PREDICT Modeling and In-Silico Screening for G-Protein Coupled Receptors

Sharon Shacham, Yael Marantz, Shay Bar-Haim, Ori Kalid, Dora Warshaviak, Noa Avisar, Boaz Inbal, Alexander Heifetz, Merav Fichman, Maya Topf, Zvi Naor, Silvia Noiman, and Oren M. Becker*
Predix Pharmaceuticals Ltd, Ramat Gan, Israel

ABSTRACT **G-protein coupled receptors (GPCRs) are a major group of drug targets for which only one x-ray structure is known (the nondrugable rhodopsin), limiting the application of structure-based drug discovery to GPCRs. In this paper we present the details of PREDICT, a new algorithmic approach for modeling the 3D structure of GPCRs without relying on homology to rhodopsin. PREDICT, which focuses on the transmembrane domain of GPCRs, starts from the primary sequence of the receptor, simultaneously optimizing multiple 'decoy' conformations of the protein in order to find its most stable structure, culminating in a virtual receptor-ligand complex. In this paper we present a comprehensive analysis of three PREDICT models for the dopamine D2, neurokinin NK1, and neuropeptide Y Y1 receptors. A shorter discussion of the CCR3 receptor model is also included. All models were found to be in good agreement with a large body of experimental data. The quality of the PREDICT models, at least for drug discovery purposes, was evaluated by their successful utilization in in-silico screening. Virtual screening using all three PREDICT models yielded enrichment factors 9-fold to 44-fold better than random screening. Namely, the PREDICT models can be used to identify active small-molecule ligands embedded in large compound libraries with an efficiency comparable to that obtained using crystal structures for non-GPCR targets. Proteins 2004;57:51–86.**
© 2004 Wiley-Liss, Inc.

Key words: modeling; docking; in silico screening; GPCR; structure based drug discovery

INTRODUCTION

G-protein-coupled receptors (GPCRs) are an important class of membrane embedded proteins involved in diverse biological functions including smell, taste, sight, and the regulation of appetite, digestion, blood pressure, reproduction, and inflammation.[1] GPCRs share a characteristic topology, consisting of an extracellular N-terminal segment, seven transmembrane helices (TMs), three exoloops (ECL), three cytoloops (ICL), and a C-terminal segment. GPCRs bind to a wide variety of molecules, including ions, amino acids, peptides, lipids, and nucleotides. Upon ligand binding, the receptor is activated leading to conformational changes that allow the receptor's cytosolic domains to interact with an intracellular heterotrimeric G-protein. The human genome contains ~800 putative GPCRs; approximately half of these are olfactory or visual receptors while the remaining have other life functions. The involvement of GPCRs in a wide variety of pathologies, such as neuronal, cardiovascular, endocrine, gastrointestinal, inflammatory, and other diseases and the diversity of ligands and activation mechanism make GPCRs ideal candidates for pharmaceutical intervention. Their medical importance is reflected by the fact that ~50% of marketed drugs act through modulation of the GPCR functions.[2]

Because GPCRs are relatively large membrane-embedded proteins, techniques such as X-ray crystallography and NMR have encountered difficulty trying to solve the 3D structure of GPCRs. Despite technical improvements over the years, the number of membrane protein structures solved is still very small, reflecting the unsolved challenges in expression, purification, crystallization, and analysis of membrane proteins. With the exception of the bovine light receptor rhodopsin, for which the X-ray structure was determined to 2.8Å resolution,[3] the atomic level structures of other GPCRs, in particular drugable GPCRs, are unknown. The only other available structure of a 7TM protein is that of bacteriorhodopsin, which was solved to a resolution of 2.5Å.[4] However, bacteriorhodopsin is not a GPCR and its 3D structure is significantly different from that of rhodopsin.

In the absence of experimentally determined structures for GPCRs, computational protein modeling was expected to fill the gap. However, having no experimental structure with which to validate such approaches, model reliability becomes a serious question. Agreement with known experimental information, primarily site-directed mutagenesis data, is viewed as an essential requirement for model reliability, though this may not be sufficient. In recent

Maya Topf's present address is Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry and California Institute for Quantitative Biomedical Research, University of California, San Francisco, CA 94143.

Zvi Naor's present address is Department of Biochemistry, Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel.

*Correspondence to: Oren M. Becker, Predix Pharmaceuticals Ltd, SAP Building, 3 Hayetzira Street, Ramat Gan 52521, Israel.
E-mail: Becker@predixpharm.com

Received 18 November 2003; Revised 2 March 2004; Accepted 18 March 2004

Published online 29 June 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20195

© 2004 WILEY-LISS, INC.

years, the model's success in in silico screening is being regarded as a good measure of the accuracy, reliability, and utility in drug discovery.[5,6]

The most straightforward approach to determining the 3D structure of a GPCRs is homology modeling, in which the new model is built based on the 3D structure of a homologous protein.[7,8] Homology modeling has proven effective for many globular proteins, and in recent years automatic model-building based on homology has become a routine technique implemented in most molecular graphics software packages.[7] However, the success of this technique is highly dependent on the extent of sequence homology between the template protein(s) and the modeled protein as well as on the number of related known structures that can be used for optimizing the template.[9] The existence of only one experimental nondrugable GPCR structure (bovine rhodopsin) and the poor sequence homology between this structure and that of drugable human GPCRs[10,11] limits the use of homology modeling as a basis for structure-driven drug design.

Nonetheless, the relative simplicity of homology modeling has led to the creation of many homology-based GPCR models using both bacteriorhodopsin and rhodopsin as templates.[12–16] Rhodopsin-based homology models, using either the Cα trace of the low-resolution data[17,18] or the recent 2.8Å X-ray structure,[3] have been developed for dopamine D2 receptor,[19] opioid receptors,[20] and a β2 adrenergic receptor[21] and others. Unfortunately, the quality of these models has not been validated by in silico screening in the context of drug discovery. In recent years, new approaches that attempt to improve the straightforward homology modeling techniques were developed. Bizzants et al.[6] optimized rhodopsin-based models by generating protein-ligand complexes with known agonists or antagonists bound to the receptor. The ligand-bound models of several receptors generated this way, including dopamine D3, vasopressin V1a, β2 adrenergic and δ opioid receptors, proved successful in computational screening tests. In another approach, Vaidehi et al.[22] used only the helical centers of the rhodopsin structure for predicting the 3D structures of rhodopsin and of the β2 adrenergic receptor. Though the root mean square distance (rmsd) between this rhodopsin model and the X-ray structure was only 3.0Å at the TM region, such a measure of quality may be flawed because the TM center coordinates from the X-ray structure were used as an input to the model generation process. The binding mode of epinephrine docked into the β2 adrenergic receptor model that was generated by Vaidehi et al.[22] was in good agreement with experimental data. Similarly, Filizola et al.[23] modeled the structure of GPCRs without using a template structure, using only limited data derived from the electron density map of rhodopsin (helical center coordinates and tilt orientations). Using this method, the authors created a model for bovine rhodopsin and manually docked 11-cis-retinal into the helical bundle. In the model structure, the location of key residues was in good agreement with experimental data. Yet, not all models based on rhodopsin structure agree with experimental results. Archer et al.[24] recently showed that the rhodopsin structure can not be directly applied for modeling of the CCK1 receptor. Bizzant et al.[6] showed that the homology-based structure of the muscarinic M1 receptor performed poorly in computational screening vs. their other models, and recently Manivet et al.[25] showed that bacteriorhodopsin, rather then rhodopsin, serves as a better template for modeling the structure of the 5-HT2B receptor.

The new NMR structure of the second intracellular loop of the α2a receptor[26] and the second extracellular loop of the κ opioid receptor[27] strengthen the notion that structural differences in the TMs as well as in the extra- and intracellular domains are responsible for the diversity of ligand selectivity and signal transduction mechanisms exhibited by GPCRs. Some models using bacteriorhodopsin appeared to be consistent with structure activity relationships of known ligands for the target receptor (e.g., AT1 receptor model[28] or CCR2b model[29]). However, the structure of bacteriorhodopsin significantly differs from that of rhodopsin and most bacteriorhodopsin-based models were later invalidated by experimental studies or by the structure of rhodopsin.[3] Several attempts were also made to apply de novo techniques for modeling GPCRs,[30] but none of these methods were able to determine with sufficient accuracy the native or near native conformations of GPCRs necessary for structure-based drug discovery.

In this report, we describe for the first time the details of the PREDICT algorithm,[5,31,32] a de novo approach for modeling the 3D structure of any GPCR that is not based on homology to known structures of rhodopsin or bacteriorhodopsin. The PREDICT method relies on the primary sequence of the receptor itself and on structural constraints imposed by the membrane environment, e.g., the general orientation of the TM domains along the membrane normal and the hydrophobic/hydrophilic character of the phospholipid regions. PREDICT is based on physico-chemical properties of a single sequence and, therefore, is substantially different from existing modeling approaches that rely on known structures or multiple sequence alignment. Following the detailed description of the algorithm and a comparison to the rhodopsin structure, we present a comprehensive analysis of the PREDICT models for three GPCR receptors: dopamine D2, neurokinin NK1, and neuropeptide Y Y1. Some discussion of the CCR3 receptor model is also included. All models were found to be in a good agreement with a large body of experimental data. The quality of the PREDICT models for drug discovery purposes was validated by their successful utilization in virtual screening. Virtual screening using all three PREDICT models yielded significant enrichment factors over random screening, being able to identify known small-molecule ligands embedded in a large library of random drug-like compounds with an efficiency that is comparable to those obtained using crystal structures of non-GPCR targets.[33,34]

MATERIALS AND METHODS

Outline of the PREDICT Modeling Approach

The PREDICT algorithm[5,31,32] is a de novo modeling approach for packing transmembrane helical bundles,

such as in GPCRs. Without relying on sequence similarity to any other protein of known structures, the algorithm predicts the native 3D conformation of a protein using only its amino acid sequence and physioco-chemical properties of the membrane environment. The concept of structural "decoys" is employed to ensure that the algorithm identifies the correct structure. Rather than optimizing a single starting structure and risk ending up trapped in a local minimum, PREDICT generates many alternative possible conformations and optimizes them simultaneously. This allows for an efficient search throughout the receptor's conformation space enabling the method to overcome local minima as it aims to distinguish between stable conformations and unstable folds.

The core of the PREDICT algorithm consists of two main steps: generation of decoys, followed by optimization and scoring of the decoys. In the first step the algorithm automatically produces hundreds to thousands of possible TM conformations (decoys). In this step, a coarse 2D grid search is conducted over the receptor conformational space, followed by initial optimization of the hydrophobic moments and protein-protein interactions in each decoy conformation, and generation of the initial 3D structures. In the second step, all the decoys are optimized and the most stable structures are identified according to the PREDICT energy function. Each 3D decoy structure is subjected to a multilevel optimization procedure, including optimization of helix orientations, helix vertical alignments (relative to the other helices and to the membrane/water boundary), helix positions, π-stacking of aromatic residues, and helical tilts. This optimization is coupled to a continuous annealing relaxation of side chain rotamers. A reduced representation of the side chain structure[35] and side chain rotamers[36] allows PREDICT to balance computational efficiency and accuracy.

For efficiency, the PREDICT algorithm is used consecutively twice: (1) first PREDICT is used to coarsely search the protein's conformation space and distinguish stable and unstable regions on the protein's energy landscape. (2) In the second application, PREDICT is used to finely search the stable, narrow conformational regions identified in the coarse search. This two-tier modeling approach improves computational efficiency while also ensuring an effective coverage of the protein's conformation space. The final model is further optimized using molecular dynamics simulations and virtual complexation with a known ligand. Namely, the resulting modeling procedure includes the following four major steps:

1. First "coarse" application of the PREDICT algorithm.
2. Second "fine" application of the PREDICT algorithm.
3. Molecular dynamics refinement
4. Generation of a virtual receptor-ligand complex

PREDICT thus represents a comprehensive multistep approach to GPCR modeling. While a few of the features in PREDICT (e.g., hydrophobic moments) have been previously used in other modeling methods, the algorithm includes many new and unique features that have not been reported before. Among these are: Monte Carlo–based side-chain rotamer optimization coupled to each and every other optimization step, a novel tilting algorithm that explores all possible 3- and 4-helical tilts along arches, the parallel optimization of many decoy structures, and the two-tier approach to the search problem (i.e., two sequential applications of PREDICT), which enables an efficient search throughout conformational space.

General Assumptions for PREDICT GPCR Modeling

Our approach to 3D modeling of seven TM proteins benefits from several observations and assumptions to constrain the modeling problem. These include:

1. The TMs are amphiphatic and have a length of ~20–30 residues.
2. The loops connecting the helices are relatively short[17] indicating that (1) the helices are packed in an antiparallel orientation and (2) the TMs are arranged in a sequential topology, so that the TM order along the sequence is also their order in the folded structure.
3. The TM helices are arranged in a counter-clockwise manner when viewed from the extracellular side as was shown for rhodopsin[3] and bacteriorhodopsin.[4]
4. Being embedded in a hydrophobic environment suggests that hydrophilic side-chains of the amphiphatic TM helices will favor the interior part of the protein, creating a "closed" structure in which the membrane-exposed surface is hydrophobic.

Determining the Seven TMs

Most modeling procedures require a precise definition of the sequences of the TM regions, which is usually obtained from multiple sequence alignments using different statistical approaches such as PHDhtm[37] or TMPRED.[38] Though identification of transmembrane regions is routinely performed, the precise determination of the helix sequences remains fuzzy. Namely, it is uncertain where the helices start and where they end. For this reason, the PREDICT algorithm does not to rely on this initial definition of the TM region. All it requires is a semi-fuzzy identification of the helical domain, around which it takes a sequence that is longer than the expected TM itself (stretches of 28 amino acids or longer). Within these longer stretches of amino acids, the algorithm later identifies the actual TM domain using its 3D optimization and energy-scoring methods. The resulting algorithm is thus essentially independent of the method used for determining the TM domains. Any such method can be used. To verify this, three TM identification methods were evaluated: PHDhtm,[39] TMPRED,[38] and a third method that combines a general identification of the TMs by Swissprot followed by a per-helix sequence alignment to Baldwin's patterns of conserved residues.[40] We found that the final PREDICT optimized model was not sensitive to the TM identification method used, as all three methods led to the same results. For modeling class A GPCRs, we selected the third method as the default method in PREDICT since it enables the incorporation of

experimental data into the general TM identification. Modeling GPCRs from other subfamilies (class B and C GPCRs) will require the use of the appropriate conservation motifs that are unique to those families. These can be derived, for example, from the PRINTS fingerprints compendium.[41] In the case of orphan receptor, algorithms such as TMHMM[42] or PHDhtm[39] can be used to guide the identification of the helices from the protein sequence.

Numbering Scheme

The general numbering scheme of Ballesteros et al.[43] is used throughout this report to identify positions in GPCR sequences independent of the residue numbering in each particular sequence. Each residue in a TM segment has a unique identifier composed of the helix number (1 to 7) and a position number. After defining the most conserved residue in each TM segment as position 50, the other amino acids are identified with increasing numbers toward the TM C-terminus, and decreasing numbers toward the TM N-terminus. For example, residue Trp386 of the dopamine D2 receptor is also designated 6.48, indicating that it is two residues upstream from the conserved Pro in TM6.

Generating Multiple Decoys in 2D

In the first step of the modeling process, the helices are treated as two-dimensional dials, formed by projection of perpendicular ideal helices with radius 5–6 Å[44] onto the membrane plane. Simple geometrical rules are used to systematically generate all "reasonable" close-packing conformations (decoys) in 2D. The assumption that the seven helices form a closed structure of simple topology is implemented by limiting the maximal diameter of the molecule to 5-times the diameter of a single helix and limiting the maximal distance between any two neighboring helices to no more than 4-times this diameter. This value was selected to ensure the sequential topology and to prevent penetration of phospholipids into the protein packed core. An iterative grid search is used to systematically generate all 2D conformations of the 7 TMs that abide by the above rules. The grid search is conducted on the angles that are formed between every three adjacent helices. The search angle values are variable and can be optimized according to the search requirements (grid steps of 15° are usually used for the coarse search and of 6° are used for the fine search).

To further reduce the number of initial 2D conformational decoys, secondary constraints based on experimentally known interactions between non-adjacent helices, such as in the case of correlated double mutations,[45] can be applied. These constraints are introduced as a vehicle through which experimental system-specific information can be used in the modeling. For example, Zhou et al.[46] suggested direct hydrogen bonding between Asn from TM2 and Asp from TM7 of the GnRH receptor. These two residues are interchanged in the GnRH receptor and are highly conserved, as Asp in TM2 and Asn in TM7, in the GPCR superfamily. The distance between the Cα atoms of these residues in the rhodopsin structure is 9.3Å, indicating close proximity between the helices (although not a direct hydrogen bonding). Thus, for many class A GPCR during the generation of the initial 2D conformations, one can add a secondary requirement that the centers (in 2D) of TM2 and TM7 must be no more than 20 Å apart, discarding all closed-packed conformations that do not conform to this requirement. Constraining distance in this manner focuses the search to conformations in which these two helices are fairly close, but it does not force a direct interaction between residues of the two helices. Other knowledge-based secondary constraints can be added in a system-specific manner. These secondary constraints should be used only when the experimental data justifies it. When such information does not exist (e.g., non-class A GPCRs or orphan GPCRs) or when the experimental data contradict the constraint, these secondary constraints can be omitted altogether.

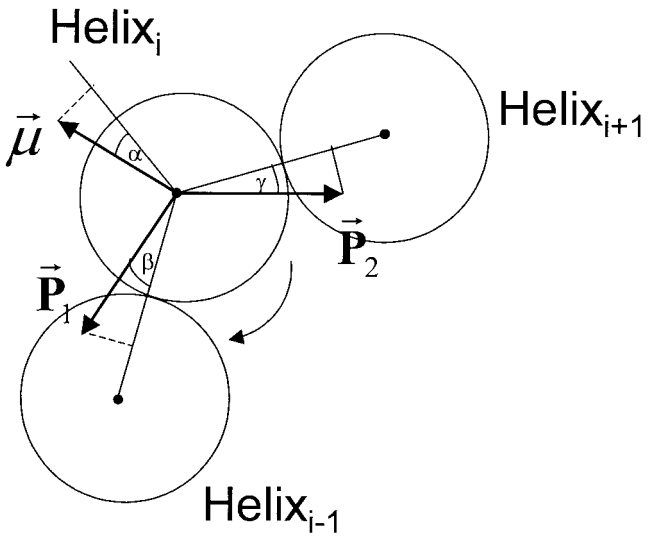
Optimizing Helical Rotational Orientation in 2D

The relative rotational orientation of the helices in each of the acceptable close packed 2D conformations is optimized. A common approach for orienting TM helices involves assigning a hydrophobicity moment to each TM helix.[47,48] These vectors, which utilize various hydrophobicity scales, are then directed towards the lipid membrane orienting the helices accordingly. However, studies on membrane proteins demonstrated that hydrophobic moments are not sufficient for determining the solvent-accessible surface of TM helices.[49,50] Aromatic-aromatic interactions are known to play a significant role in stabilizing both globular and membrane proteins.[51–53] Some studies have indicated that more then 60% of Phe residues participate in aromatic stacking and about 80% of these involve more then two aromatic residues. Moreover, aromatic residues in alpha helices participate in both intra-helical and inter-helical π interactions,[53] thus affecting the rotational orientation of the TMs. The PREDICT algorithm accounts for these effects during decoy generation, in addition to a more comprehensive treatment as part of the PREDICT energy function. A novel three-vector method ($\vec{\mu}$, $\vec{\mathbf{P}}_1$, $\vec{\mathbf{P}}_2$) is used for orienting the TM helices by balancing hydrophobicity vector $\vec{\mu}$, which accounts for membrane-helix intra-helical aromatic interactions, with two vectors $\vec{\mathbf{P}}_1$, $\vec{\mathbf{P}}_2$ reflecting inter-helical aromatic interactions (Scheme 1).

Hydrophobicity Vector

The membrane-exposed surface of each TM helix is defined by the arc formed between the center of helix i and the centers of its two neighboring helices, $i+1$ and $i-1$. To describe hydrophobic interactions with the membrane, a projection of the hydrophobic vector ($\vec{\mu}$) onto the exposed-surface arc bisector is calculated (Scheme 1), where the hydrophobic moment $\vec{\mu}$ is defined as:

$$\vec{\mu} = \frac{1}{m}\sum_{i=1}^{m} H_i \vec{\mathbf{S}}_i \qquad (1)$$



Scheme 1. A schematic representation of the three vectors approach ($\vec{\mu}$, $\vec{\mathbf{P}}_1$, $\vec{\mathbf{P}}_2$for rotational optimization of the TM helices, where helices *i*-1, *i* and *i*+1 indicate any three neighboring helices. Rotation of helix *i* in the 2D plane is performed to maximize the weighted sum of these three vectors projected on the defined axes (Eq. 5).

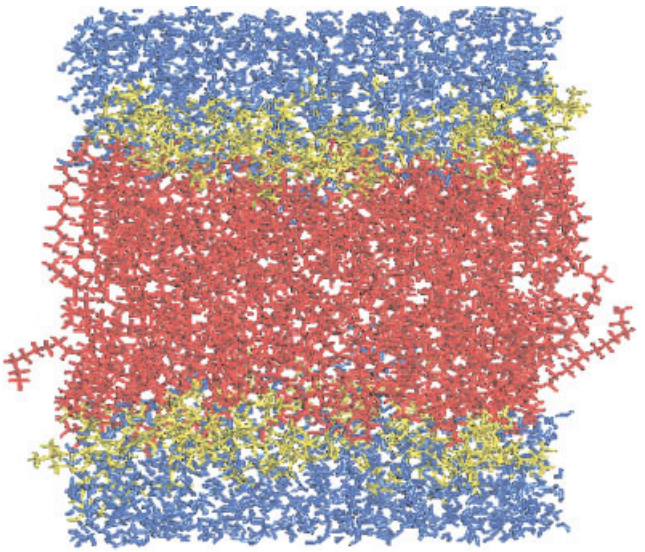


Fig. 1. An atomistic picture of a phospholipid bilayer. Depicted is a snapshot from a molecular dynamics simulations of a 1,2-dipalmitoyl-3-sn-phosphatidylcholine, DPPC.[83,84] The three regions of the membrane are color coded: The hydrophobic core formed by the phospholipid hydrocarbon tails is shown in red, the polar (or charged) interfaces on both sides which are formed by the phospholipid head groups are shown in yellow, and the water regions are shown in blue.

where m is the length of the helix, H_i is the hydrophobicity of the i-th residue and $\vec{\mathbf{S}}_i$ is a unit vector directed from the helix axis to the C_α atom of residue i.

As mentioned above, secondary structure predictions of TM helices are complicated by the vague definition of the helix end points. In part, this is because the cell membrane is a complex environment constructed of three biophysically different regions. The hydrophobic interior of the bilayer, composed of the lipid hydrocarbon chains, is separated from the surrounding aqueous environment by the hydrophilic phospholipid head-groups (e.g., phosphatidylcholine), which form the outer surfaces on both sides of the membrane (Fig. 1). Ideally, the helix's hydrophobic moment should be calculated only for lipid embedded residues. Inclusion of hydrophilic residues that are in contact with the phospholipid head groups (or even water-exposed residues) in the calculation of the hydrophobic moment $\vec{\mu}$ may lead to an error in the predicted helix orientation. To address this issue without defining the exact location of the helix boundary, the PREDICT algorithm gradually reduces the weights of the contribution of residues near the membrane surface to the hydrophobic vector. For simplicity, we applied a trapezoidal weighting mask, which linearly reduces down the contribution of the helix's terminal residues, ensuring that the hydrophobic moment is dominated by lipid-embedded residues (towards the center of the helix). The weighted hydrophobicity vector is given by:

$$\vec{\mu} = \frac{1}{m}\sum_{i=1}^{m} w_i H_i \vec{\mathbf{S}}_i \qquad (2)$$

where w_i are normalized weight factors which are set to unity for the 16 central residues of the helix and are then linearly reduced to zero over a range of six additional residues towards both ends of the helix.

As discussed above, intra-helical aromatic-aromatic interactions may play a significant role in stabilizing protein structures. Although aromatic residues are hydrophobic in nature and as such will prefer to face the lipid environment, recent potential of mean force calculations supports the conclusion that aromatic stacking is favored in an aqueous environment.[52] For GPCR modeling, this means that groups of 2–3 aromatic residues positioned one above the other along the helix (at positions i-3/i-4 relative to each other) may prefer to face the hydrophilic core of the TM bundle to optimize their intra-helical aromatic stacking. The PREDICT algorithm searches for potential instances of intra-helical aromatic residue stacking (Phe/Trp/Tyr/His) and when found automatically changes the hydrophobicity H_i of the participating residues to reflect the tendency of this aromatic cluster to face the hydrophilic interior of the membrane embedded protein.

Inter-helical interaction vectors

Two new vectors ($\vec{\mathbf{P}}_1$, $\vec{\mathbf{P}}_2$)are defined to account for aromatic interactions between neighboring helices. Both $\vec{\mathbf{P}}$ vectors are calculated in a similar way, by summing the contribution of the aromatic residues that face one or the other of the two neighboring helices,

$$\vec{\mathbf{P}}_x = \frac{1}{n_x}\sum_{i=1}^{n_x} \omega_i w_i \vec{\mathbf{S}}_i \quad x = 1, 2 \qquad (3)$$

where

$$\omega_i = \begin{cases} 1 & \text{residue } i = \text{Phe, Tyr, Trp, His} \\ 0 & \text{otherwise} \end{cases} \qquad (4)$$

n_x for $x = 1, 2$ *is the number of residues facing either helix* $i-1$ or helix $i+1$, respectively ($n_x < m$), w_i is the normalized weight factor defined in Eq. (2) and each residue can contribute to only one of these two interaction vectors.

