

Unraveling the allosteric cross-talk between coactivator peptide and ligand binding site in glucocorticoid receptor

Giuseppina La Sala¹, Anders Gunnarsson², Karl Edman², Christian Tyrchan³, Anders Hogner¹ and Andrey I. Frolov¹

1. Medicinal Chemistry, Research and Early Development, Cardiovascular, Renal and Metabolism (CVRM), BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.
2. Discovery Science, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.
3. Medicinal Chemistry, Research and Early Development, Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.

Corresponding authors:

giuseppina.lasala@astrazeneca.com

andrey.frolov@astrazeneca.com

Orcid:

Giuseppina La Sala: 0000-0001-6565-197X

Anders Gunnarsson: 0000-0002-6334-9520

Karl Edman: 0000-0002-2560-2388

Christian Tyrchan: 0000-0002-6470-984X

Anders Hogner: 0000-0003-3823-4534

Andrey I. Frolov: 0000-0001-9801-3253

Abbreviations:

AF-2 = activation function 2

AR = androgen receptor

CPMG = Carr-Purcell-Meiboom-Gill

DEX = dexamethasone

FP = fluorescence polarization

GR = glucocorticoid receptor

GR_{LBD} = glucocorticoid receptor ligand binding domain

HDX-MS = hydrogen-deuterium exchange coupled with mass spectrometry

HMR = hydrogen mass repartitioning

LBD = ligand binding domain

LBP = ligand binding pocket

MR = mineralcorticoid receptor

NHR = nuclear hormone receptor

NR3C = 3-ketosteroid nuclear receptor

PR = progesterone receptor

PRGC1= peroxisome proliferator-activated gamma coactivator 1- α

RMSD = root mean square deviation

RMSF = root mean square fluctuation

SPR = surface plasmon resonance

TIF2 = transcriptional mediators/intermediary factor 2

Keywords: Glucocorticoid Receptor, Peroxisome proliferator-activated gamma coactivator 1- α (PRGC1), Transcriptional mediators/intermediary factor 2 (TIF2), Allostery, Molecular Dynamics Simulations, Drug Discovery.

ABSTRACT

Glucocorticoid receptor (GR) is a nuclear receptor that controls critical biological processes by regulating the transcription of specific genes. There is a known allosteric cross-talk between the ligand and coregulator binding sites within the GR ligand binding domain that is crucial for the control of the functional response. However, the molecular mechanisms underlying such an allosteric control remain elusive. Here, molecular dynamics (MD) simulations, bioinformatic analysis and biophysical measurements are integrated to capture the structural and dynamic features of the allosteric cross-talk within GR. We identified a network of evolutionarily conserved residues that enables the allosteric signal transduction, in agreement with experimental data. MD simulations clarify how such network is dynamically interconnected and offer a mechanistic explanation of how the different peptides affect the intensity of the allosteric signal. This study provides useful insights to elucidate the GR allosteric regulation, ultimately, posing the foundation for designing novel drugs.

INTRODUCTION

Glucocorticoid receptor (GR), a member of the nuclear hormone receptor (NHR) family, is a transcription factor that regulates several physiological processes, such as cell differentiation, apoptosis and metabolism¹. GR agonists, such as the glucocorticoids (GCs) dexamethasone (Dex) and prednisolone, are widely prescribed to patients with inflammatory and autoimmune diseases². However, the chronic use of high-dose GCs is associated with side effects like osteoporosis, diabetes, hypertension and Cushing's syndrome³. For this reason, several efforts are dedicated to the design of new GR modulators able to separate the efficacy from the adverse effects^{4, 5, 6}. To this aim, deciphering molecular mechanisms that enable the GR function is key for the design of new therapeutics.

GR has a modular structure composed of three domains: i) the N-terminal domain, an unstructured portion containing a docking site for coregulator proteins, termed activation-function 1 (AF-1); ii) the DNA binding domain, a region containing zinc finger motifs that bind to specific DNA sequences; iii) the C-terminal ligand binding domain (LBD), containing the ligand binding pocket (LBP) and the activation-function 2 (AF-2), a shallow pocket formed by helices H3, H4 and H12 that interact with the coregulator proteins (**Figure 1**)⁷. In absence of ligands, GR is predominantly located in the cytosol, forming a complex with chaperone proteins. Upon ligand binding, the chaperone complex is partially dissolved, and GR is translocated to the nucleus. Here, GR binds to specific DNA sequences and assembles a range of coregulator proteins to ultimately induce or repress the expression of target genes^{1, 8}. Importantly, the functional response in the cell is finely tuned by the interplay of several factors, such as the ligand's chemical structure and its concentration in the cell, the presence of post-translational modifications and the cellular context^{5, 7, 9}.

Coregulators play a central role in GR signal transduction. Both coactivators and corepressors bind to the AF-2 region through a short amphipathic helical sequence, containing the LXXLL motif in coactivators or (L/I)XX(I/V)I motif in corepressors^{10, 11}. Upon ligand binding, the helix H12 undergoes a significant conformational change that alters the overall shape of the AF-2 docking site. This promotes the recruitment of coactivators or corepressors, resulting in a different biological response^{12, 13, 14, 15}. Insights of the allosteric cross-talk that bridges the LBP with the AF-2 (i.e., LBP \leftrightarrow AF-2) has been revealed by a number of structural and biophysical studies^{9, 12, 13, 16, 17, 18, 19}. In a recent study, Köhler et al. used NMR to identify a network of residues that are involved in the LBP \leftrightarrow AF-2 allosteric crosstalk and demonstrated that different ligands can use these allosteric pathways to fine-tune the GR's functional response. In addition, they showed that different coactivator peptides induce different GR_{LBD} dynamics and affect the intensity of the allosteric signal. Specifically, TIF2 promotes a dynamical exchange between alternate conformations of the helix H12, whereas in the presence of PRGC1 only one conformation of the helix H12 is populated. These findings indicate that helix H12 is a crucial element for the modulation of the AF-2 \rightarrow LBP allosteric signal²⁰.

Although these experimental studies offer important insights for the comprehension of allosteric regulation of GR_{LBD}, several questions still remain open: how are the residues in the LBP \leftrightarrow AF-2 allosteric pathways dynamically interconnected? How do the coregulators' peptide sequences affect the allosteric signal transmission network in GR_{LBD}? To address these questions, we adopted a multidisciplinary approach, integrating microsecond-long MD simulations, bioinformatic analysis and biophysical measurements, to decipher the molecular mechanisms that control the dynamic coupling between the AF-2 and the LBP. We found that evolutionarily

conserved residues in the AF-2 docking site, in the helix H12 and in the core of GR_{LBD} form a dynamically interconnected network that might have a crucial role in the modulation of the allosteric signal propagation. Our study also reveal that the peptides can affect the functionality of such network, explaining why the allosteric response is strongly dependent on the peptide sequence. Overall, our findings help the mechanistic understanding of allosteric regulation in GR and, ultimately, provide useful insights to drug design.

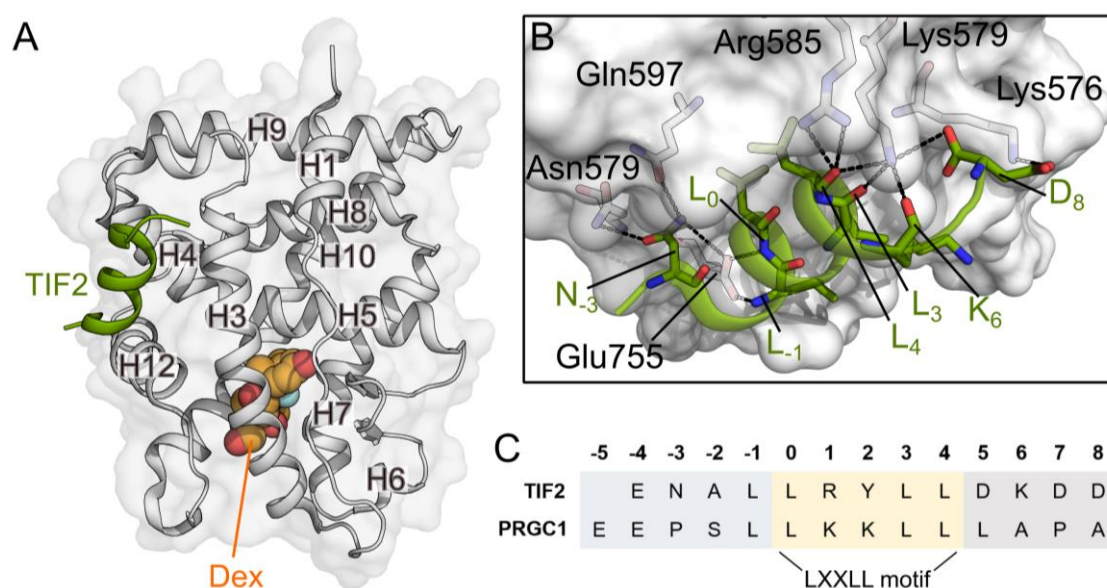


Figure 1. Structure of TIF2/GR/Dex ternary complex. A) Representation of the TIF2/GR/Dex tertiary complex (PDB code: 4UDC). The GR_{LBD} is depicted in grey cartoon and surface, the dexamethasone (Dex) with orange spheres and the TIF2 peptide in green cartoon. B) Close view of the binding of the TIF2 peptide (green cartoon) to the AF-2 surface of GR_{LBD} (grey surface). Residues involved in H-bonds are depicted in sticks, while the H-bonds are shown as dotted lines. C) Sequence alignment of TIF2 and PRGC1 peptides. The LXXLL motif is highlighted in the yellow box.

RESULTS AND DISCUSSION

TIF2 and PRGC1 peptides differently affect the intrinsic dynamics of GR_{LBD}. First, we explored how the two coregulator peptides probed in this study (i.e., TIF2 and PRGC1) alter the global dynamics of GR_{LBD}. To this aim, we examined the intrinsic dynamics of GR_{LBD} by comparing the microsecond-long all-atom classical MD simulation of TIF2/GR/Dex and PRGC1/GR/Dex ternary complexes and assessed the consistency of our outcomes by analyzing five independent replicas for each system.

The root mean square fluctuation (RMSF) analysis indicates that the most flexible regions of the TIF2/GR/Dex complex are the H9-H10-loop, H11-H12-loop and some portions in proximity of the LBP, such as the H2-H3-loop and helix H6 (**Figure 2A**). The same analysis performed on the PRGC1/GR/Dex confirmed the flexibility in both the H9-H10 and H11-H12-loops with comparable magnitude with respect to the TIF2/GR/Dex complex, whereas the region near the LBP is more rigid (**Figure 2A**).

The first three principal components (PCs) of TIF2/GR/Dex complex describe a ‘scissor’ motion, with the H11-H12-loop and the region near the LBP moving in opposite directions and a pivot point in the center of the helices H3 and H10 (**Figure 2B**). Notably, such motion is also found in the second PC of the PRGC1/GR/Dex trajectory, while the first PC describes a swing of the H5-β1 loop (referred to as ‘loop swing’ motion). The third PC of the this system describes a mixture of ‘loop swing’ motion and a ‘pendulum’ motion, where the H11-H12-loop and the region near the LBP move in the same directions (**Figure 2B**). The additional replicas confirmed the presence of the three motions in the first three PCs (**Figure S1B**). Overall the MD simulations show that the ‘scissor’ motion has a slightly higher amplitude in TIF2/GR/Dex complex (**Figure S1B**), in line with the augmented mobility of the ligand binding region observed in the RMSF analysis of TIF2/GR/Dex complex.

Taken together, both the RMSF and the PCA analysis demonstrate that the two peptides alter the dynamics of GR_{LBD} differently and that the overall motility of GR_{LBD} is reduced in presence of the PRGC1 peptide. Importantly, these findings correlate well with recent NMR spectroscopy data that indicates restricted motion of the GR_{LBD} due to a lack of relaxation dispersion in CPMG experiments.²⁰ A similar behavior has been recently observed for the ancestral steroid receptor (ancGR2), a GR-related protein having the 79% sequence identity with the human GR probed in our study. In fact, the combination of HDX-MS experiments and DSF-based thermal denaturation has evidenced that the presence of PRGC1 peptide rigidifies the ancGR2_{LBD} more efficiently than TIF2, independently from the ligand bound to the LBP¹⁶. Notably, the peptides drive the flexibility’s change in a region that overlaps with the portion of the protein important for the ligand binding and unbinding²¹. As such, this might provide a possible explanation of the peptide-dependent variation of ligand kinetics observed in the study performed by Pfaff *et al.*¹⁹.

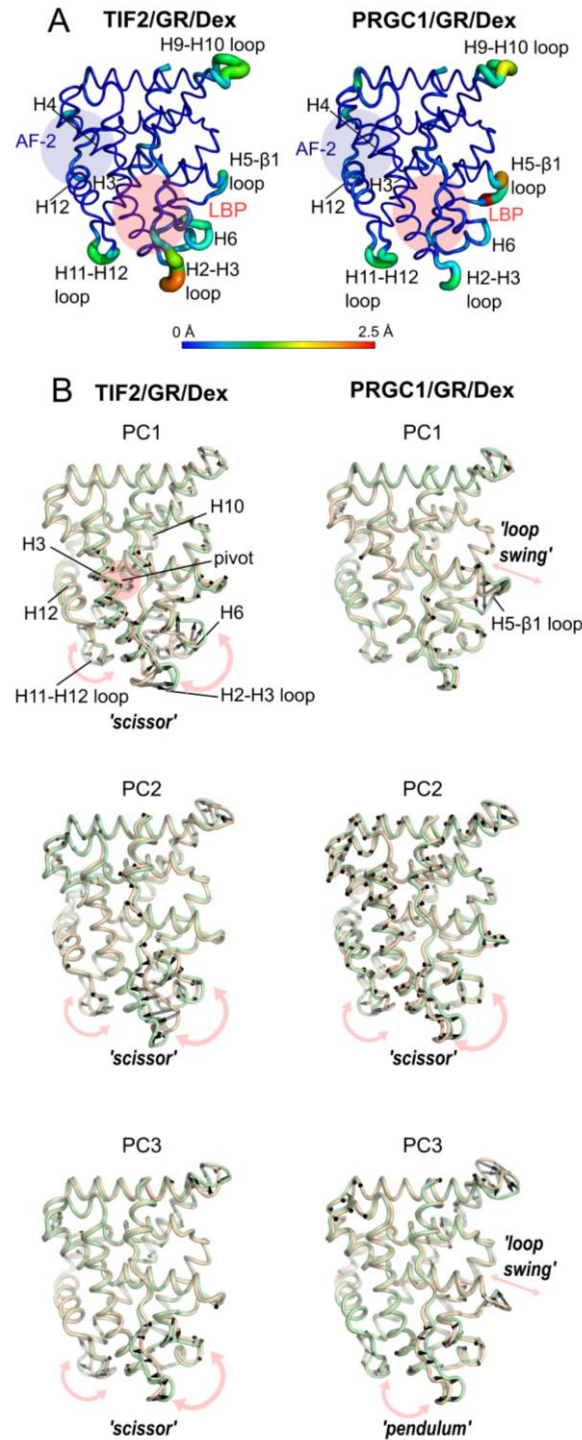


Figure 2. Intrinsic dynamics of TIF2/GR/Dex and PRGC1/GR/Dex complexes. A) The calculated B-factor of TIF2/GR/Dex (left) and PRGC1/GR/Dex (right) is plotted on the GR_{LBD} structure using a putty representation. B) The PC1, PC2 and PC3 of TIF2/GR/Dex trajectory (left) and PRGC1/GR/Dex (right) are represented on GR_{LBD} by back arrows of sizes proportional to the amplitude of motion. The pink arrows highlights the most predominant motion for each PC, which can be a 'scissor', a 'pendulum' or a 'loop swing' motion.

The LBP↔AF-2 allosteric signaling pathway. We conducted further investigation to elucidate how the allosteric signal propagates in between the AF-2 and the LBP. The goals are to determine the relevant residues involved in the allosteric signal transmission, clarify how they are dynamically interconnected and how the allosteric coupling differs in between the TIF2/GR/Dex and PRGC1/GR/Dex complexes. Towards this aim, we applied the network theory, motivated by a number of successful cases where this technique has been used to elucidate the allosteric mechanisms in biomolecular systems^{22, 23, 24, 25}. In this approach, the protein is represented as a network, where the nodes are subunits of the proteins (e.g., the C α carbons) and the edges between two nodes are weighted by their degree of “interdependence”. Although the “interdependence” can be computed in several ways (see reviews²⁶ and ²⁷), in this study the correlated motion values are used²² (see Materials and Methods). Nodes that are highly interconnected (i.e., have highly correlated motions) are clustered together in communities, which communicate with each other through a few paths. The nodes and edges involved in these paths are dubbed “critical” and are considered key for the propagation of the allosteric signals within the network^{22, 24, 28}.

