Quasi-Atomistic Receptor Surface Models: A Bridge between 3-D QSAR and Receptor Modeling

Angelo Vedani,* Max Dobler,[†] and Peter Zbinden[‡]

Contribution from the Biographics Laboratory 3R, Missionstrasse 60, 4055 Basel, Switzerland

Received November 20, 1997. Revised Manuscript Received February 19, 1998

Abstract: A "quasi-atomistic receptor model" refers to a three-dimensional receptor surface, populated with atomistic properties (hydrogen bonds, salt bridges, hydrophobic particles, and solvent) mapped onto it. In contrast to other 3D-QSAR approaches, an algorithm developed at our laboratory allows for the adaptation of the receptor-surface defining envelope to the topology of the individual ligand molecules. In addition, it includes *H-bond flip-flop particles* which can simultaneously act as H-bond donors and H-bond acceptors toward different ligand molecules, binding to the surrogate within a pharmacophore hypothesis. Such particles mimic amino-acid residues able to engage in differently directed H-bonds at the true biological receptor. Ligand−receptor interaction energies are evaluated using a directional force field for hydrogen bonds and salt bridges. On the basis of a series of ligand molecules with individually adapted receptor envelopes, the software Quasar allows a family of receptor models to be generated using a genetic algorithm combined with cross-validation. Our concept has been used to derive semiquantitative structure−activity relationships for the β2-adrenergic, aryl hydrocarbon, cannabinoid, neurokinin-1, and sweet-taste receptor as well as for the enzyme carbonic anhydrase. The receptor surrogates for these systems are able to predict free energies of ligand binding for independent sets of test ligand molecules within 0.4−0.8 kcal/mol (RMS) of the experimental value.

Introduction

Quantitative structure−activity relationship (QSAR) is an area of computational research which builds mathematical or virtual models to explain the biological activity for a series of compounds using topological and physicochemical ligand data. The idea behind QSAR is that molecular properties can be correlated with biological activity. Of particular interest for drug-design purposes are three-dimensional models, including atomistic and surface constructs of the hypothetical binding site, as they provide intuitive receptor surrogates.

In the absence of an experimental structure (X-ray or NMR) of the macromolecular receptor, whole models derived from the three-dimensional structure of a closely related homologue provide a high level of surrogate realism (cf., for example, ref 1). Binding-site models such as peptidic pseudoreceptors can approach the quality of homology models if carefully validated. The philosophy underpinning the pseudoreceptor concept is to engage the bound species in sufficient, specific noncovalent binding so as to mimic the essential ligand−macromolecule interactions at the true biological receptor (see, for example, refs 2−6). Although, in general, the sequence and arrangement of the building blocks of a pseudoreceptor (amino acid residues, metal ions, solvent) and its natural counterpart bear only little resemblance, the receptor and surrogate should accommodate a series of ligands in a relatively similar binding sense.[4,7]

A receptor-surface model provides essential information about the hypothetical receptor site by means of a three-dimensional envelope 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.[8] Various algorithms to generate such three-dimensional receptor surrogates have been described and validated.[8−13] Of these, the software GERM[10] is probably closest in philosophy to our approach; however, it calculates the ligand interactions toward an averaged receptor model and by means of a nondirectional force field. Moreover, it lacks H-bond flip-flop particles and the possibility to simulate solvation effects.

Although the validity of both atomistic and surface models has been demonstrated for a wealth of systems, two properties of biological receptors have so far not been directly simulated with such techniques: *receptor adaptation* and *H-bond flip-flop*. As both approaches determine ligand−receptor interaction energies toward an averaged receptor, subtle effects associated with the adaptation of the receptor shape to the individual ligand molecules able to bind to it remain unaddressed. At the true

[†] Laboratory for Organic Chemistry, ETH Zürich, 8092 Zürich, Switzerland.

[‡] Present address: Discovery Technologies AG, Gewerbestrasse 16, 4123 Allschwil, Switzerland.

(1) Sali, A.; Overington, J. P.; Johnson, M. S.; Blundell, T. L. Trends *Biochem. Sci.* **1990**, *15*, 235−240.

(2) Hong, J.-L.; Namgoong, S. K.; Bernardi, A.; Still, W. C. *J. Am. Chem. Soc.* **1991**, *113*, 5111−5112.

(3) Höltje, H. D.; Anzali, S. *Pharmazie* **1993**, *47*, 691−698.

(4) Snyder, J. P.; Rao, S. N.; Koehler, K. F.; Vedani, A. In *3D QSAR in Drug Design*; Kubinyi, H., Ed.; ESCOM Science Publishers: Leiden, 1993; pp 336−354.

(5) Vedani, A.; Zbinden, P.; Snyder, J. P.; Greenidge, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4987−4994.

(6) Zbinden, P.; Dobler, M.; Folkers, G.; Vedani, A. *QSAR*, in press.

(7) Snyder, J. P.; Rao, S. N.; Koehler, K. F.; Pellicciari, R. In *Trends in Receptor Research*; Angeli, P., Gulini, U., Quaglia, W., Eds.; Elsevier Science Publishers: Amsterdam, 1992; pp 367−403.

(8) Hahn, M. *J. Med. Chem.* **1995**, *38*, 2080−2090.

(9) Srivastava, S.; Richardson, W. W.; Bradley, M. P.; Crippen, G. In *3D QSAR in drug design: Theory, Methods and Applications*; Kubinyi, H., Ed.; Escom: Leiden, 1993; pp 80−116.

(10) Walters, D. E.; Hinds, R. M. *J. Med. Chem.* **1994**, *37*, 2527−2536.

(11) Hahn, M.; Rogers, D. *J. Med. Chem.* **1995**, *38*, 2091−2102.

(12) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.

(13) Cramer, R. D.; DePriest, S. A.; Patterson, D. E. In *3D QSAR in drug design: Theory, Methods and Applications*; Kubinyi, H., Ed.; Escom: Leiden, 1993; pp 443−485.

