Published in final edited form as:

Biochemistry. 2015 February 24; 54(7): 1534–1541. doi:10.1021/bi501152d.

Mapping Allostery through Computational Glycine Scanning and Correlation Analysis of Residue–Residue Contacts

Quentin R. Johnson[†,‡], **Richard J. Lindsay**[‡,§], **Ricky B. Nellas**[‖], **Elias J. Fernandez**[§], and **Tongye Shen**[†,§,*]

[†]UT-ORNL Graduate School of Genome Science and Technology, Knoxville, Tennessee 37996, United States

[‡]Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, United States

[§]Department of Biochemistry and Cellular & Molecular Biology, University of Tennessee, Knoxville, Tennessee 37996, United States

[‖]Institute of Chemistry, University of the Philippines Diliman, Quezon City 1101, Philippines

Abstract

Understanding allosteric mechanisms is essential for the physical control of molecular switches and downstream cellular responses. However, it is difficult to decode essential allosteric motions in a high-throughput scheme. A general two-pronged approach to performing automatic data reduction of simulation trajectories is presented here. The first step involves coarse-graining and identifying the most dynamic residue–residue contacts. The second step is performing principal component analysis of these contacts and extracting the large-scale collective motions expressed via these residue–residue contacts. We demonstrated the method using a protein complex of nuclear receptors. Using atomistic modeling and simulation, we examined the protein complex and a set of 18 glycine point mutations of residues that constitute the binding pocket of the ligand effector. The important motions that are responsible for the allostery are reported. In contrast to conventional induced-fit and lock-and-key binding mechanisms, a novel “frustrated-fit” binding mechanism of RXR for allosteric control was revealed.

Graphical Abstract

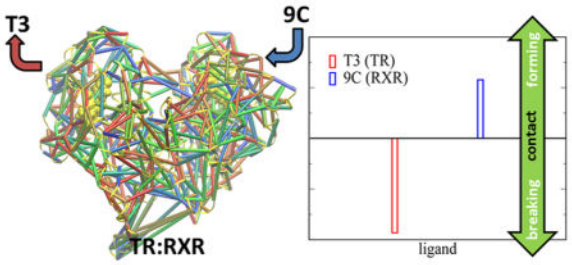
[*]Corresponding Author: tshen@utk.edu. Phone: (865) 974-4088. Fax: (865) 974-6306.

Notes

The authors declare no competing financial interest.

Supporting Information

One table and eight figures. This material is available free of charge via the Internet at http://pubs.acs.org.



Many biomolecules, especially signaling proteins, may have more than one accessible conformational state under physiological conditions. The barrier(s) separating these states is not prohibitively high.[3,4] Thus, these macromolecules are able to switch their conformations upon sensing environmental perturbations. There are a variety of such perturbations, such as interaction with another macromolecule,[1,5] interaction with a ligand,[6] solvent conditions[7–10]-(polarity and pH), physical environment[11–15] (salt, pressure, temperature, and light), or post-translational modifications.[13,16]

The conformational changes induced by these perturbations serve important biological functions. The changes can be local (centered around a few residues) or global (extensive and convoluted, involving the rearrangement of the whole molecule). The amplitude of these changes can range from subtle root-mean-square deviation changes to partial unfolding. One of the most well-studied cases of conformational switches is allostery.[1,17,18] A classic example of allostery involves the cooperative or anticooperative affinity of ligand binding pockets; i.e., binding of one ligand (effector) at one pocket may induce conformational changes that promote or inhibit the binding of the second ligand at a second pocket.[19]

Much success has been achieved in understanding allostery over the last century,[20] but questions about the mechanistic action of allostery, especially on negative allostery, remain, i.e., how one ligand can negatively influence a remote binding or active site.[21,22] Although there are textbook examples of thoroughly studied positive allosteric systems such as hemoglobin, negative allosteric regulation is infrequently reported. One possibility is that negative allostery is rarely used in nature's design of biochemical pathways. Another reason could be that identifying negative allostery with conventional methods is difficult. Indeed, allosteric mechanisms can have a variety of themes, and a specific mechanism can be painstaking to decode.[23] Given a particular state of a protein (or protein complex) that possesses multiple binding sites, can one predict whether the binding is cooperative or anticooperative? Do all negative (or positive) allosteric mechanisms share an underlying theme? Even more interestingly, some allosteric proteins (such as the one we will examine below) can display both positive or negative allostery, depending on the allosteric target. How do such molecular switches function? The answers to these questions could lead to future bioengineering and design of a new class of signaling proteins.

Several methods have been used to examine the intricate regulation and remote communication between binding sites.[6,24–28] Most methods so far have focused on the changes of elements of allostery (such as residue nativeness or residue–residue contacts) at the "mean-field" level. Typically, a group of potentially important elements of allostery are

first identified either through experimental mutation or through computational study. Then, further analysis using network or phylogenetic tree-type classification is applied to reconnect these isolated elements.[29] Because the process of identifying important elements is often performed and recorded independently, such as alanine scanning of individual residues, the connection between these isolated elements is ignored. What is missing from these approaches is the dynamic correlation between these elements. Indeed, as described by bioinformatics approaches, this aspect can be examined by analyzing the correlation between these elements, i.e., the coevolution of protein residues.[27,30] Similarly, coarse-grained models of proteins have been studied computationally to examine the local folding states (whether a residue is native or unfolded) and how the "melting" state of one residue affects another.[24,25] In such models, the element of allostery is the physical status or chemical identity of each residue.

