Cell, Vol. 116, 417–429, February 6, 2004, Copyright ©2004 by Cell Press

# Structural Determinants of Allosteric Ligand Activation in RXR Heterodimers

**Andrew I. Shulman, Christopher Larson, David J. Mangelsdorf,* and Rama Ranganathan**
Howard Hughes Medical Institute
Department of Pharmacology
University of Texas Southwestern Medical Center
5323 Harry Hines Boulevard
Dallas, Texas 75390

## Summary

**Allosteric communication underlies ligand-dependent transcriptional responses mediated by nuclear receptors. While studies have elucidated many of the components involved in this process, the energetic architecture within the receptor protein that mediates allostery remains unknown. Using a sequence-based method designed to detect coevolution of amino acids in a protein, termed the statistical coupling analysis (SCA), we identify a network of energetically coupled residues that link the functional surfaces of nuclear receptor ligand binding domains. Functional analysis of these predicted residues demonstrates their participation in an allosteric network that governs the ability of heterodimeric receptors to activate transcription in response to ligand binding by either partner. Interestingly, mutation of a single network residue can discriminate between receptor activation by endocrine, dietary, and synthetic agonists. These results reveal a structural network required for RXR heterodimer allosteric communication and suggest that the specificity of ligand response and permissivity coevolved to enable signal discrimination.**

## Introduction

Allosteric communication, the long-range energetic interaction of sites on proteins, is fundamental for many biological processes. Examples include ligand activation of G protein-coupled receptors (Menon et al., 2001), regulation of catalytic activity by phosphorylation in tyrosine kinases (Young et al., 2001), and ligand-dependent transcription in nuclear receptors (Mangelsdorf and Evans, 1995). The core biological role of each of these proteins is not defined by the independent activity of functional surfaces, but from the efficient coupling of two or more molecular interactions to produce regulated and stimulus-dependent output responses.

An excellent model system for studying allosteric communication is the ligand binding domain (LBD) of the nuclear receptors. These receptors are modular transcription factors that recognize response elements in target promoters, typically as homo- or heterodimers, with a highly conserved DNA binding domain (Mangelsdorf and Evans, 1995). Transcriptional activation is mediated by the LBD, a complex domain that contains four structurally distinct but functionally linked surfaces: (1) the ligand binding pocket, which interacts with diverse lipophilic small molecules; (2) a cofactor binding surface, which binds to regulatory protein complexes that modulate transcriptional activity; (3) an activation function 2 (AF2) helix, which mediates ligand-dependent transactivation; and (4) a dimerization surface, which mediates interaction with partner LBDs. Nuclear receptor activity requires a complex allosteric interaction between all four LBD functional surfaces. For example, ligand binding induces a conformational change in the cofactor binding site and AF2 helix that leads to exchange of corepressors for coactivators and transcriptional activation. In nuclear receptors that form heterodimers with the retinoid X receptors (RXRs), ligand binding also affects the stability and propagation of signals across the heterodimerization interface, indicating that the ligand binding pocket and dimerization interface are in some way energetically linked (Cheskis and Freedman, 1996; Thompson et al., 1998). This allosteric coupling allows ligands of one member of an RXR heterodimer to regulate the activity of its partner LBD, a phenomenon termed the "phantom ligand effect" (Schulman et al., 1997; Willy and Mangelsdorf, 1997). Thus, the LBDs of nuclear receptors are complex allosteric signaling domains that are able to integrate multiple molecular interactions at structurally distinct sites to modulate transcriptional activation.

How can we understand the structural principles underlying signal transmission in protein domains such as the nuclear receptor LBD? Despite the fundamental nature of this problem and considerable structural and biochemical analyses, the molecular basis for long-range communication remains poorly understood in even the best-studied proteins. For example, atomic structures of LBDs in complex with ligands, coactivator peptides, and several heterodimeric partners have provided key information about the chemistry of individual interactions, but have not explained the mechanisms that energetically couple these interactions. Structural analyses of many other proteins share similar limitations. The essence of the problem is the difficulty of predicting conformational change in protein structures; that is, how the effect of a perturbation (such as ligand binding) at any given site in a protein propagates through the protein structure to affect other sites. Fundamentally, this problem is due to our inability to estimate the energetic value of interactions between amino acid residues in even high-resolution crystal structures. An understanding of allosteric coupling in proteins requires first identifying the residues that energetically participate in the communication processes, and then ultimately understanding the physical mechanism by which they interact.

One approach to identifying such residues is large-scale systematic mutagenesis. Techniques such as alanine scanning mutagenesis have been applied successfully to map the importance of individual sites within proteins (Wells, 1991), but do not provide information about the interaction between sites. Other methods, such as double mutant cycle analysis (Carter et al., 1984), are well suited to mapping interactions between amino acids, but practical issues limit these studies to

*Correspondence: davo.mango@utsouthwestern.edu

only small regions of proteins. To map amino acid interaction wholesale, we recently introduced an alternative method called the statistical coupling analysis (SCA) that is based on simple rules of molecular evolution (Lockless and Ranganathan, 1999). The core principle of this method is that evolution represents a large-scale experiment in mutagenesis with selection for function, and that the functional interaction of two residues in a protein drives the coevolution of the two positions. This coevolutionary constraint can be extracted from statistical analysis of a large and diverse multiple sequence alignment of a protein family by selecting a subset of sequences in which a fixed amino acid appears at a specified position and then assessing the effect of this statistical perturbation on the amino acid distribution at other sites (Lockless and Ranganathan, 1999). Previous studies have shown that this sequence-based analysis is well correlated with thermodynamic coupling energies measured through mutagenesis and is consistent with the known allosteric mechanism in several classically studied protein families (Suel et al., 2003).

In this study, we applied the SCA to map the global energetic architecture of the nuclear receptor LBD. Remarkably, the SCA identified a sparse network of coevolving residues that connects the four main functional surfaces of the LBD through a specific set of core contacts. Mutagenesis studies confirm the role of this network in mediating allosteric communication in the LBD and demonstrate that a subset of core residues carry information between functional surface sites. Importantly, structural and sequence conservation analyses have not suggested functional roles for these core residues. In the RXR heterodimeric receptors, this network reveals the molecular basis for the phantom ligand effect by identifying previously unknown residues that participate in this phenomenon. In addition, this analysis shows how distinct ligands might differentially engage a receptor's allosteric mechanism to produce biologically diverse responses.

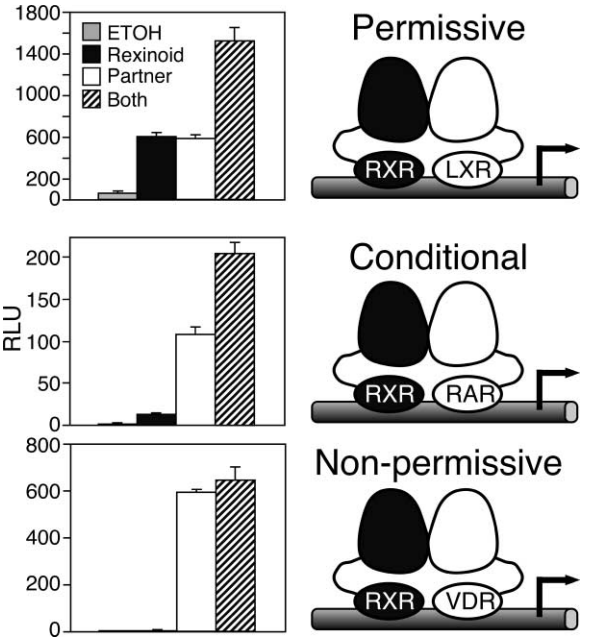


Figure 1. RXR Heterodimers Demonstrate Differential Response to Ligands

Top: activation of a tk-LXREx3-luc reporter by CMX-hLXR$\alpha$ in HEK293 cells (which express endogenous RXR) and by treatment with RXR agonist LG268 (10 nM) and LXR agonist T0901317 (1 $\mu$M) (Repa et al., 2000). Middle: activation of a tk-TREx2-luc reporter by CMX-hRAR$\alpha$ in HEK293 cells and by treatment with LG268 (10 nM) and the RAR agonist TTNPB (10 nM) (Kurokawa et al., 1994). Bottom: activation of a tk-mSppx3-luc reporter by CMX-hVDR in HEK293 cells and by treatment with LG268 (100 nM) and 1, 25 $(OH)_2$ vitamin D3 (100 nM). RLU, relative light units.

## Results

### Differential Ligand Response in RXR Heterodimers

Nuclear receptor heterodimers integrate the signaling capacity of two LBDs into a combined transcriptional response that satisfies precise requirements for the control of gene expression in many physiological systems. The ability of the common heterodimeric partner, RXR, to bind endogenous ligands, such as 9-*cis* retinoic acid, and a variety of synthetic agonists (called rexinoids) imparts RXR heterodimers with the potential to be activated by both RXR agonists and ligands for the partner receptor (Chawla et al., 2001). RXR heterodimers exhibit three modes of activation that reveal the existence of a ligand-mediated allosteric pathway. Examples of all three modes of activation are shown in Figure 1, where the response of a representative set of RXR heterodimers to their cognate ligands was tested in a standard cotransfection assay. In the first example, the RXR/liver X receptor (LXR) heterodimer exhibited dual ligand permissivity since it could be activated by rexinoid, LXR agonist, or both agonists in a more than additive fashion (Figure 1, top). In the second example, the RXR/retinoic acid receptor (RAR) heterodimer demonstrated conditional permissivity since full response to rexinoid occurred only in the presence of an RAR agonist (Figure 1, middle). The third example is represented by the nonpermissive RXR/vitamin D receptor (VDR) heterodimer, which cannot be activated by rexinoid either in the presence or absence of VDR ligand (Figure 1, bottom). Previous work has shown that in the context of each of these heterodimers, RXR retains the ability to bind ligand (Cheskis and Freedman, 1996; Germain et al., 2002; Thompson et al., 1998). How then is RXR activity controlled by its binding partner? Although the ultimate "read-out" governing permissive versus nonpermissive ligand activation involves the differential binding of cofactor proteins (Germain et al., 2002), the proximal events governing transactivation of the heterodimeric complex remain unknown. Importantly, the studies above imply that these proximal events include not only ligand binding, but also the engagement of a mechanism for coupling ligand binding to transactivation across the heterodimerization interface.

