Perturbation-Response Scanning Reveals Ligand Entry-Exit Mechanisms of Ferric Binding Protein

Canan Atilgan*, Ali Rana Atilgan

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

Abstract

We study apo and holo forms of the bacterial ferric binding protein (FBP) which exhibits the so-called ferric transport dilemma: it uptakes iron from the host with remarkable affinity, yet releases it with ease in the cytoplasm for subsequent use. The observations fit the "conformational selection" model whereby the existence of a weakly populated, higher energy conformation that is stabilized in the presence of the ligand is proposed. We introduce a new tool that we term perturbation-response scanning (PRS) for the analysis of remote control strategies utilized. The approach relies on the systematic use of computational perturbation/response techniques based on linear response theory, by sequentially applying directed forces on single-residues along the chain and recording the resulting relative changes in the residue coordinates. We further obtain closed-form expressions for the magnitude and the directionality of the response. Using PRS, we study the ligand release mechanisms of FBP and support the findings by molecular dynamics simulations. We find that the residue-by-residue displacements between the apo and the holo forms, as determined from the X-ray structures, are faithfully reproduced by perturbations applied on the majority of the residues of the apo form. However, once the stabilizing ligand (Fe) is integrated to the system in holo FBP, perturbing only a few select residues successfully reproduces the experimental displacements. Thus, iron uptake by FBP is a favored process in the fluctuating environment of the protein, whereas iron release is controlled by mechanisms including chelation and allostery. The directional analysis that we implement in the PRS methodology implicates the latter mechanism by leading to a few distant, charged, and exposed loop residues. Upon perturbing these, irrespective of the direction of the operating forces, we find that the cap residues involved in iron release are made to operate coherently, facilitating release of the ion.

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 [6,7]. To explore the presence or the absence of such "shifts in the energy landscapes," [8] one needs to perturb the protein structure, and observe the response [9]. The perturbation may be in the form of changing the environmental factors (e.g. changes in ionic concentration [10]), or may target specific locations on the structure itself, either through chemically modifying the residues (inserting mutations) [11] or by inducing site-specific perturbations (e.g. as is done in single molecule experiments [12], or through ligand binding). Ubiquitous post-translational modifications are also possible. The response may be measured directly, as a change in the overall conformation of the protein [13], or indirectly, e.g., through determining the kinetic parameters, and proposing kinetic models that explain the observations.[14,15] 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 [16,17], or to tailor enzymes serving as "materials."[18]

Linear response theory (LRT) has been recently used to study conformational changes undergone by proteins under selected external perturbations [19]. This approach has recently been applied to the study of the conformational switching upon phosphorylation [20]. In this study, we develop a toolkit that we term perturbation-response scanning (PRS) which is based on sequential application of LRT to study the origins of structural changes undergone by protein molecules. Similar approaches have been adopted in other work, whereby the perturbations on residues are introduced by modifying the effective force constants [21] or distances [22] between contacting pairs.

PRS relies on systematically applying forces at singly selected residues and recording the linear response of the whole protein. The response is quantified as both the magnitude of the displacements undergone by the residues, and their directionality. Closed form expressions that summarize the theoretical implications of the PRS technique in the limit of a large number of perturbations introduced at a given residue are provided. We note that we have previously studied the stability of proteins using a similar sequential perturbation-response approach, based on inserted displacements followed by energy minimization of the system [9,23]. Therein we have also shown that the response of the

Author Summary

Upon binding ligands, many proteins undergo structural changes compared to the unbound form. We introduce a methodology to monitor these changes and to study which mechanisms arrange conformational shifts between the liganded and free forms. Our method is simple, yet it efficiently characterizes the response of proteins to a given perturbation on systematically selected residues. The coherent responses predicted are validated by molecular dynamics simulations. The results indicate that the iron uptake by the ferric binding protein is favorable in a thermally fluctuating environment, while release of iron is allosterically moderated. Since ferric binding protein exhibits a high sequence identity with human transferrin whose allosteric anion binding sites generate large conformational changes around the binding region, we suggest mutational studies on remotely controlling sites identified in this work.

system is within the linear regime for local distortions of atoms up to ca. 1.5 Å, despite the large local forces brought about [9].

Using PRS, we analyze the ferric binding protein A (FBP) as an example system, and describe alternative approaches that may have evolved in the structure to control function. The validity of the methodology is supported by molecular dynamics (MD) simulations. 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 the protein 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 [24]. The protein is made up of two domains characteristic of periplasmic Fe^{+3} binding family as well as the host protein transferrin. 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 [25]. 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 [26]. 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 [25,27,28]. Another involves the direct action of chelators on the ion [25]. 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 [29].

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 [30]. 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 [31]. Despite the resemblance, the iron binding/release kinetics in the two lobes differ. It has been implicated that there may be several approaches used for iron release in transferrins, including chelating agents and synergistic anions acting directly around the ferric binding site [32–35]. 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 [10,36,37]. 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 the current work, we study FBP in detail due to an extensive literature on the iron uptake mechanisms of this and evolutionarily related proteins; moreover, the molecular dynamics (MD) results of the apo structure have previously been analyzed by perturbing a singly selected residue with linear response theory [19]. We develop the PRS scheme such that, (i) we systematically apply LRT by scanning every residue on the protein so as to discriminate between residues that have major contributions to the biologically significant displacements, measured by x-ray experiments; (ii) we provide closed-form expressions for both the magnitude and the directionality of the response; (iii) we carry out further analysis of the response to uncover regions in the protein where coherent responses occur. Our findings are in agreement with a model where iron uptake by the protein is a favored process in the fluctuating environment, while iron release is specifically managed through several mechanisms including chelation and allosteric control. Furthermore, our findings suggest additional locations on the protein surface, far from the binding site, for allosteric control of iron release. The observations fit the "conformational selection" model whereby the existence of a weakly populated, higher energy conformation that is stabilized in the presence of the ligand is proposed [38,39]; the hypothesis has recently been supported by NMR experiments by studying the protein structural ensemble of up to microseconds [40].

Methods

The new tool introduced in this work for the analysis of remote control strategies utilized by proteins is based on applying forces at a given residue as a perturbation, 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 (PRS). Below, we first review the theory and then outline the details of the PRS technique. Finally, we describe the MD simulations.

Linear response theory. Here we present a derivation of how a structure may be manipulated by external forces [41,42]. We construct the protein as a residue network of N nodes that are 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).

In the notation used, **r** and **f** refer to the bond and internal force vectors along the edge connecting any two nodes, respectively. On the other hand, **R** and **F** are vectors on the nodes and are referred to as the position and external force vectors, respectively. There are m interactions pertaining to each residue (Figure 1, as an example, schematically illustrates the interactions for a residue that has six interactions, i.e., $m = 6$), and a total of M interactions for the system of N residues. In the absence of an external force acting on the system, the equilibrium condition for each residue, i, necessitates that the summation of the internal, residue-residue
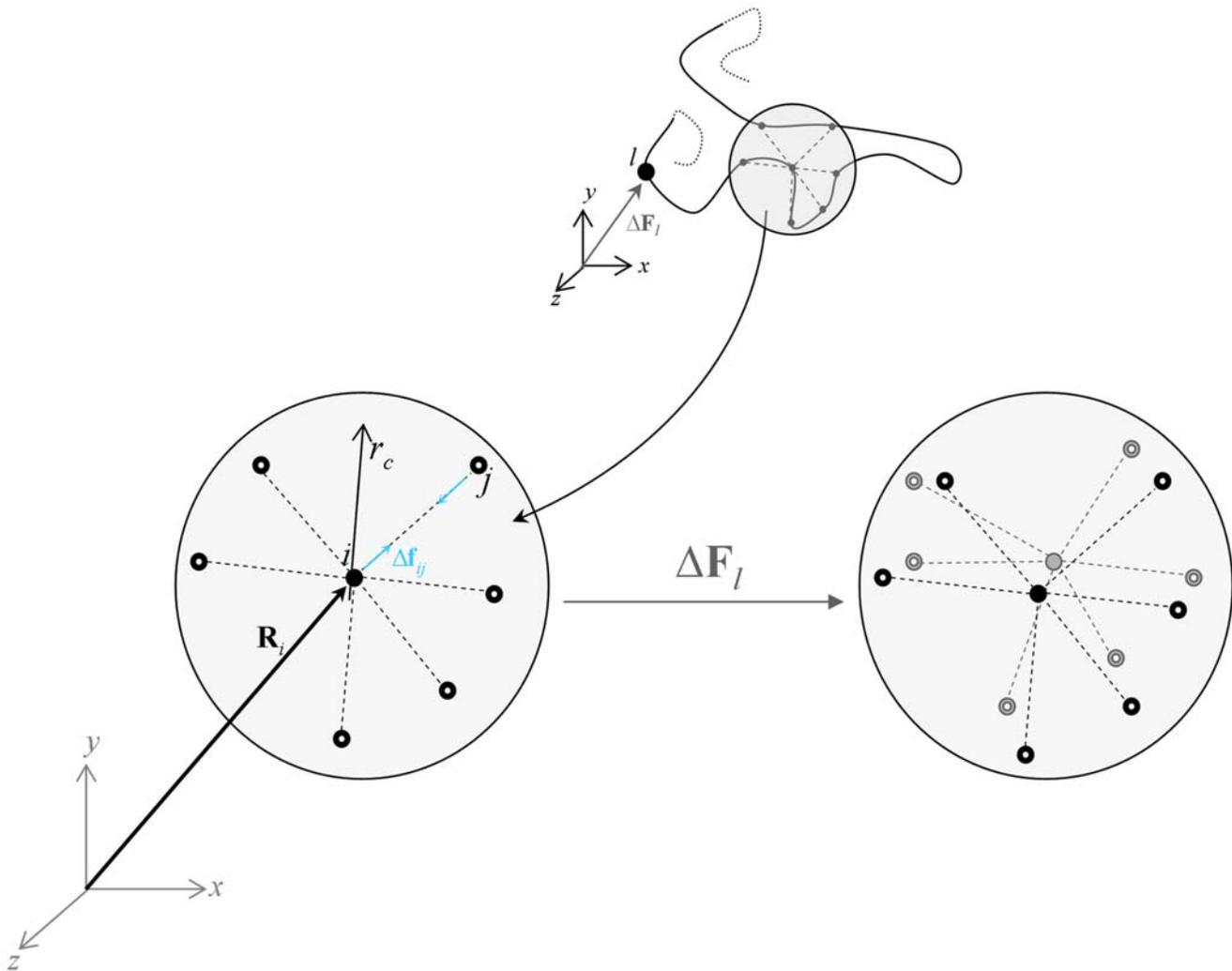
Figure 1. Free-body diagram of a residue. Excerpted from the protein chain (upper panel), scheme depicting the free body diagram of a $C_{\alpha i}$ atom coordinated by $C_{\alpha j}$'s within a cut-off radius r_c (lower left). $\Delta \mathbf{f}_{ij}$ denotes the interaction force between i and j. Under 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.
doi:10.1371/journal.pcbi.1000544.g001

