JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

An Automated System for the Analysis of G Protein-Coupled Receptor Transmembrane Binding Pockets: Alignment, Receptor-Based Pharmacophores, and Their Application

Nicole A. Kratochwil, Pari Malherbe, Lothar Lindemann, Martin Ebeling, Marius C. Hoener, Andreas Mhlemann, Richard H. P. Porter, Martin Stahl, and Paul R. Gerber

J. Chem. Inf. Model., **2005**, 45 (5), 1324-1336• DOI: 10.1021/ci050221u • Publication Date (Web): 03 September 2005

An Automated System for the Analysis of G Protein-Coupled Receptor Transmembrane Binding Pockets: Alignment, Receptor-Based Pharmacophores, and Their Application

Nicole A. Kratochwil,* Pari Malherbe, Lothar Lindemann, Martin Ebeling, Marius C. Hoener, Andreas Mühlemann, Richard H. P. Porter, Martin Stahl, and Paul R. Gerber

Pharmaceuticals Division, F. Hoffmann-La Roche Ltd, CH-4070 Basel, Switzerland

Received May 30, 2005

G protein-coupled receptors (GPCRs) share a common architecture consisting of seven transmembrane (TM) domains. Various lines of evidence suggest that this fold provides a generic binding pocket within the TM region for hosting agonists, antagonists, and allosteric modulators. Here, a comprehensive and automated method allowing fast analysis and comparison of these putative binding pockets across the entire GPCR family is presented. The method relies on a robust alignment algorithm based on conservation indices, focusing on pharmacophore-like relationships between amino acids. Analysis of conservation patterns across the GPCR family and alignment to the rhodopsin X-ray structure allows the extraction of the amino acids lining the TM binding pocket in a so-called ligand binding pocket vector (LPV). In a second step, LPVs are translated to simple 3D receptor pharmacophore models, where each amino acid is represented by a single spherical pharmacophore feature and all atomic detail is omitted. Applications of the method include the assessment of selectivity issues, support of mutagenesis studies, and the derivation of rules for focused screening to identify chemical starting points in early drug discovery projects. Because of the coarseness of this 3D receptor pharmacophore model, however, meaningful scoring and ranking procedures of large sets of molecules are not justified. The LPV analysis of the trace amine-associated receptor family and its experimental validation is discussed as an example. The value of the 3D receptor model is demonstrated for a class C GPCR family, the metabotropic glutamate receptors.

INTRODUCTION

The first high-resolution X-ray crystal structure of a class A member of the G protein-coupled receptor (GPCR)[1−16] family, bovine rhodopsin, was reported in the year 2000,[17−20] heralding a new era of structure-based GPCR modeling. Earlier homology models based on the 3D structure of bacteriorhodopsin, a non-GPCR seven-transmembrane (7TM) bacterial retinal protein, were revisited.[21,22] In the meantime, it has become common practice to model class A GPCRs on the basis of the X-ray crystal structure of bovine rhodopsin and to refine the generated homology models using extensive molecular mechanics/molecular dynamics protocols.[23,24] To obtain high-quality homology models suitable for docking experiments, additional, work-intensive refinement cycles are needed that explicitly take into account binding studies with highly potent ligands.[7,21,23,25−35]

This kind of approach is very valuable for optimizations of medicinal chemistry series in advanced projects on well-characterized targets, where mutational data and structure−activity relationships for the ligands are available and can be incorporated into validation cycles of the models. In early-stage GPCR projects, however, where often only the natural ligands are known and data from site-directed mutagenesis are missing, a traditional GPCR homology model with its focus on atomic details may turn out to be of only limited use. At this early stage of the drug discovery process, the emphasis is on finding chemical tools for establishing in vitro receptor assays, assessing species differences and selectivity issues, and cross-linking information within GPCR gene families.

GPCRs from different subfamilies share several common features. The binding site of *cis*-retinal in the rhodopsin structure is located in a pocket within the TM domain of the protein. This ligand binding pocket, formed by an interplay of the 7TM helices and by part of the extracellular loop 2 (EC2), is likely to be part of the general GPCR architecture, and it is assumed that this conserved 7TM architecture of GPCRs provides a generic binding pocket within the transmembrane region for hosting agonists, antagonists, and allosteric modulators. Following this simplistic view of GPCR architecture, one expects to find both well-conserved sequence positions required for the common overall 3D structure and highly variable sequence positions responsible for the interactions with the wide range of chemically different GPCR ligands and effectors. Indeed, clear preferences for certain TM helix and loop positions to be involved in ligand binding have been repeatedly detected across the GPCR family, and many of them have been shown to correspond to positions in close proximity to *cis*-retinal, an inverse agonist of rhodopsin, in the bovine rhodopsin crystal structure. By means of a combination of entropy and variability measures, and correlation analysis of the GPCR sequences, Oliveira et al.[36] showed that several highly variable sequence positions, when mapped onto the X-ray crystal structure of bovine rhodopsin, were found to be close to the binding site of *cis*-retinal. Bondensgaard et al.[37] measured the degree of variability in the 7TM domain of

* To whom correspondence should be addressed. Phone: +41-61-6872079. Fax: +41-61-6887408. E-mail: nicole_a.kratochwil@roche.com.

10.1021/ci050221u CCC: $30.25
Published on Web 09/03/2005

111 class A GPCRs by employing a similar, entropy-based method. For the ligand binding site, they report several highly variable residues on TM3 and TM5, close to the ligand binding pocket, whereas more conserved ones are found in a subpocket spanned by positions between TM5 and TM6. Detailed information about the population of amino acids at conserved and variable helix positions was compiled by Mirzadegan et al.,[38] who compared the sequences of 270 GPCR receptors and listed the frequencies of the 20 natural amino acids at the different 7TM helix positions. In addition, a statistical analysis of the GRAP mutant database[39] shows that TM helix positions with high variability are among the most frequently studied sequence positions by site-directed mutational analysis and are, in addition, very often cited to be involved in ligand binding and activation.[22,40−43]

We investigated if an in-depth sequence analysis of only the transmembrane ligand binding pocket of GPCRs (as opposed to standard approaches based on the complete receptor sequences) would allow the gain of comparable or even new insights into in vitro GPCR pharmacology. We developed a fast, automated procedure for extracting from a GPCR sequence the transmembrane ligand binding pocket, an amino acid vector consisting of those amino acids most likely to be involved in ligand interactions. This vector we refer to as a ligand pocket vector (LPV). We show how to analyze and interpret LPVs and apply the technique to explain in vitro pharmacology results, using the trace amine-associated receptor family as an example.

To extend the purely sequence-based approach so that it would support elementary analyses in molecular design, we established an automated method to translate LPVs into a 3D receptor pharmacophore model. These models allow a visual analysis of the general ligand binding properties of any GPCR without the need to concentrate on atomic details of side chain orientations. The aim of this approach is to generate ideas about possible ligand pharmacophores, which represent the complement to the generated 3D receptor pharmacophores, and to set up rules for the selection of compounds to be screened in focused screening campaigns.

We present an example of the 3D receptor pharmacophore model applied to a class C GPCR family, the metabotropic glutamate receptors, and point out a possible way of converting these 3D pharmacophore models into a full homology model.

The approach presented here is based on a simplistic view of GPCR receptors. The following limitations, valid for all current GPCR modeling approaches,[33] are explicitly acknowledged.

First, we assume that the conserved 7TM architecture of GPCRs provides a generic binding pocket within the TM region for hosting agonists, antagonists, and allosteric modulators. Second, the often very low sequence identity between bovine rhodopsin and the target GPCR receptors, typically less than 30%, is a further challenge despite the conserved 7TM architecture. The existing patterns of highly conserved residues in GPCRs, however, help to guide the alignments, so this is less of a constraint for the method developed here than it would be for a full modeling approach. Third, the availability of only one X-ray crystal structure of a GPCR receptor is a severe restriction. It would be unreasonable to try and characterize a whole protein family and to fully understand and predict ligand−protein interactions on the basis of just one structure. Also, rhodopsin is a special case among GPCRs in that it covalently binds its ligand. It is the inactive form of the receptor that is represented by the X-ray structure, with no direct hints of what the active form might look like. Finally, recent biochemical and biophysical studies supports a mechanistic view of signal transduction wherein the receptor functions as an oligomer.[44] The cooperative interactions within such an oligomeric array may be critical for the propagation of an external signal across the cell membrane and to the G protein and are not taken into account in the modeling enterprise, where the receptor is handled as a monomer.

Nevertheless, to bridge the gap between the limited availability of experimentally validated data and a high demand to quickly support new GPCR drug discovery projects, an elementary approach focusing on basic aspects may be adequate. The usefulness of our approach in various aspects of GPCR research could already be demonstrated. Here, we give a detailed description of the new methodology and briefly discuss some of its applications. Detailed descriptions of the application examples can be found elsewhere.[45−49]

RESULTS AND DISCUSSION

Functional Conservation Profiles of TM Helices in Class A GPCRs. To assess the variability in sequence conservation in GPCR TM regions, the TM segments of the GPCR sequences were extracted and aligned against the corresponding helices of the opsin family. Details of the automated alignment procedure are given in the Materials and Methods section.

Briefly, for all the registered GPCR subfamilies in SWISS−PROT, the sequence segments of the seven TM helices of the family members were retrieved according to the annotated TM boundaries and aligned separately for each helix and subfamily. Each aligned helix block of a GPCR subfamily was then aligned against the helix profile of the opsin family by an algorithm that ensures that no gaps are introduced into the TM regions. The helix boundaries were then modified according to the X-ray crystal structure of bovine rhodopsin. Parts of the EC2 from these GPCRs were also aligned, in this case, using as an anchoring point the cysteine involved in the highly conserved disulfide bridge between TM3 and EC2. The derived alignments for each TM helix and the EC2 loop for the entire class A were then further analyzed.

As a measure of the conservation of each alignment position, an index F_p was calculated that indicates the degree of conservation of pharmacophore features within the amino acid distribution profiles for each position p (Materials and Methods, eq 1). For the description of the 20 naturally occurring amino acids, topological pharmacophore descriptors[50] were generated for the isolated amino acids in their amide forms and the resulting descriptors converted to a matrix of similarity values.