Optimization of helix orientation

The algorithm balances the hydrophobic moment $\vec{\mu}$ with these inter-helical interaction vectors by considering the projections of both $\vec{\mathbf{P}}$ vectors along the lines connecting helix i with its adjacent helices. The final orientation of the helix is the one that maximizes the sum S of the three vector projections,

$$S = w_\mu \vec{\mu} \cos \alpha + w_{p1} \vec{\mathbf{P}}_1 \cos \beta + w_{p2} \vec{\mathbf{P}}_2 \cos \gamma \qquad (5)$$

where w_μ, w_{p1} and w_{p2} are the weights of the $\vec{\mu}$, $\vec{\mathbf{P}}_1$ and $\vec{\mathbf{P}}_2$ vectors, respectively, and α, β, and γ are the angles between the $\vec{\mu}$, $\vec{\mathbf{P}}_1$ and $\vec{\mathbf{P}}_2$ vectors and the projection axes (Scheme 1). The S score is maximized by rotating the helix along its axis using 2° steps.

Choosing hydrophobic scales

The value of the hydrophobicity moments depends on the choice of the hydrophobicity scale used of which there are several to choose (e.g., the scales of Kyte and Doolitle[54] or Eisenberg et al.[48]). Cornette et al.[55] compared the effectiveness of 38 hydrophobicity scales in identifying α-helical and β-strand periodicity and generated a "consensus scale" that optimizes secondary structure predictions of amphipathic α-helices. In this consensus scale, the Arg residue is treated as if it were a hydrophobic residue and not hydrophilic as expected. This assignment stems from the observation that arginine and lysine residues tend to appear close to the ends of the TM helices and facing the membrane.[56] However, this apparent "hydrophobicity" is, in fact, just a reflection of the ionic interactions between the Arg side-chain and the phospholipid head-groups that cause the Arg to face the membrane. These interactions may serve to anchor the TM helices in the membrane as suggested by NMR studies of other membrane proteins.[57] On the other hand, Arg residues located along the helix deeper into the membrane behave as expected, orienting away from the hydrophobic lipids. The PREDICT energy function (see below) accounts for this distinction using a three-layer membrane representation (Fig. 1). The effect of Arg residues on the hydrophobic moment is implicitly reduced by the down weighting of residues near the helical boundaries in the calculation. Ee found that the Eisenberg scale,[48] which treats Arg residues as very polar, value gave the best results in our calculations.

From 2D Dials to a Reduced 3D Structure

A three-dimensional skeleton, which includes the protein backbone atoms N, Cα, and Cβ, is constructed from each rotationally oriented 2D helical structure (decoy) prepared in the previous steps. These atoms serve as the scaffold to which a reduced representation of the amino acid side chains will be added. The reduced representation used here was originally developed by Herzyk and Hubbard for use with dynamic simulated annealing.[35] This reduced representation preserves the overall shape of the side chains as much as possible and, more importantly, it allows for side chain rotamer motion. Except for glycine (which has no side chain), all residues are classified into three groups having one, two, or three virtual side chain atoms. Side chains are added to the helical backbone in their reduced representation using the virtual bonds and angles of Herzyk and Hubbard, and the rotamer dihedral angles that characterize each amino acid. For each side chain, the center of the first virtual atom ($C\alpha'$) is set at the backbone $C\alpha$ position. Then the center of mass of the N main contributing atoms included in that specific second virtual atom, $C\beta_{com}$, is calculated according to their real coordinates (x,y,z) and masses (m):

$$C_{\beta_{com(x)}} = \frac{\sum_{i=1}^{N} m_i x_i}{\sum_{i=1}^{N} m_i} \quad C_{\beta_{com(y)}} = \frac{\sum_{i=1}^{N} m_i y_i}{\sum_{i=1}^{N} m_i} \quad C_{\beta_{com(z)}} = \frac{\sum_{i=1}^{N} m_i z_i}{\sum_{i=1}^{N} m_i} \qquad (6)$$

Next, the second virtual atom ($C\beta'$) is set in the same direction as $C\beta_{com}$ (at the direction of $C\alpha$-$C\beta_{com}$ vector), but with different length from $C\alpha$, according to the appropriate virtual bond. If an amino acid has only two virtual atoms, only the $C\beta'$ coordinates have to be calculated. However, larger amino acids, which have more than two virtual atoms, require an additional calculation. The parameters that determine the centers of the third and the fourth virtual atoms are the rotamer dihedral angles that characterize the specific amino acid (χ_1, χ_2), the appropriate virtual bonds length, and the appropriate angles based on the Herzyk and Hubbard model. N coordinates of the native protein are also used in order to set the $C\alpha$ chirality and consequently $C\beta'$ chirality.

Rotamers

To explore the different possible rotamers for each amino acid, the algorithm uses the backbone-independent rotamer library of Ponder and Richards,[36] which refers to the individual populations of rotamers without reference to the local backbone conformation angles (ϕ and ψ). Since the maximal number of virtual atoms in the Herzyk and Hubbard representation is four, they can represent only the χ_1 and χ_2 side chain dihedral angles (first and second dihedral angles along the side chain) the values of which are taken from the above rotamer library. The use of a rotamer library enhances efficiency because side-chain dihedral angles are not evenly distributed; rather they tend to adopt a limited set of preferred angles. Such rotamer libraries are usually derived from statistical analysis and clustering of side-chain conformations in known structures of proteins.[58] The Ponder and Richards library, which is used in this study, includes average dihedral angles, variances, and their frequencies.[36]

The PREDICT Energy Function

Once the initial 3D decoy structures are generated, the algorithm continues to optimize each of these structures in an attempt to minimize the PREDICT energy function. Clearly, the energy function is a key component in reliably distinguishing correct folds from incorrect ones. The efficiency of the energy function is highly dependent on

details present in the studied system. Although energy estimations for reduced representation models are usually based on effective residue-residue interaction energies, evidence suggests that a very simple contact energy function is not sufficient for separating the native folds from less stable decoys.[59–61] The PREDICT energy score includes a distance-dependence function f_{ij} superimposed on the residue-based contact energy function,

$$E_{ij} = \varepsilon_{ij} \cdot f_{ij} \tag{7}$$

where ε_{ij} are the Miyazawa and Jernigan[62] contact energies between residue i and j and f_{ij} is the distance function described hereafter.

Distance function

The distance dependence function f_{ij} has the general shape of a soft n-m potential function, allowing the algorithm to reject structures with significant side-chain overlaps as well as to identify lack of interactions between residues that are too far apart.

$$f_{ij} = \sum_{k=\alpha}^{Mi} \sum_{l=\alpha}^{Mj} \frac{1}{M_i \cdot M_j} \cdot \left[a \cdot \left(\frac{R_{eff}(C'_{ik}) + R_{eff}(C'_{jl})}{R_{C'_{ik}C'_{jl}}} \right)^m - b \cdot \left(\frac{R_{eff}(C'_{ik}) + R_{eff}(C'_{jl})}{R_{C'_{ik}C'_{jl}}} \right)^n \right] \tag{8}$$

where $R_{eff}(C'_{ik})$ and $R_{eff}(C'_{jl})$ are the type-dependent effective radii of virtual atoms k and l of residues i and j positioned at coordinates C'_{ik} and C'_{jl}, respectively, and $R_{C'_{ik}C'_{jl}}$ is the distance between these two virtual atoms. This distance function is based on summing the interaction energies between all possible pairs of virtual atoms (C'_{ik}, C'_{jl}), except for two neighboring C_α virtual atoms that do not contribute to the interaction. The number of interactions between every two residues depends on the number of virtual atoms in each residue (M_i, M_j).

The powers m and n, $m>n$, determine the shape of the distance function f_{ij} of Eq. 8, where $m = 12$ and $n = 6$ represents an atomic Lenard-Jones 6-12 potential. After testing various functions, we found that the best results were obtained using $m = 6$ and $n = 4$, making the potential much softer then the Lenard-Jones potential, as may be expected from the multi-atomic nature of the virtual "atoms." We found that higher powers on the repulsive part overemphasized the repulsion, making the attractive interactions effectively negligible, while smaller powers were less effective in distinguishing different structures. The 4th power on the attractive part is set to ensure that the characteristic range of these interactions is about twice the sum of the residue radii (which range from 2.0 Å for Gly to 5.0 Å for Trp[63]). The constants a and b also play a role in defining the depth of the potential well in Eq. 8. It was found that the values $a = 5$ and $b = 6$ best describe the interaction potential within the current reduced representation of the system.

Energetic contribution due to Aromatic Interaction

Aromatic residues (Phe, Tyr, Trp, and His) have the highest propensity to be in the tightly packed TM domain[52] and are of great importance in stabilizing protein structure. Aromatic residues are widely distributed and highly conserved among GPCRs, especially in TM6 and TM7. Furthermore, not only do most of these aromatic side-chains participate in aromatic interactions, they also tend to form aromatic clusters.[53,64] In the context of catecholamine receptor. it has been suggested that binding of an agonist to such an aromatic cluster may be involved in receptor activation.[65]

Hydrophobic and aromatic interactions are favored by the Miyazawa and Jernigan[53,62] contact energy table on which our energy function is based. The lowest contact energy in that table is −1 for Leu-Leu interactions, with aromatic-aromatic interactions scoring between −0.99 (Phe-Phe) and −0.42 (Tyr-His). For comparison, the least favorable contact energy in the table is −0.016 for Lys-Lys interaction and that the average interaction value is −0.42. The basic Miyazawa and Jernigan contact energies, ε_{ij}, do not account for the added stability from clusters of aromatic residues; therefore, we added an aromatic-clustering factor (λ_{arom}) to the energy function as a user defined parameter, α, allowing the user to increase the relative weight of aromatic-aromatic interactions. This modifies the interaction energy as follows,

$$\varepsilon'_{ij} = \varepsilon_{ij} \cdot \lambda_{arom} \tag{9}$$

where

$$\lambda_{arom} = \begin{cases} \alpha > 1 & i \text{ and } j = \text{Phe, Tyr, Trp, His} \\ 1 & \text{otherwise} \end{cases} \tag{10}$$

Energetic contribution due to cation-π interactions

Cation-π interactions are recognized as an important non-covalent binding interaction in α-helical peptides.[66] Recent studies showed that the average energetic contribution of Arg-Phe, Arg-Tyr, and Lys-Phe is approximately 6.5 kcal/mol in membrane proteins.[66] Burely and Petsko[51] found that the average distance between positively charged amino groups and the ring centroids of Phe, Tyr, or Trp, where they make van der Waals contact with the π-electrons, is approximately 6 Å. To allow the introduction of cation-π contributions to the energy score a second user defined parameter, β, was added so that the residue-residue interaction energy could be scaled by a cation-π factor λ_{cat},

$$\varepsilon'_{ij} = \varepsilon_{ij} \cdot \lambda_{cat} \tag{11}$$

where

$$\lambda_{cat} = \begin{cases} \beta > 1 & i = Arg, Lys \text{ and } j = \text{Phe, Tyr, Trp, His} \\ 1 & \text{otherwise} \end{cases} \tag{12}$$

Polar interactions

Hydrophobic interactions play a significant role in stabilizing TM helices in the nonpolar membrane. Yet, several studies have pointed to specific polar interactions implicated in driving TM helix association,[67,68] especially in the

hydrophilic core of GPCRs. Since the Mayzawa and Jernigan contact interactions table favors hydrophobic interactions, we used a polar factor (λ_{polar}) to strengthen the contribution of such polar-polar interactions to the overall energy score of the receptor conformation.

$$\varepsilon'_{ij} = \varepsilon_{ij} \cdot \lambda_{polar} \tag{13}$$

where

$$\lambda_{polar} = \begin{cases} \gamma > 1 & i = Arg, Lys, Asp, Asn, Gly, Gln \\ 1 & \text{otherwise} \end{cases} \tag{14}$$

Interaction with the membrane

The effect of positive and negative charged amino acid on the helix conformation may differ according to their position along the helix. For example, Arg and Lys both have very long flexible side-chains with an appreciable nonpolar segment connected to the terminal charged moiety, allowing them to bury a substantial fraction of their nonpolar surface area in the membrane while still exposing the charged moiety to surrounding water.[69] Therefore, it can be assumed that Arg and Lys positioned at the central portion of a TM helix will face the "inside" of the helical bundle adopting favorable interactions with residues from other helices, while Arg and Lys positioned towards the end of the helix will face "outside" adopting energetically favorable interactions with the membrane (the phospholipid head-groups). On the other hand, Asp and Glu have only short side chains with little amphiphatic character and thus will have unfavorable energetically interactions with the lipophilic membrane. To accommodate the different interactions of the amino acids with the different regions of the membrane, we define a membrane interaction $E_{membrane}$term as:

$$E_{membrane}(\mathrm{Res}_i) = \\ = \begin{cases} a < -1 & \mathrm{Res}_i = Arg, Lys, Asp, Glu \text{ and } 6\text{Å} < Z_i < 24\text{Å} \\ b > 1 & \mathrm{Res}_i = Arg, Lys \text{ and } (Z_i < 6\text{Å or } Z_i > 24\text{Å}) \\ 1 & \text{otherwise} \end{cases} \tag{15}$$

where Z_i is the position of residue i along the Z-coordinate, normal to the membrane plane and assuming the membrane is 30Å thick.

Overall energy function

The overall energy function used for optimizing the 3D conformations of all decoy structures, therefore, includes two parts, a residue-residue interaction term and a single residue term reflecting its interaction with the membrane,

$$E = \sum_{i,j} E_{\mathrm{int}}(\mathrm{Res}_i, \mathrm{Res}_j) + \sum_i E_{membrane}(\mathrm{Res}_i) \tag{16}$$

where Res_i and Res_j are the two interacting residues and their interaction energy $E_{int}(Res_i, Res_j)$ is given by

$$E_{\mathrm{int}}(\mathrm{Res}_i, \mathrm{Res}_j) = \varepsilon_{ij} \cdot \lambda_{arom} \cdot \lambda_{cat} \cdot \lambda_{polar} \cdot f_{ij} \tag{17}$$

Optimization of the 3D Structures

Each 3D structure is subjected to four different types of self-consistent optimization processes: optimization of the vertical alignment of the helices relative to each other and to the membrane surface, refinement of the inside-out distributions of the residues of each helix, optimization of the position of the helix center in x-y plane, and assignment of helical tilt angles, performed in a sequential manner. A self-consistent iterative process is used for each optimization process, sequentially optimizing one helix after the other, scanning the possible range of parameters for each helix. Each optimization attempt is followed by Monte Carlo re-optimization of side-chain packing in the resulting transient conformation. The process iterates repeatedly through this sequential seven-helix cycle until the energy of the conformation converges over a complete iteration cycle. In the receptors modeled this far, the energy converged within 3 to 5 iteration cycles. Parallel optimization of all four processes simultaneously was tested but did not yield significant improvement over the sequential optimization process while at the same time requiring significantly greater computer resources.

Monte-Carlo optimization of side-chain packing

Each conformation generated during this multilevel optimization process is subject to a Monte Carlo optimization of side-chain packing. In fact, one of the main reasons for using the reduced representation and the rotameric representation was to enable efficient optimization of local side-chain packing, which is done using a Monte Carlo search in rotamer space. Each Monte Carlo move includes a random selection of up to three interacting residues among the seven TM helices. The rotamers of the selected amino acids are randomly changed until the interaction energy between these residues reaches a minimum. A Metropolis criterion[70] determines whether the move is accepted based on the energy differences ΔE_i between the energy of the trial configuration E_i, which is the lowest energy among all possible rotamers of that structure, and the energy of the last accepted configuration $E_{i\text{-}1}$. If ΔE_i is negative, the move is accepted; otherwise, the trial conformation is evaluated according to the Boltzmann distribution $e^{-\Delta Ei/RT_n}$ where the temperature of the system T_n is either fixed or a varying annealing temperature. The trial configuration is accepted if:

$$p < e^{-\Delta Ei/RT_n} \tag{18}$$

where p is a random number in the interval [0,1]. The search continues through a large number of trials in order to allow convergence of the energy.

Vertical alignment optimization

The vertical positioning of the helices relative to each other and relative to the membrane surface can play a major role in stabilizing the molecule. Due to the imprecise definition of the TM boundaries, optimal a priori vertical positioning of the helices is not possible, leading to possible mismatches resulting in loss (or gain) of stabilizing interactions. Shift of a few Å along the z-axis (perpendicular to the

membrane surface) may result in the gain or loss of a favorable intramolecular polar interaction or can bring a polar charged residue into favorable contact with the hydrophilic head group region of the membrane and contribute to the anchoring of the receptor in the membrane. To determine the vertical alignment of the helices, PREDICT uses an iterative self-consistent procedure in which the vertical position of each individual helix is optimized sequentially within a ±5 Å vertical range (approximately one helical turn) and iterating this process through all TM helices until convergence is reached for the whole protein. As with other steps in PREDICT, short side chain rotamers minimization procedure is used to reoptimize the side chain packing against neighboring helices between iterations. All proteins tested by us converge within 2 to 3 iterations around the molecule.

Refinement of rotational orientation of the helices

The inside-out distributions of the helical residues were determined in the 2D helical rotation optimization step discussed above. However, this preliminary orientation does not take into account the details of the 3D structure nor the effect of side chain packing. We introduce local refinements to the helix orientations to allow the structure to adopt energetically favorable intramolecule interactions. In this step, we allow the hydrophobicity vector to rotate ±15° around its initial direction as determined in the 2D rotation optimization step. As with previous PREDICT stages, this step was performed in an iterative self-consistent procedure coupled to Monte Carlo side chain optimization on each trial conformation.

Optimization of the Position of the Helix Center

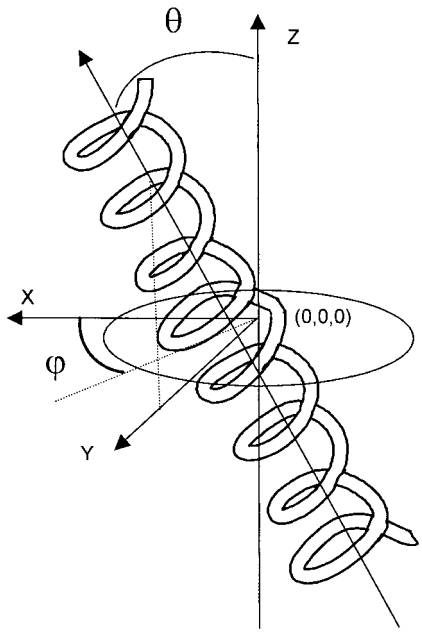
The initial 2D geometries were constructed assuming fixed distance between the centers of neighboring helices. This assumption was based on the high helical packing value that is characteristic of membrane helical bundles. Gly and Ala have the highest packing values in membrane proteins while the large aromatic residues (Trp, Tyr and Phe) have the lowest values.[71] However, in GPCRs there are several known intrahelical interactions that involve aromatic residues, indicating that these residues are located at the interaction interface between helices and may disrupt the tight helical packing, resulting in higher distance between two neighboring helices.[65,72,73] Moreover, it is known that the length of the extra- and intracellular loops, connecting the helices, varies between the different GPCRs, thereby permitting the distance between two neighboring helices to vary. Changes in the distances between neighboring helices may notably affect the overall helical bundle, and may therefore result in gain or loss of important intramolecular interactions. PREDICT employs an iterative self-consistent procedure in which each helix is moved step by step in circle of 1 Å radius around its initial center. A minimization procedure is performed after each iteration to optimize the rotamers of the involved residues, similar to the optimization steps described above.

Assignments of inter-helical tilts

One of the most important structural features of helix-helix packing in both globular and transmembrane proteins, is their tendency to be tilted relative to each other thus increasing the inter-helical interface.[74] Such tilts are observed in the X-ray structures of bacteriorhodopsin[4] and rhodopsin[3] and, in a survey of 45 TM helices, Bowie[75] found that the distribution of helix packing angles for TM membranes (~+21°) vary considerably from those in soluble proteins (~−35°). The same study also found that the larger tilt angles are more likely than that for smaller tilt angles. One of the reasons for the restricted packing angles found for membrane proteins is their relatively large size (average length of 26.4 amino acids compared to 12 amino acids for globular proteins), which result in large packing interfaces. Efficient packing is clearly one of the important forces stabilizing TM helices. Helix-helix interactions with the helices nearly parallel to each other (but not necessarily to the membrane plane) are much more favored in membrane proteins, which likely contribute to a more efficient packing density.[76] Given the density of the TM bundle and the presence of membrane environment, the likelihood of a single independent helical tilt is fairly small. Rather, a correlated tilt of groups of 3 to 4 helices together along an arc seems much more likely. These arc-shaped tilts minimize steric hindrance otherwise introduced by tilting single helices. Indeed, in both bacteriorhodopsin and rhodopsin, groups of 3–4 helices with significantly tilted angles were found while the other helices adopted only slight tilting angles.[4,10]

The tilting of three or four neighboring helices together (though not necessarily at the same angle or in the same direction) is included in the PREDICT modeling procedure. The primary tilting angles for each group of helices is determined according to the angles of the arc that they are forming (see below). Optimal individual tilting angles for each of the seven helices are then determined using an iterative randomized selection of helices and their tilting angles. In each step, the side chains in the new tilted position are minimized using the Metropolis Monte Carlo simulation described above. The tilt refinement procedure, which is comparatively time consuming, is usually performed only on the lowest energy conformations.

The coordinate system used for introducing helical tilts has the membrane normal oriented along the Z-axis (positive Z-axis pointing toward the intracellular side), with its origin corresponding to the middle of the membrane-spanning region of the helix. The tilt of the helix relative to the Z-axis is defined by angle θ and the interhelical-crossing angle is defined by angle φ (Scheme 2). The φ angle was calculated based on the arc formed by the centers of 3 or 4 neighboring helices. For each combination of three out of the four tilted helices, the circumcenter coordinates and the circumradii of the circumscribed circle are calculated by solving the following set of three equations (Scheme 3):



Scheme 2. The coordinate system used for tilting the TM helices, where the helix tilt-angle θ is defined relative to the membrane normal (the Z-axis) and the φ angle is defined on the XY plane.

$$\begin{cases} (X_1 - X_0) + (Y_1 - Y_0) = r^2 \\ (X_2 - X_0) + (Y_2 - Y_0) = r^2 \\ (X_3 - X_0) + (Y_3 - Y_0) = r^2 \end{cases} \tag{19}$$

The φ angle for helix *i* for a combination of three helices is calculated as:

$$\varphi_i = a \tan\left(\frac{dX}{dY}\right) = a \tan\left(\frac{X_i - X_o}{Y_i - Y_o}\right) \tag{20}$$

Where X_i, Y_i are the helices coordinates and X_o, Y_o are the circumcenter coordinates.

For four-helical tilts the φ_i for each helix *i* is calculated as the average of the three-helical φ_j angles corresponding to the three possible combinations of three-helical arcs in which helix *i* participates:

$$\varphi_i = \frac{1}{3} \sum_{j=1}^{3} \varphi_j \tag{21}$$

For tilting the helices the α, β, and γ rotation angles around the X, Y, and Z axes are calculated from θ and φ and the Euler rotation matrix **A** is used to rotate the helix accordingly:

$$\mathbf{A} = \begin{pmatrix} \cos\gamma\cos\beta\cos\alpha - \sin\gamma\sin\alpha & \cos\gamma\cos\beta\sin\alpha + \sin\gamma\cos\alpha & -\cos\gamma\sin\beta \\ -\sin\gamma\cos\beta\cos\alpha - \cos\gamma\sin\alpha & -\sin\gamma\cos\beta\sin\alpha + \cos\gamma\cos\alpha & \sin\gamma\sin\beta \\ \sin\beta\cos\alpha & \sin\beta\sin\alpha & \cos\beta \end{pmatrix} \tag{22}$$

Each φ angle is then optimized by iterative grid search from $\varphi = -10°$ to $\varphi = +10°$ where Monte Carlo optimization of the side chain rotamers is performed at each intermediate conformation.

Intermediate Model Selection

The above process is performed twice, once in the first "coarse" pass of PREDICT and again in the second "fine" application of the algorithm. Each time, once the initial conditions are defined, the above process is essentially automatic resulting in a short-list of the 15 most-stable conformations according to the PREDICT energy score. This list of conformations is further reduced using an automated principle component analysis clustering procedure, ending with <5 possible models (typicallt between 1 and 3 models).

Only when moving from one major modeling step to the next, the method allows for a manual selection of the model(s) that will continue in the modeling process. This occurs at the end of the first coarse application of PREDICT (before moving on to the fine modeling phase) and after the second fine application of PREDICT (before the between the subsequent molecular dynamics refinement). Although this selection is done manually, it is based on relatively objective criteria following a detailed structural analysis of the models and a detailed comparison of the model binding-pocket to experimental data (mutational data and structure activity analysis with various ligands). In principle, this comparison and model selection could have been automated as well. When there is no available data on which to base model selection, such as in the case of orphan receptors, the modeling process would either continue with the most stable conformation or alternatively a small number of alternative models would be generated.

