

# SCIENTIFIC REPORTS



OPEN

## Allosteric Pathways in the PPAR $\gamma$ -RXR $\alpha$ nuclear receptor complex

Clarisse G. Ricci<sup>1</sup>, Rodrigo L. Silveira<sup>1</sup>, Ivan Rivalta<sup>2,3</sup>, Victor S. Batista<sup>3</sup> & Munir S. Skaf<sup>1</sup>

Received: 17 August 2015  
Accepted: 07 December 2015  
Published: 29 January 2016

Understanding the nature of allostery in DNA-nuclear receptor (NR) complexes is of fundamental importance for drug development since NRs regulate the transcription of a myriad of genes in humans and other metazoans. Here, we investigate allostery in the peroxisome proliferator-activated/retinoic X receptor heterodimer. This important NR complex is a target for antidiabetic drugs since it binds to DNA and functions as a transcription factor essential for insulin sensitization and lipid metabolism. We find evidence of interdependent motions of  $\Omega$ -loops and PPAR $\gamma$ -DNA binding domain with contacts susceptible to conformational changes and mutations, critical for regulating transcriptional functions in response to sequence-dependent DNA dynamics. Statistical network analysis of the correlated motions, observed in molecular dynamics simulations, shows preferential allosteric pathways with convergence centers comprised of polar amino acid residues. These findings are particularly relevant for the design of allosteric modulators of ligand-dependent transcription factors.

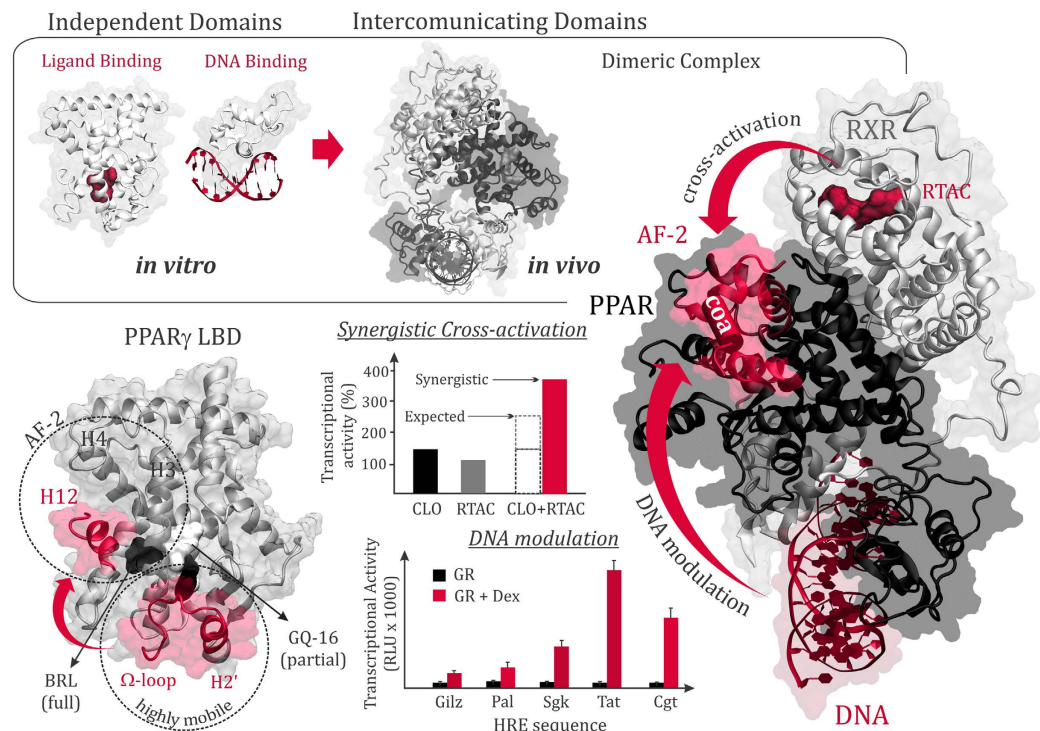
Understanding the molecular origin of allosteric mechanisms in protein and DNA complexes is a challenge of great current interest<sup>1,2</sup> because of its importance to cellular regulation and the design and development of allosteric drugs<sup>3</sup>. Despite recent efforts<sup>4–8</sup>, allostery remains a biophysical enigma since little is known about the molecular mechanisms that trigger long-range conformational changes in biomolecular systems<sup>9,10</sup>. However, it is well known that fluctuations between conformational substates<sup>11,12</sup> are key for activation and functional regulation under various physiological conditions<sup>13</sup>. Here, we focus on the molecular origin of allosteric pathways in nuclear receptors (NRs) that regulate the transcription of a myriad of genes in humans and other metazoans<sup>14,15</sup>.

NRs are proteins that attach to specific sequences of DNA known as Hormone Response Elements (HRE), forming homodimers or heterodimers with the promiscuous Retinoic X Receptor (RXR), depending on the NR family type<sup>14,16</sup>. Each NR monomer comprises three major domains: (i) the amino terminus domain (also known as the A/B region) containing a transactivation domain; (ii) the small well-conserved DNA Binding Domain (DBD) that recognizes and targets specific HREs, and (iii) a larger globular carboxy-terminal Ligand Binding Domain (LBD) with the ligand binding pocket (LBP) and the main molecular surfaces involved in dimerization and coactivator recruitment<sup>14,15</sup>. The ‘hinge’ is a more variable, unstructured, fragment that connects the DBD to the LBD. Whether the hinge has other specific functional roles remains unclear<sup>16</sup> and is here explored by our computational analysis. It is clear that NRs are modular in nature and that the isolated domains display intrinsic activities *in vitro*. In addition, there is compelling evidence showing that different domains influence one another in their native environment, adding an extra level of complexity to NR modulation<sup>17,18</sup>. For instance, it was recently shown that the dynamics of LBDs is affected by their interaction with DNA<sup>19</sup>. More importantly, ligand-induced activation is dependent on the anchoring DNA sequence<sup>20,21</sup>. Although these studies support the existence of synergism and communication between DBDs, LBDs and DNA, significant structural differences between isolated, complexed domains and DNA bound dimers have not been detected that could explain such behavior, suggesting that dynamic or ensemble allostery might play a role in NRs function.

Moreover, in the non-permissive type of heterodimers, heterodimerization and interaction with DNA can impair ligand binding by RXR<sup>22</sup>, suggesting communication between the two subunits. In the permissive type of heterodimers – in which the RXR fully retains its ability to bind ligands – occupancy of RXR's binding pocket leads to activation of the other NR subunit, either in the absence (transactivation) or in the presence of its own *bona fide* ligand (synergism)<sup>17,23–25</sup>, which also supports the existence of functional inter-LBD cooperation.

The latter scenario is common to Peroxisome Proliferator-Activated Receptors (PPARs), which are activated by 9-*cis*-retinoic acid (RTAC) as well as by their own ligands (e.g., fibrate drugs, thiazolidinediones and fatty acid derivatives from diet and metabolism)<sup>8,24,26</sup>. PPARs are expressed as three subtypes ( $\alpha$ ,  $\gamma$ , and  $\delta$ ) with distinct

<sup>1</sup>Institute of Chemistry, University of Campinas-UNICAMP, Cx. P. 6154, Campinas SP 13084-862, Brazil. <sup>2</sup>Université de Lyon, CNRS, Laboratoire de Chimie, École Normale Supérieure de Lyon, UMR 5182, 46 Allée d'Italie, 69364 Lyon, Cedex 07, France. <sup>3</sup>Department of Chemistry, Yale University, P.O. Box 208107, New Haven, CT 06520-8167, United States. Correspondence and requests for materials should be addressed to I.R. (email: ivan.rivalta@ens-lyon.fr) or V.S.B. (email: victor.batista@yale.edu) or M.S.S. (email: skaf@iqm.unicamp.br)



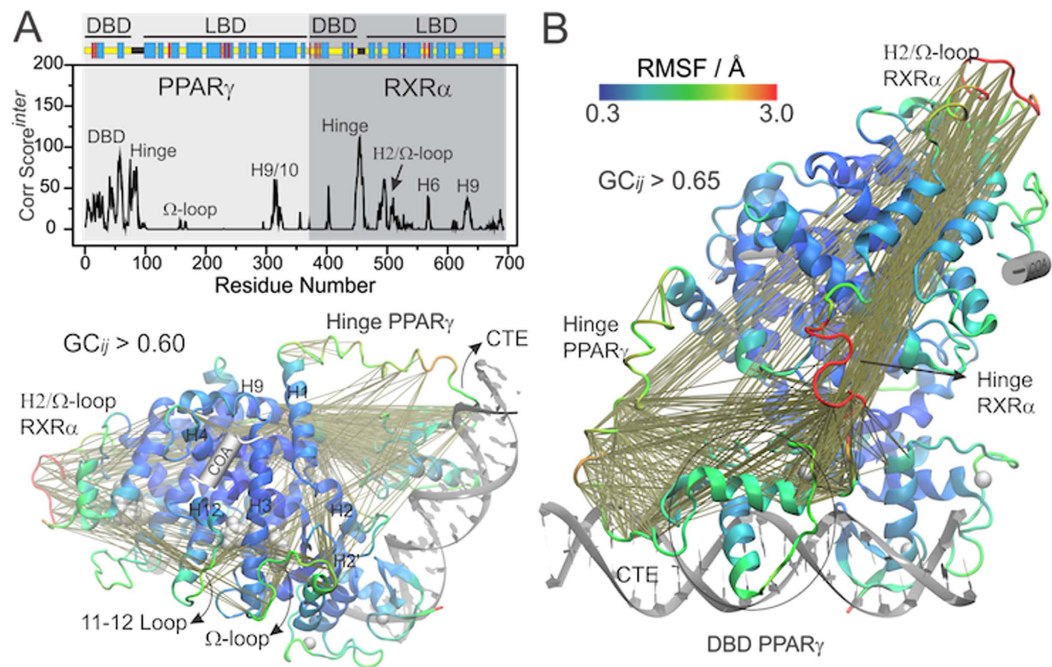
**Figure 1. Allosteric communications in the PPAR $\gamma$ -RXR  $\alpha$  nuclear receptor (NR).** NRs are modular proteins formed by ligand (LBD) and DNA (DBD) binding domains that display intrinsic activity *in vitro*. *In vivo*, they form dimeric complexes, such as the PPAR $\gamma$ -RXR $\alpha$  complex, that bind to DNA and show allosteric interdomains communication (top and right panels). RXR ligands, such as 9-*cis*-retinoic acid (RTAC), can cross-activate PPAR by inducing conformational changes in its activation function-2 (AF-2) region, comprising the helix 12 (H12) in the LBD. Synergistic cross-activation is known to take place when PPAR is bound to one of its own *bona fide* ligands, such as clofibric acid (CLO, illustrative histograms reproduced from ref. 23). As observed for other NRs (such as glucocorticoid, estrogen and vitamin D receptors), transcriptional activity can be modulated by the DNA sequence (illustrative histograms reproduced from ref. 21), suggesting allosteric communications between the DNA and the AF-2 region. The PPAR $\gamma$  LBD (bottom left panel) differs from other NRs by having an extra helix (H2') followed by an extremely large and flexible  $\Omega$ -loop. Full agonists, such as rosiglitazone (BRL), activate the AF-2 region by forming a direct interaction with the H12, while partial agonists (such as GQ-16) do not make direct contact with H12, indicating alternative allosteric mechanisms that may involve the  $\Omega$ -loop and H2' flexible regions.

although overlapping tissue distributions and impact on metabolism<sup>24</sup>. In particular, subtype  $\gamma$  is a major regulator of glucose and lipid metabolism and has received significant attention due to its implication in modern-life metabolic conditions such as diabetes, obesity, and cardiovascular diseases<sup>27,28</sup>.

