Perturbation-response scanning: Ligand entry-exit mechanisms of ferric binding protein*

Canan Atilgan and Ali Rana Atilgan

Faculty of Engineering and Natural Sciences, Sabanci University, 34956 Istanbul, Turkey

e-mail: canan@sabanciuniv.edu
telephone: +90 (216) 4839523
telefax: +90 (216) 4839550

Keywords: stability analysis, ligand binding models, allostery, linear response theory, elastic network models, hinge-bending motions.

*All computer programs used in the analyses are available upon request.

ABSTRACT

A new tool, termed perturbation-response scanning (PRS), for the analysis of remote control strategies utilized by proteins is introduced, and its predictive power is shown on the ligand release mechanisms of haemophilus influenzae Ferric binding protein (FBP). The approach relies on systematic use of computational perturbation/response techniques based on linear response theory. By sequentially exerting directed forces on single-residues along the chain and recording the resulting relative changes in the residue coordinates, we find that for a large number of the cases, residue-by-residue displacements as determined from the X-ray structures are faithfully reproduced for the apo FBP. Once a stabilizing ligand is integrated to the system in the holo form, only a few residues remain that are particularly successful in reproducing the experimental displacements. It is possible to manipulate the bound form of the protein towards the unbound form by either directly perturbing the Fe binding residues, or by controlling the distant loop residues that show large displacements upon binding. The latter are charged residues, providing binding locations for ions which are known to influence Fe^{+3} release kinetics. By perturbing any one of these residues, the tip of the cap that opens the exit of the Fe^{+3} is made to operate coherently, irrespective of the direction of the perturbation. The analysis, based on the structural features of the protein, provides alternative strategies for ferric iron release.

INTRODUCTION

Functional proteins are complex structures, which may remain mainly unmodified as a result of a multitude of mutations,[1] yet may have their energy surface go through significant changes upon perturbing highly specific regions.[2-4] The various accessible states populated may be manipulated by inducing short and long-range conformational changes in the structure,[5] alternatively, a dynamical control may take place without any significant structural variation.[6] To explore the presence or the absence of such "shifts in the energy landscapes," one needs to perturb the protein structure, and observe the response.[7] The perturbation may be in the form of changing the environmental factors such as ionic concentration,[8] or may target specific locations on the structure itself, either through chemically modifying the residues (inserting mutations)[9] or by inducing site-specific perturbations as is done in single molecule experiments.[10] The response may be measured directly, as a change in the overall conformation of the protein,[11] or indirectly, e.g., through determining the kinetic parameters, and proposing kinetic models that explain the observations.[12,13] The purpose in such work is to understand and therefore control the response of the protein for a plethora of reasons, including, but not limited to, the design of efficient drugs,[14,15] or to tailor enzymes serving as "materials."[16] In this study, we analyze Ferric binding protein A (FBP), using computational perturbation/response techniques,[7,17,18] and describe alternative approaches that may have evolved to modify its energy surface and control its interactions. The perturbation-response scanning (PRS) technique developed here is generalizable to the study of the structural changes undergone by protein molecules.

FBP is involved in the shuttling of Fe^{+3} from the mammalian host to the cytoplasm of pathogenic bacteria. To make iron unavailable to such pathogens, host organisms have iron transport systems such as transferrin that tightly sequester the ion. Pathogens have developed strategies to circumvent this approach, one of them being the development of receptors for the iron transport proteins of the host. FBP resides in the periplasm, and receives iron from these receptors to eventually deliver it to the cytosol[19] via a sequence of events that are thought to involve the reduction of the Fe^{3+} to Fe^{2+} and release of a synergistic phosphate anion necessary for tight binding[20,21] as well as the loading of the Fe^{+3} into the FBP binding site.[22] The protein is made up of two domains characteristic of periplasmic

Fe^{+3} binding family and transferrin, the latter being a blood plasma protein for iron ion delivery. These host/pathogen iron uptake proteins are thought to be distantly related through divergent evolution from an anion binding function.

Fe^{+3} is bound to FBP with remarkable affinity, with association constants on the order of 10^{17} – 10^{22} M^{-1} depending on the measurement conditions.[23] It was recently shown that a relatively high affinity of iron binding is required for the removal of iron from transferrin, and its transport across the periplasm.[24] Yet, this high affinity poses a Fe^{+3} transport dilemma, suggesting another necessary step for the release of the ion. It is of interest to understand how Fe^{+3} is eventually released from the binding site for subsequent use by the pathogen. One mode of action that has been suggested involves the control of the Fe^{+3} release kinetics by the exchange of synergistic anions forming relatively stable intermediates.[23,25,26] Another involves the direct action of chelators on the ion.[23] It has also been shown that mutants of FBP that are defective in binding the synergistic anion are still capable of donating iron, suggesting the possibility of still other alternative mechanisms for the process.[27]

FBP is referred to as bacterial transferrin due to the similarities with transferrin in the structural folds, the highly conserved set of iron-coordinating residues, and their usage of a synergistic anion.[21] They do, however, differ in size, transferrin being made up of two-lobes having high sequence identity with each other (e.g. 45 % in human transferrin). Each lobe itself is comparable to FBP in size, fold, and iron binding location. In transition from the open to the closed form, only one of the sub-domains in each lobe undergoes significant reorientation, similar to FBP.[28] Despite the resemblance, the iron binding/release kinetics in the two lobes differ. The kinetics in the N- [29] and C-lobes[30,31] have been studied separately, and their cooperativity has been discussed.[32,33] In these studies it has been implicated that there may be several approaches used for iron release in these proteins, including chelating agents and synergistic anions acting directly around the ferric binding site.[30,34-36] Additionally, it has been shown that chloride and other ion concentrations are effective on the kinetics, and it has been proposed that allosteric anion binding sites that trigger large conformational changes exist.[8,31,37] Based on the similarities between FBP and transferrin, it is of interest to find out if these routes also exist for FBP, and if they do, what the details of the mechanisms are. It is also of

significance to determine possible binding locations on the surface as well as to understand the physical origin of such control.

In this study, we perform a stability analysis on FBP in its apo- and holo-forms, and show that the apo form is structurally amenable to switch to the holo form, whereas the holo form will change to the apo conformation under much restricted conditions. These findings are in agreement with a model where iron uptake by the protein is a thermodynamically and kinetically favored process, and iron release is managed through several mechanisms including chelation and allosteric control. Several locations for allosteric control of iron release are proposed. The overall response of residues to perturbations is quantified using both the size and the direction of the response.

The manuscript is organized as follows: In the **Methods** section, we first describe the linear response theory underlying the PRS technique, and then go on to describe PRS in detail. We further describe an algorithm by which the directionality of the response may be solicited to get further insight into the mechanisms. In the **Results and Discussion** section, we first outline the generalized results on the apo- and holo-FBP, as well as a hypothetical "holo-like" structure which has the holo conformation, but is stripped of the Fe ion. We finally present evidence for alternative mechanisms by which Fe ion dissociation may be manipulated locally and remotely in the holo FBP. Implications for cooperativity in proteins are discussed under **Conclusions**.

METHODS

The new tool introduced in this work for the analysis of remote control strategies utilized by proteins is based on inserting forces as perturbations on a given residue, and recording the displacements of all the residues as the response. Since the procedure is repeated sequentially for all the residues in the protein, we term the technique, perturbation-response scanning. Below, we first review the theory on which the linear response nature of the relationship between the forces and the displacements is based. We then outline the computational details of the PRS technique. Finally, we describe an algorithm that assesses the directional preferences of the responses.

