X_j - X_i)(Y_j - Y_i)/\boldsymbol{R}_{ij}^2$.

The correlations between fluctuations can be derived from equation (3) if $\gamma^{-1}\mathbf{H}$ replaces $\mathbf{\Gamma}$, which yields the well-known inverse dependence of the covariance matrix $\mathbf{C}$ and the Hessian matrix. The mode frequencies in the ANM (or GNM) are given by $\omega_i = (\gamma\lambda_i)^{1/2}$, where λ_i is the ith eigenvalue of $\gamma^{-1}\mathbf{H}$ (or $\mathbf{\Gamma}$), for $1 \le i \le 3N - 6$ (or $N - 1$). Let $\boldsymbol{C}_{ij}$ designate the ijth superelement of $\boldsymbol{C}$, i.e. the 3×3 matrix of the correlations between the components of $\Delta\boldsymbol{R}_i$ and $\Delta\boldsymbol{R}_j$. $\boldsymbol{C}_{ij}$ is expressed as a sum over $3N - 6$ non-zero modes, as:

$$C_{ij} = \langle \Delta\boldsymbol{R}_i \Delta\boldsymbol{R}_j^{\mathrm{T}} \rangle = (k_{\mathrm{B}}T/\gamma) \sum_k \lambda_k^{-1} [(\boldsymbol{u}_k^{\mathrm{ANM}})(\boldsymbol{u}_k^{\mathrm{ANM}})^{\mathrm{T}}]_{ij} \quad (12)$$

Here $(\boldsymbol{u}_k^{\mathrm{ANM}})$ is the kth eigenvector found by diagonalizing $\gamma^{-1}\mathbf{H}$, and the term $[(\boldsymbol{u}_k^{\mathrm{ANM}})(\boldsymbol{u}_k^{\mathrm{ANM}})^{\mathrm{T}}]_{ij}$ designates the ijth superelement of the matrix enclosed in square brackets. The ms fluctuations of residue i, or corresponding B-factor are found from:

$$\langle (\Delta\boldsymbol{R}_i)^2 \rangle = \sum_k [(\Delta\boldsymbol{R}_i)^2]_k = \mathrm{tr} C_{ii}$$
$$B_i = (8\pi^2/3)\ \mathrm{tr}\ C_{ii} \quad (13)$$

where tr designates the trace of $\mathbf{C}_{ii}$, and the contribution of mode k is calculated from:

$$[(\Delta\boldsymbol{R}_i)^2]_k = (k_{\mathrm{B}}T/\gamma)\lambda_k^{-1}\ \mathrm{tr}[(\boldsymbol{u}_k^{\mathrm{ANM}})(\boldsymbol{u}_k^{\mathrm{ANM}})^{\mathrm{T}}]_{ii} \quad (14)$$

For more details, the reader is referred to Atilgan *et al.*[39]

Conformational changes induced by global modes and reconstruction of all atom models

In as much as the fluctuations are symmetric with respect to the equilibrium coordinates $\{\boldsymbol{R}_1^0, \boldsymbol{R}_2^0, \ldots, \boldsymbol{R}_N^0\}$, two sets of deformed conformations are obtained by varying the size of the fluctuations driven by the global modes:

$$\boldsymbol{R}_i(\pm s) = \boldsymbol{R}_i^0 \pm s\Delta\boldsymbol{R}_i^{(1)} = \boldsymbol{R}_i^0 \pm s\lambda_1^{-1/2}[u_1^{\mathrm{ANM}}]_i \quad (15)$$

Here s is a parameter that scales the size of the deformation induced by mode 1, and $[\boldsymbol{u}_1^{\mathrm{ANM}}]_i$ is the ith super element (3D-vector) of the eigenvector $\boldsymbol{u}_1^{\mathrm{ANM}}$. Theoretically s is equal to $(k_{\mathrm{B}}T/\gamma)^{1/2}$; however, in the absence of quantitative information on γ, it is left as a variable to match the changes observed in the different crystal forms. Reconstruction of all atoms model was done by applying equation (15) to each atom, i.e. we add/subtract the fluctuation vector $s\Delta\boldsymbol{R}_i^{(1)}$ for calculating the positions of all atoms belonging to amino acid i. The obtained deformed structure was then minimized using the program mini_pwl of the MOIL package[79] with a tolerance of 0.2 kcal/mol.