In Figure 1, the functional conservation indices F_p are displayed as a function of helix positions for the class A GPCR helices 1−7. As expected, relative maxima of these indices are found at positions widely recognized as well-conserved (and referred to as ".50" in the Ballesteros−Weinstein numbering scheme[51]). It is interesting to note that

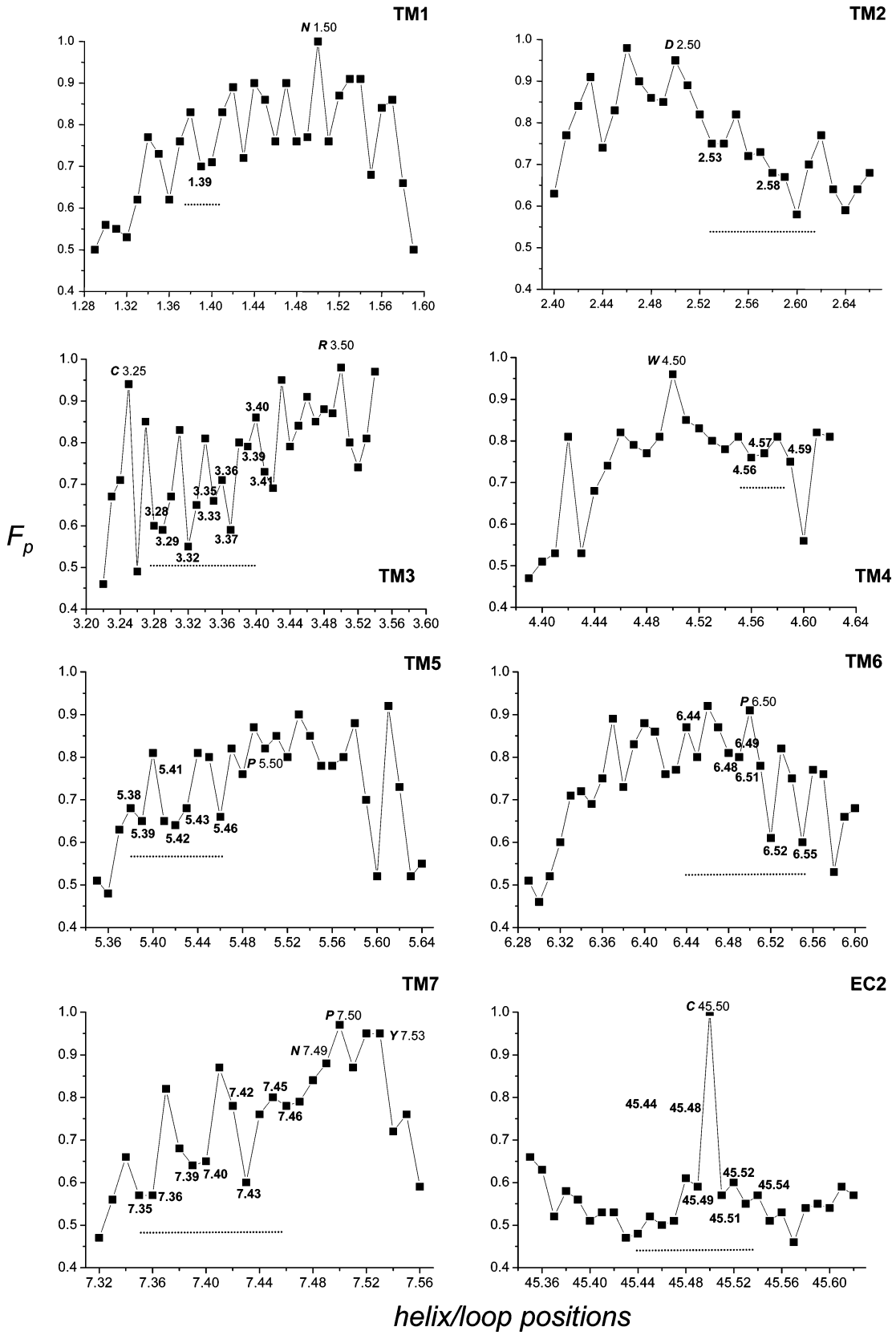

Figure 1. Functional conservation indices F_p for helices 1−7 of the class A GPCRs plotted as a function of TM helix position (in the Ballesteros−Weinstein helix numbering scheme) and for the extracellular loop EC2 as a function of loop position. In analogy to the Ballesteros−Weinstein helix numbering scheme[51] for TM helices, EC2 is labeled "45" to indicate its location between helices 4 and 5. The highly conserved cysteine, thought to be disulfide-bonded, was given the index 45.50, and the residues within EC2 were then indexed relative to this ".50" position. The BW annotation is specified for positions with local minima in F_p values and for selected helix positions. Labeled positions highlighted by a dotted line indicate helix positions with low functional conservation, for which interactions with GPCR ligands have repeatedly been reported in the literature.

the chosen anchoring point in helix 5, a proline at 5.50, is less conserved compared to the reference points of the other helices. For all the helices (less pronounced for helix 4), the F_p values show a clear pattern, with local maxima of the functional conservation index occurring at every third to fourth helix position. These maxima of F_p indicate the conservation of the hydrophobic function at these positions and reflect a preference for hydrophobic amino acids at helix positions directed toward the membrane, which is confirmed by a structural analysis of the rhodopsin X-ray structure. The less-pronounced pattern for helix 4 suggests a special environment and role of helix 4 within the GPCR architecture. Comparisons across the whole GPCR family show that helix positions with low F_p values indicate positions that are very variable in terms of pharmacophore properties. Helix positions with very low F_p values are highlighted in Figure 1. The highlighted helix positions are often within a distance of 6 Å from *cis*-retinal in the X-ray crystal structure of bovine rhodopsin and are frequently cited in the GRAP mutant database[39] as influencing ligand activity. Thus, minima in conservation tend to represent areas within the helices for which interactions with ligands binding to the TM region have been reported in the literature. In comparison to the graphs for the 7 helices, F_p values for the EC2 loop are low, and the lack of a repetitive pattern is striking. Only the position of the cysteine conserved in the EC2 loop of all class A GPCRs gives rise to an F_p value of 1.0. The functional conservation analysis across all helix positions for the whole class A GPCR family is in good agreement with the 7TM architecture being conserved for the whole family. Furthermore, the well-conserved minima in pharmacophore conservation support the hypothesis of favored ligand interaction sites within the TM regions of the receptors across the whole class A GPCR family. All further analyses described here focus on positions with low F_p values in Figure 1.

A Similarity Analysis of Class A GPCRs based on LPVs versus Whole Sequence Alignment. Sequence positions characterized by very low F_p values were combined in a so-called LPV. The selected helix and EC2 loop positions are listed in the Ballesteros−Weinstein numbering scheme[51] in Table 1. To see if GPCRs could be differentiated on the basis of these 35 sequence positions only, the LPVs for all class A GPCRs were generated and a matrix of pairwise similarity values were calculated. The similarity metric used for LPVs, $S_{LPV,st}$, is a weighted sum of amino acid similarity values for each LPV position (Material and Methods, eq 4). Weights for each position of the LPV (Table 1) were assigned according to the functional conservation profiles from the preceding section (Figure 1), with low F_p values translating into high index weights and higher F_p values into lower weights. All positions within EC2 received a low weight (see Materials and Methods for details).

For several subfamilies of the class A GPCR receptors, the similarity values were also calculated on the basis of the full-length GPCR sequence alignment from GPCRDB.[5,52] For both the LPV and the whole-sequence comparisons, topological pharmacophore descriptors[50] of the individual amino acids formed the basis of the amino acid similarity values, as in the alignment procedures described above.

Figure 2 compares the hierarchical tree presentations of the calculated similarity matrices for six class A families (acetylcholine receptors [ACM], serotonin receptors [5H], dopamine receptors [DDR], nucleotide receptors [AA and P2Y], and neuropeptide Y receptors [NYR]) on the basis of both the full-length GPCR sequence alignment[5,52] and the LPVs. The tree topologies agree with those obtained using BLOSUM62 as an amino acid similarity matrix instead of the pharmacophore description-based matrix used here (data not shown).

Table 1. LPV and Index Weights, w_i, for the Calculation of the Similarity Values between Pairs of LPVs[a]

index number	w_i	index number	w_i
1.38	0	6.44	0
1.39	0	6.48	0
2.58	0	6.51	0
2.61	0	6.52	0.8
3.28	0.8	6.55	1.0
3.29	0.8	7.38	0
3.32	1.0	7.39	0.8
3.33	0	7.4	0.8
3.35	0	7.42	0
3.36	0	7.43	1.0
3.37	0.8	45.44	0.6
3.4	0	45.48	0.6
4.56	0	45.49	0.6
5.42	1.0	45.5	0
5.43	0.8	45.51	0.6
5.46	0.8	45.52	0.6
5.47	0	45.54	0.6
5.51	0		

[a] The LPV is a list of amino acids assumed to form the general TM binding pocket.

Figure 2 demonstrates that the well-known grouping of the GPCR subfamilies can be reproduced, although only 35 positions were considered in the LPVs compared to the complete sequences used in the GPCRDB alignment. Thus, the positions included in the LPVs are already sufficient to serve as characteristic fingerprints for the GPCR subfamilies and do not introduce artificial relationships. An advantage of using the LPVs compared to the whole sequence alignment is the strong reduction of noise, giving rise to a sharp separation of the subfamilies.

Applications of the LPV Methodology. In the following sections, two applications of the LPV analysis approach are discussed from Roche drug discovery projects. The first example deals with the reclassification of trace-amine-associated GPCR receptors and serves to compare classical phylogenetic analysis and the LPV method. The second application of the LPV methodology demonstrates how LPVs could support site-directed mutational analysis studies in order to work toward a better understanding of GPCR tuning at a molecular level. Both studies are discussed in detail elsewhere,[45−49] and only the application of the LPV methodology is highlighted below.

