Induced Fit—The Key for Understanding LSD Activity? A 4D-QSAR Study on the 5-HT_{2A} Receptor System

Daniel Streich, Margareta Neuburger-Zehnder and Angelo Vedani*

Institute of Inorganic Chemistry, University of Basel, Spitalstrasse 51, 4052 Basel, Switzerland, Biographics Laboratory 3R, Missionsstrasse 60, 4055 Basel, Switzerland

Abstract

Using a 4D-QSAR approach (software *Quasar*) allowing for multiple-conformation, orientation and protonation-state ligand representation as well as for the simulation of induced-fit phenomena, we have validated a family of receptor surrogates for the 5-HT_{2A} receptor system. The evolution was based on a population of 200 receptor models and simulated during 6,000 cross-over steps, corresponding to 30 generations. It yielded a cross-validated r^2 of 0.951 for the 23 ligands of the training set and a predictive r^2 of 0.859 for the 7 ligands of the test set. In this simulation, all ligand molecules were represented by four different conformers, obtained from a Monte-Carlo search in implicit aqueous solution. A series of six *scramble tests* (with an average predictive r^2 of -1.05) indicate a high sensitivity of the surrogate family towards the biological data.

The quantitative analysis of the contribution of the individual functional groups to the free energy of ligand binding, $\Delta G°$, reveals that the key factors for strong binding—and hence activity—are the ligand desolvation energy and the costs associated with induced fit, the adaptation of the receptor-binding site to the ligand topology. While the ammonium functionality is essential for recognition, its contribution to $\Delta G°$ is not favorable due to a high desolvation energy; important groups are rather methoxy and halide substituents. For most ligand molecules, the evolution does not select the lowest-energy conformer, contrary to previous assumptions in a corresponding 3D-QSAR study.

1 Introduction

In a recent article, we have described generation and validation of a surrogate for the 5-HT_{2A} receptor based on a 3-D QSAR approach [1]. This receptor is known to act as the biological target for a series of hallucinogenic substances, including substituted phenylalkylamines, tryptamines and LSD [2]. A prerequisite for a hallucinogenic effect is an agonistic binding mode to the high-affinity state of the receptor. In addition, a protonated N atom (the aminoethylside-chain of the phenyl-alkylamines and tryptamines or ring D in LSD; cf. Figure 1) is also mandatory, where by a primary amine yields maximal activity. Hydroxylation and methylation of position 2 of the phenylalkylamines or position 5 of the tryptamines combined with hydrophobic substituents (methyl, halides) in position 4 of the phenylalkylamines and position 7 of the tryptamines leads to a significantly increased activity towards the 5-HT_{2A} receptor. The receptor also exhibits stereoselectivity for the *R*-enantiomer of the phenylalkylamines and the *S*-enantionmer of the tryptamines. Pharmacology and structure-activity relationships are reviewed, for example, in Refs. 3–9. A large body of 3D-QSAR studies on this system exist [10–19] due to the fact that the experimental structure of the 5-HT_{2A} receptor is not presently available. In addition, homology models based on the structure of bacteriorhodopsin [20] were constructed—unfortunately, these studies are based on subsets (phenylalkylamines or tryptamines) only [17–19, 21].

In our previous study, we used a 3D-QSAR concept to construct a receptor surrogate based on 23 compounds defining the training set. It yielded a cross-validated r^2 of 0.954 and a predictive r^2 of 0.860. Subsequently, a series of 53 hypothetical congeneric compounds was evaluated with some of the ligands showing a predicted activity close to the experimental value of LSD [1]. Inherent with the 3D-QSAR approach, each agonist molecule was represented by a single conformation derived from a conformational search.

* To receive all correspondence

Key words: 4D-QSAR, multiple ligand representation, genetic algorithms, quasi-atomistic receptor models, induced-fit simulation, 5-HT_{2A} receptor

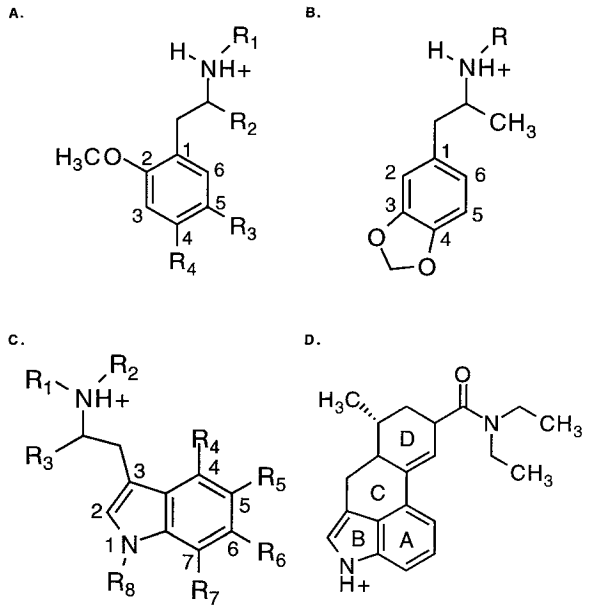

Figure 1. Conformer selection within the 5-HT_{2A} data set. A: phenylalkylamines/hallucinogens, B: phenylalkylamines/entactogens, C: tryptamines, D: ergoline (LSD).

A single representation, however, is susceptible to both systematic and conceptual bias. As 4D-QSAR techniques became available in the meantime, we felt that a re-evaluation of the 5-HT_{2A} receptor surrogate using a multiple-ligand representation is timely.

Quantitative structure-activity relationships (QSAR) is an area of computational research which builds atomistic or virtual models to predict quantities such as the binding affinity, the toxicity, or pharmacokinetic parameters of a given molecule. The idea behind QSAR is that structural features can be correlated with biological activity. Of particular interest for the biomedical research are QSAR based on three-dimensional models (3D-QSAR) because they allow for the simulation of directional forces—hydrogen bonds, metal-ligand contacts, polarization effects, and the interaction between electric dipoles. Such forces are known to play a key role for both molecular recognition and selective ligand binding.

While at the true biological receptor only one ligand molecule can bind at the time, a QSAR study is typically based on a series of ligand molecules binding "simultaneously" to an averaged receptor model. This leads to problems when subtle effects associated with the adaptation of the receptor to the individual ligand molecules or conformationally flexible H-bond moieties (see, for example, Ref. 22) are essential for ligand activity. A more fundamental problem in QSAR is associated with the mutual alignment of the ligand molecules—i.e. the choice of the bioactive conformation. Typically, the ligand alignment is derived from conformational-search studies combined with cluster-analyzing protocols—without direct means to control the result. If the construction of a receptor model is based on incorrect bioactive conformations, the resulting surrogate is hardly of any use for predictive purposes. While the alignment problem has long been recognized [23, 24], only the more recently developed 4D-QSAR technologies would seem to provide satisfactory solutions [25–30].

