5D-QSAR: The Key for Simulating Induced Fit?

Angelo Vedani* and Max Dobler

Biographics Laboratory 3R, Friedensgasse 35, CH-4056 Basel, Switzerland

Received July 30, 2001

In this journal we recently reported the development and the validation of a four-dimensional (4D)-QSAR (quantitative structure−activity relationships) concept, allowing for multiple conformation, orientation, and protonation state representation of ligand molecules. While this approach significantly reduces the bias with selecting a bioactive conformer, orientation, or protonation state, it still requires a "sophisticated guess" about manifestation and magnitude of the associated local induced fit—the adaptation of the receptor binding pocket to the individual ligand topology. We have therefore extended our concept (software Quasar) by an additional degree of freedom—the fifth dimension—allowing for a multiple representation of the topology of the quasi-atomistic receptor surrogate. While this entity may be generated using up to six different induced-fit protocols, we demonstrate that the simulated evolution converges to a single model and that 5D-QSAR—due to the fact that model selection may vary throughout the entire simulation—yields less biased results than 4D-QSAR where only a single induced-fit model can be evaluated at a time. Using two bioregulators (the neurokinin-1 receptor and the aryl hydrocarbon receptor), we compare the results obtained with 4D- and 5D-QSAR. The *NK-1* receptor system (represented by a total of 65 antagonist molecules) converges at a cross-validated r^2 of 0.870 and a predictive r^2 of 0.837; the corresponding values for the *Ah* receptor system (represented by a total of 131 dibenzodioxins, dibenzofurans, biphenyls, and polyaromatic hydrocarbons) are 0.838 and 0.832, respectively. The results indicate that the formal investment of additional computer time is well-returned both in quantitative and in qualitative values: less-biased boundary conditions, healthier (i.e., less inbred) model populations, and more accurate predictions of new compounds.

Introduction

Quantitative structure−activity relationships (QSAR) is an area of computational research that builds atomistic or virtual models to predict quantities such as the binding affinity, the toxicity, or the pharmacokinetic parameters of existing or hypothetical molecules. The idea behind QSAR is that structural features can be correlated with biological activity. Of particular interest in biomedical research are QSAR based on three-dimensional models (3D-QSAR)[1−5] because they allow for the simulation of directional forces: hydrogen bonds, metal−ligand contacts, polarization effects, and the interaction between electric dipoles—quantities known to play a key role for both molecular recognition and selective binding.[6−8]

While at the true biological receptor only one ligand molecule binds at the time, a QSAR study is typically based on a series of ligand molecules binding "simultaneously" to the atomistic or virtual receptor surrogate. In 3D-QSAR—where each ligand molecule is represented by a single, 3D entity—the identification of the bioactive conformation, orientation, and, possibly, the protonation state is a crucial step in the procedure. If the underlying pharmacophore hypothesis is based on incorrect assumptions, the resulting surrogate is hardly of any use for predictive purposes. While the alignment problem has long been recognized,[1−3,9] only the more recently developed 4D-QSAR technologies would seem to provide decent solutions.[9−15] In explicit 4D-QSAR approaches, the ligands of both training and test set are provided as an ensemble of conformations, orientations, and protonation states. The most likely bioactive representation is then genetically evolved from this reservoir using a Boltzmann-weighted selection criterion.[12−15] An adequate simulation of conformationally flexible H-bond donor or acceptor moieties at the true biological receptor, able to engage in differently directed H bonds with dissimilar ligand molecules (i.e., Ser, Thr, Tyr, Cys, His, Asn, and Gln residues), has become possible with the introduction of H-bond flip-flop particles in quasi-atomistic receptor models.[12,16] Inhibitor-dependent H-bond flip-flop has been observed, for example, in purine nucleoside phosphorylase.[17]

Unfortunately, even with 4D-QSAR, a major unknown persists, manifestation (and magnitude) of the induced fit—the ligand-induced adaptation of the binding site to the topology of the small molecule. Ligand-dependent induced fit has been experimentally detected, for example, in the family of the serine proteases.[18] In the context of quasi-atomistic binding site models,[12−15] it would seem necessary to emphasize that the simulation of a receptor-to-ligand adaptation is limited to rather small shifts (rms < 4.0 Å), which implies that more pronounced movements cannot presently be accounted for—particularly, if the large differences are a consequence of the ligand-induced triggering mechanism. Fortunately, the genetic algorithm typically fails under such boundary conditions, i.e., the evolution comes forth only very slowly, if at all.

* To whom correspondence should be addressed. Tel.: +41-61-261-4256. Fax: +41-61-261-4258. E-mail: admin@biograf.ch.

10.1021/jm011005p CCC: $22.00
Published on Web 04/18/2002

Table 1. Properties of Quasi-Atomistic Particles Used in Quasar

particle (property)	nonbonded potential type[a]	electric charge	well depth of nonbonded function (kcal/mol)
hydrophobic, neutral	6/12		−0.024[b]
hydrophobic, positive	6/12 + electrostatics	+0.10	−0.09[b]
hydrophobic, negative	6/12 + electrostatics	−0.10	−0.09[b]
hydrogen-bond donor	10/12		−5.0/−4.1/−2.3[c]
hydrogen-bond acceptor	10/12		−5.0/−4.1/−2.3[c]
salt bridge, positive	10/12 + electrostatics	+0.25	−5.0/−4.1/−2.3[c]
salt bridge, negative	10/12 + electrostatics	−0.25	−5.0/−4.1/−2.3[c]
H-bond flip-flop[d]	10/12		−5.0/−4.1/−2.3[c]
surface solvent	symmetric 10/12[e]		−0.97/−0.80/−0.46[c,f]
void (shallow pockets)			

[a] The values *i,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] This function adapts the form $E(r) = A/r^{12} - C/r$.[6] The coefficients A and C are calculated according to $A = -\epsilon\cdot(r_i + r_j)^{12}$ and $C = -2\cdot\epsilon\cdot(r_i + r_j)$,[6] respectively, and with $\epsilon = (\epsilon_i\cdot\epsilon_j)^{1/2}$. The given figure represents ϵ_j; r_i and r_j correspond to the van der Waals radii of the two involved atoms. [c] 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. [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 $r = r°$) is used. This represents a possible approximation to a mobile solvent. [f] As the virtual particles are different in radius than a water molecule, the associated energy must be corrected for different volumes: $E = (2\cdot r_{vp}/2.75)^3\cdot E_0$; e.g., for $r_{vp} = 0.8$ Å → $E = 0.197\cdot E_0$. The 2.75 Å correspond to a mean O−H···O H-bond distance.

In the Quasar concept,[12−15] local induced fit can be simulated by mapping a "mean envelope" (surrounding all ligands of the training set at van der Waals distance) on to a transiently generated "inner envelope", which snugly accommodates the individual ligand molecule. This may be achieved isotropically (linearly), anisotropically (steric, electric, H-bond, or lipophilicity-potential scaled), or through energy minimization.[12] As the magnitude of induced fit cannot be estimated in the absence of the true biological receptor, it has previously been necessary to perform several simulations based on different induced fit hypotheses.

We have, therefore, extended our concept[12−15] by a fifth dimension—a multiple representation of induced-fit hypotheses. These model hypotheses can be generated using different protocols for adapting the mean envelope to each ligand molecule used in the QSAR study. The current version of Quasar allows up to six different modes, each with an induced-fit degree varying from 0.0 to 1.0—0.0 implying no induced fit and 1.0 representing a perfect adaptation. The energetic cost for receptor-to-ligand adaptation is estimated from the mean → individual envelope shift. Using this new approach, we have then compared the results of 4D- and 5D-QSAR for the neurokinin-1 receptor and the aryl hydrocarbon receptor, respectively—two large systems previously studied with Quasar.[12−16]

Materials and Methods

A quasi-atomistic binding site surrogate refers to a high level of model abstraction. The essential information about the hypothetical receptor site is provided by means of a 3D envelope, which surrounds the ligand molecules at van der Waals distance and which is populated with properties mapped onto its surface. The topology of this surface mirrors the 3D shape of the binding site; the mapped properties represent other information of interest, such as hydrophobicity, partial charge, electrostatic potential, and hydrogen-bonding propensity. A variety of algorithms to generate and validate binding site models have been described.[4−5,9−16,19−22] While most approaches are based on a 3D-QSAR concept, more recent algorithms allow for a multiple representation of the ligand conformation (4D-QSAR). Significant contributions to this field include the construction of 3D models using a 4D formalism,[10] the use of a genetic neural network,[11] the Catalyst concept,[9,22] the Almond approach,[23] the utilization of self-organizing molecular field analysis,[24] and the Quasar concept.[12−15]

The Quasar concept developed at our laboratory allows for a multiple representation of the ligand topology (conformations, orientations, stereoisomers, and protomers, referred to as the fourth dimension in QSAR)[12] as well as a multiple representation of induced-fit hypotheses (the fifth dimension). Both ensembles are available throughout the entire simulation, and Boltzmann criteria are used for selecting the most powerful combinations.[12] This approach reduces the bias associated with the choice of the bioactive conformation, the ligand alignment, and the induced-fit model. Quasar allows also for H-bond flip-flop and accounts for solvation phenomena. The technical details of model construction in Quasar are published elsewhere[12−15] and shall therefore only be summarized here.

