doi:10.1016/j.jmb.2008.03.010 J. Mol. Biol. (2008) 378, 1155–1173

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Molecular Dynamics Simulation of the *Escherichia coli* NikR Protein: Equilibrium Conformational Fluctuations Reveal Interdomain Allosteric Communication Pathways

Michael J. Bradley[1,2], Peter T. Chivers[2] and Nathan A. Baker[2]*

[1]*Graduate Program in Molecular Biophysics, Washington University in St. Louis, St. Louis, MO 63110, USA*

[2]*Department of Biochemistry and Molecular Biophysics, Washington University in St. Louis, St. Louis, MO 63110, USA*

Received 26 November 2007; received in revised form 5 March 2008; accepted 7 March 2008
Available online 14 March 2008

Escherichia coli NikR is a homotetrameric Ni^{2+}- and DNA-binding protein that functions as a transcriptional repressor of the NikABCDE nickel permease. The protein is composed of two distinct domains. The N-terminal 50 amino acids of each chain forms part of the dimeric ribbon–helix–helix (RHH) domains, a well-studied DNA-binding fold. The 83-residue C-terminal nickel-binding domain forms an ACT (aspartokinase, chorismate mutase, and TyrA) fold and contains the tetrameric interface. In this study, we have utilized an equilibrium molecular dynamics simulation in order to explore the conformational dynamics of the NikR tetramer and determine important residue interactions within and between the RHH and ACT domains to gain insight into the effects of Ni^{2+} on DNA-binding activity. The molecular simulation data were analyzed using two different correlation measures based on fluctuations in atomic position and noncovalent contacts together with a clustering algorithm to define groups of residues with similar correlation patterns for both types of correlation measure. Based on these analyses, we have defined a series of residue interrelationships that describe an allosteric communication pathway between the Ni^{2+}- and DNA-binding sites, which are separated by 40 Å. Several of the residues identified by our analyses have been previously shown experimentally to be important for NikR function. An additional subset of the identified residues structurally connects the experimentally implicated residues and may help coordinate the allosteric communication between the ACT and RHH domains.

Edited by D. Case

Keywords: allostery; NikR; molecular dynamics; correlations; fluctuations

Introduction

Escherichia coli uses nickel ions under anaerobic conditions for the activity of its Ni–Fe hydrogenase enzymes, where the metal assists in the reversible oxidation of diatomic hydrogen.[1] The bacterium acquires nickel via the NikABCDE nickel permease.[2] Excess Ni^{2+} accumulation by this pathway is controlled in part via transcriptional repression of *nikABCDE* expression by the nickel-dependent NikR repressor protein.[3,4] NikR is one of two nickel-responsive repressors in *E. coli*. The other is RcnR, which is structurally unrelated to NikR and controls expression of the RcnA nickel and cobalt efflux protein.[5]

Structural characteristics

NikR is a homotetramer composed of two distinct domains[6] (see Fig. 1). The N-terminal 50 amino acids of each chain contribute to two dimeric domains of the ribbon–helix–helix (RHH) family,[7] a known and well-studied DNA-binding fold.[8] The C-terminal 83 amino acids of each chain form a tetramer composed of four βαββαβ domains that together contain the high-affinity nickel-binding sites (K_d ~1 pM).[6,9,10]

*Corresponding author. E-mail address: baker@ccb.wustl.edu.

Abbreviations used: RHH, ribbon–helix–helix; ACT, aspartokinase, chorismate mutase, and TyrA; MD, molecular dynamics; PCA, principal component analysis; MSA, multiple sequence alignment.

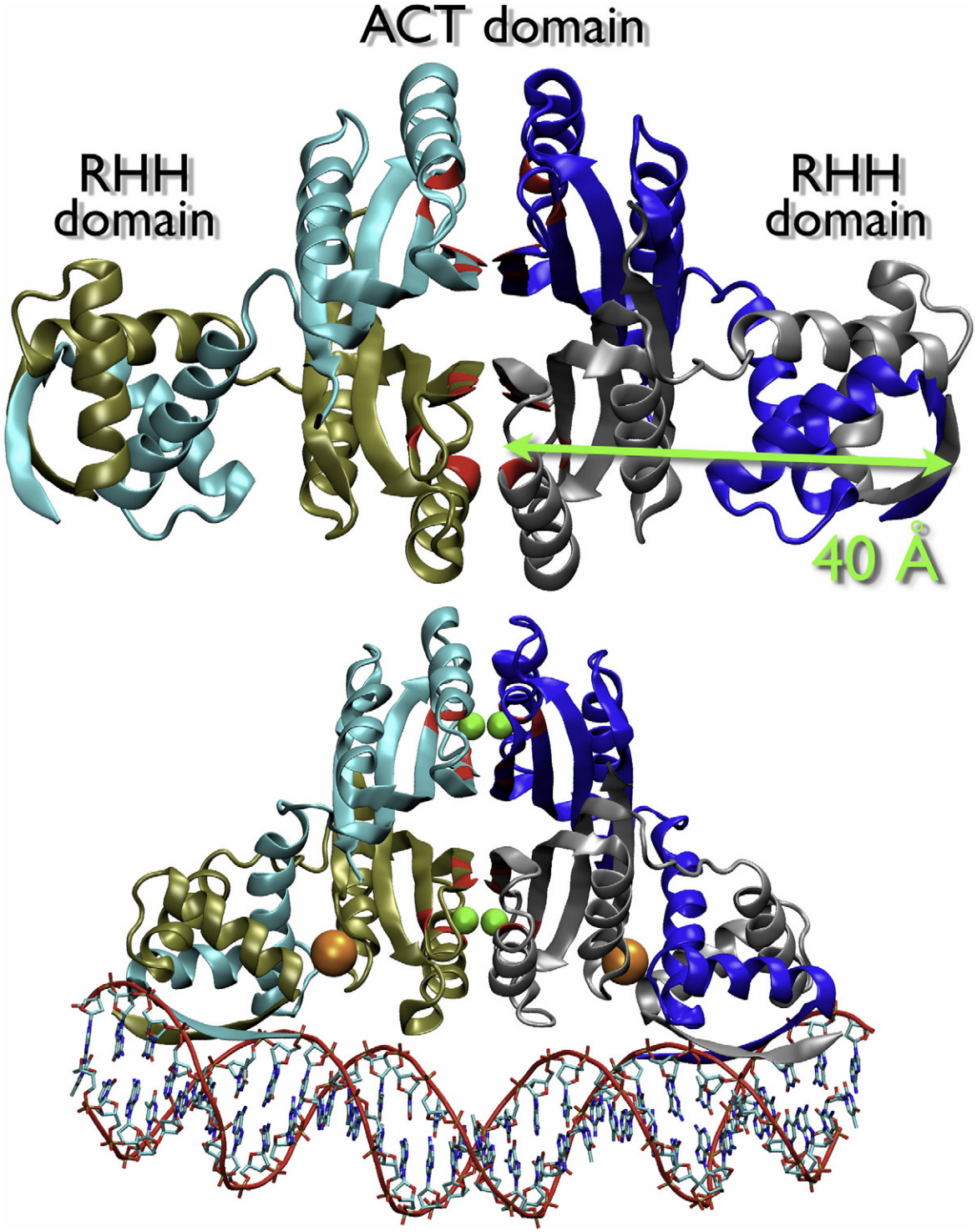

Fig. 1. Comparison of *apo* and Ni^{2+}–DNA-bound NikR X-ray crystal structures. *Apo* structure was generated from PDB ID 1Q5V with missing atoms built in and energy minimized. Ni^{2+}–DNA-bound NikR is from PDB ID 2HZV. Protein chains are in "new cartoon" rendering and colored cyan, blue, tan, and silver. Ni^{2+}-binding site residue positions (His76, His87, His89, Cys95) are colored red. DNA double-helix backbone is outlined in "tube" rendering and colored red while nucleotides are in "bond" rendering and colored by atom name. Ni^{2+} atoms are shown as green spheres. K^+ atoms are shown as orange spheres. All molecular images generated using VMD 1.8.6†.[11]

This domain is structurally homologous to the ACT (aspartokinase, chorismate mutase, TyrA)-fold family, which is a structural motif that typically binds small molecules that allosterically control enzymatic activity.[12] The NikR protein forms an interesting union of the DNA-binding RHH and the regulatory ACT domains to provide allosteric control of DNA binding through Ni^{2+} binding and

† http://www.ks.uiuc.edu/Research/vmd/

thereby allowing NikR to act as a transcriptional repressor that responds to elevated intracellular Ni^{2+} concentrations.[4,6,13] A number of residues for NikR activity have been identified through structural and functional studies (Table 2).

Functional sites

Ni^{2+}-binding sites

Metal binding is required for measurable association of *E. coli* NikR to *nik* operator DNA *in vitro*.[6,9,14] Metal binding stabilizes *E. coli* NikR against both heat and chemical denaturation[9,10] as well as proteolytic cleavage.[15] This stabilization is likely related to a nickel-specific conformational change of the protein. Changes in backbone amide hydrogen/deuterium exchange[16] show the largest effect with nickel, which binds with square planar coordination geometry.[6,17,18] Other first-row transition metals bind with different coordination geometries and exhibit hydrogen/deuterium exchange different from the Ni^{2+}-bound protein.[16]

The X-ray structure of the *E. coli* NikR C-domain[6] first identified the high-affinity nickel-binding sites, located at the interface between two NikR dimers (see Fig. 1). The Ni^{2+} sites are roughly ~40 Å distant from the β sheet in the N-terminus that forms sequence-specific DNA contacts (see Fig. 1). This structure confirmed the square planar geometry first reported by X-ray absorption spectroscopy.[17] The nickel ligands come from the side chains of His87, His89, Cys95, and His76 bound directly to the metal. These residues are conserved in alignments of NikR sequences from over 82 bacterial and archaeal species (Fig. 8). Mutation of Cys95 to Ala eliminates the nickel-responsive DNA-binding activity of NikR[17] and is functionally inactive *in vivo*,[13] equivalent to deletion of the *nikR* gene. The X-ray structure also revealed additional conserved residues (Tyr60, His62, Gln75, Glu97). Mutation of Glu97 also eliminates the Ni-responsive DNA-binding activity of NikR.[17] The square planar Ni site has also been identified in structures of the *Helicobacter pylori*[19] and *Pyrococcus horikoshii*[20] NikR orthologs.

DNA-binding sites

NikR was identified[7] because of its sequence homology to the Arc repressor, the archetypical RHH DNA-binding protein.[21–23] The isolated N-terminal domain of NikR retains weak, nickel-independent DNA-binding activity.[7] Mutation of Arg3 to Ala also abrogates DNA binding. Schreiter *et al.*[18] have determined the co-crystal structure of EcNikR with an operator DNA fragment. This structure confirms the expected specific RHH–DNA interactions via the β-sheet. This structure also identified nonspecific contacts between the ACT domain and the DNA.

Low-affinity metal-binding site

There is evidence for a functional second nickel-binding site in *E. coli* NikR that results in increased DNA-binding affinity[4,9,14] and increased repression of NikABCDE expression[5,13] than is observed with just high-affinity Ni-binding site occupancy. The structural location of this site is ambiguous. An X-ray structure of NikR from *P. horikoshii*[20] reported a second Ni^{2+} site at the N- and C-terminal domain interface (COO^- of Glu30 and Asp34 and main-chain C=O of Ile116, Gln118, and Val121). In contrast, the *E. coli* NikR–DNA co-crystal structure reported a K^+ ion bound using the same residues. Mutation of either Glu 30[20] or Asp34[18] eliminates the increased DNA affinity with increased Ni^{2+}.

