Published in final edited form as:

J Mol Graph Model. 2016 November ; 70: 140–152. doi:10.1016/j.jmglm.2016.10.004.

Homology modeling of a Class A GPCR in the inactive conformation: a quantitative analysis of the correlation between model/template sequence identity and model accuracy

Stefano Costanzi^{a,b,*}, Matthew Skorski^a, Alessandro Deplano^a, Brett Habermehl^a, Mary Mendoza^a, Keyun Wang^a, Michelle Biederman^a, Jessica Dawson^a, and Jia Gao^a

^aDepartment of Chemistry, American University, Washington, DC 20016, USA

^bCenter for Behavioral Neuroscience, American University, Washington, DC 20016, USA

Abstract

With the present work we quantitatively studied the modellability of the inactive state of Class A G protein-coupled receptors (GPCRs). Specifically, we constructed models of one of the Class A GPCRs for which structures solved in the inactive state are available, namely the β_2 AR, using as templates each of the other class members for which structures solved in the inactive state are also available. Our results showed a detectable linear correlation between model accuracy and model/template sequence identity. This suggests that the likely accuracy of the homology models that can be built for a given receptor can be generally forecasted on the basis of the available templates. We also probed whether sequence alignments that allow for the presence of gaps within the transmembrane domains to account for structural irregularities afford better models than the classical alignment procedures that do not allow for the presence of gaps within such domains. As our results indicated, although the overall differences are very subtle, the inclusion of internal gaps within the transmembrane domains has a noticeable beneficial effect on the local structural accuracy of the domain in question.

Graphical abstract

*Address correspondence to: Stefano Costanzi, Ph.D., Department of Chemistry, American University, 4400 Massachusetts Ave, NW, 20016 Washington, DC, USA, Tel: +1-202-885-1722, costanzi@american.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

SUPPORTING INFORMATION

The following material is available as supporting information: 1) Sequence alignments; 2) RMSD values calculated with respect to the α carbons; 3) Correlation between RMSD values calculated with respect to the α carbons and sequence identity values; 4) Comparison of the two alignment strategies – average, for α carbons; 5) Comparison of the two alignment strategies – individual models for α carbons and heavy atoms.

Keywords

Homology modeling; sequence alignment; G protein-coupled receptors; GPCRs; β_2 adrenergic receptor

INTRODUCTION

G protein-coupled receptors (GPCRs) are membrane bound proteins that function as cellular receivers for extracellular stimuli. Specifically, GPCRs act as molecular switches that can be turned on by chemicals known as agonists. Upon binding, agonists shift the receptors' conformational equilibrium from the inactive (non-signaling) state to one of several possible active (signaling) states. In turn, once activated, the receptors initiate a cascade of events that ultimately results in a physiological response. Agonists that activate GPCRs include a wide range of natural ligands, for instance neurotransmitters and hormones. Beyond responding to their physiological ligands, GPCRs can also be targeted with synthetic molecules for pharmacological purposes. Indeed, GPCRs are targeted by many of the marketed and leisure drugs, several of which act directly on the receptors, functioning either as agonists that stimulate their activity, or as blockers that prevent the stimulation of the receptors by natural agonists [1-4].

GPCRs constitute a large superfamily of proteins that, in the human species, comprises about 1000 members. The superfamily is divided into five different families, the largest of which is the so-called Rhodopsin family, also known as Class A or Family I, which comprises about 84% of all GPCRs. Structurally, GPCRs are constituted by a single polypeptide chain that spans seven times the plasma membrane with seven transmembrane (TM) domains, connected by three extracellular loops (EL) and three intracellular loops (IL). The N-terminus is in the extracellular space, while the C-terminus is in the cytosol. Given the pharmaceutical relevance of GPCRs, there is great interest in the experimental elucidation and the theoretical modeling of their three-dimensional (3D) structures, which can be employed as platforms to discover or design new ligands that can modulate their activity. In the year 2000, the structure of a GPCR, namely bovine rhodopsin, was solved for the first time through X-ray crystallography. The first structures of a nonrhodopsin GPCR were not solved until 2007, when the β_2 adrenergic receptor (β_2 AR) yielded for the first time to the efforts of crystallographers. Following the 2007 breakthrough and a wealth of subsequent developments, structural studies in the field of GPCRs experienced a rapid

expansion [4-8]. To date, a total of 147 structures have been solved for a total of 33 receptors, 28 of which belonging to Class A (see our searchable table, with references, at: www.costanziresearch.com/p/table.html).

Although GPCR structural studies are in rapid expansion, a thorough experimental characterization of all the receptors is likely unfeasible due to the large size of the superfamily. In this context, we and other authors have shown evidence that the available experimental structures of GPCRs can be used as templates for the construction of three-dimensional models of the other family members through homology modeling [9-16]. Notably, in 2008, we demonstrated for the first time that accurate models of GPCRs can be constructed through homology modeling, by comparing models and experimental structures of the β_2 AR [9]. Prior to then, such a study would not have been feasible for the lack of more than one crystallized receptor. Later the same year, a series of blind assessments started, where models built and submitted to an ad hoc committee before the crystallization of GPCRs were compared to the subsequently unveiled structures. The excellent results that we obtained in the first of such assessments, which was focused on the adenosine A_{2A} , confirmed the level of accuracy that we had shown in the 2008 article for the β_2 AR models is indeed achievable [10]. Further assessments highlighted that, as one could intuitively predict, the modeling efforts afford better results when the modeled receptors are closely related to the templates [17, 18].

With the present work we aimed at adding a quantitative layer to the field of GPCR homology modeling. Specifically, we focused our study on the modellability of Class A GPCRs in the inactive state, since more than half of the solved GPCR structures – namely 80 of them – reflect the inactive state of 24 different Class A GPCRs, solved in complex with either antagonists or inverse agonists (**Figure 1**). To this extent, we constructed models of one of the 24 Class A GPCRs for which structures solved in the inactive state are available, namely the β_2 AR, using as templates each of the other 23 Class A GPCRs for which structures solved in the inactive state are also available (**Figure 1**).

The main question that we posed is the following: is there a quantifiable linear correlation between the accuracy of the models and the model/template sequence identity? As it turned out, our results indicated that there is indeed a detectable linear correlation between model accuracy and model/template sequence identity, suggesting that the accuracy of the homology models that can be built for a given receptor can be generally forecasted on the basis of the available templates. Another question that we posed relates to the strategy employed for the alignment of the seven TM domains. In particular, we asked whether sequence alignments that allow gaps within the TM domains to account for predicted bulges, tight coils, bends or other structural irregularities afford better models than the classical alignment procedures that do not allow for the presence of gaps within such domains. As our results indicated, the overall differences are very subtle, although the inclusion of internal gaps within the TM domains does appear to have a beneficial effect on the local structural quality of the domain in question.

