JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

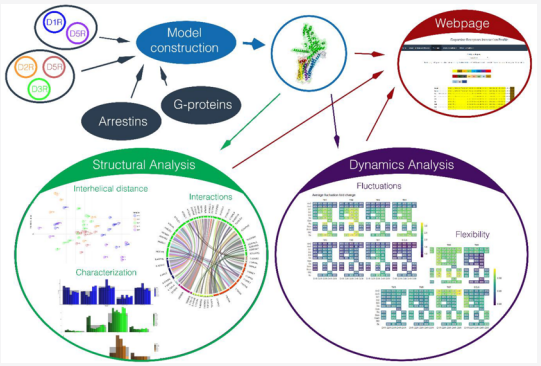
Understanding the Binding Specificity of G-Protein Coupled Receptors toward G-Proteins and Arrestins: Application to the Dopamine Receptor Family

A. J. Preto, Carlos A. V. Barreto, Salete J. Baptista, José Guilherme de Almeida, Agostinho Lemos, André Melo, M. Nátalia D. S. Cordeiro, Zeynep Kurkcuoglu, Rita Melo, and Irina S. Moreira*

Cite This: *J. Chem. Inf. Model.* 2020, 60, 3969−3984 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: G-Protein coupled receptors (GPCRs) are involved in a myriad of pathways key for human physiology through the formation of complexes with intracellular partners such as G-proteins and arrestins (Arrs). However, the structural and dynamical determinants of these complexes are still largely unknown. Herein, we developed a computational big-data pipeline that enables the structural characterization of GPCR complexes with no available structure. This pipeline was used to study a well-known group of catecholamine receptors, the human dopamine receptor (DXR) family and its complexes, producing novel insights into the physiological properties of these important drug targets. A detailed description of the protein interfaces of all members of the DXR family (D1R, D2R, D3R, D4R, and D5R) and the corresponding protein interfaces of their binding partners (Arrs: Arr2 and Arr3; G-proteins: Gi1, Gi2, Gi3, Go, Gob, Gq, Gslo, Gssh, Gt2, and Gz) was generated. To produce reliable structures of the DXR family in complex with either G-proteins or Arrs, we performed homology modeling using as templates the structures of the β2-adrenergic receptor (β2AR) bound to Gs, the rhodopsin bound to Gi, and the recently acquired neurotensin receptor-1 (NTSR1) and muscarinic 2 receptor (M2R) bound to arrestin (Arr). Among others, the work demonstrated that the three partner groups, Arrs and Gs- and Gi-proteins, are all structurally and dynamically distinct. Additionally, it was revealed the involvement of different structural motifs in G-protein selective coupling between D1- and D2-like receptors. Having constructed and analyzed 50 models involving DXR, this work represents an unprecedented large-scale analysis of GPCR-intracellular partner interface determinants. All data is available at www.moreiralab.com/resources/dxr.



INTRODUCTION

G-Protein coupled receptors (GPCRs) share a number of common structural features—they are composed by seven transmembrane domains (TM1-7), connected to each other either by intracellular loops (ICL1-3) in the cytoplasm or extracellular loops (ECL1-3) outside the cell. In addition, they also possess another common motif that runs parallel to the membrane interface, namely Helix-8 (H8), which was shown to participate in modulating the interaction between the receptor and its intracellular partners such as PDZ domain-containing proteins.[1,2] The amino and carboxyl termini of class A GPCRs reside in the extracellular and intracellular parts of the cell, respectively.[3,4] These two regions together with the ICL3, which links TM5 and TM6, exhibit the highest variability in terms of amino acid sequence among this GPCR class.

Dopamine receptors (DXRs), which are a family of class A GPCRs that endogenously bind dopamine,[5] were the subject of our study. Dopamine is a catecholaminergic neuromodulator involved in a plethora of physiological functions in the central nervous system, such as voluntary movements, event prediction, sleep, attention, learning, and memory.[6,7] Additionally, it is also involved in the peripheral nervous system through the modulation of hormones, cardiovascular and renal functions, and the sympathetic system. Dopamine and other related ligands exert their physiological and pharmacological effects through the activation of five closely related subtypes of DXR complexes. These are split into two major subclasses—D1-like receptors (D1R and D5R) and D2-like receptors (D2R, D3R, and D4R)—based on their ligand and G-protein subtype specificity, anatomical distribution, and physiological effects. GPCRs interact with both G-proteins and arrestins (Arrs), leading to a rich and diverse cross talk between different signaling pathways. Arrs are a small family of

Received: April 13, 2020
Published: July 21, 2020

Table 1. Functional Binding Partners of Each Member from the DXR Family[a]

	G-protein									arrestin	
	Gi1	Gi2	Gi3	Go	Gob	Gz	Gq	Gs	Gt2	Arr2	Arr3
D1R	22	22	22	23			24, 25	22, 23		26	26
D2R	27	27		28		29, 30				31	31, 32
D3R	33	33	33	33, 34		29, 30	25, 33, 35	30, 36		37	26, 37
D4R	38	38	38	38	39	29, 30			40	41	41
D5R						39	25, 39, 42	39		43	43

[a]Arrs and Gi/o, Gq/11, and Gs subfamily members were reported as interacting partners of DXR. The functional partners of D2R correspond to the D2R short isoform, which is used in the study. The numbers in the table correspond to the literature references.

intracellular scaffolding proteins, initially recognized as being solely responsible for GPCR desensitization. However, they were recently implied in G-protein independent signaling pathways as well.[8−10] Moreover, regarding nonvisual Arrs, which are particularly involved in interactions with different GPCRs classes, only two structures of Arr2, one complexed with neurotensin receptor-1 (NTSR1)[11] and another with muscarinic 2 receptor (M2R),[12] were solved so far. Nonetheless, the orientation of Arr2 in relation to each of those receptors is significantly different, suggesting that different bioactive Arr conformations can be implicated in GPCR activation.[11,12] Understanding structural determinants that mediate the differential coupling between a specific DXR subtype and different G-protein and arrestin subtypes may reveal the underlying molecular mechanisms that regulate several physiological and pharmacological outcomes associated with the DXR family. As such, differentiating specific structural motifs formed at protein−protein interfaces (PPIs) of DXR-G-protein and DXR-Arr complexes can help us delineate the molecular mechanism of functional selectivity, which is key to understand and modulate cellular events such as receptor internalization, recycling, trafficking to the endosome-lysosome pathway for proteolytic degradation, etc.

To model and characterize all PPIs of documented DXR-Arr complexes (D1R-Arr2_6pwc, D1R-Arr3_6pwc, D1R-Arr2_6u1n, D1R-Arr3_6u1n, D2R-Arr2_6pwc, D2R-Arr3_6pwc, D2R-Arr2_6u1n, D2R-Arr3_6u1n, D3R-Arr2_6pwc, D3R-Arr3_6pwc, D3R-Arr2_6u1n, D3R_Arr3_6u1n, D4R-Arr2_6pwc, D4R-Arr3_6pwc, D4R-Arr2_6u1n, D4R-Arr3_6u1n, D5R-Arr2_6pwc, D5R-Arr3_6pwc, D5R-Arr2_6u1n, D5R-Arr3_6u1n) as well as DXR-G-protein complexes (D1R-Gi1, D1R-Gi2, D1R-Gi3, D1R-Go, D1R-Gq, D1R-Gslo, D1R-Gssh, D2R-Gi1, D2R-Gi2, D2R-Go, D2R-Gz, D3R-Gi1, D3R-Gi2, D3R-Gi3, D3R-Go, D3R-Gq, D3R-Gslo, D3R-Gssh, D3R-Gz, D4R-Gi1, D4R-Gi2, D4R-Gi3, D4R-Go, D4R-Gob, D4R-Gt2, D4R-Gz, D5R-Gq, D5R-Gslo, D5R-Gssh, D5R-Gz), several software programs and web platforms were used, namely MODELLER,[13] high ambiguity driven protein−protein DOCKing (HADDOCK),[14] Bio3d,[15] InterProSurf,[16] biocomplexes contact maps (COCOMAPS),[17] conservation surface mapping (ConSurf),[18,19] the pocket volume measurer (POVME),[20] and elastic network modeling (ENM).[21] We assessed evolutionary conservation as well as structural (intermolecular interactions, salt bridges, hydrogen bonds, and solvent accessibility) and dynamic (fluctuations and flexibility) features. This allowed us to extensively characterize the DXR-partner interfaces and to understand the molecular determinants responsible for specific binding of DXR complexes to their cognate G-protein and Arr subtypes.

METHODS

Throughout this section we explain the methods used for each of the computational pipeline steps.

DXR Binding Partners. Functional binding partners of all members from the DXR family found in the literature were listed in Table 1.

Homology Modeling. Proteins were modeled using homology modeling with the MODELLER package.[13] This software allows the construction of 3D models for a protein with no known structure by using one or more proteins with structural information as templates. A key feature of this method is the utilization of templates that share high sequence homology with the modeled structure.[44] This methodology is still vital to study GPCRs at a molecular level as it overcomes the scarcity of available GPCR experimental structures. The β2-adrenergic receptor (β2AR) (PDB-ID: 3SN6, chain R[45]) was chosen as the template for G-protein activated DXRmon (the monomeric structure of the DXR) with UniProt[46] sequence IDs P21728 and P21918 for D1R and D5R, respectively. The rhodopsin-Gi complex (PDB-ID: 6CMO, chain A[47]) was chosen as the template for G-protein activated DXRmon with UniProt[46] sequence IDs P14416, P35462, and P21917 for D2R, D3R, and D4R, respectively. For Arr complexes, two structures complexed with Arr2 were used as templates for Arr-activated DXRmon for all DXR members: the NTSR1-Arr2 complex (PDB-ID: 6PWC, chain R[11]) and the M2R-Arr2 complex (PDB-ID: 6U1N, chain R[12]). These two structures differ drastically on the orientation of Arr2 in relation to the receptor, but the structural information available points to the coexistence of these two conformations;[11,12] therefore, both were included in this study. An Ala_n linker was added to connect TM5 and TM6, which were modeled with an extended helical segment (beyond the membrane) up to the linker, making the intracellular extension of these helices similar to that observed in the crystal structure of the β2AR-Gs complex (PDB-ID: 3SN6[45]). Clustal Omega[48] was used to generate Multiple Sequence Alignments (MSA) for all sequences in all three protein families−DXR, G-proteins, and Arrs. Monomer models were obtained using MODELLER;[13] the final alignments are available at the web server. TMs were defined as well as disulfide bonds. One hundred models were created for each query sequence. The 20 best models from each query were selected using discrete optimized protein energy score,[49] MODELLER objective function,[50] ProSA-web server Z-score,[51] and ProQ scores (LGScore and MaxSub using PSIPRED results).[52] The top models were then visually inspected using PyMOL[53] software where the intramembrane domain axis-ICL3 distance was taken into account. Model evaluation was further performed with root mean-squared deviation (RMSD) measurements of all the dopamine

https://dx.doi.org/10.1021/acs.jcim.0c00371
J. Chem. Inf. Model. 2020, 60, 3969−3984

receptors' with PyMOL software;[53] these are available on the "overall complex RMSD", "receptor RMSD", and "GPCR motif RMSD", on the STRUCTURE ANALYSIS tab.

The crystal structure of the Gs-protein (PDB-ID: 3SN6, chain A[45]) was used as a template for the construction of the 3D models of Gslo and Gssh, while the cryo-EM structure of the Gi-protein (PDB-ID: 6CMO, chain A[47]) was used for construction of the 3D models of Gi1, Gi2, Gi3, Go, Gob, Gq, Gt2, and Gz. The human NTSR1-Arr2 complex (PDB-ID: 6PWC, chain A[11]) and the M2R-Arr2 complex (PDB-ID 6U1N, chain A[12]) were both used as templates for the 3D models of Arrs. The accession codes of query sequences of Gi1, Gi2, Gi3, Go, Gob, Gq, Gssh, Gslo, Gt2, and Gz used for homology modeling were P63096, P04899, P08754, P09471-1, P09471-2, P50148, P63092-2, P63092-1, P19087, and P19086, respectively. The accession codes of Arr2 and Arr3 query sequences used for homology modeling were P49407-1 and P32121, respectively.

The DXR 3D selected models were then embedded in a lipid bilayer membrane of POPC:cholesterol (9:1 ratio) using the CHARMM-GUI web server.[54−56] The systems were subjected to an equilibration protocol to further relax the structure of the receptors. The equilibration was performed in GROMACS 2018.4[57−59] with the CHARMM36 force field.[60] The protocol consisted of 50 ns of equilibration with sequentially decreasing constraint forces until only Cα of DXR were constrained. The structure from the last snapshot was then selected to perform the HADDOCK[14] docking protocol.

Structure Refinement with HADDOCK. The refinement interface in the HADDOCK[14] web platform was used to perform protein structure refinement in an explicit solvent representation. To construct 3D models of DXR complexes-G-protein complexes, the models of DXR complexes and G-proteins were aligned based on the corresponding crystal structures of the β2AR-Gs complex (PDB-ID: 3SN6[45]) or the Rhodopsin-Gi (PDB-ID: 6CMO[47]), and the models for DXR-Arrs were aligned based on the two aforementioned structures: NTSR1-Arr2 complex (PDB-ID: 6PWC[11]) and M2R-Arr2 complex (PDB-ID: 6U1N[12]). These complexes were submitted to the refinement interface of the HADDOCK web server, which enables the movement of side chains and backbone at the interface, and the best model for each protein−protein complex was used in subsequent analyses.

