ATP Binding Enables Broad Antibiotic Selectivity of Aminoglycoside Phosphotransferase(3′)-IIIa: An Elastic Network Analysis

Silke A. Wieninger[1], Engin H. Serpersu[2] and G. Matthias Ullmann[1]*

[1]*Computational Biochemistry/Bioinformatics, University of Bayreuth, Universitätsstr. 30, BGI, 95447 Bayreuth, Germany*

[2]*Department of Biochemistry and Cellular and Molecular Biology, The University of Tennessee, Knoxville, TN 37996, USA*

The bacterial enzyme aminoglycoside phosphotransferase(3′)-IIIa (APH) confers resistance against a wide range of aminoglycoside antibiotics. In this study, we use the Gaussian network model to investigate how the binding of nucleotides and antibiotics influences the dynamics and thereby the ligand binding properties of APH. Interestingly, in NMR experiments, the dynamics differ significantly in various APH complexes, although crystallographic studies indicate that no larger conformational changes occur upon ligand binding. Isothermal titration calorimetry also shows different thermodynamic contributions to ligand binding. Formation of aminoglycoside–APH complexes is enthalpically driven, while the enthalpic change upon aminoglycoside binding to the nucleotide–APH complex is much smaller. The differential effects of nucleotide binding and antibiotic binding to APH can be explained theoretically by single-residue fluctuations and correlated motions of the enzyme. The surprising destabilization of β-sheet residues upon nucleotide binding, as seen in hydrogen/deuterium exchange experiments, shows that the number of closest neighbors does not fully explain residue flexibility. Additionally, we must consider correlated motions of dynamic protein domains, which show that not only connectivity but also the overall protein architecture is important for protein dynamics.

Introduction

Pathogenic bacteria increasingly evolve mechanisms to evade the effects of antibiotics in use. Infection with multidrug-resistant strains has become a serious health problem, especially in hospitals.[1] A common strategy of bacterial resistance is chemical modification of antibiotics, as catalyzed by aminoglycoside phosphotransferases. Aminoglycosides are a large group of structurally diverse antibiotics that bind to the 30S ribosome and prevent proper bacterial protein translation.[2] The modification enzyme aminoglycoside phosphotransferase(3′)-IIIa (APH) catalyzes the MgATP-dependent phosphorylation of the 3′-OH or the 5″-OH of aminoglycosides and thereby decreases the antibiotic's affinity for ribosomal RNA.[3,4] APH has been extensively studied by kinetic and thermodynamic means because it has the broadest substrate selectivity among the APH isozymes, phosphorylating more than 10 different

---

Corresponding author. E-mail address: Matthias.Ullmann@uni-bayreuth.de.

Abbreviations used: APH, aminoglycoside phosphotransferase(3′)-IIIa; H/D, hydrogen/deuterium; ITC, isothermal titration calorimetry; ENM, elastic network model; GNM, Gaussian network model; ANM, anisotropic network model; PDB, Protein Data Bank.

aminoglycosides of various sizes and structures.[5] How APH achieves such a broad substrate selectivity is still not understood. As reported earlier by one of us, ligand-dependent changes in the dynamic behavior of APH were detected by NMR, hydrogen/deuterium (H/D) exchange experiments,[6] and isothermal titration calorimetry (ITC),[7] and gave insights into the mechanism of the substrate promiscuity of APH. While aminoglycoside binding to the apoenzyme induces an enhanced conformational stability of APH and a reduction of entropy, the enthalpic change upon aminoglycoside binding to the nucleotide–APH complex is about 20 kcal mol^{-1} less negative, which is compensated for by a smaller entropic penalty. Also, different aminoglycosides show differing effects on the dynamic behavior of APH.

In this work, we use the elastic network model (ENM) to analyze the molecular basis of the differential effects of nucleotide binding and antibiotic binding to APH. The dynamics of proteins has been successfully studied by such coarse-grained protein models with a simple energy function.[8–14] The ENM describes the protein as a collection of mass points connected by springs and allows the determination of single-residue fluctuations and collective protein motions based on purely topological constraints arising from the three-dimensional protein structure. Despite its simplicity, the ENM was shown to predict harmonic protein motions well and to be comparable to full-atom normal-mode analysis.[15] Obvious advantages of coarse-grained methods are speed and memory requirements. Moreover, minimalist models can reveal patterns that would be hidden in complex and detailed data. In our study, we apply the Gaussian network model (GNM), a specific variant of the ENM.[16]

ENMs are often applied to proteins with known alternative conformations (e.g., an open ligand-free state and a closed ligand-bound state). Based on different crystal structures, changes in dynamics[17–19] and conformational changes[15,20–27] upon ligand binding are investigated. Our work differs in that known X-ray structures of APH crystallized in various binary and ternary complexes,[28,29] except for loop regions, are very similar to each other. Thus, the differences in dynamics mainly arise from the presence or the absence of ligand, and not from the conformational change induced by the binding of the ligand. Generally, it is assumed that ligand binding leads to reduced H/D exchange due to steric exclusion of solvent.[30,31] The surprising experimental observation that nucleotide binding destabilizes β-sheet residues[6] contradicts this idea and leads to the identification of dynamic protein domains. On the other hand, the calculation of theoretical *B*-factors affirms that residue flexibility decreases with the number of closest neighbors. We introduce connectivity *B*-factors to investigate the importance

of connectivity on flexibility. From our analysis, we conclude that the number of closest spatial neighbors is important for dynamics but, by far, is not its only determinant. Also the overall architecture of the protein influences the dynamics of the protein. By comparing the binding of the two antibiotics kanamycin A and neomycin B to APH, we show that a small change in the number of nodes composing the elastic network leads to detectable differences in protein dynamics.

Results and Discussion

Identification of dynamic protein domains

Visual inspection of APH structures (Fig. 1a) does not allow us to divide the protein into domains. However, the dynamic properties of the protein suggest the existence of several domains. One possibility of identifying such dynamic domains is to search for the most rigid parts of the molecule.[34–36] We follow a different approach, where the definition of domains is based on correlations of fluctuations calculated by the GNM.[37] The correlation plot of the apo form of APH is shown in Fig. 1c, where the degree of correlation is indicated by the intensity of colors. Based on these correlations, three dynamic domains of APH are identified using the algorithm described in Methods. Residues 2–91 comprising the N-terminal α-helix, a five-stranded antiparallel β-sheet, and a second α-helix define domain I. One can easily recognize a large area of positive correlations of these first residues to residues lying in the same domain, while they are clearly anticorrelated to the rest of the enzyme. Residues 92–129 and 180–249 define domain II, which is the largest domain. It is the core domain that is connected to the two other domains and contributes to the binding of both substrates (i.e., ATP and the aminoglycosides). Finally, a helix–loop–helix section spanning residues 130–179, together with the C-terminal helix of residues 250–264, defines domain III, which also shows typical positive intradomain correlations and negative interdomain correlations. Figure 1a shows the dynamic domain structure of APH and illustrates how the nucleotide substrate binds in the cleft formed by domains I and II, while the aminoglycoside substrate binds between all three domains. As several X-ray structures of APH crystallized in the presence of different ligands are available, we can test the effect of structural differences on the identification of the dynamic domains. In the following, APH structures will be indicated by *X*, followed by their substrate content. For example, *X* [APH-Nuc-Neo] denotes the ternary complex structure of APH crystallized in the presence of MgADP
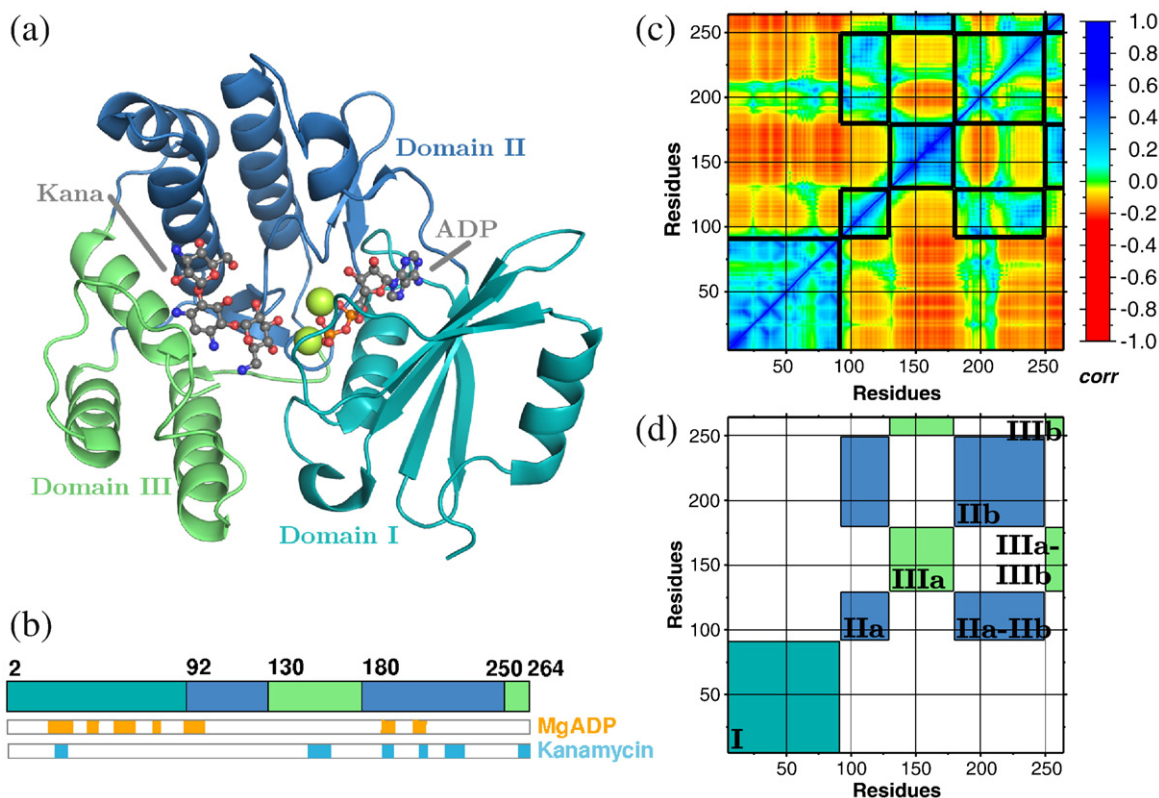
Fig. 1. Definition of the dynamic domains of APH based on correlations calculated using the GNM. (a) Structural representation of the domains. Domain I (cyan), domain II (blue), and domain III (green) of APH are shown with the substrates MgADP and kanamycin A bound between the domains. Domain I consists of an N-terminal α-helix, a five-stranded antiparallel β-sheet, and a second α-helix. Core domain II is the largest and is involved in the binding of both substrates. Domain III consists of a helix–loop–helix motif and a C-terminal α-helix. The image was produced using PyMOL.[32] (b) The bar indicates the sequential assignment of residues to the protein domains and the positions of residues that build the substrate binding sites. (c) Correlated motions of apo-APH (X[APH]). The enzyme is divided into three domains according to the correlations between the residues. Correlations between residues of the same domain are mostly positive, while the intersections between different APH domains show negative correlations. The image was produced using GMT.[33] (d) The principal correlation pattern of APH is used for the classification of residues in dynamic domains. Areas of positive correlation along the diagonal arise from sequential residues that belong to the same domain. Off-diagonal areas of positive correlations indicate that the corresponding residues belong to the same domain, although they are not sequential. Domains II and III consist of two nonsequential parts, each labeled (a) and (b).