Expansion of the Structures to Full Atomistic Models

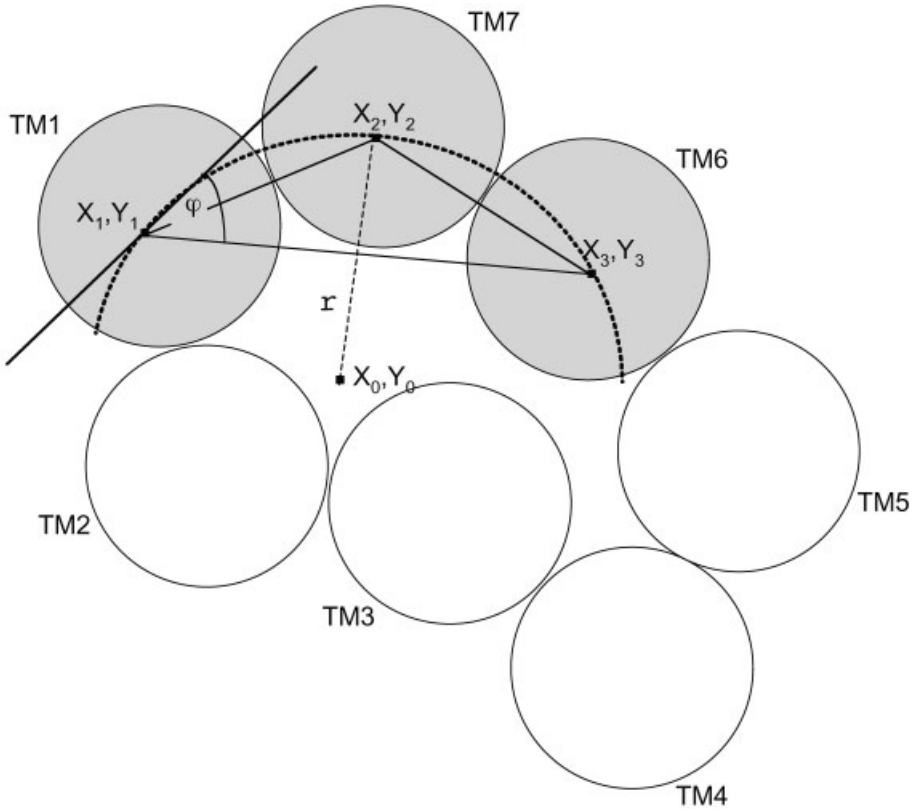
The above procedures determine in an optimized model structure, which is still at the reduced representation level. The next step is to expand this model into a detailed all-atom model. Because the reduced representation model includes the receptor's backbone as well as an optimized set of side chain rotamers (for χ_1 and χ_2), the model's expansion into an all-atom representation is straightforward and can be done using any molecular modeling program, such as the InsightII software.[77]

Structure Refinement

The resulting all-atom model is subsequently subjected to molecular mechanics refinement using the molecular mechanics CHARMm software[78] and the CFF force field[79,80] with a distance-dependent dielectric function, $\epsilon(r)$. The refinement process includes an initial minimization followed by a molecular dynamics simulation of the free receptor and ends in the generation of a virtual protein-ligand complex model. This final ligand-bound receptor model is the structure used for in silico screening.

Initial minimization

A two-step minimization protocol is performed on the initial all-atom model. During the first minimization step,



Scheme 3. Definition of an arc connecting the centers of three neighboring helices that is used in the calculation of helix tilts. The arc is part of the circumscribed circle of the triangle that is defined by the three helical centers.

the backbone is kept fixed and only the side chains are minimized stopping after 5,000 steepest descent steps and 2,000 steps of conjugate gradient, or until the rmsd gradient of the potential energy reaches a desired level. In a second minimization step, the backbone constraints are removed and all atoms, backbone, and side chains alike are minimized using 500 steepest descent steps.

Molecular dynamic refinements and introduction of helical kinks

The minimized model is then subjected to a molecular dynamics simulation. The simulation is designed to: (1) relax unfavorable local side chain interactions that may persist after the minimization step, and (2) enable the protein helices to adopt proline- or glycine-induced kinks if appropriate. As we will discuss, this last issue is not trivial since not all proline residues in TM helices induce kinks. Proline-induced kinks, if present, are very important protein structure and may function as molecular switches in cell membrane signaling.[81] Molecular dynamics simulations by Visiers et al.[82] were found useful for understanding the structural biology of these conformational switches.

The following simulation setup was used to enable the molecular dynamics process to introduce Pro-kinks into the model. Soft harmonic constraints were imposed on the φ, ψ, and ω backbone dihedral angles of all amino acids except for Pro residues and for the two residues flanking the Pro residue on both sides. Depending on the size of the binding site a 12–15Å water sphere, containing explicit TIP3P water molecules, was added onto the binding cavity subject to a stochastic solvent boundary.[39,78] Water molecules that overlap the protein within 2.8 Å were removed and the remaining water molecules were relaxed using 4 picoseconds (ps) of Langevin dynamics at 300K (fixed protein). The protein simulation starts with a 20-ps heating period from 0 to 300K subject to the above constraints. Over a period of 60 ps at 300K, the φ, ψ, and ω harmonic dihedral angle constraints are gradually reduced. Finally the simulation continues for an additional 200 ps at 300K with only minimal constraints applied to these dihedral angles (harmonic force constant = 30). To ensure that the TM helices, which are truncated at the membrane surface, do not "unwind" during the simulation, simple harmonic distance constraints are applied throughout the simulation to mimic the alpha-helical i, $i+4$ carbonyl-amine hydrogen bonds near the extra- and intra- cellular sides of the helices. The various constraints and the relatively short length of the simulation (280 ps) permit the model to undergo local refinements and introduce kinks, but at the same time ensures that the final structure solved by PREDICT is not dramatically changed by the dynamics simulation. These constraints are to accurate and efficient structure determination in the absence of both an explicit

membrane environment[83,84] and the extra- and intracellular parts of the receptor.

Creating virtual receptor-ligand complexes

Because all high-throughput docking procedures must use a rigid protein structure, in silico screening gives the best results when using a ligand-bound conformation of the target protein. To generate the receptor-ligand complex that will be used for in silico screening, a potent small-molecule ligand is placed inside the binding pocket in an initial orientation that agrees with proposed key interaction in the site. Clearly, any such initial placement suffers from a strong bias, which we address by subjecting the virtual complex to a free-energy-perturbation-like simulation designed to allow the ligand to reorient itself in the site in response to specific ligand-protein interactions. In this simulation, the interaction between the ligand and protein, $V_{ProtLig}$, is weighted by a λ factor of varying value,

$$V(\lambda) = V_{Prot} + V_{Lig} + \lambda V_{ProtLig} \tag{23}$$

During the course of the simulation, λ is gradually scaled up from 0 (no interaction) to 1 (full interaction) allowing the ligand and the protein's side chains gradually to adjust to the presence of each other. The dynamics protocol starts with a 10-ps heat-up period from 0 to 300K at a very small λ value (λ = 0.05). This is followed by five consecutive 20-ps steps with increased λ values (λ = 0.05, 0.2, 0.3, 0.4, and 0.5). In each 20-ps λ-step, the receptor-ligand complex is subjected to a simulated-annealing protocol comprised of five "sawtooth" annealing cycles, where in each cycle the temperature is linearly reduced from 1,000 to 300K over a 4-ps time period (total of 20 ps per λ-step), This is followed by three steps of 10-ps equilibration simulation at 300K with higher λ values (λ = 0.6, 0.7, and 1.0). To prevent the ligand from drifting out of the binding site when the value of λ is small, we introduce a weak constraint tethering the ligand to the site. This constraint is defined as a weak NOE constraint between one of the ligand atoms representing the "center of the ligand" and the "center of the binding site." This constrain is defined as:

$$E(R) = \begin{cases} 0 & R \le R_{cutoff} \\ 0.5 \cdot K_{cutoff} \cdot (R - R_{cutoff})^2 & R > R_{cutoff} \end{cases} \tag{24}$$

where R is the distance between the "center of the ligand" and the "center of the binding site,", K_{cutoff} is the force constant and R_{cutoff} is the cutoff distance. In the current procedure, K_{cutoff} was set to 1 kcal/mol and R_{cutoff} was 3 Å. Harmonic constraints are also imposed on the helical backbone to maintain their shape. At the end of the simulation, the ligand is removed from the complex leaving a well-defined binding pocket.

Algorithm performance

The full two-stage PREDICT algorithm is efficient, consuming approximately 3 days on a 60 PentiumIII processor Linux cluster, with the first "coarse modeling" step taking one day and the second "fine modeling" step taking two days. The molecular dynamics refinement and virtual complexation simulations take approximately 2 weeks to complete on an SGI Octane computer.

In Silico Screening Against the Receptor Model

The optimized model, in its ligand-bound conformation, can now be used for in silico screening aiming to identify potent new drug candidates from large compound databases. Prior to this step, the accuracy and reliability of the protein model should be estimated by taking a random drug-like library to which a small number of known ligands were added and screening it in silico against the receptor model. Success in such a screening test suggests that that model can be regarded as reliable at least in the context of drug discovery. The details of the screening and scoring processes used in this study are reported elsewhere (Marantz et al., forthcoming) hence only the outline of these processes will be given here.

Compound libraries

A set of known small-molecule ligands with Ki values better than 1 μM was identified for each of the modeled GPCR receptors. Very similar compounds were removed from this set to maximize diversity. 2D structures were expanded to 3D using the CONCORD software.[85] Gasteiger atom types and protonation state were determined using a SLN (Sybyl Line Notation) script[86] and the library was stored in a multi-mol2 format. Depending on the receptor, between 18 and 47 ligands were included in the libraries of known-ligands.

The random compounds screening libraries were extracted from the Advanced Chemical Directory (ACD v.2001-1; more than 1,000,000 compounds)[87] and the Tripos LeadQuest library (~80,000 compounds).[88] To ensure that only relatively drug-like compounds are included in the screening libraries, the following criteria were applied: (1) Molecular weight within the 200 to 550 range, (2) compounds do not contain Na, Ca, Mg, Hg, Cu, B, Li, or Fe atoms, (3) the number of S and P atoms in the molecule is no more than 3 for each of them. These compounds were also required to fulfill the Lipinsky rules for ClogP, number of donors, and number of acceptors,[89] although these too do not guarantee drug-likeness. A simple algorithm was used to obtain maximum diversity and a wide distribution of MW in the library. Two such Random Screening Libraries were used:

1. Random Library 1 included 10,000 drug-like compounds selected according to the above criteria. The average properties of this library are specified in Table I.As can be seen, the library is well within the Lipinsky definitions of drug-like compounds. Furthermore, the properties of this library are similar to those of known dopamine D2 receptor ligands.
2. Random Library 2 included 6,200 drug-like compounds selected to have average properties similar to those of ligands that bind to the peptide receptors such as the NK1 receptor. These compounds were selected with the following set of criteria: molecular weight between 400 and 650, number of donors <5, number of acceptors

TABLE I. Properties of Random Library 1 and of Known Active Compounds for the D2 Receptor

Library	N	MW	ClogP	Ndonors	Nacceptors
Random Library 1	10,000	330 ± 80	3.7 ± 2.1	0.9 ± 0.9	2.8 ± 1.4
D2 ligands[a]	43	361 ± 114	4.0 ± 1.5	1.8 ± 0.8	2.0 ± 1.5

[a]Details in Table III.

TABLE II. Properties of Random Library 2 and of Known Active Compounds for the NK1 Receptor

Library	N	MW	ClogP	Ndonors	Nacceptors
Random Lib 2	6,200	442 ± 39	5.0 ± 1.9	1.1 ± 1.0	3.6 ± 1.3
NK1 ligands[a]	26	520 ± 78	5.1 ± 1.2	1.8 ± 1.3	4.9 ± 3.2

[a]Details in Table VI.

between 1 and 12 and ClogP <6.5. The average properties of this random library, which are given in Table II, compare well with the average properties of known NK1 ligands.

For each library, the 2D structures were converted to a multi-mol2 format using the protocol described above. Screening libraries were obtained by adding one of the known-ligand libraries to one of the random libraries.

Screening protocol

Screening of the compound library against the receptor models was performed with the DOCK4.0 software[90] and a special screening and scoring protocol described elsewhere (Marantz et al., forthcoming). In short, the SPHGEN software[90] was used to generate the site filling DOCK spheres. Since DOCK does not account for the flexibility of ring scaffolds, we use the HYPERION program (Shacham et al., unpublished data) to identify multiple scaffolds in a single compound and break it down into separate fragments (scaffolds, substituents, and linkers). CONFORT[91] is used for adding conformations for each scaffold and them HYPERION invoked to rebuild the compound from its fragments. This procedure results in multiple copies of every compound, reflecting the different 3D conformations associated with its various scaffolds. The compounds are then docked into the receptor and scores using DOCK4.0.[90] To ensure convergence of the scores, each compound realization (with one of the specific anchor conformation) is docked 8 times using different random number seeds. The best (lowest) Dock-score was used to compare the different ligands. At the end of the docking process, the compounds were sorted based on their best docking score.

Evaluation of virtual screening

Performance evaluation for the virtual screening is done by the following two measures: "enrichment factor," which is related to the common measure of "hit rate," and "coverage."

1. Enrichment factor

An enrichment factor measures the success of virtual screening relative to random screening (which is equivalent to experimental HTS) by comparing the virtual screening hit rate with the hit rate of random screening. A hit-rate is the fraction of active compounds found in the Screening Library by whatever method. The baseline hit-rate H_{rnd} is that associated with random screening (i.e., HTS) which is defined as

$$H_{rnd} = M/N \tag{25}$$

where M is the number of active compounds in the N compound screening-library. In our case, M is the number of known ligands seeded into the N compound library of random drug-like compounds described above (e.g., N = 10,000). Unlike random screening, virtual screening ranks the compounds in the screening library according to some score. The hit-rate associated with virtual screening $H_{vs}^{p}$is thus defined as

$$H_{vs}^{p} = M^{p}/N' \tag{26}$$

where M^{p} is the number of active compounds (i.e., known ligands), equivalent to the first p percent of all active compounds, found within the top N' ranked compounds.

The enrichment factor, which measures the relative success of virtual screening over random screening, is defined as ratio between the virtual screening hit-rate H_{vs}^{p} and the baseline random hit-rate H_{rnd},

$$Enrichment^{p} = H_{vs}^{p}/H_{rnd} \tag{27}$$

For example, for M = 20 known binders seeded into a random library of N = 10,000 compounds the H_{rnd} is 0.002. If following virtual screening, 50% of these 20 compounds are ranked among the top 200 compounds, the hit rate $H_{vs}^{50\%}$ is 0.05 and the enrichment factor $Enrichment^{50\%}$ is 25; i.e., the virtual screening (at the 50% point) is 25-fold more efficient than random screening. The enrichment factor usually deteriorates as the percent p of known actives found approaches 100%, indicating that some compounds are missed. The maximal theoretical enrichment factor is reached when H_{vs} is 1 (i.e., the M top ranked compounds are all active compounds, i.e. $M^{p} = N^{p} = M$) and, therefore, the maximal theoretical enrichment factor equals $1/H_{rnd}$.

2. Coverage

Coverage is defined as the number of active compounds M^q that are included within the top q percent of the ranked library vs. the total number of active compounds in that library M:

$$Coverage^q = M^q/M \quad (28)$$

This measure highlights how many of the known actives are found within a small portion q of the ranked library. The complementary measure, 1–$Coverage^q$, reflects the number of known actives that are missed. A $Coverage^{10\%}$ value of 85% means that 85% of the known active ligands are found within the top 10% of the ranked library after virtual screening. The coverage is directly related to the application of virtual screening for drug discovery. If after docking most of the known ligands are concentrated within a small fraction of the original library, it means that when screening for new compounds the actual in vitro assays can be performed on this small fraction of the library rather than on the full compound dataset, with the obvious speed-up and efficiency benefits.

Validation of the PREDICT Algorithm: Bovine Rhodopsin

The structure prediction algorithm was first validated against the bovine rhodopsin structure, which is currently the only GPCR for which the crystal structure is available (at a resolution of 2.8 Å).[3] The PREDICT algorithm was used to create an all-atom model of bovine rhodopsin that was further optimized using 300 ps of molecular dynamics simulation without the retinal. The results of this model, which was in good agreement with the overall structure of the TM domains in the rhodopsin crystal structure (Cα-rmsd 3.2Å, excluding TM4) was reported previously.[31,32] The 11-cis retinal was then attached to the side chain of Lys296. The conformation of the protonation state of the reinal was adopted from the 1f88 rhodopsin X-ray structure.[10] To refine the conformation of the retinal-bound rhodopsin model, the optimization process describe in Materials and Methods: Structure Refinement was used with a harmonic distance constraint between the CD1 atom of Phe212(5.46) and C4 atom of the β-ionone was used (harmonic force constant = 3). A moderate rigid body rotation of the helices (perpendicular to the membrane plan) was used to optimize the amino acids position relative to the membrane and the helical bundle. The resulting PREDICT rhodopsin model was in good agreement with both the overall structure of the TM domains in the rhodopsin crystal structure as well as the fine details of the binding crevice. The Cα-rmsd between the retinal-bound PREDICT rhodopsin model and the TM region of rhodopsin crystal structure (PDB: 1F88) was 2.9 Å [Fig. (2a)]. The heavy-atom rmsd of 11-cis retinal between the model and the crystal structure was 0.9 Å. The location and side chain orientation of most key residues known to be involved in retinal binding are nicely reproduced in the model, adopting a conformation similar to that in the X-ray structure [Fig. 2(b)]. These residues include Glu113, Leu125, Met207, Phe208, His211, Phe212, and Lys296.[92] Some differences were observed with the side chains orientation of the two aromatic residues Trp265 and Tyr268 in TM6 that were found to play a role in retinal binding. These residues are pointing into the binding site in the X-ray structure where their aromatic rings can create π-π interaction. In our rhodopsin model, Trp265 face the binding site; however, the side chain of Tyr269 is direct more towards TM7. This can be partially explained by the two compensatory π-π interactions that are seen in the model, of Trp265 with Phe261 and Tyr269 with Phe293 (TM7). The overall Cα rmsd value is comparable with the rmsd obtained by Vaidehi et al.[22] even though their rhodopsin model, unlike PREDICT, was specifically based on the X-ray rhodopsin structure to which it was later compared. PREDICT reached similar results without relying on the rhodopsin X-ray structure in any way.

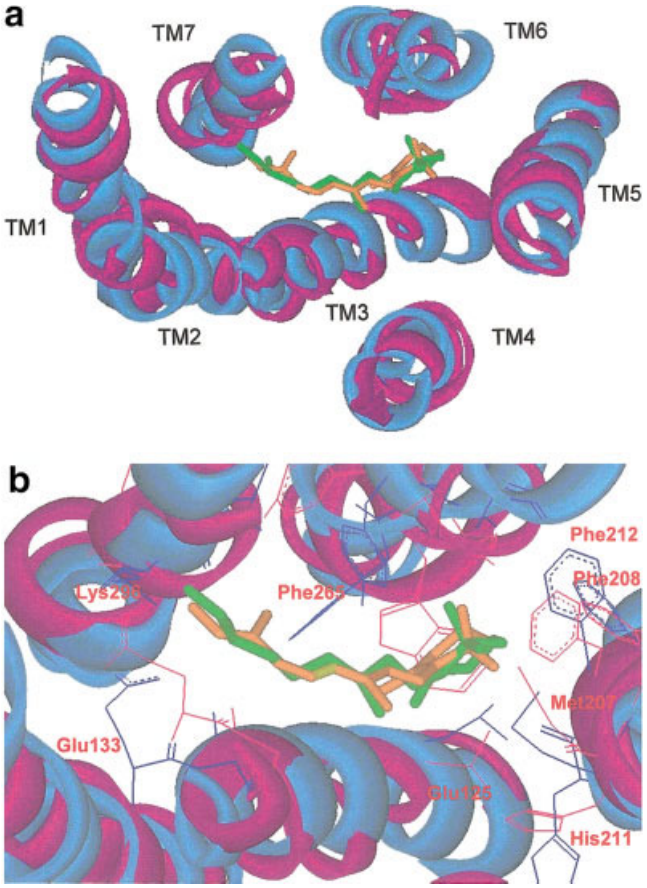


Fig. 2. **a:** The PREDICT model of bovine rhodopsin (purple) overlaid on the 2.8-Å resolution X-ray structure of bovine rhodopsin (PDB: 1F88, blue)[3] viewed from the extracellular side. **b:** A focus on the retinal binding-site, showing the key retinal-binding residues.

Pro and Gly kinks

As discussed above, TM helices often contain proline residues. A Pro residue within the TM domain disrupts the hydrogen bond network that stabilizes the helical structure and often introduces a kink that effectively breaks the helix into two segments. Because the kink region lacks a

hydrogen bond, it no longer exhibits the ideal α-helical 3.6 amino-acid periodicity. Instead, the kink is often accompanied by a twist in the helix orientation before or after the Pro position. However, measurement of the degree of twists associated with helical kinks was shown to be highly variable with a wide energy minimum.[93] In the rhodopsin structure, TMs 1, 4, and 6 are kinked, while the presence of Pro residues in TMs 5 and 7 led to unusual kinked conformations.[3] Both the degree of the helical kinks and the twist angles were successfully reproduced in our bovine rhodopsin PREDICT model. Gly residues located in a TM helix can cause a similar kink effect because the absent side chain increases backbone flexibility, which in turn may lead to disruptions in the helix' hydrogen bond network within the helix.[94] Both types of kinks, which are observed in the X-ray structure, were successfully reproduced in the PREDICT model, e.g., the Pro(1.48) induced inward-bend in TM1 of the rhodopsin crystal structure.[3] PREDICT was also successful in introducing kinks induced by Thr and Gly residues, For example, most GPCRs possess a conserved Pro(2.58) residue at TM2. Although this Pro does not exist in the crystal structure of rhodopsin, TM2 is still substantially bent. This bend is induced by the Thr-Gly motif (GGxTT) that is found in TM2. It was shown that Thr can induce a bend in an ideal alpha helix by creating hydrogen bonds to the carbonyl backbone of the proceeding turn.[95] This bend in TM2 appears also in the PREDICT model of rhodopsin and is very similar in magnitude and orientation to the one observed in the crystal structure, indicating that Thr and Gly kinks can be reproduced by the model.

RESULTS

In this study, we describe a new modeling algorithm (PREDICT) and report results of structural modeling and in silico screening results for three G-protein coupled receptors: Dopamine D2 (D2), neuropeptide Y Y1 (NPY Y1), and neurokinin NK1. These receptors represent two main groups of Class A GPCRs: biogenic amine receptors (D2) and peptide receptors (NK1, NPY Y1). A brief discussion is also included on the PREDICT model of the chemokine receptor CCR3. For each of these systems, we will first discuss the details of the 3D models, focusing on their ligand binding sites, followed by a report of the in silico screening results obtained using these receptor models.

Modeling the Human Dopamine D2 Receptor

Dopamine is the predominant catecholamine neurotransmitter in mammalian brains where it controls a variety of functions including locomotor activity, cognition, emotion, positive reinforcement, food intake, and endocrine regulation. Human dopamine D2 receptor contains 414(short)/443(long) amino acids. It is located in Caudate/putamen, nucleus accumbens, olfactory tubercle, and cerebral cortex and is involved in motor function and behavior (brain), control of prolactin γMSH secretion from pituitary and cardiovascular function.[19]

Coarse PREDICT modeling

The first step in the PREDICT modeling procedure is to generate multiple closed-packed 7TM conformations in 2D. Two simple constraints were used as secondary filters to limit the number of two-dimensional conformations for D2, leading to the generation of 462 out of the 10^7 possible conformations. These constraints were: (1) Class A GPCR conserved residues ASP (2.50) at TM2 and Asn (7.49) at TM7 are known to be in spatial proximity allowing them to form a network of hydrogen bonds. This spatial proximity was found in several other receptors, including HT2A, GnRH and μ opioid receptors.[46,96–98] In the dopamine D2 receptor, these two residues correspond to Asp80 and Asn422, respectively. This information led to the application of a secondary filter constraining the distance between TM2 and TM7 to no more than 17 Å. (2) Three conserved residues, Asp114 (3.33) in TM3, Ser194(5.43), and Ser197(5.46) both in the TM5,[19,99–101] are implicated by site-directed mutagenesis experiments as being involved in ligand binding. This information led us to apply an additional secondary filter constraining the distance between TM3 and TM5 to no more than 17 Å. Application of this filter reduced the total number of selected 2D conformations from 1,212 (without the two filters) to 462 conformations.

The rotational orientation of the helices in each of these 462 2D conformations was then optimized as described above. Figure 3 shows six representative 2D packing arrangements of the D2 receptor TM helices. After optimization, the first two conformations (47, 52) yielded low-energy structures. The next two conformations (4, 201) led to medium energy structures and the bottom two conformations (9, 329) turned out to be associated with unstable non-physical structures. A substituted-cysteine accessibility method (SCAM) study that was performed in order to determine the D2 receptor residues that face the binding crevice found that three highly conserved aromatic residues (Trp6.48, Phe6.51, Phe6.52) and a histidine in TM6 (His6.60) face the water-accessible surface of the helix and participate in ligand binding.[102] As seen in Figure 3, the rotation optimization procedure used by the algorithm takes into account the presence of aromatic clusters and positions these residues at the interior of the helical bundle facing the binding crevice or facing other helices for possible intrahelical interactions. Regarding the conserved Asp-Asn residue pair in TM2 and TM7, while Asn422(7.49) is positioned on the intramolecular water-accessible area of TM7, Asp80(2.50) was oriented in all conformations pointing towards TM3 (Fig. 3). Such an orientation of the two residues does not allow for direct hydrogen bonding interaction to be present in the final all-atom model. Rather, this conformation leads to a more complicated hydrogen network, which is in agreement with experimental data that will be discussed later.

