

Proteochemometrics Modeling of the Interaction of Amine G-Protein Coupled Receptors with a Diverse Set of Ligands

MARIS LAPINSH, PETERIS PRUSIS, TORBJÖRN LUNDSTEDT, and JARL E. S. WIKBERG

Departments of Pharmaceutical Biosciences (M.L., P.P., J.E.S.W.) and Pharmaceutical Chemistry (T.L.), Uppsala University, Uppsala, Sweden; and Melacure Therapeutics AB, Uppsala Science Park, Uppsala, Sweden (T.L.)

Received June 25, 2001; accepted March 8, 2002

This article is available online at <http://molpharm.aspetjournals.org>

ABSTRACT

We have evaluated the proteochemometrics approach in the analysis of the interactions of a diverse set of organic ligands with subtypes of serotonin, dopamine, histamine, and adrenergic receptors. As used herein, proteochemometrics exploits affinity data for series of organic amines binding to wild-type amine G protein-coupled receptors, correlating it to descriptions and cross-description derived from the primary amino

acid sequences of the receptors and the computed structures of the organic compounds. We show that after appropriate data preprocessing, statistically valid models that have good external predictive ability can be created. Evaluation of the models gave important quantitative insight into the mode of interactions of the amine G protein-coupled receptors with their ligands.

Molecular recognition is central to all activities performed by living organisms. A detailed understanding of the mechanisms involved is the key to understanding the function of the cells, and their ability to respond to foreign substances. Whereas the recent cloning of the entire genomes of species constitutes a framework to the understanding of cell functions, effective computational tools are needed to be able to describe and analyze molecular interactions of the cell constituents. In the past, the main efforts have been directed toward 3D computational methods (Sansom, 1998; Flower, 1999; Wishart et al., 1999). However, there is a bottleneck in deriving accurate 3D structures of macromolecules, thus imposing restrictions on the use of 3D structure methods.

In two recent articles, we developed a new technology, termed proteochemometrics, and showed it to be useful for analyzing protein-ligand interactions (Lapinsch et al., 2001; Prusis et al., 2001). The approach is based on a unified description of the physicochemical properties of the primary amino acid sequences of proteins, and the description of the physicochemical properties of the ligands that may interact with the proteins. Our two previous studies were made on chimeric and mutated melanocortin and α_1 -adrenergic receptors; both of these receptors are G protein-coupled receptors (GPCRs) and their primary amino acid sequences show fairly high homology. Using the proteochemometrics approach in these cases yielded very robust models.

The aim of the present study was to develop the proteo-

chemometrics technology further and apply it to receptors of wider diversity. To this end, we considered using wild-type GPCRs from one of the subclasses of the GPCR superfamily. Amine GPCRs contain about 260 known members, 42 of which are human (<http://www.gpcr.org>; Horn et al., 1998). We here used a subset of the amine GPCRs for which suited functional activity data were available. The receptors used were subtypes of dopamine, serotonin, histamine, and α - and β -adrenergic receptors, and we demonstrate here the creation of successful proteochemometrics models for their binding of a wide diversity of organic compounds.

Materials and Methods

Principal Component Analysis

PCA is useful for extracting information from large data sets and it was used as a data-preprocessing method. PCA approximates a multivariate data matrix by finding a correlation structure that is present even if particular variables are not correlated. Data reduction is obtained by projecting the data onto a lower dimensionality variable space, called latent variables or principal components (PCs) (Wold et al., 1987). The significance of extracted components was assessed by cross-validation, as described previously (Eriksson et al., 1997). Principal component analysis was performed using SIMCA-P (Umetrics AB, Umeå, Sweden).

Partial Least-Squares Projections to Latent Structures

PLS is an extension of PCA that finds the relationship between a matrix of predictor variables, X , and a matrix of

Financial support was obtained from the Swedish MRC (04X-05975) and TFR (230-2000-291), and Melacure Therapeutics.

ABBREVIATIONS: 3D, three-dimensional; GPCR, G protein-coupled receptor; PLS, partial least-squares projections to latent structures; PCA, principal component analysis; TM, transmembrane; GRIND, grid independent descriptors.

dependent variables, Y , and was used for building proteochemometrics models. PLS has two objectives: to approximate the X and Y data matrices, and to maximize the correlation between them. Whereas the extraction of PLS components is performed stepwise and the importance of a single component is assessed independently, a regression equation relating each Y variable with the X matrix is created. For a detailed description of the PLS method, see Wold (1995). PLS analysis was performed using SIMCA-P (Umetrics AB).

The goodness of fit of PLS models was assessed by calculating R^2Y , and the predictive ability by calculating Q^2 using cross validation, as described previously (Eriksson et al., 1999). The R^2Y indicates the explained variance of the dependent variable Y . Provided that a sufficient number of predictor variables are used the R^2Y value would approach unity. However, the model in such a case may simply have accumulated chance correlations. Hence, the Q^2 value is the main criterion for assessing the quality of a model. Quantitative structure-activity relationship modeling comprising biological data is generally considered acceptable if Q^2 is higher than 0.4 (Lundstedt et al., 1998). Cross-validation was also used to assess the significance of each extracted PLS component; their extraction was terminated when a $Q^2_{\text{component}}$ below the threshold level 0.097 was encountered (SIMCA; Umetrics AB). Models were additionally validated by response permutations, as described previously (Eriksson et al., 1997).

Models were created both without and with variable selection, the latter being done by excluding descriptors with negligible variable importance in projection (VIP). VIPs characterize the contribution of descriptors for explaining Y accumulated over all model dimensions and were calculated as described previously (Eriksson et al., 1999). Terms with larger VIPs are the most relevant for explaining Y and models were improved by iteratively excluding descriptors with VIPs < 0.5 to 0.7 , essentially as described previously (Lapinsh et al., 2001).

Source of Data

The affinities of 23 organic compounds binding to 21 amine GPCRs were collected from the literature. The receptors were

cloned human and rat GPCRs representing subtypes of dopamine, serotonin, α - and β -adrenergic, and histamine receptors (Table 1). The 23 organic compounds included candidate and approved drugs, most of which were typical and atypical antipsychotics, as well as some that showed their highest affinities for adrenergic and histamine receptors (see Fig. 1 for a list of the compounds). The data consisted of 222 pK_i values for 20 compounds binding to 18 different amine GPCRs reported in five articles by Millan et al. (1998, 1999, 2000a,b,c), and 161 pK_i values for 11 compounds binding to 15 different amine GPCRs reported by Schotte et al. (1996). The pK_i values covered a range of more than 5 logarithmic units. As seen from Fig. 1, the pK_i values for some ligand-receptor combinations were missing, because all ligands had not been tested on all the receptors. Fifty-one of the observations overlapped between the data from the Millan and Schotte laboratories, allowing us to judge the consistency of the biological measurements; the residual standard deviation for the overlapping affinity values was 0.42 pK_i units. Still, for some cases, the difference was as large as 1 pK_i unit. For the overlapping data, we used the average of the reported pK_i values. In a few cases, the drug binding was reported to have a pK_i value < 6 . For these cases, we arbitrarily assigned pK_i a value of 5.5. The pK_i values used in the present study are shown in Fig. 1. The sequences for the receptors were retrieved from the Swiss-Prot database (Bairoch and Apweiler, 2000) under the accession identifiers listed in Table 1.

Generation of Physicochemical Descriptions of Receptors and Compounds

Description of Receptors. The physicochemical description of the amine GPCRs was derived from the amino acid sequences of their seven transmembrane (TM) regions. Restricting the analysis to the TM helices seemed reasonable because 3D modeling and mutagenesis studies have indicated that the amine binding sites of amine GPCRs are situated within the dihedral cleft located between these helices (Bikker et al., 1998; Jacoby et al., 1999). The amino acid sequences of seven transmembrane (TM1–TM7) regions were first aligned, using the alignments retrieved from the GPCRDB database (Horn et al., 1998). The amino acid positions corresponding to TM1 to TM7 were then selected (using the numbering of the A1AA human):