Theory. Here we summarize the theory used in this work; a generalized derivation of how the structure may be manipulated by external forces, as well as how the architecture intrinsically controls the response appears in literature.[18]

We perceive the protein as a residue network of N nodes centered on the C_α atoms. Any given pair of nodes are assumed to interact via a harmonic potential, if they are within a cut-off distance r_c of each other (Figure 1).

[Insert Figure 1 here]

In the notation used, **r** and **f** refer to the distance and force vectors along the interaction direction (local frame), respectively, whereas **R** and **F** are those on the nodes, expressed with respect to the absolute frame. There are m interactions pertaining to each residue (m = 6 in Figure 1), and a total of M interactions in the system of N residues. In the absence of an external force on the system, each residue, i, is in equilibrium such that

$$\mathrm{b}\Delta\mathrm{f}_i=0 \qquad (1)$$

where the $3\times m$ coefficient matrix **b** consists of the direction cosines of each interaction force, its row indices being x, y, or z. $\Delta\mathbf{f}_i$ is an $m\times 1$ column matrix of forces in the local frame. Generalizing equation 1 to the whole system of N nodes and a total of M interactions, the forces along the line of interactions may be projected to the absolute frame by

$$\mathrm{B}\Delta\mathrm{f}=0 \qquad (2)$$

with the $3N\times M$ direction cosine matrix **B** and the $M\times 1$ column matrix of forces. It is straightforward to generate the matrix **B** from the topology of the native structure [i.e. the protein data bank (PDB) file [38]] with a specified r_c.

In the presence of an external force, $\Delta\mathbf{F}$, acting on the system (figure 1), the new equilibrium condition dictates,

$$\mathbf{B}_{3N\times M}\Delta\mathbf{f}_{M\times 1} = \Delta\mathbf{F}_{3N\times 1} \qquad (3)$$

It is possible to express the M distortions of the bonds connecting a pair of residues in the local frame,

$\Delta\mathbf{r}$, in terms of the positional displacements, $\Delta\mathbf{R}$, of the residues in the absolute frame using the projection:[18]

$$\mathrm{B}^{T}_{M\times 3N}\Delta\mathrm{R}_{3N\times 1} = \Delta\mathrm{r}_{M\times 1} \tag{4}$$

Within the scope of an elastic network of residues connected to their neighbors with springs, the forces are related to the distortions in the local frame through Hooke's law by

$$\mathrm{K}_{M\times M}\Delta\mathrm{r}_{M\times 1} = \Delta\mathrm{f}_{M\times 1} \tag{5}$$

where the coefficient matrix **K** is diagonal. Inasmuch as the native structures are stabilized predominantly by homogeneous tertiary contacts rather than specific interactions[39] we take the entries of **K** to be equivalent in this work. Note that we also check this assumption by comparing sample calculations using the current methodology with those from the Hessian obtained from an all atom approach, and find that the level of accuracy in both methods are equivalent (see **Results and Discussion** for details).

Thus, rearranging equations 3-5, one gets the forces necessary to induce a given displacement profile:

$$\left(\mathrm{BKB}^{\mathrm{T}}\right)\Delta\mathrm{R}=\Delta\mathrm{F} \tag{6}$$

On the other hand, one may choose to perturb to a single or a set of residues, and follow the response of the residue network through,

$$\left(\mathrm{BKB}^{\mathrm{T}}\right)^{-1}\Delta\mathrm{F}=\Delta\mathrm{R} \tag{7}$$

where the $\Delta\mathbf{F}$ vector will contain the components of the externally inserted force vectors on the selected residues. The ($\mathbf{BKB^T}$) matrix is equivalent to the Hessian[40] and its inverse has six zero eigenvalues, corresponding to the translational and rotational degrees of freedom of the system. Note that this formulation is equivalent to the development in reference 17.

Perturbation-response scanning. Our detailed PRS analysis is based on a systematic application of equation 7. We apply a force on the C_α atom of each residue by forming the $\Delta\mathbf{F}$ vector in such a way that all the entries, except those corresponding to the residue being perturbed, are equal to zero. For a selected residue *i*, the random force q_i is (q_x, q_y, q_z) so that the external force vector is constructed as

$$(\Delta \mathrm{F})^{\mathrm{T}} = \{0\ 0\ 0 \ldots q_x\ \ q_y\ q_z \ldots 0\ 0\ 0\}_{1\times 3N} \quad (8)$$

We then compute the resulting $(\Delta \mathbf{R})^i$ vector of the protein through equation 7. We scan the protein by consecutively perturbing each residue, and record the expected displacements as a result of the linear response of the protein. In all the calculations based on equation 7, we report the averages over five independent runs where the forces are applied randomly, with the magnitudes chosen uniformly in the interval [-1, 1] (arbitrary units due to the linear response nature of the analysis). Note that although the displacements are recorded as vectors, we average over the magnitude of the response; the directionality of the response will be further considered in the next subsection.

Thus, the predictions of the average displacement of each residue, *j*, as a response of the system to inserted forces on residue *i*, $(\Delta R_j)^i$ are compared with the experimental conformational changes between the original and the target PDB structures (e.g. the apo and the holo forms), ΔS_j. The goodness of the prediction is quantified as the Pearson correlation coefficient for each perturbed residue, *i*, as follows:

$$C_i = \frac{\sum_{j=1}^{N}\left[(\Delta R_j)^i - (\overline{\Delta R})^i\right](\Delta S_j - \overline{\Delta S})}{(N-1)\sigma_R \sigma_S} \quad (9)$$

where the overbar indicates the average. σ_R and σ_s are the respective standard deviations of calculated and experimental displacements. A pseudo-algoritm describing the PRS is given below.

[Insert Algorithm A here]

Analysis on the directionality of the response. The PRS technique produces both the magnitude and the directionality of the response. However, the averaging over several independent runs focuses on capturing the relative magnitudes of the displacements and disregards the directionality information. For a set of perturbations whose vectorial sum is zero, the response on every residue is also expected to have a zero vectorial sum due to the linear response nature of the theory (equation 7). Note that if the collection of forces applied on a specific residue is independent and large in number, they will appear in a spherically symmetric set of directions. The responses, however, need not be recovered in the same manner; e.g. they may be distributed along a line or in a plane so that the net response is still

zero. Deviations from such a spherically symmetric distribution of responses hint at the roles of certain residues in the remote control of the ligand, as will be shown in the final part of the **Results and Discussion** section.

For an analysis that probes the directionality of the recorded responses, we proceed as follows: We first concentrate on those residues for which the Pearson correlation between experiment and theory is large (i.e. $C_i > 0.9$). Amongst them, we further locate those residues, *i*, that are distant from the ligand binding site, *l*, (i.e. the distance between them, $d_{il} >> r_c$). For the selected residue, *i*, *k* forces are inserted such that $\Sigma_k \Delta\mathbf{F}_k^i = 0$; *k* is large and ensures that a spherically symmetric region around *i* is covered (in the current work we take $k = 50$). The sum of the responses on each residue *j* is zero, $\Sigma_k \Delta\mathbf{R}_k^j = 0$. The results are visualized as vector plots on the protein structure (as will be exemplified in figure 5). A pseudo-algoritm describing the analysis on the directionality of response to selected perturbations is given below.

[Insert Algorithm B here]

Systems studied and residue network construction. We analyze the haemophilus influenzae ferric-binding protein in apo and holo forms (PDB codes: 1D9V and 1MRP, respectively). We perturb the protein in three different initial conformations: (i) the apo form; (ii) the holo form with only the protein coordinates (the Fe ion stripped from the data); (iii) the holo form where the Fe is treated as an additional node of the network.