In this study, we address the mechanism of allostery from the viewpoint of contact events (between residues) collected from atomistic simulations. Thus, the elements of allostery are not the status of individual residues, but the status of contacts (residue–residue interactions). The focus will be the dynamic correlation between the contact events, i.e., when residues *i* and *j* form a contact, whether residues *k* and *l* also form a contact. We first construct atomistic models and obtain simulation trajectories. We then extract the discrete contact degrees of freedom (DOFs) from simulation snapshots via coarse-graining and identify the important dynamic contacts. Finally, a statistical method (principal component analysis) is applied to these contact DOFs to identify the allosteric mechanism.

The method developed here is applied to a heterodimeric nuclear receptor (NR) complex. The nuclear receptor superfamily is a well-known group of signaling proteins that is responsible for sensing the presence of important small molecules for the downstream events of transcription. As the common subunit of heterodimers of the NR superfamily, promiscuous protein retinoid X receptor (RXR) displays essential allosteric regulation of other members in the superfamily. Previous experimental studies showed that the ligand binding status of RXR influences the ligand binding status of the thyroid hormone receptor (TR),[31] the constitutive androstane receptor (CAR),[32] and the liver X receptor (LXR).[33] For instance, the TR–RXR heterodimer can achieve full activity only upon binding of the 3,3′,5-triiodo-L-thyronine (T3) ligand to TR, and transactivation levels drop in the presence of the RXR ligand 9-*cis*-retinoic acid (9C), a clear case of negative cooperativity between the two binding sites.[31] On the other hand, RXR positively regulates the binding activity of CAR. Thus, how this promiscuous protein RXR can have different functions is quite intriguing. An illustration of the TR–RXR complex (ligand binding domains) is shown in Figure 1. Below, we will report how binding of 9C will induce the release of T3. We also will report an interesting "frustrated fit" between the binding pocket and its endogenous ligand, 9C.

SYSTEMS AND METHODS

Atomistic Modeling of the TR–RXR Complex and Glycine Scanning

Besides directly investigating the system of the wild-type protein complex TR(T3)–RXR(9C), a series of single-residue glycine mutants were constructed to enhance the sampling of conformational space. Traditionally, alanine scanning is used as a technique to

determine the contributions of individual amino acids to a protein's structure, function, and binding activity.[34] Alanine is an ideal residue for minimizing native residue–residue interactions for experimental methods, as it is relatively inert and has minimal steric and electrostatic impact.[35] In this study, a more drastic glycine mutation is used to scan the binding pocket. Glycine residues are completely devoid of side chains, and thus eliminate most residue–ligand contact interactions. As this is a perturbation study with the native structure as the initial state, the results are not impacted by potential ill effects (such as protein instability encountered in experiments) because they are not shown in a relatively short-time all-atom simulation.

As listed in Figure 1 and Table S1 of the Supporting Information, the 18 residues selected to be scanned form the binding site of RXR and can potentially form direct interactions with ligand 9-*cis*-retinoic acid (9C). The selection was made by directly inspecting the crystal structure and representative snapshots of the simulation of the wild-type protein. This set of 18 glycine mutants can perturb interactions between RXR and its ligand, 9C. These side chain-deleting mutants provide an ensemble of perturbed ligand-docked (holo) structures that are closer to the configurations of apo systems. Note, however, this particular residue–ligand interaction may not be completely removed by these point mutants. For example, in mutant A327G, glycine still can hold a backbone hydrogen bond with the carboxyl end of ligand 9C.

We also reason that, rather than comparing the snapshots from one long-time simulation of the wild type alone, an ensemble of glycine mutants will sample the representative populated configurations more efficiently in a given short time. Here we demonstrate that this perturbation scheme provides an important ensemble of configurations and utilizes them to detect the allosteric mechanism. Furthermore, statistical analysis of this perturbative ensemble of structures reveals important allosteric information about ligand binding and ligand crosstalk. In our design of the simulation procedure, we did not use the traditional method of directly studying the apo form (void of 9C) versus the ligand-bound form of the complex. The concern is that there is no crystal structure of the apo form. The structural changes between the apo form and the holo form are known to be drastic. If we start the simulation from the holo form of the structure with the ligand completely removed, it would take an unrealistic time for the system to reach the relaxed apo form. Rather, we focus on a set of more subtle perturbations of the holo form and perform short-time simulations. Thus, this scheme has the potential of being developed into a high-throughput method.

Simulation Setup

The starting structure for each of the 19 protein complexes was constructed from Protein Data Bank (PDB) entry 3UVV[22] via the xleap module of AMBER 10.[36] The force fields for protein and water are AMBER ff99SB[37] and TIP3P, respectively. Besides solvent, there are four separate entities in each system: protein TR (internal index 1–259, corresponding to standard residue index 147–405), protein RXR (internal index 260–491, corresponding to standard index 227–458), ligand T3 (3,3′,5-triiodo-L-thyronine, internal index 492), and ligand 9C (9-*cis*-retinoic acid, internal index 493). We use the standard residue index below unless otherwise specified.

Nuclear receptors have a common fold, the so-called α-helical sandwich,[38,39] which is mostly constructed from 11 or 12 helices. The loop (residues 195–202) that connects the first two helices of TR and the corresponding loop for RXR (residues 247–260) were missing in the reported PDB structure; these regions were modeled via ModLoop.[40] Ligands T3 and 9C were modeled using the antechamber module of AMBER 10.[36] The charges of the ligands were obtained from previous results[41] or from *ab initio* calculations.[42]

The size of each system is roughly that of the wild type (61625 atoms, including 17 sodium ions and 17906 waters). Note that each system size can vary slightly because of glycine mutation, but all the systems have a box size of ~88 Å × 98 Å × 94 Å. The energy-minimized wild-type complex was used as the starting structure for every mutant studied here. From this structure, 18 single-point glycine mutant complexes were constructed with the xleap module.