TABLE 1
Swiss-Prot identifiers of amine GPCRs used

Receptor	ID	Accession Number
D ₁	DADR_HUMAN	P21728
D ₂	D2DR_HUMAN	P14416
D ₃	D3DR_HUMAN	P35462
D ₄	D4DR_HUMAN	P21917
5HT _{1A}	5H1A_HUMAN	P08908
5HT _{1B}	5H1B_HUMAN	P28222
5HT _{1D}	5H1D_HUMAN	P28221
5HT _{1E}	5H1E_HUMAN	P28566
5HT _{1F}	5H1F_HUMAN	P30939
5HT _{2A}	5H2A_HUMAN	P28223
5HT _{2B}	5H2B_HUMAN	P41595
5HT _{2C}	5H2C_HUMAN	P28335
5HT ₆	5H6_RAT	P31388
5HT ₇	5H7_RAT	P32305
α_{1A}	A1AA_HUMAN, A1AA_RAT	P35348, P43140
α_{2A}	A2AA_HUMAN, A2AA_RAT	P08913, P22909
α_{2B}	A2AB_HUMAN	P18089
α_{2C}	A2AC_HUMAN	P18825
β_1	B1AR_HUMAN	P08588
β_2	B2AR_HUMAN	P07550
H ₁	HH1R_HUMAN	P35367

25 to 49, 62 to 86, 98 to 122, 145 to 166, 185 to 206, 273 to 292, and 309 to 328. The sequences were subsequently coded using the five z-scale descriptors, z1 to z5, derived by Sandberg et al. (1998). The Sandberg descriptors are the principal components of 26 different physicochemical measured and calculated properties of amino acids, and represent essentially hydrophobicity/hydrophilicity (z1), steric/bulk properties and polarizability (z2), polarity (z3), and electronic effects (z4 and z5) of the amino acids. In this way, a total of 795 variables characterizing the 159 selected amino acids confined in seven different descriptor blocks (herein termed R1–R7) representing seven TM regions were derived for each receptor.

Description of Organic Compounds. The 3D structures of the organic compounds were modeled in their protonated forms by using the Sybyl 6.5 software (Tripos Inc., St. Louis, MO). Electrostatic charges were calculated using the Gasteiger-Hückel method. A set of conformers was generated for each compound by randomly assigning angles to rotatable bonds. (For compounds that contained asymmetric nitrogen atoms and that thus were capable of becoming protonated from both sides, two sets of conformations were created.) Each conformer was subjected to simulated annealing using the initial temperature 1000°K, 1000-fs plateau time, and exponential annealing to 300°K in 1000 fs. Thereafter, the energy was minimized using the Tripos force field by using a conjugate gradient of 0.05, or a maximum of 10,000 iterations as termination criteria. Although at least fifteen conformers were always created for each compound, the generation of conformers was terminated when 10 successive cycles had not found any new conformer of lowest energy. In most cases, convergence under this criterion was reached after the generation of 20 to 50 conformers. Only the conformations that had energies of less than 4 kcal/mol above the lowest energy conformer were used.

The conformations were coded into GRid INdependent Descriptors (GRIND) (Pastor et al., 2000) using the ALMOND software (<http://miasrl.com>). GRINDs represent alignment independent descriptors that relate to the ability of a compound to form favorable interactions with independent pharmacophoric groups. The generation of these descriptors involves several steps. Molecular interaction fields (MIFs) are first calculated on grid points surrounding the molecule. The grid nodes showing the energetically most favorable interactions with the probes are then selected and the distances between the selected nodes representing the same type of MIFs (autocorrelograms) or different types of MIFs (cross-correlograms) are calculated. Descriptors are thereafter generated by calculating the products of the energy values for grid point pairs that show the maximal value within specified distance ranges ('smoothing windows'). We

used the three default probes, DRY, O, and N1, for computing the MIFs. These probes represent hydrophobic, hydrogen bond acceptor, and hydrogen bond donor groups, respectively, and will be referred to as DRY, O, and N1 interactions. Default parameters were also used for grid spacing (i.e., 0.5 Å), as well as for the number of extracted nodes (i.e., 100). Moreover, the default width (0.8 Å) was used for the smoothing window. Thus, according to the selected settings, 52 descriptors were obtained for each of the three autocorrelogram (i.e., DRY-DRY, O-O, and N1-N1) and three cross-correlogram (i.e., DRY-O, DRY-N1, and O-N1) variable blocks, giving a total of 312 descriptors confined in six different descriptor blocks. (These descriptor blocks are herein termed L1–L6.)

To identify any eventual outliers of the conformers, the conformational space of each compound was screened by applying PCA on the descriptor data. The significant principal components (PCs) were thus extracted from the descriptor data of all conformers of each compound and the distance to the PC model (i.e., DModX) was assessed for each conformer (Eriksson et al., 1999). Conformers with DModX larger than the critical distance for a membership probability of 5% or less were considered being outliers. The analysis showed that the lowest energy conformer for each compound, in all cases, did not exceed the critical distance, and were accordingly selected as the representative of the compound. Using the data set of lowest energy conformation of each compound made it possible to remove 76 descriptors because of their lack of variance (i.e., the descriptors had zero values).

Data Preprocessing

Principal Component Analysis of Descriptor Blocks. The data were preprocessed by use of PCA. To preserve the interpretability of the models, we applied PCA separately to portions of the descriptors rather than applying it over the whole data set. PCA was thus applied to each of the L1 to L6 block of descriptors for organic compounds. This seemed reasonable as the information given by different probes represent the capacity to form distinct types of noncovalent interactions. All variables were mean-centered and 10 principal components were extracted for each of the six blocks. Thus, preprocessing the L1 to L6 blocks reduced the total number of descriptors from 236 to 60. For each block, 96 to 98% of the total variance of the original data were explained by the principal components.

In a similar fashion, PCA was applied separately on the seven R1 to R7 descriptor blocks of the receptors. Fifteen principal components

	D ₁	D ₂	D ₃	D ₄	5HT _{1A}	5HT _{1B}	5HT _{1D}	5HT _{1E}	5HT _{1F}	5HT _{2A}	5HT _{2B}	5HT _{2C}	5HT ₆	5HT ₇	α _{1A}	α _{2A}	α _{2B}	α _{2C}	β ₁	β ₂	H ₁
Amperozide	6.7	6.3	6.4	5.6	<6					8.2		5.6			7.4	5.9					
Clozapine	7.8	6.9	6.6	7.4	6.9	6.4	6.2	6.4	6.9	8.4	8.8	7.3	8.0	7.3	8.9	7.0	7.7	7.9	5.3	5.8	9.6
Fluparoxan	<6	<6			6.8	<6	<6			<6			<6		6.3	8.1	8.9	8.8			
Fluspirilene			9.0	8.4	8.5	7.3	6.0	5.3	5.7		8.0					6.2	7.3	7.2	6.2	6.3	7.9
GGR218231	<6	7.2	9.0	5.0	6.8	<6	<6			<6	<6	<6			<6	6.4					
Haloperidol	8.5	9.1	8.3	8.4	5.8	6.6	5.3	5.3		6.9	<6	5.0	<6	6.6	8.2	5.9	6.2	5.9	5.3	5.3	6.1
L741626	6.1	8.4	7.2	6.5	<6	<6	<6			6.5	6.2	<6			6.6	5.9					
MDL100,907	6.6	<6	4.9	6.2	5.0					9.2		7.6			6.6	5.0					
Ocaperidone	7.6	9.2	8.6	8.5	8.0					10.0		8.0			9.7	7.7					
Olanzapine	7.9	7.9	7.4	7.6	5.7	6.2	6.3	5.7	6.5	8.8	8.1	8.1	8.0	7.0	7.7	6.1	6.7	6.7	5.3	5.3	9.2
ORG5222	8.8	8.8	8.8	8.7	8.2	8.4	8.1	8.0		9.8	9.8	9.6	9.0	9.1	9.2	8.2	9.3	7.9	5.3	5.3	9.8
Pipamperone		6.9	6.6	8.3	5.6	6.8	6.2	5.3	5.3	8.3						6.1	7.5	6.5	5.3	5.3	5.6
Raclopride	5.0	8.9	8.8	5.4	5.1					6.7		5.0			5.0	5.0					
Risperidone	7.7	8.4	7.9	8.0	6.5	8.0	6.9	5.9	5.9	9.3		7.8			9.3	7.6	8.1	8.0	5.3	5.3	7.6
S16924		7.5	7.3	8.0	8.7					9.0	8.4	7.7	7.6	7.1	8.7	6.8	7.4	7.0	5.0	5.7	
S18327	7.2	7.1	7.0	8.1	7.4					8.5		6.4			9.0	7.0					
S33084	6.3	7.5	9.6	5.7	<6	6.1	6.9			6.0	6.8	7.1			6.9	<6					
Seroquel	6.7	6.5	6.4	5.7	6.5	5.7	5.3	5.9	5.6	6.8	7.0	6.2	5.5	6.6	8.0	5.3	7.0	6.5	5.3	5.3	8.7
Sertindole	7.7	8.4	8.1	7.7	6.5	7.2	7.0		6.4	6.4		8.7			9.9	6.4	6.3	6.3	5.3	5.3	6.9
Tiospirone	7.4	8.7	8.6	8.3	8.4					9.3					9.6	6.7					
Yohimbine		6.4	5.4		7.3	6.8	7.6			<6		<6			6.7	8.3	8.7	9.6	<6	<6	
Ziprasidone	8.5	8.3	8.0	7.6	8.4	9.0	8.3	6.4		9.2	8.4	8.7	7.6	8.2	8.9	6.6	7.3	7.1	5.6	5.3	7.8
Zotepine		8.1	8.2	7.4	6.5	7.3	7.2	6.5		8.6						6.7	8.2	6.8	5.4	5.6	9.2

Fig. 1. Binding affinity (pK_i) of compounds for human and rat amine GPCRs used in the present study. Gray shading indicate that data are from a rat receptor; all other data were from human receptors.

were extracted from each block, resulting in the coverage of between 94 and 98% of the original variance of the data. Although in some cases the last components were not significant, these PCs were still retained because they reduced the unexplained variance of some particular descriptors. Preprocessing of the R1 to R7 blocks reduced the number of descriptors from 795 to 105.