### Statistical Coupling Analysis of the Nuclear Receptor Alignment

At a minimum, the physical network of amino acids that mediates nuclear receptor transactivation should energetically link four distantly positioned functional surfaces: the dimerization interface, the coactivator binding

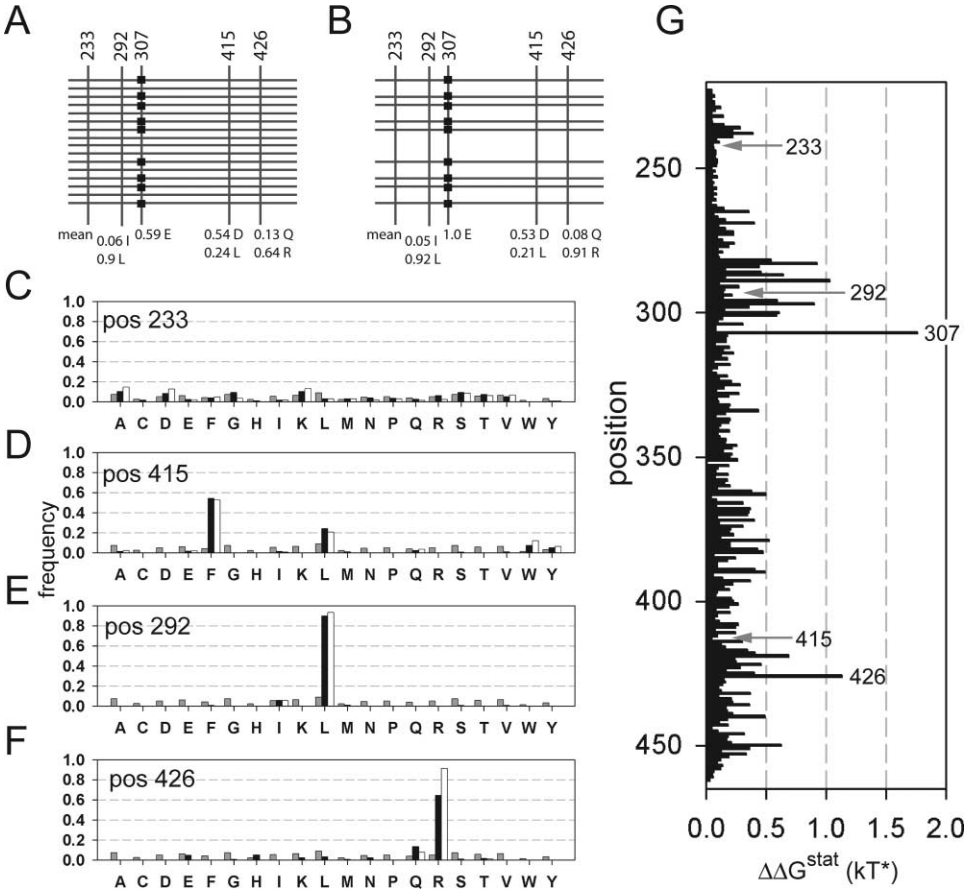


Figure 2. The Statistical Coupling Analysis in an Alignment of 560 Nuclear Hormone Receptors

(A) A schematic representation of the alignment, showing the frequencies of the dominant amino acids at five example sites, 233 (mean), 292 (6% Ile, 90% Leu), 307 (59% Glu), 415 (54% Asp, 24% Leu), and 426 (13% Gln, 64% Arg). Residue numbering is arbitrarily based on the primary sequence of human RXR$\alpha$ (accession # 10862707). Sequences with Glu at position 307 are indicated with black boxes

(B) The subalignment after making a perturbation at position 307 (restricting it to Glu). Since the parent alignment showed 59% Glu at 307, the subalignment now contains $0.59 \times 560 = 330$ sequences.

(C–F) The frequency distribution of amino acids at the indicated position in the full alignment (black bars) and 307E subalignment (white bars). For comparison, the mean frequencies of amino acids in 36,498 proteins in the nonredundant database are shown (gray bars). (C) Position 233 is not conserved or coupled. (D) Position 415 is moderately conserved but is not coupled to 307E. (E) Position 292 is conserved but is not coupled to 307E. (F) Position 426 is both conserved and coupled to 307E.

(G) The complete set of statistical coupling values for the 307E perturbation for all sites *i* ($\Delta\Delta G^{stat}_{i,307\ E}$) in the multiple sequence alignment. For calculation of $\Delta G^{stat}$ and $\Delta\Delta G^{stat}$ values, see Experimental Procedures.

surface, the AF2 helix, and the ligand binding pocket. Since three-dimensional structural and biochemical analyses have not elucidated such networks of long-range interactions, we reasoned that the SCA technique would provide an innovative approach, based on its successful application to other large protein families (Suel et al., 2003). The SCA is based on two simple postulates of molecular evolution. First, the lack of evolutionary constraint at any position should cause the observed amino acid frequencies at that site to approach their mean values in all proteins. As a corollary, the evolutionary constraint (conservation) at any site *j* is the degree to which the observed distribution deviates from these mean frequencies and is measured in an energy-like statistical parameter called $\Delta G^{stat}_j$. Second, the energetic coupling of two positions, regardless of structural location, should drive the coevolution of the two positions. In the SCA, the coevolution of two sites is measured by performing a statistical perturbation experiment on a multiple sequence alignment, where a change is introduced to the frequency of an amino acid at a test site *i*, and the impact of this perturbation for amino acid *x* at another site *j* is measured in another energy-like statistical parameter, $\Delta\Delta G^{stat,x}_{i,j}$ (Lockless and Ranganathan, 1999). Calculated for all sites *j*, this analysis represents a mapping of how the perturbation at *i* is felt by all other sites and is an evolutionary prediction of the global pattern of thermodynamic coupling for test site *i*. For any specific perturbation, the SCA analysis should demonstrate the following classes of outcomes at other sites: (1) unconserved sites, which by definition show little evolutionary constraint, should stay unconserved upon perturbation, (2) some conserved sites that are important but energetically independent are expected to remain similarly conserved, and (3) some conserved sites are expected to undergo significant redistri-

bution of their amino acid distribution, indicating that they are evolutionarily and functionally coupled. We applied the SCA technique using an alignment that included the LBDs of all members of the nuclear receptor superfamily (560 sequences). The primary sequence of human RXR$\alpha$ was chosen arbitrarily to number the amino acids in the alignment. To illustrate the SCA method, consider a perturbation at position 307 in the multiple sequence alignment (Figure 2). Fixing the amino acid identity of test site 307 to only glutamic acid resulted in a subalignment containing 59% (330 sequences) of the full alignment, (Figure 2B). Figures 2A–2F show examples of how the perturbation at 307 affects the evolutionary outcome at four other sites. Position 233 is unconserved ($\Delta G^{stat}_{233} = 0.12kT^*$) since its amino acid distribution (black bars, Figure 2C) is close to the average found in 36,498 proteins in the nonredundant database (gray bars). As expected, position 233 shows little change to its distribution upon making the 307E perturbation (compare black and white bars) and is therefore not coupled ($\Delta\Delta G^{stat}_{233,307\ E} = 0.05kT^*$). Positions 415 and 292 are moderately conserved ($\Delta G^{stat}_{415} = 1.1kT^*$, $\Delta G^{stat}_{292} = 1.88kT^*$) since their amino acid frequencies differ greatly from the mean values. However, these positions also show little change to their distributions upon making the 307E perturbation and are uncoupled ($\Delta\Delta G^{stat}_{415,307\ E} = 0.126kT^*$,$\Delta\Delta G^{stat}_{292,307\ E} = 0.153kT^*$). In contrast, position 426, located at the dimerization interface and a direct packing neighbor to position 307, is not only conserved ($\Delta G^{stat}_{426} = 1.31kT^*$) but also shows a significant redistribution of its amino acid frequencies upon the 307E perturbation. Thus, this position is evolutionarily coupled to 307 ($\Delta G^{stat}_{426,307\ E} = 1.13kT^*$).

The complete set of statistical coupling values for the 307E perturbation for all sites $i$ ($\Delta\Delta G^{stat}_{i,307\ E}$) gives a global mapping of how this perturbation impacts evolution at other sites and represents an evolution-based prediction of the thermodynamic interactions between 307 and these other sites (Figure 2G). While the majority of positions in the LBD show no significant coupling, a sparse subset of positions has high $\Delta\Delta G^{stat}_{i,307\ E}$ values that indicate evolutionary coupling to position 307.