For each system, minimization, heating, and equilibration (for approximately 1 ns) were conducted and followed by a 20 ns *NPT* production run at 1 bar and 300 K using NAMD2.7.[43] The Langevin thermostat was used to regulate temperature; long-range interactions were treated using the particle mesh Ewald method,[44] and all bonds involving hydrogen atoms were constrained using the SHAKE[45] algorithm. A 2 fs time step was used for the simulation with snapshots taken and trajectories generated every 1 ps, producing 20000 snapshots for each system (and a total of 380000 snapshots). These snapshots, containing Cartesian coordinates of all the atoms, are the starting data sets for expressing the conformations in terms of contact degrees of freedom.

From the simulation snapshots, we have applied the computer program CAMERRA (Computation of Allosteric Mechanism by Evaluating Residue–Residue Associations) to perform the analysis of the allosteric motions that we report below. CAMERRA with sample data and results is available upon request.

RESULTS AND DISCUSSION

Selecting Important Contact Degrees of Freedom for Principal Component Analysis

To simplify the multidimensionality of the configurational space and to extract the allosteric motion, a data reduction scheme is desired to locate important degrees of freedom (DOFs). Often, principal component analysis is used for such tasks using biological systems.[46] Traditionally, Cartesian coordinates are used as the inputs for principal component analysis (PCA) in biomolecular simulations.[47] PCs of Cartesian coordinates resemble soft "vibrational" modes of atom positions. However, protein motion is extremely complex,[2] and expressing structural data in other formats, such as internal DOFs[48] and contact DOFs,[16,49] can have some advantage, depending on the type of motions (such as backbone rotation, vibration, and folding). In the case described in this work, allostery will be examined from the viewpoint of residue–residue interaction. Thus, the PC modes reported here correspond to the vibrational motion of contacts (contact forming and breaking). Residue–residue contact DOFs are convenient for describing large-scale motions of proteins such as protein folding,[1] and these can be an efficient way of coarse-graining atomistic information about large biomolecular systems, as well.

First, certain DOFs are removed from the system to produce coarse-grained representations of the residue level contact interactions. Here, each snapshot, initially expressed as the Cartesian coordinates of all the atoms, is condensed down to the information that indicates whether there is any direct contact between residues. We use discrete u_{ij} values of 1 and 0 to indicate whether contacts are formed. A contact is deemed to be formed between two residues when the distance between any atoms of said residues is less than 4.2 Å; in this case, $u_{ij} = 1$. Otherwise, $u_{ij} = 0$.

A protein (complex) of N residues has residue–residue contacts on the order of $O(N^2)$, which is typically larger than the number of Cartesian DOFs, even expressed at atomic resolution. Thus, one needs more numbers to express a particular configuration using contact DOFs. Fortunately, a protein has a much more sparse contact matrix if it is confined to a particular well-folded basin. If each residue has a maximum m possible native contacts (practically, m is a small number on the order of 10), there are $m \times N$ total contact DOFs. Besides those contacts that are never formed, the contact DOFs that are always formed should also be excluded. Neither of these two types of contact DOFs will be able to contribute to the important motions of the protein complex or reveal the intricate interactions between contact formations. For comparison, the luxury of limiting the contact DOFs to that of the native basin was absent in a previous study of folding and structure prediction.[49]

We first examined all potential contacts and ranked the percentage of formations during the simulation. Figure 2 shows the ranked contact curves (RCC), defined as the individual contact probability $p(i)$ sorted by its rank i. Note that all contacts (including the N diagonal portion of the contact matrix that gives a constant 1) are included in the RCC. The x-axis value J_{50} at which the curve intercepts with a y of 0.5 indicates the number of contacts that are formed ≥50% of the time during the simulation; i.e., $p(J_{50}) = 0.5$.

On the basis of the sigmoidal nature of the RCCs, a sharp “transition zone” separates a left region where contacts were always formed (near 100%) and a right one where contacts were never formed during the simulation. The most interesting contact DOFs are the contacts in the transition zone. Because these contacts are not always formed, their motions will be most important in describing the conformational dynamics of the protein complex. The combined RCC, which is constructed from all snapshots, is also displayed here. We then select the interesting contact DOFs, defined as those formed between 20 and 80% of the combined RCC for the analysis below. For this system, the “20/80” cutoff leaves a manageable amount of DOFs to perform the second stage of data reduction analysis, i.e., PCA studies. The total number of contact DOFs used in PCA is $J \equiv J_{20} - J_{80}$. As illustrated in Figure 2, the combined RCC provides the following cutoff values: $J_{20} = 3386$ and $J_{80} = 2897$. We generate a total of 489 DOFs shown as colored cylinders in Figure 2b. Note that they are broadly distributed across the complex.

It is important to point out the RCC of the combined data, i.e., the collection of all the snapshots from 18 glycine mutants, is not a direct average of individual RCC curves. We observe that the combined RCC has a more gradual decay with an increasing contact index. The combined RCC is softened with a broader tail and a less dramatic drop. The longer tail of the combined RCC is due to the variety of contacts that are formed in the combined

ensemble. A highly formed (near 100%) contact in a particular ensemble is not guaranteed to form in all other ensembles. Thus, the combined RCC is below the individual RCCs at this region.

Besides being used to select the important contacts to include in the analysis, RCCs can also provide information about the level of compactness of each mutant. In general, an RCC shift toward the right and up indicates more contacts are formed and thus the protein is more compact. Interestingly, all mutants show certain enhancement of contact formation compared to that of the wild-type complex, as shown in Figure 2c. This may indicate that the wild-type complex has certain frustration caused by steric interactions that are alleviated by glycine mutation. One can use J_{50} as a measurement of the compactness of a structure. Another way to measure the level of contact formation is to evaluate the mean contacts formed per residue. Here, each residue has ~10 contacting neighbors (including sequential neighbors) on average.

