Quantitative Binding Site Model Generation: Compass Applied to Multiple Chemotypes Targeting the 5-HT_{1A} Receptor

Ajay N. Jain,* Nomi L. Harris, and John Y. Park

Arris Pharmaceutical Corporation, 385 Oyster Point Boulevard, South San Francisco, California 94080

Received October 3, 1994®

We present enhancements to the Compass algorithm that automatically deduce interchemotype relationships and generate predictive quantitative models of receptor binding based solely on structure–activity data. We applied the technique to a series of compounds assayed for 5-HT_{1A} binding. A model was constructed from 20 compounds of two chemotypes and used to predict the affinities and bioactive conformation of 35 new compounds, most of which had new underlying scaffolds and/or functional groups. The model's mean error of prediction was 0.5 log units (essentially the assay resolution), even on quite divergent series. The predictions are supported by an interpretable hypothesis for the binding determinants of the receptor and the geometric relationships of the chemotypes.

Introduction

Structure–activity data is expensive and time-consuming to generate. Consequently, effective exploitation of such data is of keen interest to those engaged in drug discovery. This work was prompted by three papers that investigated several series of compounds assayed for binding to the 5-HT_{1A} and D_2 receptors. Lin *et al.*[1,2] reported two series of potent conformationally constrained tricyclic compounds, and Chidester *et al.*[3] discussed the impact of these new compounds on pharmacophores for the receptors proposed by others.[4–7] Through an in-depth analysis of the structure–activity data along with small-molecule X-ray crystallographic data and extensive modeling, Chidester *et al.* were able to refine the hypotheses about binding determinants for each receptor.[3]

The Compass algorithm[8,9] analyzes structure–activity data to construct a *quantitative binding site model*—a mathematical characterization of the ligand-receptor binding interface. In addition to predicting affinities, the algorithm predicts the bioactive conformations and alignments of molecules, which can be displayed to show the structural requirements for binding and the geometrical relationship between chemotypes. Compass has been described previously in two applications.[8,9] It has three characteristics that differentiate it from other computational techniques.[10–16] First, its molecular representation is based on a dense sampling of surface shape and potential hydrogen bonding interfaces independent of internal molecular connectivity. Second, Compass is able to automatically select from multiple conformations and orientations of molecules, perhaps *the* critical issue in this application. Third, Compass employs a nonlinear computational model of the relationship between molecular features and activity. Compass's numerical prediction accuracy depends on the combination of these three characteristics.[8,9,17]

We present enhancements to the Compass algorithm that facilitate construction of a 5-HT_{1A} binding site model. The compounds analyzed in this paper span several chemotypes, and the problem of ascertaining the relative geometric relationships between the compounds in their bound states is complex. Two procedures were added to Compass in the course of this work: an automatic shape-based procedure for generating initial molecular alignments and a procedure for forming an initial hypothesis for bioactive conformations of molecules based on global constraints. A Compass model was constructed from 20 molecules of two chemotypes (see Figure 1 for examples) and used to predict the affinities and bioactive conformations of 35 new compounds, most of which had underlying scaffolds and/or pendant functionalities not present in the learning set. The model's mean prediction error was 0.5 log units, even on quite divergent series. Given that the majority of new compounds were racemic mixtures and assuming that the assays were at least slightly noisy, this level of predictive performance error is essentially indistinguishable from noise. The predictions are supported by interpretable hypotheses for the binding determinants of the compounds and the relationships between different chemotypes.

Methods

Algorithmic Approach. Compass has been described in more detail in previous papers,[8,9] but a brief review is given here. It is followed by descriptions of two algorithmic enhancements. The Experimental Section contains additional details not central to understanding the algorithms.

Review of Compass Compass takes molecular structures and assay values as input. The output is a quantitative model of the ligand–receptor interface, which is used on novel molecules to predict biological activity, bioactive conformation, and alignment and provide graphical guidance for design.

The procedure of building a model begins by conducting a search for low-energy conformations of a set of molecules. This provides a pool of energetically accessible shapes for each molecule. Compass is allowed to choose from conformations that are within 5.0 kcal of the minimum for each isomer. The computed conformational energy is *not* used as an input to Compass since the conformations of small molecules interacting with proteins are typically significantly higher in energy than their computed minimum energy conformations, even for high-affinity interactions. This procedure provides a broad sample of potential bioactive conformations without relying too heavily on the accuracy of force-field computations. The conformations are placed in rough initial orientations by an automated alignment procedure that seeks to maximize the similarity of each conformation to the conformations of active molecules. A specific conformation and orientation is called

* Author to whom all correspondence should be addressed.
® Abstract published in *Advance ACS Abstracts*, March 15, 1995.

0022-2623/95/1838-1295$09.00/0

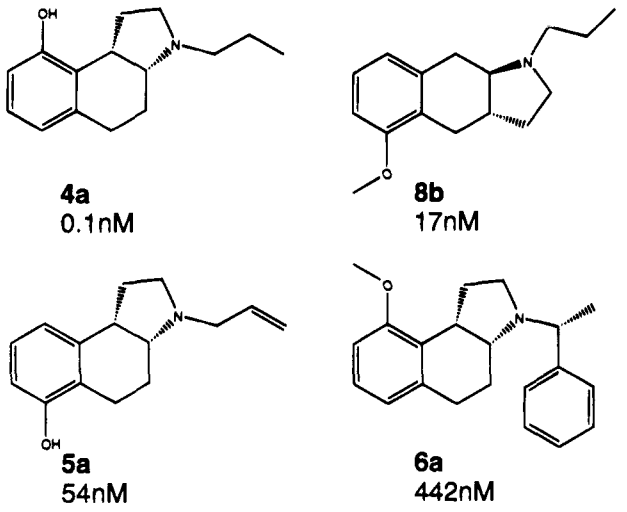

Figure 1. Example molecules (see Table 1 for the full data set).

a pose. From these starting poses, an initial model of activity is constructed by training a neural network (a complex nonlinear function with adjustable parameters). The model is trained with a numerical characterization of surface shape and potential hydrogen bonding interfaces as input and experimentally observed activity as its desired output. Choice among poses is interleaved with incremental neural network training to allow the network to examine all the available poses during training. The intermediate model is used to generate improved molecular poses for each molecule. The model is refined using the new poses, and the process iterates until it converges on a best pose for each molecule within the model. The resulting model is applied to new molecules using the analogous procedure—in essence "docking" new molecules into Compass's quantitative binding site model.

Shape-Based Molecular Alignment. Prior work with Compass relied on initial alignments that were generated in an *ad hoc* fashion.[8,9] For example, in the steroid application predicting carrier protein affinities,[8] it was sufficient to initially align the molecules based on the carbons forming the standard steroid backbone (paralleling the work of Cramer *et al.*[15]). While Compass is able to refine the initial alignments adaptively as a model is constructed, a process that is key to making accurate predictions,[8,9,17] gross misalignments can present difficulties. Compass's learning algorithm is capable of generating complex models of binding by forming disjunctions of chemotype-specific constraints. Such models will not yield accurate predictions for candidate molecules designed to test relationships between chemotypes.

In the work reported here, there are multiple active chemotypes that lack an obvious alignment when one considers the two-dimensional representation of these compounds. The full data set is presented in Table 1. Molecule names that are appended with an "a" indicate the α configuration of the nitrogen as drawn, and "b" indicates β. Molecule names lacking the a or b suffix indicate racemic mixtures. Figure 1 shows some examples of compounds in the series. There are two tempting superpositions of **8b** onto **4a** (the most active molecule in the series) that align the six-membered rings and align either the nitrogens or the oxygens. This is also true for **5a**, in which case even the geometry of the nitrogen relative to the ring is conserved in the nitrogen-based alignment. However, when one considers the alignment problem in three dimensions, other orientations appear to be considerably more plausible.

