0026-895X/10/7702-149–158$20.00
MOLECULAR PHARMACOLOGY
Vol. 77, No. 2
60103/3552764
Mol Pharmacol 77:149–158, 2010
Printed in U.S.A.

MOL PHARM
MOLECULAR PHARMACOLOGY

A Novel Multilevel Statistical Method for the Study of the Relationships between Multireceptorial Binding Affinity Profiles and In Vivo Endpoints[S]

Jana Selent, Anna Bauer-Mehren, Laura López, María Isabel Loza, Ferran Sanz, and Manuel Pastor

Research Unit on Biomedical Informatics (GRIB), IMIM-Hospital del Mar, Universitat Pompeu Fabra, Barcelona, Spain (J.S., A.B.-M., L.L., F.S., M.P.); and Departamento de Farmacología, Facultad de Farmacia, Universidad de Santiago de Compostela, Santiago de Compostela, Spain (M.-I.L.)

Received August 6, 2009; accepted November 9, 2009

ABSTRACT

The present work introduces a novel method for drug research based on the sequential building of linked multivariate statistical models, each one introducing a different level of drug description. The use of multivariate methods allows us to overcome the traditional one-target assumption and to link in vivo endpoints with drug binding profiles, involving multiple receptors. The method starts with a set of drugs, for which in vivo or clinical observations and binding affinities for potentially relevant receptors are known, and allows obtaining predictions of the in vivo endpoints highlighting the most influential receptors. Moreover, provided that the structure of the receptor binding sites is known (experimentally or by homology modeling), the proposed method also highlights receptor regions and ligand-receptor interactions that are more likely to be linked to the in vivo endpoints, which is information of high interest for the design of novel compounds. The method is illustrated by a practical application dealing with the study of the metabolic side effects of antipsychotic drugs. Herein, the method detects related receptors confirmed by experimental results. Moreover, the use of structural models of the receptor binding sites allows identifying regions and ligand-receptor interactions that are involved in the discrimination between antipsychotic drugs that show metabolic side effects and those that do not. The structural results suggest that the topology of a hydrophobic sandwich involving residues in transmembrane helices (TM) 3, 5, and 6, as well as the assembling of polar residues in TM5, are important discriminators between target/antitarget receptors. Ultimately, this will provide useful information for the design of safer compounds inducing fewer side effects.

The classic formulation in which a disease is associated to a single biological receptor is clearly an oversimplification, valid only for exceptional situations. Recent advances in genomics, proteomics, and systems biology depict a far more complex scenario, much less comfortable to work with, in which the biochemical mechanisms of diseases and therapies involve many different biological receptors linked in complex interaction pathways, which are usually poorly understood. The best possible therapies for many complex diseases will probably require a deep and detailed understanding of those relationships. However, because of the extreme complexity of the involved mechanisms, we will not achieve such detailed understanding in the immediate future. Hence, there is a need for new methodological approaches that are able to advance in this way. In this work, we propose a novel method for exploiting pharmacologic information obtained for sets of currently available drugs that are known to interact with multiple related receptors. This method is intended to improve our understanding of the biological mechanisms implicated in the pharmacological treatments but also to provide hints about how the chemical structure of the studied drugs could be improved. It is based on the sequential building of

This work was supported by the Spanish Ministry of Science and Innovation [Grant SAF2005-08025-C03], the Instituto de Salud Carlos III for projects RETICS HERACLES [Grant RD06/0009] and COMBIOMED [Grant RD07/0067], and the Departament d'Innovació, Universitat i Empresa de la Generalitat de Catalunya. The Research Unit on Biomedical Informatics (GRIB, InstitutMunicipal d'Investigació Mèdica de l'Hospital del Mar y la Universitat Pompeu Fabra) is the Biomedical Informatics Node of the Spanish Institute of Bioinformatics (INB).

J.S. and A.B.-M. have contributed equally to the studies presented in this manuscript and should be considered co-first authors.

Article, publication date, and citation information can be found at http://molpharm.aspetjournals.org.

doi:10.1124/mol.109.060103.

[S] The online version of this article (available at http://molpharm.aspetjournals.org) contains supplemental material.

ABBREVIATIONS: PCA, principal component analysis; PLS-R, partial least square regression; APD, antipsychotic drug; 5-HT, 5-hydroxytryptamine; GPCR, G-protein coupled receptor; LV, latent variable; WG, weight gain; TM, transmembrane helix.

multivariate statistical methods, such as principal component analysis (PCA) and partial least square regression (PLS-R) to describe the connection between in vivo observations, profiles of binding affinities for multiple receptors, and structural information. Here, we describe the method using as case study the investigation of common adverse effects of antipsychotic drugs (APD).

Schizophrenia is a good example of a complex disease of unknown pathogenesis. Recent research (González-Maeso et al., 2008; Altar et al., 2009) confirms the complexity of the mechanisms involved in the disease. Nevertheless, highly valuable information can be obtained by studying the currently available antipsychotic drugs. APDs have a proven clinical efficacy, confirmed by more than 50 years of clinical use. A growing body of evidence confirms that the therapeutic effects of APDs are complex and cannot be ascribed to a single receptor. Hence, the traditional concept of a single pharmacological receptor must be expanded to a whole set of biomolecules (receptorome) that are putatively involved in the pharmacological effect of the drugs (Roth et al., 2004). The optimum binding profile refers to the affinities of a drug to a set of receptors, which results in a highly efficient drug that lacks undesirable side effects. In the field of APDs, the idea was suggested by the early work of Meltzer (Roth et al., 1998), and the so called Meltzer Index (a defined D_2/5-HT_{2A} binding affinity ratio) can be regarded as a first rough attempt of characterizing the optimal multireceptorial binding profile of atypical APDs. However, apart from D_2 and 5-HT_{2A} receptors, there is no clear agreement about what other receptors are relevant for the pharmacological effects of APDs and much less about which is the optimum binding affinity profile.

The application of multivariate statistical methods in this particular field is not new. Silvestre and Prous (2005) published a study in which the application of statistical methods provided robust evidence about the association of M_3 receptor binding affinity and diabetes type 2 risk. Later, PCA was used on a matrix of binding affinity profiles for a mixed set of nine clinically used APDs and three new drug candidates to identify similarity patterns of the compounds (Lange et al., 2007). Our method presented here has the advantage of introducing for the first time a multilevel approach, able to interconnect different types of description of actors (ligands and receptors) that are involved in the pharmacological effects. The chosen example (APDs and metabolic side effects) is relatively well known, because we expect to compare our results with information already available in the literature. Nevertheless, some of the results that we obtain are original and suitable for being experimentally tested and validated.

Materials and Methods

Clinical/In Vivo Data Collection as Input for Statistical Model S1. We gathered information on adverse effects caused by APDs because the attempt to collect clinical data representing the efficacy of APDs was not fruitful for a reasonably large set of APDs. Metabolic adverse effects have received attention in different studies (Farwell et al., 2004; Silvestre and Prous, 2005; O'Neill, 2005; Reynolds et al., 2006) and they include weight gain, hyperglycemia, and induction of type 2 diabetes. A bibliographic search allowed us to find several indexes of these adverse side effects for the considered drugs, which are listed in Table 1.

Binding Affinity Data Collection as Input for Statistical Models S1 and S2. A preliminary list of receptors that are potentially important for the antipsychotic effects as well as for side effects was gathered from bibliographic sources (Silvestre and Prous, 2005). This set mainly comprises receptors belonging to the G-protein coupled receptor (GPCR) superfamily. To avoid missing values, receptors for which only few binding data were available were discarded from the initial set. Moreover, for the two drugs sulpiride and pimozide, we used the reported average muscarinic affinity of M_1, M_2, and M_3, in which the individual affinities were missing. The final set includes the following receptors: dopamine (D_2, D_3, and D_4), serotonin (5-HT_{1A}, 5-HT_{2A}, 5-HT_{2C}, 5-HT_6, and 5-HT_7), histamine (H_1), adrenergic (α_1 and α_2), and cholinergic muscarinic (M_1, M_2, and M_3). The 5-HT_3 channel was also considered. The series of compounds includes structurally diverse agents comprising representatives for typical and atypical APDs. In total, the set consists of 25 diverse chemotypical APDs (Table 1 and Scheme 1).