SCA is subject to several constraints, which limit the number of positions that are suitable for the perturbation analysis. Specifically, perturbations at sites must produce subalignments that retain sufficient size and diversity so that coupling observed between sites reflects evolutionary constraint during evolution and not just historical relationships that dominate small and highly similar subalignments. Supplemental Figure S1 (http://www.cell.com/cgi/content/full/116/3/417/DC1) shows an empirical method to determine a subalignment size cutoff for determining the number of allowed positions for perturbation analysis. In the LBDs, we conservatively chose a minimum subalignment cutoff size of 250 sequences (45% of the full alignment), yielding a total of 40 perturbations that can be subjected to SCA (Supplemental Figure S1b on *Cell* website).

### Identification of a Putative Nuclear Receptor Allosteric Network

The total SCA for the LBD family is shown as a matrix of $\Delta\Delta G^{stat}$ values that combines the data for all 40 perturbations. In essence, the matrix is a concatenation of perturbation experiments like those shown in Figure 2G. The matrix is organized with positions (N terminus to C terminus) on the multiple sequence alignment as rows (top to bottom) and perturbation experiments as columns. Thus, each element is the statistical coupling value for one position given perturbation at one test site, and each column represents the overall coupling of one test site over all LBD positions. The matrix is a global representation of interactions between pairs of residues, where the interactions are expressed not as physical distances or thermodynamic couplings, but as the coupled evolution of many pairs of sites.

To look for networks of coevolving residues in the LBD, we used two-dimensional hierarchical clustering (Getz et al., 2000) to identify residues that demonstrate similar patterns of statistical coupling (Figure 3B). This analysis is analogous to that used for DNA microarrays; instead of clustering genes with similar expression profiles in many array experiments, we used clustering to reveal positions in proteins that display similar patterns of coevolution in many perturbation experiments. As shown in Figure 3B, a small group of positions form a primary cluster (marked in red) that shows high amplitude coupling and a similar pattern of coevolution. Isolation of this primary cluster revealed a group of 27 residues that form a self-consistent unit; they show a similar pattern of statistical coupling and they couple to each other (Figure 3C). It is also worth noting that most positions evolved nearly independently and show little to no coupling to any perturbation. These findings are consistent with the idea that intramolecular energetic networks in proteins are dependent on relatively small numbers of positions and that most positions are only weakly coupled to these networks (Suel et al., 2003).

The network includes sites that define the four well-characterized functional surfaces of the LBD: the heterodimerization interface, the cofactor binding surface, the AF2 helix, and the ligand binding pocket (Figures 3D and 3E, Table 1). The human RXR$\alpha$ LBD was chosen as a reference scaffold to visualize the network because RXR is the integral common partner in heterodimeric signaling and has a well-characterized LBD structure (Bourguet et al., 1995; Egea et al., 2000). Indeed, residues selected by the analysis include those that have been structurally observed to interact directly with heterodimerization partners (positions 379, 393, 419, and 426), participate in ligand recognition (position 305), and contact coactivator and corepressor helix motifs (positions 284, 297, 301, 302, and 450) (Gampe et al., 2000; Xu et al., 2002). Other SCA-selected residues do not contribute directly to known functional surfaces and are not obvious from examining LBD structures. These residues are located in the core and form a largely contiguous pathway of physical interactions linking the functional surfaces (Figures 3D and 3E; Supplemental Movie S1 on *Cell* website). For example, 18 of the 27 residues are located within van der Waals interaction radius of the bound 9-*cis* retinoic acid ligand, a bound coactivator peptide, or other network residues (Figure 3D). These positions form a web of physical interactions that link the heterodimerization interface to the ligand binding pocket and the cofactor binding surface. Residues such as E307 have no reported function (i.e., they do not

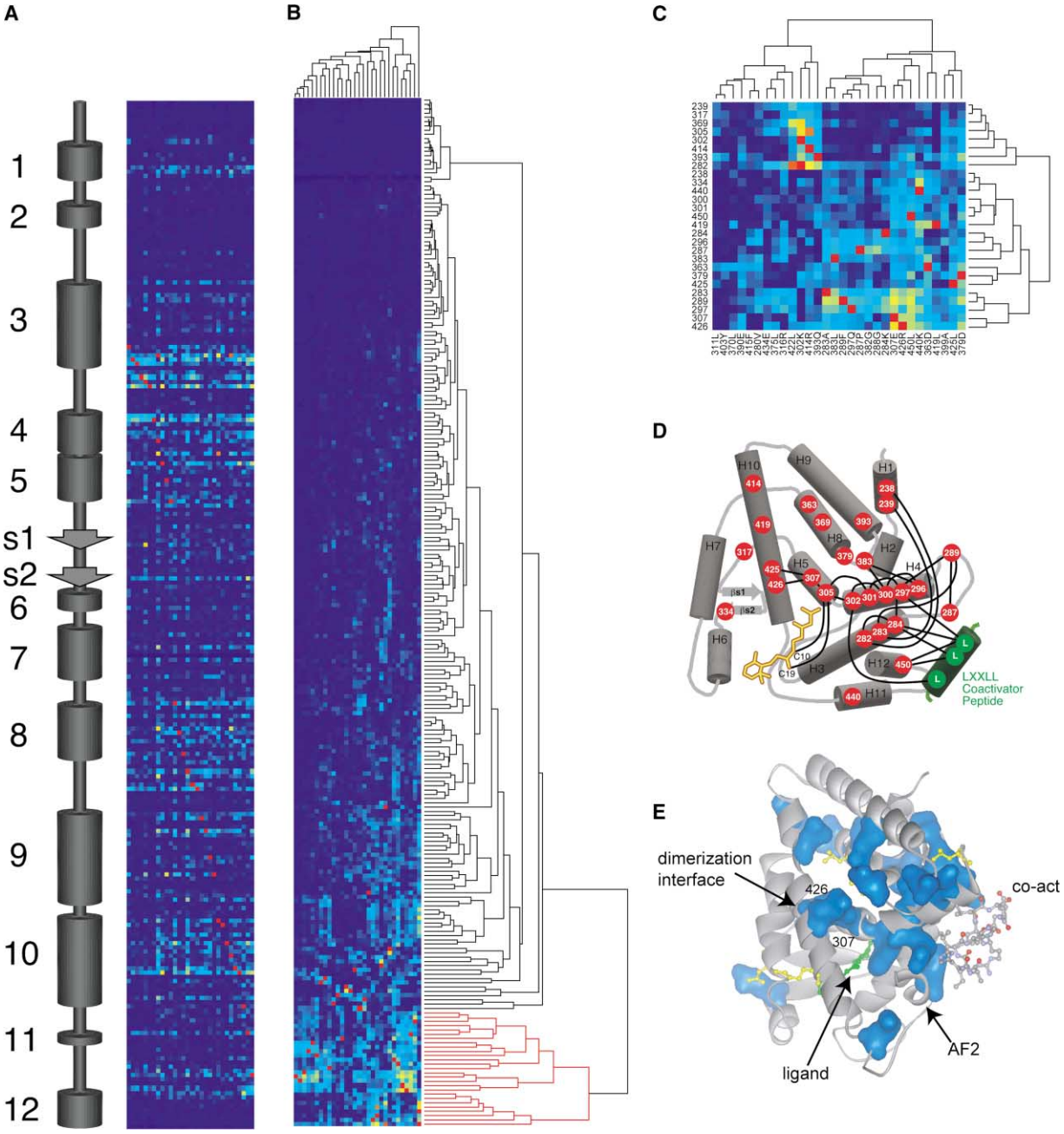


Figure 3. SCA Reveals a Putative Nuclear Hormone Receptor Allosteric Network

(A) Matrix of $\Delta\Delta G^{stat}$ values reporting the coevolution of many pairs of positions in an alignment of 560 members of the LBD family. Rows in the matrix represent LBD positions (N-terminal to C-terminal as shown by the secondary structure schematic to the left of the matrix), and columns represent perturbation experiments as described in the text. Self-coupling is shown as red pixels. The color scale varies linearly from blue (0 *kT**) to red (1.5 *kT**). *kT** is an arbitrary energy unit (Lockless and Ranganathan, 1999).

(B) Two-dimensional hierarchical clustering of the matrix reveals that most positions show little coupling to any perturbation and that a small subset of positions (the cluster marked in red) shows a similar pattern of high $\Delta\Delta G^{stat}$ values.

(C) Extraction and re-clustering of the 27 primary positions (red cluster in [B]) show that these positions form a self-consistent cluster; they show a similar pattern of coupling to perturbations, and are related by perturbations within the group itself (note the clustering of the red pixels). Thus, these positions evolutionarily interact with each other.

(D and E) A schematic diagram and tertiary structure mapping of the 27 primary cluster positions on the RXR$\alpha$ LBD, solved in complex with 9-*cis* retinoic acid ligand, coactivator peptide, and a heterodimeric partner (PPAR$\gamma$, not shown) (Gampe et al., 2000). In (D), a line is drawn between any two cluster positions that are located within physical contact distance on the solved structure. In (E), a van der Waals surface (blue) is drawn around the clustered residues. Coupled residues occur at all four functional surfaces of the LBD (the heterodimerization interface, the cofactor binding surface, the activation function 2 helix [AF2], and the ligand binding pocket) and form a network linking these surfaces through the protein core. See Supplemental Movie S1 on *Cell* website for more details. The figure was prepared using ViewerPro (Accelrys Inc.).