Allostery and structural changes

Despite the available structural and biochemical data for NikR, it is not clear how Ni^{2+} binding activates the protein for DNA binding. Recent crystal structures of full-length *E. coli* NikR with Ni^{2+} and with both Ni^{2+} and DNA bound[18] suggest that Ni^{2+} binding is not correlated with a single conformational state of the protein; i.e., Ni^{2+} binding does not appear to "prearrange" the protein into a conformation similar to the DNA-bound state. As such, it is not obvious from crystal structure comparisons how nickel binding activates the protein for DNA binding. Conformational heterogeneity has also been observed in crystal structures of the *H. pylori* and *P. horikoshii* NikR orthologs.[19,20] Solution studies suggest that at least a small change in the average conformation of NikR must accompany Ni^{2+} binding,[15,16] indicating that perturbations of the protein's conformational flexibility and dynamics may play an important role in activation of DNA binding. Recently, Cui and Merz used coarse-grained elastic network models to probe the range of conformational motions available to the NikR tetramer.[24] They concluded that large-scale conformation changes were essential to transform *apo*NikR into a DNA-binding state but were unable to identify the specific residue-level changes needed to effect this change.

A series of both experimental and simulation studies support the notion of a preexisting equilibrium of protein conformational states in the absence of bound ligands.[25–30] Additionally, recent work using coarse-grained modeling of several enzymes supports a connection between the conformational fluctuations of the *apo* protein and the motions involved in functionally important conformational changes.[31,32] A helpful framework for this idea is provided by the fluctuation–dissipation theorem, which states that the response of a system to small perturbations involves fluctuation pathways that are present at equilibrium prior to the perturbation.[33]

We have utilized an equilibrium molecular dynamics (MD) simulation in order to explore the conformational dynamics of the *E. coli* NikR tetramer‡. Molecular dynamics methods have been used

‡ Hereafter, discussion of NikR in this manuscript refers to the *E. coli* protein, unless otherwise noted.

with a variety of other proteins to find residue interactions that are involved in conformational changes.[34–37] The goal of our *apo*NikR simulation and analysis is to provide insight into conformational fluctuations and local interactions involved in the structural changes regulating the protein's DNA-binding affinity. Analyses of contact and positional correlations within and between residues of the ACT and RHH domains have defined networks of residues that connect the Ni^{2+}- and DNA-binding sites and are likely important for allosteric control of NikR function. Several of the residues identified by these analyses have been shown experimentally to be important for NikR function (Ni^{2+} or DNA binding). An additional subset of these residues provides a structural link between experimentally identified residues and may help coordinate the allosteric communication between the ACT and RHH domains. Given the structural similarity of the Ni^{2+}-binding ACT domain in NikR with other small-molecule-binding ACT domains in a variety of other proteins, understanding how ACT domain control works in NikR could help elucidate a common regulatory mechanism for several ACT-containing systems.

Results

Equilibration measures

Molecular dynamics of *apo*NikR was performed as described in Methods and Theory. Before analyzing

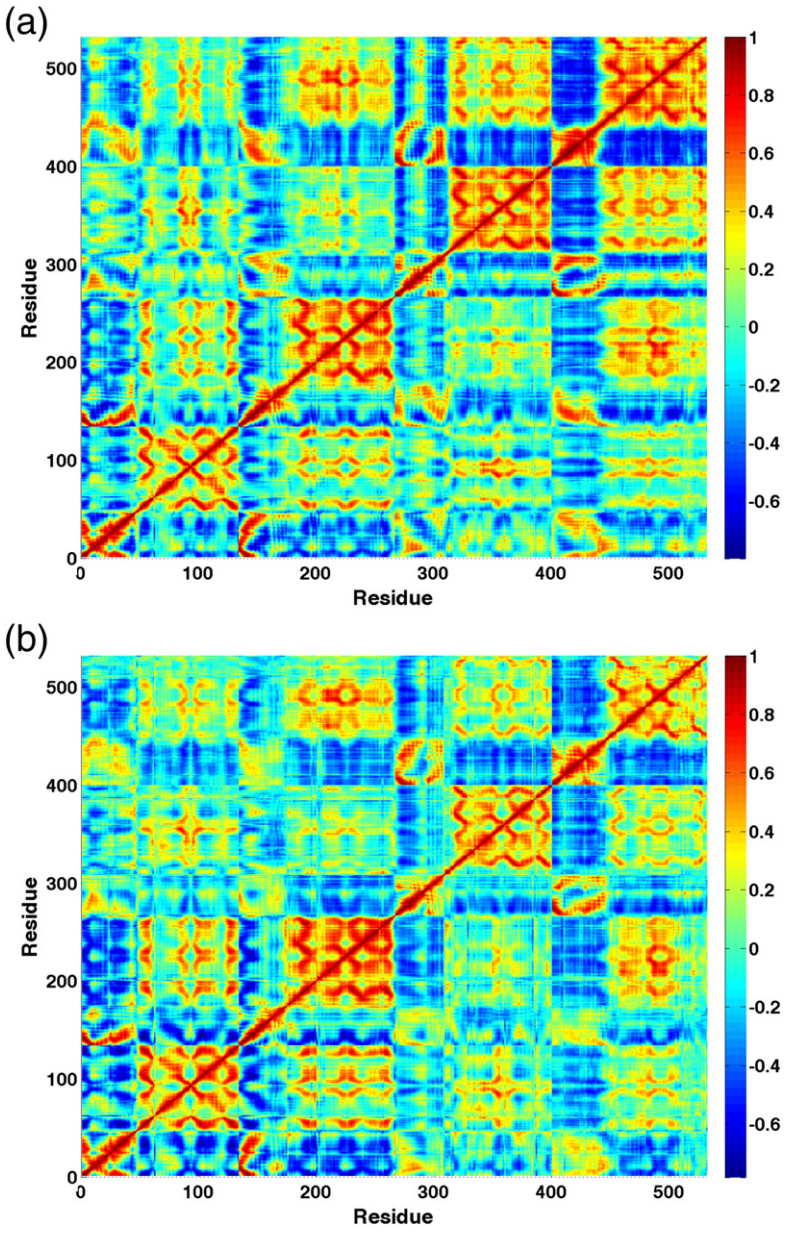

Fig. 2. C^{α} correlation matrices. Last 71 ns (a) and last 51 ns (b). The data from the matrix that include earlier nonequilibrated structural change has systematically larger (anti-)correlations within two of the four subunits. Figure generated using MATLAB 7.4 (The MathWorks, Natick, MA).

the conformational fluctuations of the NikR tetramer (532 residues) observed during the MD simulation, it was important to assess whether the simulation achieved equilibrium or steady-state sampling. The evolution of the RMSD relative to the starting structure has been used to estimate equilibration and simulation stability in other systems.[38,39] An initial rapid rise was observed in RMSD that leveled off around 10 ns (see Supplementary Fig. S1), but with substantial continued fluctuation between 2.5 and 4.5 Å. An examination of the dihedral potential energy revealed a change over the first 30 ns of the simulation (see Supplementary Figs. S2 and S3). At least part of the dihedral energy drift appeared to correlate with changes in secondary structure over time (see Supplementary Fig. S4). Using secondary-structure assignments from DSSP[40] as implemented in the ptraj module of AMBER 9.0,[41] this secondary-structure change appeared to be caused by a loss of α-helix with a concomitant gain in 3_{10}-helix content, mainly in helix B of the RHH domain and helix C of the ACT domain. Taken together, these data showed a clear systematic drift in the MD trajectory that leveled off at ~30 ns.

The conformational effect of this systematic drift was also apparent in the correlation matrix of atomic positions. Correlation matrices were calculated for the α carbon position of each residue (532 total; 532 × 532 matrix). For comparison, the correlation matrix without the first 10 ns (change in RMSD) was compared with the matrix calculated without the first 30 ns (systematic drift in potential energy). Clearly, inclusion of the simulation data between 10 to 30 ns has a strong effect on the correlation matrix (see Fig. 2), leading to the conclusion that the systematic drift apparent over the first 30 ns of the simulation gives rise to correlated motions due to concerted structural changes rather than equilibrium fluctuations of the system. Therefore, only the last 51 ns of the simulation were used for subsequent analyses.

Principal component analysis

Principal component analysis (PCA) yields a set of modes (eigenvectors) that represent a nonredundant set of motions observed in the MD trajectory for a set of selected atoms.[42–45] PCA was carried out on the simulation data using the α carbon of each of the 532 residues in the NikR tetramer. This analysis yields $3N$ total modes, where N is the number of atoms included, giving a total of 1596 modes. The first mode can be interpreted as the principal axis of the largest atomic fluctuations represented in the covariance matrix, e.g., the direction of maximum variation in conformation observed over the course of the molecular simulation. Each subsequent mode represents the next largest principal axis of atomic fluctuations orthogonal to all previous axes. Every PCA mode is also associated with an eigenvalue, this eigenvalue corresponds to the amplitude of fluctuations along that mode. Therefore, each eigenvalue, divided by the sum of all eigenvalues, represents the relative contribution of a mode to the total conformational variance observed during the simulation.

To place the *apo*NikR MD simulation in the context of experimentally determined structural transitions for *E. coli* NikR, PCA provided a means to determine whether the observed equilibrium conformational fluctuations resemble the transformation between *apo*NikR and DNA-bound Ni^{2+}NikR. The observed PCA modes, v_i, were compared with a "structural change vector," Δx, defined by aligning the DNA-bound crystal structure with the minimized *apo*NikR structure (e.g., the starting MD conformation described in the Methods and Theory section) and calculating the change in position of each alpha carbon. A "completeness" test was performed to ascertain if Δx could be adequately described using a subset of the PCA modes as a basis set [Eqs. (3)–(6) in Methods and Theory]. The full basis set of 1596 PCA modes gave high overlap [cos θ, Eq. (5)] of 0.999 and low error of 0.034 [ε, Eq. (6)] when used to represent the Δx observed in X-ray studies (see Table 1). From this PCA basis set, a minimal set of modes was identified that represented the observed displacement between *apo* and DNA-bound NikR. Table 1 lists the mode overlaps between Δx and the "top 10" PCA modes based on the magnitude of the eigenvalues. The first and largest PCA mode accounts for 41.7% of the total motion, but has only a small overlap [cos(θ) = 0.125] with the observed changes in the X-ray structure. This largest mode represents an asymmetric twisting of one RHH dimer relative to the rest of the NikR structure. This twisting motion occurs along the long axis of the NikR molecule. The second PCA mode accounts for 13.2% of the total motion and has a strong overlap [cos(θ) = 0.811]. This second mode is a highly symmetric "flapping" motion of both RHH dimers relative to the ACT

Table 1. PCA mode overlap with the NikR structural change vector

Mode index	Eigenvalue	% of Motion	cos(θ)	$\alpha^{(i)}$	Error[a]
1	590.948	41.7	−0.125	−19.14	0.992
2	187.356	13.2	−0.811	−121.4	0.573
3	129.591	9.1	0.157	23.66	0.551
4	78.895	5.6	−0.307	−46.13	0.462
5	44.271	3.1	0.150	23.18	0.441
6	39.941	2.8	−0.044	−6.52	0.439
7	34.617	2.4	0.211	31.75	0.387
8	24.542	1.7	−0.138	−20.75	0.364
9	17.087	1.2	−0.017	−2.56	0.364
10	12.447	0.9	−0.020	−3.20	0.364
First 10	(sum) 1159.695	81.9	0.932	—	0.364
All 1596	(sum) 1416.344	100	0.999	—	0.034

As described in the text, the PCA modes describe a vector displacement of all C^{α} positions and are indexed by decreasing magnitude (eigenvalue). The structural change vector is calculated by comparing the C^{α} positions of the minimized *apo*NikR crystal structure (PDB ID 1Q5Y with missing atoms built in) with the Ni^{2+}–DNA-bound NikR crystal structure (PDB ID 2HZV). Equations for the vector overlap cos(θ), $\alpha^{(i)}$, and the error can be found in the Methods and Theory section.