Input Data for Statistical Model S3. The molecular descriptors used in the statistical analysis require obtaining structural models of the ligand-receptor complexes. Hence, structural models for 14 of the receptors mentioned above (excluding 5-HT_3, which does not belong to the same structural family) were built using a standard homology modeling protocol based on the novel template of the β_2 adrenergic receptor, which was previously published by our group (Selent et al., 2008). A molecule of clozapine was inserted in the binding site of all the receptors, using the docking software GOLD (Verdonk et al., 2003). Clozapine was docked into the active site of each receptor by defining a 15-Å region centered on the Asp3.32, a residue conserved in all aminergic receptors and known to be crucial for ligand interaction (Selent et al., 2008). To simulate an induced fit mechanism, the best docking solution was subjected to an optimization protocol using the molecular modeling program MOE (Molecular Operating Environment; Chemical Computing Group, Montreal, QC, Canada). In a first step, a brief energy minimization of the complexes was carried out using the standard protocol of the LigX tool in MOE (all atom force field MMF94x) considering the receptor residues in an 8-Å radius around clozapine. The complex was further refined by means of a 200-ps molecular dynamics simulation (force field MMF94x, 300 K, time step 2 fs) and was subsequently energy minimized by applying gradient minimization until the root-mean-square gradient was lower than 0.001 kcal/mol Å.

For the selection of residues to be considered in the optimization, we started listing all the residues located within a radius of 8 Å from any atom of the ligand as measured in the structural model of the clozapine-D_2 receptor complex (obtained as described above), with the exception of the residues located at the extracellular loop 2. The 58 residues selected for the D_2 receptor were also selected for the rest of the receptors (Fig. 4a) based on the complete multiple sequence alignment of the 14 GPCRs, which was previously used for the generation of the homology models (Selent et al., 2008). For the statistical analysis (S3), the extracted sequences were described using five amino acid descriptors (Gottfries, 2006) representing charge, molecular weight, lipophilicity, as well as rigidity and flexibility of the amino acid side chains.

Statistical Modeling. The proposed method operates by building linked models using two multivariate statistical methods: PCA and PLS-R. PCA is a multivariate analysis tool for data supervision and dimensionality reduction. The method has been described elsewhere (Wold et al., 1987). In brief, it computes an approximated lower-dimension representation of the original data matrix X, in terms of the product of two matrices: the matrix of objects T (scores) and the matrix of variables P (loadings). In the matrix T, every object is represented by a small number of new variables (principal components), which are orthogonal linear combinations of the original variables, chosen to explain as best as possible the variance present in X. PLS-R (Wold et al., 2001) is a multivariate linear regression analysis technique particularly well suited for matrices in which the number of variables is larger than the number of objects. In PLS-R,

TABLE 1
Receptor binding affinity for 25 antipsychotics to 15 receptors

Compound[a]	5-HT$_{1A}$	5-HT$_{2A}$	5-HT$_{2C}$	5-HT$_3$	5-HT$_6$	5-HT$_7$	D$_2$	D$_3$	D$_4$	α_1	α_2	Target[b]	H$_1$	M$_1$	M$_2$	M$_3$	Antitarget[c]	dOR[d]	WG	Diabetes (3 Levels)	WG	Y_{S2}
																			kg			
1	5.60	5.51	5.00		6.00	5.52	7.20	7.70	5.62	5.59	6.21	*6,00*	4.20	3.55	3.35	5.00	*4.03*	1.52	−0.73	1	1	−1.89
2	5.00	5.32	5.00	4.60	5.30	5.30	7.61	7.45	5.66	4.43	4.32	*5,54*	4.14	4.37	4.37	4.37	*4.31*			1	1	−1.82
3	5.16	5.28	5.07	5.00	5.30	4.44	6.64	5.72	5.46	4.76	5.16	*5,30*	5.09	4.83	4.83	4.83	*4.90*	1.15	1.29	1	1	−1.49
4	5.65	6.90	5.68	5.00	5.90	6.24	6.94	6.46	6.39	6.74	6.82	*6,37*	6.24	4.82	5.62	4.82	*5.38*		0.13		1	−1.89
5	5.71	6.96	5.13	5.08	5.26	6.49	8.70	8.36	8.38	8.04	5.62	*6,87*	6.07	5.51	5.28	5.15	*5.50*	1.71	0.78	1	1	−1.37
6	5.60	7.39	5.30	5.43			9.07	8.05	7.86	7.26	5.57	*7,01*	6.01	5.12	5.74	5.15	*5.51*		0.4		1	−1.75
7	6.43	9.36	7.53	5.08	5.80	8.16	8.38	8.00	7.89	7.86	7.46	*7,69*	7.86	5.18	5.21	5.07	*5.83*	1.21	1.89	1	2	−0.71
8	8.47	7.86	7.10	6.20	6.17	8.00	8.71	8.35	6.41	7.52	7.15	*7,57*	7.56	5.17	5.45	5.33	*5.88*		0.71	1	1	−1.57
9	6.41	7.43	5.80	5.53	7.15	9.30	8.17	8.27	7.49	6.86	6.07	*7,30*	6.45	5.86	5.86	5.71	*5.97*	1.4	−3.11	2	1	−2.01
10	6.45	9.49	8.62	5.50	8.27	7.55	8.29	8.25	7.82	7.65	6.39	*7,88*	6.59	6.20	5.56	5.57	*5.98*	1.15	2.93	1	3	0.14
11	7.66	9.06	8.30	5.55	7.22	8.18	8.10	8.03	7.68	8.15	6.89	*7,93*	7.64	5.97	5.60	5.25	*6.12*		0.16	1	1	−1.77
12	5.83	7.76	6.01	5.37	7.42	8.10	9.05	8.85	7.05	8.02	6.31	*7,44*	7.68	5.49	5.66	5.84	*6.17*	1.7	0.52	1	2	−0.87
13	6.04	7.05	5.87	5.73	6.58	7.81	9.38	6.73	6.44	7.72	7.02	*7,06*	8.28	5.55	5.61	5.24	*6.17*	1.55	2.85	1	3	0.30
14	4.92	7.99	6.42	5.92	6.84	6.54	8.79	9.35	6.88	7.62	6.02	*7,14*	7.20	5.89	5.66	6.00	*6.19*	1.55	0.34	1	1	−1.57
15	5.70	8.25	6.88	5.90	7.77	7.64	9.01	8.67	6.71	8.00	6.42	*7,51*	8.10	5.77	5.54	5.73	*6.29*	1.65	3.97	3	3	1.88
16	5.23	8.14	6.91	5.87	6.83	5.17	8.65	8.74	8.27	8.08	5.77	*7,18*	7.72	6.11	5.85	6.04	*6.43*	1.58	3.24	2	3	1.03
17	6.51	6.66	5.77	5.67	6.19	6.51	6.31	6.38	5.70	7.68	5.88	*6,36*	7.89	6.86	6.20	5.40	*6.59*	0.98	2.55	1	2	−0.63
18	5.57	8.43	7.74	6.72	7.52	7.06	7.62	7.74	8.21	7.43	6.17	*7,35*	8.25	6.95	6.28	6.59	*7.02*	1.51	0.7	1	2	−0.91
19	5.75	8.43	7.81	6.11	7.71	7.45	8.14	8.47	7.49	8.84	6.55	*7,66*	8.10	7.21	6.56	7.22	*7.27*	3.25	2.8	3	3	2.28
20	6.50	8.58	8.54	6.37	8.93	8.23	8.13	7.95	7.66	8.19	6.89	*7,96*	8.71	7.74	6.85	7.14	*7.61*		2.32	2	3	1.28
21	6.30	8.20	6.80		6.42	7.14	7.72		7.87	8.70	5.80	*7,22*	8.74	7.80	7.17	7.05	*7.69*	2.91	1.68	2	2	0.62
22	6.69	7.72	7.28	5.89	7.24	7.00	7.96	8.00	7.80	8.06	6.42	*7,42*	7.77	8.19	7.58	7.61	*7.79*	3.43	3	3	3	2.42
23	6.86	9.52	8.35	6.40	8.52	8.25	8.25	8.72	9.19	9.00	6.73	*8,34*	9.05	7.58	7.10	7.66	*7.85*	3.31		2	3	1.79
24	6.59	8.22	7.69	7.09	8.00	7.11	6.83	6.47	7.48	8.01	7.15	*7,36*	8.70	8.22	7.08	7.59	*7.90*	4.21	4.23	3	3	3.12
25	5.76	8.35	7.98	7.15	8.06	6.68	7.69	7.46	7.49	6.94	6.55	*7,30*	8.85	8.15	7.32	7.36	*7.92*	3.8	3.83	3	3	2.82