interaction forces must be zero for each residue. Therefore,

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

where the $3 \times m$ coefficient matrix $\mathbf{b}$ consists of the direction cosines of each force representing the residue-residue interaction. The row indices of $\mathbf{b}$ are $x, y,$ or z. Here $\Delta \mathbf{f}_i$ is an $m \times 1$ column matrix of forces aligned in the direction of the bond between the two interacting residues. For instance, in Figure 1, residue i has six contacts; and, thus, $\Delta \mathbf{f}_i$ is a 6×1 column matrix. Following the example outlined in Figure 1, equation 1 sums up the projection of these six forces on the $x, y,$ and z-axes. This algebra gives rise to three independent equations involving six unknown interaction forces, which are the residual interaction forces of residue i with its contacting neighbors. One can write the equilibrium condition (equation 1) for each residue. This results in a total of N sets of equations, each of which involves the summation of forces in three respective directions. Consequently, generalizing equation 1 to the whole system of N nodes and a total of M interactions, one can write the following algebraic system of a total of $3N$ number of equations consisting of

M number of unknown residue-residue interaction forces

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

with the $3N \times M$ direction cosine matrix $\mathbf{B}$ and the $M \times 1$ column matrix of residue-residue interaction forces, $\Delta \mathbf{f}$. It is straightforward to generate the matrix $\mathbf{B}$ from the topology of the native structure (i.e., the protein data bank (PDB) file[43]) with a specified r_c. As an example, apo FBP has 309 residues and a total number of 1542 interactions when the cut-off distance of 8 Å is selected.

In the presence of an external force, $\Delta \mathbf{F}$ (Figure 1), the equilibrium consideration for each residue dictates that the summation of the residue-residue interaction forces for each residue must be equal to the external, applied force on the same residue. Then, the equation 2 may be cast into the following form

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

Under the action of external forces, each residue experiences a displacement, $\Delta \mathbf{R}$, which is termed the positional displacement

vector. Moreover, the bond distance between any two residues changes in the amount of $\Delta\mathbf{r}$ in accord with the positional displacements of the two residues which participate in the bonding. Therefore, there must be compatibility between the total of $3N$ number of positional displacements and the changes that take place in the intra-residual distances, a total of M number of distortions. This compatibility is very similar to the form given in equation 3 [42]:

$$\mathbf{B}^T_{M \times 3N} \Delta\mathbf{R}_{3N \times 1} = \Delta\mathbf{r}_{M \times 1} \qquad (4)$$

Within the scope of an elastic network of residues that are connected to their neighbors with linear-elastic springs, the residual interaction forces, $\Delta\mathbf{f}$, are related to the changes in the contact distances, $\Delta\mathbf{r}$, through Hooke's law by

$$\mathbf{K}_{M \times M} \Delta\mathbf{r}_{M \times 1} = \Delta\mathbf{f}_{M \times 1} \qquad (5)$$

where the coefficient matrix $\mathbf{K}$ is diagonal. Inasmuch as the native structures are stabilized predominantly by homogeneous tertiary contacts rather than specific interactions [44] we take the entries of $\mathbf{K}$ to be equivalent in this work. Note that we validate this assumption by comparing the residue cross-correlations obtained from the simplified Hookean potential in equation 5 with those from all atom MD simulations (see the subsection **Molecular Dynamics simulations** below, and **Results and Discussion** for details).

Thus, rearranging equations 3–5, one gets the forces necessary to induce a given point-by-point displacement of residues:

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

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

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

where the $\Delta\mathbf{F}$ vector will contain the components of the externally applied force vectors on the selected residues. The $(\mathbf{BKB^T})$ matrix is equivalent to the Hessian [41] and its inverse has six zero eigenvalues, corresponding to the global translational and rotational degrees of freedom of the system. The elements of the inverse of the Hessian, $\mathbf{G} = \mathbf{H}^{-1}$, may be used to predict the auto- and cross-correlations of residues. $\mathbf{G}$ may be viewed as an $N \times N$ matrix whose ijth element is the 3×3 matrix of correlations between the x-, y-, and z-components of the fluctuations $\Delta\mathbf{R}_i$ and $\Delta\mathbf{R}_j$ of residues i and j; i.e.,

$$\mathbf{G}^{ij} = \begin{bmatrix} \langle \Delta X_i \Delta X_j \rangle & \langle \Delta X_i \Delta Y_j \rangle & \langle \Delta X_i \Delta Z_j \rangle \\ \langle \Delta Y_i \Delta X_j \rangle & \langle \Delta Y_i \Delta Y_j \rangle & \langle \Delta Y_i \Delta Z_j \rangle \\ \langle \Delta Z_i \Delta X_j \rangle & \langle \Delta Z_i \Delta Y_j \rangle & \langle \Delta Z_i \Delta Z_j \rangle \end{bmatrix} \qquad (8)$$

The cross-correlations between residue pairs are obtained from the trace of its components:

$$\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle = tr\left(\mathbf{G}^{ij}\right) \qquad (9)$$

Equation 9 has been shown to reproduce the cross-correlations obtained from MD simulations and molecular mechanics [9,23]. In this work, we shall not be directly interested in the correlations, but rather shall use $\mathbf{G}$ as a kernel to predict the response of other residues to applied perturbations on selected ones as we discuss next.

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 force $\Delta\mathbf{F}^i$ is $\left(\Delta F^i_x \; \Delta F^i_y \; \Delta F^i_z \right)$ so that the external force vector is constructed as

$$(\Delta\mathbf{F})^\mathbf{T} = \left\{ 0\,0\,0 \ldots \Delta F^i_x \, \Delta F^i_y \, \Delta F^i_z \ldots 0\,0\,0 \right\}_{1 \times 3N} \qquad (10)$$

The direction of the applied force vector deserves special attention. Here we choose the forcing direction randomly, attributing no bias due to the specific contact topology or the solvent exposed nature of the residue being perturbed. The forcing directions are uniformly distributed within a sphere enveloping the residue; therefore, the forcing may well be termed isotropic. It is definitely possible to favor specific directions leading to anisotropy in forcing, since there are no intrinsic constraints in the methodology dictating the opposite. A plausible forcing scenario for contact with a ligand, similar to one in [19] may also be conceived to determine the associated conformational changes. For specific applications, such as pulling experiments in which the application point and the direction of the forcing is identified (see, e.g. [45]), we can simulate the experimental loading conditions. If one does not explicitly know where such forces arrive at, however, the response of the different locations of the molecule under the action of a virtually applied isotropic force field is monitored by scanning the whole protein.