The new alignment procedure relies on Compass's surface-based features.[8,9] These features are computed with respect to a set of reference points uniformly placed on spheres of 6.0 and 9.0 Å in radius centered at the origin of a Cartesian reference frame. Figure 2 shows a molecule inside the set of reference points. The features are uniformly distributed in space. At each point, three distances are computed to the

Table 1. Structural and Assay Data

name	body	config[a]	R_1	R_2	R_3	R_4	K_i (nM)[b]
1a	A	α	allyl		OCH_3		0.2[c]
1b	A	β	allyl		OCH_3		49
2a	A	α	*n*-propyl		OCH_3		0.2
2b	A	β	*n*-propyl		OCH_3		186
3a	A	α	allyl	OCH_3			236
3b	A	β	allyl	OCH_3			333
4a	A	α	*n*-propyl		OH		0.1
4b	A	β	*n*-propyl		OH		3
5a	A	α	allyl	OH			54
5b	A	β	allyl	OH			1000
6a	A	α	mbz[b]		OCH_3		442
7b	A	β	mbz[d]		OCH_3		317
8a	B	α	*n*-propyl		OCH_3		1000
8b	B	β	*n*-propyl		OCH_3		17
9a	B	α	allyl		OCH_3		1000
9b	B	β	allyl		OCH_3		41
10a	B	α	*n*-propyl		OH		470
10b	B	β	*n*-propyl		OH		36
11a	B	α	allyl		OH		470
11b	B	β	allyl		OH		56
12	A		allyl				31
13	A		cpm[e]	OCH_3			1000
14	A		*n*-propyl				37
15	A		*n*-propyl	OCH_3			1000
16	A		allyl		OH		0.3
17	A		*n*-propyl	OH			1000
18	A		*n*-propyl		OH	α-CH_3	1
19	A		*n*-propyl		OH	β-CH_3	16
20	B		benzyl		OCH_3		127
21	B				OCH_3		29
22	B		cpm		OCH_3		29
23	B		cpm		OH		69
24	B		allyl	OCH_3			104
25	C		allyl				11
26	C		allyl		OCH_3		4.7
27	C		allyl	OCH_3			16
28	C		cpm	OCH_3			66
29	C		*n*-propyl				44
30	C		*n*-propyl	OCH_3			66
31	C		allyl		OH		4.9
32	C		allyl	OH			16
33	C		*n*-propyl	OH			338
34	D		*n*-propyl		OCH_3		1000
35	D		allyl		OCH_3		1000
36	D		benzyl		OCH_3		1000
37	D		*n*-propyl		OH		1000
38	D				OCH_3		1000
39	D		cpm		OCH_3		721
40	D		allyl	OCH_3			145
41	D		*n*-propyl	OCH_3		α-CH_3	384
42	D		*n*-propyl	OCH_3		β-CH_3	11
43	E		H				0.3
44	E		CH_3				5.9
45	F		1[f]				2.2
46	F	α	0[f]				24.8

[a] The α and β indications refer to the orientation of the nitrogen as drawn; for body type A, α = *cis* 3a*S*, β = *cis* 3a*R*; body B, β = *trans* 3a*S*, α = *trans* 3a*R*; body F, α = *R* (numbering conventions from original papers: a molecules,[1] b molecules,[2] and c molecules[3]). [b] K_i values are the first reported from the references except where noted. [c] K_i is from cloned CHO cells. [d] Mbz = (*R*)-α-methylbenzyl group. [e] Cpm = cyclopropylmethyl group. [f] This is (*n*) for body type F (**45** has one methylene).

molecule: the distance to the van der Waals surface, the distance to the nearest potential hydrogen-bond donor, and the distance to the nearest potential hydrogen-bond acceptor. This results in a set of 252 distance features. The hydrogen-bonding features have an associated directional component that is represented as a scalar strength between 0 and 1. The strength captures the degree to which a vector going from the origin to the reference point is coincident with the direction of the hydrogen-bond participant (see Figure 3). This creates a singular maximum for the possible interaction and eliminates degeneracies inherent in a pure distance feature. A constraint on a pure distance feature yields a spherical region

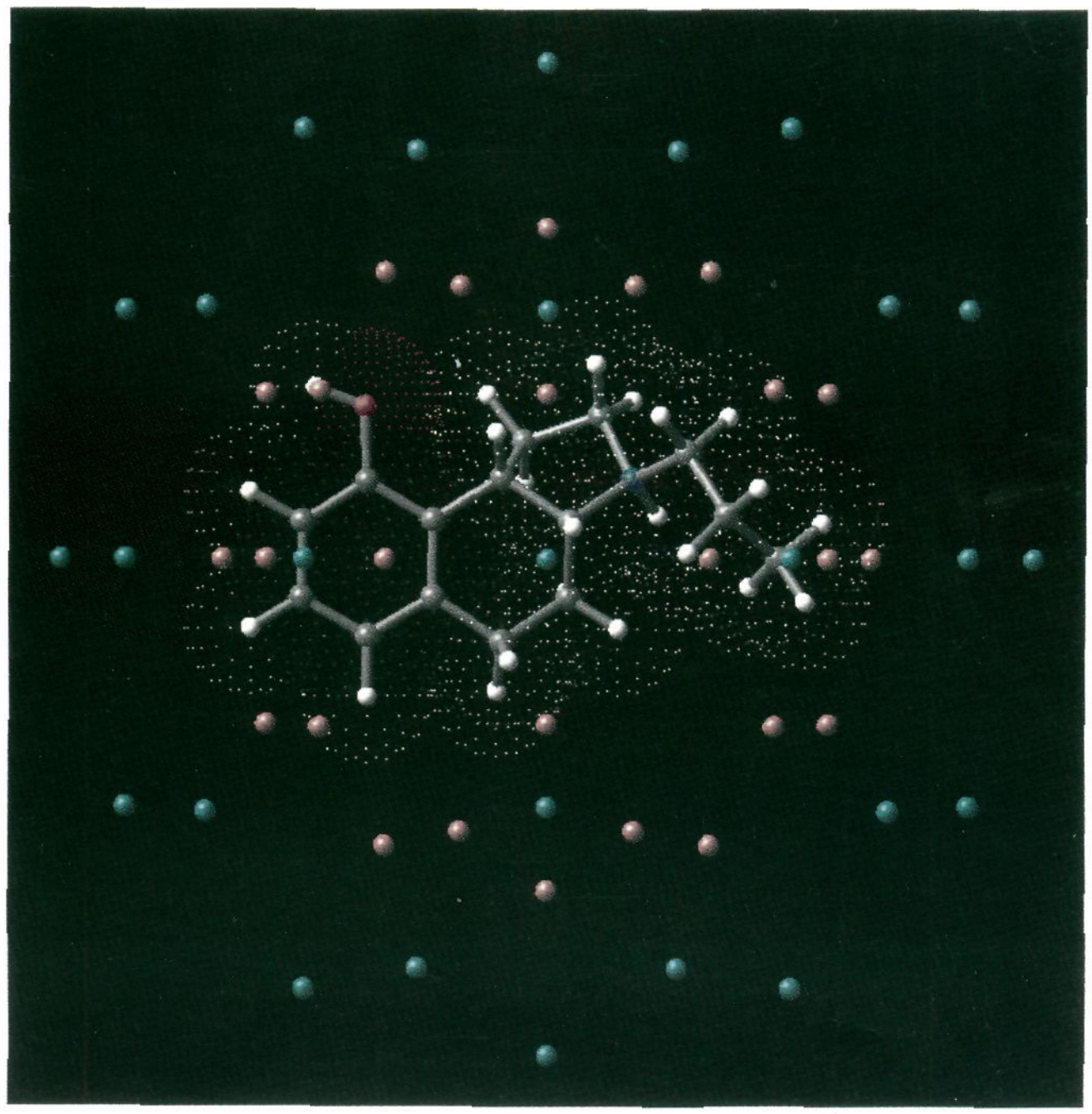

Figure 2. Molecule **4a** shown in feature reference points. The red spheres are at 6.0 Å and the blue at at 9.0 Å.