In the next step of the modeling procedure, the 462 2D rotationally optimized conformations were expanded into the 462 corresponding reduced-representation initial 3D protein models. As described, each amino acid is represented in this model by two to four spheres and a side

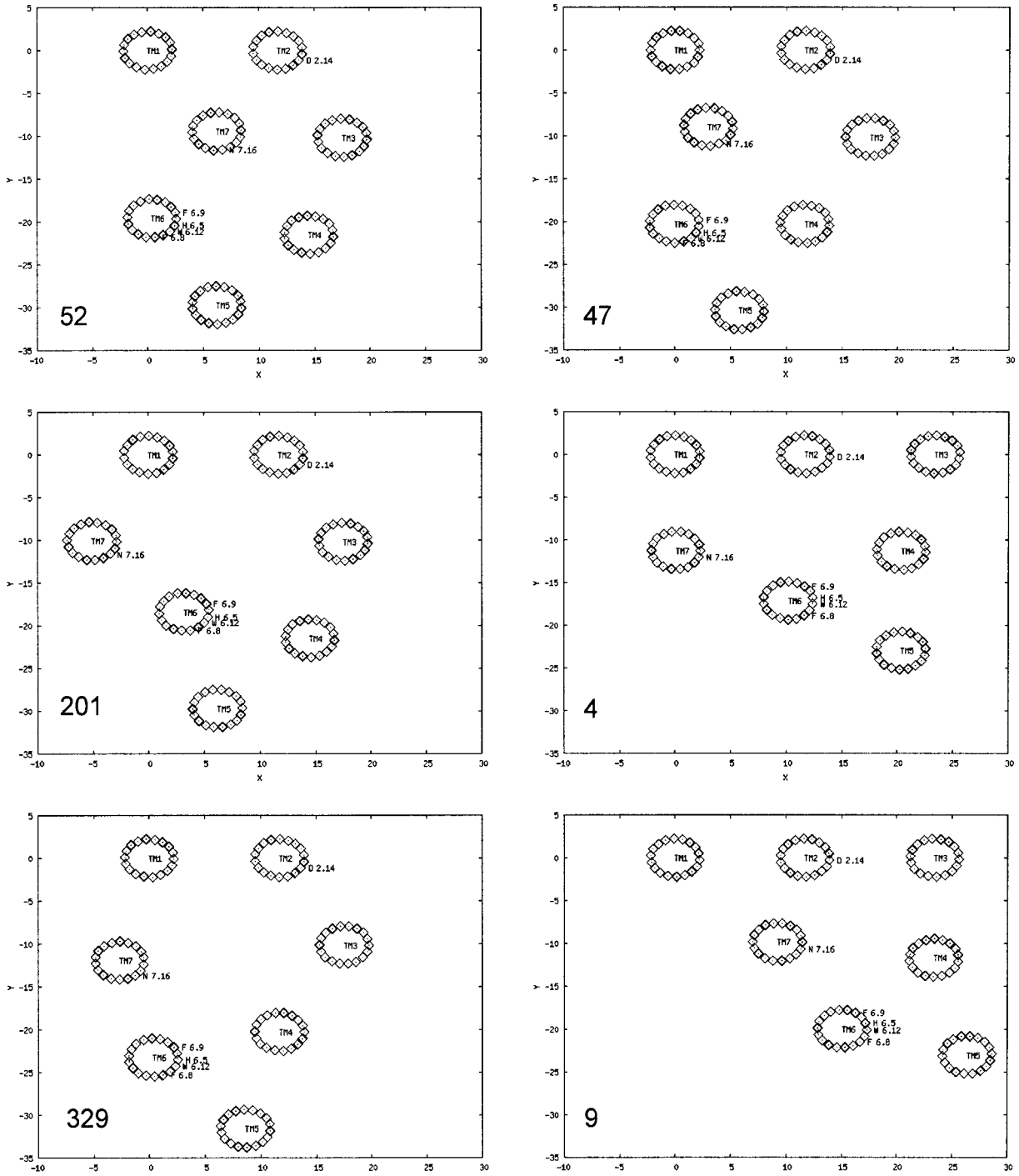


Fig. 3. A representative set of six 2-dimensional initial packing configurations (out of the 462 PREDICT "decoys") for the D2 receptor leading to 3D models of different stability (conformation numbers are indicated). Conformations 52 and 47 are among the initial conformations that resulted in stable low-energy 3D structures; the two conformations reflect a similar arrangement of the helices (conformation 52 was later used as the template for the second phase of the modeling). Conformations 4 and 201 represent conformations that resulted in structures with mid-level energy scores. Conformations 9 and 329 resulted in unstable 3D structures.

chain rotameric conformation, which is determined by the dedicated Monte Carlo rotamer optimizer. Each initial 3D model (decoy) is then refined through a series of optimization procedures that include vertical alignments of the helices relative to each other and to the membrane surface, optimization of the distribution of polar and nonpolar residues of the amphiphatic helices, as well as small adjustments of the position of helical centers. The optimization procedure is driven by the PREDICT energy function of Eq. 16, yielding an energy score for each of the 462

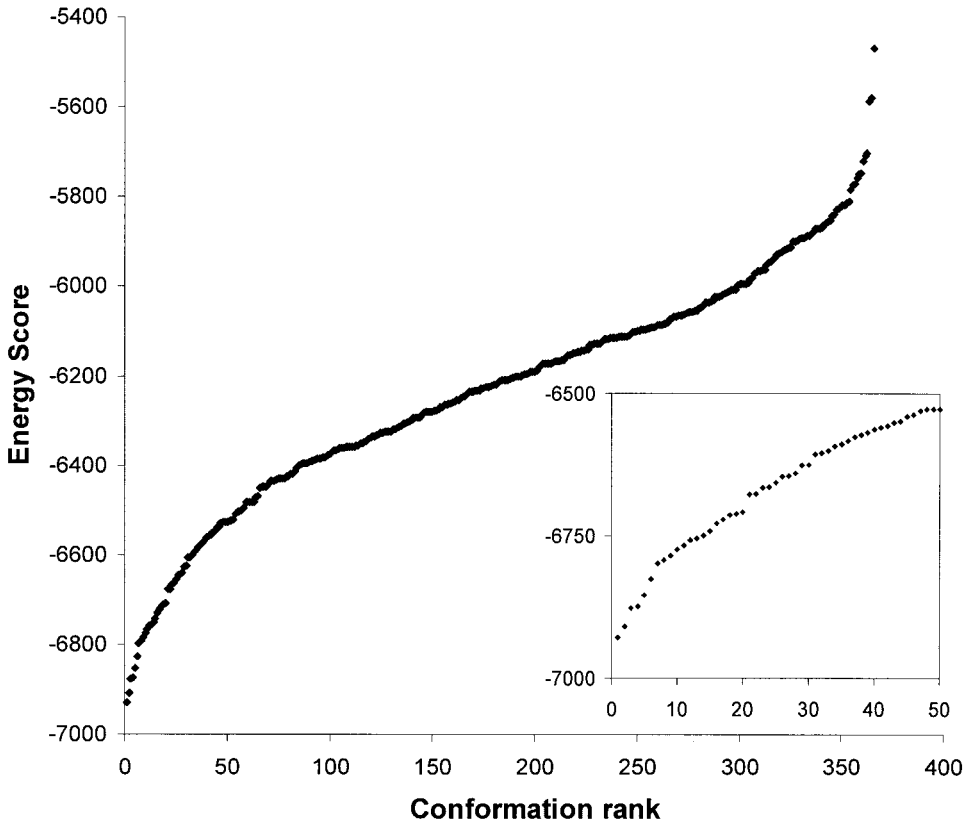


Fig. 4. An energy score profile calculated by PREDICT for a diverse set of possible initial 3D structural models ("decoys") for the D2 receptor. **Inset:** Enlargement of the 50 best scoring decoys. A small number of conformations are clearly identified as superior (lower energy scores) to the other decoys.

optimized structures. The energy-score profile of the resulting 3D structures (Fig. 4) demonstrates that only a few of the optimized 3D structures have significantly better energy scores compared to the optimized versions of all others possible structures (the decoys). This is confirmed by a Z-score functional analysis of the energy distribution that measures the rendering of the raw scores in mean units of standard deviation of the lowest energy scores.[103,104] The average Z-score for the best 15 structure was -2.3, with the Z-score of the best structure at -2.6, indicating that the PREDICT energy function of Eq. 16 is successful in discriminating between low-energy structures and the rest of the conformational decoys.

Naturally, one would expect to find similar structures among the top scoring structures. To analyze the similarity among these structures, we calculate the pairwise rms distances between the best 15 scored conformations. Principal component analysis (PCA) was applied to the rmsd matrix[105] leading to the projection on the first and second principal eigenvectors that is shown in Figure 5. The four clusters shown on the projection map indicate groups of similar structure.

To select the best "coarse" structures that will serve as "lead models" in the second application of PREDICT, the putative binding sites in the 15 lowest energy structures were studied. In particular, the size of the putative binding site and the position of key residues known to participate in ligand binding were examined in each of these models. For example, in model-52 (i.e., the optimized model stemming from conformation 52), which is the lowest energy model in its cluster (Fig. 5), one can see a putative binding pocket between TM2, 3, 4, 5, and 6. Furthermore, even at in this early stage of the modeling process, one could see that this pocket was large enough to incorporate small molecules. In addition, in model-52 the Cα atoms of Asp114(3.32) and Ser194(5.43) were positioned, as expected, facing the binding site. Residue Ser197(5.46), on the other hand, was pointing in this initial model more towards the protein interior mainly because of the ~100° angle between the two Ser residues. This is to be expected in an initial model, but as will be discussed later, small change to the helical bundle will enable the side chain rotamer of Ser197(5.46) to reach the binding pocket in the final high-resolution model. Residue Phe389(6.51), which is known to be involved in binding,[102] is also located in the model's binding pocket.

A similar analysis was performed for the lower energy model in each of the four clusters identified in Figure 5. It was found that the locations of key interacting residues in model-52 were significantly more concurred with published data than in the other three cluster-representing models (models 102, 167, and 5, respectively). Therefore, only model-52 was selected as a "lead model" for the second round of PREDICT modeling, in which a higher resolution

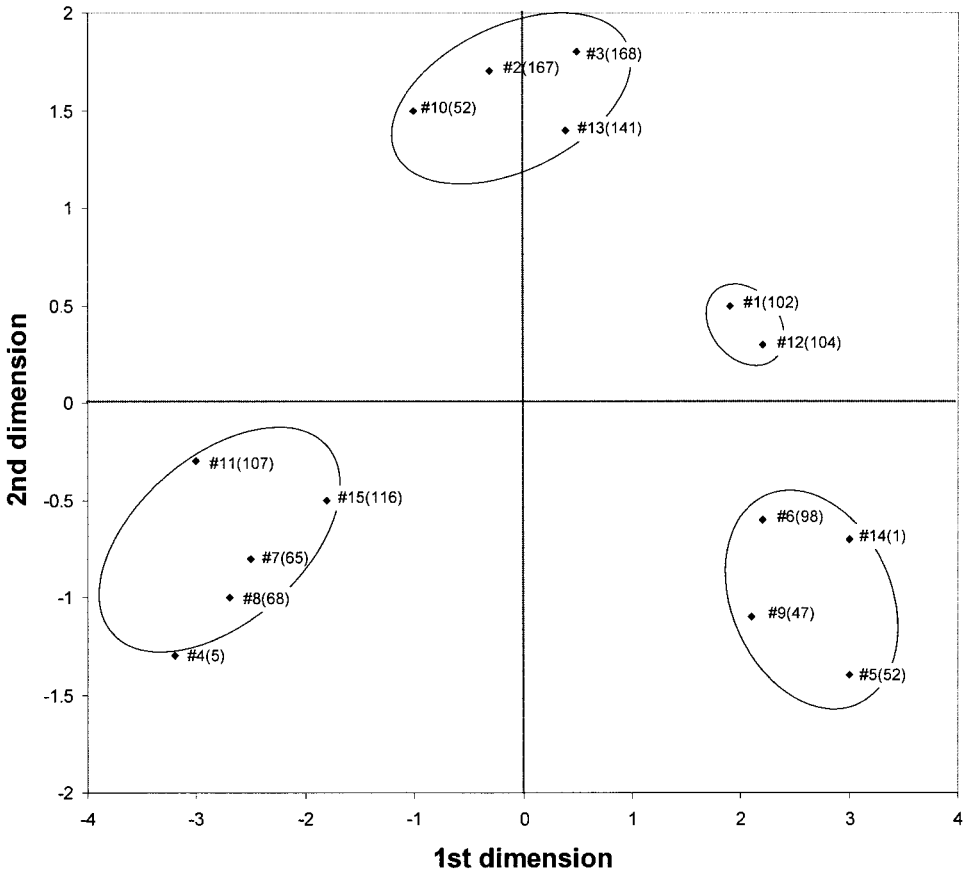


Fig. 5. A projection of the 15 best optimized 3D structures, generated for the D2 receptor in the initial coarse modeling step, on a two-dimensional principal coordinate map. Each conformation is identified by its relative rank among the "best" 15 structures. The conformation numbers are indicated in parentheses.

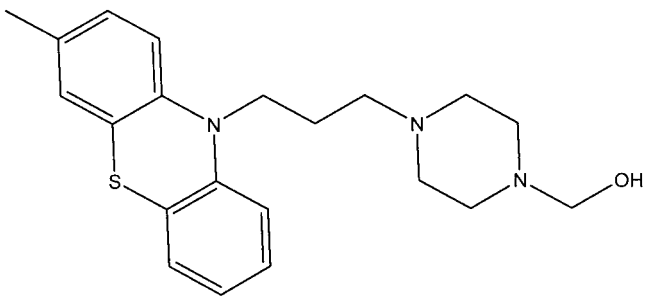
grid will be used to generate many conformations in the vicinity of this coarse "lead model."

Fine PREDICT modeling

In the second phase of the PREDICT modeling procedure, 2,251 2D conformations were generated around the structure of lead-model-52. These 2,251 2D conformations were expanded to reduced representation 3D structures following which they were subjected to the PREDICT multilevel optimization procedure. The PREDICT tilting algorithm was then applied to the 15 structures with lowest PREDICT energy score. The rmsd matrix between the structures was calculated and a clustering analysis, similar to that used in the first phase, was performed. Two of these models (model-695 and model-2186) were selected as the final PREDICT models based on the PCA clustering results, the energy scores, the size of the binding pockets, and the agreement with experimental data.

Model refinement and the virtual protein-ligand complex

The two final reduced representation models were expanded into the corresponding all-atom models and subjected to a minimization protocol. After minimization, it was found that the binding pocket in the minimized model-2186 was too small for ligand binding, leaving model-695 as the final PREDICT human dopamine D2 model. This model was then subjected to the molecular dynamics protocol discussed above to refine side chain orientation and to introduce helical kinks where necessary. A virtual receptor-ligand complex was created with fluphenazine (Scheme 4), which is a potent D2 antagonist with Ki = 0.32–4 nM. This compound was selected for the virtual complex because it included several key moieties: a tri-cyclic moiety, a piperazine, and a hydroxyl. Experimental information on the interactions of fluphenazine and the D2 receptor were used to guide the initial positioning of



Scheme 4. Fluphenazine, a dopamine D2 receptor antagonist.

the compound in the binding site. The initial conformation included interaction between the Asp114(3.33) residue and the N4 of the compound's piperizine ring and an aromatic interaction between Phe389(6.51) and the one of aromatic rings of the phenotiazin scaffold of fluphenazine. This initial D2–fluphenazine virtual complex was then refined by the virtual complexation simulations protocol described in Materials and Methods. The weak NOE constraint, which prevents the compound from leaving the site when the value of λ is small, was placed between the second carbon of the propyl chain in fluphenazine and the geometrical center of the binding pocket (calculated by CHARMm). The fluphenazine ligand was then removed from the model leaving a set of 3D coordinates that are ready for virtual screening. The rmsd between the resulting ligand-bound structure and the model structure before the molecular dynamics refinement was 2.77Å. The rmsd to the TM domains of the X-ray bovine rhodopsin was 6.3Å, this is mainly due to a different tilt of TM3. Close inspection showed that the overall conformation of the model is quite similar to the rhodopsin X-ray structure, except for TM3, which is much more tilted in the rhodopsin structure. This has a major contribution to the relatively large rmsd between the two structures. The side chain orientation of residues that are known to play a role in ligands binding to the D2 receptor and participate in retinal contact in the corresponding position in the rhodopsin structure are generally in good agreement (residues Ser193(5.42), Ser197(5.46), Phe198(5.47), Trp386(6.48), Phe389(6.51), and His394(6.55) in the dopamine D2 structure and their corresponding residues Met207, His211, Phe212, Trp265, Tyr268, and Ala272 in the rhodopsin structure.

Analysis of the Dopamine D2 Receptor Model

The following analysis of the final PREDICT human dopamine D2 receptor model will initially focus on the characteristics of the binding site and then proceed to discuss the structural features of the whole TM domain.

Binding site characterization

The binding site cavity in the final D2 receptor model is located within the TM domain, between TMs 2, 3, 5, and 6, and is faced by the highly conserved dopamine D2 receptor residues, Asp114(3.33), Ser194(5.43), Ser197(5.46), and Phe389(6.51) (Fig. 6). As can be seen from Figure 6(c), the binding pocket is localized within the extracellular one-third of the TM domain. Site-directed mutagenesis and affinity-labeling studies of the D2 receptor have shown that D2 agonists bind differently than D2 antagonists.[92] In the model, both binding sites are part of the same binding cavity and will be discussed here.

Antagonists binding site

Site-directed mutagenesis and affinity-labeling studies of the D2 receptor, recently reviewed in Shi and Javitch,[92] show that antagonists bind between TM 2, 3, 4, 5, 6, and 7. Key residues, which were experimentally discovered to affect antagonist binding, are as follows. TM2: Val91(2.61); TM3: the conserved Asp114(3.32) and Cys118(3.37), which is located one helical turn downstream from it. Cys118 is part of the binding crevice although it is not critical for the binding itself[106] TM5: Phe189(5.38), Phe198(5.47) and Ser197(5.46); TM6: Trp386(6.48), Phe389(6.51), and His394(6.55). In particular, residue Phe389(6.51) is highly conserved in all catecholamine neurotransmitters receptors. In addition to the polar Asp in TM3 and one of the three Ser residues in TM5, the D2 antagonist binding-pocket contains many aromatic residues forming the aromatic cluster essential for binding of D2 antagonists.[19,99,102]

As can be seen in Figure 6(b), all of the key residues map well in our D2 model and form an elongated antagonist binding-site. As expected, this site is characterized by a very large aromatic cluster composed of Cys118(3.37), Phe198(5.47), Trp386(6.48), Phe389(6.51), and His394(6.55). According to the model, the residue Tyr426(7.53) is also an integral part of this cluster. The model also shows that residue Cys118 points to the interior of the protein structure and becomes an integrated part of the hydrophobic/aromatic antagonist binding site, in agreement with experimental findings. Residue Val91(2.61), which was experimentally found to affect antagonist binding,[107] lines the outer region of this hydrophobic/aromatic site and partially faces TM3. The model suggests that the adjacent residue Val92(2.62) may have similar importance. Residue Phe189(5.38), which also affects antagonists binding, is approximately two helical turns above the main cluster and faces the binding cervix in the model. In addition to this cluster, the antagonist binding-site also includes residue Asp114(3.32), which is located in its midst and is responsible for anchoring the antagonists' amine moieties. Residue Ser197(5.46) is located at the far end of the binding-site and can interact with polar moieties that are sometimes present in these antagonists.

Figure 7(a) shows the docked conformation of D2 antagonist fluphenazine (Scheme 4) within the D2 binding pocket highlighting the specific interactions between it and key residues. As expected, a primary interaction exists between one of the piperazine nitrogen atoms of fluphenazine and Asp114 in TM3 (2.7 Å distance). Although Ser197 faces the binding pocket in the model, the docked conformation of fluphenazine is such that its hydroxyl group is a bit too distant for hydrogen bonding with Ser197 (5 Å distance). Nonetheless, given the conformational flexibility of this part of fluphenazine, it is possible for this compound to adopt a conformation in which the hydroxyl moiety will interact with Ser197. The tri-cyclic moiety of fluphenazine is nestled within the aromatic pocket near TM6 and TM7. According to our D2 model, residue Phe389 plays a formative role in this binding site by defining a hydrophobic-aromatic region into which the characteristic tri-cyclic moiety of fluphenazine docks.

The same binding mode is also observed for other tri-cyclic amine D2 antagonists, such as chlorpromazine (Scheme 5) as shown in Figure 7(b). Similar to fluphenazine binding, the tri-cyclic moiety chlorpromazine inter-

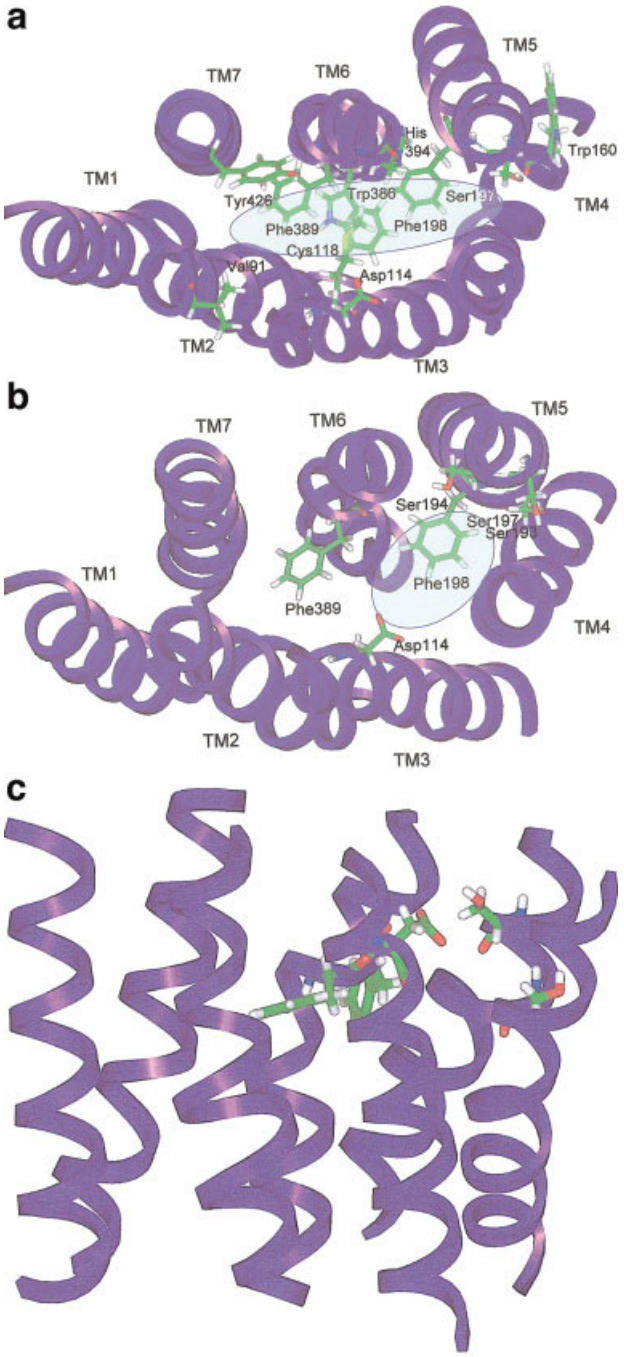


Fig. 6. The PREDICT model of the dopamine D2 receptor highlighting the respective antagonist and agonist binding sites. The D2 antagonist binding-site (**a**) and the D2 agonist binding-site (**b**). Both viewed from the extracellular side. Highlighted are residues known experimentally to affect antagonist or agonist binding, respectively. The elipsoides schematically depict the antagonist or agonist binding-site, respectively. **c:** Side view of the D2 receptor TM domain.