We then compute the resulting $(\Delta\mathbf{R})$ vector of the protein through equation 7, as we explain in detail in the following. Let the elements of $\mathbf{G}$ in equation 8 be g_{lm} where l and m denote the indices for the second order partial differential of the total energy with respect to the directionality (l, $m = x_j$, y_j or z_j; j is the residue index.) When a force is applied only at residue i, equation 7 in expanded form becomes:

$$\begin{bmatrix} g_{x_1x_1} & g_{x_1y_1} & g_{x_1z_1} & \cdots & \cdots & \cdots & g_{x_1x_N} & g_{x_1y_N} & g_{x_1z_N} \\ g_{y_1x_1} & g_{y_1y_1} & g_{y_1z_1} & \cdots & \cdots & \cdots & g_{y_1x_N} & g_{y_1y_N} & g_{y_1z_N} \\ g_{z_1x_1} & g_{z_1y_1} & g_{z_1z_1} & \cdots & \cdots & \cdots & g_{z_1x_N} & g_{z_1y_N} & g_{z_1z_N} \\ \vdots & \vdots & \vdots & & \cdots & & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & & \cdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & & & \vdots & \vdots & \vdots \\ g_{x_Nx_1} & g_{x_Ny_1} & g_{x_Nz_1} & \cdots & \cdots & \cdots & g_{x_Nx_N} & g_{x_Ny_N} & g_{x_Nz_N} \\ g_{y_Nx_1} & g_{y_Ny_1} & g_{y_Nz_1} & \cdots & \cdots & \cdots & g_{y_Nx_N} & g_{y_Ny_N} & g_{y_Nz_N} \\ g_{z_Nx_1} & g_{z_Ny_1} & g_{z_Nz_1} & \cdots & \cdots & \cdots & g_{z_Nx_N} & g_{z_Ny_N} & g_{z_Nz_N} \end{bmatrix}_{3N \times 3N} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \Delta F^i_x \\ \Delta F^i_y \\ \Delta F^i_z \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{3N \times 1} = \begin{bmatrix} \Delta R^1_x \\ \Delta R^1_y \\ \Delta R^1_z \\ \vdots \\ \vdots \\ \vdots \\ \Delta R^N_x \\ \Delta R^N_y \\ \Delta R^N_z \end{bmatrix}_{3N \times 1} \qquad (11)$$

The right-hand-side vector gives the displacement recorded in all the residues in response to a perturbation at a selected residue i,

$$\begin{bmatrix} \Delta R^1_x \\ \Delta R^1_y \\ \Delta R^1_z \\ \vdots \\ \vdots \\ \vdots \\ \Delta R^N_x \\ \Delta R^N_y \\ \Delta R^N_z \end{bmatrix} = \begin{bmatrix} \Delta F^i_x g_{x_1x_i} + \Delta F^i_y g_{x_1y_i} + \Delta F^i_z g_{x_1z_i} \\ \Delta F^i_x g_{y_1x_i} + \Delta F^i_y g_{y_1y_i} + \Delta F^i_z g_{y_1z_i} \\ \Delta F^i_x g_{z_1x_i} + \Delta F^i_y g_{z_1y_i} + \Delta F^i_z g_{z_1z_i} \\ \vdots \\ \vdots \\ \vdots \\ \Delta F^i_x g_{x_Nx_i} + \Delta F^i_y g_{x_Ny_i} + \Delta F^i_z g_{x_Nz_i} \\ \Delta F^i_x g_{y_Nx_i} + \Delta F^i_y g_{y_Ny_i} + \Delta F^i_z g_{y_Nz_i} \\ \Delta F^i_x g_{z_Nx_i} + \Delta F^i_y g_{z_Ny_i} + \Delta F^i_z g_{z_Nz_i} \end{bmatrix} \qquad (12)$$

Thus, a perturbation operates on the $3N{\times}3$ super-column of $\mathbf{G}$ that belongs to the residue being perturbed. The response on a specific residue k due to this perturbation on i is the vector $\Delta\mathbf{R}^{ki}$

$$
\Delta\mathbf{R}^{ki}=\begin{bmatrix} \Delta F_x^i g_{x_kx_i}+\Delta F_y^i g_{x_ky_i}+\Delta F_z^i g_{x_kz_i} \\ \Delta F_x^i g_{y_kx_i}+\Delta F_y^i g_{y_ky_i}+\Delta F_z^i g_{y_kz_i} \\ \Delta F_x^i g_{z_kx_i}+\Delta F_y^i g_{z_ky_i}+\Delta F_z^i g_{z_kz_i} \end{bmatrix}
$$
$$
=\begin{bmatrix} g_{x_kx_i} & g_{x_ky_i} & g_{x_kz_i} \\ g_{y_kx_i} & g_{y_ky_i} & g_{y_kz_i} \\ g_{z_kx_i} & g_{z_ky_i} & g_{z_kz_i} \end{bmatrix}\begin{bmatrix} \Delta F_x^i \\ \Delta F_y^i \\ \Delta F_z^i \end{bmatrix} \quad (13)
$$

We shall denote the $3{\times}3$ matrix of correlations between i and k by $\mathbf{G}^{ki}$ (see also equation 8).

One may further obtain a theoretical limit on the displacements, as we shall show below. The elements and properties of $\mathbf{G}^{ki}$ may be used to predict the average response of the system. First we apply a large collection, say r, of forces in random directions on residue i with mean zero and variance $\overline{F}^2$.

The average displacement on residue k in the x-direction is

$$
\langle\Delta R_x^{ki}\rangle=\frac{1}{r}\sum_r\left[g_{x_kx_i}\left(\Delta F_x^i\right)_r+g_{x_ky_i}\left(\Delta F_y^i\right)_r+g_{x_kz_i}\left(\Delta F_z^i\right)_r\right]
$$
$$
=\frac{1}{r}\left[g_{x_kx_i}\sum_r(\Delta F_x^i)_r+g_{x_ky_i}\sum_r(\Delta F_y^i)_r+g_{x_kz_i}\sum_r(\Delta F_z^i)_r\right] \quad (14)
$$
$$
=0
$$

since the application of forces in an uncorrelated fashion dictate $\sum_r\left(\Delta F_x^i\right)_r=0$ and similarly for the y- and z-directions. Here the average over any variable a is defined as $a=\sum_r a/r$. The average *magnitude* of the responses, on the other hand, is related to the applied forces and the elements of $\mathbf{G}$ through,

$$
\langle(\Delta\mathbf{R}^{ki})^2\rangle=\langle\Delta\mathbf{R}^{ki}\cdot\Delta\mathbf{R}^{ki}\rangle=\left\langle\left(\Delta F_x^i g_{x_kx_i}+\Delta F_y^i g_{x_ky_i}+\Delta F_z^i g_{x_kz_i}\right)^2\right\rangle
$$
$$
+\left\langle\left(\Delta F_x^i g_{y_kx_i}+\Delta F_y^i g_{y_ky_i}+\Delta F_z^i g_{y_kz_i}\right)^2\right\rangle \quad (15)
$$
$$
+\left\langle\left(\Delta F_x^i g_{z_kx_i}+\Delta F_y^i g_{z_ky_i}+\Delta F_z^i g_{z_kz_i}\right)^2\right\rangle
$$

In the expansion of the above expression, the average over any of the cross terms is zero, e.g.

$$
\left\langle\left(\Delta F_x^i g_{x_kx_i}\right)\left(\Delta F_y^i g_{x_ky_i}\right)\right\rangle=\left\langle\left(\Delta F_x^i\Delta F_y^i\right)\left(g_{x_kx_i}g_{x_ky_i}\right)\right\rangle
$$
$$
=\left(g_{x_kx_i}g_{x_ky_i}\right)\left\langle\Delta F_x^i\Delta F_y^i\right\rangle=0 \quad (16)
$$

Here, the elements of $\mathbf{G}$ factor out, as they are constants. Moreover, $\left\langle\Delta F_x^i\Delta F_y^i\right\rangle=0$ since the forces are applied in an

uncorrelated manner. The self terms, on the other hand, survive to yield:

$$
\left\langle(\Delta\mathbf{R}^{ki})^2\right\rangle=\left\langle\left(\Delta F_x^i\right)^2\right\rangle\left[(g_{x_kx_i})^2+(g_{x_ky_i})^2+(g_{x_kz_i})^2\right]
$$
$$
+\left\langle\left(\Delta F_y^i\right)^2\right\rangle\left[(g_{y_kx_i})^2+(g_{y_ky_i})^2+(g_{y_kz_i})^2\right] \quad (17)
$$
$$
+\left\langle\left(\Delta F_z^i\right)^2\right\rangle\left[(g_{z_kx_i})^2+(g_{z_ky_i})^2+(g_{z_kz_i})^2\right]
$$

with $\left\langle\left(\Delta F_x^i\right)^2\right\rangle=\left\langle\left(\Delta F_y^i\right)^2\right\rangle=\left\langle\left(\Delta F_z^i\right)^2\right\rangle=\overline{F}^2/3$ we have

$$
\left\langle(\Delta\mathbf{R}^{ki})^2\right\rangle=\frac{\overline{F}^2}{3}\sum_{l=x_k,y_k,z_k}\sum_{m=x_i,y_i,z_i}(g_{lm})^2 \quad (18)
$$

Thus, the average expected displacement of a residue k due to the bombardment of another residue i may be calculated from the sum of the squares of the elements of the $\mathbf{G}^{ki}$ matrix.

Correlations between predicted and experimental x-ray structures. Using PRS, we scan the protein by consecutively perturbing each residue, and record the associated displacements as a result of the linear response of the protein. We define ΔR_i^k as the theoretical predictions of the displacement of each residue, k, as a response of the system to applied forces on residue i (equation 13). Similarly, ΔS_k are the displacements between the apo and the holo forms obtained from the PDB structures (these are also referred to as the experimental conformational changes or experimental displacements throughout the text). Thus, ΔR_i^k are compared with ΔS_k and the goodness of the prediction is quantified as the Pearson correlation coefficient for each perturbed residue i:

$$
C_i=\frac{\sum_{j=1}^N\left[(\Delta R_k)^i-(\overline{\Delta R})^i\right](\Delta S_k-\overline{\Delta S})}{(N-1)\sigma_R\sigma_S} \quad (19)
$$

where the overbar indicates the average. σ_R and σ_S are the respective standard deviations of calculated and experimental displacements. Gerstein and coworkers have demonstrated that when comparing two structures, the results from a selected subset of the residues may be more informative [46]. Thus, in calculating the correlations with the x-ray experiments, first a best-fit between the fixed domains of the predicted and the experimental structures are made. This is due to a generally used convention to interpret displacements in multi-domain proteins so as to accentuate the response of the moving domain to see the biological relevance of the observed changes. These calculations are repeated over at least five scans. The overall PRS scheme is summarized in Algorithm A, shown in Figure 2.