The protein has two domains, and upon binding one moves relative to the other as shown in figure 2. Unless otherwise specified, the final structures are superimposed on the fixed domain of the initial structure (residues 83 – 87, 102 – 225, 277 – 307), before the displacements are computed. Thus, we ensure that the response that is reflected on the collective motion of the fixed domain is removed. Note, however, that this is done only to achieve a quantitative comparison with the all-atom study in reference 17 and the overall conclusions of the study are not affected by this choice. When no superimposition is applied, the values of the correlations, C_i (equation 9) are lower in value, but their relative ordering does not change.

[Insert Figure 2 here]

For all the structures, we use a cut-off distance of r_c = 8.0 Å on the C_α atoms of the protein (equation 1), and the system exhibits six zero eigenvalues. We have also verified that the main conclusions of the study are not affected for a range of values between 7.0 – 8.5 Å.[41] At the cut-off distance of 8.0 Å, there are 1542, 1576, and 1587 interactions for initial conformations (i) – (iii), respectively. Thus, the holo form of the protein is a tighter structure, with the presence of Fe introducing few additional interactions with the protein.

RESULTS and DISCUSSION

Based on linear response theory, we apply perturbations at selected points along the chain to diagnose the dynamic response of the haemophilus influenzae FBP. The protein is known to have a Fe^{3+} binding location, and the structure of the holo form is also known. The overall RMSD between the two forms is 2.48 Å. FBP has two domains, termed as the fixed and moving domains, respectively (in figure 2 the two structures are superimposed on the fixed domain residues: yellow apo-, purple holo-FBP). The difference in the contact numbers of the two forms is shown in the bottom panel of figure 2. The main motion in the protein is hinge, with the moving domain residues closing up on the Fe binding location and the total number of contacts in the protein increasing from 1542 to 1576. For 111 residues, the contact number remains unchanged; others lose/gain up to three contacts, with the exception of residue 293 having four additional contacts (figure 2). Thus, the two structures have distinct features; yet, the conformational change is mainly governed by a hinge motion around the Fe-binding location. It is therefore of interest to see the extent to which the two conformations are inter-convertible, so as to describe the mechanisms governing the high affinity iron binding and its selective release.

The force input profile is not localized around the ferric binding region. It was previously shown that insertion of forces on residue 57 near the ferric binding site (Fe ion is shown in red in figure 2), yields the expected atomic displacements with an all atom method using the Hessian obtained from MD simulations.[17] Therein, the correlation coefficient between the theoretical and experimental

relative displacements, computed from the displacement vectors between the holo and apo forms, was reported to be 0.95. Using equation 7 with the residue-based coarse graining (see the Methods for details), we get a correlation coefficient of 0.96 as an average over five independent perturbations. The relative magnitudes of the residue-by-residue displacement vectors between the experimental apo – holo structures after superimposing their fixed domains is shown in the bottom curve of figure 3a. Example cases of a single perturbation on residue 57 placed in a randomly chosen direction on the apo and holo structures are shown respectively in the middle and upper curves of the same figure.

[Insert Figure 3 here]

The close agreement of the residue-by-residue displacements between the current methodology and the all-atom approach in reference 17 justifies the assumptions that (i) the Hessian obtained from the elastic network adequately describes the system; (ii) it suffices to take the contacts to be homogeneous (i.e. the **K** matrix in equation 7 is identity).

Whether residue 57 is unique in reproducing the conformational response of the protein, or that there are other residues leading to the same result is investigated next. To understand the degree to which such a perturbation is to be selectively introduced in the protein, we first seek to understand the type of force input profile that would lead to the experimentally observed displacements. If the locations of residues whose perturbation lead to the conformational change are highly tailored along the chain, we expect to find a few spikes in the force-response profile; in other words, the $\Delta\mathbf{F}$ vector in equation 2 is expected to be sparse. Thus, we apply equation 6 using the experimental displacement vector, $\Delta\mathbf{S}$ and the magnitudes of the resulting force profile are shown in figure 3b. We find that the inserted forces do not cluster around the few residues near the Fe^{+3} binding region; instead, they are distributed throughout the protein. In particular, peaks arise at several locations that are both in the moving and the fixed domains. Moreover, only one of the peaks in the force response coincides with a residue that is a direct neighbor of the Fe (residue 195). The resulting picture implies that selection of the proper locations for the applied forces which lead to the conformational change may be more complicated than focusing on those located near the reactive center. We therefore scan the whole protein by sequentially applying forces on each residue, and measuring the response (hence the name PRS), as

reported in the next subsection.

Non-reciprocal response to localized perturbations throughout the protein uncovers landscape properties. We perform the residue-by-residue scan on the protein as described in Algorithm A. Starting from each of the initial conformations, (i) apo form, (ii) holo form without the Fe, and (iii) holo form with the Fe ion as an additional node in the network, we compute $\Delta\mathbf{R}$ for each residue and record the response obtained. Typical responses to the perturbation of residue 57 for (i) and (ii) are shown in figure 3a, middle and top curves, respectively. The correlation coefficient, C_{57} (equation 9), between these curves and the experimental curve is 0.95 and 0.92, respectively for these example cases.

In figure 4 we display the resulting correlation coefficients C_i between the calculated and the experimental data for each residue. Note that each point on these figures is the result of the comparison of the displacements of the 309 residues in response to a perturbation inserted at the selected residue, *i*, such as that obtained from the correlation between the middle and bottom curves in figure 3a. We find that there is a plethora of residues whose perturbation leads to the desired "final" structure. This is the most prominent for the perturbations performed on the apo structure (i), whereby *all* exerted forces lead to displacements well correlated with experiments, with the worst perturbation having a correlation of 0.6 ± 0.1 (figure 4a). In fact, on average, perturbing 169 of the 309 residues (55 %) of this protein lead to displacements that are correlated with a coefficient of 0.90 or better with experiments. The reverse scenario (ii) also gives results that are highly correlated with experiments, albeit with some reduction in the number of residues with the highest correlations (64 of the 309; i.e. 21 %, with correlation larger than 0.90; figure 4b). Note that the initial structures for (i) and (ii) are different from each other, the latter having 34 more contacts in total (figure 2). When we modify the input structure (ii) to include the Fe ion leading to input structure (iii), only 24 out of 309 residues (8 %) that perform better than 0.90 correlation remain (figure 4c). In all the simulations, we find that the standard error on the mean increases with decreasing correlations, as determined from the averages of five perturbations in randomly selected directions. For correlations > 0.90, it is less than 0.05 for (i), 0.04 for (ii), and 0.02 for (iii); the largest are 0.2 for the least correlated values.

[Insert Figure 4 here]

In perturbing the residues of the apo FBP (i), the residues that give the worst correlations are 105 and 205-207, all in the fixed domain of the protein. Residue 105 is in the core of the β sheet structure located in this domain and 205-207 are at the turn adjoining a helix to a β strand. Additional residues with the largest deviations between the experiments and predicted values of the structural differences are due to perturbations in the fixed domain. Finally, residues 23 and 249, located in the core of the moving domain also lead to poorer predictions. These relatively low correlation responses (C_i = 0.58 – 0.75) are not due to high coordination numbers of the involved residues which span a wide range of 4 – 13 contacts, implying a more intricate set of interactions leading to these results.

The results in figure 4b show that in general, the correlation patterns in response to forces exerted on the residues of the holo form (ii) are the same; i.e. those with lower correlations are similar in both situations. However, the values of the correlations are somewhat lower (as small as 0.3 for residue 105). The number of residues whose perturbation leads to close-to-perfect correlations (C_i > 0.90) is significantly reduced, although they are still distributed evenly throughout the protein.