S0002-7863(97)03976-0 CCC: $15.00
Published on Web 04/28/1998

biological receptor, amino-acid residues bearing a conformationally flexible H-bonding function (i.e., Ser, Thr, Tyr, Cys, His, Asn, and Gln) can engage in differently directed H-bonds with different ligand molecules—an effect that can also not be simulated with an averaged receptor, simultaneously binding a series of ligand molecules in a virtual experiment. For example, ligand-dependent H-bond flip-flop has been observed in small-molecule inhibitor complexes of the enzyme purine nucleoside phosphorylase.[14]

In a previous paper, we have described a method to generate a peptidic pseudoreceptor—a miniprotein—about any molecular framework of interest, e.g., a pharmacophore. The concept was validated by constructing receptor surrogates for the enzyme human carbonic anhydrase, the dopaminergic receptor, and the β2-adrenergic receptor.[5] More recently, we developed an algorithm to allow for *pharmacophore equilibration*, an iterative positional, orientational, and conformational relaxation of the ligand molecules during receptor optimization. To avoid problems resulting from *functional-group obscuring* (e.g., the occurrence of a H-bond donor and a H-bond acceptor in close vicinity within the pharmacophore), we have devised a technique referred to as *receptor-mediated ligand alignment*.[6] Problems associated with the adaptation of a receptor to an individual ligand topology—albeit small in size—and H-bond flip-flop may still jeopardize otherwise reasonable pseudoreceptors as long as they represent an averaged receptor entity. The splitting of a pharmacophore−receptor system into n (1:1) complexes, followed by individual mutual optimization may be considered as a functional work-around. However, an estimation of the energy associated with receptor adaptation is hardly possible with a binding-site model, due to the lack of a residual entity, defining the amount of tolerable displacement as well as accounting for a substantial fraction of the total energy of the system.

Methods

The quasi-atomistic-modeling concept (software Quasar) developed at our laboratory allows the construction of a receptor surface model—a three-dimensional envelope, populated with atomistic properties at uniformly distributed discrete positions—about any molecular framework of interest, e.g., a pharmacophore. In contrast to other 3-D QSAR techniques, our approach includes receptor envelopes, individually adapted to the topology of the various ligand molecules as well as *H-bond flip-flop particles* (mimicking Ser, Thr, Tyr, Cys, His, Asn, and Gln residues at the true biological receptor) engaging in differently directed hydrogen bonds with different ligand molecules, simultaneously binding to the receptor.[15] In Quasar, a family of receptor-surface models is generated by means of a genetic algorithm combined with cross-validation. The construction of such a family of receptor models includes the following steps.

1. Construction of Individually Adapted Receptor Envelopes. First, the training set of ligand molecules is surrounded by virtual particles (e.g., radius $r^\circ = 0.8$ Å, well-depth $\epsilon^\circ = -0.024$ kcal/mol, no electrical charge) defining a van der Waals surface. Next, this envelope is optimized by means of energy minimization. Optionally, the ligand conformational space within this primordial envelope may be scanned with a Monte Carlo search protocol. We refer to this entity as the "averaged receptor envelope". Then, each ligand molecule of both the training and test sets is optimized as a 1:1 ligand−receptor complex, starting from the averaged receptor envelope, allowing for individual adaptation of the receptor envelope to the very ligand topology. Each "lattice point" of the emerging individual receptor envelope is coupled to the corresponding point of the averaged receptor envelope by means of a weak positional constraint, typically 0.1−0.5 kcal/(mol·Å^2). This ensures a minimal deformation but, simultaneously, allows any van der Waals repulsions between ligands and the envelope, possibly caused by the averaged receptor envelope, to be overcome. During model evaluation (cf. below), the associated energy is fully considered. Typical RMS deviations for corresponding points on the envelope lie in the range of 0.05−0.2 Å, with associated deformation energies ranging from 0.25 to 3.0 kcal/mol.

Table 1. Properties of Receptor-Surface Particles Used in Quasar

particle	nb potential type[a]	electronic charge	well depth of nb function (kcal/mol)
salt bridge, positive	10/12 + elec	+0.25	−4.95/−4.07/−2.33[b]
salt bridge, negative	10/12 + elec	−0.25	−4.95/−4.07/−2.33[b]
H-bond donor	10/12		−4.95/−4.07/−2.33[b]
H-bond acceptor	10/12		−4.95/−4.07/−2.33[b]
hydrophobic, positive	6/12 + elec	+0.1	−0.09[c]
hydrophobic, negative	6/12 + elec	−0.1	−0.09[c]
H-bond flip-flop[d]	10/12		−4.95/−4.07/−2.33[b]
surface solvent	symmetric 10/12[e]		−0.97/−0.80/−0.46[b,f]

[a] The values i and j refer to the attractive and repulsive coefficients of the nonbonded potential function used for the ligand−receptor interaction. The general form of this potential is $E(r) = A/r^i - C/r^j$. [b] Values for −O−H···Y, >N−H···Y, and −S−H···Y H-bond interactions, respectively, where "Y" denotes a virtual H-bond acceptor. Identical values are used for the X···O, X···N, and X···S arrangement where "X" denotes a virtual H-bond donor. [c] This function adopts the form $E(r) = A/r^{12} - C/r^6$. The coefficients A and C are calculated according to $A = -\epsilon(r_i + r_j)^{12}$ and $C = -2\epsilon(r_i + r_j)^6$, respectively, and with $\epsilon = (\epsilon_i\epsilon_j)^{1/2}$. The given figure represents ϵ_i; r_i and r_j correspond to the van der Waals radii of the two involved atoms. [d] H-bond flip-flop particles can adapt their property (H-bond donor or acceptor) to each ligand molecule within the pharmacophore, depending on its interacting functional group. [e] To avoid repulsive forces between surface solvent and any ligand molecule, a symmetric 10/12 potential (mirrored at $r = r^\circ$) is used. This represents a possible approximation to a mobile solvent. [f] As the virtual particles may be different in radius from a water molecule, the associated energy must be corrected for different volumes: $E = (2r_{vp}/2.75)^3E^\circ$; e.g., for $r_{vp} = 0.8$ Å, $E = 0.197E^\circ$.

2. Generation of an Initial Family of Parent Structures. Points on the receptor surface are randomly populated with atomistic properties (cf. Table 1), optionally observing a minimal distance between two points occupied by H-bonding particles.[16] Potential H-bonding sites may be restricted to points on the receptor envelope which are located within a reasonable distance and at a favorable orientation with respect to any H-bond donor or acceptor moiety of the ligand molecules defining the training set. Likewise, positions suited to host a H-bond flip-flop particle are defined at spatial regions where donor and acceptor moieties cluster appropriately. To identify the position on the receptor surface yielding optimal interactions with ligand functional groups, we make use of a vector concept, based on the directionality of hydrogen bonds (cf. ref 5 and references therein). For all systems discussed in this paper, we used an initial population of 200 different models. If there is experimental or other evidence for a solvent-accessible receptor cavity, parts of the receptor envelope may be assigned as representing solvent (cf. the carbonic anhydrase simulation, below). Alternatively, regions may be defined as being purely hydrophobic or nonexistent (void). Such assignments remain unaltered throughout the entire simulation.