Structurally, the PPAR $\gamma$ -LBD displays the usual three-layer  $\alpha$ -helix fold common to other NRs (Fig. 1), although it differs from others for having an extremely large LBP and an additional helix between helices 2 and 3, named helix 2'<sup>29–32</sup>. Helix 2' (H2') is followed by a lengthy loop commonly referred to as  $\Omega$ -loop and its preceding helix (henceforth referred to as  $\Omega$ -region) comprise the most conformationally flexible part of the LBD, as revealed by crystallographic B-factors and molecular dynamics simulations<sup>31,33,34</sup>.

The classic activation mechanism proposes that ligands activate NRs by stabilizing the C-terminal helix (helix 12) in a favorable position to harbor coactivator molecules<sup>16,18,35</sup>. Such a mechanism is consistent with the observation that PPAR $\gamma$ 's full agonists typically activate the activation function-2 (AF-2) region, interacting with helix 12 (H12) by means of a hydrogen bonding with Tyr473<sup>29,30</sup>. However, studies with partial agonists have shown that: i) such ligands do not interact directly with H12<sup>36–41</sup>, and ii) their transcriptional activity is not correlated exclusively with structural or dynamic changes in H12<sup>39,41,42</sup>. Therefore, other regions of the LBD may play roles in ligand-induced activation, such as the  $\Omega$ -loop, which appears to be particularly important for PPAR $\gamma$  activation by fatty acids and flavonoids<sup>33,43</sup>. An aspect that remains unclear is whether the  $\Omega$ -loop operates via an allosteric mechanism involving helix 12.

In marked contrast with ligand-induced activation, the allosteric mechanisms by which DNA or RXR ligands modulate PPAR $\gamma$  activity remain poorly understood, partially due to a lack of structural knowledge on how different NR domains are organized in a quaternary structure. Recent breakthroughs, however, have yielded the structures of functional NR complexes bound to DNA<sup>44–48</sup>, opening up the possibility of investigating inter-domain communications at the molecular level. Herein, we apply molecular dynamics (MD) simulation methods in conjunction with the community network analysis of correlated motions<sup>6,49–51</sup> to characterize interdependent



**Figure 2. Large-scale inter-domain correlated motions on the full-length complex.** (A) Correlation score per residue, computed for inter-domain correlations with  $GC_{ij}^{inter} > 0.6$ . (B) Localization of the most intense inter-domain correlations in the three-dimensional structure. The heterodimer  $PPAR\gamma$ - $RXR\alpha$  was colored according to backbone RMSF, while other components are represented in silver. For the sake of clarity, the bottom left panel only shows inter-domain correlations that involve  $PPAR\gamma$  residues.

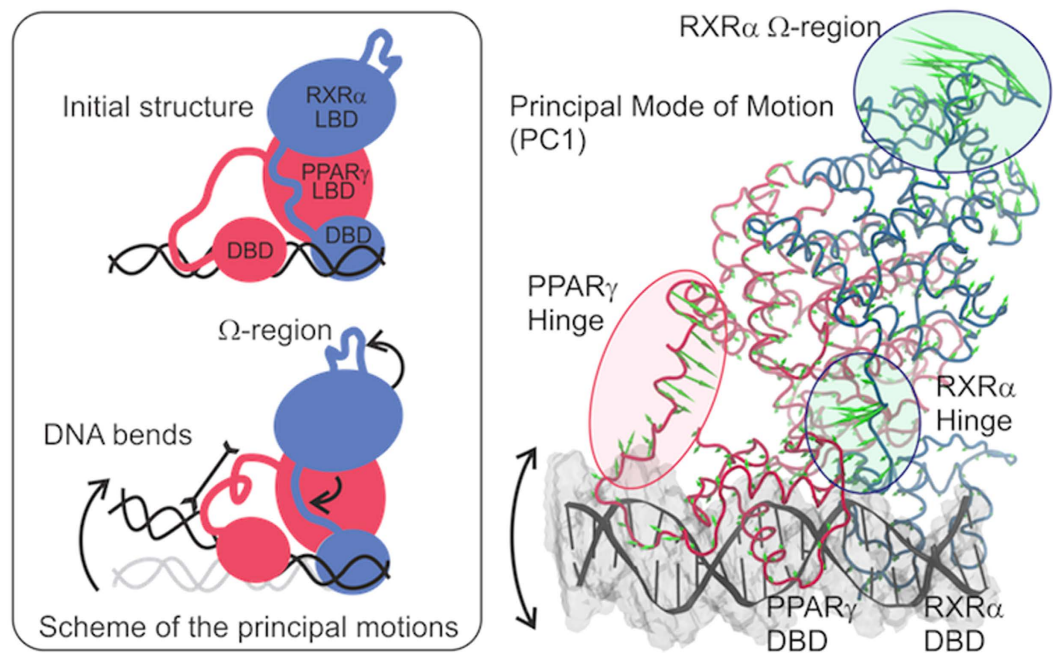
motions and allosteric pathways responsible for inter-domain communications in the full-length  $PPAR\gamma$ - $RXR\alpha$  complex bound to DNA.

### Inter-Domain Correlated Motions

To quantify interdependent motions taking place in the  $PPAR\gamma$ - $RXR\alpha$ -DNA complex, we computed inter-residue generalized correlation (GC) coefficients from the MD trajectories. GC analysis produces a matrix of pair-wise correlation coefficients,  $GC_{ij}$ , which measure the extent to which the position (or motion) of residue  $i$  restricts the range of positions (or motions) available to residue  $j$ . The resulting correlation matrix of the  $PPAR\gamma$ - $RXR\alpha$  complex (Figure S1A) shows that almost all residues in the complex are correlated to one another to some extent ( $GC > 0.2$ ), reflecting the fact that the residues are all bound – either covalently or by non-bonding contacts – in the same quaternary structure and that large protein regions display rigid body translations. This is especially the case of intra-domain correlations, which arise from rigid body motions of a particular domain within the  $PPAR\gamma$ - $RXR\alpha$ /DNA molecular architecture. Since each subunit is formed by a DBD-LBD pair connected by a disordered link (hinge), if one domain translates relative to the others, such motion is expected to produce collective changes of the correlation coefficients for all pairs of residues belonging to that domain. Moreover, the sum of correlation coefficients (see ‘correlation scores’ in Methods) accumulated over that domain would be proportional to the number of its residues. Indeed, we observe such pattern for the internal correlations of residues belonging to domains that undergo translations during the MD trajectories (Figure S2), suggesting that most of the intra-domain correlations in the  $PPAR\gamma$ - $RXR\alpha$  complex arise from rigid body-like motions. Because detection of such ‘trivial correlations’ makes it difficult to identify which correlations are truly representative of allosteric mechanisms, we focused our attention on the most intense correlations, which comprise also interdependent conformational transitions in flexible regions.

To filter long-distance and non-trivial correlations, we computed a correlation score per residue that accumulates all intense and inter-domain correlations ( $CG_{ij}^{inter} > 0.6$ ) displayed by each residue  $i$  (see Methods for details). Figure 2 shows the distribution of such scores along the primary structure (panel A), with the most intense correlations depicted as lines in the three-dimensional structure of the full-length complex (panel B). Interestingly, the most intense inter-domain correlations were not equally distributed along the structure, but rather concentrated in specific regions of the complex that comprise the  $PPAR\gamma$ -DBD, conformationally flexible parts (hinges and  $\Omega$ -loops) and helices that participate in dimerization contacts (H9/10).

Regions that concentrate strong inter-domain correlations are also highly mobile, indicating that they either undergo large conformational changes or execute rigid body like translations within the complex (see RMSF color scale in Fig. 2B). For the  $PPAR\gamma$ -DBD, the latter is likely the case, since DBDs are known for being internally rigid domains<sup>52,53</sup>. In particular, both the  $\Omega$ -loops and their preceding helices (helix 2 in  $RXR\alpha$  and helix 2' in  $PPAR\gamma$ ) show large amplitude correlated motions with one another and with  $PPAR\gamma$ -DBD/CTE (C-terminal



**Figure 3. Principal mode of motion (PC1) reveals translations of PPAR $\gamma$ -DBD and coupled conformational changes in the hinges and  $\Omega$ -regions.** PC1 is represented schematically (left panel) or by arrows of sizes proportional to the amplitude of motion (right panel). DNA occupancy map is shown as a grey surface in the right panel.

extension). The RXR $\alpha$ -hinge also appears to be correlated to the motions of the PPAR $\gamma$ -DBD and CTE. In contrast, RXR $\alpha$ -DBD is less mobile and displays only minor correlations with the LBDs.

### Intrinsic dynamics and mobility analysis

We next set out to examine what are the main motions performed by the complex as portrayed by a principal component analysis (PCA). Figure 3 illustrates the dynamics of the complex along the first principal mode of motion (PC1), where the arrows indicate both the direction and the relative amplitude of motions. We found that the dynamics of the hinges and  $\Omega$ -regions are coupled to the motions of PPAR $\gamma$ -DBD/CTE, which appears to move along with the DNA fragment.

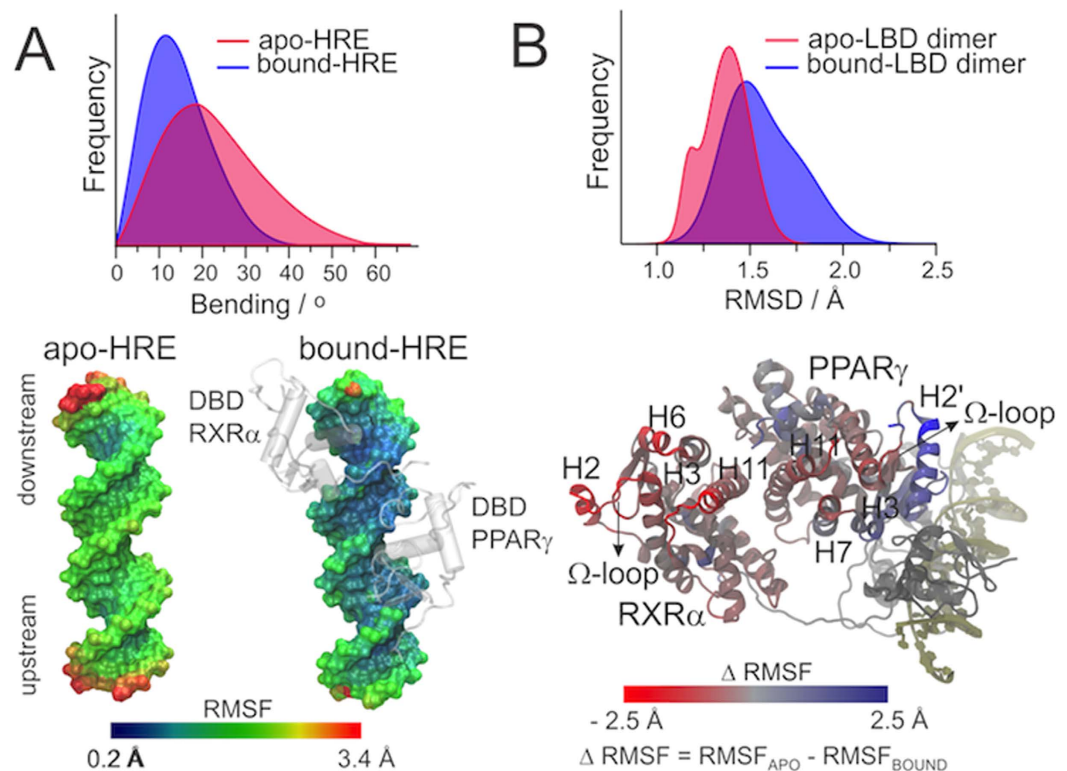
To investigate whether DNA induces such coupled motions, we additionally simulated the PPAR $\gamma$  Hormone Response Element as an isolated DNA stretch (apo-HRE). Conformational analysis revealed that apo-HRE is intrinsically prone to bend, reaching curvatures of up to 50 degrees (histograms in Fig. 4A). When interacting with the NR heterodimer, the bound-HRE has its mobility restricted, particularly in the downstream region, which hosts the RXR $\alpha$ -DBD and on top of which sit the LBDs (bottom panel in Fig. 4A). The upstream region, which hosts the PPAR $\gamma$ -DBD/CTE domain, retains residual motions that are sufficient to promote DNA bending of up to 30 degrees. It is thus likely that the translations displayed by PPAR $\gamma$ -DBD are induced by intrinsic bending motions of the HRE itself. The fact that RXR $\alpha$ -DBD lies on a more constrained region of the HRE is consistent with its reduced mobility as compared to PPAR $\gamma$ -DBD (see Fig. 2).