The J_{50} and mean contact per residue are strongly correlated. Both of these properties indicate an intriguing fact that the binding pocket of each contact-deleting mutant is able to accommodate the endogenous ligand better and makes the complex more compact in comparison to the wild type. For the extreme case, the removal of residue H435 can increase the number of global contacts by ~1% relative to that of the wild type. This indicates that the binding of ligand 9C creates steric frustration (especially at the ring end of the ligand), which forces the molecule to adopt an ensemble of more flexible conformations. Thus, the binding pocket of T3 does not form well and decreases the affinity for T3, which is consistent with the experimental evidence that upon binding 9C, RXR changes conformation and TR becomes more “loose” and dynamic.[31] Experimental evidence showed a more dispersed electron density in the holo system (with both ligands) than in the apo system (without 9C).[31] The trend is quite pronounced for TR, which leads to the conclusion that TR becomes more flexible upon the binding of 9C to RXR.[31] This conclusion is also supported by the computed radius of gyration analysis, which indicates that mutants are more compact than the wild type (data shown in Figure S3 of the Supporting Information).

Principal Components and Projections of Conformations onto Them

In the case presented in this work, PCA is performed on the selected total of 489 contact DOFs introduced in the previous section. We first calculate the elements of the covariance matrix (size of $J \times J$) $\mathbf{C}_{\alpha\beta} \equiv \mathbf{C}_{i;j,k;l}$, which describes the level of correlation between contact events u indexed by $\alpha = 1, 2, 3, \ldots, J$ (between residues i and j) with the event $\beta = 1, 2, 3, \ldots, J$ (between residues k and l). Thus, covariance matrix $\mathbf{C}$ is defined as $\mathbf{C}_{ijkl} = \langle (u_{ij} - \langle u_{ij} \rangle) \times (u_{kl} - \langle u_{kl} \rangle) \rangle$. We use a collection of snapshots (380000 frames) from 19 mutant and wild-type simulations. The results presented below are based on this combined covariance matrix.

For PCA, covariance matrix $\mathbf{C}$ is then diagonalized, and transformation matrix $\mathbf{U}$ is obtained, which yields $\mathbf{U}^{\mathrm{T}}\mathbf{C}\mathbf{U} = \Lambda$, where matrix $\mathbf{U}$ gives a recipe to linearly transform the contact DOFs to another set of DOFs and matrix Λ is diagonal. The significance of the transformed DOFs, termed principal components (PCs), is that they are directly ranked by their importance according to eigenvalue Λ. The top PCs (containing the largest Λs) provide the major modes of protein dynamics in terms of contacts. Ideally, the fluctuation of

residue–residue contact and detachment events contains only the largest few PCs for the dimension reduction, and the sorted eigenvalue Λ_i drops with index i. The 65th largest PC is less than one-tenth of the largest. The relatively slow decay of eigenvalues suggests that quite a few modes contribute to protein motion. However, it does not change the ranking of the top PCs. These top PCs still hold the most likely motions. Although the decrease is not extremely rapid, the top 13 PCs together contribute more than half of the whole amplitude. We will focus on PC1 and PC2 below.

We display the top two PCs (normalized eigenvectors) in Figure 3. Principal component 1 (PC1) is of immediate interest because it represents the largest mode of movements and will likely exhibit the most interesting information. Here, in Figure 3a, the normalized eigenvector $PC1_{ij}$ is shown on a residue–residue contact map, and the values of these 489 contacts are color-coded. The black dashed lines demarcate the intrachain contact region and the interchain region. Figure 3b contains the same information about PC1 but focuses on the contacts between the protein and the ligand. Although there are a variety of positive and negative values for each contact along PC1, the overall contact with ligand T3 is positively correlated along PC1 while 9C is negatively correlated. This means that the modes of ligand–protein interaction are anticorrelated and is an indication that when one ligand is drawn closer to its binding site, the second ligand will escape from its binding site. Our negative allostery results indicate the binding of T3 is weakened once 9C is bound. This is consistent with the earlier experimental study that examined the binding affinity and reported an increase in the dissociation constant and a decrease in the binding constant of the T3 pocket in the TR(T3)–RXR(9C) complex in comparison to those of the TR(T3)–RXR(ø) complex.[31]

Besides the two-dimensional plot (Figure 3a) and the ligand plot (Figure 3b), a cylinder representation is also used to display the information about PC1 in a three-dimensional view (front and back), as shown in Figure 3d. Additional side views are presented in Figures S5 and S6 of the Supporting Information. Here the contacts are again visualized as cylinders, similar to that of the mean contact plot in Figure 2b. All three representations indicate the dynamics of the PC1 mode. Overall, the most dramatic dynamics of PC1 center around the binding pocket of T3, which is remarkable given that all the mutations directly probe the 9C binding of the RXR side. Previously, it was reported that electron density for the T3 binding site in TR is markedly discontinuous in the TR(T3)–RXR(9C) complex.[31] In this PC1 mode, there is one main region of interchain contact. Helix H11 of RXR and helix H10′ of TR seem to hold many of the important interchain (interfacial) contact pairs. Also, the contact between helix H6′ of TR and the turn region (residues 377–383) of RXR contributes to the interfacial interaction between TR and RXR.

Similarly, Figure 3c is a contact map view of the PC2 value for each residue–residue contact. While PC2 may be less important than PC1, there is some useful information in this mode, which will be discussed in the PC projection below. The values in PC2 mode have smaller amplitudes, as expected.

Figure 4 depicts the conformation (contact states) of all systems studied projected onto the top two principal components (presented in Figure 3). Projecting the dynamic snapshots of

all systems onto these two PCs may enhance our understanding of the propagation of the allosteric control utilized by the TR–RXR complex. In this case, PC1 distinguishes the wild type from the mutants, as the wild type is located on the extreme positive end of PC1 in Figure 4, reaching values well above 4.0. Opposite to the wild type on the extreme negative end of PC1 lies mutant R316, reaching values of less than –4.0.