[a] For modes 1–10 the error is cumulative from mode 1 up to the specified mode.

tetramer that clearly resembles the conformational change necessary to transform the *apo*NikR crystal structure into the DNA-bound Ni^{2+}NikR structure (see Fig. 3). Together, the first 10 modes account for 81.9% of the total motion. A "reconstructed vector" that consists of the first 10 modes reweighted by their respective $\alpha^{(i)}$ also has a high overlap of 0.932 with Δx. In other words, the conformational fluctuations represented by the first 10 PCA modes can be used to provide a reasonable representation of a

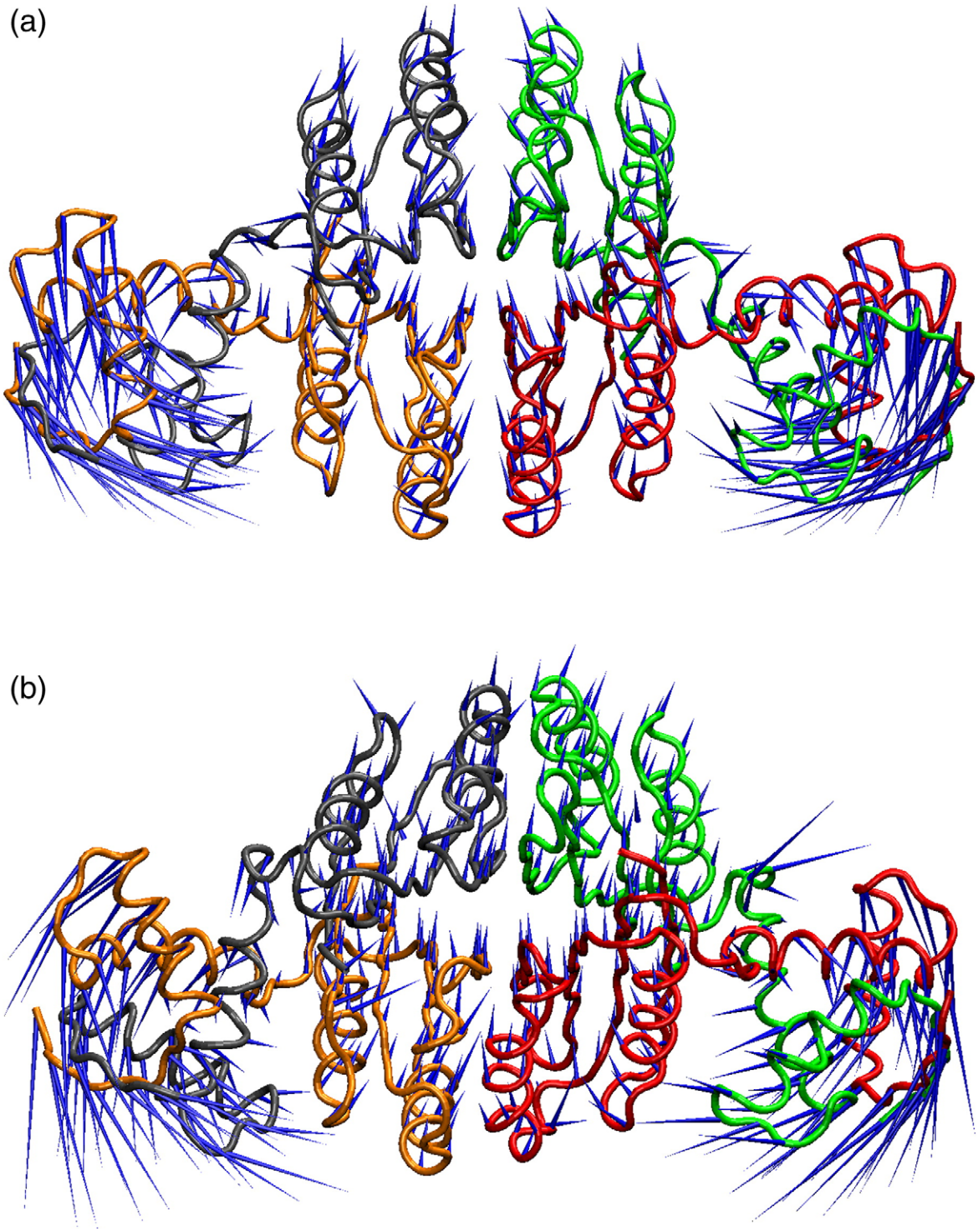

Fig. 3. X-ray crystal structure and PCA mode displacement visualizations. (a) The minimized starting structure is shown in tube representation colored by chain. The blue "porcupine needles"[36] indicate the direction of displacement in going from the *apo* to the DNA-bound Ni^{2+}NikR crystal structure, denoted as Δx in the text. (b) The MD average structure from the last 51 ns is shown in tube representation colored by chain. The blue porcupine needles indicate the direction of displacement based on the second PCA mode, which has the highest overlap with Δx.

conformational change that is expected to be functionally relevant.

Correlation-matrix-based residue clustering

An additional set of observables of interest in the simulation includes correlations in the equilibrium fluctuations of atomic position between sets of residues. A compact representation of these data is the covariance matrix, which for a chosen set of atoms expresses the variances (matrix diagonal) and covariances (off-diagonals) in Cartesian coordinates. In three-dimensional space, the covariance matrix is $3N \times 3N$ for N selected atoms. The scalar correlation matrix is instead $N \times N$ and simply contains the mean-squared correlations [range −1.0 to 1.0, see Eq. (8) in Methods and Theory] between all N atoms in the off-diagonals and 1.0 along the diagonal. The scalar correlation matrix can be visualized to observe how motions in different regions of the protein are correlated (see Fig. 2). By using the magnitude of the (anti-)correlation of α carbon atoms as an effective distance between residues (see Methods and Theory) protein residues were clustered into groups whose motions were correlated with Ni^{2+}- and DNA-binding site residues (see Fig. 4) where selected clusters are plotted as blocks of color on the NikR structure (see also Supplementary Table S1). Three clusters that connect the Ni^{2+}- and DNA-binding sites are shown (red, green, and blue) as well as an additional cluster contained within one RHH dimer (orange). This analysis allows visualization of dynamic correlations as structural connections between groups of residues that could be important for NikR activation upon nickel binding.

Noncovalent contact correlations

One observable that reports on equilibrium conformational fluctuations is the making and breaking of noncovalent bonds or "contacts" in residue–residue interactions. We hypothesize that significant correlations between contact fluctuations indicate energetic connections between regions of the protein that are not apparent from a static structure. Using contact definitions defined in the Methods and Theory section, correlation statistics between all i,j pairs of contacts were calculated according to the ϕ_{ij} binary correlation measure. By utilizing the connection between ϕ_{ij} and χ^2 (see Methods and Theory), correlations could be classified as "significant" at a 95% confidence interval, leading to the discovery of networks of residues with significantly correlated contacts that connect the Ni^{2+} and DNA binding domains of NikR. Furthermore, certain residues are categorized as highly connected "hubs" in these networks (residues marked with a "#" in Table 2).

The clusters selected based on the "depth-first" criteria for noncovalent contact correlations (see Methods and Theory) are shown mapped onto the NikR structure in Fig. 5. Three key regions had a high concentration of these residues (see Fig. 5b): helix D (residues 60–65), the turn into beta strand 5 (residues 118–122), and helix B (residues 27–42). Residues associated with the selected contacts irrespective of chain ID are mapped onto the structure. These three

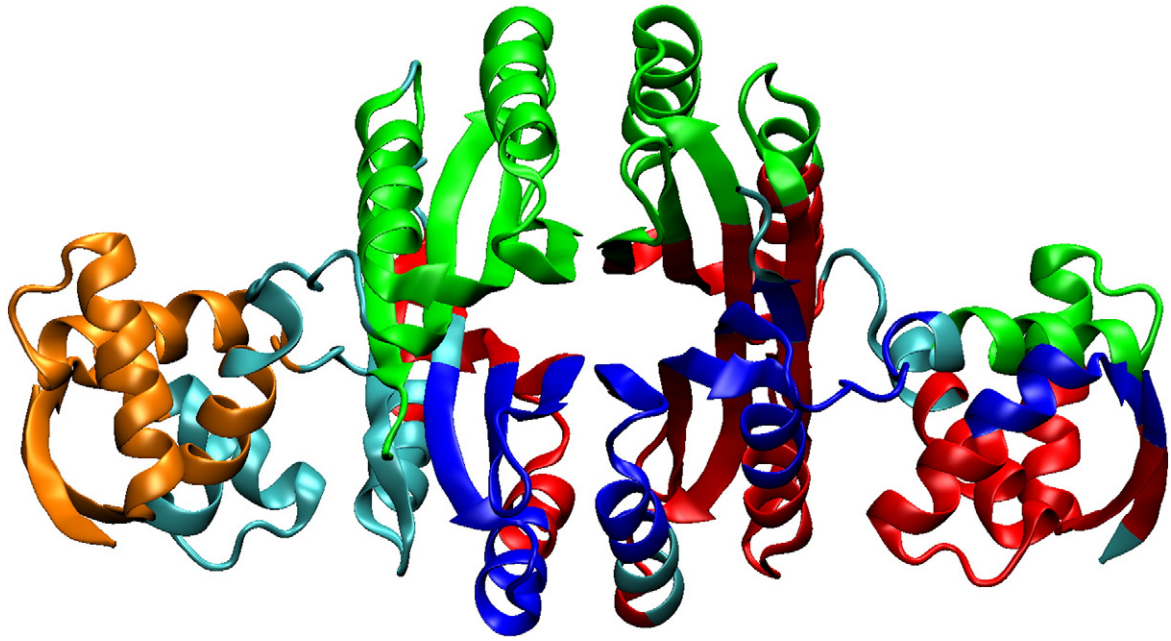

Fig. 4. Residue clusters based on the correlation matrix. Clusters generated using the UPGMA algorithm were selected based on including either Ni^{2+} or DNA-binding site residues or both. These clusters imply groups of residues with concerted conformational fluctuations that could be important for interdomain communication. Different clusters are indicated by color, except cyan, which indicates residues not found in the selected clusters.