1. Construction of Receptor Envelopes. Induced fit may be simulated by adapting a van der Waals surface (constructed about all ligands defining the training set) to the topology of each ligand molecule of training, test, and prediction set. This is achieved by mapping this surface on to a transiently generated inner envelope, which snugly accommodates the individual ligand molecule. This procedure, mimicking a local induced fit, can be performed isotropically (linearly), anisotropically (field-scaled),[25] or through energy minimization.[12] The rms deviation from the mean → inner envelope is used to estimate the energy associated with the receptor-to-ligand adaptation (cf. eq 2). Typical rms shifts are in the range of 0.4−2.5 Å and yield "induced-fit energies" of 0.2−6.0 kcal/mol.

2. Generation of an Initial Family of Parent Structures. Points on the receptor surface are then randomly populated with atomistic properties (Table 1). While the distributed properties are identical for all ligand molecules, their exact location on the envelope varies slightly (rms fluctuations range from 0.5 to 1.5 Å with maximal individual shifts as large as 3.5 Å) depending on the very ligand molecule.

If there is experimental or other evidence for a solvent accessible receptor cavity, parts of the envelope may be assigned to represent solvent. Alternatively, regions may be defined as being purely hydrophobic in nature or nonexistent (void), allowing for shallow binding pockets. Such assignments can be static in nature or dynamically evolved.[12−15]

3. Evolution of a Model Family. Using a genetic algorithm (for a detailed description, see, for example, refs 20 or 26), the initial family of receptor models is evolved simulating crossover events. At each crossover step, there is a small probability (typically 0.01−0.02) of a transcription error, which is expressed by a random mutation. Thereafter, those two individuals of the population with the highest lack-of-fit (rms of $\Delta G_{pred}° - \Delta G_{exp}°$ obtained from a cross-validation, augmented by three penalty terms for the total number and types of properties, the difference between any other model, and the selectivity within the ensemble of conformations/orientations/

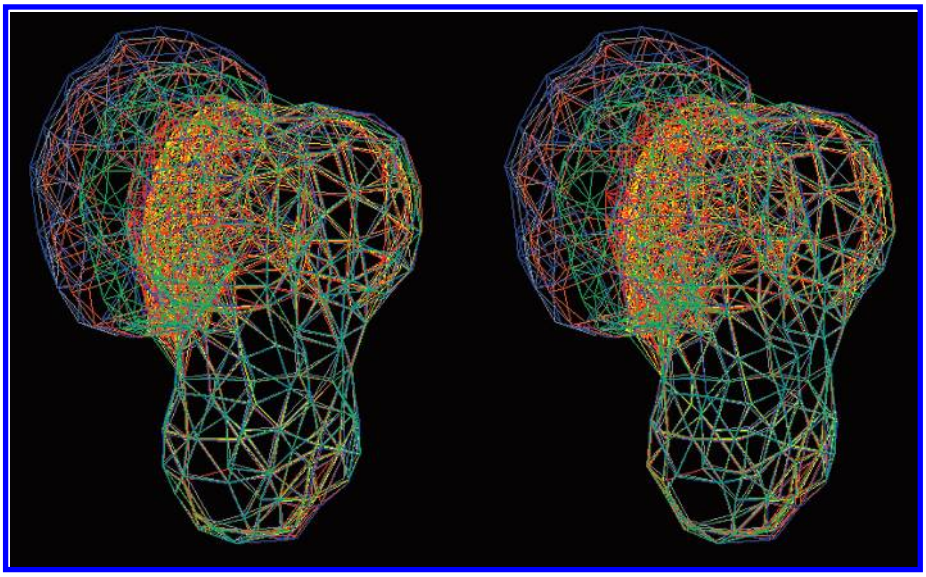

Figure 1. Stereoview of the envelope selection (induced fit hypotheses) for an antagonist molecule. Color coding: gray = field energy minimized mode; yellow = linear mode (contoured at the 75% level); red = steric field mode; orange = electrostatic field mode; blue = H bond field mode; green = lipophilicity potential mode.

protonation states, respectively, cf. eq 1) are discarded. This process is repeated until a target cross-validated r^2 (typically 0.85−0.95) or the experimental accuracy of the binding data (typically 0.1−0.24 kcal/mol, corresponding to an uncertainty of 20−50%) is reached.

$$\mathrm{LoF} = \mathrm{rms}\,[\Delta G^\circ_{\mathrm{pred}} - \Delta G^\circ_{\mathrm{exp}}]/\{1.0 - (p_{\mathrm{part}} + p_{\mathrm{diff}} + p_{\mathrm{sele}})/3.0\} \quad (1)$$

where p_{part} is the penalty for models with relatively many properties mapped onto their surface and $\{p_{\mathrm{part}} | 0.0 < p_{\mathrm{part}} < 1.0\}$, ≈0.3 = optimal situation, ≈0.99 = worst case scenario; p_{diff} is the penalty for relative model similarity when compared with all other models and $\{p_{\mathrm{diff}} | 0.0 \le p_{\mathrm{diff}} \le 1.0\}$, 0.0 = maximal model dissimilarity, 1.0 = all models identical; and p_{sele} is the penalty for unspecific selection of the conformer/orientomer/protomer ensemble and $\{p_{\mathrm{sele}} | 0.0 \le p_{\mathrm{sele}} \le 1.0\}$, 0.0 = only a single conformer selected, 1.0 = all conformers are selected equally frequent.

4. Estimation of Relative Free Energies of Ligand Binding. In our concept,[12−15] we have combined the approach of Blaney et al.[27] with a method of Still et al.[28] for estimating ligand solvation energies and a term to correct for the loss of entropy upon receptor binding following Searle and Williams.[29]

$$E_{\mathrm{bdg}} \approx E_{\mathrm{lig-rec}} - T\Delta S_{\mathrm{bdg}} - E_{\mathrm{solv,lig}} - \Delta E_{\mathrm{int,lig}} - E_{\mathrm{env.adapt,lig}} \quad (2)$$

where $E_{\mathrm{lig-rec}}$ is the force field energy of the ligand−receptor interaction,[12,30−31;32] $T\Delta S_{\mathrm{bdg}}$ is the change in ligand entropy upon receptor binding,[29] $E_{\mathrm{solv,lig}}$ is the ligand desolvation energy,[28] $\Delta E_{\mathrm{int,lig}}$ is the change in ligand internal energy upon receptor binding, and $E_{\mathrm{env.adapt,lig}}$ is the energy uptake required for adapting the receptor envelope (cf. above).

When using a multiple ligand representation, the interactions of all conformations, orientations, and protonation states are calculated toward all members of the receptor−model family. The contribution of an individual entity to the total energy is determined using a normalized Boltzmann distribution.[12−15]

$$E_{\mathrm{bdg,tot}} = \sum E_{\mathrm{bdg,ind}} \cdot \exp\,(-w_i \cdot E_{\mathrm{bdg,ind}}/E_{\mathrm{bdg,ind,lowest}}) \quad (3)$$

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

Free energies of ligand binding, $\Delta G^\circ_{\mathrm{prd}}$, are then predicted by means of a linear regression between $\Delta G^\circ_{\mathrm{exp}}$ and E_{bdg} (cf. eq 2) using the ligand molecules of the training set

$$\Delta G^\circ_{\mathrm{prd}} = |a| \cdot E_{\mathrm{bdg}} + b \quad (4)$$

Slope and intercept of eq 4 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. As in Quasar, the receptor surrogate is represented by a family of models (typically 100−1000), and this approach allows for a more subtle scaling of the ligand−receptor interactions.

5. Analysis of the Model Family. A mandatory criterion to validate a family of receptor models is their ability to predict relative free energies of ligand binding for an external set of test ligand molecules, not used during model construction (e.g., its rms deviation or the predictive r^2 value). A more serious challenge to a model family is the so-called scramble test (cf. ref 26). Here, the binding data (i.e., $\Delta G^\circ_{\mathrm{exp}}$) of the training set are randomly scrambled with respect to the true biological values, and the simulation is repeated under otherwise identical conditions. If, under these circumstances, the ligands of the test set are still predicted correctly (i.e., a predictive $r^2 > 0.5$), the model is worthless, as it is not sensitive toward the biological data ($\Delta G^\circ_{\mathrm{exp}}$).

In contrast to the previously published 4D-QSAR studies using the Quasar technology,[12−15] our most recent concept allows for a simultaneous evaluation of an ensemble of induced-fit hypotheses (5D-QSAR), thus reducing the bias with the choice of the adaptation mechanism. Presently, up to six protocols may be selected as follows: (i) a linear mode, scalable from 0.0 to 1.0 and typically applied at the 0.75 level, with 0.0 referring to "no induced fit" and 1.0 to a maximal adaptation to the topology of the individual molecule; (ii−iv) adaptations based on the steric, electrostatic, and H-bond field, respectively; (v) an energy minimization along the steric field vectors;[33] and (vi) an adaptation based on the molecular lipophilicity potential.[34,35] The induced fit in protocols ii−iv and vi is executed proportional to the effective field acting on the mean envelope (the outer surface), followed by a constrained minimization, to restore equal separations between neighbored grid points defining the adapted surface. While the linear mode behaves isotropically and depends solely on molecular shape, all other options account for molecular properties (presently the steric field, the electric charge distribution, H-bond donors or acceptors and the lipophilicity potential), yielding an anisotropic induced-fit model. Figure 1 shows an example of the induced-fit models that result from the protocols described above.