Evolutionary Characterization of Complex Interfaces. The degree of positional conservation of specific amino acid residues has been linked to their importance in protein structure and function. Thus, the determination of the conserved amino acid positions among GPCR and GPCR-partners' family members may uncover the relevance of each position to structure and function of the receptor. ConSurf[18,19,61] was deployed to probe evolutionarily conserved position-specific amino acids and to identify structurally and functionally important regions within the proteins. This algorithm, coupled with its corresponding database ConsurfDB, takes as the basis MSA of target proteins and corresponding homologues (which are determined in the same pipeline where proteins with less than 50 homologues are discarded and homologues with over 95% redundancy are removed). A phylogenetic tree was calculated from the MSA, and evolutionary rates were assigned to each position in the alignment. The final grades range from a scale of 1 to 9, 1 as the lowest, 5 as intermediate, and 9 as highly conserved positions.

Structural and Dynamical Characterization of Protein−Protein Interfaces. InterProSurf[16] and COCOMAPS[17] were applied to fully characterize the most relevant structural features shared by all aforementioned DXR-G-protein and DXR-Arr complexes. Considering InterProSurf, complex related results regarding total, polar, and apolar area/energy as well as the number of surface and buried atoms were considered. COCOMAPS was used to predict hydrogen bonds (HB) and solvent accessible surface area (SASA), the latter being used to define a broad range of interfacial residues. In-house Python scripts were also constructed to perform additional analysis, such as salt bridges (SB) identification (considering a 4 Å threshold), detection of α-carbon interactions (using a 8 Å cutoff), and residue content by single amino acid or amino acid group. In order to perform binding pocket analysis, we deployed POVME[20] combined with an in-house script in which the pocket parameters were defined taking the starting point as the geometrical center between the residues 3.50, 6.30, and 7.53 to a sphere of 10 Å radius and considering three contiguous points as the threshold to merge adjoining areas. The binding pocket setup was deployed with a 0.5 Å grid spacing.

To comparatively characterize the flexibility of TMs and the H8, we performed a coarse-grained normal-mode analysis (NMA)[62] as implemented in Bio3d[15] to correctly assess changes in the flexibility of TMs and H8 for all DXRs upon the binding of different partners. We calculated the Bhattacharya coefficient (BC), a measure adapted from the Bhattacharya distance,[63] between each dopamine receptor in a complex−DXRcomp−and its respective monomeric structure−DXRmon. The BC, as first described in Fuglebakk et al.[64] and implemented in Bio3d,[15] is used to compare protein flexibility between conserved protein regions and is calculated for selected rows and columns of the covariance matrices (i.e., those pertaining to each of the different structural motifs−TM1-7 and H8) of both DXRcomp and DXRmon. This allowed us to quantify changes in protein flexibility upon binding that were normalized by the monomeric structure. Additionally, we computed the fold change in fluctuation for all residues in D_XR upon partner binding and averaged these values for each structural motif. Multidimensional scaling (MDS),[65] which consists of applying eigenvalue decomposition on the double-centered Euclidean distance matrix, was performed using the normalized flexibility of each structural motif as a vector. This allows us to assess in a two-dimensional space how different DXRcomp clusters are based on their identity (D1-like vs D2-like) and on the identity of their partners (Arr vs G-protein and within-G-protein calculations).

To better understand how well our characterization was able to discriminate DXRcomp, we calculated their structural phylogeny−a hierarchy that allowed us to relate all complexes. Considering that the features used herein to characterize different DXR complexes exist in different scales and dimensions, calculating a structural phylogeny for DXRcomp cannot consider all features in an equal manner. As such, we calculated a tree for each set of structural characteristics−namely InterProSurf, COCOMAPs, interhelical distances, and HB/SB−dynamic characteristics BC and fluctuation−and the composition of the interface in amino acid proportions. Considering these seven trees, each of which describes the structural phylogeny of all DXRcomp for a specific character-

https://dx.doi.org/10.1021/acs.jcim.0c00371
J. Chem. Inf. Model. 2020, 60, 3969−3984

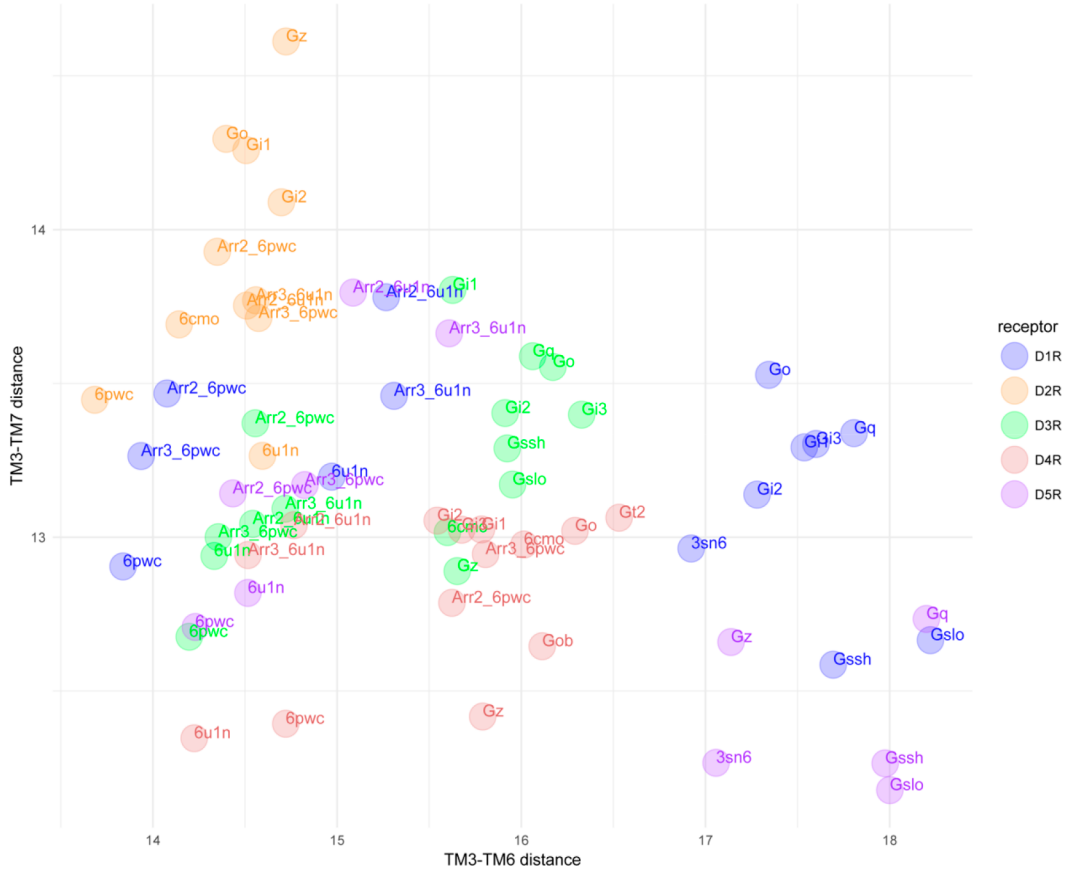


Figure 1. Interhelical distances plotted according to TM3-TM6 and TM3-TM7 distances. Residues 3.50, 6.30, and 7.53 were used for TM3, TM6, and TM7, respectively. Points are colored by type of receptor and labeled according to the specific partner. Points labeled as 3sn6, 6cmo, 6pwc, and 6u1n represent each receptor structure before coupling with partner. The distances were measured in Å.

istic, an average tree was calculated to yield the final structural phylogeny for all DXRcomp. This tree was calculated by minimizing the quadratic path difference between the final tree and the other seven trees as implemented in the phytools package[66] for the R language.[67] To visualize the final tree, the ape package[68] was used.

Web-Server Implementation. The amount of data generated in these analyses was too extensive to be clearly detailed and listed herein. As such we restricted it to the main important conclusions, and full information about the results can be found online at http://www.moreiralab.com/resources/dxr/. Static data visualization presented in this paper was performed using *ggplot2*,[69] *circlize*,[70] *seaborn*,[71] and *matplotlib*.[72] Web site construction was performed using *shiny*.[73] All intermediate calculations not previously referred were performed using either Python or R.

RESULTS

For prompt display of the results, a fully dynamic web server was constructed and can be easily accessed at www.moreiralab.com/resources/dxr. Throughout the Results section, references to the web site contents were made when appropriate, and the sections' names and subtitles have corresponding matches in the provided web site.

Structural Analysis. *Interhelical Distance.* In agreement with the procedure reported by Sandhu et al., we measured the distance between residues 3.50 and 7.53 (TM3-TM7) to understand the levels of GPCR activation in the different complexes (Figure 1).[74] In line with the previous step, we also measured the interhelical distance between residues 3.50 and 6.30 (TM3-TM6) and plotted them against each other; the opening of the binding crevice expected upon activation was seen for both DXR-Arrs and DXR-G-protein complexes, which tends to lead to higher TM3-TM6 distances in comparison to DXR alone. Arrs coupling tends to promote lower TM3-TM6 distances, as already observed.[75] Arrs and G-proteins tend to be clustered differently, depending on the DXR bound. In D1-like receptors, Gs is isolated from Gi and Gq subfamilies. Gz is the most isolated partner among the Gi/o subfamily in all DXR. Indeed, Gs in complex with D1-like receptors led to higher TM3-TM6 (around 17−18 Å) and lower TM3-TM7 distances (around 12−13 Å). Conversely, DXR-Gi complexes tend to promote lower TM3-TM6 (around 14−17 Å) and higher TM3-TM7 (around 13−15 Å) distances, especially if D2-like receptors are involved. Complexes with Arrs form different clusters depending on the template used. 6pwc derived models cluster within 14−15 Å in the TM3-TM6 distance, apart from D_4R complexes which are closer to 16 Å. Regarding the TM3-TM7 distance, complexes are more dispersed. Complexes modeled with 6U1N successfully differentiate D1-like and D2-like complexes. Regarding the TM3-TM6 distance, D1-like complexes fall into the 15.0−15.5 Å interval, while D2-like receptors fall in the 14.5 Å line. In the TM3-TM5 distance, D1-like receptors also showed higher values than D2-like, except for D2R-complexes that have the same values as D1-like complexes.

Electrostatic Distance. The electrostatic distance analysis of the different partners with DXR clearly (exemplified for D1R

https://dx.doi.org/10.1021/acs.jcim.0c00371
J. Chem. Inf. Model. 2020, 60, 3969−3984

in Figure 2) shows the separation of three groups: Arrs, Gs-proteins, and Gi-proteins. Furthermore, Gq is usually isolated either by itself or in a smaller group with Gz. Gz is usually isolated in the same manner. Gi1, Gi2, Gi3, and Go are usually closely related, forming two subgroups inside the Gi-proteins, in which they can alternate position depending on the receptor. Generally, Gi1 is more closely related with Gi3. The same happens between Gi2 and Go. Regarding the two different templates used to model the Arrs (6pwc and 6u1n), they lead to models with similar electrostatic distance profiles, commonly clustered in the same branch of the tree. It should be noted, however, that at D3R, this branch is shared with Gs-proteins, and in D4R, Gob appears closely related to complexes modeled with 6U1N. In D5R, there is the most striking dissimilarity, with complexes modeled with 6pwc being in a different branch from all the remaining complexes modeled.

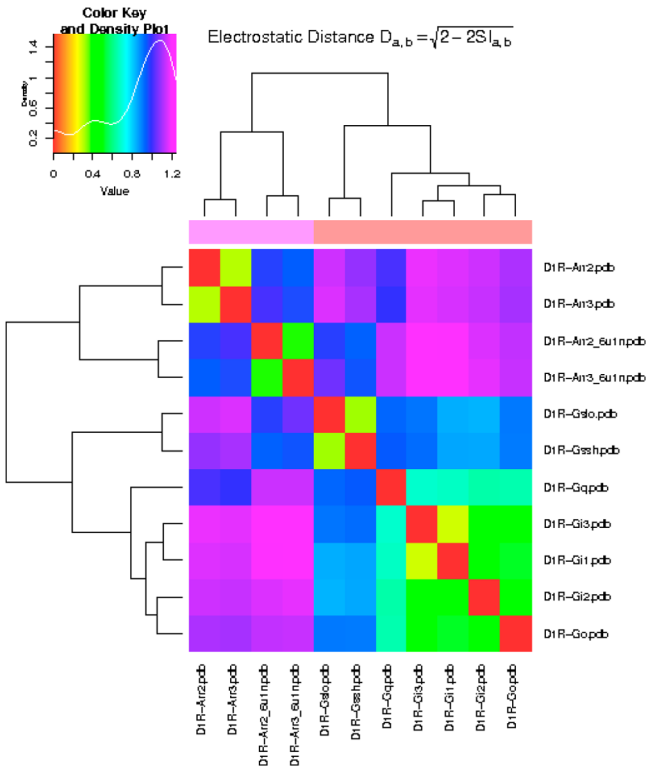


Figure 2. Example of an electrostatic distance plot as presented in the web server for each of the five dopamine receptors, in this case for D1R. The plot was attained from the PIPSA web server.[76]

Interaction Plots. Throughout the manuscript, binding motifs were listed using one-code amino acid nomenclature with the Ballesteros-Weinstein numbering scheme[77] (X.50 defines the most conserved residue in each TM domain) or the residue number in case of belonging to an ICL or the binding partner in superscript. Upon analysis, the one letter code "X" was used to define a wild-card amino acid, implying the position can represent any of the 20 amino acids. Figure 3 exemplifies an interaction plot of D1R-Gslo. The additional interaction plots retrieved for each DXR complex are available on the web site. All DXR interact with Arrs through contacts on ICL2, ICL3, and TM6. Whereas ICL1 and H8 seem to be always involved in interactions with Arrs_6pwc, these substructures only interact (and at the least extension) with some DXR at 6U1N derivative complexes. ICL1 is only involved in interactions with D5R for both Arrs_6u1n and D2R and D4R when complexed with Arr3_6u1n. H8 establishes one interaction through $\mathbf{N}^{8.43}$ with the finger loop of Arrs for all DXR but D3R-Arr3 and D4R. Conversely, TM2