The community analysis showed that both TIF2/GR/Dex and PRGC1/GR/Dex complexes are composed of six communities, as depicted in **Figures 3A** and **3C**. The size and spatial localization of the communities are comparable in the two systems, except for the communities 1 and 3 that together form the LBP. In fact, in TIF2/GR/Dex system the LBP is split in two distinct communities (1 and 3), while in the PRGC1/GR/Dex the LBP is formed almost exclusively by community 1 (**Figures 3A** and **3C**). This indicates that the two peptides cause a different dynamical reorganization of GR_{LBD}, where the PRGC1 promotes a more efficient interconnection among the residues of the LBP. This high interconnection, in turn, likely improves the overall stability of the LBP, in agreement with previous RMSF and PCA analysis. A coarse grained representation of the community network is depicted in **Figures 3B** and **3D**. The spatial assembly of the communities suggests that the allosteric signal between AF-2 and LBP can be transmitted through two possible short paths, namely ‘route 1’ and ‘route 2’ (dark green and purple dashed lines in **Figures 3A-D**). The former route has already been identified by Liu et al.^{16, 18} in the study of a related GR protein (ancGR2) and directly connects community 2 and community 1, passing through the helix H12 and N-term of helix H3. Our analysis revealed a higher edge betweenness (see Materials and Methods) between these two communities in the PRGC1/GR/Dex (**Figures 3B** and **3D**), indicating that in the human GR_{LBD} form, the information in ‘route 1’ flows in a more efficient way when PRGC1 is bound, similarly as observed in ancGR2 protein¹⁶. In their studies, the authors also propose that Leu753, Gly567 and Asn564 (human GR numbering) are important nodes for the signal transmission in ‘route 1’¹⁸. Importantly, these residues are also identified by our network analysis, confirming the reliability of our approach (**Figures 3E** and **3F**).

The ‘route 2’ is a slightly longer path that connects community 2 and community 1, passing through community 6 (**Figures 3A-D**). In contrast to the pathway proposed by Liu et al., this route involves all the structural elements of the AF-2 surface, such as the helix H12, the helix H4 and the C-term of helix H3 and, for this reason, it might also be considered a plausible allosteric signal transmission route. In both TIF2/GR/Dex and PRGC1/GR/Dex systems the two structural modules that form the AF-2 surface (i.e., community 2 and 6) are interconnected via two critical edges located in helix H12 and helix H4 (Ile756 - Gln597 and Ile756 - Leu596 for TIF2/GR/Dex and PRGC1/GR/Dex, respectively. **Figures 3E** and **3F**). From helix H4, a chain of critical nodes located in the H4/H5 kink enables the signal propagation up to the helix H5 located in proximity of the LBP (Leu596 and Trp600 in TIF2/GR/Dex and Ser599, Phe602, Leu603 and Ala607 in PRGC1/GR/Dex. **Figures 3E** and **3F**). Thus, we hypothesize that these critical nodes and nearest neighbors are important for the propagation of the allosteric signal

in GR_{LBD}. As a support of this hypothesis, a recent study reported that mutations of four amino acids that correspond to or are nearest neighbors to the critical nodes found in our analysis (i.e., Met752Ile, Phe602Ser, Tyr598Gln and Met604Thr) have an impact on the stabilization of either the agonist or antagonist conformation of GR.¹⁴ To further support this hypothesis, a bioinformatic analysis was conducted to assess the evolutionary conservation and co-evolution of the critical nodes within the 3-ketosteroid nuclear receptor (NR3C) family. Indeed, both evolutionarily conserved and co-evolving residues are known to maintain the structural integrity and to regulate the function of the proteins^{29, 30, 31, 32, 33} and they might coincide with the critical nodes identified in the MD-generated network analysis. **Figure S2A** shows an high conservation of the residues forming the ‘route 2’. Also, we found that the critical nodes in ‘route 2’ are part of a cluster of co-evolving residues (**Figure S2B**). Taken together these analysis indicates that critical nodes identified in our MD simulations are important for GR’s structure and function, supporting our hypothesis that they are key to conduct the signal between the AF-2 and the LBP.

To further validate our approach, we related our findings to a recent NMR experiment aimed at probing the dynamical allosteric network of residues in the LBD of TIF2/GR/Dex ternary complex²⁰. Often, the critical nodes and edges are experimentally validated through NMR spectroscopy, demonstrating that computational methods and NMR studies can be used synergistically to clarify the allosteric mechanisms in proteins^{23, 28, 34}. By means of ¹³C and ¹⁵N CPMG relaxation dispersion experiments, Köhler *et al.* identified a set of residues that undergo to a conformational exchange between alternate configurations on a millisecond time scale. These residues connect the LBP, the AF-2 and tau2 site and their slow concerted dynamical motion was thought to make a key contribution to mediating the allosteric regulation in GR_{LBD}. To compare these experimental findings with the outcomes from our MD simulations, we mapped on GR_{LBD} the critical nodes identified by the dynamical network analysis that show also ¹³C and ¹⁵N CPMG relaxation dispersion signal. Nine critical nodes exactly match the NMR experiments (**Figure 3G**, following referred with *), while eight critical nodes are shifted of either +/- 1 (**Figure 3H**, following referred with •) or +/- 2 (**Figure 3I**, following referred with ••) positions respect to NMR experiments. Matching residues were identified in the tau2 site (Glu540*, Pro541*, Glu542* and Val543*) and in the C-terminal half of helix H3 (Trp577* and Ala574*), a region of the protein proposed to be involved in the allosteric signaling transmission linking the AF-2 to tau2 site. Moreover, matching residues were identified in the helices H4, H5 and in the N-terminal half of H12 (Leu 596*, Gln597*, Trp600*, Leu753* Ile756* and Ile757*). Such a network of amino acids is proposed to be involved in the transmission of AF-2 ↔ LBP allosteric signal. Overall, this analysis highlights a remarkable consistency between the NMR experiments and MD simulations, evidencing the reliability of our computational approaches. Despite our MD simulations only accessing dynamical motions on relatively short timescales compared to the CPMG relaxation dispersion experiments, it is interesting that our sampling can qualitatively recognize the same ‘hotspot’ regions detected in these experiments. In addition, our network analysis shows that the allosteric pathway is similar between the TIF2/GR/Dex and PRGC1/GR/Dex, inasmuch as the critical nodes in ‘route 2’ of the resulting networks mostly overlap (**Figures 3E and 3F**). This is largely consistent with the NMR experiments which demonstrate a common signaling pathway between the two systems, albeit with a diverse intensity in the signal propagation.²⁰

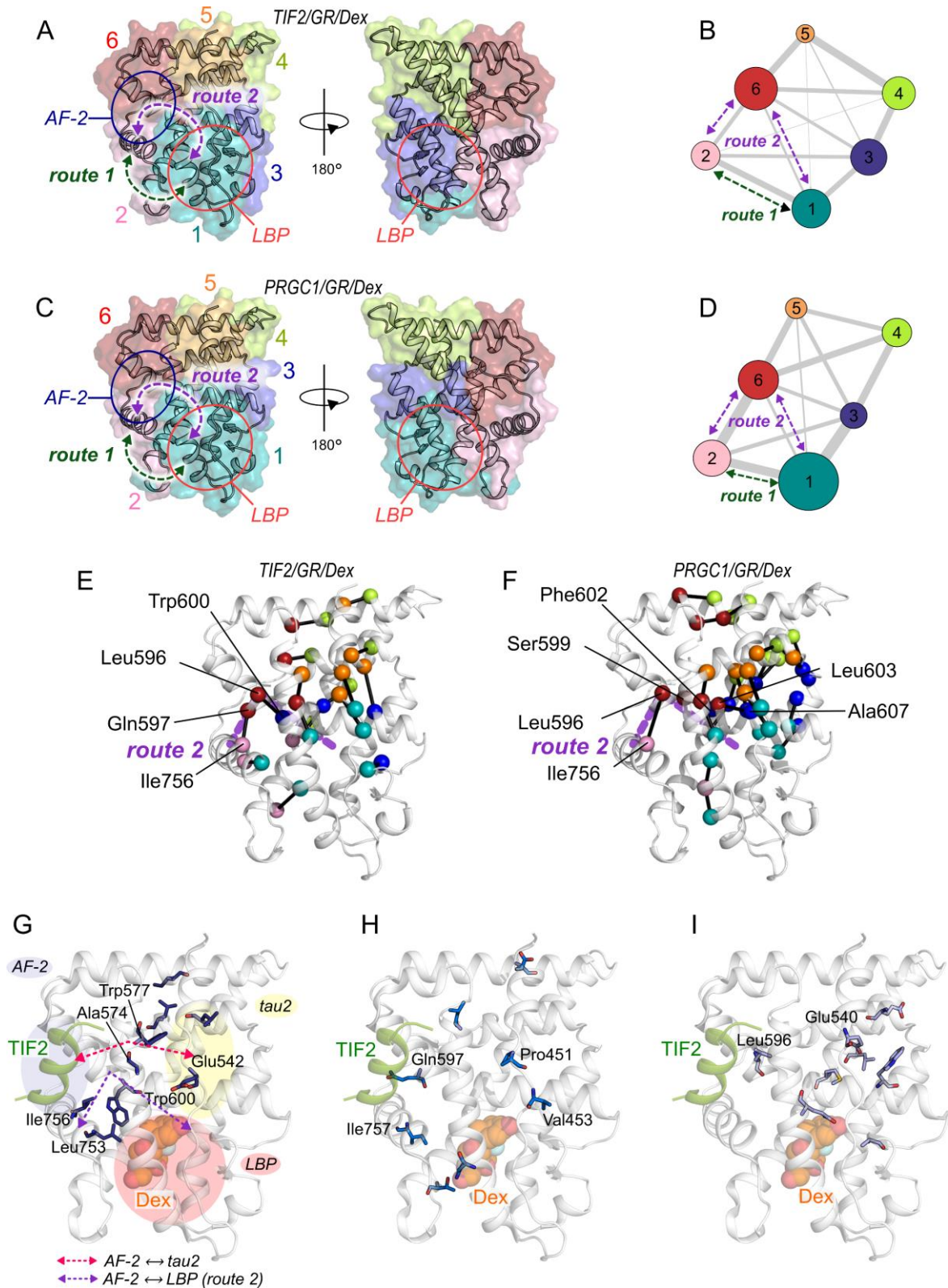


Figure 3. Representation of communities and critical nodes. A, C) The communities identified by the network analysis are depicted as surfaces with different colors for both TIF2/GR/Dex (A) and PRGC1/GR/Dex (C) systems. The ligand binding pocket (LBP) and the AF-2 surface are circled in red and blue, respectively. The ‘route 1’ and the ‘route 2’ are depicted as green and purple dashed arrows, respectively. B, D) Communities’ network representations for TIF2/GR/Dex (B) and PRGC1/GR/Dex (D). Each node represents a community, whose size is proportional to the number of residues within the community. The width of the edges is proportional to the number of shortest paths passing through this connection (see Materials and Methods). The thicker the line, the stronger the communication between the two communities. The ‘route 1’ and the ‘route 2’ are depicted as green and purple dashed arrows, respectively. E, F) The critical nodes of the network of both TIF2/GR/Dex (E) and PRGC1/GR/Dex

(F) systems are represented as spheres, whereas the critical edges are represented as black lines. The ‘route 1’ is represented as purple dashed lines. G, H and I) Critical nodes found in the dynamical network analysis of TIF2/GR/Dex that exactly match the NMR CPMG relaxation dispersion experiments (G), or shifted with respect of it of +/- 1 position (H) or +/- 2 positions (I) are represented as sticks. The crimson and the violet dashed arrows highlights the AF-2 \leftrightarrow tau2 and the AF-2 \leftrightarrow LBP (or ‘route 2’) allosteric pathways.

Conserved residues in the AF-2 surface and in helix4/5 kink orchestrate the allosteric signaling pathway.

Understanding the mechanistic details of the LBP \leftrightarrow AF-2 allosteric cross-talk is of paramount importance in the field of drug design. Indeed, using the ligand pharmacophore to modulate the AF-2 surface is a common strategy to tune the GR functional response⁵. Therefore, we next explored the details of how residues within ‘route 2’ are dynamically interconnected, how the allosteric signal is propagated at atomic level, and, ultimately, whether the bound peptide can influence the allosteric signal transduction. After a visual inspection of our MD trajectories of both TIF2/GR/Dex and PRGC1/GR/Dex complexes, we found three dynamically interconnected residues in ‘route 2’ (i.e., Gln597, Trp600 and Gln760, referred to as ‘Q₅₉₇W₆₀₀Q₇₆₀-Triad’) that bridge the structural regions of the AF-2 surface. Importantly, these residues belong to the set of critical nodes and evolutionarily conserved residues highlighted in the previous analysis. In both MD trajectories (i.e., TIF2/GR/Dex and PRGC1/GR/Dex), the Gln760 side chain is oriented either in the ‘up’ conformation, where it approaches the Gln597 side chain and can establish an H-bond (Gln597-Gln760 H-bond), or in the ‘down’ conformation, where it approaches the Trp600 side chain and can establish an H-bond (Trp600-Gln760 H-bond) (**Figure 4A**). Both H-bonds connect the helix H12 with the core of GR_{LBD}. As depicted in **Figure 4A**, the ‘down’ state is more populated in the PRGC1/GR/Dex system, being occupied for the ~54% of the simulated time, compared to ~40% in the TIF2/GR/Dex system. Conversely, the ‘up’ state is more populated in the TIF2/GR/Dex system, being populated for the ~55% of the simulated time, compared to ~31% in the PRGC1/GR/Dex system. This trend is also reflected by the Gln597-Gln760 and Trp600-Gln760 H-bond occupancy. Both the main MD simulations and the replicas show a statistically significant increase in the occupancy of the Trp600-Gln760 H-bond in presence of the PRGC1 peptide (i.e., the H-bond occupancy in the TIF2/GR/Dex system is 7.5% and $5.4 \pm 1.5\%$ for the main MD simulation and the five replicas, respectively; whereas the H-bond occupancy in PRGC1/GR/Dex is 13.4% and $16.5 \pm 4.1\%$ for the main simulation and the five replicas, respectively (**Figure S3**)). Conversely, the occupancy of the Gln597-Gln760 H-bond is slightly higher in the presence of the TIF2 peptide, however the difference is not statistically relevant with the collected data (**Figure S3**). Thus, this analysis demonstrates that the peptides affect the occupancy of Trp600-Gln760 H-bond, which is intensified in presence of the PRGC1 peptide.

At this point, we investigated the kinetic aspect of the Trp600-Gln760 H-bond formation to determine whether the two peptides also influence the stability over the time of this interaction. To do so, we firstly plotted the formation of the Trp600-Gln760 H-bond with time (**Figure 4B**). The plot shows a qualitative difference between the two systems, namely the Trp600-Gln760 H-bond appears more transient in presence of TIF2. To quantify this difference, we built a kinetic model discretizing our trajectories in macro-states and measuring the transitions between them. Then, we estimated the likelihood of the Trp600-Gln760 H-bond to remain over time in the two systems. To this aim, a transition matrix was built on the following four macro-states: (state 1) the side chain/side chain Gln597-Gln760 H-bond is formed; (state 2) side chain/side chain Trp600-Gln760 H-bond is formed; (state 3 – or *intermediate 1*) both backbone/side chain Gln597-Gln760 and side chain/side chain Trp600-Gln760 H-

bonds are formed and (state 4 – or *intermediate 2*) no H-bonds are formed (**Figure S4**). Although four states are needed to discretize the trajectories in an accurate way, for the purposes of this work we focused only on the metastability of state 2 (Trp600-Gln760 H-bond is formed), which is measured by the diagonal term of the matrix, (see Materials and Methods) ³⁵. We found that the likelihood of the system to remain in state 2 (Trp600-Gln760 H-bonds is formed) is higher in presence of the PRGC1 peptide, both in the main 4 μ s-long MD trajectories as in the replicas. In fact, the probability of the Trp600-Gln760 H-bond to remain formed in the next frame is 31.0% and $32 \pm 4\%$ for the main simulation and the five replicas of the PRGC1/GR/Dex system, respectively, and 18.8% and $22 \pm 1\%$ for the main simulation and the five replicas of the TIF2/GR/Dex system, respectively (**Figures S5** and **S6**).