Trace-Amine-Associated Receptors.[45,46] Trace amines are endogenous compounds structurally related to classical biogenic amines that act on the recently discovered trace-amine-specific GPCR receptors (TAs, TARs, TAARs).[45,46,54,55] This GPCR subfamily includes 9, 16, and 19 members in human, mouse, and rat species, respectively. So far, however, a clear functional relationship between trace amines and the trace-amine-associated receptors could only be detected for the TAAR receptors 1 and 4. The TAARs are attractive targets for drug development because of reports that several psychoactive compounds act directly on TAAR1 and because

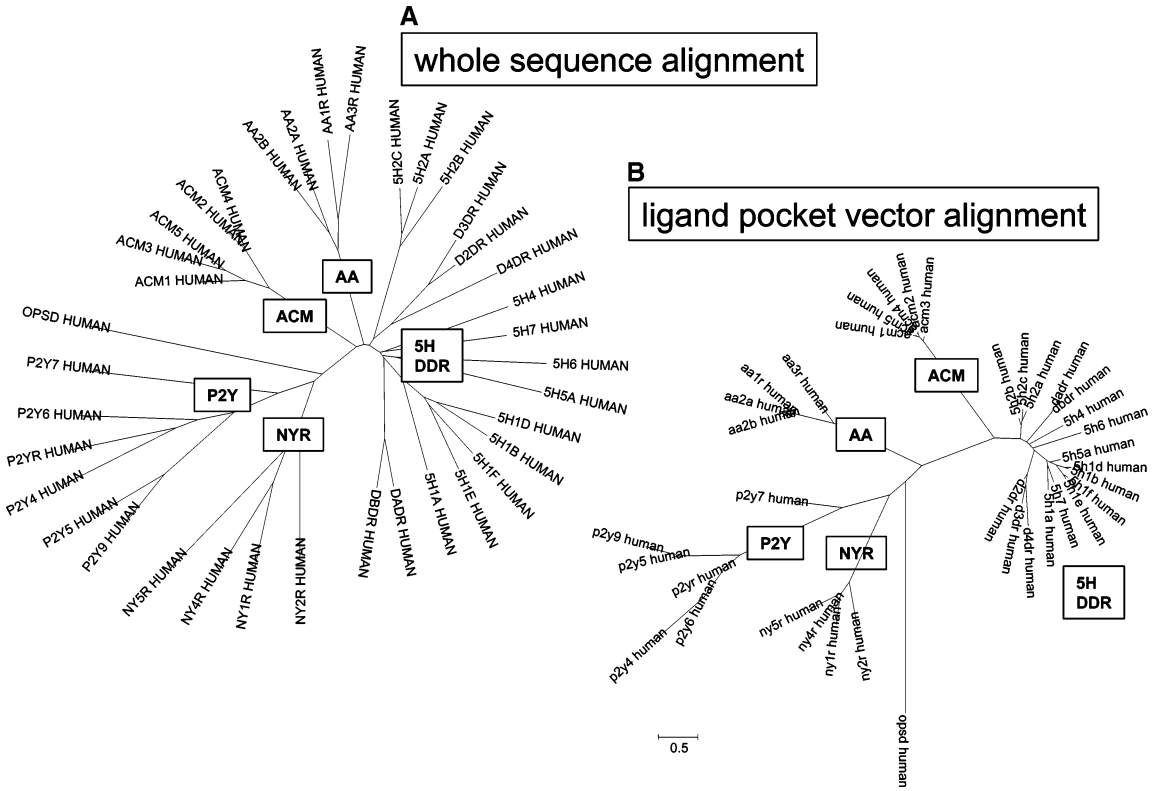

Figure 2. Tree representations (UPGMA clustering)[53] of similarity matrices based on topological pharmacophore descriptions, for six class A families—acetylcholine receptors (ACM), serotonin receptors (5H), dopamine receptors (DDR), nucleotide receptors (AA and P2Y), and neuropeptide Y receptors (NYR). (A) Analysis based on full-length GPCRDB[5,52] sequence alignment. (B) Analysis based on LPVs only. The reduction of noise in the LPV tree compared to the GPCRDB tree is remarkable, while the general relationships detected between the GPCRs are independent from the method applied.

of findings relating trace amines to depression and schizophrenia.[46,56,57] When targeting any member of the TAAR GPCR family, it is important to elucidate relationships between the different members and to identify potential secondary targets to be included in a panel of functional assays. Both the phylogenetic analysis and the new pharmacophore similarity analysis of LPVs provided compelling evidence for three separate TAAR subgroups, with the first subgroup formed by TAARs 1−4, the second subgroup represented by TAAR5, and the third subgroup containing TAARs 6−9. Probably, only subgroup 1 (TAARs 1−4), especially TAAR1 and TAAR4, represents specific receptors for trace amines.[45,46] While the LPV-based analysis and the full-length sequence-based phylogenetic analysis agreed in the overall family topology (compare Figure 3), one interesting difference was obtained. The LPV analysis showed a higher degree of similarity between human and mouse TAAR1 receptors as opposed to the human and rat TAAR1 receptors, which is not apparent from the phylogenetic analysis (see Figure 3). This result of the LPV analysis was confirmed by experimental pharmacological data on the rat, mouse, and human receptor by determining the EC50 values of four trace amines and four biogenic amines. The pharmacological profiles of the mouse and human TAAR1 receptors were found to be very similar to each other but different from the profile for the rat receptor, in agreement with the finding obtained from the LPV analysis. This example clearly demonstrates how the analysis of LPVs can provide new insights for the prediction and explanation of species differences and GPCR relationships in terms of in vitro pharmacology.

Metabotropic Glutamate Receptors.[47−49] Metabotropic glutamate receptors play central roles as modulators of both glutamatergic and other neurotransmitter systems in the central nervous system. They belong to class C GPCRs and differ from class A GPCRs by an exceptionally long extracellular amino-terminal domain (500−600 amino acids) that contains a "venus flytrap" module for agonist binding.[58] The recent discovery of allosteric modulators, which bind to the 7TM regions of the mGlu receptors, has raised interest in these GPCRs as attractive targets for the therapeutic treatment of psychiatric and neurological disorders. The negative allosteric modulator of mGlu5, 2-methyl-6-(phenyl)-pyridine (MPEP), has been shown to display anxiolytic and antidepressant activity in various rodent test models.[59] It has been demonstrated that MPEP interacts with the 7TM region of the mGlu5 receptor. Besides the negative modulator MPEP of mGlu5, a subtype-selective, highly potent, negative allosteric modulator EM-TBPC {1-ethyl-2-methyl-6-oxo-4-(1,2,4,5-tetrahydro-benzo[*d*]azepin-3-yl)-1,6-dihydro-pyrimidine-5-carbonitrile} is known for the rat mGlu1 receptor.[60] In addition, mGlu5 receptors have also been reported to be amenable to positive allosteric modulation by benzaldazine analogues such as 3,3′-difluorobenzaldazine (DFB)[61,62] With these three chemical tools at hand, the LPV methodology was applied with the goal to obtain novel insights for the modulation of these receptors.

The alignment of the mGlu receptor helix segments to the Class A opsin family cannot be reliably guided by sequence similarity measures but, rather, has to rely on the functional conservation profiles introduced here. Using the functional conservation index alignment method, an unambiguous

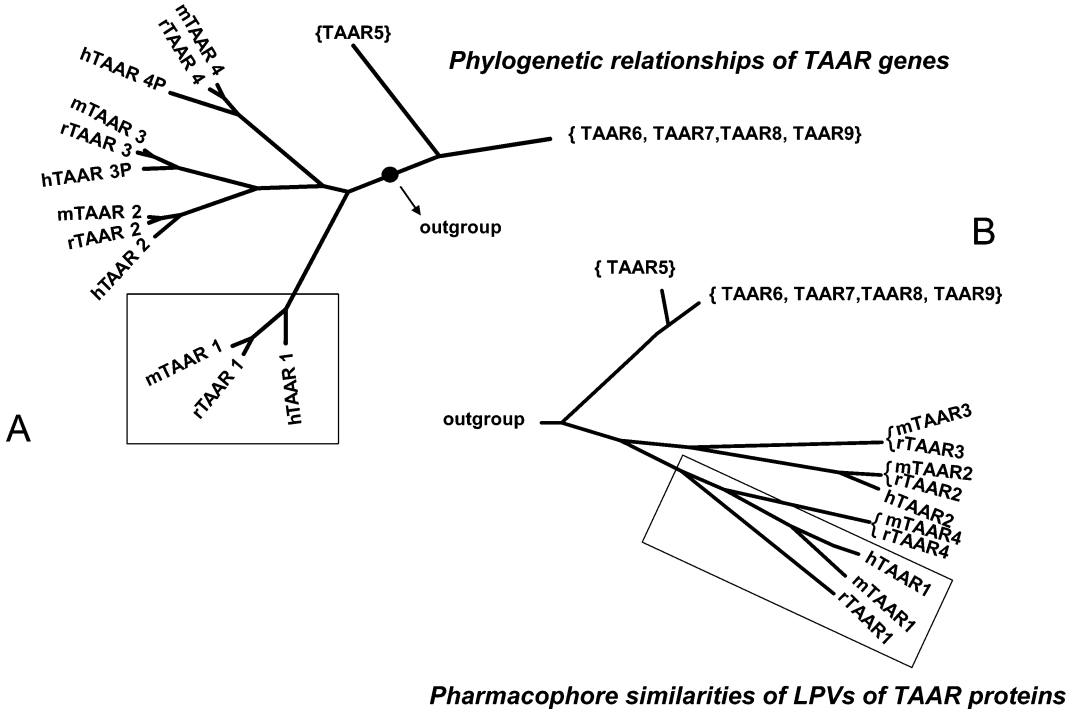

Figure 3. (A) Phylogenetic relationship of TAAR genes based on their DNA sequences. (B) Pharmacophore similarities of the LPVs of TAAR proteins (hierarchical tree representation). Pseudogenes (hTAAR3P and hTAAR4P) are not included in the LPV analysis, because no functional proteins are coded from pseudogenes. Branches for the subgroups of TAAR5 and TAAR6, TAAR7, TAAR8, and TAAR9 are not shown. The squares highlight the found differences between the two analysis methods for TAAR1. h: human. m: mouse. r: rat. P: pseudogenes.