A high correlation coefficient between the residue-by-residue patterns of the experimental and the predicted displacements merely suggests that the resulting motions are *directed* towards the holo form. Whether it will actually inter-convert to that structure depends on the energy barrier between the two forms. It may further be that the holo form may only be stabilized in the presence of its ligand. It is highly probable that following an applied perturbation on any given residue, the structure will originally tend towards the holo conformation, but will return to the average apo conformation before reaching the holo state due to the incessant bombardment from the surroundings. Yet, as depicted in figure 4a, most of the perturbations inserted at a variety of residues distributed throughout the protein lead to very high correlations with the experimental results. This implies that the conformational changes attempted by the apo form of the protein in its dynamical environment are narrowly distributed, and tuned towards the holo form. These results further suggest that the apo and holo forms are probably closely located along the free energy landscape.

The Fe ion is known to bind FBP with very high affinity, assisted by the coordinating phosphate ion.

Through the residue-by-residue perturbation analysis (figure 4c) we find that in case (iii), the number of residues that lead to a highly correlated $\Delta\mathbf{R}$ profile with the crystal structure are further reduced. Unlike in (ii), those with high C_i values are not distributed throughout the structure, but are clustered in two α-helices in the fixed domain, and on loops in the moving domain. Thus, once the Fe ion is bound, the structure becomes stabilized such that it can be moved towards the apo form only through perturbing specific residues. In fact, direct force insertion on the Fe ion itself leads to an average correlation coefficient of 0.88.

To display the reduction in the number of residues that lead to favorable conformational change upon perturbation, the correlations in figure 4a-c are ranked according to their size, and are plotted in figure 4d. The holo structure is also shown in the inset, where the residues whose perturbation leads to displacements with the largest C_i [i.e., those that lie above the dotted line for (iii)] are colored respectively in yellow. Also marked in orange are residues that lead to the lowest correlations, all located in the fixed domain (residues 105, 124, 278; ball and stick representation).

In passing we note that the current PRS approach provides information on the inherent motions of proteins that cannot be uncovered by a normal modal analysis (NMA). In the latter, the projection of the overall motion on the slowest modes has been found to predict the experimental displacements in some studies,[42-44] but it has been discussed in detail that the success of this approach is limited by the degree of collectivity displayed by the conformational change.[44,45] We wish to emphasize that although both methods utilize the Hessian in their respective analyses, NMA assumes that the most collective modes of motion carry information on global structural changes, and involves a principal component analysis of the Hessian. PRS, however, does not discriminate between the modes, and utilizes the whole Hessian as a kernel that projects selected perturbations as displacements. Furthermore, NMA does not provide a methodology for revealing specific residues that are operative in the execution of the observed motions, unless it is coupled with other tools.[46,47] In the next subsection, we rely on the directionality information provided by the PRS to uncover the mechanisms that dictate protein functionality.

Local and remote modulation of Fe ion dissociation. To provide further understanding of how the

protein operates structurally, we turn to the few residues that give high correlations in the presence of the ligand (colored yellow in the inset to figure 4d). These include residues that are either in the fixed domain that support the ferric binding region or those that are located in the moving domain loops. Thus, it is possible to manipulate the bound form of the protein towards the unbound form by either directly perturbing the Fe-binding residues, or by controlling the distant flexible loops. If there is a directionality preference in the response of residues, they should additionally be revealed by the current analysis. Such a preference may be imposed by the excluded volume, or may be the result of close coupling between the motions of subsets of nodes.

In the previous subsection we have shown that in holo FBP, singly placed forces on residues listed under the caption to figure 4 display the experimental displacement profile to a very good precision. The relative magnitudes of the displacements are correctly captured by this procedure. On the other hand, the recorded response also has a directionality that is concealed in these correlations. One measure of directionality is the overlap, defined as the cosine of the angle between the calculated and measured direction of motion. However, this definition eliminates information on the magnitude of the motion, and even if the applied methodology captures the essential motions of key residues, many others that have no preferred direction of motion blur the statistics. Therefore, the results from a selected subset of the residues may be more informative, as demonstrated by Gerstein and coworkers.[48] In fact, we find that the statistics collected over the perturbations applied to each residue over five independent runs result in overlap distributions expected from random directionality (results not shown). We therefore perform the following alternative analysis (see Algoritm B): For a selected residue, e.g., from those that have the highest correlations ($C_i > 0.9$), we introduce a large collection of perturbations in directions that are spherically symmetric around it, so that their vectorial sum is zero. For each perturbation, we monitor the resulting response in the residues directly contacting the Fe^{+3}.

The residues for which $C_i > 0.9$ and are far from the Fe ion ($d_{i\text{-Fe}} >> r_c$) are D47, D52, both at the tips of distant loops as well as the loop spanning 232-236. The results for the perturbations on D47 (C_i = 0.91, $d_{47\text{-Fe}}$ = 28.5 Å) are shown in figure 5, where the inserted forces are in red and the responses are in orange. Many of the residues that have high displacements are found to move in a plane, due to

constraints imposed by chain connectivity. Others, usually those with small magnitudes, move in a spherical region, whereas a few show movements along a line; response sets intermediate of a line and a plane are also observed. We further focus on the responses of the 11 residues whose C_α atoms are within 8 Å of the ion; the volume they take up in the protein is shown with the shaded area, and is zoomed in the inset. Three are located in the moving domain (residues 8, 9 and 57), and the remaining are in the fixed domain. The former operate in a coherent fashion at the tip of the cap that opens the exit of the Fe ion. Here, we loosely define coherence as the tendency of residues to move along parallel lines. The fixed domain residues, on the other hand, respond to these perturbations in directions that occur in a plane normal to the chain axis. The coherence is also obtained for forces applied on residues 52 and 232-236 ($C_i > 0.9$ and $d_{i\text{-Fe}} > 21$ Å). On the other hand, directly perturbing the Fe ion, as well as other residues for which $C_i > 0.9$ and $d_{i\text{-Fe}} \approx r_c$, destroys this coherence of the cap residues; i.e., they move in independent directions. Thus, local cooperativity emerges by perturbing specific distantly located residues.

[Insert Figure 5 here]

Our findings suggest the existence of two alternative mechanism of Fe ion release in FBP: The ion is either is closely controlled by synergistic anions and chelators acting in the vicinity of the ion, or remote control by distant charged residues located in solvent exposed loops is achieved (e.g. D47, D52, K234) due to their observed ability to mechanically control the cap over the ligand binding region. For FBP, the former type of control has been evidenced by a plethora of experiments where the exchange of synergistic anions forming relatively stable intermediates or the direct action of chelators on the ion have been observed.[23,25,26] There are no mutational studies on FBP directly implicating the distant residues mentioned in latter scenario. However, it was recently shown that *H. influenza* strains expressing mutant proteins that are defective in binding the phosphate anion are capable of donating iron, calling for mechanisms of iron transport that do not involve a synergistic anion.[27] Furthermore, the kinetic effect of chloride and perchlorate (which does not coordinate to Fe^{+3}) has called for anion binding sites on the surface of FBP, similar to those found in transferrin.[25] Allosteric anion binding sites that trigger large conformational changes located at the surface[8,31,37] have been experimentally

determined in the structurally and functionally similar protein transferrin. In particular, R124 located in the N-lobe of the latter has been found to control iron release rate by anchoring synergistic anions.[34] Structural alignment[49] of transferrin with FBP shows that this position is equivalent to F142 located in the helix supporting the Fe binding region; the latter is amongst the residues that result in the highest correlations with the experimental data upon perturbation. Similarly, K206 which provides anion binding sites in human transferrin N-lobe holds an equivalent position to that of E193 in FBP, the latter also showing high displacement correlation following its perturbation.