Taken together, both H-bond frequency analysis and the time stability analysis support the hypothesis that the two peptides influence the dynamics of 'Q₅₉₇W₆₀₀Q₇₆₀-Triad, affecting the properties of the Trp600-Gln760 H-bond. The PRGC1 peptide consistently increases both the occupancy and the stability of the Trp600-Gln760 H-bond in our MD simulations. Since this H-bond directly connects the helix H12 with the core of GR_{LBD}, the stronger Trp600-Gln760 H-bond present in the PRGC1/GR/Dex system might strengthen the anchoring of the helix H12 crystallographic agonist-like conformation. On the contrary, the weaker Trp600-Gln760 H-bond present in the TIF2/GR/Dex system might impair the crystallographic agonist-like helix H12 conformation. Notably, these results provide a partial mechanistic explanation of the recent CPMG relaxation dispersion experiments published by Köhler and coworkers. They report a single conformation of the helix H12 in presence of the PRGC1 peptide in contrast to a dynamical exchange between alternative helix H12 conformations in presence of TIF2 ²⁰.

A stable Trp600-Gln760 H-bond might result in a rigidification of the Trp600 side chain, a critical node identified in the network analysis and well-conserved residue among the NR3C family (**Figure S2A**). In our main MD simulations the distribution of Trp600 χ_1 torsional angle in the PRGC1/GR/Dex complex oscillates about $71.5^\circ \pm 8.5^\circ$, whereas in TIF2/GR/Dex the dihedral angle oscillates about $64.7^\circ \pm 9.2^\circ$ (**Figure 4C**). Although the difference between the two distributions is not very pronounced, it is still noticeable that the Trp600 side chain deviates more from the starting X-ray structure (Trp600 $\chi_1 = 77.9^\circ$ in PDB 4UDC) in the case of TIF2/GR/Dex complex, being more prone to swing out (see yellow sticks in **Figure 4C**). A similar behavior occurs in our replica MD simulations (**Figure S7**). The importance of Trp600 plasticity for nuclear receptors' function has been demonstrated for both androgen receptor (AR) and progesterone receptor (PR) (Trp741 and Trp755 in AR and PR numeration, respectively) ¹⁴. For example, MD simulations integrated with structural data proposed that the binding of some antagonist ligands causes a rotation of the Trp741 side chain, fostering a perturbation of the helical conformation of H12 (and, in turn, the AF-2 surface) in AR ³⁶. Similarly, X-ray structure of PR in complex with the mixed profile ligand Org3H shows a ligand-induced rotation of the Trp755 that, in turn, destabilizes the agonist-like conformation of the helix H12 ³⁷. Taken together, these studies suggest that the plasticity of this conserved tryptophan, which is strategically located in between LBP and the AF-2 docking site, might dynamically influence the spatial organization of the helix H12, propagating the bi-directional AF-2 \leftrightarrow LBP allosteric signal. Unlike the studies of AR and PR, we do not observe a significant rotation of the Trp600. This is likely due to the fact that we compared the effect of two strong coactivators. Here, a large helix H12 conformational change is not expected, because the agonist-like conformation is always preferred, as demonstrated by X-ray structural data ¹⁶. ²¹. However, we propose that the different flexibility of Trp600 captured in our MD simulations might represent

one of the possible mechanical switches used to fine-tune the structural state of the helix H12 and, consequently, the functional response, as observed in NMR experiments of TIF2/GR/Dex and PRGC1/GR/Dex complexes ²⁰.

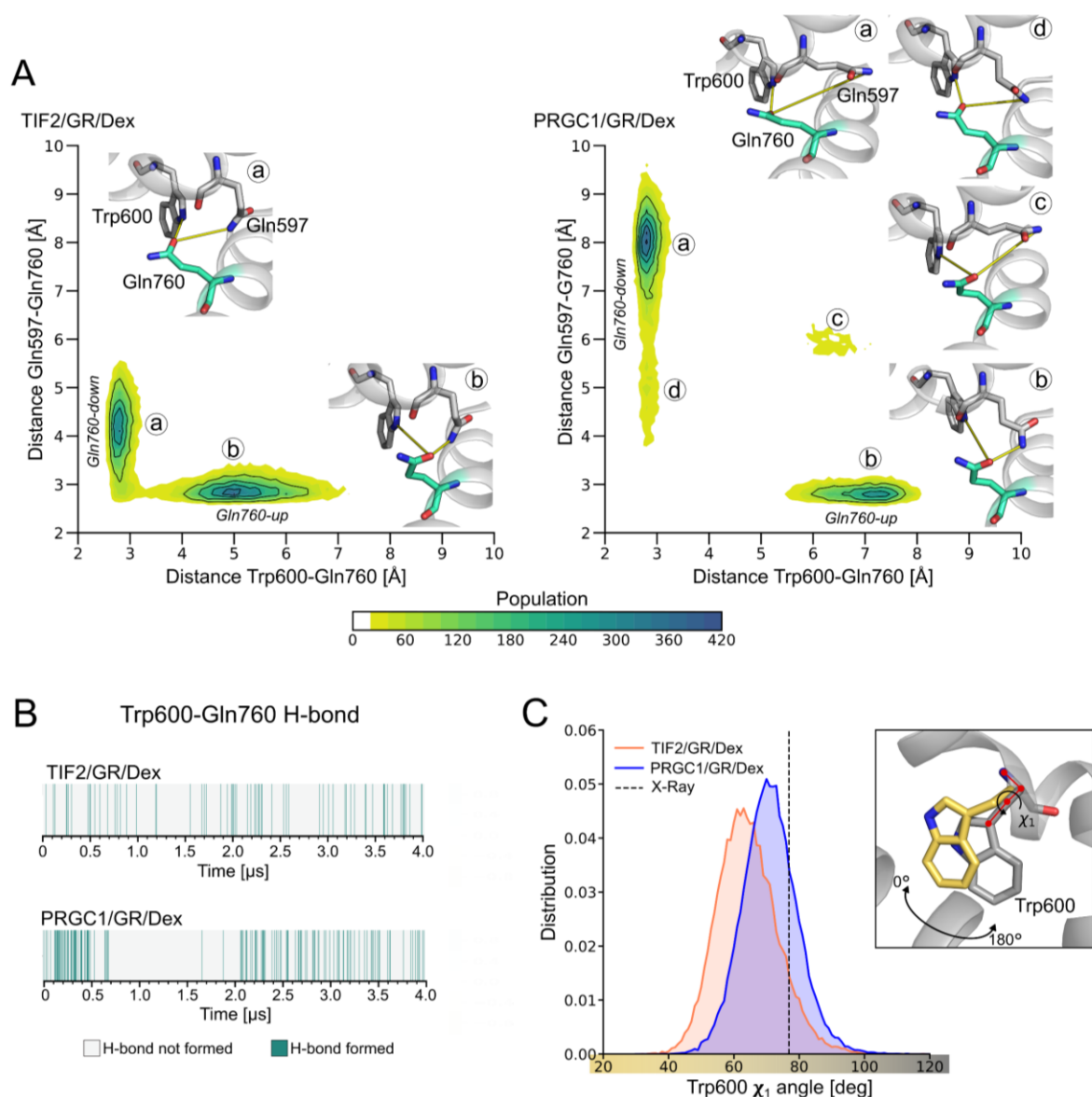


Figure 4. Dynamic of the Q₅₉₇W₆₀₀Q₇₆₀-Triad. A) 2D plot of TIF2/GR/Dex (left) and PRGC1/GR/Dex (right) MD trajectories' points projected along the distance between the Trp600 N ϵ 1 atom and Gln760 O ϵ 1 atom (i.e., Distance Trp600-Gln760, X axes) and along the distance between the Gln597 N ϵ 2 atom and Gln760 O ϵ 1 atom (i.e., Distance Gln597-Gln760, Y axes). The insets represent the local minima (a, Gln760-down) and (b, Gln760-up) and the intermediate states (c) and (d). In the two systems the local minima *a* and *b* are differently separated in the XY space due to a spatial rearrangement of the Gln597 side chain (see main text and **Figure 5B**). B) Heat map displaying H-bond presence (dark cyan lines) with time in Trp600-Gln760 for both TIF2/GR/Dex and PRGC1/GR/Dex. C) Trp600 χ_1 torsional angle distribution in TIF2/GR/Dex (orange) and PRGC1/GR/Dex (blue). The dashed black line measures the dihedral angle in the X-ray 4UDC. The insert represents the two limits of Trp600 configurations adopted in our MD simulations.

The peptide sequences affect the dynamical allosteric network. At this point, we carried out further investigations to establish what causes the diverse dynamical behavior of the Q₅₉₇W₆₀₀Q₇₆₀-Triad in the two studied systems and why the Trp600-Gln760 H-bond is more populated and stable in presence of the PRGC1 peptide. Because the coregulator peptides bind in proximity to the Q₅₉₇W₆₀₀Q₇₆₀-Triad, we hypothesized that the peptide sequence can directly alter the dynamical interconnection of the triad, especially affecting the Trp600-Gln760 H-bond.

A visual inspection of the public available X-ray structures of GR revealed that the coregulator TIF2 peptide can directly interact with the Q₅₉₇W₆₀₀Q₇₆₀-Triad through a H-bond established with Gln597 in helix H4. Such H-bond has a total occurrence of 37.5% among the 32 collected X-ray structures and involves either the N₃ or E₄ residues of TIF2 (**Table S1**). In our 4- μ s long MD simulations of the TIF2/GR/Dex system, after ~60 ns the distance between N₃ and Gln597 shortened (**Figure 5A**), allowing the formation of a stable H-bond (the overall occupancy is ~98% over the simulated time). This H-bond is formed also in other replicas MD simulations (**Table S2**). Importantly, the formation of Gln597-N₃ H-bond in the TIF2/GR/Dex system is concomitant with the re-orientation of the Gln597 side chain. In fact, the Gln597 χ_2 torsional angle is mainly distributed between ~30° and ~80°, departing from the crystallographic value (155° in PDB code 4UDC) and pointing, instead, towards the N₃ side chain of the TIF2 peptide (referred to as the ‘inward’ conformation, see **Figure 5B** and **5C**). On the other hand, the PRGC1 peptide does not interact with Gln597, because the position -3 is occupied by a proline (instead of asparagine as in TIF2). This maintains the Gln597 side chain in the crystallographic ‘outward’ orientation, being the Gln597 χ_2 torsional angle centered between ~150° and ~-150° (**Figure 5B** and **5C**). A similar behavior is also observed in replica MD simulations (**Figure S8**). The population distribution of the Gln760-up and -down states in **Figure 4A** further supports the observed differences of populations of the ‘inward’ and ‘outward’ orientations of the Gln597 side chain.

Based on these observations, we hypothesized a molecular mechanism that might explain why the dynamical behavior of the Q₅₉₇W₆₀₀Q₇₆₀-Triad depends on the peptide sequence (**Figure 5C**). The stabilization of the ‘inward’ conformation of the Gln597, mediated by the TIF2 H-bond, fosters the up \leftrightarrow down flip of Gln760 side chain. (**Figure 4A**) As a result, the Trp600-Gln760 H-bond is weakened thermodynamically (i.e. the overall H-bond occupancy is lower compared to the PRGC1/GR/Dex system) and becomes less stable kinetically (i.e., the Trp600-Gln760 H-bond is more transient compared to the PRGC1/GR/Dex system). Thus, the weak Trp600-Gln760 H-bond confers more flexibility on the Trp600 side chain which, in turn, might alter stability of the agonist-like configuration of H12 by a similar mechanism as observed in the other nuclear receptors, AR and PR^{36, 37}. Conversely, the stabilization of the ‘outward’ conformation of the Gln597, as found in the presence of the PRGC1 peptide, might create a steric hindrance that disfavors the up \leftrightarrow down transitions of the Gln760 side chain (**Figure 4A**). Also, the formation of cooperative H-bonds between Gln597 and Gln760 and between Gln597 and N₃ might discourage the up \leftrightarrow down transitions of the Gln760 side chain. This strengthens the Trp600-Gln760 H-bond thermodynamically and makes it less transient kinetically. Here, the Trp600 is also closer to the crystallographic orientation and this, in turn, might further favor the stabilization of H12 in the agonist-like configuration.

Overall, this mechanism provides a plausible explanation for the experimental data provided by Köhler and co-workers. The TIF2 peptide impairs the formation of a competent Trp600-Gln760 H-bond. This weakens the structural contact between the helix H12 and the core of GR_{LD} and causes, in turn, a possible destabilization of the crystallographic agonist-like configuration of the helix H12. Instead, the PRGC1 peptide promotes the

formation of a stable Trp600-Gln760 H-bond. This strengthens the connection between the helix H12 and the core of GR_{LBD} that may cause a stabilization of the crystallographic agonist-like conformation of the helix H12.

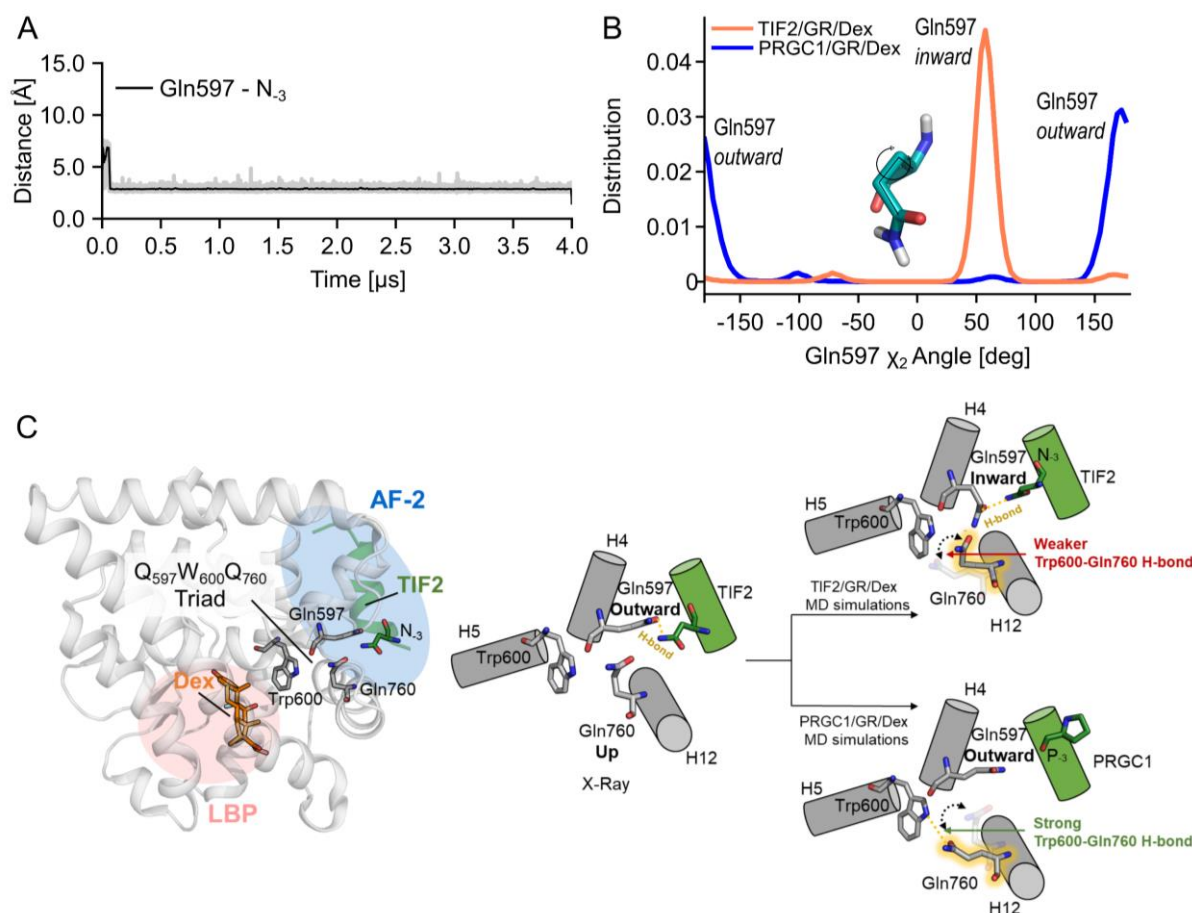


Figure 5. Mechanism that controls the dynamics of ‘Q₅₉₇W₆₀₀Q₇₆₀-Triad’. A) Distance with time between Gln597 O ϵ 1 atom and N₃ N δ 2 atom in the TIF2/GR/Dex trajectory (grey lines) and the value averaged over time (black line). B) Distribution of the Gln597 χ_2 dihedral angle in TIF2/GR/Dex (orange line) and PRGC1/GR/Dex (blue line). C) Left: representation of GR_{LBD} (white cartoon) in complex with dexamethasone (orange sticks) and the TIF2 peptide (green cartoon), (PDB 4UDC). The LBP and the AF-2 regions are highlighted with pink and blue circles, respectively. The ‘Q₅₉₇W₆₀₀Q₇₆₀-Triad’ is represented as grey sticks. Right: representation of the ‘Q₅₉₇W₆₀₀Q₇₆₀-Triad’ mechanism as described in the main text. On the left, the starting point (i.e., PDB 4UDC X-ray structure) is depicted. On the top-right, the TIF2/GR/Dex mechanism is shown: the Trp600-Gln760 H-bond is poorly populated, resulting in a weaker interaction. Here, the N₃ of the TIF2 peptide interacts through an H-bond with the Gln597 side chain, locking it in the ‘inward’ conformation. On the bottom-right, the PRGC1/GR/Dex mechanism is shown: the Trp600-Gln760 H-bond is more populated, resulting in a stronger Trp600-Gln760 H-bond. Here, the P₃ of the PRGC1 peptide cannot interact with the Gln597 side chain that occupies, in turn, the ‘outward’ conformation.