Utility of combining GNM and ANM

GNM offers the advantage of doing calculations in an N-dimensional space. Matrix diagonalization is the most time-consuming part of the computations. It scales with the third power of the size of the matrix, approximately. Thus GNM calculations are 2–3 orders of magnitude faster than ANM. Secondly, ANM assumes that the displacements in the x, y and z-directions are all controlled by spring forces, whereas GNM assumes that the magnitudes of the displacements, only, are subject to elastic forces. The fluctuation amplitudes derived by the GNM are in this respect more realistic, as they are devoid of any constraints imposed on the components of the fluctuation vector. Interestingly, even the highest frequency modes of GNM have been shown to be physically meaningful (as opposed to "noise" in NMA), which could be related to folding nuclei and conserved sites.[33] Finally, a third advantage of the GNM is the possibility of interpreting the results in terms of statistical thermodynamics concepts. The H/^{2}H exchange free energies (and protection factors) under mildly denaturing conditions are well-accounted for by the entropic contribution to the free energy calculated with the GNM.[80] Likewise the present analysis reveals that the entropic forces play a significant role in driving the passage between the T and R2 forms.

On the other hand, GNM cannot predict the directions of the motions, or the components of the fluctuation vectors $\Delta\boldsymbol{R}_i(\mathbf{1} \le i \le N)$. These are found by the ANM. In principle, it is possible to "map" the GNM modes to those found by ANM by comparing the square fluctuations $[\Delta\boldsymbol{R}_i \cdot \Delta\boldsymbol{R}_i]_k$, predicted for mode k' by the GNM (equation (10)) with the product $([\Delta\boldsymbol{R}_i]^{(k)} \cdot [\Delta\boldsymbol{R}_i]^{(k)})$ found from ANM (see equation (14)) for mode k for all (k, k') pairs. The pair that yields the minimal difference between the two profiles $(1 \le i \le N)$ corresponds to the equivalent modes. Here, we focused on the slowest GNM mode, as the most cooperative mechanism of motion, also confirmed for many systems to be physically meaningful and robust. We identified the corresponding ANM directional vectors by examining the 20 slowest mode profiles found by ANM. This analysis yielded the second slowest ANM mode for the T form (1a3n), and the first for the R2 form (1bbb).

Correlation between ANM slow modes and the displacement vectors between the T and R2 forms

In principle, the deformation undergone by individual residues during the passage from the T to the R2 state does not necessarily conform to the displacement vectors obtained by a linear interpolation between the two states. Nevertheless, it might be of interest to examine to what extent these displacement vectors correlate with those induced by the individual modes. For a comparative analysis, we examined the correlation cosine between these displacement vectors and the eigenvectors associated with the slowest 20 modes obtained by ANM for both the T and R2 forms. Figure 9 displays the results from this correlation. Consistent with the previous

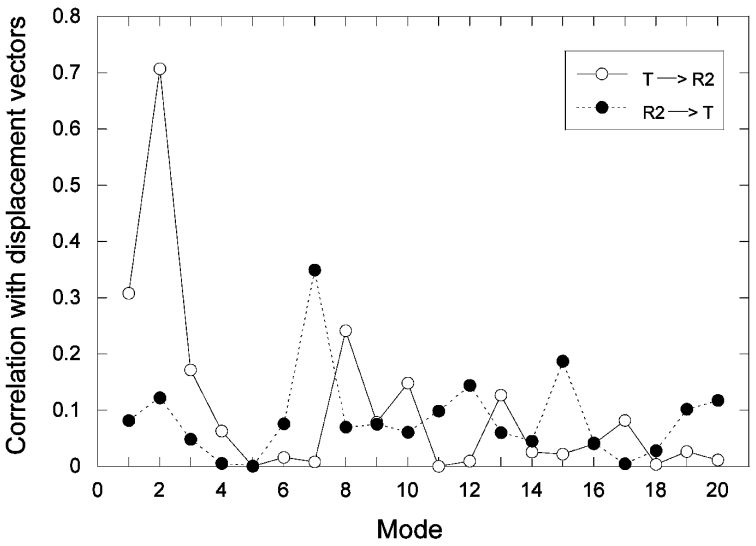

Figure 9. Correlation between the displacement vectors associated with the $T \rightleftarrows R2$ transition (deduced from the comparison of the residue coordinates in the two PDB structures) and the slowest ANM modes obtained for the T state (open circles) and R2 state (filled circles). The ordinate is the correlation cosine between the displacement vectors and the eigenvectors corresponding to the 20 slowest modes.