3. Evolution of a Model Family. Using a genetic algorithm (for a detailed description, see, for example, ref 17) the initial family of receptor models is evolved using both crossover and mutation events. When two parents are selected, those with an already better fit are more likely to be selected for a crossover event than "weaker individuals".

(14) Montgomery, J. A.; Secrist, J. A., III. *Perspect. Drug Discovery Des.* **1994**, *2*, 205−220.

(15) Vedani A.; Zbinden P. *Pharm. Acta. Helv.*, in press.

(16) Such a constraint might be meaninful when using the receptor surrogate for drug-design purposes as the true biological receptor H-bond donors and H-bond acceptors are never observed at a distance closer than 2.4 Å.

(17) Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854−866.

Table 2. Summary of Receptor Models as Generated by Quasar[a]

biological system	number of ligands in training and test set	number of parents/ crossovers	training set: correlation coefficients cvd q^2	training set: correlation coefficients r	training set: RMS deviation (kcal/mol) [factor in k]	test set (predictions): RMS deviation (kcal/mol) [factor in k]	test set (predictions): maximal deviation (kcal/mol) [factor in k]
β2-adrenergic receptor	13/6	200/5000	0.847	0.946	0.358 [1.8]	0.829 [4.2]	1.15 [7.2]
aryl hydrocarbon receptor	12/6	200/5000	0.892	0.961	0.353 [1.8]	0.686 [3.2]	1.13 [7.0]
carbonic anhydrase	8/5	200/5000	0.815	0.944	0.380 [1.9]	0.423 [2.1]	0.645 [3.0]
cannabinoid receptor	18/10	200/5000	0.622	0.829	0.693 [3.3]	0.843 [4.3]	1.88 [25]
neurokinin-1 receptor	21/10	200/5000	0.819	0.923	0.483 [2.3]	0.651 [3.1]	1.28 [9.0]
sweet-taste receptor	17/8	200/5000	0.749	0.905	0.287 [1.6]	0.733 [3.5]	−1.44 [12]

[a] All data reflect quantities averaged over the 200 models.

At each crossover step, there is a small probability (typically 0.01−0.1) of a transcription error which is expressed by means of a random mutation. Only those children are retained which differ by a minimal amount (typically 5−10% of all populated points) from any parent. Thereafter, those two individuals of the population with the highest *lack-of-fit* (RMS of $\Delta G°_{pred}$ vs $\Delta G°_{exp}$ obtained from a leave-one-out cross-validation, combined with the weighted sum of populated points on the receptor surface; cf., for example, ref 16) are discarded. This process is repeated until a target q^2 value[13] (typically 0.9) or the limiting number of crossover steps (typically 5000) is reached.

4. Estimation of Free Energies of Ligand Binding. In our concept,[5,6,15] we have combined the approach of Blaney et al.[18] with a method of Still et al.[19] for estimating ligand solvation energies and a term to correct for the loss of entropy upon receptor binding following Searle and Williams:[20]

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

The term $\Delta E_{int,lig}$ accounts for the change—typically an increase—of the ligand internal energy while bound to the receptor surrogate from a strain-free reference conformation. This would seem necessary as the internal energy of a ligand molecule may increase while maximizing its interaction with the receptor. Blaney's approximation is based on the assumption that all ligands are equally buried within the receptor and, hence, differences in the solvation energy of the ligand−receptor complexes become negligible.[18] For systems where the ligands expose a different fraction of their surface from a solvent-accessible binding site, it is possible to define a solvent-accessible part on the receptor envelope and, thus, correct for such a situation (cf. below). $\Delta E_{env\ adapt,lig}$ is associated with the energy uptake upon modifying the mean receptor envelope to an entity, individually adapted to each ligand molecule. To determine the ligand−receptor interaction energy, $E_{lig-rec}$, we make use of a directional force field.[5,6,15,21,22] Free energies of ligand binding, $\Delta G°_{pred}$, are then obtained by means of a linear regression between $\Delta G°_{exp}$ and E_{bdg} using the ligand molecules of the training set:

$$\Delta G°_{pred} = |a|E_{bdg} + b \quad (2)$$

The slope and intercept of (2) are inherent to a given receptor model and are also applied to predict the 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[5,6,15] rather than apply a universal function for the various protein−ligand systems.

5. Analysis of the Model Family. The most powerful criterion to validate a family of receptor models is their ability to predict free energies of ligand binding for an external set of test ligand molecules not used during model construction. Other criteria include the cross-validated q^2 value,[13] the *lack-of-fit* for the ligands of the training set, and the uniformness of the distribution of the properties mapped onto the receptor envelope, e.g., larger hydrophobic pockets or solvent-accessible regions.

To select a training set from the available biological data that spans parameter space homogeneously, we have adapted a method developed by Marengo and Todeschini.[23] Their algorithm was originally developed for applications to distance-based experimental design with the aim to select a fraction of the most different compounds from a given set of molecules by means of the maximal "minimum distance". In our approach, the minimum distance between two molecules is computed as a weighted function of electrostatic and van der Waals interactions—determined at points of a common surface.[6] This allows for an unbiased selection of the most dissimilar molecules from an ensemble of ligands to be used as the training set during pseudoreceptor construction.

Results

We have applied our concept to derive semiquantitative structure−activity relationships for a series of biological systems, including the β2-adrenergic, aryl hydrocarbon, cannabinoid, neurokinin-1, and sweet-taste receptor. In addition, we have derived a model for the enzyme carbonic anhydrase, as a three-dimensional structure is available for this globular protein and its active site is known to be solvent accessible.[24] The summary of the results is given in Table 2. As examples, we shall discuss the β2-adrenergic receptor and the enzyme human carbonic anhydrase in detail. The three-dimensional coordinates of all receptor models described in this paper—except for the neurokinin-1 ligand set which is proprietary—are available for distribution (biograf@dial.eunct.ch).