Peptide binding affects the organization of the AF-2 docking site. Further analysis were performed to shed light on the mechanistic details of peptide binding. To do so, we monitored the Q₅₉₇W₆₀₀Q₇₆₀-Triad's behavior on the peptide-free GR_{LBD} system (i.e. GR/Dex, see Materials and Methods) and compared the results with the GR peptide-bound form (i.e., TIF2/GR/Dex and PRGC1/GR/Dex complexes). In contrast with both the peptide-bound systems, the intermediate state of the Gln760 side chain is more populated in GR/Dex both in the main MD simulations and in the five independent replicas (~39%, ~5% and ~15% in GR/Dex, TIF2/GR/Dex and PRGC1/GR/Dex main 4 μ s MD simulations, respectively, and 34.7 \pm 3.2%, 15.9 \pm 1.2% and 17.2 \pm 6.4% in GR/Dex, TIF2/GR/Dex and PRGC1/GR/Dex replicas, respectively). As a consequence, both the Trp600-Gln760 and Gln597-Gln760 H-bonds are less formed or have a comparable formation time in GR/Dex compared to both the peptide-bound systems (**Figure S3**). In addition, similarly to the PRGC1/GR/Dex, in GR/Dex the Gln597 side chain mainly occupies the “outward” conformation (**Figure S8**). These data show that the binding of TIF2 and PRGC1 peptides shifts the populations of the Q₅₉₇W₆₀₀Q₇₆₀-Triad in a specific manner. The binding of the PRGC1 peptide does not affect the orientation of the Gln597 sidechain and intensifies the Trp600-Gln760 H-bond. In contrast, the TIF2 induces a reorganization of Gln597 sidechain, which might cause the weakening of the Trp600-Gln760 H-bond. Overall, these results suggest that the structural organization of the AF-2 in the peptide-free form is more competent for the PRGC1 peptide binding. On the other hand, the optimal binding of the TIF2 peptide requires a reorganization of the residues in the AF-2 docking site.

The ‘-2’ peptide position plays an important role for the coregulator binding. Ligand binding impacts the helix H12 structure and stabilizes the AF-2 binding site. The conformation of the AF-2 surface will, in turn, drive coregulator recruitment to DNA and elicit a specific biological response^{14, 15}. Interestingly, experimental studies show that coregulator peptides with high binding affinity to GR_{LBD}, such as PRGC1, more efficiently restrict the conformational space of the helix H12 compared to other coregulator peptides, such as TIF2²⁰. Here, we investigated the dynamical aspect of the helix H12/peptide interaction to evaluate how different peptide sequences stabilize the helix H12.

The analysis of all X-ray structures of GR in complex with TIF2 deposited in the protein data bank (PDB) revealed that the coactivator peptide directly interacts through a H-bond with the well conserved Glu755 in helix H12 in the 62.5% of the collected structures (**Table S1**). This interaction is part of the ‘charge clamp’, which is important for coregulator binding across the NHR family³⁸. We found that in the TIF2/GR/Dex trajectories the H-bond between the Glu755 and the peptide is transient. In fact, the crystallographic H-bonds between Glu755 and the backbone of both L₋₁ and L₀ are lost at the beginning of the MD simulations (**Figure S9**). On the other hand, in the PRGC1/GR/Dex system the Glu755 establishes a more favorable H-bond with the side chain of the S₋₂, being occupied for ~37% of the simulated time. These results are consistent within our replicas (**Table S2 and S3**). Based on these data, we hypothesized that the position -2 of the peptide might be crucial for an efficient formation of the ‘charge clamp’ that helps maintaining the peptide steadily anchored to the AF-2 surface. To confirm this hypothesis, we performed SPR biosensor experiments to quantitatively characterize the binding affinity of the peptides and determine if the residue in position -2 is important for the peptide/GR_{LBD} interaction, as revealed in our MD simulations. First, we compared the binding affinity of the wild type peptides and we found that PRGC1

is more affine towards GR_{LBD} compared to TIF2 (K_d of $0.8 \pm 0.1 \mu\text{M}$ and $2.0 \pm 0.2 \mu\text{M}$ for PRGC1 and TIF2, respectively), in line with previous studies ²⁰ (**Figure 6**). Upon replacing the alanine with a serine in TIF2 (TIF2mut/GR/Dex), we observed a gain of peptide affinity with respect to the wild type form (K_d of $0.8 \pm 0.1 \mu\text{M}$) (**Figure 6**). This gain might be due to the formation of a new H-bond between the Glu755 and the S₋₂. Finally, the mutation of the serine to alanine in PRGC1 (i.e., PRGC1mut/GR/Dex) causes a drop of the binding affinity with respect to the wild type (K_d of $2.5 \pm 0.2 \mu\text{M}$) (**Figure 6**).

Taken together, this experimental evidence corroborates our hypothesis that position -2 is important for an efficient binding of the coregulators peptides and that the serine is strategic for a competent formation of the ‘charge clamp’ with Glu755. In other words, the PRGC1 peptide, which bears a serine in position -2, is more affine towards GR_{LBD} compared to the TIF2 peptide. This, in turn, might rigidify the helix H12, explaining why the CPMG relaxation dispersion experiments revealed only one populated conformation in presence of the PRGC1 peptide ²⁰. Notably, further studies report that the position -2 influences the peptide binding, supporting our data which show that the S₋₂ is an optimal residue for an efficient peptide binding ^{19, 39, 40}. For example, in a study carried out by Broekema and colleagues, the authors assessed the binding affinity of ~3700 coregulators towards the GR protein. They found that the serine in position -2 is conserved among the 57 coregulators that display the highest binding affinity versus GR ³⁹ (**Figure S10**). Finally, it is worth noting that in GR, ligand and the coregulators work together to elicit a specific allosteric signal ²⁰. Here, we show that the different magnitude of the allosteric response in TIF2/GR/Dex and PRGC1/GR/Dex complexes likely depends also on the different strength of the peptide/GR interaction.

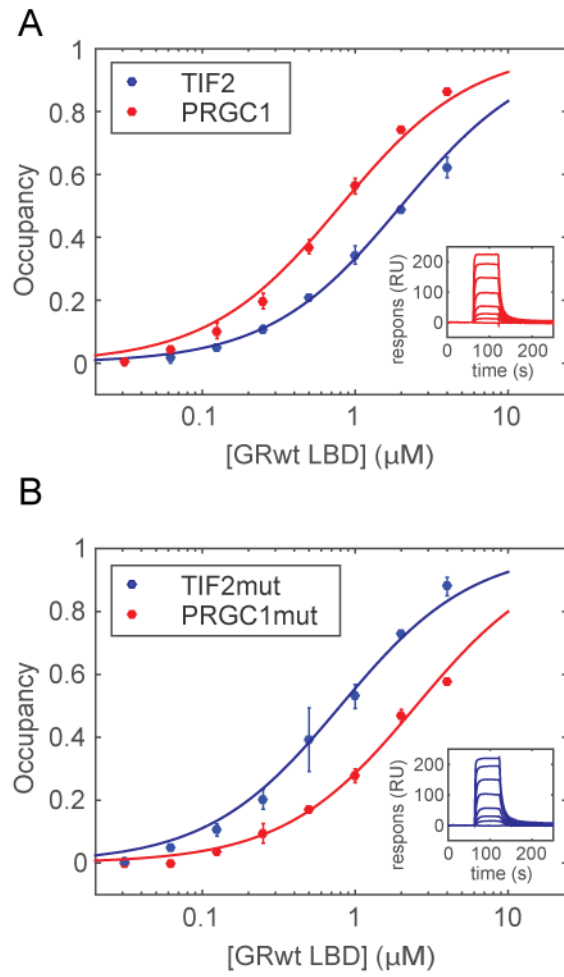


Figure 6. SPR biosensor experiments characterizing binding affinities to coregulator peptides. Concentration-response curves of wild type GR_{LBD} (GRwt LBD) binding to immobilized PRGC1 and TIF2 (A) or corresponding mutants (B). Inset show example sensorgrams. The error bars indicate one standard deviation for three consecutive concentration-response. Steady state binding levels were fitted to 1:1 binding model (solid lines) to determine affinities.

CONCLUSIONS

A number of structural and biophysical studies demonstrate that GR_{LBD} is allosterically controlled by the interplay of coregulator peptides and ligands, that bind respectively to the AF-2 docking site and to the LBP. Notably, such cross-talk is a fundamental mechanism to fine-tune the protein functional response in GR. In this context, the motility of helix H12, a structural element that bridges the two binding sites, is deemed a key feature for the modulation of the allosteric signal.²⁰ In this study, we adopted a multidisciplinary approach to unravel the molecular mechanisms that control the allosteric signal transmission between the AF-2 docking site and the LBP. In addition, we aimed to determine how the coregulators' sequences affect signal transduction, as proposed in a recent NMR study²⁰.

We performed microsecond-long MD simulations of GR_{LBD} bound to dexamethasone and either TIF2 or PRGC1 coregulator peptides. In agreement with the recently published ¹³C and ¹⁵N CPMG relaxation dispersion experiments, we observed a different flexibility of the region in proximity of the ligand binding site between the two systems, proving that MD simulations capture long-range allosteric effects. Using community network analysis, we identified a novel AF-2 ↔ LBP allosteric signal propagation route that involves a set of residues identified as hotspots by CPMG relaxation dispersion experiments (or in proximity to those)²⁰. Evolutionarily conserved Trp600 is the key node that connects AF-2 and LBP. In our MD simulations, we found that Trp600 is involved in a H-bond network with evolutionarily conserved residues, forming the so called Q₅₉₇W₆₀₀Q₇₆₀-Triad. Such network connects the structural elements of the AF-2 (helices H4 and H12) to LBP (helix H5) and the Trp600 is strategically located in between AF-2 and LBP. In this context, we speculate that the Trp600-Gln760 H-bond is a key interaction that helps keep the helix H12 anchored to the core of the GR_{LBD}, stabilizing the agonist-like conformation. Therefore, the Trp600-Gln760 H-bond might be important to control the alternative conformations of the helix H12, which lead to different levels of GR activation.²⁰

Statistical analysis of MD trajectories allowed us to deduce a plausible molecular mechanism that explains how the allosteric signal affects the motility of the helix H12 differently in the two TIF2/GR/Dex and PRGC1/GR/Dex complexes, as observed in recent CPMG experiments²⁰. Firstly, we observed that the coregulator peptides affect the dynamics of the Q₅₉₇W₆₀₀Q₇₆₀-Triad, by interfering with the conformational state of Gln597, which, in turn, influences the kinetic and thermodynamic stability of the Trp600-Gln760 H-bond. According to our MD simulations, the proline (P₋₃) in the PRGC1 peptide strengthens the Trp600-Gln760 H-bond compared to the asparagine (N₋₃) in the TIF2 peptide. Thus, the PRGC1 stabilizes the helix H12 more efficiently in the agonist conformation as found in CPMG experiments²⁰. Secondly, we identified a different interaction pattern between the coregulators and the helix H12. The side chain of S₋₂ of PRGC1 peptide forms an H-bond with the evolutionarily conserved Glu755 in helix H12, whereas this interaction is missing in the TIF2 peptide because the serine is replaced with an alanine (A₋₂). This H-bond is part of the “charge clamp”, which is crucial for coregulator binding across the NHR family³⁸. We propose that the tighter interaction between the PRGC1 peptide and the helix H12, confirmed by SPR experiments, contributes to augmented rigidity of the helix H12. Therefore, our data suggests that the amino acid in position -2 of the peptide plays an important role in controlling the motility of helix H12, which ultimately affects the GR functional response. In addition, our study offers further details on the mechanism of the coregulators binding. We found that the peptide-free form of GR is more competent for the

PRGC1 binding, while, on the other hand, the binding of TIF2 requires a reorganization of the residues in the AF-2 docking site.

Overall, our work contributes to the understanding of allosteric regulation of GR protein and by clarifying the molecular mechanism involved in the LBP \leftrightarrow AF-2 bidirectional allosteric cross-talk. Despite the difference in timescale, we show a significant overlap between the CPMG experiments and our outcomes, demonstrating that MD simulation is a powerful tool that complements experimental observations, providing valuable details about allosteric regulation at an atomic level. Finally, our study poses the foundation for designing novel GR drugs able to fine-tune the protein's function. Indeed, according to the proposed model, adjusting the ligand interaction with Trp600 can be a viable strategy to modulate the level of agonism in GR.

MATERIALS AND METHODS

Molecular dynamics simulations. In this study we used three molecular systems, namely TIF2/GR/Dex, PRGC1/GR/Dex and GR/Dex complexes. The first two consist of the wild type ternary complex formed by the coactivator peptide TIF2 or PRGC1, the ligand binding domain of glucocorticoid receptor (GR_{LBD}) and dexamethasone (Dex). The latter consists of GR_{LBD} in complex with Dex. The TIF2/GR/Dex system was built starting from the X-ray structure PDB code: 4UDC²¹ and replacing both the Ser602 and Asp638 with Phe and Cys, respectively, to get the wild-type form of the protein. The PRGC1/GR/Dex system was built using Maestro software (Schrödinger, LLC) starting from the TIF2/GR/Dex complex and modeling the PRGC1 peptide based on the 3D coordinates of TIF2. Notably, the obtained 3D coordinates of PRGC1 overlap well with the recently released X-ray structure of ancestral variant of GR in complex with PRGC1¹⁶ (RMSD of 0.6 Å). The GR/Dex was built by removing the TIF2 peptide from the TIF/GR/Dex system.

Both GR_{LBD} and peptides were parameterized using the Amber14SB force field⁴¹, because accurately represent protein ensembles.⁴² Dex was parametrized using the GAFF force field⁴³ after computing the point charges at the HF/6-31G* level of theory. Each complex was embedded in a TIP3P water box⁴⁴ and neutralized by adding Na⁺ and Cl⁻ ions. The complexes were simulated with a time step of 2 fs and LINCS algorithm⁴⁵ was used to constrain covalent bonds involving hydrogen atoms. All simulations were performed using a short-range neighbor list cutoff of 10 Å, whereas long-range electrostatics were treated with the Particle Mesh Ewald method using a Fourier grid spacing of 1.2 Å and a cubic spline interpolation⁴⁶. The Parrinello–Rahman barostat⁴⁷ and the velocity-rescaling thermostat⁴⁸ were employed to keep the pressure of 1 bar and a temperature of 310 K. All the simulations performed in this study were performed with Gromacs 2016 engine⁴⁹. The following simulation protocol was employed for each system. First, the systems were subjected to energy minimization, then they were slowly heated in three steps of 50 ps each to reach a target temperature of 310 K, applying a harmonic restraint to the backbone atoms. At this point, an NPT ensemble was sampled for 250 ps to adjust the system's density and the restraints were gradually released. For each system, we collected ~4 μs of MD simulations in the NPT ensemble. Five replicas for each of the three systems were carried out to enhance the sampling and improve the statistical significance and reproducibility of our results. The length of the replica varies from 700 ns to 2 μs. We used the hydrogen mass repartitioning (HMR) scheme⁵⁰ and the same protocol as previously described. The first 50 ns

from each trajectory were discarded to perform the next analysis, as we considered this time still part of the equilibration phase.