and neomycin, whereas X[APH] denotes the crystal structure of apo-APH. All further abbreviations are given in Table 1. Ligand binding changes the interactions in the elastic network, as will be demonstrated in Influence Of Substrate Binding on Correlated Motions. Therefore, we remove the ligand nodes from the structures for the calculation of dynamic domains. Model complexes of the structures with changed substrate content are indicated by M (Table 1). The classification in dynamic domains, calculated for all structures, is given in Table S1 of Supplemental Data. The principal domain pattern is conserved across all structures, although some loop nodes that lie on the border of the domains are assigned differently. In $M_{kan}[APH]$ and $M_{neo}[APH]$, domain I additionally contains nodes 92–95. In $M_{nuc}[APH]$, nodes 92–95

and 197–202 additionally belong to domain I. But for all structures, three domains were identified. This is remarkable because the number of domains was not obvious a priori.

Influence of substrate binding on correlated motions

The influence of substrate binding on dynamics and thermodynamics has been studied by H/D exchange and ITC. We compare our theoretical predictions with these experimental data. In H/D exchange, residues located in flexible regions of proteins are more frequently exposed to solvent; therefore, their exchangeable protons usually display reduced protection in H/D exchange experiments. H/D exchange data are available for all

Table 1. Crystal structures and derived model complexes of APH

Abbreviation	Description
X[APH]	Apo-APH (PDB ID: 1J7I)[28]
X[APH-Nuc]	APH–MgAMPPNP complex (PDB ID: 1J7U)[28]
X[APH-Nuc-Kan]	APH–MgADP–kanamycin complex (PDB ID: 1L8T)[29]
X[APH-Nuc-Neo]	APH–MgADP–neomycin complex (PDB ID: 2B0Q)[29]
M_{kan}[APH]	X[APH-Nuc-Kan], MgADP and kanamycin deleted
M_{kan}[APH-Nuc]	X[APH-Nuc-Kan], kanamycin deleted
M_{kan}[APH-Kan]	X[APH-Nuc-Kan], nucleotide deleted
M_{neo}[APH]	X[APH-Nuc-Neo], MgADP and neomycin deleted
M_{neo}[APH-Neo]	X[APH-Nuc-Neo], nucleotide deleted
M_{nuc}[APH]	X[APH-Nuc], nucleotide deleted

complexes of APH.[6] Briefly, H/D exchange studied by NMR revealed that APH exchanges all of its backbone amides within 15–20 h of exposure to D_2O, which is extremely unusual for a 31-kDa protein. At least a certain set of amide protons, particularly those buried in the hydrophobic core of the protein or those lying on the β-sheets of a structured protein of this size, is expected to remain unexchanged upon exposure to D_2O for a much longer period of time. The presence of bound nucleotide only marginally affects the H/D exchange profile and allows protection of <10 residues. In contrast to these findings, 30–40% of the backbone amides are protected against exchange for more than 96 h in binary enzyme–aminoglycoside complexes. The binary enzyme-aminoglycoside complex shows the highest level of protection among all complexes, and addition of nucleotide to this complex renders several additional residues exchangeable, some of which are surprisingly located in the center of the largest β-sheet of the enzyme that also interacts with the nucleotide. In general, β-sheet residues are highly protected due to backbone hydrogen-bond interactions.

To understand the differential effects of antibiotic and nucleotide binding on the dynamics of the enzyme, we compare correlated motions of APH in the presence and in the absence of ligands. Since no X-ray structure of the binary APH–antibiotic complex is available, the effect of nucleotide binding to the binary complex cannot be investigated directly. Instead, we can stepwise remove substrates from a ternary complex structure. This approach has the advantage of neglecting the minor structural differences between the ligand-bound form and the apo form of the enzyme and therefore allows for an easier extraction of relevant information on the role of substrate binding because differences in the correlations can arise solely from the presence or the absence of substrate nodes. Nevertheless, the structural differences between the apo form and the ligand-bound form may play a role in enzyme dynamics and function, as will be discussed later on. Superpositioning the backbone of APH in several complexes shows that the structural differences between different complexes are rather small, with

RMSD values ranging from 0.7 Å for the comparison between X[APH-Nuc-Kan] and X[APH-Nuc-Neo] to 1.7 Å for the comparison between X[APH] and X[APH-Nuc]. Therefore, we first analyze the correlation differences for X[APH-Nuc-Kan] with different substrates bound, and then discuss if the results are conferrable to the different X-ray structures.

In Fig. 2, the correlation plot for X[APH-Nuc-Kan] is compared to the correlation plots of the same structure but with only one of the substrates bound or with no substrate bound at all. To clearly illustrate the differences between the structures with different substrate contents, we subtract these plots from the plot of the ternary complex. Three areas are highlighted in the correlation plot of the ternary complex, as they notably change their correlations upon substrate removal. The largest area (residues 2–91) coincides with domain I. When the nucleotide is bound to APH, the correlations of the residues of domain I to the rest of the protein are less negative, which can be seen as positive correlation differences in plot b. In contrast, the correlations within domain I are less positive than in the absence of nucleotide, resulting in negative correlation differences. The reduced correlation agrees with H/D exchange data showing that addition of the nucleotide to the APH–antibiotic complex rendered several amides exchangeable.[6] ITC data showed that binding of the aminoglycoside to the APH–nucleotide complex was accompanied with a significantly less reduction in enthalpy (15–25 kcal mol^{-1}) compared to the binding of aminoglycosides to apo-APH.[7] The much smaller reduction in enthalpy was attributed to the enthalpic penalty of breakage or weakening of hydrogen bonds within the large β-sheet upon ATP binding, which rendered several amides exchangeable. Our correlation data also support this conclusion since, without nucleotide, domain I (including the β-sheet) can move quite independently as a rigid body, away from the other domains. This behavior manifests as a higher correlation of the residues of domain I to themselves. Binding of the nucleotide connects domain I to domain II, which may cause buckling of the large β-sheet. Consistent with this interpretation, the largest difference in the correlations with and without nucleotide is found for
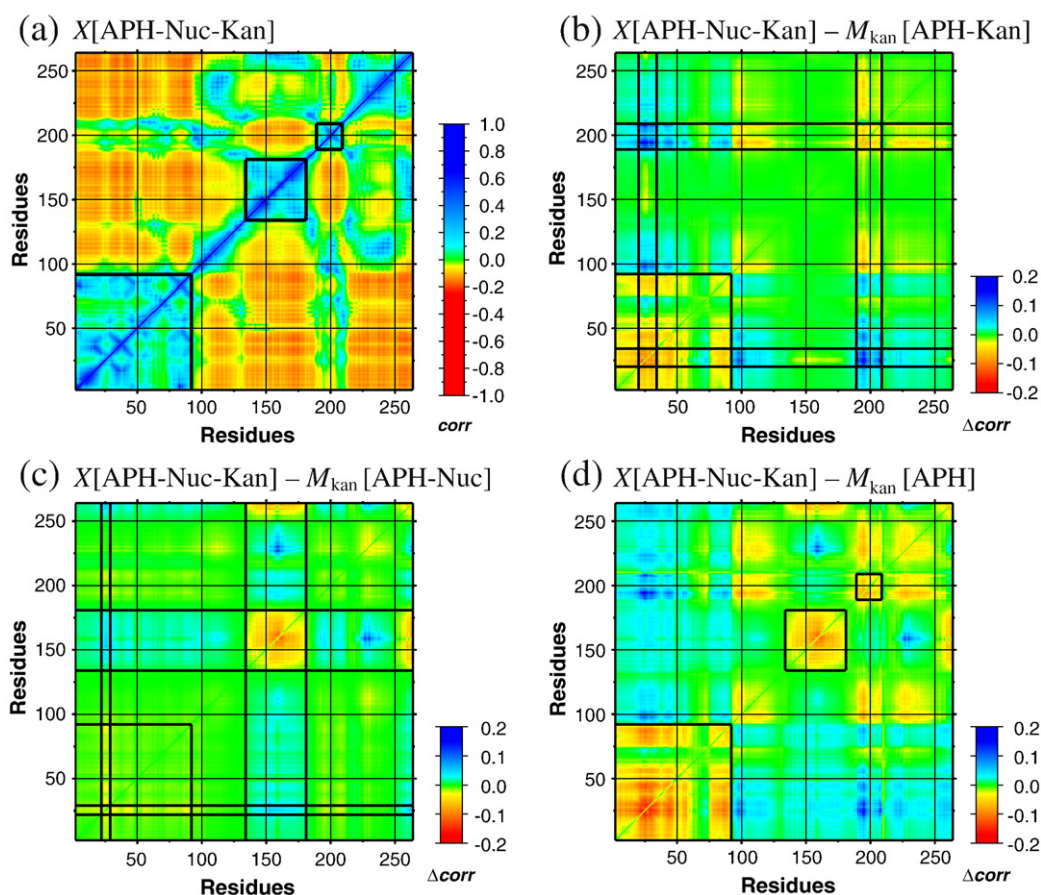
Fig. 2. Correlated motions of APH calculated for the structure of *X*[APH-Nuc-Kan] and comparison to model complexes derived from *X*[APH-Nuc-Kan]. (a) Correlation plot of *X*[APH-Nuc-Kan]. Residues of the three highlighted, positively correlated areas notably change their correlations upon substrate removal. (b) The difference correlation plot of the original *X*[APH-Nuc-Kan] complex minus correlations of the model complex M_{kan}[APH-Kan] shows the effect of nucleotide binding to the binary APH–kanamycin complex. The positive correlations within domain I and those of the nucleotide binding region of domain II are lower in the presence of nucleotide, leading to negative correlation differences. The largest difference in the correlations with and without nucleotide is found for loop residues 23–27, which connect the first two strands of the large β-sheet and interact directly with the nucleotide. (c) The difference correlation plot of the original *X*[APH-Nuc-Kan] complex minus correlations of the model complex M_{kan}[APH-Nuc] shows that the presence of kanamycin leads to a reduced correlation within domain III. The largest difference occurs for the long loop (residues 152–166) that closes over kanamycin. (d) The difference correlation plot of the original *X*[APH-Nuc-Kan] complex minus correlations of the model complex M_{kan}[APH] shows that the effects of the binding of both substrates are nearly additive.