In contrast, HRE binding has a global effect of enhancing LBD dynamics, as shown by the RMSD distribution of the PPAR $\gamma$ -RXR $\alpha$ -DNA complex (bound-LBD dimer) in comparison to the PPAR $\gamma$ -RXR $\alpha$  dimer (apo-LBD dimer) (histograms in Fig. 4B). Despite such global effect, some particular regions of the heterodimer were stabilized after interaction with the HRE, including PPAR $\gamma$ -H2' and part of its  $\Omega$ -loop (bottom panel in Fig. 4B). Therefore, interaction with the HRE appears to induce different effects on each of the LBDs: while it reduces the mobility of PPAR $\gamma$ -H2', it enhances the motions in RXR $\alpha$   $\Omega$ -region. Still, the fact that absolute RMSF of PPAR $\gamma$ -H2' remains high in the bound state (up to 5 Å in the C-terminal extremity) suggests that the interaction with DNA does not necessarily stabilize it, but rather induces a more coherent type of motion.

### Intra-domain correlations within PPAR $\gamma$ LBD

A generalized correlation analysis was also performed to examine the non-trivial intra-domain correlations within the PPAR $\gamma$ -LBD and their dependence with the LBD oligomerization state (Fig. 5). Comparison of the PPAR $\gamma$ -LBD as a monomer with PPAR $\gamma$ -LBD as part of the full-length complex revealed that most correlations exhibited by the monomer are lost upon heterodimerization and interaction with DNA. However, they are replaced by numerous and strong correlations that do not appear in the monomer and which connect the  $\Omega$ -region to the helix 12 and its preceding loop (Fig. 5A).

Comparison of intra-domain correlation scores (Fig. 5B) confirms that the native oligomeric state induced a remarkable increase in the accumulated scores of residues lying close to the LBP (Fig. 5C), many of which have



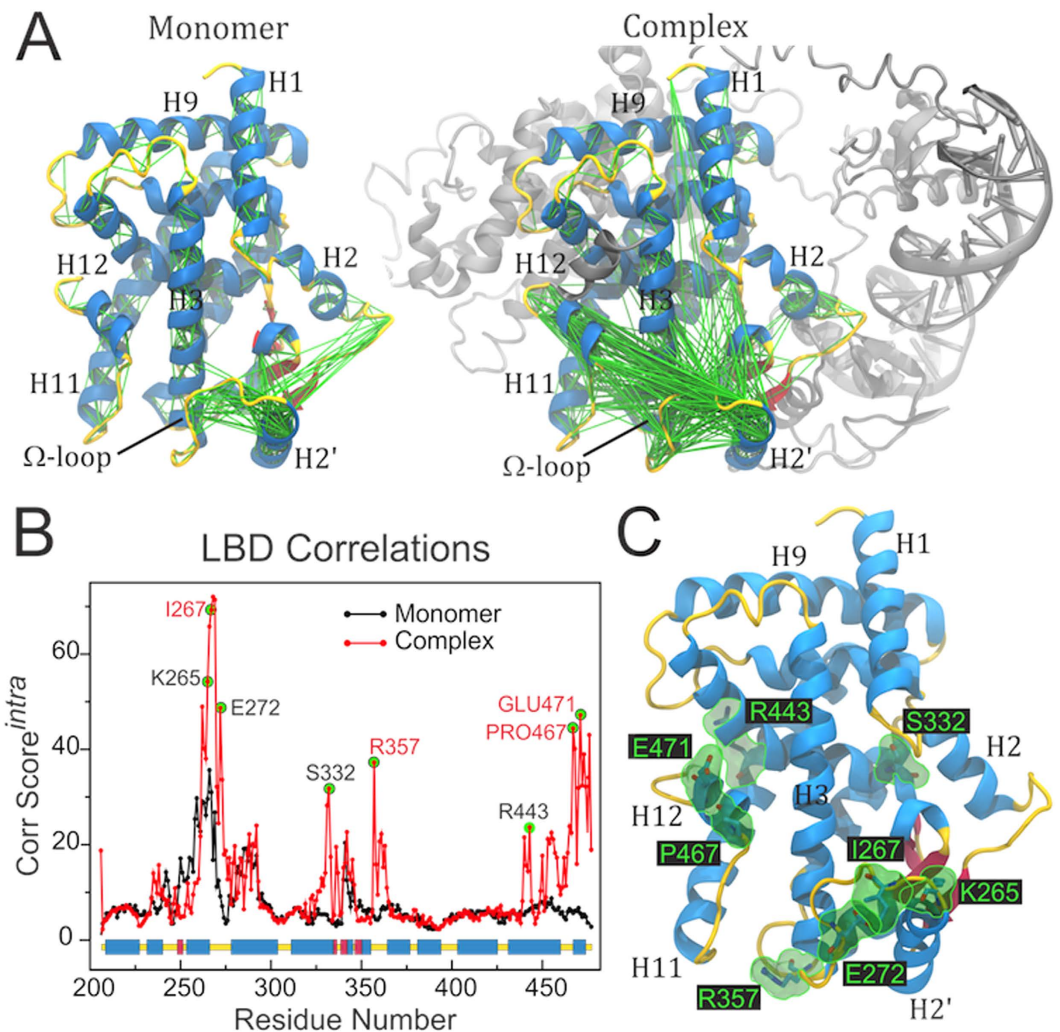
**Figure 4. Mobility analysis comparing apo- or bound- states of the HRE (A) and the LBD dimer (B).** The HRE mobility is shown in terms of bending (histograms) and RMSF (bottom panel). For the sake of clarity, only the DBDs are shown in the bound-HRE state. The LBD dimer mobility is shown in terms of RMSD distribution (histograms) and differential RMSF (bottom panel). Differential RMSF was computed by subtracting the RMSF in the bound state from the RMSF in the apo state.

either polar or charged side-chains. Interestingly, at least four of these residues have been indicated to be relevant for PPAR $\gamma$  function, according to experimental or theoretical studies (Table 1). A point mutation at I267, located at the  $\Omega$ -loop, was recently found to abolish PPAR $\gamma$  activation by 15-deoxy- $\delta^{12,14}$ -prostaglandin J $_2$ <sup>43</sup>. Since I267 has a very prominent role in the correlation network, its functional importance might be related to a key participation in intra-domain allosteric mechanisms, which could be the case of other residues belonging to the  $\Omega$ -loop. R357 lies at the bottom part of PPAR $\gamma$ -LBD (see Fig. 5C), where it forms salt bridges with E276 and E460, locking the loop 6–7, the bottom part of  $\Omega$ -loop, and the loop 11–12 together. Computational molecular simulations have suggested that this region comprises an important dissociation pathway by which ligands may escape the ligand binding pocket<sup>54</sup>. Therefore, concerted motions involving R357 might be a mechanism of controlling ligand dissociation kinetics. Two other residues whose correlations emerged in the native oligomeric state lie at helix 12 and are engaged in important hydrophobic contacts (P467) or electrostatic interactions (E471) that hold the coactivator peptide in place<sup>29,55</sup>. Allosteric modulation of these residues by distant regions such as the  $\Omega$ -loop can thus directly impact PPAR $\gamma$  ability to recruit coactivator proteins.

### Residual Local Frustration

To quantify conformational residual local frustration in the crystallographic structure of the full-length complex, we applied the frustratometer algorithm<sup>56</sup>. The frustratometer measures how ‘frustrated’ a residue is by checking how conformational/mutational changes in its vicinity shift the residue energetics<sup>13</sup>. According to the energy landscape theory, while minimally frustrated contacts are important to stabilize the folded core of the protein, local clusters of highly frustrated contacts might have evolved so that proteins can easily adapt their structure and modulate their function in response to their environment<sup>13</sup>.

Figure 6 shows the distribution of minimally (green) and highly (red) frustrated contacts along the quaternary fold. As expected, DBDs along with the inner core of LBDs and dimerization regions are dense with minimally frustrated contacts. The PPAR $\gamma$ -hinge, in contrast, was found to be the most frustrated region of the complex, suggesting that it works as a ‘molecular spring’ that can change shape in response to DNA distortions<sup>48</sup>. Highly frustrated contacts were also observed in the RXR $\alpha$ -hinge, both the  $\Omega$ -regions and other helices and loops near the proteins’ surface (H1, H2, loop 3–4, H6 to H7, loop 11–12, and H12). Interestingly, some of the highly frustrated clusters map well to highly correlated regions (compare to Fig. 2), reinforcing that frustrated regions might have additionally evolved to undergo interdependent conformational changes. Since flexible regions are separated by minimally frustrated contacts at the folded core of the complex, in the next section we investigate how information is transmitted along the structure through interdependent motions.



**Figure 5.** Correlation network within PPAR $\gamma$ -LBD depends on the oligomeric state. (A) Intense correlations ( $GC_{ij}^{intra} > 0.6$ ) displayed by PPAR $\gamma$ -LBD as a monomer (left) or as part of the full-length complex (right). (B) Correlation score per residue, computed for  $GC_{ij}^{intra} > 0.6$ . Residues written in red are known to be important for function. (C) Localization of some of the most intensely correlated residues on the LBD structure (marked in green in (B)).

Residue	Location	Evidence supporting functional importance
I267	$\Omega$ -loop	I267A abolishes 15d-PGJ <sub>2</sub> activation (ref. 43)
R357	Loop6-7	Forms salt bridge that locks a ligand exit pathway (ref. 54)
P467	H12	P467L attenuates ligand binding and coactivator recruitment (ref. 55)
E471	H12	Participates in the charge clamp (ref. 29)

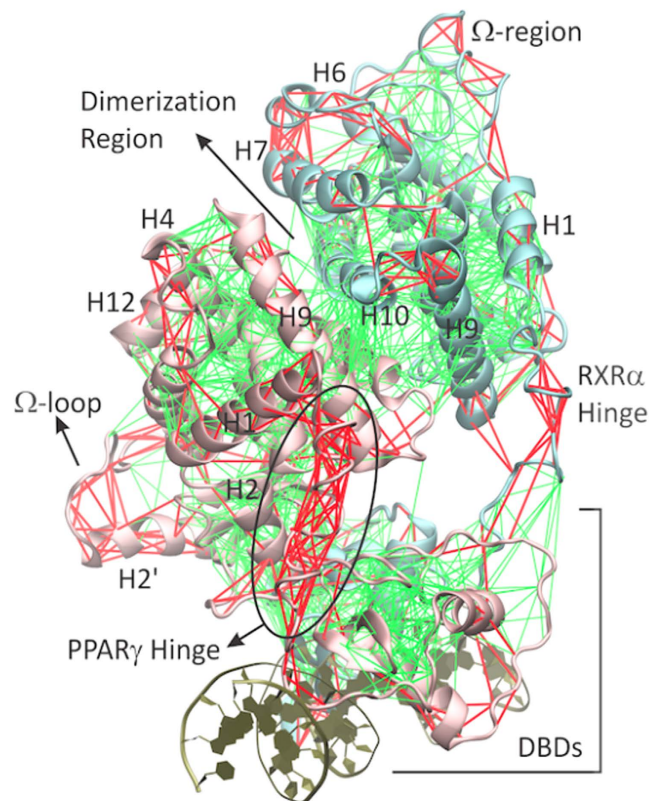
**Table 1.** Residues with high correlation scores and demonstrated (or suggested) importance for PPAR $\gamma$  function.