During parent generation, all selected induced-fit models are evaluated and the entity with the lowest lack-of-fit value is selected. This implies that for each model of the surrogate family *n* conformers (typically 4−16) and *k* induced-fit models (2−6) are evaluated, representing a genuine 5D-QSAR approach.[36] A 5D-QSAR simulation may, however, not simply be substituted by *n* 4D-QSAR runs as the 5D approach supports a dynamic interchange of the induced fit models—a behavior that is frequently observed. Nonetheless, the evolution converges to a single induced fit mode in most cases (cf. Figures 4 and 7).

Results and Discussion

1. Investigated Systems. The neurokinin-1 (*NK-1*) receptor is one of a family of neuroreceptors involved in various signal−transduction pathways, including nociception, nausea, bronchioconstriction, vasodilation, and visceral smooth muscle contraction.[37,38] The *NK-1* receptor binds the undecapeptide neurotransmitter substance P (SP) with a binding affinity in the 0.05−0.5 nM range.[39] A link between transmission of pain, the induction of inflammatory processes as a result of noxious stimuli, and the release of SP has been established. These observations suggest that SP receptor antagonists may be of significant therapeutic use in the treatment of a wide range of clinical conditions, ranging from arthritis, migraine, and asthma to postoperative pain and nausea.[40−41]

In a recent 4D-QSAR study based on 50 + 15 antagonist molecules (training and test set), we have validated a surrogate for the *NK-1* receptor. The evolution was based on a population of 500 receptor models and simulated during 40 000 crossover steps, corresponding to 80 generations. It yielded a cross-validated r^2 of 0.887 for the 50 ligands of the training set (represented by a total of 218 conformers and protonation states) and a predictive r^2 of 0.834 for the 15 ligands of the test set (represented by a total of 70 conformers and protonation states).[12] A series of three scramble tests (with an average predictive r^2 of −0.416) demonstrated the sensitivity of the surrogate toward the biological data. Using this model, the activities of twelve new compounds—four of which were subsequently synthesized and tested—have been predicted.[12] For most of the *NK-1* antagonists, the genetic algorithm selected a single entity (out of the up to 12 conformers or protomers) to preferably bind to the receptor surrogate. Most important, the evolution converged at an identical protonation scheme for all antagonists. This indicates that 4D-QSAR techniques can reduce the bias associated with the choice of the bioactive conformation as each ligand molecule may be represented by an ensemble of conformations, orientations, protonation states, and enantiomers.

2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD) and related compounds represent serious environmental health hazards whose effects include tumor promotion, dermal toxicity, immunotoxicity, developmental and reproductive toxicity, and induction and inhibition of various enzyme activities. TCDD also induces differentiation changes affecting, for example, the human epidermis—manifesting itself as chloracne. There is strong evidence that the toxicity is mediated by the aryl hydrocarbon (*Ah*) receptor, a regulatory element involved in the mammalian metabolism of xenobiotics.[42−46]

In two earlier studies, we used 3D- and 4D-QSAR concepts to construct receptor surrogates for a total of 121 polysubstituted dibenzodioxins, dibenzofurans, and biphenyls as well as polyaromatic hydrocarbons.[13,47] As the relative orientation of the various (quasi-symmetric) toxins in the binding site cannot unambiguously be assumed, we allowed for up to four different orientations for each toxin in the 4D study. The latter yielded a cross-validated r^2 of 0.857 for the 91 ligands of the training set (represented by a total of 348 different orientations) and a predictive r^2 of 0.795 for the 30 test compounds (113 orientations total). No test compound was predicted false positive or false negative (the threshold was a factor of 12.6 in K: a range of 3.2×10^5 in K split into the five toxicity classes). A negative scramble test (predictive r^2: −0.066) demonstrated the sensitivity toward the biological data. An analysis of the individual orientations contributing to the final model family suggests that the mutual orientation cannot be trivially derived (from 2D or 3D data) and that a multiple representation of each compound significantly reduces the bias associated with the ligand alignment. The fact that better results are achieved when using 4D-QSAR supported these arguments.

The observation that 4D-QSAR concepts may be used for predicting the toxicity of known or hypothetical substances—if a receptor-mediated phenomenon can be assumed underlying the adverse effects—has spawned a new concept for testing larger batches of compounds for a potential toxicity by computational approaches: we are presently establishing an Internet Laboratory for estimating the harmful potential of any given compound—chemical, toxin, or drug—in silico.[48]

2. Boundary Conditions Used in the 5D-QSAR Simulations. The preparation of the data input (model building, conformational search, selection of the ligand representation, atomic partial charge model, and calculation of the ligand solvation energy) is described elsewhere.[12,13] For both systems—the *NK-1* and the *Ah* receptor—we selected all induced-fit models presently available in Quasar: a linear induced fit scaled to 0.75, the four field-based modes (steric, electrostatic, H bond, and lipophilicity), and the minimization along the steric field lines. Induced fit simulated using the steric field as determinant typically yields the tightest model: for the *NK-1* receptor, it leads to an rms induced fit of 1.9 Å (1.2−2.2 Å) and for the *Ah* receptor to corresponding values of 1.8 Å (1.6−2.2 Å). When using the minimization along the field lines instead, the effect is minimal (both *NK-1* and *Ah*: 1.2 Å). It is likewise interesting to analyze the variation of the induced fit over the different ligand molecules and conformers/protomers/orientomers. The maximum individual variation of a grid point (on the surface) triggered by a ligand molecule varies on a much narrower scale: for the *NK-1* receptor from 0.32 to 0.59 Å and for the *Ah* receptor from 0.44 to 0.65 Å. To allow for a true comparison, we used as identical settings as possible for the simulated evolution (random number generator, population size, mutation rate, radii, cross-validation groups, and weights; cf. refs 12 and 47) as well as identical computer operating systems (*NK-1*: Silicon Graphics; *Ah*: Macintosh) to avoid divergent evolutions as a consequence of different boundary conditions and rounding errors. In contrast hereto, the

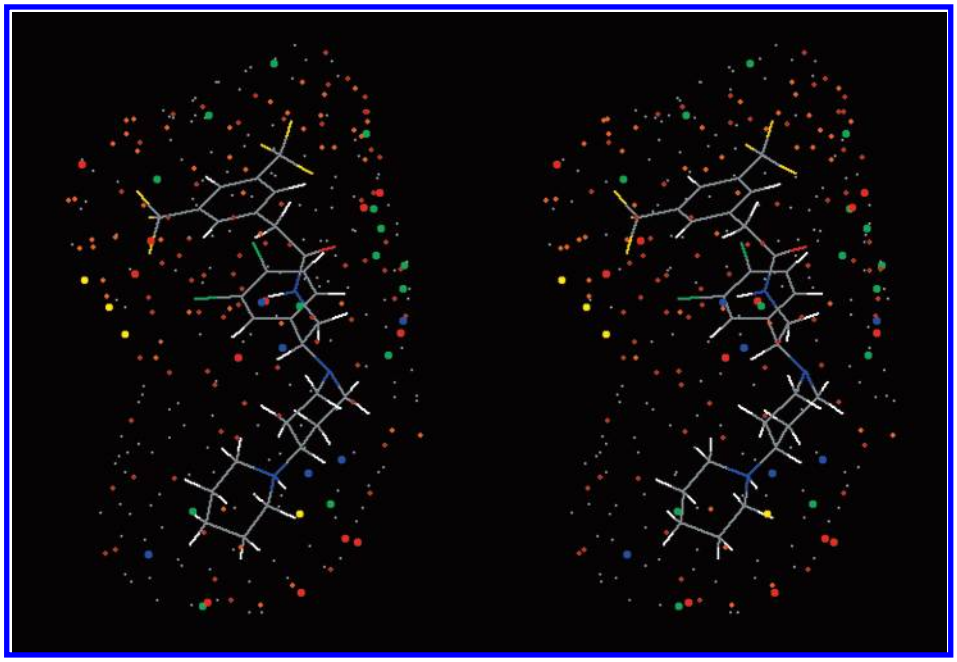

Figure 2. Stereoview of the surrogate for the *NK-1* receptor. Color coding of the mapped properties: red = positively charged salt bridge; blue = negatively charged salt bridge; yellow = H bond donor; green = H bond acceptor; light brown = positively charged hydrophobic; dark brown = positively charged hydrophobic; gray = neutral hydrophobic.