loop residues 23–27, which connect the first two strands of the large β-sheet and interact directly with the nucleotide. In the absence of the nucleotide, not only the correlations of the strands of the large β-sheet but also the correlations of the nucleotide binding region of domain II (which is the second area of large correlation differences) are higher.

In contrast to the removal of the nucleotide, the removal of kanamycin hardly affects the residues of the large β-sheet, except residues 25–27 on the loop, which are the only residues of domain I that interact directly with both substrates. In the absence of antibiotic, they are more correlated to the rest of the β-sheet, which is seen as a narrow red band in Fig. 2c. But removal of kanamycin leads to a large change in

the correlations of domain III residues. Correlation of the helix–loop–helix to itself and the C-terminal helix becomes stronger, while it becomes more anti-correlated to the rest of the protein. The largest change in correlations occurs for the long loop of domain III (residues 152–166), which forms a lid that closes over kanamycin. Residues 155–162 and residues 225–232 of domain II bind kanamycin directly. These residues are positively correlated to each other in the presence of kanamycin, while they are slightly anticorrelated in its absence. This behavior leads to high positive correlation differences between the kanamycin-bound state and the free state (observable as blue spots in Fig. 2c). Obviously, substrate binding always leads to less

correlation within the binding region. So why do nucleotide binding and antibiotic binding act so differently on APH mobility? The decisive difference is the high level of flexibility of the antibiotic binding loop, which will be discussed in Comparison of Experimental and Theoretical *B*-Factors, while the β-sheet is a stable protein region by itself. Nucleotide binding to the apoenzyme can still increase the stability of some residues, which are protected against H/D exchange in the binary complex. When nucleotide and antibiotic are bound, β-sheet residues 25–27 are connected to both other domains. The β-sheet undergoes frustration,[38] resulting in a weakening of hydrogen bonds.

The removal of both substrates from X[APH-Nuc-Kan] results in correlation differences that are a combination of the effects of the removal of the single substrates. As can be discerned from Fig. 2b, c, and d, the effects of substrate binding are nearly additive, with a deviation between −0.05 and +0.05 (Fig. S1, Supplemental Data). This additivity implies that nucleotide removal from M_{kan}[APH-Nuc] would yield results analogous to those of nucleotide removal from X[APH-Nuc-Kan], just as the effects of kanamycin removal from M_{kan}[APH-Kan] are nearly equal to those of kanamycin removal from X[APH-Nuc-Kan].

For comparison between different ENMs, we calculated the substrate-dependent correlations for X[APH-Nuc-Kan] using the anisotropic network model (ANM), a directional version of the GNM[39] (Fig. S2, Supplemental Data). The classification in dynamic protein domains is less obvious from the correlations calculated with the ANM because no large areas of only positive correlation or only negative correlation exist. But the substrate-dependent correlation differences are very similar to the differences obtained with the GNM, giving us greater confidence on the method. The observation that correlations calculated with the GNM are more suitable for the identification of dynamic domains is in line with previous studies showing that the GNM reproduces experimental *B*-factors better than the ANM.[10]

We showed that the addition of a small number of ligand nodes to the elastic network can lead to important changes in the dynamics. The ligand binding sites seem to be prominent positions in the protein that allow for substrate-adjustable flexibility. To see if their location at the interface of dynamic domains is the decisive factor for this ability, we calculate the influence of adding pseudo-substrate nodes on the surface of X[APH]. Previous perturbation studies investigated the effect of additional nodes or arbitrarily changed force constants on the normal modes that describe a functional conformational transition.[40–42] As APH does not undergo large-scale conformational changes, we analyze the perturbing effect of additional nodes on the correlations of fluctuations.
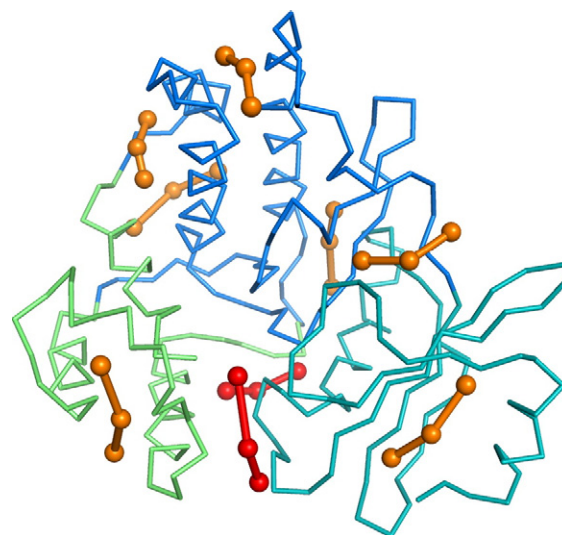

Fig. 3. Nine positions of the pseudo-substrate on the surface of X[APH]. The pseudo-substrate consists of three nodes and is located once on the surface of each domain (domain I: cyan; domain II: blue; domain III: green) and twice at the interface between each domain pair. Pseudo-substrates shown in orange have only minor and localized effects on the correlations. Pseudo-substrates shown in red lie between domain I and domain III and have a large effect on the correlations of motions.

A pseudo-substrate consisting of three nodes is placed in nine different clefts on the protein surface. It is located once on the surface of each domain and twice at the interface between each domain pair. The positions of all pseudo-substrates are shown in Fig. 3. The difference correlation plots that show the correlation change upon pseudo-substrate binding are shown in Fig. S3 of Supplemental Data. Oftentimes, only nodes that gain a direct connection due to pseudo-substrate binding show an increase in their correlation. The size of this increase (0.1–0.2) is comparable to the correlation differences due to real ligand binding, but the changes are much more localized. Only for both positions at the interface between domain I and domain III do the additional nodes have a remarkable effect on the correlations, resembling the effect of ligand binding. One characteristic of real ligand binding is that the correlation between some nodes decreases, which is a less localized effect than the increase in correlation between nodes with added connections. The pseudo-substrates at the interface between domain I and domain III lead to a decrease in the correlations within each domain and to an increase in the correlation between the two domains. The fact that pseudo-substrate binding between domain II and the other domains has no large effect on the correlations demonstrates that not all sites that lie between two dynamic domains are sensitive to perturbations. Perturbation-sensitive sites must lie

between nodes that are clearly anticorrelated when no substrate is bound. Virtually all pairs of nodes of domains I and III fulfill this condition (Fig. 1c), while several nodes of domain II are uncorrelated to parts of the other domains.

To analyze the effect of structural changes on correlations, we calculated the correlations of the various APH complexes directly based on available crystal structures. Figure 4a shows the correlation differences between X[APH-Nuc] and X[APH]. According to our previous results, correlation within the large β-sheet region is clearly higher in X[APH] than in X[APH-Nuc]. But the correlations of loop residues 23–27 to the rest of the β-sheet are more positive in the nucleotide-bound form, in contrast to their much higher correlation to themselves in X[APH-Nuc-Kan] upon removal of