### Weighted protein network and preferential allosteric pathways

To compute the relative contribution of different types of residues to the overall correlation network, we devised a ‘contribution score’, computed for each residue type  $i$  as the difference between the *accumulated score* ( $AS_i$ ) and the *expected score* ( $ES_i$ ), where  $AS_i$  is the sum of the correlation coefficients involving at least one residue of type  $i$  and  $ES_i$  is computed as:

$$ES_i = TS \times \left( \frac{N_i}{N} \right) \quad (1)$$

where  $TS$  is the sum of all generalized coefficients in the complex, *total score*,  $N_i$  is the number of residues of type  $i$  and  $N$  is the total number of residues in the complex. Therefore, the  $ES_i$  depends only on the number of



**Figure 6. Conformational residual local frustration analysis.** Minimally frustrated contacts (green lines) are concentrated in the DBDs, the folded core of the LBDs, and the hetero-dimerization surface. Highly frustrated contacts (red lines) are clustered in the hinges,  $\Omega$ -loops, and in surface helices and loops.

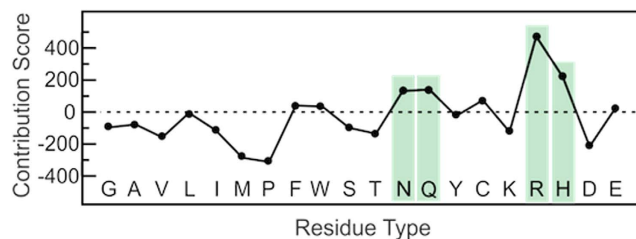
occurrences of each residue type in the structure. It reflects the relative contribution that would be expected if every type of amino acid contributed with the same intensity to the overall network of correlations, irrespective of the side-chain nature. By subtracting the  $ES_i$  from the observed  $AS_i$ , we obtain a metric of the deviations from this behavior, which we call *contribution score*. While positive values reflect a more important contribution than would be expected if only the backbone atoms mattered for conveying information, negative values reflect an unexpectedly low contribution to the overall network.

It is evident from Fig. 7 that aliphatic and apolar amino acids tend to contribute less to the allosteric communication than would be expected from their population in the primary sequence. On the other hand, arginines display an unexpected high contribution, followed by histidines and, to a lesser extent, glutamines and asparagines. Such intense participation in allosteric transmission could be either related to their location in the quaternary fold or to the very nature of their polar side chains.

We also used a weighted protein-network approach to identify the ‘shortest’ pathways connecting highly correlated residues that lie far apart in the three-dimensional structure. In this approach, the whole structure is represented by a set of nodes (i.e. residues) connected by edges whose *length* ( $w_{ij}$ ) depends on the GC coefficients (see Methods for more details). Based on such representation, the shortest pathways are those that simultaneously minimize the spatial distance (i.e. the number of nodes involved in the pathway) and maximize the allosteric efficiency (i.e. the intensity of the GC coefficients along the pathway).

This approach was applied to find the preferential pathways connecting the  $\Omega$ -regions and PPAR $\gamma$ -DBD/CTE. Interestingly, we found that the ‘shortest’ allosteric pathways avoid the more rigid core of the complex by going through solvent exposed loops and helices and by making use of highly frustrated contacts (Fig. 8, left panels). The PPAR $\gamma$ -hinge, despite being highly frustrated, does not comprise any important allosteric pathway. Therefore, DNA bending motions appear to translate into distant conformational changes by means of PPAR $\gamma$ -DBD translations, not via the PPAR $\gamma$ -hinge.

Additionally, while communication within the same or even neighboring secondary structures are usually degenerated, distinct secondary structures communicate by means of key interactions that generally involve at least one polar or charged residue (Fig. 8, right panels). This is particularly evident in case of communications between the two NR subunits, which take advantage of hydrogen bonds connecting i) PPAR $\gamma$  Q451 (H11) to RXR $\alpha$  E434 (H11), ii) RXR $\alpha$  K395 (H9) to PPAR $\gamma$  Q430 (L9-10), and iii) RXR $\alpha$  E217 (hinge) to PPAR $\gamma$  N151 (DBD).



**Figure 7.** Contribution Score computed for each residue type. Positive (or negative) values reflect residue types that contribute more (or less) to the correlation network than would be expected from their population in the primary sequence.

## Community Network and Vertex Betweenness

We have applied a community network analysis<sup>6,49,50</sup> to identify groups (or ‘communities’) of residues that are closely correlated. Such communities are correlated with each other through key amino acid residues that establish contacts critical for long-range allosteric mechanisms<sup>6</sup>.

The resulting communities obtained from our network analysis of MD simulations are depicted in Fig. 9A. The weighted protein-network is divided into communities based on the flow of allosteric information that passes through each pair of nodes (or ‘edge’). This is measured by the edge betweenness parameter, defined as the number of shortest pathways that pass through an edge. By looking at the residue pairs with the highest edge betweenness, we could also identify the strongest and best conserved contacts involved in inter-community communications, displayed in Fig. 9B.

We found that DBDs behave as individual communities (communities 0 and 4 in Fig. 9A), while LBDs display some dissociation between their ‘upper’ (communities 2 and 5 for PPAR $\gamma$  and RXR $\alpha$ , respectively) and ‘lower’ halves (communities 3 and 6, for PPAR $\gamma$  and RXR $\alpha$ , respectively), which appear as distinct though strongly connected communities. (The ‘upper’ and ‘lower’ parts are defined relative to the LBD orientation used in Fig. 1). Such community structure is consistent with NR’s modular nature, since DBDs, lower- and upper-LBDs have distinct and well-defined functions: DNA binding, ligand binding, and dimerization, respectively. Interestingly, the connection between upper-LBDs (communities 2 and 5) is stronger than the connections between upper and lower parts within each LBD (communities 2 and 3 for PPAR $\gamma$  and communities 5 and 6 for RXR $\alpha$ ), which highlights how effective the dimerization contacts are. Dimerization contacts also allow for communication between the two ligand binding pockets, as indicated by the sizeable interconnectivity of the lower LBD communities (connection between communities 3 and 6), with information flowing through the PPAR $\gamma$ -Y477/RXR $\alpha$ -E434 H-bond interaction connecting the two H11 helices (see Fig. 9B, bottom panel). Interestingly, the PPAR $\gamma$ -hinge appears as an extremely isolated community (community 1), displaying only weak communications with PPAR $\gamma$  DBD and LBD. This is consistent with the fact that the PPAR $\gamma$ -hinge appears as the most frustrated region in the complex. The RXR $\alpha$ -hinge, on the other hand, takes part in two communities that are mainly formed by PPAR $\gamma$ -DBD and RXR $\alpha$ -LBD.

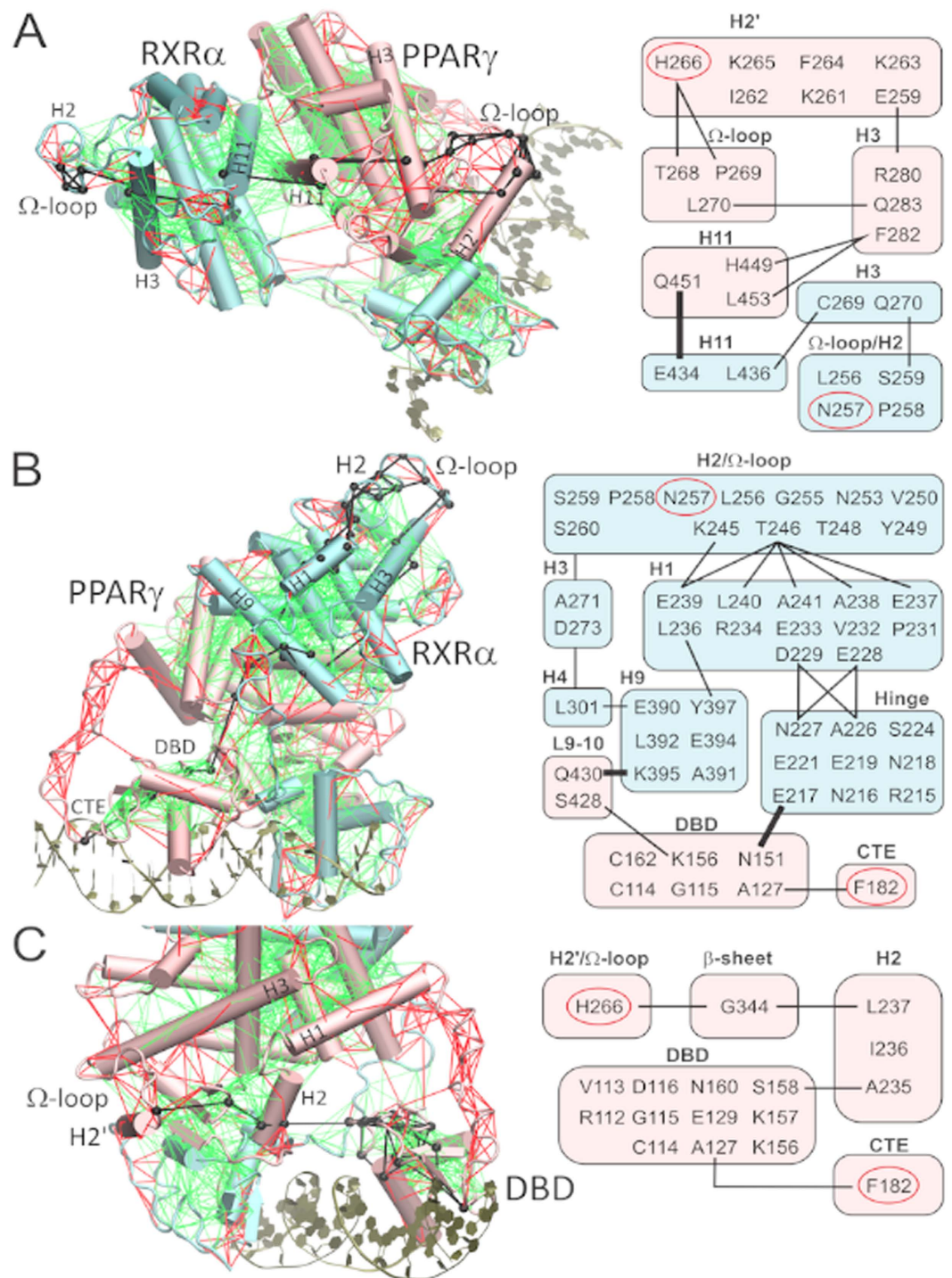
Edge betweenness analysis revealed a strong flow of allosteric information between PPAR $\gamma$ -DBD, PPAR $\gamma$ -LBD, and RXR $\alpha$ -hinge (Fig. 9B, top panel). A closer inspection reveals that it involves mainly electrostatic interactions between i) PPAR $\gamma$  Arg153 or His155 (DBD) and PPAR $\gamma$  Glu427 (loop 9–10) and ii) PPAR $\gamma$  Ser429 (loop 9–10) and RXR $\alpha$  Glu395 (hinge). Again, medium/long-chain charged or polar residues appear to be fundamental in conveying allosteric information between different subunits. In particular, PPAR $\gamma$  9–10 loop appears to form an important allosteric center mediating communication between PPAR $\gamma$ -DBD and both LBDs. Another important edge is formed by PPAR $\gamma$  Tyr477 (H12) and RXR $\alpha$  Glu434 (H11) (Fig. 9B, bottom panel). Analyses of the MD trajectories confirmed the existence of solvent-mediated hydrogen bond interactions linking Glu434 side-chain to Tyr477 backbone. Such residues displaying high betweenness are likely to play important roles in allosteric mechanisms and should be good candidates for mutagenesis studies.