Table 2. NikR residues selected from computational analyses

Residue	Experimental significance	Computational significance	Sequence conservation
Arg3	Specific DNA binding[7]	A,B	High (0.83)
Thr5	Specific DNA binding[17]	B	High (0.62)
Thr7	Specific DNA binding[17]	A,B,*	High (0.81)
Asp9	Unknown	A,B,#	Moderate (0.58)
Arg22	Unknown	A,B,*	Moderate (0.45)
Asn27	Nonspecific DNA contact[17]	A,B	Moderate (0.64)
Arg28	Nonspecific DNA contact[17]	A,B	High (1.00)
Ser29	Nonspecific DNA contact[17]	A,B,*	High (1.00)
Glu30	Low-affinity metal site[17,19]	A,B	High (0.81)
Arg33	Nonspecific DNA contact[17]	A,B	High (0.87)
Asp34	Low-affinity metal site[17,19]	A,B	Moderate (0.71)
Arg37	Unknown	A,B,#	High (0.85)
Gln42	Unknown	A,B,*	Low (0.41)
Tyr58	Ni^{2+} site H-bond network[6]	B	High (0.60)
Tyr60	Close proximity to Ni^{2+} site[6]	A,B	High (0.86)
His62	Close proximity to Ni^{2+} site[6]	A,B,*,#	High (0.83)
Lys64	Nonspecific DNA contact[17]	A,B	Low (0.38)
Arg65	Nonspecific DNA contact[17]	A,B,*	Moderate (0.52)
Ser69	Unknown	A,B,*	Moderate (0.51)
Gln75	Ni^{2+} site H-bond network[6]	A,B	High (0.74)
His76	High-affinity Ni^{2+}-binding site[6]	A,B,*	High (1.00)
His87	High-affinity Ni^{2+}-binding site[6]	A,B	High (1.00)
His89	High-affinity Ni^{2+}-binding site[6]	B	High (1.00)
Cys95	High-affinity Ni^{2+}-binding site[6]	A,B	High (1.00)
Glu97	Reduced Ni^{2+} and DNA binding upon mutation[16]	A,B	High (0.98)
Gln109	Unknown	A,B,*	Moderate (0.49)
Asp114	Unknown	A,B,*	Moderate (0.47)
Ile116	Low-affinity metal site[17,19]	B	Low (0.39)
Gln118	Low-affinity metal site[17,19]	A,B,#	Moderate (0.45)
Arg119	Nonspecific DNA contact[17]	A,B	High (0.62)
Val121	Low-affinity metal site[17,19]	A,B	High (0.82)

Computational significance denoted by A, contact correlation clusters that include Ni^{2+}- and DNA-binding site residues; B, domain-spanning correlation matrix clusters; *, top 10 residues in number of contact correlations within the selected clusters; #, top 4 residues in total number of significant contact correlations. Sequence conservation is denoted by position conservation scores from Scorecons given in parentheses (see Methods and Theory section). Those residues with "unknown" experimental significance have not yet been tested and represent positions for which mutation is predicted to alter NikR function.

regions contain several evolutionarily conserved residues as determined by multiple sequence alignment (MSA) (see Methods and Theory) and include residues known experimentally to be important for NikR function (see Table 2). A visual representation of the MSA is included in Fig. 8. Several of the noncovalent contacts that are correlated with Ni^{2+}- and DNA-binding site contacts (see Fig. 5a) bridge the three regions highlighted in Fig. 5b and directly connect these regions to binding site residues. This is illustrated in Fig. 7, where a subset of residues from Table 2 are mapped onto the NikR structure along with noncovalent contacts selected by our clustering method. The top 10 residues, irrespective of chain ID, that "own" the largest number of contact correlations in the selected clusters (listed Supporting Information) are marked with "*" in Table 2. These residues are highly interconnected within the network of contact correlations selected by our analysis based on binding site residues. All of the residues marked with either "*" or "#" in Table 2 are included in Fig. 7. Through this analysis, specific residue–residue contacts distributed between the ACT and RHH domains have been identified that connect to both the Ni^{2+}- and DNA-binding sites and could be important for allosteric control in NikR.

Discussion

We have used an MD simulation, together with new noncovalent contact and position correlation clustering methods, to investigate the mechanism of allostery in the NikR protein. Our hypothesis is that the dynamics observed in the *apo*NikR MD simulation contain functionally relevant conformational fluctuations. The lack of Ni^{2+} in these simulations prevents us from drawing detailed conclusions about the specific role of Ni^{2+}as compared to other transition metal ligands. However, our overall approach and its relevance to NikR function is supported by comparison of the dominant modes of motion from PCA with the conformational change necessary to transform the *apo*NikR crystal structure into the DNA-bound Ni^{2+} NikR structure. The new contact and position correlation methods are utilized to find clusters of residues that share similar correlation patterns with Ni^{2+}- and DNA-binding site residues. This results in the identification of a network of residue interactions that connect the two types of NikR-binding sites and further highlights individual residues that could be important allosteric communication links. Several of these residues are evolutionarily conserved among members of the NikR

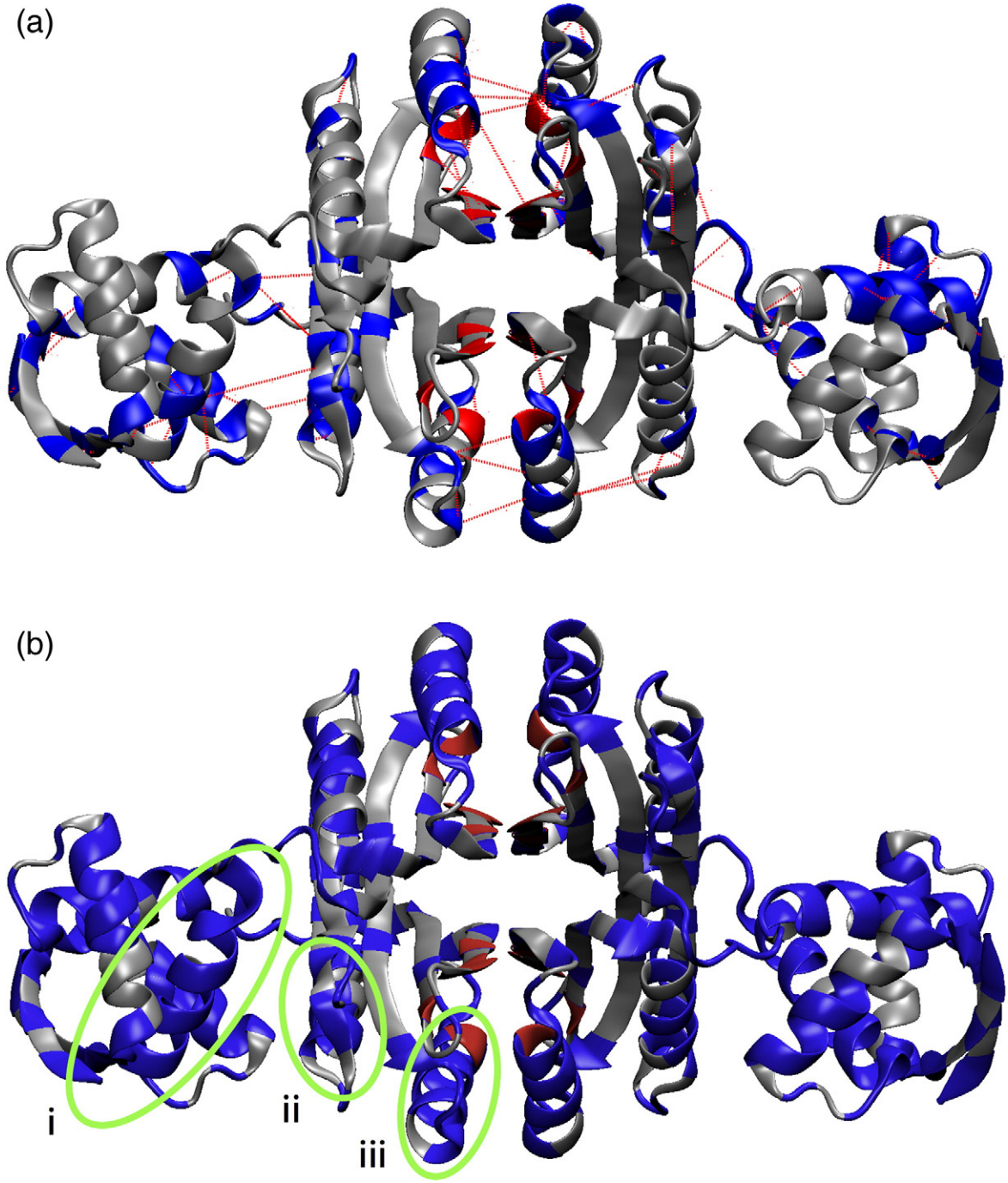

Fig. 5. Protein regions identified by noncovalent contact correlations. (a) Residues colored blue were selected using the UPGMA clustering criteria of finding the smallest clusters that include both Ni^{2+}- and DNA-binding site residues. The noncovalent contacts selected by this method are shown as red dashes. (b) Color the same as in (a) except residues are colored based on residue number irrespective of chain, thereby emphasizing the symmetry of the tetramer. The nickel-binding residues are colored red (only His76, His87, and Cys95 were found in the UPGMA clusters). The remaining nickel-binding residue, His89, is colored yellow. The three green ovals each highlight one example of the three regions with a concentration of correlated contacts: (i) helix B (residues 27–31, 33, 34, 37–48), (ii) the end of helix D and turn leading into strand 5 (residues 117–123), and (iii) end of strand 2 and helix C (residues 62–66 and 68–80).

family; those residues found in the Ni^{2+}-binding domain could represent important control points that are common elements in ACT domain control of biological activity.

Correlation analysis methodology

While previous investigators have considered patterns in correlation matrices obtained from

molecular simulations,[46,47] the methods presented here combine automatic clustering of residues based on both position and contact correlations with further refinement using functional information about allosterically linked sites of the molecule. Inclusion of multiple functional sites in the cluster selection method can result in relatively large clusters, which parse the NikR structure into regions that have significant connections manifested by the observed correlation. While these residue clusters are likely involved in interdomain communication, they do not show individual residue interactions that could be important for the allosteric mechanism. The most significant methodological development in this work is the analysis of noncovalent contact correlations to measure fluctuations in "interaction space." This contact correlation analysis approach provides a more local view of interactions, which complements the global motions typically identified with PCA or the position clustering methods described in this article. The interpretation of the results of these position and contact correlation analyses is based on the hypothesis that such correlations imply structural and energetic connections between residues important for changes in NikR conformational distributions due to Ni^{2+}. This interpretation is supported by the observation that several identified residues have functional consequences when mutated, appear to be important in the crystal structure of the NikR–DNA complex, and/or are evolutionarily conserved (see Fig. 6 and Table 2). The usefulness of this approach is further supported by a related recent study of allosteric protein structures that analyzed local changes in contacts to successfully identify residue interaction networks.[48] The residues identified by our method do not provide direct physical information about the series of events that generate the structural changes in NikR; however, they do provide a starting point for subsequent experimental and computational studies designed to specifically determine the pathways for allosteric changes in the protein.