Analysis on the directionality of the response. 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, may be distributed along a line or in a plane so that the net response is still zero. Thus, although the perturbations are isotropic, the response may well be anisotropic. 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 **Results and Discussion** section.

For an analysis that probes the directionality of the recorded responses, we proceed as follows: We first concentrate on those

Begin PRS with $i = 1$

Apply random force on residue i; $\Delta \mathbf{F}^i = (\Delta F_x{}^i, \Delta F_y{}^i, \Delta F_z{}^i)$:
$(\Delta \mathbf{F})^{\mathrm{T}} = \{0\ 0\ 0\ \ldots\ \Delta F_x{}^i\ \Delta F_y{}^i\ \Delta F_z{}^i \ldots\ 0\ 0\ 0\}$

Compute the displacement vector $\Delta \mathbf{R}$ in response to $\Delta \mathbf{F}$ (eq. 7)

Repeat on residue i for m independent runs

Compute the average magnitude of the displacements for all residues j due to the forces introduced on i, $(\Delta R_j)^i$

Record the Pearson correlation, C_i, between the average displacements and experimental conformational change

$i = i + 1$

Exit PRS for $i > N$

Figure 2. Algorithm A describing the overall PRS scheme.
doi:10.1371/journal.pcbi.1000544.g002

residues for which the Pearson correlation between experimental and theoretical displacements is large. Amongst them, we further locate those residues, i, that are distant from the ligand binding site, l, (i.e., the distance between them, $d_{il} \gg r_c$). For the selected residue, i, k forces are applied such that $\Sigma_k \Delta F_k{}^i = 0$; k is large and ensures that a spherically symmetric region around i is covered. 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 the subsection **Local and remote modulation of Fe ion dissociation**).

One may further analyze the eigenvalue structure of the $\mathbf{G}^{ki}$ submatrix to have an understanding on the nature of the response by decomposing it using the transformation

$$\mathbf{G}_{3 \times 3}^{ki} = \mathbf{U}_{3 \times 3} \Lambda_{3 \times 3} \mathbf{U}_{3 \times 3}^T \qquad (20)$$

The three orthogonal vectors defined by $\mathbf{U}$, $\mathbf{u}_j$, give the three principal axes of the line of action of residue k in response to perturbations in i. The size of the associated elements, λ_j, provide the contribution, $p_j = \lambda_j / \sum_1^3 \lambda_j$, of each vector to the overall response to the perturbation $\Delta \mathbf{F}^i$ (equation 13). Thus, if there is one dominant eigenvalue in $\mathbf{G}^{ki}$, i.e., $p_1 \approx 1$, then no matter what the values of the elements of $\Delta \mathbf{F}^i$, they will be projected on the associated eigenvector, $\mathbf{u}_1$. Therefore, the collection of responses $\Delta \mathbf{R}^{ki}$ to a number of perturbations will all be collected in a line along $\mathbf{u}_1$. Similarly, if two eigenvalues dominate, i.e. $p_1 + p_2 \approx 1$, then the collection of responses will occur in the plane defined by $\mathbf{u}_1$ and $\mathbf{u}_2$.

The degree of collectivity of the response of a group of neighboring residues to a perturbation on i may also be measured using equation 20. If the response of the neighbors possesses collectivity, then various symmetries in their action may be expected. For example, if the residues collectively move in a line to open a cap, not only each is expected to have a single dominant eigenvalue, but also the eigenvectors belonging to these dominant

eigenvalues are to be parallel; i.e.,

$$\mathbf{u}_1^{k_1 i} \cdot \mathbf{u}_1^{k_2 i} = \cos \theta \approx 1 \qquad (21)$$

where θ is the angle between the two eigenvectors, and k_1 and k_2 are the two residues whose responses are being compared. The directionality analysis is summarized in Algorithm B, shown in Figure 3.

Systems studied and residue network construction. We analyze FBP in detail, using both PRS and MD. In addition, the PRS methodology is applied to several other cases to demonstrate how the approach may be used. These are included in Text S1.

The apo and holo forms of FBP have PDB codes 1D9V and 1MRP, respectively; in the latter case the Fe ion 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 4a. 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 by equation 7 or 11. 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 19 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 19) are lower in value, but their relative ordering does not change; for example, the largest correlations in the apo form reduce from 0.98 to 0.90 and those in the apo form reduce from 0.89 to 0.80.

As a cut-off distance, we seek the smallest value for which the experimental B-factors are well-reproduced (equation 9 with $i = j$) and the results are the same over a range of values. We find that a cut-off distance of $r_c = 8.0$ Å on the C_α atoms of the protein (equation 1) is suitable. 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; the system exhibits six zero eigenvalues at all these cut-off distances. At $r_c = 8.0$ Å, there are 1542 and 1587 interactions for the apo and holo initial conformations, respectively.

for a selected residue i with
$C_i >$ threshold and $d_{il} \gg r_c$

Apply random force on residue i; $\Delta \mathbf{F}^i = (\Delta F_x^i, \Delta F_y^i, \Delta F_z^i)$:
$(\Delta \mathbf{F})^T = \{0\ 0\ 0 \ldots \Delta F_x^i\ \Delta F_y^i\ \Delta F_z^i \ldots 0\ 0\ 0\}$

Repeat on residue i
for k independent
directions

Compute the displacement vectors $\Delta \mathbf{R}$ in response to $\Delta \mathbf{F}^i$ (eq. 7)

Plot all collected vectors on the protein structure
and examine their directionality
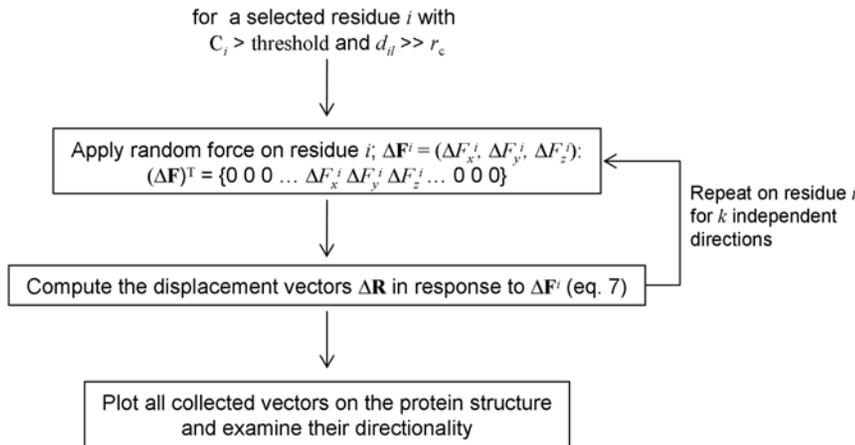
Figure 3. Algorithm B describing the directionality analysis.
doi:10.1371/journal.pcbi.1000544.g003

Molecular dynamics simulations. To provide a basis for comparison with results from an all-atom approach, we have performed MD simulations on both the apo and the holo forms of FBP in water. The systems are prepared using the VMD 1.8.6 program with solvate plug-in version 1.2.[47] The NAMD package is used to model the dynamics of the protein – water systems.[48] The apo form is neutral with 309 amino acids, and an exogenous phosphate ion. The holo form additionally has the Fe^{+3} ion. Three chloride ions are added to achieve charge neutrality in the latter system. The protein is soaked in a solvent box such that
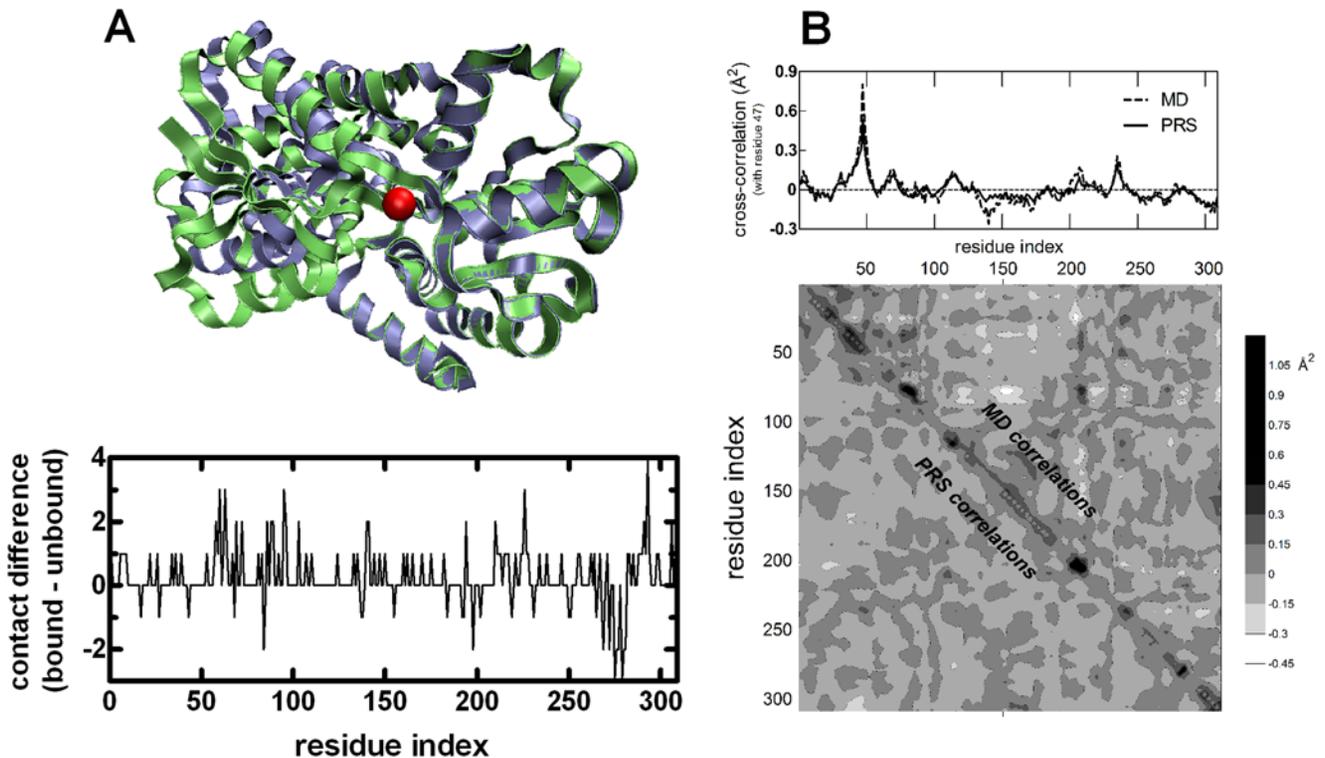
Figure 4. Analysis of ferric-binding protein. A Upper panel displays the haemophilus influenzae ferric-binding protein in apo (purple; PDB code: 1D9V) and holo forms (green; PDB code: 1MRP). The two structures are superimposed on the fixed domain (residues 83–87, 102–225, 277–307). The Fe^{3+} ion is shown as a red sphere. 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. **B** Contour map comparing the residue cross-correlations obtained from the detailed MD simulations of 10 ns duration (upper-right part of the map) and the coarse-grained model (lower-left part of the map) for holo-FBP. White and light gray are negative correlations, dark gray and black are positive correlations. Also displayed at the top of the figure are the detailed cross-correlations of residue 47 from the coarsened protein model (solid lines) and MD simulations (dashed lines). The value of zero is marked for each case to better discriminate positively and negatively correlated regions. The Pearson correlations between the data are 0.74 for the whole map and 0.94 for the selected residue, respectively.
doi:10.1371/journal.pcbi.1000544.g004