The β2-adrenergic receptor is a member of the class of G protein-coupled receptors. The rational design of both potent and selective β2-adrenergic agonists—of particular interest for the clinical treatment of asthma[25]—would be facilitated by the availability of the three-dimensional structure of the receptor-binding pocket. In the past decade, efforts have been undertaken in order to derive three-dimensional models for integrated membrane proteins (see, for example, ref 26). Much attention

(18) Blaney, J. M.; Weiner, P. K.; Dearing, A.; Kollman, P. A.; Jorgensen, E. C.; Oatley, S. J.; Burridge, J. M.; Blake. J. F. *J. Am. Chem. Soc.* **1982**, *104*, 6424−6434.

(19) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.

(20) Searle, M. S.; Williams, D. H. *J. Am. Chem. Soc.* **1992**, *114*, 10690−10697.

(21) Vedani, A.; Huhta, D. W. *J. Am. Chem. Soc.* **1990**, *112*, 4759−4767.

(22) 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 nonelectrostatic 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 (compulsorily 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 5 and 21.

(23) Marengo, E.; Todeschini, R. *Chemom. Intell. Lab. Syst.* **1992**, *16*, 37−44.

(24) Kannan, K. K.; Ramanadham, M.; Jones, T. A. In *Biology of the carbonic anhydrases*; Tashian, R. E., Hewett-Emmett, D., Eds.; *Ann. N. Y. Acad. Sci.* **1984**, *429*, 49−60.

(25) Main, B. G. In *Comprehensive Medicinal Chemistry*; Emmett, J. C., Ed.; Pergamon Press: Oxford, 1993; pp 187−228.

(26) Kontoyianni, M.; Lybrand, T. P. *Perspect. Drug Discovery Des.* **1993**, *1*, 291−300.

has been concentrated upon the building of β2-adrenergic receptor models in which the X-ray structure of bacteriorhodopsin[27] was used either directly or indirectly, but some controversy still exists over the validity of such homology models.[26,28]

In an earlier study, we derived an atomistic binding-site model for the β2-adrenergic receptor by means of pseudoreceptor modeling.[5] Using nine adrenergic agonists as a training set, we have constructed a peptidic surrogate consisting of 15 amino acid residues. The model was capable of reproducing relative free energies of ligand binding for an external set of six test ligands within an RMS value of 0.6 kcal/mol of the experimental value, corresponding to an uncertainty factor of 2.7 in the binding affinity. This surrogate, however, was constructed about a "static" pharmacophore model. More recent techniques use pharmacophore equilibration and scanning of receptor space using appropriate Monte Carlo search protocols.[6]

To select a structurally and topologically most diverse training set from the 19 β2-adrenergic antagonists on the basis of the minimum-distance approach (cf. above), we have used the electrostatic and van der Waals field within a primordial receptor envelope. The atomic partial charge model (MNDO/ESP: partial charges fitted to the electrostatic potential) for the ligand molecules was derived using MOPAC 6.0;[29] free energies of ligand solvation were calculated using a semianalytical approach following Still and co-workers.[19] Experimental free energies of ligand binding, $\Delta G°_{exp}$, were taken from ref 30. The alignment of the ligand molecules is described in ref 5.

A primordial envelope was constructed about 13 ligand molecules, defining the training set: FEN, TER, AH3, CLB, 2CL, SAL, SKF, TBF2, SYN, AH2, NOR, NIS, and ISOP (for the systematic names, cf. Table 3). The envelope comprised 201 virtual particles with a radius of 0.8 Å and was subsequently energy minimized, yielding the averaged envelope, representing the inner lining of a mean receptor cavity. Next, the ligand molecules of the training set were minimized within this envelope, identifying a final position, orientation, and conformation. Finally, the ligand molecules of the test set (ISO NAB, ORC, ADR, DU3, and DU2) were similarly allowed to relax within the mean envelope using the same settings as for the training set. Starting from this mean envelope, the envelope of each ligand was then allowed to individually relax and to adjust to the topology of the very ligand molecules. As at the true biological receptor adaptation to the ligand topology is associated with a change in receptor energy, we couple the individual envelope *in status nascendi* by a weak positional constraint of 0.25 kcal/(mol·Å^2) to the mean receptor envelope. This led to RMS shifts ranging from 0.105 to 0.135 Å, corresponding to energies of 0.557−0.914 kcal/mol. The largest envelope deformation for a ligand molecule of the test set was observed for ligand ISOP (the enantiomer of ligand ISO), with the aminoethanol C atom "*S*" configured, while the stereochemistry at the corresponding C atom of all other ligand molecules corresponds to an "*R*" configuration.

Using an initial population of 200 receptor models, the system (comprising the 13 ligand molecules defining the training set; cf. above) was allowed to evolve for 5000 crossover cycles. The transcription-error rate was set to 0.05, and the minimal difference between two receptor models was required to be larger than 0.05 (i.e., 10 out of 201 particles). For the best individual model, the simulation yielded a cross-validated q^2 of 0.852, and a classical r value for the linear regression of 0.947. A test set of six ligand molecules was then used to validate the model family. The RMS deviation of experimental and predicted free energies of ligand binding for the best model was 0.854 kcal/mol, corresponding to an uncertainty factor of 4.3 in the binding constant.

The largest individual deviation was observed for ligand DU3: its experimental and predicted free energies of ligand binding differ by 1.19 kcal/mol, corresponding to an uncertainty factor of 7.8 in the binding constant. When averaging was done over all 200 models, the RMS deviation of experimental and predicted free energies of ligand binding was calculated to be 0.829 kcal/mol, corresponding to an uncertainty factor of 4.2 in the binding constant. The largest individual deviation was observed for ligand ISO: its experimental and predicted free energies of ligand binding differ by 1.15 kcal/mol, corresponding to an uncertainty factor of 7.2 in the binding constant.