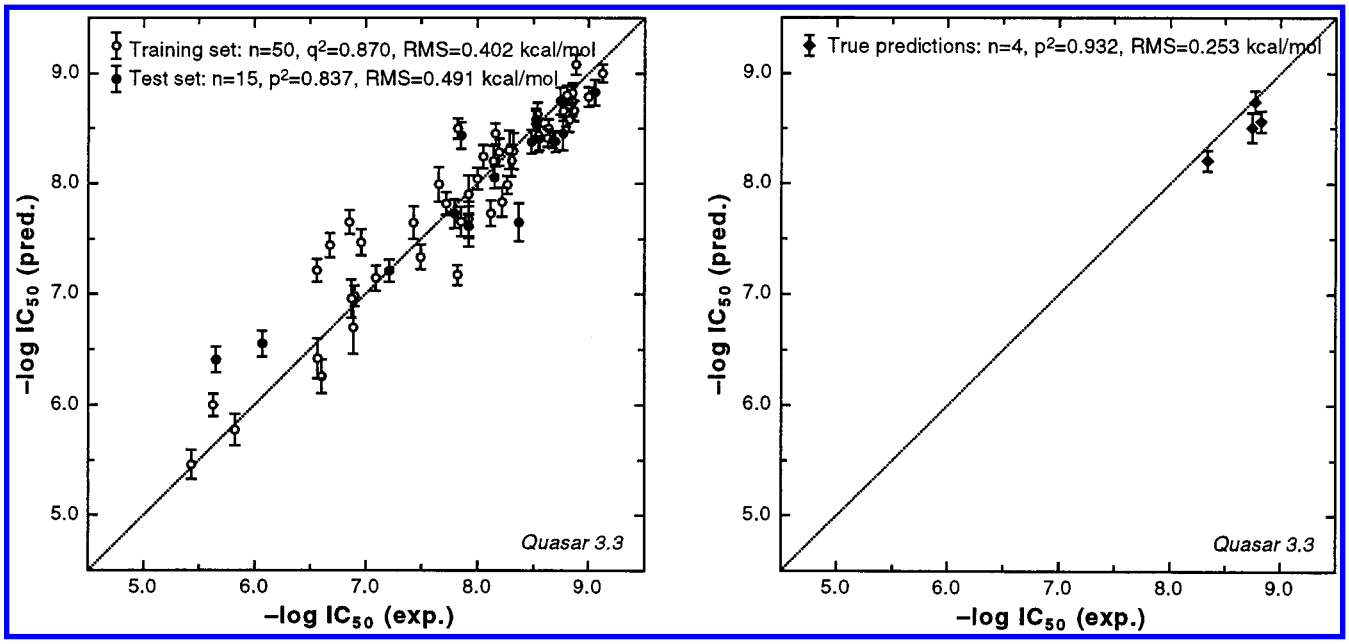

Figure 3. Graphical comparison of experimental and predicted binding affinities for the *NK-1* receptor: training and test set (left) and true predictions (right).

4D-QSAR simulations were based on a single induced-fit model at the time (*NK-1*: field energy minimized, linear, steric, and electrostatic; *Ah*: field energy minimized, linear, and steric) without the possibility of model crossover during the simulated evolution.

3. Results Obtained for the *NK-1* Receptor System. For the *NK-1* receptor, we used a population of 500 receptor models, a transcription error rate was set to 0.02, and the system was allowed to evolve for 30 000 crossover cycles, corresponding to 60 generations. In contrast hereto, the 4D simulation[12] had to run for 80 generations to achieve a comparable cross-validated r^2. The simulation reached a cross-validated r^2 of 0.870 (4D-QSAR: 0.887) and a predictive r^2 of 0.837 (4D-QSAR: 0.834). These quantities reflect values averaged over the 500 models—which, among themselves, differ in 39% [36−43%] (4D-QSAR: 25% [22−32%]) of the mapped 338 properties (4D-QSAR: 319), indicating that the model population evolved using 5D-QSAR is significantly less-affected by inbreed. The cross-validation was based on five groups comprising 10 ligands each (leave-10-out). A stereo representation of the quasi-atomistic receptor surrogate is depicted in Figure 2; experimental and calculated IC_{50} values are compared in Figure 3. A general comparison of the boundary conditions and results is given in Table 2.

The rms deviation for the 50 ligand molecules of the training set of 0.40 kcal/mol corresponds to an uncertainty factor of 2.0 in the IC_{50} (4D-QSAR: 0.37 kcal/mol, factor 1.9), the maximal individual deviation is 1.07

Table 2. Comparison of 4D/5D-QSAR Boundary Conditions and Results (cf. Text)

parameter	*NK-1* (4D-QSAR)	*NK-1* (5D-QSAR)	*Ah* (4D-QSAR)	*Ah* (5D-QSAR)
size of training set	50	50	91	91
size of test set	15	15	30	30
size of true predictions	4	4	4	4
size of population	500	500	200	250
crossovers/generations	40 000/80	30 000/60	24 000/120	50 000/200
cross-validated r^2 (q^2)	0.887	0.870	0.857	0.838
predictive r^2 (p^2)	0.834	0.837	0.795	0.832
true predictions (x^2)	0.778	0.932	0.684	0.854
scramble test (p^2) [n]	−0.416 [3]	−0.061 [5]	−0.066 [1]	−0.040 [5]
rms training set	0.374 ± 0.132 [1.9 in K]	0.402 ± 0.168 [2.0 in K]	0.564 ± 0.158 [2.6 in K]	0.609 ± 0.156 [2.8 in K]
max training set	1.144 ± 0.093 [7.1 in K]	1.069 ± 0.147 [6.3 in K]	1.440 ± 0.152 [12 in K]	1.690 ± 0.084 [18 in K]
rms test set	0.508 ± 0.133 [2.4 in K]	0.491 ± 0.167 [2.3 in K]	0.764 ± 0.175 [3.7 in K]	0.676 ± 0.218 [3.2 in K]
max test set	1.153 ± 0.185 [7.2 in K]	1.011 ± 0.157 [5.7 in K]	1.453 ± 0.156 [12 in K]	1.305 ± 0.097 [9.4 in K]
rms true predictions	0.513 ± 0.118 [2.4 in K]	0.253 ± 0.145 [1.5 in K]	0.724 ± 0.169 [3.5 in K]	0.468 ± 0.190 [2.2 in K]
max true predictions	0.843 ± 0.134 [4.3 in K]	0.349 ± 0.129 [1.8 in K]	0.885 ± 0.143 [4.6 in K]	0.718 ± 0.217 [3.4 in K]
mapped properties	319	338	158	190
identical properties	0.746	0.610	0.870	0.436
conformer selection	0.907	0.919	0.761	0.868

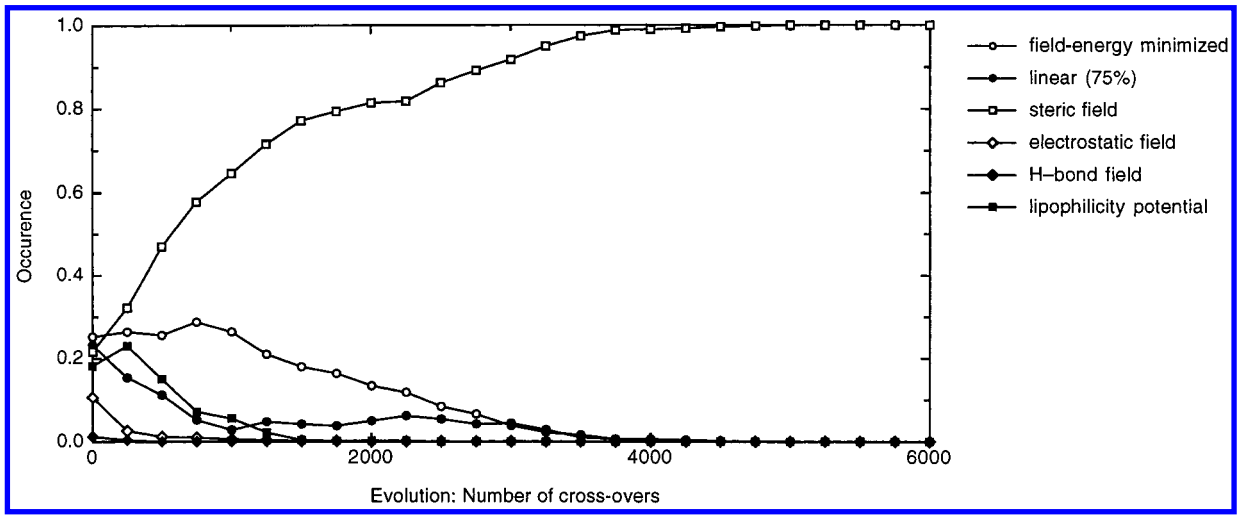

Figure 4. Evolution of the induced fit model selection in the *NK-1* receptor simulation.

kcal/mol, corresponding to a factor 6.3 in the IC_{50} (4D-QSAR: 1.1 kcal/mol, 7.1 in the IC_{50}). The analysis of the protomer distribution reveals that the evolution converged at an identical protonation scheme for 86% of the antagonist molecules. While all antagonists prefer the piperazine ring closer to the amide bond protonated—the majority of the exceptions is characterized by a bulky aromatic substituent in a position to this ring. The 65 antagonist molecules used in this study are described and depicted in ref 12.