2 Methods

The present study is based on a family of genetically evolved quasi-atomistic receptor models [25–37]—receptor-surface constructs which represent a high level of model abstraction. Here, the essential information about the hypothetical receptor site is provided by means of a three-dimensional envelope, which surrounds the ligand molecules of the training set at the van-der-Waals distance and is populated with properties mapped onto its surface. The shape of the surface represents information about the steric nature of the receptor site; the associated properties represent other information of interest, such as hydrophobicity, partial charge, electrostatic potential, and hydrogen-bonding propensity [32]. Various algorithms to generate and validate receptor-surface models have been described [31, 33–35]. In this context, "quasi-atomistic" refers to the utilisation of properties such as hydrogen bonds, salt-bridges, and virtual solvent. The forces associated with these particles are directional in nature and allow for a subtle evaluation of ligand-receptor interactions.

The *Quasar* software is based on a true 4D-QSAR concept—the fourth dimension being the possibility to represent each molecule by an ensemble of conformations, orientations, protonation states, and enantiomers throughout the entire simulation, thereby reducing the bias associated with the choice of the bioactive conformation and the ligand alignment. In contrast to most other approaches in the field, our concept allows also for the simulation of local induced fit and H-bond flip-flop. The technical details of the construction of a family of receptor models in *Quasar* are published [25–27]; therefore, only the derivation of free energies of ligand binding are discussed at this point.

In our concept we have combined the approach of Blaney *et al*. [36] with a method of Still *et al*. [37] for estimating ligand solvation energies and a term to correct for the loss of entropy upon receptor binding following Searle and Williams [38]:

$$E_{bdg} \approx E_{lig\text{-}rec} + T\Delta S_{bdg} - E_{solv.,lig} + \Delta E_{int,lig} + E_{env.adapt.,lig} \quad (1)$$

$E_{lig\text{-}rec}$: Force-field energy ($E \leq 0.0$) of the ligand-receptor interaction [39–40]

$T\Delta S_{bdg}$: Change in ligand entropy ($T\Delta S \geq 0.0$) upon receptor binding [38]

$E_{solv.,lig}$: Ligand desolvation energy [37]

$\Delta E_{int.,lig}$: Change in ligand internal energy ($\Delta E \geq 0.0$) upon receptor binding [31]

$E_{env.adapt.,lig}$: Energy uptake ($E \geq 0.0$) required for modifying the receptor enevelope [25–27]

When using a multiple ligand representation, the interactions of all conformations, orientations, and protonation-states are calculated towards all members of the receptor-model family. The contribution of an individual entity to the total energy is determined using a normalized Boltzmann distribution [25–27]:

$$E_{bdg,tot} = \sum E_{bdg,ind} \cdot \exp(-w_i \cdot E_{bdg,ind}/E_{bdg,ind,lowest}) \quad (2)$$

where $w_i = (\sum E_{bdg,ind}/E_{bdg,ind,lowest})^{-1}$ is the normalizing factor

Free energies of ligand binding, ΔG°_{prd}, are then predicted by means of a linear regression between ΔG°_{exp} and E_{bdg} (cf. Eq. (1)) using the ligand molecules of the training set:

$$\Delta G^{\circ}_{prd} = |a| \cdot E_{bdg} + b \quad (3)$$

Slope and intercept of Eq. (3) are inherent to a given receptor model and are subsequently applied to predict the relative binding energy of ligand molecules different from those in the training set. In contrast to other methods, we calibrate each receptor system with a training set, rather than to apply an universal function for the various receptor systems.

3 Results and Discussion

The 5-HT_{2A} receptor system represents an attractive biological system for the more recently developed 4D-QSAR technology [25–27] as the included molecules belong to three different classes of receptor agonists—phenylalkylamines, tryptamines and ergolines (LSD). Although the selected data set would seem to be small in size, the ligands comprising it span a range of almost four orders of magnitude in the inhibition constant. The three-dimensional structures of the 30 ligand molecules used in this study (Figure 4; see also Ref. 1) were generated using the *MacroModel* 6.5 software [41] and subsequently refined in implicit aqueous solvent and evaluated using the AMBER force field [42].

The conformational search was performed on the cationic ligands using a Monte-Carlo method [43] as implemented in *MacroModel* 6.5 [41]. All conformers within 25 kcal/mol of the global minimum were collected and minimized to low gradient, leading to 6–84 conformers (26 on the average) per ligand molecule. Although there is no limitation in *Quasar* for the number of different conformations of a ligand molecule, the induced-fit simulation may lead to unrealistic results if a too diverse data set is used [25, 26]. For each agonist molecule, four low-energy conformers representing a

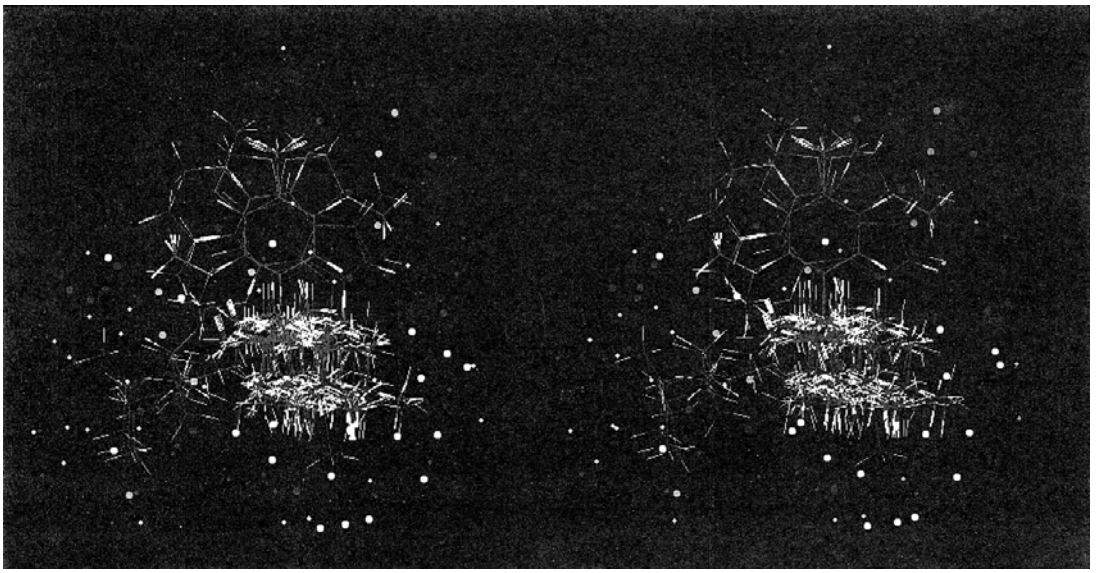

Figure 2. Stereoscopic view of the surrogate for the 5-HT_{2A} receptor. Color-coding scheme: red = positively charged salt bridge; blue = negatively charged salt bridge; yellow = H-bond donor; green = H-bond acceptor; saddle brown = positively charged hydrophobic; chocolate brown = negatively charged hydrophobic; pink = H-bond flip-flop; grey = neutral hydrophobic. The depicted model corresponds to the most frequently occurring property, averaged over the 200 receptor models.