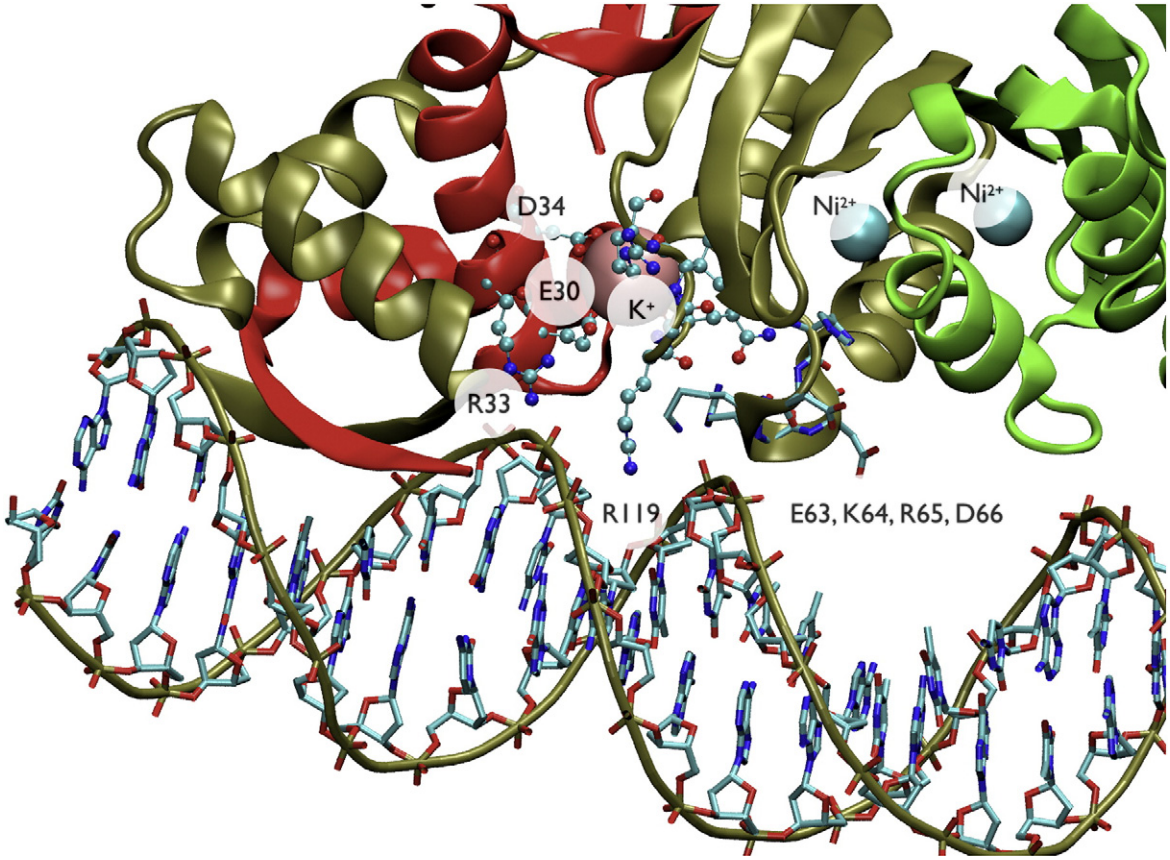

Fig. 6. Close-up view of NikR–DNA interactions from the crystal structure (PDB ID 2HZV[18]). Protein chains are in "cartoon" rendering and colored red, tan, and green. Ni^{2+} atoms are shown as cyan spheres, while K^+ is shown in pink. The DNA backbone is outlined in "tube" rendering and colored tan, while nucleotides are in "bond" rendering and colored by atom type. Protein side chains for residues 63–66 are in "bond" rendering and colored by atom type, while residues 30, 33, 34, and 118–122 are in "CPK" rendering and colored by atom type. The labeled protein residues correspond with the three groups identified in Fig. 5 from MD analysis of *apo*NikR conformational dynamics. All three regions contain residues that make nonspecific contacts with the DNA phosphate backbone. In addition, the residues flanking R119 from one region along with residues E30 and D34 from another region form the cation-binding site that apparently helps stabilize the DNA-bound conformation.

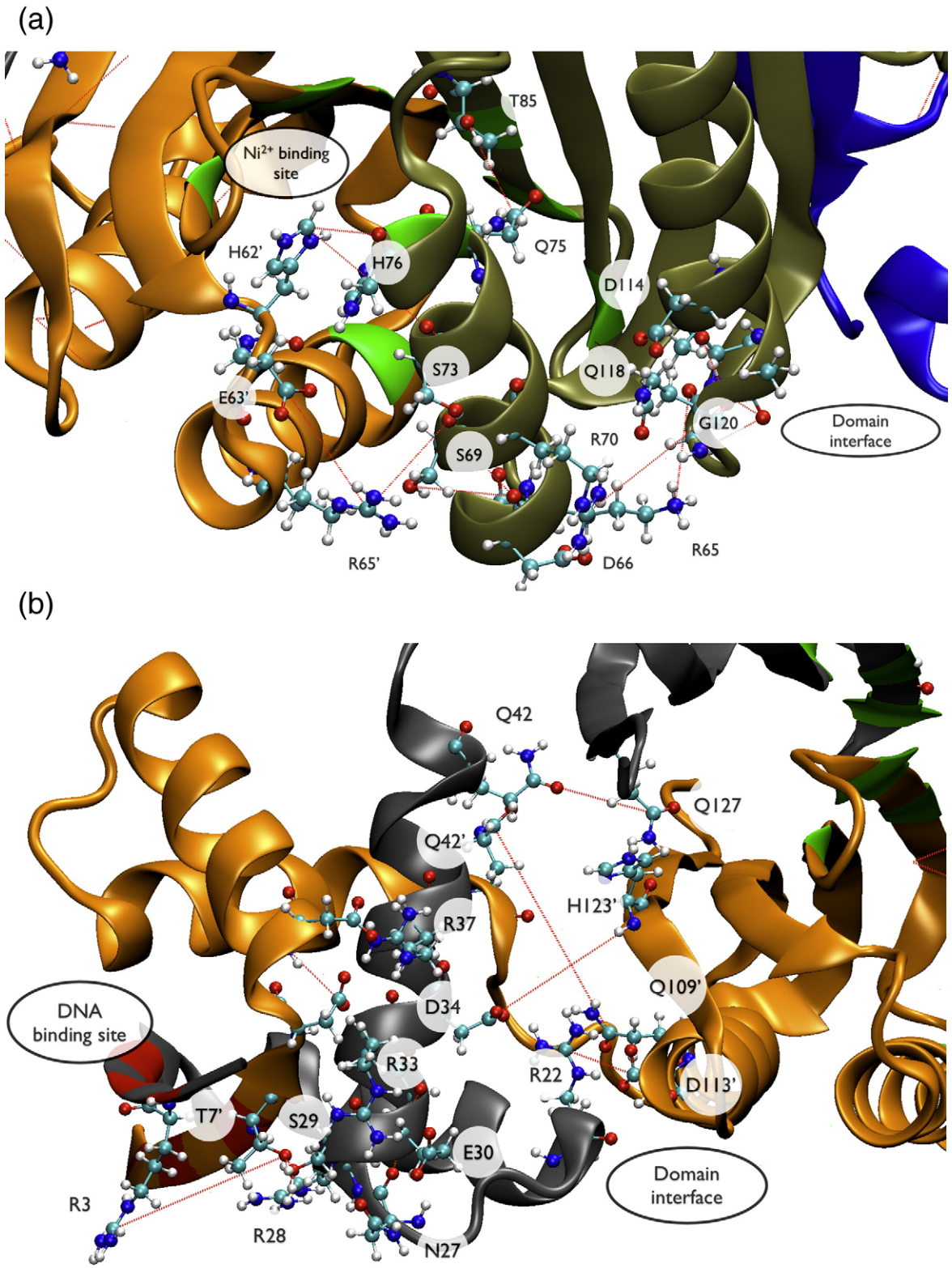

Fig. 7. This figure shows subsets of residues from Table 2 mapped onto the NikR structure: (a) residues that connect the Ni^{2+}-binding site to the RHH/ACT interface and (b) the RHH domain and RHH/ACT interface. Individual protein chains are colored gray, orange, tan, and blue. Ni^{2+}-binding residues are colored green; residues that make sequence-specific DNA contacts are colored red. Red dashes indicate correlated noncovalent contacts selected by UPGMA clustering (see Methods and Theory). Selected residues are shown in CPK rendering and labeled by residue type/number.

Identification of residues implicated in NikR allostery

We observe groups of noncovalent contact correlations between three regions of the NikR structure identified in Fig. 5 and the Ni^{2+}- and DNA-binding sites. When taken together with the domain-spanning clusters identified from the positional correlation matrix, a picture emerges of a pathway connecting the two functional sites that could transfer energy and information such as nickel-binding site occupancy. This pathway is made up of the residues listed in Table 2 and mapped on the NikR structure in Fig. 7. The pathway includes several residues with known experimental importance combined with an additional subset that structurally connects these residues with the Ni^{2+}- and DNA-binding sites.

The PCA results support the idea that the *apo*NikR simulation contains functionally relevant conformational fluctuations by demonstrating that the displacement necessary to go from the *apo* to the DNA-bound X-ray crystal structures is well represented by the second PCA mode and by a weighted combination of the first 10 PCA modes. While the simulation does not undergo full conformational transitions between states similar to the *apo*, Ni^{2+}-bound, and Ni^{2+}–DNA-bound crystal structures, it does sample local fluctuations that are likely important to the allosteric transitions. The connection between *apo*-NikR conformational fluctuations and functionally relevant residue interactions is further borne out by a close inspection of NikR–DNA interactions from the X-ray crystal structure in Fig. 6. In this figure, the labeled protein residues correspond with the three groups identified in Fig. 5 from our noncovalent contact correlation analyses. All three groups contain residues that make nonspecific contacts with the DNA phosphate backbone. In addition, the residues flanking R119 from one group along with residues E30 and D34 from another group form the cation-binding site that helps stabilize the DNA-bound conformation.[18] Thus, our analyses of equilibrium conformational fluctuations of the *apo* structure have identified residues that are apparently important in nickel-activated DNA binding.

Other studies have introduced the idea of shifting a preexisting conformational equilibrium in allosteric systems,[34–37] including the original Monod *et al.* model of protein allostery.[49] The fluctuation–dissipation theorem[33] supports the idea that the *apo* protein ensemble should include fluctuations that are involved in shifting the conformational ensemble to the (de)activated state upon ligand binding. More recent experimental[25] and computational[27,50] studies further support this view of protein allostery. In this study, correlations between residues resulting from fluctuations in atomic position and noncovalent contacts are interpreted as reporting on important interactions for the allosteric mechanism of NikR.

The residues in Table 2 and Fig. 7 reveal a number of interesting sites for further experimental analysis. The computational significance column of Table 2 indicates which of the analyses presented in this study found at least one instance of each residue. Several of the residues were identified by more than one computational analysis, and those with unknown experimental significance were included in the table based on being highly interconnected in the contact correlation analysis. Those residues marked with "#" in Table 2 had the largest total number of significant contact correlations; these can be thought of as "hubs" in the interaction network that were selected in an unbiased way without assuming knowledge of interactions that are important for NikR function. The association of hub residues with function in Table 2 suggests that it is possible to apply our contact correlation method without prior knowledge of important functional sites. The "unweighted pair group method with arithmetic mean" (UPGMA) clustering method used here does not depend on such knowledge and simply provides a hierarchical tree of significant contact correlations, which can be parsed in a variety of ways. However, because the Ni^{2+}- and DNA-binding site residues are known, that information can also be used to select a subset of "important" contact correlations. The residues marked with "*" in Table 2 had the largest number of these correlations selected based on the functional criteria described in the Clustering section of Methods and Theory. Residues marked with both "*" and "#" could be particularly important for transducing the Ni^{2+} binding signal and therefore of significant interest for future experimental study. Residues that have "unknown" experimental significance in Table 2 have not yet been tested but are positions at which mutations are *predicted* to alter NikR function. The majority of the residues in Table 2 are also conserved or have conservative mutations for other NikR orthologs with known structures.[19,20] A representation of the MSA used to determine conservation is provided by the sequence logo in Fig. 8. The sequence logo shows that the NikR family has several highly conserved residues. For the purpose of selecting mutagenesis targets, our MD approach provides additional data to help develop hypotheses for why residues in different regions of the protein are conserved. The analyses presented here provide additional rationale for interpreting the effects of NikR mutations.