It is interesting to point out mutation R316 is expected to cause the most drastic changes to the binding interaction, because the arginine to glycine mutation removes the salt bridge interaction between the side chain of arginine and the carboxylic group of ligand 9C. We suggest that PC1 resembles the docking mode of the carboxyl end of ligand 9C. Additionally, L433 and L436 have similar values along PC1 and PC2. These two residues have the same chemical properties and are also found only three residues apart, which explains their similar behavior along these modes. Interestingly, A327 and R316, which are found in a very similar locale, have very different PC1 values. This can be explained by the difference in their chemical properties. Compared to the mutations near the ring of 9C, those residues near the carboxyl end of 9C tend to have lower PC1 values.

The wild-type TR–RXR complex shows relatively high values of PC2, as well. PC2 is used here as another way to differentiate mutant from wild-type conformations. Mutant Q275 shows the lowest values along PC2, while F313 shows the highest values. Mutants C432, L433, H435, and L436 (all similar locales) show very similar PC2 values. One may wonder whether 20 ns simulations are sufficiently long to obtain convergent PCA results. To examine the dynamics aspect of protein motion projected onto the top PCs, we split the same data into two halves (10 ns each) as shown in Figure S8 of the Supporting Information. We can see that most systems are relaxed to a stationary ensemble and show little change between the two halves. There are a couple exceptions that indicate transient motions. Note that our perturbation scheme aims to generate a set of diverse and slightly perturbed structures, not necessarily fully equilibrated ensembles.

CONCLUSIONS

We present a method that isolates important allosteric modes from glycine mutant simulations by performing PCA on selected dynamic contacts. This two-stage data reduction scheme [(1) selecting dynamic contacts based on RCC and (2) rendering important contact collective modes by PCA] is efficient in terms of computational cost and maintains the atomistic interactions of the system. The method is applied to reveal the interactions between ligands of the TR–RXR NR complex. Ligand 9C binds RXR, which invokes frustration of the binding site, and this frustration is propagated through the interface of the TR–RXR complex. This “frustrated fit” eventually leads to a decrease in the level of contact between TR and its ligand, T3. Understanding the allostery of RXR with its endogenous ligand will serve as a model system for studying rational drug design of antagonists that control the regulation pathway. Besides other NRs in the superfamily, the method presented here can be generally useful in the study of many other allosteric systems and large-scale motions of flexible proteins.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

Acknowledgment is made to the Donors of the American Chemical Society Petroleum Research Fund (52616-DNI6) for partial support of this research. Q.R.J. was supported by a National Science Foundation-funded graduate fellowship program SCALE-IT. E.J.F. is supported in part by National Institutes of Health Grant DK097337-01. Support from the JDRD program of Science Alliance at UT-ORNL is also acknowledged.

We acknowledge the computational support provided by the UT-ORNL Center for Molecular Biophysics and by allocations of advanced computing resources (Kraken at NICS and STAMPEDE at TACC) provided by the National Science Foundation.

ABBREVIATIONS

RXR	retinoid X receptor
TR	thyroid hormone receptor
9C	9-*cis*-retinoic acid
T3	3,3′,5-triiodo-L-thyronine
CAR	constitutive androstane receptor
NR	nuclear receptor
PCA	principal component analysis
PC	principal component
DOF	degree of freedom

References

1. Fersht, A. Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding. W. H. Freeman; New York: 1999.
2. Frauenfelder, H.; Deisenhofer, J.; Wolynes, PG. Simplicity and Complexity in Proteins and Nucleic Acids. Dahlem University Press; Berlin: 2000.
3. Miyashita O, Onuchic JN, Wolynes PG. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. Proc Natl Acad Sci USA. 2003; 100:12570–12575. [PubMed: 14566052]
4. Tripathi S, Portman JJ. Allostery and folding of the N-terminal receiver domain of protein ntrc. J Phys Chem B. 2013; 117:13182–13193. [PubMed: 23961720]
5. Taylor, ME.; Drickamer, K. Introduction to Glycobiology. Oxford University Press; New York: 2003.
6. Hardy JA, Wells JA. Searching for new allosteric sites in enzymes. Curr Opin Struct Biol. 2004; 14:706–715. [PubMed: 15582395]
7. Colombo MF, Rau DC, Parsegian VA. Protein solvation in allosteric regulation: A water effect on hemoglobin. Science. 1992; 256:655–659. [PubMed: 1585178]
8. Nellas RB, Johnson QR, Shen T. Solvent-induced α- to 3_{10}-helix transition of an amphiphilic peptide. Biochemistry. 2013; 52:7137–7144. [PubMed: 24066804]