Table 1. Experimental and predicted K_i values. Top: training set, bottom: test set

Ligand molecule	K_i (exp)	K_i (calc)
LSD	2.5×10^{-9}	2.8×10^{-9}
SS7	1.0×10^{-8}	1.3×10^{-8}
SS8	3.5×10^{-8}	3.3×10^{-8}
SS3	6.0×10^{-8}	6.1×10^{-8}
SS5	1.5×10^{-7}	7.4×10^{-8}
S22	1.7×10^{-7}	4.3×10^{-8}
SS9	2.1×10^{-7}	1.9×10^{-7}
S10	2.6×10^{-7}	3.2×10^{-7}
S16	3.1×10^{-7}	3.3×10^{-7}
S23	3.6×10^{-7}	2.1×10^{-7}
S18	4.0×10^{-7}	8.3×10^{-7}
S20	4.8×10^{-7}	4.6×10^{-7}
S30	5.7×10^{-7}	4.9×10^{-7}
SS1	8.2×10^{-7}	1.2×10^{-6}
S24	1.1×10^{-6}	1.4×10^{-6}
S19	1.3×10^{-6}	1.3×10^{-6}
S21	2.5×10^{-6}	2.1×10^{-6}
S12	3.1×10^{-6}	3.7×10^{-6}
S28	3.4×10^{-6}	2.8×10^{-6}
S13	5.0×10^{-6}	3.3×10^{-6}
S31	5.5×10^{-6}	3.2×10^{-6}
S27	1.3×10^{-5}	1.1×10^{-5}
S29	1.6×10^{-5}	1.5×10^{-5}
SS4	2.4×10^{-8}	6.9×10^{-8}
SS6	3.4×10^{-8}	5.2×10^{-8}
S11	1.1×10^{-8}	7.1×10^{-8}
S14	3.0×10^{-7}	2.7×10^{-7}
SS2	1.0×10^{-7}	9.3×10^{-7}
S17	1.2×10^{-7}	2.7×10^{-6}
S26	3.5×10^{-6}	1.8×10^{-6}

major conformational class were then selected using the *PrGen* software [44]. To allow for a comparison with our previous 3D-QSAR study, we have applied the identical ligand-alignment protocol (cf. Ref. 1).

Next, partial charge models were calculated using the CM-1 algorithm (as implemented in AMSOL 5.4 [45]. Solvation energies were calculated based on the GB/SA algorithm [37]. Relative energies of the various ligand conformers were determined by full minimization using the AMBER force field and implicit aqueous solvent. A training set comprising 23 molecules (92 conformers total) representing the largest possible diversity from the whole data set was then defined following [1]. The remaining seven ligands (28 conformers total) served as the test set.

The mean envelope (cf. Refs. 25–27) was generated using all ligands of the training set; the individual envelopes were then generated using a field-based, anisotropic algorithm [25; 46]. The resulting "induced fit" ranged from 1.5–2.3 Å (rms) with associated energies of 2.2–5.3 kcal/mol. Using an initial population of 200 receptor models and a dynamic transcription-error rate [47], the system was allowed to evolve for 6,000 cross-over cycles, corresponding to 30 generations. The simulation reached a cross-validated r^2 of 0.951 and a predictive r^2 of 0.859. These quantities reflect values averaged over the 200 models—which, among themselves, differ in 26–35 % of the mapped 204 properties. The cross-validation was based on three groups comprising 7–9 ligands (leave-*n*-out). A stereo representation of the receptor surrogate is depicted in Figure 2; experimental and calculated K_i

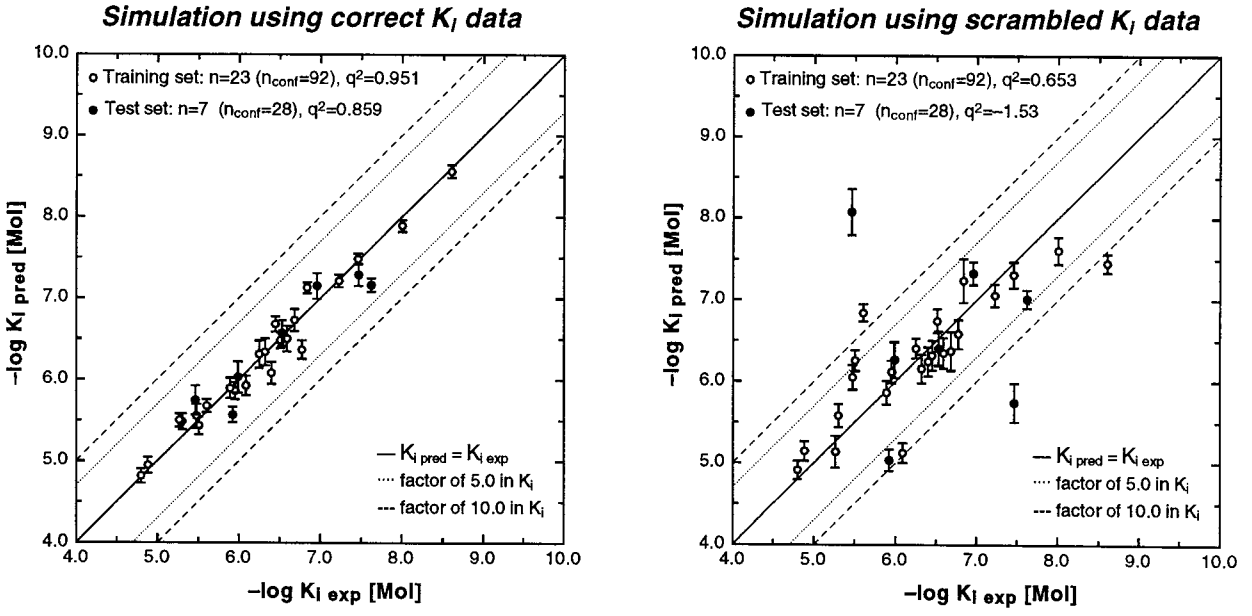

Figure 3. Graphical comparison of experimental and predicted inhibition constants for the 5-HT_{2A} receptor: simulation using correct K_i data (left), simulation using scrambled K_i data (right). The error bar corresponds to the variation of the predicted K_i value over the 200 individual receptor models.