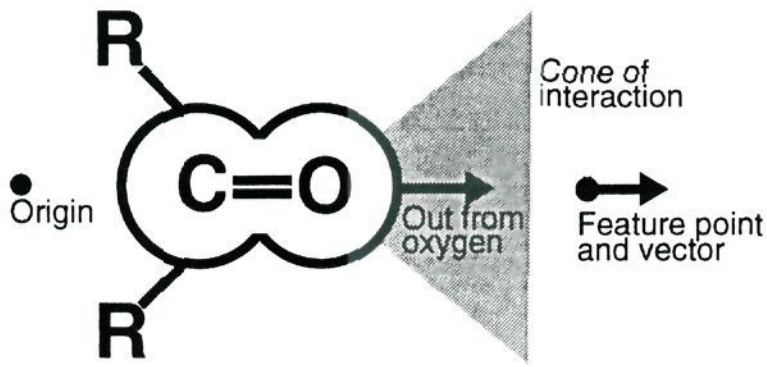

Figure 3. Directionality constraint in hydrogen bonding.

in space instead of a patch. The directional component restricts the portion of the sphere to a patch with a singular maximum. A more detailed accounting of lone-pair directionality has not been implemented. For uniformity, steric features have an associated constant strength of 1. The strengths are used to weight the functions of the feature values (more details can be found in the Experimental Section).

One can align the conformations of different molecules to maximize their mutual similarity according to Compass's features. A similarity measure m between some pose p and a reference pose r is defined as follows:

$$m(p,r) = \sum_i \left(\frac{e^{(-(f_i(p)-f_i(r))^2)/\sigma} s_i(p) s_i(r)}{\sum_j s_i(r)} \right)$$

A pose is specified by a conformation and six orientation parameters. The function f_i computes the ith feature value of a pose; the function s_i computes the strength of the feature. The exponential has a maximum when features from the poses correspond and falls off toward zero as they diverge. The parameter σ controls the steepness of the decay. Maximizing m by varying the orientation parameters of p aligns the shape and polar functionality of p to that exhibited by the reference pose r. The strength term in the numerator enforces a matching directionality constraint on potential hydrogen bonds. The sum in the denominator normalizes for the total strength of the reference pose, which may differ for different poses based on variations in hydrogen-bonding geometries relative to the feature reference points. The similarity function m is maximized using multiple initial starting alignments coupled with gradient descent on the derivative of m with respect to the orientation parameters of p (more details can be found in the Experimental Section). Note that when differences between molecular shapes are large, the contribution to the similarity function approaches zero, so parts of molecules may be allowed to protrude to maximize m if other areas come into correspondence.

Figure 4 shows mutual alignment of the molecules from Figure 1 computed to maximize m. The reference molecule displayed (gray body) is **4a**, which is the most active in the set. Panel A shows a conformation of **8b** in pink. The "bottom" of the left view is the "front" of the right view. Note that the proton on the nitrogen of **8b** comes into tight alignment with that of the reference molecule (marked with an arrow), as does the oxygen. Observing the positions of the hydrogens, it is clear that the shapes of the molecules are quite homologous. The alignment shown achieves a higher similarity score than that for the ring-based superpositions suggested by the two-dimensional reasoning alluded to earlier. Panel B shows **5a** with the reference molecule. In this case, as with **8b**, a flipped and rotated orientation maximizes the similarity measure.

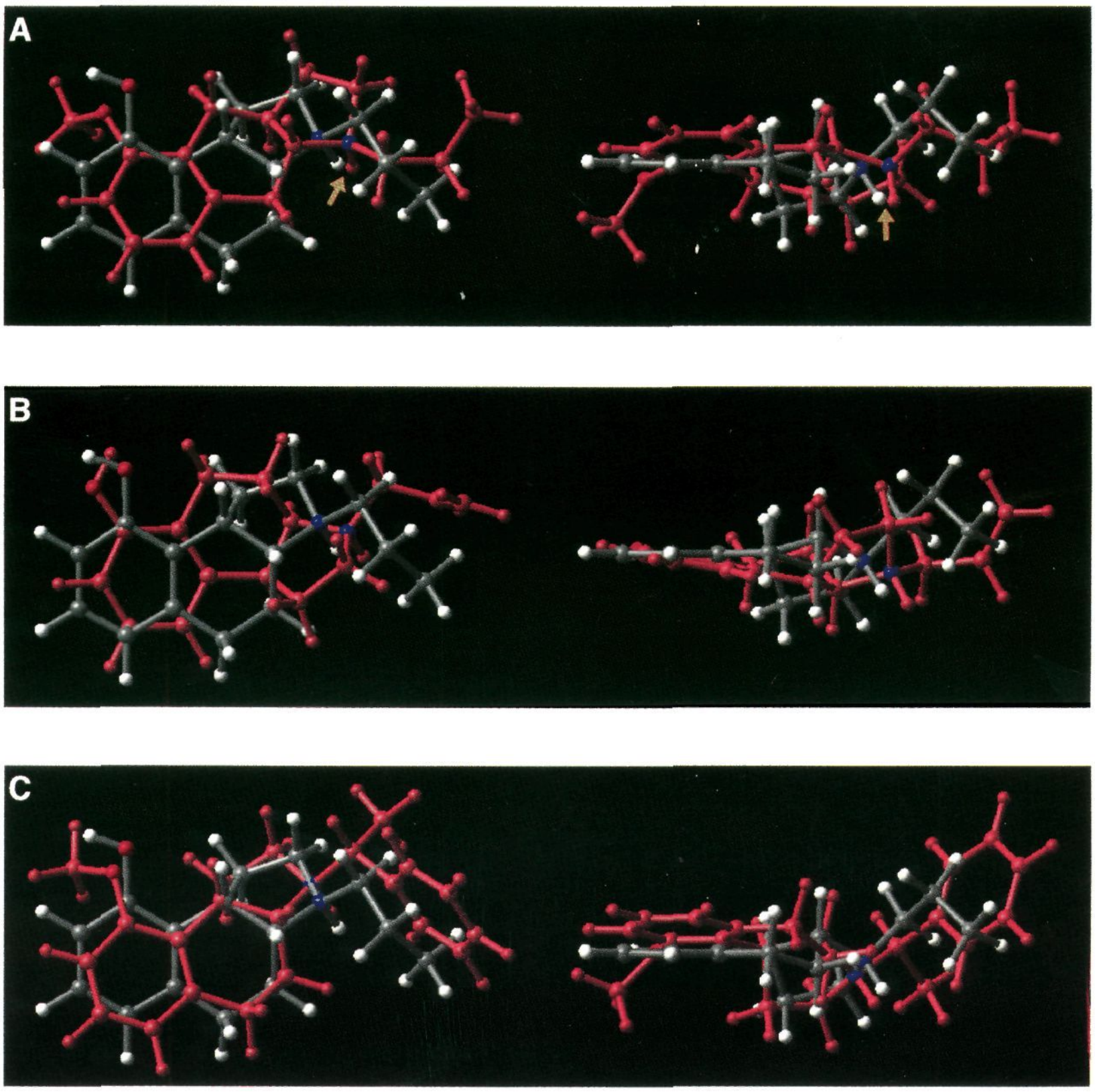

Figure 4. Alignment of molecules to **4a** (gray). Panel A shows **8b**, panel B shows **5a**, and panel C shows **6**.