[a] For the generic names of the compounds, see Scheme 1.
[b] Average binding affinity of GPCRs in the target list (5-HT$_{1A}$, 5-HT$_{2A}$, 5-HT$_{2C}$, 5-HT$_6$, 5-HT$_7$, D$_2$, D$_3$, D$_4$, α_1, and α_2).
[c] Average binding affinity of GPCRs in the antitarget list (H$_1$, M$_1$, M$_2$, and M$_3$).
[d] Mean logistic regression odds ratio for antipsychotic-induced new-onset type-2 diabetes versus no-treatment group.

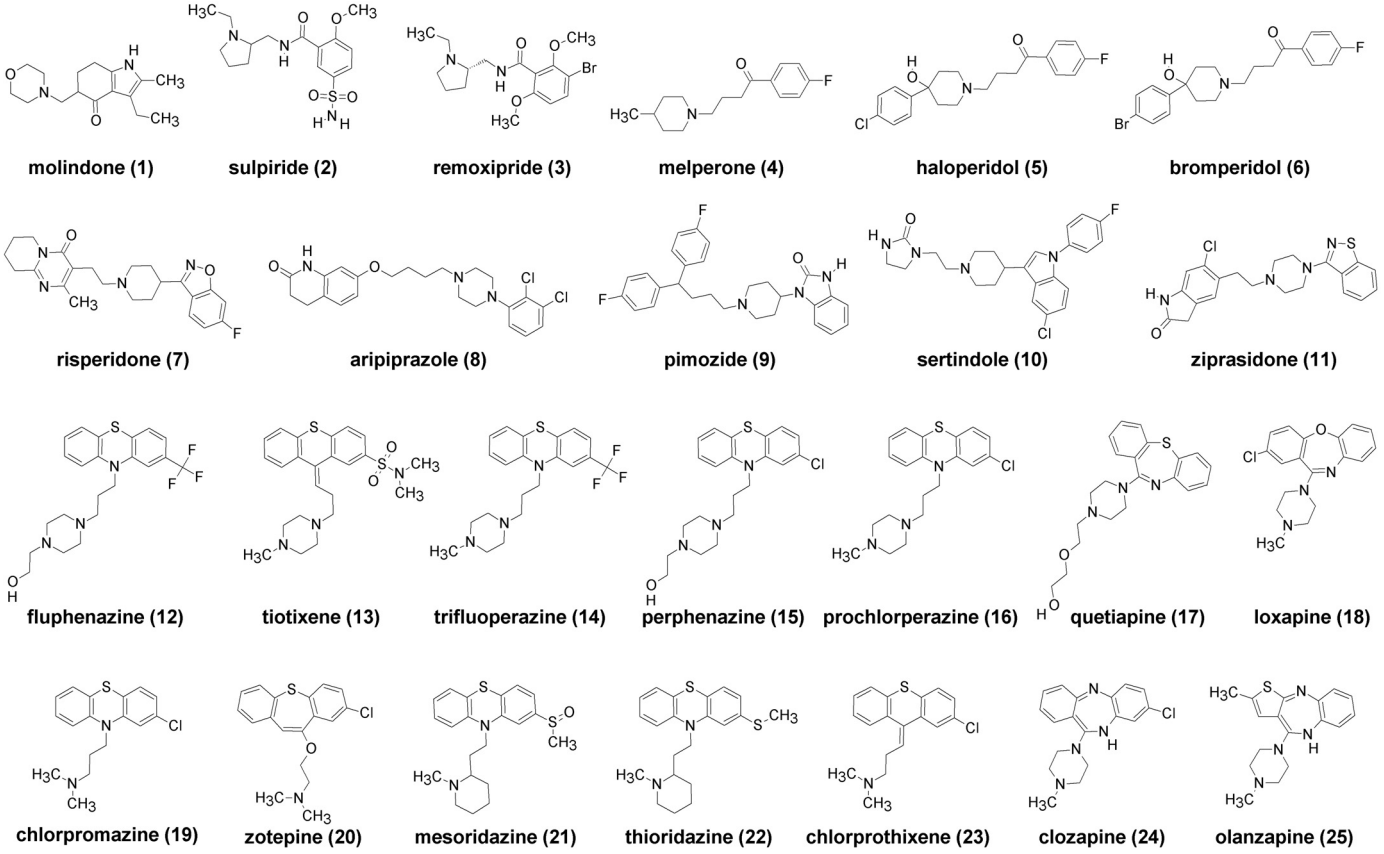
the original descriptor matrix is projected into a lower dimensionality space, which is defined by a set of a few latent variables (LV). Every LV is a linear combination of the original variables that describes the object differences and, at the same time, exhibits a high correlation with the outcome variable *Y*. PLS-R models are usually built extracting successive LVs, where each LV increases the total amount of explained *Y* variance. LVs are added until the predictive ability of the model reaches a maximum, which is monitored by successive internal cross-validation tests.

One of the advantages of the PLS-R models is that they can be analyzed to extract information about both the objects and the variables. The objects can be represented graphically in highly informative "scores plots" according to the LV values. In addition, the original variables can be represented according to their overall contribution to the model. The "coefficient plots" provide information about which *X* variables exhibit a relevant association with the outcome variable Y, with which sign, and whether this association is statistically significant (according to resampling techniques).

In this work, all PCA and PLS-R models were built using SIMPCA-P (Version 11.5) software (http://www.umetrics.com/default.asp/pagename/software_simcap/c/3) in Windows XP workstations. The values of the X variables were centered and scaled to unit variance unless it is explicitly stated otherwise in the text. Resampling methods, which are implemented in SIMCA-P, were used to test statistical significance, and the confidence intervals were computed at 95% confidence level. The original data matrices used for generating all models (S1, S2, and S3) are provided in the Supplemental Data.

Results

General Approach. The steps of our method are summarized in Fig. 1. The process starts by collecting a first data matrix (X_{S1}) containing clinically or in vivo relevant data (e.g., therapeutic efficacy, side effects, or animal experiments) for a reasonably large set of compounds. In cases in which no single variable can be used to describe the clinical/in vivo endpoint under study, PCA can be used to condense several variables providing incomplete or indirect information into a suitable single score. This condensing step (S1) is required only if the clinical/in vivo data comprise several correlated variables. In the next step (S2), the same set of drugs is characterized by gathering their binding affinities for a collection of putatively relevant receptors. The resulting matrix (X_{S2}) is analyzed by PLS-R analysis using as outcome variable Y_{S2}, which consists of the scores computed in S1. The coefficients that the PLS-R analysis assigns to the different receptors indicate whether the measured binding affinity correlates directly or indirectly, significantly or non-significantly, with the outcome variable. Here, statistically significant coefficients identify the receptors that are likely to play a relevant role in the clinical/in vivo endpoint. In a third step (S3), the structures of the ligand-perturbed receptors are incorporated as new level of description. In this model, the *X* matrix (X_{S3}) includes variables describing comparable structural features of the receptor binding sites, whereas the *Y* vector (Y_{S3}) corresponds to the S2 coefficients. The S3 model highlights the binding site regions and features that differentiate the related receptors. Because the receptors were previously associated to the clinical/in vivo effect (see models S1 and S2), the outcome of model S3 is susceptible to being exploited in the design of selective compounds with (or without) the clinical/in vivo effect or the score summarizing several of them.