METHODOLOGY

Templates

As templates, we used the crystal structures of the 23 Class A GPCRs solved in the inactive conformation at the time when the homology modeling work was performed. Whenever more than one structure reflecting the inactive state was available for a given receptor, the highest resolution structure (as of September 30, 2015) was selected. A list of the template receptors, indicating the PDB ID and the structural resolution follows: A_{2A} (4EIY, 1.80 Å) [19], AT₁ (4YAY, 2.90 Å) [20], turkey β₁ (4BVN, 2.1 Å) [21], CCR5 (4MBS, 2.71 Å) [22], CXCR4 (3ODU, 2.50 Å) [23], D₃ (3PBL, 2.89 Å) [24], δ (4N6H, 1.80 Å) [25], mouse δ (4EJ4, 3.40 Å) [26], FFA1 (4PHU, 2.33 Å) [27], H₁ (3RZE, 3.10 Å) [28], κ (4DJH, 2.90 Å) [29], LPA₁ (4Z35, 2.9 Å), [30] M₂ (3UON, 3.00 Å) [31], rat M₃ (4U15, 2.80 Å) [32], mouse μ (4DKL, 2.80 Å) [33], NOP (4EA3, 3.10 Å) [34], OX₂ (4S0V, 2.50 Å) [35], P2Y₁ (4XNV, 2.20 Å) [36], P2Y₁₂ (4NTJ, 2.62 Å) [37], PAR1 (3VW7, 2.20 Å) [38], cattle rhodopsin (1U19, 2.2 Å) [39], squid rhodopsin (2Z73, 2.5 Å) [40], S1P₁ (3V2Y, 2.80 Å) [41]. Prior to the modeling, all the non-receptor components were expunged from the structure. Moreover, whenever more than one receptor monomer was present, all but the first one were expunged from the structure.

Alignment of the TMs domains, strategy 1: no gaps within the TMs

The sequence alignment of the templates and the target receptor was performed through a manual procedure based on the identification of the seven TMs, in analogy to what we described in a recent methodology article [12]. The entire alignment procedure was conducted with the Swiss-PDB Viewer (Deep View, <http://spdbv.vital-it.ch>) [42]. For practical purposes, an initial automatic sequence alignment was first obtained with the Blosom 62 matrix, with penalties for the opening and the extension of gaps set to 4 and 3, respectively. This initial alignment served only as a starting guide for the following manual alignment procedure, in which the automatic positioning of the gaps was overridden so as to position the gaps in structurally compatible regions. In particular, the initial alignment was first manually adjusted to ensure the matching of the motifs that identify each of the 7 TMs. For each sequence alignment, the boundaries of the 7 membrane-spanning helical domains were assigned according to the three-dimensional structure of the template (see supporting information, sequence alignment files). No internal gaps were allowed within the boundaries of the membrane-spanning domains; hence any internal gap possibly generated by the initial automatic alignment was manually removed.

Alignment of the TMs, strategy 2: structure-based sequence alignment with gaps allowed within the TMs

In addition to the alignment described in the paragraph above, an alternative alignment of the templates and the target receptor was produced by aligning the TMs following the structure-based alignment tool available at the GPCRdb website (<http://gpcrdb.org/alignment/targetselection>) and described by Isberg et al. in a recent article [6]. This alignment procedure allows the presence of gaps within the TMs, to account for the crystallographically observed presence of structural irregularities. The alignments of the

TMs extracted from the GPCRdb website were indeed carefully inspected prior to their employment. No corrections were needed.

Alignment of residues in the second extracellular loop (EL2)

The β_2 AR features three Cys residues in its EL2 domain, at positions 184, 190 and 191. Cys 191 is involved in the formation of a disulfide bridge with Cys 106 in TM3 – an analogous bridge is found in many other Class A GPCRs. Cys 184 and Cys 190 are involved in the formation of a second disulfide bridge internal to EL2 [43]. Notably, the presence of these two disulfide bridges in the structure of the β_2 AR had been hypothesized long before the solution of the crystal structure of the receptor. Hence, their accounting in the alignment as well as the modeling procedure is fully justified [43]. After the alignment of the TM domains, we manually aligned Cys 191 to the corresponding residue Cys that connects EL2 and TM3 in the template, whenever present. Moreover, we arranged the alignment of the two remaining Cys residues in order to match them with residues close enough to allow the construction of a disulfide bridge in the model without structural distortions. As in all other cases, the actual viability of the alignment was tested after the construction of the models through the evaluation of the quality of the structural parameters. Alternative alignments were tested until a viable solution was found (see below, construction of homology models). Notably, for the purposes of our sequence and structural comparisons we divided the EL2 domain into two subdomains: EL2a, defined as the portion of EL2 upstream of Cys191, and EL2b, defined as the portion of EL2 from Cys 191 on.

Alignment of loops and termini

Once the TMs and the Cys residues in EL2 were aligned as indicated above, the remainder of the receptor was aligned by accounting for differences in sequence length by placing gaps in the loop domains or the termini. For domains that had fewer residues in the modeled receptor than in the template (*deletions in the modeled receptor*), the alignment was adjusted in a way that the physical distance between the α carbons of the last aligned residue upstream of the gap and the first aligned residue downstream of the gap was lower than a given cutoff dependent on the number of non-aligned residues in the sequence of the modeled receptor found between the aligned extremities (**Figure 2**). Whenever possible, each residue of the modeled receptor was aligned with a residue of the template, thus minimizing the number of non-aligned template residues. However, when a solution that respected the set cutoff values could not be found, template residues were taken out of the alignment, one by one, until a solution that respected the cutoff values could be found (**Figure 2**). Conversely, for domains that had more residues in the modeled receptor than in the template (*insertions in the modeled receptor*), the alignment was adjusted in a way that the additional residues of the modeled receptor were inserted between two residues of the template that formed a sharp cusp or turn (as judged by visual inspection), and avoiding the insertion within elements of secondary structure.

Overhangs at the termini

If there were overhangs in the *N-terminal* and *C-terminal* domains of the template (i.e. residues in the sequence of the template that go beyond the termini of the modeled

sequence), these were expunged from the structure of the template prior to the construction of the models. Although non-aligned residues are not used as templates for the construction of the model, they influence the model building process by providing “excluded areas” with which the model under construction cannot overlap. Expunging the overhangs prevents this situation.