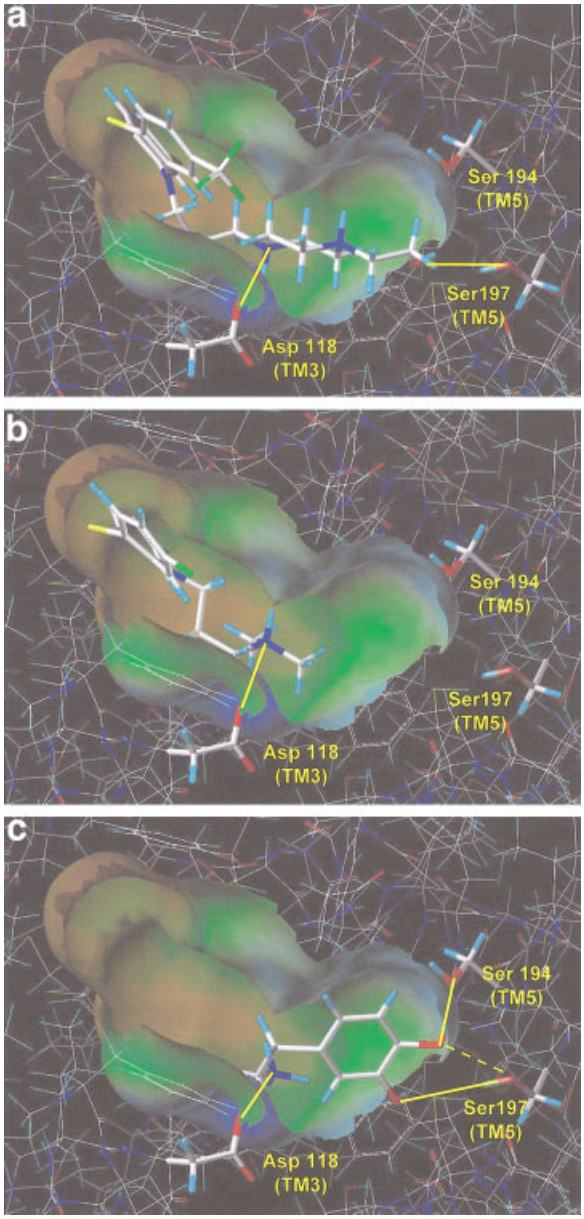
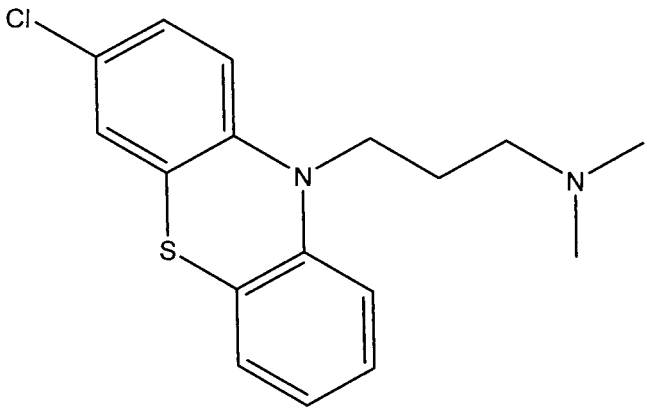


Fig. 7. (**a**) Fluphenazine (scheme 4), (**b**) chloropromazine (scheme 5), and (**c**) dopamine (Scheme 6) docked in the binding site of the dopamine D2 receptor model. The surface defining the binding site cavity is colored by lipophilicity, blue-green for hydrophilic regions and brown for hydrophobic regions. Specific interactions between the ligands and key protein residues are highlighted.

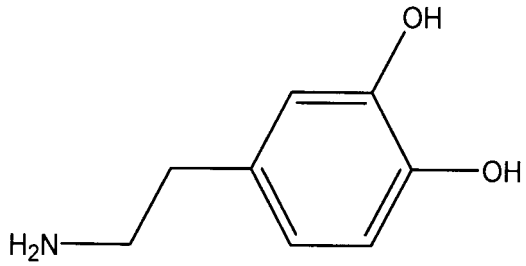
acts with the aromatic cluster and its tertiary amine moiety interacts with Asp114 (3.8 Å distance). Because chlorpromazine is shorter than fluphenazine, it does interact with Ser197, which lies at the far end of the pocket.

Agonist binding site

While antagonists were shown to interact with many D2 residues, the binding of D2 agonists is much more localized. It was experimentally found that such agonists interact mainly with residues on TMs 3, 5, and 6.[92] Key residues, which were experimentally found to affect agonist binding, include Asp114(3.32) in TM3, the three Ser residues in TM5-Ser193(5.42), Ser194(5.43), and Ser197(5.46), Phe198(5.47) in TM5, and Phe389(6.51) in TM6. Most of these residues—Asp114, Ser197, Phe198,



Scheme 5. Chloropromazine, a dopamine D2 receptor antagonist.



Scheme 6. Dopamine, a dopamine D2 receptor agonist.

and Phe389—are shared by the antagonist binding-site. Ser193 and Ser194 are unique for the agonist site.

As can be seen in Figure 6(c), all key residues listed map well in our D2 model, forming a localized agonist binding-site, significantly smaller than and partially overlapping the antagonist site described previously. This site has three polar interaction points: Asp114, which interacts with the amine moiety of the catecholamine agonist,[108,109] and two Ser interaction points, Ser194 and Ser193/Ser197, which interact with the two catecholamine hydroxyl moieties. This is in agreement with data showing that Ser197 is a hydrogen bond donor interacting with the hydroxyl groups of the dopamine catechol moiety.[100,110] The aromatic residues play a role in anchoring the agonist aromatic scaffold.

Figure 7(c) depicts the binding mode of dopamine (scheme 6), which is a small catecholamine agonist, docked into the D2 receptor binding-pocket. It is seen that dopamine fits well within the agonist binding site defined in Figure 6(b). Its primary amine group interacts with Asp114 (3.5 Å distance), and its two hydroxyl moieties of the catechol moiety are oriented towards the serine residues of TM5, in agreement with experimental data.[100,110] One of the two hydroxyl groups was placed near Ser194, at a 3.4 Å distance, somewhat between this serine and Ser197 at a distance of only 4.8 Å from it. In this specific docked conformation, the second hydroxyl, while pointing in the direction of Ser197, was a bit too far from the serine hydroxyl. A partial rotation of the compound within the binding pocket will bring these moieties to within interaction distance. It is clearly seen that dopamine does not interact with the aromatic cluster that binds the tri-cyclic moietied of the D2 antagonists, with the exception of Phe389, which interacts with the aromatic catachol moiety.

Overall dopamine D2 structure

Good agreement between the PREDICT model and experimental data is obtained not only within the binding pocket but also with regard to the rest of the structure.

In most vertebrate opsins, there is a conserved Pro residue in position 1.48 of TM1. In the crystal structure of bovine rhodopsin, this residue creates the inward bend of TM1. This Pro residue, however, is not conserved among other members of class-A GPCRs, suggesting that in its absence TM1 may orient more outwards, away from the binding crevice. Although biogenic amine receptors have a highly conserved Gly(1.49) residue in TM1, which may have introduced a bend in that helix, such an effect is not observed in our dopamine D2 model. In this model, TM1 remains close to linear and does not participate in the binding site. This observation is consistent with experimental data obtained from Substituted Cysteine Accessibility Method (SCAM) studies, which showed that the extracellular portion of TM1 is not part of the binding pocket.[101]

In the dopamine D2 model, TM2 exhibits a kink induced by the conserved Pro89(2.59) residue that is absent in bovine rhodopsin. The orientation of this Pro-kink differs from the kink observed in TM2 of the rhodopsin X-ray structure, possibly because the kink in rhodopsin is induced by Thr residues, which cause a wide range of perturbations in the φ and ψ backbone dihedral.[95,111] The role of the conserved Asp(2.50) in TM2 was examined in many GPCRs and was found to be associated with G-protein coupling, ligand binding, Na+ sensitivity, and more.[94,112] In the bovine rhodopsin structure, the distance between the conserved Asp(2.50) residue in TM2 and the conserved Asn(7.49) in TM7 is 9.03Å, which is not close enough for a direct hydrogen bonding but suggests the possibility of a hydrogen bonding network. Several studied showed that while a single mutation for one of these residues reduced or abolished ligand binding, a reciprocal mutation of both residues restored binding affinities.[46,96] In our D2 model, the Cα distance between the conserved Asp80(2.50) residue in TM2 and the conserved Asn422(7.49) in TM is 12.3Å. Although there is no direct experimental information regarding this distance in the dopamine receptor family, the 12.3Å distance in the D2 model suggests that these residues participate in a hydrogen bond interaction network similar to that of rhodopsin.

Although TM3 does not contain any Pro residues, it is slightly bent in the model. This is probably due to the conserved CTXS motif at positions 3.37–3.40, which can hydrogen bond to the helix backbone. A similar effect is seen in the rhodopsin structure, although the orientation of the bend is different. Since TM3 is found to be involved in receptor activation for several GPCRs,[1] it is possible that different orientations of the twist are associated with different functional state of the receptors.

TM4 contains a Trp residue, Trp160(4.50), that is highly conserved among class A GPCRs. In the PREDICT D2, model this residue faces outwards towards the hydrophobic membrane environment [Fig. 6(c)]. This is an atypical situation for a conserved residue, which would typically be involved either in intra-protein interaction or in ligand binding. Using the cysteine cross-linking method, Guo et al.[113] reported recently that TM4 lies at the dimer-interface of the D2 receptor homodimer. Therefore, it is quite possible that the membrane exposed Trp160(4.50), which favors aromatic interactions, contributes to stabilizing the D2 dimer. Interestingly, this result is in agreement with the position of Trp4.50 in the rhodopsin structure, where it is also exposed to the lipids environment. Recently, Liang et al.[114] determined by atomic force microscopy the organization of rhodopsin in native membranes obtained from wild-type mouse photoreceptors and opsin isolated from photoreceptors of Rpe65−/− mutant mice, which do not produce the chromophore 11-cis-retinal. Their results indicate that the inter-dimeric contact is likely to involve TM4 and TM5. Therefore, it seems that Trp4.50 has a functional role at the interface of GPCR dimerization.

The two Ser residues in TM5, Ser197(5.46) and Ser194(5.43), were discussed above in the context of the binding pocket. Using the model, we found that both residues point towards the binding crevice, but Ser197(5.46) is located close to TM3. In bovine rhodopsin, there is a His residue at position 5.46 where Ser197 is located in the D2 receptor sequence. This His211(5.46) residue participates in a strong electrostatic interaction with Glu122(3.37) at TM3. Since TM5 of rhodopsin possesses a very unusual kink with an unwinding around His211(5.46), this electrostatic interaction may be responsible for pertubing the helix conformation to form this unusual kink. It is possible that in the D2 receptor Ser197(5.46) has a similar structural role of affecting the kink conformation of TM5 through hydrogen bonding.

TM6 was found to be involved in receptor activation of several GPCRs including rhodopsin, β2 adrenergic receptor, and the opioid receptor family.[115–117] A mutagenesis study of the β2 adrenergic receptor that was combined with molecular dynamics simulation suggested that a rotation and/or tilting of TM6 may be involved in agonist activation. However, the mechanism of receptor activation differs among the GPCRs. For example, it has been proposed that activation of rhodopsin and biogenic amine GPCRs involves the disruption of a conserved Arg3.50–Glu6.30 ionic interaction between TM3 and TM6.[116] The orientation of TM3 and TM6 in the PREDICT D2 model indeed allows for a direct interaction between Arg132(3.50) and Glu368(6.30). On the other hand, opioid receptors have at position 6.30 a Leu residue, which cannot participate in such an interaction. A site-directed mutagenesis study that changed Leu6.30 to an Asp resulted in a constitutive active receptor, suggesting that differences in the local geometry of the cytoplasmic ends of TM5 and TM6 may be responsible for differences in receptor activation mechanisms.[118] In the PREDICT D2 receptor model, there is a pronounced Pro-kink at position 6.50 of TM6 causing the extracellular portion of the helix to be part of the binding crevice. This portion of the helix was discussed above in the context of the antagonist binding pocket; it contains the aromatic residues Trp386(6.48), Phe389(6.51), and His398(6.60), which participate in the aromatic cluster that lines the binding pocket. The orientation of the TM6 helix in the model is in good agreement with experimental results from a SCAM study showing that the three conserved aromatic residues and a histidine are exposed to the water-accessible surface of the binding cavity.[102]

The conserved NPXXY motif in TM7 was also found to be involved in the activation mechanism of many GPCRs.[97,119,120] Analysis of the rhodopsin structure, experimental data derived from metal ion activation, site-directed mutagenesis, and computational modeling of other GPCRs suggest that this NPXXY motif forms a flexible hinge region that may function as a conformational switch mediating receptor activation.[10,73,81,121] In our D2 receptor model, there is indeed a kink in TM7, which is induced by Pro7.50, which is part of the NPXXY motif, bending the helix in a way that allows for hydrogen bonding between TM7 and TM6. This result agrees with the hydrogen bond network analysis done of rhodopsin in the ground state, which showed that TM7 is the only helix that is connected by hydrogen bonds to TM6.[122] That study also showed that in rhodopsin, TM7 is hydrogen bonded to most of the other TMs except for TM4 and TM5. This kind of interaction network can explain the key functional role of TM7 in receptor activation, which is highly associated with the putative conformational changes in TM6.

In summary, the D2 receptor model bound to fluphenazine, which was generated by the PREDICT process agrees in all its details with known experimental data related to the orientation of the helices and to interactions within the binding site cavity, both for agonist binding and for antagonist binding.

Virtual Screening Using the D2 Receptor Model

A set of 43 known dopamine D2 agonists and antagonists with Ki values better than 1 μM was collected from the literature (Table III),of which 22 compounds had a Ki value lower than 10 nM. These compounds were based on more then 6 agonist scaffolds and more than 10 antagonist scaffolds (approximately half of the antagonists belong to the tri-cyclic amine class). The compounds exhibited a wide spread of physiochmical properties (Table I): MW ranges from 153 to 654 (average 361 ± 114), ClogP ranges from 0.2 to 6.6 (average 4.0 ± 1.5), number of acceptors range from 0 to 6 (average 1.8 ± 0.8), and number of donors range from 1 to 4 (average 2.0 ± 1.5). Of these, 10 compounds were D2 agonists and 33 compounds were antagonists The 43 known D2 ligands were seeded to the 10,000 compound Random Library 1, which has average properties very similar to those of the D2 ligands (Table I). The joined library was screened against the receptor model according to the docking protocol detailed in Materials and Methods. In preparation for DOCK, the binding-site was described by 22 SPHGEN spheres, energy and contact

TABLE III. Known Dopamine D2 Receptor Ligands

Ligand name	Mechanism	Ki (nM)	Reference
Bromocriptine	Agonist	0.62–8.2	154–156,110,164
Lisuride	Agonist	1.4	155
N-propylnorapomorphine	Agonist	2.9	155
Pergolide	Agonist	19	156,164
Apomorphine	Agonist	30–63	154–156,110
7-OH-DPAT	Agonist	60–158	155
Talipexole	Agonist	390	155
Dopamine	Agonist	549–1,000	110,154–156
Quinpirole	Agonist	600–1,400	154,155,156,164
Pramipexole	Agonist	790	155
Spiperone	Antagonist	0.16	159
Pipotiazine	Antagonist	0.2	156
Risperidone	Antagonist	0.3	157
Fluphenazine	Antagonist	0.32–4	157,159
Haloperidol	Antagonist	0.35–4	158,154,159,156,157
Pimozide	Antagonist	0.39–29	156,154,160
Thioridazine	Antagonist	0.4–26	159,157
Prochlorperazine	Antagonist	0.4–7	159,156
Butaclamol	Antagonist	0.44–0.86	154
Thioproperazine	Antagonist	0.45	156
Domperidone	Antagonist	0.58–1.3	156
Chloropromazine	Antagonist	0.66–19	154,156,157,159
Trifluoperazine	Antagonist	0.96–2.8	159,157
Raclopride	Antagonist	0.64	110
Sertindole	Antagonist	0.95	157
Amisulpride	Antagonist	1.3	156
Perphenazine	Antagonist	1.4	159
Olanzapine	Antagonist	3.7	157
Sulpiride	Antagonist	3–27	159
Loxapine	Antagonist	5.2–70	159,161
Molindone	Antagonist	6–120	159
Chlorprothixene	Antagonist	8	159
Carpipramine	Antagonist	8–70	156
Iloperidone	Antagonist	21.4	162
Remoxipride	Antagonist	30–198	156,163
Clozapine	Antagonist	38–180	154,156,157,159
Perlapine	Antagonist	60	157
Melperone	Antagonist	88	157
Promazine	Antagonist	160	159
Seroquel	Antagonist	186–263	154
Doxepin	Antagonist	360	160
Femoxetine	Antagonist	590	160
Imipramine	Antagonist	620	160

scoring grids of 0.3-Å spacing were generated using a united-atom representation of the protein, and a distance-dependent dielectric constant of 4r with a 15-Å cutoff. The docking of the compound library was repeated 10 times (each with a different random seed) with the best docking score for each compound taken as the final docking score. DOCK parameters were tuned according to the structural features specific to the virtual D2 receptor-fluphenazine complex.

Figure 8 shows the enrichment graph for the known D2 antagonists and for the known D2 agonists. First we will discuss the virtual screening of D2 antagonists. The enrichment factor for identifying 50% of the known D2 antagonists, $Enrichment^{50\%}$, is 17-fold higher than random screening, and all of these compounds were ranked within the top 3% of the dock-score ranked library (i.e., Coverage3% = 50%). $Coverage^{10\%}$ for the D2 antagonists, i.e., the percent of known antagonists within the top 10% of the DOCK ranked library, equals 85%. These results demonstrate that the PREDICT model was able to discriminate between D2 antagonists and nonactive drug-like compounds for this target.

Since the virtual complex of the D2 receptor model was created with a D2 antagonist, fluphenazine, it was questionable whether virtual screening with known D2 agonists would be successful at all, more so as we saw that the agonist binding-site is only a portion of the whole D2 pocket. Despite the small number of D2 agonists that used the enrichment factor for identifying 50% of these known D2 agonists, $Enrichment^{50\%}$, is 9-fold higher than random

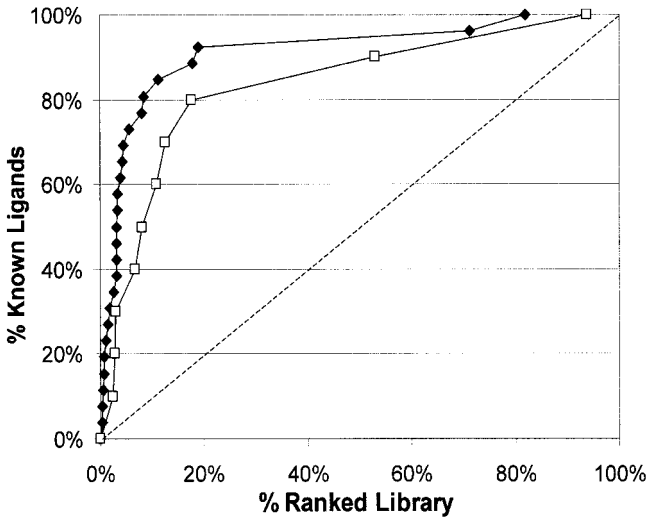


Fig. 8. Enrichment graph for in silico screening of known D2 antagonists (dark diamonds) and D2 agonists (open squares) embedded in the 10,000 compound random drug-like library 1 using the PREDICT model of the dopamine D2 receptor. Enrichment at 50% is 17- and 9-fold better than random screening (dashed line) for D2 antagonists and agonists, respectively.

screening, and all of these compounds were ranked within the top 6% of the dock-score ranked library (i.e., Coverage6% = 50%). $Coverage^{10\%}$ for these D2 agonists is 70%. Namely, despite the above considerations, the model was also successful in screening for D2 agonists, albeit to a lesser degree than for antagonists. It is possible that if the virtual protein-ligand complex had been created with a D2 agonist, the virtual screening results would have turned out more successful for agonists. Table IVsummarizes these screening results. It is worth noting that the Ki values of most D2 agonists are significantly higher than those of the D2 antagonists that may have also contributed to the lower agonist enrichment. Interestingly, the two least potent D2 agonists in our library, quinpirole and pramipexole, were also ranked last among the docked agonists, and the most potent D2 agonist, bromocriptine, was ranked among the top two docked agonists when sorted by the docking score.

Modeling the Human NPY Y1 Receptor

The D2 receptor model that was discussed above is a representative of the biogenic amine receptor class. The second GPCR model that will be discussed here in detail is of the NPY Y1 receptor, which is a representative of peptide receptors. Neuropeptide Y (NPY) receptors are class A GPCRs, which mediate several physiological responses, such as blood pressure, food intake, sedation, and memory retention. Six subtypes of NPY receptors have been cloned, showing a large degree of sequence variation.[123,124] The NPY Y_1 (Y1) receptor is expressed in the central and the peripheral nervous system and plays a role in the in vivo release of oxytocin,[125] and the release of LHRH from hypothalamic tissue.[126] It is 384 amino acids long and possesses very low homology to rhodopsin (19% overall and 26% at the TM region). The native ligand of the Y1 receptor is neuropeptide Y (NPY), which is a 36 amino acid peptide. NPY interacts with residues on the ECLs and in the TM domain of the Y1 receptor.[127] However, small molecule antagonists bind primarily within the TM domains. For example, compound BIBP3226 (Scheme 7), which was designed to mimic the C-terminus of the NPY peptide, binds within the TM domains,[12] much like the C-terminus of the NPY peptide itself.[128]

We used the PREDICT modeling process to generate a 3D model of the NPY Y1 receptor. Briefly, in this case 614 2D conformations were generated for the initial coarse modeling step. Each of these 614 conformations were rotationally optimized and expanded into 3D reduced representation structures. Each putative structure was then optimized for vertical positioning of the helices, rotational orientation of the amino acids in the helix and the position of the helical centers, simultaneously optimizing side chain rotamer packing. The 15 structures that scored best according to the PREDICT energy function (Eq. 16) were then clustered using a principal component analysis projection. Based on PREDICT energy score, agreement with experimental data, size of the putative binding crevice, and other general physiochemical properties of the structure, one representative structure was selected from each of the three clusters that were identified. Approximately 1,000 conformations were generated around each of the three lead structures for the second step of fine PREDICT modeling, following the protocol described in Materials and Methods. At the end of the modeling process, a single optimized model was obtained from each of the original lead-models. One of these models was easily discarded due to discrepancies in the locations of key residues at the binding crevice. Therefore, only the lowest energy models from the other two lead structures (lead-model-170, lead-model-65) were expanded into full all-atom models and subjected to molecular dynamics refinement. Only one of these two structures, stemming from lead-model-170, had a binding site big enough to accommodate the ligands. This model was selected as the final NPY Y1 receptor model. To model the antagonist bound conformation of the Y1 receptor, we introduced the potent small compound antagonist BIBP3226 (Ki=5.1 nM, Scheme 7) into the binding cavity using the molecular dynamics protocol describe in Materials and Methods. The initial orientation was with the guanidinium group in the site near Asp287(6.59) and the diphenyl group pointing outwards. The weak NOE constraint, which prevents the compound from leaving the site when the value of λ is small, was placed between the carbon joining the two amide groups of BIBP3226 and the geometrical center of the binding site (calculated by CHARMm).

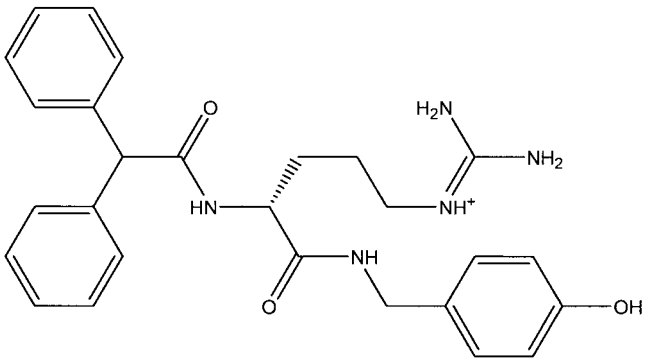
Analysis of the NPY Y1 Receptor Model

A comparison of the PREDICT Y1 receptor-ligand complex to mutagenesis data for BIBP3226 at the Y1 receptor showed excellent agreement between the two. The PREDICT model was also compared to a proposed NPY Y1 receptor model published by Du et al.[12] based on their mutagenesis data using the low-resolution structure of

TABLE IV. In Silico Screening Hit Rates and Enrichment Factors Using 3D GPCR Models Generated by PREDICT

Receptor	Library	No. known ligands	Coverage$^{10\%}$	Enrichment$^{50\%}$
D3 receptor: antagonists	Random Lib 1	33	85%	17-fold
D2 receptor: agonists	Random Lib 1	10	70%	9-fold
NPY receptor	Random Lib 1	18	100%[a]	44-fold
NK1 receptor	Random Lib 2	26	80%	20-fold

[a]All the NPY ligands were found within the top 8% of the ranked library.



Scheme 7. BIBP3226, a NPY Y1 receptor antagonist.

bovine rhodopsin[129] as a template. As will be discussed below, the PREDICT model agrees with the experimental data better than the previously proposed rhodopsin-based model.

The binding crevice of the Y1 receptor model is located between TM 3, 4, 5, 6, and 7 (Fig. 9). The overall conformation of BIBP3226 in the binding site is such that the guanidinium group faces TMs 5 and 6, the phenol is positioned near TM6 and TM7, and the diphenylmethyl group is generally positioned between TM 3, 6, and 7. Several hydrogen bonds between the ligand and the receptor are suggested by the 3D model, including two hydrogen bonds between the argininamide backbone of BIBP3226 and residues Asn283(6.55) and Asp287(6.59) in TM6. Asp287 is known to be crucial for agonist binding and is thought to interact with the guanidinum group of Arg33 of NPY.[12] To validate these interactions, we docked into the NPY Y1 pocket another guanidinuim derivative antagonist, LY-357897. It was found that in this compound too, the gunaydyl group of the ligand fitted into the negative charged cleft between these two residues, creating strong electrostatic interactions with both. This agrees with the experimental finding that mutation of Asn283(6.55) to Ala resulted in a reduced binding affinity for BIBP3226.[128,130]

Residue Phe286(6.58) at TM6, which is one helical turn above Asn283, was also found to be important for BIBP3226 binding.[128,130] The distance in our model between the guanidinium group and Phe286 allows for cation-π interactions between the two moieties.

According to the model, residue Gln219(5.46) in TM5 is also located within a hydrogen-bond distance from the argininamide backbone of BIBP3226. Du et al.[12]found that mutation of Gln219 resulted in a reduced affinity for BIBP3226 and NPY binding. However, their rhodopsin-based model could not account for this interaction between BIBP3226 and the receptor.[12] The new PREDICT model, on the other hand, agrees with the hypothesis that the Gln219 side chain is located within the binding crevice but distant from the aromatic moieties of the ligands. Note that another important hydrogen bond suggested by the model is between the hydroxyl group of BIBP3226 and Gln120(3.33) in TM3. Mutagenesis data indicates that residue His306(7.38) is important for the binding of NPY and BIBP3226.[12] This interaction was not reproduced in the rhodopsin-based model of Du et al.[12] in which His306 is not part of the binding pocket. The PREDICT model, on the other hand, clearly shows His306 facing the binding pocket, with the histidine imidazole ring forming an aromatic interaction with the diphenyl moiety of BIBP3226. Residue Tyr211 (5.37) is also of interest in this context. Mutation of this residue to Ala in the human Y1 receptor retained wild-type affinity for NPY but showed no affinity for BIBP3226.[128] In the PREDICT model, this residue is located at the bottom of the binding pocket and is involved in aromatic stacking with the phenol ring of BIBP3226, explaining the observed effect on BIBP3226 binding.