Cross-Description of Ligand-Receptor Complexes. After having mean centered all the descriptors, cross-terms of variables were generated by calculating the products for each variable characterizing the ligand and each variable characterizing the receptor. Centering of these cross-terms was then performed and the deviation from the mean of each of these cross-terms were additionally calculated, generating a further set of descriptors. [The latter descriptors were obtained to account for possible nonlinear relationships between cross-terms and the biological data; e.g. the possibility exists that only one end of a physicochemical property scale is correlated with the biological response. See Eriksson et al. (2000) and references therein for a further discussion related to this topic.] Thus, in this way, 42 blocks of cross-descriptors (R1L1 to R7L6) containing cross-terms and cross-term deviations from the means (i.e., 300 descriptors in each block amounting to a total of 12,600 descriptors) were obtained.

Centering and Scaling. Descriptors were mean-centered before the PCA and PLS and they were scaled to unit variance before the PLS.

Estimation of Contribution of Descriptor Blocks for Ligand-Receptor Affinity

Formulas were developed for estimation of the contribution of descriptor blocks for receptor-ligand affinity. All the descriptors and the corresponding pK_i value for each binding experiment (ligand-receptor combination) is called "an observation", which is expressed mathematically by a vector of numbers o . Similarly, we can express each descriptor by a vector of numbers d . The particular numerical value that a descriptor d has for an observation o is defined as "the variable" x_{od} . Thus, the variable x_{od} represents the intersection of vectors o and d .

The descriptors can be grouped into sets or "descriptor blocks" (B_p) sharing some common property p . Examples of p can be TM region, probe atom of MIFs, or all cross-terms between one TM region and one probe atom. The descriptor block of observation o (b_{op}) would then be the set of all variables that simultaneously belong to the observation o and descriptor block B_p .

The contribution of each descriptor d for explaining the drug binding affinity could be assessed from its PLS coefficient C_d (Eriksson et al., 1999). Although PLS uses a latent variable framework, and although the PLS components are extracted stepwise, the regression coefficients show the overall relationship of the predictor variables to the dependent ones. Accordingly, we can calculate the contribution (Δ_{od}) of each variable x_{od} to the pK_i value for each observation o by using Eq. 1:

$$\Delta_{od} = \Delta(pK_i)_{od} = C_d \times (x_{od} - \bar{x}_d) \quad (1)$$

where Δ_{od} is the contribution of variable x_{od} to the pK_i value of observation o ; C_d is the PLS coefficient of descriptor d ; x_{od} is the value of descriptor d for observation o ; and \bar{x}_d is the mean value of all variables of descriptor d .

Because the contributions of single variables are additive, they can be used to calculate the contribution (Δ_{op}) of any descriptor block b_{op} to the pK_i values for each observation o according to Eq. 2:

$$\Delta_{op} = \sum_{d \in b_{op}} \Delta_{od} \quad (2)$$

where Δ_{op} is the contribution of descriptor block b_{op} to the pK_i value of observation o and Δ_{od} is the contribution of variable x_{od} to the pK_i value of observation o .

Estimation of Affinity and Selectivity Contributions

To gain a detailed understanding of each particular compound's descriptors and cross-term contributions for its affinity and selectivity, we elaborated two contribution estimates: contribution of property p for affinity of compound s (α_{sp}) and contribution of property p for selectivity of compound s (σ_{sp}). The contribution for affinity α_{sp} is the average of contributions Δ_{op} of all observations that include compound s as ligand. To find the contribution for selectivity, Δ_{op} values are plotted versus the calculated pK_i values for each observation o that includes compound s as ligand. Thus, 21 values for each receptor included in the data set are plotted. The slope of the trendline of this plot corresponds to the contribution for selectivity σ_{sp} . Examples of α_{sp} and σ_{sp} are shown in Fig. 4.

The contribution of every single amino acid within every separate receptor sequence for explaining the affinity of every particular compound could be assessed by first replacing the actual values of the z-scale descriptors of the particular amino acid by their mean values for the given sequence position in the studied amine receptors (i.e., in the present case by replacing with zeros, because variables were mean centered). The principal components were then computed for the receptor with the thus modified amino acid and the cross-descriptions were derived for the ligand. The binding affinities were thereafter calculated from the PLS model and the affinity contribution, Δ_{on} , of the particular amino acid was estimated from Eq. 3:

$$\Delta_{on} = (pK_i)_o - (pK_i)_{on} \quad (3)$$

wherein $(pK_i)_o$ and $(pK_i)_{on}$ are the predicted affinity for the ligand-receptor combination o without and with, respectively, the modification in sequence position n .

Cluster Analysis

Cluster analysis of compounds was performed based on the above selectivity measures, σ_{sp} . Because the overall selectivity of the different compounds were not the same, we scaled the σ_{sp} measures of a compound by the standard deviation of the calculated affinity values of that compound for all the receptors. Thus, multiplying a σ_{sp} measure with the standard deviation for calculated affinities resulted in a non-normalized σ_{sp} measure in pK_i units. Hierarchical clustering was performed simultaneously using both the σ_{sp} and non-normalized σ_{sp} measures by applying Ward reciprocal nearest neighbors linkage. Cluster analysis and cluster significance analysis was performed using Tsar software, as described previously (Tsar 3.3. for Windows; Accelrys, San Diego, CA).

Results

Creation of a Proteochemometrics Model

We first tried PLS analysis using only descriptors of receptors and organic compounds (i.e., without using any cross-

description). A PLS model with one significant latent variable was obtained describing $R^2Y = 0.41$ of the variance of Y and having a predictive ability $Q^2 = 0.31$ (using seven cross-validation groups). Extraction of additional components increased R^2Y to 0.60. However these components were insignificant because the Q^2 value for none of them was positive. Thus, the linear combination of the receptors' and compounds' descriptor blocks could only partially explain the data, something which could have been anticipated from the extreme nonlinearity of the receptor-drug 'affinity surface' (see Fig. 1).

Accordingly, in our next attempt, we used the cross-description as well as the receptors' and compounds' descriptor blocks in the modeling. To get an objective assessment of the predictive ability of obtained PLS model, cross validation was performed using five to nine cross-validation groups. Thus, after extracting five PLS components, the Q^2 values ranged between 0.60 and 0.67 (at an explained Y variance $R^2Y = 0.96$), whereas after extracting seven components ($R^2Y = 0.98$), Q^2 ranged between 0.67 and 0.73. However, for some components, the $Q^2_{\text{component}}$ was below the 0.097 limit, thus indicating that the particular component might explain the variance of Y by a chance correlation. Moreover, despite the increased Q^2 values in the seven-component model, the extremely high R^2Y indicated the presence of an overfit. Hence, variable selection was performed, which removed descriptors with little correlation to Y (see *Materials and Methods*), and it resulted in further improvement of the predictive ability of the model. Thus, a five-component model was obtained using 3335 X variables. Cross validation with seven groups gave an estimate of $Q^2 = 0.75$. Moreover, a reduced margin between the Q^2 and the $R^2Y = 0.92$ was obtained. The results of the model are shown graphically in Fig. 2, where the observed versus calculated pK_i values are plotted.

To explore the robustness of the model, cross validation was performed using lower numbers of cross-validation groups. Cross-validation with three groups indicated similar predictive ability ($Q^2 = 0.78$) as obtained above, whereas

cross-validation with two groups gave a Q^2 value of 0.64. Thus, even when only half of the data were included in the model, it showed reasonable predictive ability.