With respect to the induced-fit hypothesis, the simulated evolution converged at the steric field model.[49] In contrast hereto, the best results in the 4D-QSAR simulation were obtained when using the electrostatic field-based approach,[12] a model that already became extinct after the second generation during the 5D-QSAR simulation (cf. Figure 4). The results using the steric induced-fit model in the 4D-QSAR simulation were of lesser quality (cross-validated $r^2 = 0.861$, predictive $r^2 = 0.713$), suggesting that early induced-fit hypothesis crossover during the simulated evolution is a critical process for diversity within the model family.

A total of 15 compounds (not used for model construction) were then selected for testing the predictive power of the receptor surrogate. For the 5D-QSAR simulation, a predictive r^2 of 0.837 was obtained, which compares to a value of 0.834 in the 4D-QSAR experiment. On the average, the predicted IC_{50} value of the test ligands deviates by 0.491 kcal/mol (a factor of 2.3 in the IC_{50}) from the experiment (4D-QSAR: 0.508 kcal/mol, 2.4 in the IC_{50}); the maximal observed deviation is 1.01 kcal/mol (a factor of 5.7 in the IC_{50}) from the experiment (4D-QSAR: 1.15 kcal/mol, 7.2 in the IC_{50}).

The process of choosing the "bioactive conformer" shows that both the 5D-QSAR and the 4D-QSAR approach are about equally selective: 0.919 (with 1.0 corresponding to a 100% selection of a single entity out of the conformer/protomer ensemble and 0.0 to an unspecific selection, cf. above) and 0.907, respectively. When analyzing the similarity of the 500 receptor models generated, they show a greater variation of the properties in the 5D-QSAR approach: on the average, only 61.0% of the mapped properties are identical throughout all surrogates while 39.0% still differ (4D-QSAR: 74.6% identical properties).

Also of great interest are the standard deviations associated with the calculated IC_{50} values: On the average, the predicted free energies of ligand binding obtained with 5D-QSAR vary slightly more over the 500 receptor surrogates—the ligands of the training set by 0.168 kcal/mol (4D-QSAR: 0.132 kcal/mol), those of the test set by 0.167 kcal/mol (4D-QSAR: 0.133 kcal/mol). While all values are well within the experimental error,[12] the observed differences are mainly a conse-

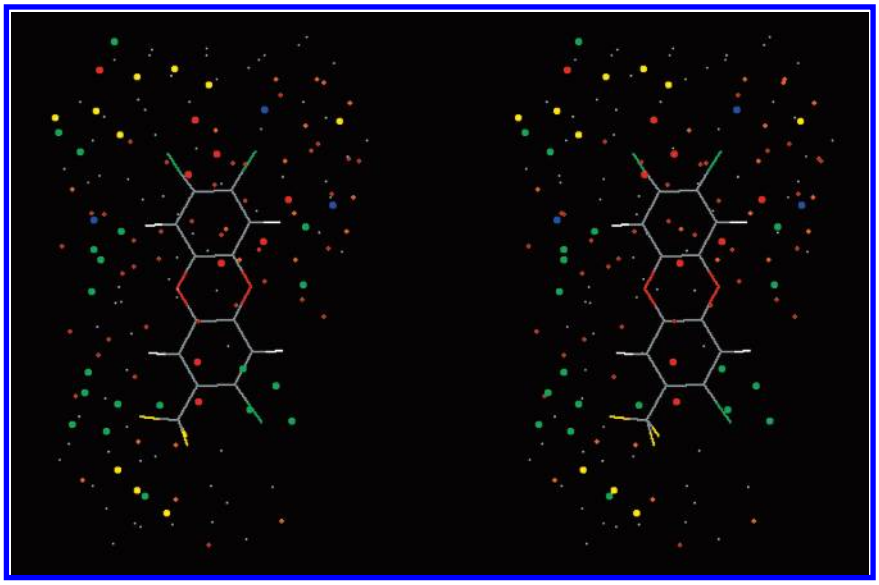

Figure 5. Stereoview of the surrogate for the aryl hydrocarbon receptor. Color-coding scheme of the mapped properties: see caption to Figure 2. This model has been obtained by explicitly allowing for a partially open binding pocket. On the average, 18% of the total surface represents open space (lower right).

quence of the model similarity among the surrogate family (5D: 61.0%; 4D: 74.6%). This demonstrates that 5D-QSAR can achieve an identical predictive power but with a more diverse population—an indicator of a good model family fitness.[50]

Next, the validity of the model family was assessed by a series of five scramble tests. The resulting predictive r^2 values (on the average: −0.01; 4D-QSAR: −0.44) demonstrate the sensitivity of the surrogate family toward the experimental IC_{50} data, for which it should establish a quantitative structure−activity relationship (cf. Table 2).

The ultimate test for any receptor model is "true predictions"—ligand molecules that were not part of either training or test set and that, preferably, are synthesized and tested after the prediction. In our previous study, we used the surrogate to predict the affinity of four new compounds, synthesized thereafter.[12] The predictive r^2 value obtained with 5D-QSAR (0.932) suggests a significantly higher quality of the model (4D-QSAR: 0.778). The least precise prediction deviates 0.35 kcal/mol from the experiment (a factor 1.8 in the IC_{50}); in contrast hereto, the results obtained with the 4D-QSAR simulation show a maximal deviation of 0.84 kcal/mol (a factor 4.3 in the IC_{50}). Details are given in Table 2 and Figure 3.

4. Results Obtained for the *Ah* Receptor System. For the *Ah* receptor, we used a population of 250 receptor models, a transcription error rate was set to 0.02, and the system was allowed to evolve for 50 000 crossover cycles, corresponding to 200 generations. The simulation reached a cross-validated r^2 of 0.838 (4D-QSAR: 0.857) and a predictive r^2 of 0.832 (4D-QSAR: 0.795). These quantities represent values averaged over the 250 models—which, among themselves, differ in 54% [50−64%] (4D-QSAR: only 13% [11−17%]) of the mapped 190 properties (4D-QSAR: 158), indicating that the population evolved using 5D-QSAR is substantially less-affected by inbreed. The cross-validation was based on four groups comprising 22−23 ligands each. A stereo representation of the receptor surrogate is depicted in Figure 5; experimental and calculated IC_{50} values are compared in Figure 6. A general comparison of the boundary conditions and results is given in Table 2. The 131 toxins used in this study are described in ref 47.

The rms deviation for the 91 ligand molecules of the training set of 0.61 kcal/mol corresponds to an uncertainty factor of 2.8 in the dissociation constant (4D-QSAR: 0.56 kcal/mol, a factor 2.6 in K), the maximal individual deviation is 1.7 kcal/mol, corresponding to a factor 18 in the dissociation constant (4D-QSAR: 1.4 kcal/mol, 12 in the IC_{50}). The clearly larger deviations—when compared to the *NK-1* system, cf. above—may be explained by the fact that the *Ah* data set includes four different substance classes (dibenzodioxins, dibenzofurans, biphenyls, and PAHs) while the *NK-1* data set comprises a single class. Moreover, the *Ah* data set spans a range of 3.1×10^5 in K_i, the *NK-1* system 4.9×10^3 in IC_{50}; a factor 63 difference.

A total of 30 compounds (not used for model construction) was then employed for testing the predictive power of the receptor surrogate. For the 5D-QSAR simulation, a predictive r^2 of 0.832 was obtained, which compares to a value of 0.795 in the 4D-QSAR experiment. On the average, the predicted binding affinities of the test ligands deviate by 0.676 kcal/mol (a factor of 3.2 in the binding affinity) from the experiment (4D-QSAR: 0.764 kcal/mol, factor 3.7); the maximal observed deviation is 1.31 kcal/mol (a factor of 9.4 in K) from the experiment (4D-QSAR: 1.45 kcal/mol, 12 in K).

The evolution of the induced-fit model is depicted in Figure 7; as in the 4D-QSAR study, the best results were obtained using the field energy minimized-based approach.[47] In the 5D-QSAR study, two crossovers (i) lipophilicity-based potential to field energy minimized after approximately one generation and (ii) lipophilicity-based potential to steric field shortly thereafter) are observed.