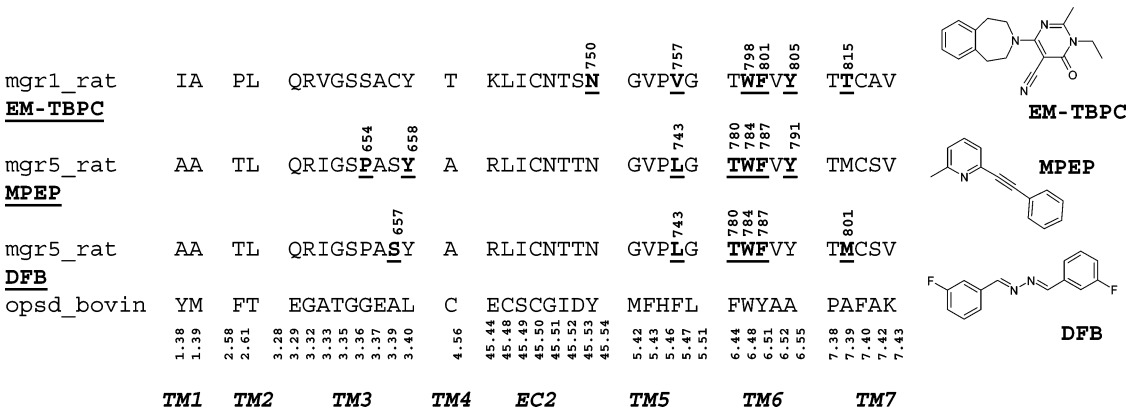

Figure 4. LPVs of the metabotropic glutamate receptors mGlu1 and mGlu5. The Ballesteros−Weinstein numbering scheme of the amino acids in the LPVs is indicated to facilitate the comparison with other GPCRs. Critical residues involved in the allosteric modulation of the mGlu receptors are shown in bold and underlined. EM-TBPC is a negative allosteric modulator of rat mGlu1. MPEP and DFB are negative and positive allosteric modulators of rat mGlu5, respectively.

alignment of the helices was found and the LPVs of the metabotropic glutamate receptor mGlu5 and mGlu1 (SWISS−PROT ID: mgr5_rat and mgr1_rat) were derived and are presented in Figure 4. To test whether the amino acids compiled in the LPVs really are essential for the action of the three different modulators of the metabotropic glutamate receptors, a site-directed mutational analysis was performed.

Indeed, residues were identified within TM3, TM5, TM6, and TM7 that are crucial for MPEP binding to the rat mGlu5 receptor and that are all part of the automatically generated LPV. Three mutations (Y658$^{3.40}$V, W784$^{6.48}$A, and F787$^{6.51}$A) completely disrupt [^{3}H]MPEP binding and block MPEP-mediated inhibition of the quisqualate-induced intracellular calcium ([Ca^{2+}]$_i$) response.[47] Figure 4 summarizes the critical amino acids (bold and underlined) within the LPV, thus mapping the binding regions of the modulators. Since DFB partially displaces the binding of the allosteric inhibitor [^{3}H]MPEP[61] and since MPEP attenuates the action of DFB,[62] the relationship between the MPEP and DFB binding sites in the TM region was investigated. Three mutations (S657$^{3.39}$C, T780$^{6.44}$A, and M801$^{7.39}$T) caused a complete loss of the DFB-mediated increase in functional response.[49] In the binding study of the allosteric modulator of rat mGlu1, [^{3}H]-EM-TBPC, three mutations (F801$^{6.51}$A, Y805$^{6.55}$A, and T815$^{7.39}$M) completely abolished [^{3}H]-EM-TBPC binding to the mGlu1 receptor. The conversion of valine 757 to a leucine (V757$^{5.47}$L) or to an alanine (V757$^{5.47}$A) further led to a dramatic reduction in [^{3}H]-EM-TBPC affinity by about 13-fold in both mutants. Moreover, the conversion of tryptophan 798 to a phenylalanine (W798$^{6.48}$F) or to a tyrosine (W798$^{6.48}$Y) significantly increased the binding affinity of [^{3}H]-EM-TBPC by 9- and 5-fold, respectively.

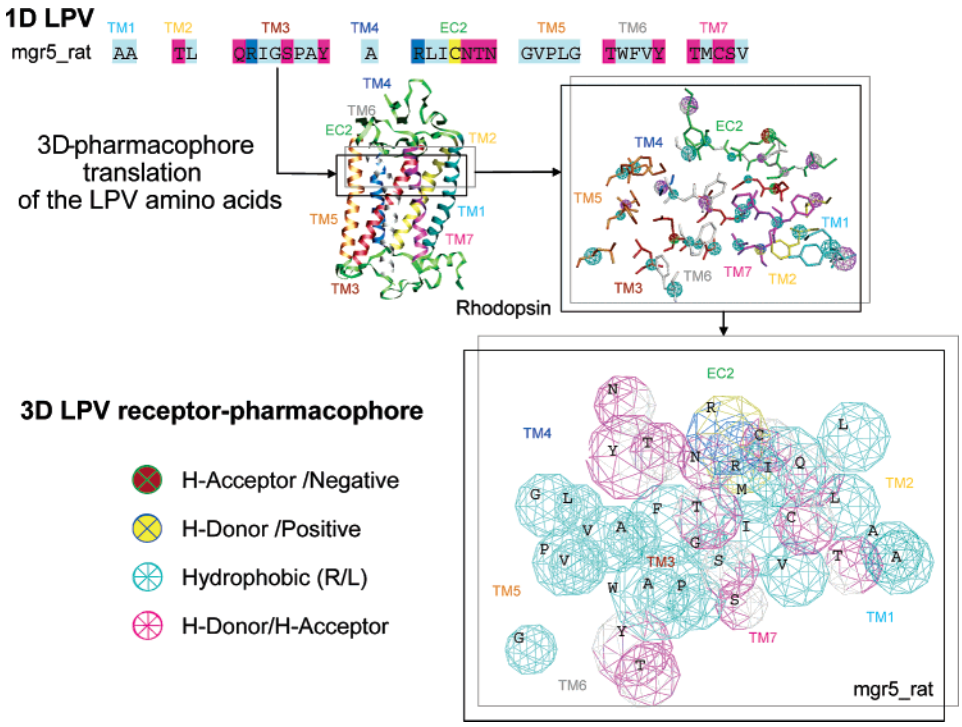

Figure 5. Translation of the 1D LPV into a 3D receptor pharmacophore. Pharmacophore features describing the approximate size and interaction types of the amino acids in the 1D LPV are placed near the Cα positions of the corresponding amino acids in the rhodopsin structure to build a 3D receptor pharmacophore model. As a representative example, the rat metabotropic glutamate receptor 5 (SWISS−PROT: mgr5_rat) was chosen. Hydrophobic amino acids are represented by blue pharmacophores; amino acids with H-donor/acceptor functionalities by magenta pharmacophores; amino acids with H-bond acceptor functionalities and negative charge, such as D and E, by red and green ones; respectively; and amino acids with H-bond donor functionalities and positive charge, such as K and R, by yellow and blue ones, respectively.

In summary, six residues (N750$^{45.54}$, V757$^{5.47}$, W798$^{6.48}$, F801$^{6.51}$, Y805$^{6.55}$, and T815$^{7.39}$) are crucial in the binding of EM-TBPC to rmGlu1.[48] These experimental site-directed mutational studies provide strong evidence for overlapping binding sites of the three allosteric modulators. Despite the inevitably simplistic picture of the GPCR function used here, the described experimental findings demonstrate the practical usefulness of the LPV methodology even for class C GPCR receptors: the LPV methodology provides suggestions for mutational analysis studies and accelerates the mapping of binding sites of GPCR ligands in the TM region.

Converting LPVs to 3D Pharmacophore Representations. Since the amino acids of the LPV line the putative general TM binding pocket of all GPCRs, a translation of the LPV to a 3D model of the binding site is a logical extension of this method. The goal in attempting this translation was not to create full homology models but to concentrate on the general topology of the binding site and the approximate relative orientation of charged, donor/acceptor, and neutral side chains. Details on the generation of the 3D receptor-pharmacophore models based on the LPVs are described in the Materials and Methods section and are outlined in Figure 5.

Each 3D receptor pharmacophore model consists of 35 pharmacophore features, representing the 35 amino acids of the LPV. Pharmacophore features are spherical and are colored according to the pharmacophore properties of the amino acids. These are classified into six categories: L, aliphatic; R, aromatic; D, H-bond donor; A, H-bond acceptor; P, positively charged; and N, negatively charged. The C α positions of the X-ray crystal structure of bovine rhodopsin (1f88)[18] are used to orient the LPV pharmacophore features in space. Feature centers and radii are positioned such as to optimally cover the accessible side chain orientations of the respective amino acids in α helices. These values were derived from the statistical analysis of a rotamer library.[63] Features with H-bond donor or H-bond acceptor properties are assigned a cone pointing toward the most likely direction of the H bond. The angle for the cone opening indicates a measure for the direction of the H bond. A large opening angle indicates a high uncertainty, whereas a small opening angle represents a high confidence for predicting the H-bond direction.