there is at least a 5 Å layer of solvent in each direction from any atom of the protein to the edge of the box. The simulated protein-water complexes have 5156 and 5368 water molecules, respectively for the apo and holo forms. The CharmM27 force field parameters are used for protein and water molecules [49].

The binding site parameters are as follows: The exogenous phosphate is modeled in the $H_2PO_4^-$ state using the parameters reported in detail in the literature [50]. For the Fe^{+3} ion, an effective van der Waals interaction term in addition to electrostatics is included in the spirit done for other ions in the literature [51]. Since the parameters for Fe^{+3} do not appear in literature, we have self-consistently parameterized them so that the six liganded coordination within 2.0 ± 0.2 Å average distance of the ion [52] is maintained after energy minimization and 200 ps long MD simulations. The optimal values of the Lennard-Jones parameters were found to be -0.1 kcal/mol for well-depth and 2.6 Å for the separation at the minimum.

Long range electrostatic interactions were calculated using particle mesh Ewald (PME) method [53]. The cutoff distance for non-bonded van der Waals interactions was set to 12 Å with a switching function cutoff of 10 Å. Rattle algorithm was used to fix the bond lengths to their average values. During the simulations, periodic boundary conditions were used and the equations of motion were integrated using the Verlet algorithm with a step size of 2 fs [54]. Temperature control was carried out by Langevin dynamics with a dampening coefficient of 5/ps and pressure control was attained by a Langevin piston. Volumetric fluctuations were preset to be isotropic in the NPT runs.

Both systems were first subjected to energy minimization with the conjugate gradients algorithm until the gradient tolerance was less than 10^{-2} kcal/mol/Å. 500 ps MD runs in the NVT ensemble at 310 K were carried out on the resulting systems. The final structures were then run in the NPT ensemble at 1 atm and 310 K until volumetric fluctuations were stable to maintain the desired average pressure. This process required 500 ps long MD runs at the end of which the average volume is maintained at 196900 ± 700 and 203300 ± 600 Å^3 in the apo and holo structure runs, respectively. Finally, the runs in the NPT ensemble were extended to a total of 10 ns. The coordinate sets were saved at 2 ps intervals for subsequent analysis.

The RMSD of the trajectories were calculated (Figure S1). For the holo form stabilized by the Fe ion, equilibration is reached within 500 ps, with the value 1.4 ± 0.1 Å averaged over the remaining 9.5 ns trajectory. On the other hand, the RMSD of the apo form displays larger fluctuations, and the point of equilibration is harder to judge for this form. The RMSD is 1.6 ± 0.3 Å averaged over the last 9.5 ns trajectory, and 1.8 ± 0.2 averaged over the last 5.0 ns. The results presented in this manuscript are those obtained from the last 5.0 ns of the simulations for both forms of FBP. Note that, we have repeated the analyses for the last 9.5 ns of the trajectories, without any significant difference in the results.

The correlations between residue pairs derived from the MD trajectories are of particular interest. The snapshots recorded during the MD simulations are organized in the fluctuation trajectory matrix of order $3N\times T$, [23]

$$\Delta\mathbf{R} = \begin{bmatrix} \Delta\mathbf{R}_1(t_1) & \Delta\mathbf{R}_1(t_2) & \cdot & \Delta\mathbf{R}_1(t_T) \\ \Delta\mathbf{R}_2(t_1) & \Delta\mathbf{R}_2(t_2) & \cdot & \Delta\mathbf{R}_2(t_T) \\ \Delta\mathbf{R}_3(t_1) & \Delta\mathbf{R}_3(t_2) & \cdot & \Delta\mathbf{R}_3(t_T) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \Delta\mathbf{R}_N(t_1) & \Delta\mathbf{R}_N(t_2) & \cdot & \Delta\mathbf{R}_N(t_T) \end{bmatrix} \quad (22)$$

The $3N\times3N$ correlation matrix is then calculated by $\mathbf{A} = \Delta\mathbf{R}\Delta\mathbf{R}^T$, where T is the transpose. The cross-correlation between residue pairs are then computed through equations 8 and 9, where $\mathbf{A}$ substitutes for $\mathbf{G}$.

Note that a each nanosecond of an MD simulation takes ca. 9.8 hours on a server with 2 GB memory and eight CPUs each with 2.4 GHz quadcore architecture. Five complete scans of the protein with the PRS method uses three minutes on a single CPU of the same server.

Results

Based on linear response theory, we apply perturbations at selected points along the chain to diagnose the response of 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 4a the two structures are superimposed on the fixed domain residues: purple apo-, green holo-FBP). The difference in the contact numbers of the two forms is shown in the bottom panel of figure 4a. 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. The Fe ion brings in only 11 of the new contacts. 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 4a). The conformational change is mainly governed by a hinge motion around the Fe-binding location; yet, the two structures have distinct features. 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.

Validation of the coarse grained model. Since the methodology employs the Hessian obtained from a coarse-grained potential in equation 7, we first determine the extent to which this potential represents the actual interactions between residue pairs. We compare the correlations obtained from the network model with those from the MD simulations for both the apo and the holo forms (by applying equations 8 and 9 for the former, and equations 22, 8, and 9 for the latter). We find that the coarsened protein model reproduces the residue-wise MD correlations; the Pearson correlations between the data are 0.76 and 0.74, respectively for the apo and holo forms. In Figure 4b we display a comparative map of the correlations obtained by the two approaches for holo FBP. A cross-section through a sample residue, Asp47 that resides on a flexible loop on the surface of the moving domain, is also displayed at the top of the figure. The two approaches agree well, including the regions of negative correlations, which lie below the solid and dashed lines for the respective cases of coarse-grained and MD results, and corroborate the results in reference 19. Note that therein equation 7 was computed for perturbations introduced to Glu57 only. In the next subsection, we elaborate on the results obtained for this residue, which directly contacts the ion in the holo structure.

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 5. It was previously shown that applying forces on residue 57 near the ferric binding site yields the expected atomic displacements.[19] 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. Example cases of the computed displacements using equation 7 are shown in the middle and
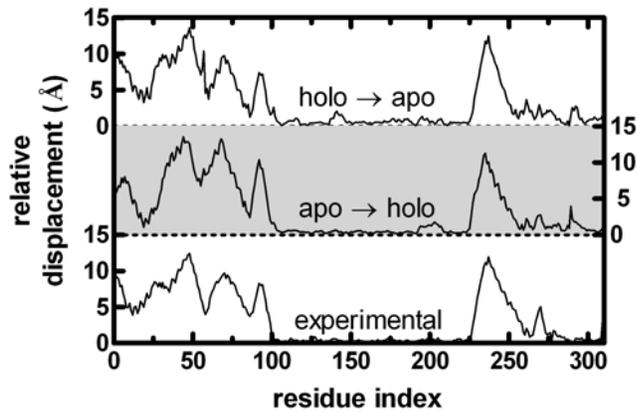
Figure 5. Relative displacements of residues between the apo and holo forms (x-ray), and typical responses to a given force perturbation on residue 57 in the apo and holo forms. The latter two curves are nudged for ease of comparison; their baselines are shown by dashed lines. Since the calculated displacement is proportional to the imposed force in LRT and therefore may be rescaled by a proportionality constant, the magnitude of the force is adjusted so as to make the average displacement the same as that of the x-ray experiment.
doi:10.1371/journal.pcbi.1000544.g005

upper curves of Figure 5. Here, a single perturbation on residue 57 is placed in a randomly chosen direction on the apo and holo structures, respectively, and the fixed domains of the resulting structures are superimposed. The correlation coefficient, C_{57} (equation 19), between these curves and the experimental curve is 0.95 and 0.92, respectively for these example cases. The close agreement of the residue-by-residue displacements between the current methodology and the all-atom approach in reference 19 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). We next investigate whether residue 57 is unique in reproducing the conformational response of the protein by performing PRS.