The model was further validated by response permutations, which gave a high value for the R^2Y intercept (0.61). This shows that a high R^2Y can be obtained with random descriptor data. This is not surprising, however, and can be explained by the large number of descriptor variables used in the model. However, the Q^2 -intercept obtained was -0.04 , thus showing that random descriptor data could not yield models with any predictive ability. Thus, we conclude that the present approach provided a valid model.

In the following, the model that was based on cross-description and receptors' and compounds' descriptor blocks will be referred to as 'the model' and will be the one used in all subsequent analysis unless otherwise stated.

Assessment of External Predictability

To further validate the modeling approach, we created models in the same way as above (i.e., with variable selection) on subsets of the data. This was done to assess the external predictability of the approach. Thus, we excluded each fifth observation from the data and created a model from the remaining data. A five-component model using 3002 X -variables was obtained on the reduced data set, which had essentially the same statistical validity ($R^2Y = 0.93$, $Q^2 = 0.75$) as the model obtained on the full data set. The model obtained on the reduced data set was then used to predict the excluded data; the standard deviation of errors for prediction for the observations not included in model building was 0.69 pK_i units. The correlation of predicted versus observed pK_i values obtained from this analysis is shown in Fig. 2.

Interpretation of the Model

Contribution of TM Regions for Binding of Organic Compounds. To analyze the contribution of each of the TM helices of each of the studied receptors to the binding of each of the 23 studied compounds, we used the PLS regression equation of the model. Thus, the contribution Δ_{op} of descriptor blocks of each drug-receptor combination were calculated (see *Materials and Methods* for details). Results for the Δ_{op} of the TM regions of each receptor for binding series of compounds are shown graphically in Fig. 3. Inspection of the figure reveals different aspects of the drug receptor interactions. First, the overall contribution to affinity of a drug of a certain region can be assessed by inspecting the deviation of the Δ_{op} from the average affinity for all drugs over all receptors (i.e., the zero line). For example, for the β_2 receptor, it can be seen that the TM regions 2, 3, 4, 6, and 7 are responsible for the overall low drug affinities of the 23 compounds. This contrasts, for example, with the α_{1a} receptor, in which these regions show major positive contributions to the affinity, thus explaining the high affinities of the majority of compounds for this receptor. Moreover, for both the α_{1a} and β_2 receptors, the TM regions 1 and 5 contribute little to the drug affinities.

It can also be seen from the graphs of Fig. 3 that the spreads of Δ_{op} of the different TM regions for the various ligands differ largely between different TM regions and different receptors. When the spread is small for a TM region, it does not discriminate much between different drugs. A large span, of course, indicates conversely that a TM region was

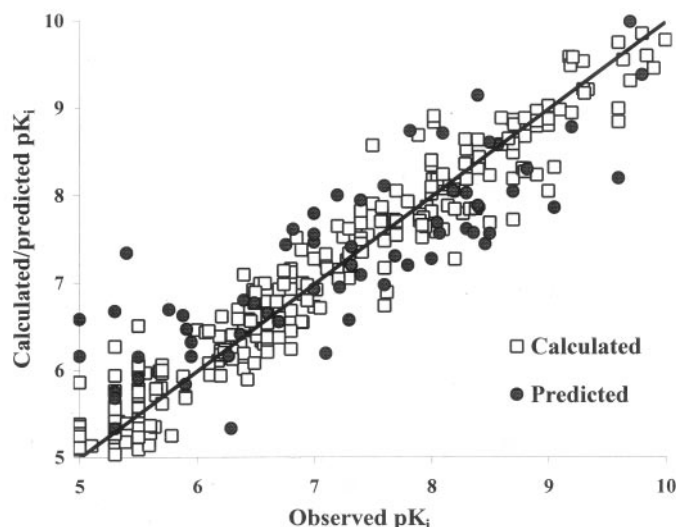


Fig. 2. Correlation of calculated and predicted pK_i versus observed pK_i values derived by the proteochemometrics model for amine GPCRs. The goodness of fit is indicated by the correlation of the calculated versus observed pK_i values, whereas external predictive ability is indicated by the correlation of the predicted versus observed pK_i values.

important in affording selective binding of compounds. Thus, the average level for a TM region could be taken as the overall contribution of that region for the drug binding, whereas the deviation from that average would indicate its importance in the creation of selective interactions. Moreover, a closer inspection of Fig. 3 reveals both similarities and differences in patterns of contribution for TM-regions of the different receptors. For example, the D_2 and D_3 receptors show similar patterns, whereas the pattern for the D_4 receptors is considerably different from the D_2 and D_3 receptors.

Contribution of Distinct Types of Noncovalent Interactions for Selective Binding of Organic Compounds.

To bring the analysis further, we attempted to identify the importance of different types of noncovalent interactions between compounds and receptor regions for determining the variation in the compounds' ability to bind to the receptors by calculating Δ_{op} values for cross-descriptor blocks between each of the seven TM helices and each of the DRY, O, and N1 molecular interaction fields (see *Materials and Methods* for details). The sets of Δ_{op} values obtained for each compound for its interactions with different receptors could be used thereafter in the characterization and comparisons of compounds.

In the further analysis, we aimed to create an approach that could separate the importance of descriptor blocks for their contribution to creation of selectivity of a compound from their importance for contribution to overall affinity for the given set of receptors. To visualize the different types of contributions, we plotted the pK_i values of each compound versus the Δ_{op} values for each TM region and molecular interaction field. An example of this is given in Fig. 4 for the interactions of sertindole with TM2 of the amine receptors. Thus, one measure affording characterization of the data would be the average Δ_{op} value for the compound for all receptors (indicated by the horizontal lines in Fig. 4). This average would thus give an estimate for the average ability of the compound to bind to that TM region of all the receptors by virtue of a particular molecular interaction type.

However, upon inspecting the plots of the type shown in Fig. 4, it can be seen that for some descriptor blocks, the Δ_{op} values strongly and positively correlate to the pK_i values, whereas for others, the correlation is weaker. The presence of

a clear trend would thus indicate that differences in the chemical properties of the TM regions of the receptors influenced differently the binding affinities. By contrast, no clear trend indicates that chemical differences in the region do not play a dominant role for creation of the variation of the compounds' affinities.

Accordingly, the compounds could be compared by using two types of normalized measures; the average Δ_{op} for each TM region and molecular interaction (herein termed contribution for affinity, α_{sp}), and by the slopes of the trend lines for Δ_{op} versus pK_i values, for each TM region and molecular interaction (herein termed contribution for selectivity, σ_{sp}). It shall be observed that the sum of all α_{sp} values for a compound is composed of the deviation of the calculated pK_i value from the mean of the all the calculated pK_i values of all the compounds in the data set. Moreover, the sum of σ_{sp} for a compound from all descriptor blocks would correspond to unity.

To visualize all α_{sp} and σ_{sp} values, we used radar plots. An example of such a plot given in Fig. 5 for sertindole. Figure 5, top, shows the radar plot for the α_{sp} values. As seen the analysis indicates that the major contributors for creation of affinity are N1 type interactions. Figure 5, bottom, shows the plot for σ_{sp} values. As can be seen, the pattern of σ_{sp} values is distinctly different from that of α_{sp} values. According to the analysis, hydrophobicity (i.e., DRY-type) of interactions at TM2, 6, and 7 is important. Inspection of the α_{sp} plots for all compounds indicated, however, that the measure did not delineate very clearly the different receptor regions. Rather, the analysis showed that some interaction type dominated over the other interaction types for the particular compounds, without any of the TM regions being particularly favored. These results would thus indicate that α_{sp} measures characterize mainly properties of the compound, whereas σ_{sp} values afford characterization of ligand-receptor properties.