analysis, the second slowest ANM mode obtained for the T state shows a high correlation (>0.7) with the actual displacement vectors. The eigenvectors obtained for the R2 state, on the other hand, show in general a weaker correlation with the displacement vectors, except for the seventh mode, the contribution of which is not as strong as the leading few modes, confirming that the passage R2 → T is not as favorable as the T → R2 passage. Finally, we note that the passage T → R2 seems to be favored to some extent by modes 2 and 8.

Acknowledgements

We thank Drs Ho, Lukin and Thirumalai for insightful comments during the course of this work. C.X. is grateful to N. Alpay Temiz, Basak Isin and LeeWei Yang for sharing their expertise in GNM and ANM methods.

References

1. Dickerson, R. E. & Geis, I. (1983). *Hemoglobin: Structure, Function, Evolution, and Pathology*, Benjamin/Cummings, Menlo Park, CA.
2. Monod, J., Wyman, J. & Changeux, J.-P. (1965). On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12**, 88–118.
3. Koshland, D. E., Jr, Nemethy, G. & Filmer, D. (1966). Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, **5**, 365–385.
4. Perutz, M. F. (1970). Stereochemistry of cooperative effects in haemoglobin. *Nature*, **228**, 726–739.
5. Eaton, W. A., Henry, E. R., Hofrichter, J. & Mozzarelli, A. (1999). Is cooperative oxygen binding by hemoglobin really understood? *Nature Struct. Biol.* **6**, 351–358.
6. Eigen, M. (1976). Kinetics of reaction control and information transfer in enzymes and nucleic acids. In *Fast React. Primary Processes Chem. Kinet.* In Proceedings of Nobel Symposium, Fifth (Claesson, S., ed.), pp. 333–369, Almqvist and Wiksell, Stockholm, Sweden.
7. Fersht, A. (1999). Conformational change, allosteric regulation, motors and work. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W.H. Freeman, New York pp. 289–323..
8. Yuan, Y., Simplaceanu, V., Lukin, J. A. & Ho, C. (2002). NMR investigation of the dynamics of tryptophan side-chains in hemoglobins. *J. Mol. Biol.* **321**, 863–878.
9. Ackers, G. K., Dalessio, P. M., Lew, G. H., Daugherty, M. A. & Holt, J. M. (2002). Single residue modification of only one dimer within the hemoglobin tetramer reveals autonomous dimer function. *Proc. Natl Acad. Sci. USA*, **99**, 9777–9782.
10. Chang, C. K., Simplaceanu, V. & Ho, C. (2002). Effects of amino acid substitutions at beta 131 on the structure and properties of hemoglobin: evidence for communication between alpha 1 beta 1- and alpha 1 beta 2-subunit interfaces. *Biochemistry*, **41**, 5644–5655.
11. Silva, M. M., Rogers, P. H. & Arnone, A. (1992). A third quaternary structure of human hemoglobin A at 1.7 A resolution. *J. Biol. Chem.* **267**, 17248–17256.
12. Srinivasan, R. & Rose, G. D. (1994). The T-to-R transformation in hemoglobin: a reevaluation. *Proc. Natl Acad. Sci. USA*, **91**, 11113–11117.
13. Zoete, V., Michielin, O. & Karplus, M. (2002). Relation between sequence and structure of HIV-1 protease inhibitor complexes: a model system for the analysis of protein flexibility. *J. Mol. Biol.* **315**, 21–52.
14. Mouawad, L., Perahia, D., Robert, C. H. & Guilbert, C. (2002). New insights into the allosteric mechanism of human hemoglobin from molecular dynamics simulations. *Biophys. J.* **82**, 3224–3245.
15. Elber, R. & Karplus, M. (1987). Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. *Science*, **235**, 318–321.
16. Guilbert, C., Perahia, D. & Mouawad, L. (1995). A method to explore transition paths in macromolecules. Applications to hemoglobin and phosphoglycerate kinase. *Comparative Phys. Commun.* **91**, 263–273.
17. Zaloj, V. & Elber, R. (2000). Parallel computations of molecular dynamics trajectories using the stochastic