Hydroxyl torsions were not searched, and the hydroxyls are not oriented in the same direction. This does not affect Compass's perception of the oxygens, since the acceptor distance features are not affected by occlusion. The orientation of the hydrogens does affect the feature computation, but we have assumed that the models will learn to accommodate a certain degree of variation of this type. An orientation of the molecules that superimposes the rings at the expense of misaligning the hydroxyls of **5a** and **4a** is a local maximum in the similarity measure. The last panel shows **6a**'s alignment to **4a**. Note that the pendant phenyl is allowed to protrude. Procedures such as rms minimization of differences between surface points would tend to shift the molecule as a whole. From an initial alignment produced using compass's measure, it is easy for the learning system to localize areas of difference in otherwise similar molecules, whereas for rms-type alignments the differences are spread more globally.

Bioactive Pose Hypothesis Based on Global Constraints. Given a potentially large pool of poses for each molecule (averaging 50 in this case), Compass must evolve a binding site model while it chooses the set of bioactive poses. A hypothesis for how chemotypes are related to one another that results in a parsimonious binding site model is preferable to a hypothesis that results in a complex disjunctive binding site model. In previous work, Compass relied on the emergence of such a hypothesis from a neural network with small initial weights.[8,9] Iteration of network refinement with pose choice was sufficient in those cases to arrive at simple nondisjunctive hypotheses for the bioactive poses. For cases in which either there is significant molecular flexibility or there are multiple plausible alignments of chemotypes, it is not possible to rely on the emergence of a hypothesis that accounts well for global constraints.

We have added a procedure that seeks to find a seed hypothesis for the bioactive conformations of a series of molecules by selecting a set of poses (one per molecule) that maximizes the joint similarity of all molecules to the most active ones. Poses for molecules are selected in decreasing potency rank. Once a pose is selected for a molecule, its other poses are ignored in this computation. For a molecule m, each of its poses is compared to the poses of all other molecules using the similarity measure defined earlier, and the best matching pose for each is marked. The match scores are summed, weighted by the potency of the matched molecules ($-\log K_i$). The pose of m that maximizes this sum is selected. The procedure continues until all molecules have a selected pose. This initial hypothesis of a bioactive pose set is used for the first few iterations of neural network training. After this initial reliance on a fixed set of poses, the entire pool of possible poses is made available to the neural network model for the remaining training process.

In practice, after model construction is complete, the poses

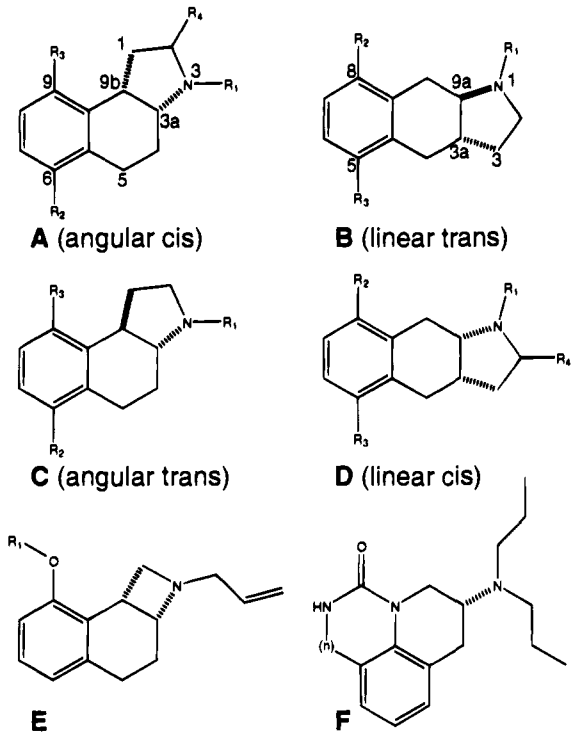

Figure 5. The six body types referred to in Table 1.

that maximize the model's output (the predicted bioactive poses) are different from those that resulted from the initial hypothesis, but they tend to conform to the basic shape of the seed hypothesis. When constructing multiple models from different subsets of molecules, the resulting models show stable behavior with respect to overall patterns of pose selection.

Structure–Activity Data. The structure–activity data used in this study was taken from three papers[1-3] and is presented in Table 1 (Figure 5 shows the body types of the molecules referenced in the table). The table is grouped into four sections: the 20 resolved compounds used to construct the Compass model (all from the A and B body types), racemic mixtures with the A and B body types, mixtures with C and D body types (which differ in ring fusion geometry from A and B), and molecules with body types E and F (which differ by more than ring fusion geometries). Following earlier work, all amine nitrogens were protonated and both α and β configurations were considered for asymmetric nitrogens. The protonation state of the molecules may be affected in binding to the receptor, but we assumed that this effect would consistently affect the compounds under consideration, hence our consistent treatment of protonation.

Most of the compounds were assayed for 5-HT_{1A} binding using bovine hippocampus, but for some, the assays were done using mammalian receptors expressed in cloned CHO cells. The assays are not completely commensurable, typically related by ±0.5 log units,[1,2] but this is a relatively minor source of noise since it only occurs for a small number of compounds. There is a more serious complication with the interpretation of the binding data for the mixtures. Given some racemic mixture in which one enantiomer is active (say 1 nM) and the other is not (say > 1000 nM), one would expect that the binding affinity of the mixture would be very close to 2 nM. Assuming independence of binding, the mixture affinity should always be between 1.0 and 2.0 times the affinity of the more active enantiomer.[18] Even for cases in which the resolved enantiomers are not pure, this ratio should hold. In such cases, the observed activity of the active enantiomer will be slightly shifted toward the mixture affinity, and the inactive enantiomer will be shifted (occasionally very significantly) toward the mixture affinity. Table 2 shows binding data for nine racemic mixtures along with their resolved enantiomers. Data are presented on compounds for which the K_i of the mixture and each enantiomer had better than 1 mM affinity. The ratio

Table 2. Assays of Resolved Compounds versus Mixtures

name	$isomer_1$[a]	$isomer_2$	mixture	mix./act.
1	0.2	49	1.2	6.0
2	0.2	186	4.4	22.0
3	236	333	252	1.1
4	0.1	3.0	1.4	14.0
10	36	470	62	1.7
11	56	470	105	1.9
3 (D_2)	25	339	42	1.7
5 (D_2)	49	259	198	4.0
10 (D_2)	2	917	17	8.5

[a] Affinities are nM.

Table 3. Comparison of Formulas for Predicting Mixture Affinities (Errors Shown in Brackets)

name	$isomer_1$[a]	$isomer_2$	log space	linear	actual
1	9.7	7.3	8.5 [0.4]	9.4 [0.5]	8.9
2	9.7	6.7	8.2 [0.2]	9.4 [1.0]	8.4
3	6.6	6.5	6.6 [0.0]	6.6 [0.0]	6.6
4	10.0	8.5	9.3 [0.4]	9.7 [0.8]	8.9
10	7.4	6.3	6.9 [0.3]	7.2 [0.0]	7.2
11	7.3	6.3	6.8 [0.2]	7.0 [0.0]	7.0
3 (D_2)	7.6	6.5	7.0 [0.4]	7.3 [0.1]	7.4
5 (D_2)	7.3	6.6	7.0 [0.3]	7.1 [0.4]	6.7
10 (D_2)	8.7	6.0	7.4 [0.4]	8.4 [0.6]	7.8

[a] All values are in $-\log K_i$ units.

of the mixture affinities to those of the active enantiomers ranges from 1.1 to 22.0.