Table 1. Components and Sequence Comparison of the Nuclear Receptor Allosteric Network

| | | | | Permissive[e] | | | Non-Permissive[e] | | |
|---|---|---|---|---|---|---|---|---|---|
| Residue[a] | Location[b] | Domain[c] | Reported[d] | RXRα | LXRα | PPARα | VDR | RARα | TRβ |
| 238 | H1 | CBS core | No | A | A | A | M | A | A |
| 239 | H1 | CBS core | No | E | A | Y | D | H | H |
| 282 | H3 | CBS core | Yes | W | F | F | F | F | F |
| 283 | H3 | CBS core | Yes | A | A | A | A | A | A |
| 284 | H3 | CBS surface | Yes | K | K | K | K | K | K |
| 287 | loop 3-4 | CBS core | No | P | P | P | P | P | P |
| 289 | loop 3-4 | CBS core | Yes | F | F | F | F | F | F |
| 296 | H4 | CBS core | No | D | D | D | D | D | D |
| 297 | H4 | CBS surface | Yes | Q | Q | Q | Q | Q | Q |
| 300 | H4 | CBS core | No | L | L | L | L | L | L |
| 301 | H4 | CBS surface | Yes | L | L | L | L | L | L |
| 302 | H4 | CBS surface | Yes | R | K | K | K | K | K |
| 305 | H5 | LBP | Yes | W | A | V | A | C | C |
| 307 | H5 | core | No | E | E | E | E | D | E |
| 317 | loop 5-βs1 | core | No | S | Y | M | F | Y | Y |
| 334 | βs2 | HD core | No | R | R | R | V | R | R |
| 363 | H8 | surface | No | D | N | D | H | D | D |
| 369 | H8 | core | No | C | L | L | L | L | L |
| 379 | loop 8-9 | HD surface | Yes | D | D | D | D | D | D |
| 383 | loop 8-9 | CBS core | No | L | V | L | V | L | L |
| 393 | H9 | HD surface | Yes | R | Q | Q | Q | Q | Q |
| 414 | H10 | HD core | No | R | M | L | L | M | F |
| 419 | H10 | HD surface | Yes | L | L | L | I | L | L |
| 425 | H10 | HD core | No | L | L | L | L | L | L |
| 426 | H10 | HD surface | Yes | R | R | R | R | R | R |
| 440 | H11 | HD, near AF2 | Yes | K | R | K | S | K | K |
| 450 | H12 | AF2 | Yes | F | L | L | L | L | L |

[a] Human RXRα numbering for the indicated positions in the multiple sequence alignment.
[b] Secondary structural location of the indicated residue in the canonical nuclear receptor LBD structure (Bourguet et al., 1995). H—helix; βs—beta sheet.
[c] Physical relationship of the indicated residue to LBD functional domains (Gampe et al., 2000). CBS—cofactor binding surface; LBP—ligand binding pocket; HD—heterodimerization interface; AF2—activation function 2 helix.
[d] Refers to whether or not the indicated residue has been previously reported to be important in nuclear receptor ligand activation in biochemical or structural studies.
[e] Amino acid identities (single letter code) are shown for the corresponding position in each of the indicated human receptors.

participate in ligand recognition, cofactor binding, or heterodimerization) and would not be expected to be involved in allosteric coupling based on available structural data alone. The SCA, however, predicts that these sites act as energetic connections, linking the functional surfaces to mediate permissivity.

Sequence comparison of the 27 SCA-predicted positions in a representative set of permissive and nonpermissive receptors revealed no clear amino acid determinants that might differentiate the functional subtypes (Table 1). This is consistent with the observation that energetically coupled residues typically demonstrate significant conservation (Figure 2F) and suggests that permissivity is not conferred by a small number of moderately conserved positions, but instead relies on the conformational context of a large number of residues that do not demonstrate high-amplitude statistical coupling.

### Mutation of Network Residues Selectively Disrupts Ligand Permissivity

The function of the predicted energetic network was tested by transiently expressing site-directed RXR and LXR mutants with an ADH-LXREx2-luc reporter in *Drosophila* SL2 cells. Because the readout from this assay is transcriptional activation, it represents a combined quantitative assessment of the heterodimerization, DNA binding, cofactor recruitment, and transactivation potential of the receptors and is the best single parameter for testing the overall function of the network. SL2 cells lack endogenous RXR and LXR and therefore provide a null background for the investigation of mutant receptors. RXR transfection alone showed a minimal but reproducible induction (2.2 ± 0.6 fold) by the synthetic RXR agonist LG268 (Figure 4A). Likewise, transfection with LXR alone resulted in a moderate induction by the synthetic LXR agonist T1317 (3.0 ± 0.6 fold), but no response to rexinoid was detected without RXR cotransfection (Figure 4A). This basal LXR activity in the absence of RXR was likely due to heterodimerization with USP, the *Drosophila* RXR homolog, which is not activated by rexinoids. Cotransfection of RXR and LXR confirmed that the RXR/LXR heterodimer is permissive for activation by both rexinoid and LXR agonist (Figure 4A).

In contrast, cotransfection of RXR with an LXR mutated at the SCA-predicted position 296 (E296A, which corresponds to RXR E307) resulted in a heterodimer that has now lost its ability to respond to LG268 (2.1 ± 0.5 fold, a value similar to the basal, Figure 4A). The RXR/LXR E296A mutant still responded like the wild-type heterodimer to LXR agonist and showed a substantially increased response to combined treatment with

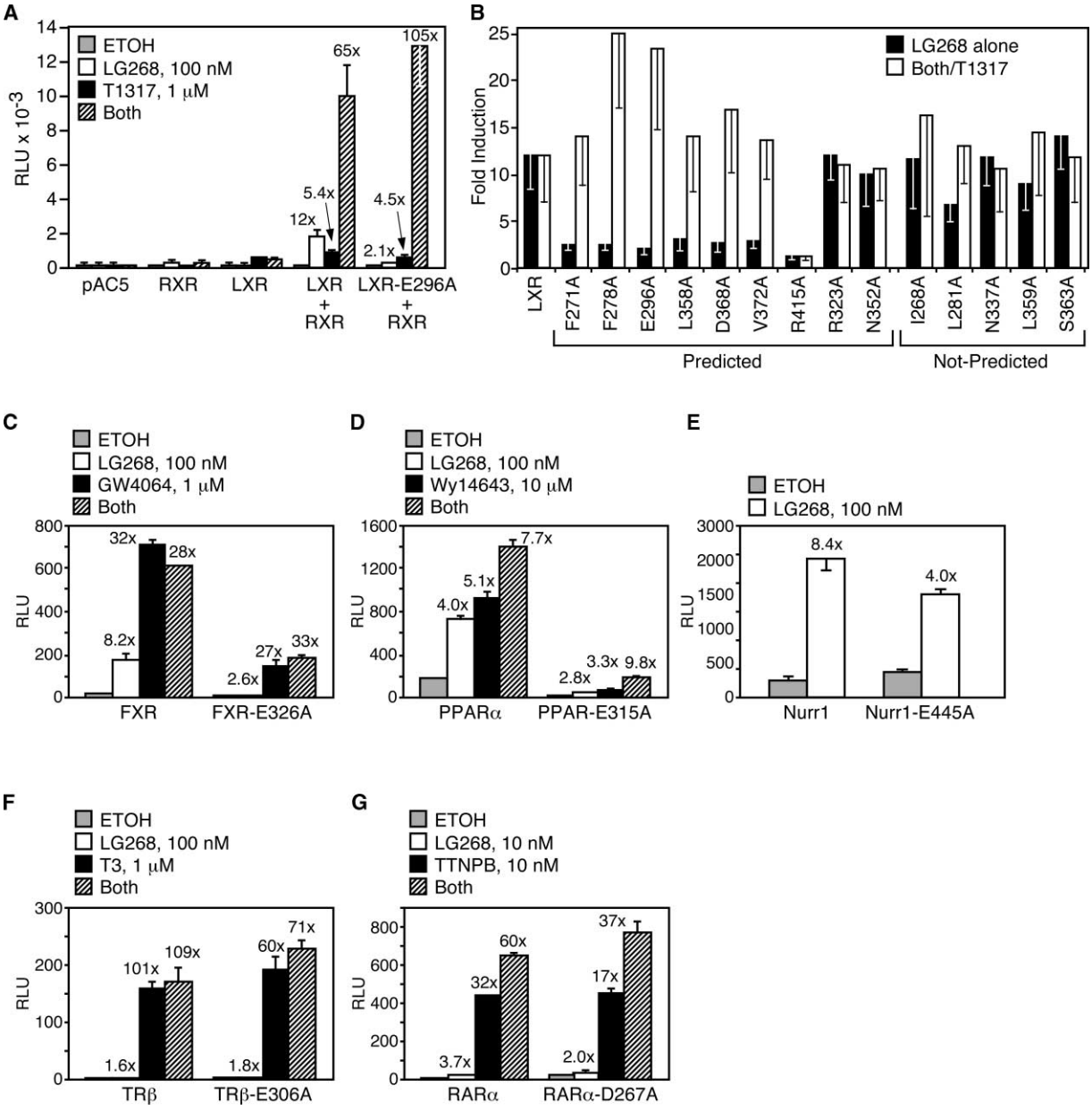


Figure 4. Mutation of Energetic Network Residues Selectively Disrupts Ligand Permissivity

(A) hLXRα and E296A activation of an ADH-LXREx2-luc reporter stimulated with LG268 (100 nM) and T0901317 (1 μM) in SL2 cells.

(B) Permissivity of SCA predicted and not predicted LXR mutants (data calculated from Table 2). Permissive mutants respond to LG268 alone (black bars).

Conditionally permissive mutants respond to LG268 only in the background of T1317 treatment (ratio of both ligands/T1317; white bars). Transfection of permissive (C–E), nonpermissive (F), and conditional (G) receptors and point mutants in HEK 293T cells. Cells were stimulated with the RXR agonist LG268 at the indicated concentration, a specific agonist for the partner receptor, or both ligands. Fold induction values are shown for (C)–(G).