9. Ehrlich LS, Liu T, Scarlata S, Chu B, Carter CA. HIV-1 capsid protein forms spherical (immature-like) and tubular (mature-like) particles in vitro: Structure switching by pH-induced conformational changes. Biophys J. 2001; 81:586–594. [PubMed: 11423440]
10. Barrera FN, Fendos J, Engelman DM. Membrane physical properties influence transmembrane helix formation. Proc Natl Acad Sci USA. 2012; 109:14422–14427. [PubMed: 22908237]
11. Wroblowski B, Díaz JF, Heremans K, Engelborghs Y. Molecular mechanisms of pressure induced conformational changes in bpti. Proteins: Struct, Funct, Bioinf. 1996; 25:446–455.
12. Schlesinger MJ. Heat shock proteins. J Biol Chem. 1990; 265:12111–12114. [PubMed: 2197269]
13. Ha J, Loh SN. Protein conformational switches: From nature to design. Chemistry. 1990; 18:7984–7999. [PubMed: 22688954]
14. Imamoto Y, Kamikubo H, Harigai M, Shimizu N, Kataoka M. Light-induced global conformational change of photoactive yellow protein in solution. Biochemistry. 2002; 41:13595–13601. [PubMed: 12427020]
15. Vreede1 J, Juraszek J, Bolhuis PG. Predicting the reaction coordinates of millisecond light-induced conformational changes in photoactive yellow protein. Proc Natl Acad Sci USA. 2010; 107:2397–2402. [PubMed: 20133754]
16. Latzer J, Shen T, Wolynes PG. Conformational switching upon phosphorylation: A predictive framework based on energy landscape principles. Biochemistry. 2008; 47:2110–2122. [PubMed: 18198897]
17. Hill, T. Cooperativity Theory in Biochemistry: Steady-State and Equilibrium Systems. 1. Springer; Berlin: 2011.
18. Cui Q, Karplus M. Allostery and cooperativity revisited. Protein Sci. 2008; 17:1295–1307. [PubMed: 18560010]
19. Pauling L. The oxygen equilibrium of hemoglobin and its structural interpretation. Proc Natl Acad Sci USA. 1935; 21:186–191. [PubMed: 16587956]
20. Hill R. Chemical nature of haemochromogen and its carbon monoxide compound. Proc R Soc B. 1926; 100:419–430.
21. Stevens S, Sanker S, Kent C, Zuiderweg E. Delineation of the allosteric mechanism of a cytidylyltransferase exhibiting negative cooperativity. Nat Struct Biol. 2001; 8:947–952. [PubMed: 11685240]
22. Putcha BK, Wright E, Brunzelle JS, Fernandez EJ. Structural basis for negative cooperativity within agonist-bound TR:RXR heterodimers. Proc Natl Acad Sci USA. 2012; 109:6084–6087. [PubMed: 22474364]
23. Ferreon AC, Ferreon JC, Wright PE, Deniz AA. Modulation of allostery by protein intrinsic disorder. Nature. 2013; 498:390–394. [PubMed: 23783631]
24. Pan H, Lee JC, Hilser VJ. Binding sites in *Escherichia coli* dihydrofolate reductase communicate by modulating the conformational ensemble. Proc Natl Acad Sci USA. 2000; 97:12020–12025. [PubMed: 11035796]
25. Freire E. The propagation of binding interactions to remote sites in proteins: Analysis of the binding of the monoclonal antibody d1.3 to lysozyme. Proc Natl Acad Sci USA. 1999; 96:10118–10122. [PubMed: 10468572]
26. Sethy A, Tian J, Derdeyn CA, Korber B, Gnanakaran S. A mechanistic understanding of allosteric immune escape pathways in the HIV-1 envelope glycoprotein. PLoS Comput Biol. 2013; 9:e1003046. [PubMed: 23696718]
27. Shulman AI, Larson C, Mangelsdorf DJ, Ranganathan R. Structural determinants of allosteric ligand activation in RXR heterodimers. Cell. 2004; 116:417–429. [PubMed: 15016376]
28. Jackson, MB. Molecular and Cellular Biophysics. 1. Cambridge University Press; Cambridge, U.K: 2006.
29. Newman, M. Networks: An introduction. 1. Oxford University Press; Oxford, U.K: 2010.
30. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. Science. 1999; 286:295–299. [PubMed: 10514373]
31. Putcha BDK, Fernandez EJ. Direct inter-domain interactions can mediate allosterism in the thyroid receptor. J Biol Chem. 2009; 284:22517–22524. [PubMed: 19561066]

32. Wright E, Busby SA, Wisecarver S, Vincent J, Griffin PR, Fernandez EJ. Helix 11 dynamics is critical for constitutive androstane receptor activity. Structure. 2011; 19:37–44. [PubMed: 21220114]
33. Shulman AI, Larson C, Mangelsdorf DJ, Ranganathan R. Structural determinants of allosteric ligand activation in RXR heterodimers. Cell. 2004; 116:417–429. [PubMed: 15016376]
34. Moreira IS, Fernandes PA, Ramos MJ. Hot spots: A review of the protein-protein interface determinant aminoacid residues. Proteins: Struct, Funct, Bioinf. 2007; 68:803–812.
35. DeLano WL. Unraveling hot spots in binding interfaces: Progress and challenges. Curr Opin Struct Biol. 2002; 12:14–20. [PubMed: 11839484]
36. Case, DA.; Darden, TA.; Cheatham, TEI.; Simmerling, CL.; Wang, J.; Duke, RE.; Luo, R.; Crowly, M.; Walker, RC.; Zhang, W.; Merz, KM.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvry, I.; Wong, KF.; Pasani, F.; Vanicek, J.; Wu, X.; Brozell, SR.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, DH.; Seetin, MG.; Sagui, C.; Babin, V.; Kollman, PA. AMBER 10. University of California; San Francisco: 2008.
37. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins: Struct, Funct, Bioinf. 2006; 65:712–725.
38. Moras D, Gronemeyer H. The nuclear receptor ligand-binding domain: Structure and function. Curr Opin Cell Biol. 1998; 10:384–391. [PubMed: 9640540]
39. Bourguet W, Ruff M, Chambon P, Gronemeyer H, Moras D. Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-α. Nature. 1995; 375:377–382. [PubMed: 7760929]
40. Fiser A, Sali A. Modloop: Automated modeling of loops in protein structures. Bioinformatics. 2003; 19:2500–2501. [PubMed: 14668246]
41. Martinez L, Webb P, Polikarpov I, Skaf MS. Molecular Dynamics Simulations of Ligand Dissociation from Thyroid Hormone Receptors: Evidence of the Likeliest Escape Pathway and Its Implications for the Design of Novel Ligands. J Med Chem. 2006; 49:23–26. [PubMed: 16392786]
42. Valiev M, Bylaska EJ, Govind N, Kowalski K, Straatsma T, van Dam HJJ, Wang D, Nieplocha J, Apra E, Windus TL, de Jong WA. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. Comput Phys Commun. 2010; 181:1477–1489.
43. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem. 2005; 26:1781–1802. [PubMed: 16222654]
44. Darden T, York D, Pedersen L. Particle mesh ewald: An n.log(n) method for ewald sums in large systems. J Chem Phys. 1993; 98:10089–10092.
45. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. J Comput Phys. 1977; 23:327–341.
46. Jolliffe, IT. Principal component analysis. Springer-Verlag; New York: 1986.
47. Garcia AE. Large-amplitude nonlinear motions in proteins. Phys Rev Lett. 1992; 68:2696–2699. [PubMed: 10045464]
48. Altis A, Nguyen P, Hegger R, Stock G. Dihedral angle principal component analysis of molecular dynamics simulations. J Chem Phys. 2007; 126:244111. [PubMed: 17614541]
49. Shen T, Zong C, Hamelberg D, McCammon JA, Wolynes PG. The folding energy landscape and phosphorylation: Modeling the conformational switch of the NFAT regulatory domain. FASEB J. 2005; 19:1389–1395. [PubMed: 16126906]