Construction of the homology models

Following the sequence alignment, for each template 10 alternative models were built with the “automodel” Class of the Modeller program [44]. The parameters for the variable target function method (VTFM) optimization were set as follows: library_schedule = autosched.slow; deviation = 4; max_var_iterations = 300. The parameters for the molecular dynamics optimization are set as follows: md_level = refine.slow. The formation of two disulfide bridges in the β_2 AR models – one between Cys 106 and Cys 191, the other between Cys 184 and Cys 190 – was enforced in the modeling procedure. As mentioned, the presence of these two disulfide bridges in the structure of the β_2 AR had been hypothesized long before the solution of the crystal structure of the receptor. Hence, their accounting in the alignment as well as the modeling procedure is fully justified [43]. After their construction, the models were evaluated on the basis of the Modeller's Discrete Optimized Protein Energy (DOPE) method to identify the top-scoring model. The structural viability of each residue of the top-scoring model was then assessed through Modeller's probability density function (PDF). If the PDF analysis of the top-scoring model showed values higher than a threshold set at 0.06 for one or more residues of the modeled receptor, lower ranking models were progressively inspected, discarding however all those with a global DOPE exceeding that of the top scoring model of more than 1% of the value of the latter. If a model with a satisfactory global DOPE and per-residue PDF values contained within the threshold was found, this was accepted as the final model. Otherwise, the total number of alternative models built was incremented from 10 to 100, and the results were inspected to identify a viable solution. In the case that this could not be found, the sequence alignment was modified to relieve the problem causing the structural abnormality of the models. Note: the application of the last two measures (i.e. increasing the number of built models or changing the alignment) was not necessary for the construction of any of the models reported in this article.

Reference structure

The accuracy of the homology models was evaluated by calculating the RMSD values between the coordinates of the models and a reference structure of the β_2 -adrenergic receptor co-crystallized in its inactive state in complex with the inverse agonist carazolol (2RH1) [45, 46]. In particular, the RMSD calculations were performed on the entire structure as well as portions relative to individual domains and regions of the receptor, all defined on the basis of the reference structure (**Figure 3**). One of the regions taken into account was the binding cavity of the β_2 -adrenergic receptor, defined as the residues located within 5 Å from the co-crystallized ligand in the reference structure. Specifically, the binding cavity region comprised the following residues: Trp 109, Thr 110, Asp 113, Val 114, Val 117, Thr 118, Phe 193, Thr 195, Tyr 199, Ala 200, Ser 203, Ser 204, Ser 207, Trp 286, Phe 289, Phe 290, Asn 293, Tyr 308, Asn 312, Tyr 316 (**Figure 3**). As mentioned, we split

the EL2 domain into two subdomains: EL2a, defined as the portion of EL2 upstream of Cys191, and EL2b, defined as the portion of EL2 from Cys 191 on.

Root mean square deviation (RMSD) of the atomic coordinates calculations

The RMSD calculations were performed through Modeller [44], by means of an “in-house” written Python script. Prior to the calculation of the RMSD values, the script superposed the α carbons of each homology model with those of the reference structure. The operation was performed through the “superpose” module of Modeller [44], with the “superpose_refine” and “rms_cutoff” parameters set to True and 5, respectively, in order to repeat the fitting procedure by including in the calculation only the equivalent atom pairs with an RMSD lower than 5 Å, until there was no change in the number of included pairs. Subsequently, two sets of RMSD calculations were performed, one relative to the α carbons only, and the other relative to all the heavy atoms (non-hydrogen atoms) of the receptor, including backbone and side-chains. Both sets of calculations were performed on the entire structure as well as individuals domains (see the *reference structure* paragraph for the definition of the domains).

Sequence identity calculations

The sequence alignments of each template with the target receptor used for the construction of the models were also subjected to a sequence analysis through Modeller [44], by means of an “in-house” written Python script. The sequence identities were calculated with respect to the entire sequence alignment as well as portions relative to individual domains and regions of the receptor, defined on the basis of the structure of the template receptors (Supporting Information, Sequence Alignments, page 4-26). The calculations took into account the presence of gaps by considering alignment positions containing a residue in one sequence and a gap in the other as non-matching residues.

Plots and correlation analyses

The plotting of the data and the calculation of the correlations were done through the Microsoft® Excel® for Mac 2011, version 14.5.4. Specifically, the correlations were calculated with the linear regression model implemented in the “trendline” function of Excel, and were expressed in terms of the square of the correlation coefficient (R^2).

Phylogenetic tree

A phylogenetic tree based on sequence similarity was constructed on the basis of the ungapped sequence alignment of the transmembrane domains of 24 Class A GCPRs for which X-ray structures have been solved in complex with antagonists or inverse agonists. A distance matrix reflecting the pairwise fractions of non-identical residues among the receptors was obtained with Clustal X, accounting for the positions with gaps and without correcting for multiple substitutions. The resulting pairwise distance matrix was used to generate the phylogenetic tree by the program Fitch of the PHYLIP package, version 3.65, applying the Fitch-Margoliash method, setting the power property to 2 [47]. Finally, the phylogenetic tree was plotted with the graphical viewer FigTree, version 1.4.0 (<http://tree.bio.ed.ac.uk/software/figtree/>) [48]. In the figure, each template receptor was marked

with a label inversely proportional to the RMSD between the resulting β_2 AR homology model and the reference β_2 AR, calculated with respect to all the heavy atoms (font size = $120/\text{RMSD}$).

RESULTS

Homology models constructed without allowing gaps in the sequence alignment of the TM domains

As indicated in the introduction, in this study we constructed 23 homology models of the β_2 AR using as templates the crystal structures of each Class A GPCRs that has been solved in complex with either antagonists or inverse agonists (**Figure 1**). Whenever, more than one structure crystallized with an agonist or an inverse agonist was solved for a given receptor, we selected the highest resolution structure. As thoroughly described in the methodology session, the models were built on the basis of a manual sequence alignment procedure similarly to the one that we described in a recent methodology article [12]. Briefly, the procedure is based on the ungapped alignment of the 7 TM domains, and the placement of gaps in the interconnecting extracellular and intracellular loops to account for differences in sequence length between the templates and the modeled receptor. The pairwise sequence alignments of each of the templates with the β_2 AR are given in the Supporting Information (Supporting Information, Sequence Alignments, page 4-26).

After their construction, we calculated the RMSD values between each model and the reference β_2 AR structure – see methodologies for further details. Plots showing the RMSD values, with respect to the heavy atoms, calculated for the entire structure, the TM domains, the loops, and the set of amino acid residues that line the binding cavity of the β_2 AR in the reference structure are shown in panel a of **Figure 4**. Moreover, plots for the individual TMs and the individual loops and the C-terminus are provided in panels b-d of **Figure 4**. Plots analogous to those shown in **Figure 4**, calculated with respect to the α carbons, are provided in the Supporting Information (Supporting Information, RMSD values with respect to the α carbons, page 28-30). Note that we did not calculate RMSD values for the N-terminus and IL3, because these domains are not visible in the β_2 AR reference structure.

As shown by the plots in **Figure 4** and pictorially illustrated in **Figure 5**, the models are endowed with overall RMSD values, calculated with respect to the heavy atoms, ranging from 2.56 Å (β_1 AR-based model) to 7.41 Å (P2Y₁-based model). The RMSD values are generally lower for the TM domains (1.50 Å to 4.21 Å, average 3.12 Å) than for the ELs (2.24 Å to 12.4 Å, average 8.86 Å) or the ILs (4.55 Å to 9.15 Å, average 6.30 Å). Moreover, the RMSD values relative to the cavity are very close to those relative to the TM domains (0.950 Å to 4.61 Å, average 2.71 Å) and closely follow their trend. Within the TM domains, the lowest RMSD values were generally detected for TM3 (1.26 Å to 3.09 Å, average 3.26 Å), while the highest were generally detected for TM1 (1.28 Å to 5.28 Å, average 3.81 Å). Within the ELs, the lowest RMSD values were generally detected for EL1 (1.86 Å to 10.2 Å, average 3.19 Å), closely followed by EL2b (1.71 Å to 6.59 Å, average 4.84 Å), *i.e.* the portion of the EL2 downstream the conserved Cys residue that connects EL2 to TM3 (Cys 191). Within the ILs, the lowest RMSD values were generally detected for IL1 (1.54 Å to 9.47 Å, average 2.16 Å).