Comparison of Compounds. Thus, according to the above results, the profiles resulting from α_{sp} values and σ_{sp} values of all compounds would afford characterization of the compounds' behavior toward all the receptors included in analysis. Using these measures, it might thus be possible to arrange compounds according to similarities or dissimilarities in their binding modes. However, using the α_{sp} values, a

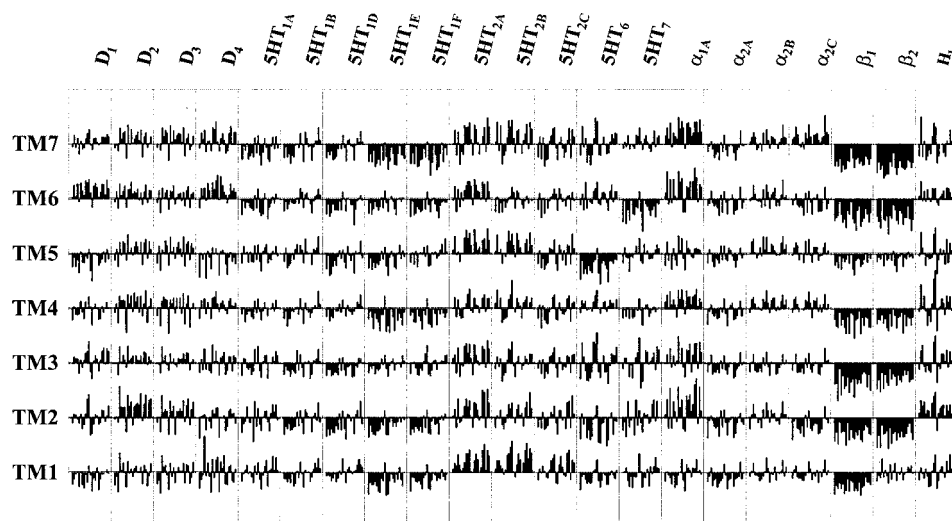


Fig. 3. Contribution of TM regions in amine GPCRs for creation of receptor-ligand affinity according to the proteochemometrics model. Shown are data for all amine receptors included in the analysis; for each receptor, the bars represent the Δ_{op} values for 23 compounds for each indicated TM region, plotted in the same order as the compounds are listed in Fig. 1. Moreover, the seven horizontal lines shown for each TM region represent the mean pK_i value over the whole data set. Distances between horizontal lines correspond to one pK_i unit.

major discriminating factor would to a large extent lead to the arrangement of compounds into high- and low-affinity binders, which of course would be a quite trivial result. By contrast, the σ_{sp} values have removed the contribution of average affinity and instead give information on how much the chemical variation in different regions of the receptors contributes to the variation of the drugs' affinities.

We therefore attempted to cluster the drugs by using the σ_{sp} measure (see *Materials and Methods* for details). The resulting dendrogram revealed four clusters at 50% amalgamation distance (Fig. 6); the centers of the clusters were represented by 1) haloperidol, 2) clozapine, 3) tiospirone, and 4) yohimbine. Cluster significance analysis for four clusters showed 0.91, 1.00, 1.00, and 0.70 probability at a 95% confidence level.

In Fig. 7, the radar plots of the σ_{sp} measures for the center compounds are shown. As can be seen from the figure, the

compounds show both similarities and dissimilarities in their interactions with the receptors. Thus, haloperidol, clozapine, and tiospirone, but not yohimbine, show high influence from the DRY-interactions with TM6 and TM7. By contrast, for yohimbine, the most prominent feature is O-interactions with TM6 and TM7. Besides the TM6 and TM7 DRY-interactions, tiospirone also shows prominent DRY-interactions with TM2 and TM4. For all compounds, the N1-interactions show lower importance. Moreover, the overall profiles of haloperidol and clozapine were quite similar.

Assessment of the Contribution of Single Amino Acids for Ligand Affinity. The contribution of every single amino acid included in the analysis for every single receptor sequence for explaining the binding affinity of every one of the studied organic compounds could be assessed by using Eq. 3 (see *Materials and Methods*). Its use is illustrated in Fig. 8, which depicts the affinity contributions of amino acids in A1AA_HUMAN (α_{1a}), D2DR_HUMAN (D_2), and 5HT_{2A}_HUMAN (5HT_{2A}) receptors for the binding of every one of the 23 compounds evaluated herein. (Numbering of compounds on the ordinates in Fig. 8 corresponds to the numbering in Fig. 1.) Inspection of Fig. 8 reveals that relatively few sequence positions show a major contribution in explaining the compound affinities. It is also apparent from the figure that some of the amino acids in the respective receptor contribute mainly with high affinity, whereas some other amino acids contribute with low affinity, over the whole compound set. However, other amino acid positions show large variations in their affinity contributions for the different compounds. The latter positions may be regarded as the ones that segregate the compounds into high- and low-affinity binders for the particular receptor. The

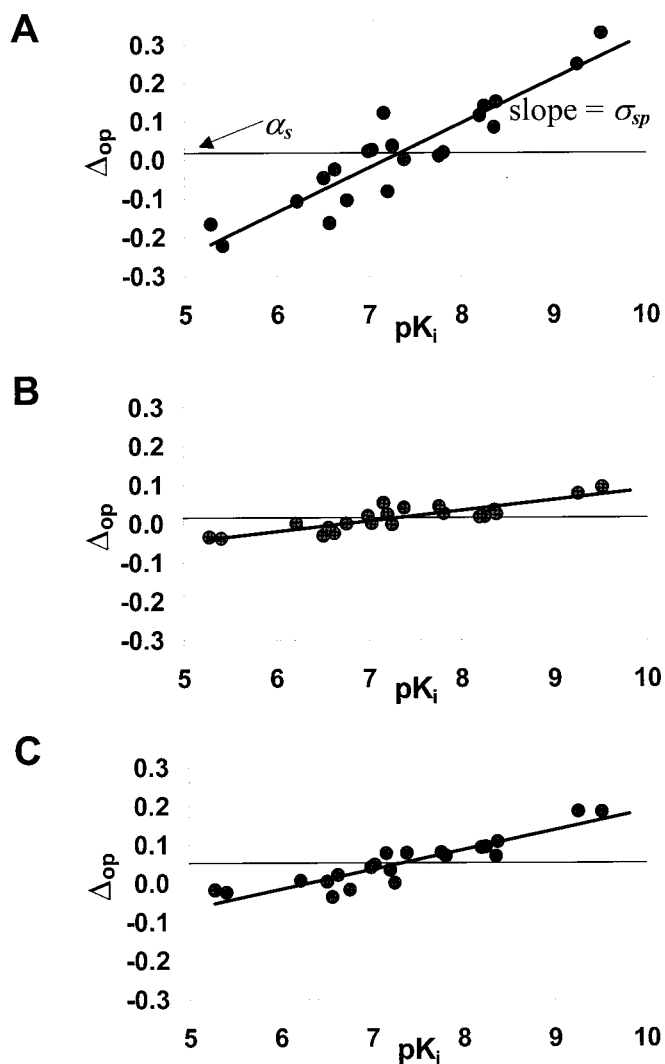


Fig. 4. Plots of Δ_{op} values for TM2 region for different molecular interactions fields versus the calculated pK_i for the binding of sertindole to 21 amine GPCRs derived from the proteochemometrics model. The horizontal lines in each represent the average Δ_{op} for the respective interaction and correspond to the α_{sp} measure. Also shown are regression lines, from which slopes corresponding to the σ_{sp} measures were derived (see text for details). represent the Δ_{op} values for DRY (A), O (B), and N1 (C) fields; the α_{sp} values for A-C were 0.018, 0.016, and 0.054, respectively. The corresponding σ_{sp} values are 0.115, 0.027, and 0.051, respectively.

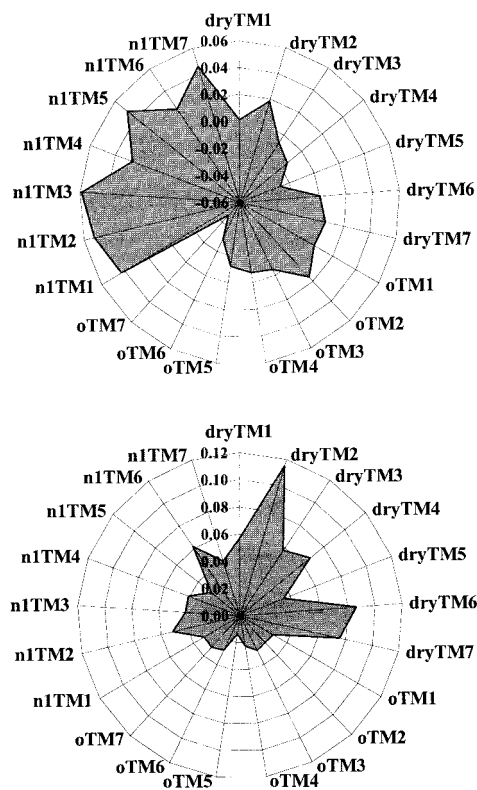


Fig. 5. Radar-plot representation of α_{sp} values (top) and σ_{sp} values (bottom) for sertindole, for the DRY (dry), O (o), and N1 (n1) interactions with each of the seven TM regions of the amine GPCRs.