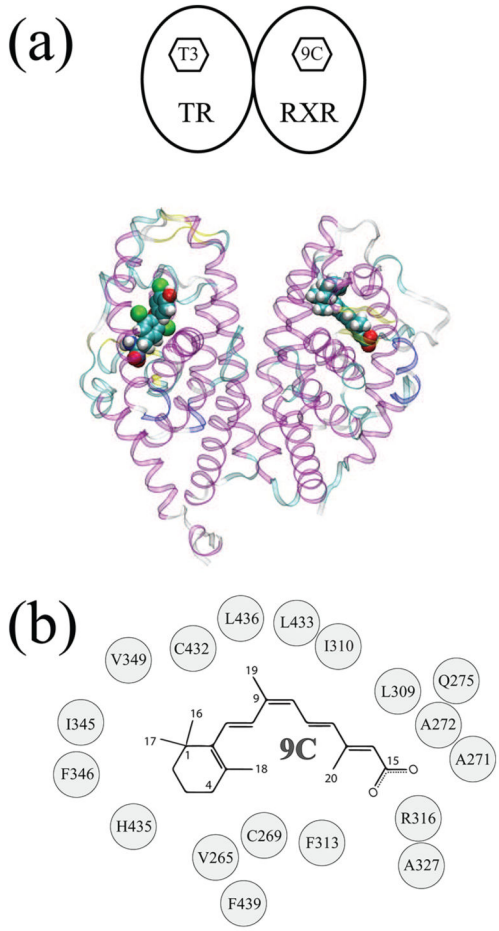


Figure 1.

Several depictions of the structure of the RXR–TR complex with varying levels of detail. (a) Cartoon rendering of the structure of the RXR(9C)–TR(T3) complex. (b) Interaction between the RXR binding pocket and ligand 9C. The amino acid residues used to construct glycine mutants in this study are listed.

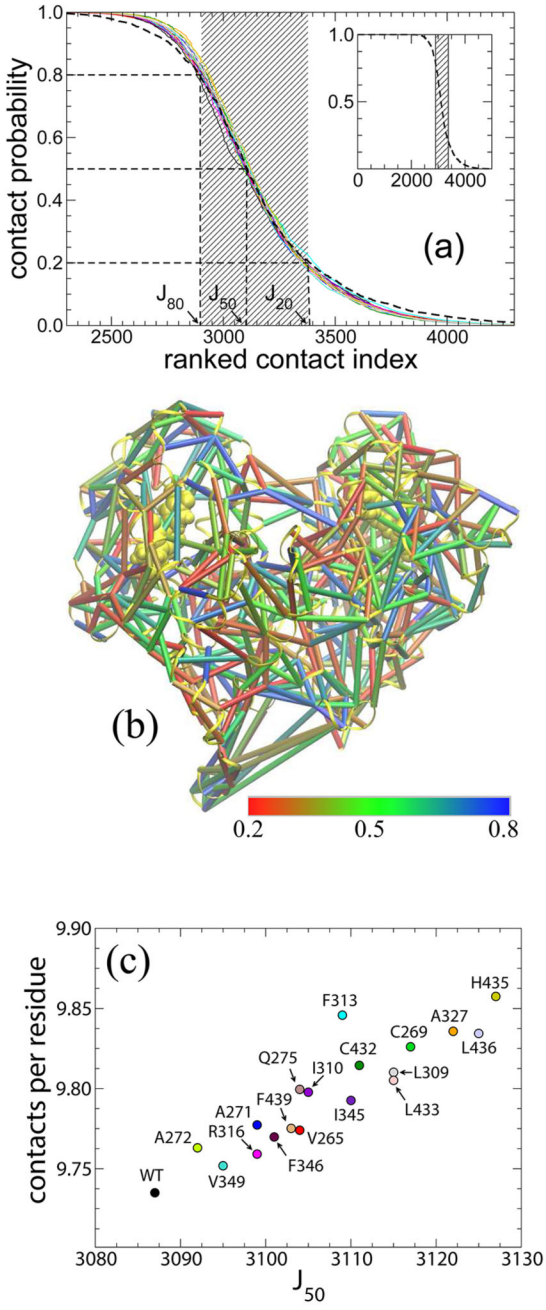


Figure 2.