As pictorially illustrated in **Figure 6**, the RMSD values calculated with respect to the heavy atoms indicate that the model constructed on the basis of the β_1 AR resulted to be the most accurate with respect to the overall structure (2.56 Å), the TM domains (1.50 Å), as well as the ligand binding cavity (0.950 Å). Conversely, as pictorially illustrated in **Figure 7**, the model constructed on the basis of the P2Y₁ receptor resulted to be the least accurate with respect to the overall structure (7.41 Å), while the one constructed on the basis of the P2Y₁₂ was the least accurate with respect to the TM domains (4.21 Å), as well as the ligand binding cavity (4.61 Å).

Correlation between RMSD values and sequence identity

After the construction of the β_2 AR models and the calculation of the RMSD values with respect to the reference β_2 AR, we proceeded with the study of the correlation between said RMSD values and the sequence identity between template receptors and the β_2 AR. In particular, for the RMSD value relative to the entire structure and each of the domains, we studied the correlation with the sequence identity detected between the template and the β_2 AR with respect to the entirety of the sequence alignment, the alignment of the TM domains, and the alignment of the domain in question. For instance, for TM1, we calculated the following correlations: RMSD(TM1) vs. % identity across entire sequence alignment; RMSD(TM1) vs. % identity within the TMs; RMSD(TM1) vs. % identity within TM1. The results, in terms of R^2 values, are shown in **Figure 8** for the RMSD calculated with respect to the heavy atoms. An analogous figure for the RMSD values calculated with respect to the α carbons are provided in the Supporting Information (Supporting Information, Correlation between RMSD and sequence identity – bar graphs for α carbons, page 31-32).

As the data reported in **Figure 8** illustrate, the percentage of sequence identity calculated with respect of the TM domains shows a good correlation not only with the RMSD values calculated with respect to the TM domains but also with those calculated with respect to the overall structure. In particular, as illustrated by the scatter plots shown in **Figure 9**, for both correlations we detected an R^2 higher than 0.7. **Figure 9** also provides the linear correlation equations linking the overall RMSD with the percentage of sequence identity within the TM domains (panel a) and the RMSD relative to the TM domains with the percentage of sequence identity within the TM domains (panel b). In good agreement with the detected correlation between the accuracy of the models and the percentage of sequence identity between the templates and the β_2 AR, the β_1 AR, which is the template that yielded the most accurate models, is also the one that showed the highest sequence identity with the β_2 AR (65% of identity with respect TM domains). Conversely, the least accurate models were those based on the P2Y₁, P2Y₁₂, and the FFA1 receptors, which are the furthest templates from the β_2 AR (identity with respect to the alignment of the TM domains, respectively, for both receptors: 19% for P2Y₁ and P2Y₁₂, 16% for FFA1).

As it can be noticed from **Figure 8**, with the exception of EL2a, i.e. the portion of EL2 upstream of Cys 191, the R^2 values found for the individual domains are rather low. Moreover, it is worth noting that, in majority of the cases, the individual domains show a higher R^2 with respect to overall percentage of identity or the percentage of identity in the

TM domains rather than with respect to the percentage of identity within the domain in question.

Structure-based sequence alignment as an alternative strategy upon which to base the construction of homology models

As explained in the introduction, through this work we also compared the accuracy of models constructed following a classical alignment strategy that does not allow for the presence of gaps within the TM domains with respect to that of models built on the basis of structure-based alignments that do allow for the presence of gaps within the TM domains. Specifically, we followed the structure-based alignment generated by the server available at the GPCRdb website (<http://gpcrdb.org/alignment/targetselection>) [6], which places gaps within the TM domains of the aligned receptors to account for the crystallographically observed presence of structural irregularities.

All the sequence alignments are provided in the Supporting Information (Supporting Information, Sequence Alignments, page 4-26), where each page features, side-by-side, the alignments of the β_2 AR sequence with the same template receptor obtained following the two strategies.

As the results illustrated below indicate, although the introduction of gaps within the alignments of TMs yielded models with an overall accuracy similar to that of the models constructed without gaps in TMs, placing them in the TMs led to a substantially higher accuracy in the domains where the gaps are placed.

Concerning the overall picture, as mentioned, a comparison of the average accuracy of the models constructed with a sequence alignment strategy that does not allow gaps within the TM domains, versus those constructed with a structure-based sequence alignment is provided in **Figure 10**. Specifically, the figure compares the average RMSD values between the models and the reference β_2 AR structure, calculated with respect to the heavy atoms, obtained with the two alignment techniques. As shown in **Figure 10**, there are no substantial differences in the average accuracy of all the models obtained following the two strategies. An analogous plot calculated with respect to the α carbons, supporting the same conclusions, is provided in the supporting information (Supporting Information, Comparison of average RMSD values obtained with the two alignment techniques, page 33-34).

However, as mentioned, a closer look at the individual models shows that allowing internal gaps within the TM domains consistently yielded models with a higher local accuracy in the domain where the gaps are found. Plots comparing the accuracy of the models constructed following both strategies, on the basis of each of the templates, are given in the Supporting Information (Supporting Information, Comparison of RMSD values for the individual models obtained with the two alignment techniques, page 35-43), with black arrows indicating the TM domains within which the structure-based alignment places internal gaps. As it can be noticed, the presence of internal gaps within a specific domain consistently yielded more accurate models for the domain in question. By way of example, three specific cases are illustrated in **Figures 11-13**. As shown in **Figure 11**, the β_2 AR features an extra residue within the sequence of the TM2 domain with respect to the κ opioid receptor (panel

a). As evidenced by the superposition of the structures of the β_2 AR and the κ opioid receptor, the presence of this internal gap is due to a wider helix turn in the TM2 domain of the β_2 AR with respect to the same domain of the κ opioid receptor (panel b). The comparison of the RMSD values relative to the β_2 AR reference structure detected for the models constructed following the two strategies revealed that allowing for the presence of a gap within the TM2 domain resulted in lower RMSD values (panel c). The difference between the models obtained following the two strategies is pictorially described by the structures shown in panels d and e. In particular, the superposition of the β_2 AR reference structure and the model constructed without allowing gaps within the TM domains indicates that the model has a narrower helix turn in the TM2 domain with respect to the reference structure (panel d). Conversely, the superposition of the β_2 AR reference structure and the model constructed on the basis of the structure-based sequence alignment indicates that the presence of a gap within the sequence of the κ opioid receptor yielded a helix turn with a geometry close to that of the reference structure. Similar conclusions can be drawn from the results shown in **Figures 12-13**. In particular, as **Figure 12** illustrates, the histamine H_1 receptor features a narrow turn in TM4, which is not present in the β_2 AR. By accounting for the difference between the two structures through the insertion of a gap in the β_2 AR sequence, the local accuracy of the model increased. Similarly, as **Figure 13** illustrates, the β_2 AR features a subtle bent in TM5 that is not present in the LPA1 receptor. As expected, the insertion of a gap in the LPA1 sequence, does not introduce in the β_2 AR model the bent that is absent in the template. However, it introduces in the β_2 AR model a bulge that mitigates the local accuracy in the model above and below the bent.