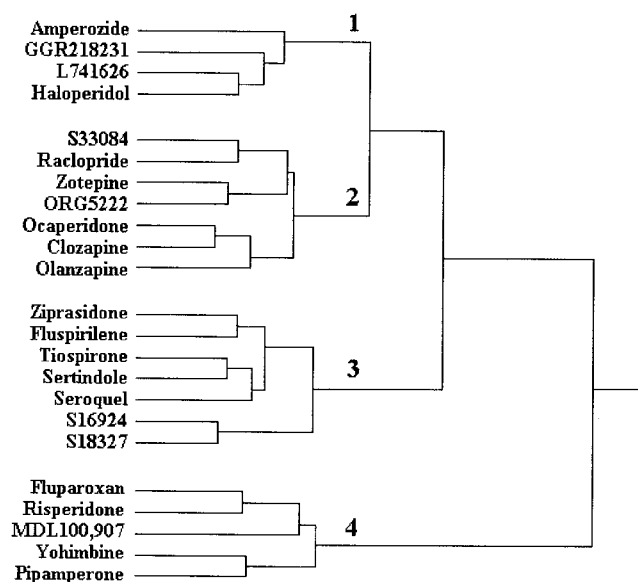


Fig. 6. Hierarchical clustering of organic compounds based on σ_{sp} measures (see *Materials and Methods* for details).

identical evaluation of all the 21 receptors using Eq. 3 yielded similar but distinct patterns for every one of the receptors (data not shown).

Discussion

We have here developed further our proteochemometrics approach for the analysis of protein-ligand interactions. We previously applied it to chimeric MC₁/MC₃ receptor subtypes showing about 45% sequence homology, as well as to chimeric

and point mutated α_1 -adrenoceptor subtypes showing about 45 to 50% sequence homology. In the present study, we obtained models using a set of amine GPCRs having fairly low sequence homology; the overall homology to the consensus sequence of the amine GPCRs used ranged from about 20 to 30%. For the TM regions, the sequence homology to the consensus sequence of amine GPCRs ranged from 35 to 45%.

The validity of our present proteochemometrics model for the amine receptors is assured from cross-validation, validation by response permutation and external prediction. Thus, all state-of-the-art methods for model validation in PLS showed the validity of the model. Moreover, its quality as assessed by Q^2 was very high when one considers the nature of the biological measurements used in the collection of the data. Thus, also in the present case proteochemometrics has proven a very sturdy method.

The goal of our study was to develop a method that could make predictions for novel drug-receptor combinations (i.e., before experimental testing) and to create models that are interpretable in a physicochemical sense. It seemed plausible that the use of physicochemical descriptors of both receptors and ligands would make it possible to create models that could be applied onto new proteins as well as new ligands. Our results using external prediction indicate that this is indeed feasible. Moreover, the proteochemometrics approach could be applied onto wild-type receptors, thus avoiding the necessity to construct mutated or chimeric receptors.

One very important finding was that creating a model using descriptors of receptors and ligands only resulted in inferior models, whereas when the cross-descriptions were included the models improved drastically. It is conceivable that our models are mainly based on the complementarity of ligand-receptor descriptions, rather than being just linear

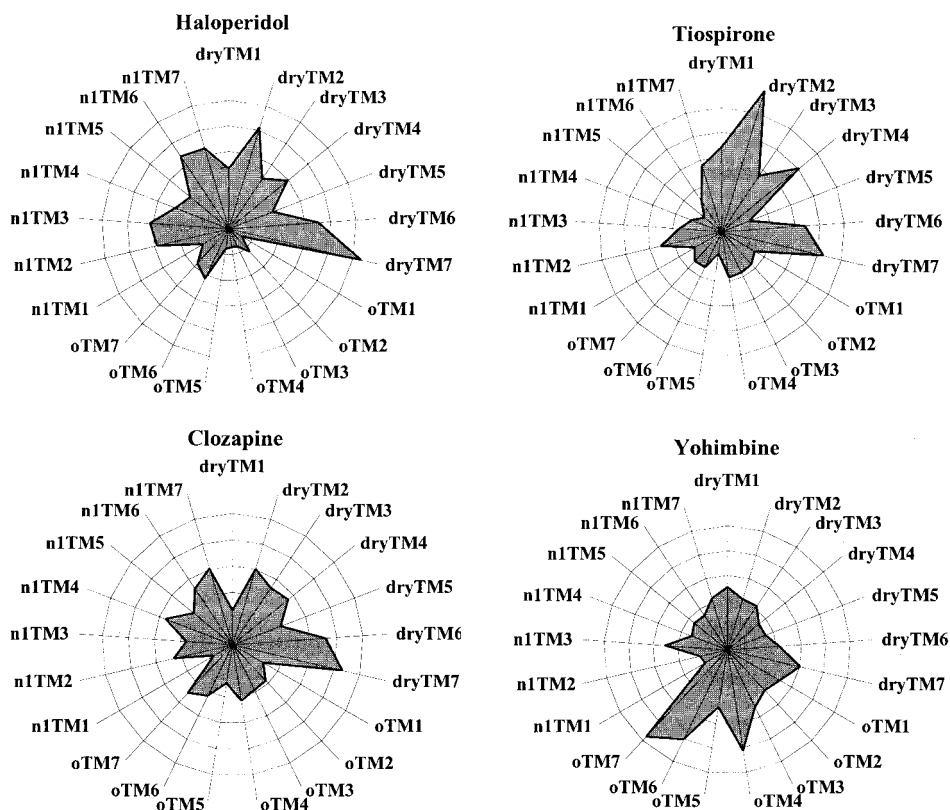


Fig. 7. Radar-plot representation of σ_{sp} values of haloperidol, tiospirone, clozapine, and yohimbine. Scaling is identical to that in Fig. 5, bottom.

combinations of separate receptor and ligand descriptions. It seems quite reasonable that the combined use of receptor and ligand descriptors condenses the over all information confined in the separate ligand and receptor descriptions and in their cross-interaction, thus providing an explanation why the present good models are possible to obtain.

However, the approach of using cross-descriptions necessitated preprocessing of the data. Thus, if cross-descriptions had been calculated on the original descriptors of the present data set almost 400,000 cross-terms would have had to be generated. It was obvious that any calculation on such a large descriptor set would be computationally very difficult. Accordingly, we considered PCA for data reduction. However, applying PCA over the whole data set would make the models less useful for interpretations. We therefore elected to perform PCA on the data blocks. It seemed reasonable to apply PCA separately onto the data blocks representing each separate molecular interaction field and receptor region, because this would later make it simple to derive tools for

analysis of the importance of ligand chemical property types and receptor regions for the ligand binding.

We were here able to separate two types of contributions that are involved in the receptors' binding of ligands. These were the α_{sp} values and σ_{sp} values. The α_{sp} values characterize the interactions that contribute to the creation of the affinity of a ligand to a set of receptors. The σ_{sp} values describe the importance of particular interactions for creation of ligand specificity for a set of receptors. Information derived by these measures would be immediately useful (e.g., in ligand design, if one wanted to create new molecules with higher affinity, or improved selectivity). A very interesting finding was that the chemical properties and topology involved in creation of affinity and selectivity seem to be different for the present series of compounds and receptors (compare Figure 5). This in turn would indicate the feasibility to optimize affinity and selectivity properties of ligands independently. Moreover, the knowledge might be useful in