nucleotide. This observation can be explained by structural differences between X[APH-Nuc] and X[APH]. Their effect is derived by subtracting the correlations of the apoenzyme from the correlations of X[APH-Nuc] in the absence of substrate, which is denoted as the model complex M_{nuc}[APH] (Table 1). Figure 4b shows that structural changes in loop residues lead to correlation differences whose size is comparable to the correlation differences due to the removal of ligand nodes. But these differences are very localized, in contrast to correlation differences due to changes in the node number, which affect larger regions. The latter are given by the difference correlation plot X[APH-Nuc] − M_{nuc}[APH], which would be obtained by subtracting plot b of Fig. 4 from plot a and looks very similar to plot b of Fig. 2.
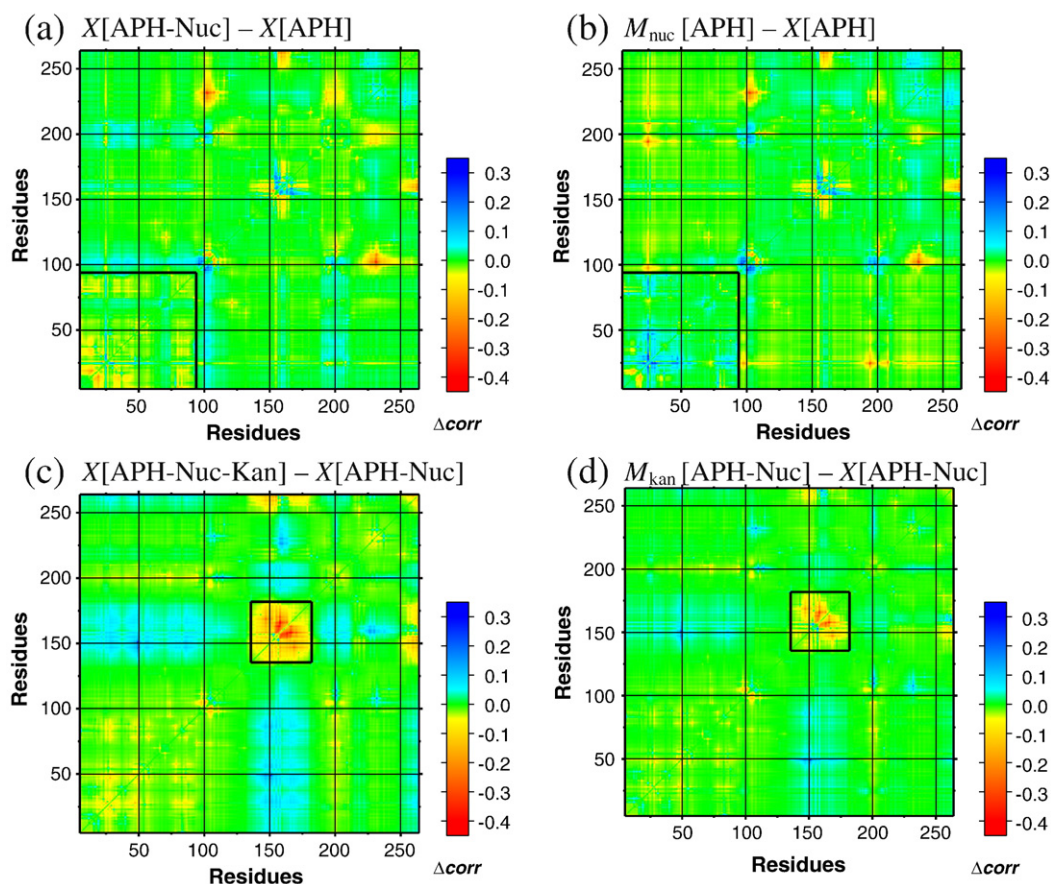

Fig. 4. Comparison of the correlated motions of APH calculated for different crystal structures. (a) The difference correlation plot of X[APH-Nuc] minus X[APH] shows that nucleotide binding leads to a reduced correlation within the large β-sheet (highlighted area). But in contrast to Fig. 2b, the correlations of loop residues 23–27 to the rest of the β-sheet are more positive in the nucleotide-bound form. (b) Correlation differences due to structural changes between X[APH-Nuc] and X[APH] are shown as the difference correlation plot of M_{nuc}[APH] minus X[APH]. The correlations of loop residues 23–27 to domain I are more positive in M_{nuc}[APH] than in X[APH]; thus, the structural differences partially compensate for the effect of nucleotide removal. (c) The difference correlation plot of X[APH-Nuc-Kan] minus X[APH-Nuc] shows an enhanced correlation of domain III (highlighted area) upon kanamycin removal. (d) Correlation differences due to structural changes between X[APH-Nuc-Kan] and X[APH-Nuc]—shown as the difference correlation plot of M_{kan}[APH-Nuc] minus X[APH-Nuc]—further enhance the effect of antibiotic binding.

Loop residues 23–27 occupy alternative positions in the different structures. In X[APH], they block the nucleotide binding site and thereby make connections to the other side of the nucleotide binding site. Therefore, the lack of nucleotide in X[APH] leads to less correlation within domain I than in M_{nuc}[APH], which can be observed as a blue cross in Fig. 4b. The difference correlation plot X[APH-Nuc-Kan] – X[APH-Nuc] (Fig. 4c) shows the pattern characteristic of the removal of antibiotic also seen in Fig. 2c. In this case, the structural changes even enhance the increased correlation within the antibiotic binding loop upon kanamycin removal, as indicated by the difference correlation plot between M_{kan}[APH-Nuc] and X[APH-Nuc] in Fig. 4d. This increased correlation is due to the opening of the long loop in domain III in the structure without bound antibiotic (X[APH-Nuc]). Thus, the connectivity of this loop to domains I and II of the protein is reduced, in addition to the effect of the substrate release. While the enzyme amplifies the effect of missing antibiotic (indicated by an equal sign in the highlighted area of plots c and d of Fig. 4), it partially compensates for the absence of nucleotide, which is seen as opposite signs in the β-sheet region of plots a and b. The alternative positions of loop residues 24–26 and 152–166 show that connections must be broken to accommodate nucleotide as ligand, while additional connections arise from antibiotic binding by closing of the loop consisting of residues 152–166. These theoretical findings explain the differential effects of antibiotic binding and nucleotide binding on structural stability. The reduction of favorable enthalpy upon aminoglycoside binding to the enzyme in the presence of nucleotide, in comparison to aminoglycoside binding to apo-APH, has two contributions. As explained above, nucleotide binding hinders the rigid-body movement of domain I. Additionally, the structural differences suggest that favorable interactions between domain I and domain II must be broken to accommodate a nucleotide as ligand. In contrast, the strengthened interactions between the antibiotic binding loop and the rest of the protein upon antibiotic binding decrease the enthalpy.

Comparison of experimental and theoretical *B*-factors

The stabilizing effect of aminoglycoside binding to APH observed from H/D exchange experiments[6] can be explained by theoretical *B*-factors calculated by the GNM. Regions with high *B*-factors are flexible and usually display reduced protection against H/D exchange. Figure 5a, which compares experimental *B*-factors from crystallographic studies of different complexes and calculated *B*-factors of the same structures, shows a general agreement
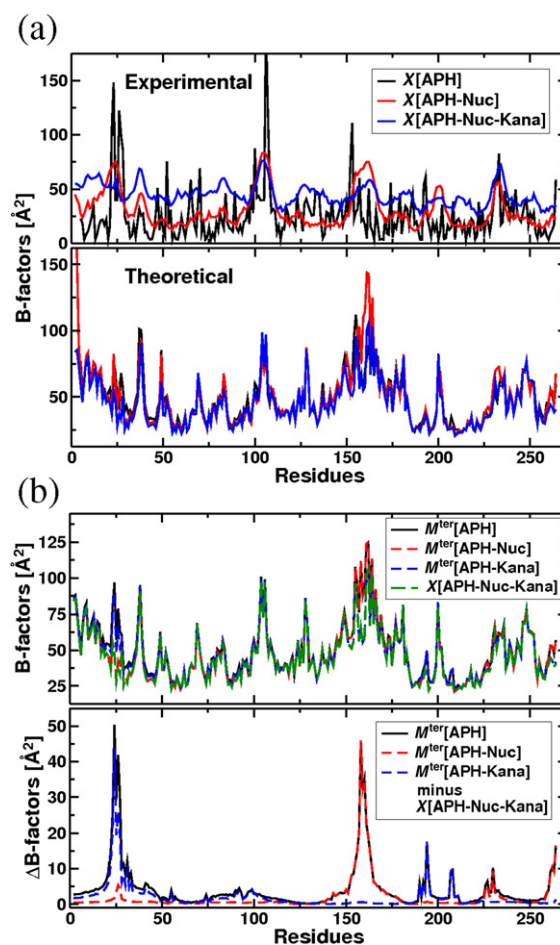

Fig. 5. Comparison of the experimental and theoretical *B*-factors of APH. (a) Experimental and theoretical *B*-factors of the structures X[APH], X[APH-Nuc], and X[APH-Nuc-Kan] show good agreement. Residues Asp153-Asp162 have the highest *B*-factors in X[APH-Nuc], both in theory and in experiment. By increasing the flexibilities of these residues near the antibiotic binding region, nucleotide binding could facilitate the binding of a variety of aminoglycosides. (b) Theoretical *B*-factors of the model complexes M_{kan}[APH], M_{kan}[APH-Nuc], and M_{kan}[APH-Kan] in comparison to the theoretical *B*-factors of the structure X[APH-Nuc-Kan]. The theoretical *B*-factors differ in protein regions that are involved in substrate binding: the nucleotide binding region (residues 20–32) and the antibiotic binding region (residues 153–162). Bottom: To show the differences more clearly, we subtracted the theoretical *B*-factors of X[APH-Nuc-Kan] from the theoretical *B*-factors of M_{kan}[APH], M_{kan}[APH-Nuc], and M_{kan}[APH-Kan]. The removal of ligand nodes always results in higher flexibilities.

between theory and experiment. In X[APH], we observe a larger deviation between experimental *B*-factors and theoretical *B*-factors, which could be caused by the low resolution of X[APH] (3.2 Å) compared to those of X[APH-Nuc] (2.4 Å), X[APH-Nuc-Kan] (2.4 Å), and X[APH-Nuc-Neo] (2.7 Å). In

the experiment, several residues of X[APH] are highly flexible, which could result from partial unfolding flexible. In comparison, the experimental B-factors of the binary and ternary complexes have broader peaks, consistent with collective motions. The calculated B-factors for all three structures, however, are very similar and only differ in certain protein regions that participate in ligand binding. Generally, the elastic network becomes more rigid when more connections are present due to ligand binding, but structural differences between the different complexes cause some residues to become more flexible in the antibiotic-bound form, and even more flexible in the nucleotide-bound form. There is a maximum of both experimental and theoretical B-factors in X[APH-Nuc] around residue Glu24, that is not observed when both substrates are bound. This loop of domain I interacts directly with the nucleotide. Residues Thr23-Met26 show different backbone conformations in all three structures. In X[APH-Nuc], the distance between the loop and the nucleotide is larger than that in X[APH-Nuc-Kan], leading to a higher level of flexibility in this region. In contrast, in X[APH], loop residues Glu24-Met26 are shifted towards the nucleotide binding site, which may render them less flexible than expected due to missing connections via the nucleotide substrate. One of the largest regions with high B-factors is that between Asp153 and Asp162, which is clearly visible in both experimental and computational data. Specifically, X[APH-Nuc] shows generally high B-factors for all of the residues in this region (Fig. 5). In fact, the largest theoretical B-factors of X[APH-Nuc] (neglecting the N-terminus) belong to this antibiotic binding loop of domain III. By increasing the flexibility of the residues near the antibiotic binding site, nucleotide binding could facilitate the binding of a variety of aminoglycosides. These results are also consistent with kinetic data suggesting that the binding of the nucleotide is followed by the binding of the aminoglycoside for catalysis to occur.[43]