(C) hFXR and hRXRα activation of a tk-FXREx4-luc reporter stimulated by GW4064 (1 μM).

(D) mPPARα activation of a tk-PPREx3-luc reporter stimulated by WY14643 (10 μM) (Schmidt et al., 1992).

(E) Gal4-mNurr1 and hRXRα activation of a tk-MH100x4-luc reporter stimulated by LG268.

(F) hTRβ activation of a tk-TREx2-luc reporter stimulated by $T_3$ (1 μM).

(G) hRARα activation of the tk-TREx2-luc reporter stimulated by TTNPB (10 nM).

both receptors' ligands (Figure 4A). Consistent with a multiplicative effect of combined ligand treatment, wild-type RXR/LXR responded to LG268 in the presence and absence of T1317 with equivalent efficacy (Figure 4B). While the LXR E296A mutant is unresponsive to LG268 alone, in the background of T1317 treatment, LG268 induction exceeded that of the wild-type receptor (23.5 ± 8.8 fold, Figure 4B). This finding implies that the E296A mutation effectively converted RXR/LXR from a permissive to a conditional heterodimer. Importantly,

the ability of the LXR mutant to respond equally well to both T1317 and LG268 in combination with T1317 indicates that the mutation specifically impaired activation by LG268 alone without globally disrupting protein stability or heterodimer function. It is of further significance to note that three-dimensional structural analysis does not suggest that E296 is more important than other nearby residues in mediating RXR/LXR function. Crystal structures of LXR$\alpha$ and LXR$\beta$ show that E296 is one of the residues that resides within the core of the LBD between the ligand binding pocket and the heterodimerization interface (Svensson et al., 2003; Williams et al., 2003). Residues in this vicinity have the potential to transmit energetic information between these two functional surfaces.

To determine if mutation of LXR position 296 selectively affects permissive RXR heterodimers, we mutated the analogous position in a representative set of permissive and nonpermissive RXR heterodimers. Mutation of this site in the permissive heterodimers farnesoid X receptor (FXR), peroxisome proliferator-activated receptor $\alpha$ (PPAR$\alpha$), and nuclear-related receptor 1 (Nurr1) resulted in decreased LG268 response when applied independently, indicating loss of permissivity (Figures 4C–4E). Specifically, the mutation rendered FXR nonpermissive (no gain of LG268 response in the background of FXR agonist, Figure 4C) and PPAR$\alpha$ conditionally permissive (Figure 4D). Although basal transcriptional activity of the FXR E326A and PPAR$\alpha$ E315A mutants was diminished, in both cases ligand induction by permissive partner agonists and by both ligands remained intact. Consistent with a variable effect of the mutation on basal transcription, the Nurr1 mutant demonstrated increased constitutive activity and decreased induction by LG268 (Figure 4E). Further analysis of ligand induction by mutants analogous to LXR E296A in the nonpermissive RXR/thyroid hormone receptor $\beta$ (TR$\beta$) heterodimer (Figure 4F) and the conditional RXR/RAR$\alpha$ heterodimer (Figure 4G) showed no significant differences from wild-type. For both TR$\beta$ E306A and RAR$\alpha$ D267A, basal activity was increased, resulting in a slight decrease in RXR partner ligand fold induction despite transcriptional activity levels that exceeded those of the wild-type heterodimer. These data demonstrate that position 296 is not generally required for ligand response and that mutation of this site does not globally destabilize the protein. Instead, position 296 appears to act selectively to mediate permissivity.

### The SCA-Predicted Network Mediates Permissivity

To further test the hypothesis that the SCA-predicted network mediates allostery in permissive heterodimers, we mutated a number of LXR residues both predicted and not predicted by the SCA and assayed their ligand response in the RXR/LXR heterodimer. Sites not predicted by the SCA were selected by the criteria that they occur as direct packing neighbors to predicted residues and are similarly conserved (Figure 2 and Supplemental Movie S1 on *Cell* website). Thus, mutation of these sites represents a stringent test of the hypothesis that the predicted network selectively mediates allosteric communication despite packing with chemically similar residues. Remarkably, seven of nine predicted LXR mutants tested showed loss of permissivity (F271A, F278A, E296A, L358A, D368A, V372A, and R415A, Figure 4B and Table 2). Six of the mutations, including E296A (discussed above), convert the permissive RXR/LXR heterodimer to a conditional heterodimer. Mutation at one site that makes a direct contact at the dimerization interface (R415A [RXR 426]) showed a complete loss of LXR function, recapitulating the phenotype of a truncation mutant lacking the AF2 helix (Figure 4B and Table 2). The two predicted mutants that fail to show a phenotype (R323A [RXR 334] and N352A [RXR 363]) are included among the seven predicted residues that are physically isolated from the rest of the network (Figure 3D, see Discussion below). Despite spatial proximity to network residues and to known functional sites, none of the mutations at the five not predicted sites showed ligand responses that differ significantly from wild-type (I268A, L281A, N337A, L359A, and S363A, Figure 4B and Table 2). These results demonstrate the improved predictive power of SCA when compared to structural analysis or random alanine scanning mutagenesis in identifying residues involved in allosteric communication.

Analysis of allosteric network positions in RXR affords an opportunity to compare the contributions of RXR and LXR to RXR/LXR heterodimer permissivity. Interestingly, mutation of allosteric network positions in RXR revealed little impact on LG268 responses when tested in heterodimeric complex with LXR (Table 2). Of the eleven SCA-predicted RXR mutants tested, three mutants were nonpermissive and the remainder showed a wild-type phenotype. While W305A and R426A were loss-of-function mutants that resembled the RXR AF2 truncation mutant, F289A exhibited a conditionally permissive phenotype. One of the not predicted RXR mutations caused a loss of function (D347A) and the remaining five mutants retained wild-type ligand response (Table 2). It is of interest to note, however, that two of these mutants (L279A and L370A), along with the predicted mutants D379A and R426A, did demonstrate a loss of rexinoid-induced transactivation when tested in the context of an RXR homodimer (data not shown). The finding that mutation of the allosteric network in the permissive partner had a greater effect on RXR agonist response than mutation of the same residues in RXR suggests that permissive partners are functionally dominant in RXR heterodimers.

### The Allosteric Network Contributes Specificity to Ligand Response

The mechanism by which nuclear receptors respond differentially to structurally distinct agonists has emerged as a question of great physiological and pharmacological interest. Nonpermissive RXR heterodimers primarily function as receptors for endocrine ligands while permissive receptors serve as sensors for lipid-derived products of intermediary metabolism (Chawla et al., 2001). This observation suggests an interesting hypothesis: the allosteric network predicted by the SCA may be selectively required for mediating metabolic signaling and may be silent in mediating responses to endocrine ligands. The recent finding that the secondary bile acid lithocholic acid (LCA) is an endogenous ligand for VDR makes it possible to test this hypothesis (Makishima et al., 2002). If the allosteric network selects be-

Table 2. Ligand Activation of LXRα and RXRα Point Mutants

| | Mutant | LG268 | T1317 | LG268 + T317 |
|---|---|---|---|---|
| | No Receptor | 1.02 ± 0.23 | 0.99 ± 0.21 | 0.84 ± 0.20 |
| | LXR | 1.10 ± 0.17 | 2.95 ± 0.57 | 2.78 ± 0.57 |
| | RXR | 2.15 ± 0.59 | 0.96 ± 0.19 | 2.20 ± 0.68 |
| | RXR + LXR | 12.10 ± 3.77 | 5.37 ± 1.66 | 64.98 ± 18.58 |
| LXR Mutants | Predicted | | | |
| | F271A (282) | 2.69 ± 0.67 | 8.47 ± 1.92 | 119.79 ± 35.81 |
| | F278A (289) | 2.56 ± 0.61 | 2.47 ± 0.53 | 61.94 ± 15.09 |
| | E296A (307) | 2.14 ± 0.48 | 4.46 ± 1.23 | 105.01 ± 26.26 |
| | R323A (334) | 12.24 ± 2.70 | 6.82 ± 1.36 | 76.65 ± 23.28 |
| | N352A (363) | 10.14 ± 3.36 | 7.46 ± 1.53 | 79.39 ± 18.87 |
| | L358A (369) | 3.15 ± 1.03 | 4.27 ± 1.30 | 60.22 ± 17.07 |
| | D368A (379) | 2.82 ± 0.90 | 5.87 ± 1.95 | 100.47 ± 24.20 |
| | V372A (383) | 3.14 ± 0.98 | 7.25 ± 1.76 | 100.09 ± 22.54 |
| | R415A (426) | 1.40 ± 0.46 | 1.01 ± 0.19 | 1.48 ± 0.35 |
| | Not Predicted | | | |
| | I268A (279) | 11.80 ± 5.34 | 5.35 ± 2.59 | 87.78 ± 39.79 |
| | L281A (292) | 6.85 ± 1.71 | 7.30 ± 1.43 | 96.44 ± 22.33 |
| | N337A (347) | 12.06 ± 3.13 | 6.85 ± 2.09 | 73.56 ± 23.38 |
| | L359A (370) | 9.21 ± 2.85 | 7.59 ± 2.44 | 110.76 ± 44.27 |
| | S363A (374) | 14.13 ± 3.63 | 5.19 ± 1.65 | 61.89 ± 15.07 |
| | ΔAF2 truncation | 1.54 ± 0.50 | 1.13 ± 0.37 | 1.83 ± 0.73 |
| RXR Mutants | Predicted | | | |
| | E239A | 10.54 ± 3.38 | 4.95 ± 1.50 | 57.01 ± 15.48 |
| | W282A | 10.60 ± 2.81 | 4.30 ± 1.23 | 59.13 ± 20.33 |
| | F289A | 3.88 ± 1.53 | 6.25 ± 1.58 | 40.30 ± 11.91 |
| | W305A | 1.23 ± 0.34 | 7.08 ± 2.03 | 8.09 ± 2.76 |
| | E307A | 7.88 ± 1.39 | 6.82 ± 1.65 | 47.60 ± 13.22 |
| | R334A | 14.07 ± 2.73 | 5.12 ± 1.40 | 65.34 ± 10.62 |
| | D363A | 10.52 ± 2.49 | 4.64 ± 0.98 | 48.43 ± 11.29 |
| | C369A | 15.71 ± 4.72 | 6.51 ± 1.95 | 75.39 ± 23.34 |
| | D379A | 7.18 ± 2.69 | 5.02 ± 2.00 | 47.86 ± 19.40 |
| | L383A | 17.15 ± 14.95 | 4.04 ± 3.52 | 57.38 ± 50.54 |
| | R426A | 1.13 ± 0.27 | 4.20 ± 1.00 | 4.09 ± 1.11 |
| | Not Predicted | | | |
| | L279A | 13.53 ± 4.19 | 5.24 ± 1.52 | 52.54 ± 14.06 |
| | L292A | 14.79 ± 3.65 | 6.79 ± 1.91 | 71.37 ± 19.93 |
| | D347A | 1.11 ± 0.34 | 3.78 ± 1.14 | 3.57 ± 1.10 |
| | L370A | 9.74 ± 2.85 | 5.25 ± 1.33 | 60.78 ± 13.76 |
| | V374A | 7.44 ± 2.99 | 5.57 ± 1.77 | 50.48 ± 16.21 |
| | K431A | 18.22 ± 13.35 | 6.08 ± 4.30 | 79.92 ± 61.29 |
| | ΔAF2 truncation | 1.47 ± 0.62 | 4.09 ± 1.64 | 5.63 ± 2.35 |