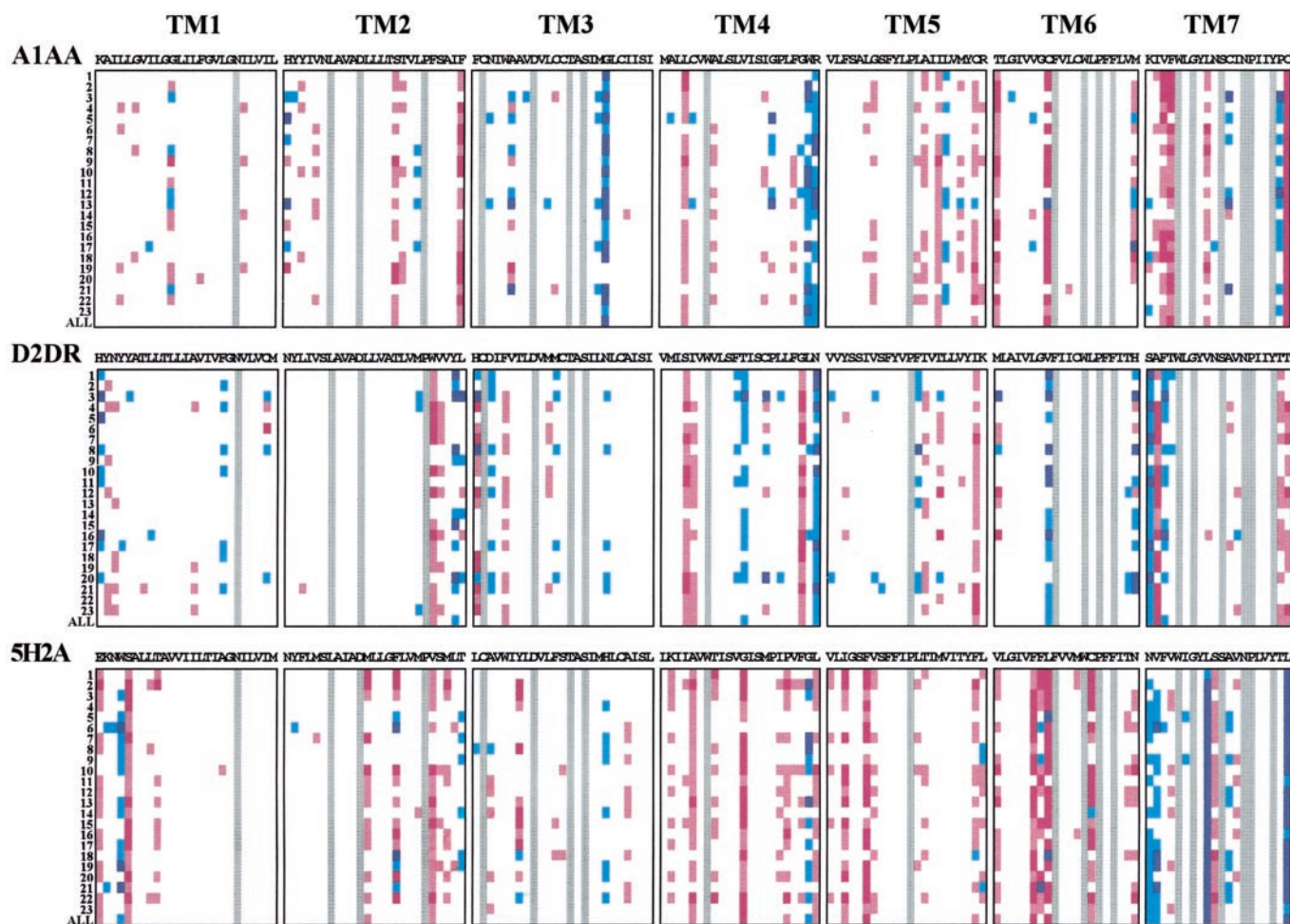


Fig. 8. Contribution of receptor sequence positions for explaining the respective binding affinity of 23 organic compounds to A1AA_HUMAN (A1AA; human α_1), D2DR_HUMAN (D2DR; human D_2), and 5H2A_HUMAN (5H2A; human 5HT_{2A}) GPCRs, according to the proteochemometrics model. Compounds are numbered in their alphabetical order, in the same way as shown in Fig. 1. Each amino acid position is color-coded according to its contribution to (relative) ligand-receptor affinity (determined using Eq. 3) as follows: dark blue, $\Delta_{on} < -0.1$ pK_i units; light blue, $-0.1 \leq \Delta_{on} \leq -0.05$ pK_i units; light red, $0.05 \leq \Delta_{on} \leq 0.1$ pK_i units; and dark red, $\Delta_{on} > 0.1$ pK_i units. White coding indicates amino acids that contribute only marginally in explaining binding affinity (i.e., $-0.05 < \Delta_{on} < 0.05$ pK_i units). Gray coding indicates amino acids that were conserved in the whole data set (as well they are conserved in all amine GPCRs), and whose contribution for explaining the differences in compound affinities accordingly can not be computed by the proteochemometrics model. The aligned amino acid sequences from transmembrane regions TM1 to TM7 included in the analysis are indicated on top of each.

optimizing other properties of a ligand (such as, for example, its membrane penetrability).

Another important point of the present analysis is that information is gained from multiple regions of the receptors. It is likely that the present analysis include both direct and indirect receptor interactions (e.g., indirect interactions might account for conformational changes, helix packing, and alike). Such effects may be very important in creation of drug affinity and selectivity and would be difficult to analyze using a 3D receptor modeling approach.

The analysis further indicates that the compounds explore different chemical spaces in the receptors in their creation of affinities. Moreover, because it was possible to cluster compounds into four groups according to their selectivity determining properties, it seems that members within each of these groups explore common chemical spaces. A closer inspection of the radar plots of Fig. 7 shows that TM7 and TM6 are important for all compounds, albeit for different molecular interactions (see *Results*). For some compounds, the TM2 and TM4 regions also show important DRY-interactions. Further inspections of the plots reveal other differences, as well as similarities between the groups.

An additional important result of the present study was the development of an approach that could assess the contribution of every single amino acid in the receptor for the affinity of every single compound included in the model. The data generated from this analysis (shown in part in Fig. 8) showed that very distinct roles exist for particular amino acid positions for the ligand affinity and selectivity of the different GPCRs.

Previous 3D modeling and mutagenesis studies localize the catecholamine binding sites somewhere in between TM3, TM4, TM5, TM6, and TM7 of the adrenoceptors (Bikker et al., 1998). A conserved aspartate of TM3 was invariably suggested to act as the counterion to the protonated catecholamine. Conserved serines of the TM5 were also suggested to be involved in hydrogen bondings to the hydroxygroups of the catechol ring. Similar placements of the natural amines as in the adrenoceptors have also been proposed for the dopamine and serotonin in their receptors (Bikker et al., 1998). However, some ambiguity does exist in the roles of amino acids of other TM regions in the binding of the natural amines (see Bikker et al., 1998). Some studies have also been devoted to the docking of antagonist ligands into amine receptors. The picture emerging from these studies is also not entirely clear and sometimes even quite contradictory; e.g., modeling and mutagenesis studies suggest the placement of the binding pocket of pirenzepine in alternative binding pockets of the muscarinic receptors (Fanelli et al., 1995; Bourdon et al., 1997) and it is not an easy task to resolve such ambiguities using 3D modeling. It has also been suggested that antagonist ligands for amine receptors may bind to alternative binding pockets in the receptors (Schwartz 1994; Jacoby et al., 1999). Thus, according to the so-called three-binding site hypothesis, amine ligands bind at three spatially distinct binding regions, all of which are supposed to be located between the seven helix transmembrane domains of the amine GPCRs. Moreover, depending on ligand size and shape of the ligand, the binding is proposed to more or less overlap these three binding regions (Jacoby et al., 1999). Obviously, as the hypothesis is based on 3D modeling, it is restricted to the correctness of the docking procedures. The possibility

even exists that a ligand might show distinctive interactions at multiple binding pockets within a receptor.

By contrast, the present approach constitutes a purely operational approach for the spatial and chemical mapping of binding regions. The proteochemometrics approach thus eliminates the ambiguity and problems intrinsic to modeling and docking. However, the present modeling approach and the 3D modeling gain partly different types of information. The proteochemometrics modeling is thus dependent on the existence of variations in chemical properties of the receptors to reveal an interaction; e.g., the conserved aspartic residue of TM3, presumed to serve as a counter-ion to the protonated nitrogen of the amines, is conserved in all the GPCRs. This is probably the reason that the TM3 does not show any importance in the proteochemometrics modeling. On the other hand, the proteochemometrics modeling seems to be capable of revealing indirect effects arising from distant regions in a receptor; e.g., TM6, as well for some compounds TM2, show important effects from the present models, whereas 3D modeling have attributed these regions minor importance. Here it may be mentioned that one of our previous studies, where we applied proteochemometrics modeling onto chimeric and point-mutated α_1 -adrenoceptors, the TM2 region also showed up as an important specificity-determining region (Lapinsh et al., 2001).

Although most previous mutational studies concentrated on conserved receptor positions (not possible to analyze with the present model), a few of them assessed nonconserved amino acids. Results from such studies are consistent with the present analysis. Thus, a study by Hamaguchi et al. (1996) ascertains the importance of F86 in the final helical turn of TM2 of the human α_{1A} -adrenoceptor (A1AA_HUMAN). Mutation of this residue to methionine resulted in a large loss of affinity for the α_{1A} -adrenoceptor-selective dihydropyridine compounds studied, the effect of this single mutation being similar or larger than of other mutations the TM2 region, in complete accord with the major contribution of the position revealed by the analysis presented in Fig. 8. In another study, M292 at the end of TM6 in the human α_{1A} -adrenoceptor was mutated to leucine, which led to increased affinities for some phenylethylamines, whereas it decreased it for oxymetazoline (Hwa et al., 1996). In our model, this amino acid shows up as one of those that discriminates the ligands in their affinities for the α_{1A} -adrenoceptor (Fig. 8). Thus, the experimental data of Hwa et al. (1996) lend full support to the proteochemometrics modeling approach.