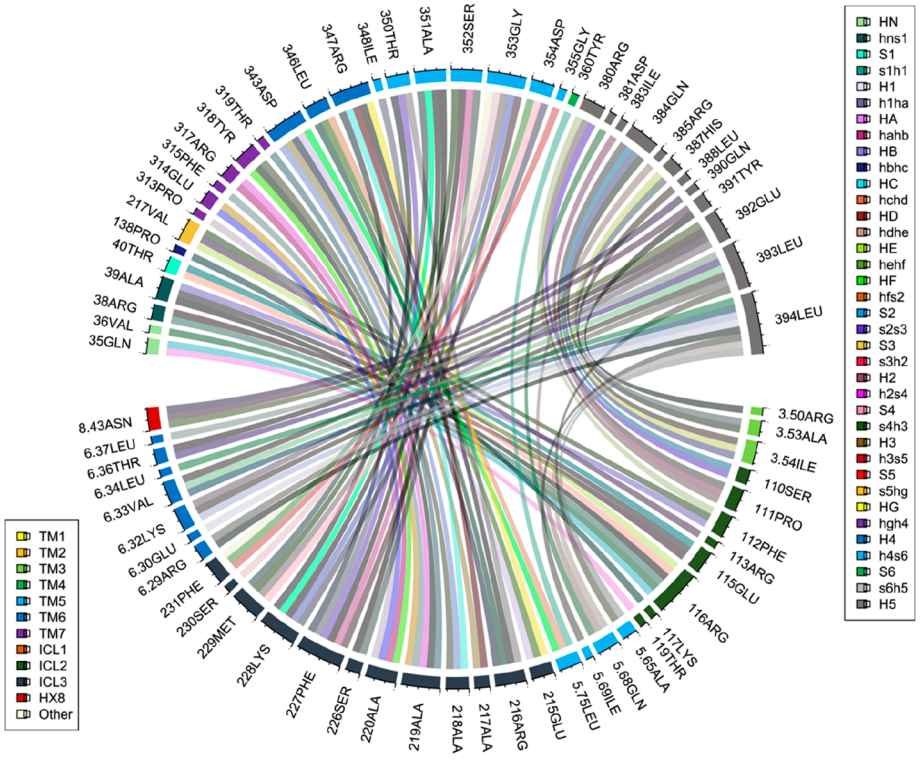


Figure 3. Example of an interaction plot as presented in the web server for each of the complexes, in this case for D1R-Gslo.

https://dx.doi.org/10.1021/acs.jcim.0c00371
J. Chem. Inf. Model. 2020, 60, 3969−3984

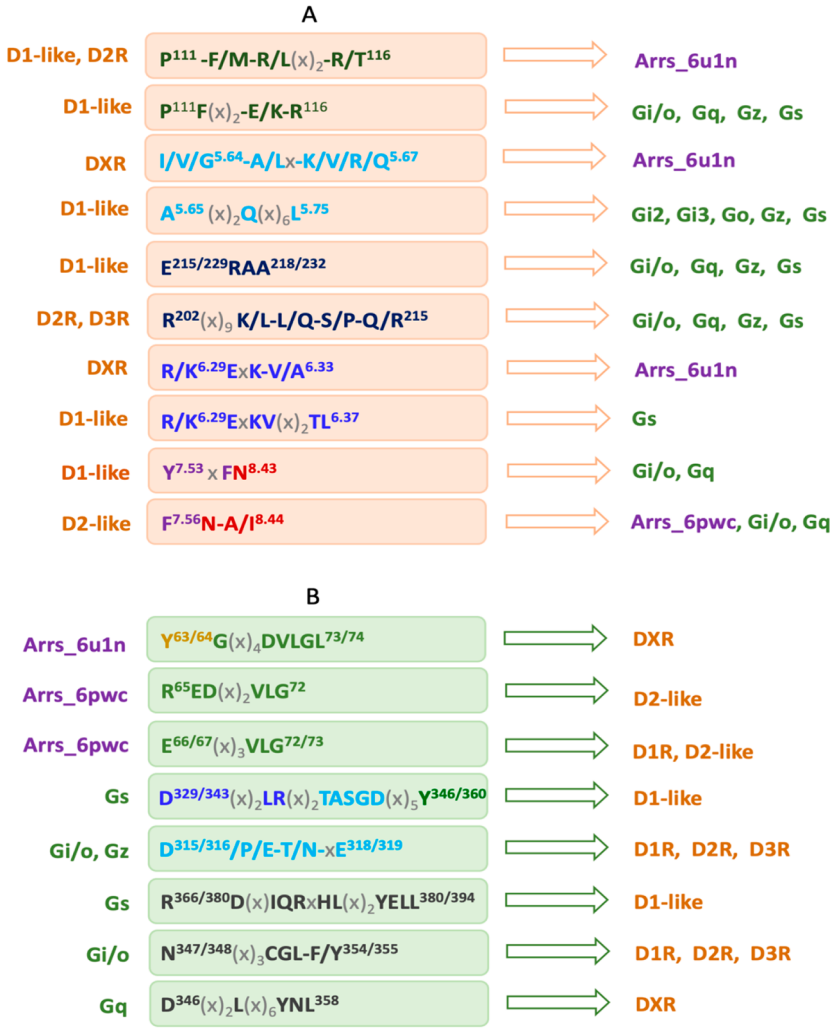


Figure 4. Key relevant interaction patterns observed at the DXR interface (A) and its partners, Arrs and G-proteins interfaces (B). The sequence motifs were colored according to the colors used to identify the structural motifs of DXR and the partners at our web site. The x represents positions that can be filled with any type of amino acid.

and TM3 are only involved in interactions at DXR-Arrs_6u1n complexes. While Arrs' C-loop derived from 6pwc is involved in interactions with ICL2, the same domain interacts with TM5 at all DXR-Arrs_6u1n complexes, with the exception of D3R-Arr2 that interacts with both ICL2 and TM5. Moreover, the finger loop interacts with ICL1, ICL3, TM6, and H8 (all DXR-Arrs) or TM2, TM3, and TM5 (DXR-Arrs_6u1n). Contrary to Arrs_6pwc, Arrs_6u1n establishes interactions with the ICL2 of all DXR through the lariat loop ($\mathbf{RG}^{286/287}$ for all DXR complexes but D3R_Arr2 that only interact through $\mathbf{G}^{286}$). Unlike D1-like receptors, D2-like receptors also interact with all Arrs' finger loops through TM7 ($\mathbf{F}^{7.56}$), with the exception of D4R-Arr2_6u1n.

TM5 ($\mathbf{L/K/T}^{5.75}$) is also involved in interactions at all DXR-Arrs complexes, with the exception of D1R-Arr2_6pwc. This domain is particularly important at DXR-Arrs_6u1n, with a higher number of interactions. Interestingly, a specific interaction involving the middle loop of Arr2_6pwc ($\mathbf{E}^{134}$) and the H8 structural motif ($\mathbf{D}^{8.45}$) or Arrs_6u1n ($\mathbf{G}^{133/134}$) and the ICL2 ($\mathbf{T}^{116}$) is observed at D_5R-Arr2_6pwc and D2R-Arrs complexes, respectively.

Analysis of the main interacting structural motif patterns involved in DXR-Arrs complexes showed that DXR receptors mainly interact with the $\mathbf{E}^{66/67}\mathbf{(x)_3VLG}^{72/73}$ (Arrs_6pwc) or $\mathbf{G}^{64/65}\mathbf{(x)_4DVLGL}^{73/74}$ (Arrs_6u1n) finger loop motif, with the exception of D5R which only contacts $\mathbf{VLG}^{72}$ and $\mathbf{VL}^{72}$ residues when bound to Arr2_6pwc and Arr3_6pwc, respectively. $\mathbf{R}^{65}\mathbf{ED(x)_2VLG}^{72}$ is a common interaction pattern for Arr2_6pwc bound to D2-like receptors. Concerning Arrs_6U1N, which establishes a higher number of interactions compared to Arrs_6pwc, some interesting interaction patterns involving different DXR subdomains were also disclosed, namely with TM3 ($\mathbf{R}^{3.50}\mathbf{(x)_2A\text{-}I/V}^{3.54}$, except for D2R_Arr3_6u1n), ICL2 ($\mathbf{P}^{111}\mathbf{\text{-}F/M\text{-}R/L(x)_2\text{-}R/T}^{116}$, except for D3R and D4R complexes), TM5 ($\mathbf{I/V/G}^{5.64}\mathbf{\text{-}A/Lx\text{-}K/V/R/Q}^{5.67}$), and TM6 ($\mathbf{R/K}^{6.29}\mathbf{ExK\text{-}V/A}^{6.33}$). For D2-like receptors, TM5 is particularly interacting with Arrs_6u1n, through the interaction pattern $\mathbf{V/G}^{5.64}\mathbf{Lx\text{-}R/Q\text{-}R/W(x)_6\ K/T}^{5.75}$.

Regarding the coupling of G-proteins with DXR, some particular pattern interactions were also highlighted. G-Proteins binding generally involved a higher number of interactions with DXR in comparison to Arrs binding, in particular Arrs_6pwc. However, this is not true for D2R and D4R, in which Arrs_6u1n tends to establish a higher number of interactions in comparison with G-proteins and Arrs_6pwc. DXR interacts with Gi/o and Gq through TM2, ICL2, TM3,

https://dx.doi.org/10.1021/acs.jcim.0c00371
J. Chem. Inf. Model. 2020, 60, 3969−3984

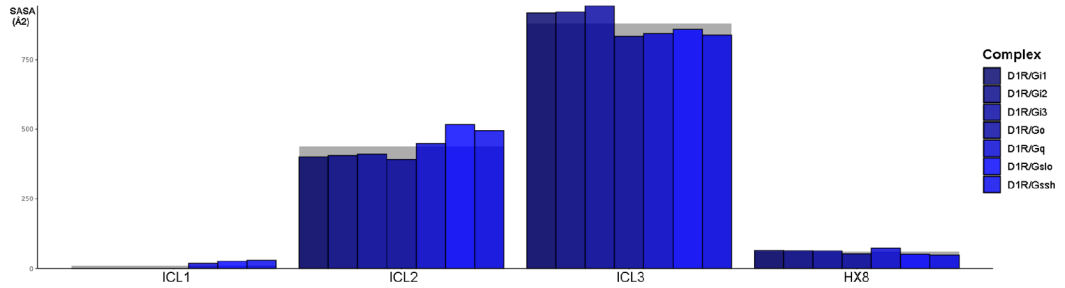


Figure 5. Example of a bar plot used for interface characterization, in this case, for the D1R-G-protein associated complexes, regarding the SASA at all substructures. Plots of this type were constructed for all the data regarding both interface characterization and interaction percentages and are available at the web server.

TM5, ICL3, TM6, TM7, and H8, with the exception of D2R that did not establish any interaction through ICL2 or TM5 when complexed with Gi/o. Interactions with ICL1 ($\mathbf{T}^{37/38/39}$) were only observed at D2R-like-G-protein complexes, namely Gi1 (D2R and D3R), Gi2 (D4R), Gi3 (D3R and D4R), and Go/oB (D4R).

On the other hand, the main interactions established by Gs with DXR involved TM3, ICL2, TM5, ICL3, TM6, and H8. Unlike Gi/o and Gq, no interactions were established between Gs and TM2 or TM7, with the exception of Gs bound to D3R that contacted TM7 through $\mathbf{F}^{7.56}$.

In all DXR-G-protein complexes, G-proteins' H5 was the main interacting domain, being involved in contacts with ICL1, TM2, TM3, TM5, TM6, TM7, and H8 of all DXR. Alongside with H5, hns1 and S3 were the main structural motifs involved in interactions between G-proteins and ICL2 of D1-like receptors. Noteworthy, Gs established more interactions with ICL2 at DXR than Gi/o, through HN, hns1, hbhc, and S3 (D1-like receptors) or hbhc, HN, and HB (D3R) motifs. It is important to mention that the hbhc finger loop was only involved in interactions with ICL2 at DXR-Gs complexes. Furthermore, ICL3 interacted with H4 and h4s6 of G-proteins at all DXR complexes, with the exception of D4R that only interacted with H4. In fact, among all DXRs, D4R was the receptor that established the lowest number of interactions with each G-protein.

Moreover, despite being an important interacting structural motif in both DXR-G-protein and DXR-Arrs_6pwc complexes, H8 seems to be particularly relevant in D2-like receptors, interacting with the H5 domain through the common interaction motif $\mathbf{F}^{7.56}\mathbf{N\text{-}A/I}^{8.44}$, which involves the TM7-H8 boundary. For Gi/o and Gq, this motif was extended to $\mathbf{F}^{7.56}\mathbf{N\text{-}A/I\text{-}E}^{8.45}$. It is also key that $\mathbf{N}^{8.43}$ was involved in interactions with both Arrs and G-proteins in all DXR complexes, with the exception of Arr2_6u1n bound to D4R and Arr3_6u1n bound to D3R and D4R or even Gz and Gt2 that do not interact with TM7 nor H8. Noteworthy, the D1R receptor interacts with the TM7-H8 boundary by the specific sequence motif $\mathbf{Y}^{7.53}\mathbf{xFN}^{8.43}$ when complexed with Gi/o and Gq.