DISCUSSION

As mentioned in the introduction, through this controlled homology modeling study, we aimed at investigating on the presence of a correlation between the accuracy of homology models of GPCRs and the sequence identity between the modeled receptors and templates employed for their construction. Moreover, we aimed at investigating on the differences in the accuracy of models resulting from sequence alignment techniques that do not allow for the presence of gaps within the TM domains versus those resulting from sequence alignment techniques that do allow for the presence of gaps within the TM domains. We conducted our analyses by using as a case study the β_2 AR, a prototypical GPCR for which a wealth of structural information is available. In particular, we focused our analysis on the construction of models reflecting the inactive state of the β_2 AR, given the fact that a large share of the available experimental GPCR structures were solved in complex with inverse agonists or antagonists (www.costanziresearch.com/p/table.html).

As outlined in the Methodology and the Results sections, whenever multiple structures solved in complex with antagonists or inverse agonists were available for a receptor, we chose as the template the structure endowed with the highest resolution. In doing so, we also made sure that the highest resolution structure did not missing domains with respect to the lower resolution structures available for the same receptor. In general, if the highest resolution structure available for a receptor is complemented by more complete lower resolution structures, the construction of a hybrid structure in which the missing portions in

the highest resolution structure are extracted from the lower resolution structures could be a viable strategy for the generation of an optimal template.

Indeed, we observed a detectable linear correlation between the percentage of sequence identity and the accuracy of the obtained models. In particular, as indicated in **Figure 9**, the sequence identity measured within the TM domains showed a significant correlation with the accuracy of the models within the TM domains as well with the overall accuracy of the models, with a R^2 values higher than 0.7 in both cases. The R^2 detected for correlation between sequence identity and the accuracy of the residues lining the ligand-binding cavity in the β_2 AR was relatively low. In particular, as **Figure 8** illustrates, the detected R^2 value was slightly below 0.4 for the correlation between the RMSD of the binding cavity and the sequence identity calculated with respect to the all the TM domains (red bar in the Cavity bin of Figure 8). Moreover, the R^2 value did not increase substantially when the correlation was calculated between the RMSD of the binding cavity and the sequence identity of the binding cavity itself (green bar in the Cavity bin of Figure 8). However, as shown in **Figure 4**, the accuracy of the models in the ligand-binding cavity region, was always rather close to the accuracy of the models in the TM domains. Hence, predicting the accuracy of the TM domains is sufficient to get a good estimate of the accuracy ligand-binding cavity as well. Perhaps this is not surprising, since the TM domains make up for a substantial part of the residues that line the binding cavity for the β_2 AR. Hence, it is reasonable to expect that this observation will remain valid for all GPCRs that also feature a ligand-binding cavity embedded within the TM domains.

In good agreement with the linear correlation detected between the accuracy of the models and the percentage of sequence identity between the templates and modeled receptor, the β_1 AR, which is the template that yielded the most accurate models, is also the one that showed the highest sequence identity with the β_2 AR (59% and 65% of identity with respect to the entire alignment and the alignment of the TM domains, respectively). Conversely, the P2Y₁ and P2Y₁₂ receptors, which are among the furthest templates (16% and 19% of identity with respect to the entire alignment and the alignment of the TM domains, respectively, for both receptors). Notably, the FFA1 receptor, which is the only template slightly further from the β_2 AR than P2Y₁ and P2Y₁₂ in terms of percentage of sequence identity, also yielded a model endowed with a significant structural deviation with respect to the reference structure of the β_2 AR, ranking right below the P2Y₁-based and P2Y₁₂-based models in terms of highest RMSD values.

Although models of GPCRs have been classically built on the basis of sequence alignments that do not allow for the presence of gaps within the TM domains, a comparison of the solved crystal structures indicates that the TM domains of GPCRs feature different bulges, tight coils, bents or other structural irregularities that would be better modeled by allowing for the presence of gaps within the TM domains. Hence, we compared the results obtained through a classical alignment strategy that does not allow for the presence of gaps within the TM domains with those obtained through the structure-based alignment tool available at the GPCRdb website (<http://gpcrdb.org/alignment/targetselection>) [6]. According to our results, as indicated by **Figures 11-13**, the structure of the domains where the gaps are placed significantly benefits from the insertion of gaps. This being said, it must be kept in mind

that, in the case study in question, the structure-based alignments provided by the GPCRdb website were experimentally derived, since experimental structures are available not only for the templates, but also for the target receptor. This situation would not arise in a real-life homology modeling study, when the structure of the target receptor is unknown. However, concerning the global accuracy of the models, it is worth noting that, on average, the models built without allowing gaps within the TM domains are very close in accuracy to those constructed on the basis of a structure-based sequence alignment that does allow for the presence of gaps within the TMs (**Figure 10**). The results shown in **Figure 10** are perhaps not surprising, since significant local deviations detected when comparing the modeled receptor with one template will be averaged out if the same deviation is not present with respect to the majority of the templates. However, they are certainly encouraging, because, as mentioned, it is not always straightforward to predict where to add gaps in a real-life homology modeling situation in which the structure of the target receptor is unknown.

CONCLUSIONS

With the present work we quantitatively studied the modellability of the inactive state of Class A GPCRs, by constructing models of the β_2 AR as a case study. Our results show a detectable linear correlation between model/template sequence identity and model accuracy, suggesting that the likely accuracy of the homology models that can be built for a given receptor can be forecasted on the basis of the available templates. We also probed whether sequence alignments that allow gaps within the TM domains to account for structural irregularities afford better models than the classical alignment procedures that do not allow for the presence of gaps within the TMs. As our results indicate, although the overall differences are very subtle, the inclusion of internal gaps within the TM domains has a noticeable beneficial effect on the structural quality of the domain in question.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

Research reported in this publication was supported by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under award number R15GM119084. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

This research was also supported by: American University; the Schwartz Fellowship of the Chemistry Department; and the Mathias Fellowship of the College of Arts and Sciences. Computing resources were provided by the American University High Performance Computing System, which is funded in part by a grant from the National Science Foundation (BCS-1039497).