Effect of node connectivity on flexibility

The decreased flexibility of residues due to substrate binding to APH results from the increased number of connections. The atomic mean square displacements were shown to be essentially determined by variations in local packing density,[44] which is defined as the number of nonhydrogen atoms within a spherical region. Also for ENMs, it was suggested that residue fluctuations are majorly influenced by connectivities.[45] In order to examine the influence of bound substrates on elastic network properties, we use X[APH-Nuc-Kan] for calculating B-factors, but with different substrates bound to it (Fig. 5b). The original complex with both substrates bound always yields the lowest B-factors. Removal

of the substrates from the ternary complex leads to less connections and higher flexibilities. As expected, the removal of the nucleotide enhances the flexibility of residues 17–32, which are near the nucleotide binding site and also show a peak of experimental B-factors in X[APH]. The removal of kanamycin enhances the flexibility of residues 151–168, which are on the loop of domain III and are involved in antibiotic binding. From these results, it seems that the connectivities are the determinant for molecular flexibility.

In the following, we want to examine mathematically if the flexibilities can be soundly approximated by node connectivities, or if the overall protein architecture plays an important role, too. In the GNM, the diagonal elements of the Kirchhoff matrix Γ_{ii} give the number of connections of node i and are a direct measure of local packing density or connectivity. Let us assume that the motions of a node would be fully determined by its connectivity. Then the off-diagonal elements of the Kirchhoff matrix can be set to 0, and the inverse Kirchhoff matrix exists and is diagonal, with elements $\hat{\Gamma}_{ii}^{-1} = \frac{1}{\Gamma_{ii}}$ (i.e., the B-factors are inversely proportional to residue connectivities) (Eq. (11)). In reality, the diagonal elements of the pseudo-inverse of the full Kirchhoff matrix are:

$$\tilde{\Gamma}_{ii}^{-1} \geq \frac{1}{\Gamma_{ii}} \left(\frac{N-1}{N} \right)^2 \qquad (1)$$

(for derivation, see Supplemental Data). It follows that, for a large number of nodes N, the off-diagonal elements of the full Kirchhoff matrix make a positive contribution to the B-factors. For simplicity, the theoretical B-factors according to the original description will be referred to as total B-factors, and the B-factors calculated from the diagonal Kirchhoff matrix will be referred to as connectivity B-factors. We compare the total B-factors to the connectivity B-factors, whereby we obtain the contribution of the off-diagonal elements of the Kirchhoff matrix to the total B-factors as discrepancy between the two approaches. Figure 6a compares the total B-factors and the connectivity B-factors calculated for X[APH-Nuc-Kan]. The overall distribution of maxima is reproduced by the connectivity B-factors. The linear correlation coefficient (Eq. (12)) between the total B-factors and the connectivity B-factors is about 0.9 for all structures. But the magnitudes of the B-factors differ. Figure 6b illustrates that only residues that are within the cutoff radius of ligand nodes can change their connectivity B-factors, depending on ligand content. The substrate-dependent changes in the total B-factors of all other residues are solely caused by off-diagonal elements of the Kirchhoff matrix. As the B-factors of apo-APH are always higher than the B-factors of binary and ternary
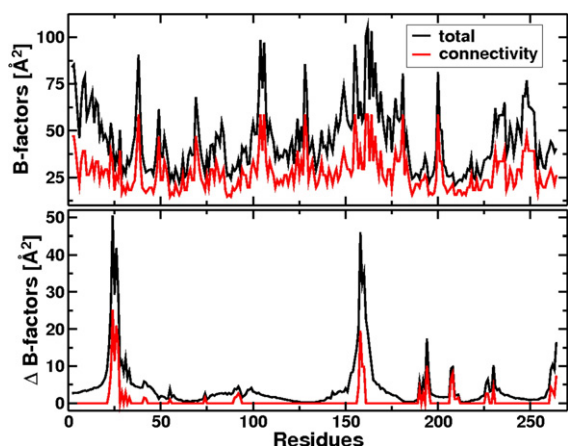
Fig. 6. Comparison between total *B*-factors and connectivity *B*-factors. Total *B*-factors are calculated using the full Kirchhoff matrix, and connectivity *B*-factors are calculated from the diagonal Kirchhoff matrix only. Top: Total and connectivity *B*-factors, calculated for X[APH-Nuc-Kan], show that maxima are reproduced by the connectivity *B*-factors but the magnitudes of the *B*-factors differ. The correlation between total *B*-factors and connectivity *B*-factors is 0.93. Bottom: The *B*-factors of X [APH-Nuc-Kan] are subtracted from the *B*-factors of the same structure, but without substrate (M_{kan}[APH]). This calculation is performed once for the total *B*-factors and once for the connectivity *B*-factors. Substrate-dependent changes in connectivity *B*-factors can only occur for nodes within the cutoff radius of the substrate nodes (red curve), while many residues change their total *B*-factors upon substrate binding (black curve).

complexes, these off-diagonal contributions are also higher if less edges exist in the network. Obviously, not only residues bound directly to the substrates are influenced by substrate binding. Such nonlocal effects would be overlooked if off-diagonal elements of the Kirchhoff matrix are neglected.

Comparison between kanamycin-bound APH and neomycin-bound APH

The crystal structures of two APH–nucleotide–antibiotic complexes are solved.[29] They contain kanamycin A or neomycin B as ligand, which differ in size and spatial orientation. Neomycin consists of four rings, and its 2-deoxystreptamine ring, which is common to all aminoglycosides, is 4,5-disubstituted, whereas kanamycin is 4,6-disubstituted and has only three rings. In H/D exchange experiments, differences between the two complexes can be seen, too. Neomycin-bound APH shows much greater solvent protection compared to kanamycin-bound APH.[6] In the following, we analyze the relationship between theoretical *B*-factors and H/D exchange times. A good correlation between residue flexibilities calculated by the GNM and H/D exchange rates was shown for several proteins.[46] When the *B*-

factors are high, the H/D exchange should be fast. Accordingly, in the neomycin-bound structure, residues have lower theoretical *B*-factors than in the kanamycin-bound structure. For calculations, we use the model complexes M_{kan}[APH-Kan] and M_{neo}[APH-Neo]. A comparison of the calculated *B*-factors for both binary complexes is shown in Fig. 7, where the *B*-factors of the APH–neomycin complex were subtracted from those of the corresponding kanamycin complex. There are much more positive values than negative values for this difference, indicating that the residues in the kanamycin-bound complex are more flexible, in agreement with the H/D exchange data. *B*-factor differences sorted by their size are also shown in Fig. 7 to demonstrate this behavior more clearly. Seventy-five residues have lower *B*-factors in the kanamycin-bound complex than in the neomycin complex, while the *B*-factors of the remaining residues are equal to or larger in the kanamycin-bound complex. The H/D exchange times and theoretical *B*-factors of kanamycin-bound and neomycin-bound APH are shown in Fig. S4 of Supplemental Data. Note that H/D exchange times could only be measured for about half of the residues. Also, for 74 residues in the kanamycin complex and for 65 residues in the neomycin complex, hydrogen exchange occurred faster than the acquisition of the first spectrum (3–4 min of delay in starting data acquisition). Both complexes show that there is a general agreement between the H/D exchange data and the
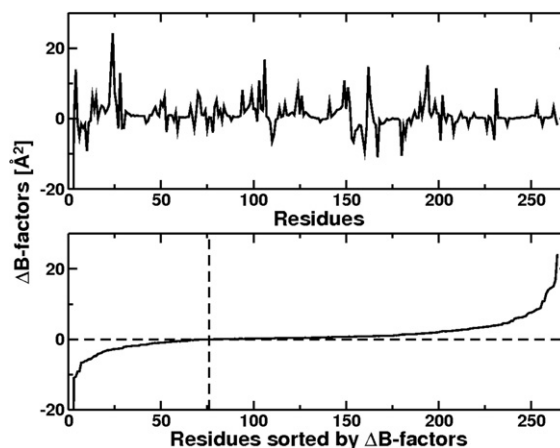

Fig. 7. Comparison between the theoretical *B*-factors of the two different binary APH–antibiotic complexes. Top: *B*-factors of M_{kan}[APH-Kan] minus *B*-factors of M_{neo} [APH-Neo]. Bottom: The *B*-factor differences are sorted by size. The dotted line indicates a *B*-factor difference of 0, showing that only 75 of 263 residues have higher *B*-factors in the kanamycin complex than in the neomycin complex. All negative *B*-factor differences add up to −266.6 Å², whereas all positive *B*-factor differences add up to 461.8 Å².

calculated *B*-factors where protected amides display lower *B*-factors and vice versa. However, there are also aminoglycoside-dependent and aminoglycoside-independent exceptions to this. Such discrepancies can arise from limitations of the harmonic model, which does not imply large-amplitude motions and local unfolding. These processes could be responsible for the large number of APH residues with fast H/D exchange. However, correlated motions can also be a reason for H/D exchange times shorter than expected from residue flexibilities. For example, Val57 has a very low theoretical *B*-factor, but it exchanges its proton before the acquisition of the first spectrum. It is located at the beginning of the connecting helix between domain I

and domain II, and its motion is not correlated to one of the three domains. As already shown for the β-sheet residues, localization between different correlated domains can lead to distortion and reduced stability in H/D exchange. The discrepancy between *B*-factors and H/D exchange for several residues suggests that a combination of residue flexibilities and correlation of motions is important for understanding protection in H/D exchange experiments.