Flexibility analysis. The root mean square fluctuation (RMSF) was performed to analyze the portions of the protein that fluctuate most over the MD trajectories. The RMSF values of the C α atoms were computed with the *gmx rmsf* tool integrated in Gromacs 2016 software⁴⁹. To remove the noise, the N- and C- terminal tails, as well as the coregulator peptide sequences were removed from the analysis. The principal component analysis (PCA) was carried out to capture the essential motions of the protein during the MD simulations⁵¹. The computation was performed using *gmx covar* and *gmx ana eig* tools integrated in Gromacs 2016⁴⁹. Also, here, the analysis was performed on the C α atoms, excluding the N- and C- terminal tails, as well as the coregulator peptide sequences. The first 20 modes of the TIF2/GR/Dex and PRGC1/GR/Dex main trajectories covered the 60.5% and 58.3% of the total variance, respectively (**Figure S1A**). Instead, the first 20 modes of the merged trajectories of TIF2/GR/Dex and PRGC1/GR/Dex replicas simulations covered the 64.0% and 65.0% of the total variance, respectively (**Figure S1A**). As expected, the same number of modes (i.e., 20) covers a higher variance when a combination of multiple trajectories (i.e., replicas) is analyzed⁵². Since the study of 20 modes is of difficult interpretation, we focuses our analysis on fewer modes. We adopted the Cattell criterion to identify the optimal number of modes to analyze as it has been demonstrated that this method works nicely in proteins. In brief, we constructed the eigenvalue scree plot (**Figure S1A**) and identified the “kink” in the curves, which in our case is at mode=3.⁵²

Allosteric network analysis. The dynamical network analysis was performed to study the transmission of the allosteric signal in GR_{LBD}. The network is defined as a set of nodes (i.e., the C α atoms of GR_{LBD}) that are connected by edges. The weight (w_{ij}) of the edges is the probability of information transfer across the nodes i and j is and calculated according to the following formula:

$$w_{ij} = -\log(|C_{ij}|)$$

where C_{ij} is the pairwise correlations defined by:

$$C_{ij} = \frac{\langle \Delta r_i(t) \cdot \Delta r_j(t) \rangle_t}{\sqrt{\langle \|\Delta r_i(t)\|^2 \rangle_t} \sqrt{\langle \|\Delta r_j(t)\|^2 \rangle_t}}$$

where $r_i(t)$ and $r_j(t)$ are the position of the atoms corresponding to the i^{th} and j^{th} node as a function of time t . $\Delta r_i(t) = r_i(t) - \langle r_i(t) \rangle_t$, where $\langle \rangle_t$ indicates the time ensemble average. Two nodes are considered connected if the heavy atoms are within a distance cutoff of 5.5 Å during the simulation time, whereas they are not considered as connected if they are the nearest neighbors in sequence. The value of this cutoff is chosen accordingly the work of Ricci *et al.*²⁴, albeit it has been demonstrated that different values lead to minor changes in the community distribution of the network²². A community is formed by a set of nodes having a stronger connection within the nodes of the community itself than to nodes belonging to other communities. Since in this work we applied the dynamic correlation to measure the interconnection between nodes, we can infer that a community is composed by residues that move in a concerted fashion. We used the Girvan-Newman algorithm to identify the communities of the network⁵³. In detail, this algorithm uses the “edge betweenness” criterion to identify communities, which is defined as the number of shortest paths that cross an edge. It iteratively removes edges with the highest betweenness and recalculate again the betweenness of the remaining edges until all edges have been removed. The

modularity parameter Q is used to determine the goodness of the community structure. The best division occurs when nodes within a community are highly intra-connected and, in the same time, the different communities are poorly inter-connected. In typical biological networks, the optimal modularity score Q ranges from 0.4 to 0.7²². In our case, Q is equal to 0.57 and 0.55 for TIF2/GR/Dex and PRGC1/GR/Dex, respectively. Different communities can be interconnected by few edges. The edges with the highest betweenness are called ‘critical’ and are deemed to be important for the propagation of the allosteric signal²². The dynamic cross correlation map has been calculated with *cpptraj*⁵⁴, while community network analysis and data visualization has been carried out with *NetworkView*⁵⁵.

Analysis of structural data from MD simulations. Distances and dihedral angles were computed via the *driver* module integrated in Plumed 2⁵⁶. The H-bonds were calculated with the *gmx hbond* module in Gromacs 2016⁴⁹ and post-processed with the *readHBmap.py* script. The “down” population includes all the configurations having a Trp600-Gln760 distance < 3.5 Å; the “up” population includes all the configurations having a Trp600-Gln760 distance > 3.5 Å and a Gln597-Gln760 distance < 3.5 Å; the “intermediate” population includes the remaining configurations (see **Figure 4A**).

Calculation of transition matrix. The transition matrix was calculated to provide a description of the transition process and kinetics relative to the $Q_{597}W_{600}Q_{760}$ -Triad. The final aim is to assess the stability over time of the Trp600-Gln760 H-bond during the MD simulations. To do so, we firstly defined four non-overlapping discrete states that described the transition process in the $Q_{597}W_{600}Q_{760}$ -triad: (1) side chain/side chain Gln597-Gln760 H-bond is formed; (2) side chain/side chain Trp600-Gln760 H-bond is formed; (3) both backbone/side chain Gln597-Gln760 and side chain/side chain Trp600-Gln760 H-bonds are formed; (4) no H-bonds are formed (**Figure S4 and S5**). Then, we assigned at each frame our MD trajectories the corresponding state, as described previously, to obtain a one-dimensional time series. This was used as input to calculate the row-normalized transition matrix T , where T_{ij} corresponds to the likelihood that state i changes to state j within a certain lag time τ .⁵⁷ The diagonal term of the matrix T_{ii} represents, instead, the metastability of a state i . The lag time τ was 100 ps. The transition matrix was computed and visualized through an in-house python script.

Conserved residues analysis. We used BLAST to search the UniprotKB database for GR sequences belonging to species other than human. After removal of all isoforms, a total 61 sequences were collected and aligned using Clustal Omega using default settings⁵⁸ (see Supporting Information for a detailed list of UniProt codes used in the alignment). WebLogo webserver was used to display patterns in sequence conservation⁵⁹. We employed GREMLIN pseudolikelihood method to study the co-evolution of residue-residue pairs in GR⁶⁰. To do so, we employed the online server (<http://gremlin.bakerlab.org/>), giving as input the FASTA sequence of GR_{LBD}. The multiple sequence alignment was performed with HHblits, using an e-value of 1E-10. The alignment was then filtered by removing the sequences that cover less than 75% of the query sequence (i.e., GR) and that contain over 75% of gaps.

Surface Plasmon Resonance. All SPR measurements were run on a BIAcore 3000 (GE healthcare) using a running buffer of 10 mM HEPES, 50 mM NaCl, 0.05% Tween 20, pH 7.6. To avoid mass transport limitations biotinylated coregulators (Thermo Fisher) were immobilized on a low-density chip (SAHC30M Xantec) at ~50-60 RU. The wild type GR_{LBD} was produced according to²⁰, incubated with Dexamethasone at 50 μ M and injected at increasing concentrations for 1 min in the running buffer. Importantly, to maximize accuracy for comparison, each peptide pair was immobilized on the same chip and exposed to identical protein concentrations. Steady state

binding levels were fitted to 1:1 binding model to determine affinities. The peptide sequences employed in SPR measurements are KENALLRYLLDKDDTK, KENSLRYLLDKDDTK, EEPsLLKKLLAPANT and EEPALLKKLLAPANT for TIF2, TIF2mut, PRGC1 and PRGC1mut, respectively.

Statistical analysis. The five replica MD simulations generated in this study for both TIF2/GR/Dex and PRGC/GR/Dex systems were used to support the observations of the main 4- μ s long MD simulations. The Mann-Whitney U-test was performed to estimate the statistical significance of H-bond population and H-bond time stability (**Figures S3 and S6**). The results are presented as the means \pm SEM.

ACKNOWLEDGMENTS

We thank Dr. Matti Lepistö for helpful discussion and Dr. Jon Paul Janet for critical reading of the manuscript. G.L.S is a fellow of the AstraZeneca R&D postdoc program.

AUTHOR CONTRIBUTIONS

G.L.S., K.E. C.T. A.H and A.I.F have conceived and designed the work; G.L.S performed bioinformatic analysis, MD simulations and relative analysis; A.G. performed SPR experiments and analysis; all authors have interpreted the data and drafted the manuscript.

DECLARATION OF INTERESTS

The authors are employees of AstraZeneca.

References

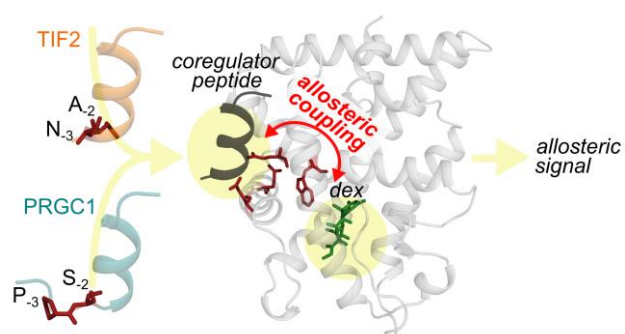
- (1) Bamberger, C. M.; Schulte, H. M.; Chrousos, G. P. Molecular Determinants of Glucocorticoid Receptor Function and Tissue Sensitivity to Glucocorticoids. *Endocr. Rev.* **1996**, *17* (3), 245–261. <https://doi.org/10.1210/edrv-17-3-245>.
- (2) Rhen, T.; Cidlowski, J. A. Antiinflammatory Action of Glucocorticoids - New Mechanisms for Old Drugs. *N. Engl. J. Med.* 2005. <https://doi.org/10.1056/NEJMr050541>.
- (3) Schäcke, H.; Döcke, W. D.; Asadullah, K. Mechanisms Involved in the Side Effects of Glucocorticoids. *Pharmacology and Therapeutics*. 2002, pp 23–43. [https://doi.org/10.1016/S0163-7258\(02\)00297-8](https://doi.org/10.1016/S0163-7258(02)00297-8).
- (4) Buttgereit, F. Can We Shift the Benefit–risk Ratio of Glucocorticoids? *The Lancet Rheumatology*. 2020, pp e5–e6. [https://doi.org/10.1016/S2665-9913\(19\)30138-9](https://doi.org/10.1016/S2665-9913(19)30138-9).
- (5) Ripa, L.; Edman, K.; Dearman, M.; Edenro, G.; Hendrickx, R.; Ullah, V.; Chang, H. F.; Lepistö, M.; Chapman, D.; Geschwindner, S.; et al. Discovery of a Novel Oral Glucocorticoid Receptor Modulator (AZD9567) with Improved Side Effect Profile. *J. Med. Chem.* **2018**, *61* (5), 1785–1799. <https://doi.org/10.1021/acs.jmedchem.7b01690>.
- (6) Hegelund Myrbäck, T.; Prothon, S.; Edman, K.; Leander, J.; Hashemi, M.; Dearman, M.; Edenro, G.; Svanberg, P.; Andersson, E. M.; Almquist, J.; et al. Effects of a Selective Glucocorticoid Receptor Modulator (AZD9567) versus Prednisolone in Healthy Volunteers: Two Phase 1, Single-Blind, Randomised Controlled Trials. *Lancet Rheumatol.* **2020**, *2* (1), e31–e41. [https://doi.org/10.1016/S2665-9913\(19\)30103-1](https://doi.org/10.1016/S2665-9913(19)30103-1).
- (7) Nicolaidis, N. C.; Galata, Z.; Kino, T.; Chrousos, G. P.; Charmandari, E. The Human Glucocorticoid Receptor: Molecular Basis of Biologic Function. *Steroids*. 2010, pp 1–12. <https://doi.org/10.1016/j.steroids.2009.09.002>.
- (8) Oakley, R. H.; Cidlowski, J. A. The Biology of the Glucocorticoid Receptor: New Signaling Mechanisms in Health and Disease. *Journal of Allergy and Clinical Immunology*. 2013, pp 1033–1044. <https://doi.org/10.1016/j.jaci.2013.09.007>.
- (9) Weikum, E. R.; Knuesel, M. T.; Ortlund, E. A.; Yamamoto, K. R. Glucocorticoid Receptor Control of Transcription: Precision and Plasticity via Allostery. *Nat. Rev. Mol. Cell Biol.* **2017**, *18* (3), 159–174. <https://doi.org/10.1038/nrm.2016.152>.
- (10) Heery, D. M.; Kalkhoven, E.; Hoare, S.; Parker, M. G. A Signature Motif in Transcriptional Co-Activators Mediates Binding to Nuclear Receptors. *Nature* **1997**, *387* (6634), 733–736. <https://doi.org/10.1038/42750>.
- (11) Hu, X.; Lazar, M. A. The CoRNR Motif Controls the Recruitment of Compressors by Nuclear Hormone Receptors. *Nature* **1999**, *402* (6757), 93–96. <https://doi.org/10.1038/47069>.
- (12) Frego, L.; Davidson, W. Conformational Changes of the Glucocorticoid Receptor Ligand Binding Domain Induced by Ligand and Cofactor Binding, and the Location of Cofactor Binding Sites Determined by Hydrogen/Deuterium Exchange Mass Spectrometry. *Protein Sci.* **2006**, *15* (4), 722–730. <https://doi.org/10.1110/ps.051781406>.
- (13) Schoch, G. A.; D’Arcy, B.; Stihle, M.; Burger, D.; Bär, D.; Benz, J.; Thoma, R.; Ruf, A. Molecular Switch in the Glucocorticoid Receptor: Active and Passive Antagonist Conformations. *J. Mol. Biol.* **2010**, *395* (3), 568–577. <https://doi.org/10.1016/j.jmb.2009.11.011>.
- (14) Okafor, C. D.; Colucci, J. K.; Ortlund, E. A. Ligand-Induced Allosteric Effects Governing SR Signaling. *Nucl. Recept. Res.* **2019**, *6*, 1–17. <https://doi.org/10.32527/2019/101382>.
- (15) Alves, N. R. C.; Pecci, A.; Alvarez, L. D. Structural Insights into the Ligand Binding Domain of the Glucocorticoid Receptor: A Molecular Dynamics Study. *J. Chem. Inf. Model.* **2020**, *60* (2), 794–804. <https://doi.org/10.1021/acs.jcim.9b00776>.
- (16) Liu, X.; Wang, Y.; Ortlund, E. A. First High-Resolution Crystal Structures of the Glucocorticoid Receptor Ligand-Binding Domain-Peroxisome Proliferator-Activated γ Coactivator 1- α Complex with Endogenous and Synthetic Glucocorticoids. *Mol. Pharmacol.* **2019**, *96* (4), 408–417. <https://doi.org/10.1124/mol.119.116806>.