path approach. *Computer Phys. Commun.* **128**, 118–127.
18. Berendsen, H. J. & Hayward, S. (2000). Collective protein dynamics in relation to function. *Curr. Opin. Struct. Biol.* **10**, 165–169.
19. Abseher, R. & Nilges, M. (2000). Efficient sampling in collective coordinate space. *Proteins: Struct. Funct. Genet.* **39**, 82–88.
20. Goodman, J. L., Pagel, M. D. & Stone, M. J. (2000). Relationships between protein structure and dynamics from a database of NMR-derived backbone order parameters. *J. Mol. Biol.* **295**, 963–978.
21. Wand, A. J. (2001). Dynamic activation of protein function: a view emerging from NMR spectroscopy. *Nature Struct. Biol.* **8**, 926–930.
22. Tolman, J. R., Al Hashimi, H. M., Kay, L. E. & Prestegard, J. H. (2001). Structural and dynamic analysis of residual dipolar coupling data for proteins. *J. Am. Chem. Soc.* **123**, 1416–1424.
23. Frauenfelder, H. & McMahon, B. (1998). Dynamics and function of proteins: the search for general concepts. *Proc. Natl Acad. Sci. USA*, **95**, 4795–4797.
24. Stock, A. (1999). Relating dynamics to function. *Nature*, **400**, 221–222.
25. Ma, B., Wolfson, H. J. & Nussinov, R. (2001). Protein functional epitopes: hot spots, dynamics and combinatorial libraries. *Curr. Opin. Struct. Biol.* **11**, 364–369.
26. Brooks, B. & Karplus, M. (1983). Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl Acad. Sci. USA*, **80**, 6571–6575.
27. Levitt, M., Sander, C. & Stern, P. S. (1985). Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* **181**, 423–447.
28. Gibrat, J. F. & Go, N. (1990). Normal mode analysis of human lysozyme: study of the relative motion of the two domains and characterization of the harmonic motion. *Proteins: Struct. Funct. Genet.* **8**, 258–279.
29. Amadei, A., Linssen, A. B. & Berendsen, H. J. (1993). Essential dynamics of proteins. *Proteins: Struct. Funct. Genet.* **17**, 412–425.
30. Krebs, W. G., Alexandrov, V., Wilson, C. A., Echols, N., Yu, H. & Gerstein, M. (2002). Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. *Proteins: Struct. Funct. Genet.* **48**, 682–695.
31. Romo, T. D., Clarage, J. B., Sorensen, D. C. & Phillips, G. N., Jr (1995). Automatic identification of discrete substates in proteins: singular value decomposition analysis of time-averaged crystallographic refinements. *Proteins: Struct. Funct. Genet.* **22**, 311–321.
32. Bahar, I., Erman, B., Haliloglu, T. & Jernigan, R. L. (1997). Efficient characterization of collective motions and interresidue correlations in proteins by low-resolution simulations. *Biochemistry*, **36**, 13512–13523.
33. Bahar, I., Atilgan, A. R. & Erman, B. (1997). Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* **2**, 173–181.
34. Haliloglu, T., Bahar, I. & Erman, B. (1997). Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* **79**, 3090–3093.
35. Bahar, I., Atilgan, A. R., Demirel, M. C. & Erman, B. (1998). Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. *Phys. Rev. Lett.* **80**, 2733–2736.