Combining our present approach with experimental studies may in the future bring mutational analysis to much higher level. By using proteochemometrics, it would technically be a simple matter to separately assess the importance of each of the physicochemical amino acid properties (i.e., the five z-scales) at each of the positions in the receptor sequences and accordingly select a proper amino acid at a relevant position for mutation to reveal the role of a specific physicochemical effect. This approach contrasts to most previous mutational studies where amino acids were mutated to alanine in a more or less nonsystematic fashion. Mutating to alanine may in many cases have led to drastic changes of several or all of the z-scale properties of the native amino acid, something that in turn could have caused a drastic effects on the protein function (e.g., causing both direct and indirect effects) leading to difficulties in interpretation of the data.

In conclusion, we have here been able to create valid pro-

teochemometrics models for analysis of wild-type amine GPCRs. The models give insight into the mode of interaction of synthetic amine ligands with these receptors. These models may find use not only in ligand design but also in prediction of properties of novel related receptors. The resolution of the present models is of course limited to the richness of the underlying data. However, because the models are soft and inclusive, one may just include data from chimeric and mutated amine receptors, as well as more measurements from additional ligands to improve the models. We predict that the future systematic use of the proteochemometrics approach will greatly increase its resolution (e.g., allowing detailed quantitative mapping of the chemical interactions of the GPCRs with their ligands).

References

- Bairoch A and Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* **28**:45–48.
- Bikker JA, Trumpp-Kallmeyer S, and Humblet C (1998) G-Protein coupled receptors: models, mutagenesis, and drug design. *J Med Chem*. **41**:2911–2927.
- Bourdon H, Trumpp-Kallmeyer S, Schreuder H, Hoffack J, Hibert M, and Wermuth CG (1997) Modelling of the binding site of the human m1 muscarinic receptor: experimental validation and refinement. *J Comput-Aided Mol Des* **11**:317–332.
- Eriksson L, Johansson E, Kettaneh-Wold N, and Wold S (1999) Introduction to multi- and megavariable data analysis using projection methods (PCA/PLS). Umetrics AB, Umeå, Sweden.
- Eriksson L, Johansson E, Lindgren F, and Wold S (2000) GIF-PLS: modeling of non-linearities and discontinuities in QSAR. *Quant Struct-Act Relat* **19**:1–11.
- Eriksson L, Johansson E, and Wold S (1997) Quantitative structure-activity relationship model validation, in *Quantitative Structure-Activity Relationships in Environmental Sciences - VII* (Chen F, Schuurmann G eds), p 381–397, SETAC Press, Pensacola, Florida.
- Fanelli F, Menziani MC, and De Benedetti PG (1995) Molecular dynamics simulations of m3-muscarinic receptor activation and QSAR analysis. *Bioorg Med Chem* **3**:1465–1477.
- Flower DR (1999) Modelling G-protein-coupled receptors for drug design. *Biochim Biophys Acta* **1422**:207–234.
- Hamaguchi N, True TA, Saussy DL Jr, and Jeffs PW (1996) Phenylalanine in the second membrane-spanning domain of alpha 1A-adrenergic receptor determines subtype selectivity of dihydropyridine antagonists. *Biochemistry* **35**:14312–14317.
- Horn F, Weare J, Beukers MW, Hörsch S, Bairoch A, Chen W, Edvardsen Ø, Campagne F, and Vriend G. (1998) GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Res* **26**:277–281.
- Hwa J, Graham RM, and Perez DM (1996) Chimeras of α_1 -adrenergic receptor subtypes identify critical residues that modulate active state isomerization. *J Biol Chem* **271**:7956–7964.
- Jacoby E, Fauchere JL, Raimbaud E, Ollivier S, Michel A, and Spedding M (1999) A three binding site hypothesis for the interaction of ligands with monoamine G protein-coupled receptors: implications for combinatorial ligand design. *Quant Struct-Act Relat* **18**:561–572.
- Lapins M, Prusis P, Gutcaits A, Lundstedt T, and Wikberg JES (2001) Development of proteochemometrics: a novel technology for the analysis of drug-receptor interactions. *Biochem Biophys Acta* **1525**:180–190.
- Lundstedt T, Seifert E, Abramo L, Thelin B, Nyström Å, Pettersen J, Bergman B (1998) Experimental design and optimization. *Chemom Intell Lab Syst* **42**:3–40.
- Millan MJ, Brocco M, Rivet JM, Audinot V, Newman-Tancredi A, Maioffs L, Queriaux S, Despaux N, Peglion JL, and Dekeyne A (2000a) S18327 (1-[2-[4-(6-fluoro-1,2-benzisoxazol-3-yl)piperid-1-yl]ethyl]3-phenyl imidazolin-2-one), a novel, potential antipsychotic displaying marked antagonist properties at α_1 - and α_2 -adrenergic receptors: II. Functional profile and a multiparametric comparison with haloperidol, clozapine, and 11 other antipsychotic agents. *J Pharmacol Exp Ther* **292**:54–66.
- Millan MJ, Gobert A, Newman-Tancredi A, Audinot V, Lejeune F, Rivet JM, Cussac D, Nicolas JP, Muller O, and Lavielle G. (1998) S 16924 ((R)-2-[1-[2-(2,3-dihydrobenzo[1,4]dioxin-5-yloxy)-ethyl]-pyrrolidin-3-yl]-1-(4-fluoro-phenyl)-ethanone), a novel potential antipsychotic with marked serotonin (5-HT)1A agonist properties: I. Receptorial and neurochemical profile in comparison with clozapine and haloperidol. *J Pharmacol Exp Ther* **286**:1341–1355.
- Millan MJ, Gobert A, Newman-Tancredi A, Lejeune F, Cussac D, Rivet JM, Audinot V, Dubuffet T, and Lavielle G. (2000b) S33084, a novel, potent, selective, and competitive antagonist at dopamine D3-receptors: I. Receptorial, electrophysiological and neurochemical profile compared with GR218,231 and L741,626. *J Pharmacol Exp Ther* **293**:1048–1062.
- Millan MJ, Newman-Tancredi A, Audinot V, Cussac D, Lejeune F, Nicolas JP, Cogé F, Galizzi JP, Boutin JA, Rivet JM, et al. (2000c) Agonist and antagonist actions of yohimbine compared with fluparoxan at 2-adrenergic receptors (AR)s, serotonin (5-HT)1A, 5-HT1B, 5-HT1D and dopamine D2 and D3 receptors. Significance for the modulation of frontocortical monoaminergic transmission and depressive states. *Synapse* **35**:79–95.
- Millan MJ, Schreiber R, Monneyron S, Denorme B, Melon C, Queriaux S, and Dekeyne A (1999) S-16924, a novel, potential antipsychotic with marked serotonin1A agonist properties. IV. A drug discrimination comparison with clozapine. *J Pharmacol Exp Ther* **289**:427–436.
- Pastor M, Cruciani G, McLay I, Pickett S, and Clementi S (2000) Grid-Independent Descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. *J Med Chem* **43**:323–3243.
- Prusis P, Muceniece R, Andersson P, Post C, Lundstedt T, and Wikberg JES (2001) PLS modeling of chimeric MS04/MSH-peptide and MC1/MC3-receptor interactions reveals a novel method for analysis of ligand receptor interactions. *Biochem Biophys Acta* **1544**:350–357.
- Sandberg M, Eriksson L, Jonsson J, Sjöström M, and Wold S (1998) New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *J Med Chem* **41**:2481–2491.
- Sansom MSP (1998) Models and simulations of ion channels and related membrane proteins. *Curr Opin Struct Biol* **8**:237–244.
- Schotte A, Janssen PF, Gommeren W, Luyten WH, Van Gompel P, Lesage AS, De Loore K, and Laysen JE. (1996) Risperidone compared with new and reference antipsychotic drugs: in vitro and in vivo receptor binding. *Psychopharmacology* **124**:57–73.
- Schwartz TW (1994) Locating ligand-binding sites in 7TM receptors by protein engineering. *Curr Opin Biotechnol* **5**:434–444.
- Wishart W, Bremner DH, and Sturrock KR (1999) Molecular modelling of 5-hydroxytryptamine receptors. *Recept Channels* **6**:317–335.
- Wold S (1995) PLS for linear modelling, in *Chemometric Methods in Molecular Design* (van de Waterbeem D ed) vol 2, pp. 195–218, VCH Verlagsgesellschaft, Weinheim.
- Wold S, Esbensen K, and Geladi P (1987) Principal component analysis. *Chemom Intell Lab Syst* **2**:37–52.

Address correspondence to: Dr. Jarl Wikberg, Pharmaceutical Pharmacology, Uppsala University, Box 591, BMC, SE-751 24 Uppsala, Sweden. E-mail: jarl.wikberg@farmbio.uu.se