Scheme 1. Chemical structures and generic names of the 25 antipsychotic drugs that were used in this study.

Statistical Model 1. S1 aims at obtaining a quantitative score summarizing several clinical/in vivo observations. In the considered example, we collected indicators for metabolic side effects of APDs. None of these indicators was suitable to be used as only descriptor of metabolic side effects for several reasons: 1) they were compiled from diverse sources and 2) some of them [diabetes and weight gain (WG)] were obtained from qualitative scales. Thus, a PCA (S1 in Fig. 1) was carried out, using as X matrix the four variables related to metabolic disturbances, which were centered and scaled to unit variance. Because of the intrinsic robustness to the presence of a small number of missing values of the nonlinear iterative partial least squares algorithm (Wold et al., 1987), the analysis allowed the incorporation of all the APDs listed in Table 1 without further treatment. The first principal component explains 74.2% of the X matrix variance and the scores vector (see Y_{S2} in Table 1) provides a convenient way to describe quantitatively the tendency of the diverse APDs to produce metabolic adverse effects.

Statistical Model 2. S2 allows us to describe the relationship between multiple binding affinities and metabolic side effects. We collected from bibliographic sources (see *Materials and Methods*) the binding affinities of a set of 25 clinically useful compounds for a selection of 15 receptors (plus the 5-HT_3 channel) that are assumed to be involved in the APD pharmacologic effects. The final binding affinity values (see Table 1), were centered (but not scaled) and used to build a PLS-R model (S2), in which the scores obtained in S1 were used as outcome variable Y_{S2} (Fig. 1). The quality of the obtained model is reasonable (2LV, $r^2 = 0.70$, $q^2_{LOO} = 0.60$). This is also shown in Fig. 2a, depicting the observed versus the predicted values.

Statistical Model 3. The sequence information of the ligand-receptor complexes was extracted as reported under *Materials and Methods* and was used as an additional level of description in S3. We built a new PLS-R model (S3) using the set of the most relevant residues in the ligand-perturbed binding site that were numerically described using a set of five descriptors as X matrix and the coefficient values of S2 as Y_{S3} matrix (Fig. 2b). Nonsignificant coefficients of S2 were replaced by zero. The obtained model has a very good quality (2LV, $r^2 = 0.99$, $q^2_{LOO} = 0.83$). As can be seen in Fig. 2c, the first LV clearly discriminates between the muscarinic receptors (on the right) and the rest of the receptors (on the left), with the exception of H_1, which is located between the clusters. This finding correlates well with the reported phylogenetic analysis (Surgand et al., 2006), which underlines the fact that the H_1 receptor contains structural features of both the target (left) and the antitarget (right) clusters. Moreover, the plot (Fig. 2c) reflects the evolutionary difference between the 5-HT_6 and the target cluster by the second LV.

Structural Models of the Clozapine-Receptor Complex. The novel high resolution X-ray structure of the β_2 adrenergic receptor (Protein Data Bank ID 2rh1) is an excellent template for modeling the APD-receptor complexes because 1) the binding pocket shows a high sequence homology with respect to the GPCRs of interest (up to 60%) and 2) the published structure was cocrystallized with the tricyclic ligand carazolol, which is structurally very similar to APDs, producing WG and metabolic side effects (e.g., clozapine). Therefore, it can be expected that such drugs will adopt a similar orientation within the same binding site (Selent et al., 2008). The obtained complex structures suggest that the bulky tricyclic clozapine-like compounds (compounds 12–25; Table 1 and Scheme 1) are accommodated in a binding pocket located between transmembrane helices (TM) 3, 5, and 7 for all 14 GPCRs, which is in agreement with the position of the tricyclic carazolol observed in the template. The polar and hydrophobic key interactions found for all 14 GPCR complexes are coincident with those described in our previous work (Selent et al., 2008) and are summarized in Fig. 3: 1) Polar interactions: salt bridge between the ligand protonated nitrogen and Asp3.32; ligand nitrogen NH with polar residues in TM5; ligand amid nitrogen with Ser/Cys3.36. 2) Hydrophobic interactions: hydrophobic sandwich between residues 3.33 on the one side and 5.47, 6.51, and 6.52 on the other; aromatic interaction (edge to face) of the ligand tricyclic system with Trp6.48.

To interpret the PLS-R model S3, we chose the structural models of clozapine-D_2 receptor (as target) and clozapine-M_1/H_1 receptors (as antitargets) as archetypical representa-

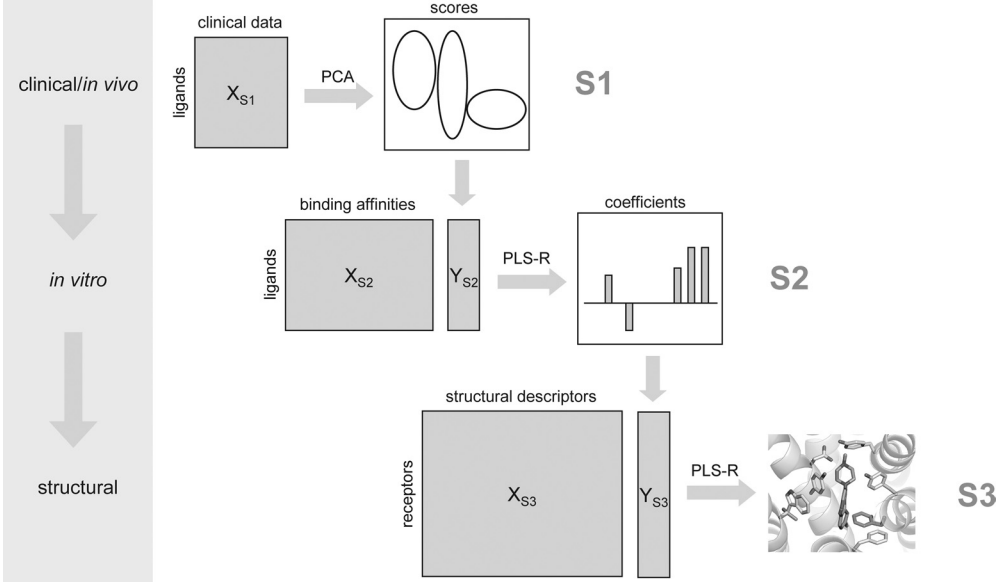


Fig. 1. The proposed method comprises three statistical models, S1, S2, and S3. In S1, the clinical/in vivo data are condensed into a quantitative score by applying PCA. This step is required only if the clinical/in vivo data comprise several correlated variables. PLS-R is used in model S2 to detect receptors being correlated with the clinical/in vivo data. A third level of description is introduced in model S3. PLS-R is applied to detect structural features of the ligand-receptor complexes that are associated to the clinical/in vivo effects through the linkage of models S1, S2, and S3.

tives of the 14 generated complexes. These complexes are used to illustrate the principal features of ligand interactions in target and antitarget receptors. The PLS-R coefficient plot of S3, shown in Fig. 2d, describes the positions and physicochemical properties of the ligand-receptor complexes that have a higher ability to discriminate between targets and antitargets or, in other words, that are more characteristic of each receptor type. In our analysis, the physicochemical properties of ligand-receptor interactions within a 8 Å radius of clozapine were used (Fig. 4a). Thus, the coefficient plot describes which of these residues contributes either directly or indirectly to the ligand binding. Herein, according to the aforementioned ligand-receptor models, positions 3.33, 5.43, 6.51 correspond to direct contacts with the ligand, whereas 3.35, 5.41, 5.51 correspond to indirect effects in ligand binding (Fig. 2d and 4b).