The simplistic TM binding pocket pharmacophores can serve several purposes. They can be used to understand mutational data in a 3D context and, thus, help to refine the understanding of binding modes of reference ligands. Once an approximate binding orientation is found, the receptor pharmacophore can guide chemical optimization by pointing out potential exit vectors (both direction and type of substituent) on the compound. In our hands, the combination of site-directed mutagenesis studies with the LPV method have proven valuable in testing and refining hypotheses for ligand binding and in assisting further in the optimization process of a drug candidate. For early-stage projects, and especially when no ligand information is available, the method can help to guide focused screening experiments to arrive at chemical tools to further characterize receptors. The method allows to quickly assess the requirements for ligand binding: in its simple form, a ligand-based pharmacophore hypothesis is the inverse of the receptor-based pharmacophore. It is clear, however, that the coarseness of the

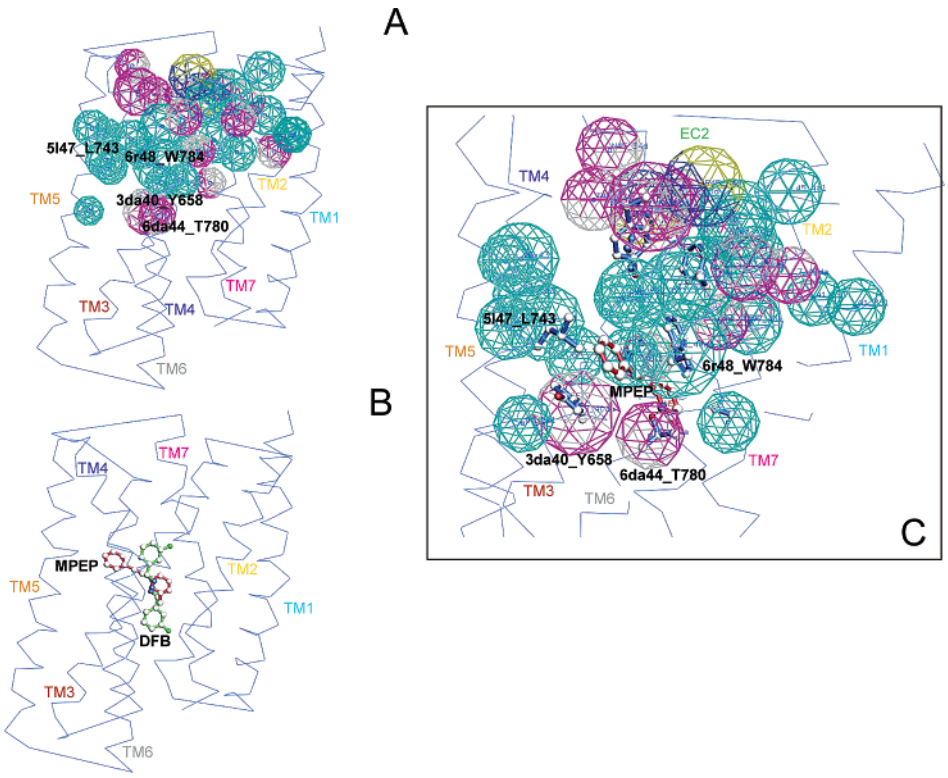

Figure 6. 3D receptor pharmacophore model of a TM binding pocket of the rat metabotropic glutamate receptor 5 (mgr5_rat) with postulated binding modes of the negative and positive allosteric modulators, 2-methyl-6-(phenylethynyl)-pyridine (MPEP) and DFB, respectively. (A) Receptor pharmacophore model of the TM binding pocket of the rat metabotropic glutamate receptor 5 (mgr5_rat), which is superimposed on the rhodopsin 7TM C_α skeleton. Loops are removed, and the helices are directly connected. This artificial 7TM C_α skeleton serves as crude grid for the 7TM architecture. (B) Possible binding modes of MPEP (red) and DFB (green) within the 7TM region of the receptor emphasizing the experimentally found overlapping binding sites of the two modulators near the TM6. (C) Important amino acids for MPEP binding. Rotamer conformations are shown for the essential determinants and the pharmacophore representations are superimposed.

pharmacophore model does not justify direct automated ranking procedures of large sets of molecules.

The experimentally well-characterized system of the rat metabotropic glutamate receptor 5 (SWISS−PROT ID: mgr5_rat) may serve as an illustrating example for the application of the 3D receptor-pharmacophore models. In Figure 6A, the pharmacophores of the rat metabotropic glutamate receptor 5 are superimposed on the C_{alpha} skeleton of the 7TM helices, with the intermittent loop regions omitted. The receptor model gives valuable information on the approximate 3D distribution of the pharmacophores. On the basis of site-directed mutagenesis, the two allosteric modulators can be roughly positioned within the 7TM regions. Figure 6B shows possible binding modes of the two negative and positive allosteric modulators, MPEP and DFB, respectively. In agreement with the mutational analysis experiments, the orientation of MPEP includes interactions with the aromatic network of $Y658^{3.40}$, $W784^{6.48}$, and $F787^{6.51}$ and a hydrogen bond with $T780^{6.44}$, which could prevent the movement of TM6 relative to TM3 and stabilize the inactive conformation of the receptor. Since DFB partially displaces the binding of the allosteric inhibitor [^{3}H]MPEP[61] and the DFB enhancer effect was completely absent in $T780^{6.44}$A mutant receptors, $T780^{6.44}$ might serve as a polar anchor for both modulators. A mutation of $W784^{6.48}$, resulting, on one hand, in a dramatic increase of the DFB-mediated potentiation and, on the other hand, in a loss of MPEP-mediated inhibition, supports the hypothesis that $W784^{6.48}$ plays a central role in controlling the activation states of the receptor. The model confirms an overlapping but distinct binding site to MPEP and emphasizes the key role of $W784^{6.48}$ in TM6 mediating the action of both allosteric modulators.[49] The TM6 helix is a well-characterized region responsible for the activation of GPCRs.[22] This, together with the fact that the binding pockets of negative (MPEP) and positive (DFB) modulators are overlapping in TM6, suggests a similar mechanism for blocking and enhancing activities in mGlu receptors. In Figure 6C, the essential amino acids for MPEP binding are shown in a distinct rotamer conformation with their pharmacophore representations superimposed. Tyrosine $Y658^{3.40}$ and threonine $T780^{6.44}$ are enclosed in two donor−acceptor features, where the hydroxyl groups of tyrosine and threonine, respectively, are pointing in the direction of the cones of the features. Tryptophan $W784^{6.48}$ and leucine $L743^{5.47}$ are represented by hydrophobic features without any cones. The degree of uncertainty of the receptor model depends on the availability of experimental data and mirrors the confidence in the predictive power attributed to models of this kind.

CONCLUSIONS

A detailed sequence analysis of TM helices across the entire class A of GPCRs (more than 1000 receptors) identifies positions with a pronounced variability within the TM helix regions. TM helix positions with low conservation indices were combined into LPVs, ignoring highly conserved TM helix positions with hydrophobic amino acid substitution

patterns, or engaged in key helix contacts. Because of the well-documented role of EC2 for ligand binding specificity,[37,64,65] amino acids from this loop close to the conserved disulfide bridge between TM3 and EC2 were also included in the LPVs. In general, the similarity analysis of these LPVs reveals the same GPCR receptor family structures as found by sequence-based phylogenetic analysis, as demonstrated both for a diverse set of class A GPCRs and, more specifically, for the trace amine-associated GPCR subfamily. In addition, the similarity analysis was able to predict receptor pharmacologies not apparent from the phylogenetic analysis of the trace amine-associated receptor family.[45,46] This is attributable both to a reduction of noise achieved by focusing on the LPV and to the use of a pharmacophore-based metric in the similarity analysis presented here.

It is well documented in the literature that sequence motifs can be sufficient as fingerprints for classification of GPCR subgroups. Huang[64] recently published that the TM3 helix position (3.32) and the TM7 helix position (7.40) are sufficient for classifying the aminergic GPCRs and for predicting novel members of this class among orphan GPCR sequences. Prusis et al.[66−69] reported a proteochemometrics approach to identify amino acid motifs important for ligand binding to aminergic GPCR receptors from within the TM regions. They described the aligned 7TM helices of the aminergic GPCRs by physicochemical descriptors and applied a multivariate analysis to explore the interactions of the TM helix amino acids of the α_{1a}-, α_{1b}-, and α_{1d}-adrenoreceptors with 4-piperidyl oxazole antagonists. Jacoby[70] reports that the monoamine-related GPCR sequences can be unambiguously identified by the presence of the TX(DE)R(IF) motif in the TM3 helix (note that this motif is not involved in the LPVs introduced here). Furthermore, for aminergic GPCRs, he compares the results of the standard phylogenetic analysis of the 7TM domain to phylogenetic relationships derived from amino acid fingerprints identified by mutational analysis for three different 5HT1A receptor ligands. He states that this strategy to analyze sequence similarities only for regions contributing to the ligand binding sites clearly enhances the identification of putative similarities for orphan GPCR receptors to well-characterized GPCRs, providing more starting points in the search for potential ligands. This study is, thus, in excellent agreement with the results presented here.

Ligand pharmacophore models have proven to be among the most successful approaches for finding new leads in the pharmaceutical industry.[71] Receptor-based pharmacophores based on target X-ray crystal structures have recently started to attract the attention of researchers.[72,73] This paper presents a new receptor-based pharmacophore approach for GPCRs by translating the LPVs into a 3D pharmacophore presentation of the TM pocket, which, in comparison to classical homology models, puts very little emphasis on atomic details.

Explicit GPCR homology models and GPCR receptor pharmacophore models have a number of things in common. Both approaches require the assumption that the 7TM architecture is common to all GPCRs. This has been experimentally verified. All studies on GPCR modeling, so far, indicate that sequence identity does not have to be as high as that for typical soluble enzymes to arrive at a reasonable model, since the common fold, the membrane environment, and the high conservation of certain key residues essentially defines the general form.[74] Thus, in contrast to comparative modeling in other protein families, there is no absolute need for sequence identity above 30% between the model sequence and the template structure for accurate GPCR models,[75−78] because the identification of the sequences as GPCRs is not in question, and therefore, the 7TM fold of rhodopsin is a valid structural template for the transmembrane region of any GPCR. The location of the 7TM helices and their alignment to the helices of the rhodopsin is the main task in comparative modeling of the 7TM of GPCRs. We address this aspect very carefully with the conservation index-based sequence alignment methods. In applications discussed here, the 7TM alignments have been confirmed by mutational analysis, leading to the detection of critical residues in these helices for ligand binding. We have also shown that the developed method is not limited to class A GPCR receptors, where existing patterns of highly conserved residues in the transmembrane region of the GPCRs help to guide the 7TM alignments, but is also valid for class C GPCR receptors, where no such pattern is observed. Jiang et al.[79] have recently reported the successful identification of the critical determinants for the binding of an allosteric modulator, lactisole, for the taste receptor, T1R3, which is another class C GPCR receptor. This study is another experimentally validated example for a successful 7TM mapping of a class C GPCR to the rhodopsin template and is in agreement with the findings reported here for the metabotropic glutamate receptors.