Concerning D1R-G-protein complexes, and as shown in Figure 4, several common interaction motifs of D1-like receptors when complexed with G-proteins were revealed, namely $\mathbf{P}^{111}\mathbf{F}(\mathrm{x})_2\mathbf{\text{-}E/K\text{-}R}^{116}$ (ICL2), $\mathbf{E}^{215/229}\mathbf{RAA}^{218/232}$ (ICL3), and $\mathbf{A}^{5.65}(\mathrm{x})_2\mathbf{Q}(\mathrm{x})_6\mathbf{L}^{5.75}$ (TM5), except for D1R_Gi1 and D5R_Gq complexes. Interestingly, the interactions at TM5 were especially frequent at D1-like receptors and were completely absent for D2R when coupled with all G-proteins and D3R-Gssh and D4R-Gt2 complexes. On the other hand, the $\mathbf{R}^{202}(\mathrm{x})_9\mathbf{K/L\text{-}L/Q\text{-}S/P\text{-}Q/R}^{215}$ motif was identified as a common interaction pattern involving ICL3 of both D2R and D3R when complexed with G-proteins. Moreover, $\mathbf{R/K}^{6.29}\mathbf{ExKV}(\mathrm{x})_2\mathbf{TL}^{6.37}$ (TM6) was a common interaction motif of D1-like receptors when complexed with Gs.

Taking into account G-protein structural motifs, $\mathbf{D}^{329/343}(\mathrm{x})_2\mathbf{LR}(\mathrm{x})_2\mathbf{TASGD}(\mathrm{x})_5\mathbf{Y}^{346/360}$ appeared as a common interaction pattern of D1-like-Gs-proteins. It is mainly represented by polar residues that constitute H4, h4s6, and S6 residues. On the other hand, $\mathbf{D}^{315/316}\mathbf{/P/E\text{-}T/N\text{-}xE}^{318/319}$ (h4s6 loop) is a common motif for Gi/o and Gz bound to D1R, D2R, and D3R. Noteworthy, with the exception of the D4R-Gt2 complex, D4R does not interact with the h4s6 loop. Moreover, $\mathbf{R}^{366/380}\mathbf{DxIQRxHL}(\mathrm{x})_2\mathbf{YELL}^{380/394}$, $\mathbf{N}^{347/348}(\mathrm{x})_3\mathbf{CGL\text{-}F/Y}^{354/355}$, and $\mathbf{D}^{346}(\mathrm{x})_2\mathbf{L}(\mathrm{x})_6\mathbf{YNL}^{358}$ were also relevant interaction patterns for Gs- (with the D3R-Gs exception), Gi- (with the D4R-Gi1, D4R-Gi2, and D4R-Go exceptions), and Gq-proteins, respectively. It is important to note that Gz and Gt2, which are the members with the largest phylogenetic distance among Gi-proteins studied,[78] show the most different interaction pattern among Gi-proteins. Gt2 and Gz were the partners that are engaged in a lower number of contacts with DXR (in both cases D4R), only establishing a total of seven interactions involving hns1 (Gt2), h4s6 (Gt2), H4, and H5 (both Gt2 and Gz).

Interface Characterization. The number of surface atoms was consistently lower for Arrs-bound complexes (regardless of the template used in the modeling) in comparison to G-protein-bound, particularly in the case of D2-like receptors. An example of a plot for interface characterization is shown in Figure 5 for D1R-G-protein associated complexes. The number of buried atoms was usually higher for Arrs (particularly for Arrs_6u1n) and Gs-proteins than the remaining. D2-like complexes with Arrs_6pwc showed higher SASA at ICL1, ICL2, and H8. For D1-like complexes, the same was only true for ICL1 and H8. It is noteworthy to mention that arrestin-involving complexes modeled with 6U1N had higher SASA values at ICL1 and H8 in comparison with the ones modeled with 6pwc, which had higher values for substructures ICL2 and ICL3. Complexes with G-proteins displayed higher SASA values at ICL3. When considering G-proteins, complexes of D1-like with Gs and Gq tended to have higher SASA values at ICL1. Gi/o subfamily complexes showed higher SASA values at ICL3, apart from D3R. Gz complexes displayed very low SASA values for all substructures analyzed, particularly at ICL1 and H8.

The number of HB and SB was significantly lower for complexes with Arrs_6pwc, with the exception of D2R-Arr2. Arrs were the only partners that displayed HB/SB at ICL1 for all DXR complexes. For Arrs_6pwc complexes, this was only

https://dx.doi.org/10.1021/acs.jcim.0c00371
J. Chem. Inf. Model. 2020, 60, 3969−3984

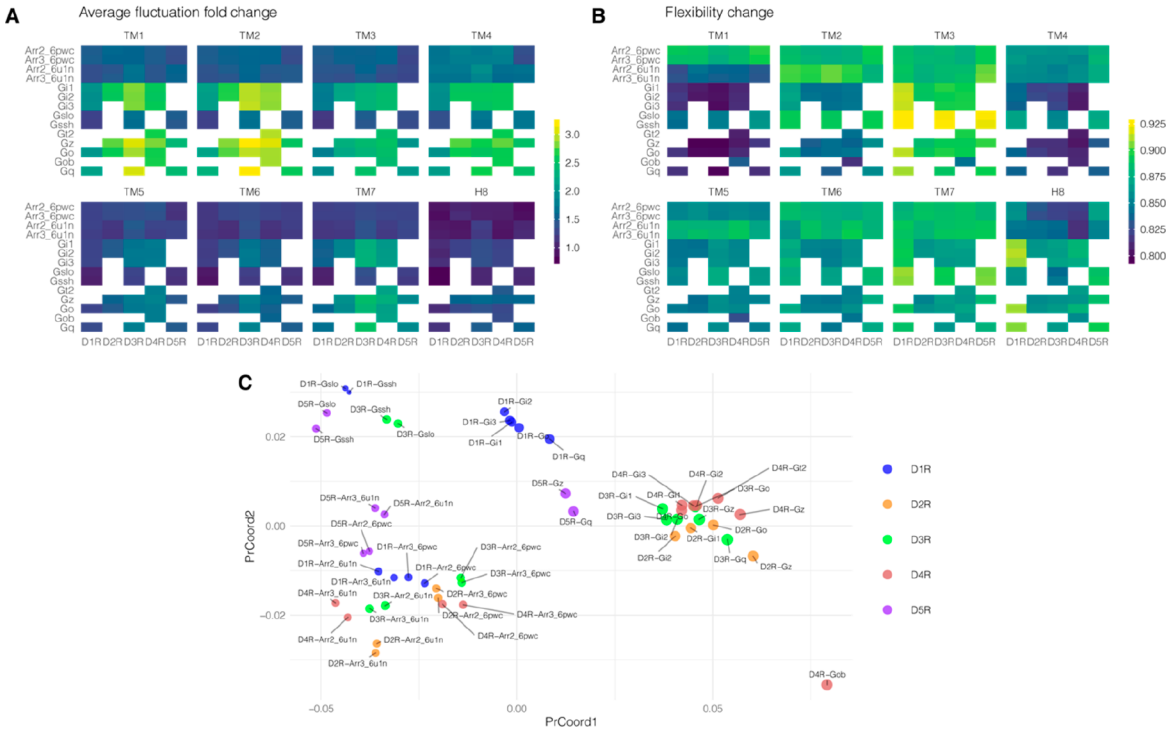


Figure 6. Dynamic analysis of the biologically relevant DXR complexes and their structural motifs regarding changes in flexibility and fluctuation. These results were obtained through a coarse-grained NMA considering only the alpha carbons performed on all monomeric and complexed DXR. A - Flexibility changes for TM1-7 and H8, as calculated by the Bhattacharyya distance to the respective monomer of each DXR. B - Average fluctuation fold change between DXR in complex and their respective monomer. C - Multidimensional scaling of the flexibility change values. The size of each dot is proportional to the average fold change in fluctuation values between bound and unbound DXR.

observed for D2R and D5R complexed to Arr3. For Arrs_6u1n complexes, no HB/SB at ICL1 were observed for D4R-Arr3 or D3R. Regarding the DXR in complex with the G-proteins, the overall number of HB/SB was usually higher in complexes with Gs-proteins and notably lower for Gz and Gq associated receptors. Gi-proteins usually had more HB/SB at ICL3 than the remaining complexes with either other G-proteins or Arrs, with the exception of D4R-Arrs_6u1n complexes. Considering D1-like receptors, only Arrs and Gs displayed HB/SB at ICL1. On the other hand, Arrs tend to have a lower amount of HB/SB at ICL2 and ICL3 than G-proteins for D2-like receptors, with the exception of D2R (at ICL2) and D4R-Arrs_6u1n (at ICL3). No HB/SB were observed at H8 for D1-like receptors. Arr3_6pwc was the only partner that had HB/SB for all D2-like receptors. For D4R, there were no HB/SB at ICL1 or H8 except for the complexes with Arrs.

Regarding the binding pocket, we considered its location starting in the centered geometrical point between residues 3.50, 6.30 and 7.53, for all the DXR complexes under scope, thus, focusing on the intracellular modifications induced by the interaction with the respective partners. Binding pocket volume ($Å^3$) was then analyzed. Compared with G-proteins, complexes with Arrs_6pwc generally displayed higher binding pocket volumes, with the exception of D1R. Considering all complexes with Arrs as partners, D2-like receptors had slightly lower binding pocket volumes, particularly for D2R and D4R and D3R_6u1n. Complexes with Go displayed higher binding pocket volume than the remaining Gi/o proteins, excepting for D2R.

Interaction Percentages. The interacting residues at DXR-partner showed different physicochemical properties. Regarding the interacting residues at Arrs-DXR complexes, Leu was the most predominant residue at DXR interfaces, except for D3R and D4R complexes modeled with 6pwc, in which both Lys and Arg were the most frequent residues. On the other hand, Arg was the most frequent residue at Arrs interfaces, with the exception of D2R, in which Lys (Arr2-D2R) and both Arg and Lys (Arr3-D2R and Arrs-D2R_6u1n) appeared as the most predominant one. His, Met, and Trp were absent for all DXR interfaces, as well as Cys for Arrs interfaces.

Nonpolar aliphatic residues were the most predominant residues at DXR interfaces of Arrs-DXR complexes, with the exception of Arrs-D2-like complexes modeled with 6pwc, in which basic positive residues (Arrs-D3R) or both basic positive and nonpolar aliphatic residues (Arr2-D2R and Arrs-D4R) were more frequent. In general, we observed that, together, basic positive and nonpolar aliphatic appeared in more than 50% at the Arrs interfaces. Nonpolar aliphatic residues were the most predominant for complexes involving D1-like receptors and D4R and Arrs_6pwc, while basic positive and polar uncharged were the most frequent ones for D2R-Arrs_6pwc and D3R-Arrs_6pwc, respectively. Regarding DXR-Arrs_6u1n complexes, nonpolar aliphatic residues were the most frequent for D3R and D5R, while basic positives are the most representative for D1R and D4R. Both nonpolar aliphatic and polar uncharged are the most frequent residues for D2R-Arrs_6u1n.

Concerning the interactions residues for Gi1-3-DXR complexes, negatively charged residues represented the most predominant type of residues at DXR. On the other hand, basic positively charged residues, alongside with nonpolar aliphatic, were the most representative of Gi1-3 (Arg as the top one). These results suggest a clear complementarity between these two binding partners.

https://dx.doi.org/10.1021/acs.jcim.0c00371
J. Chem. Inf. Model. 2020, 60, 3969−3984

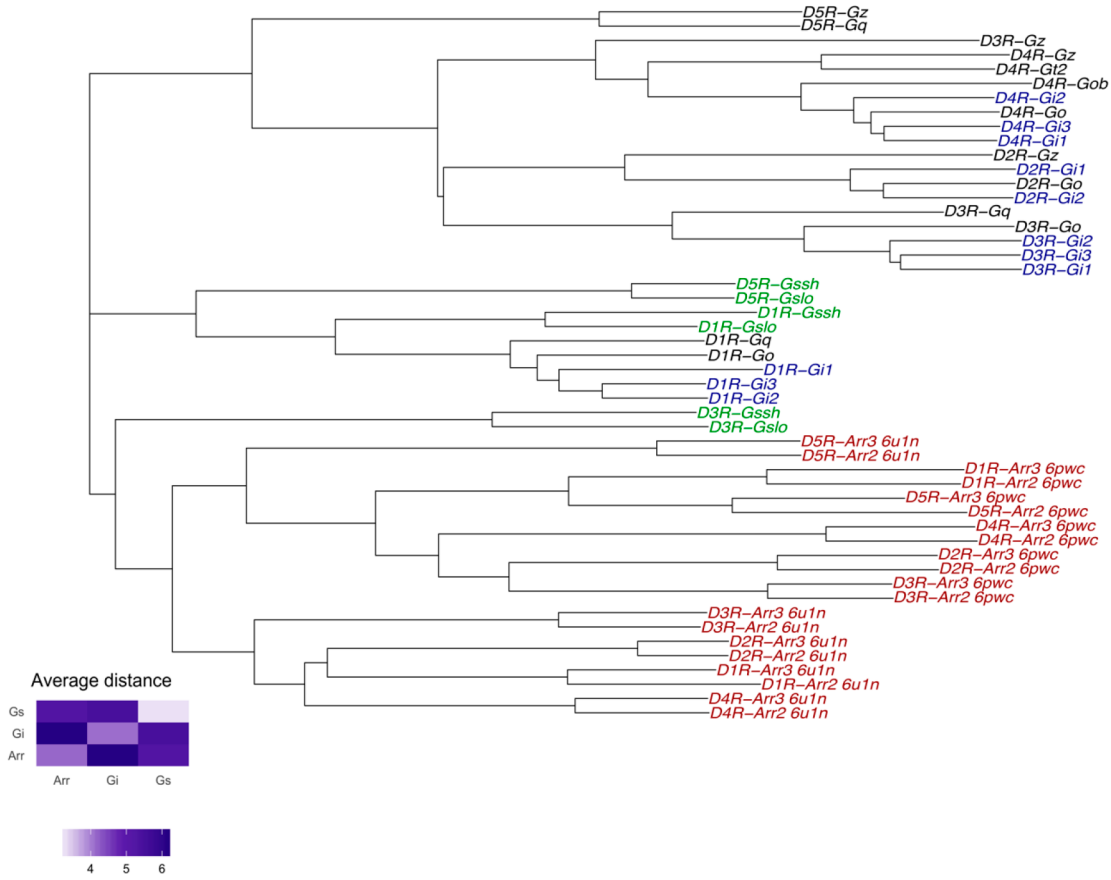


Figure 7. The structural phylogeny of DXR. This tree is the average of seven trees obtained for different sets of DXR complex characteristics (InterProSurf, COCOMAPS, HB/SB, interhelical distance, interface composition, BC, and fluctuation values) and was obtained by minimizing the quadratic path difference between the final tree and seven initial trees, allowing us to also calculate the branch lengths. In the bottom left corner, the distances between complexes bound to three relevant DXR complex groups (arrestin-bound (Arr), Gi-bound (Gi), and Gs-bound (Gs)) are represented in a heatmap, showing that intracomplex group distances are quite smaller than intercomplex group distances.