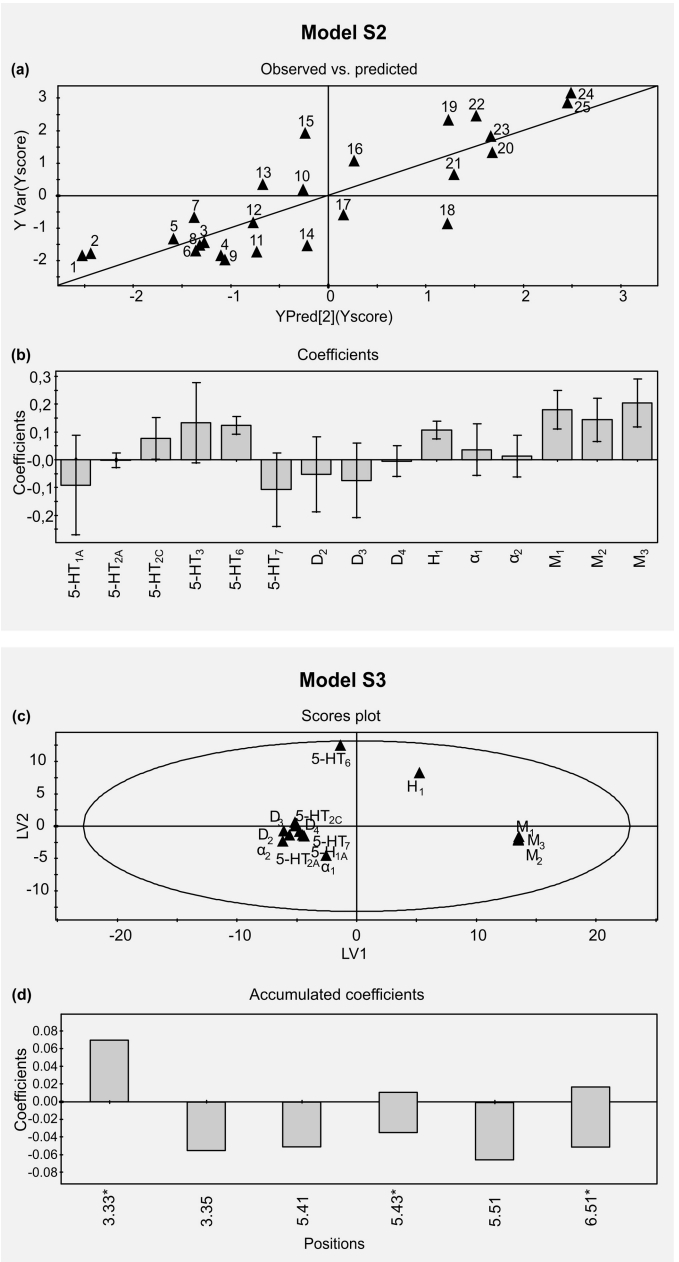


Fig. 2. a, observed versus predicted values depicting the good quality of model S2. b, coefficient plot of model S2: receptors that exhibit a statistically relevant contribution to the metabolic side effect stick out (M_1, M_2, M_3, H_1, 5-HT_6, and 5-HT_{2C}). c, PLS scores plot of model S3: LV1 discriminates the muscarinic receptors (on the right) from the rest of the receptors (on the left) except for H_1, which accommodates an intermediate position between both clusters. The 5-HT_6 receptor clearly separates from the main clusters by LV2. d, coefficient plot of model S3: the five coefficient values representing the five amino acid descriptors were accumulated for each position maintaining their algebraic sign. Only the most important residues are depicted; they can correspond to either indirect or direct ligand interactions (the latter are highlighted with an asterisk).

Coefficients that stand out with high values in the coefficient plot (Fig. 2d) correspond to hydrophobic residues at positions 3.33 and 6.51. These positions are of particular importance because they are individually conserved in target receptors as Val3.33/Phe6.51 and in antitarget receptors as Tyr3.33/Tyr6.51 (Fig. 4a). Based on our structural complexes (Fig. 5, a and b), the hydrophobic position 3.33 is part of a hydrophobic sandwich (3.33, 4.57, 5.47, 6.51, and 6.52). In the antitarget receptors, position 3.33 is conserved as Tyr3.33, allowing strong aromatic-aromatic interaction (edge to face) with the aromatic tricyclic pharmacophore of clozapine-like drugs. A result of this interaction is that the ligand is pushed tightly against interacting residues in TM5 and TM6 (5.47, 6.48, 6.51, and 6.52). The position 3.33 is stabilized through interaction with hydrophobic residues in position 4.57 (Trp/Val). In the target receptors, position 3.33 contains a valine, and the hydrophobic stabilization through position 4.57 (mainly polar Ser/Thr) is absent. Moreover, the modeled clozapine complexes show that Tyr/Phe6.51 flanks the binding pocket close to the piperazine ring of the ligand (edge to face) (Fig. 5, a and b), thus affecting the interaction patterns of target and antitarget receptors. In the antitargets, the clozapine complex is reinforced by a polar interaction with the phenolic group present in Tyr6.51, allowing more ligand contacts. Furthermore, the S3 outcome reveals position 5.43 (Fig. 2d). According to our models, this position forms part of the hydrophilic region in TM5, which interacts with the polar distal border of the tricyclic system of clozapine-like ligands (Fig. 5, c and d). An interesting fact is that target receptors have mainly polar residues in positions Cys/Ser/Thr5.43 and Ser5.42, whereas antitarget receptors are characterized by a nonpolar Ala5.43 with a polar Thr5.42 in the vicinity (Fig. 4a). The visual inspection of our structural models did not reveal the reasons of the high importance attributed by S3 to position 5.43. The use of static receptor structures prob-

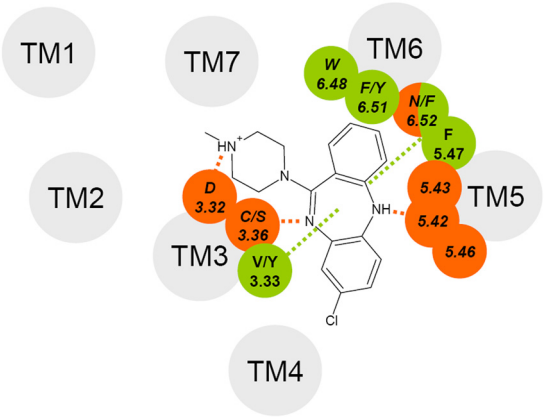


Fig. 3. Key interactions for clozapine-like ligand-receptor complexes: polar (orange) and hydrophobic (green).

ably did not allow representation of the fine details of the ligand-receptor interaction at this region, as well as any TM5 kink, rotation, or bending that could lead to a closer ligand interaction.

Apart from directly interacting receptor positions, the S3 model also reveals as influential positions 3.35, 5.41, and 5.51, which are not in direct contact with the ligand. Residue 3.35 is located between TM2 and TM3, oriented toward the exterior of the receptor in a position that does not allow ligand contacts. Similar circumstances are found for position 5.41 that accommodates a space between TM4 and TM5 and position 5.51 that occupies a region between TM5 and TM6 (Fig. 4b).

Connection between Chemotypes and Binding Profile. The binding data of the studied set of drug candidates reveals that some chemotypes are related to a certain binding profile regarding target and antitarget receptors (Table 1 and Scheme 1). Herein, compounds with a monoaromatic ring connected by an aliphatic linker to the protonated nitrogen (compounds 1–6), possess low average binding affinity for the antitarget receptors. Risperidone-like molecules (compounds 7 and 9–11), in which a biaromatic moiety is connected to a monoaromatic moiety by a flexible linker that contains the basic nitrogen, show enhanced binding affinity to the antitarget receptors. Moreover, clozapine-like molecules with a triaromatic scaffold, which is linked to a chain/ring structure containing the basic nitrogen (compounds 12–25), are characterized by enhanced antitarget binding affinities.