## Discussion

In recent years, multidomain structures of NR complexes provided us with the first insights on how quaternary architecture influences NR function<sup>44–48</sup>. However, only a few structures are available, which capture the average conformation adopted by these complexes and, hence, information on the role of dynamics for allosteric mechanisms in physiological conditions is still rather limited. Here we report on the dynamics of the first X-ray structure of a full-length NR complex by means of molecular dynamics simulations. We used a combination of generalized correlation coefficients, principal component analysis (essential dynamics), residual local frustration analysis, and community networks to investigate correlated motions that enable distant regions of PPAR $\gamma$ -RXR $\alpha$ -DNA complex to allosterically communicate.

Generalized correlation analysis revealed that distant mobile regions of the complex undergo large-scale interdependent motions. Altogether, conformationally flexible parts of both LBDs appear to have their dynamics connected to PPAR $\gamma$ -DBD/CTE, but not to RXR $\alpha$ -DBD. Importantly, these mobile regions encompass the  $\Omega$ -loops and their preceding helices ( $\Omega$ -regions), which are strongly correlated to PPAR $\gamma$ -DBD and its C-terminal extension, inserted in the DNA minor groove. The  $\Omega$ -loops comprise the most flexible parts of the LBDs and have been recently suggested to be important modulators of PPAR function in addition to or in combination with H12<sup>33,34,43</sup>.

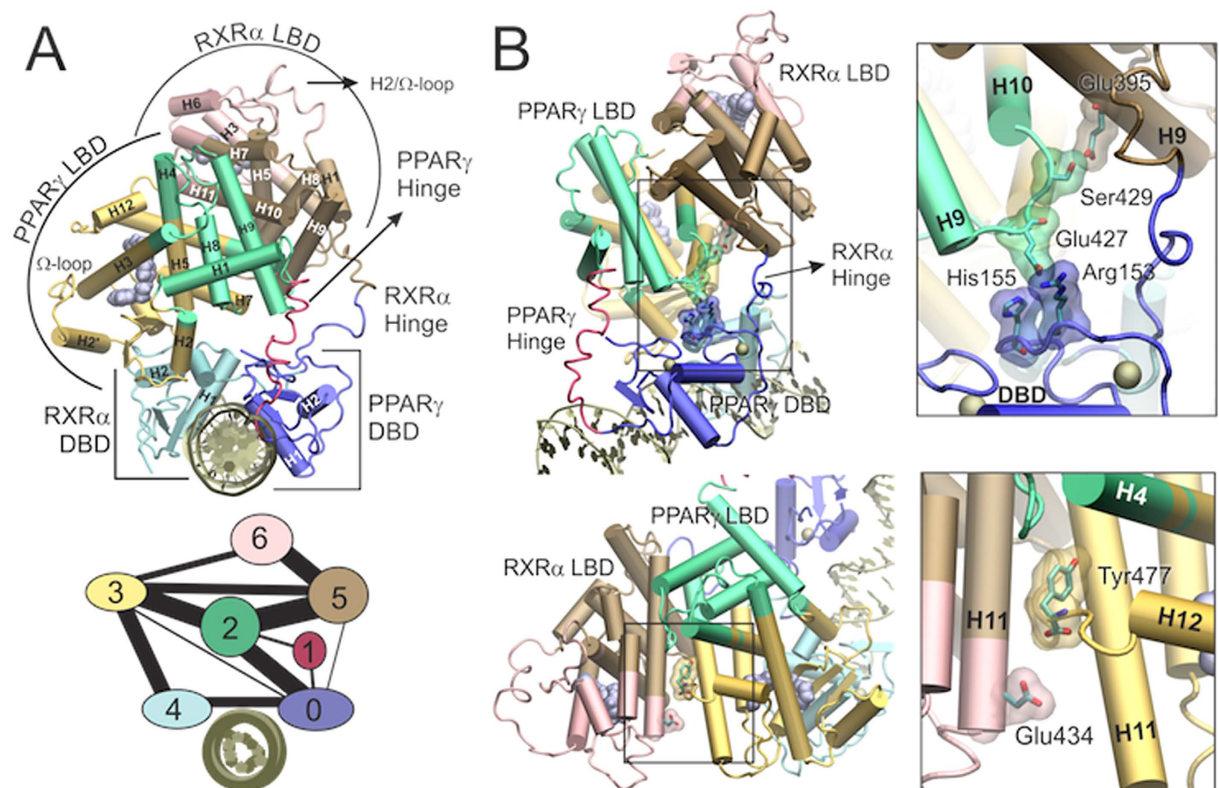




**Figure 8.** Shortest pathways connecting (A) RXR $\alpha$   $\Omega$ -loop to PPAR $\gamma$   $\Omega$ -loop, (B) RXR $\alpha$   $\Omega$ -loop to PPAR $\gamma$  –CTE and (C) PPAR $\gamma$   $\Omega$ -loop to PPAR $\gamma$ -CTE. Left panels indicate the shortest pathways as black lines in the three-dimensional structure, along with highly (red) and minimally (green) frustrated contacts from residual frustration analysis. Right panels schematize the shortest pathways connecting different secondary structures, with inter-domain communications highlighted in bold.

Therefore, our findings suggest that the  $\Omega$ -loops could work as important molecular switches that modulate PPAR $\gamma$  function in response to the dynamics of PPAR $\gamma$ -DBD.

Essential dynamics of the full-length complex confirmed that PPAR $\gamma$ -DBD and its CTE undergo rigid-body translations relative to the LBDs. This is in agreement with recent crystallographic data on the RXR $\alpha$ -LXR $\beta$  complex showing significant differences in the relative positions of the DBDs within the asymmetric unit of the crystal<sup>48</sup>. In our simulations, such translations appear to be induced by the HRE, which is intrinsically prone to

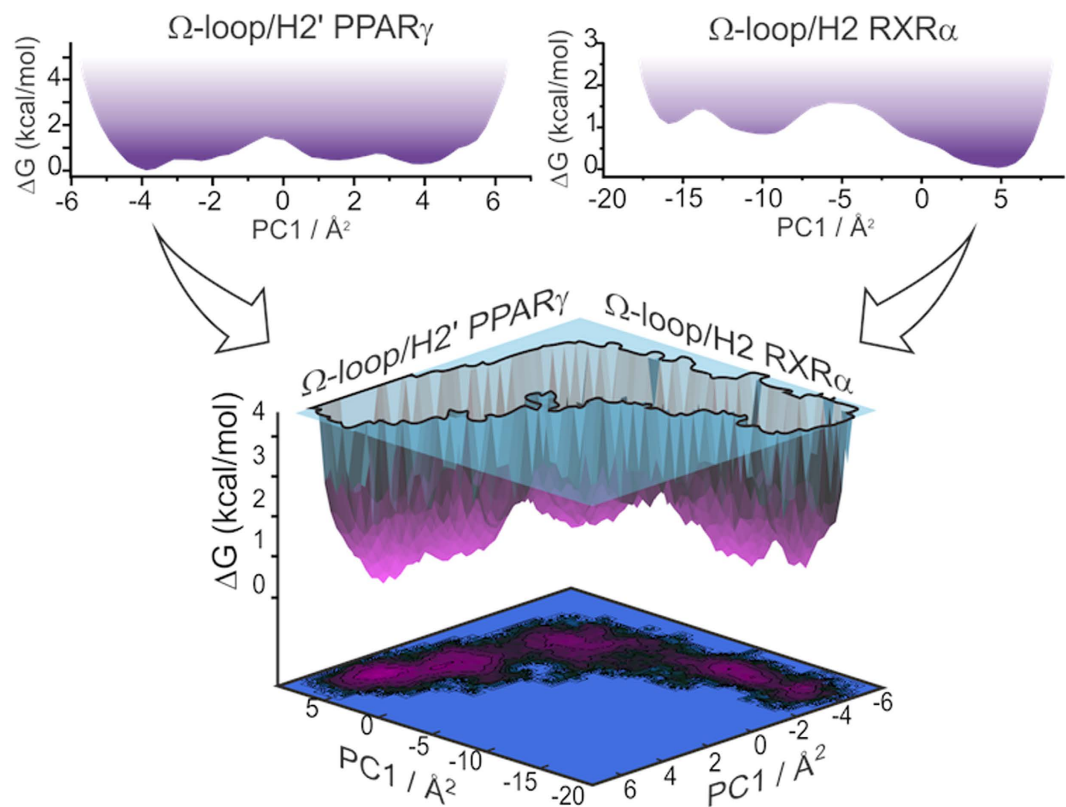


**Figure 9. Community network analysis.** (A) Community structure of the full-length complex displayed in the three-dimensional structure (top panel) or in schematic two-dimensional representation (bottom panel). The two-dimensional view of the communities depicts the relative size of the communities (number of residues) as colored circles of varying areas and the relative interconnectivity strength as lines of varying thicknesses. (B) Residues displaying the largest edge betweenness.

bending. Although the comparative analysis of the mobility indicated that the HRE loses motional freedom upon interaction with the full-length complex, the NR-bound HRE retains residual bending motions that are sufficiently large to translate the PPAR $\gamma$ -DBD/CTE domain and also to induce and control the dynamics of distant regions of the LBD dimer, including the  $\Omega$ -regions. Thus, consistent with the hypothesis of DNA being an active player on transcription<sup>1,20,21</sup>, our simulations suggest there is a flux of conformational information flowing from DNA to the LBDs, which is mediated by the PPAR $\gamma$ -DBD. Since small differences within a DNA sequence or in its flanking regions can produce large effects on its intrinsic flexibility, HRE dynamics – not its average conformation – could be the hidden information that is interpreted by bound NRs and translated into different levels of transcriptional activity, as previously reported<sup>20,21</sup>.

The RXR $\alpha$ -DBD, in turn, appears to have a much less prominent role than PPAR $\gamma$ -DBD on reading DNA dynamics, probably owing to its placement in a more constrained region of the HRE. Such observation is consistent with a recent Hydrogen/Deuterium Exchange study on the VDR-RXR complex, which has shown that RXR-DBD dynamics is less affected by DNA binding as compared to its partner DBD<sup>19</sup>. Despite the fact that the RXR $\alpha$ -DBD contribution for allosteric communication is likely to differ in other quaternary architectures, in PPAR $\gamma$ -RXR $\alpha$  complex its role appears to be limited to anchoring the complex to DNA, while the PPAR $\gamma$ -DBD works both on anchoring and allowing sequence-specific motions of the HRE to affect the dynamics of the two LBDs and, possibly, their function.

Within PPAR $\gamma$ -LBD, motions appear to be strongly dependent on the native oligomeric state, which queue the  $\Omega$ -region and H12 to move concertedly. Such coherence could be enforced either by i) steric constraint imposed by other domains of the complex, ii) DNA-induced motions or iii) a combination of both. At any rate, the native oligomeric environment appears to consolidate an allosteric mechanism by which the conformation and dynamics of the  $\Omega$ -loop could modulate H12 function, consistently with previous reports<sup>34,43</sup>. Our correlation analysis suggests that the intra-domain correlation network is mainly formed by residues surrounding the LBP, some of which have been suggested or proven to affect PPAR $\gamma$  function. While some of these residues directly interact with coactivator peptides (Pro467<sup>55</sup> and Glu471<sup>29</sup>), it is less evident how Ile267<sup>43</sup> or Arg357<sup>54</sup>, which are removed from H12, could modulate PPAR $\gamma$  activity. We suggest that their functional importance is strongly related to their participation in intra-domain allosteric mechanisms, by which they could influence H12 dynamics (Ile267) or ligand dissociation kinetics (Arg357). Since PPAR $\gamma$  depends both on dimerization with RXR $\alpha$  and on binding to DNA to be transcriptionally active, other residues whose correlations increase in the native state are likely important for function and may be good candidates for novel mutagenesis studies.



**Figure 10. Two-dimensional energy landscapes corresponding to the dynamics of the two  $\Omega$ -regions projected along their principal modes of motion (PC1).** Flexible regions such as the  $\Omega$ -loops correspond to flat areas in one-dimensional landscapes (top panels) that combine to form multidimensional canyons (bottom panels). Such canyon topography means that the conformation adopted by the RXR $\alpha$   $\Omega$ -loop restricts the range of conformations that can be adopted by the PPAR $\gamma$   $\Omega$ -loop and vice-versa.