36. Demirel, M. C., Atilgan, A. R., Jernigan, R. L., Erman, B. & Bahar, I. (1998). Identification of kinetically hot residues in proteins. *Protein Sci.* **7**, 2522–2532.
37. Keskin, O., Jernigan, R. L. & Bahar, I. (2000). Proteins with similar architecture exhibit similar large-scale dynamic behavior. *Biophys. J.* **78**, 2093–2106.
38. Doruker, P., Atilgan, A. R. & Bahar, I. (2000). Dynamics of proteins predicted by molecular dynamics simulations and analytical methods. Application to α-amylase inhibitor. *Proteins: Struct. Funct. Genet.* **40**, 512–524.
39. Atilgan, A. R., Durrell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O. & Bahar, I. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **80**, 505–515.
40. Isin, B., Doruker, P. & Bahar, I. (2002). Functional motions of influenza virus hemagglutinin: a structure-based analytical approach. *Biophys. J.* **82**, 569–581.
41. Keskin, O., Bahar, I., Flatow, D., Covell, D. G. & Jernigan, R. L. (2002). Molecular mechanisms of chaperonin GroEL–GroES function. *Biochemistry*, **41**, 491–501.
42. Flory, P. J. (1976). Statistical thermodynamics of random networks. *Proc. Roy. Soc. ser. A*, **351**, 351–380.
43. Tirion, M. M. (1996). Large amplitude elastic motions in proteins from a single-parameter atomic analysis. *Phys. Rev. Lett.* **77**, 1905–1908.
44. Doruker, P., Jernigan, R. L. & Bahar, I. (2002). Dynamics of large proteins through hierarchical levels of coarse-grained structures. *J. Computer Chem.* **23**, 119–127.
45. Isin, B., Doruker, P. & Bahar, I. (2002). Functional motions of influenza virus hemagglutinin: a structure-based analytical approach. *Biophys. J.* **82**, 569–581.
46. Hinsen, K. (1998). Analysis of domain motions by approximate normal mode calculations. *Proteins: Struct. Funct. Genet.* **33**, 417–429.
47. Kitao, A. & Go, N. (1999). Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struct. Biol.* **9**, 164–169.
48. Sinha, N. & Nussinov, R. (2001). Point mutations and sequence variability in proteins: redistributions of preexisting populations. *Proc. Natl Acad. Sci. USA*, **98**, 3139–3144.
49. Tama, F., Gadea, F. X., Marques, O. & Sanejouand, Y. H. (2000). Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins: Struct. Funct. Genet.* **41**, 1–7.
50. Flory, P. J. (1976). Statistical thermadynamics of random networks. *Proc. Roy. Soc. London*, **351**, 351–380.
51. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucl. Acids Res.* **28**, 235–242.
52. Shaanan, B. (1983). Structure of human oxyhaemoglobin at 2.1 Å resolution. *J. Mol. Biol.* **171**, 31–59.
53. Tame, J. R. & Vallone, B. (2000). The structures of deoxy human haemoglobin and the mutant Hb Tyralpha42His at 120 K. *Acta Crystallog. D. Biol. Crystallog.* **56 (Pt 7)**, 805–811.
54. Miyazawa, S. & Jernigan, R. L. (1996). Residue–residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* **256**, 623–644.
55. Kundu, S., Melton, J. S., Sorensen, D. C. & Phillips, G. N., Jr (2002). Dynamics of proteins in crystals: comparison of experiment with simple models. *Biophys. J.* **83**, 723–732.