A subset of the residues from Table 2 are highlighted in Fig. 7. Figure 7a shows a close-up view of residues that connect the Ni^{2+}-binding site with the RHH/ACT domain interface. H76 is a Ni^{2+}-binding residue that forms a correlated contact with H62′ across the tetrameric interface. H62′ in turn contacts S73 back across the tetrameric interface. S73 is the $i+4$ residue to S69 along helix C. E63′ is covalently connected to H62′ and contacts R65′. R65′ and R65 form nonspecific phosphate backbone contacts with DNA[18] and also span the tetrameric interface to contact S69. In addition, R65 forms a correlated contact with G120, a highly conserved residue at the RHH/ACT interface. R70 is covalently attached to

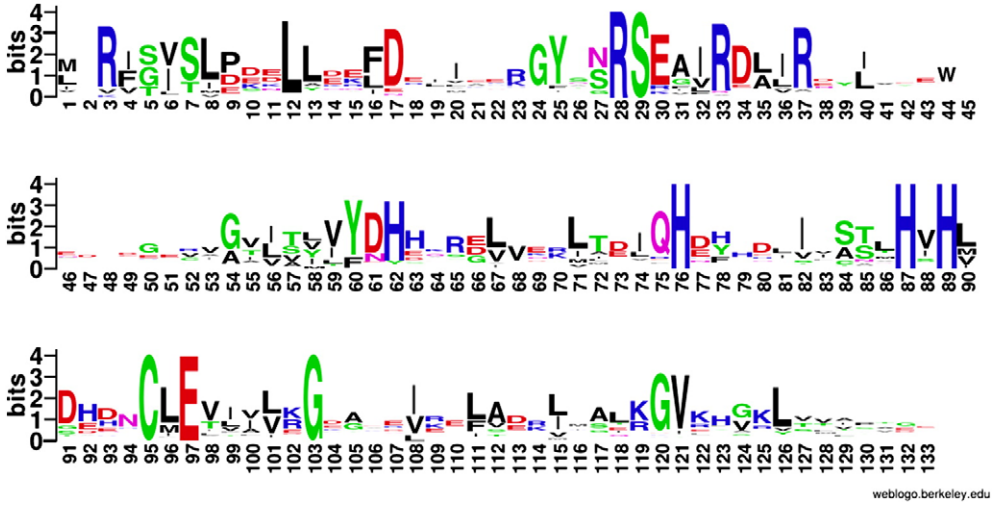

Fig. 8. Sequence logo for the NikR family. This logo represents an MSA containing 82 sequences, numbered according to the *E. coli* NikR sequence. For each position, the total height (in bits) of the residue letters indicates the degree of conservation at that position.[51] Note that this sequence entropy measure of conservation is different from the method used for Table 2, which takes into account substitution matrices. This figure was generated using WebLogo.[52]

S69 and forms a correlated contact with Q118 at the RHH/ACT domain interface. Q118 is also involved in forming a nonspecific cation-binding site that stabilizes the DNA-bound conformation.[18] This network of interactions is likely to be important to communicate the Ni^{2+} binding signal to the RHH/ACT interface.

Figure 7b depicts the DNA-binding site residues colored in red. Residue S29, near the N-terminus of helix B, contacts residue T7 and could help orient this DNA-binding group. R33 and R37, both near the middle of helix B, also make contacts with either DNA-binding residues or D9 and D11, residues at the junction between the DNA-binding β-strands and the N-terminal end of helix A. On the other face of helix B, D34 and N42 both make contacts that span the domain interface to H123 and N127, respectively. R22, at the C-terminal end of helix A, also forms a domain interface spanning contact with D113. E30 forms several contacts over the course of the simulation that tie together the N-terminal end of helix B with the C-terminal end of helix A and thus could help transmit effects from the RHH/ACT interface across the RHH domain to the DNA-binding site.

Of the residues identified in this study, contacts at the interface of the ACT and RHH domains are of obvious interest given the necessary interaction of these domains in Ni^{2+}-induced conformational change. However, residues within RHH domain helices also merit attention. Some of the contact correlations include backbone–backbone H bonds running along helices A and B, which might suggest a concerted, rigid-body response in transferring the Ni^{2+} binding status to the DNA-binding interface, and *vice versa*. Perturbing interactions between residues outside the DNA and Ni^{2+}-binding sites may therefore uncouple the DNA binding response under saturating Ni^{2+} conditions. Such perturbations will be the subject of future mutagenesis studies.

Implications for the ACT domain family

To our knowledge, this study is the first reported atomic-scale MD simulation of a protein containing an ACT domain. This domain family is found in a variety of contexts with essentially no conserved sequence homology between ACT proteins with different functions.[12] However, the extraordinary degree of structural similarity between the regulatory domains of these proteins leads us to hypothesize that there could be a common regulatory mechanism of this fold regardless of associated "biological function" domains. A structural alignment of the ACT domains from *E. coli* NikR and *E. coli* D-3-phosphoglycerate dehydrogenase (PGDH) (data not shown) shows that some experimentally identified residues that are important for PGDH function[53–56] align well with residues listed in Table 2. Therefore, the ACT domain residues identified in this study in the context of NikR function might also represent key positions for transferring allosteric effects in many ACT domain-containing systems. Several comparative studies of protein fold families have fruitfully determined common mechanisms of action and overlapping control points in approximately congruent structures.[57–61] However, in some proteins with similar folds, even those with the same biological function in different organisms, slight differences in sequence appear to generate differences in molecular mechanism.[62–64] These observations suggest that the analyses of NikR in this study may be useful for guiding computational and experimental work in other ACT domain proteins; however, such work should be undertaken with caution and awareness of the complex relationship between sequence and function.

Methods and Theory

Molecular dynamics simulation

The NikR MD simulation utilized the AMBER 8.0 molecular modeling package[41,65] with the ff99 force field.[66] The starting structure for simulation was based on the X-ray crystal structure of the *apo*NikR tetramer (PDB ID 1Q5V)[6] with missing backbone atoms reconstructed by symmetry between the NikR monomers (generously provided by Eric Schreiter). The WHAT IF molecular modeling package was used to rebuild missing amino acid side chains and eliminate steric overlap through geometry optimization.[67] The minimized structure was solvated in a periodic truncated octahedron simulation box of ~28,000 TIP3P water molecules,[68] providing a minimum of 10 Å of water between the protein surface and any periodic box edge. Sodium and chloride ions were added to neutralize the total system and achieve a salt concentration of ~150 mM (as estimated by the simulation ion mole fraction and bulk water molarity). AMBER 8.0 was used to first energy-minimize solvent and ions using 5000 steps of steepest descent followed by 5000 steps of conjugate gradient minimization. The entire system was then energy-minimized using the same procedure. Following minimization, the entire system was heated in 50 K increments up to 298 K with 10 ps of isobaric–isothermal (NpT) MD equilibration per temperature step. The production simulation was conducted at 298 K and 1 atm of pressure with the NpT ensemble using the Berendsen thermostat[69] with 1.0-ps coupling frequency and the Berendsen barostat[69] with 0.2-ps coupling frequency. The trajectory was calculated with 2-fs time steps using SHAKE[70] constraints on hydrogen–heavy atom bonds. The total production simulation length is 81.78 ns, of which the first 30.78 ns were discarded as an "equilibration period." This extensive relaxation/equilibration period was necessary due to drift in the potential energy, which correlates with changes in other observables (see Results for a description of criteria used for equilibration). Snapshots were retained every 10 ps for analysis.

Simulation analyses

Equilibration measures

The sander module of AMBER provides energy output in a text file that was parsed to obtain energies as a function of time. The ptraj module of AMBER 9.0 was used to write out the backbone RMSD of atomic position relative to the starting structure. The DSSP algorithm[40] implemented in the ptraj module of AMBER 9.0 was used to calculate relative secondary-structure content for each snapshot as a function of time. Energy, RMSD, and secondary-structure plots *versus* time (see Supplementary Material) were all generated using the Xmgrace software package§ and used to assess simulation equilibration.

Principal component analysis

As it relates to MD, PCA involves diagonalization of the positional covariance matrix $\underline{\underline{C}}$ to identify an orthogonal set of eigenvectors or "modes" describing directions of maximum variation in the observed conformational distribution.[42–45] The elements of $\underline{\underline{C}}$ in Cartesian coordinate space are defined as follows:

$$c_{ij} = \left\langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \right\rangle \tag{1}$$

where x_i and x_j are atomic coordinates and the $\langle \ldots \rangle$ denote trajectory averages. Note that the protein structures from the trajectory are superimposed to a reference structure to remove overall translational and rotational motion prior to the calculation of $\underline{\underline{C}}$. PCA diagonalization of the covariance matrix involves the following eigenvalue problem:

$$\underline{\underline{C}}\,\underline{u}^{\alpha} = \lambda^{\alpha}\underline{u}^{\alpha} \tag{2}$$

for the eigenvectors $u^{(\alpha)}$ and the eigenvalues $\lambda^{(\alpha)}$.[71] As with related methods such as singular value decompositions[72] and isomaps,[73] one motivation for PCA is to reduce the dimensionality of the MD trajectory data and provide a concise way to visualize, analyze, and compare large-scale collective motions observed over the course of the simulation. In particular, eigenvectors with the largest eigenvalues provide the biggest contributions to the observed covariance. The "essential modes" from a PCA analysis are usually a selection of these eigenvectors and associated eigenvalues that collectively account for a large percentage of the total observed motion.[71]

In this work, PCA modes are leveraged to identify large-scale *apo*NikR motions that are similar to structural transitions between the available NikR crystal structures. The PCA modes can be defined as displacement vectors from the average structure for the MD simulation. PCA for α carbon atoms was carried out using the ptraj module of Amber 9.0.[41] Using MATLAB 7.4 (The MathWorks, Natick, MA), we compared PCA modes (v_i) with a vector (Δx) describing the displacement of α carbon atoms from the minimized starting *apo*NikR structure (e.g., PDB entry 1Q5V[6] with missing atoms built in) to the DNA-bound Ni^{2+}NikR structure determined by X-ray crystallography. This vector was calculated following alignment of the minimized and crystal structures with the MD average structure to remove rotation and translation of the center of mass. We began by determining whether Δx can be reasonably represented in the vector space described by various subsets of PCA modes. This is accomplished by calculating the weight factor ($\alpha^{(i)}$) for each PCA mode as follows:

$$\alpha^{(i)} = \frac{v_i \cdot \Delta x}{||v_i||^2} \tag{3}$$

and subsequently using a subset S of modes to calculate a reconstructed vector ($\tilde{v}$) as follows:

$$\tilde{v} = \sum_{i \in S} \alpha^{(i)} v_i \tag{4}$$

The similarity between Δx and each PCA mode (v_i or $\tilde{v}$) is calculated by their overlap as measured by the angle between the two vectors:

$$\cos(\theta) = \frac{v \cdot \Delta x}{||v|| \cdot ||\Delta x||} \tag{5}$$

Furthermore, the relative error in the reconstructed vectors can be computed as:

$$\varepsilon = \frac{||\Delta x - \tilde{v}||}{||\Delta x||} \tag{6}$$

This error simply provides a measure of how well the displacement due to the recalculated vector recapitulates the observed crystallographic Δx and thus allows us to

§ http://plasma-gate.weizmann.ac.il/Grace/

assess the number of modes required to accurately represent the conformational change.

Clustering

To define groups of residues with similar correlation patterns, a UPGMA clustering algorithm was used.[74,75] To do this, an effective "distance" d_{ij} between contacts/residues i and j was defined based on a correlation measure, c_{ij}, as follows:

$$d_{ij} = \sqrt{1.0 - c_{ij}^2} \tag{7}$$

These distances between contacts/residues represent the strength of the relationship between them, with the "closest" contacts/residues having the largest magnitude correlations. In this work, we use this methodology for both the correlation matrix of atomic positions and of noncovalent contacts (see below). With an effective distance defined between all contacts/residues, it is possible to cluster them using a hierarchical agglomerative approach such as UPGMA.[74] This method yields a "treelike" representation of correlated residues in which individual residues are the "leaves" and clusters of residues are defined by different branch points in the tree at different "tree heights." With this type of data representation, one must select a level of the tree hierarchy for defining clusters of residues (see below).