It is equally important to point out the differences between explicit GPCR modeling and the presented GPCR receptor pharmacophore approach. The receptor pharmacophores represent a rough "road map" indicating how the essential features for ligand binding might be positioned in space. In an explicit homology model, individual rotamer positions are assigned to all amino acids, and the model is then typically refined by molecular dynamics or stepwise relaxation steps folding the side chains around a number of known ligands. While there are various techniques described in the literature to arrive at state-of-the-art homology models,[23,24] they all stress the need of additional information from highly potent ligands[7,23−32,34,35] and of several refinement cycles. There is no reference point by which the accuracy of such models can be judged, except for consistency with known rules about protein folding. As a result, each model fits to the ligands used in the study and to the binding mode implied for these ligands. Despite the substantial progress in building homology models and their proven usefulness to identify new ligands, independently of the level of sophistication of the techniques used, there always remains the risk of a mismatch between optimized rotamers and ligand binding requirements, leading to the rejection of ligands in virtual screening experiments. Receptor pharmacophores are an attempt to avoid atomic details tending to mask that inherent uncertainty. Nevertheless, the receptor pharmacophores are not meant to be used for virtual screening; they are too coarse-grained to justify the use of any automated ranking procedure. Rather, they give quick access to approximate representations of essentially all GPCR transmembrane binding sites useful to understanding relationships between GPCR subfamilies and selectivity issues between receptor subtypes, to initiating site-directed mutagenesis studies, and to guiding chemical optimization. They complement and help to validate

ligand-based pharmacophore models. Furthermore, in connection with the experiment, they are valuable in the generation of general ideas for the modification and optimization of ligands. By providing all relevant information for all GPCRs in the same format, the method allows researchers to test the limits of the hypothesis that all GPCRs share a transmembrane ligand binding pocket for ligands.

MATERIAL AND METHODS

Here, each step of an automated procedure is described for the alignment of GPCR sequences and for the generation of 3D pharmacophore descriptions of GPCR ligand binding pockets. First, the TM segments of a GPCR sequence are predicted and aligned against the helices of the opsin family by a functional conservation index alignment algorithm. This procedure replaces any interactive knowledge-based refinement of the alignment required after using standard bioinformatics alignment tools.[26] The opsin family is used as a reference because of the available X-ray crystal structure of bovine rhodopsin.[18] In a second step, amino acids of the aligned sequence positions involved in ligand binding are extracted. They form an ordered sequence of residues referred to as a LPV. In a third step, pharmacophore features describing the approximate size and interaction types of the amino acids in the LPV are placed onto the Cα positions of the corresponding amino acids in the rhodopsin structure to build a 3D receptor pharmacophore model. In the following, each of these three key steps is explained in detail.

1. Sequence Preparation and Alignment within Family. From any given input sequence, TM helix[80] sections are extracted as stored in SWISS−PROT and as predicted by a hidden Markov model.[81] If there are discrepancies between the two predictions, boundary positions are averaged. At this level, small errors in the predicted helix coordinates do not affect the alignment, as the predicted TMs are used only to be able to extract roughly the TM regions for the separate alignment of the helices; helix boundaries are later corrected on the basis of the alignment to the rhodopsin structure. In addition, sequence segments of the EC2 and a list of all cysteine residues within this loop are extracted. The extremely well-conserved disulfide bridge between TM3 and EC2 in almost all of the GPCRs and the ensuing structural constraints for the amino acids in the neighborhood of this disulfide bridge make it possible to include these positions in the analysis. For each family of GPCRs, the extracted sequences are then aligned by standard bioinformatics alignment tools. A high gap penalty is applied to avoid gaps within the multiple TM regions.

2. Blockwise Alignment to Opsin Family Sequences. The multiple-sequence alignment for helices 1−7 of a particular family is aligned against the one of the opsin family by means of a functional conservation index. The input of the program consists of two blocks of multiply aligned sequences for a single helix region, one block for any GPCR family, the other one for the opsin family; the output is an optimized relative alignment of these two helix blocks. First, sequence positions containing end gaps are removed from the two sequence blocks. The two blocks are then shifted against each other. For each position p of the alignment, the functional conservation index F_p is calculated according to eq 1.

$$F_p = \frac{\sum_{k=1}^{20}\sum_{j=1}^{20} S_{kj} n_k m_j}{\sum_{k=1}^{20} n_k \sum_{j=1}^{20} m_j} \quad (1)$$

Here, n and m are amino acid counts in the two sequence blocks and S_{kj} is a similarity index for amino acids j and k. The values of S_{kj} take on values between 0 and 1 and can, for example, be chosen from one of the widely used similarity matrices such as BLOSUM62[82] or PHAT_T75_B73.[83] Similarity values S_{kj} are employed on the basis of topological pharmacophores,[50] which were directly calculated for isolated amino acid pairs in their amide form. Furthermore, similarity values could also be calculated from conventional amino acid descriptors[84] by means of eq 2.

$$S_{kj} = \frac{\sum^{nz} z_k z_j}{2\left(\sqrt{\sum^{nz} z_k z_k}\sqrt{\sum^{nz} z_j z_j}\right)} + 0.5 \quad (2)$$

In eq 2, z represents the amino acid descriptors and nz is the number of such descriptors used. A marked increase was noted in the sharpness of the alignment profiles with the similarity values derived from topological pharmacophores. For the classification of GPCR sequences, BLOSUM62 would have been an equally good choice.

From the values of F_p for each amino acid position, a weighted sum F_s is calculated for the entire window of common sequence positions in the block alignment. Where F_s reaches its maximum, the alignment between the two sequence blocks is optimal.

$$F_s = \frac{\sum_{p=1}^{\mathrm{SQL}} F_p w_p}{\sum_{p=1}^{\mathrm{SQL}} w_p} \quad (3)$$

Here, SQL denotes the sequence length of the window of common sequence positions and w_p is a position weight. Position weights were introduced in order to align sequences of no homology to the opsin family, such as those of class B and C GPCRs. For the helix alignments of class A GPCRs against the opsin family, the introduction of position weights did not change the alignment because of already sharp F_s profiles due to the highly conserved amino acids within the transmembrane helices. Position weights for all subfamilies were, therefore, set to 1.0 for class A GPCRs. For the alignment of class B and C GPCRs, however, different position weights for the single classes were crucial for the alignment process. Positions of well-conserved amino acids in the class A helix profiles had to be labeled with high position weights for deriving an ambiguous alignment of class B and C receptors. The values for the seven helix classes for the alignments of class B and C GPCRs, which were derived by using all class A helix profiles as training

Table 2. Average Position Vectors of Amino Acid Side Chain Atoms Carrying Pharmacophore Properties, Expressed as Directions (Unit Vectors) and Lengths l_v Relative to the Local Coordinate System Defined for Each Amino Acid, and Parameters for the Pharmacophore Representation of Amino Acids

amino acid	atoms (*d* CA-)	*x*	*y*	*z*	l_v	category	*d*	r_1	ϕ
A	CB	0.000	0.000	1.000	1.530	L	0.8	2.4	0.0
C	SG	0.172	−0.794	0.583	0.736	DA	1.4	2.3	74.1
D	OD1, OD2	0.099	−0.893	0.439	0.679	AN	1.6	2.3	82.2
E	OE2, OE1	−0.247	−0.630	0.736	0.660	AN	2.1	2.6	84.8
F	CD1, CE1, CZ, CE2, CD2, CG	0.077	−0.250	0.965	0.700	R	0.0	3.1	0.0
G		0.000	0.000	0.000	0.000	L	0.0	2.0	0.0
H	ND1, NE2	0.328	−0.415	0.849	0.774	DA	0.0	3.3	68.7
I	CD1, CG2, CG1	−0.101	−0.591	0.800	0.621	L	1.5	3.2	0.0
K	NZ	0.553	0.014	0.833	0.751	DP	2.0	3.4	72.1
L	CD1, CD2	−0.225	−0.800	0.556	0.716	L	1.7	2.8	0.0
M	SD, CE	−0.227	−0.713	0.663	0.720	L	2.1	3.0	0.0
N	OD1, ND2	0.854	0.219	0.472	0.687	DA	1.6	2.4	81.0
P	CD	0.937	0.111	0.331	1.000	L	0.0	2.4	0.0
Q	NE2, OE1	0.738	0.125	0.663	0.662	DA	2.1	2.7	84.6
R	NH1, NH2, NE	0.532	−0.036	0.846	0.551	DP	2.8	3.1	100.3
S	OG	0.192	−0.189	0.963	0.545	DA	1.2	2.2	101.0
T	OG1	0.491	−0.514	0.703	0.752	DA	1.3	2.5	71.9
V	CG1, CG2	−0.312	−0.502	0.807	0.610	L	1.3	2.6	0.0
W	NE1, CD1, CE3, CZ3, CH2, CZ2	0.361	−0.124	0.924	0.613	R	0.0	3.7	0.0
Y	OH	−0.436	−0.460	0.773	0.723	DA	0.0	3.2	75.9

sets and optimizing the sharpness of the F_s profiles, are given in the Supporting Information.

The final alignment of each TM helix block against the opsin TM helix block is annotated using the Ballesteros−Weinstein nomenclature for conserved helix amino acid residues within class A as reference points. These positions are defined as[51] 1.50: N55, 2.50: D83, 3.37: E122, 4.50: W161, 5.50: P215, 6.50: P267, and 7.50: P303. Because the Ballesteros−Weinstein numbering scheme does not include the annotation of loops, the label "45" was given to the EC2 loop to indicate the location between the helices 4 and 5. The highly conserved cysteine, assumed to be disulfide-bonded, was given the index number 45.50, and the residues within EC2 are then indexed relative to this position.

3. Generation and Comparison of LPVs. Once the alignment of a GPCR sequence of interest relative to the rhodopsin structure is finalized, a list of 35 amino acids potentially involved in ligand binding and activation of the GPCR receptor[22,39] are extracted. This ordered sequence of amino acids is being referred to as the LPV. The list of amino acid indices is given in Table 1 in Ballesteros−Weinstein notation. The procedure described thus far was used to generate LPVs for each SWISS−PROT GPCR entry.