Values indicate fold ligand induction by RXR agonist LG268 (100 nM), LXR agonist T1317 (1 μM), and both ligands (LG268 + T1317) for LXRα and RXRα receptors mutated at positions either predicted or not predicted by the SCA. Corresponding RXR position numbers for the LXR mutants are labeled in parentheses. Data represent the combination of triplicate values from at least three independent experiments. Root mean square errors are shown for each fold induction. Transfections were performed as in Figure 4A. ΔAF2 truncation—activation function 2 helix truncation mutant.

tween a receptor's response to endocrine and metabolic agonists, then mutation of the VDR network should only affect the response to LCA. Indeed, mutations in the predicted residues (F244A, E269A, V346A, and R391A, which correspond to RXR positions W282, E307, L383, and R426, respectively) responded to the classical endocrine ligand, $1,25(OH)_2$-vitamin $D_3$, similar to wild-type VDR (Figure 5A), but were completely insensitive to LCA (Figure 5B). The not predicted VDR mutant S208A responded to both vitamin D and LCA, demonstrating the positional specificity of mutations in discriminating endocrine and metabolic agonists.

Analysis of FXR activation by the endogenous bile acid chenodeoxycholic acid (CDCA), the synthetic agonist GW4064 (Maloney et al., 2000), and the newly developed and chemically distinct agonist fexaramine revealed that although the agonists are all efficacious activators of FXR in vitro, each ligand directs unique programs of gene expression in microarray experiments (Downes et al., 2003). To further test the idea that residues in the allosteric network may mediate differential responses to multiple ligands, we assayed the impact of network mutations in FXR on the response to CDCA, GW4064, and fexaramine. While all these agonists induced robust activation of wild-type FXR, only GW4064 was able to significantly activate the FXR allosteric network mutants F301A (RXR W282) and E326A (RXR E307, Figure 5C). The E326A and I398A (RXR L383) mutants also showed an increased response to GW4064. In contrast, I398A had a wild-type response to CDCA and fexaramine and R441A (RXR R426) demonstrated wild-type responses to all three ligands. These data demonstrate that mutation of network positions results in ligand activation phenotypes ranging from loss of response to a

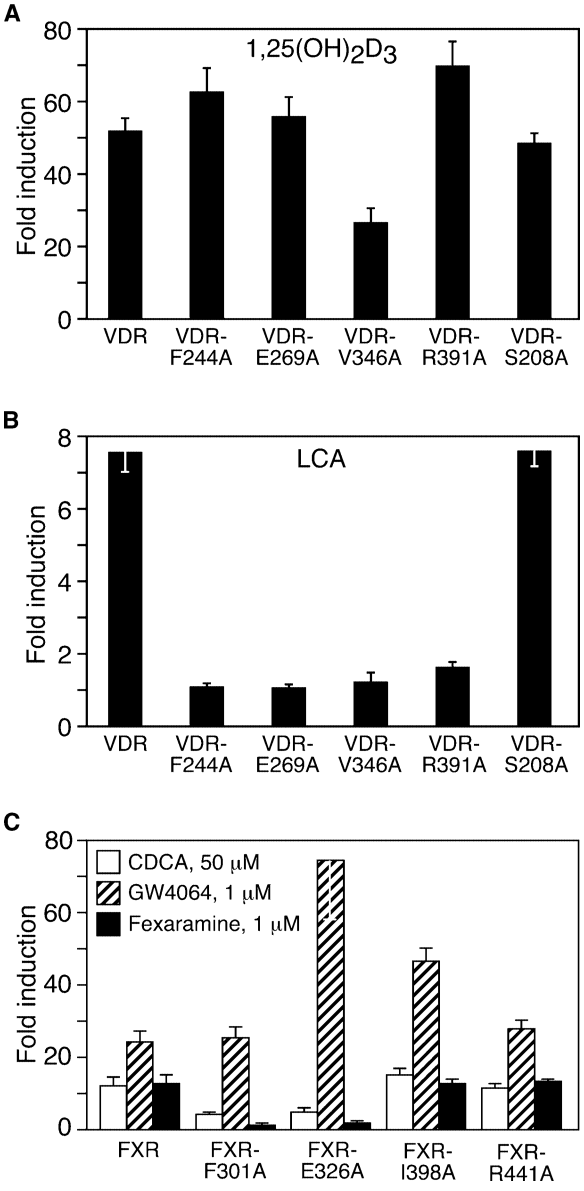


Figure 5. The Nuclear Receptor Allosteric Network Imparts Specificity to Ligand Response

(A and B) Activation of the tk-hCyp3A4 ER6x3-luc reporter by cotransfection of CMX-hRXR$\alpha$ and CMX-hVDR or hVDR mutants in HEK 293T cells (Makishima et al., 2002). Cells were stimulated with 1,25$(OH)_2$-vitamin $D_3$ (100 nM) or lithocholic acid (LCA, 30 $\mu$M). (A) Fold induction by 1,25$(OH)_2$-vitamin $D_3$. (B) Fold induction by LCA. (C) Activation of the tk-FXREx4-luc reporter by cotransfection of CMX-hRXR$\alpha$ and CMX-hFXR or hFXR mutants in HEK 293T cells. Cells were stimulated with chenodeoxycholic acid (CDCA, 50 $\mu$M), GW4064 (1$\mu$M), or fexaramine (1 $\mu$M) (Downes et al., 2003).

particular agonist to selective increases in ligand sensitivity. The location of the mutated positions outside of the ligand binding pocket in both VDR and FXR emphasizes that the effects on ligand response come from perturbation of a distributed energetic network rather than from direct alteration of ligand binding. Taken together, the finding that network mutations differentially impact endocrine and dietary signaling in VDR and selectively discriminate between FXR agonists suggests that the allosteric network identified by the SCA underlies the ligand specificity of nuclear receptor function.

## Discussion

### Recognition of Allostery in Nuclear Receptors

Allosteric communication by nuclear receptor heterodimers allows multiple ligand-mediated pathways to be integrated into a transcriptional response. The "phantom ligand effect," the ability of ligand-induced allosteric signal transmission by permissive nuclear receptors to activate the LBD of unliganded heterodimeric partners, is a particularly dramatic example of functional engagement of the allosteric network and is a biologically critical, but poorly understood, aspect of nuclear receptor action (Schulman et al., 1997). While biochemical experiments have identified interaction partners and structural studies have delineated the functional surfaces that mediate these interactions, the characteristics of the energetic architecture that connects these physically distant surfaces and enables concerted allosteric action by nuclear receptors remain unknown. Since crystal structures provide a limited view of protein energetics, and traditional structure-function approaches, such as random mutagenesis, cannot feasibly identify functional coupling of residues in a complex protein, the task of identifying the residues that comprise the energetic architecture in the nuclear receptor LBD has remained a challenging problem. We recently developed the SCA (statistical coupling analysis), a computational method to detect functionally significant coevolution of pairs of residues in a multiple sequence alignment (Lockless and Ranganathan, 1999). The SCA method differs fundamentally from other alignment-based methods, such as Evolutionary Trace, which reveal conservation at individual sites and have no predictive value for coevolving sites (Lichtarge et al., 1996). This method has been validated by the demonstration that SCA can identify residues that have been experimentally implicated in the allosteric mechanism of several well-studied protein families (Suel et al., 2003). In this work, we employed SCA to identify a small ($\sim$12% of total) but contiguous network of 27 evolutionarily coupled amino acids that define and link the functional surfaces of the LBD. Half of these residues are known to participate directly in LBD interactions, an important validation of the SCA method. The remaining predicted residues, which have no known function in RXR heterodimer activation, are candidate molecular bridges in the allosteric mechanism. It is important to note that these residues cannot be predicted through analysis of three-dimensional structural data or sequence conservation alone. Mutation of these residues in permissive heterodimeric partners (e.g., LXR E296A) abrogated response to RXR agonist in the absence of partner ligand. The finding that these mutated network residues still retain full responsiveness to the partner ligand and to costimulation by RXR agonist plus partner ligand argues that such mutations specifically impair the allosteric communication required for permissivity without globally disrupting receptor folding or stability. The finding that conditional and nonpermissive receptors, which do not display rexi-

noid allostery, are unaffected by mutation confirms that the allosteric network is a structural determinant of permissivity. In several cases, mutation of evolutionarily coupled residues increased or decreased RXR heterodimer basal activity (Figure 4). Consistent with this finding, we detected mutants with altered DNA binding and solution heterodimerization affinities in in vitro assays (data not shown). The ability of agonists to induce mutant heterodimer activity with efficacy comparable to the wild-type proteins indicates that basal transcriptional activity is separable from ligand-dependent response.