Table 2. Ligand–receptor energy partitioning (all values are given in kcal/mol). Top: training set, bottom: test set

Ligand	E(tot)	E(vdw)	E(elec)	E(HBond)	E(pol)	E(solv)	TΔS	E(int)	E(Ind.Fit)	ΔG°(calc)
LSD	−27.368	−13.372	−4.172	−4.535	−5.289	−10.296	−2.10	0.649	2.228	−12.095
SS7	−31.954	−10.601	−5.841	−6.711	−8.801	−12.546	−3.50	0.598	4.161	−11.148
SS8	−30.759	−10.675	−4.847	−6.923	−8.314	−12.467	−3.50	0.000	4.215	−10.577
SS3	−30.426	−10.290	−6.008	−5.557	−8.571	−12.003	−3.50	0.597	4.135	−10.191
SS5	−30.264	−10.790	−4.751	−6.489	−8.234	−12.484	−3.50	0.000	4.208	−10.072
S22	−27.827	−13.882	−5.257	−0.963	−7.726	−11.862	−2.10	0.048	4.820	−8.997
SS9	−32.199	−11.885	−4.047	−6.941	−9.326	−14.526	−3.50	0.514	4.148	−9.511
S10	−26.926	−11.303	−2.324	−5.609	−7.691	−10.195	−3.50	0.192	3.852	−9.186
S16	−29.642	−11.475	−5.213	−5.235	−7.719	−12.836	−2.80	0.167	4.682	−9.157
S23	−28.105	−13.366	−5.555	−2.955	−6.229	−11.475	−2.80	0.053	4.337	−9.441
S18	−26.886	−13.376	−6.472	−1.150	−5.888	−11.078	−2.10	0.119	4.998	−8.591
S20	−28.765	−13.184	−4.589	−3.617	−7.376	−12.444	−2.80	0.047	4.520	−8.953
S30	−31.620	−11.472	−6.250	−5.683	−8.215	−14.580	−2.80	0.048	5.279	−8.913
SS1	−28.859	−10.610	−5.555	−4.733	−7.961	−12.561	−2.80	0.636	4.489	−8.373
S24	−26.818	−14.848	−3.286	−1.740	−6.943	−10.523	−4.20	0.103	3.707	−8.284
S19	−26.689	−13.976	−4.858	−0.563	−7.291	−11.504	−2.80	0.491	3.563	−8.332
S21	−26.422	−14.172	−3.846	−1.232	−7.172	−11.390	−2.10	0.085	4.820	−8.026
S12	−27.583	−11.810	−4.617	−3.014	−8.142	−12.735	−2.10	0.297	4.773	−7.679
S28	−26.415	−12.331	−2.920	−3.405	−7.759	−11.983	−2.10	0.130	4.356	−7.847
S13	−27.405	−11.922	−5.498	−1.765	−8.220	−12.763	−2.10	0.056	4.735	−7.750
S31	−26.859	−15.349	−2.757	−0.977	−7.777	−11.682	−2.80	0.077	4.525	−7.775
S27	−27.012	−12.034	−3.283	−3.334	−8.361	−13.101	−2.10	0.105	4.714	−6.992
S29	−25.399	−13.512	−1.925	−2.187	−7.775	−11.965	−2.10	0.217	4.315	−6.802
SS4	−30.953	−10.902	−5.604	−5.771	−8.675	−12.560	−3.50	0.598	4.175	−10.120
SS6	−32.374	−10.634	−5.762	−7.340	−8.639	−13.387	−3.50	0.751	4.439	−10.297
S11	−32.405	−10.337	−6.492	−7.378	−8.199	−13.030	−3.50	1.362	4.412	−10.102
S14	−31.327	−11.942	−6.702	−4.275	−8.408	−13.932	−2.80	0.000	5.314	−9.281
SS2	−30.283	−11.673	−5.808	−3.505	−9.297	−13.264	−2.80	1.091	4.609	−8.520
S17	−26.643	−13.837	−4.398	−1.268	−7.141	−11.627	−2.10	0.092	4.952	−7.873
S26	−28.097	−11.035	−4.862	−3.843	−8.358	−13.115	−2.10	0.067	4.694	−8.121

values are compared in Table 1; a graphical comparison is shown in Figure 3.

The rms deviation for the 23 ligand molecules of the training set of 0.22 kcal/mol corresponds to an uncertainty factor of 1.4 in K_i, the maximal individual deviation (compound **S22**, 0.57 kcal/mol) to a factor 2.5 in K_i. The corresponding rms for the seven ligands of the test set is 0.37 kcal/mol (1.8 in K_i) and the maximal individual deviation 0.65 kcal/mol (compound **SS4**; 2.9 in K_i). Finally, the validity of the model family was assessed by a series of six scramble tests. The resulting predictive r^2 values of −1.04/−1.53/−1.71/+0.10/−0.73/−1.40 (average: −1.05) demonstrate the sensitivity of the surrogate family towards the experimental K_i data, for which it should establish a quantitative structure-activity relationship (Figure 3). The ligand-receptor interactions have been interpreted by analyzing the contribution of each functional group towards ΔG°, a new option in *Quasar* V2.5 [48]:

$$\Delta G^{\circ}_{calc.} = E_{lig\text{-}rec}/slope - T\Delta S_{bdg} - E_{solv.,lig} - \Delta E_{int.,lig} - E_{env.adapt.,lig} \quad (4)$$

As derived above, these contributions are Boltzmann-weighted with respect to the four conformers included for each ligand and averaged over the 200 receptor models. While the enthalpic contributions range from −25.4 kcal/mol (**S29**) to −32.4 kcal/mol (**S11**), the major effects towards ΔG° are observed from ligand desolvation and induced-fit (Table 2). Ligand desolvation energies range from −51.0 kcal/mol (**S10**) to −72.9 kcal/mol (**S30**), an immense variation to be compensated by ligand-receptor interactions [49]. The energy associated with the simulated induced fit ranges from 2.2 kcal/mol (**LSD**) to 5.3 kcal/mol (**SS2**). The changes in entropy upon receptor binding (2.1–4.2 kcal/mol) and the internal strain relative to the lowest-energy conformer (0.0–1.4 kcal/mol) are rather small quantities. The strong interaction of **LSD** with the 5-HT_{2A} receptor can be rationalized by the second-lowest desolvation energy of the series (−10.3 kcal/mol) and the lowest induced-fit energy (+2.2 kcal/mol) −1.9 kcal/mol lower than the next-potent compound **SS7**. Its ligand-receptor interaction energy (−27.4 kcal/mol) is rather modest in the series. A significant contribution thereof originates from polarization (−5.3 kcal/mol) and H-bonds (−4.5 kcal/mol); the remaining components are electrostatic (−4.2 kcal/mol) and hydrophobic in nature (−13.4 kcal/mol).