Interestingly, alanine mutants at positions Tyr100 and Asp104, which are located at ECL1 at the top of TM2 of the

TABLE V. Known Small-Molecule NPY Y1 Receptor Ligands

Ligands name	Ki (nM) (PYY)	Reference
BIBO-3304	0.38[a]	165
Sub-Bimid	0.052	166
BIBP-3226-3	0.075	167
LY-357897	0.75	168
BIBP-3226-56	1.7	167
BIBP-3226	4.6	169
BIBP-3226-61	7	167
J-104870	1.3[b]	170
J-115814	1.4–1.9	171
H-409/22	5.6	172
Guanylurea	31[a]	173
US 6218408	Binder[c]	Patent
WO 9614307	Binder[c]	Patent
WO 9803493	Binder[c]	Patent
WO 9803492	Binder[c]	Patent
WO 9803494	Binder[c]	Patent
WO 9719911	Binder[c]	Patent
WO 9915498	Binder[c]	Patent

[a]IC50 value.
[b]Kd value.
[c]Exact Ki value not published.

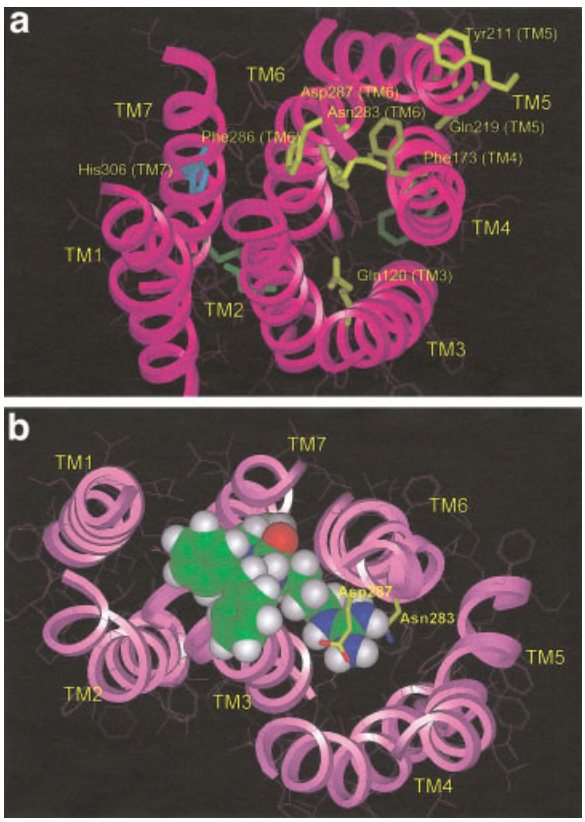


Fig. 9. (**a**) The PREDICT model of the NPY Y1 receptor highlighting key residues, viewed from the extracellular side; (**b**) A proposed binding mode for BIBP326 (Scheme 7, sphere representation) docked in the binding site of the NPY Y1 receptor model. Residues Asn283 (6.55) and Asp287 (6.59) that were implicated in BIBP3226 binding are shown in a stick representation.

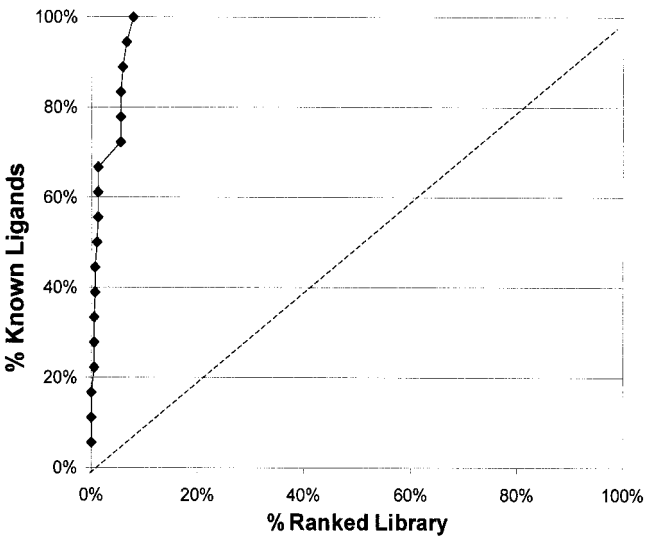


Fig. 10. Enrichment graph for in silico screening of 18 known NPY Y1 antagonists embedded in a 10,000 random compound library against the PREDICT model of the NPY Y1 receptor. Enrichment at 50% is 44-fold better than random screening (dashed line).

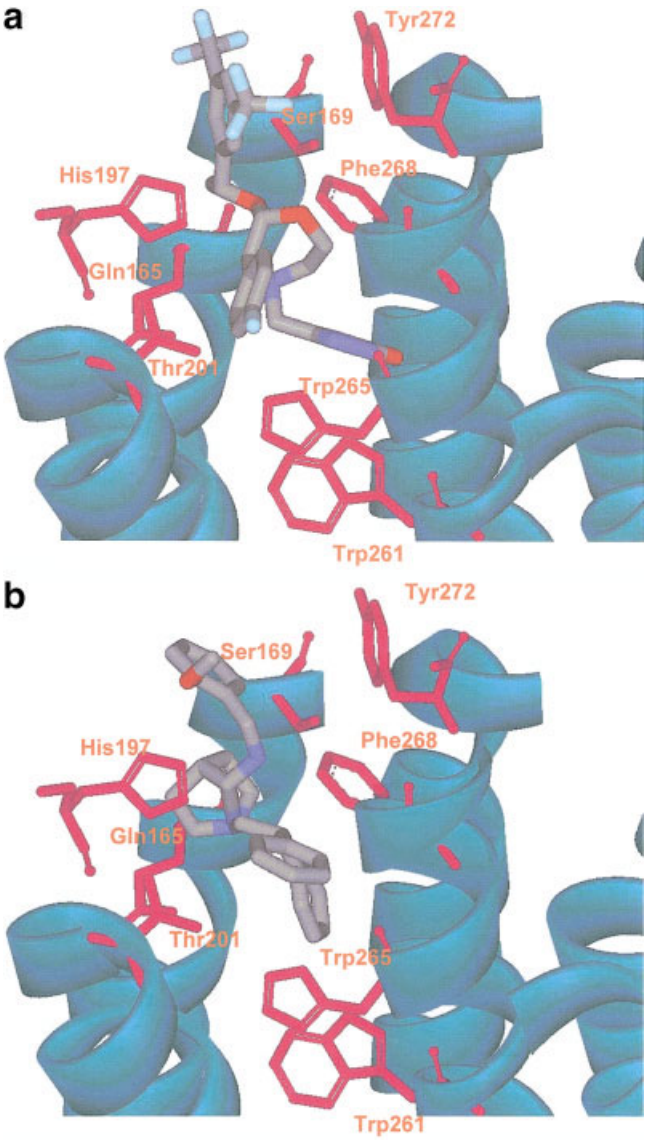
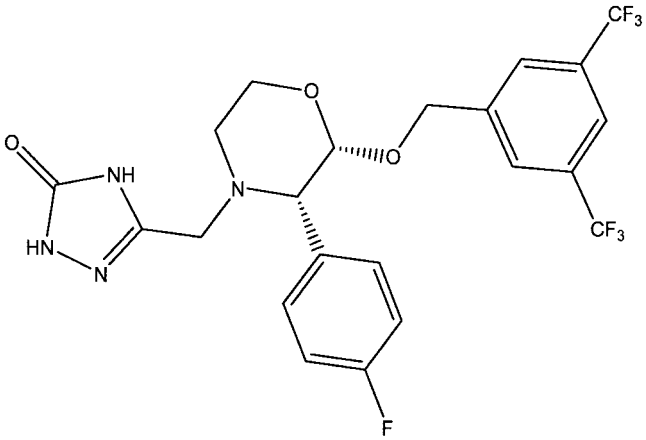


Fig. 11. (**a**) Aprepitant (Scheme 8) and (**b**) CP-96345 (Scheme 9) docked in the binding-site of the NK1 receptor model. Protein residues that are implicated in specific interaction with the ligands, as well as several other key residues, are highlighted.

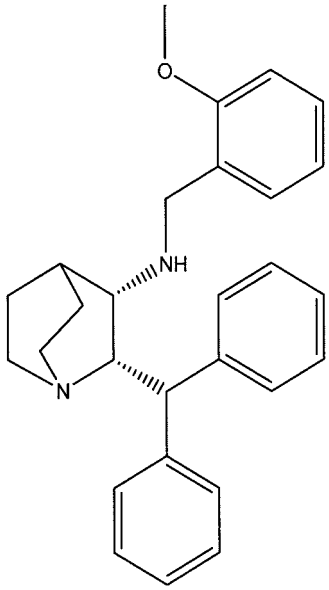
human Y1 receptor, had no effect on BIBP3226 binding, but significantly reduced binding of the indigenous NPY peptide.[128] Although these ECL1 residues are not included in the model, by extending the helix and base of the position of Tyr100 in the helical wheel, one can hypothesize from the model that this residue faces a small crevice formed between TMs 1, 2, 6, and 7. It is proposed that this small secondary site may be involved in the binding of the N-terminus of the NPY peptide. Since BIBP3226 mimics the interactions formed by the C-terminus of the peptide, it is not surprising that mutations of Tyr100 and Asp104 residues had no affect on BIBP3226 binding.

Virtual Screening Using the NPY Y1 Receptor

To evaluate whether the Y1 receptor model can be used for in silico screening, a set of 18 known potent NPY Y1



Scheme 8. Aprepitant (MK869), a NK1 receptor antagonist.



Scheme 9. CP-96345, a NK1 receptor antagonist.

antagonists was collected from the literature (Table V). These known antagonists satisfy the requirement of structural diversity, based on more then 6 different scaffolds and exhibiting a broad range of physiochmical properties: MW ranges from 360 to 593, ClogP ranges from 0.5 to 7.5 (average 4.6 ± 2.3), number of acceptors range from 1 to 6 (average 3.0 ± 1.2), and number of donors range from 1 to 7 (average 3.7 ± 2.0). These known ligands were added to the 10,000 compound Random Library 1 (Table I). The compounds on this library have chemophysical properties similar to the NPY ligands (e.g., ClogP, number of acceptors), although the average size of the compounds in this library is smaller. The compounds were then docked into the Y1 receptor model according to the screening protocol described above. Ten docking repetitions were performed for each compound and the best docking score was used. The average dock score of the known ligands was −72 kcal, while the average docking score of the random compounds was −26 kcal, clearly indicating that the model favors the ligands over random compounds. The enrichment graph for this in silico screening is shown in Figure 10. The enrichment factor for identifying 50% of the known ligands, $Enrichment^{50\%}$, is 44-fold higher than random screening, and all of these compounds were ranked within the top 1.1% of the dock-score ranked library (i.e., $Coverage^{1\%} = 50\%$). In this system 100% percent of the ligands were identified within the top 8% of the ranked library ($Coverage^{8\%} = 100\%$). It is possible that these results are somewhat shifted upwards because of the smaller average size of the random compounds. However, we believe that they clearly indicate that the human Y1 model can be used to discriminate known human Y1 receptor drugs from random drug-like compounds using our in-silico approach.

Modeling the Human NK1 Receptor

The three structurally related tachykinins, substance P (SP), neurokinin A (NKA), and neurokinin B, are widely distributed in the central and peripheral nervous systems exerting their physiological effects by binding to the NK1, NK2, and NK3 receptors, respectively.[131] Substance P plays a role in many biological processes and is involved is CNS disorders, transmission of pain, regulation of immune response, dural inflammation in migraine, cytotoxin-induced emesis, cystisis, asthma, etc. Based on the wide distribution of NK1 receptors and their diverse pathophysiological roles, it has been anticipated that NK1 receptor antagonists may have several therapeutic applications at central and peripheral level. And in other peptide receptors, binding of the indigenous peptide, the 11 amino acid Substance P, to the receptor involves TM residues as well as residues on first and second extracellular loops (Asn23, Gln24, and Phe25 from ECL1; Asp96 and His108 from ECL2).[132] However, it has been shown by structure activity studies that the binding site of nonpeptide antagonists lies primarily within the TM domain, mainly near TMs 4, 5, and 6.[133] This suggests that a 3D PREDICT model of this receptor's TM domain can be used for virtual screening of small molecule NK1 antagonists.

The PREDICT modeling process described above was used to generate the 3D model of the NK1 receptor. Based on the extensive site-directed mutagenesis studies of the NK1 receptor that mapped major interaction points of nonpeptide antagonists to TMs 4, 5, and 6,[133] the initial distance between TM4 and TM6 was constrained to be between no more than 22 Å. This is in addition to the Class A constraint, which restricts the initial distance between TM2 and TM7 to be no more than 20 Å. This resulted in a set of approximately 1,000 initial 2D conformations. Each of these conformations was rotationally optimized and expanded into 3D reduced representation structures. Each putative structure was then further optimized as described above. After scoring the structures with the PREDICT energy function and clustering the best 15 structures using a principal component analysis projection we selected model-961 as a template structure for the second, finer, modeling step (based on its energetic score, agreement with experimental data, and the physiochemical

properties of the binding cavity). Based on lead-model-961, approximately 1,200 conformations were generated in the second step of fine PREDICT modeling. Two of these models (model-82 and model-233) were selected as the final PREDICT models based on the PCA clustering results, the energy scores, the size of the binding pockets, and the agreement with experimental data. Both models were further refined using the molecular dynamics protocol described in Materials and Methods. During this simulation, a gap between helices 3 and 4 of model-82 was observed causing the rotamers of key residues in TM4 to point outside the binding crevice. Hence, this model was discarded. On the other hand, the binding site in model-233 was big enough to accommodate the ligands and the key residues at TMs 4, 5, and 6 were all oriented toward the binding site. This model was, therefore, selected as the final NK1 receptor model.

A virtual receptor-ligand complex was created with aprepitant (Scheme 8), a potent NK1 antagonist (Ki = 0.9 nM), which was recently approved for the treatment of acute and delayed nausea and vomiting associated with chemotherapy. Aprepitant was designed by Merck based on the quinuclidine series of NK1 antagonists from Pfizer (representative compound CP-99994[134]). Hence, experimental information on the interactions of quinuclidine antagonists with the NK1 receptor was used to guide the initial positioning of aprepitant in the binding site. The initial binding mode was then refined by the virtual complexation simulations protocol described in Materials and Methods, with a NOE constraint defined between the carbon at position 3 of the morpholine ring of aprepitant and the geometrical center of the binding site. The resulting virtual complex was in good agreement with experimental data on the interactions of the similar quinuclidine antagonists with the receptor (described below). However, the side chain of Gln165(4.60), which was found to play a key role in the binding of quinuclidine antagonists, probably through hydrogen bonding with the OH or N group at position C3, is facing TM5 and not the binding crevice. The distance between the O atom at position 2 of the morpholine ring of aprepitant (which is in the equivalent position to the NH group in CP 96345 (Scheme 9)[138]) and the N atom of the Gln165(4.60) amide in the virtual complex, is 6.7Å, which is too large for direct hydrogen bonding. However, the distance between this oxygen and the OH group of Ser169(4.64) is 4.2 Å, which is sufficient for hydrogen bonding. Studies showed that different NK1 antagonists can bind to the same region of the receptor although the specific intermolecular interactions vary within this region.[135] For example, CP 96345 interacts strongly with Gln165(4.60) but has only a weak interaction with Ser169(4.64) while the opposite was shown for compound RP 67580. To further optimize the interaction with Gln165(4.60), the rotamer of this side chain was rotated to face the binding pocket and aprepitant was then removed from the model leaving a set of 3D coordinates that are ready for virtual screening.

Analysis of the NK1 Receptor Model

The binding site of Substance P for the receptor involves most of the TMs as well as residues within the first and second extracellular domains (see Fong et al.[132] for a review). On the other hand, site-directed mutagenesis studies showed that small molecule antagonists bind to the NK1 receptor in a site that includes mainly the top portions of TM5 and TM6 with additional interaction points present on TM3 and TM4.[136–141] Mutations that affect the binding of small molecule antagonists failed to affect the binding of the natural ligand, suggesting that different domains of the receptor are involved in the binding of the native ligand and the nonpeptide antagonists.

Aromatic residues play an important role in the binding of nonpeptide antagonists to the NK1 receptor through amino-aromatic interactions or by a cluster of aromatic rings. In the NK1 receptor model, several aromatic residues (including His197(5.39), Trp261(6.48), Phe264(6.51), His265(6.52), Phe268(6.55), and Tyr272(6.59)) at the extracellular portion of TMs 5 and 6 are facing the binding crevice and can participate in ligand binding. Site-directed mutagenesis studies demonstrated that the benzhydryl group at position 2 of the quinuclidine ring of CP 96354, a potent and selective NK1 antagonist,[138] participates in an amino aromatic interaction with residue His197(5.39) of the NK1 receptor. This interaction is also reproduced in the PREDICT NK1 model in which the distance between the N atom of His197 and the C4 atom of the benzhydryl group is 3.1Å. In the aprepitant-NK1 virtual complex, His197(5.39) interacts with the 4 fluorobenzyl ring at position 3 of the morpholine, which mimics the benzhydryl substituents in the quinuclidine antagonists. His197 was found to participate in binding of Zn^{2+} ion when two other residues, Glu193(5.35) and Tyr272(6.59), were also mutated to histidine.[142] In the double-mutated receptor (E193H, Y272H), the binding site for CP96345 was virtually exchanged with a submicromolar affinity of the tridentate Zn^{2+} site.[142] The side chains orientation of His197(5.39) and Tyr272(6.59) and the volume between them can account for the accommodation of a Zn^{2+} ion and for a volume exclusion affect of the ion on the binding of CP 96354. Although Glu193(5.35) was not included in the TM5, it is located one helical wheel above His197 and, therefore, would probably face the same microenvironment of His197 and Y272. A single mutation of Tyr272 to Ala impaired the binding of LY303870 to the NK1 receptor but did not affect the binding of several other nonpeptide antagonists including CP 96345. In agreement with that, in the PREDICT NK1 model, Tyr272 is located with the binding cavity but is not involved in a direct interaction with CP 96345 [Fig. 11(b)]. The binding mode of LY303870 docked to the NK1 receptor model includes interactions with both Tyr272(5.59) and His197(5.39) (data not shown).

His265(6.52) was also shown to be involved in the binding of the quinuclidine antagonists to the receptor. Analysis of structure activity relationships of the quinuclidine antagonists with the wild-type and H265A mutant receptor suggest an amino aromatic interactions of the His

residue with the 3,5 substituted benzyl moiety at the 3 position of the quinuclidine ring. This interaction is reproduced in the CP 96354 compound docked into the receptor structure. The 3,5 disubstituted benzyl ring of aprepitant also interacts with His265(6.52) in the aprepitant-NK1 receptor complex, indicating the similar binding modes that two the compounds possess in the receptor binding pocket. The orientation of the histidine hydrogens and the position of the benzyl ring suggest an amino-aromatic interaction.

Another aromatic residue in TM6, Phe268(6.55), was also found to be involved in the binding of several nonpeptide antagonist.[134] Mutating Phe268(6.55) to an Ala resulted in a reduced affinity for all the nonpeptide antagonists that were tested. However, the binding of the ligands to the F268W or F268H mutated receptors was almost not affected, indicting the importance of an aromatic side chain in this position.[134] In the aprepitant-NK1 complex model, the 4-fluoro phenyl ring at position 3 of the morpholine is part of an aromatic cluster with both His265 and Phe268 [Fig. 11(a)]. The 3,5 substituted benzyl substituents of CP 96345 are also involved in π-π stacking with His265and Phe268 in the docked conformation of CP 96345. Trp261(6.48) in TM6 is positioned at the base of the binding pocket and also participates in this aromatic clustering, indicating the importance of aromatic residues in the binding of small molecule antagonists to the NK1 receptor.

Two additional residues that play a role in the binding of nonpeptide antagonists to the NK1 receptor are Gln165(4.60) and Ser169(4.64), positioned at the top of TM4. In the final receptor model, the two residues face the binding crevice; however, the orientation of their side chain can create specific interactions with different ligands. For example, the NH group of CP 93645 is positioned in a hydrogen bonding distance from the amide group of Gln165(4.60) (3.7Å distance). However, its distance from the hydroxyl group of Ser169(4.64) is too far to create a direct hydrogen bond (4.75Å distance). On the other hand, the binding mode of L-703776 allows for direct hydrogen bonding with the two residues (distances between the O atom at position C3 of L-703776 and the N of the amide group of Gln165(4.60) or the OH group of Ser165(4.64) are 3.83 and 4.3Å, respectively). This is in agreement with several studies that indicated that different NK1 antagonists could bind to the same region of the receptor although the specific intermolecular interactions vary within this region.[135] For example, CP 96345 interacts strongly with Gln165(4.60) but has only a weak interaction with Ser169(4.64) while the opposite was shown for RP 67580. Thr201(5.43) located at the top of TM5 can also hydrogen bond with both Gln165(4.60) and Ser169(4.64) in the NK1 receptor model. Mutations of Thr201(5.43) to Ala or Ile were found to impair the binding of two NK1 antagonists, CAM4092 and LY303870, while they had no effect on the binding of other antagonists including CP 96345 and RP 67580. The close proximity of Thr201(5.43) to Gln165(4.60) and Ser169(4.64) can create a network of hydrogen bonds that can account for differences in sensitivity of the various antagonists to mutations of these residues.

TABLE VI. Known Small-Molecule NK1 Receptor Ligands

Molecules	Ki (nM)	Reference
L741671	0.03	174
GR203040	0.05	175
L743310	0.06	174
Comp9d	0.12	176
SDZnkt343	0.16	177
CP99994	0.28	134
CP96345	0.31	134
PD154075	0.35	178
LY303870	0.38	134
SR140333	0.44	134
WO0216324	0.5	Patent
CGP49823	0.69	134
FK888	1.1	134
CAM4092	3.5	134
SR140603	4.2	179
Comp3	0.19[a]	180
Comp43	0.24[a]	181
TAK637	0.45[a]	182
PD174424	0.46[a]	183
PD160226	0.73[a]	183
Comp3a	0.8[a]	184
Aprepitant (MK869)	0.9[a]	180
Comp1t	1.4[a]	184
NKP608	2.6[a]	185
Comp32	2.8[a]	181
MDL105212a	3.69[a]	186

[a]IC50 value.

Virtual Screening Using the NK1 Model

A set of 26 analogs were selected from the literature to test the ability of the NK1 receptor model to discriminate between known ligands and random drug-like compounds. All the 26 ligands were potent NK1 antagonists with Ki values better than 10 nM (Table VI).These antagonists satisfy the requirement of structural diversity, belonging to more than 12 different scaffold families, and exhibit a broad range of physiochmical properties: MW ranges from 298 to 640 (average 530 $\pm$ 78), ClogP ranges from 2.1 to 6.7 (average 5.1 $\pm$ 1.2), number of acceptors range from 1 to 11 (average 5.9 $\pm$ 3.2), and number of donors range from 0 to 4 (average 1.8 $\pm$ 21.3). These known ligands were added to the 6,200 compound Random Library 2 (Table II), which was created to have chemophysical properties similar to those of the NK1 ligands (Table II, see Materials and Methods). The compounds were then docked into the NK1 receptor model according to the screening protocol described above. Ten docking repetitions were performed for each compound and the best docking score was used.

The enrichment graph for this in silico screening is shown in Figure 12. The enrichment factor for identifying 50% of the known ligands, $Enrichment^{50\%}$, is 20-fold higher than random screening, and all of these compounds were ranked within the top 2.5% of the dock-score ranked library (i.e., $Coverage^{2.5\%} = 50\%$). The four best scored ligands were CAM4092, SDZ NKT 343, aprepitant, and

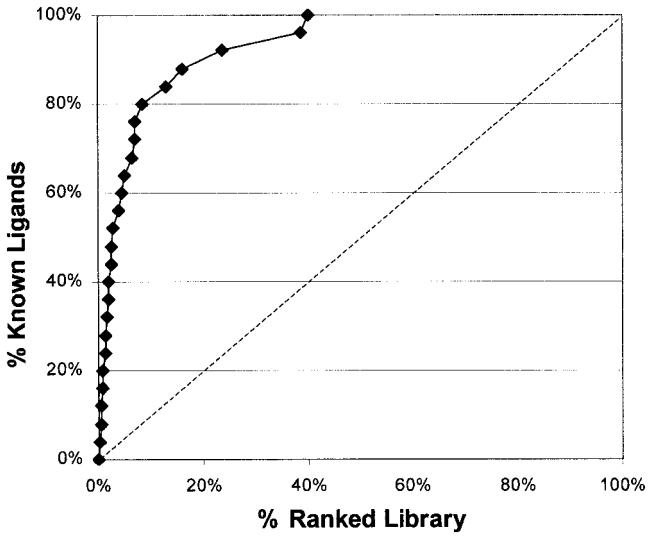


Fig. 12. Enrichment graph for in silico screening of 26 known NK1 antagonists embedded in the 6,200 compound Random Library 2 against the PREDICT model of the NK1 receptor. Enrichment at 50% is 20-fold better than random screening (dashed line).