Although the majority of residues exist in a physically connected network, several explanations could account for the finding that some predicted residues are physically isolated in the RXR LBD structure (Figure 3D). The isolated residues might mediate direct interaction with additional associated factors (e.g., AF2 position 450 is isolated in the absence of cocrystallized coactivator peptide); these residues might contact allosteric network positions in other nuclear receptor family members; or SCA-selected, isolated residues might represent "false positives" (as appears to be the case for LXR 323 and 352, Figure 4B).

The statistical mapping used in this study to identify a putative nuclear receptor energetic architecture cannot in itself provide a physical mechanism of interaction between residues. However, the finding of physical connectivity between residues in the predicted network suggests that mechanical coupling could efficiently propagate local perturbations in ligand binding, heterodimerization, and cofactor recruitment between distant functional surfaces. The experimental data presented support this hypothesis and are consistent with studies that have characterized downstream consequences of ligand-induced engagement of the allosteric network. The demonstration that RXR can bind ligand and recruit coactivators when complexed with apo-RAR through the application of biochemical techniques in combination with mass spectrometry demonstrates that RXR is competent to mediate allostery, even in the context of a conditionally permissive heterodimer (Germain et al., 2002). That differential cofactor recruitment by permissive and nonpermissive heterodimers might read the receptor conformation specified by the allosteric network into a transcriptional output is a logical prediction of our model.

### Allosteric Control and the Evolution of Signal Discrimination

RXR heterodimeric partners include adopted orphan receptors that recognize dietary lipids at low affinity and endocrine receptors that mediate high-affinity responses to hormonal lipids (Chawla et al., 2001). RXR heterodimers that sense dietary lipids and activate feedforward metabolic cascades to maintain the homeostatic levels of potentially harmful intermediates (e.g., LXRs and PPARs) are permissive to RXR action. By contrast, the endocrine receptors (e.g., VDR and TRs) are nonpermissive RXR partners that govern crucial metabolic and developmental events and mediate negative-feedback control of hormone synthesis at distant sites. One attractive hypothesis that is consistent with our data is that the nonpermissive RXR partners TR and VDR functionally diverged from permissive receptors as they acquired the ability to recognize high-affinity hormonal ligands. RAR, the conditionally permissive RXR partner, may represent an evolutionary intermediate; it recognizes dietary-derived lipids (i.e., vitamin A) similar to permissive receptors, but regulates morphogenesis and development in the mode of endocrine receptors.

The finding that RXR function is relatively unaffected by mutation of its allosteric network is consistent with its subordinate role as the common partner of both permissive and nonpermissive receptors (Table 2). This differential sensitivity of RXR and its heterodimeric partners to mutation illustrates the conditional importance and context dependence of network positions for LBD function. In this way, we hypothesize that ligand binding specificity and allosteric communication coevolved to allow nuclear receptors to differentially couple rexinoid signaling with the regulation of intermediary metabolism and endocrine physiology.

The identification of LCA as an endogenous, lipid-derived VDR agonist (Makishima et al., 2002) provides a unique opportunity to test the role of the allosteric network in a receptor's response to endocrine and dietary signaling. The demonstration that VDR allosteric network mutants respond to vitamin D (Figure 5A) but are unresponsive to LCA (Figure 5B) confirms that the bound ligand is an integral part of the allosteric network. Additionally, this result suggests that LCA might represent a "vestigial" lipid-derived VDR ligand that retains a requirement for RXR allosteric contribution to the stabilization of an active conformation. In this way, the ability of the allosteric network to be differentially engaged by physiologically distinct ligands may have evolved to exploit the potential for regulatory complexity afforded by nuclear receptor heterodimers.

### Allostery and Improved Pharmacological Specificity

The development of synthetic ligands that induce therapeutically desirable target genes in pathologically compromised tissues without activating programs of gene expression that underlie potentially debilitating adverse reactions is a necessary step in improving nuclear receptor-targeted drugs. Synthetic agonists that bind to the same receptor can direct dramatically different programs of gene expression. Most striking is the recent observation that synthetic FXR agonists, which bind to the receptor with comparable affinity and activate known target genes, differentially induce and repress a large number of genes in microarray experiments (Downes et al., 2003). We observed that FXR mutants have a range of phenotypes including specific loss and augmentation of response to FXR agonists, demonstrating that network mutants have variable effects on allosteric control. The finding that FXR R441A (RXR R426) responds effectively to the three ligands tested is consistent with the ability of VDR R391A (RXR R426) to respond to vitamin D. These data suggest that the biologically distinct activities of the FXR agonists may be due in part to differential engagement of the heterodimeric allosteric network. The physical mechanism of signal transmission in FXR and the possibility that the molecular context of FXR target gene promoters might encode this specificity awaits further investigation.

## Perspectives

The use of SCA to identify a coupled network of amino acids that are shown to be involved in allosteric ligand regulation of RXR heterodimers raises the question of the importance of this network in other members of the family, such as the steroid hormone and orphan receptors. The possibility that engagement of the allosteric network in these receptors might depend on a particular receptor's mechanism of activation or on posttranslational modification (e.g., phosphorylation) warrants future experimentation. Another promising outcome of this work may be its application to receptor pharmacology. For example, the demonstration that alternate ligands induce state-dependent changes in a receptor's allosteric network could aid the discovery of synthetic agonists with improved pharmacological specificity. An interesting future test of this concept might be to screen receptors containing various network mutations for unique pharmacophores that would give selective transcriptional responses.

### Experimental Procedures

#### Sequence Alignments

Sequences for the nuclear hormone receptor alignment were collected from the nonredundant database using 50 rounds of PSI-BLAST (Altschul et al., 1997) against human RXR$\alpha$. Significant matches (e-score $< 0.001$) were initially aligned with ClustalW (Thompson et al., 1994) and manually adjusted with standard structure-based alignment techniques (Doolittle et al., 1996).

#### Statistical Coupling Analysis (SCA)

The statistical coupling matrix is assembled from the 40 site-specific perturbations that satisfy the criteria of the analysis as described in Supplemental Figure S1. Calculation of statistical coupling was carried out as previously described (Lockless and Ranganathan, 1999; Suel et al., 2003). Briefly, each position $j$ in the multiple sequence alignment is described as a 20 element vector of individual amino acid frequencies (e.g., $\overrightarrow{f_j} = [f_j^{ala}, f_j^{cys}, f_j^{asp}, \ldots, f_j^{tyr}]$). The frequency vector is then converted to a vector of statistical energies ($\overrightarrow{\Delta G_j^{stat}} = [\Delta G_j^{ala}, \Delta G_j^{cys}, \Delta G_j^{asp}, \ldots, \Delta G_j^{tyr}]$), where each term is the value for amino acid $x$ at site $j$ and is given by $\Delta G_j^x = kT^* \ln P_j^x$. $P_j^x$ is the binomial probability of observing amino acid $x$ at site $j$ given its mean frequency in all natural proteins. To measure coupling between a perturbation at $i$ and any site $j$, we calculate the difference energy vector, $\overrightarrow{\Delta G_{j,i}^{stat}} = \overrightarrow{\Delta G_j^{stat}} - \overrightarrow{\Delta G_{j|i}^{stat}}$, where $\overrightarrow{\Delta G_j^{stat}}$ is the statistical energy vector of site $j$ in the parent alignment and $\overrightarrow{\Delta G_{j|i}^{stat}}$ is that of site $j$ in the subalignment derived from the perturbation at $i$. The scalar coupling energy ($\Delta\Delta G_{j,i}^{stat}$) is the magnitude of this difference vector and reports the combined effect of perturbation at $i$ on all amino acids at position $j$. If sites $i$ and $j$ are evolutionarily independent, the coupling energy is zero and is consistent with lack of interaction, but if the coupling energy is non-zero, the two sites interact to the extent measured by $\Delta\Delta G_{j,i}^{stat}$. Calculation of $\Delta\Delta G_{j,i}^{stat}$ for all sites $j$ given a perturbation at $i$ is a mapping of how all sites in the protein are affected by perturbing $i$ (Figure 2G). Code implementing this algorithm and sample datasets are available upon request. Hierarchical clustering was performed using MATLAB 6.5 (The Mathworks, Inc.) using the cityblock distance metric and complete linkage. Details are provided on the authors' website (http://www.hhmi.swmed.edu/Labs/rr).