It appears that the independence assumption breaks down in these data. Since the differences between the affinity of an enantiomer to its racemic mixture can be up to 1.3 log units, this raises a difficulty in evaluating the prediction of affinities for individual resolved isomers when only the racemate's binding affinity is known. However, there is an approximate empirical relationship between the isomer and mixture affinities. A simple formula that captures this phenomenon equates the predicted mixture affinity to the mean in log space of the individual isomer affinities. Table 3 shows the predictions resulting from this formula, compared with the linear formula, which asymptotes to 2.0 times the active isomer affinity as the disparity increases and is 1.0 when the affinities are equal.[18] The mean error in the predicted mixture affinity by the mean log-space formula is 0.3 log. The maximum error is 0.4 log. The other formula yields 0.4 log mean error and 1.0 log max error. In the following, except where noted, predicted mixture affinities are computed using the mean log-space formula. Raw data are also presented for predictions on the individual enantiomers of the novel compounds.

Results

Two types of experiments were conducted using models constructed with the 20 resolved compounds of A and B body types: internal cross-validation estimation of predictive ability and predictions on novel compounds of body types A–F. The following summarizes the numerical performance.

Internal Prediction Tests. Cross-validation was used to estimate the predictive performance of Compass using the compounds and associated binding data from the first section of Table 1. A 10-fold cross-validation was performed in which 10 Compass models were constructed, each by holding out a different pair of molecules. Figure 6 shows a log plot of predicted versus actual affinities resulting from the holdout tests for all 20 compounds. Table 4 shows the estimated predictive ability with three different metrics, which are presented for the full set of 20 molecules and without molecule **6a**, which was a significant outlier. It protruded more relative to the other compounds than any other holdout

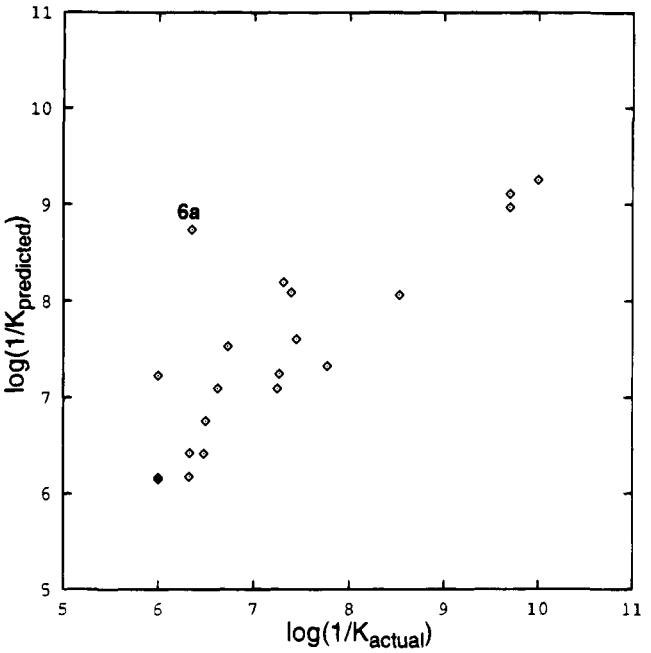

Figure 6. Plot of actual versus predicted affinities in cross-validation experiment.

Table 4. Cross-Validation Performance

	error[a]	PRCC[b]	PRCC[c]	cv r^2
all molecules	0.53	0.84	0.90	0.63
without **6a**	0.44	0.89	0.94	0.81

[a] log units. [b] $\Delta = 0$ log. [c] $\Delta = 0.5$ log.

Table 5. Performance on Novel Compounds

	n	error(log)	PRCC[a]	PRCC[b]	PRCC[c]
all	35	0.55	0.78	0.83	0.88
types A, B	13	0.59	0.73	0.77	0.85
types C, D	18	0.51	0.81	0.80	0.86
types E, F	4	0.72 [0.42][d]	1.00 [0.83]	1.00 [1.00]	1.00 [1.00]

[a] $\Delta = 0$ log. [b] $\Delta = 0.5$ log. [c] $\Delta = 1.0$ log. [d] Numbers in brackets were computed using the linear formula.

molecule. The mean absolute error of prediction was 0.53 log units (a factor of about 3 in binding affinity). The cross-validated r^2 was 0.63 (0.81 without **6**). The PRCC is the pair ranking correlation coefficient. It is the estimated probability that two novel molecules whose actual assays differ by more than Δ will be correctly ranked (discussed in more detail in the Experimental Section). The Δ factor allows for more realistic evaluation of performance given some amount of assay noise. The PRCC (0 log units) was 0.84, and with $\Delta = 0.5$, PRCC = 0.90. The result is highly statistically significant ($p < 0.001$ based on PRCC).

Novel Compounds. A Compass model was constructed using all 20 of the molecules from the first section of Table 1 and was used to predict the affinities of the remaining 35 compounds, none of which had been used in any manner in model construction. For cases in which the compounds were racemic mixtures, predictions were made for each enantiomer. Table 5 shows the numerical performance, both overall and broken down by groups of body types. Table 6 gives the raw predictions for the compounds. Predictions for the racemic compounds of body types A–D were computed by the mean log-space formula described earlier. For body types E and F, which differ significantly from the small set used to develop the mixture prediction formula, performance metrics were also computed using the linear formula and are given in brackets. Figure 7 shows a log plot of predicted versus actual affinities (the marked points will be discussed later).

Table 6. Prediced Affinities for Novel Compounds

name	isomer$_1$	isomer$_2$	predicted	actual
12	7.35	6.66	7.01	7.51
13	6.47	6.32	6.40	6.0
14	8.68	6.67	7.68	7.43
15	6.54	6.24	6.39	6.0
16	9.60	8.86	9.23	9.52
17	6.71	6.14	6.43	6.0
18	9.08	7.96	8.52	9.0
19	9.08	6.37	7.73	7.80
20	6.20	6.25	6.23	6.90
21	6.24	6.11	6.18	7.54
22	6.74	6.27	6.51	7.54
23	6.21	6.25	6.23	7.16
24	7.32	7.61	7.47	6.98
25	7.99	7.20	7.60	7.96
26	8.27	8.39	8.33	8.33
27	6.74	6.19	6.47	7.80
28	6.72	7.87	7.30	7.18
29	7.72	8.30	8.01	7.36
30	6.44	7.93	7.19	7.18
31	9.08	9.35	9.21	8.31
32	7.32	6.35	6.84	7.80
33	7.62	7.47	7.55	6.47
34	6.23	6.60	6.42	6.0
35	6.21	6.64	6.43	6.0
36	6.11	6.32	6.22	6.0
37	6.42	6.23	6.21	6.0
38	6.14	6.50	6.32	6.0
39	6.38	6.63	6.51	6.14
40	6.83	6.46	6.65	6.84
41	6.55	6.48	6.52	6.42
42	6.24	6.49	6.37	7.96
43	9.32	7.23	8.28 [9.02][b]	9.52
44	8.50	6.93	7.72 [8.30]	8.23
45	7.87	7.57	7.72 [7.87]	8.66
46	7.42	–	7.42	7.61

[a] All values are in log space. [b] The bracketed numbers were computed using the linear formula.