The general behavior in selecting a "bioactive orientation" shows the 5D-QSAR approach (mean fraction of the most frequency selected orientation = 0.868; with 1.0 corresponding to a 100% selection of a single entity

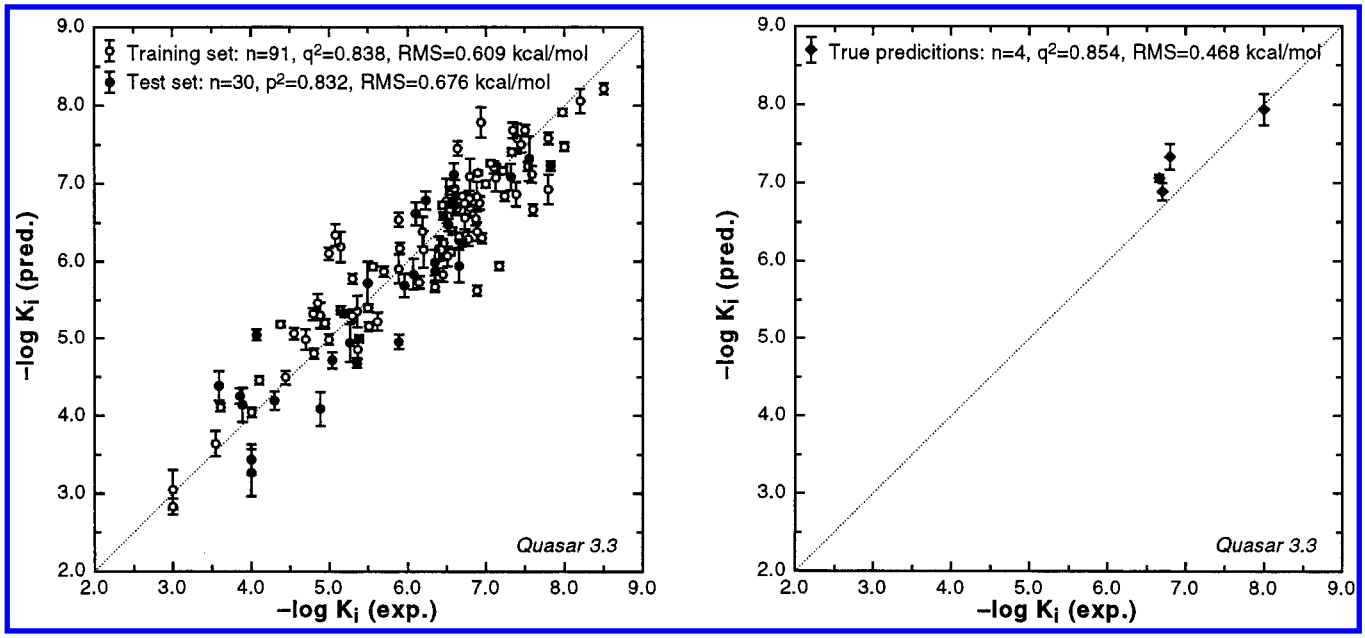

Figure 6. Graphical comparison of experimental and predicted binding affinities for the *Ah* receptor: training and test set (left) and true predictions (right).

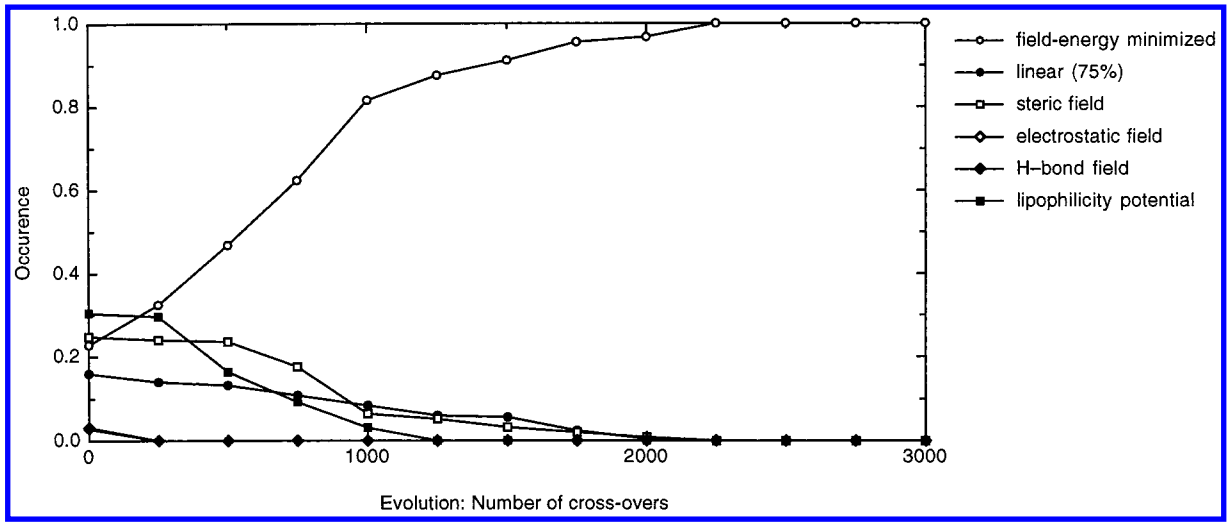

Figure 7. Evolution of the induced fit model selection in the *Ah* receptor simulation.

out of the orientational ensemble and 0.0 to an unspecific selection, cf. above) identifies a sharper selection as the 4D-QSAR simulations did (0.761). When analyzing the similarity of the 250 receptor models generated, they show a greater variation of the properties in the 5D-QSAR approach: on the average, 43.6% of the mapped properties are identical throughout all surrogates while 56.4% still differ (4D-QSAR: only 13.0% differ); both trends suggest a healthier population. The 5D-QSAR requires slightly more properties (190 on the average, cf. Table 1) than the 4D-QSAR simulation (158).

The standard deviations associated with the predicted free energies of ligand binding (IC_{50}) obtained with 5D-QSAR vary slightly more over the 250 receptor surrogates—training set: 0.156 kcal/mol (4D-QSAR: 0.158 kcal/mol); test set: 0.218 kcal/mol (4D-QSAR: 0.175 kcal/mol).

Next, the validity of the models was assessed by a series of five scramble tests. The resulting predictive r^2 values (on the average: −0.040; 4D-QSAR: −0.066) demonstrate the sensitivity of the surrogate family toward the experimental binding affinity, for which it should establish a quantitative structure−activity relationship. (cf. Table 2).

As for the *NK-1* system, we have used four new molecules (true predictions) to challenge the receptor model: 1,2,7,8-tetrachlorodibenzo-*p*-dioxin, 1,7,8-trichlorodibenzo-*p*-dioxin, 7,8-benzo-2,3-dichlorodibenzo-*p*-dioxin, and 1,3,6,7,8-pentachlorodibenzofuran. Ligand generation, conformational search, and superposition are described in ref 13. The predictive r^2 value obtained with 5D-QSAR (0.854) suggests a significantly higher quality of the model (4D-QSAR: 0.684). The least precise prediction deviates 0.72 kcal/mol from the experiment (a factor 3.4 in the K_i); in contrast hereto, the results obtained with the 4D-QSAR simulation show a maximal deviation of 0.89 kcal/mol (a factor 4.6 in the K_i). Details are given in Table 2 and Figure 6.

5. Comparison of 4D-QSAR and 5D-QSAR. The performance of 5D-QSAR can further be evaluated by analyzing the residuals ($|\Delta G_{pred}° - \Delta G_{exp}°|$) of the test

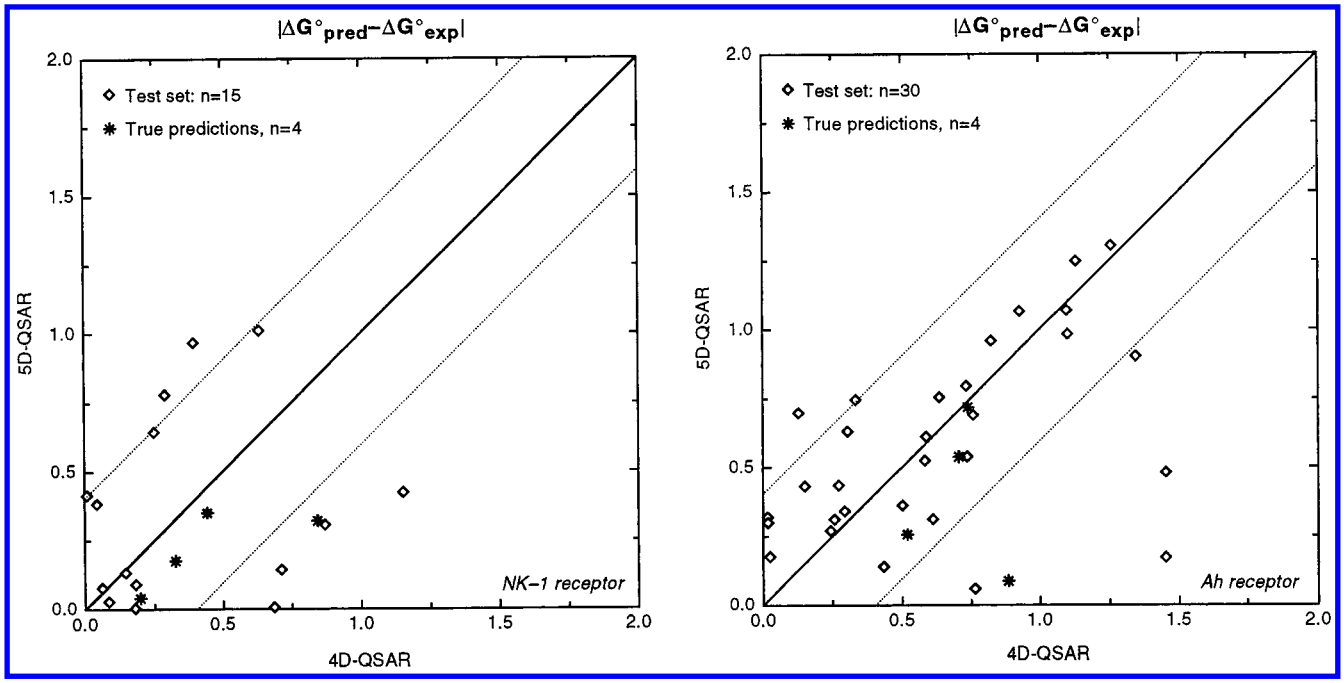

Figure 8. Comparison of residuals ($|\Delta G°_{pred} - \Delta G°_{exp}|$) of 5D-QSAR and 4D-QSAR for ligands of the test and prediction set. Points below the diagonal indicate a better performance of 5D-QSAR; points above show that 4D-QSAR is superior. Dashed lines are drawn at ±0.404 kcal/mol (a factor of 2.0 in IC_{50} or K_i). Left, neurokinin-1 receptor; right, aryl hydrocarbon receptor.