- (17) Ricketson, D.; Hostick, U.; Fang, L.; Yamamoto, K. R.; Darimont, B. D. A Conformational Switch in the Ligand-Binding Domain Regulates the Dependence of the Glucocorticoid Receptor on Hsp90. *J. Mol. Biol.* **2007**, *368* (3), 729–741. <https://doi.org/10.1016/j.jmb.2007.02.057>.
- (18) Liu, X.; Wang, Y.; Gutierrez, J. S.; Damsker, J. M.; Nagaraju, K.; Hoffman, E. P.; Ortlund, E. A. Disruption of a Key Ligand-H-Bond Network Drives Dissociative Properties in Vamorolone for Duchenne Muscular Dystrophy Treatment. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, 1–9. <https://doi.org/10.1073/pnas.2006890117>.
- (19) Pfaff, S. J.; Fletterick, R. J. Hormone Binding and Co-Regulator Binding to the Glucocorticoid Receptor Are Allosterically Coupled. *J. Biol. Chem.* **2010**, *285* (20), 15256–15267. <https://doi.org/10.1074/jbc.M110.108118>.
- (20) Köhler, C.; Carlström, G.; Gunnarsson, A.; Weininger, U.; Tångeford, S.; Ullah, V.; Lepistö, M.; Karlsson, U.; Papavoine, T.; Edman, K.; et al. Dynamic Allosteric Communication Pathway Directing Differential Activation of the Glucocorticoid Receptor. *Sci. Adv.* **2020**, *6* (29), eabb5277. <https://doi.org/10.1126/sciadv.abb5277>.
- (21) Edman, K.; Hosseini, A.; Bjursell, M. K.; Aagaard, A.; Wissler, L.; Gunnarsson, A.; Kaminski, T.; Köhler, C.; Bäckström, S.; Jensen, T. J.; et al. Ligand Binding Mechanism in Steroid Receptors: From Conserved Plasticity to Differential Evolutionary Constraints. *Structure* **2015**, *23* (12), 2280–2290. <https://doi.org/10.1016/j.str.2015.09.012>.
- (22) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. Dynamical Networks in tRNA:Protein Complexes. *Proc. Natl. Acad. Sci.* **2009**, *106* (16), 6620–6625. <https://doi.org/10.1073/pnas.0810961106>.
- (23) Rivalta, I.; Sultan, M. M.; Lee, N.-S.; Manley, G. A.; Loria, J. P.; Batista, V. S. Allosteric Pathways in Imidazole Glycerol Phosphate Synthase. *Proc. Natl. Acad. Sci.* **2012**, *109* (22), E1428–E1436. <https://doi.org/10.1073/pnas.1120536109>.
- (24) Ricci, C. G.; Silveira, R. L.; Rivalta, I.; Batista, V. S.; Skaf, M. S. Allosteric Pathways in the PPAR γ 3-RXR α Nuclear Receptor Complex. *Sci. Rep.* **2016**, *6* (1), 1–17. <https://doi.org/10.1038/srep19940>.
- (25) Palermo, G.; Ricci, C. G.; Fernando, A.; Basak, R.; Jinek, M.; Rivalta, I.; Batista, V. S.; McCammon, J. A. Protospacer Adjacent Motif-Induced Allostery Activates CRISPR-Cas9. *J. Am. Chem. Soc.* **2017**, *139* (45), 16028–16031. <https://doi.org/10.1021/jacs.7b05313>.
- (26) Wagner, J. R.; Lee, C. T.; Durrant, J. D.; Malmstrom, R. D.; Feher, V. A.; Amaro, R. E. Emerging Computational Methods for the Rational Discovery of Allosteric Drugs. *Chem. Rev.* **2016**, *116* (11), 6370–6390. <https://doi.org/10.1021/acs.chemrev.5b00631>.
- (27) Schueler-Furman, O.; Wodak, S. J. Computational Approaches to Investigating Allostery. *Curr. Opin. Struct. Biol.* **2016**, *41*, 159–171. <https://doi.org/10.1016/j.sbi.2016.06.017>.
- (28) East, K. W.; Newton, J. C.; Morzan, U. N.; Narkhede, Y. B.; Acharya, A.; Skeens, E.; Jogl, G.; Batista, V. S.; Palermo, G.; Lisi, G. P. Allosteric Motions of the CRISPR-Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. *J. Am. Chem. Soc.* **2020**, *142* (3), 1348–1358. <https://doi.org/10.1021/jacs.9b10521>.
- (29) Lockless, S. W.; Ranganathan, R. Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science (80-.)*. **1999**, *286* (5438), 295–299. <https://doi.org/10.1126/science.286.5438.295>.
- (30) Süel, G. M.; Lockless, S. W.; Wall, M. A.; Ranganathan, R. Evolutionarily Conserved Networks of Residues Mediate Allosteric Communication in Proteins. *Nat. Struct. Biol.* **2003**, *10* (1), 59–69. <https://doi.org/10.1038/nsb881>.
- (31) Anishchenko, I.; Ovchinnikov, S.; Kamisetty, H.; Baker, D. Origins of Coevolution between Residues Distant in Protein 3D Structures. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (34), 9122–9127. <https://doi.org/10.1073/pnas.1702664114>.
- (32) Hjortness, M. K.; Riccardi, L.; Hongdusit, A.; Zwart, P. H.; Sankaran, B.; De Vivo, M.; Fox, J. M. Evolutionarily Conserved Allosteric Communication in Protein Tyrosine Phosphatases. *Biochemistry* **2018**, *57* (45), 6443–6451. <https://doi.org/10.1021/acs.biochem.8b00656>.

- (33) Reynolds, K. A.; McLaughlin, R. N.; Ranganathan, R. Hot Spots for Allosteric Regulation on Protein Surfaces. *Cell* **2011**, *147* (7), 1564–1575. <https://doi.org/10.1016/j.cell.2011.10.049>.
- (34) East, K. W.; Skeens, E.; Cui, J. Y.; Belato, H. B.; Mitchell, B.; Hsu, R.; Batista, V. S.; Palermo, G.; Lisi, G. P. NMR and Computational Methods for Molecular Resolution of Allosteric Pathways in Enzyme Complexes. *Biophys. Rev.* **2020**, *12* (1), 155–174. <https://doi.org/10.1007/s12551-019-00609-z>.
- (35) Riccardi, L.; Nguyen, P. H.; Stock, G. Construction of the Free Energy Landscape of Peptide Aggregation from Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2012**, *8* (4), 1471–1479. <https://doi.org/10.1021/ct200911w>.
- (36) Duan, M.; Liu, N.; Zhou, W.; Li, D.; Yang, M.; Hou, T. Structural Diversity of Ligand-Binding Androgen Receptors Revealed by Microsecond Long Molecular Dynamics Simulations and Enhanced Sampling. *J. Chem. Theory Comput.* **2016**, *12* (9), 4611–4619. <https://doi.org/10.1021/acs.jctc.6b00424>.
- (37) Lusher, S. J.; Raaijmakers, H. C. A.; Vu-Pham, D.; Kazemier, B.; Bosch, R.; McGuire, R.; Azevedo, R.; Hamersma, H.; Dechering, K.; Oubrie, A.; et al. X-Ray Structures of Progesterone Receptor Ligand Binding Domain in Its Agonist State Reveal Differing Mechanisms for Mixed Profiles of 11 β -Substituted Steroids. *J. Biol. Chem.* **2012**, *287* (24), 20333–20343. <https://doi.org/10.1074/jbc.M111.308403>.
- (38) Jin, L.; Li, Y. Structural and Functional Insights into Nuclear Receptor Signaling. *Adv Drug Deliv Rev* **2010**, *62* (13), 1218–1226. <https://doi.org/10.1016/j.addr.2010.08.007>.
- (39) Broekema, M. F.; Hollman, D. A. A.; Koppen, A.; Van Den Ham, H. J.; Melchers, D.; Pijnenburg, D.; Ruijtenbeek, R.; Van Mil, S. W. C.; Houtman, R.; Kalkhoven, E. Profiling of 3696 Nuclear Receptor-Coregulator Interactions: A Resource for Biological and Clinical Discovery. *Endocrinology* **2018**, *159* (6), 2397–2406. <https://doi.org/10.1210/en.2018-00149>.
- (40) Wu, Y.; Chin, W. W.; Wang, Y.; Burris, T. P. Ligand and Coactivator Identity Determines the Requirement of the Charge Clamp for Coactivation of the Peroxisome Proliferator-Activated Receptor γ . *J. Biol. Chem.* **2003**, *278* (10), 8637–8644. <https://doi.org/10.1074/jbc.M210910200>.
- (41) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>.
- (42) Petrović, D.; Wang, X.; Strodel, B. How Accurately Do Force Fields Represent Protein Side Chain Ensembles? *Proteins Struct. Funct. Bioinforma.* **2018**, *86* (9), 935–944. <https://doi.org/10.1002/prot.25525>.
- (43) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174. <https://doi.org/10.1002/jcc.20035>.
- (44) 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.* **1983**, *79* (2), 926–935.
- (45) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472. [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H).
- (46) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.
- (47) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190. <https://doi.org/doi:http://dx.doi.org/10.1063/1.328693>.
- (48) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 14101. <https://doi.org/doi:http://dx.doi.org/10.1063/1.2408420>.
- (49) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- (50) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics

- through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11* (4), 1864–1874. <https://doi.org/10.1021/ct5010406>.
- (51) Amadei, A.; Linssen, A. B. M.; de Groot, B. L.; van Aalten, D. M. F.; Berendsen, H. J. C. An Efficient Method for Sampling the Essential Subspace of Proteins. *J. Biomol. Struct. Dyn.* **1996**, *13* (4), 615–625. <https://doi.org/10.1080/07391102.1996.10508874>.
 - (52) David, C. C.; Jacobs, D. J. Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. *Methods Mol. Biol.* **2014**, *1084*, 193–226. https://doi.org/10.1007/978-1-62703-658-0_11.
 - (53) Girvan, M.; Newman, M. E. J. Community Structure in Social and Biological Networks. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (12), 7821–7826. <https://doi.org/10.1073/pnas.122653799>.
 - (54) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095. <https://doi.org/10.1021/ct400341p>.
 - (55) Eargle, J.; Luthey-Schulten, Z. NetworkView: 3D Display and Analysis of Protein Á RNA Interaction Networks. *Bioinforma. Appl.* **2012**, *28* (22), 3000–3001. <https://doi.org/10.1093/bioinformatics/bts546>.
 - (56) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604–613.
 - (57) Prinz, J. H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schtte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011**, *134* (17), 174105. <https://doi.org/10.1063/1.3565032>.
 - (58) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. <https://doi.org/10.1038/msb.2011.75>.
 - (59) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14* (6), 1188–1190. <https://doi.org/10.1101/gr.849004>.
 - (60) Kamisetty, H.; Ovchinnikov, S.; Baker, D. Assessing the Utility of Coevolution-Based Residue-Residue Contact Predictions in a Sequence- and Structure-Rich Era. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (39), 15674–15679. <https://doi.org/10.1073/pnas.1314045110>.

Table of Contents (TOC)/Abstract Graphic



Supplemental Information

Unraveling the allosteric cross-talk between coactivator peptide and ligand binding site in glucocorticoid receptor

Giuseppina La Sala¹, Anders Gunnarsson², Karl Edman², Christian Tyrchan³, Anders Hogner¹ and Andrey I. Frolov¹

1. Medicinal Chemistry, Research and Early Development, Cardiovascular, Renal and Metabolism (CVRM), BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.
2. Discovery Science, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.
3. Medicinal Chemistry, Research and Early Development, Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.

Corresponding authors:

giuseppina.lasala@astrazeneca.com

andrey.frolov@astrazeneca.com

1. List of UniProt codes employed for multiple sequence alignment (MSA) of 3-ketosteroid nuclear receptors.

O73673, P04150, P06537, P06536, O46567, P59667, O13186, Q6XLJ0, Q5R9P5, P49843, Q9N1U3, P49115, P79269, P79686, Q95267, P49844, P08235, Q8VII8, P22199, Q9IAC6, Q3YC04, Q9N0W8, Q4JM28, Q29131, Q91573, Q9BDJ7, P79404, Q9TT90, O97776, P49699, P10275, O97952, P19091, O97960, O97775, Q6QT55, Q9GKL7, P15207, Q7T1K4, P06401, P07812, Q00175, Q63449, Q9GLW0, P06186, A7X8B7, A7X8B3, A7X8C7, A7X8B5, A7X8D2, A7X8C4, A7X8C2, A7X8B9, A7X8D4, A7XW25, A7XW16, A7XW20, A7X8C9, P79373, Q8AYI2, Q28590.

2. Multiple sequence alignment of 57 peptide sequences belonging to Cluster II in the work of Broekema et al. ¹. Logo sequence representation is depicted in Figure SI 10.

>TIF1A_747_769

ESRPQNANYPRSILTSLLLNSSQ----

>NR0B1_136_159

---GEDHPRQGSILYSLLTSSKQTHVA

>ZNHI3_89_111

-LQNLKNLGESATLRSLLLNPFLR---

>PRGC1_130_155

---PPQEAEEPSLLKKLLLAPANT---

>NCOA1_1421_1441

SGPQTPQAQQKSLLQQLLTEA-----

>PPRC1_1159_1181

QAFISEIGIEASDLSSLLEQFEK----

>PRGC2_338_358

-----AEFSILRELLAQDVLCDVS

>NR0B2_106_128

-TFEVAEAPVPSILKKILLEEPSS---

>JHD2C_2054_2076

-PLVSQNNEQGSTLRDLLTTTAGK---

>NCOA3_725_747

EQLSPKKKENNALLRYLLDRDDP----

>NCOA2_733_755

-EPVSPKKKENALLRYLLDKDDTK---

>TREF1_168_190

TQSAVMDGAPDSALRQLLSQKPM----

>LCOR_40_62
 TTSPTAATTQNPVLSKLLMADQD----
 >BRD8_254_276
 TVAASPAASGAPTLSRLLEAGPT----
 >UBE3A_649_671
 RDLGDSHPVLYQSLKDLLEYEGN----
 >HAIR_745_767_C755S/C759S
 EDRAGRGLPSPSLSELLASTAV----
 >NRIP1_253_275_C263S
 PATSPKPSVASSQLALLSSEAH----
 >MED1_632_655
 VSSMAGNTKNHPMLMNLLKDNPAQ---
 >MED1_591_614
 HGEDFSKVSQNPILTSLLQITGNG---
 >NRIP1_488_510
 KNSKLNSHQKVTLLQLLLGHKNE----
 >NRIP1_1055_1077
 EKDSPRLTKTNPILYYMLQKGGN----
 >NRIP1_368_390
 -RNNIKQAANNSLLLHLLKSQTIP---
 >NR0B1_68_90_C69S
 -FSGKDHPRQGSILYSMLTSAKQT---
 >NCOA2_677_700
 STHGTSLKEKHILHRLQLDSSSP---
 >NCOA1_677_700
 PSSHSSLTERHKILHRLQEGSPS---
 >NCOA3_673_695
 -MHGSSLQEKHRILHKLLQNGNSP---
 >WIP1_119_141
 ESIYIHNIKDMKLLKTLLDIPAN----
 >NCOA3_609_631
 -QRGPLESKGHKLLQLLTCSSDD---
 >TIP60_476_498
 LSEDIVDGERAMLKRLLRIDSK----

>NCOA3_609_631_C627S
 -QRGPLESKGHKLLQLLTSSSD---
 >NCOA2_628_651
 GQSRLHDSKGQTKLLQLLTTKSDQ---
 >NCOA1_737_759
 -ASKKKESKDHQLRLRYLLDKDEKD---
 >TRXR1_132_154
 GHGPTLKAYQEGRLQKLLKMNGP----
 >PROX1_57_79
 SVVQHADGEKSNVLRKLLKRANS----
 >UBE3A_396_418
 DDEEPIPESELTLQELLGEERR----
 >NCOA6_1479_1501
 -LVSPAMREAPTSLSQLLDNSGAP---
 >PR285_1105_1127
 QQARLYENLPPAALRKLLRAEPE----
 >ILK_131_153
 KYGEMPVDKAKAPLRELLRERAE----
 >NRIP1_924_946
 -RSWARESKSFNVLKQLLLSENCV---
 >PELP1_446_468
 AGMLQGGASGEALLTHLLSDISPA---
 >NCOA6_875_897
 -PVNKDVTLTSPLLVNLLQSDISA---
 >NR0B1_1_23
 -MAGENHQWQGSILYNMLMSAKQT---
 >EP300_69_91
 -GMVQDAASKHKQLSELLRSGSSP---
 >CBP_57_80
 GNLVPDAASKHKQLSELLRGGSGS---
 >NCOR1_2376_2398
 SSTGSTQFPYNPLTMRMLSSTPP----
 >PPRC1_151_173
 DSELLVSPREGSSLHKLLTLSRT----

>NSD1_894_916
 SSQNHPIEPDYKFSTLLMMLKD----
 >MLL2_4175_4197
 LLAGPRSEAGHLLLQKLLRAKNV----
 >NRBF2_128_150
 PEIQGIFDRDPDTLLYLLQQKSE----
 >NCOA1_620_643
 SDGDSKYSQTSHKLVQLLTTTAEQ---
 >PRGC2_146_166
 ---PAPEVDELSLLQKLLLATSYP---
 >NRIP1_120_142
 VDSVPGKGQDSTLLASLLQSFSS----
 >NRIP1_121_143_P124R
 -DSVRKGKQDSTLLASLLQSFSSR---
 >NRIP1_924_946_C945S
 -RSWARESKSFNVLKQLLLSENSV---
 >NRIP1_805_831
 VSPQDFSFSKNGLLSRLLRQNQDSYL-
 >IKBB_277_299
 -PLGSAMLRPNPILARLLRAHGAP---
 >NR0B2_9_31_C9S/C11S
 -SPSQGAASRPAILYALLSSSLKA---