The LPVs form the basis of a new similarity metric for comparing GPCR sequences. The similarity metric $S_{\mathrm{LPV},st}$ for pairs (*s*, *t*) of LPVs is the weighted sum of amino acid similarities S^i_{st} at each LPV position *i* (eq 4). The values S^i_{st} are derived from the similarity values (see eq 2) of the pair of amino acids that are at position *i* for LPV sequences *s* and *t*.

$$S_{\mathrm{LPV},st} = \frac{\sum_{i=1}^{\mathrm{SQL}} S_{st}{}^{i} w_i}{\sum_{i=1}^{\mathrm{SQL}} w_i} \qquad (4)$$

The index weights w_i (Table 1) were derived from an analysis of functional conservation profiles of class A GPCR helix residues (details in Results section). Positions with minima in the F_p values were assigned a weight of 1.0, positions with low F_p values were assigned a weight of 0.8, and positions within the extracellular loop (EC2) were labeled with a weight of 0.6. Positions giving rise to medium to high F_p values are neglected for the similarity analysis by setting the index weights w_i to zero in order to reduce noise. Note that the weights used here have the opposite effect of the ones used for the alignment. For the alignment, positions with high conservation indices received higher weight, whereas here, the focus is on the variable sequence positions.

4. Building 3D Receptor Pharmacophore Models. For each of the 20 amino acid side chains, an approximate spherical pharmacophore presentation was generated. A pharmacophore feature is defined by three coordinates, a radius, and one or two of six category labels: L for aliphatic, R for aromatic, D for H-bond donors, A for H-bond acceptors, P for positively charged side chains, and N for negatively charged side chains. Amino acids with H-bond donor or acceptor properties are attributed a cone indicating the most likely direction of the H bond. A local coordination system was defined in which the unit vector in the direction CA−CB was introduced as the *z* axis. The unit vector orthogonal to the *z* axis in direction N was used as the *x* axis, and the vector orthogonal to both vectors served as the *y* axis. This local coordinate system was defined relative to the absolute coordinate system by the corresponding Euler angles θ, ϕ, and ψ. From standard rotamer conformations, the positions of relevant atoms for each amino acid side chain were averaged by taking into account the relative frequencies of the rotamers found in α helices from a recently published rotamer library.[63] These average positions are listed in Table 2, where all values are reported within the local coordinate system introduced above. They were used as starting points for the empirical generation of an optimum spherical representation of the amino acid side chains. Both the radius of the features and the shift of the feature center along the unit vector in Table 2 were modified to optimally cover the relevant side chain orientations, leading to a value of *d* along the unit vector differing from the original value l_v. The radius

of the feature is chosen such that the modeled surfaces of the different rotamers of the amino acid are optimally represented by the feature volumes (r_1 in Table 2).

To build a 3D receptor model from such features, the Cartesian coordinates of the C_α carbons in the rhodopsin structure are used unchanged and pharmacophore features for each side chain are generated with the center at distance d from C_α in the direction of the unit vector of Table 2, and with radius r_1.

Features with H-bond properties are attributed an additional cone angle indicating the variability of the H-bond direction. The direction vector for the cone of a feature with H-bond acceptor or donor properties was chosen as the unit vector of Table 2. The angle for the cone opening, ϕ, depends on the length of the average vector of atom positions in Table 2 (eq 5).

$$\phi = 2(-70.565l_v + 89.0) \quad (5)$$

Here, l_v is the length of the average position vector in Table 2. A length of 0 gives rise to a large opening angle of close to 180°, indicating a high uncertainty for the direction of the H bond. A length of 1, on the other hand, gives rise to a small opening angle of 37°.

ACKNOWLEDGMENT

We thank Clemens Broger, Martin Strahm, and Wolfgang Guba for interesting and highly valuable discussions as well as for their continuous support of this work.

Supporting Information Available: A table of the values for the seven helix classes for the alignments of class B and C GPCRs. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Okada, T.; Palczewski, K. Crystal structure of rhodopsin: implications for vision and beyond. *Curr. Opin. Struct. Biol.* **2001**, *11*, 420−426.
(2) Ji, T. H.; Grossmann, M.; Ji, I. G Protein-coupled receptors. *J. Biol. Chem.* **1998**, *273*, 17299−17302.
(3) Kenakin, T. Drug efficacy at G protein-coupled receptors. *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 349−379.
(4) Kenakin, T. Efficacy at G-protein-coupled receptors. *Nat. Rev.* **2002**, *1*, 103−110.
(5) Horn, F.; Weare, J.; Beukers, M. W.; Hörsch, S.; Bairoch, A. GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Res.* **1998**, *26*, 275−279.
(6) Sakmar, T. P. Structure of rhodopsin and the superfamily of seven-helical receptors: the same and not the same. *Curr. Opin. Cell Biol.* **2002**, *14*, 189−195.
(7) Flower, D. R. Modelling G-protein-coupled receptors for drug design. *Biochim. Biophys. Acta* **1999**, *1422*, 207−234.
(8) Angers, S.; Salahpour, A.; Bouvier, M. Dimerization: An emerging concept for G protein-coupled receptor ontogeny and function. *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 409−435.
(9) Soudijn, W.; van Wijngaarden, I.; Ijzerman, A. P. Allosteric modulation of G protein-coupled receptors. *Curr. Opin. Drug Discovery Dev.* **2002**, *5*, 749−755.
(10) Schöneberg, T.; Schultz, G.; Gudermann, T. Structural basis of G protein-coupled receptor function. *Mol. Cell. Endocrinol.* **1999**, *151*, 181−193.
(11) Christopoulos, A.; Kenakin, T. G Protein-coupled receptor allosterim and complexing. *Pharmacol. Rev.* **2002**, *54*, 323−374.
(12) Spence, P.; Bard, J.; Jones, P.; Betty, M. The identification of G-protein coupled receptors in sequence databases. *Expert Opin. Ther. Pat.* **1998**, *8*, 235−247.
(13) Attwood, T. K.; Findlay, J. B. Fingerprinting G-protein-coupled receptors. *Prot. Eng.* **1994**, *7*, 195−203.
(14) Attwood, T. K.; Croning, M. D. R.; Gaulton, A. Deriving structural and functional insights from a ligand-based hierarchical classification of G protein-coupled receptors. *Prot. Eng.* **2001**, *15*, 7−12.
(15) Kolakowski, L. F. J. GPCRDb: a G-protein-coupled receptor database. *Recept. Channels* **1994**, *2*, 1−7.
(16) Karchin, R.; Karplus, K.; Haussler, D. Classifying G-protein coupled receptors with support vector machines. *Bioinformatics* **2002**, *18*, 147−159.
(17) Ballesteros, J.; Palczewski, K. G protein-coupled receptor in drug discovery: Implications from the crystal structure of rhodopsin. *Curr. Opin. Drug Discovery Dev.* **2001**, *4*.
(18) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **2000**, *289*, 739−745.
(19) Teller, D. C.; Okada, T.; Behnke, C. A.; Palczewski, K.; Stenkamp, R. E. 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.
(20) Fotiadis, D.; Liang, Y.; Filipek, S.; Saperstein, D. A.; Engel, A. Rhodopsin dimers in native disc membranes. *Nature* **2003**, *421*, 127−128.
(21) Archer, E.; Maigret, B.; Escrieut, C.; Pradayrol, L.; Fourmy, D. Rhodopsin crystal: new template yielding realistic models of G-protein-coupled receptors? *Trends in Pharmacol. Sci.* **2003**, *24*, 36−40.
(22) Ballesteros, J. A.; Shi, L.; Javitch, J. A. 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.
(23) Becker, O. M.; Shacham, S.; Marantz, Y.; Noiman, S. Modeling the 3D structure of GPCRs: Advances and application to drug discovery. *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 353−361.
(24) Hillisch, A.; Pineda, L. F.; Hilgenfeld, R. Utility of homology models in the drug discovery process. *Drug Discovery Today* **2004**, *9*, 659−669.
(25) Bissantz, C.; Bernard, P.; Hilbert, 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.
(26) Boeckler, F.; Lanig, H.; Gmeiner, P. Modeling the similarity and divergence of dopamine D2-like receptors and identification of validated ligand−receptor complexes. *J. Med. Chem.* **2005**, *48*, 694−709.
(27) Cavasotto, C.; Orry, A. J. W.; Abagyan, R. A. Structure-based identification of binding sites, native ligands and potential inhibitors for G-protein coupled receptors. *Proteins: Struct., Funct., Genet.* **2003**, *51*, 423−433.
(28) Evers, A.; Klebe, G. Ligand-supported homology modeling of G-protein-coupled receptor sites: models sufficient for successful virtual screening. *Angew. Chem., Int. Ed.* **2004**, *43*, 248−251.
(29) Evers, A.; Klabunde, T. Structure-based Drug Discovery Using GPCR Homology Modeling: Successful Virtual Screening for Antagonists of the Alpha 1A Adrenergic Receptor. *J. Med. Chem.* **2005**, *48*, 1088−1097.
(30) Gershengorn, M. C.; Osman, R. Minireview: Insights into G protein-coupled receptor function using molecular models. *Endocrinology* **2001**, *142*, 2−10.
(31) Moro, S.; Deflorian, F.; Spalluto, G.; Patorin, G.; Cacciari, B. Demystifying the three-dimensional structure of G protein-coupled receptors (GPCRs) with the aid of molecular modeling. *Chem. Commun.* **2003**, 2949−2956.
(32) Müller, G. Towards 3D structures of G protein-coupled receptors: A multidisciplinary approach. *Curr. Med. Chem.* **2000**, *7*, 861−888.
(33) Oliveira, L.; Hulsen, T.; Hulsik, D. L.; Paiva, A. C. M.; Vriend, G. Heavier-than-air flying machines are impossible. *FEBS Lett.* **2004**, *564*, 269−273.
(34) Schafferhans, A.; Klebe, G. Docking ligands onto binding site representations derived from proteins built by homology modelling. *J. Mol. Biol.* **2001**, *307*, 407−427.
(35) Vaidehi, N.; Floriano, W. B.; Trabanino, R.; Hall, S. E.; Freddolino, P. Predition of structure and function of G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12622−12627.
(36) Oliveira, L.; Paiva, A. C. M.; Vriend, G. Correlated Mutation Analysis on Very Large Sequence Families. *ChemBioChem* **2002**, *3*, 1010−1017.
(37) Bondensgaard, K.; Ankersen, M.; Thøgersen, H.; Hansen, B. S.; Wulff, B. S. Recognition of privileged structures by G-protein coupled receptors. *J. Med. Chem.* **2004**, *47* (4), 888−899.
(38) Mirzadegan, T.; Benko, G.; Filipek, S.; Palczewski, K. Sequence analysis of G-protein-coupled receptors: similarities to rhodopsin. *Biochemistry* **2003**, *42*, 2759−2767.
(39) Kristiansen, K.; Dahl, S. G.; Edvardsen, Ø. A database of mutants and effects of site-directed mutagenesis experiments on G-protein coupled receptors. *Proteins: Struct., Funct., Bioinf.* **1996**, *26*, 81−94.