Non-reciprocal response to localized perturbations throughout the protein uncovers landscape properties. We now perform the residue-by-residue scan on the protein. Starting from each of the initial conformations, (i) apo form and (ii) 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 as outlined in the subsection **Correlations between predicted and experimental structures**. In figure 6 we display the resulting correlation coefficients C_i between the calculated and the experimental data for every residue (equation 19). 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 applied at the selected residue, *i*, such as that obtained from the correlation between the middle and bottom curves in figure 5. We find that there is a plethora of residues whose perturbation leads to the holo structure when applied on the apo structure, whereby *all* applied forces lead to displacements well correlated with those from x-ray structures, with the worst perturbation having a correlation of 0.6 ± 0.1 (figure 6a). 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. When we perform PRS on the holo form as the input structure, only 24 out of 309 residues (8%) that perform better than 0.90 correlation remain (figure 6b). To display the reduction in the number of residues that
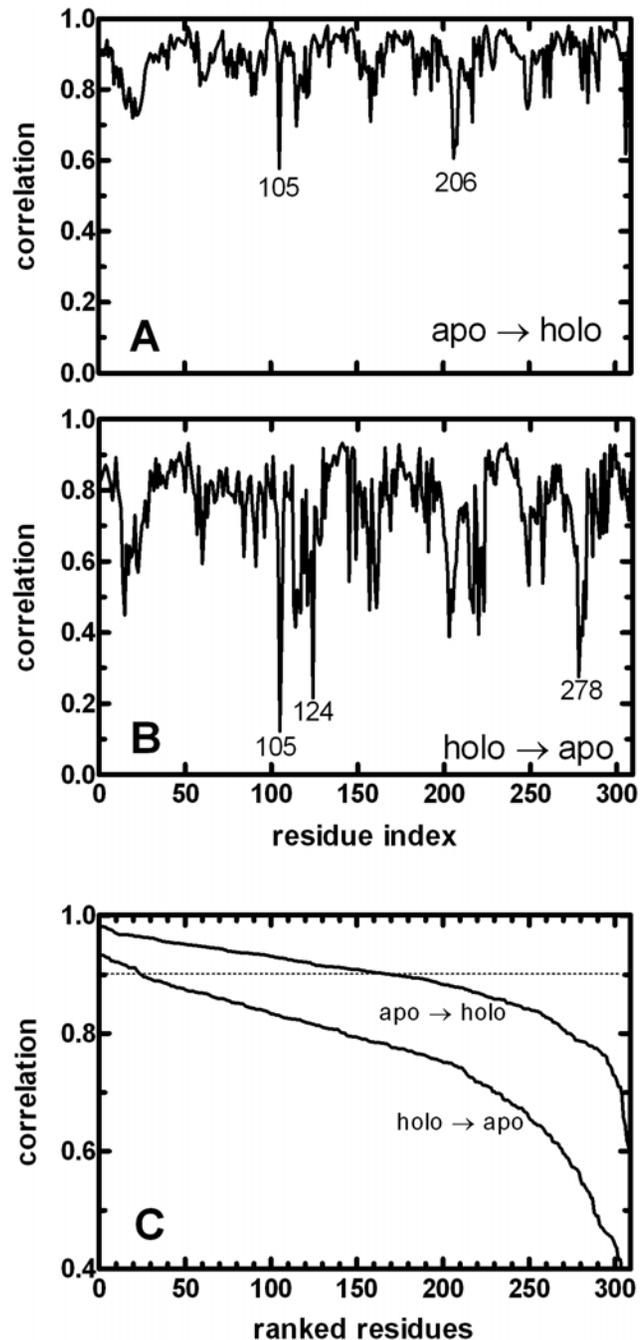
Figure 6. Displacement vectors between the perturbed and initial structures of ferric-binding protein. The displacements are compared with those of the crystal structures, for **A** the apo, and **B** the holo form as the initial structure. In **C**, data in **A** and **B** are sorted from larges to smallest. Residues 47, 52, 130, 139–144, 147, 148, 166, 174, 186, 226, 232–236, 293, 296, 298, 299 give the highest and 105, 124, 278 the lowest correlations in the holo structure. The standard error on the mean increases with decreasing correlations, as determined from the averages of five perturbations in randomly selected directions. For correlations greater than 0.90, it is less than 0.05 for the apo, and 0.02 for the holo form; the largest errors on the mean are 0.2 for the least correlated values in both cases.
doi:10.1371/journal.pcbi.1000544.g006

lead to favorable conformational change upon perturbation, the correlations in figure 6a and b are ranked according to their size, and are plotted in figure 6c.

In perturbing the residues of the apo FBP, 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.

In perturbing the residues of the holo FBP, the number of residues that leads to a highly correlated ΔR profile with the crystal structure is reduced. The residues whose perturbation leads to displacements with the largest C_i [i.e., those that lie above the dotted line] and the lowest C_i are listed in the caption to Figure 6. All of the latter are located in the fixed domain. The Fe ion is known to bind FBP with very high affinity, assisted by the coordinating phosphate ion. 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. Direct force application at the Fe ion itself leads to an average correlation coefficient of 0.88.

The high C_i values resulting from the perturbation of many residues on the apo structure merely suggest that the 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. The existence of a weakly populated, higher energy conformation that is stabilized in the presence of the ligand, so that the equilibrium is shifted towards this minor species has been experimentally observed,[40] supporting the "conformational selection" hypothesis.[38] This viewpoint of "shifts in the energy landscapes" was put championed a decade ago by Nussinov and coworkers [39,55]. 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 6a, most of the perturbations applied at a variety of residues distributed throughout the protein lead to very high correlations with the displacements obtained from x-ray structures. 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. In the next subsection, we rely on the directionality information provided by the PRS to uncover the mechanisms that dictate protein function.

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. 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 coupling between the motions of subsets of nodes.

In the previous subsection we have shown that in holo FBP, singly placed forces on the residues listed under the caption to figure 6 reproduce the displacement profile from x-ray structures with very high 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 [46]. In fact, we find that overlap distributions over all the residues are close to those of a random distribution of angles (results not shown). We therefore perform the following alternative analysis (see Algorithm B in Figure 3): For a selected residue, for instance, 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 directional response is also analyzed analytically by applying equations 20 and 21.

The residues for which $C_i > 0.9$ and are far from the Fe ion ($d_{i\text{-Fe}} \gg r_c$) 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 7a, where the applied forces are in red and the responses are in orange. We observe that 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; three are located in the moving domain (residues 8, 9 and 57), and the remaining are in the fixed domain. The volume they take up in the protein is shaded in figure 7a. This region is magnified in figure 7b, where the recorded responses to selected perturbations are shown. The upper left panel of figure 7b focuses on the data in figure 7a to show that the moving domain residues contacting the Fe ion operate in a coherent fashion at the tip of the cap that opens the exit of the Fe ion. Here, we define coherence as the tendency of residues to move along parallel lines. The local eigenvector analysis described in the Methods section (equations 20 and 21) confirms that upon perturbing these highly correlating residues, the inverse of the Hessian operates on a single dominant eigenvalue ($p_1 > 0.8$) whose associated eigenvectors are parallel to each other (cos $\theta > 0.95$). 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). We also perform the same analysis to the correlation matrix obtained from MD simulations (equations 22, 8 and 9). The results are summarized in Table 1 for perturbations on residues 47 and 52 where both PRS and MD predict coherent response with a single dominant eigenvalue and parallel eigenvectors.