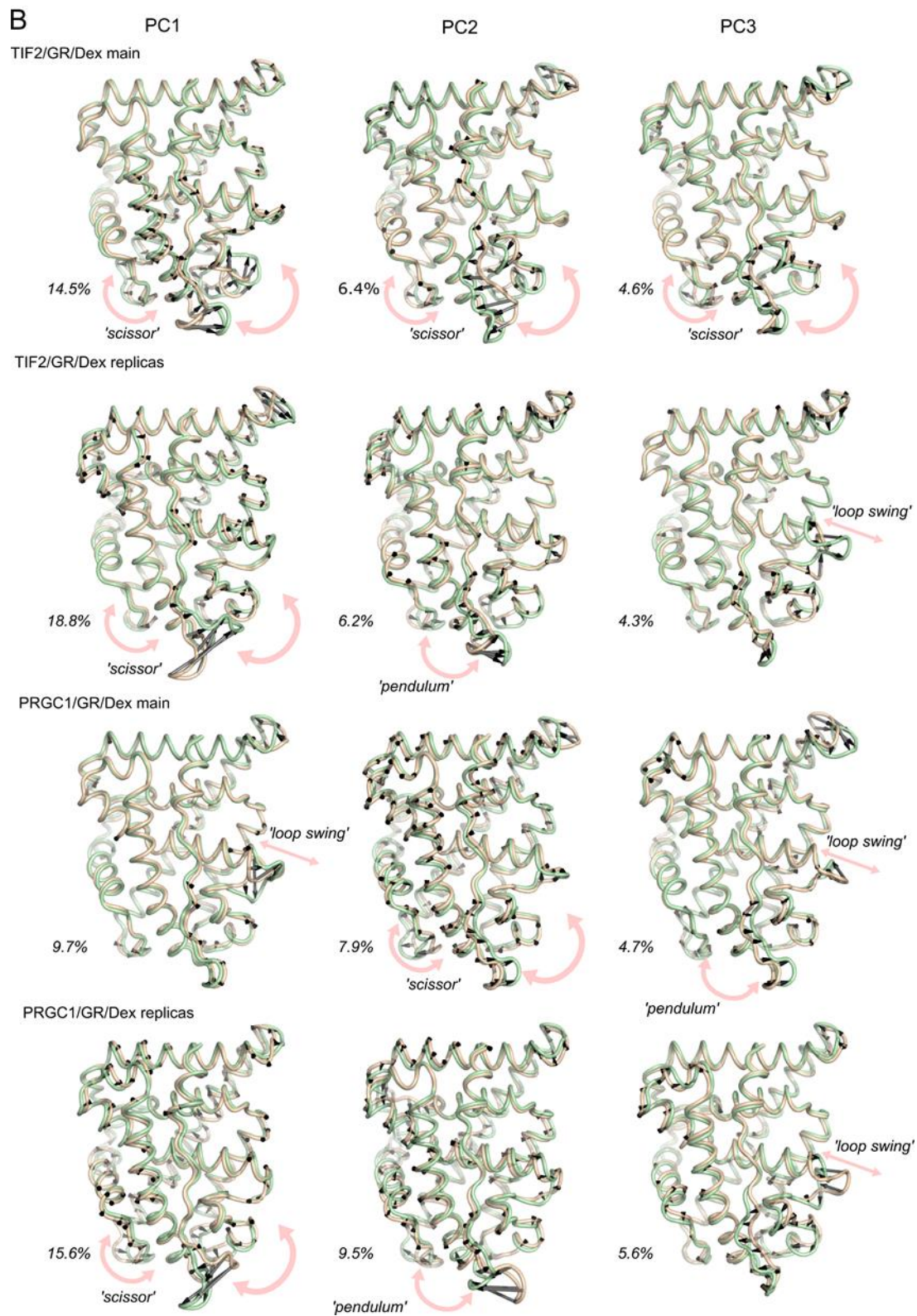
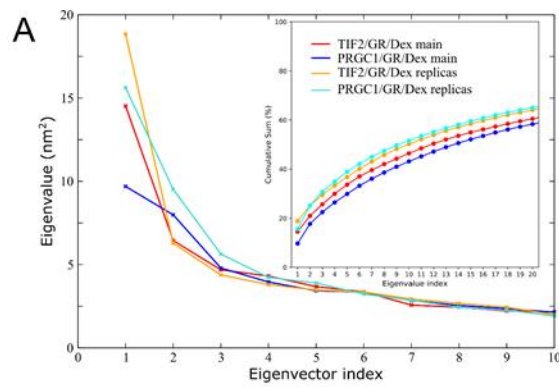


Figure S1. A) Eigenvalue Scree plot for first 10 modes of TIF2/GR/Dex main (red) and replica trajectories (orange) and PRGC1/GR/Dex main (blue) and replica trajectories (cyan). The inset show the cumulative proportion of the total variance generated from a PCA for the first 20 modes. B) The PC1, PC2 and PC3 of TIF2/GR/Dex main trajectory, TIF2/GR/Dex replicas, PRGC1/GR/Dex main trajectory and PRGC1/GR/Dex replicas are represented on GR_{LBD} by back arrows of sizes proportional to the amplitude of motion. The pink arrows highlights the most predominant motion for each PC, which can be a ‘scissor’, a ‘pendulum’ or a ‘loop swing’ motion. The variance captured for each mode is expressed in percentage.

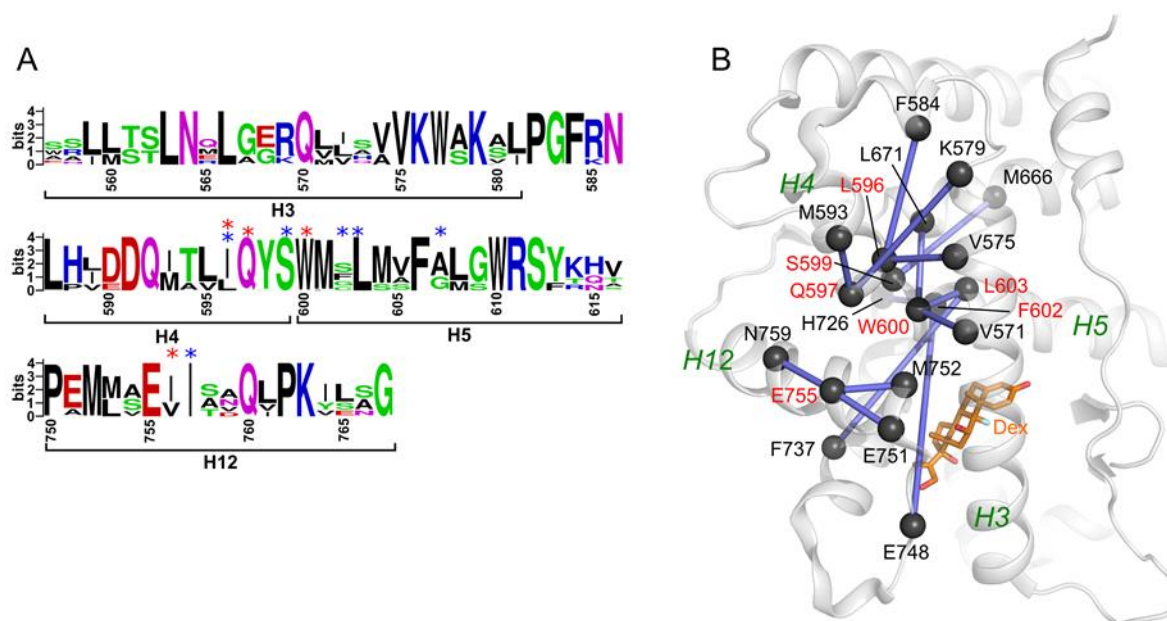


Figure S2. A) Sequence logo alignment of AF-2 region of 3-ketosteroid nuclear receptors. The asterisks indicate the critical nodes found in the dynamical network analysis of TIF2/GR/Dex (red) and PRGC1/GR/Dex (blue). B) Pairs of coevolving residues in AF-2 region identified with the GREMLIN pseudolikelihood method are mapped onto 3D structure of GR_{LBD} (PDB: 4UDC). Only pairs of residues having a GREMLIN score > 1 have been shown. Peptide has been removed for clarity. Dexamethasone (Dex) is represented as orange sticks. Red labels indicate residues present in ‘route 2’. E755 has been highlighted because forms the ‘charge clamp’, an interaction key for the peptide binding.

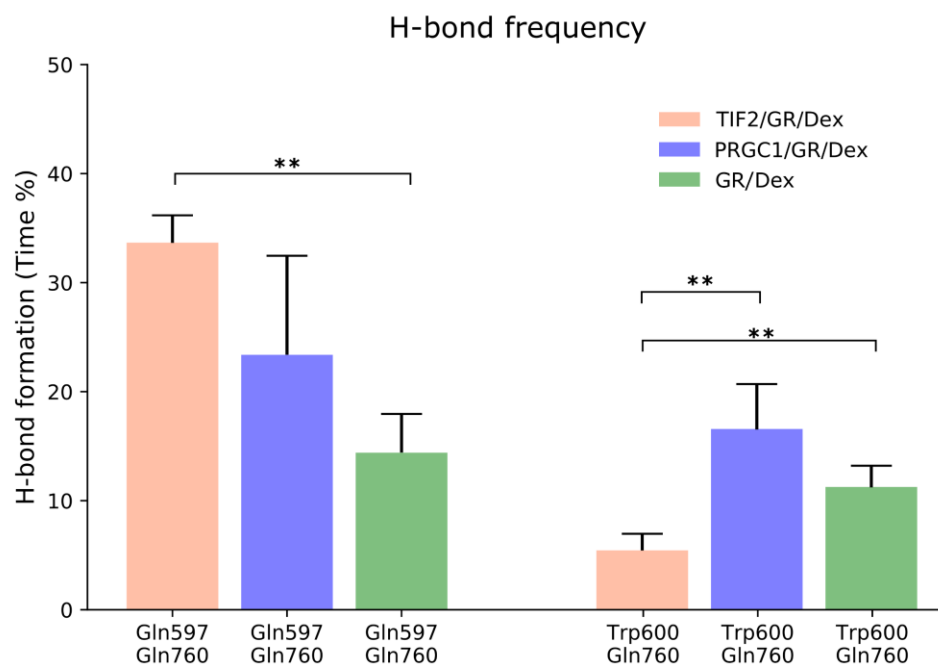


Figure S3. Frequency of formation of the Gln597-Gln760 H-bond and the Trp600-Gln760 H-bond along 5 replica MD simulations of TIF2/GR/Dex (orange), PRGC1/GR/Dex (blue) and GR/Dex (green). Error bars correspond to the standard error. (* indicates $p=0.05$ and ** indicates $p<0.05$, if there are not notations, then $p>0.05$. Mann-Whitney test U is performed)

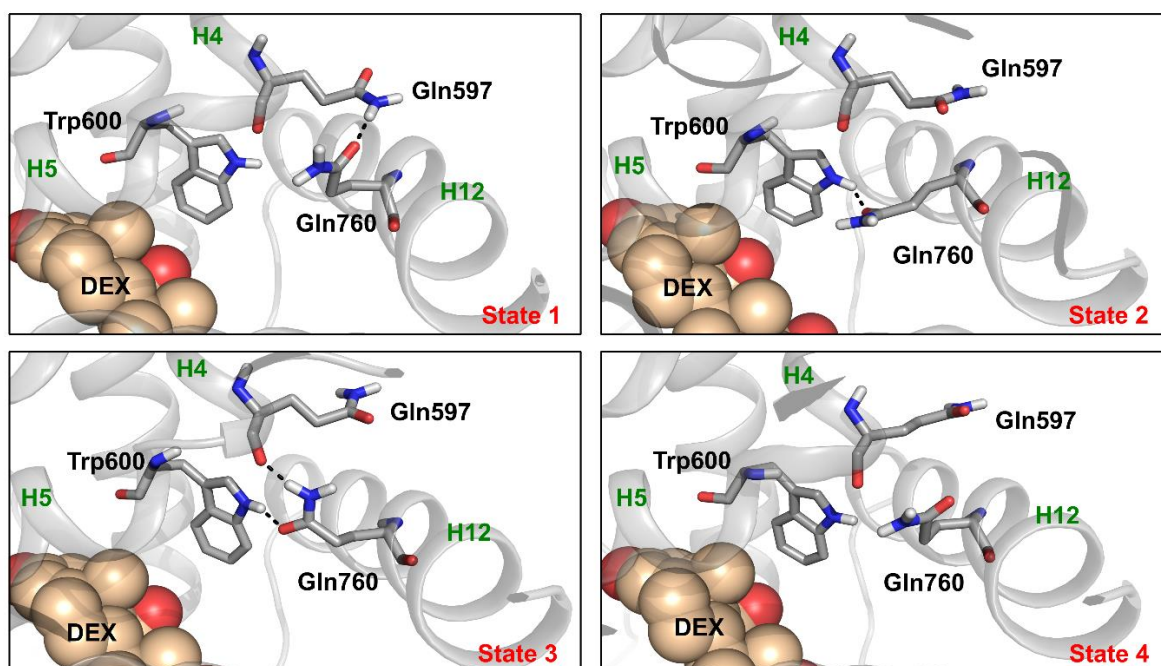


Figure S4. Frames extracted from MD simulations showing the four discrete states described by the transition matrix: state 1 (only side chain/side chain Gln597-Gln760 H-bond is formed); state 2 (only side chain/side chain Trp600-Gln760 H-bond); state 3 (both backbone/side chain Gln597-Gln760 H-bond and side chain/side chain Trp600-Gln760 H-bond are formed) and state 4 (no H-bonds are formed). Dexamethasone is represented as orange spheres, while helices H4, H5 and H12 are depicted as cartoons.

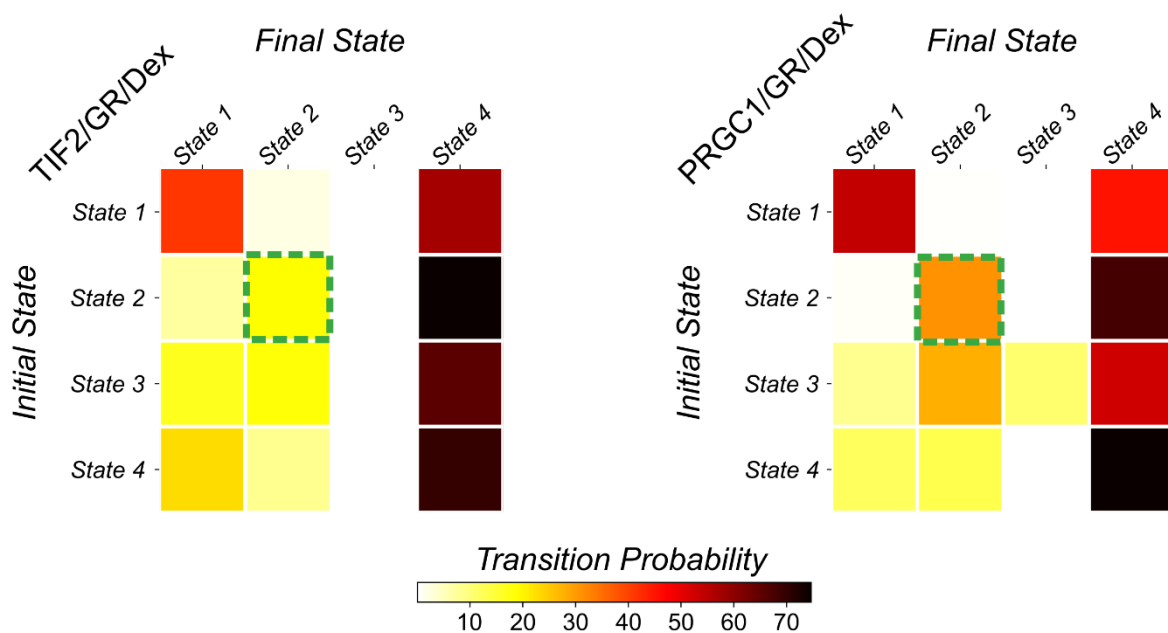


Figure S5. Transition matrix for TIF2/GR/Dex (left) and PRGC1/GR/Dex (right). The four states reported in the main text are labelled as state 1 (only side chain/side chain Gln597-Gln760 H-bond is formed); state 2 (only side chain/side chain Trp600-Gln760 H-bond); state 3 (both backbone/side chain Gln597-Gln760 H-bond and side chain/side chain Trp600-Gln760 H-bond are formed) and state 4 (no H-bonds are formed). The green dotted box highlights the portion of the transition matrix investigated in this work, which is the likelihood of the system to remain in state 2.