Table 3. Relative conformer energies (in kcal/mol) and selected Fraction. Top: training set, bottom: test set. Values given in italics represent the 3D-QSAR study [1]. The most frequently selected fraction is shown in bold face

Ligand	nth cf; E(rel)	Fraction #1	nth cf; E(rel)	Fraction #2	nth cf; E(rel)	Fraction #3	nth cf; E(rel)	Fraction #4	nth cf; E(rel)	Fraction 3D
LSD	1 : 0.000	0.000	2 : 0.645	**0.974**	3 : 0.789	0.026	4 : 0.813	0.000	*1 : 0.000*	*1.0*
SS7	1 : 0.000	0.000	2 : 0.359	0.000	3 : 0.598	**1.000**	4 : 0.741	0.000	*3 : 0.598*	*1.0*
SS8	1 : 0.000	**1.000**	2 : 0.382	0.000	3 : 0.597	0.000	4 : 0.764	0.000	*3 : 0.597*	*1.0*
SS3	1 : 0.000	0.000	2 : 0.358	0.000	3 : 0.597	**1.000**	4 : 0.741	0.000	*3 : 0.597*	*1.0*
SS5	1 : 0.000	**1.000**	2 : 0.406	0.000	3 : 0.598	0.000	4 : 0.813	0.000	*3 : 0.598*	*1.0*
S22	1 : 0.000	**0.595**	2 : 0.048	0.005	4 : 0.119	0.399	6 : 0.191	0.001	*4 : 0.119*	*1.0*
SS9	1 : 0.000	0.133	5 : 0.406	0.021	9 : 0.598	**0.846**	13 : 0.789	0.000	*9 : 0.598*	*1.0*
S10	1 : 0.000	0.189	2 : 0.144	0.016	3 : 0.239	**0.796**	4 : 0.335	0.000	*1 : 0.000*	*1.0*
S16	1 : 0.000	0.000	2 : 0.144	0.000	3 : 0.167	**0.999**	4 : 0.430	0.001	*4 : 0.430*	*1.0*
S23	1 : 0.000	**0.359**	2 : 0.048	0.332	3 : 0.071	0.002	5 : 0.119	0.307	*5 : 0.119*	*1.0*
S18	1 : 0.000	0.000	4 : 0.119	**0.987**	7 : 0.143	0.013	8 : 0.215	0.000	*8 : 0.215*	*1.0*
S20	1 : 0.000	0.002	2 : 0.047	**0.995**	5 : 0.071	0.000	7 : 0.143	0.003	*7 : 0.143*	*1.0*
S30	1 : 0.000	0.000	2 : 0.048	1.000	3 : 0.191	0.000	5 : 1.003	0.000	*3 : 0.191*	*1.0*
SS1	1 : 0.000	**0.823**	2 : 0.454	0.000	4 : 1.099	0.065	5 : 5.016	0.113	*1 : 0.000*	*1.0*
S24	1 : 0.000	0.081	3 : 0.071	**0.602**	4 : 0.095	0.003	5 : 0.191	0.314	*3 : 0.071*	*1.0*
S19	1 : 0.000	0.006	3 : 0.215	0.028	5 : 0.406	0.000	7 : 0.502	**0.966**	*5 : 0.406*	*1.0*
S21	1 : 0.000	0.281	2 : 0.048	0.204	3 : 0.143	**0.483**	4 : 0.191	0.032	*4 : 0.191*	*1.0*
S12	1 : 0.000	0.001	2 : 0.167	**0.860**	3 : 1.099	0.139	4 : 1.194	0.000	*2 : 0.167*	*1.0*
S28	1 : 0.000	0.006	2 : 0.024	**0.733**	3 : 0.406	0.000	4 : 0.430	0.261	*1 : 0.000*	*1.0*
S13	1 : 0.000	**0.953**	2 : 0.191	0.000	3 : 1.123	0.004	4 : 1.218	0.042	*2 : 0.191*	*1.0*
S31	1 : 0.000	0.112	2 : 0.048	0.252	4 : 0.096	**0.466**	5 : 0.120	0.042	*1 : 0.000*	*1.0*
S27	1 : 0.000	0.051	2 : 0.024	**0.867**	3 : 1.004	0.000	4 : 1.028	0.082	*2 : 0.024*	*1.0*
S29	1 : 0.000	0.023	2 : 0.024	**0.499**	3 : 0.406	0.029	4 : 0.430	0.448	*1 : 0.000*	*1.0*
SS4	1 : 0.000	0.000	2 : 0.406	0.000	3 : 0.598	**1.000**	4 : 0.813	0.000	*3 : 0.598*	*1.0*
SS6	1 : 0.000	0.458	2 : 0.502	0.000	3 : 1.266	0.000	4 : 1.386	**0.542**	*1 : 0.000*	*1.0*
S11	1 : 0.000	0.000	2 : 0.478	0.000	3 : 1.218	0.000	4 : 1.362	**1.000**	*2 : 0.478*	*1.0*
S14	1 : 0.000	**1.000**	2 : 0.143	0.000	3 : 0.167	0.000	4 : 0.263	0.000	*4 : 0.263*	*1.0*
SS2	1 : 0.000	0.002	2 : 0.478	0.000	3 : 1.027	0.089	4 : 1.099	**0.909**	*1 : 0.000*	*1.0*
S17	1 : 0.000	0.238	2 : 0.024	0.017	4 : 0.120	**0.663**	5 : 0.144	0.081	*4 : 0.120*	*1.0*
S26	1 : 0.000	0.000	2 : 0.024	**0.957**	3 : 1.004	0.000	4 : 1.028	0.043	*2 : 0.024*	*1.0*

This indicates that while ligand-receptor interactions are, of course, essential for ligand binding, determining effects are often a consequence of ligand desolvation and induced fit—two terms that are often not accounted for in QSAR studies.

The intra-ligand analysis (Figure 4) indicates that groups important for a high affinity are more apolar in nature (methoxy, methyl, nitro, and halide substituents). While the ammonium functionality common to all agonist molecules is essential for binding, its contribution to ΔG° is not favorable due to its high desolvation energy (-10.6 for **SS5** to -3.0 kcal/mol for **S10**) and a modest enthalpic contribution towards the surrogate family (-11.5 for **S12** to -1.5 kcal/mol for **LSD**). Nonetheless, it would seem that a major contribution towards the ΔG° of LSD stems from the "less unfavorable desolvation energy" of its ammonium group—apparently better shielded in the rigid LSD than in the conformationally flexible phenylalkylamines and tryptamines agonist molecules. Still, the major effect originates in the energy associated with the receptor-to-ligand adaptation (induced fit). This could be explained in terms that the binding pocket optimally fits LSD, while all other molecules would have to pay a higher price in energy for this adaptation.

How do the results from 4D-QSAR compare with our earlier study based on 3D-QSAR [1]? This can be best discussed by analyzing the selected conformers. Out of the 30 agonists used in this study, the genetic algorithm identified ten to select a single conformer by 100 % (Table 3). Thereof, only four find the lowest-energy conformer relevant for binding to the 5-HT_{2A} receptor. The others prefer energetically less favorable entities which, however, are in a position to engage in stronger interactions with the receptor surrogate: two select the second while the remaining four select the fourth-lowest entity. Another seven ligands select a single conformer with a preference >90 %, 10 with a preference >50 %, while the remaining three ligands (**S23**, **S21**, **S31**) calculate their mean interaction significantly from more than one conformer. In the 3D-QSAR study [1], only 10 of the 30 conformers were identified "correctly" (cf. Table 3)—suggesting an error rate of 67 % in selecting the bioactive conformer.