On the other hand, directly perturbing the Fe ion, as well as other local 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 a much wider range of directions as exemplified by perturbing Fe ion and residue 57 in the right hand side panels of figure 7b. These do not have dominant eigenvalues in the cap residues (Table 1) and the angles between the eigenvectors that belong to the largest eigenvalue of pairs of residues
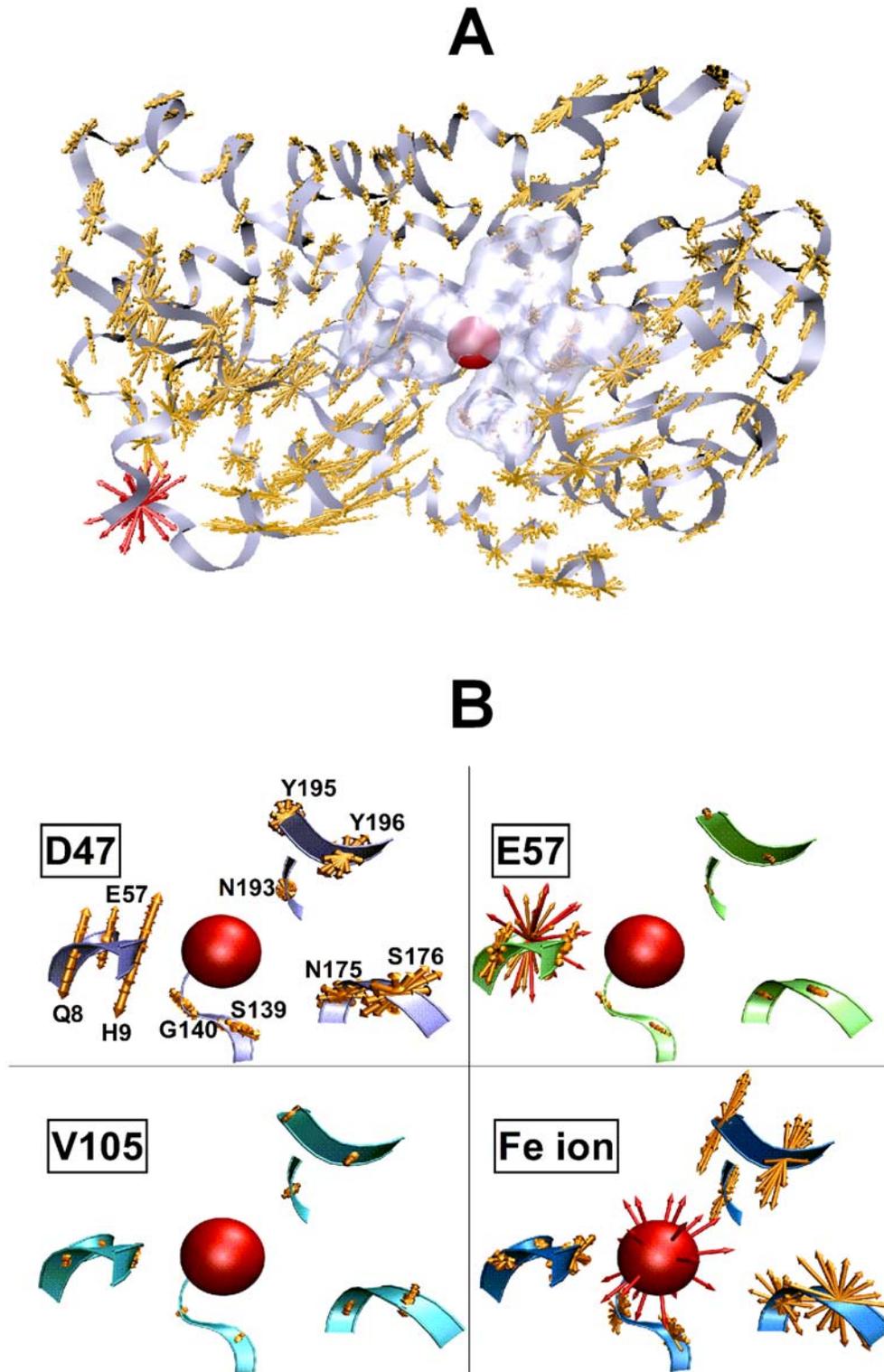
Figure 7. Remote modulation of Fe ion dissociation. A The response of the protein to forces exerted on residue $j = 47$ in different directions (forces are shown with the red arrows.) The collection of all the displacements is shown in orange; responses on the three neighbors along the chain in either direction ($44 \leq j \leq 50$) are not shown for clarity. The volume taken up by the first neighbors of the Fe^{+3} is shaded. **B** The region of the first neighbors of the Fe^{+3} is magnified to display the response to perturbations applied at selected residues. All vector lengths are relative to the magnitude of the same size of unit response. **Upper left** figure magnifies the results in **A**. The Fe^{+3} ion contacting residues in the fixed domain (located on the right hand side) respond incoherently, moving mainly within a plane. Residues Q8, H9 and E57, 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. Perturbations on regions in the binding domain (exemplified by **E57** and **Fe ion**) also induce large changes in this region, the former in the moving domain, the latter in both domains, but coherence in cap residues no longer exists. Distant, non-controlling residues (e.g. **V105**) cannot induce large enough motions in this region, although coherence of the cap residues may exist (see Table 1). Figures are drawn with the VMD software [47].
doi:10.1371/journal.pcbi.1000544.g007

Table 1. Response of cap residues 8, 9, and 57 to perturbations on selected residues.

perturbed residue, i	fractional contribution of the highest eigenvalue p_1 of the response matrix G^{ki} (k: 8/9/57)[†]		angle θ between average response vectors of residue pairs 8–9/8–57[‡]	
	PRS	MD	PRS	MD
D47	0.95/0.92/0.81	0.78/0.83/0.60	9/16	5/28
D52	0.88/0.90/0.81	0.91/0.95/0.79	9/14	4/18
E57	0.54/0.56/0.50	0.77/0.86/0.57	36/38	4/16
Fe ion	0.56/0.67/0.54	0.74/0.81/0.59	25/36	5/28
V105	0.73/0.77/0.62	0.89/0.88/0.92	7/11	6/19

[†]As computed from equation 20.
[‡]Computed through equation 21; relative response of cap residues in the moving domain, k, to perturbations in residue i is coherent if θ is small and p_1 is large.
doi:10.1371/journal.pcbi.1000544.t001

8, 9 and 57 are larger than $30°$, making the response in this region far from coherent. Analyses of MD results predict similar coherence in the correlations amongst the binding region residues compared to those between this region and the distant controlling sites.

In the lower left panel of Figure 7b, we display the response of the binding region residues to a collection of perturbations on residue 105, which gives the lowest correlations in holo FBP (recall Figure 6c). We find that the magnitude of the induced response is insignificant, although the cap residues act more coherently compared to local perturbations, as shown in Table 1; MD correlations also corroborate this tendency.

Thus, by jointly focusing on the specific distantly located residues which (i) invoke a large amount of correlations in the whole protein, and (ii) induce local cooperativity in the binding domain, one may be able to uncover sites that remotely control function in the protein. Interestingly, the same analysis carried out on the N-lobe of human transferrin implies K278 in that protein as an allosteric controller (see Text S1, section E). This residue resides in a bulging loop, and does not have a structurally aligned counterpart on bacterial FBP. These observations open the way to the design of agents that specifically target the pathogen, while the functionality of the host is not altered.

Discussion

In this study we introduce the PRS methodology based on systematically perturbing all residues of a protein, and classifying both the magnitude and the directionality of the recorded linear response. The approach is unique in that the cross-correlations between residues are processed as input for further predictions on the system behavior, unlike other methods where they are the final results. Closed-form expressions for the magnitude and the directionality of the response are also provided. The protein FBP is studied in detail, and results using PRS for additional cases are presented in Text S1. All computer programs used in the analyses are available upon request. The findings on FBP are also supported by MD simulations.

For the particular case of FBP, our analysis suggest the existence of two alternative mechanism of Fe ion release in FBP: (i) Local control of the ion by synergistic anions and chelators acting in the binding groove, and (ii) remote control by ions acting on distant charged residues located in solvent exposed loops (*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 [25,27,28]. 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.[29] 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 [27]. Allosteric anion binding sites that trigger large conformational changes located at the surface [10,36,37] have been 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 [32]. Structural alignment [56] 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 (figure 6).

The results resolve the so-called Fe^{+3} transport dilemma: The protein is ready to uptake the ion in the apo structure (figure 6a), but it is necessary to perturb highly specific locations along the chain for its release in holo FBP (figures 6b and 7). This mode of action provides a mechanism for recent NMR observations of ligand binding whereby the energy landscape of the free protein is made-up of a set of coupled low-free energy states [5,40,57,58]. Therein, ligand binding is considered as the source for shifts in the landscapes [38,39]. PRS confirms this view for the particular case of FBP, and further provides the mechanisms on how the ligand binding region may be manipulated, as outlined by scenarios (i) and (ii).

In summary, the PRS methodology introduces an efficient approach to determine regions in the protein that mechanically moderate binding region motions. It may therefore be used to determine candidate sites for mutational studies. In a forthcoming study, the biological implications from the results of PRS will be presented on set of twenty proteins that display various types of motions such as shear, hinge, allosteric, partial refolding as well as more complex protein motions, as classified in the Database of Macromolecular Motions [59].

Although studies such as the current one help estimate allosteric sites, they do not provide information on pathways along which the remote communication takes place. As future work, it is of interest to locate these paths. Robust techniques to predict them using evolutionary information [60] or the specificity of the

interactions within the residue networks [61] have been developed. Cross-correlation information from MD simulations have been used with residue network properties [62] to generate information on remote communication pathways [60,63,64]. Complemented with point mutation studies, such analyses will not only aid the protein design process, but will also uncover the physics of remote communication between residues.

Supporting Information

Figure S1 The RMSD values of the 10 ns long MD trajectories. The holo form is stabilized by the ligand, while the apo form displays larger fluctuations.