The structural models show that triaromatic cycles are tightly fixed in a sandwich formed by residues in positions 3.33, 5.47, and 6.52 (Fig. 5e). The binding site is delimited on the top by Tyr/Phe6.51 and on the bottom by Val/Ile3.40, and the triaromatic rings can fill the pocket completely, making contacts to top and bottom residues at the same time. The reduction of the ring size to a mono- or biaromatic moiety results in an overall decrease of target and antitarget binding affinities (Table 1, Scheme 1) as a result of the following effects (Fig. 5, e and d): 1) loss of the contact with the bottom residue Val/Ile3.40 and 2) loss of the hydrophobic interaction with Phe5.47. The described correlation between binding affinity and the extension of the aromatic ring system is more pronounced in antitarget receptors. It would seem that the antitarget Tyr3.33 residue establishes stronger hydrophobic interaction with the ligand aromatic moieties than the Val3.33 present in target receptors.

Discussion

The proposed method aims to exploit data describing, at diverse levels, the drug-receptor interactions for a set of drugs. These levels comprise 1) the clinical/in vivo data of the

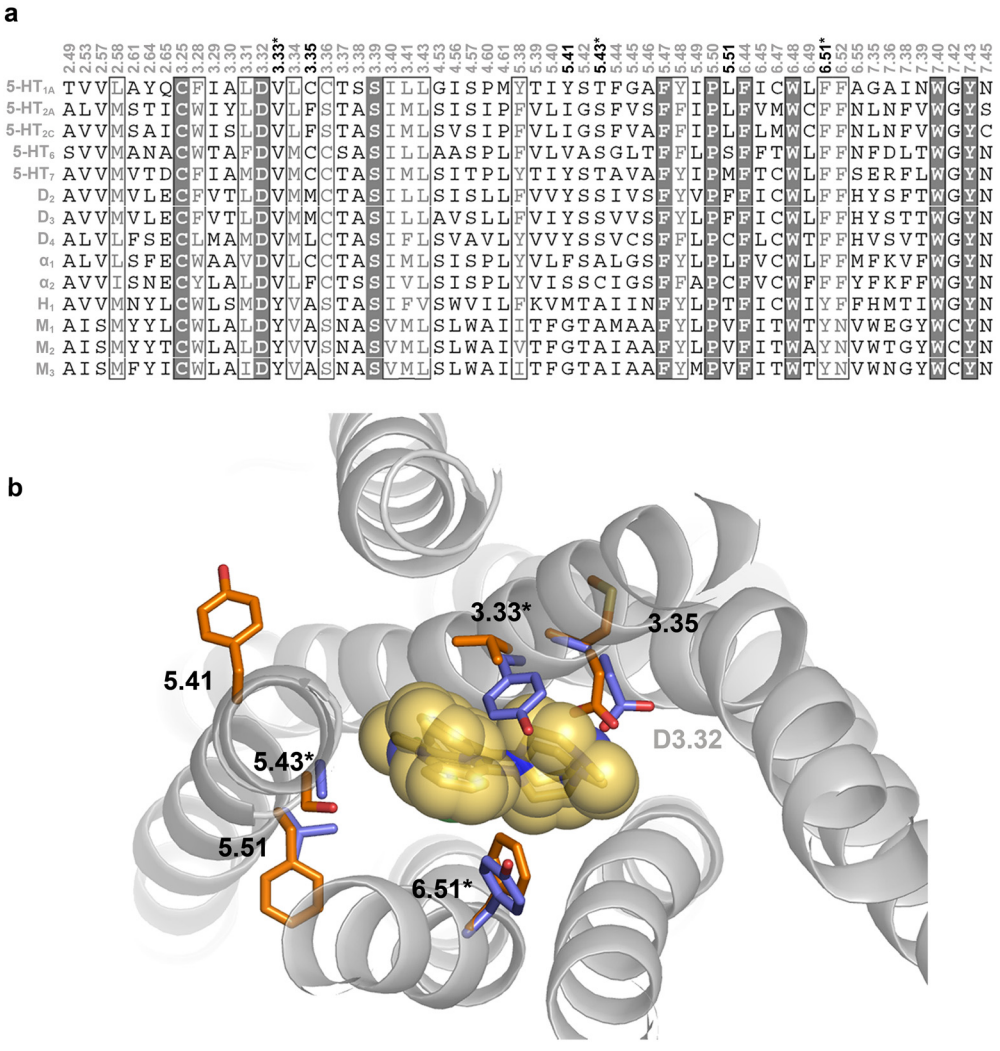


Fig. 4. a, multiple sequence alignment of residues placed at a distance shorter than 8 Å from clozapine in clozapine-target and clozapine-antitarget complexes. Direct (asterisk highlighted) and indirect interacting residues that emerge in the statistical model S3 (Fig. 2d) are depicted in bold. b, superimposition of the structural models for the complexes of clozapine with the D_2 target (orange) and the M_1 antitarget (blue) with direct (asterisk highlighted) and indirect residue contributions to ligand binding and/or receptor function.

drugs, 2) the in vitro binding affinities of the drugs to receptors and 3) the molecular structures of the ligand-receptor complexes. Three different models (S1–S3) are built to detect the associations between the different levels (Fig. 1). In the considered example, the model S1 describes quantitatively the tendency of APDs to produce metabolic adverse effects. The coefficients of model S2 (Fig. 2b) suggest that only a few receptors exhibit a statistically relevant contribution to the metabolic side effect: the muscarinic series (M_1, M_2 and M_3), the H_1, the 5-HT_{2C}, and 5-HT_6. Therefore, these receptors can be provisionally included into a shortlist of "antitargets" for the APD metabolic side effects. Published experimental results provide support for this finding. WG effects of antipsychotics have been ascribed primarily to interactions with H_1 (Kim et al., 2007), 5-HT_{2C} (Reynolds et al., 2006), and muscarinic receptors, in particular the M_3 receptor (O'Neill, 2005). Nevertheless, there is also evidence about the contribution of 5-HT_{2C} to the beneficial effects of atypical APDs, prompting us to move it to the target list (Lieberman et al., 1998; Millan et al., 1998; Reynolds et al., 2005). On the other hand, 5-HT_6 was excluded from the short list after further analysis because a comprehensive literature search revealed no direct evidence for APD-associated metabolic side effects. In fact, it has been suggested that blockade of the 5-HT_6 receptor leads to a cognitive enhancement in schizophrenia (Liu and Robichaud, 2009). In the interpretation of S2, it must be kept in mind that statistical association is not equivalent to causation. The outcome of S2 can be seen as a score that can be exploited in qualitative terms as a way to identify the set of targets more likely to be associated with the studied in vivo endpoints. This information can be exploited directly, applying traditional methods (e.g., quantitative structure-activity relationship for one or multiple receptors, as in Wang et al. (2008) aiming to obtain more selective compounds for the right target combination. Alternatively, the S2 scores can be used quantitatively, as input in a new model (S3) that allows us to highlight which residues and which kinds of interactions are characteristic of either type of receptors and therefore useful for the design of selective ligands. For simplicity, the structures of D_2 and M_1/H_1 have been used as representative of the target and antitarget receptors classes, respectively.