JG worked on this research as a student intern at American University, in the course of her undergraduate studies at the University of Pittsburgh. AD worked on this research as a student intern at American University, in the course of his doctoral studies at the Department of Life and Environmental Sciences of the University of Cagliari, Italy. AD gratefully acknowledges the Sardinia Regional Government for the financial support of his PhD scholarship (P.O.R. Sardegna F.S.E. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2007-2013 - Axis IV Human Resources, Objective 1.3, Line of Activity 1.3.1).

REFERENCES

1. Pierce K, Premont R, Lefkowitz R. Seven-transmembrane receptors. *Nat Rev Mol Cell Biol.* 2002; 3:639–50. [PubMed: 12209124]
2. Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there? *Nat Rev Drug Discov.* 2006; 5:993–6. [PubMed: 17139284]
3. Lefkowitz RJ. A brief history of g-protein coupled receptors (nobel lecture). *Angew Chem Int Ed Engl.* 2013; 52:6366–78. [PubMed: 23650015]
4. Kobilka B. The structural basis of g-protein-coupled receptor signaling (nobel lecture). *Angew Chem Int Ed Engl.* 2013; 52:6380–8. [PubMed: 23650120]
5. Jacobson KA, Costanzi S. New insights for drug design from the x-ray crystallographic structures of g-protein-coupled receptors. *Mol Pharmacol.* 2012; 82:361–71. [PubMed: 22695719]
6. Isberg V, de Graaf C, Bortolato A, Cherezov V, Katritch V, Marshall FH, et al. Generic GPCR residue numbers - aligning topology maps while minding the gaps. *Trends Pharmacol Sci.* 2015; 36:22–31. [PubMed: 25541108]
7. Costanzi S, Wang K. The GPCR Crystallography Boom: Providing an Invaluable Source of Structural Information and Expanding the Scope of Homology Modeling. *Adv Exp Med Biol.* 2014; 796:3–13. [PubMed: 24158798]
8. Costanzi S, Siegel J, Tikhonova I, Jacobson K. Rhodopsin and the others: a historical perspective on structural studies of G protein-coupled receptors. *Curr. Pharm. Des.* 2009; 15:3994–4002. [PubMed: 20028316]
9. Costanzi S. On the applicability of GPCR homology models to computer-aided drug discovery: a comparison between in silico and crystal structures of the beta2-adrenergic receptor. *J Med Chem.* 2008; 51:2907–14. [PubMed: 18442228]
10. Michino M, Abola E, 2008 Participants, G. Brooks C.r. Dixon J, Moulton J, et al. Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nat Rev Drug Discov.* 2009; 8:455–63. [PubMed: 19461661]
11. Costanzi S. Modelling G protein-coupled receptors: a concrete possibility. *Chim Oggi.* 2010; 28:26–30. [PubMed: 21253444]
12. Costanzi S. Homology modeling of class a g protein-coupled receptors. *Methods Mol Biol.* 2012; 857:259–79. [PubMed: 22323225]
13. Cavasotto CN, Palomba D. Expanding the horizons of G protein-coupled receptor structure-based ligand discovery and optimization using homology models. *Chem Commun (Camb).* 2015; 51:13576–94. [PubMed: 26256645]
14. Cavasotto CN. Homology models in docking and high-throughput docking. *Curr Top Med Chem.* 2011; 11:1528–34. [PubMed: 21510834]
15. Bhattacharya S, Lam AR, Li H, Balaraman G, Niesen MJ, Vaidehi N. Critical analysis of the successes and failures of homology models of G-protein coupled receptors. *Proteins.* 2012
16. Carlsson J, Coleman RG, Setola V, Irwin JJ, Fan H, Schlessinger A, et al. Ligand discovery from a dopamine D3 receptor homology model and crystal structure. *Nat Chem Biol.* 2011; 7:769–78. [PubMed: 21926995]
17. Kufareva I, Katritch V, Participants of GD, Stevens RC, Abagyan R. Advances in GPCR modeling evaluated by the GPCR Dock 2013 assessment: meeting new challenges. *Structure.* 2014; 22:1120–39. [PubMed: 25066135]
18. Kufareva I, Rueda M, Katritch V, Stevens RC, Abagyan R. Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure.* 2011; 19:1108–26. [PubMed: 21827947]
19. Liu W, Wacker D, Gati C, Han GW, James D, Wang D, et al. Serial femtosecond crystallography of G protein-coupled receptors. *Science.* 2013; 342:1521–4. [PubMed: 24357322]
20. Zhang H, Unal H, Gati C, Han GW, Liu W, Zatspein NA, et al. Structure of the Angiotensin receptor revealed by serial femtosecond crystallography. *Cell.* 2015; 161:833–44. [PubMed: 25913193]

21. Miller-Gallacher JL, Nehme R, Warne T, Edwards PC, Schertler GF, Leslie AG, et al. The 2.1 Å resolution structure of cyanopindolol-bound beta1-adrenoceptor identifies an intramembrane Na⁺ ion that stabilises the ligand-free receptor. *PLoS One*. 2014; 9:e92727. [PubMed: 24663151]
22. Tan Q, Zhu Y, Li J, Chen Z, Han GW, Kufareva I, et al. Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. *Science*. 2013; 341:1387–90. [PubMed: 24030490]
23. Wu B, Chien EY, Mol CD, Fenalti G, Liu W, Katritch V, et al. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science*. 2010; 330:1066–71. [PubMed: 20929726]
24. Chien EY, Liu W, Zhao Q, Katritch V, Han GW, Hanson MA, et al. Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science*. 2010; 330:1091–5. [PubMed: 21097933]
25. Fenalti G, Giguere PM, Katritch V, Huang XP, Thompson AA, Cherezov V, et al. Molecular control of delta-opioid receptor signalling. *Nature*. 2014; 506:191–6. [PubMed: 24413399]
26. Granier S, Manglik A, Kruse AC, Kobilka TS, Thian FS, Weis WI, et al. Structure of the delta-opioid receptor bound to naltrindole. *Nature*. 2012; 485:400–4. [PubMed: 22596164]
27. Srivastava A, Yano J, Hirozane Y, Kefala G, Gruswitz F, Snell G, et al. High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875. *Nature*. 2014; 513:124–7. [PubMed: 25043059]
28. Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, et al. Structure of the human histamine H1 receptor complex with doxepin. *Nature*. 2011; 475:65–70. [PubMed: 21697825]
29. Wu H, Wacker D, Mileni M, Katritch V, Han GW, Vardy E, et al. Structure of the human kappa-opioid receptor in complex with JDTic. *Nature*. 2012; 485:327–32. [PubMed: 22437504]
30. Chrencik JE, Roth CB, Terakado M, Kurata H, Omi R, Kihara Y, et al. Crystal Structure of Antagonist Bound Human Lysophosphatidic Acid Receptor 1. *Cell*. 2015; 161:1633–43. [PubMed: 26091040]
31. Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, et al. Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature*. 2012
32. Thorsen TS, Matt R, Weis WI, Kobilka BK. Modified T4 Lysozyme Fusion Proteins Facilitate G Protein-Coupled Receptor Crystallography. *Structure*. 2014; 22:1657–64. [PubMed: 25450769]
33. Manglik A, Kruse AC, Kobilka TS, Thian FS, Mathiesen JM, Sunahara RK, et al. Crystal structure of the mu-opioid receptor bound to a morphinan antagonist. *Nature*. 2012
34. Thompson AA, Liu W, Chun E, Katritch V, Wu H, Vardy E, et al. Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature*. 2012; 485:395–9. [PubMed: 22596163]
35. Yin J, Mobarec JC, Kolb P, Rosenbaum DM. Crystal structure of the human OX2 orexin receptor bound to the insomnia drug suvorexant. *Nature*. 2015; 519:247–50. [PubMed: 25533960]
36. Zhang D, Gao ZG, Zhang K, Kiselev E, Crane S, Wang J, et al. Two disparate ligand-binding sites in the human P2Y1 receptor. *Nature*. 2015; 520:317–21. [PubMed: 25822790]
37. Zhang K, Zhang J, Gao Z-G, Zhang D, Zhu L, Han GW, et al. Structure of the human P2Y12 receptor in complex with an antithrombotic drug. *Nature*. 2014; 509:115–8. [PubMed: 24670650]
38. Zhang C, Srinivasan Y, Arlow DH, Fung JJ, Palmer D, Zheng Y, et al. High-resolution crystal structure of human protease-activated receptor 1. *Nature*. 2012; 492:387–92. [PubMed: 23222541]
39. Okada T, Sugihara M, Bondar AN, Elstner M, Entel P, Buss V. The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J. Mol. Biol.* 2004; 342:571–83. [PubMed: 15327956]
40. Murakami M, Kouyama T. Crystal structure of squid rhodopsin. *Nature*. 2008; 453:363–7. [PubMed: 18480818]
41. Hanson MA, Roth CB, Jo E, Griffith MT, Scott FL, Reinhart G, et al. Crystal structure of a lipid G protein-coupled receptor. *Science*. 2012; 335:851–5. [PubMed: 22344443]
42. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 1997; 18:2714–23. [PubMed: 9504803]