In Go-DXR complexes, Asn was the most predominant residue for DXR, while Arg was the most frequent residue at the partner interface. Nonpolar aliphatic residues were the most representative for DXR. This type of residue, alongside with basic positive, was the most frequent for G-protein, as well.

Regarding Gq-DXR complexes, Asp was also the most frequent residue in all DXR, being Arg and Ala are the most predominant ones in Gq. Similar to Go, in both Gq and DXR systems, nonpolar aliphatic residues represented the most abundant group.

Despite being the member with a larger phylogenetic distance among Gi-protein subtypes, Gz showed a similar pattern of interacting residues with all other Gi members analyzed. Therefore, acid negatively charged residues were the most abundant group in all DXR, being Glu was the most predominant residue. In a complementary way, basic positively charged residues were the most representative at Gz in Gz-DXR complexes, and Arg was the most frequent residue.

Among all G-protein-DXR complexes analyzed, Gs was the G-protein that revealed the most different pattern interfaces. In fact, in both Gslo-DXR and Gssh-DXR, polar uncharged residues represented the most abundant group for D3R and D5R interfaces, while basic positive residues were the most representative at the D1R interface. However, Arg, alongside Asp, was among the most frequent residue types at all DXR interfaces. On the other hand, nonpolar aliphatic residues were the most representative group at partner interfaces.

Noteworthy, some interesting findings were also disclosed by analyzing specific interacting residues of G-proteins coupled to a specific DXR. For example, Gly and Trp were only present at G-protein interfaces in D3R-G-protein and D4R-G-protein complexes, respectively.

Dynamic Analysis. All complexes were analyzed using NMA, and fluctuation values were calculated for all relevant structural motifs. The average fluctuation fold changes were calculated as the fold change between the receptor in complex and its monomeric template. Considering that fluctuation can be seen as a proxy for protein conformational flexibility, an increase in the average fold change indicates that, upon binding to its partner, the DXR undergoes conformational changes that increase the motility of the relevant structural motif. These values allow us to see distinct changes between relevant complex groups (Figure 6A): in particular, (i) DXR-Arr complexes, for both templates, showed the least fluctuation increase for all TM segments and even decreases in H8; (ii) complexes of Gi/o and Gq subfamilies have higher values of fluctuation throughout all the DXR family; (iii) D1-like receptors showed fairly lower increases in fluctuation when compared to D2-like receptors; and (iv) considering all DXR family members, D3R had, in general, higher average fluctuation values.

https://dx.doi.org/10.1021/acs.jcim.0c00371
J. Chem. Inf. Model. 2020, 60, 3969−3984

The dynamics of DXR when in complex with different partners was analyzed with regards to the flexibility of each TM1-7 and H8, calculated as the BCs between the DXR when in complex and their respective DXR in monomeric form. It is worth noting the BC does not work as a distance but rather as a coefficient of similarity, and as such, for values equal to 1, we expect very high similarity (or a very small distance), while for values below 1 we can expect decreased similarity as the BC decreases toward 0.

As evidenced in Figure 6B, DXR-Gs and DXR-Arrs complexes showed a fairly high BC, which is indicative of smaller changes in flexibility upon binding. Additionally, differences were also observed for D1-like receptors when compared with D2-like receptors—particularly, on TM2, TM3, TM6, TM7, and H8, with smaller flexibility changes observable in D1-like receptors as compared with D2-like receptors.

To assess whether changes in flexibility could be associated with the utilization of different templates, we compared the BCs for all motifs between the monomeric DXR used to model Arr complexes and the monomeric DXR used to model G-proteins in all cases. For all cases, the BC between both monomeric DXR was above 0.92, with the sole exception being TM1 in D3R and H8 in D1R and D5R, which had fairly smaller values (below 0.90), corresponding to 7.5% of the analyzed motifs. Provided that this is a computational analysis based on modeling, we decided to be as conservative as possible on any *posthoc* analysis and, while we present results for TM1-7 and H8, those for TM1 and H8 should be considered slightly erratic, despite following the general expected pattern for TM1-7 and H8. As such, we removed TM1 and H8 to calculate the MDS (Figure 6C).

All these changes in the different structural motifs end up contributing to the very clear distinctions when looking to the MDS of the flexibility changes. Figure 6C shows the first two components for each biologically relevant DXR complex. As expected, we observed a very evident clustering of Arrs complexes, with Gs-protein complexes (D1R, D3R, D5R) fairly isolated from other complexes.

What was also fairly evident is the clustering of Gi/o and Gq subfamily complexes, with clear distinctions between D1-like and D2-like complexes. D4R-Gob showed distinctly different flexibility changes upon binding when compared with any other complex, making conclusions regarding its comparative flexibility changes complicated.

Structural Phylogeny. The structural phylogenetic of all DXR considered the analyzed characteristics that were useful in grouping biologically distinct groups—InterProSurf, CO-COMAPS, HB/SB, interhelical distance, interface composition, BC, and fluctuation values—to calculate phylogenetic trees. The average of these trees was calculated and depicted in Figure 7. It becomes important to address how the expected groupings become apparent in the tree, namely the grouping of DXR-Arr complexes, of DXR-Gi-bound and of DXR-Gs-bound receptors, and how within the latter two we can still distinguish D1- and D2-like receptors, with the exceptions of D5R-Gz and D5R-Gq, and D3R-Gs complexes that are in an isolated clade inside D2-like and arrestin branches, respectively. DXR-Arr complexes diverge earlier on in the tree, which signals the fairly distinct conformational changes in DXR upon Arr binding. Inside the arrestin branch, two different clusters are present corresponding to the two different templates used. Two different clades hold all Gi/Go/Gob/Gz/Gq/Gs/-protein-bound DXR, with a common grouping of D1-like and D2-like, with the exception of D3-Gs complexes and D5-Gi/o- or D5-Gq- family complexes. Inside of the D1-like clade, there is a separation between the D1R branch (that includes both Gi/o and Gs-bound complexes) and D5-Gs. Similarly, inside the D2-like clade each receptor (D2R, D3R, D4R) forms a different branch.

In general, this analysis allows us to present a summarized picture of our computational findings and provide, in an unbiased way, a biologically relevant grouping that was obtained by combining the structural and dynamical changes in DXR upon partner binding and its interface composition.

DISCUSSION

GPCRs mediate several cellular pathways by coupling to two major cytoplasmic family players in signal transduction, G-proteins and Arrs.[79,80] Positional changes of TM5, TM6, and TM7 in relation to TM3 have been described upon GPCR activation, namely the inward movement of TM7 and the outward movement of TM6, creating a cavity in the intracellular surface of the receptor where Arrs and G-proteins bind to.[81,82] The interhelical distance between TM3-TM6 against TM3-TM7 analysis showed that Gs induces a higher TM3-TM6 and lower TM3-TM7 distances when bound to DXR in comparison to Arrs, Gi, and Gq, as stated in studies regarding other GPCR subfamilies in complex with those partners.[45,81,83] Interestingly, for D3R, which couples primarily to Gi, Gs displays a similar profile in terms of interhelical distances between Gi and Gq. This was not observed, however, for D1-like receptors, which primarily coupled to Gs. Indeed, these receptors showed different interhelical distances for TM3-TM6 and TM3-TM7 among diverse G-protein families, especially Gs and Gi. While D1-like receptors couple with Gq (both D1R and D5R) or Gs (D1R), only D3R is able to interact with secondary coupling G-proteins, being the most promiscuous within the D2-like subfamily. Among the Gi family, Gz is the partner that presents the most different TM3-TM6 and TM3-TM7 profile distances, which can be explained by its largest phylogenetic distance among Gi-proteins studied.[78] In line with interhelical distance results, the electrostatic distance showed a strong separation of three groups at DXR complexes: Arrs and Gs- and and Gi-proteins. Moreover, Gq is usually isolated, as expected, or in a smaller group with Gz, often isolated in the same manner. This also agrees with the most different profile shown by Gz among the Gi-protein subfamily. The differences between the three more debated groups through this paper (Arr, Gi, and Gs) became even more evident when we look at the average distances within the phylogenetic tree. Figure 7 aggregates relevant information extracted throughout this work, allowing the depiction of a larger landscape: the structural phylogeny of DXR binding reveals biologically relevant groups. These groups are in agreement with the literature, and as such, our structural phylogeny analysis corroborates the strength of our modeling protocol and further validates the need to take a deeper look at the individual structural analysis results.

The main interaction motifs involved in DXR-partner complexes could also attest differences in the binding of the most studied partner classes, Arrs, Gi, and Gs, when complexed to D1- and/D2-like receptors (Figure 8). Despite the differences between the two major GPCR partner groups, a high degree of shape complementarity has been reported between both G-proteins and Arrs (in particularly visual arrestins) and the intracellular receptor surface, which includes

https://dx.doi.org/10.1021/acs.jcim.0c00371
J. Chem. Inf. Model. 2020, 60, 3969−3984

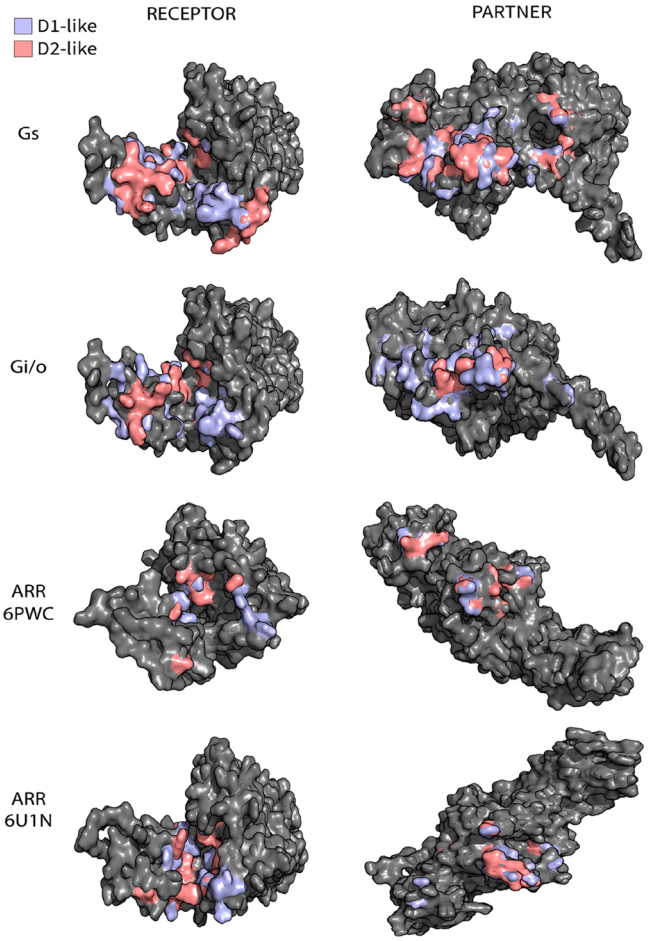


Figure 8. Representation of the surface interaction for D1- and D2-like receptors bound to the most distinct partner groups: Gs, Gi, and Arrs (Arr_6pwc and Arrs_6u1n). This figure highlights the common interaction motifs for each DXR-partner group above-mentioned. D2-like interactions represented in Gs are only for D3R and not the whole subfamily, as well as D1-like interactions in Gi which are only for the D1R-Gi/o complex.