Remarkably, strongly correlated regions such as the  $\Omega$ -regions and PPAR $\gamma$ -DBD/CTE are not spatially proximal, hinting at how complex is the mechanism by which allosteric information travels along the heterodimer. Moreover, in the space between such highly correlated regions lies the folded core of the complex, formed by minimally frustrated residues. While highly frustrated residues display unfavorable contacts that allow for large amplitude motions without significant energy cost, minimally frustrated residues are so comfortably placed that even small scale motions are energetically expensive<sup>13</sup>. Therefore, it would be naïve to assume that communication between distant and flexible regions occurs by a single (or even a few) allosteric pathway, which would require large-scale motions in the minimally frustrated core. Instead, we propose that communication between largely flexible regions must arise as a cooperative effect emerging from an extensive and complex network of small/medium amplitude motions throughout the structure.

This is equivalent to saying that the system's conformational energy landscape has been shaped to have multidimensional canyons as shown in Fig. 10. Therefore, some of the highly frustrated regions in the complex appear to have evolved to display interdependent motions, as previously proposed by Ferreira *et al.*<sup>57</sup>. Considering the large number of global conformations that would be available if the motions of two highly frustrated regions were completely independent of each other, such canyon-like topography might have evolved to restrict unproductive combinations of local conformations and, thus, accelerate transitions between functionally active states in proteins and other macromolecules.

We also found that polar amino acids have a more prominent role in the correlation network than would be expected from their frequency in the primary structure, thus suggesting the existence of preferential allosteric pathways by means of electrostatic interactions. This is particularly true for arginines, which is consistent with their recently described role as the most effective heat diffusers in PPAR $\gamma$  and in other proteins<sup>58</sup>. Weighted protein-network analysis also revealed that while internal correlations tend to be distributed in several degenerated pathways, they funnel down into key electrostatic and polar interactions when it comes to inter-domain communication.

As a general trend, the shortest correlation pathways connecting strongly correlated regions exploit frustrated regions at the surface of the complex, avoiding the minimally frustrated core. They comprise medium- or long-chain polar and charged residues near the surface of the proteins, especially at the PPAR $\gamma$  9–10 loop. This region represents a ‘convergence zone’, connecting the PPAR $\gamma$  DBD to both LBDs. A similar ‘convergence center’ was recently reported for the HNF-4 $\alpha$  NR complex, where single-residue modifications were found to significantly alter DNA binding and NR transcriptional activity<sup>45</sup>. Arg91, located at the upstream DBD, protrudes into

the HNF4- $\alpha$  convergence center, similarly to what was observed for Arg153 in the DBD of PPAR $\gamma$ . In PPAR $\gamma$ , Arg153 protrudes towards loop 9–10, forming a salt bridge that was found to have a very strong betweenness in the weighted network. Additionally, the strong interconnectivity between RXR $\alpha$  Glu434 (H11) and PPAR $\gamma$  Tyr477 (12), mediated by solvent molecules, could be important for the asymmetric permissiveness observed in the PPAR $\gamma$ -RXR $\alpha$  heterodimers, by which RXR $\alpha$  ligands can activate PPAR $\gamma$ <sup>25</sup>. Our results thus support the idea of important ‘convergence centers’ for propagating allosteric information, as in the compact folds reported by Chandra *et al.*<sup>44,45</sup>. This interpretation is apparently inconsistent with an alternative ‘open’ architecture for the intact complex proposed on the basis of low-resolution structural methods<sup>59</sup>. A thorough discussion about the architecture of PPAR-RXR and other full-length NRs is found in Rastinejad *et al.*<sup>60</sup>.

The upper half of the PPAR $\gamma$  LBD appeared as the central community in the complex network, consistently with its central position in the asymmetric quaternary organization reported by Chandra *et al.*<sup>44</sup>. While the RXR $\alpha$ -hinge appeared as part of the PPAR $\gamma$ -DBD community, PPAR $\gamma$ -hinge appeared as an independent and poorly connected community. It is thus likely that PPAR $\gamma$ -hinge has evolved as a highly frustrated region that easily adapts its length to buffer the impact of DNA bending modes on the heterodimer stability. This view stands in contrast with previous hypotheses of hinges working as ‘rigid spacers’<sup>46,47</sup> that have been more recently denied<sup>60</sup>, and in agreement with the suggestion that the RXR $\alpha$ -hinge plays an important functional role modulating the relative distance between DBDs and LBDs in the RXR $\alpha$ -LXR $\beta$  complex<sup>48</sup>.

## Conclusions

We have elucidated the allosteric pathways in DNA-nuclear receptor (NR) complexes, based on the first X-ray structure of a full-length DNA/NR complex. The analysis of correlated motions in the DNA/PPAR $\gamma$ -RXR $\alpha$  complex and in the isolated PPAR $\gamma$ -LBD monomer reveals highly correlated motions between distant mobile regions of the complex, with the most prominent inter-domain correlations found at the conformationally flexible hinges,  $\Omega$ -loops and helices that participate in dimerization contacts (H9/10) of PPAR $\gamma$ -DBD. We found that the oligomerization state affects the correlated motions within the PPAR $\gamma$ -LBD. The  $\Omega$ -region motions involved in inter-domain correlations also affect those of helix-12, i.e. the target of PPAR $\gamma$ 's full agonists. Remarkably, such concerted motions are absent in the PPAR $\gamma$ -LBD monomer.

The reported analysis provides a rigorous interpretation of experimental observations from mutagenesis studies and suggests other amino acid residues expected to be crucial for PPAR $\gamma$ 's transcriptional activities. The principal component analysis of MD trajectories supports the central role played by highly correlated regions during allosteric modulation of DNA/NR complexes. In fact, the analysis of essential dynamics indicates that, upon formation of the NR-DNA complex, residual DNA motions (with bending angle up to 30 degrees) translate the PPAR $\gamma$ -DBD/CTE domain and affect the dynamics of distant regions, including the  $\Omega$ -loop in the LBD dimers. In response to DNA distortions, NR hinges change in shape, functioning as ‘molecular springs’, while the RXR $\alpha$ -DBD remains relatively rigid and displays only minor correlations with the LBDs. This suggests that the RXR $\alpha$ -DBD anchors the NR heterodimer to the DNA, while the flux of conformational information between DNA and the LBD domains is established through the PPAR $\gamma$ -DBD. These results support the hypothesis that DNA dynamics (with specific promoter sequence) play an active role in transcription activities.

Our residual frustration analysis revealed that highly correlated regions map well with some highly (mobile) frustrated clusters, indicating that flexible regions could undergo interdependent conformational changes associated with allostery. However, these flexible regions are physically separated by minimally frustrated contacts at the folded core of the complex, whose motions are energetically expensive. Our weighted network, based on mutual information of correlated protein motions, provides clear pathways for communication between distant flexible regions. We find evidence that allostery avoids the more rigid core of the complex and makes use of highly frustrated contacts, passing through solvent exposed loops and helices. Our community network analysis also shows that distinct secondary structures communicate through polar amino acid residues, with convergence centers linking different subunits of the NR complex, such as PPAR $\gamma$  9–10 loop and RXR $\alpha$  (H11)-PPAR $\gamma$  (H12). Notably, we found that the PPAR $\gamma$ -hinge, despite being the most frustrated region of the complex, it is not involved in any important allosteric pathway and consequently remains as an isolated community in the complex network.

We conclude that sequence-dependent DNA bending motions are correlated with the ligand binding domains of the nuclear receptor by translation of one DNA binding domain (PPAR $\gamma$ -DBD). Allosteric pathways involve highly frustrated clusters and specific polar amino acid residues. Future experiments might provide further characterization supporting the reported insights on allosteric pathways in DNA-NR complexes, based on the generalized correlation analysis, essential dynamics, residual local frustration analysis, protein network theory and community network analysis. These methodologies already provide valuable insights on structure/function relations sites that could lead to new strategies in the development of allosteric drugs that target protein-DNA complexes, especially considering that no such drugs have yet been uncovered to date<sup>16,44</sup>. These methods are expected to be useful, in general, for other systems where allosteric processes are critical, including allostery in prokaryotic immune systems with regularly interspaced short repetitions of base sequence (CRISPR-Cas9), currently emerging as a revolutionary tool in structural and molecular biology<sup>61</sup>.

## Methods

**Molecular Dynamics Simulations.** Four different systems were simulated: full-length complex, PPAR $\gamma$ -RXR $\alpha$  LBD heterodimer, PPAR $\gamma$ -LBD as a monomer and apo-HRE. The initial coordinates were taken from crystallographic structures deposited in the Protein Data Bank with the corresponding entries: 3ZDY (full-length complex)<sup>44</sup> and 1FM6 (heterodimer and monomer)<sup>32</sup>. The initial structure of apo-HRE was also taken from PDB 3ZDY after removal of the non-nucleic parts. The  $\Omega$ -loops (residues 242 to 265 in PPAR $\gamma$  and 260 to 275 in RXR $\alpha$ ) – which were missing in the full-length structure – were taken from 1FM6 after an alignment of the

structures had been performed with LovoAlign<sup>62</sup>. Hydrogen atoms were added and protonation states of ionizable groups were estimated using the server H++<sup>63</sup> at pH 7.

Complete simulation boxes were built with VMD<sup>64</sup>. Structures were solvated with TIP3P water molecules<sup>65</sup> within rectangular boxes such that the solute molecules (protein, DNA and ligands) were at least 20 Å from the box boundaries. Sodium and chloride ions were added in order to make the systems electrically neutral at physiological concentration (0.154 mol/L). The ions were placed at favorable positions as determined by electrostatic potential calculations with MeadIonize VMD plugin. Complete systems consisted of approximately 160,000 atoms (full-length complex), 85,000 atoms (LBD heterodimer), 60,000 atoms (PPAR $\gamma$ -LBD monomer) and 40,000 (apo-HRE).

Simulations were performed with NAMD by applying periodic boundary conditions and an integration time-step of 2.0 fs. All bonds involving non-polar hydrogens were constrained at their equilibrium length using the SETTLE algorithm, as implemented in the NAMD code<sup>66</sup>. Pressure (1 atm) and temperature (310 K) were controlled using the Langevin/Nosé-Hoover barostat and the Langevin thermostat, respectively.

CHARMM parameters<sup>67</sup> were applied for protein, DNA and 9-*cis*-Retinoic Acid, while Rosiglitazone parameters were obtained from Hansson *et al.*<sup>68</sup>. A 12 Å cutoff with smooth switching function starting at 10 Å was used for van der Waals interactions and Particle Mesh Ewald method was applied to evaluate electrostatic interactions, as implemented in NAMD.

Minimization and equilibration procedures consisted of i) 2000 gradient conjugate minimization steps and ii) 200 ps of molecular dynamics simulations, keeping the solute molecules fixed (proteins, DNA and ligands). Thereafter, these procedures were repeated keeping only the protein  $\alpha$ -carbons and the DNA backbone fixed. As a last equilibration step, a 2 ns simulation without restraints was performed. After these procedures, 3 independent production trajectories were generated for each system (full-length complex, PPAR $\gamma$ -RXR $\alpha$  LBD heterodimer, PPAR $\gamma$ -LBD as a monomer and apo-HRE), lasting 40 ns each. Here, we assume that the motions in the first 40 ns are those crucial for determining the allosteric mechanism. This assumption is corroborated by the converge study reported in the Supporting Information (Fig. S1), based on longer MD simulations involving: i) simulation time extension of the three original trajectories from 40 ns to 120 ns; and ii) addition of other three independent simulations of 100 ns; for a total simulation time of 0.66 microseconds.