Table 3. Energies Associated with the Receptor Model for the β2-Adrenergic Receptor As Generated by Quasar[a]

Training Set (13 Molecules)

ligand[b]	ΔG_{exp}	$\Delta E_{env\ ada}$	ΔG_{pred} (**b**)	$\Delta\Delta G_{p-e}$ (**b**)	ΔG_{pred} (**a**)	$\Delta\Delta G_{p-e}$ (**a**)
FEN	−11.66	0.834	−11.58	0.08	−11.69	−0.03
TER	−11.66	0.675	−11.07	0.60	−11.05	0.61
CLB	−11.35	0.659	−11.27	0.08	−11.27	0.08
AH3	−11.21	0.557	−11.26	−0.05	−11.27	−0.06
2CL	−10.78	0.741	−10.20	0.58	−10.26	0.52
SAL	−10.70	0.593	−10.80	−0.10	−10.73	−0.03
SKF	−10.50	0.678	−11.07	−0.57	−10.97	−0.47
TBF2	−10.34	0.717	−10.31	0.03	−10.39	−0.05
SYN	−9.76	0.699	−10.12	−0.36	−10.01	−0.25
AH2	−9.60	0.692	−9.69	−0.09	−9.61	−0.01
NOR	−8.64	0.656	−8.17	0.47	−8.14	0.50
NIS	−8.57	0.884	−9.17	−0.60	−9.23	−0.66
ISOP	−8.16	0.914	−8.24	−0.08	−8.33	−0.17

Test Set (6 Molecules)

ligand[c]	ΔG_{exp}	$\Delta E_{env\ ada}$	ΔG_{pred} (**b**)	$\Delta\Delta G_{p-e}$ (**b**)	ΔG_{pred} (**a**)	$\Delta\Delta G_{p-e}$ (**a**)
ISO	−11.60	0.730	−10.51	1.09	−10.45	1.15
NAB	−11.06	0.746	−10.68	0.38	−10.61	0.45
ORC	−10.70	0.802	−9.87	0.83	− 9.87	0.83
ADR	−10.11	0.813	−9.65	0.46	−9.62	0.49
DU3	−9.69	0.682	−10.88	−1.19	−10.79	−1.08
DU2	−9.13	0.749	−9.97	−0.84	−9.82	−0.69

[a] All energies are given in kilocalories per mole. Columns marked with (**a**): average over 200 models. Columns marked with **(b)** best receptor model. [b] Training set: FEN, 1-(3,4-dihydroxyphenyl)-2-(isopropyl-*p*-hydroxyphenyl)aminoethanol (*Fenoterol*); TER, 1-(3,4-dihydroxyphenyl)-2-*tert*-butylaminoethanol; AH3, 1-(3-amido-4-hydroxyphenyl)-2-*tert*-butylaminoethanol; CLB, 1-(4-amino-3,5-dichlorophenyl)-2-*tert*-butylaminoethanol (*Clenbutarol*); 2CL, 1-(2-chlorophenyl)-2-*tert*-butylaminoethanol; SAL, 1-(4-hydroxy-3-hydroxymethylphenyl)-2-*tert*-butylaminoethanol (*Salbutamol*); SKF, 1-(4-hydroxy-3-aminomethyl-phenyl)-2-*tert*-butylaminoethanol; TBF2, 1-(3,5-dihydroxyphenyl)-2-*tert*-butylaminoethanol (*Terbutaline*); SYN, 1-(4-hydroxyphenyl)-2-isopropylaminoethanol (*N-isopropylnorsynephrine*); AH2, 1-(4-hydroxy-3-hydroxymethylphenyl)-2-isopropylaminoethanol; NOR, 1-(3,4-dihydroxyphenyl)-2-aminoethanol (*Norepinephrine*); NIS, 1-(3-hydroxyphenyl)-2-isopropylaminoethanol (*N-isopropylnorphenylephrine*); ISOP, (+)-1-(3,4-dihydroxyphenyl)-2-isopropylaminoethanol (*Isoproterenol*). [c] Test set: ISO, (−)1-(3,4-dihydroxyphenyl)-2-isopropylaminoethanol (*Isoproterenol*); NAB, 1-(4-amino-3,5-dichlorophenyl)-2-isopropylaminoethanol; ORC, 1-(3,5-dihydroxyphenyl)-2-isopropylaminoethanol (*Orciprenaline*); ADR, 1-(3,4-dihydroxyphenyl)-2-methylaminoethanol (*Epinephrine*); DU3, 1-(4-hydroxy-3-aminophenyl)-2-*tert*-butylaminoethanol; DU2, 1-(4-hydroxy-3-aminopheny)-2-isopropylaminoethanol.

(27) Henderson, R.; Baldwin, J. M.; Ceska, T. A.; Zemlin, F.; Beckmann, E.; Downing, K. M. *J. Mol. Biol.* **1990**, *213*, 899−923.
(28) Hoflack, J.; Trumpp-Kallmeyer, S. *TIPS* **1994**, *15*, 7−9.
(29) Stewart, J. J. P. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1−105. Distributed by QCPE, University of Indiana, Bloomington, IN (Program 455).
(30) Donné-Op den Kelder, G. M.; Bultsma, T.; Timmerman, H. *J. Med. Chem.* **1988**, *31*, 1069−1079.

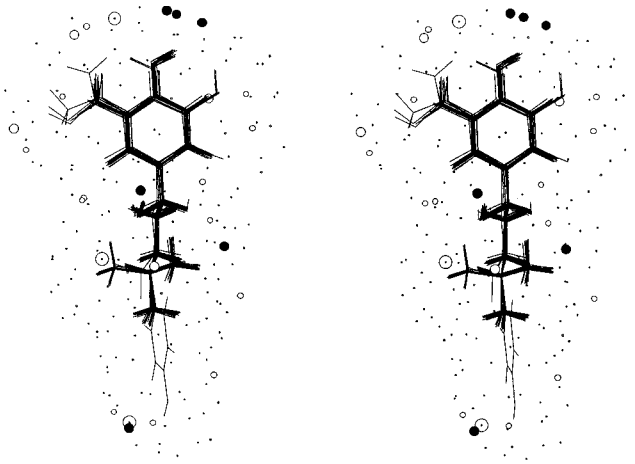

Figure 1. Stereoscopic view of the best surface model of the β2-adrenergic receptor. Positively charged salt bridges are shown as large open circles, negatively charged salt bridges as large filled circles. H-bond donors are shown as medium-sized open circles, H-bond acceptors as medium-sized filled circles. H-bond flip-flop particles are shown as large open circles with a central dot. Charged hydrophobic particles are represented as small open circles, uncharged hydrophobic particles as dots.

Throughout the 200 models, the variation of $\Delta G°_{pred}$ of the ligands of the training set ranges from 0.290 to 0.674 kcal/mol (corresponding to an uncertainty factor of 1.6−3.4 in the binding constant), within the test set from 0.325 to 0.723 kcal/mol (1.7−3.5 in k). Thus, for the β2-adrenergic receptor, $\Delta G°_{pred}$ values averaged over all 200 models do not differ significantly from the values predicted by the best individual model. Details are given in Table 3; the best individual model is shown in Figure 1.