56. Yuan, Y., Simplaceanu, V., Lukin, J. A. & Ho, C. (2002). NMR investigation of the dynamics of tryptophan side-chains in hemoglobins. *J. Mol. Biol.* **321**, 863–878.
57. Barrick, D., Ho, N. T., Simplaceanu, V., Dahlquist, F. W. & Ho, C. (1997). A test of the role of the proximal histidines in the Perutz model for cooperativity in haemoglobin. *Nature Struct. Biol.* **4**, 78–83.
58. Barrick, D., Ho, N. T., Simplaceanu, V. & Ho, C. (2001). Distal ligand reactivity and quaternary structure studies of proximally detached hemoglobins. *Biochemistry*, **40**, 3780–3795.
59. Kim, H. W., Shen, T. J., Sun, D. P., Ho, N. T., Madrid, M. & Ho, C. (1995). A novel low oxygen affinity recombinant hemoglobin (alpha96val → Trp): switching quaternary structure without changing the ligation state. *J. Mol. Biol.* **248**, 867–882.
60. Kim, H. W., Shen, T. J., Sun, D. P., Ho, N. T., Madrid, M., Tam, M. F. *et al.* (1994). Restoring allosterism with compensatory mutations in hemoglobin. *Proc. Natl Acad. Sci. USA*, **91**, 11547–11551.
61. Rivetti, C., Mozzarelli, A., Rossi, G. L., Henry, E. R. & Eaton, W. A. (1993). Oxygen binding by single crystals of hemoglobin. *Biochemistry*, **32**, 2888–2906.
62. Bettati, S., Mozzarelli, A., Rossi, G. L., Tsuneshige, A., Yonetani, T., Eaton, W. A. & Henry, E. R. (1996). Oxygen binding by single crystals of hemoglobin: the problem of cooperativity and inequivalence of alpha and beta subunits. *Proteins: Struct. Funct. Genet.* **25**, 425–437.
63. Inaba, K., Ishimori, K., Imai, K. & Morishima, I. (2000). Substitution of the heme binding module in hemoglobin alpha- and beta-subunits. Implication for different regulation mechanisms of the heme proximal structure between hemoglobin and myoglobin. *J. Biol. Chem.* **275**, 12438–12445.
64. Elber, R., Roitberg, A., Simmerling, C., Goldstein, R., Li, H., Verkhiver, G. *et al.* (1994). MOIL: a program for simulation of macromolecules. *Comparative Phys. Commun.* **91**, 159–189.
65. Mouawad, L. & Perahia, D. (1996). Motions in hemoglobin studied by normal mode analysis and energy minimization: evidence for the existence of tertiary T-like, quaternary R-like intermediate structures. *J. Mol. Biol.* **258**, 393–410.
66. Eaton, W. A., Henry, E. R. & Hofrichter, J. (1991). Application of linear free energy relations to protein conformational changes: the quaternary structural change of hemoglobin. *Proc. Natl Acad. Sci. USA*, **88**, 4472–4475.
67. Janin, J. & Wodak, S. J. (1985). Reaction pathway for the quaternary structure change in hemoglobin. *Biopolymers*, **24**, 509–526.
68. Gibson, Q. H. (1959). The photochemical formation of a quickly reacting form of hemoglobin. *Biochem. J.* **71**, 293–303.
69. Hopfield, J. J., Shulman, R. G. & Ogawa, S. (1971). An allosteric model of hemoglobin. I. Kinetics. *J. Mol. Biol.* **61**, 425–443.
70. Hofrichter, J., Sommer, J. H., Henry, E. R. & Eaton, W. A. (1983). Nanosecond absorption spectroscopy of hemoglobin, elementary processes in kinetic cooperativity. *Proc. Natl Acad. Sci. USA*, **80**, 2235–2239.
71. Eaton, W. A., Henry, E. R., Hofrichter, J. & Mozzarelli, A. (1999). Is cooperative oxygen binding by hemoglobin really understood? *Nature Struct. Biol.* **6**, 351–358.
72. Thomas, A., Field, M. J., Mouawad, L. & Perahia, D. (1996). Analysis of the low frequency normal modes of the T-state of aspartate transcarbamylase. *J. Mol. Biol.* **257**, 1070–1087.
73. Perahia, D. & Mouawad, L. (1995). Computation of low-frequency normal modes in macromolecules: improvements to the method of diagonalization in a mixed basis and application to hemoglobin. *Computer Chem.* **19**, 241–246.
74. Goldbeck, R. A., Esquerra, R. M. & Kliger, D. S. (2002). Hydrogen bonding to Trp beta37 is the first step in a compound pathway for hemoglobin allostery. *J. Am. Chem. Soc.* **124**, 7646–7647.
75. Ranson, N. A., Farr, G. W., Roseman, A. M., Gowen, B., Fenton, W. A., Horwich, A. L. & Saibil, H. R. (2001). ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell*, **107**, 869–879.
76. Xu, Z., Horwich, A. L. & Sigler, P. B. (1997). The crystal structure of the asymmetric GroEL–GroES–(ADP)7 chaperonin complex. *Nature*, **388**, 741–750.
77. Eichinger, B. E. (1972). Elasticity theory. I. Distribution functions for perfect phantom networks. *Macromolecules*, **5**, 496–505.
78. Kloczkowski, A., Mark, J. E. & Erman, B. (1989). Chain dimensions and fluctuations in random elastomeric networks. 1. Phantom Gaussian networks in the undeformed state. *Macromolecules*, **22**, 1423–1432.
79. Elber, R., Roitberg, A., Simmerling, C., Goldstein, R., Li, H., Verkhiver, G. *et al.* (1994). MOIL: a program for simulation of macromolecules. *Comparative Phys. Commun.* **91**, 159–189.
80. Bahar, I., Wallqvist, A., Covell, D. G. & Jernigan, R. L. (1998). Correlation between native-state hydrogen exchange and cooperative residue fluctuations from a simple model. *Biochemistry*, **37**, 1067–1075.

Edited by M. Levitt

(Received 29 April 2003; received in revised form 29 July 2003; accepted 11 August 2003)