The receptors studied here exhibit a large degree of homology, and the binding affinities for certain compounds are often highly correlated because of the structural similarity of their binding sites. This means that receptors can be included in the aforementioned antitarget shortlist for two reasons: 1) because they are really involved in the metabolic side effect or 2) because they interact with the studied APD set similar to another (unknown) receptor that, in turn, is truly involved in the metabolic side effect. The second reason may explain the outcome regarding the 5-HT_6 receptor in

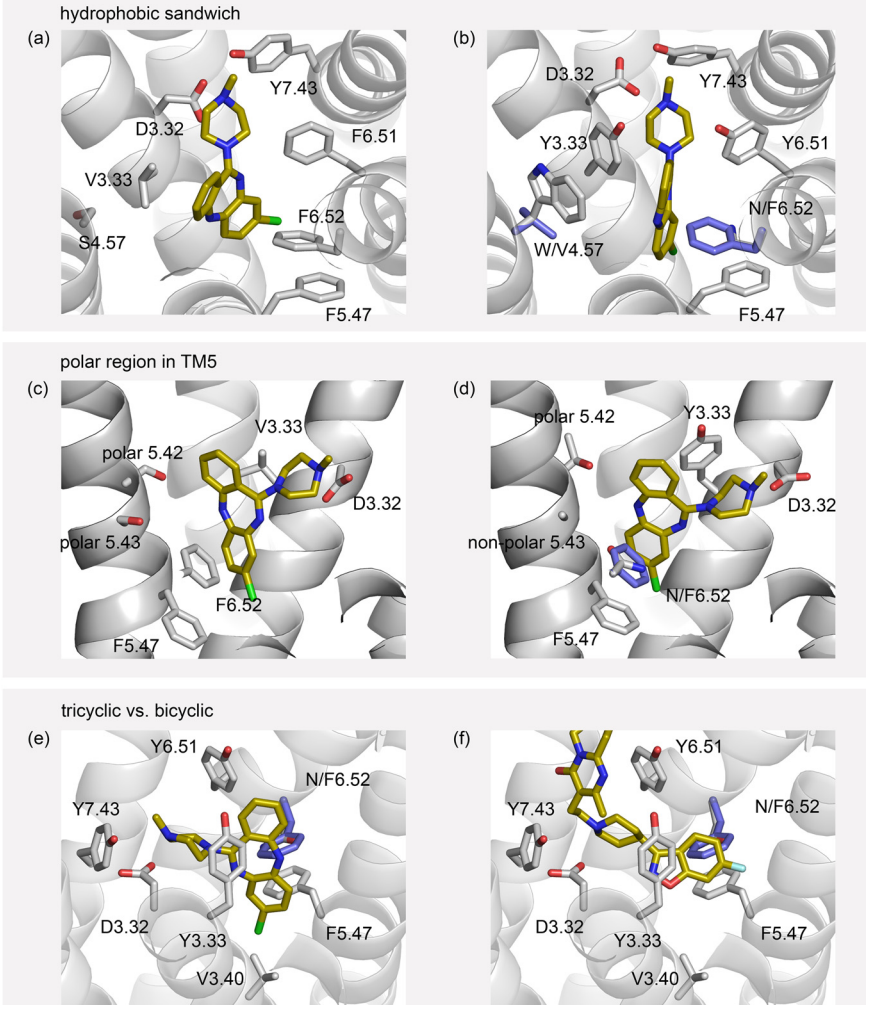


Fig. 5. Hydrophobic ligand-receptor interactions with residues 3.33 and 6.51 as hydrophobic hotspots in discrimination of target (a) versus antitarget receptors (b). Polar ligand-receptor interactions with residues 5.42 and 5.43 as polar hotspots in discrimination of target (c) versus antitarget receptors (d). Binding mode of triaromatic (e) and biaromatic (f) compounds in antitarget receptors (M_1 and H_1).

model S2. This confusion is an inherent limitation of statistical methods but can be solved by incorporating additional experimental data.

Receptors showing similar binding spectrum to the set of APDs appeared clustered in model S2 and are also evolutionary closely related (Surgand et al., 2006). This underlines the idea that the phylogenetic distance of receptors comes along with a similar binding spectrum of different drug chemotypes.

The results of S3 highlight residue positions contributing directly or indirectly to ligand binding (Fig. 4b). We want to stress that our structural models were used only to rationalize the statistical outcome of the model and to validate the results obtained by contrasting the residues and interactions highlighted by the model with those obtained by other authors using experimental or structure-based methods. The proposed method cannot explain the details of the ligand-receptor interaction; its usefulness is more related to its ability to point out structural regions potentially involved in relevant biological outcomes. The highest coefficients of model S3 (Fig. 2d) correspond to the directly interacting residues in positions 3.33 and 6.51 that are differentially conserved in the hydrophobic sandwich of the target receptors (Val3.33/Phe6.51) and the antitarget receptors (Tyr3.33/Tyr6.51) (Fig. 4a). The importance of position 3.33 in antagonist binding is supported by mutagenesis experiments (Lu and Hulme, 1999). The effect of the Y3.33V substitution on binding affinity can roughly be seen by comparing the average target (Val3.33) with the antitarget binding affinity of the H_1 receptor (Tyr3.33), which is, among the antitarget receptors, structurally the closest related to the target receptors (also reflected by the scores plot in Fig. 2c). Herein, the binding affinity of clozapine-like ligands (compounds 12–25) is lower for target receptors than for the antitarget H_1 receptor (target versus H_1; Table 1). Apparently, the Val3.33 hydrophobic-hydrophobic interaction in target receptors does not offset the stronger Tyr3.33 aromatic-aromatic interaction found in antitarget receptors (Fig. 5, a and b). This observation is in agreement with experimental data showing that the aromatic-aromatic interaction found for Tyr3.33 is not purely hydrophobic but also contains an electrostatic component (Waters, 2002). Regarding position 6.51, both target and antitarget receptors interact with the ligand through the aromatic ring of residue Tyr/Phe6.51. However, the phenolic group of Tyr6.51 in the antitarget receptors (Phe6.51 in target) might stabilize the position of the aromatic ring by polar interaction with neighboring polar residues and improve the strength of the ligand-receptor complex. With respect to the polar interactions, position 5.43 emerges in the coefficient plot of the S3 model (Fig. 2d), indicating its importance in the discrimination between targets and antitargets. The polar residues in TM5 have been repeatedly described as significant for ligand binding. For example, 5.42, 5.43, and 5.46 were proposed to play a major role in binding selectivity toward the aromatic moieties of endogenous ligands (Kobilka, 2004). Regarding WG and metabolic side effects, the model recognized that antitarget receptors are characterized by a nonpolar Ala5.43 and a polar Thr5.42 in vicinity, whereas target receptors possess two polar residues in both positions with few exceptions (Cys/Ser/Thr5.43 and Ser5.42) (Fig. 4a).

Apart from these direct ligand interactions, the S3 coefficient plot (Fig. 2d) also highlights a few residues that are faced toward the exterior of the receptor: 3.35, 5.41, 5.51 (Fig. 4b). Even if our structural models suggest that these positions are not in direct contact with clozapine-like ligands, it is still possible that they affect binding affinity and/or receptor function. It is remarkable that experimental studies proved that mutation of position 3.35 can lead to a constitutively active α_{1b}-adrenergic receptor, possibly because of enhanced translational movement of TM3 with significant alteration of ligand affinity (Porter et al., 1996; Hwa et al., 1997). Positions 4.51 and 5.51 are clearly exposed toward the lipid bilayer (Fig. 4b), and mutagenesis experiments were not able to detect a significant effect on ligand affinity. However, it is noteworthy that both residues are located in the putative dimerization interface of GPCRs (Kota et al., 2006). In fact, Filizola and Weinstein (2005) demonstrate that mutation of 4.51 significantly affects formation of opsin dimers (Kota et al., 2006). A growing body of evidence supports the capital importance of receptor dimerization in the function of membrane receptors (Gurevich and Gurevich, 2008). All in all, there is experimental evidence that the nondirectly interacting residue positions, highlighted in the PLS-R coefficient plot, are involved in receptor activation (3.35) and in receptor dimerization (4.51, 5.51). A detailed analysis of their potential role is outside the scope of this work, but it should be stressed that the described method would allow detection of both direct and indirect ligand-receptor interactions.