UPGMA clustering depth-first selection

After clustering, functional selection criteria were applied to identify clusters containing a network of contacts/residues known to be important for NikR function. In particular, the smallest clusters containing at least one contact/residue from the Ni^{2+}-binding site (His76, His87, His89, Cys95) and at least one sequence-specific contact/residue from the DNA-binding site (Arg3, Thr5, or Thr7). This cluster-selection step is considered depth first and selects clusters that have the strongest relationship among members while still maintaining at least one contact/residue from each of the Ni^{2+}- and DNA-binding sites, in contrast to a "breadth-first" selection that requires all Ni^{2+}- and DNA-binding site contacts/residues are included. The depth-first approach selects a subset of total contacts/residues that have the strongest relationship with binding site contacts, producing a subset of NikR residues that form our selected groups of contacts/residues, which indicates positions in the structure that are likely involved in a communication network between the Ni^{2+}- and DNA-binding sites. Functional information plays an important role in these cluster definitions. However, as described in the Results section, it is also possible to use our correlation analysis to identify "highly connected" residues without the need for functional information.

UPGMA clustering "completion" step for cluster selection

An additional step of cluster selection was used when clustering residues based on the correlation matrix (see below) to ensure that all Ni^{2+} and DNA-binding residues were accounted for. Upon identification of the initial domain-spanning clusters by the depth-first approach, the largest clusters that contained at least one Ni^{2+}- or one DNA-binding site residue and no residues that had been found in the depth-first step were then identified. This accounts for the remaining binding-site residues at the same level in the tree hierarchy as the clusters identified in the depth-first cluster-selection step described above.

Correlations in atomic position

The scalar correlation matrix was calculated across all α carbon atoms of the NikR tetramer using the ptraj module of AMBER 9.0.[41] The elements of this matrix, s_{ij}, assign a value between −1.0 and 1.0 that indicates the degree to which the fluctuations of atom i are correlated with those of atom j over the course of the MD trajectory according to the following equation[76]:

$$s_{ij} = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{\sqrt{\langle \Delta r_i \cdot \Delta r_i \rangle \langle \Delta r_j \cdot \Delta r_j \rangle}} \tag{8}$$

where Δr_i and Δr_j are the displacement vectors for atoms i and j and the $\langle \ldots \rangle$ denotes trajectory averages.

Correlation patterns apparent in this matrix were analyzed (see Fig. 2) to define groups of NikR residues with similar correlation patterns. This grouping utilized the UPGMA clustering algorithm described above including the depth-first selection criteria and the "completion" step. The residues included in each cluster are listed in Supplementary Table S1.

Correlations in noncovalent contacts

Noncovalent contacts, including hydrogen bonds and salt bridges, were analyzed for correlations to help identify the residues involved in potential communication networks connecting the Ni^{2+}-binding sites to the DNA-binding domains. Nonpolar contacts were omitted from the correlation analysis for two major reasons. First, one of the goals of this study is to suggest positions for mutagenesis; alteration of a nonpolar contact is often more likely to adversely affect protein stability. Second, nonpolar contacts are more numerous throughout the protein structure and therefore more difficult to uniquely define as binary variables for correlation purposes. Individual contacts were treated as binary variables that were either "on" or "off" for each snapshot from the *apo*NikR MD trajectory. A useful measure of correlation between binary variables is the ϕ correlation metric,[77] a binary variant of the standard Pearson correlation‖. The ϕ correlation can range between −1.0 and +1.0, indicating complete negative and positive correlation, respectively. A ϕ value of 0.0 indicates lack of correlation. The calculation of ϕ values for all possible pairs of noncovalent contacts across the NikR MD trajectory proceeded as follows:

(1) A list of observed contacts was generated for each MD snapshot using the PDB2PQR¶ program.[78] For hydrogen bonds, lists of H-bond donor (D) and acceptor (A) heavy atoms were defined based on the following criteria: D to A distance ≤3.4 Å and the A–H–D angle ≤30°.[79] For salt bridges, both positively and negatively charged heavy atoms were defined by considering amino acid side chains that carry a formal charge. Salt bridges were then assigned whenever both a positively charged atom and a negatively charged atom were ≤4.0 Å.[80] To remove

‖http://www.amstat.org/publications/jse/v5n3/falk.html

¶http://pdb2pqr.sourceforge.net/

redundancy in counting multiple interactions between residues, contacts were defined and counted as side chain–side chain, side chain–backbone, or backbone–backbone. Contact lists for each snapshot only allowed one instance of each type of contact between any two residues within each snapshot.

(2) The contact lists for all snapshots were parsed in order to populate the contact occupancy matrix B, an $N \times N$ matrix where N is the total number of unique contacts observed throughout the entire simulation. The diagonal of B stores the total number of snapshots in which each contact is observed, and the i,j off-diagonal elements store the total number of snapshots in which both contacts i and j were observed.

(3) For each i,j contact pair, the following frequencies were calculated using B: n_{00}, the number of times both contacts i and j were off; n_{10}, the number of times contact i was on while contact j was off; n_{01}, the number of times contact i was off while contact j was on; n_{11}, the number of times both contacts i and j were on. Given these frequencies, the ϕ_{ij} correlation value was calculated from[77]:

$$\phi_{ij} = \frac{(n_{00}n_{11} - n_{10}n_{01})}{\sqrt{(n_{00}+n_{10})(n_{00}+n_{01})(n_{11}+n_{10})(n_{11}+n_{01})}} \quad (9)$$

The significance of a correlation value is generally dependent on the number of independent observations. In the case of ϕ correlation, there is a simple relationship that provides an effective chi-squared value for significance[77]:

$$\chi^2 = \phi^2 N \quad (10)$$

where N is the number of independent data observations used to compute the correlation. Because MD snapshots are intrinsically correlated over varying time scales depending on the observable of interest, N was estimated separately for each i,j pair of observed contacts. The number of times that each contact was both made and broken over the course of the simulation (representing the number of on and off states) was defined as n_i and used in the following equation:

$$N = \min(n_i, n_j)$$

For this definition of N, the χ^2 value should depend on the number of independent states observed for the variable of interest. In the current work, we approximate the number of independent states by the number of times a noncovalent contact is made or broken. The ϕ correlations between i,j pairs were considered significant when χ^2_{ij} is greater than or equal to the threshold value for the 95% confidence interval using 1 degree of freedom for the binary nature of the data.[77,81]

The set of statistically significant contact correlations gives a large data set with 91,934 elements for the 532-residue NikR tetramer. To parse these data and find correlations that may be important for NikR function, the contact correlation data were used as input for the UPGMA clustering algorithm described above. Only the depth-first cluster-selection step was used here. This cluster selection finds a smaller number of total contacts, which have the strongest relationship with binding site contacts. A subset of NikR residues form our selected groups of contacts (see Fig. 5 and Table S2 in Supplementary Material) and are likely involved in a communication network between the Ni^{2+}- and DNA-binding sites.

Multiple sequence alignment and positional conservation

The *E. coli* NikR sequence (FASTA sequence from PDB ID 1Q5V) was used as the input for the National Center for Biotechnology Information BLASTP[a] server.[82] The search was performed against the nonredundant database with default parameters except that sequences with an E value less than 1.0 were retained. An MSA was constructed using CLUSTALW 1.81[83] with the BLOSUM substitution matrix series and otherwise all default parameters. This alignment was hand-pruned to remove sequences that were significantly shorter than *E. coli* NikR or that introduced large gaps into the alignment. The resulting sequences were realigned with CLUSTALW. Sequences were removed from this alignment if they did not have the His76, His87, His89, and Cys95 (*E. coli* NikR numbering) Ni^{2+}-binding residues. The remaining sequences were realigned as above and the resulting MSA was further pruned such that no pair of sequences was greater than 80% identical, followed by an additional realignment. This process yielded a final MSA containing 82 sequences with an average sequence identity of 30.3%. The final MSA was used as input for the Web-based Scorecons[b] program[84] to calculate the positional conservation score for each position in the MSA. The valdar01 scoring method was used with the BLOSUM45 substitution matrix. The positional sequence conservation values are reported in parentheses in Table 2 under the "Sequence conservation" heading. Each position was then assigned a "high" (≥ 0.6), "moderate" (0.45–0.59), or "low" (≤ 0.44) level of sequence conservation.

Acknowledgements

This work was supported by National Science Foundation (NSF) grant MCB-0520877, by an NSF Graduate Research Fellowship to M.J.B., and by the Molecular Biophysics Training Grant (T32 GM008492). We thank Eric Schreiter (now at the University of Puerto Rico) for providing an initial *apo*NikR structure with missing backbone density built in by symmetry; Rohit Pappu and Matt Wyczalkowski for helpful suggestions about simulation analysis and bootstrap methodology. We would also like to thank the anonymous reviewers for their helpful comments.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2008.03.010

[a] http://www.ncbi.nlm.nih.gov/blast/

[b] http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/valdar/scorecons_server.pl