A comparison of the cross-validated r^2 of 0.951 (3D-QSAR: 0.954) and the predictive r^2 of 0.859 (3D-QSAR: 0.860) would not seem to indicate any obvious advantage of selecting the computationally more demanding 4D-QSAR approach, except that it was already achieved after 6,000

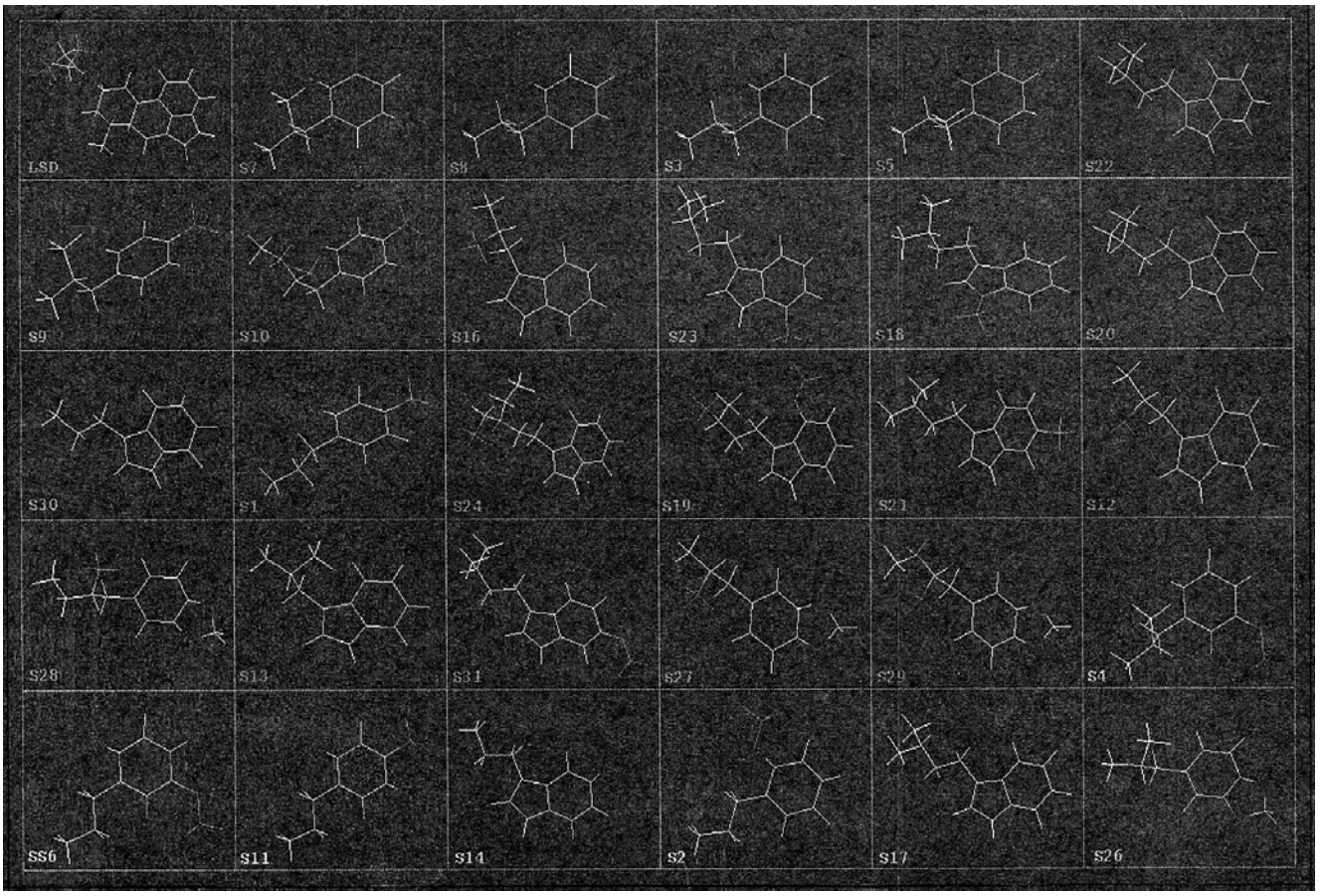

Figure 4. Contribution of the 5-HT_{2A} agonist towards ΔG° of ligand binding. Color-coding scheme: blue = good, green = fair, white = residual, yellow = poor, red = bad. The values are averaged over the four conformers and all 200 receptor models.

cross-over cycles [51], rather than after 8,000 in the 3D-approach [1]. This cross validated r^2 of the *scramble text*—an criterion for the sensitivity of the model towards the biological data—is a more critical parameter. The value obtained for the 4D-QSAR study (−1.05) indicates a strong sensitivity, while the value of the 3D-QSAR experiment (+0.14) is just acceptable. Other parameters suggesting that a model based on a single-conformer selection is more difficult to obtain include the slope (3D: 0.133; 4D: 0.210), the number of mapped properties (3D: 290; 4D: 199) and the maximal deviation of an individual ligand molecule (3D: 0.74 kcal/mol; 4D: 0.65 kcal/mol).

But the key issue is not the question whether the 4D-QSAR approach yields better values or comparable values at shorter simulation times. It is rather the question if we shall accept a biased selection of the bioactive conformers (3D-QSAR) if more potent approaches are available (4D-QSA, cf. Ref. [25–30]), significantly reducing this bias. These new technologies would also seem less susceptible to the *right answer* (*predictive r^2 value*) *for the wrong reason* (*pharmacophore hypothesis, conformer selection*) tag, still "attached" to older QSAR concepts as the more recent algorithms allow for a multiple-conformation [25–29], orientation [27], protonation [25], and enantiomer [52] representation of the ligand molecules. In this context, it is worth mentioning that the simulated evolution does not simply select the lowest-energy conformer due to the lowest (i.e. non-existent) internal strain (see Table 3)—it rather would suggest that a subtle balance between internal strain, induced, fit, and ligand-receptor interaction energy (cf. Table 2) leads to the observed selection pattern. In general, genetic algorithms react very sensitively to incorrect mechanisms [28], biological data (cf. above) or bio*inactive* conformations which implies that little can be lost when using 4-D QSAR instead of 3-D QSAR, but much gained as the boundary conditions (here the Boltzmann-weighted ensemble of lignad representations, cf. Eq. 3) are controlled by the evoluation itself.

4 Conclusions

In absence of an experimentally determined receptor structure, 4D-QSAR techniques provide an elegant approach to the estimation of free energies of ligand binding. The *Quasar* concept developed at our laboratory not only allows for the

representation of the ligand molecules by an ensemble of conformation, orientations, and protonation states but also takes local induced fit and H-bond flip-flop into account. This approach has been used to predict the K_i values of phenylalkylamines, tryptamines and ergolines (LSD) binding to the 5-HT_{2A} receptor. The results indicate that while direct enthalpic contributions are essential for ligand binding, major effects are often a consequence of ligand desolvation and induced fit—mechanisms that are often ignored in QSAR studies.