The distribution of "quasi-atomistic particles" on the receptor surface can be best described as a hydrophobic barrel (lining the pharmacophore) with 14 specific H-bonding sites interacting with the ammonium N atom, the ethanolamine O atom, and the various substituents (hydroxyl, amine) of the aromatic ring. In addition, two sites engage in a hydrogen bond with the phenolic ring present in ligand FEN. Of particular interest are the position and function of the three H-bond flip-flop particles present. One flip-flop particle, observed in all 200 models, bridges the ammonium (a pure donor) and ethanolamine (acting as acceptor) functionalities common to all ligand molecules. This particle simulates a bifunctional group as found, for example, in Ser, Thr, or Tyr residues at the true biological receptor. However, such a functionality can also be simulated by an averaged receptor model. The second flip-flop particle, observed in 199 out of 200 models, is located near the para-position of the aromatic ring. At this position, it functions as a true flip-flop particle, as ligands CLB and NAB have an amine substitution at the ring 4-position while most others (except ORC, TBF2, NIS, and 2CL) have a hydroxyl substituent. The orientation of this hydroxyl group—within our pharmacophore hypothesis—is defined by the ligand molecules including a 3,4-dihydroxy substitution, forming an intramolecular R−O−H···O hydrogen bond (e.g., ligand TER). The H-bond flip-flop particle in our quasi-atomistic model is able to simultaneously combine both donor and acceptor properties, i.e., to act as a H-bond acceptor toward the ligand molecules NAB and CLB, while acting as a donor toward most others (e.g., ISO; cf. Figure 2). The third flip-flop particle, only observed in 13 out of 200 models, simply acts as a H-bond acceptor toward the phenolic hydroxyl group of ligand FEN.

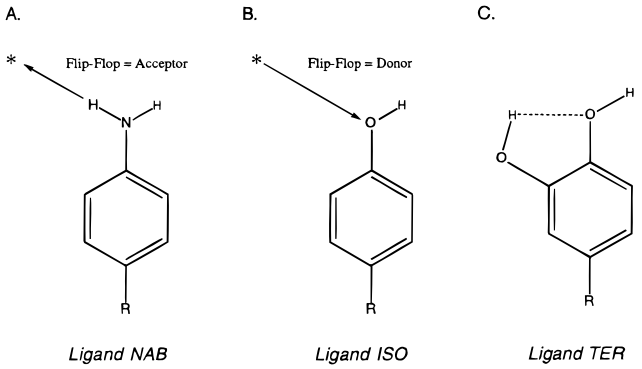

Figure 2. (A) H-bond flip-flop particle acting as a H-bond acceptor (toward ligand NAB). (B) H-bond flip-flop particle acting as a H-bond donor (toward ligand ISO). (C) Determination of the orientation of the 4-hydroxyl group within the pharmacophore by ligand TER.

We have also performed a simulation without including particles capable of H-bonding, i.e., only allowing for neutral and charged hydrophobic particles, to analyze the impact of directional interactions. The corresponding simulation, however, did not exceed a cross-validated q^2 of 0.456 (largest individual deviation of a ligand molecule of the test set 1.60 kcal/mol), demonstrating that, at least for the selected β2-adrenergic agonists, directional hydrogen bonds seem to be mandatory to reproduce the experimental binding affinities.

Carbonic anhydrase, a zinc-containing enzyme, is an extremely efficient catalyst of the reversible hydration of carbon dioxide. The crystal structure of the native enzyme has been determined to a resolution of 2.0 Å by Kannan and co-workers.[24] The structure of the complex between carbonic anhydrase and the sulfonamide inhibitor 2-acetamido-1,3,4-thiadiazole-5-sulfonamide was determined to a resolution of 3.0 Å.[31] Clinical applications of the inhibition of carbonic anhydrase focus on the treatment of glaucoma, epilepsy, and acute mountain sickness.[32]

The construction of a receptor surrogate was based on a total of 13 sulfonamide inhibitors, 8 defining the training set and 5 representing the test set. The protonation state and stereochemistry of the sulfonamide group were adapted from refs 33 and 34. The alignment of the ligands is described in ref 5.

The atomic partial charge model (CM-1 charges) for the ligand molecules and the free energies of ligand solvation were obtained using the AMSOL 5.4 software package.[35] Experimental free energies of ligand binding, $\Delta G°_{exp}$, were derived from thermodynamic and kinetic data.[36−38] In contrast to the other systems used in this study, an experimental structure is available for both native and complexed human carbonic anhydrase.[24,31] A quasi-atomistic model for this enzyme was used to investigate the benefit of particles simulating a solvent-accessible binding pocket. The active site in human carbonic

(31) Kannan, K. K.; Vaara, I.; Notstrand, B.; Lowgren, S.; Borell, A.; Fridborg, K.; Petef, M. In *Drug Action at the Molecular Level;* Roberts, C. G. K., Ed.; McMillan: London, 1977; pp 73−91.
(32) *Biology and Chemistry of the Carbonic Anhydrases*; Tashian, R. E., Hewett-Emmett, D., Eds. *Ann. N. Y. Acad. Sci.* **1984**, *429*.
(33) Mukherjee, J.; Rogers, J. I.; Khalifah, R. G.; Everett, G. W., Jr. *J. Am. Chem. Soc.* **1987**, *109*, 7232−7233.
(34) Everett, G. W., Jr. Personal communication on results from ^{15}N NMR studies, 1989.
(35) Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 629−666. Distributed by QCPE, University of Indiana, Bloomington, IN (Program 606).
(36) Taylor, P. W.; King, R. W.; Burgen, A. S. V. *Biochemistry* **1970**, *9*, 2368−2376.
(37) Kakeya, N.; Aoki, A.; Kamada, A.; Yata, N. *Chem. Pharm. Bull.* **1969**, *17*, 1010−1014.