the engagement of common interaction structural motifs in the binding of these partners to GPCR, namely at TM3, ICL2, TM5, and TM6.[81] Nevertheless, the intervention of other substructures and the organization of binding interfaces, in particular specific interacting residues, is still far from fully documented in the literature. In the case of nonvisual Arrs, this is particularly relevant since only two complexes of Arr2 complexed to different GPCRs (one complexed with NTSR1 and other with M2R) have been recently solved. Regarding the NTSR1-Arr2 complex, Yin et al. reported that Arr2, when complexed to the nonvisual GPCR NTSR1, displays a rotation of 90° when compared to the visual arrestin−rhodopsin complex.[11] Nonetheless, they share common interaction structural motifs at the arrestin-receptor interface that involves the finger loop and the C-loop domains. In fact, Arr2 contacts with NTSR1 through its finger loop (by interacting with ICL1 and TM6) and C-loop (through the interaction with ICL2). This is in line with our observations at all DXR-Arrs interfaces that pointed out the finger loop as being the most interacting domain. This is true not only for NTSR1-Arrs (DXR-Arrs_6pwc) complexes but also for M2R-Arrs (DXR-Arrs_6u1n), even though the orientation of Arr2 complexed with M2R is significantly different from NTSR1-Arrs; however, it is similar to the visual arrestin-rhodopsin complex.[12] Indeed, the common interaction motifs $\mathbf{E^{66/67}(x)_3VLG^{72/73}}$ (Arrs_6pwc) and $\mathbf{Y^{63/64}G(x)_4DVLGL^{73/74}}$ (Arrs_6u1n), which were reported as a key interaction points, are involved in contacts with different D_XR structural domains, namely ICL1, ICL2, ICL3, and TM6 (Arrs_6pwc) or TM2, TM3, ICL2, TM5, and TM6 (Arrs_6u1n). Interestingly, the $\mathbf{VLG^{72/73}}$ motif is always involved in interactions with TM6 and H8 at DXR-Arrs, independently of which Arrs conformations are involved. Exceptionally, for D5R-Arr3_6pwc only $\mathbf{VL^{72}}$ is involved in interactions with TM6 and H8. It is also noteworthy that at DXR-Arrs_6u1n, $\mathbf{D^{69/70}}$ alongside $\mathbf{V^{70/71}}$ also interacts with residues of the highly conserved DRY motif, namely $\mathbf{T^{2.39}}$ ($\mathbf{D^{69/70}}$) and $\mathbf{D^{3.50}}$ ($\mathbf{D^{69/70}}$ and $\mathbf{V^{70/71}}$). This is in line with the M2R-Arr2 complex, in which $\mathbf{D^{69}}$ is positioned in order to establish hydrogen bonds or salt bridges with $\mathbf{N^{2.39}}$ and $\mathbf{R^{3.50}}$.[12] Moreover, and similarly to what is observed by Yin et al.[11] and Carpenter et al.,[81] both Arrs' finger loops at D2-like-Arrs complexes interact with the TM7-H8 boundary, namely through the $\mathbf{F^{7.56}N{-}I/A^{8.44}}$(DXR-Arrs_6pwc) or $\mathbf{F^{7.56}N^{8.43}}$(DXR-Arrs_6u1n) sequence motifs, with the exception of Arr3_6u1n complexed with D3R and D4R which interact only with $\mathbf{F^{7.56}}$ or D4R-Arr_6u1n that do not interact with TM7 nor H8. At D1-like receptors, these interactions are also established with a few H8 residues (D5R-Arrs_6pwc) or only through $\mathbf{N^{8.43}}$ (D1R-Arrs_6pwc and D1-like-Arrs_6u1n), thus indicating some binding selectivity among D1-like-Arrs and D2-like-Arrs. Altogether, these results attest to the important role of TM7-H8 for Arrs binding to DXR-Arrs (in particular to D2-like-Arrs), in a similar way to NTSR1-Arr2 or rhodopsin-Arr, in which H8 is reported as crucial for high-affinity binding of Arrs to the activated rhodopsin.[84,85] The interactions between ICL1 and TM6 with the finger loop domain, which are present at all DXR-Arrs_6pwc complexes, are also in agreement with Yin et al. results.[11] Contacts between ICL3 of DXR and the loop between the top B strands ($\mathbf{K^{157/158}}$-I/S) are also in line with NTSR1-Arr2 interface binding.[11] It is important to mention that both Arr2 and Arr3 present a similar binding interface at DXR complexes, as reported by Yin et al. regarding the Arrs-NTSR1 binding interaction. The same is also observed for both Arr2 and Arr3 at DXR-Arrs_6u1n complexes. Interestingly, Arrs modeled with 6pwc establish a low number of interactions with DXR, in comparison with Arrs_6u1n and G-proteins. On the other hand, the Arrs-6u1n displays a higher surface interaction compared to Arrs_6pwc and G-proteins, which could explain the establishment of a larger amount of interactions, without ruling out the possibility of both conformations existing under different biological circumstances. Among all DXR complexes, the number of buried atoms was usually higher for Arrs and even higher for Gs, which is in line with the literature reporting Gs as having a greater buried surface area compared to Arrs.[84]

Concerning G-protein binding interfaces, the analyses of DXR complexes showed that TM3, ICL2, TM5, ICL3, TM6, and H8 were generally involved in interactions with G-proteins. This is in line with other studies of class A GPCR coupled with different G-proteins.[74,83,86] Moreover, and as shown in these studies, H5 of each G-protein is the most interacting motif in DXR-G-proteins complexes, corroborating the key role of this motif (namely C-terminal H5 amino acids) in G-protein binding. In fact, Gs interacts through the common sequence pattern $\mathbf{R^{366/380}D(x)IQRxHL(x)_2YELL^{380/394}}$ at

https://dx.doi.org/10.1021/acs.jcim.0c00371
J. Chem. Inf. Model. 2020, 60, 3969−3984

TM3, ICL2, TM5, TM6, and H8 at DXR-like receptors, with the exception of D3R that only interacts through the sequence $\mathbf{L}^{374/388}\mathbf{xQYELL}^{380/394}$ and is involved in contacts with the receptor at TM5, TM6, and TM7-H8. Besides the nonexistent interaction at TM3, the amount of interactions at TM5 in D3R-Gs ($\mathbf{R}^{5.68}$ for Gslo and no interaction at TM5 for Gssh) was lower in comparison with D1-like-Gs complexes, which interact through the common sequence motif $\mathbf{A}^{5.65}(\mathrm{x})_2\mathbf{QI}(\mathrm{x})_5\mathbf{L}^{5.75}$ at hgh4 (D5R-Gssh), H4 (D1R and D5R-Gslo), and S6, in addition to H5. Despite being usually reported as an important interacting domain for different G-protein subfamilies bound to GPCRs, TM5 establishes more contacts at GPCR complexes that primarily couple to Gs (e.g., β2AR) in comparison with those ones that primarily couple to Gi (CB1R and A1R).[74,83] This can be an explanation for the lower number of interactions at TM5 of D3R when in complex with Gs, since that is a primary Gi coupling receptor. Indeed, D1-like receptors establish a higher number of interactions at TM5 for G-proteins (even for secondary partner Gi/o and Gq), while D2-like receptors, which primarily couple to Gi, present an opposite pattern, establishing, in general, fewer interactions with TM5 when complexed with G-proteins, being contacts at TM5 nonexistent at D2R-G-proteins. Furthermore, Gi/o and Gq also interact with several residues through H5 domain with TM2, TM3, TM5, TM6, and TM7-H8 at DXR, having as a common interaction pattern subfamily $\mathbf{N}^{347/348}(\mathrm{x})_3\mathbf{CGL\text{-}F/Y}^{354/355}$ and $\mathbf{D}^{346}(\mathrm{x})_2\mathbf{L}(\mathrm{x})_6\mathbf{YNL}^{358}$ (respectively), except for D4R, that interacts through $\mathbf{L}^{348/349}(\mathrm{x})_2\mathbf{CGL\text{-}F/Y}^{354/355}$.

Our results are consistent with those reported by Sandhu et al., suggesting that the involvement of physicochemically different amino acids at the C-terminal of H5 is able to establish a huge number of interactions at different domains of DXR and may be directly involved in selectivity binding of different G-proteins to this receptor subfamily. Particularly, and in line with all results from structural and dynamics analysis, Gs is the partner that reveals the most different pattern interface, in which polar uncharged residues represented the most abundant group at the DXR interface, although Arg and Asp are the most common amino acids, and nonpolar aliphatic residues were the most representative group at partner interfaces. Interestingly, Gz showed a similar pattern of interacting residues with all other Gi members, even though it established fewer interactions within all DXR complexes. Another interesting interacting motif with a different pattern for D1-like and D2-like receptors was H8, in particular the turn at the TM7-H8 boundary. D2-like receptors interacted with H5 through the common sequence motif $\mathbf{F}^{7.56}\mathbf{N\text{-}A/I}^{8.44}$, which was extended to $\mathbf{F}^{7.56}\mathbf{N\text{-}A/I\text{-}E}^{8.45}$ for Gi/o and Gq couplings. Interactions involving the C-terminal of TM7 and the N-terminal of H8 were mainly reported to Gi-coupled receptors, being present in a low number or even absent for Gs-coupled proteins.[74,83] This agrees with the abundance of interactions at D2-like receptors that primarily coupled to Gi, in comparison with D1-like receptors, namely D5R, which interacted only through $\mathbf{N}^{8.43}$. D1R interacts through a specific sequence motif, involving some residues of the conserved NxxPY pattern, namely $\mathbf{Y}^{7.53}\mathbf{xFN}^{8.43}$. Nevertheless, this motif was not involved at D1R-Gs, suggesting the engagement of this sequence pattern at the binding interface of D2-like receptors that primarily couples to Gi-proteins. Noteworthy, although ICL1 seems to be involved in a few interactions (through $\mathbf{T}^{37/38/39}$) with C-terminal H5 (mainly $\mathbf{C}^{351/352}$) at DXR complexes, only D2-like receptors are involved through ICL1 in interactions with Gi/o. This finding is in line with Sandhu et al. that states that ICL1 only established interactions with C-terminal H5 in Gi-coupled receptors but not in Gs-receptors under investigation. It is also imperative to mention that ICL2 and ICL3 play an important role at PPIs for all DXR complexes. While in D1-like receptors, ICL2 was involved in interactions with a common sequence pattern among G-proteins: $\mathbf{P}^{111}\mathbf{F}(\mathrm{x})_2\mathbf{\text{-}E/K\text{-}R}^{116}$, in D2-like receptors this structural domain established few (D3R and D4R) or zero interactions (D2R all complexes but D2R-Gz). On the other hand, ICL3 showed different common interaction patterns between D1-like and D2-like receptors (except for D4R), namely $\mathbf{E}^{215/229}\mathbf{RAA}^{218/232}$ and $\mathbf{R}^{202}(\mathrm{x})_9\mathbf{K/L\text{-}L/Q\text{-}S/P\text{-}Q/R}^{215}$, respectively. D4R, among the D2-like only established few interactions through ICL3. The analysis of ICL2 and ICL3 suggests that these structural motifs are important in the coupling selectivity between D1-like and D2-like receptors. This is in agreement with studies that reported the influence of these ICLs in partner selective binding at GPCRs.[86]

Altogether, the structural and dynamic analysis performed here showed a strong separation of DXR complexes into three main groups: Arrs, Gs, and Gi-bonded. Gq presents an interaction and dynamics profile closer to the Gi subfamily, in particular Gz, the member with the most dissimilar binding interface and dynamics features among the Gi subfamily. Moreover, it is also shown that different structural motifs are involved in G-protein selective coupling between D1- and D2-like receptors, being H5, namely its C-terminal a key motif for G-protein selective coupling among DXR.

This work contributes to a deeper knowledge about the structural determinants involved in the binding specificity of G-proteins and Arrs with DXRs, giving further insight into the principles of molecular recognition between those major partner players and the DXR family. As far as we know, this constitutes the first large-scale computational study involving DXRs and their main intracellular effectors, representing an important step in the understanding of GPCR coupling specificity.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.0c00371.

All 50 models involving DXR family receptor and functional effectors identified in Table 1 (each complex identified by name of DXR receptor and corresponding effector) (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Irina S. Moreira − *Department of Life Sciences, University of Coimbra, 3000-456 Coimbra, Portugal; CNC - Center for Neuroscience and Cell Biology, University of Coimbra, Faculty of Medicine, 3004-504 Coimbra, Portugal; University of Coimbra, Center for Innovative Biomedicine and Biotechnology, Faculty of Medicine, 3004-504 Coimbra, Portugal;* orcid.org/0000-0003-2970-5250; Email: irina.moreira@cnc.uc.pt

https://dx.doi.org/10.1021/acs.jcim.0c00371
J. Chem. Inf. Model. 2020, 60, 3969−3984

Authors

A. J. Preto − *CNC - Center for Neuroscience and Cell Biology, University of Coimbra, Faculty of Medicine, 3004-504 Coimbra, Portugal; University of Coimbra, Center for Innovative Biomedicine and Biotechnology, Faculty of Medicine, 3004-504 Coimbra, Portugal; Institute for Interdisciplinary Research, University of Coimbra, 3030-789 Coimbra, Portugal;* orcid.org/0000-0003-4203-2230

Carlos A. V. Barreto − *CNC - Center for Neuroscience and Cell Biology, University of Coimbra, Faculty of Medicine, 3004-504 Coimbra, Portugal; University of Coimbra, Center for Innovative Biomedicine and Biotechnology, Faculty of Medicine, 3004-504 Coimbra, Portugal; Institute for Interdisciplinary Research, University of Coimbra, 3030-789 Coimbra, Portugal;* orcid.org/0000-0003-1459-7680

Salete J. Baptista − *CNC - Center for Neuroscience and Cell Biology, University of Coimbra, Faculty of Medicine, 3004-504 Coimbra, Portugal; University of Coimbra, Center for Innovative Biomedicine and Biotechnology, Faculty of Medicine, 3004-504 Coimbra, Portugal; Centro de Ciências e Tecnologias Nucleares, Instituto Superior Técnico, Universidade de Lisboa, 2695-066 Bobadela, Portugal*

José Guilherme de Almeida − *CNC - Center for Neuroscience and Cell Biology, University of Coimbra, Faculty of Medicine, 3004-504 Coimbra, Portugal; European Bioinformatics Institute EMBL-EBI, Hinxton, Cambridgeshire CB10 1SD, United Kingdom*

Agostinho Lemos − *CNC - Center for Neuroscience and Cell Biology, University of Coimbra, Faculty of Medicine, 3004-504 Coimbra, Portugal; GIGA Cyclotron Research Centre In Vivo Imaging, University of Liège, 4000 Liège, Belgium;* orcid.org/0000-0001-6956-1736

André Melo − *REQUIMTE/LAQV, Departamento de Química e Bioquímica, Faculdade de Ciências da Universidade do Porto, 4169-007 Porto, Portugal*

M. Nátalia D. S. Cordeiro − *REQUIMTE/LAQV, Departamento de Química e Bioquímica, Faculdade de Ciências da Universidade do Porto, 4169-007 Porto, Portugal;* orcid.org/0000-0003-3375-8670

Zeynep Kurkcuoglu − *Bijvoet Center for Biomolecular Research, Faculty of Science − Chemistry, Utrecht University, 3584, CH, Utrecht, The Netherlands*

Rita Melo − *CNC - Center for Neuroscience and Cell Biology, University of Coimbra, Faculty of Medicine, 3004-504 Coimbra, Portugal; University of Coimbra, Center for Innovative Biomedicine and Biotechnology, Faculty of Medicine, 3004-504 Coimbra, Portugal; Centro de Ciências e Tecnologias Nucleares, Instituto Superior Técnico, Universidade de Lisboa, 2695-066 Bobadela, Portugal;* orcid.org/0000-0003-1056-1007

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.0c00371

Author Contributions

A.J.P., C.A.V.B., S.J.B., J.G.d.A., and A.L. performed model construction. C.A.V.B. and S.J.B. performed model equilibration with molecular dynamics. C.A.V.B. and A.J.P. performed model submission and selection for HADDOCK models. A.J.P. constructed and deployed the structural analysis pipeline. J.G.d.A. and Z.K. constructed and deployed the dynamics analysis pipeline. A.J.P. and J.G.d.A. performed web server implementation. A.J.P., C.A.V.B., S.J.B., and J.G.d.A. analyzed all results. All authors wrote the paper, and R.S.M. and I.S.M. led the research.