Such a surrogate could be used to identify new, more active compounds—an intention that the authors don't even think of. Rather, the technology should benefit the quick assessment of the binding affinity of related compounds in the pharmaceutical sector and for governmental or regulatory bodies. For this purpose, the coordinates of both ligands and surrogate are made freely available to non-profit institutions. They can be requested via http://www.biograf.ch

4 Acknowledgment

The authors are indebted to Professor Karl-Artur Kovar for initializing this interesting project, to Dr. Meike Schulze-Alexandru for a most constructive cooperation during the 3D-QSAR studies, to Dr. Daniel R. McMasters (formerly at the Biographics Laboratory) for the efficient PERL scripts significantly facilitating conformational-search analysis, and to Professor Max Dobler for developing part of the *Quasar* algorithm.

Financial support from the *Margaret and Francis Fleitmann Foundation* (Luzern, Switzerland) and the *Foundation for Animal-free Research* (Zürich, Switzerland) is gratefully acknowledged.

References

[1] Schulze-Alexandru, M., Kovar, K.-A., and Vedani, A., Quasi-atomistic Receptor Surrogates for the 5-HT_{2A} Receptor: A QSAR Study on Hallucinogenic Substances, *Quant. Struct.-Act. Relat. 18*, 548–560 (1999).

[2] Glennon, R.A., Investigations with Serotonin and Hallucinogenic Agents, *http://www.phc.vcu.edu/rag/serotonin/index.html* (1998).

[3] Nichols, D.E., Structure-activity Relationships of Phenylethylamine Hallucinogens, *J. Pharm. Sci. 70*, 839–849 (1981).

[4] Glennon, R.A., and Gessner, P.K., Serotonin Receptor Binding Affinities of Tryptamine Analogues, *J. Med. Chem. 22*, 428–432 (1979).

[5] Nichols, D.E., Studies of the Relationship Between Molecular Structure and Hallucinogenic Activity, *Pharmacol. Biochem. Behav. 24*, 335–340 (1986).

[6] Oberlender, R., Ramachandran, P.V., Johnson, M.P., Huang, X., and Nichols, D.E., Effect of a Chiral 4-alkyl Substituent in Hallucinogenic Amphetamines, *J. Med. Chem. 38*, 3593–3601 (1995).

[7] Huang, X., Marona-Lewicka, D., Pfaff, R.C., and Nichols, D.E., Drug Discrimination and Receptor Binding Studies of N-Isopropyl Lysergamide Derivatives, *Pharmcol. Biochem. Behav. 47*, 667–673 (1994).

[8] Nichols, D.E., and Glennon, R.A., in: Jacobs, B.L., (Ed.), Hallucinogens: Neurochemical, Behavioral and Clinical Perspectives, Raven Press, New York 1984, pp. 95–141.

[9] Shulgin, A.T., in: Wolf, M.E., (Ed.), Burger's Medicinal Chemistry, part III, 4th Edition, Wiley, New York 1981, 1109–1137.

[10] Westkaemper, R.B., and Glennon, R.A., Approaches to Molecular Modeling Studies and Specific Applications to Serotonin Ligands and Receptors, *Pharmacol. Biochem. Behav. 40*, 1019–1031 (1991).

[11] Hibert, M.F., Trumpp-Kallmeyer, S., Bruinvels, A., and Hoflack, J., Three-dimensional Models of Neurotransmitter Binding to G-Protein Coupled Receptors, *Mol. Pharmacol. 40*, 8–15 (1991).

[12] Trumpp-Kallmeyer, S., Hoflack, J., Bruinvels, A., and Hibert, M.F., Modeling of G-Protein Coupled Receptors: Application to Dopamine, Adrenaline, Serotonin, Acetylcholine, and Mammalian Opsin Receptors, *J. Med. Chem. 35*, 3448–3462 (1992).

[13] Pardo, L., Bakkesteros, J.A., Osman, R., and Weinstein, H., On the Use of the Transmembrane Domain of Bacteriorhodopsin as a Template for Modeling the Three-dimensional Structure of Guanine Nucleotide-binding Regulatory Protein-coupled Receptors, *Proc. Natl. Acad. Sci. USA 89*, 4009–4012 (1992).

[14] Hoeltje, H.-D., and Jendretzki, U.K., Construction of a Detailed Serotoninergic 5-HT_{2A} Receptor Model, *Arch. Pharm. (Weinheim) 328*, 584–597 (1995).

[15] Zhang, D., and Weinstein, H., Signal Transduction by a 5-HT_2 Receptor: a Mechanistic Hypothesis from Molecular Dynamics Simulations on the Three-Dimensional Model of the Receptor Complexed to Ligands, *J. Med. Chem. 36*, 934–938 (1993).

[16] Clare, B.W., Structure-Activity Correlations for Psychotomimetics. 1. Phenylalkylamines: Electronic, Volume, and Hydrophobicity Parameters, *J. Med. Chem. 33*, 687–702 (1990).

[17] Clare, B.W., The Frontier Orbital Phase Angels: Novel QSAR Descriptors for Benzene Derivatives Applied to Phenylalkylamine Hallucinogens, *J. Med. Chem. 41*, 3845–3856 (1998).

[18] Gomez-Jeria, J.S., and Morales-Lagos, D.R., Quantum-chemical Approach to the Relationship between Molecular Structure and Serotonin Receptor Binding Affinity, *J. Pharm. Sci. 73*, 1725–1728 (1984).

[19] Gupta, S.P., Singh, P., and Bindal, M.C., QSAR Studies on Hallucinogens, *Chem. Rev. 83*, 633–649 (1983).

[20] Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E., and Downing, K.H., Model for the Structure of Bacteriorhodopsin Based on High-resolution Electron Cryomicroscopy, *J. Mol. Biol. 213*, 899–929 (1990).

[21] Beuerle, G., and Kovar, K.-A., and Schulze-Alexandru, M., Three-Dimensional Quantitative Structure-activity Relationships of Hallucinogenic Phenylalkylamine and Tryptamine Derivatives: Studies Using Comparative Molecular Field Analysis (CoMFA), *Quant. Struct.-Act. Relat. 16*, 447–458 (1997).

[22] Montgomery, J.A., and Secrist III, J.A., PNP Inhibitors, *Persp. Drug Discov. Design 2*, 205–220 (1994).

[23] Kubinyi, H., QSAR and 3D QSAR in Drug Design. 2. Applications and Problems, *Drug Discovery Today 2*, 538–546 (1997).

[24] Kubinyi, H., Folkers, G., and Martin, Y.C., QSAR: Current State, Scope, and Limitations, *Perspectives in Drug Discovery and Design 12*, 3–23 (1998).