CONCLUSIONS

In recent years a kinetic model of ligand binding has been emerging whereby the energy landscape of the free protein is made-up of a set of coupled low-free energy states[50] which is modulated by bound ligands to create preferred kinetic pathways.[51] A mechanism on the details of how the system manages this scenario is provided by the current PRS analysis. It points to specific locations along the protein structure that are mechanically supported by the protein scaffold to control various events such as binding. Independent of the direction of the applied force on these specific residues, the scaffold transduces the incoming energy to remote regions in an organized fashion (figure 5). However, whether these residues are actually utilized in the cell environment depends on their amenability to attract factors that are capable of applying the necessary forces. The accessibility of these residues and their chemical/electrostatic properties that allow for binding of controlling agents determine their utilization.

For the particular case of FBP, the apo structure accommodates a strong tendency towards the holo form in the absence of the ligand, which is uncovered by an analysis that is performed on all of its residues (figure 4a). The process is not reversible in that the ligand stabilizes the holo form and the need arises for an action orchestrated by few residues occupying key locations along the structure to reveal concerted deformations that may eventually displace the ligand (figure 5). The so-called Fe^{+3} transport dilemma is thus resolved by the readiness of the protein to uptake the ion in the apo structure, but the necessity to perturb highly specific locations along the chain for its release in holo FBP.

Whether an apo(holo) structure has an intrinsic propensity towards the holo(apo) form (thermodynamic control; e.g., figure 4a), or it requires perturbations of specific sites for conformational change (kinetic control; e.g. figures 4c and 5), it is the network of interactions in the specific fold that leads to these modifications.[52] Specificity comes into play while the polypeptide chain folds to form the scaffold for the mechanical control structure, and when providing the chemical constituents along the surface to maintain a balance on the kinetics. The paths along which the transduction takes place are important for designing proteins that will perform new functions on old folds. Robust techniques to predict these paths using the network structure of the residues as well as the specificity of their interactions, coupled with point mutation studies, will not only aid the design process but also will uncover the physics behind the communication between remote residues.[53,54]

Acknowledgments. We thank Z. Nevin Gerek and S. Banu Ozkan for stimulating discussions. This work was partially supported by the Scientific and Technological Research Council of Turkey Project No. 106T522.

REFERENCES

1. Zhang XJ, Wozniak JA, Matthews BW. Protein Flexibility and Adaptability Seen in 25 Crystal Forms of T4 Lysozyme. J Mol Biol 1995;250(4):527-552.
2. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, Skalicky JJ, Kay LE, Kern D. Intrinsic dynamics of an enzyme underlies catalysis. Nature 2005;438(7064):117-121.
3. Gunasekaran K, Ma BY, Nussinov R. Is allostery an intrinsic property of all dynamic proteins? Proteins-Structure Function and Bioinformatics 2004;57(3):433-443.
4. Volkman BF, Lipson D, Wemmer DE, Kern D. Two-state allosteric behavior in a single-domain signaling protein. Science 2001;291(5512):2429-2433.
5. Tang C, Schwieters CD, Clore GM. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. Nature 2007;449(7165):1078-1082.
6. Zhuravleva A, Korzhnev DM, Nolde SB, Kay LE, Arseniev AS, Billeter M, Orekhov VY. Propagation of dynamic changes in barnase upon binding of barstar: An NMR and computational study. J Mol Biol 2007;367(4):1079-1092.
7. Baysal C, Atilgan AR. Coordination topology and stability for the native and binding conformers of chymotrypsin inhibitor 2. Proteins-Structure Function and Genetics 2001;45(1):62-70.
8. Hamilton DH, Turcot I, Stintzi A, Raymond KN. Large cooperativity in the removal of iron from transferrin at physiological temperature and chloride ion concentration. Journal of Biological Inorganic Chemistry 2004;9(8):936-944.
9. Blaber M, Baase WA, Gassner N, Matthews BW. Alanine Scanning Mutagenesis of the Alpha-Helix-115-123 of Phage-T4 Lysozyme - Effects on Structure, Stability and the Binding of Solvent. J Mol Biol 1995;246(2):317-330.
10. Min W, English BP, Luo GB, Cherayil BJ, Kou SC, Xie XS. Fluctuating enzymes: Lessons from single-molecule studies. Accounts of Chemical Research 2005;38(12):923-931.

11. Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, Fenn T, Pozharski E, Wilson MA, Petsko GA, Karplus M, Hubner CG, Kern D. Intrinsic motions along an enzymatic reaction trajectory. Nature 2007;450(7171):838-844.

12. Prytkova TR, Kurnikov IV, Beratan DN. Coupling coherence distinguishes structure sensitivity in protein electron transfer. Science 2007;315(5812):622-625.

13. Velyvis A, Yang YR, Schachman HK, Kay LE. A solution NMR study showing that active site ligands and nucleotides directly perturb the allosteric equilibrium in aspartate transcarbamoylase. Proceedings of the National Academy of Sciences of the United States of America 2007;104(21):8815-8820.

14. Atilgan C, Aviyente V. Hybrid usage of computational tools in drug synthesis. Current Computer-Aided Drug Design 2007;3(2):149-159.

15. Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature 2007;450(7172):1001-1009.

16. Atilgan C, Aykut AO, Atilgan AR. How a vicinal layer of solvent modulates the dynamics of proteins. Biophysical Journal 2008;94(1):79-89.

17. Ikeguchi M, Ueno J, Sato M, Kidera A. Protein structural change upon ligand binding: Linear response theory. Physical Review Letters 2005;94(7):078102.

18. Yilmaz LS, Atilgan AR. Identifying the adaptive mechanism in globular proteins: Fluctuations in densely packed regions manipulate flexible parts. Journal of Chemical Physics 2000;113(10):4454-4464.

19. Chen CY, Berish SA, Morse SA, Mietzner TA. The Ferric Iron-Binding Protein of Pathogenic Neisseria Spp Functions as a Periplasmic Transport Protein in Iron Acquisition from Human Transferrin. Molecular Microbiology 1993;10(2):311-318.

20. Bekker EG, Creagh AL, Sanaie N, Yumoto F, Lau GHY, Tanokura M, Haynes CA, Murphy MEP. Specificity of the synergistic anion for iron binding by ferric binding protein from Neisseria gonorrhoeae. Biochemistry 2004;43(28):9195-9203.

21. Dhungana S, Taboy CH, Anderson DS, Vaughan KG, Aisen P, Mietzner TA, Crumbliss AL. The influence of the synergistic anion on iron chelation by ferric binding protein, a bacterial

transferrin. Proceedings of the National Academy of Sciences of the United States of America 2003;100(7):3659-3664.

22. Gabricevic M, Anderson DS, Mietzner TA, Crumbliss AL. Kinetics and mechanism of iron(III) complexation by ferric binding protein: The role of phosphate. Biochemistry 2004;43(19):5811-5819.

23. Dhungana S, Anderson DS, Mietzner TA, Crumbliss AL. Kinetics of iron release from ferric binding protein (FbpA): Mechanistic implications in bacterial periplasm-to-cytosol Fe3+ transport. Biochemistry 2005;44(28):9606-9618.

24. Khan AG, Shouldice SR, Kirby SD, Yu RH, Tari LW, Schryvers AB. High-affinity binding by the periplasmic iron-binding protein from Haemophilus influenzae is required for acquiring iron from transferrin. Biochemical Journal 2007;404:217-225.

25. Boukhalfa H, Anderson DS, Mietzner TA, Crumbliss AL. Kinetics and mechanism of iron release from the bacterial ferric binding protein nFbp: exogenous anion influence and comparison with mammalian transferrin. Journal of Biological Inorganic Chemistry 2003;8(8):881-892.