Funding

This work was funded by the European Regional Development Fund (ERDF), through the Centro 2020 Regional Operational Programme under project CENTRO-01-0145-FEDER-000008: BrainHealth 2020, and through the COMPETE 2020 - Operational Programme for Competitiveness and Internationalisation and Portuguese national funds via FCT - Fundação para a Ciência e a Tecnologia, under projects POCI-01-0145-FEDER-031356, PTDC/QUI-OUT/32243/2017, and UIDB/04539/2020. A.J.P. and C.A.V.B. were also supported by FCT through Ph.D. scholarships SFRH/BD/144966/2019 and SFRH/BD/145457/2019, respectively. I.S.M. was funded by the FCT Investigator program - IF/00578/2014 (cofinanced by European Social Fund and Programa Operacional Potencial Humano). A.M. and M.N.C. were supported by UIDB/50006/2020 with funding from FCT/MCTES through national funds.

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

Arr, arrestin; Arr2, arrestin 2; Arr3, arrestin 3; BCs, Bhattacharya coefficients; β2AR, β2-adrenergic receptor; COCOMAPS, biocomplexes contact MAPS; ConSurf, conservation surface mapping; DXR, dopamine receptor; D1R, dopamine receptor 1; D2R, dopamine receptor 2; D3R, dopamine receptor 3; D4R, dopamine receptor 4; D5R, dopamine receptor 5; D1-5R, dopamine receptors 1, 2, 3, 4, and 5; D1-like, dopamine receptors 1 and 5; D2-like, dopamine receptors 2, 3, and 4; DXRcomp, dopamine receptor in a complex; DXRmon, dopamine receptor monomeric structure; ENM, elastic network modeling; GPCR, G-protein coupled receptor; Gssh, Gαs short form; Gslo, Gαs long form; HADDOCK, high ambiguity driven protein−protein DOCKing; H8, Helix-8; HB, hydrogen bonds; ICL1-3, intracellular loops 1 to 3; ICL1, intracellular loop 1; ICL2, intracellular loop 2; ICL3, intracellular loop 3; ECL1-3, extracellular loops 1 to 3; ECL1, extracellular loop 1; ECL2, extracellular loop 2; ECL3, extracellular loop 3; MSA, multiple sequence alignments; MDS, multidimensional scaling; NTSR1, neurotensin receptor-1; PDB, The Protein Data Bank; PDB-ID, Protein Data Bank identifier; POVME, pocket volume measurer; PPIs, protein−protein interfaces; RMSD, root mean-squared deviation; SASA, solvent accessible surface area; SB, salt bridges; TM1-7, transmembrane domains 1-7

■ REFERENCES

(1) Arango-Lievano, M.; Sensoy, O.; Borie, A.; Corbani, M.; Guillon, G.; Sokoloff, P.; Weinstein, H.; Jeanneteau, F. A GIPC1-Palmitate Switch Modulates Dopamine Drd3 Receptor Trafficking and Signaling. *Mol. Cell. Biol.* **2016**, *36* (6), 1019−1031.

(2) Sensoy, O.; Weinstein, H. A Mechanistic Role of Helix 8 in GPCRs: Computational Modeling of the Dopamine D2 Receptor Interaction with the GIPC1−PDZ-Domain. *Biochim. Biophys. Acta, Biomembr.* **2015**, *1848* (4), 976−983.

(3) Kobilka, B. K. G Protein Coupled Receptor Structure and Activation. *Biochim. Biophys. Acta, Biomembr.* **2007**, *1768* (4), 794−807.

(4) Ding, X.; Zhao, X.; Watts, A. G-Protein-Coupled Receptor Structure, Ligand Binding and Activation as Studied by Solid-State NMR Spectroscopy. *Biochem. J.* **2013**, *450* (3), 443−457.

(5) Girault, J.-A.; Greengard, P. The Neurobiology of Dopamine Signaling. *Arch. Neurol.* **2004**, *61* (5), 641−644.
(6) Creese, I.; Burt, D. R.; Snyder, S. H. Dopamine Receptor Binding Predicts Clinical and Pharmacological Potencies of Antischizophrenic Drugs. *Science* **1976**, *192* (4238), 481−483.
(7) Sariñana, J.; Kitamura, T.; Künzler, P.; Sultzman, L.; Tonegawa, S. Differential Roles of the Dopamine 1-Class Receptors, D1R and D5R, in Hippocampal Dependent Memory. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (22), 8245−8250.
(8) Ma, L.; Pei, G. β-Arrestin Signaling and Regulation of Transcription. *J. Cell Sci.* **2007**, *120* (2), 213−218.
(9) Luttrell, L. M.; Gesty-Palmer, D. Beyond Desensitization: Physiological Relevance of Arrestin-Dependent Signaling. *Pharmacol. Rev.* **2010**, *62* (2), 305−330.
(10) Smith, J. S.; Rajagopal, S. The β-Arrestins: Multifunctional Regulators of G Protein-Coupled Receptors. *J. Biol. Chem.* **2016**, *291* (17), 8969−8977.
(11) Yin, W.; Li, Z.; Jin, M.; Yin, Y.-L.; de Waal, P. W.; Pal, K.; Yin, Y.; Gao, X.; He, Y.; Gao, J.; Wang, X.; Zhang, Y.; Zhou, H.; Melcher, K.; Jiang, Y.; Cong, Y.; Zhou, X. E.; Yu, X.; Xu, H. E. A Complex Structure of Arrestin-2 Bound to a G Protein-Coupled Receptor. *Cell Res.* **2019**, *29* (12), 971−983.
(12) Staus, D. P.; Hu, H.; Robertson, M. J.; Kleinhenz, A. L. W.; Wingler, L. M.; Capel, W. D.; Latorraca, N. R.; Lefkowitz, R. J.; Skiniotis, G. Structure of the M2Muscarinic Receptor−β-Arrestin Complex in a Lipid Nanodisc. *Nature* **2020**, *579* (7798), 297−302.
(13) Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinforma.* **2016**, *54* (1), 5.6.1−5.6.37.
(14) van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428* (4), 720−725.
(15) Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3d: An R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* **2006**, *22* (21), 2695−2696.
(16) Negi, S. S.; Schein, C. H.; Oezguen, N.; Power, T. D.; Braun, W. InterProSurf: A Web Server for Predicting Interacting Sites on Protein Surfaces. *Bioinformatics* **2007**, *23* (24), 3397−3399.
(17) Vangone, A.; Spinelli, R.; Scarano, V.; Cavallo, L.; Oliva, R. COCOMAPS: A Web Application to Analyze and Visualize Contacts at the Interface of Biomolecular Complexes. *Bioinformatics* **2011**, *27* (20), 2915−2916.
(18) Armon, A.; Graur, D.; Ben-Tal, N. ConSurf: An Algorithmic Tool for the Identification of Functional Regions in Proteins by Surface Mapping of Phylogenetic Information. *J. Mol. Biol.* **2001**, *307* (1), 447−463.
(19) Ashkenazy, H.; Abadi, S.; Martz, E.; Chay, O.; Mayrose, I.; Pupko, T.; Ben-Tal, N. ConSurf 2016: An Improved Methodology to Estimate and Visualize Evolutionary Conservation in Macromolecules. *Nucleic Acids Res.* **2016**, *44* (W1), W344−W350.
(20) Wagner, J. R.; Sørensen, J.; Hensley, N.; Wong, C.; Zhu, C.; Perison, T.; Amaro, R. E. POVME 3.0: Software for Mapping Binding Pocket Flexibility. *J. Chem. Theory Comput.* **2017**, *13* (9), 4584−4592.
(21) Kurkcuoglu, O.; Turgut, O. T.; Cansu, S.; Jernigan, R. L.; Doruker, P. Focused Functional Dynamics of Supramolecules by Use of a Mixed-Resolution Elastic Network Model. *Biophys. J.* **2009**, *97* (4), 1178−1187.
(22) Uh, M.; White, B.; Sidhu, A. Alteration of Association of Agonist-Activated Renal D1A Dopamine Receptors with G Proteins in Proximal Tubules of the Spontaneously Hypertensive Rat. *J. Hypertens.* **1998**, *16* (9), 1307−1313.
(23) Kimura, K.; White, B. H.; Sidhu, A. Coupling of Human D-1 Dopamine Receptors to Different Guanine Nucleotide Binding Proteins. EVIDENCE THAT D-1 DOPAMINE RECEPTORS CAN COUPLE TO BOTH Gs AND Go. *J. Biol. Chem.* **1995**, *270* (24), 14672−14678.
(24) Mannoury la Cour, C.; Vidal, S.; Pasteau, V.; Cussac, D.; Millan, M. J. Dopamine D1 Receptor Coupling to Gs/Olf and Gq in Rat Striatum and Cortex: A Scintillation Proximity Assay (SPA)/Antibody-Capture Characterization of Benzazepine Agonists. *Neuropharmacology* **2007**, *52* (3), 1003−1014.
(25) Odagaki, Y.; Kinoshita, M.; Ota, T. Dopamine-Induced Functional Activation of Gαq Mediated by Dopamine D1-like Receptor in Rat Cerebral Cortical Membranes. *J. Recept. Signal Transduction Res.* **2019**, *39* (1), 9−17.
(26) Del'Guidice, T.; Lemasson, M.; Beaulieu, J.-M. Role of Beta-Arrestin 2 Downstream of Dopamine Receptors in the Basal Ganglia. *Front. Neuroanat.* **2011**, *5*, 58.
(27) Grünewald, S.; Reiländer, H.; Michel, H. In Vivo Reconstitution of Dopamine D2S Receptor-Mediated G Protein Activation in Baculovirus-Infected Insect Cells: Preferred Coupling to Gi1 versus Gi2. *Biochemistry* **1996**, *35* (48), 15162−15173.
(28) Jiang, M.; Spicher, K.; Boulay, G.; Wang, Y.; Birnbaumer, L. Most Central Nervous System D2 Dopamine Receptors Are Coupled to Their Effectors by Go. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (6), 3577−3582.
(29) Leck, K. J.; Blaha, C. D.; Matthaei, K. I.; Forster, G. L.; Holgate, J.; Hendry, I. A. Gz Proteins Are Functionally Coupled to Dopamine D2-like Receptors in Vivo. *Neuropharmacology* **2006**, *51* (3), 597−605.
(30) Obadiah, J.; Avidor-Reiss, T.; Fishburn, C. S.; Carmon, S.; Bayewitch, M.; Vogel, Z.; Fuchs, S.; Levavi-Sivan, B. Adenylyl Cyclase Interaction with the D2 Dopamine Receptor Family; Differential Coupling to Gi, Gz, and Gs. *Cell. Mol. Neurobiol.* **1999**, *19* (5), 653−664.
(31) Beaulieu, J.-M.; Espinoza, S.; Gainetdinov, R. R. Dopamine Receptors − IUPHAR Review 13. *Br. J. Pharmacol.* **2015**, *172* (1), 1−23.
(32) Gurevich, E. V.; Gainetdinov, R. R.; Gurevich, V. V. G Protein-Coupled Receptor Kinases as Regulators of Dopamine Receptor Functions. *Pharmacol. Res.* **2016**, *111*, 1−16.
(33) Newman-Tancredi, A.; Cussac, D.; Audinot, V.; Pasteau, V.; Gavaudan, S.; Millan, M. J. G Protein Activation by Human Dopamine D3 Receptors in High-Expressing Chinese Hamster Ovary Cells: A Guanosine-5′-O-(3-[35S]Thio)-Triphosphate Binding and Antibody Study. *Mol. Pharmacol.* **1999**, *55* (3), 564−574.
(34) Zaworski, P. G.; Alberts, G. L.; Pregenzer, J. F.; Im, W. B.; Slightom, J. L.; Gill, G. S. Efficient Functional Coupling of the Human D3 Dopamine Receptor to Go Subtype of G Proteins in SH-SY5Y Cells. *Br. J. Pharmacol.* **1999**, *128* (6), 1181−1188.
(35) Jose, P. A.; Soares-da-Silva, P.; Eisner, G. M.; Felder, R. A. Dopamine and G Protein-Coupled Receptor Kinase 4 in the Kidney: Role in Blood Pressure Regulation. *Biochim. Biophys. Acta, Mol. Basis Dis.* **2010**, *1802* (12), 1259−1267.
(36) Ilani, T.; Fishburn, C. S.; Levavi-Sivan, B.; Carmon, S.; Raveh, L.; Fuchs, S. Coupling of Dopamine Receptors to G Proteins: Studies with Chimeric D2/D3 Dopamine Receptors. *Cell. Mol. Neurobiol.* **2002**, *22* (1), 47−56.
(37) Xu, W.; Reith, M. E. A.; Liu-Chen, L.-Y.; Kortagere, S. Biased Signaling Agonist of Dopamine D3 Receptor Induces Receptor Internalization Independent of β-Arrestin Recruitment. *Pharmacol. Res.* **2019**, *143*, 48−57.
(38) Kazmi, M. A.; Snyder, L. A.; Cypess, A. M.; Graber, S. G.; Sakmar, T. P. Selective Reconstitution of Human D4 Dopamine Receptor Variants with Giα Subtypes. *Biochemistry* **2000**, *39* (13), 3734−3744.
(39) Sidhu, A.; Niznik, H. B. Coupling of Dopamine Receptor Subtypes to Multiple and Diverse G Proteins. *Int. J. Dev. Neurosci.* **2000**, *18* (7), 669−677.
(40) Yamaguchi, I.; Harmon, S. K.; Todd, R. D.; O'Malley, K. L. The Rat D4 Dopamine Receptor Couples to Cone Transducin (Gαt2) to Inhibit Forskolin-Stimulated CAMP Accumulation. *J. Biol. Chem.* **1997**, *272* (26), 16599−16602.
(41) Deming, J. D.; Shin, J.; Lim, K.; Lee, E.-J.; Van Craenenbroeck, K.; Craft, C. M. Dopamine Receptor D4 Internalization Requires a

https://dx.doi.org/10.1021/acs.jcim.0c00371
J. Chem. Inf. Model. 2020, 60, 3969−3984