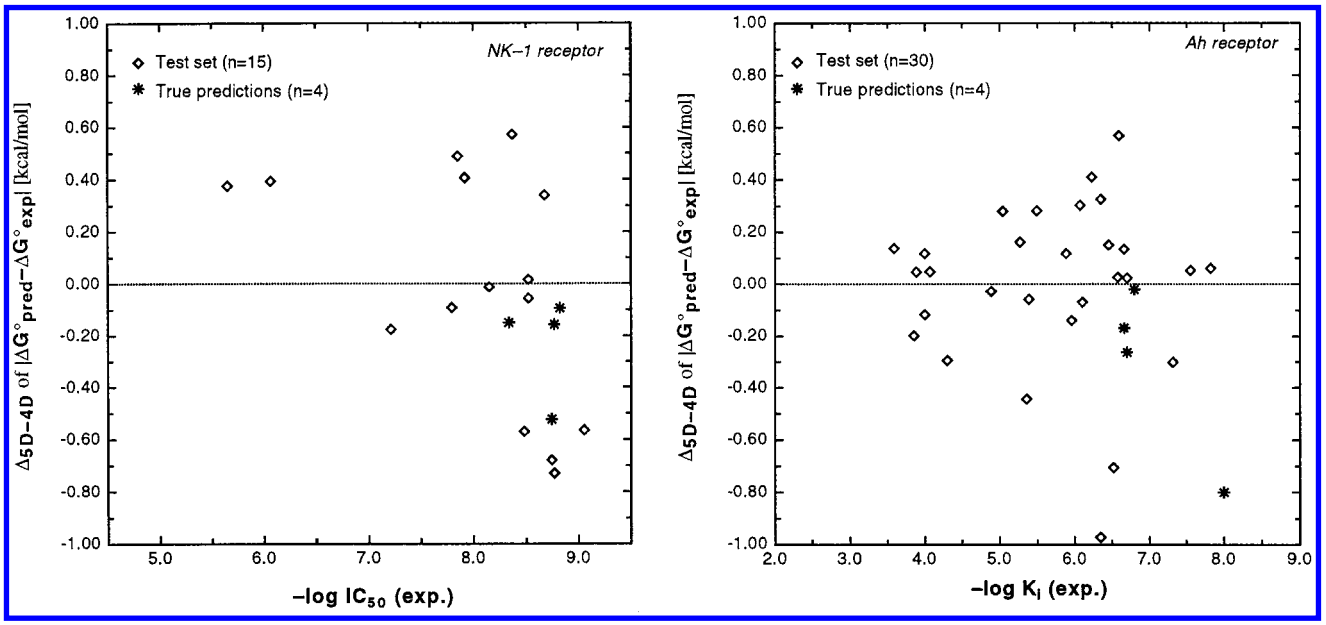

Figure 9. Comparison of relative residuals ($\Delta|\Delta G°_{pred} - \Delta G°_{exp}|$) of 5D-QSAR and 4D-QSAR as a function of the ligand affinity (IC_{50} or K_i). Negative values (below the dotted line) indicate a better performance of 5D-QSAR; positive values (above the dotted line) show that 4D-QSAR is superior. Left, neurokinin-1 receptor; right, aryl hydrocarbon receptor.

ligands. Figure 8 compares these entities with those obtained by 4D-QSAR. When selecting those ligands off by more than a factor of 2.0 from the experiment (0.404 kcal/mol) for the *NK-1* receptor, only two ligands "fail" in the 5D-QSAR simulation but five do in the 4D experiment. All true predictions are calculated more accurately in 5D-QSAR. A similar picture is obtained for the *Ah* receptor with the effects even more pronounced. Figure 9 analyzes the influence of the ligand affinity on the residuals. Clearly for the *NK-1* receptor system, 5D-QSAR performs better for high-affinity ligands ($IC_{50} < 10^{-8}$) including all true predictions while for the *Ah* receptor system ($K_i < 5.0 \times 10^{-7}$) the effect is somewhat less-pronounced but still includes all true predictions. This suggests that 5D-QSAR results in a model population with higher predictive power for high-affinity compounds—a desirable effect in drug-design applications.

Conclusions

In absence of an experimentally determined receptor structure, multidimensional QSAR techniques provide an elegant approach to the estimation of free energies of ligand binding. The 5D-QSAR concept Quasar developed at our laboratory not only allows for the representation of the ligand molecules by an ensemble of conformations, orientations, and protonation states (the fourth dimension) but also permits to simultaneously

evaluate up to six different local induced-fit protocols (the fifth dimension). The latter may significantly differ in shape and magnitude of the effect but is restricted to local phenomena ranging from zero to maximal 4.0 Å in rms shift. To estimate free energies of ligand binding, Quasar uses a directional force field and takes solvation phenomena, internal strain, and changes in entropy during receptor binding into account.[12] The new approach has been used to predict the IC_{50} values of neurokinin-1 antagonists (featuring both conformational flexibility as well as multiple protonation states) and to binding affinities of toxins binding to the Aryl hydrocarbon receptor (quasi-symmetric molecules subject to orientational degrees of freedom).

The results indicate that the use of multidimensional QSAR reduces the bias associated with the selection of ligand conformation and alignment (4D) and the choice of a suitable induced fit model (5D). The simulations performed up to date[51,52] indicate that the technique is capable of identifying a single active conformer and a single induced fit model and does not prefer a larger selection of lesser-contributing entities. Most important in our view, the binding affinities of new molecules (true predictions) are predicted more accurately with 5D QSAR. The 3D coordinates of all models discussed in this paper as well as information on the Quasar software may be obtained at http://www.biograf.ch.

Acknowledgment. Financial support from the Margaret and Francis Fleitmann Foundation (Lucerne, Switzerland) and the Foundation for Animal-Free Research (Zurich, Switzerland) is gratefully acknowledged.