GR203040, each belonging to different chemical classes. This suggests that although the NK1 receptor model was virtually complexed with one antagonist, aprepitant, the structure is general enough to allow accurate identification of diverse set of ligands (not just aprepitant-like compounds). This is, of course, essential for drug discovery purpose in order to identify new novel chemical entities. In this system 80% of the ligands were identified within the top 10% of the ranked library (Coverage$^{10\%}$ = 80%), indicating that the NK1 model can be used to discriminate known human NK1 receptor drugs from random drug-like compounds using our in-silico approach.

DISCUSSION

PREDICT Modeling Vs. Homology Modeling

The availability of a low-resolution structure for bovine rhodopsin opened a new vista for virtual drug discovery of GPCRs using the structure as a template for homology modeling. The specific arrangement of the seven TM helices stabilized by a series of intramolecular interactions mediated by several backbone and side-chain atoms seemed to be conserved among the class A receptors. In comparison to bacteriorhodopsin, which for many years was the only seven TM protein with a known structure (albeit not a GPCR), rhodopsin represents a much better template for homology-based models for class A receptors. Nonetheless, there are several reasons why rhodopsin is still not an ideal template for modeling drugable GPCRs. First, rhodopsin is an atypical nondruggable GPCR limiting its relevance to drug discovery. Second, it is a light receptor with its natural ligand bound to it at the ground state functioning as an inverse agonist. Third, rhodopsin possesses a relatively low sequence similarity to other GPCRs. In general, the success of homology modeling is determined to a high extent by the degree of sequence homology between the target and the template structures. In particular, clear structural similarity exists for highly homologous proteins with sequence identity >30%[143–145] while a lower sequence homology reflects divergence in the protein structures.[126,127] The sequence identity between bovine rhodopsin and other non-opsin GPCRs is very low, even within the TM domains (e.g., homology to human dopamine D2 receptor is only ~25% for the TM domains and ~19% for the entire sequence).[19] Therefore, the ability of rhodopsin-based homology modeling to yield realistic models for all class-A GPCRs is questionable.

Regardless of these issues, many rhodopsin-based homology models of GPCRs have been created.[11–20] Though many of these models have not been fully validated, several studies suggest that rhodopsin can be used to generate models that are accurate enough for the interpretation of experimental results.[19,146,147] Recently, Bissantz et al.[6] created structural models for five different GPCRs based on the structure of rhodopsin as a template. The modeling procedure in that study included alignment of the receptor sequences to the rhodopsin sequence, creation of the 3D model using the rhodopsin X-ray structure as a template, and optimization of the structure by energy minimization. These models were agonist-bound or antagonist-bound depending on the type of ligand that was inserted into the binding pocket to create the virtual receptor-ligand complex. The ligand-bound structural models were screened against a library of 1,000 random "drug-like" compounds seeded with 10 known binders of the target receptors. The virtual screening results obtained with these models were quite good, with optimal hit rates in the range of 10 to 40%. The poorer hit rate obtained for the M1 muscarinic receptor may indicate that rhodopsin is not a good template for modeling this receptor. The enrichment factors and coverage values achieved by the PREDICT algorithm for the receptor models discussed herein (Table IV) indicate that a rhodopsin-independent approach is also suitable for in silico screening and GPCR drug discovery, with enrichment factors comparable to those obtained when using crystal structure for non-GPCR targets,[33,34] In a recent study, Vaidehi et al.[187] used the rhodopsin structure as a template for modeling the dopamine D3 receptor. This study included homology modeling of the D3 receptor based on the X-ray structure of rhodopsin, refinement of the structure in the presence of an explicit lipid bilayer and water environment, followed by a combined computational screening approach using the model and a ligand based pharmacophore. Using this approach, 8 out of 20 compounds selected for testing in the D3 receptor binding assay had Ki values better than 1 μM.

Though the rhodopsin structure proved useful for modeling specific biogenic amine receptors, many other attempts that also used rhodopsin as a template were not successful in creating reliable models. Archer et al.[24] generated a model for the CCK1 receptor based on the rhodopsin structure as a template using Insight II Homology, Discover, and Biopolymer modules (Accelrys, San Diego, CA). Docking of the CCK1 to the receptor model resulted in a

binding mode that was not compatible with the available experimental data.[24] Manivet et al.[25] created two distinct models of the 5-HT2B receptor, one using rhodopsin and the other using bacteriorhodopsin structures as templates. In their study, the bacteriorhodopsin-based model of the 5-HT2B receptor showed better agreement to available site-directed mutagenesis results than did their rhodopsin-based model.

The differences between PREDICT and rhodopsin-based homology modeling is best exemplified in the case of chemokine receptors like the CCR3 receptor that is involved in inflammatory responses. The binding of chemokines to their receptors involves mainly the receptor N-terminus and its extracellular loops. However, studies have shown that the C-terminus of the chemokines as well as small-molecule antagonists bind within the TM domain of the receptor.[148] Site-directed mutagenesis of CCR5 suggest that TAK799, a small-molecule CCR5 antagonist, binds between TMs 1, 2, 3, and 7. TM5 and TM6, which are critical for binding small biogenic amines such as dopamine and serotonin in their respective GPCRs, are less important for binding nonpeptide antagonists in the CCR5 receptor. This finding is also relevant for the CCR3 receptor because TAK799 is also known to have a high affinity for CCR3, and the human CCR3 sequence is 62% homologous to the CCR5 receptor sequence. This receptor was modeled using PREDICT methodology resulting in a 3D model that both agrees with experimental data and was successful at in silico screening. To illustrate the difference between the PREDICT model and homology modeling, we used the SWISS-MODEL 7TM interface[149] to create a rhodopsin-based homology model for this receptor. This model was subsequently minimized using the CHARMm22 force field.[78]. A simple comparison of the two models showed that the PREDICT model is in much better agreement with the mutation data. In particular, the binding pocket that has been experimentally determined to lie between TM 1, 2, 3, 6, and 7 is clearly seen in the PREDICT model, leading to its successful application in in silico screening. Conversely, the homology-based model depicts a putative binding pocket between TMs 3, 4, 5, 6, and 7, corresponding to where retinal moieties would fit in the original rhodopsin structure. This is in contradiction to the experimental data that clearly show the involvement of TMs1 and 2 in this binding site. PREDICT, which is independent of the rhodopsin structure, can identify binding pockets and structural features even when they are novel or significantly different from those seen in rhodopsin. Moreover, attempts to dock TAK-799 to a rhodopsin-based CCR3 homology model results in a very extended conformation for the ligand, such as that reported by Paterlini[150] In the PREDICT CCR3 model, on the other hand, TAK-799 adopts a conformation in which the aromatic ring of the benzyl ammonium pyran group forms π interaction with Tyr40 of TM1. In a recent study, Berkhout et al.[151] generated a rhodopsin-based homology model for the CCR2 receptor and manually docked five known CCR2 ligands to the receptor, including TAK-799. The binding modes were similar to the one proposed for the published rhodopsin-based CCR5 model of Paterlini,[150] namely extended ligand conformations. In both studies,[150,151] the ligands were docked into the model structures explicitly according to site-directed mutagenesis data. Thus, the reported agreement with the such data is not surprising. It should be noted that, unlike the PREDICT model, neither study of the homology models sought to validate their structures with virtual screening.

TM Domain

The results presented above establish the quality of the PREDICT GPCR models and justify their use for virtual screening and drug discovery. It is important to note that these models focus only on the TM domain of the receptor and include neither loops nor terminal arms, which are much more flexible and hence harder to model. Although many endogenous ligands (e.g., peptides and proteins) bind to the extracellular loops and the N-terminal, it is well established that small-molecule ligands also bind completely or partly within the TM domain.[5] A large amount of data suggest that this is the case not only with biogenic amine receptors, but also in the case of peptide receptors, which do not necessarily mimic the binding of the endogenous peptide agonist.[1] Mutational mapping of ligand-binding sites in many peptide receptors, such as in the tachykinin, angiotensin, opioid, CCK/gastrin, neurotensin, and endothelin systems (see Gether[1] for review), has located the binding of nonpeptide antagonist of many GPCRs within the TM domain. For example, mutation of residues His197(5.39) and His265(6.52) at TM5 and TM6 of the NK1 receptor to Ala did not affect the binding affinity for [125-I]-SubstanceP,[133] suggesting that these residues are not directly involved in binding of the peptide. Nonetheless, mutations of these residues are important determinants in the binding of the small molecule quinuclidine and trypthophan antagonists.[132,135] Both quinuclidine and trypthophan antagonists are competitive antagonists of Substance P, suggesting that pharmacologically competitive binding does not necessarily imply that the ligands bind to the receptor in the same way. Therefore, for the purpose of small-molecule drugs discovery the incorporation of loop conformations is of secondary importance. The apparent success of the PREDICT models in in silico screening against both biogenic amine receptors and peptide receptors validates this argument. In view of recent progress with regard to loop modeling,[138] we are currently implementing a loop structure algorithm to supplement PREDICT and be added to the final model of the TM domain.

GPCR Oligomerization

The traditional mechanism of ligand binding and signal transduction assumes that the receptors participate in these processes in the monomeric forms. It is now clear that this simplistic view does not account for the full mechanism that is associated with GPCR activation as GPCR dimmers and oligomers may play a role in signal transduction. However, the possible functional role of

receptor oligomerization is still unclear for GPCRs. Numerous studied have showed that GPCRs can form homo- and hetro-dimers and that cooperativity between GPCR binding site exists (for a review, see George et al.[152]). Countering examples such as the negative cooperativity that was demonstrated for the human bradykinin B2 receptor can be thought of as receptor desensitization, where the binding of ligand to one receptor accelerates the dissociation of another ligand from the oligomeric coupled receptor.[153] It is possible that the increased population of binding sites in a given population of receptors can create an increased gradient of receptor-mediated signaling and, therefore, oligomerization can play a role in signal amplification. However, to the extent of our knowledge, there is no evidence to date that receptor oligomerization affects receptor binding. Therefore, to date we have not introduced receptor oligomerization into the modeling procedure or into the virtual screening process. Note, however, that in principle the PREDICT algorithm can be modified for modeling the 14TM GPCR dimers, as a straightforward extension of the process used for modeling the 7TM monomers. This extension has not been done to date.

CONCLUSIONS

A new approach for modeling the receptor-bound conformation of GPCRs based on their primary sequences alone was presented in this study. Four different GPCR models were discussed, reflecting the method's capability to model biogenic amine receptors, peptide receptors, and chemokine receptors. The models were shown to agree well with available experimental data and reliably distinguish known ligands from random "drug-like" compounds using in silico screening with excellent enrichment factors of up to 44-fold better than random screening. The described methodology is almost fully automated and can be used irrespective of the level of sequence homology to other GPCRs, the existence of known 3D structures, or any other experimental data. Experimental information, when available, can be readily incorporated in the modeling process, thus optimizing the resulting models. PREDICT models have been successfully used for in silico screening, leading to the discovery of novel lead compounds, as well as in lead optimization of drug candidates, the first of which has recently started phase I clinical trials in humans (Becker et al., forthcoming).

ACKNOWLEDGMENTS

The authors acknowledge assistance from Nili Schutz, Prasad Ramakrishna, and Chris Penland.

REFERENCES

1. Gether U. Uncovering molecular mechanisms involved in activation of G protein-coupled receptors. Endocr Rev 2000;21:90–113.
2. Drews J. Drug discovery: a historical perspective. Science 2000; 287:1960–1964.
3. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Migano M. Crystal structure of rhodopsin: A G protein-coupled receptor. Science 2000;289:739–745.
4. Pebay-Peyroula E, Rummel G, Rosenbusch JP, Landau EM. X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. Science 1997;277: 1676–1681.
5. Becker OM, Shacham S, Marantz Y, Noiman S. Modeling the 3D structure of GPCRs: advances and application to drug discovery. Curr Opin Drug Discov Dev 2003;6:353–361.
6. Bissantz C, Bernard P, Hibert M, Rognan D. Protein-based virtual screening of chemical databases. II. Are homology models of G-protein coupled receptors suitable targets? Proteins 2003;50: 5–25.
7. John B, Sali A. Comparative protein structure modeling by iterative alignment, model building and model assessment. Nucleic Acids Res 2003;31:3982–3992.
8. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
9. Fiser A, Sanchez R, Melo F, Sali A. Comperative protein structure modeling. In: Becker OM, MacKerell AD, Roux B, Watanabe M, editors. Computational biochemistry and biophysics. New York: Marcel Dekker; 2001. p 275–312.
10. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le TI, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Migano M. Crystal structure of rhodopsin: A G protein-coupled receptor. Science 2000;289:739–745.
11. Probst WC, Snyder LA, Schuster DI, Brosius J, Sealfon SC. Sequence alignment of the G-protein coupled receptor superfamily. DNA Cell Biol 1992;11:1–20.
12. Du P, Salon JA, Tamm JA, Hou C, Cui W, Walker MW, Adham N, Dhanoa DS, Islam I, Vaysse PJ, Dowling B, Shifman Y, Boyle N, Rueger H, Schmidlin T, Yamaguchi Y, Branchek TA, Weinshank RL, Gluchowski C. Modeling the G-protein-coupled neuropeptide Y Y1 receptor agonist and antagonist binding sites. Protein Eng 1997;10:109–117.
13. Herzyk P, Hubbard RE. Automated method for modeling seven-helix transmembrane receptors from experimental data. Biophys J 1995;69:2419–2442.
14. Joseph MP, Maigret B, Bonnafous JC, Marie J, Scheraga HA. A computer modeling postulated mechanism for angiotensin II receptor activation. J Protein Chem 1995;14:381–398.
15. Livingstone CD, Strange PG, Naylor LH. Molecular modelling of D2-like dopamine receptors. Biochem J 1992;287:277–282.
16. Prusis P, Schioth HB, Muceniece R, Herzyk P, Afshar M, Hubbard RE, Wikberg JE. Modeling of the three-dimensional structure of the human melanocortin 1 receptor, using an automated method and docking of a rigid cyclic melanocyte-stimulating hormone core peptide. J Mol Graph Model 1997;15: 307–317.
17. Baldwin JM, Schertler GF, Unger VM. An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. J Mol Biol 1997;272:144–164.
18. Unger VM, Hargrave PA, Baldwin JM, Schertler GF. Arrangement of rhodopsin transmembrane alpha-helices. Nature 1997; 389:203–206.
19. Ballesteros JA, Shi L, Javitch JA. Structural mimicry in G protein-coupled receptors: implications of the high-resolution structure of rhodopsin for structure-function analysis of rhodopsin-like receptors. Mol Pharmacol 2001;60:1–19.
20. Strahs D, Weinstein H. Comparative modeling and molecular dynamics studies of the delta, kappa and mu opioid receptors. Protein Eng 1997;10:1019–1038.
21. Green SA, Spasoff AP, Coleman RA, Johnson M, Liggett SB. Sustained activation of a G protein-coupled receptor via "anchored" agonist binding. Molecular localization of the salmeterol exosite within the 2-adrenergic receptor. J Biol Chem 1996;271: 24029–24035.
22. Vaidehi N, Floriano WB, Trabanino R, Hall SE, Freddolino P, Choi EJ, Zamanakos G, Goddard WA, III. Prediction of structure and function of G protein-coupled receptors. Proc Natl Acad Sci USA 2002;99:12622–12627.
23. Filizola M, Perez JJ, Carteni-Farina M. BUNDLE: a program for building the transmembrane domains of G-protein-coupled receptors. J Comput Aided Mol Des 1998;12:111–118.
24. Archer E, Maigret B, Escrieut C, Pradayrol L, Fourmy D. Rhodopsin crystal: new template yielding realistic models of G-protein-coupled receptors? Trends Pharmacol Sci 2003;24:36–40.
25. Manivet P, Schneider B, Smith JC, Choi DS, Maroteaux L, Kellermann O, Launay JM. The serotonin binding site of human

and murine 5-HT2B receptors: molecular modeling and site-directed mutagenesis. J Biol Chem 2002;277:17170–17178.
26. Chung DA, Zuiderweg ER, Fowler CB, Soyer OS, Mosberg HI, Neubig RR. NMR structure of the second intracellular loop of the alpha 2A adrenergic receptor: evidence for a novel cytoplasmic helix. Biochemistry 2002;41:3596–3604.
27. Zhang L, DeHaven RN, Goodman M. NMR and modeling studies of a synthetic extracellular loop II of the kappa opioid receptor in a DPC micelle. Biochemistry 2002;41:61–68.
28. Underwood DJ, Strader CD, Rivero R, Patchett AA, Greenlee W, Prendergast K. Structural model of antagonist and agonist binding to the angiotensin II, AT1 subtype, G protein coupled receptor. Chem Biol 1994;1:211–221.
29. Mirzadegan T, Diehl F, Ebi B, Bhakta S, Polsky I, McCarley D, Mulkins M, Weatherhead GS, Lapierre JM, Dankwardt J, Morgans D, Jr, Wilhelm R, Jarnagin K. Identification of the binding site for a novel class of CCR2b chemokine receptor antagonists: binding to a common chemokine receptor motif within the helical bundle. J Biol Chem 2000;275:25562–25571.
30. Taylor WR, Jones DT, Green NM. A method for alpha-helical integral membrane protein fold prediction. Proteins 1994;18:281–294.
31. Becker OM, Shacham S, Topf M, Naor Z. A method and system for prediction of a 3D structure of a transmembrane protein. WO 02/15106.
32. Shacham S, Topf M, Avisar N, Glaser F, Marantz Y, Bar-Haim S, Noiman S, Naor Z, Becker OM. Modeling the 3D structure of GPCRs from sequence. Med Res Rev 2001;21:472–483.
33. Oshiro C, Bradley EK, Eksterowicz J, Evensen E, Lamb ML, Lanctot JK, Putta S, Stanton R, Grootenhuis PD. Performance of 3D-database molecular docking studies into homology models. J Med Chem 2004;47:764–767.
34. Shoichet BK, McGovern SL, Wei B, Irwin JJ. Lead discovery using molecular docking. Curr Opin Chem Biol 2002;6:439–446.
35. Herzyk P, Hubbard RE. A reduced representation of proteins for use in restraint satisfaction calculations. Proteins 1993;17:310–324.
36. Ponder JW, Richards FM. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. J Mol Biol 1987;193:775–791.
37. Rost B, Casadio R, Fariselli P, Sander C. Transmembrane helices predicted at 95% accuracy. Protein Sci 1995;4:521–533.
38. Hofmann K, Stoffel W. TMbase: a database of membrane spanning proteins segments. Biol Chem 1993;374:166.
39. Rost B, Fariselli P, Casadio R. Topology prediction for helical transmembrane proteins at 86% accuracy. Protein Sci 1996;5:1704–1718.
40. Baldwin JM. The probable arrangement of the helices in G protein-coupled receptors. EMBO J 1993;12:1693–1703.
41. Attwood TK, Bradley P, Flower DR, Gaulton A, Maudling N, Mitchell AL, Moulton G, Nordle A, Paine K, Taylor P, Uddin A, Zygouri C. PRINTS and its automatic supplement, prePRINTS. Nucleic Acids Res 2003;31:400–402.
42. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001;305:567–580.
43. Ballesteros JA, 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. San Diego, CA: Academic Press; 1995. p 366–428.
44. Cohen FE, Richmond TJ, Richards FM. Protein folding: evaluation of some simple rules for the assembly of helices into tertiary structures with myoglobin as an example. J Mol Biol 1979;132:275–288.
45. Gobel U, Sander C, Schneider R, Valencia A. Correlated mutations and residue contacts in proteins. Proteins 1994;18:309–317.
46. Zhou W, Flanagan C, Ballesteros JA, Konvicka K, Davidson JS, Weinstein H, Millar RP, Sealfon SC. A reciprocal mutation supports helix 2 and helix 7 proximity in the gonadotropin-releasing hormone receptor. Mol Pharmacol 1994;45:165–170.
47. Eisenberg D, Weiss RM, Terwilliger TC. The helical hydrophobic moment: a measure of the amphiphilicity of a helix. Nature 1982;299:371–374.
48. Eisenberg D, Weiss RM, Terwilliger TC. The hydrophobic moment detects periodicity in protein hydrophobicity. Proc Natl Acad Sci USA 1984;81:140–144.
49. Cronet P, Sander C, Vriend G. Modeling of transmembrane seven helix bundles. Protein Eng 1993;6:59–64.
50. Stevens TJ, Arkin IT. Are membrane proteins "inside-out" proteins? Proteins 1999;36:135–143.
51. Burley SK, Petsko GA. Amino-aromatic interactions in proteins. FEBS Lett 1986;203:139–143.
52. Aravinda S, Shamala N, Das C, Sriranjini A, Karle IL, Balaram P. Aromatic-aromatic interactions in crystal structures of helical peptide scaffolds containing projecting phenylalanine residues. J Am Chem Soc 2003;125:5308–5315.
53. Burley SK, Petsko GA. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. Science 1985;229:23–28.
54. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157:105–132.
55. Cornette JL, Cease KB, Margalit H, Spouge JL, Berzofsky JA, DeLisi C. Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. J Mol Biol 1987;195:659–685.
56. von Heijne G. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. Nature 1989;341:456–458.
57. Yeagle PL. 31P nuclear magnetic resonance studies of the phospholipid-protein interface in cell membranes. Biophys J 1982;37:227–239.
58. Dunbrack RL Jr. Rotamer libraries in the 21st century. Curr Opin Struct Biol 2002;12:431–440.
59. Vendruscolo M, Najmanovich R, Domany E. Can a pairwise contact potential stabilize native protein folds against decoys obtained by threading? Proteins 2000;38:134–148.
60. Park BH, Huang ES, Levitt M. Factors affecting the ability of energy functions to discriminate correct from incorrect folds. J Mol Biol 1997;266:831–846.
61. Park B, Levitt M. Energy functions that discriminate X-ray and near native folds from well-constructed decoys. J Mol Biol 1996;258:367–392.
62. Miyazawa S, Jernigan RL. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J Mol Biol 1996;256:623–644.
63. Gregoret LM, Cohen FE. Novel method for the rapid evaluation of packing in protein structures. J Mol Biol 1990;211:959–974.
64. Meurisse R, Brasseur R, Thomas A. Aromatic side-chain interactions in proteins. Near- and far-sequence His-X pairs. Biochim Biophys Acta 2003;1649:85–96.
65. Shi L, Liapakis G, Xu R, Guarnieri F, Ballesteros JA, Javitch JA. Beta2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch. J Biol Chem 2002;277:40989–40996.
66. Gromiha MM. Influence of cation-pi interactions in different folding types of membrane proteins. Biophys Chem 2003;103:251–258.
67. Zhou FX, Cocco MJ, Russ WP, Brunger AT, Engelman DM. Interhelical hydrogen bonding drives strong interactions in membrane proteins. Nat Struct Biol 2000;7:154–160.
68. Cosson P, Lankford SP, Bonifacino JS, Klausner RD. Membrane protein association by potential intramembrane charge pairs. Nature 1991;351:414–416.
69. Segrest JP, De Loof H, Dohlman JG, Brouillette CG, Anantharamaiah GM. Amphipathic helix motif: classes and properties. Proteins 1990;8:103–117.
70. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equations of state calculations by fast computing machine. J Chem Phys 1953;21:1087–1091.
71. Eilers M, Shekar SC, Shieh T, Smith SO, Fleming PJ. Internal packing of helical membrane proteins. Proc Natl Acad Sci U S A 2000;97:5796–5801.
72. Lin JC, Parrish W, Eilers M, Smith SO, Konopka JB. Aromatic residues at the extracellular ends of transmembrane domains 5 and 6 promote ligand activation of the G protein-coupled alpha-factor receptor. Biochemistry 2003;42:293–301.
73. Prioleau C, Visiers I, Ebersole BJ, Weinstein H, Sealfon SC. Conserved helix 7 tyrosine acts as a multistate conformational switch in the 5HT2C receptor. Identification of a novel "locked-on" phenotype and double revertant mutations. J Biol Chem 2002;277:36577–36584.