Correlated motions also differ in the kanamycin-bound and neomycin-bound forms of APH, although the overall correlation patterns are similar. The most obvious differences in the correlations between M_{kan}[APH-Kan] and M_{neo}[APH-Neo] can be seen for residues 157–162 of domain III (Fig. 8d).
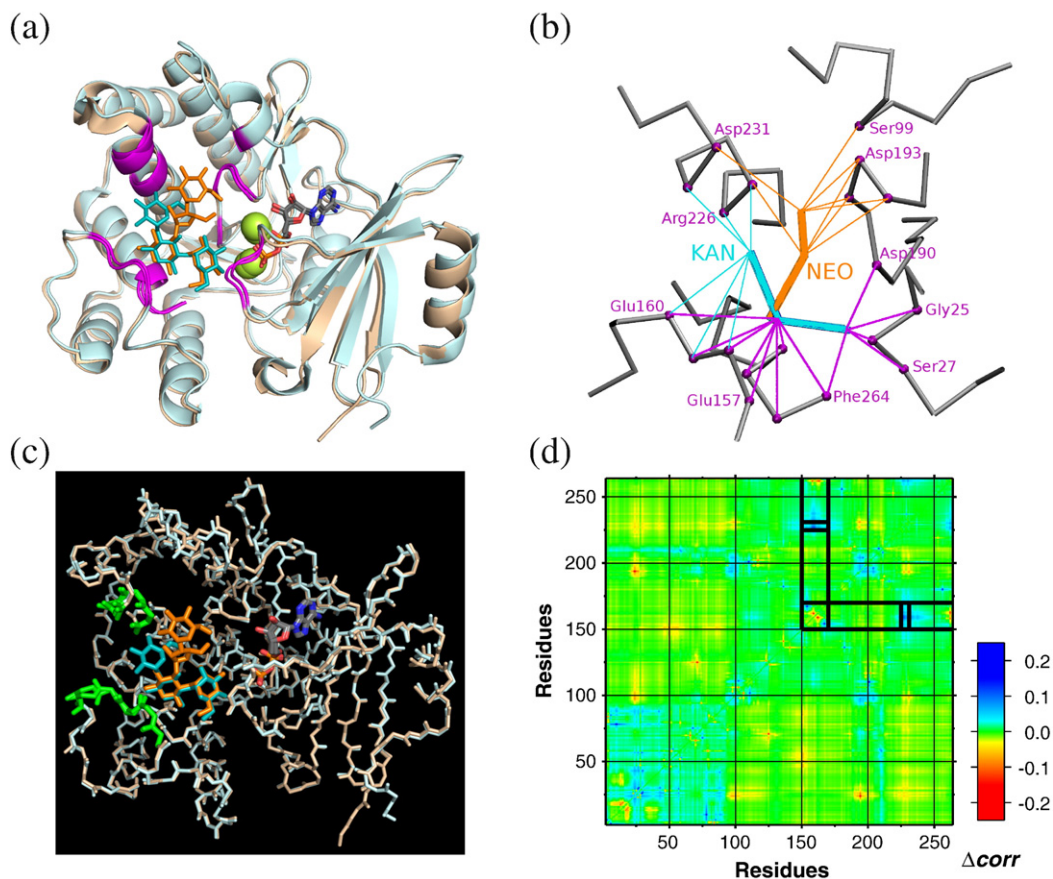

Fig. 8. Comparison between kanamycin binding and neomycin binding to APH. (a) Superimposed structures of the kanamycin-bound form (pale blue) and the neomycin-bound form (pale orange) of APH. Kanamycin is shown in cyan, and neomycin is shown in orange. Enzyme regions that are connected to any antibiotic are depicted in magenta. The image was produced using PyMOL.[32] (b) Connections of C$^\alpha$ nodes (magenta) to the antibiotic nodes of kanamycin (cyan) and neomycin (orange). Edges common to the elastic network of both antibiotics are shown as thicker magenta lines. The image was produced using VMD.[47] (c) Superimposed structures of kanamycin-bound APH (pale blue) and neomycin-bound APH (pale orange). The residues that are anticorrelated in the APH–neomycin complex, although uncorrelated in the APH–kanamycin complex, are highlighted in green. (d) Difference correlation plot of M_{kan}[APH-Kan] minus M_{neo} [APH-Neo]. The positive correlations of residues 157–162 of domain III to each other and to the C-terminal residues are higher in the neomycin-bound form. The correlations between residues 157–162 of domain III and residues 226–230 of domain II are approximately 0 in M_{kan}[APH-Kan] because they are strongly connected over kanamycin. With neomycin, the two stretches are anticorrelated.

The positive correlation of these residues to each other and to the C-terminal residues is higher in the neomycin-bound form. The correlations between residues 157 to 162 of domain III and residues 226 to 230 of domain II are approximately 0 in M_{kan}[APH-Kan] because these residues are strongly connected over kanamycin (Fig. 8c). Without antibiotic or with neomycin instead, the two stretches are anticorrelated. The correlation plots of M_{kan}[APH-Kan] and M_{neo}[APH-Neo] are given in Fig. S5 of Supplemental Data. The differences in residue flexibilities and correlations between kanamycin-bound APH and neomycin-bound APH arise from the individual connections of the two antibiotics to the enzyme. Based on a cutoff radius of 8 Å, neomycin and kanamycin are both connected to residues Gly25-Ser27, Glu157-Glu160, Asp190, and Asp261-Phe264, as well as to two to three residues of the Arg226-Asp231 stretch. However, residues Gly192-Ser194 of domain II are only connected to neomycin but not to kanamycin (Fig. 8b).

The observation that the kanamycin–APH and neomycin–APH complexes show different dynamics correlates with previous thermodynamic studies, which showed that the change in heat capacity (ΔC_p) is significantly different between the kanamycin–APH complex and the neomycin–APH complex.[48] Enthalpy change for the binding of kanamycin to APH showed a nonlinear dependence on temperature between 21 °C and 37 °C and yielded an unusually high ΔC_p value above 30° (−3.5 kcal mol^{-1} deg^{-1}), while the enthalpy change dependence for the binding of neomycin to APH remained linear at the same temperature range, yielding a significantly lower ΔC_p (−1.8 kcal mol^{-1} deg^{-1}). We note that both of these values are too high to be simply explained by the burial of hydrophobic surfaces. A large fraction of these unusually large ΔC_p values were attributed to conformational changes on the enzyme, which are different for the kanamycin-bound and neomycin-bound forms. Furthermore, even though water was not included in these calculations, the dynamic properties of different complexes may shed some light on the very unusual effects of solvent on enzyme–antibiotic complexes, where the binding enthalpy of kanamycins to APH is more favorable in H_2O than in D_2O, while the binding of neomycins is enthalpically more favored in D_2O.[48] Thus, it is highly likely that the antibiotic-dependent differences in the correlated motions of residues 157–162 of domain III and residues 226–230 of domain II coinciding with the region of the aminoglycoside–enzyme interface (Fig. 8c)—where the structural differences between kanamycins and neomycins are manifested—may cause differential interactions of protein side chains and/or aminoglycosides with the solvent yielding such a strong contrast in solvent effects on the binding enthalpy of kanamycins *versus* neomycins.

Conclusions

Our theoretical analysis of APH dynamics helps in understanding several findings from previous ITC and NMR studies in solution that could not be explained by static X-ray structures of the apo form and the ligand-bound form of the enzyme. In our experiment, addition of nucleotide to the APH–antibiotic complex rendered several amides exchangeable, which led to the conclusion that the association of the side chains of Lys44 and Tyr42 with nucleotide causes the sheet to stretch or buckle.[6] Although the coarse-grained model does not consider side-chain atoms at all, the results from the correlation analysis exactly match this conclusion. Despite its simplicity, the GNM also reproduces single-residue *B*-factors well. Substrate binding leads to lower flexibilities of residues in and around the binding sites, which are reflected by higher levels of protection in H/D experiments. Theoretical data also supplement experimental data where resonances of the apo form and the nucleotide-bound form of the enzyme could not be assigned. As nucleotide binding leads to lower theoretical *B*-factors of the N-terminal strands of the β-sheet, one can assume that the few residues, which are protected upon nucleotide binding to apo-APH, are near the ATP binding site. Finally, the differential effects of kanamycin binding and neomycin binding on thermodynamic data are reflected in residue flexibilities and correlated motions.

The GNM also provides more general information on protein architecture. Small changes in the node number through addition of a ligand can have important consequences on dynamics, if they occur at prominent positions, like between anticorrelated domains. We would also like to underline the importance of the off-diagonal elements of the Kirchhoff matrix, which exert additional constraints on the motions of nodes and are essential for both residue flexibilities and correlations of motions. While the overall pattern of residue flexibilities may be approximated by the number of nearest neighbors, important details that make the difference between the ligand-free state and the ligand-bound state can be lost, especially because residues that are not in the direct neighborhood of ligands also change their flexibilities. Also, correlations and anticorrelations of motions are important in explaining experimental results, as shown for residues of the large β-sheet of APH. This β-sheet shows a decreased correlated motion after nucleotide binding and thus a higher probability of hydrogen-bond breaking in the backbone, in contrast to expectations based on node connectivities.