References

(1) Kubinyi, H. QSAR and 3-D QSAR in drug design. 1. Methodology. *Drug Discovery Today* **1997**, *2*, 457−467.
(2) Kubinyi, H. QSAR and 3-D QSAR in drug design. 2. Applications and problems. *Drug Discovery Today* **1997**, *2*, 538−546.
(3) Kubinyi, H.; Folkers, G.; Martin, Y. C. QSAR: Current state, scope, and limitations. *Perspect. Drug Discovery Des.* **1998**, *12*, 3−23.
(4) Snyder, J. P.; Rao, S. N.; Koehler, K. F.; Vedani, A. Pseudoreceptors. In *3D QSAR in Drug Design*; Kubinyi, H., Ed.; ESCOM Science Publishers: Leiden, 1993; pp 336−354.
(5) Srivastava, S.; Richardson, W. W.; Bradley, M. P.; Crippen, G. Three-dimensional receptor modeling using distance geometry and Voronoi polyhedra. In *3D-QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; Escom: Leiden, 1993; pp 80−116.
(6) Murray-Rust, P.; Glusker, J. P. Directional hydrogen bonding to sp^2 and sp^3 hybridized O atoms and its relevance to ligand-macromolecule interactions. *J. Am. Chem. Soc.* **1984**, *106*, 1018−1025.
(7) Vedani, A.; Dunitz, J. D. Lone-pair directionality in H-bond potential functions for molecular mechanics calculations. *J. Am. Chem. Soc.* **1985**, *107*, 7653−7658.
(8) Baker, E. N.; Hubbard, R. E. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97−179.
(9) Sprague, P. W.; Hoffmann, R. Catalyst pharmacophore models and their utility as queries for searching 3D databases. In *Computer-Assisted Lead Finding and Optimization*; van de Waterbeemd, H., Testa, B., Folkers, G., Eds.; VCH: Weinheim/Germany, 1997; pp 223−240.
(10) 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.* **1997**, *119*, 10509−10524.
(11) So, S. S.; Karplus, M. Three-dimensional quantitative structure−activity relationships from molecular similarity matrixes and genetic neural networks. *J. Med. Chem.* **1997**, *40*, 4347−4359.
(12) Vedani, A.; Briem, H.; Dobler, M.; Dollinger, K.; McMasters, D. R. Multiple conformation and protonation-state representation in 4D-QSAR: The neurokinin-1 receptor system. *J. Med. Chem.* **2000**, *43*, 4416−4427.
(13) Vedani, A.; Dobler, M. Multidimensional QSAR in drug research: Predicting binding affinities, toxicity, and pharmacokinetic parameters. *Progress in Drug Research*; Jucker, E., Ed.; Birkhäuser: Basel/Boston/Berlin, 2000; pp 105−135.
(14) Vedani, A.; McMasters, D. R.; Dobler, M. Multi-conformational ligand representation in 4D-QSAR: Reducing the bias associated with ligand alignment. *Quant. Struct.-Act. Relat.* **2000**, *19*, 149−161.
(15) Streich, D.; Neuburger-Zehnder, M.; Vedani, A. Induced fit−The key for understanding LSD activity. A 4D-QSAR study on the 5-HT2A receptor system. *Quant. Struct.-Act. Relat.* **2000**, *19*, 565−573.
(16) Vedani, A.; Dobler, M.; Zbinden, P. Quasi-atomistic receptor surface models: A bridge between 3-D QSAR and receptor modeling. *J. Am. Chem. Soc.* **1998**, *120*, 4471−4477.
(17) Montgomery, J. A.; Secrist, J. A., III. PNP inhibitors. *Perspect. Drug Discovery Des.* **1994**, *2*, 205−220.
(18) Hernandez, M. A.; Powers, J. C.; Glinski, J.; Oleksyszyn, J.; Vijayalakshmi, J.; Meyer, E. F. Effect of the 7-amino substituent on the inhibitory potency of mechanism-based isocoumarin inhibitors for porcine pancreatic and human neutrophil elastases: a 1.85-Å X-ray structure of the complex between porcine pancreatic elastase and 7-[(N-tosylphenylalanyl)amino]-4-chloro-3-methoxyisocoumarin. *J. Med. Chem.* **1992**, *35*, 1121−1129.
(19) Hahn, M. Receptor surface models. 1. Definition and construction. *J. Med. Chem.* **1995**, *38*, 2080−2090.
(20) Walters, D. E.; Hinds, R. M. Genetically evolved receptor models: A computational approach to construction of receptor models. *J. Med. Chem.* **1994**, *37*, 2527−2536.
(21) Hahn, M.; Rogers, D. Receptor surface models. 2. Application to quantitative structure−activity studies. *J. Med. Chem.* **1995**, *38*, 2091−2102.
(22) Kurogi, Y.; Guner, O. F. Pharmacophore modeling and three-dimensional database searching for drug design using "Catalyst". *Curr. Med. Chem.* **2001**, *8*, 1035−1055.
(23) Pastor, M.; Cruciani, G.; McLay, I.; Picket, S.; Clementi, S. Grid-indpendent descriptors (GRIND): a novel class of alignment-independent three-dimensional descriptors. *J. Med. Chem.* **2000**, *43*, 3233−3243.
(24) Robinson, A. U.; Winn, P. J.; Lyne, P. D.; Richards, W. G. Self-organizing molecular field analysis: A tool for structure−activity studies. *J. Med. Chem.* **1999**, *42*, 573−583.
(25) The actual field excerted by all atoms of a given ligand molecule on to its envelope is determined using a nondirectional force field as—at this point—no atomic properties are deposited on the surface.
(26) Rogers, D.; Hopfinger, A. J. Genetic function approximation to generate a family of QSAR equations using genetic algorithms. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854−866.
(27) Blaney, J. M.; Weiner, P. K.; Dearing, A.; Kollman, P. A.; Jorgensen, E. C.; Oatley, S. J.; Burridge, J. M.; Blake. J. F. Molecular mechanics simulation of protein−ligand interactions: Binding of thyroid analogues to prealbumin. *J. Am. Chem. Soc.* **1982**, *104*, 6424−6434.
(28) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(29) Searle, M. S.; Williams, D. H. The cost of conformational order: Entropy changes in molecular associations. *J. Am. Chem. Soc.* **1992**, *114*, 10690−10697.
(30) Vedani, A.; Zbinden, P.; Snyder, J. P.; Greenidge, P. A. Pseudoreceptor modeling: The construction of three-dimensional receptor surrogates. *J. Am. Chem. Soc.* **1995**, *117*, 4987−4994.
(31) Vedani, A.; Huhta, D. W. A new force field for modeling metalloproteins. *J. Am. Chem. Soc.* **1990**, *112*, 4759−4767.
(32) 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 (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 ref 31.
(33) The latter approach searches for an energy minimum while moving along the field lines—thus, corresponding to a trajectory minimization.
(34) Furet, P.; Sele, A.; Cohen, N. C. 3D molecular lipophilicity potential profiles: A new tool in molecular modeling. *J. Mol. Graphics* **1988**, *6*, 182−189.
(35) Ghose, A. K.; Crippen, G. M. Atomic physicochemical parameters for three-dimensional structure-directed structure−activity relationships I: Partition coefficients as a measure of hydrophobicity. *J. Comput. Chem.* **1986**, *7*, 565−577.

(36) While simulating the evolution, the crossover procedure may be sped up by reducing the number of evaluated induced fit models, depending on their actual frequency—i.e., a model with low frequency is selected fewer times than a model with high frequency. By setting the speed factor to 0.0, all models are fully evaluated at each crossover step, a speed factor of 1.0 would represent a minimal model evaluation. The default is set at 0.9 implying that an induced fit model "extinct" at a given time is still tested at a 10% frequency.
(37) Regoli, D.; Boudon, A.; Fauchère, J.-L. Receptors and antagonists for substance P and related peptides. *Pharmacol. Rev*. **1994**, *46*, 551−599.
(38) Quartera, L.; Maggi, C. A. The tachykinin NK-1 receptor. Part I: Ligands and mechanisms of cellular activation. *Neuropeptides* **1997**, *31*, 537−563.
(39) Maggi, C. A.; Schwartz, T. W. The dual nature of the tachykinin NK-1 receptor. *Trends Pharmacol. Sci.* **1997**, *18*, 351−355.
(40) Takeuchi, Y.; Shands, E. F. B.; Beusen, D. D.; Marshall, G. R. Derivation of a three-dimensional pharmacophore model of substance P antagonists bound to the NK-1 receptor. *J. Med. Chem*. **1998**, *41*, 3609−3623.
(41) Ladduwahetty, T.; Baker, R.; Cascieri, M. A.; Chambers, M. S.; Haworth, K.; Keown, L. E.; MacIntyre, D. E.; Metzger, J. M.; Owen, S.; Rycroft, W.; Sadowsky, S.; Seward, E. M.; Shepheard, S. L.; Swain, C. J.; Tatters, F. D.; Watt, A. P.; Williamson, D. W.; Hargreaves, R. J. N-Heteroaryl-2-phenyl-3-(benzyloxy)-piperidines: A novel class of potent orally active human NK-1 antagonists. *J. Med. Chem*. **1996**, *39*, 2907−2914.
(42) Putzrath, R. M. Estimating relative potency for receptor-mediated toxicity: reevaluating the toxicity-equivalence factor (TEF) model. *Regul. Toxicol. Pharmacol.* **1997**, *25*, 68−78.
(43) Safe, S.; Krishnan, K. Cellular and molecular biology of Aryl hydrocarbon (Ah) receptor-mediated gene expression. In *Archives of Toxicology*, Suppl. 17; Degen, D. H., Seiler, J. P., Bentley, P., Eds.; Springer: Berlin, 1995; pp 116−124.
(44) Okey, A. B.; Riddick, D. S.; Harper, P. A. The Ah receptor: Mediator of the toxicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) and related compounds. *Toxicol. Lett.* **1994**, *70*, 1−22.
(45) Rappe, C. Environmental concentrations and ecotoxicological effects of PCDDs, PCDFs and related compounds. *Organohalogen Compd, Dioxin* **1993**, *12*, 163−170.
(46) Whitlock, J. P., Jr. Mechanistic aspects of dioxin action. *Chem. Res. Toxicol.* **1993**, *6*, 754−763.
(47) Vedani, A.; McMasters, D. R.; Dobler, M. Genetische Algorithmen im 3D- QSAR: Verwendung multipler Wirkstofforientierungen zur verbesserten Voraussage der Toxizität. *ALTEX* **1999**, *16*, 140−143.
(48) Vedani, A.; Dobler, M. Internet laboratory for predicting harmful effects triggered by drugs and chemicals. *ALTEX* **2001**, *18*, 110−114.
(49) As in the Quasar concept, the receptor surrogate is represented by a family of models,[12] each of these models can independently select the most predictive induced-fit model. When generating the initial population (by randomly distributing the available properties on the van der Waals surface, cf. Table 1), each model is analyzed for all selected induced fit models. Using a Boltzmann criterion, the model with the lowest lack-of-fit is accepted (cf. eq 1).
(50) A more diverse (i.e., less inbred) population is an indicator that the system has not been overtrained—if the evolution would be continued endlessly, the population would consist of clones of the best individual. A diverse population facilitates a more accurate prediction of new molecules ("true predictions") as the system does not purely mirror the training set. Therefore, the lack-of-fit function in Quasar rewards diverse models (cf. eq 1). A smaller amount of inbreed also indicates that the selected induced fit hypothesis—obtained linearly or via crossover—describes the situation at the true biological receptor more precisely as more solutions (i.e., the distribution of properties, representing flexible side chains of the binding site) yield high correlations.
(51) http://www.biograf.ch/GIFS/Predictions.gif.
(52) Vedani, A.; Dobler. Multidimensional QSAR: Moving from three- to five-dimensional concepts. *Quant. Struct.-Act. Relat*., in press.

JM011005P