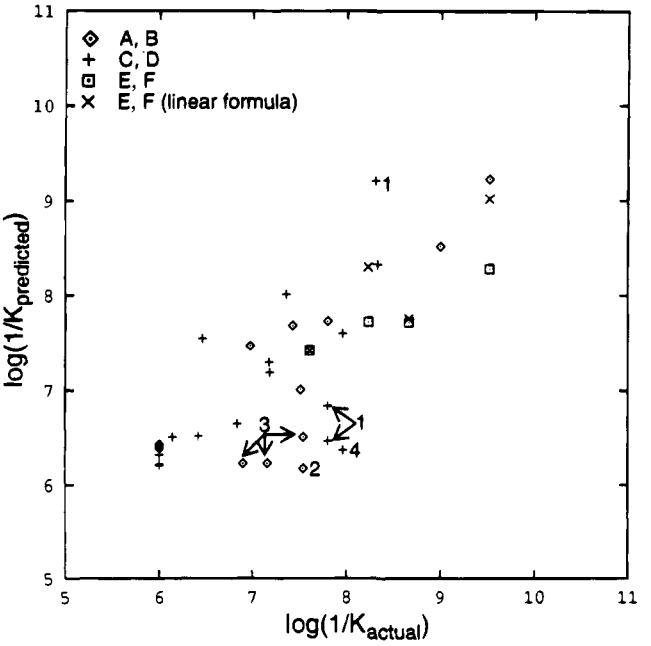

Figure 7. Plot of actual versus predicted affinities for the 35 novel compounds.

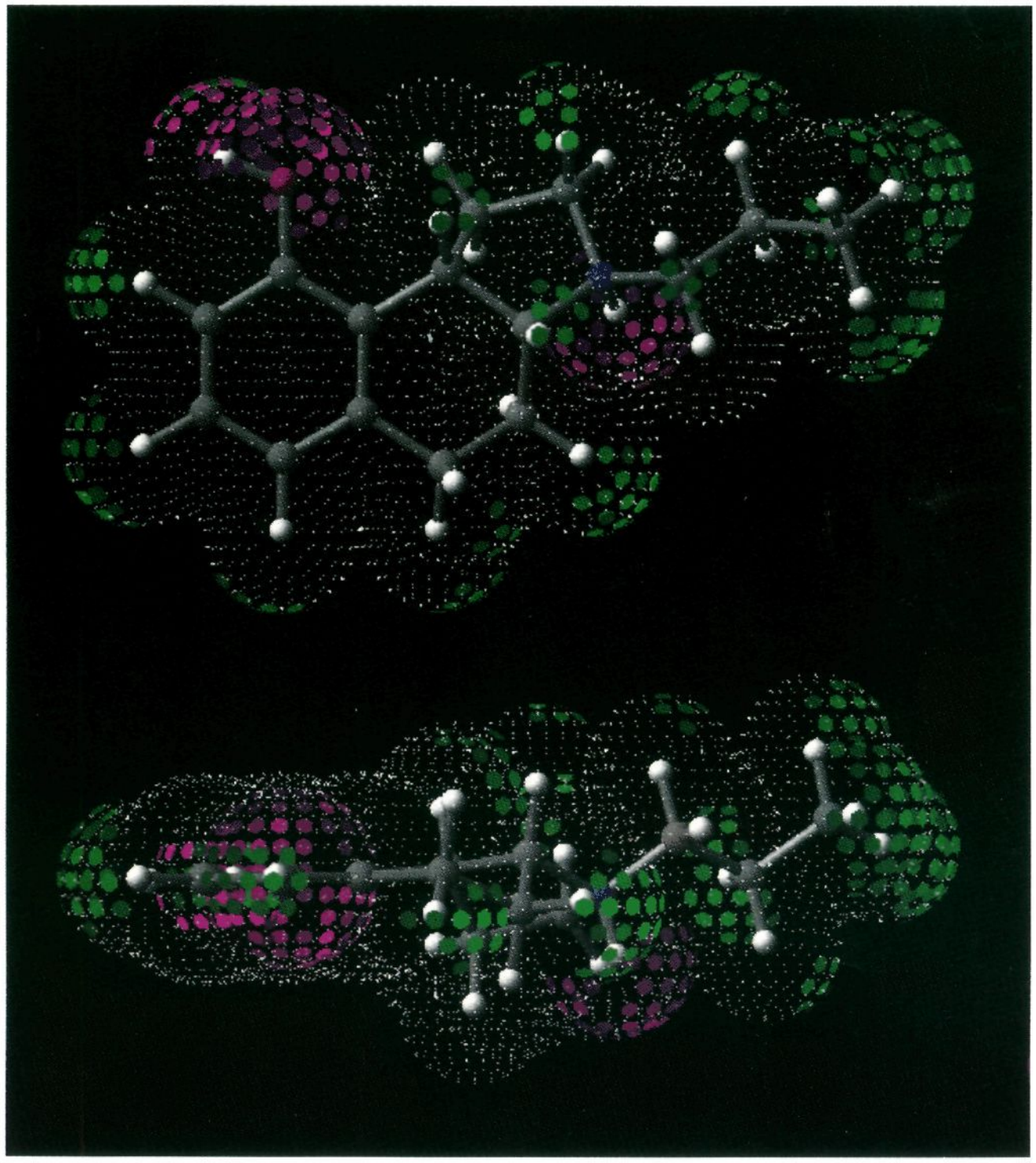

Figure 8. Molecule **4a** in Compass's 5-HT_{1A} quantitative binding site model.

The overall mean error of prediction was 0.55 log units, matching very well with the expectation from the cross-validation experiment. The performance is highly statistically significant ($p < 0.001$). The active enantiomers of the two most active compounds (**16** and **43**, from body types A and E, respectively) were predicted as such, and the top five predicted compounds, representing multiple chemotypes, include four out of five of the most active compounds.

Discussion

The central focus of this paper is to show that given a small quantity of high quality data, without using any crystallographic data and minimizing the time, cost, and potential bias of extensive analysis by hand, it is possible to build a predictive quantitative model of receptor−ligand interaction that extrapolates across chemotypes and produces insight into the structural requirements of productive binding. The foregoing shows that Compass is able to produce predictive models, but more careful analysis is required to ascertain what the model says about how the molecules bind and to what extent the model actually captures the specific binding determinants in this system and can be used productively in molecular design.

Visualizing the Model. Figure 8 shows the predicted bioactive conformation of **4a** with its interaction to Compass's binding site model painted on the molecular surface as spots. Green spots indicate favorable steric interactions, and purple spots indicate favorable polar interactions (spots on ligand hydrogens represent donor interactions and others represent acceptors). The brightness of the spots is computed by decomposing Compass's neural network model into its basis functions and calculating the contribution arising from each surface feature. The model judges all of the potential polar contacts to be significant and places significant emphasis on the steric contribution of the aliphatic N substituent. Note that little steric significance is ascribed to the "bottom" of the molecule (in the upper view).

Compass chose this conformation and alignment of **4a** by seeking to find bioactive poses of all active molecules that were similar to each other *and* were simultaneously different from the shapes that inactive molecules could adopt. This is essentially the same process that modelers of SAR must go through, but by being automated Compass can examine molecular relationships more rigorously. The choice of conformation was complicated by the nitrogen inversion phenomenon.