From our analysis, we conclude that substrate-adjustable flexibility is a key factor in explaining the

substrate promiscuity of the enzyme. The arrangement of the protein into three dynamic domains with different rigidities allows for the tuning of dynamics upon ligand binding. The interplay between residue flexibilities and correlated motions helps in understanding the unequal characters of enthalpic ligand binding and entropic ligand binding. A comparison of theoretical *B*-factors determined for X[APH-Nuc] and X[APH] indicates that nucleotide binding allows for the binding of diverse aminoglycoside substrates by enhancing the flexibility of residues around the antibiotic binding site, which also shows that the rather small structural changes upon nucleotide binding are important for enzyme function. Our conclusion does not contradict the proposal of two alternative subsites in the antibiotic binding pocket that were suggested as a means for substrate promiscuity.[29] Instead, both structural properties and dynamic properties must be considered to understand the ability of APH to adapt to varying ligands.

Methods

Gaussian network model

Proposed by Bahar *et al.*,[16] the GNM was inspired by the elasticity theory of random polymer networks by Flory[49] and Kloczkowski *et al.*[50] and by the single-parameter Hookean potential used in normal-mode analysis by Tirion.[51] The protein is modeled as an elastic network, often with one node per residue, whose position is defined by C$^\alpha$ coordinates. Nodes *i* and *j* are connected if their equilibrium distance $|\mathbf{R}_{ij}°|$ (i.e., the distance in the crystal structure) is smaller than a cutoff distance d_{cut}. Here, $\mathbf{R}_{ij}°$ is the vector that connects the equilibrium positions of nodes *i* and *j* ($\mathbf{R}_{ij}° = \mathbf{R}_i° - \mathbf{R}_j°$). The fluctuations $\Delta\mathbf{R}_{ij}$ are defined as a difference between the instantaneous separation vector $\mathbf{R}_{ij} = (x_j - x_i, y_j - y_i, z_j - z_i)$ and the equilibrium distance vector:

$$\Delta\mathbf{R}_{ij} = \mathbf{R}_{ij} - \mathbf{R}_{ij}° = \Delta\mathbf{R}_j - \Delta\mathbf{R}_i \quad (2)$$

The GNM makes two assumptions on the distribution of fluctuations about the equilibrium distances. First, the probability distribution of all fluctuations is isotropic:

$$P(\Delta\mathbf{R}_{ij}) = P(\Delta x_{ij}, \Delta y_{ij}, \Delta z_{ij}) = p(\Delta x_{ij})p(\Delta y_{ij})p(\Delta z_{ij}) \quad (3)$$

Second, the fluctuations $\Delta\mathbf{R}_{ij}$ are assumed to obey a Gaussian distribution:

$$P(\Delta\mathbf{R}_{ij}) = \left(\frac{\gamma^*}{\pi}\right)^{\frac{3}{2}} \exp(-\gamma^* \Delta\mathbf{R}_{ij} \cdot \Delta\mathbf{R}_{ij}) \quad (4)$$

with normalization constant γ^* and zero mean. Substitution of the distribution function $P(\Delta\mathbf{R}_{ij})$ into the elastic free-energy change ΔA associated with the fluctuation $\Delta\mathbf{R}_{ij}$ [$\Delta A = -k_B T \ln P(\Delta\mathbf{R}_{ij})$] yields the harmonic potential

$k_B T \gamma^* \Delta\mathbf{R}_{ij} \cdot \Delta\mathbf{R}_{ij}$ and a force constant $\gamma = 2k_B T \gamma^*$. The potential energy of the whole network is given by:

$$V = \frac{\gamma}{2} \sum_{i,j}^{N} \Gamma_{ij} \left((\Delta x_i - \Delta x_j)^2 + (\Delta y_i - \Delta y_j)^2 + (\Delta z_i - \Delta z_j)^2 \right) \quad (5)$$

where Γ_{ij} is the *ij*th element of the Kirchhoff matrix Γ defined by:

$$\Gamma_{ij} = \begin{cases} -1 & \text{if } i \neq j \text{ and } d_{ij} \leq d_{cut} \\ 0 & \text{if } i \neq j \text{ and } d_{ij} > d_{cut} \\ -\sum_{k,k \neq i} \Gamma_{ik} & \text{if } i = j \end{cases} \quad (6)$$

Since the potential of Eq. (5) is not a harmonic function of the distances between atoms, in the GNM, not only changes in interresidue distances but also any change in the direction of the interresidue vector $\mathbf{R}_{ij}°$ is penalized. Within the densely packed environment of proteins, orientational deformations may be as important as distance changes. The expectation values of the residue fluctuations $\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_i \rangle$ and their covariances $\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle$ are evaluated from the diagonal and off-diagonal elements of the pseudo-inverse Kirchhoff matrix $\tilde{\Gamma}^{-1}$, respectively:

$$\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_i \rangle = \frac{3k_B T}{\gamma} \left(\tilde{\Gamma}^{-1}\right)_{ii}; \quad \langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle = \frac{3k_B T}{\gamma} \left(\tilde{\Gamma}^{-1}\right)_{ij} \quad (7)$$

The inverse Kirchhoff matrix does not exist because one eigenvalue is equal to 0. Therefore, the pseudo-inverse of the Kirchhoff matrix is determined using nonzero modes.[39] For a network with *N* nodes, *N*−1 nonzero eigenvalues λ_i, and eigenvectors u_i, the pseudo-inverse is given by:

$$\tilde{\Gamma}^{-1} = \sum_{i=2}^{N} \frac{1}{\lambda_i} u_i u_i^T \quad (8)$$

The residue fluctuations are related to crystallographic *B*-factors:

$$B_i = \frac{8\pi^2}{3} \langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_i \rangle = \frac{8\pi^2 k_B T}{\gamma} \left(\tilde{\Gamma}^{-1}\right)_{ii} \quad (9)$$

The correlations corr are normalized covariances with values ranging from −1 to +1:

$$\text{corr}(i, j) = \frac{\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle}{\left(\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_i \rangle \langle \Delta\mathbf{R}_j \cdot \Delta\mathbf{R}_j \rangle\right)^{\frac{1}{2}}} \quad (10)$$

A correlation of 0 indicates independent motion, whereas negative values indicate anticorrelation.

Determination of dynamic domains

The nodes of the elastic network are grouped into dynamic domains based on the correlations of fluctuations. The goal is to combine nodes whose correlation to each other is high. The first step is the identification of core clusters, whose nodes are sequential. The search starts with node 1 and adds nodes 2, 3, and so on, to the first cluster if the correlation between these nodes and node 1 is higher than a cutoff value. In general, a node *i* is added to the cluster if its correlation to at least one of

the cluster nodes is higher than the cutoff value and if the previous node $i-1$ also belongs to the cluster. If node i does not meet these criteria, it initiates the next core cluster. The cutoff value is chosen large enough such that many nodes are not assigned to clusters after the first round. The search for core clusters is repeated with a smaller cutoff value. Nodes that still do not belong to a core cluster after the second round are added to clusters, irrespective of the sequential node order. Therefore, a single node is assigned to the cluster whose mean correlation change upon node addition is most favorable. The mean correlation $corr_k$ of a cluster k is:

$$corr_k = \frac{\sum_{i,j \neq i} corr(i,j)}{n} \tag{11}$$

where the sum runs over all nodes i and j, which belong to cluster k, and n is the number of node pairs. The second step of the classification procedure is the combination of the core clusters to larger clusters. The two core clusters with the highest mean correlation between their nodes are merged into one cluster. This procedure is repeated as long as clusters with positive mean correlation between their nodes exist. An advantage of the method is that the number of domains does not have to be specified before the calculation, but the algorithm stops automatically when the mean correlation between all pairs of domains is negative.

Computational details

APH structures with and without substrates are taken from the Protein Data Bank (PDB) (PDB IDs: 1J7I, 1J7U,[28] 1L8T, and 2B0Q[29]). In the structure of apo-APH, the first three residues are not resolved. For comparisons between apo-APH and other complexes, these first residues are deleted in the longer structures. For the protein, the C^α coordinates are used as nodes of the network. ADP and AMPPNP are modeled as three nodes, replacing adenine, ribose, and the phosphate groups. Antibiotics are modeled as one node per amino sugar ring. Thus, kanamycin A is represented by three nodes, and neomycin B is represented by four nodes. The coordinates of the nodes modeling nucleotide and antibiotic are calculated as the geometrical center of the atoms that are represented by the nodes. Each magnesium ion is represented by one node. Edges were added to the network if the distance between two nodes was less than a certain cutoff distance. To determine the appropriate cutoff distance, we calculated the theoretical B-factors for cutoff distances of 6 Å, 7 Å, 8 Å, and 9 Å at a temperature of 300 K and compared these values to experimental B-factors by means of the linear correlation coefficient ρ[10] (Table S2, Supplemental Data). The linear correlation coefficient between two samples with values x_i and y_i is given by:

$$\rho = \frac{\sum(x_i - x)(y_i - y)}{\sqrt{\sum(x_i - x)^2 \sum(y_i - y)^2}} \tag{12}$$

where x and y denote the mean values.

We used a force constant of 0.2 kcal $(mol\ Å^2)^{-1}$ for all connected pairs of nodes. However, the value of the force constant only influences the absolute scale of the fluctuations but has no effect on the correlation coefficient or on the correlation matrices. A cutoff radius of 8 Å gave a good correlation and is a reasonable compromise between all available APH structures and thus was used in all reported calculations. As experimental B-factors, the B-factors of the C^α atoms are taken for the nodes replacing amino acids, and the average value of all experimental B-factors is used for the substrates. The determination of all eigenvalues and eigenvectors of the Kirchhoff matrix was performed by the LAPACK subroutine DSYEV. LAPACK version 3.2.1 was used.[52]

For the identification of dynamic domains, a cutoff value of 0.42 is used in the first round, and a cutoff value of 0.2 is used in the second round of the core cluster search. These values were adopted for the determination of the dynamic domains of all APH structures and resulted in the assignment of 11–12 core clusters in the first round and 19–22 core clusters after the second round.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2011.03.061

References

1. Howe, R. A., Thompson, P. R. & Spencer, R. C. (1996). The new threats of Gram-positive pathogens: re-emergence of things past. *J. Clin. Pathol.* **49**, 444–449.
2. Vacas, T., Corzana, F., Jiménez-Osés, G., González, C., Gómez, A. M., Bastida, A. *et al.* (2010). Role of aromatic rings in the molecular recognition of aminoglycoside antibiotics: implications for drug design. *J. Am. Chem. Soc.* **132**, 12074–12090.
3. Moazed, D. & Noller, H. F. (1987). Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature*, **327**, 389–394.
4. Boehr, D. D., Thompson, P. R. & Wright, G. D. (2001). Molecular mechanism of aminoglycoside antibiotic kinase APH(3′)-IIIa: roles of conserved active site residues. *J. Biol. Chem.* **276**, 23929–23936.
5. McKay, G. A., Thompson, P. R. & Wright, G. D. (1994). Broad spectrum aminoglycoside phosphotransferase type III from *Enterococcus*: overexpression, purification, and substrate specificity. *Biochemistry*, **33**, 6936–6944.