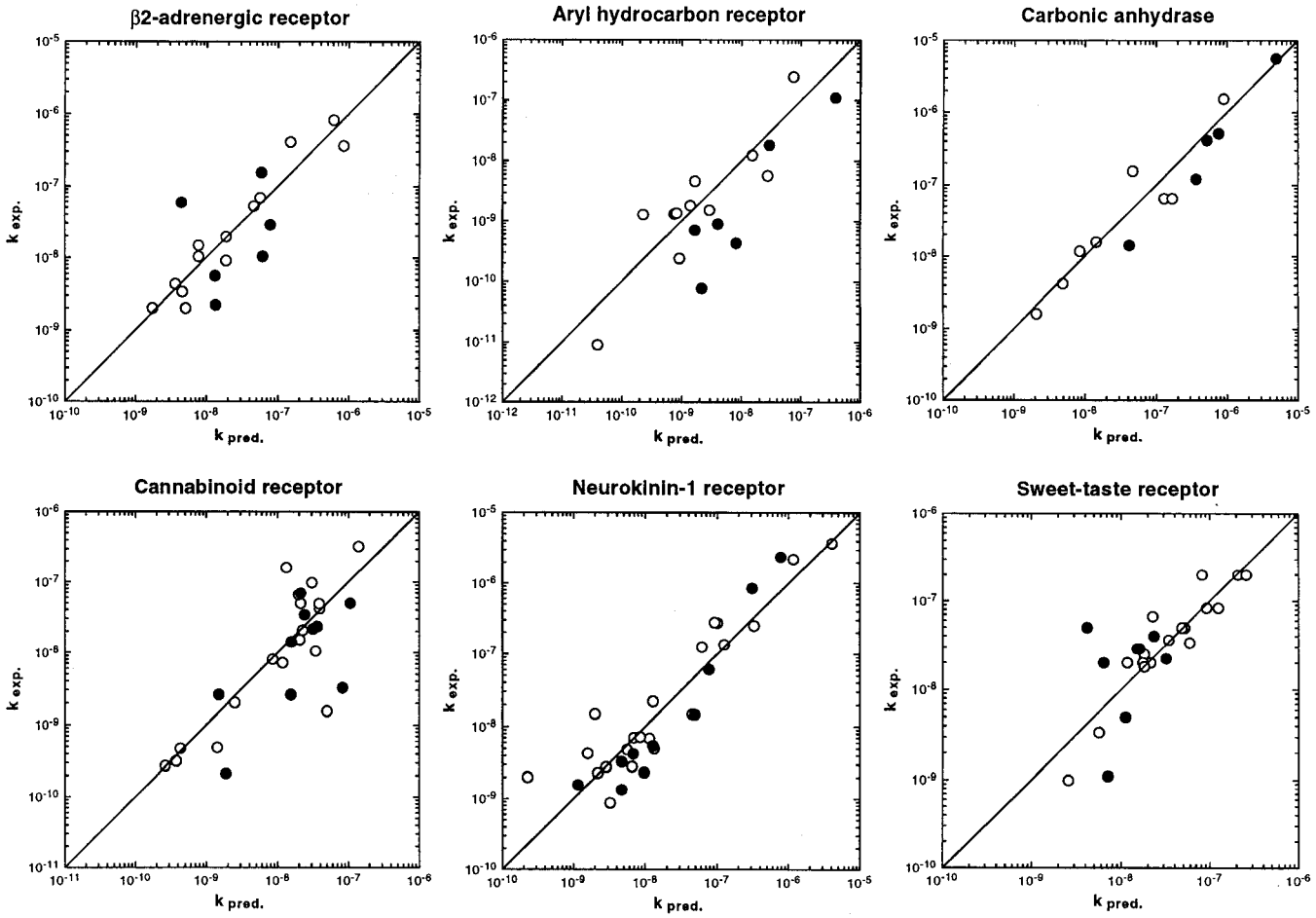

Figure 3. Experimental vs predicted dissociation constants for the β2-adrenergic, aryl hydrocarbon, cannabionoid, neurokinin-1, and sweet-taste receptor as well as for the enzyme carbonic anhydrase. Ligand molecules of the training set are represented as open circles, ligand molecules of the test as filled circles.

anhydrase is conical in shape with the catalytic zinc ion located at its apex.[24] The entrance of the active site—located 12 Å from the catalytic zinc and approximately 14 Å in diameter—is lined by the residues His 64, Asn 69, Gln 92, Ile 131, His 200, Pro 202, and Tyr 204. The position, orientation, and conformation of 2-acetamido-1,3,4-thiadiazole-5-sulfonamide (acetazolamide, a potent inhibitor of the enzyme) has been determined by means of X-ray diffraction.[31]

To analyze the impact of a solvent-accessible binding pocket on the predictive power of the quasi-atomistic models, we performed two simulations: one containing no information about a solvent-accessible portion of the binding pocket and another where 40% of the surface (at a location relative to the ligand molecules corresponding to the topology at the true enzyme) was defined to explicitly represent solvent. This region was not altered throughout the evolution while all other positions could change in character during crossover and mutation events. The results are compared in Table 4.

The simulation including a solvent-accessible surface yielded a significantly better prediction of the ligands defining the test set: an RMS deviation of 0.423 kcal/mol (corresponding to an uncertainty of a factor of 2.1 in the binding affinity) and the largest deviation for an individual ligand of 0.645 kcal/mol (3.0 in k) compare to RMS and maximal deviations of 0.639 kcal/mol (2.5 in k) and 1.15 kcal/mol (7.2 in k) for the simulation without explicit solvent.

The surrogates for the cannabinoid, neurokinin-1, and sweet-taste receptor systems were generated using settings identical with those for the β2-adrenergic receptor (cf. above). The three-dimensional coordinates of the cannabinoid pharmacophore were kindly provided by Dr. Paulette A. Greenidge (cf. ref 39). All ligand molecules were reoptimized in aqueous solution using the AMBER force field[40] as implemented in MacroModel 5.0.[41] The atomic partial charge model (MNDO-ESP) for the ligand molecules was derived using MOPAC 6.0;[29] free energies of ligand solvation were calculated using a semianalytical approach following Still and co-workers.[19] Experimental free energies of ligand binding, $\Delta G°_{exp}$, were taken from ref 42. Coordinates, alignment, and relative sweetnesses of the sweet-taste ligand set were kindly provided by Professor Lucio Merlini (Department of Organic Chemistry/DISMA, University of Milan, Italy).[43] The atomic partial charge model (MNDO-ESP) for the ligand molecules was derived using MOPAC 6.0;[29] free energies of ligand solvation were calculated using a semiana-

(38) Sprague, J. M. In *Topics in Medicinal Chemistry*, Rabinowitz, J. L., Myerson, R. M., Eds.; Interscience: New York, 1968; pp 1−63.

(39) Schmetzer, S.; Greenidge, P. A.; Kovar, K. A.; Folkers, G. *J. Comput.-Aided Mol. Des.*, in press.

(40) Weiner, S. J.; Kollmann, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.

(41) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440−467.