43. Noda K, Saad Y, Graham RM, Karnik SS. The high affinity state of the beta 2-adrenergic receptor requires unique interaction between conserved and nonconserved extracellular loop cysteines. *J. Biol. Chem.* 1994; 269:6743–52. [PubMed: 8120034]
44. Sali A, Overington J. Derivation of rules for comparative protein modeling from a database of protein structure alignments. *Protein Sci.* 1994; 3:1582–96. [PubMed: 7833817]
45. Cherezov V, Rosenbaum D, Hanson M, Rasmussen S, Thian F, Kobilka T, et al. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science.* 2007; 318:1258–65. [PubMed: 17962520]
46. Rosenbaum D, Cherezov V, Hanson M, Rasmussen S, Thian F, Kobilka T, et al. GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science.* 2007; 318:1266–73. [PubMed: 17962519]
47. Fitch WM, Margoliash E. Construction of phylogenetic trees. *Science.* 1967; 155:279–84. [PubMed: 5334057]
48. Rambaut, A. Tree Figure Drawing Tool Version 1.4.0. Institute of Evolutionary Biology, University of Edinburgh; 2006-2012.

Highlights

- β_2 AR homology models in the inactive state were built using various templates
- Model accuracy was found to be correlated with model/template sequence identity
- Allowing gaps in the alignment of the TMs improved the local accuracy of the models

Figure 1.

Phylogenetic dendrogram of class A GPCRs for which crystal structures have been solved in complex with antagonists or inverse agonists. The β_2 AR, which is the object of this homology modeling study, is highlighted with a circle. The labels of the other receptors, all of which were used as templates in this study, have a font size proportional to the accuracy of the resulting β_2 AR models (see methods for further information). All receptors are human, except when otherwise indicated (c: cattle; m: mouse; r: rat; s: squid; t: turkey).

Figure 2.

Positioning of gaps in domains where there are more residues in the modeled receptor than in the template. Panel a: Cutoff values on the distance between the α carbons of the last aligned template residue upstream of the gap and the first aligned template residue downstream of the gap. Panels c-d: three alignment examples exemplifying two acceptable solutions (b and d) and one unacceptable solution (c). Amino acid residues are schematically represented as filled circles – aligned residues are colored in blue, non-aligned residues are colored in red.

B2AR	23	DVTQERDEVW	VVGMGIVMSL	IVLAIVFGNV	LVITAIKFE	RLQTVTNYFI
B2AR	73	TSLACADLVM	GLAVVPFGAA	HILMKMWTFG	NFWCEFWTSE	DVLCVTASIE
B2AR	123	TLCVIAVDRY	FAITSPFKYQ	SLLTKNKARV	IILMVWIVSG	LTSFLPIQMH
B2AR	173	WYRATHQEAI	NCYANETCCD	FFT NQ AY AI	SS IV SF YV PL	VIM V F V Y S R V
B2AR	223	FQ E A K R Q L Q K	IDKSEGR			
B2AR	263	KFCL KE H K A L	K T L G I I M G T F	T L C N L P F F I V	N I V H V I Q D N L	I R K E V I L L N
B2AR	313	W I G Y V N S G F N	P L I Y C R S P D F	R I A F Q E L L C L	R R S S L K	

Figure 3.

Sequence of the reference structure of the β_2 -adrenergic receptor (PDB ID: 2RH1) used for the calculation of the accuracy of the models. The TM domains are colored as follow: TM1: blue, TM2: brown, TM3: cyan, TM4: green, TM5: magenta, TM6: orange, TM7: purple.

The residues that line the ligand binding cavity in the structure, defined as those located within 5 Å of the co-crystallized ligand carazolol, are in bold and highlighted in yellow. The crystallographically solved structure is composed of two segment, the first one spanning residues 23-239, the second one residues 263-348.

Figure 4.

Plots showing the RMSD values with respect to the β_2 AR reference structure (PDB ID: 2RH1), calculated relatively to the heavy atoms, detected for the homology models constructed on the basis of each of the templates. Panel a shows the RMSD values relative to all residues, TM domains, ELs, ILs, and the β_2 AR ligand-binding cavity. Panel b shows the RMSD values relative to the individual TM domains. Panel c shows the RMSD values relative to the individual loops and termini. The data refer to the models constructed

following the sequence alignment strategy that does not allow the placement of gaps within the 7 TM domains.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 5.

Structural overlap of the β_2 AR homology models constructed on the basis of each of the templates and the β_2 AR reference structure (PDB ID: 2RH1). The models are shown as cyan thin tubes, which schematically represent their backbone structure. The reference structure is shown as a ribbon, which schematically represents its backbone structure, depicted with a gradient of colors ranging from red, at the N-terminus, to purple, at the C-terminus. The data refer to the models constructed following the sequence alignment strategy that does not allow the placement of gaps within the 7 TM domains.

Figure 6.