#### Plasmids and Site-Directed Mutagenesis

Site-directed mutants of pAC5.1-hRXR$\alpha$, pAC5.1-hLXR$\alpha$, CMX-hRAR$\alpha$, CMX-hFXR, CMX-mPPAR$\alpha$, CMX-gal4-mNurr1, CMX-hTR$\beta$, and CMX-hVDR were generated using the Stratagene QuikChange Site-Directed Mutagenesis kit and were verified by DNA sequencing.

#### Cell Culture and Cotransfection Assays

HEK293 cells were maintained at 37°C, 5% $CO_2$ in DMEM containing 10% fetal bovine serum (FBS). Transfections were performed in 96-well plates in media containing 10% dextran-charcoal-stripped FBS by calcium phosphate coprecipitation as previously described (Lu et al., 2000; Makishima et al., 1999). *Drosophila* SL2 cells were maintained in Schneider *Drosophila* Medium (Gibco) containing 5% dextran-charcoal-stripped FBS and were transfected in 96-well plates by calcium phosphate coprecipitation as previously described (Harmon et al., 1995; Janowski et al., 1996).

#### Acknowledgments

We thank Drs. M. Brown, J. Goldstein, A. Gilman, and S. McKnight for comments on the manuscript; Dr. M. Makishima and members of the Mango and Ranganathan labs for helpful suggestions. This work was funded by the Howard Hughes Medical Institute (D.J.M. and R.R.), the Robert A. Welch Foundation (D.J.M. and R.R.), the Mallinckrodt Foundation (R.R.), a National Institutes of Health Pharmacological Sciences Training Grant (A.I.S.), and a Medical Scientist Training Program Grant (A.I.S. and C.L.). D.J.M. and R.R. are investigators of the Howard Hughes Medical Institute.

Received: October 17, 2003
Revised: December 16, 2003
Accepted: December 23, 2003
Published: February 5, 2004

### References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Bourguet, W., Ruff, M., Chambon, P., Gronemeyer, H., and Moras, D. (1995). Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-alpha. Nature *375*, 377–382.

Carter, P.J., Winter, G., Wilkinson, A.J., and Fersht, A.R. (1984). The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (Bacillus stearothermophilus). Cell *38*, 835–840.

Chawla, A., Repa, J.J., Evans, R.M., and Mangelsdorf, D.J. (2001). Nuclear receptors and lipid physiology: Opening the X-files. Science *294*, 1866–1870.

Cheskis, B., and Freedman, L.P. (1996). Modulation of nuclear receptor interactions by ligands: kinetic analysis using surface plasmon resonance. Biochemistry *35*, 3309–3318.

Doolittle, R., Abelson, J.N., and Simon, M.I. (1996). Computer Methods for Macromolecular Sequence Analysis (San Diego, CA: Academic Press).

Downes, M., Verdecia, M.A., Roecker, A.J., Hughes, R., Hogenesch, J.B., Kast-Woelbern, H.R., Bowman, M.E., Ferrer, J.L., Anisfeld, A.M., Edwards, P.A., et al. (2003). A chemical, genetic, and structural analysis of the nuclear bile acid receptor FXR. Mol. Cell *11*, 1079–1092.

Egea, P.F., Mitschler, A., Rochel, N., Ruff, M., Chambon, P., and Moras, D. (2000). Crystal structure of the human RXRalpha ligand-binding domain bound to its natural ligand: 9-cis retinoic acid. EMBO J. *19*, 2592–2601.

Gampe, R.T.J., Montana, V.G., Lambert, M.H., Miller, A.B., Bledsoe, R.K., Milburn, M.V., Kliewer, S.A., Willson, T.M., and Xu, H.E. (2000). Asymmetry in the PPARgamma/RXRalpha crystal structure reveals the molecular basis of heterodimerization among nuclear receptors. Mol. Cell *5*, 545–555.

Germain, P., Iyer, J., Zechel, C., and Gronemeyer, H. (2002). Co-regulator recruitment and the mechanism of retinoic acid receptor synergy. Nature *415*, 187–192.

Getz, G., Levine, E., and Domany, E. (2000). Coupled two-way clustering analysis of gene microarray data. Proc. Natl. Acad. Sci. USA *97*, 12079–12084.

Harmon, M.A., Boehm, M.F., Heyman, R.A., and Mangelsdorf, D.J. (1995). Activation of mammalian retinoid X receptors by the insect growth regulator methoprene. Proc. Natl. Acad. Sci. USA *92*, 6157–6160.

Janowski, B.A., Willy, P.J., Devi, T.R., Falck, J.R., and Mangelsdorf, D.J. (1996). An oxysterol signalling pathway mediated by the nuclear receptor LXR alpha. Nature *383*, 728–731.

Kurokawa, R., DiRenzo, J., Boehm, M., Sugarman, J., Gloss, B., Rosenfeld, M.G., Heyman, R.A., and Glass, C.K. (1994). Regulation of retinoid signalling by receptor polarity and allosteric control of ligand binding. Nature *371*, 528–531.

Lichtarge, O., Bourne, H.R., and Cohen, F.E. (1996). An evolutionary trace method defines binding surfaces common to protein families. J. Mol. Biol. *257*, 342–358.

Lockless, S.W., and Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. Science *286*, 295–299.

Lu, T.T., Makishima, M., Repa, J.J., Schoonjans, K., Kerr, T.A., Auwerx, J., and Mangelsdorf, D.J. (2000). Molecular basis for feedback regulation of bile acid synthesis by nuclear receptors. Mol. Cell *6*, 507–515.

Makishima, M., Okamoto, A.Y., Repa, J.J., Tu, H., Learned, R.M., Luk, A., Hull, M.V., Lustig, K.D., Mangelsdorf, D.J., and Shan, B. (1999). Identification of a nuclear receptor for bile acids. Science *284*, 1362–1365.

Makishima, M., Lu, T.T., Xie, W., Whitfield, G.K., Domoto, H., Evans, R.M., Haussler, M.R., and Mangelsdorf, D.J. (2002). Vitamin D receptor as an intestinal bile acid sensor. Science *296*, 1313–1316.

Maloney, P.R., Parks, D.J., Haffner, C.D., Fivush, A.M., Chandra, G., Plunket, K.D., Creech, K.L., Moore, L.B., Wilson, J.G., Lewis, M.C., et al. (2000). Identification of a chemical tool for the orphan nuclear receptor FXR. J. Med. Chem. *43*, 2971–2974.

Mangelsdorf, D.J., and Evans, R.M. (1995). The RXR heterodimers and orphan receptors. Cell *83*, 841–850.

Menon, S.T., Han, M., and Sakmar, T.P. (2001). Rhodopsin: structural basis of molecular physiology. Physiol. Rev. *81*, 1659–1688.

Repa, J.J., Turley, S.D., Lobaccaro, J.A., Medina, J., Li, L., Lustig, K., Shan, B., Heyman, R.A., Dietschy, J.M., and Mangelsdorf, D.J. (2000). Regulation of absorption and ABC1-mediated efflux of cholesterol by RXR heterodimers. Science *289*, 1524–1529.

Schmidt, A., Endo, N., Rutledge, S.J., Vogel, R., Shinar, D., and Rodan, G.A. (1992). Identification of a new member of the steroid hormone receptor superfamily that is activated by a peroxisome proliferator and fatty acids. Mol. Endocrinol. *6*, 1634–1641.

Schulman, I.G., Li, C., Schwabe, J.W., and Evans, R.M. (1997). The phantom ligand effect: allosteric control of transcription by the retinoid X receptor. Genes Dev. *11*, 299–308.

Suel, G.M., Lockless, S.W., Wall, M.A., and Ranganathan, R. (2003). Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat. Struct. Biol. *10*, 59–69.

Svensson, S., Ostberg, T., Jacobsson, M., Norstrom, C., Stefansson, K., Hallen, D., Johansson, I.C., Zachrisson, K., Ogg, D., and Jendeberg, L. (2003). Crystal structure of the heterodimeric complex of LXRalpha and RXRbeta ligand-binding domains in a fully agonistic conformation. EMBO J. *22*, 4625–4633.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.

Thompson, P.D., Jurutka, P.W., Haussler, C.A., Whitfield, G.K., and Haussler, M.R. (1998). Heterodimeric DNA binding by the vitamin D receptor and retinoid X receptors is enhanced by 1,25-dihydroxyvitamin D3 and inhibited by 9-cis-retinoic acid. Evidence for allosteric receptor interactions. J. Biol. Chem. *273*, 8483–8491.

Wells, J.A. (1991). Systematic mutational analyses of protein-protein interfaces. Methods Enzymol. *202*, 390–411.

Williams, S., Bledsoe, R.K., Collins, J.L., Boggs, S., Lambert, M.H., Miller, A.B., Moore, J., McKee, D.D., Moore, L., Nichols, J., et al. (2003). X-ray crystal structure of the liver X receptor beta ligand binding domain: regulation by a histidine-tryptophan switch. J. Biol. Chem. *278*, 27138–27143.

Willy, P.J., and Mangelsdorf, D.J. (1997). Unique requirements for retinoid-dependent transcriptional activation by the orphan receptor LXR. Genes Dev. *11*, 289–298.

Xu, H.E., Stanley, T.B., Montana, V.G., Lambert, M.H., Shearer, B.G., Cobb, J.E., McKee, D.D., Galardi, C.M., Plunket, K.D., Nolte, R.T., et al. (2002). Structural basis for antagonist-mediated recruitment of nuclear co-repressors by PPARalpha. Nature *415*, 813–817.

Young, M.A., Gonfloni, S., Superti-Furga, G., Roux, B., and Kuriyan, J. (2001). Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. Cell *105*, 115–126.