26. Heymann JJ, Weaver KD, Mietzner TA, Crumbliss AL. Sulfate as a synergistic anion facilitating iron binding by the bacterial transferrin FbpA: The origins and effects of anion promiscuity. Journal of the American Chemical Society 2007;129(31):9704-9712.

27. Khan AG, Shouldice SR, Tari LW, Schryvers AB. The role of the synergistic phosphate anion in iron transport by the periplasmic iron-binding protein from Haemophilus influenzae. Biochemical Journal 2007;403:43-48.

28. Wally J, Halbrooks PJ, Vonrhein C, Rould MA, Everse SJ, Mason AB, Buchanan SK. The crystal structure of iron-free human serum transferrin provides insight into inter-lobe communication and receptor binding. Journal of Biological Chemistry 2006;281(34):24934-24944.

29. Zak O, Aisen P, Crawley JB, Joannou CL, Patel KJ, Rafiq M, Evans RW. Iron Release from Recombinant N-Lobe and Mutants of Human Transferrin. Biochemistry 1995;34(44):14428-14434.

30. Mizutani K, Muralidhara BK, Yamashita H, Tabata S, Mikami B, Hirose M. Anion-mediated Fe3+ release mechanism in ovotransferrin C-lobe - A structurally identified SO42- binding site and its implications for the kinetic pathway. Journal of Biological Chemistry 2001;276(38):35940-35946.

31. Zak O, Tam B, MacGillivray RTA, Aisen P. A kinetically active site in the C-lobe of human transferrin. Biochemistry 1997;36(36):11036-11043.

32. Zak O, Aisen P. Iron release from transferrin, its C-lobe, and their complexes with transferrin receptor: Presence of N-lobe accelerates release from C-lobe at endosomal pH. Biochemistry 2003;42(42):12330-12334.

33. Zak O, Ikuta K, Aisen P. The synergistic anion-binding sites of human transferrin: Chemical and physiological effects of site-directed mutagenesis. Biochemistry 2002;41(23):7416-7423.

34. Adams TE, Mason AB, He QY, Halbrooks PJ, Briggs SK, Smith VC, MacGillivray RTA, Everse SJ. The position of arginine 124 controls the rate of iron release from the N-lobe of human serum transferrin - A structural study. Journal of Biological Chemistry 2003;278(8):6027-6033.

35. He QY, Mason AB, Nguyen V, MacGillivray RTA, Woodworth RC. The chloride effect is related to anion binding in determining the rate of iron release from the human transferrin N-lobe. Biochemical Journal 2000;350:909-915.

36. Abergel RJ, Raymond KN. Terephthalamide-containing ligands: fast removal of iron from transferrin. Journal of Biological Inorganic Chemistry 2008;13(2):229-240.

37. Harris WR, Brook CE, Spilling CD, Elleppan S, Peng W, Xin M, Van Wyk J. Release of iron from transferrin by phosphonocarboxylate and diphosphonate chelating agents. Journal of Inorganic Biochemistry 2004;98(11):1824-1836.

38. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Research 2000;28(1):235-242.

39. Bahar I, Jernigan RL. Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. J Mol Biol 1997;266(1):195-214.

40. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophysical Journal 2001;80(1):505-515.

41. Atilgan C, Inanc I, Atilgan AR. Residue Network Construction and Predictions of Elastic Network Models. Biophys J 2008;submitted.

42. Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins-Structure Function and Genetics 1998;33(3):417-429.

43. Keskin O. Binding induced conformational changes of proteins correlate with their intrinsic fluctuations: a case study of antibodies. Bmc Structural Biology 2007;7:31.

44. Yang L, Song G, Jernigan RL. How well can we understand large-scale protein motions using normal modes of elastic network models? Biophysical Journal 2007;93(3):920-929.

45. Petrone P, Pande VS. Can conformational change be described by only a few normal modes? Biophysical Journal 2006;90(5):1583-1593.

46. Zen A, Carnevale V, Lesk AM, Micheletti C. Correspondences between low-energy modes in enzymes: Dynamics-based alignment of enzymatic functional families. Protein Science 2008;17(5):918-929.

47. Carnevale V, Raugei S, Micheletti C, Carloni P. Convergent dynamics in the protease enzymatic superfamily. Journal of the American Chemical Society 2006;128(30):9766-9772.

48. Alexandrov V, Lehnert U, Echols N, Milburn D, Engelman D, Gerstein M. Normal modes for predicting protein motions: A comprehensive database assessment and associated Web tool. Protein Science 2005;14(3):633-643.

49. Shatsky M, Nussinov R, Wolfson HJ. A method for simultaneous alignment of multiple protein structures. Proteins-Structure Function and Bioinformatics 2004;56(1):143-156.

50. Vendruscolo M, Dobson CM. Dynamic visions of enzymatic reactions. Science 2006;313(5793):1586-1587.

51. Boehr DD, McElheny D, Dyson HJ, Wright PE. The dynamic energy landscape of dihydrofolate reductase catalysis. Science 2006;313(5793):1638-1642.

52. Atilgan AR, Akan P, Baysal C. Small-world communication of residues and significance for protein dynamics. Biophysical Journal 2004;86(1):85-91.

53. Atilgan AR, Turgut D, Atilgan C. Screened nonbonded interactions in native proteins manipulate optimal paths for robust residue communication. Biophysical Journal 2007;92(9):3052-3062.

54. Lockless SW, Ranganathan R. Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. Science 1999;286:295-299.

55. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. Journal of Molecular Graphics 1996;14(1):33-38.

Figure Captions.

Figure 1. Excerpted from the protain chain (upper panel), scheme depicting the free body diagram of a $C_{\alpha i}$ atom coordinated by $C_{\alpha j}$'s within a cut-off radius r_c (lower left). $\Delta\mathbf{f}_{ij}$ denotes the interaction force between *i* and *j*. Under an external force applied on residue *l*, $\Delta\mathbf{F}_l$, the residues are displaced in space (from the black to the gray nodes in the lower right). The contacting pairs are assumed not to change under this force.

Figure 2. Upper panel displays the haemophilus influenzae ferric-binding protein in apo (yellow; PDB code: 1D9V) and holo forms (purple; PDB code: 1MRP). The two structures are superimposed on the fixed domain (residues 83 – 87, 102 – 225, 277 – 307). The Fe^{3+} binding location is shown in red. Residues 9, 57, 175 and 193 are within 7 Å of the Fe atom. In addition, residues 8, 139-141, 176, 195 and 196 are in its 7 – 8 Å range. The lower panel shows the difference between the number of contacts of the ferric bound and unbound forms of FBP.

Figure 3. (a) Relative displacements of residues between the apo and holo forms (experimental), and typical responses to a given force perturbation on residue 57 in the apo and holo forms. The latter two curves are nudged for ease of comparison; their baselines are shown by dashed lines. (b) Profile of the magnitudes of the forces necessary to obtain the experimental structural difference between the apo and holo forms. Only one of the peaks (residue 195) coincides with the residues that directly contact the ion, listed in the caption to figure 1.

Figure 4. The displacement vectors between the perturbed and initial structures are compared with those of the crystal structures. The initial structures are (i) – (iii) in the text, respectively. In (d), data in parts (a) – (c) are sorted according to their magnitude; as an inset, residues that give the highest and the lowest correlations in (iii) are shown on structure, marked in yellow and orange, respectively. (Residues 47, 52, 130, 139-144, 147, 148, 166, 174, 186, 226, 232-236, 293, 296, 298, 299 in yellow, and residues 105, 124, 278 in orange).