Structural overlap of the β_2 AR homology model that yielded the lowest RMSD values with respect to the β_2 AR reference structure (PDB ID: 2RH1), calculated relatively to the heavy atoms. The β_1 AR template yielded the most accurate structures with respect to all the modeled residues (panels a and b), the seven TM domains (panel c), and the cavity (panel d). In panel a, the backbone of the model and the reference structure are schematically represented respectively as a cyan thin tube and a ribbon depicted with a gradient of colors ranging from red, at the N-terminus, to purple, at the C-terminus. In panels b, c, and d, the heavy atoms of the model and the reference structure are shown as thin tubes with green carbons and thick tubes with grey carbons, respectively. The data refer to the models constructed following the sequence alignment strategy that does not allow the placement of gaps within the 7 TM domains.

Figure 7.

Structural overlap of the β_2 AR homology models that yielded the highest RMSD values with respect to the β_2 AR reference structure (PDB ID: 2RH1), calculated relatively to the heavy atoms. The P2Y₁ template yielded the least accurate structures with respect to all the modeled residues (panels a and b). The P2Y₁₂ template yielded the least accurate models with respect with the seven TM domains (panel c), and the cavity (panel d). In panel a, the backbone of the model and the reference structure are schematically represented respectively as a cyan thin tube and a ribbon depicted with a gradient of colors ranging from red, at the N-terminus to purple, at the C-terminus. In panels b, c, and d, the heavy atoms of the model and the reference structure are shown as thin tubes with green carbons and thick tubes with grey carbons, respectively. The data refer to the models constructed following the sequence alignment strategy that does not allow the placement of gaps within the 7 TM domains.

Figure 8.

Bar graph showing R^2 values relative the linear correlation between the RMSD of the heavy atoms of the models with respect to the β_2 AR reference structure and the percentage of sequence identity between the β_2 AR and the template upon which each of the model is based. The domains for which the RMSD values were calculated are shown in the x-axis. The domains within which the percentages of sequence identity were calculated are color-coded (blue: % identity for the entire sequence alignment; red: % identity within the alignment of the TM domains; green: % identity within the domain for which the RMSD was calculated). The data refer to the models constructed following the sequence alignment strategy that does not allow the placement of gaps within the 7 TM domains.

Figure 9.

For each of the β_2 AR homology models, the RMSD of the heavy atoms relative to the β_2 AR reference structure is plotted versus the percentage of sequence identity between the β_2 AR and the template upon which the model is based. Regression analysis trend lines showing the correlation between the two parameters are shown in black. Panel a: RMSD relative to the entire structure vs. % identity within the alignment of the TM domains; Panel b: RMSD relative to the TM domains vs. % identity within the alignment of the TM domains. In each panel, the three templates that yielded the least accurate models are indicated with red labels, while the template that yielded the most accurate model is indicated with a green label. Moreover, for each panel, the R^2 values and linear correlation equations are provided.

Figure 10.

Comparison of the average accuracy of the models constructed with a sequence alignment strategy that does not allow gaps within the TM domains, versus those constructed with a structure-based sequence alignment strategy that allows gaps within the TM domains. For each of the two strategies, the plots show the RMSD of the heavy atoms of the models with respect to the β_2 AR reference structure relative to the entire model as well as individual domains.

Figure 11.

Effect of gaps within the TM domains for the accuracy of the models: comparison of the TM2 domain in the β_2 AR and the κ opioid receptor. Panel a: alignment with no gaps within the TM domains and structure-based sequence alignment (the blue box indicates the TM2 domain defined according to the reference structure of the β_2 AR). Panel b: superposition of the structures of the β_2 AR (orange ribbon) and the κ opioid receptor (cyan thin tube) – a white arrow indicates a wider helix turn in the TM2 domain of the β_2 AR with respect to the same domain of the κ opioid receptor. Panel c: comparison of the RMSD values relative to the β_2 AR reference structure (2RH1) detected for the models constructed without allowing gaps within the TM domains (blue line) versus those constructed on the basis of the structure-based sequence alignment (red line) – a black arrow highlights the RMSD value relative to TM2. Panel d: superposition of the β_2 AR structure (orange ribbon) and the model constructed without allowing gaps within the TM domains (cyan thin tube) – a white arrow indicates a narrower helix turn in the TM2 domain of the model with respect to the reference structure. Panel e: superposition of the β_2 AR structure (orange ribbon) and the model constructed on the basis of the structure-based sequence alignment (cyan thin tube) – a white arrow indicates a similar geometry between the model and the reference structure.

Figure 12.

Effect of gaps within the TM domains for the accuracy of the models: comparison of the TM4 domain in the β_2 AR and the H_1 histamine receptor. Panel a: alignment with no gaps within the TM domains and structure-based sequence alignment (the blue box indicates the TM4 domain defined according to the reference structure of the β_2 AR). Panel b: superposition of the structures of the β_2 AR (yellow/green ribbon) and the H_1 histamine receptor (cyan thin tube) – a white arrow indicates a narrower helix turn in the TM4 domain of the H_1 histamine receptor with respect to the same domain of the β_2 AR. Panel c: comparison of the RMSD values relative to the β_2 AR reference structure (2RH1) detected for the models constructed without allowing gaps within the TM domains (blue line) versus those constructed on the basis of the structure-based sequence alignment (red line) – a black arrow highlights the RMSD value relative to TM4. Panel d: superposition of the β_2 AR structure (yellow/green ribbon) and the model constructed without allowing gaps within the TM domains (cyan thin tube) – a white arrows indicates a narrower helix turn in the TM4 domain of the model with respect to the reference structure. Panel e: superposition of the β_2 AR structure (yellow/green ribbon) and the model constructed on the basis of the structure-based sequence alignment (cyan thin tube) – a white arrow indicates a similar geometry between the model and the reference structure.

Figure 13.

Effect of gaps within the TM domains for the accuracy of the models: comparison of the TM5 domain in the β_2 AR and the LPA₁ receptor. Panel a: alignment with no gaps within the TM domains and structure-based sequence alignment (the blue box indicates the TM5 domain defined according to the reference structure of the β_2 AR). Panel b: superposition of the structures of the β_2 AR (green ribbon) and the LPA₁ receptor (cyan thin tube) – a white arrow indicates a bent in TM5 of the β_2 AR that is not present in the LPA₁ receptor. Panel c: comparison of the RMSD values relative to the β_2 AR reference structure (2RH1) detected for the models constructed without allowing gaps within the TM domains (blue line) versus those constructed on the basis of the structure-based sequence alignment (red line) – a black arrow highlights the RMSD value relative to TM5. Panel d: superposition of the β_2 AR structure (green ribbon) and the model constructed without allowing gaps within the TM domains (cyan thin tube) – a white arrows indicates the absence of a bent in the TM5 domain of the model with respect to the reference structure. Panel e: superposition of the β_2 AR structure (green ribbon) and the model constructed on the basis of the structure-based sequence alignment (cyan thin tube) – a white arrow indicates a similar geometry between the model and the reference structure.