74. Chothia C. Principles that determine the structure of proteins. Annu Rev Biochem 1984;53:537–572.
75. Bowie JU. Helix packing in membrane proteins. J Mol Biol 1997;272:780–789.
76. Spencer RH, Rees DC. The alpha-helix and the organization and gating of channels. Annu Rev Biophys Biomol Struct 2002;31:207–233.
77. Insight II, Version 2000.1. San Diego: Accelrys Inc.
78. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983; 4:187–217.
79. Hwang MJ, Stockfisch TP, Hagler AT. Derivation and characterization of a class II force field, CFF93, for the alkyl functional group and alkane molecules. J Am Chem Soc 1994;116:2515–2525.
80. Maple JR, Hwang MJ, Jalkanen KJ, Stockfisch TP. Derivation of class II force fields. V. Quantum force field for amides. peptides, and related compounds. J Comp Chem 2003;19:458.
81. Visiers I, Ballesteros JA, Weinstein H. Three-dimensional representations of G protein-coupled receptor structures and mechanisms. Methods Enzymol 2002;343:329–371.
82. Visiers I, Braunheim BB, Weinstein H. Prokink: a protocol for numerical evaluation of helix distortions by proline. Protein Eng 2000;13:603–606.
83. Bachar M, Becker OM. Protein-induced membrane disorder: a molecular dynamics study of melittin in a dipalmitoylphosphatidylcholine bilayer. Biophys J 2000;78:1359–1375.
84. Bachar M, Becker OM. Melittin at a membrane/water interface: effects on water orientation and water penetration. J Chem Phys 1999;111:8672–8685.
85. CONCORD, Version 4.04 Tripos Inc. St. Louis, MO 63144.
86. SYBYL, Version 6.8. Tripos Inc. St. Louis, MO 63144.
87. ACD v.2001-1, MDL Information Systems Inc. San Leandro, CA 94577.
88. LeadQuest, Tripos Inc. St. Louis, MO 63144
89. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 1997;23:3–25.
90. DOCK, Version 4.0. San Francisco: Molecular Design Institute, University of California.
91. CONFORT, Version 4.1 Tripos Inc. St. Louis, MO 63144.
92. Shi L, Javitch JA. The binding site of aminergic G protein-coupled receptors: the transmembrane segments and second extracellular loop. Annu Rev Pharmacol Toxicol 2002;42:437–467.
93. Piela L, Nemethy G, Scheraga HA. Proline-induced constraints in alpha-helices. Biopolymers 1987;26:1587–1600.
94. Kumar S, Bansal M. Geometrical and sequence characteristics of alpha-helices in globular proteins. Biophys J 1998;75:1935–1944.
95. Gray TM, Matthews BW. Intrahelical hydrogen bonding of serine, threonine and cysteine residues within alpha-helices and its relevance to membrane-bound proteins. J Mol Biol 1984;175: 75–81.
96. Xu W, Ozdener F, Li JG, Chen C, de Riel JK, Weinstein H, Liu-Chen LY. Functional role of the spatial proximity of Asp114(2.50) in TMH 2 and Asn332(7.49) in TMH 7 of the mu opioid receptor. FEBS Lett 1999;447:318–324.
97. Flanagan CA, Zhou W, Chi L, Yuen T, Rodic V, Robertson D, Johnson M, Holland P, Millar RP, Weinstein H, Mitchell R, Sealfon SC. The functional microdomain in transmembrane helices 2 and 7 regulates expression, activation, and coupling pathways of the gonadotropin-releasing hormone receptor. J Biol Chem 1999;274:28880–28886.
98. Sealfon SC, Chi L, Ebersole BJ, Rodic V, Zhang D, Ballesteros JA, Weinstein H. Related contribution of specific helix 2 and 7 residues to conformational activation of the serotonin 5-HT2A receptor. J Biol Chem 1995;270:16683–16688.
99. Javitch JA, Fu D, Chen J. Residues in the fifth membrane-spanning segment of the dopamine D2 receptor exposed in the binding-site crevice. Biochemistry 1995;34:16433–16439.
100. Cox BA, Henningsen RA, Spanoyannis A, Neve RL, Neve KA. Contributions of conserved serine residues to the interactions of ligands with dopamine D2 receptors. J Neurochem 1992;59:627–635.
101. Shi L, Simpson MM, Ballesteros JA, Javitch JA. The first transmembrane segment of the dopamine D2 receptor: accessibility in the binding-site crevice and position in the transmembrane bundle. Biochemistry 2001;40:12339–12348.
102. Javitch JA, Ballesteros JA, Weinstein H, Chen J. A cluster of aromatic residues in the sixth membrane-spanning segment of the dopamine D2 receptor is accessible in the binding-site crevice. Biochemistry 1998;37:998–1006.
103. Bryant SH, Altschul SF. Statistics of sequence-structure threading. Curr Opin Struct Biol 1995;5:236–244.
104. Bowie JU, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. Science 1991;253:164–170.
105. Becker OM. Principal coordinate maps of molecular potential energy surfaces. J Comput Chem 1998;19:1255–1267.
106. Javitch JA, Li X, Kaback J, Karlin A. A cysteine residue in the third membrane-spanning segment of the human D2 dopamine receptor is exposed in the binding-site crevice. Proc Natl Acad Sci USA 1994;91:10355–10359.
107. Simpson MM, Ballesteros JA, Chiappa V, Chen J, Suehiro M, Hartman DS, Godel T, Snyder LA, Sakmar TP, Javitch JA. Dopamine D4/D2 receptor selectivity is determined by A divergent aromatic microdomain contained within the second, third, and seventh membrane-spanning segments. Mol Pharmacol 1999;56:1116–1126.
108. Strader CD, Sigal IS, Candelore MR, Rands E, Hill WS, Dixon RA. Conserved aspartic acid residues 79 and 113 of the beta-adrenergic receptor have different roles in receptor function. J Biol Chem 1988;263:10267–10271.
109. Trumpp-Kallmeyer S, Hoflack J, Bruinvels A, Hibert M. Modeling of G-protein-coupled receptors: application to dopamine, adrenaline, serotonin, acetylcholine, and mammalian opsin receptors. J Med Chem 1992;35:3448–3462.
110. Mansour A, Meng F, Meador-Woodruff JH, Taylor LP, Civelli O, Akil H. Site-directed mutagenesis of the human dopamine D2 receptor. Eur J Pharmacol 1992;227:205–214.
111. Ballesteros JA, Deupi X, Olivella M, Haaksma EE, Pardo L. Serine and threonine residues bend alpha-helices in the chi(1) = g(-) conformation. Biophys J 2000;79:2754–2760.
112. Savarese TM, Wang CD, Fraser CM. Site-directed mutagenesis of the rat m1 muscarinic acetylcholine receptor. Role of conserved cysteines in receptor function. J Biol Chem 1992;267: 11439–11448.
113. Guo W, Shi L, Javitch JA. The fourth transmembrane segment forms the interface of the dopamine D2 receptor homodimer. J Biol Chem 2003;278:4385–4388.
114. Liang Y, Fotiadis D, Filipek S, Saperstein DA, Palczewski K, Engel A. Organization of the G protein-coupled receptors rhodopsin and opsin in native membranes. J Biol Chem 2003;278:21655–21662.
115. Farrens DL, Altenbach C, Yang K, Hubbell WL, Khorana HG. Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin. Science 1996;274:768–770.
116. Ghanouni P, Steenhuis JJ, Farrens DL, Kobilka BK. Agonist-induced conformational changes in the G-protein-coupling domain of the beta 2 adrenergic receptor. Proc Natl Acad Sci USA 2001;98:5997–6002.
117. Huang P, Li J, Chen C, Visiers I, Weinstein H, Liu-Chen LY. Functional role of a conserved motif in TM6 of the rat mu opioid receptor: constitutively active and inactive receptors result from substitutions of Thr6.34(279) with Lys and Asp. Biochemistry 2001;40:13501–13509.
118. Huang P, Visiers I, Weinstein H, Liu-Chen LY. The local environment at the cytoplasmic end of TM6 of the mu opioid receptor differs from those of rhodopsin and monoamine receptors: introduction of an ionic lock between the cytoplasmic ends of helices 3 and 6 by a L6.30(275)E mutation inactivates the mu opioid receptor and reduces the constitutive activity of its T6.34(279)K mutant. Biochemistry 2002;41:11972–11980.
119. Zhou W, Rodic V, Kitanovic S, Flanagan CA, Chi L, Weinstein H, Maayani S, Millar RP, Sealfon SC. A locus of the gonadotropin-releasing hormone receptor that differentiates agonist and antagonist binding sites. J Biol Chem 1995;270:18853–18857.
120. Lu ZL, Saldanha JW, Hulme EC. Transmembrane domains 4 and 7 of the M(1) muscarinic acetylcholine receptor are critical for ligand binding and the receptor activation switch. J Biol Chem 2001;276:34098–34104.
121. Elling CE, Thirstrup K, Holst B, Schwartz TW. Conversion of

agonist site to metal-ion chelator site in the beta(2)-adrenergic receptor. Proc Natl Acad Sci U S A 1999;96:12322–12327.
122. Teller DC, Okada T, Behnke CA, Palczewski K, Stenkamp RE. Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs). Biochemistry 2001;40:7761–7772.
123. Cerda-Reverter JM, Martinez-Rodriguez G, Zanuy S, Carrillo M, Larhammar D. Molecular evolution of the neuropeptide Y (NPY) family of peptides: cloning of three NPY-related peptides from the sea bass (Dicentrarchus labrax). Regul Pept 2000;95:25–34.
124. Larhammar D. Structural diversity of receptors for neuropeptide Y, peptide YY and pancreatic polypeptide. Regul Pept 1996;65: 165–174.
125. Parker SL, Crowley WR. Central stimulation of oxytocin release in the lactating rat: interaction of neuropeptide Y with alpha-1-adrenergic mechanisms. Endocrinology 1993;132:658–666.
126. Kalra SP, Fuentes M, Fournier A, Parker SL, Crowley WR. Involvement of the Y-1 receptor subtype in the regulation of luteinizing hormone secretion by neuropeptide Y in rats. Endocrinology 1992;130:3323–3330.
127. Kannoa T, Kanatani A, Keen SL, Arai-Otsuki S, Haga Y, Iwama T, Ishihara A, Sakuraba A, Iwaasa H, Hirose M, Morishima H, Fukami T, Ihara M. Different binding sites for the neuropeptide Y Y1 antagonists 1229U91 and J-104870 on human Y1 receptors. Peptides 2001;22:405–413.
128. Sautel M, Rudolf K, Wittneben H, Herzog H, Martinez R, Munoz M, Eberlein W, Engel W, Walker P, Beck-Sickinger AG. Neuropeptide Y and the nonpeptide antagonist BIBP 3226 share an overlapping binding site at the human Y1 receptor. Mol Pharmacol 1996;50:285–292.
129. Schertler GF, Villa C, Henderson R. Projection structure of rhodopsin. Nature 1993;362:770–772.
130. Sautel M, Martinez R, Munoz M, Peitsch MC, Beck-Sickinger AG, Walker P. Role of a hydrophobic pocket of the human Y1 neuropeptide Y receptor in ligand binding. Mol Cell Endocrinol 1995;112:215–222.
131. Severini C, Improta G, Falconieri-Erspamer G, Salvadori S, Erspamer V. The tachykinin peptide family. Pharmacol Rev 2002;54:285–322.
132. Fong TM, Huang RR, Yu H, Strader CD. Mapping the ligand binding site of the NK-1 receptor. Regul Pept 1993;46:43–48.
133. Cascieri MA, Shiao LL, Mills SG, MacCoss M, Swain CJ, Yu H, Ber E, Sadowski S, Wu MT, Strader CD, et al. Characterization of the interaction of diacylpiperazine antagonists with the human neurokinin-1 receptor: identification of a common binding site for structurally dissimilar antagonists. Mol Pharmacol 1995; 47:660–665.
134. Holst B, Zoffmann S, Elling CE, Hjorth SA, Schwartz TW. Steric hindrance mutagenesis versus alanine scan in mapping of ligand binding sites in the tachykinin NK1 receptor. Mol Pharmacol 1998;53:166–175.
135. Fong TM, Yu H, Cascieri MA, Underwood D, Swain CJ, Strader CD. The role of histidine 265 in antagonist binding to the neurokinin-1 receptor. J Biol Chem 1994;269:2728–2732.
136. Gether U, Johansen TE, Snider RM, Lowe JA, III, Nakanishi S, Schwartz TW. Different binding epitopes on the NK1 receptor for substance P and non-peptide antagonist. Nature 1993;362:345–348.
137. Gether U, Emonds-Alt X, Breliere JC, Fujii T, Hagiwara D, Pradier L, Garret C, Johansen TE, Schwartz TW. Evidence for a common molecular mode of action for chemically distinct nonpeptide antagonists at the neurokinin-1 (substance P) receptor. Mol Pharmacol 1994;45:500–508.
138. Snider RM, Constantine JW, Lowe JA, III, Longo KP, Lebel WS, Woody HA, Drozda SE, Desai MC, Vinick FJ, Spencer RW, et al. A potent nonpeptide antagonist of the substance P (NK1) receptor. Science 1991;251:435–437.
139. Fong TM, Yu H, Strader CD. Molecular basis for the species selectivity of the neurokinin-1 receptor antagonists CP-96,345 and RP67580. J Biol Chem 1992;267:25668–25671.
140. Fong TM, Huang RR, Strader CD. Localization of agonist and antagonist binding domains of the human neurokinin-1 receptor. J Biol Chem 1992;267:25664–25667.
141. Fong TM, Yu H, Huang RR, Strader CD. The extracellular domain of the neurokinin-1 receptor is required for high-affinity binding of peptides. Biochemistry 1992;31:11806–11811.
142. Elling CE, Nielsen SM, Schwartz TW. Conversion of antagonist-binding site to metal-ion site in the tachykinin NK-1 receptor. Nature 1995;374:74–77.
143. Yang AS, Honig B. An integrated approach to the analysis and modeling of protein sequences and structures. I. Protein structural alignment and a quantitative measure for protein structural distance. J Mol Biol 2000;301:665–678.
144. Wilson CA, Kreychman J, Gerstein M. Assessing annotation transfer for genomics: quantifying the relations between protein sequence, structure and function through traditional and probabilistic scores. J Mol Biol 2000;297:233–249.
145. Yang AS, Honig B. An integrated approach to the analysis and modeling of protein sequences and structures. II. On the relationship between sequence and structural similarity for proteins that are not obviously related in sequence. J Mol Biol 2000;301:679–689.
146. Lu ZL, Saldanha JW, Hulme EC. Seven-transmembrane receptors: crystals clarify. Trends Pharmacol Sci 2002;23:140–146.
147. Shin N, Coates E, Murgolo NJ, Morse KL, Bayne M, Strader CD, Monsma FJ, Jr. Molecular modeling and site-specific mutagenesis of the histamine-binding site of the histamine H4 receptor. Mol Pharmacol 2002;62:38–47.
148. Ye J, Kohli LL, Stone MJ. Characterization of binding between the chemokine eotaxin and peptides derived from the chemokine receptor CCR3. J Biol Chem 2000;275:27250–27257.
149. SWISS-MODEL 7TM. 2003. Retrieved Internet Communication. http://www.expasy.org/swissmod/SM_GPCRInformation.html
150. Paterlini MG. Structure modeling of the chemokine receptor CCR5: implications for ligand binding and selectivity. Biophys J 2002;83:3012–3031.
151. Berkhout TA, Blaney FE, Bridges AM, Cooper DG, Forbes IT, Gribble AD, Groot PH, Hardy A, Ife RJ, Kaur R, Moores KE, Shillito H, Willetts J, Witherington J. CCR2: Characterization of the antagonist binding site from a combined receptor modeling/mutagenesis approach. J Med Chem 2003;46:4070–4086.
152. George SR, O'Dowd BF, Lee SP. G-protein-coupled receptor oligomerization and its potential for drug discovery. Nat Rev Drug Discov 2002;1:808–820.
153. Pizard A, Marchetti J, Allegrini J, Alhenc-Gelas F, Rajerison RM. Negative cooperativity in the human bradykinin B2 receptor. J Biol Chem 1998;273:1309–1315.
154. Freedman SB, Patel S, Marwood R, Emms F, Seabrook GR, Knowles MR, McAllister G. Expression and pharmacological characterization of the human D3 dopamine receptor. J Pharmacol Exp Ther 1994;268:417–426.
155. Sautel F, Griffon N, Levesque D, Pilon C, Schwartz JC, Sokoloff P. A functional test identifies dopamine agonists selective for D3 versus D2 receptors. Neuroreport 1995;6:329–332.
156. Sokoloff P, Giros B, Martres MP, Andrieux M, Besancon R, Pilon C, Bouthenet ML, Souil E, Schwartz JC. Localization and function of the D3 dopamine receptor. Arzneimittelforschung 1992;42:224–230.
157. Seeman P, Corbett R, Van Tol HH. Atypical neuroleptics have low affinity for dopamine D2 receptors or are selective for D4 receptors. Neuropsychopharmacology 1997;16:93–110.
158. Lejeune F, Newman-Tancredi A, Audinot V, Millan MJ. Interactions of (+)- and (−)-8- and 7-hydroxy-2-(di-n-propylamino)tetralin at human (h)D3, hD2 and h serotonin1A receptors and their modulation of the activity of serotoninergic and dopaminergic neurones in rats. J Pharmacol Exp Ther 1997;280:1241–1249.
159. Richelson E, Nelson A. Antagonism by neuroleptics of neurotransmitter receptors of normal human brain in vitro. Eur J Pharmacol 1984;103:197–204.
160. Cusack B, Nelson A, Richelson E. Binding of antidepressants to human brain receptors: focus on newer generation compounds. Psychopharmacology (Berl) 1994;114:559–565.
161. Seeman P, Van Tol HH. Deriving the therapeutic concentrations for clozapine and haloperidol: the apparent dissociation constant of a neuroleptic at the dopamine D2 or D4 receptor varies with the affinity of the competing radioligand. Eur J Pharmacol 1995;291:59–66.
162. Kongsamut S, Roehr JE, Cai J, Hartman HB, Weissensee P, Kerman LL, Tang L, Sandrasagra A. Iloperidone binding to human and rat dopamine and 5-HT receptors. Eur J Pharmacol. 1996 Dec 19;317(2–3):417–423.
163. Vallone D, Picetti R, Borrelli E. Structure and function of dopamine receptors. Neurosci Biobehav Rev 2000;24:125–132.
164. Millan MJ, Maiofiss L, Cussac D, Audinot V, Boutin JA, Newman-

Tancredi A. Differential actions in antiparkinson agents at multiple classes of monoaminergic receptor. I. A multivariate analysis of the binding profiles of 14 drugs at 21 native and cloned human receptor subtypes. J Pharmacol Exp Ther 2002;303: 791–804.
165. Wieland HA, Engel W, Eberlein W, Rudolf K, Doods HN. Subtype selectivity of the novel nonpeptide neuropeptide Y Y1 receptor antagonist BIBO 3304 and its effect on feeding in rodents. Br J Pharmacol 1998;125:549–555.
166. Zarrinmayeh H, Zimmerman DM, Cantrell BE, Schober DA, Bruns RF, Gackenheimer SL, Ornstein PL, Hipskind PA, Britton TC, Gehlert DR. Structure-activity relationship of a series of diaminoalkyl substituted benzimidazole as neuropeptide Y Y1 receptor antagonists. Bioorg Med Chem Lett 1999;9:647–652.
167. Zarrinmayeh H, Nunes AM, Ornstein PL, Zimmerman DM, Arnold MB, Schober DA, Gackenheimer SL, Bruns RF, Hipskind PA, Britton TC, Cantrell BE, Gehlert DR. Synthesis and evaluation of a series of novel 2-[(4-chlorophenoxy)methyl]benzimidazoles as selective neuropeptide Y Y1 receptor antagonists. J Med Chem 1998;41:2709–2719.
168. Hipskind PA, Lobb KL, Nixon JA, Britton TC, Bruns RF, Catlow J, Dieckman-McGinty DK, Gackenheimer SL, Gitter BD, Iyengar S, Schober DA, Simmons RM, Swanson S, Zarrinmayeh H, Zimmerman DM, Gehlert DR. Potent and selective 1,2,3-trisubstituted indole NPY Y-1 antagonists. J Med Chem 1997;40: 3712–3714.
169. Zhang W, Lundberg JM, Thoren P. Neuropeptide Y Y1 receptor antagonist (BIBP 3226) attenuates stress evoked tachycardia in conscious spontaneously hypertensive rats. Cardiovasc Drugs Ther 1997;11:801–806.
170. Kannoa T, Kanatani A, Keen SL, Arai-Otsuki S, Haga Y, Iwama T, Ishihara A, Sakuraba A, Iwaasa H, Hirose M, Morishima H, Fukami T, Ihara M. Different binding sites for the neuropeptide Y Y1 antagonists 1229U91 and J-104870 on human Y1 receptors. Peptides 2001;22:405–413.
171. Kanatani A, Hata M, Mashiko S, Ishihara A, Okamoto O, Haga Y, Ohe T, Kanno T, Murai N, Ishii Y, Fukuroda T, Fukami T, Ihara M. A typical Y1 receptor regulates feeding behaviors: effects of a potent and selective Y1 antagonist, J-115814. Mol Pharmacol 2001;59:501–505.
172. Lecklin A, Lundell I, Paananen L, Wikberg JE, Mannisto PT, Larhammar D. Receptor subtypes Y1 and Y5 mediate neuropeptide Y induced feeding in the guinea-pig. Br J Pharmacol 2002;135:2029–2037.
173. Hong Y, Gregor V, Ling AL, Tompkins EV, Porter J, Chou TS, Paderes G, Peng Z, Hagaman C, Anderes K, Luthin D, May J. Synthesis and biological evaluation of novel guanylurea compounds as potent NPY Y1 receptor antagonist. 217th ACS National Meeting (Anaheim, Spring 1999). Abstract MEDI-108.
174. Tattersall FD, Rycroft W, Francis B, Pearce D, Merchant K, MacLeod AM, Ladduwahetty T, Keown L, Swain C, Baker R, Cascieri M, Ber E, Metzger J, MacIntyre DE, Hill RG, Hargreaves RJ. Tachykinin NK1 receptor antagonists act centrally to inhibit emesis induced by the chemotherapeutic agent cisplatin in ferrets. Neuropharmacology 1996;35:1121–1129.
175. Beattie DT, Beresford IJ, Connor HE, Marshall FH, Hawcock AB, Hagan RM, Bowers J, Birch PJ, Ward P. The pharmacology of GR203040, a novel, potent and selective non-peptide tachykinin NK1 receptor antagonist. Br J Pharmacol 1995;116:3149–3157.
176. Dinnell K, Chicchi GG, Dhar MJ, Elliott JM, Hollingworth GJ, Kurtz MM, Ridgill MP, Rycroft W, Tsao KL, Williams AR, Swain CJ. 2-Aryl indole NK1 receptor antagonists: optimisation of the 2-aryl ring and the indole nitrogen substituent. Bioorg Med Chem Lett 2001;11:1237–1240.
177. Walpole C, Ko SY, Brown M, Beattie D, Campbell E, Dickenson F, Ewan S, Hughes GA, Lemaire M, Lerpiniere J, Patel S, Urban L. 2-Nitrophenylcarbamoyl-(S)-prolyl-(S)-3-(2-naphthyl)alanyl-N-benzyl-N - methylamide (SDZ NKT 343), a potent human NK1 tachykinin receptor antagonist with good oral analgesic activity in chronic pain models. J Med Chem 1998;41:3159–3173.
178. Boyle S, Guard S, Higginbottom M, Horwell DC, Howson W, McKnight AT, Martin K, Pritchard MC, O'Toole J, Raphy J, et al. Rational design of high affinity tachykinin NK1 receptor antagonists. Bioorg Med Chem 1994;2:357–370.
179. Emonds-Alt X, Doutremepuich JD, Heaulme M, Neliat G, Santucci V, Steinberg R, Vilain P, Bichon D, Ducoux JP, Proietto V, et al. In vitro and in vivo biological activities of SR140333, a novel potent non-peptide tachykinin NK1 receptor antagonist. Eur J Pharmacol 1993;250:403–413.
180. Harrison T, Owens AP, Williams BJ, Swain CJ, Williams A, Carlson EJ, Rycroft W, Tattersall FD, Cascieri MA, Chicchi GG, Sadowski S, Rupniak NM, Hargreaves RJ. An orally active, water-soluble neurokinin-1 receptor antagonist suitable for both intravenous and oral clinical administration. J Med Chem 2001;44:4296–4299.
181. Stevenson GI, Huscroft I, MacLeod AM, Swain CJ, Cascieri MA, Chicchi GG, Graham MI, Harrison T, Kelleher FJ, Kurtz M, Ladduwahetty T, Merchant KJ, Metzger JM, MacIntyre DE, Sadowski S, Sohal B, Owens AP. 4,4-Disubstituted piperidine high-affinity NK1 antagonists: structure-activity relationships and in vivo activity. J Med Chem 1998;41:4623–4635.
182. Natsugari H, Ikeura Y, Kamo I, Ishimaru T, Ishichi Y, Fujishima A, Tanaka T, Kasahara F, Kawada M, Doi T. Axially chiral 1,7-naphthyridine-6-carboxamide derivatives as orally active tachykinin NK(1) receptor antagonists: synthesis, antagonistic activity, and effects on bladder functions. J Med Chem 1999;42: 3982–3993.
183. Ashwood VA, Field MJ, Horwell DC, Julien-Larose C, Lewthwaite RA, McCleary S, Pritchard MC, Raphy J, Singh L. Utilization of an intramolecular hydrogen bond to increase the CNS penetration of an NK(1) receptor antagonist. J Med Chem 2001;44:2276–2285.
184. Ikeura Y, Ishichi Y, Tanaka T, Fujishima A, Murabayashi M, Kawada M, Ishimaru T, Kamo I, Doi T, Natsugari H. Axially chiral N-benzyl-N,7-dimethyl-5-phenyl-1, 7-naphthyridine-6-carboxamide derivatives as tachykinin NK1 receptor antagonists: determination of the absolute stereochemical requirements. J Med Chem 1998;41:4232–4239.
185. Papp M, Vassout A, Gentsch C. The NK1-receptor antagonist NKP608 has an antidepressant-like effect in the chronic mild stress model of depression in rats. Behav Brain Res 2000;115:19–23.
186. Kudlacz EM, Shatzer SA, Knippenberg RW, Logan DE, Poirot M, van Giersbergen PL, Burkholder TP. In vitro and in vivo characterization of MDL 105,212A, a nonpeptide NK-1/NK-2 tachykinin receptor antagonist. J Pharmacol Exp Ther 1996;277: 840–851.
187. Varady J, Wu X, Fang X, Min J, Hu Z, Levant B, Wang S. Molecular modeling of the three-dimensional structure of dopamine 3 (D3) subtype receptor: discovery of novel and potent D3 ligands through a hybrid pharmacophore- and structure-based database searching approach. J Med Chem 2003;21:4377–4392.