References

1. Vignais, P. M., Billoud, B. & Meyer, J. (2001). Classification and phylogeny of hydrogenases. *FEMS Microbiol. Rev.* **25**, 455–501.
2. Navarro, C., Wu, L. F. & Mandrand-Berthelot, M. A. (1993). The nik operon of *Eschericia coli* encodes a periplasmic binding-protein-dependent transport system for nickel. *Mol. Microbiol.* **9**, 1181–1191.
3. De Pina, K., Desjardin, V., Mandrand-Berthelot, M. A., Giordano, G. & Wu, L. F. (1999). Isolation and characterization of the *nikR* gene encoding a nickel-responsive regulator in *Escherichia coli*. *J. Bacteriol.* **181**, 670–674.
4. Chivers, P. T. & Sauer, R. T. (2000). Regulation of high affinity nickel uptake in bacteria. Ni^{2+}-dependent interaction of NikR with wild-type and mutant operator sites. *J. Biol. Chem.* **275**, 19735–19741.
5. Iwig, J. S., Rowe, J. L. & Chivers, P. T. (2006). Nickel homeostasis in *Escherichia coli*—the rcnR–rcnA efflux pathway and its linkage to NikR function. *Mol. Microbiol.* **62**, 252–262.
6. Schreiter, E. R., Sintchak, M. D., Guo, Y., Chivers, P. T., Sauer, R. T. & Drennan, C. L. (2003). Crystal structure of the nickel-responsive transcription factor NikR. *Nat. Struct. Biol.* **10**, 794–799.
7. Chivers, P. T. & Sauer, R. T. (1999). NikR is a ribbon–helix–helix DNA-binding protein. *Protein Sci.* **8**, 2494–2500.
8. Schreiter, E. R. & Drennan, C. L. (2007). Ribbon–helix–helix transcription factors: variations on a theme. *Nat. Rev. Microbiol.* **5**, 710–720.
9. Chivers, P. T. & Sauer, R. T. (2002). NikR repressor: high-affinity nickel binding to the C-terminal domain regulates binding to operator DNA. *Chem. Biol.* **9**, 1141–1148.
10. Wang, S. C., Dias, A. V., Bloom, S. L. & Zamble, D. B. (2004). Selectivity of metal binding and metal-induced stability of *Escherichia coli* NikR. *Biochemistry*, **43**, 10018–10028.
11. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD - Visual Molecular Dynamics. *J. Molec. Graphics*, **14**, 33–38.
12. Grant, G. A. (2006). The ACT domain: a small molecule binding domain and its role as a common regulatory element. *J. Biol. Chem.* **281**, 33825–33829.
13. Rowe, J. L., Starnes, G. L. & Chivers, P. T. (2005). Complex transcriptional control links NikABCDE-dependent nickel transport with hydrogenase expression in *Escherichia coli*. *J. Bacteriol.* **187**, 6317–6323.
14. Bloom, S. L. & Zamble, D. B. (2004). Metal-selective DNA-binding response of *Escherichia coli* NikR. *Biochemistry*, **43**, 10029–10038.
15. Dias, A. V. & Zamble, D. B. (2005). Protease digestion analysis of *Escherichia coli* NikR: evidence for conformational stabilization with Ni(II). *J. Biol. Inorg. Chem.* **10**, 605–612.
16. Leitch, S., Bradley, M. J., Rowe, J. L., Chivers, P. T. & Maroney, M. J. (2007). Nickel-specific response in the transcriptional regulator, *Escherichia coli* NikR. *J. Am. Chem. Soc.* **129**, 5085–5095.
17. Carrington, P. E., Chivers, P. T., Al-Mjeni, F., Sauer, R. T. & Maroney, M. J. (2003). Nickel coordination is regulated by the DNA-bound state of NikR. *Nat. Struct. Biol.* **10**, 126–130.
18. Schreiter, E. R., Wang, S. C., Zamble, D. B. & Drennan, C. L. (2006). NikR–operator complex structure and the mechanism of repressor activation by metal ions. *Proc. Natl. Acad. Sci. USA*, **103**, 13676–13681.
19. Dian, C., Schauer, K., Kapp, U., McSweeney, S. M., Labigne, A. & Terradot, L. (2006). Structural basis of the nickel response in *Helicobacter pylori*: crystal structures of HpNikR in Apo and nickel-bound states. *J. Mol. Biol.* **361**, 715–730.
20. Chivers, P. T. & Tahirov, T. H. (2005). Structure of *Pyrococus horikoshii* NikR: nickel sensing and implications for the regulation of DNA recognition. *J. Mol. Biol.* **348**, 597–607.
21. Raumann, B. E., Rould, M. A., Pabo, C. O. & Sauer, R. T. (1994). DNA recognition by beta-sheets in the Arc repressor–operator crystal structure. *Nature*, **367**, 754–757.
22. Sauer, R. T., Milla, M. E., Waldburger, C. D., Brown, B. M. & Schildbach, J. F. (1996). Sequence determinants of folding and stability for the P22 Arc repressor dimer. *FASEB J.* **10**, 42–48.
23. Brown, B. M., Milla, M. E., Smith, T. L. & Sauer, R. T. (1994). Scanning mutagenesis of the Arc repressor as a functional probe of operator recognition. *Nat. Struct. Biol.* **1**, 164–168.
24. Cui, G. & Merz, K. M. (2008). The intrinsic dynamics and function of nickel binding regulatory protein: insights from elastic network analysis. *Biophys. J.* doi:10.1529/biophysj.107.115576.
25. Volkman, B. F., Lipson, D., Wemmer, D. E. & Kern, D. (2001). Two-state allosteric behavior in a single-domain signaling protein. *Science*, **291**, 2429–2433.
26. Pan, H., Lee, J. C. & Hilser, V. J. (2000). Binding sites in *Escherichia coli* dihydrofolate reductase communicate by modulating the conformational ensemble. *Proc. Natl. Acad. Sci. USA*, **97**, 12020–12025.
27. Formaneck, M. S., Ma, L. & Cui, Q. (2006). Reconciling the "old" and "new" views of protein allostery: a molecular simulation study of chemotaxis Y protein (CheY). *Proteins*, **63**, 846–867.
28. Ghosh, A. & Vishveshwar, S. (2007). A study of communication pathways in methionyl- tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc. Natl. Acad. Sci. USA*, **104**, 15711–15716.
29. Swain, J. F. & Gierasch, L. M. (2006). The changing landscape of protein allostery. *Curr. Opin. Struct. Biol.* **16**, 102–108.
30. Velyvis, A., Yang, Y. R., Schachman, H. K. & Kay, L. E. (2007). A solution NMR study showing that active site ligands and nucleotides directly perturb the allosteric equilibrium in aspartate transcarbamoylase. *Proc. Natl. Acad. Sci. USA*, **104**, 8815–8820.
31. Chennubhotla, C. & Bahar, I. (2007). Signal propagation in proteins and relation to equilibrium fluctuations. *PLoS Comput. Biol.* **3**, e172.
32. Tobi, D. & Bahar, I. (2005). Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. USA*, **102**, 18908–18913.
33. Kubo, R. (1966). The fluctuation–dissipation theorem. *Rep. Prog. Phys.* **29**, 255–284.
34. Radkiewicz, J. L. & Brooks, C. L., III (2000). Protein dynamics in enzymatic catalysis: exploration of dihydrofolate reductase. *J. Am. Chem. Soc.* **122**, 225–231.
35. Ma, J., Sigler, P. B., Xu, Z. & Karplus, M. (2000). A dynamic model for the allosteric mechanism of GroEL. *J. Mol. Biol.* **302**, 303–313.
36. Tai, K., Shen, T., Borjesson, U., Philippopoulos, M. & McCammon, J. A. (2001). Analysis of a 10-ns

molecular dynamics simulation of mouse acetylcholinesterase. *Biophys. J.* **81**, 715–724.
37. Perryman, A. L., Lin, J.-H. & McCammon, J. A. (2004). HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. *Protein Sci.* **13**, 1108–1123.
38. Forrest, L. R., Kukol, A., Arkin, I. T., Tieleman, D. P. & Sansom, M. S. P. (2000). Exploring models of influenza A M2 channel: MD simulations in a phospholipid bilayer. *Biophys. J.* **78**, 55–69.
39. Gullingsrud, J., Kosztin, D. & Schulten, K. (2001). Structural determinants of MscL gating studied by molecular dynamics simulations. *Biophys. J.* **80**, 2074–2081.
40. Kabsch, W. & Sander, C. (1983). DSSP: definition of secondary structure of proteins given a set of 3D coordinates. *Biopolymers*, **22**, 2577–2637.
41. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M. *et al.* (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688.
42. Hayward, S., Kitao, A. & Go, N. (1995). Harmonicity and anharmonicity in protein dynamics: a normal mode analysis and principal component analysis. *Proteins*, **23**, 177–186.
43. Garcia, A. E. (1992). Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* **68**, 2696–2699.
44. Leach, A. R. (2001). *Molecular Modelling Principles and Applications*, 2nd edit. Pearson Education Limited, Essex, UK.
45. Levy, R. M., Srinivasan, A. R., Olson, W. K. & McCammon, J. A. (1984). Quasi-harmonic method for studying very low frequency modes in proteins. *Biopolymers*, **23**, 1099–1112.
46. Rod, T. H., Radkiewicz, J. L. & Brooks, C. L., III (2003). Correlated motion and the effect of distal mutations in dihydrofolate reductase. *Proc. Natl. Acad. Sci. USA*, **100**, 6980–6985.
47. Cheng, X., Ivanov, I., Wang, H., Sine, S. M. & McCammon, J. A. (2007). Nanosecond time scale conformational dynamics of the human alpha7 nicotinic acetylcholine receptor. *Biophys. J.* **93**, 2622–2634.
48. Daily, M. D., Upadhyaya, T. J. & Gray, J. J. (2007). Contact rearrangements form coupled networks from local motions in allosteric proteins. *Proteins* (EPub).
49. Monod, J., Wyman, J. & Changeux, J. P. (1965). On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12**, 88–118.
50. Arora, K. & Brooks, C. L., III (2007). Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. *Proc. Natl. Acad. Sci. USA*, **104**, 18496–18501.
51. Crooks, G. E. & Brenner, S. E. (2004). Protein secondary structure: entropy, correlations and prediction. *Bioinformatics*, **20**, 1603–1611.
52. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190.
53. Schuller, D., Grant, G. A. & Banaszak, L. (1995). Crystal structure reveals the allosteric ligand site in the V_{max}-type cooperative enzyme: D-3-phosphoglycerate dehydrogenase. *Nat. Struct. Biol.* **2**, 69–76.
54. Bell, J. K., Grant, G. A. & Banaszak, L. (2004). Multiconformational states in phosphoglycerate dehydrogenase. *Biochemistry*, **43**, 3450–3458.
55. Thompson, J. R., Bell, J. K., Bratt, J., Grant, G. A. & Banaszak, L. (2005). V_{max} regulation through domain and subunit changes. The active form of phosphoglycerate dehydrogenase. *Biochemistry*, **44**, 5763–5773.
56. Grant, G. A., Hu, Z. & Xu, X. L. (2005). Identification of amino acid residues contributing to the mechanism of cooperativity in *Eschericia coli* D-3-phosphoglycerate dehydrogenase. *Biochemistry*, **44**, 16844–16852.
57. Hall, B. M., LeFevre, K. R. & Cordes, M. H. J. (2005). Sequence correlations between Cro recognition helices and cognate OR consensus half-sites suggest conserved rules of protein–DNA recognition. *J. Mol. Biol.* **350**, 667–681.
58. Kang, S.-G. & Saven, J. G. (2007). Computational protein design: structure, function and combinatorial diversity. *Curr. Opin. Chem. Biol.* **11**, 329–334.
59. Shrivastava, I. & Bahar, I. (2006). Common mechanism of pore opening shared by five different potassium channels. *Biophys. J.* **90**, 3929–3940.
60. Dima, R. I. & Thirumalai, D. (2006). Determination of network of residues that regulate allostery in protein families using sequence analysis. *Protein Sci.* **15**, 258–268.
61. del Sol, A., Fujihashi, H., Amoros, D. & Nussinov, R. (2006). Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol. Syst. Biol.* **2**. doi:10.1038/msb4100063.
62. Scrutton, N. S., Deonorain, M. P., Berry, A. & Perham, R. N. (1992). Cooperativity induced by a single mutation at the subunit interface of a dimeric enzyme: glutathione reductase. *Science*, **258**, 1140–1143.
63. Kuo, L. C., Zambidis, I. & Caron, C. (1989). Triggering of allostery in an enzyme by a point mutation: ornithine transcarbamoylase. *Science*, **245**, 522–524.
64. Heddle, J. G., Okajima, T., Scott, D. J., Akashi, S., Park, S.-Y. & Tame, J. R. H. (2007). Dynamic allostery in the ring protein TRAP. *J. Mol. Biol.* **371**, 154–167.
65. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., Debolt, S. *et al.* (1995). Amber, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **91**, 1–41.
66. Wang, J., Cieplak, P. & Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049–1074.
67. Vriend, G. (1990). WHAT IF: A molecular modeling and drug design program. *J. Mol. Graphics*, **8**, 52–56.
68. Jorgensen, W., Chandrasekhar, J. & Madura, J. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
69. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690.
70. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341.
71. Amadei, A., Linssen, A. B. M. & Berendsen, H. J. C. (1993). Essential dynamics of proteins. *Proteins*, **17**, 412–425.
72. Doruker, P., Atilgan, A. R. & Bahar, I. (2000). Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor. *Proteins*, **40**, 512–524.

73. Plaku, E., Stamati, H., Clementi, C. & Kavraki, L. (2007). Fast and reliable analysis of molecular motion using proximity relations and dimensionality reduction. *Proteins*, **67**, 897–907.
74. Sneath, P. H. A. & Sokal, R. R. (1973). *Numerical Taxonomy.* W.H. Freeman and Company, San Francisco, CA.
75. Durbin, R., Eddy, S., Krogh, A. & Mitchinson, G. (1998). *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids.* Cambridge University Press, New York.
76. Ichiye, T. & Karplus, M. (1991). Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins*, **11**, 205–217.
77. Bailey, N. T. J. (1995). *Statistical Methods in Biology*, 3rd edit. Cambridge University Press, Cambridge, UK.
78. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson–Boltzman electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667.
79. Mezei, M. & Beveridge, D. L. (1981). Theoretical studies of hydrogen bonding in liquid water and dilute aqueous solutions. *J. Chem. Phys.* **74**, 622–632.
80. Barlow, D. J. & Thornton, J. M. (1983). Ion-pairs in proteins. *J. Mol. Biol.* **168**, 867–885.
81. Bruning, J. L. & Kintz, B. L. (1997). *Computational Handbook of Statistics*, 4th edit. Addison Wesley Longman Inc., Reading, MA.
82. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
83. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
84. Valdar, W. S. J. (2002). Scoring residue conservation. *Proteins*, **48**, 227–241.