**Mobility and conformational analysis.** Protein mobility analyses (RMSD and RMSF) were performed with a home-made analysis suite. DNA conformational analysis was performed with an adapted version of program CURVES+<sup>69</sup>. DNA bending histograms were generated with Canal<sup>69</sup>. Trajectory visualization and pictures were made with VMD<sup>64</sup>.

**Generalized Correlation Analysis.** For the analysis of correlated motions, we used the method of generalized correlations (GC), proposed by Lange *et al.*<sup>70</sup>. In comparison with the more traditional Pearson coefficients, GC analysis has the advantages of i) being independent of the relative orientation of the atomic fluctuations and ii) being able to capture non-linear correlations. GC analysis is based on the statistical concept stating that two random variables can only be considered independent if their joint probability distribution,  $p(x_i, x_j)$ , equals the product of their marginal distributions,  $p(x_i) \cdot p(x_j)$ . If the values adopted by  $x_i$  somehow restrict the range of values accessible to  $x_j$ , then the joint probability is smaller than  $p(x_i) \cdot p(x_j)$ . Such deviation thus reflects the degree of correlation between  $x_i$  and  $x_j$  and is called mutual information (MI), defined as:

$$MI [x_i, x_j] = \iint p(x_i, x_j) \ln \frac{p(x_i, x_j)}{p(x_i) \cdot p(x_j)} dx_i dx_j \quad (2)$$

MI is closely related to the concept of Shannon entropy (Equation 3), which states that the expected information content of a discrete random variable  $x$ , having a probability distribution  $p(x)$  corresponds to:

$$H [x] = \int p(x) \ln p(x) dx \quad (3)$$

Therefore, the mutual information associated to random variables  $x_i$  and  $x_j$  can be computed as in Equation 4, where  $H[x_i]$  and  $H[x_j]$  correspond to marginal entropies, and  $H[x_i, x_j]$  corresponds to the joint entropy, which can be estimated with different methods. In this work, generalized correlations were computed with the *g\_correlation* tool<sup>70</sup> built to work within GROMACS 3.3<sup>71</sup>. The *g\_correlation* tool estimates marginal ( $H[x_i]$  and  $H[x_j]$ ) and joint ( $H[x_i, x_j]$ ) entropies by means of the *k*-nearest neighbor distances algorithm<sup>72</sup> applied to the atomic positions fluctuations from MD simulations.

$$MI [x_i, x_j] = H[x_i] + H[x_j] - H[x_i, x_j] \quad (4)$$

Because MI varies from 0 to  $+\infty$ , the GC coefficients defined as in Equation 5 provide more intuitive values ranging from 0 (independent variables) to 1 (fully correlated variables)

$$GC_{ij} [x_i, x_j] = \{1 - e^{-2MI[x_i, x_j]/d}\}^{-1/2} \quad (5)$$

where  $d$  is the dimensionality of the variable  $x$ .

A complete convergence study of the correlated motions in the 40 ns time window and a comparison with longer MD simulations is provided in the Supplementary Information (Fig. S1).

**Correlation Score Function.** GC coefficients were also used to build a correlation score function that is a measure of both the number and the intensity of correlations displayed by each residue, defined as:

$$\text{Correlation Score } (i) = \sum_j^{N-1} GC_{ij} \quad (6)$$

By tracking the residues that display large correlation scores, such function helps to point out the most important regions to orchestrate correlated motions. The correlation score function can be filtered to include only intra- or inter-domain correlations. Additionally, an *intensity cutoff* parameter can be used to filter only the most intense correlations. Once computed for every residue in the structure, the correlation score can also be accumulated over a group of residues belonging to the same domain or displaying the same side chains.

**Principal Component Analysis and Free Energy Landscapes.** Essential motions of the system were obtained by Principal Component Analysis (PCA). PCA consists in diagonalizing a variance-covariance matrix of the system  $3N$  atomic positional fluctuations in order to obtain a new set of coordinates (eigenvectors) to describe the system motions. PCA allows decomposing the total motion described by the system into several independent (orthogonal) motions of varying time scales and amplitudes. Each eigenvector (or principal component, PC) has an associated eigenvalue corresponding to the mean square fluctuation contained in the system's trajectory projected along that eigenvector. By sorting the eigenvectors according to their eigenvalues, the first principal component (PC1) corresponds to the system's highest amplitude motion, generally approximated as its 'essential dynamics'. In this work, PCA was performed with a home-made program.

PCA analyses showed in Figs 5 and 10 have been performed for a single MD trajectory of 40 ns. To generate the energy landscapes shown in Fig. 10, we performed independent PCA analysis on selected regions of the complex and computed the corresponding free-energy surfaces according to Equation 7, where  $k_B$  is Boltzmann constant,  $T$  is the temperature,  $P(q)$  is an estimate of the probability along the variable  $q$  and  $P_{MAX}(q)$  is the probability of the most probable state.  $P(q)$  and  $P_{MAX}(q)$  were obtained from histograms of the MD trajectories projected along PC1.

$$\Delta G = -k_B T \ln \frac{P(q)}{P_{MAX}(q)} \quad (7)$$

**Residual Local Frustration Analysis.** Residual Local Frustration is essentially a measure of how 'comfortable' a residue is in terms of the energetics of interaction with its neighbor residues in the native structure of a protein or macromolecular complex<sup>13</sup>. To quantify local frustration in the PPAR $\gamma$ -RXR $\alpha$  complex, we submitted the pdb structure (3DZY) to the frustratometer web server (<http://www.frustratometer.tk/>). The frustratometer algorithm quantifies local frustration by applying conformational changes (or mutations) in the structure and analyzing how such changes affect the interaction energies. Residues that are systematically destabilized by random changes in their vicinities are considered minimally frustrated, while those that are systematically stabilized are considered highly frustrated<sup>56</sup>.

**Weighted Protein-Network, Shortest Pathways and Community Analyses.** The communication network is defined as a set of nodes (amino acid residues), connected by edges (residue pair connections) whose length is weighted using the generalized correlation coefficients, GC. Two nodes are considered connected if the heavy atoms are within a *distance cutoff* (5.5 Å) for at least a *frame cutoff* (75% of the MD frames analyzed). The choice of the cutoff parameters is justified at the end of this section. The edge lengths in the 'weighted protein-network' are calculated as  $w_{ij} = -\log(GC_{ij})$ , with  $GC_{ij}$  defined as in Equation 5. Such a weighted graph defines the dynamical network that contains information paths and critical edges/nodes that are crucial for communication within the complex. As indicated in the "Generalized Correlation Analysis" paragraph, the generalized correlation coefficients have been obtained as averages of three independent 40 ns simulations.

Within the weighted protein-graph, the 'shortest pathways' connecting any pair of residues have been calculated using the Floyd-Warshall algorithm<sup>73</sup>. This algorithm sums up the *lengths* ( $w_{ij}$ ) of all edges involved in different paths of nodes connecting two distant residues and identifies the pathway displaying the shortest total length. For specific residue pairs showing large long-range correlations (see Fig. 8) we have calculated all possible pathways (not only the shortest) within the weighted protein-network. The communication pathways could be very close in length to the shortest pathway (sub-optimal pathways), where the pathway length is defined as the sum of the edge lengths involved in that pathway. This likely happens when the shortest pathways involve highly correlated residues within the same or nearby secondary structures. For simplicity, we consider as sub-optimal pathways only those whose lengths are not larger than 2% of the shortest pathway length.

In the weighted network, there are groups of residues ('communities') within which the connections are dense but between which they are sparse. These local substructures can be obtained using the Girvan-Newman algorithm<sup>74</sup>, a divisive algorithm that is essentially based on the use of the *edge betweenness* as partitioning criterion. The edge betweenness measures signal traveling through a network and is defined as the number of shortest pathways that cross an edge<sup>74</sup>, providing a parameter that favors edges that interconnect communities and disfavors edges that lie within communities. The edges with the highest betweenness connect many pairs of nodes and form the link between different communities. High edge betweenness also associate with pairs of residues that are important for the communication flow within the protein-network.

The Girvan-Newman algorithm is an iterative procedure in which the edge with the highest betweenness is removed from the network and the betweenness of the remaining edges is recalculated, with communities being

progressively isolated up to the point when each node will represent a community. The best division network can be determined using the modularity parameter<sup>74</sup> in such way that nodes within a community are highly intra-connected while different communities are poorly inter-connected through few critical edges. The modularity values fall in the range from 0 to 1, with larger values indicating higher community structure quality. The optimum community structure obtained for the full-length complex (see Fig. 9) has modularity equal to 0.67, in agreement with standard modularity values found in 3D structure of proteins (0.4–0.7)<sup>74</sup>.

The choice of the *distance* and *frames cutoff* parameters is made using a criterion that guarantees convergence of the optimal community network, i.e. the Community Repartition Difference (CRD) defined as

$$CRD(c_1, c_2) = 1 - \frac{\sum_{n_i, n_j} z(n_i, n_j, c_1) z(n_i, n_j, c_2)}{\sum_{n_i, n_j} z(n_i, n_j, c_1)} \quad (8)$$

where  $z(n_i, n_j, c_i)$  is 1 if nodes  $n_i$  and  $n_j$  belong to the same community in a given partition  $c_i$  (or community structure) and 0 otherwise. The CRD provides a normalized count of pairs that are grouped together in two community structures, providing a good estimate of the similarities between different network partitions, as in the case of community structures obtained with different cutoff values.