Figure 5. The response of the protein to forces exerted on residue $j = 47$ in different directions (forces are shown with the red arrows.) The collection of all the displacements is shown in orange; responses

on the three neighbors along the chain in either direction ($44 \leq j \leq 50$) are not shown for clarity. The region of the first neighbors of the Fe^{+3} is magnified, and the volume taken up by this region is shown in cage representation. The Fe^{+3} ion contacting residues in the fixed domain (located on the right hand side) respond incoherently, moving mainly within a plane. Residues 8, 9 and 57, that are in the moving domain of the protein respond coherently, moving parallel to each other, tending to open the cap for Fe ion to exit. The coherence is also present when forces are applied to distant residues that display high correlations and is absent for the others (see text for more details). Figures are drawn with the VMD software.[55]

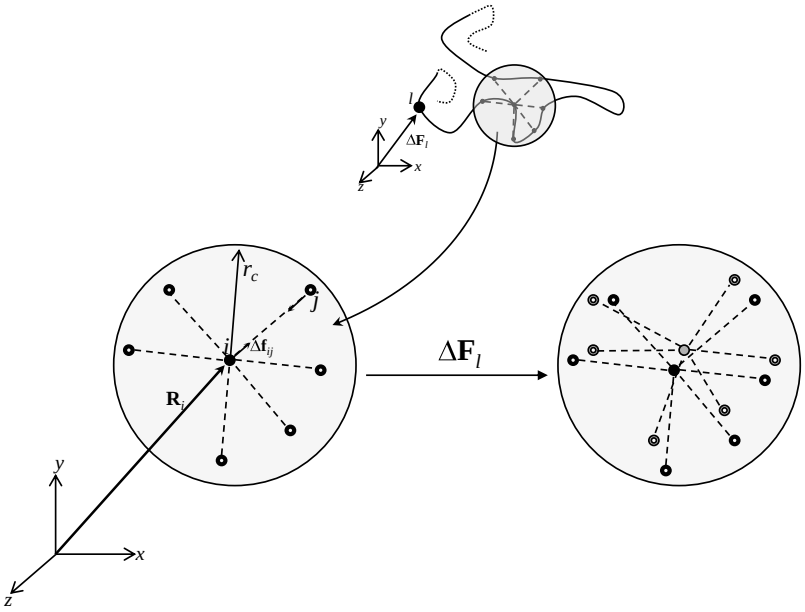


Figure 1

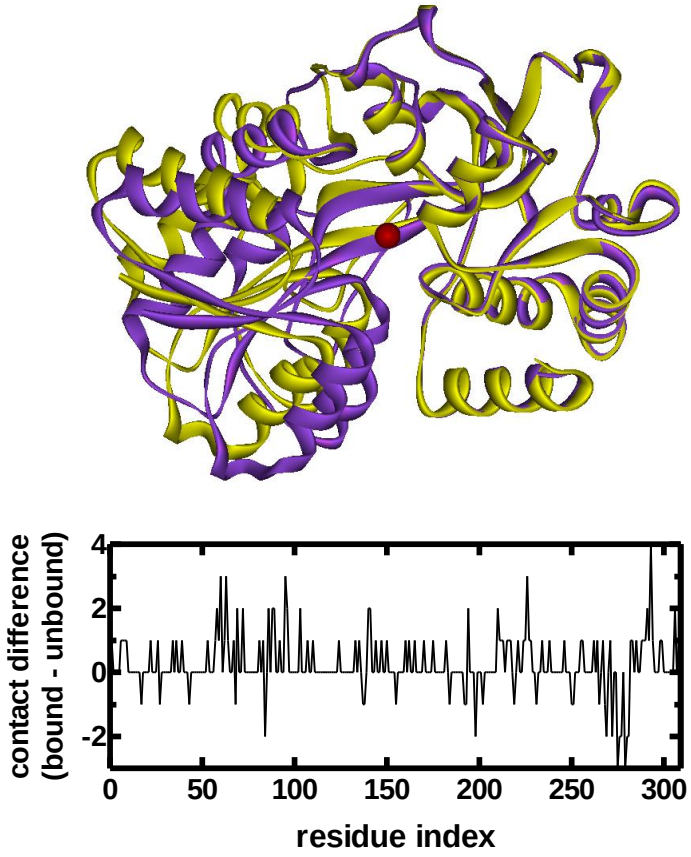


Figure 2

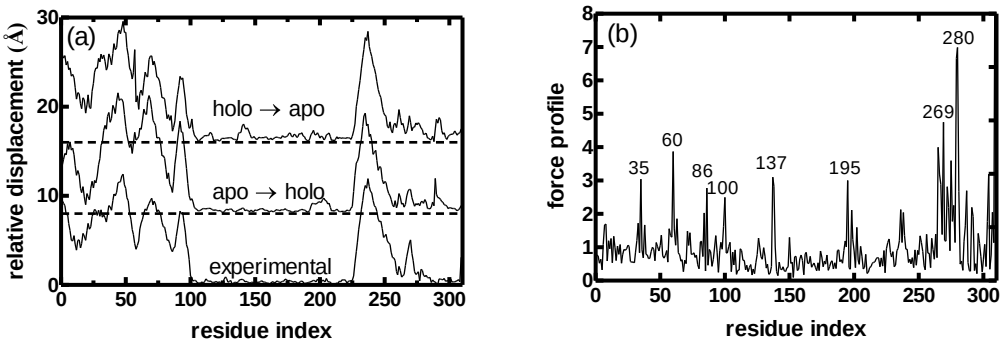


Figure 3

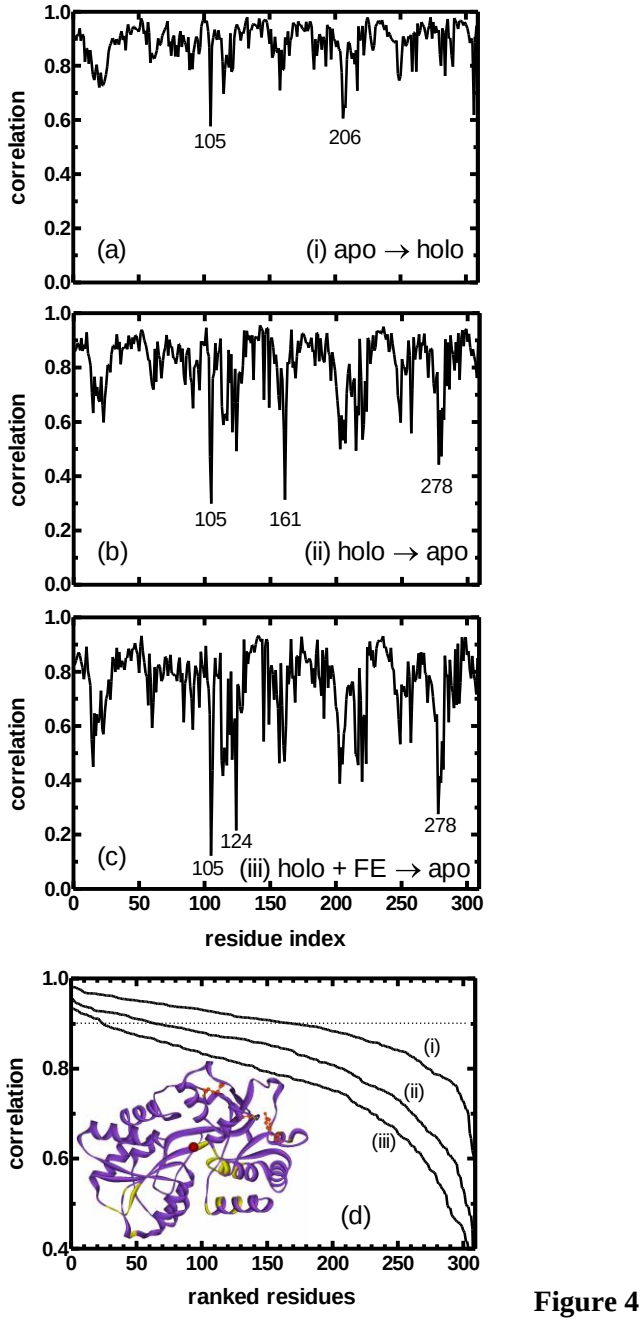


Figure 4

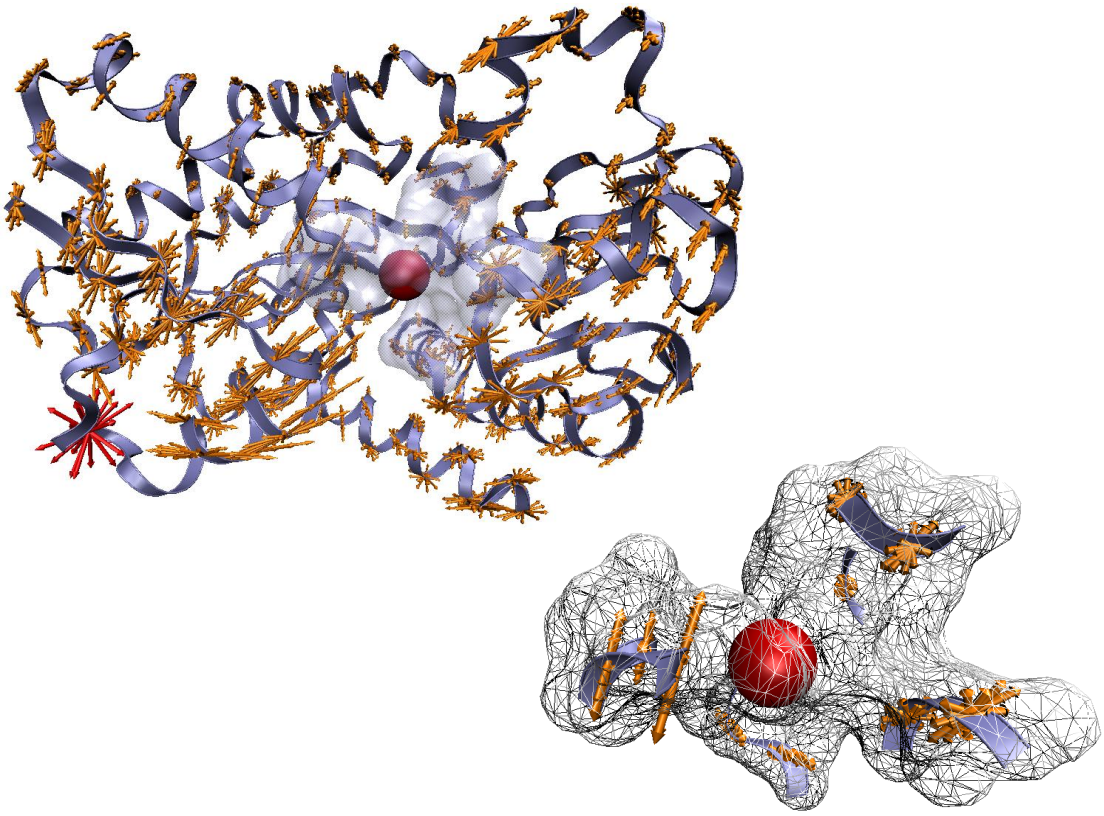
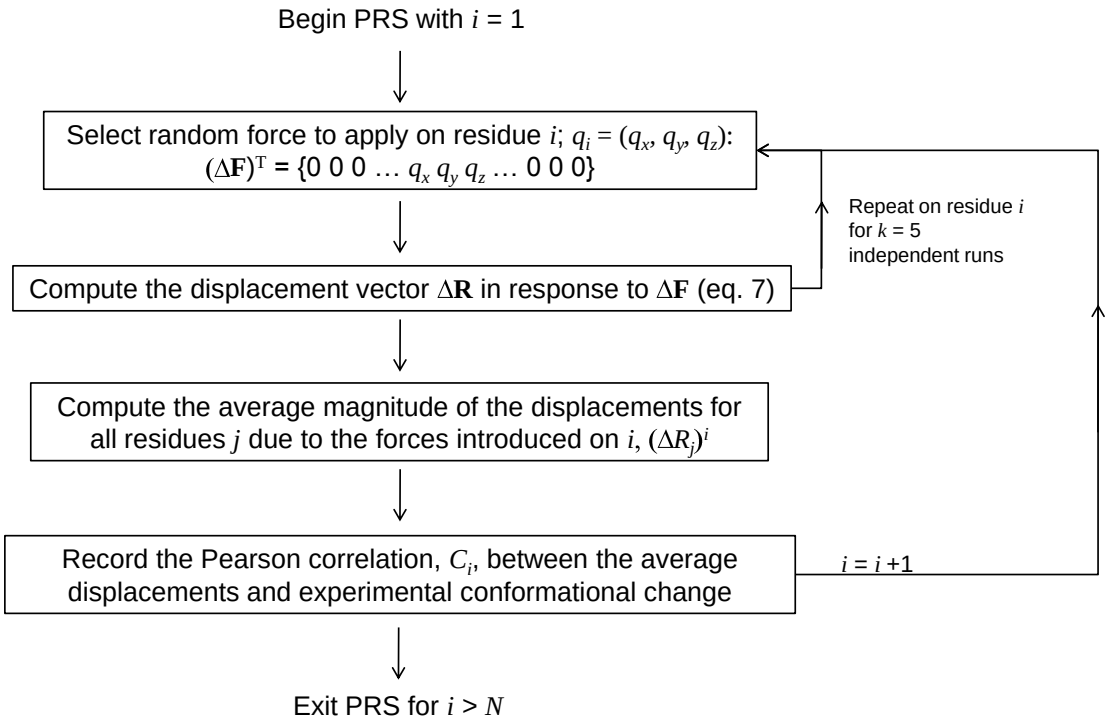
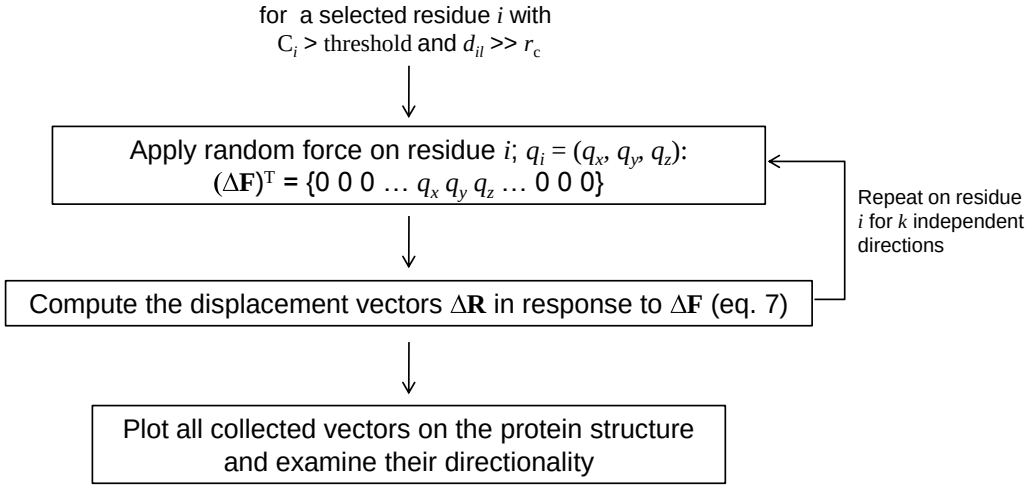


Figure 5



Algorithm A



Algorithm B