Considering the large importance of residues with aromatic properties (3.33 and 6.51) in S3, it is not surprising that in the series of considered drugs (Table 1 and Scheme 1), the average antitarget binding affinity correlates to the number of fused aromatic cycles in the drug structures. Our structural models are in agreement with this observation because a triaromatic moiety fits better into the binding pocket of the antitarget receptor than biaromatic (Fig. 5e) or monoaromatic moieties. This finding explains why tricyclic clozapine-like drugs possess high binding affinity to antitarget receptors, which are most likely responsible for the metabolic side effects. In summary, our structural findings suggest that tricyclic scaffolds such as the ones found in clozapine-like ligands might be responsible for high risk of WG and metabolic related side effects. Moreover, the obtained data indicate that 1) the architecture of the hydrophobic sandwich is of relevance; in particular, positions 3.33 and 6.51 can be considered hydrophobic hotspots for discriminating target and antitarget and 2) positions 5.42, 5.43 are promising polar hotspots in the binding pocket for improving drug selectivity between targets and antitargets. All these findings are susceptible to being exploited in terms of design of compounds with physicochemical properties complementing the aforementioned distinctive features between targets and antitargets (selective hotspots).

In this work, we used APDs and metabolic side effect as case study to illustrate the usefulness of an innovative statistical strategy for jointly analyzing different levels of information. We want to emphasize that our approach can be applied to any field in which the therapeutic or adverse effects of drugs are based on multireceptorial interactions and for which different levels of data are available. In particular, the clinical endpoint must contain sufficient variability within the series of studied drugs; i.e., the series cannot consist only of inactive or active compounds.

In conclusion, we presented a method integrating different levels of description of a set of compounds (in vivo, in vitro, molecular descriptors) that overcomes the simplistic approach (one target assumption) of traditional drug design methods. Instead of producing models that behave like "black boxes," the method proposed here can be inspected in different ways. It allows evaluating the relevance of diverse receptors or their structural features with respect to a biological property of interest. This information can ultimately be used for the design of new compounds that possess properties related to efficacy but lack features related to undesired side effects.

References

Altar CA, Vawter MP, and Ginsberg SD (2009) Target identification for CNS diseases by transcriptional profiling. *Neuropsychopharmacology* **34:**18–54.

Farwell WR, Stump TE, Wang J, Tafesse E, L'Italien G, and Tierney WM (2004) Weight gain and new onset diabetes associated with olanzapine and risperidone. *J Gen Intern Med* **19:**1200–1205.

Filizola M and Weinstein H (2005) The study of G-protein coupled receptor oligomerization with computational modeling and bioinformatics. *FEBS J* **272:**2926–2938.

González-Maeso J, Ang RL, Yuen T, Chan P, Weisstaub NV, López-Giménez JF, Zhou M, Okawa Y, Callado LF, Milligan G, et al. (2008) Identification of a serotonin/glutamate receptor complex implicated in psychosis. *Nature* **452:**93–97.

Gottfries J (2006) The drug designer's guide to selectivity. *Chemometrics Intellig Lab Syst* **83:**148–156.

Gurevich VV and Gurevich EV (2008) How and why do GPCRs dimerize? *Trends Pharmacol Sci* **29:**234–240.

Hwa J, Gaivin R, Porter JE, and Perez DM (1997) Synergism of constitutive activity in alpha 1-adrenergic receptor activation. *Biochemistry* **36:**633–639.

Kim SF, Huang AS, Snowman AM, Teuscher C, and Snyder SH (2007) From the Cover: Antipsychotic drug-induced weight gain mediated by histamine H1 receptor-linked activation of hypothalamic AMP-kinase. *Proc Natl Acad Sci U S A* **104:**3456–3459.

Kobilka B (2004) Agonist binding: a multistep process. *Mol Pharmacol* **65:**1060–1062.

Kota P, Reeves PJ, Rajbhandary UL, and Khorana HG (2006) Opsin is present as dimers in COS1 cells: identification of amino acids at the dimeric interface. *Proc Natl Acad Sci U S A* **103:**3054–3059.

Lange JH, Reinders JH, Tolboom JT, Glennon JC, Coolen HK, and Kruse CG (2007) Principal component analysis differentiates the receptor binding profiles of three antipsychotic drug candidates from current antipsychotic drugs. *J Med Chem* **50:**5103–5108.

Lieberman JA, Mailman RB, Duncan G, Sikich L, Chakos M, Nichols DE, and Kraus JE (1998) Serotonergic basis of antipsychotic drug effects in schizophrenia. *Biol Psychiatry* **44:**1099–1117.

Liu KG and Robichaud AJ (2009) 5-HT_6 antagonists as potential treatment for cognitive dysfunction. *Drug Dev Res* **70:**145–168.

Lu ZL and Hulme EC (1999) The functional topography of transmembrane domain 3 of the M1 muscarinic acetylcholine receptor, revealed by scanning mutagenesis. *J Biol Chem* **274:**7309–7315.

Millan MJ, Dekeyne A, and Gobert A (1998) Serotonin (5-HT)2C receptors tonically inhibit dopamine (DA) and noradrenaline (NA), but not 5-HT, release in the frontal cortex in vivo. *Neuropharmacology* **37:**953–955.

O'Neill MF (2005) Unusual suspect for antipsychotic-induced diabetes. *Drug Discov Today* **10:**1338–1338.

Porter JE, Hwa J, and Perez DM (1996) Activation of the alpha1b-adrenergic receptor is initiated by disruption of an interhelical salt bridge constraint. *J Biol Chem* **271:**28318–28323.

Reynolds GP, Hill MJ, and Kirk SL (2006) The 5-HT2C receptor and antipsychoticinduced weight gain–mechanisms and genetics. *J Psychopharmacol* **20:**15–18.

Reynolds GP, Templeman LA, and Zhang ZJ (2005) The role of 5-HT2C receptor polymorphisms in the pharmacogenetics of antipsychotic drug treatment. *Prog Neuropsychopharmacol Biol Psychiatry* **29:**1021–1028.

Roth BL, Meltzer HY, and Khan N (1998) Binding of typical and atypical antipsychotic drugs to multiple neurotransmitter receptors. *Adv Pharmacol* **42:**482–485.

Roth BL, Sheffler DJ, and Kroeze WK (2004) Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nat Rev Drug Discov* **3:**353–359.

Selent J, López L, Sanz F, and Pastor M (2008) Multi-receptor binding profile of clozapine and olanzapine: a structural study based on the new beta2 adrenergic receptor template. *ChemMedChem* **3:**1194–1198.

Silvestre JS and Prous J (2005) Research on adverse drug events. I. Muscarinic M3 receptor binding affinity could predict the risk of antipsychotics to induce type 2 diabetes. *Methods Find Exp Clin Pharmacol* **27:**289–304.

Surgand JS, Rodrigo J, Kellenberger E, and Rognan D (2006) A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors. *Proteins* **62:**509–538.

Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, and Taylor RD (2003) Improved protein-ligand docking using GOLD. *Proteins* **52:**609–623.

Wang XS, Tang H, Golbraikh A, and Tropsha A (2008) Combinatorial QSAR modeling of specificity and subtype selectivity of ligands binding to serotonin receptors 5HT1E and 5HT1F. *J Chem Inf Model* **48:**997–1013.

Waters ML (2002) Aromatic interactions in model systems. *Curr Opin Chem Biol* **6:**736–741.

Wold S, Esbensen K, and Geladi P (1987) Principal component analysis. *Chemometrics Intell Lab Syst* **2:**37–52.

Wold S, Sjöström M, and Eriksson L (2001) PLS-regression: a basic tool of chemometrics. *Chemometrics Intell Lab Syst* **58:**109–130.

Address correspondence to: Manuel Pastor, Research Unit on Biomedical Informatics (GRIB), IMIM-Hospital del Mar, Universitat Pompeu Fabra, Dr. Aiguader 88, E-08003 Barcelona, Spain. E-mail: manuel.pastor@upf.edu