(40) Shi, L.; Javitch, J. A. The binding site of aminergic G protein-coupled receptors: The transmembrane segments and second extracellular loop. *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 437−467.
(41) Visiers, I.; Ballesteros, J. A.; Weinstein, H. Three-dimensional representation of G protein-coupled receptor structures and mechanisms. *Methods Enzymol.* **2001**, *343*, 329−371.
(42) Gether, U. Uncovering molecular mechanisms involved in activation of G protein-coupled receptors. *Endocrine Rev.* **2000**, *21*, 90−113.
(43) Lu, Z.-L.; Saldanha, J. W.; Hulme, E. C. Seven-transmembrane receptors: crystals clarify. *Trends Pharmacol. Sci.* **2002**, *23*, 140−146.
(44) Park, P. S.-H.; Filipek, S.; Wells, J. W.; Palczewski, K. Oligomerization of G Protein-Coupled Receptors: Past, Present, and Future. *Biochemistry* **2004**, *43*, 15643−15656.
(45) Lindemann, L.; Ebeling, M.; Kratochwil, N. A.; Bunzow, J. R.; Grandy, D. K. Trace amine-associated receptors form structurally and functionally distinct subfamilies of novel G protein-coupled receptors. *Genomics* **2005**, *85*, 372−385.
(46) Lindemann, L.; Hoener, M. C. A renaissance in trace amines inspired by a novel GPCR family. *Trends Pharmacol. Sci.* **2005**, *26*, 274−281.
(47) Malherbe, P.; Kratochwil, N. A.; Zenner, M.-T.; Piussi, J.; Diener, C. Mutational analysis and molecular modeling of the binding pocket of the metabotropic glutamate 5 receptor negative modulator 2-methyl-6-(phenylethynyl)-pyridine. *Mol. Pharmacol.* **2003**, *64*, 823−832.
(48) Malherbe, P.; Kratochwil, N. A.; Knoflach, F.; Zenner, M.-T.; Kew, J. N. Molecular modeling and mutational analysis of the allosteric binding site of a novel, selective, noncompetitive antagonist of the metabotropic glutamate 1 receptor. *J. Biol. Chem.* **2003**, *278*, 8340−8347.
(49) Mühlemann, A.; Ward, N. A.; Kratochwil, N. A.; Diener, C.; Fischer, C. Key amino acid residues implicated in the positive allosteric modulation of rat mGlu5a receptors by 3,3′-difluorobenzaldazine (DFB). **2005**, submitted for publication.
(50) Gerber, P. R. Topological Pharmacophore Description of Chemical Structures using MAB−Force-Field-Derived Data and Corresponding Similarity Measures. *Fundamentals of Molecular Similarity*; Kluwer Academic/Plenum: New York, 2001; pp 67−81.
(51) Ballesteros, J. A.; Weinstein, H. Integrated methods and computational probing of structure−function relations in G protein-coupled receptors. *Methods Neurosci.* **1995**, *25*, 366−428.
(52) Horn, F.; Vriend, G.; Cohen, F. E. Collecting and Harvesting Biological Data: The GPCRDB & NucleaRDB Databases. *Nucleic Acids Res.* **2001**, *29*, 346−349.
(53) Sokal, R. R.; Michener, C. D. A statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.* **1958**, *28*, 1409−1438.
(54) Bunzow, J. R.; Sonders, M. S.; Arttamangkul, S.; Harrison, L. M.; Zhang, G. Amphetamine, 3,4-methylenedioxymethamphetamine, lysergic acid diethylamide, and metabolites of the catecholamine neurotransmitters are agonists of a rat trace amine receptor. *Mol. Pharmacol.* **2001**, *60*, 1181−1188.
(55) Borowsky, B.; Adham, N.; Jones, K. A.; Raddatz, R.; Artymyshyn, R. Trace amines: identification of a family of mammalian G protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 8966−8971.
(56) Shirkande, S.; O'Reilly, R.; Davis, B.; Durden, D.; Malcom, D. Plasma phenylethylamine levels of schizophrenic patients. *Can. J. Psychiatry* **1995**, *40*, 221.
(57) Davis, B. A.; Boulton, A. A. The trace amines and their acidic metabolites in depression−an overview. *Prog. Neuro−Psychopharmacology Biol. Psychiatry* **1994**, *18*, 17−45.
(58) Swanson, C. J.; Bures, M.; Johnson, M. P.; Linden, A.-M.; Monn, J. A. Metabotropic glutamate receptors as novel targets for anxiety and stress disorders. *Nat. Rev. Drug Discovery* **2005**, *4*, 131−144.
(59) Gasparini, F.; Lingenhohl, K.; Stoehr, N.; Flor, P. J.; Heinrich, M. 2-Methyl-6-(phenylethynyl)-pyridine (MPEP), a potent, selective and systemically active mGlu5 receptor antagonist. *Neuropharmacology* **1999**, *38*, 1493−1503.
(60) Annoura, H.; Fukunaga, A.; Uesugi, M.; Tatsuoka, T.; Horikawa, Y. A novel class of antagonists for metabotropic glutamate receptors, 7-(hydroxyimino)cyclopropa[b]chromen-1a-carboxylates. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 763−766.
(61) O'Brien, J. A.; Lemaire, W.; Chen, T.-B.; Chang, R. S. L.; Jacobson, M. A. A family of highly selective allosteric modulators of the metabotropic glutamate receptor subtype 5. *Mol. Pharmacol.* **2003**, *64*, 731−740.
(62) Goudet, C.; Gaven, F.; Kniazeff, J.; Vol, C.; Liu, J. Heptahelical domain of metabotropic glutamate receptor 5 behaves like rhodopsin-like receptors. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 378−383.
(63) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The Penultimate Rotamer Library. *Proteins: Struct., Funct., Bioinformatics* **2000**, *40*, 389−408.
(64) Huang, E. S. Construction of a sequence motif characteristic of aminergic G protein-coupled receptors. *Prot. Sci.* **2003**, *12*, 1360−1367.
(65) Shi, L.; Simpson, M. M.; Ballesteros, J. A.; Javitch, J. A. 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.
(66) Prusis, P.; Lundstedt, T.; Wikberg, J. E. S. Proteo-chemometrics analysis of MSH peptide binding to melanocortin receptors. *Prot. Eng.* **2002**, *15*, 305−311.
(67) Lapinsh, M.; Gutcaits, A.; Prusis, P.; Post, C.; Lundstedt, T. Classification of G-protein coupled receptors by alignment-independent extraction of principal chemical properties of primary amino acid sequences. *Prot. Sci.* **2002**, *11*, 795−805.
(68) Lapinsh, M.; Prusis, P.; Lundstedt, T.; Wikberg, J. E. S. Proteo-chemometrics modeling of the interaction of amine G-protein coupled receptors with a diverse set of ligands. *Mol. Pharmacol.* **2002**, *61*, 1465−1475.
(69) Lapinsh, M.; Prusis, P.; Gutcaits, A.; Lundstedt, T.; Wikberg, J. E. S. Development of proteo-chemometrics: a novel technology for the analysis of drug-receptor interactions. *Biochim. Biophys. Acta* **2001**, *1525*, 180−190.
(70) Jacoby, E. A novel chemogenomics knowledge-based ligand design strategy-Application to G protein-coupled receptors. *Quant. Struct.-Act. Relat.* **2001**, *20*, 115−123.
(71) Mason, J. S.; Good, A. C.; Martin, E. J. 3-D pharmacophores in drug discovery. *Curr. Pharm. Des.* **2001**, *7*, 567−597.
(72) Marquez, V. E.; Nacro, K.; Benzaria, S.; Lee, J.; Sharma, R. The transition from a pharmacophore-guided approach to a receptor-guided approach in the design of potent protein kinase C ligands. *Pharmacol. Ther.* **1999**, *82*, 251−261.
(73) Klebe, G. Lead identification in post-genomics: computers as a complementary alternative. *Drug Discovery Today: Technol.* **2004**, *1*, 225−230.
(74) Moro, S.; Spalluto, G.; Jacobson, K. A. Techniques: Recent developments in computer-aided engineering of GPCR ligands using the human adenosine A3 receptor as example. *Trends Pharmacol. Sci.* **2005**, *26*, 44−51.
(75) Baker, D.; Sali, A. Protein Structure Prediction and Structural Genomics. *Science* **2001**, *294*, 93−96.
(76) Chothia, C.; Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *EMBO J.* **1986**, *5*, 823−826.
(77) Ginalski, K.; Grishin, N. V.; Godzik, A.; Rychlewski, L. Practical lessons from protein structure prediction. *Nucleic Acids Res.* **2005**, *33*, 1874−1891.
(78) Vitkup, D.; Melamud, E.; Moult, J.; Sander, C. Completness in structural genomics. *Nat. Struct. Biol.* **2001**, *8*, 559−566.
(79) Jiang, P.; Cui, M.; Zhao, B.; Liu, Z.; Snyder, L. A. Lactisole Interacts with the Transmembrane Domains of Human T1R3 to Inhibit Sweet Taste. *J. Biol. Chem.* **2005**, *280*, 15238−15246.
(80) Möller, S.; Croning, M. D. R.; Apweiler, R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **2001**, *17*, 646−653.
(81) Krogh, A.; Larsson, B.; Heijne, G.; Sonnhammer, E. L. L. J. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567−580.
(82) Henikoff, S.; Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 10915−10919.
(83) Ng, P. C.; Henikoff, J. G.; Henikoff, S. PHAT: a transmembrane-specific substitution matrix. *Bioinformatics* **2000**, *16*, 760−766.
(84) Sandberg, M.; Eriksson, L.; Jonsson, J.; Sjöström, M.; Wold, S. New chemical descriptors relevant for the design of biologically active peptides. A multivariate characerization of 87 amino acids. *J. Med. Chem.* **1998**, *41*, 1.

CI050221U