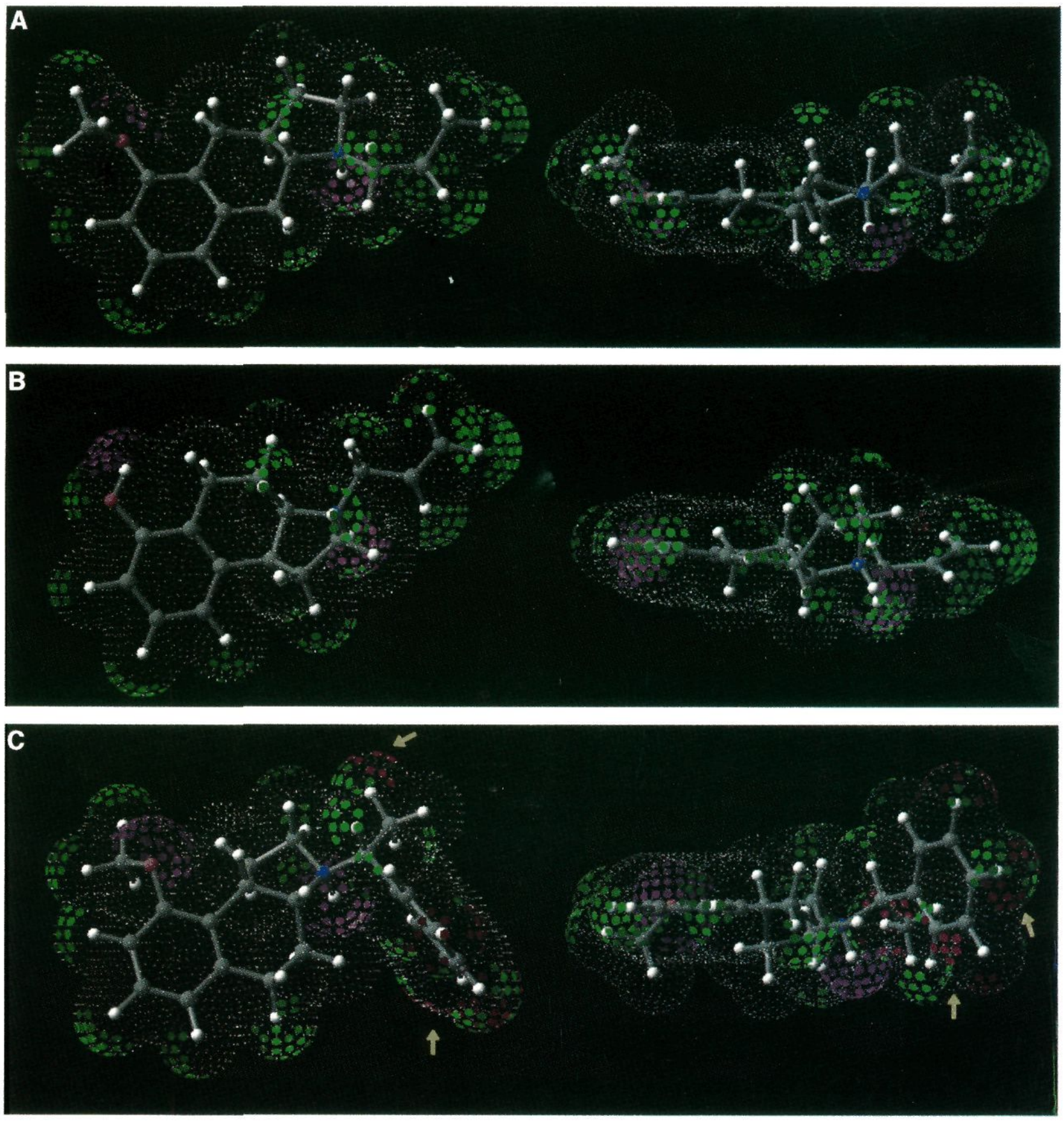

Figure 9. Molecules of different chemotypes with decreasing potency. Panel A shows **8b** (17 nM), panel B shows **5a** (54 nM), and panel C shows **6** (442 nM).

There was a choice as to which way the nitrogen's donor proton could face for all compounds with asymmetric nitrogens. Compass chose to place the proton α in the preferred conformation of the active molecules. It did this reliably under different model construction conditions (e.g., different subsets of training molecules). There is sufficient information in the small set of compounds, based on multiple chemotypes and orientations, that the consensus shape of active molecules is most consistent with this hypothesis. It is in agreement with prior work in the area,[3,6,7] but it was deduced based solely on structure–activity data.

Figure 9 shows three molecules in the same type of display as Figure 8. Panel A shows **8b**. The molecule is able to orient the protonated nitrogen appropriately, but only at the expense of pushing its oxygen too far to achieve a maximal positive contribution as seen in **4a**. It is also unable to make as favorable a set of steric contacts as **4a**. Still, the two molecules exhibit surprisingly similar shapes. Preference for this enantiomer of **8** is explained by the flipped orientation, which facilitates its ability to mimic the more active A body type. Panel B shows **5a**, the 6-hydroxyl analog of **4a**, which is predicted to bind in a flipped orientation, much the same as **8b**. It is less successful than **8b** at fitting the binding pocket. Panel C shows **6a**, the least active analog of **4a**. The red spots (highlighted by arrows) indicate an unfavorable steric protrusion. In an attempt to accommodate the excess steric bulk from the N substituent, the molecule tilts slightly and trades off improved steric fit with a suboptimal oxygen position. However, it is still unable to fit well.

Compass's quantitative binding site model has intuitive appeal, and it is consistent with the pharmacophoric model of the 5-HT_{1A} site of Chidester *et al.*[3] that was based on extensive modeling, analysis, and small mol-

ecule crystallography of a larger number of compounds than used here. The extent of the similarities is striking. However, Compass's model was based on the SAR of just 20 molecules and offers the ability to make *quantitative* assessment of novel compounds.

Novel Compounds

The experiment on the 35 compounds outside of the model construction set showed accurate predictions, with the mean predictive error of 0.55 log unit corresponding well to the expected value based on the cross-validation. This is somewhat surprising since it was necessary to make predictions on individual enantiomers to arrive at the final mixture predictions using a formula that in the average case is off by 0.3 log unit. Compass was able to find the most active molecules from multiple chemotypes.

However, numerical measures do not do a good job of assessing how well a model captures the various specific factors that govern binding in a particular site. Structural features in the new compounds differing from those present in the learning set help to assess this. There were no compounds in the model construction set with the following characteristics: (a) body types C, D, E, F; (b) no oxygen substituent; (c) no nitrogen substituent; (d) cyclopropylmethyl or benzyl substituents; (e) oxygen substituents on the same side as the nitrogen for the linear body type; and (f) substituents on the carbon adjacent to the nitrogen.

In fact, there were only 3 compounds of the 35 that had none of these novel structural features (all from body type A): **15**, **16**, and **17**. Their prediction errors were all very low (0.4, 0.3, and 0.4 log unit, respectively). Each of the novel structural features tests different aspects of the model; the following sections discuss their significance.

Body Types E and F. The ability to extrapolate to molecules with very different dispositions of chemical functionalities is the most difficult test of any method such as Compass. Figure 10 shows the predicted bioactive conformations of **43a** and **45a**. Molecule **43a** is similar to **4a** and its analogs and is predicted as such (0.5 nM). However, the allyl moiety of **43a** is presented at a significantly different angle than the aliphatic nitrogen substituents of molecules of the A body type, so the prediction is dependent on Compass's ability to systematically select molecular poses and assess shape without being substituent-based.

Molecules **45** and **46** are from a very different chemotype, and they begin to strain Compass's ability to extrapolate, with **46** predicted well, but **45** predicted somewhat low. Molecule **45a** (the more active enantiomer) was predicted to be 13 nM and was reported to be 2.2 nM (for the mixture). In the graphical display of the model, the molecule fits quite well. Both the protonated nitrogen and the carbonyl are able to satisfy the requirements learned by Compass. Interestingly, the urea proton is able to partially serve the function of the hydroxyl proton of molecules such as **4a**. The corresponding proton of **46** is unable to do this due to the smaller ring size changing its angle of presentation (not shown). Because **45a** is so different than the molecules from the model set, there are places where this molecule begins to encounter areas for which the model has learned insufficiently general constraints. The yellow area on the back side of the carbonyl indicates a weak unfavorable polar contact caused by tilting of the molecule back to accommodate its bulk. The red spots on one of the propyls are protruding into another exclusion area. Despite the somewhat low numerical prediction for **45**, this is a success, especially when one considers the intuitive fit to the model that is displayed.