## References

- Lefstin, J. A. & Yamamoto, K. R. Allosteric effects of DNA on transcriptional regulators. *Nature* **392**, 885–888 (1998).
- Motlagh, H. N., Wrabl, J. O., Li, J. & Hilser, V. J. The ensemble nature of allostery. *Nature* **508**, 331–339 (2014).
- Conn, P. J., Christopoulos, A. & Lindsley, C. W. Allosteric modulators of gpcrs: A novel approach for the treatment of cns disorders. *Nat. Rev. Drug Discovery* **8**, 41–54 (2009).
- Fidelak, J. *et al.* Dynamic correlation networks in human peroxisome proliferator-activated receptor-gamma nuclear receptor protein. *Eur. Biophys. J.* **39**, 1503–1512 (2010).
- Mino-Galaz, G. A. Allosteric communication pathways and thermal rectification in pdz-2 protein: A computational study. *J. Phys. Chem. B* **119**, 6179–6189 (2015).
- Sethi, A., Eargle, J., Black, A. A. & Luthey-Schulten, Z. Dynamical networks in trna: Protein complexes. *Proc. Natl. Acad. Sci. USA* **106**, 6620–6625 (2009).
- Manley, G., Rivalta, I. & Loria, J. P. Solution nmr and computational methods for understanding protein allostery. *J. Phys. Chem. B* **117**, 3063–3073 (2013).
- Venäläinen, T., Molnár, F., Oostenbrink, C., Carlberg, C. & Peräkylä, M. Molecular mechanism of allosteric communication in the human ppar $\alpha$ -rxr $\alpha$  heterodimer. *Proteins: Struct. Funct. Bioinf.* **78**, 873–887 (2010).
- Changeux, J.-P. Allostery and the monod-wyman-changeux model after 50 years. *Annu. Rev. Biophys.* **41**, 103–133 (2012).
- Christopoulos, A. *et al.* International union of basic and clinical pharmacology. Xc. Multisite pharmacology: Recommendations for the nomenclature of receptor allostery and allosteric ligands. *Pharmacol. Rev.* **66**, 918–947 (2014).
- Pontiggia, F. *et al.* Free energy landscape of activation in a signalling protein at atomic resolution. *Nature communications* **6**, 7284 (2015).
- Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **254**, 1598–1603 (1991).
- Ferreiro, D. U., Hegler, J. A., Komives, E. A. & Wolynes, P. G. Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. USA* **104**, 19819–19824 (2007).
- McKenna, N. J., Lanz, R. B. & O'Malley, B. W. Nuclear receptor coregulators: Cellular and molecular biology. *Endocr. Rev.* **20**, 321–344 (1999).
- Gronemeyer, H., Gustafsson, J. A. & Laudet, V. Principles for modulation of the nuclear receptor superfamily. *Nat. Rev. Drug Discovery* **3**, 950–964 (2004).
- Huang, P. X., Chandra, V. & Rastinejad, F. Structural overview of the nuclear receptor superfamily: Insights into physiology and therapeutics. *Annu. Rev. Physiol.* **72**, 247–272 (2010).
- Mangelsdorf, D. J. & Evans, R. M. The rxr heterodimers and orphan receptors. *Cell* **83**, 841–850 (1995).
- Steinmetz, A. C. U., Renaud, J. P. & Moras, D. Binding of ligands and activation of transcription by nuclear receptors. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 329–359 (2001).
- Zhang, J. *et al.* DNA binding alters coactivator interaction surfaces of the intact vdr-rxr complex. *Nat. Struct. Mol. Biol.* **18**, 556–U172 (2011).
- Hall, J. M., McDonnell, D. P. & Korach, K. S. Allosteric regulation of estrogen receptor structure, function, and coactivator recruitment by different estrogen response elements. *Mol. Endocrinol.* **16**, 469–486 (2002).
- Meijsing, S. H. *et al.* DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science* **324**, 407–410 (2009).
- Kurokawa, R. *et al.* Regulation of retinoid signaling by receptor polarity and allosteric control of ligand-binding. *Nature* **371**, 528–531 (1994).
- Kliwer, S. A., Umesono, K., Noonan, D. J., Heyman, R. A. & Evans, R. M. Convergence of 9-cis retinoic acid and peroxisome proliferator signaling pathways through heterodimer formation of their receptors. *Nature* **358**, 771–774 (1992).
- Desvergne, B. & Wahli, W. Peroxisome proliferator-activated receptors: Nuclear control of metabolism. *Endocr. Rev.* **20**, 649–688 (1999).
- Schulman, I. G., Shao, G. & Heyman, R. A. Transactivation by retinoid x receptor peroxisome proliferator-activated receptor gamma (ppar gamma) heterodimers: Intermolecular synergy requires only the ppar gamma hormone-dependent activation function. *Mol. Cell. Biol.* **18**, 3483–3494 (1998).
- Willson, T. M., Brown, P. J., Sternbach, D. D. & Henke, B. R. The ppar $\alpha$ : From orphan receptors to drug discovery. *J. Med. Chem.* **43**, 527–550 (2000).
- Lehrke, M. & Lazar, M. A. The many faces of ppar gamma. *Cell* **123**, 993–999 (2005).
- Ahmadian, M. *et al.* Ppar gamma signaling and metabolism: The good, the bad and the future. *Nat. Med.* **19**, 557–566 (2013).
- Nolte, R. T. *et al.* Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma. *Nature* **395**, 137–143 (1998).
- Xu, H. E. *et al.* Molecular recognition of fatty acids by peroxisome proliferator-activated receptors. *Mol. Cell* **3**, 397–403 (1999).
- Liberato, M. V. *et al.* Medium chain fatty acids are selective peroxisome proliferator activated receptor (ppar) gamma activators and pan-ppar partial agonists. *PLoS One* **7** (2012).
- Gampe, R. T. *et al.* Asymmetry in the ppar gamma/rxr alpha crystal structure reveals the molecular basis of heterodimerization among nuclear receptors. *Mol. Cell* **5**, 545–555 (2000).
- Puhl, A. C. *et al.* Mode of peroxisome proliferator-activated receptor gamma activation by luteolin. *Mol. Pharmacol.* **81**, 788–799 (2012).

34. Bernardes, A. *et al.* Molecular mechanism of peroxisome proliferator-activated receptor alpha activation by wy14643: A new mode of ligand recognition and receptor stabilization. *J. Mol. Biol.* **425**, 2878–2893 (2013).
35. Kallenberger, B. C., Love, J. D., Chatterjee, V. K. K. & Schwabe, J. W. R. A dynamic mechanism of nuclear receptor activation and its perturbation in a human disease. *Nat. Struct. Biol.* **10**, 136–140 (2003).
36. Oberfield, J. L. *et al.* A peroxisome proliferator-activated receptor gamma ligand inhibits adipocyte differentiation. *Proc. Natl. Acad. Sci. USA* **96**, 6102–6106 (1999).
37. Ostberg, T. *et al.* A new class of peroxisome proliferator-activated receptor agonists with a novel binding epitope shows antidiabetic effects. *J. Biol. Chem.* **279**, 41124–41130 (2004).
38. Burgermeister, E. *et al.* A novel partial agonist of peroxisome proliferator-activated receptor-gamma (ppar gamma) recruits ppar gamma-coactivator-1 alpha, prevents triglyceride accumulation, and potentiates insulin signaling *in vitro*. *Mol. Endocrinol.* **20**, 809–830 (2006).
39. Pochetti, G. *et al.* Insights into the mechanism of partial agonism - crystal structures of the peroxisome proliferator-activated receptor gamma ligand-binding domain in the complex with two enantiomeric ligands. *J. Biol. Chem.* **282**, 17314–17324 (2007).
40. Li, Y. *et al.* T2384, a novel antidiabetic agent with unique peroxisome proliferator-activated receptor gamma binding properties. *J. Biol. Chem.* **283**, 9168–9176 (2008).
41. Motani, A. *et al.* Int131: A selective modulator of ppar gamma. *J. Mol. Biol.* **386**, 1301–1311 (2009).
42. Bruning, J. B. *et al.* Partial agonists activate ppar gamma using a helix 12 independent mechanism. *Structure* **15**, 1258–1271 (2007).
43. Waku, T. *et al.* Structural insight into ppar gamma activation through covalent modification with endogenous fatty acids. *J. Mol. Biol.* **385**, 188–199 (2009).
44. Chandra, V. *et al.* Structure of the intact ppar-gamma-rxr-alpha nuclear receptor complex on DNA. *Nature* **456**, 350–U333 (2008).
45. Chandra, V. *et al.* Multidomain integration in the structure of the hnf-4 alpha nuclear receptor complex. *Nature* **495**, 394–398 (2013).
46. Rochel, N. *et al.* Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings. *Nat. Struct. Mol. Biol.* **18**, 564–U207 (2011).
47. Orlov, I., Rochel, N., Moras, D. & Klaholz, B. P. Structure of the full human rxr/vdr nuclear receptor heterodimer complex with its dr3 target DNA. *EMBO J.* **31**, 291–300 (2012).
48. Lou, X. H. *et al.* Structure of the retinoid x receptor alpha-liver x receptor beta (rxr alpha-lxr beta) heterodimer on DNA. *Nat. Struct. Mol. Biol.* **21**, 277–281 (2014).
49. Rivalta, I. *et al.* Allosteric pathways in imidazole glycerol phosphate synthase. *Proc. Natl. Acad. Sci. USA* **109**, E1428–1436 (2012).
50. VanWart, A. T., Eargle, J., Luthey-Schulten, Z. & Amaro, R. E. Exploring residue component contributions to dynamical network models of allostery. *J. Chem. Theory Comput.* **8**, 2949–2961 (2012).
51. Ribeiro, A. A. S. T. & Ortiz, V. Determination of signaling pathways in proteins through network theory: Importance of the topology. *J. Chem. Theory Comput.* **10**, 1762–1769 (2014).
52. Eriksson, M. A. L., Berglund, H., Hard, T. & Nilsson, L. A comparison of <sup>15</sup>N nmr relaxation measurements with a molecular-dynamics simulation - backbone dynamics of the glucocorticoid receptor DNA-binding domain. *Proteins: Struct. Funct. Genet.* **17**, 375–390 (1993).
53. Berglund, H., Kovacs, H., Dahlmanwright, K., Gustafsson, J. A. & Hard, T. Backbone dynamics of the glucocorticoid receptor DNA-binding domain. *Biochemistry* **31**, 12001–12011 (1992).
54. Genest, D. *et al.* Ligand-escape pathways from the ligand-binding domain of ppar gamma receptor as probed by molecular dynamics simulations. *Eur. Biophys. J.* **37**, 369–379 (2008).
55. Barroso, I. *et al.* Dominant negative mutations in human ppar gamma associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature* **402**, 880–883 (1999).
56. Jenik, M. *et al.* Protein frustratometer: A tool to localize energetic frustration in protein molecules. *Nucleic Acids Res.* **40**, W348–W351 (2012).
57. Ferreira, D. U., Hegler, J. A., Komives, E. A. & Wolynes, P. G. On the role of frustration in the energy landscapes of allosteric proteins. *Proc. Natl. Acad. Sci. USA* **108**, 3499–3503 (2011).
58. Martinez, L., Figueira, A. C. M., Webb, P., Polikarpov, I. & Skaf, M. S. Mapping the intramolecular vibrational energy flow in proteins reveals functionally important residues. *J. Phys. Chem Lett.* **2**, 2073–2078 (2011).
59. Osz, J. *et al.* Solution structures of ppar  $\gamma$ 2/rxr  $\alpha$  complexes. *PPAR Research* **2012**, 701412 (2012).
60. Rastinejad, F., Ollendorff, V. & Polikarpov, I. Nuclear receptor full-length architectures: Confronting myth and illusion with high resolution. *Trends Biochem. Sci.* **40**, 16–24 (2015).
61. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with crispr-cas9. *Science* **346**, 1077–1086 (2014).
62. Martinez, L., Andreani, R. & Martinez, J. M. Convergent algorithms for protein structural alignment. *BMC Bioinform.* **8**, doi: 10.1186/1471-2105-8-306 (2007).
63. Myers, J., Grothaus, G., Narayanan, S. & Onufriev, A. A simple clustering algorithm can be accurate enough for use in calculations of pks in macromolecules. *Proteins: Struct. Funct. Bioinf.* **63**, 928–938 (2006).
64. Humphrey, W., Dalke, A. & Schulten, K. Vmd: Visual molecular dynamics. *J. Mol. Graphics Modell.* **14**, 33–38 (1996).
65. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
66. Phillips, J. C. *et al.* Scalable molecular dynamics with namd. *J. Comput. Chem.* **26**, 1781–1802 (2005).
67. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
68. Hansson, A., Souza, P. C. T., Silveira, R. L., Martinez, L. & Skaf, M. S. Charrm force field parameterization of rosiglitazone. *Int. J. Quantum Chem.* **111**, 1346–1354 (2011).
69. Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. & Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **37**, 5917–5929 (2009).
70. Lange, O. F. & Grubmuller, H. Generalized correlation for biomolecular dynamics. *Proteins* **62**, 1053–1061 (2006).
71. Lindahl, E., Hess, B. & Van Der Spoel, D. Gromacs 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* **7**, 306–317 (2001).
72. Kraskov, A., Stogbauer, H. & Grassberger, P. Estimating mutual information. *Phys. Rev. E* **69**, doi: 10.1103/Physreve.69.066138 (2004).
73. Floyd, R. W. Algorithm-97 - shortest path. *Commun. Acm* **5**, 345–345 (1962).
74. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, doi: 10.1103/Physreve.69.026113 (2004).

## Acknowledgements

This work was supported by the São Paulo Research Foundation FAPESP (Grant #2013/08293-7). We would like to thank Prof. Richard Lavery for kindly adapting the Curves + program to handle PDB trajectories. VSB acknowledges supercomputer time from the National Energy Research Scientific Computing Center (NERSC) and support from the National Institute of Health (NIH) grant 1R01GM10621-01A1.



### Author Contributions

C.G.R., R.L.S. and I.R. performed calculations, carried out analysis and interpreted the data. V.S.B. and M.S.S. designed the study. C.G.R. wrote most part of the paper. All authors were involved in organizing the presentation of the results, contributed to the writing and discussed the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Ricci, C. G. *et al.* Allosteric Pathways in the PPAR $\gamma$ -RXR $\alpha$  nuclear receptor complex. *Sci. Rep.* **6**, 19940; doi: 10.1038/srep19940 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>