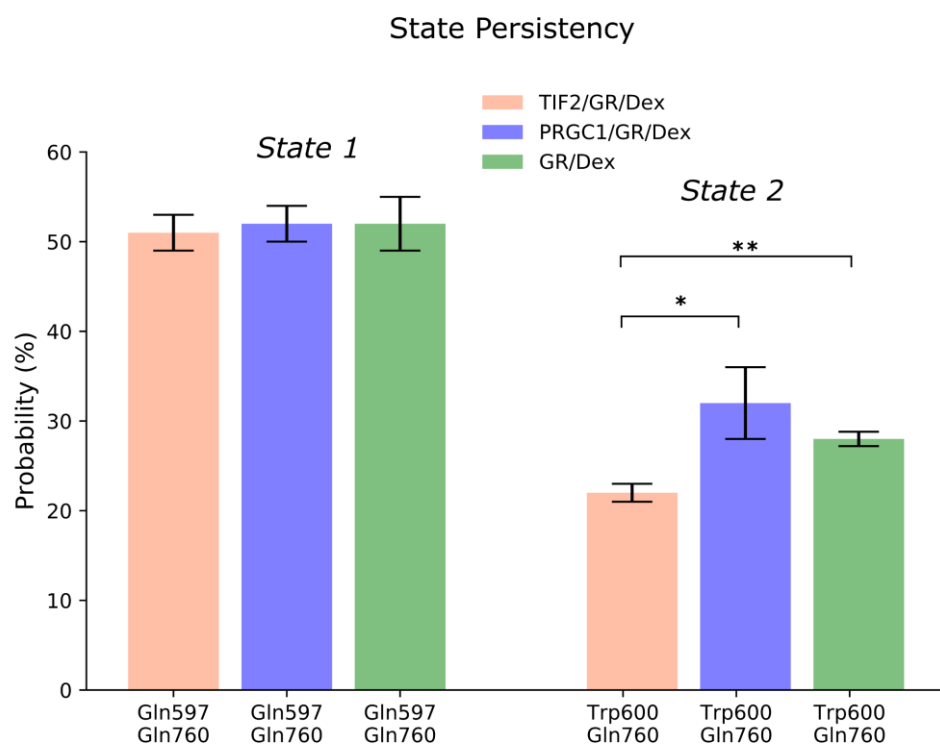


Figure S6. Probability expressed in percentage to stay in either state 1 (side chain/side chain Gln597-Gln760 H-bond) or state 2 (side chain/side chain Trp600-Gln760 H-bond). Values are computed for the 5 replica MD simulations of TIF2/GR/Dex (orange), PRGC1/GR/Dex (blue) and GR/Dex (green). Error bars correspond to the standard error. (* indicates $p=0.05$ and ** indicates $p<0.05$, if there are not notations, then $p>0.05$. Mann-Withney test U is performed)

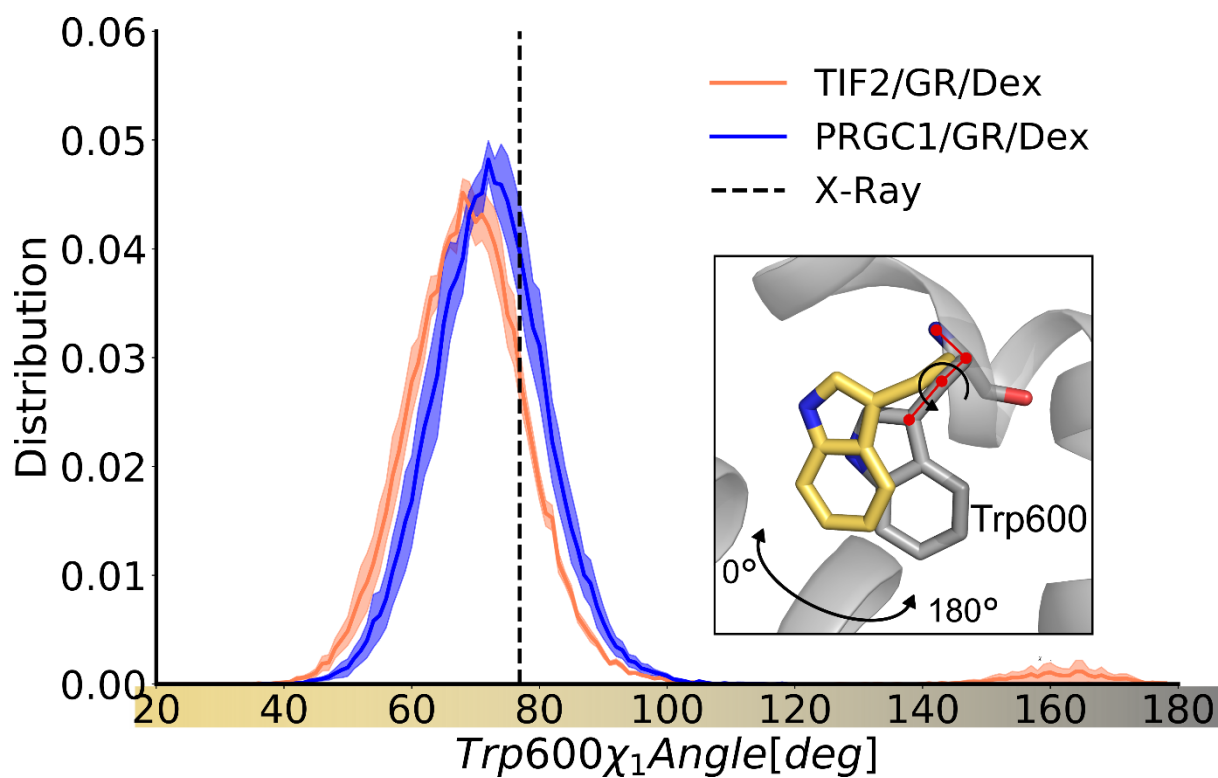


Figure S7. Trp600 χ_1 torsional angle distribution in TIF2/GR/Dex (orange) and PRGC1/GR/Dex (blue). The line represents the averaged value among the five MD replicas, while the transparent fill represent the standard error. The dashed black line measures the dihedral angle in the X-ray 4UDC. The inset represents the two limits of Trp600 configurations adopted in our MD simulations.

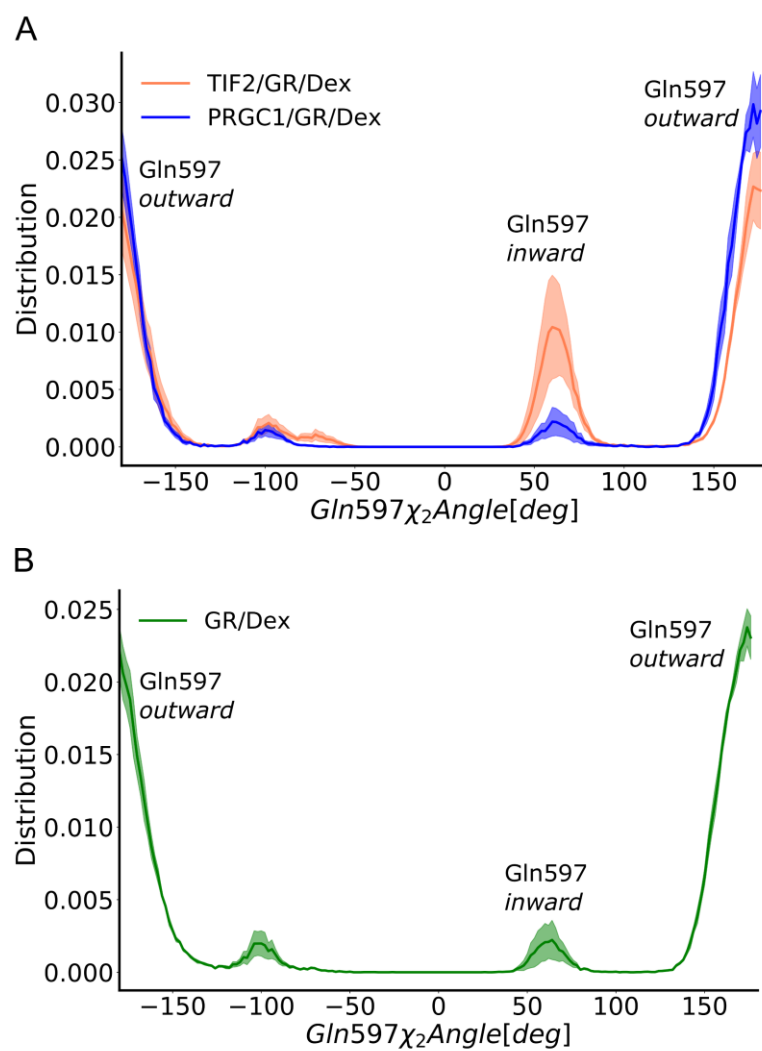


Figure S8. Distribution of Gln597 χ_2 dihedral angle in TIF2/GR/Dex (orange line) and PRGC1/GR/Dex (blue line); (A) and in GR/Dex (green line) (B). The line represents the averaged value among the five MD replica, while the transparent fill represent the standard error.

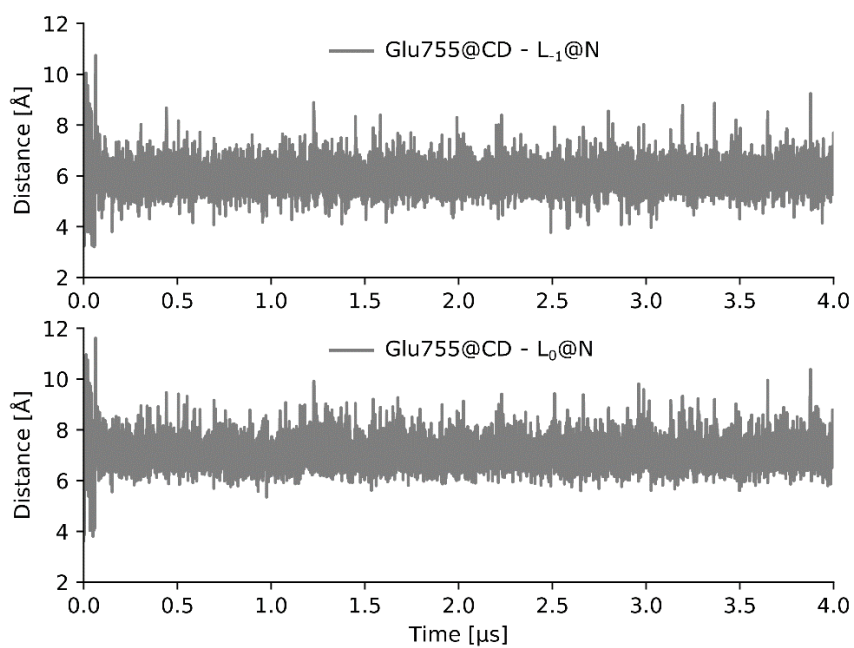


Figure S9. Distance along time between Glu755 C δ atom and L₁ N atom (upper plot) and Glu755 C δ atom and L₀ N atom (bottom plot) in TIF2/GR/Dex trajectory.

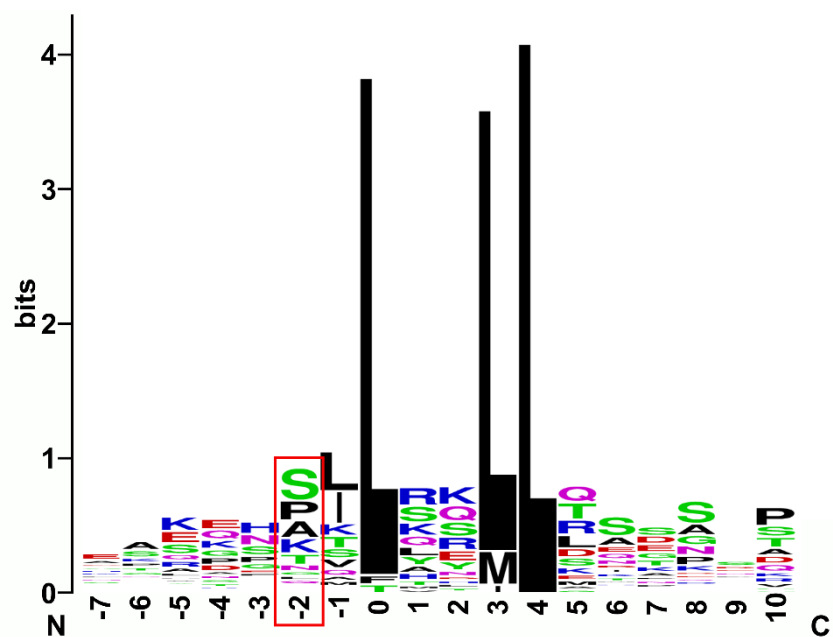


Figure S10. Sequence logo alignment of the 57 peptides sequences belonging to Cluster II in Broekema et al.¹ Position -2 highlighted as red square. WebLogo webserver² was used to obtain the LOGO representation from the multiple sequence alignment (MSA) of the 57 sequences. Graphic manipulation was performed using Inkscape.

Chains		Lys576	Thr578	Lys579	Arg585	Asp590	Gln597	Glu751	Glu755	Asn759
1M2Z	A/B									
	C/D									
1P93	A/E									
	B/F									
	C/G									
	D/H									
3CLD	A/H									
	B/C									
3E7C	A/C									
	B/D									
3K22	A/H									
	B/D									
3K23	A/D									
	B/E									
	C/F									
4CSJ										
4P6W										
4P6X	A/B									
	C/D									
	E/F									
	G/H									
	I/J									
	K/L									
4UDC										
4UDD										
5G3J										
5G5W										
5NFP										
5NFT										
6EL6										
6EL7										
6EL9										
Total										
Occurrence (%)		3,1	3,1	96,8	56,2	18,7	37,5	3,1	62,5	43,7

Table S1. List of available GR_{LBD} X-ray structures in complex with the TIF2 peptide. The first two columns indicate the PDB code and the chains (GR_{LBD} and peptide), respectively. The green box indicates the presence of an H-bond between TIF2 and the residues in the AF-2 docking site. For each interaction, the total occurrence expressed in percentage is reported in the last row.

TIF2/GR/Dex			
	H-bond	Occupancy(%)	Simulated Time (μ s)
Replica1	Gln597 - N-3	59.2	2.5
Replica2	Gln597 - N-3	82.7	0.8
Replica3	Gln597 - N-3	39.0	0.8
	Glu755 - N-3	23.9	
		23.0	
Replica4	Gln597 - N-3	/	1.6
Replica5	Gln597 - N-3	/	1.5
	Gln597 - A-2	23.8	

Table S2. The table shows the occupancy of H-bonds between TIF2 peptide and both Gln597 and Glu755. Only H-bonds with an occupancy higher than 20% are reported. The last column indicated the total simulated time for each replica.

PRGC1/GR/Dex			
	H-bond	Occupancy(%)	Simulated Time (μ s)
Replica1	Glu755 - S-2	20.3	2.0
		23.5	
Replica2	Glu755 - S-2	32.3	0.6
		38.4	
	Glu755 - L-1	24.4	
		23.0	
Replica3	Glu755 - S-2	31.7	0.7
		34.8	
	Glu755 - L-1	31.3	
		24.9	
Replica4	Glu755 - S-2	27.7	0.7
		35.6	
	Glu755 - L-1	24.0	
Replica5	Glu755 - L-1	21.5	0.7
	Glu755 - S-2	/	

Table S3. The table shows the occupancy of H-bonds between PRGC1 peptide and both Gln597 and Glu755. Only H-bonds with an occupancy higher than 20% are reported. The last column indicated the total simulated time for each replica.

References

- (1) Broekema, M. F.; Hollman, D. A. A.; Koppen, A.; Van Den Ham, H. J.; Melchers, D.; Pijnenburg, D.; Ruijtenbeek, R.; Van Mil, S. W. C.; Houtman, R.; Kalkhoven, E. Profiling of 3696 Nuclear Receptor-Coregulator Interactions: A Resource for Biological and Clinical Discovery. *Endocrinology* **2018**, *159* (6), 2397–2406. <https://doi.org/10.1210/en.2018-00149>.
- (2) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14* (6), 1188–1190. <https://doi.org/10.1101/gr.849004>.