Beta-Arrestin and a Visual Arrestin. *Cell. Signalling* **2015**, *27* (10), 2002−2013.
(42) Rashid, A. J.; O'Dowd, B. F.; Verma, V.; George, S. R. Neuronal Gq/11-Coupled Dopamine Receptors: An Uncharted Role for Dopamine. *Trends Pharmacol. Sci.* **2007**, *28* (11), 551−555.
(43) Abraham, A. D.; Neve, K. A.; Lattal, K. M. Activation of D1/5 Dopamine Receptors: A Common Mechanism for Enhancing Extinction of Fear and Reward-Seeking Behaviors. *Neuropsychopharmacology* **2016**, *41* (8), 2072−2081.
(44) Fiser, A. Template-Based Protein Structure Modeling. In *Computational Biology*; Fenyö, D., Ed.; Methods in Molecular Biology (Methods and Protocols); Humana Press: Totowa, NJ, 2010; Vol. *673*, pp 73−94, DOI: 10.1007/978-1-60761-842-3_6.
(45) Rasmussen, S. G. F.; DeVree, B. T.; Zou, Y.; Kruse, A. C.; Chung, K. Y.; Kobilka, T. S.; Thian, F. S.; Chae, P. S.; Pardon, E.; Calinski, D.; Mathiesen, J. M.; Shah, S. T. A.; Lyons, J. A.; Caffrey, M.; Gellman, S. H.; Steyaert, J.; Skiniotis, G.; Weis, W. I.; Sunahara, R. K.; Kobilka, B. K. Crystal Structure of the β 2 Adrenergic Receptor−Gs Protein Complex. *Nature* **2011**, *477* (7366), 549−555.
(46) The UniProt Consortium. UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019**, *47* (D1), D506−D515.
(47) Kang, Y.; Kuybeda, O.; de Waal, P. W.; Mukherjee, S.; Eps, N. V.; Dutka, P.; Zhou, X. E.; Bartesaghi, A.; Erramilli, S.; Morizumi, T.; Gu, X.; Yin, Y.; Liu, P.; Jiang, Y.; Meng, X.; Zhao, G.; Melcher, K.; Ernst, O. P.; Kossiakoff, A. A.; Subramaniam, S.; Xu, H. E. Cryo-EM Structure of Human Rhodopsin Bound to an Inhibitory G Protein. *Nature* **2018**, *558* (7711), 553−558.
(48) Sievers, F.; Higgins, D. G. Clustal Omega, Accurate Alignment of Very Large Numbers of Sequences. In *Multiple Sequence Alignment Methods*; Russell, D. J., Ed.; Methods in Molecular Biology (Methods and Protocols); Humana Press: Totowa, NJ, 2014; Vol. *1079*, pp 105−116, DOI: 10.1007/978-1-62703-646-7_6.
(49) Shen, M.; Sali, A. Statistical Potential for Assessment and Prediction of Protein Structures. *Protein Sci.* **2006**, *15* (11), 2507−2524.
(50) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M.; Pieper, U.; Sali, A. Comparative Protein Structure Modeling Using Modeller. *Curr. Protoc. Bioinforma.* **2006**, *15* (1), 5.6.1−5.6.30.
(51) Wiederstein, M.; Sippl, M. J. ProSA-Web: Interactive Web Service for the Recognition of Errors in Three-Dimensional Structures of Proteins. *Nucleic Acids Res.* **2007**, *35* (suppl_2), W407−W410.
(52) Wallner, B.; Elofsson, A. Can Correct Protein Models Be Identified? *Protein Sci.* **2003**, *12* (5), 1073−1086.
(53) *The PyMOL Molecular Graphics System*; Schrödinger, LLC.
(54) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29* (11), 1859−1865.
(55) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L.; MacKerell, A. D.; Klauda, J. B.; Im, W. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12* (1), 405−413.
(56) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder toward Realistic Biological Membrane Simulations. *J. Comput. Chem.* **2014**, *35* (27), 1997−2004.
(57) 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.
(58) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1−3), 43−56.
(59) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701−1718.
(60) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34* (25), 2135−2145.
(61) Goldenberg, O.; Erez, E.; Nimrod, G.; Ben-Tal, N. The ConSurf-DB: Pre-Calculated Evolutionary Conservation Profiles of Protein Structures. *Nucleic Acids Res.* **2009**, *37* (suppl_1), D323−D327.
(62) Levitt, M.; Sander, C.; Stern, P. S. Protein Normal-Mode Dynamics: Trypsin Inhibitor, Crambin, Ribonuclease and Lysozyme. *J. Mol. Biol.* **1985**, *181* (3), 423−447.
(63) Bhattacharyya, A. On a Measure of Divergence between Two Multinomial Populations. *Sankhyā Indian J. Stat. 1933−1960* **1946**, *7* (4), 401−406.
(64) Fuglebakk, E.; Echave, J.; Reuter, N. Measuring and Comparing Structural Fluctuation Patterns in Large Protein Datasets. *Bioinformatics* **2012**, *28* (19), 2431−2440.
(65) Gower, J. C. Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis. *Biometrika* **1966**, *53* (3−4), 325−338.
(66) Revell, L. J. Phytools: An R Package for Phylogenetic Comparative Biology (and Other Things). *Methods Ecol. Evol.* **2012**, *3* (2), 217−223.
(67) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria.
(68) Paradis, E.; Claude, J.; Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R Language. *Bioinformatics* **2004**, *20* (2), 289−290.
(69) Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer-Verlag: New York, 2016.
(70) Gu, Z.; Gu, L.; Eils, R.; Schlesner, M.; Brors, B. Circlize Implements and Enhances Circular Visualization in R. *Bioinformatics* **2014**, *30* (19), 2811−2812.
(71) Waskom, M.; Botvinnik, O.; O'Kane, D.; Hobson, P.; Ostblom, J.; Lukauskas, S.; Gemperline, D. C.; Augspurger, T.; Halchenko, Y.; Cole, J. B.; Warmenhoven, J.; Ruiter, J.; de Pye, C.; Hoyer, S.; Vanderplas, J.; Villalba, S.; Kunter, G.; Quintero, E.; Bachant, P.; Martin, M.; Meyer, K.; Miles, A.; Ram, Y.; Brunner, T.; Yarkoni, T.; Williams, M. L.; Evans, C.; Fitzgerald, C.; Brian; Qalieh, A. *Mwaskom/Seaborn: V0.9.0 (July 2018)*; Zenodo, 2018; DOI: 10.5281/zenodo.1313201.
(72) Caswell, T. A.; Droettboom, M.; Hunter, J.; Lee, A.; Firing, E.; Stansby, D.; Klymak, J.; Andrade, E. S.; de Nielsen, J. H.; Varoquaux, N.; Hoffmann, T.; Root, B.; Elson, P.; May, R.; Dale, D.; Lee, J.-J.; Seppänen, J. K.; McDougall, D.; Straw, A.; Hobson, P.; Gohlke, C.; Yu, T. S.; Ma, E.; Vincent, A. F.; Silvester, S.; Moad, C.; Katins, J.; Kniazev, N.; Ariza, F.; Ernest, E. *Matplotlib/Matplotlib: REL: V3.1.1*; Zenodo, 2019; DOI: 10.5281/zenodo.3264781.
(73) Chang, W.; Cheng, J.; Allaire, J. J.; Xie, Y.; McPherson, J. *Shiny: Web Application Framework for R*; 2018.
(74) Sandhu, M.; Touma, A. M.; Dysthe, M.; Sadler, F.; Sivaramakrishnan, S.; Vaidehi, N. Conformational Plasticity of the Intracellular Cavity of GPCR−G-Protein Complexes Leads to G-Protein Promiscuity and Selectivity. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (24), 11956−11965.
(75) Liu, J. J.; Horst, R.; Katritch, V.; Stevens, R. C.; Wüthrich, K. Biased Signaling Pathways in B2-Adrenergic Receptor Characterized by 19F-NMR. *Science* **2012**, *335* (6072), 1106−1110.
(76) Richter, S.; Wenzel, A.; Stein, M.; Gabdoulline, R. R.; Wade, R. C. WebPIPSA: A Web Server for the Comparison of Protein Interaction Properties. *Nucleic Acids Res.* **2008**, *36* (Web Server issue), W276−W280.
(77) Ballesteros, J. A.; Weinstein, H. Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G Protein-Coupled Receptors. In *Methods in Neurosciences*; Sealfon, S. C., Ed.; Receptor

Molecular Biology; Academic Press: 1995; Vol. *25*, pp 366−428, DOI: 10.1016/S1043-9471(05)80049-7.
(78) Syrovatkina, V.; Alegre, K. O.; Dey, R.; Huang, X.-Y. Regulation, Signaling, and Physiological Functions of G-Proteins. *J. Mol. Biol.* **2016**, *428* (19), 3850−3868.
(79) Moreira, I. S. Structural Features of the G-Protein/GPCR Interactions. *Biochim. Biophys. Acta, Gen. Subj.* **2014**, *1840* (1), 16−33.
(80) Sensoy, O.; Moreira, I. S.; Morra, G. Understanding the Differential Selectivity of Arrestins toward the Phosphorylation State of the Receptor. *ACS Chem. Neurosci.* **2016**, *7* (9), 1212−1224.
(81) Carpenter, B.; Tate, C. G. Active State Structures of G Protein-Coupled Receptors Highlight the Similarities and Differences in the G Protein and Arrestin Coupling Interfaces. *Curr. Opin. Struct. Biol.* **2017**, *45*, 124−132.
(82) Weis, W. I.; Kobilka, B. K. The Molecular Basis of G Protein−Coupled Receptor Activation. *Annu. Rev. Biochem.* **2018**, *87* (1), 897−919.
(83) Draper-Joyce, C. J.; Khoshouei, M.; Thal, D. M.; Liang, Y.-L.; Nguyen, A. T. N.; Furness, S. G. B.; Venugopal, H.; Baltos, J.-A.; Plitzko, J. M.; Danev, R.; Baumeister, W.; May, L. T.; Wootten, D.; Sexton, P. M.; Glukhova, A.; Christopoulos, A. Structure of the Adenosine-Bound Human Adenosine A 1 Receptor−G i Complex. *Nature* **2018**, *558* (7711), 559−563.
(84) Kang, Y.; Zhou, X. E.; Gao, X.; He, Y.; Liu, W.; Ishchenko, A.; Barty, A.; White, T. A.; Yefanov, O.; Han, G. W.; Xu, Q.; de Waal, P. W.; Ke, J.; Tan, M. H. E.; Zhang, C.; Moeller, A.; West, G. M.; Pascal, B. D.; Eps, N. V.; Caro, L. N.; Vishnivetskiy, S. A.; Lee, R. J.; Suino-Powell, K. M.; Gu, X.; Pal, K.; Ma, J.; Zhi, X.; Boutet, S.; Williams, G. J.; Messerschmidt, M.; Gati, C.; Zatsepin, N. A.; Wang, D.; James, D.; Basu, S.; Roy-Chowdhury, S.; Conrad, C. E.; Coe, J.; Liu, H.; Lisova, S.; Kupitz, C.; Grotjohann, I.; Fromme, R.; Jiang, Y.; Tan, M.; Yang, H.; Li, J.; Wang, M.; Zheng, Z.; Li, D.; Howe, N.; Zhao, Y.; Standfuss, J.; Diederichs, K.; Dong, Y.; Potter, C. S.; Carragher, B.; Caffrey, M.; Jiang, H.; Chapman, H. N.; Spence, J. C. H.; Fromme, P.; Weierstall, U.; Ernst, O. P.; Katritch, V.; Gurevich, V. V.; Griffin, P. R.; Hubbell, W. L.; Stevens, R. C.; Cherezov, V.; Melcher, K.; Xu, H. E. Crystal Structure of Rhodopsin Bound to Arrestin by Femtosecond X-Ray Laser. *Nature* **2015**, *523* (7562), 561−567.
(85) Kirchberg, K.; Kim, T.-Y.; Möller, M.; Skegro, D.; Raju, G. D.; Granzin, J.; Büldt, G.; Schlesinger, R.; Alexiev, U. Conformational Dynamics of Helix 8 in the GPCR Rhodopsin Controls Arrestin Activation in the Desensitization Process. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (46), 18690−18695.
(86) Flock, T.; Hauser, A. S.; Lund, N.; Gloriam, D. E.; Balaji, S.; Babu, M. M. Selectivity Determinants of GPCR−G-Protein Binding. *Nature* **2017**, *545* (7654), 317−322.

https://dx.doi.org/10.1021/acs.jcim.0c00371
J. Chem. Inf. Model. 2020, 60, 3969−3984