Residue–residue contact analysis. (a) The ranked contact curves (RCCs) display the probability of residue–residue contacts for each individual system (thin lines) and the combined data (thick dashes). Note that there are 19 colored lines; each color represents an individual system from this study. The color scheme is displayed in panel c. Positions of J_{20}, J_{50}, and J_{80} of the combined RCC are labeled. The inset presents the overall view of the combined RCC. (b) The average contact probabilities for those contacts selected (between 20 and 80% formed) are displayed as cylinders (total of 489) between residues in the complex structure. The color spectrum is from red (0.2) to green (0.5) to blue (0.8). (c) The

value of J_{50} vs the corresponding contacts per residue is displayed here for individual mutants and the wild type. The color scheme for the 19 systems is consistent throughout this work.

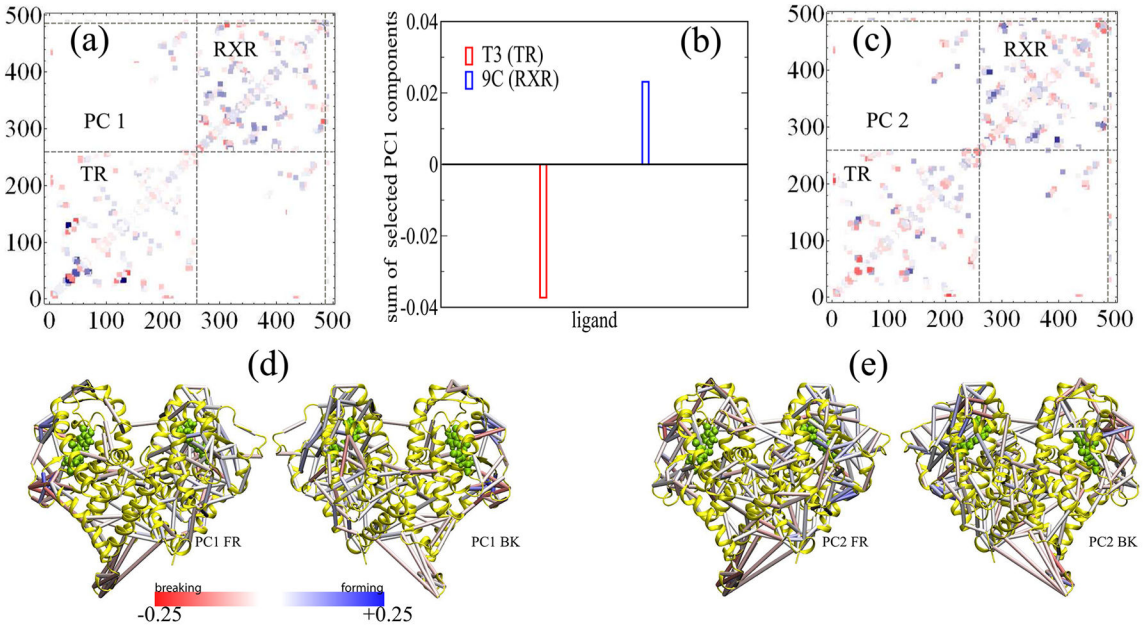


Figure 3.

(a) Eigenvector PC1 displayed on the residue–residue contact map. The internal index of the residues is used here for the axes. The dashed lines separate intra- and interprotein contacts. Another set of dashes demarcates the contacts formed between ligands (internal indices 492 and 493) and protein residues. (b) Sum of selected elements in eigenvector PC1 that contain the contact interaction between ligands (T3 and 9C) and the protein complex. The values of individual elements are listed in Figure S4 of the Supporting Information. (c) Eigenvector PC2 displayed on the contact map. Three-dimensional representations of PC1 and PC2 values are shown in panels d and e, respectively. For the sake of clarity, only those contacts with absolute values of >0.03 are explicitly displayed as colored cylinders.

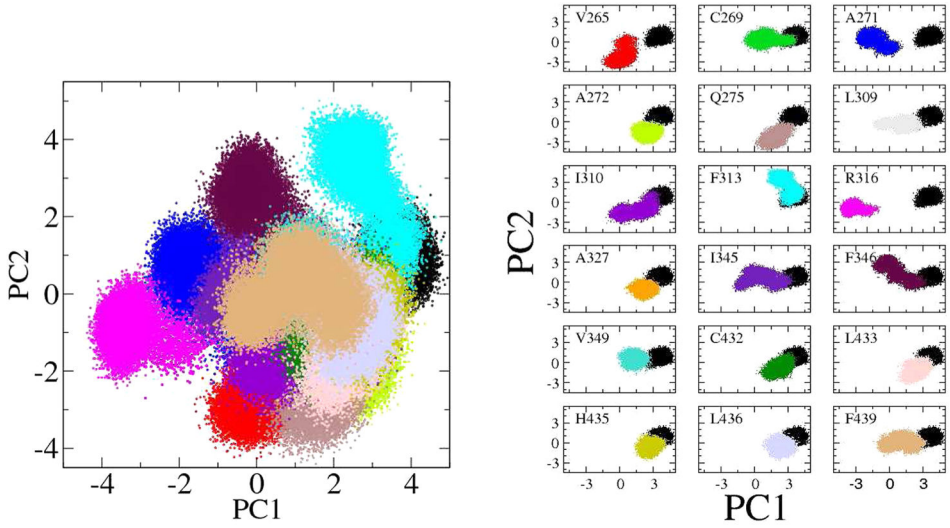


Figure 4.

Projection of all snapshots onto top PCs. The configurations of the protein–ligand complex are projected onto PC1 and PC2. Each of the 380000 points represents a specific configuration of the system obtained from a snapshot, as shown at the left. The color scheme matches that of previous figures. Each individual mutant is also displayed in a multipanel view at the right for a clear comparison with the wild type.