[25] Vedani, A., Briem, H., Dobler, M., Dollinger, K., and McMasters, D.R., Multiple Conformation and Protonation-State Representation in 4D-QSAR: The Neurokinin-1 Receptor System, *J. Med. Chem. 43*, 4416–4427 (2000).
[26] Vedani, A., McMasters, D.R., and Dobler, M., Multi-conformational Ligand Representation in 4D-QSAR: Reducing the Bias Associated with Ligand Alignment, *Quant. Struct.-Act. Relat. 19*, 149–161 (2000).
[27] Vedani, A., and Dobler, M., Multi-dimensional QSAR in Drug Research: Predicting Binding Affinities, Toxicity, and Pharmacokinetic Parameters, in: Jucker, E., (Ed.), Progress in Drug Research, Birkhäuser, Basel/Boston/Berlin 2000, 105–135.
[28] Hopfinger, A.J., Wang, S., Tokarski, J.S., Jin, B.Q., Albuquerque, M., Madhav, P.J., Duraiswami, C., Construction of 3D-QSAR Models Using 4D-QSAR Analysis Formalism, *J. Am. Chem. Soc. 119*, 10509–10524 (1997).
[29] So, S.S., and Karplus, M., Three-dimensional Quantitative Structure-activity Relationships From Molecular Similarity Matrices and Genetic Neural Networks, *J. Med. Chem. 40*, 4347–4359 (1997).
[30] Sprague, P.W., and Hoffmann, R., Catalyst Pharmacophore Models and Their Utility as Queries for Searching 3D Databases, in: van de Waterbeemd, H., and Folkers, G., (Eds.), Computer-Assisted Lead Finding and Optimization, VCH Weinheim/Germany 1997, 223–240 (1997).
[31] Vedani, A., Dobler, M., and Zbinden, P., Quasi-Atomistic Receptor Surface Models: A Bridge Between 3-D QSAR and Receptor Modeling, *J. Am. Chem. Soc. 120*, 4471–4477 (1998).
[32] Hahn, M., Receptor Surface Models. 1. Definition and Construction, *J. Med. Chem. 38*, 2080–2090 (1995).
[33] Srivastava, S., Richardson, W.W., Bradley, M.P., and Crippen, G., Three-dimensional Receptor Modeling Using Distance Geometry and Voronoi Polyhedra, in: Kubinyi, H., (Ed.), 3D-QSAR in drug design: Theory, Methods and Applications, Escom, Leiden 1993, 80–116.
[34] Walters, D.E., and Hinds, R.M., Genetically Evolved Receptor Models: A Computational Approach to Construction of Receptor Models, *J. Med. Chem. 37*, 2527–2536 (1994).
[35] Hahn, M., and Rogers, D., Receptor Surface Models. 2. Application to Quantitative Structure-Activity Studies, *J. Med. Chem. 38*, 2091–2102 (1995).
[36] Blaney, J.M., Weiner, P.K., Dearing, A., Kollman, P.A., Jorgensen, E.C., Oatley, S.J., Burridge, J.M., and Blake. J.F., Molecular Mechanics Simulation of Protein-ligand Interactions: Binding of Thyroid Analogues to Prealbumin, *J. Am. Chem. Soc. 104*, 6424–6434 (1982).
[37] Still, W.C., Tempczyk, A., Hawley, R.C., and Hendrickson, T., Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics, *J. Am. Chem. Soc. 112*, 6127–6129 (1990).
[38] Searle, M.S., and Williams, D.H., The Cost of Conformational Order: Entropy Changes in Molecular Associations, *J. Am. Chem. Soc. 114*, 10690–10697 (1992).
[39] Vedani, A., and Huhta, D.W., A New Force Field for Modeling Metalloproteins, *J. Am. Chem. Soc. 112*, 4759–4767 (1990).
[40] As a Virtual Particle (VP) in a quasi-atomistic approach has no bonding partners (i.e. unlike functional groups of real molecules it bears no lone-pairs), we apply a reduced function to determine the non-electrostatic contribution to the H-bond energy involving a VP: For the constellation Don-H···VP, we correct for nonlinearity of the Don-H···VP angle (compulsory assuming a perfect directionality at the VP). For the arrangement Acc···VP, we correct for the deviation of the virtual hydrogen bond from the closest lone pair at the acceptor fragment (angle LP-Acc···VP) and assume a perfect linearity of the hydrogen bond. Derivation and calibration of the directional function for H-bond interactions is described in Refs. 39 and 31.
[41] Mohamadi, F., Richards, N.G.J., Guida, W.C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T., and Still, W.C., MacroModel—An Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics, *J. Comput. Chem. 11*, 440–467 (1990).
[42] Weiner, S.J., Kollmann, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta Jr. S., and Weiner, P., A New Force Field for Molecular-Mechanical Simulation of Nucleic Acids and Proteins, *J. Am. Chem. Soc. 106*, 765–784 (1984).
[43] Chang, G., Guida, W.C., and Still, W.C., An Internal Coordinate Monte Carlo Method for Searching Conformational Space, *J. Am. Chem. Soc. 111*, 4379–4386 (1989).
[44] Zbinden, P., Dobler, M., Folkers, G., and Vedani, A., PrGen: Pseudoreceptor Modeling Using Receptor-Mediated Ligand Alignment and Pharmacophore Equilibration, *Quant. Struct.-Act. Relat. 17*, 122–130 (1998).
[45] Cramer, C.J., and Truhlar, D.G., AM1–SM2 and PM3–SM3 parametrized SCF solvation models for free energies in aqueous solution, *J. Comp.-Aided Molec. Design 6*, 629–666 (1992). Distributed by QCPE, University of Indiana, Bloomington, Indiana, USA (Program #606).
[46] The actual field exerted by all atoms of a given ligand molecule on to its envelope is determined using a non-directional force field. It includes terms for electrostatic, van-der-Waals and potential H-bond interactions.
[47] The dynamic mutation-rate option allows for a definition of an initial, a lower and an upper limit for the transcription-error rate expressed by random mutations. The genetic algorithm then dynamically adjust this rate—within the defined limits—to obtain an optimal ration between successful mutations and failures. In the actual simulations those values were set at 0.025; 0.02–0.10.
[48] *http://www.biograf.ch/software.html* (2000)
[49] Because of this wide range, we have applied an attenuation factor of 5.0 to reduce the influence of ligand desolvation—a variable parameter in *Quasar*. As a consequence, all desolvation energies given in the following and in Table 2 are reduced by this factor. A simulation with unattenuated solvation energy lead to clearly inferior results [50].
[50] It yielded a cross-validated r^2 of 0.715 for the 23 ligands of the training set and a mere predictive r^2 of 0.155 for the 7 ligands of the test set.
[51] The simulation was terminated at this point as the *rms* deviation of $\Delta G^{\circ}_{exp} - \Delta G^{\circ}_{exp}$ reached a value comparable to the experimental error. A continuation of the evolution would only apparently lead to better predictions as the calculation cannot be more accurate than the underlying experimental IC_{50} values.
[52] Vedani, A., and Dobler, M., The Benzodiazepine/$GABA_A$ Receptor: Modeling Racemic Mixtures in 4D-QSAR, *Quant. Struct.-Act. Relat.* (in preparation).

Received on September 18, 2000; accepted on December 11, 2000