6. Norris, A. L. & Serpersu, E. H. (2009). NMR detected hydrogen–deuterium exchange reveals differential dynamics of antibiotic- and nucleotide-bound aminoglycoside phosphotransferase 3′-IIIa. *J. Am. Chem. Soc.* **131**, 8587–8594.

7. Özen, C. & Serpersu, E. H. (2004). Thermodynamics of aminoglycoside binding to aminoglycoside-3′-phosphotransferase IIIa studied by isothermal titration calorimetry. *Biochemistry*, **43**, 14667–14675.

8. Bahar, I., Lezon, T. R., Yang, L. W. & Eyal, E. (2010). Global dynamics of proteins: bridging between structure and function. *Annu. Rev. Biophys.* **39**, 23–42.

9. Yang, L. W., Eyal, E., Chennubhotla, C., Jee, J., Gronenborn, A. M. & Bahar, I. (2007). Insights into equilibrium dynamics of proteins from comparison of NMR and X-ray data with computational predictions. *Structure*, **15**, 741–749.

10. Kundu, S., Melton, J. S., Sorensen, D. C. & Phillips, G. N. (2002). Dynamics of proteins in crystals: comparison of experiment with simple models. *Biophys. J.* **83**, 723–732.

11. Haliloglu, T. & Bahar, I. (1999). Structure-based analysis of protein dynamics: comparison of theoretical results for hen lysozyme with X-ray diffraction and NMR relaxation data. *Proteins*, **37**, 654–667.

12. Liao, J. L. & Beratan, D. N. (2004). How does protein architecture facilitate the transduction of ATP chemical-bond energy into mechanical work? The cases of nitrogenase and ATP binding-cassette proteins. *Biophys. J.* **87**, 1369–1377.

13. Suhre, K. & Sanejouand, Y. H. (2004). ElNemo: a normal mode Web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.* **32**, 610–614.

14. Nicolay, S. & Sanejouand, Y. H. (2006). Functional modes of proteins are among the most robust. *Phys. Rev. Lett.* **96**, 078104.

15. Tama, F. & Sanejouand, Y. H. (2001). Conformational change of proteins arising from normal mode calculations. *Protein Eng.* **14**, 1–6.

16. Bahar, I., Atilgan, A. R. & Erman, B. (1997). Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* **2**, 173–181.

17. Bahar, I., Erman, B., Jernigan, R. L., Atilgan, A. R. & Covell, D. G. (1999). Collective motions in HIV-1 reverse transcriptase: examination of flexibility and enzyme function. *J. Mol. Biol.* **285**, 1023–1037.

18. Kurt, N., Scott, W. R., Schiffer, C. A. & Haliloglu, T. (2003). Cooperative fluctuations of unliganded and substrate-bound HIV-1 protease: a structure-based analysis on a variety of conformations from crystallography and molecular dynamics simulations. *Proteins*, **51**, 409–422.

19. Yesylevskyy, S. O., Kharkyanen, V. N. & Demchenko, A. P. (2006). The change of protein intradomain mobility on ligand binding: is it a commonly observed phenomenon? *Biophys. J.* **91**, 3002–3013.

20. Xu, C., Tobi, D. & Bahar, I. (2003). Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin T ↔ R2 transition. *J. Mol. Biol.* **333**, 153–168.

21. Zheng, W. & Doniach, S. (2003). A comparative study of motor-protein motions by using a simple elastic-network model. *Proc. Natl Acad. Sci. USA*, **100**, 13253–13258.

22. Ahmed, A. & Gohlke, H. (2006). Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. *Proteins*, **63**, 1038–1051.

23. Yang, L., Song, G. & Jernigan, R. L. (2007). How well can we understand large-scale protein motions using normal modes of elastic network models? *Biophys. J.* **93**, 920–929.

24. Bahar, I., Chennubhotla, C. & Tobi, D. (2007). Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr. Opin. Struct. Biol.* **17**, 633–640.

25. Zheng, W., Brooks, B. R. & Thirumalai, D. (2007). Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations. *Biophys. J.* **93**, 2289–2299.

26. Zheng, W. (2009). Normal-mode-based modeling of allosteric couplings that underlie cyclic conformational transition in F(1) ATPase. *Proteins*, **76**, 747–762.

27. Ahmed, A., Villinger, S. & Gohlke, H. (2010). Large-scale comparison of protein essential dynamics from molecular dynamics simulations and coarse-grained normal mode analyses. *Proteins*, **78**, 3341–3352.

28. Burk, D. L., Hon, W. C., Leung, A. K. & Berghuis, A. M. (2001). Structural analyses of nucleotide binding to an aminoglycoside phosphotransferase. *Biochemistry*, **40**, 8756–8764.

29. Fong, D. H. & Berghuis, A. M. (2002). Substrate promiscuity of an aminoglycoside antibiotic resistance enzyme via target mimicry. *EMBO J.* **21**, 2323–2331.

30. Lee, T., Hoofnagle, A. N., Resing, K. A. & Ahn, N. G. (2005). Hydrogen exchange solvent protection by an ATP analogue reveals conformational changes in ERK2 upon activation. *J. Mol. Biol.* **353**, 600–612.

31. Polshakov, V. I., Birdsall, B. & Feeney, J. (2006). Effects of co-operative ligand binding on protein amide NH hydrogen exchange. *J. Mol. Biol.* **356**, 886–903.

32. DeLano, W. L. (2008). The PyMOL Molecular Graphics System DeLano Scientific LLC, Palo Alto, CA. http://www.pymol.org.

33. Wessel, P. & Smith, W. H. F. (1998). New, improved version of Generic Mapping Tools released. *EOS Trans. Am. Geophys. U.* **79**, 579.

34. Hayward, S., Kitao, A. & Berendsen, H. J. (1997). Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme. *Proteins*, **27**, 425–437.

35. Hinsen, K. (1998). Analysis of domain motions by approximate normal mode calculations. *Proteins*, **33**, 417–429.

36. Zheng, W. & Brooks, B. (2005). Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of an elastic network model. *J. Mol. Biol.* **346**, 745–759.

37. Yesylevskyy, S. O., Kharkyanen, V. N. & Demchenko, A. P. (2006). Hierarchical clustering of the correlation patterns: new method of domain identification in proteins. *Biophys. Chem.* **119**, 84–93.

38. Chavez, L. L., Onuchic, J. N. & Clementi, C. (2004). Quantifying the roughness on the free energy

landscape: entropic bottlenecks and protein folding rates. *J. Am. Chem. Soc.* **126**, 8426–8432.

39. Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O. & Bahar, I. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **80**, 505–515.

40. Zheng, W., Brooks, B. R., Doniach, S. & Thirumalai, D. (2005). Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved. *Structure*, **13**, 565–577.

41. Taly, A., Corringer, P. J., Grutter, T., Prado de Carvalho, L., Karplus, M. & Changeux, J. P. (2006). Implications of the quaternary twist allosteric model for the physiology and pathology of nicotinic acetylcholine receptors. *Proc. Natl Acad. Sci. USA*, **103**, 16965–16970.

42. Zheng, W. & Tekpinar, M. (2009). Large-scale evaluation of dynamically important residues in proteins predicted by the perturbation analysis of a coarse-grained elastic model. *BMC Struct. Biol.* **9**, 45.

43. McKay, G. A. & Wright, G. D. (1995). Kinetic mechanism of aminoglycoside phosphotransferase type IIIa. Evidence for a Theorell–Chance mechanism. *J. Biol. Chem.* **270**, 24686–24692.

44. Halle, B. (2002). Flexibility and packing in proteins. *Proc. Natl Acad. Sci. USA*, **99**, 1274–1279.

45. Rader, A. J., Chennubhotla, C., Yang, L. W. & Bahar, I. (2006). The Gaussian network model: theory and applications. In (Cui, Q. & Bahar, I., eds), pp. 41–64, Taylor and Francis Group, London, UK.

46. Bahar, I., Wallqvist, A., Covell, D. G. & Jernigan, R. L. (1998). Correlation between native-state hydrogen exchange and cooperative residue fluctuations from a simple model. *Biochemistry*, **37**, 1067–1075.

47. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD—Visual Molecular Dynamics. *J. Mol. Graphics*, **14**, 33–38.

48. Özen, C., Norris, A. L., Land, M. L., Tjioe, E. & Serpersu, E. H. (2008). Detection of specific solvent rearrangement regions of an enzyme: NMR and ITC studies with aminoglycoside phosphotransferase(3′)-IIIa. *Biochemistry*, **47**, 40–49.

49. Flory, P. J. (1976). Statistical thermodynamics of random networks. *Proc. R. Soc. London Ser. A*, **351**, 351–378.

50. Kloczkowski, A., Mark, J. E. & Erman, B. (1989). Chain dimensions and fluctuations in random elastomeric networks. *Macromolecules*, **10**, 1426–1432.

51. Tirion, M. M. (1996). Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **77**, 1905–1908.

52. Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J. *et al.* (1999). *LAPACK Users' Guide*, 3rd edit. Society for Industrial and Applied Mathematics, Philadelphia, PA.