Found at: doi:10.1371/journal.pcbi.1000544.s001 (0.10 MB DOC)

Text S1 Sample applications of the PRS method.
Found at: doi:10.1371/journal.pcbi.1000544.s002 (1.45 MB DOC)

Author Contributions

References

1. Zhang XJ, Wozniak JA, Matthews BW (1995) Protein Flexibility and Adaptability Seen in 25 Crystal Forms of T4 Lysozyme. Journal of Molecular Biology 250: 527–552.
2. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, et al. (2005) Intrinsic dynamics of an enzyme underlies catalysis. Nature 438: 117–121.
3. Gunasekaran K, Ma BY, Nussinov R (2004) Is allostery an intrinsic property of all dynamic proteins? Proteins-Structure Function and Bioinformatics 57: 433–443.
4. Volkman BF, Lipson D, Wemmer DE, Kern D (2001) Two-state allosteric behavior in a single-domain signaling protein. Science 291: 2429–2433.
5. Tang C, Schwieters CD, Clore GM (2007) Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. Nature 449: 1078–1082.
6. Zhuravleva A, Korzhnev DM, Nolde SB, Kay LE, Arseniev AS, et al. (2007) Propagation of dynamic changes in barnase upon binding of barstar: An NMR and computational study. Journal of Molecular Biology 367: 1079–1092.
7. Tsai CJ, del Sol A, Nussinov R (2008) Allostery: Absence of a change in shape does not imply that allostery is not at play. Journal of Molecular Biology 378: 1–11.
8. Tsai CJ, Ma BY, Nussinov R (1999) Folding and binding cascades: Shifts in energy landscapes. Proceedings of the National Academy of Sciences of the United States of America 96: 9970–9972.
9. Baysal C, Atilgan AR (2001) Coordination topology and stability for the native and binding conformers of chymotrypsin inhibitor 2. Proteins-Structure Function and Genetics 45: 62–70.
10. Hamilton DH, Turcot I, Stintzi A, Raymond KN (2004) Large cooperativity in the removal of iron from transferrin at physiological temperature and chloride ion concentration. Journal of Biological Inorganic Chemistry 9: 936–944.
11. Blaber M, Baase WA, Gassner N, Matthews BW (1995) Alanine Scanning Mutagenesis of the Alpha-Helix-115-123 of Phage-T4 Lysozyme - Effects on Structure, Stability and the Binding of Solvent. Journal of Molecular Biology 246: 317–330.
12. Min W, English BP, Luo GB, Cherayil BJ, Kou SC, et al. (2005) Fluctuating enzymes: Lessons from single-molecule studies. Accounts of Chemical Research 38: 923–931.
13. Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, et al. (2007) Intrinsic motions along an enzymatic reaction trajectory. Nature 450: 838–844.
14. Prytkova TR, Kurnikov IV, Beratan DN (2007) Coupling coherence distinguishes structure sensitivity in protein electron transfer. Science 315: 622–625.
15. Velyvis A, Yang YR, Schachman HK, Kay LE (2007) 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 104: 8815–8820.
16. Atilgan C, Aviyente V (2007) Hybrid usage of computational tools in drug synthesis. Current Computer-Aided Drug Design 3: 149–159.
17. Wells JA, McClendon CL (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature 450: 1001–1009.
18. Atilgan C, Aykut AO, Atilgan AR (2008) How a vicinal layer of solvent modulates the dynamics of proteins. Biophysical Journal 94: 79–89.
19. Ikeguchi M, Ueno J, Sato M, Kidera A (2005) Protein structural change upon ligand binding: Linear response theory. Physical Review Letters 94: 078102.
20. Latzer J, Shen T, Wolynes PG (2008) Conformational switching upon phosphorylation: A predictive framework based on energy landscape principles. Biochemistry 47: 2110–2122.
21. Sacquin-Mora S, Lavery R (2009) Modeling the Mechanical Response of Proteins to Anisotropic Deformation. Chemphyschem 10: 115–118.
22. Zheng WJ, Brooks B (2005) Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of an elastic network model. Journal of Molecular Biology 346: 745–759.
23. Baysal C, Atilgan AR (2001) Elucidating the structural mechanisms for biological activity of the chemokine family. Proteins-Structure Function and Genetics 43: 150–160.
24. Chen CY, Berish SA, Morse SA, Mietzner TA (1993) The Ferric Iron-Binding Protein of Pathogenic Neisseria Spp Functions as a Periplasmic Transport Protein in Iron Acquisition from Human Transferrin. Molecular Microbiology 10: 311–318.
25. Dhungana S, Anderson DS, Mietzner TA, Crumbliss AL (2005) Kinetics of iron release from ferric binding protein (FbpA): Mechanistic implications in bacterial periplasm-to-cytosol Fe3+ transport. Biochemistry 44: 9606–9618.
26. Khan AG, Shouldice SR, Kirby SD, Yu RH, Tari LW, et al. (2007) High-affinity binding by the periplasmic iron-binding protein from Haemophilus influenzae is required for acquiring iron from transferrin. Biochemical Journal 404: 217–225.
27. Boukhalfa H, Anderson DS, Mietzner TA, Crumbliss AL (2003) 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 8: 881–892.
28. Heymann JJ, Weaver KD, Mietzner TA, Crumbliss AL (2007) 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 129: 9704–9712.
29. Khan AG, Shouldice SR, Tari LW, Schryvers AB (2007) The role of the synergistic phosphate anion in iron transport by the periplasmic iron-binding protein from Haemophilus influenzae. Biochemical Journal 403: 43–48.
30. Dhungana S, Taboy CH, Anderson DS, Vaughan KG, Aisen P, et al. (2003) 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 100: 3659–3664.
31. Wally J, Halbrooks PJ, Vonrhein C, Rould MA, Everse SJ, et al. (2006) The crystal structure of iron-free human serum transferrin provides insight into inter-lobe communication and receptor binding. Journal of Biological Chemistry 281: 24934–24944.
32. Adams TE, Mason AB, He QY, Halbrooks PJ, Briggs SK, et al. (2003) 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 278: 6027–6033.
33. He QY, Mason AB, Nguyen V, MacGillivray RTA, Woodworth RC (2000) The chloride effect is related to anion binding in determining the rate of iron release from the human transferrin N-lobe. Biochemical Journal 350: 909–915.
34. Mizutani K, Muralidhara BK, Yamashita H, Tabata S, Mikami B, et al. (2001) 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 276: 35940–35946.
35. Abergel RJ, Raymond KN (2008) Terephthalamide-containing ligands: fast removal of iron from transferrin. Journal of Biological Inorganic Chemistry 13: 229–240.
36. Harris WR, Brook CE, Spilling CD, Elleppan S, Peng W, et al. (2004) Release of iron from transferrin by phosphonocarboxylate and diphosphonate chelating agents. Journal of Inorganic Biochemistry 98: 1824–1836.
37. Zak O, Tam B, MacGillivray RTA, Aisen P (1997) A kinetically active site in the C-lobe of human transferrin. Biochemistry 36: 11036–11043.
38. Boehr DD, Wright PE (2008) How do proteins interact? Science 320: 1429–1430.
39. Ma BY, Kumar S, Tsai CJ, Nussinov R (1999) Folding funnels and binding mechanisms. Protein Engineering 12: 713–720.
40. Lange OF, Lakomek NA, Fares C, Schroder GF, Walter KFA, et al. (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble. Science 320: 1471–1475.
41. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, et al. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophysical Journal 80: 505–515.
42. Yilmaz LS, Atilgan AR (2000) Identifying the adaptive mechanism in globular proteins: Fluctuations in densely packed regions manipulate flexible parts. Journal of Chemical Physics 113: 4454–4464.
43. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. Nucleic Acids Research 28: 235–242.

44. Bahar I, Jernigan RL (1997) Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. Journal of Molecular Biology 266: 195–214.

45. Bertz M, Wilmanns M, Rief M (2009) The titin-telethonin complex is a directed, superstable molecular bond in the muscle Z-disk. Proceedings of the National Academy of Sciences of the United States of America 106: 13307–13310.

46. Alexandrov V, Lehnert U, Echols N, Milburn D, Engelman D, et al. (2005) Normal modes for predicting protein motions: A comprehensive database assessment and associated Web tool. Protein Science 14: 633–643.

47. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. Journal of Molecular Graphics 14: 33–38.

48. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. Journal of Computational Chemistry 26: 1781–1802.

49. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, et al. (1983) CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J Comput Chem 4: 187–217.

50. Yang W, Gao YQ, Cui Q, Ma J, Karplus M (2003) The missing link between thermodynamics and structure in F1-ATPase. Proceedings of the National Academy of Sciences of the United States of America 100: 874–879.

51. Beglov D, Roux B (1994) Finite Representation of an Infinite Bulk System - Solvent Boundary Potential for Computer-Simulations. Journal of Chemical Physics 100: 9050–9063.

52. Bruns CM, Nowalk AJ, Arvai AS, McTigue MA, Vaughan KG, et al. (1997) Structure of Haemophilus influenzae Fe+3-binding protein reveals convergent evolution within a superfamily. Nature Structural Biology 4: 919–924.

53. Darden T, Perera L, Li LP, Pedersen L (1999) New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. Structure with Folding & Design 7: R55–R60.

54. Swope WC, Andersen HC, Berens PH, Wilson KR (1982) A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. J Chem Phys 76: 637–649.

55. Tsai CJ, Kumar S, Ma BY, Nussinov R (1999) Folding funnels, binding funnels, and protein function. Protein Science 8: 1181–1190.

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

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

58. Vendruscolo M, Dobson CM (2006) Dynamic visions of enzymatic reactions. Science 313: 1586–1587.

59. Flores S, Echols N, Milburn D, Hespenheide B, Keating K, et al. (2006) The database of macromolecular motions: new features added at the decade mark. Nucleic Acids Research 34: D296–D301.

60. Lockless SW, Ranganathan R (1999) Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. Science 286: 295–299.

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

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

63. Ghosh A, Vishveshwara S (2007) A study of communication pathways in methionyl-tRNA synthetase by molecular dynamics simulations and structure network analysis. Proceedings of the National Academy of Sciences of the United States of America 104: 15711–15716.

64. Ghosh A, Vishveshwara S (2008) Variations in Clique and Community Patterns in Protein Structures during Allosteric Communication: Investigation of Dynamically Equilibrated Structures of Methionyl tRNA Synthetase Complexes. Biochemistry 47: 11398–11407.