(42) Compton, D. R.; Rice, K. C.; DeCosta, R. K.; Razdan, L. S.; Melvin, L. S.; Johnson, M. R.; Martin, B. R. *J. Pharmacol. Exp. Theor.* **1993**, *265*, 218−226.

(43) Arnoldi, A.; Bassoli, B.; Merlini, L.; Ragg, E. *J. Chem. Soc., Perkin Trans.* **1993**, *1*, 1359−1366.

Table 4. Energies Associated with the Receptor-Surface Model for the Enzyme Carbonic Anhydrase As Generated by Quasar[a]

Training Set (8 Molecules)

ligand[b]	ΔG_{exp}	$\Delta E_{env\ ada}$	ΔG_{pred} (**n**)	$\Delta\Delta G_{p-e}$ (**n**)	ΔG_{pred} (**s**)	$\Delta\Delta G_{p-e}$ (**s**)
ETZA	−11.79	0.762	−11.80	−0.01	−11.64	0.15
NTS	−11.23	0.395	−11.17	0.06	−11.13	0.10
MTZ	−10.63	0.593	−10.68	−0.05	−10.81	−0.18
BAAA	−10.46	0.536	−10.43	0.03	−10.50	−0.04
NBSA	−9.64	0.508	−9.55	0.09	−9.23	0.41
DBSA	−9.64	0.413	−9.47	0.17	−9.08	0.56
SBSA	−9.13	0.431	−9.20	−0.07	−9.83	−0.70
BSA	−7.79	0.290	−8.04	−0.25	−8.10	−0.31

Test Set (5 Molecules)

ligand[b]	ΔG_{exp}	$\Delta E_{env\ ada}$	ΔG_{pred} (**n**)	$\Delta\Delta G_{p-e}$ (**n**)	ΔG_{pred} (**s**)	$\Delta\Delta G_{p-e}$ (**s**)
AAA	−10.52	0.474	−9.37	1.15	−9.89	0.63
LBSA	−9.27	0.359	−8.60	0.67	−8.63	0.64
YBSA	−8.56	0.418	−8.45	0.11	−8.42	0.14
MBSA	−8.43	0.374	−8.46	−0.03	−8.20	0.23
SAM	−7.04	0.380	−6.53	0.51	−7.11	−0.07

[a] All energies are given in kilocalories per mole. Columns marked with (**n**): normal simulation. Columns marked with (**s**): simulation including 40% solvent-accessible surface. All data represent quantities averaged over 200 models. [b] Training set: ETZA, 6-ethoxybenzothiazole-2-sulfonamide (*Ethoxzolamide*); NTS, 2-nitrothiophene-5-sulfonamide; MTZ, 2-acetimido-3-methyl-1,3,4-thiadiazole-5-sulfonamide (*Metazolamide*); BAAA, 2-butylamido-1,3,4-thiadiazole-5-sulfonamide; NBSA, 4-nitrobenzenesulfonamide; DBSA, 3,5-dichlorobenzenesulfonamide; SBSA, benzene-1,4-disulfonamide; BSA, benzenesulfonamide. [c] Test set: AAA, 2-acetamido-1,3,4-thiadiazole-5-sulfonamide (*Acetazolamide*); LBSA, 4-chlorobenzenesulfonamide; YBSA, 4-cyanobenzenesulfonamide; MBSA, 4-methylbenzenesulfonamide; SAM, 4-aminobenzenesulfonamide (*Sulfanilamide*).

lytical approach following Still and co-workers.[19] Coordinates and experimental free energies of ligand binding, $\Delta G°_{exp}$, of the neurokinin-1 antagonist molecules were kindly provided by Dr. Hans Briem (Boehringer Ingelheim KG, Ingelheim, Germany). The atomic partial charge model (CM-1 charges) and the free energies of ligand solvation were obtained using the AMSOL 5.4 software package.[35]

While the models for the neurokinin-1 and the sweet-taste receptor provide reasonable constructs (cross-validated q^2, 0.819 and 0.749; RMS deviations for the ligand molecules of the test set, 0.651 and 0.733 kcal/mol, corresponding to uncertainty factors of 3.1 and 3.5 in the binding affinity, respectively), the surrogate for the cannabinoid receptor (cross-validated q^2, 0.622; RMS deviation for the ligand molecules of the test set, 0.843 kcal/mol due to an outlier, predicted to bind 1.88 kcal/mol too weakly; factor of 25 in k) represents a less powerful model. A summary of the results on the quasi-atomistic models is given in Table 2; all plots of $\Delta G°_{pred}$ vs $\Delta G°_{exp}$ are shown in Figure 3.

Conclusions

Quasi-atomistic receptor modeling bridges 3-D QSAR and receptor modeling by populating receptor surface models with atomistic properties such as hydrogen bonds, salt bridges, hydrophobic regions, and solvent. The adaptation of a receptor to an individual ligand topology—albeit small in size—is simulated by *individually adjusted receptor envelopes*, coupled to the averaged envelope by means of a soft positional constraint. In addition, our approach includes *H-bond flip-flop particles* (mimicking conformationally mobile H-bond functionalities at the true biological receptor) as well as for solvent-accessible binding sites. A family of receptor models is evolved using a genetic algorithm, such as to reduce the influence of random errors and, simultaneously, scanning receptor space more exhaustively. The use of quasi-atomistic surrogates would seem to be advantagous when large numbers of ligand molecules are to be tested against a known or hypothetical receptor. The quasi-atomistic character of the approach yields still intuitive models for drug-design purposes.

Note Added in Proof. The most recent version of Quasar allows also for ligand−receptor polarization effects. Following the approach of Howard et al.[44] and using parameters of Applequist et al.,[45] this additional term permits a more subtle treatment of ligand molecules lacking in functional groups capable of hydrogen bonding.

Acknowledgment. We are indebted to Dr. Hans Briem, Dr. Horst Dollinger, and Dr. Herbert Köppen (Boehringer Ingelheim, Germany) for providing the neurokinin-1 ligand set as well as for most valuable discussions. We thank Dr. Paulette A. Greenidge for her contributions during the initial phase of the cannabinoid project. Financial support from the Foundation Biographics Laboratory 3R (Basel, Switzerland) and the Margaret and Francis Fleitmann Foundation (Lucerne, Switzerland) is gratefully acknowledged.

JA973976T

(44) Howard, A. E.; Chandra Singh, U.; Billeter, M.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 6984−6991.

(45) Applequist, J.; Carl, J. R.; Fung, K. K. *J. Am. Chem. Soc.* **1972**, *94*, 2952−2960.