Propyl versus Allyl Substituents. In several cases, molecules differing by only the substitution of a propyl for an allyl show markedly worse affinity. This effect has been previously explained qualitatively by a difference in how well the allyl group fits the receptor when it is presented on different scaffolds.[3] Compass's hypothesis about the bioactive conformations of the molecules yields a physical three-dimensional basis for this effect.

The only compounds that exhibit the above phenomenon from the model construction set are the pair **1b**/**2b**. This two-atom subtlety is only evidenced when an *n*-propyl group is forced into an unfavorable approach by the combination of the underlying scaffold along with its substituents. In order to see this effect with body type A, one needs to have a methoxy and an *n*-propyl on the β enantiomer. With the hydroxyl instead (**4a**/**4b**), the less active isomer is still able to fit quite well, at 3 nM for **4b** versus 186 for **2b**. With just this single datapoint, Compass was unable to deduce the precise binding constraints that account for the propyl/allyl phenomenon in all cases.

However, for molecules of body type A, Compass correctly detected the cases in which allyl/propyl differences were significant. Molecule **17** (1000 nM), the *n*-propyl derivative of training compound **5a** (54 nM), was predicted very well (just 0.4 log unit error). Molecule **15** (1000 nM), the *n*-propyl derivative of training compound **3a** (236 nM), was predicted equally well. Note that each of these was from the "flipped" orientation of body type A, whereas the training example was of the standard variety. Molecule **16**, the very potent allyl derivative of **4a**, was also predicted accurately.

The allyl/propyl distinction was not learned as well on body type C (the corresponding points are marked with a 1 in Figure 7). Molecules **27** and **32** (each having allyl substituents) were both predicted low. Both of these are predicted in flipped orientations, which places their oxygen substituents in good alignment with the most active of the molecules. Molecule **32** (hydroxyl substituted) has its more active enantiomer predicted quite well, just 0.5 log unit lower than actual for the mixture. The less active isomer is judged to present its allyl substituent poorly, resulting in the low overall prediction. Molecule **27**, the methoxy analog, shows the same pattern, but both enantiomers are predicted lower, since the extra methyl further exacerbates the poor allyl fit. In these two cases, Compass's placement of constraints to explain the poor activity of training molecule **2b** resulted in low predictions. Molecule **33** shows the opposite effect—it is predicted high. It has an *n*-propyl that is supposed to cost more than 1 log unit compared to its allyl analog (**32**).

Figure 11 shows the predicted bioactive conformations of the three subnanomolar compounds **4a**, **2a**, and **16a** (gray), **33b** (purple), and **32b** (yellow). Note that one of the active compounds is an allyl analog and two are

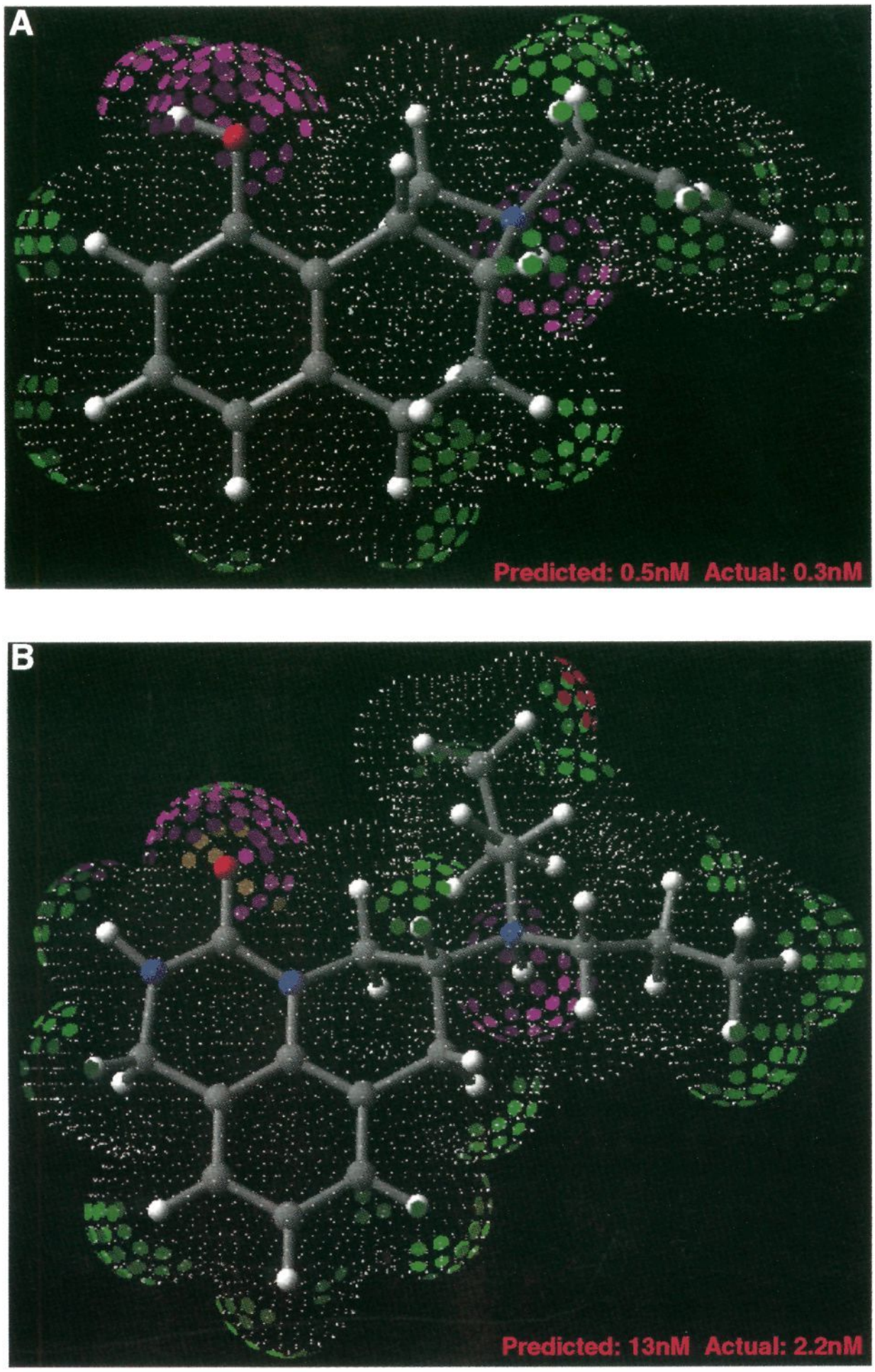

Figure 10. Potent novel compounds of body types E and F. Panel A shows **43a**, and panel B shows **45a**.

n-propyls. Despite the errors in affinity prediction, Compass's hypotheses for the bioactive conformations of the molecules yield an explanation for the differences in affinities. Molecule **32b**, the allyl analog (predicted slightly low), places its N substituent directly in the middle of the space occupied by the most active molecules, but it extends slightly (highlighted in yellow) beyond the optimal region. Molecule **33**, the *n*-propyl derivative, was predicted incorrectly to be more active than **32**. Its *n*-propyl exceeds the boundaries (highlighted in purple) exhibited by the very active molecules. Compass did not discover the exact constraints in the highlighted regions. Molecule **30**, also of body type C (the *n*-propyl analog of **27**), was predicted correctly to be moderately active.

No Oxygen Substituent. In order to accurately predict compounds lacking an oxygen substituent, the system needed to deduce how much the oxygen functionality was "worth" based only on differential activities exhibited by compounds in the model construction set displaying different oxygen positions. The most active of the new molecules lacking an oxygen substituent was of body type C (**25**). It was predicted within 0.4 log of actual. The mean error on the four molecules lacking oxygen substituents was 0.45 log, with a maximum of 0.6. Despite lacking a pure example of oxygen deletion,

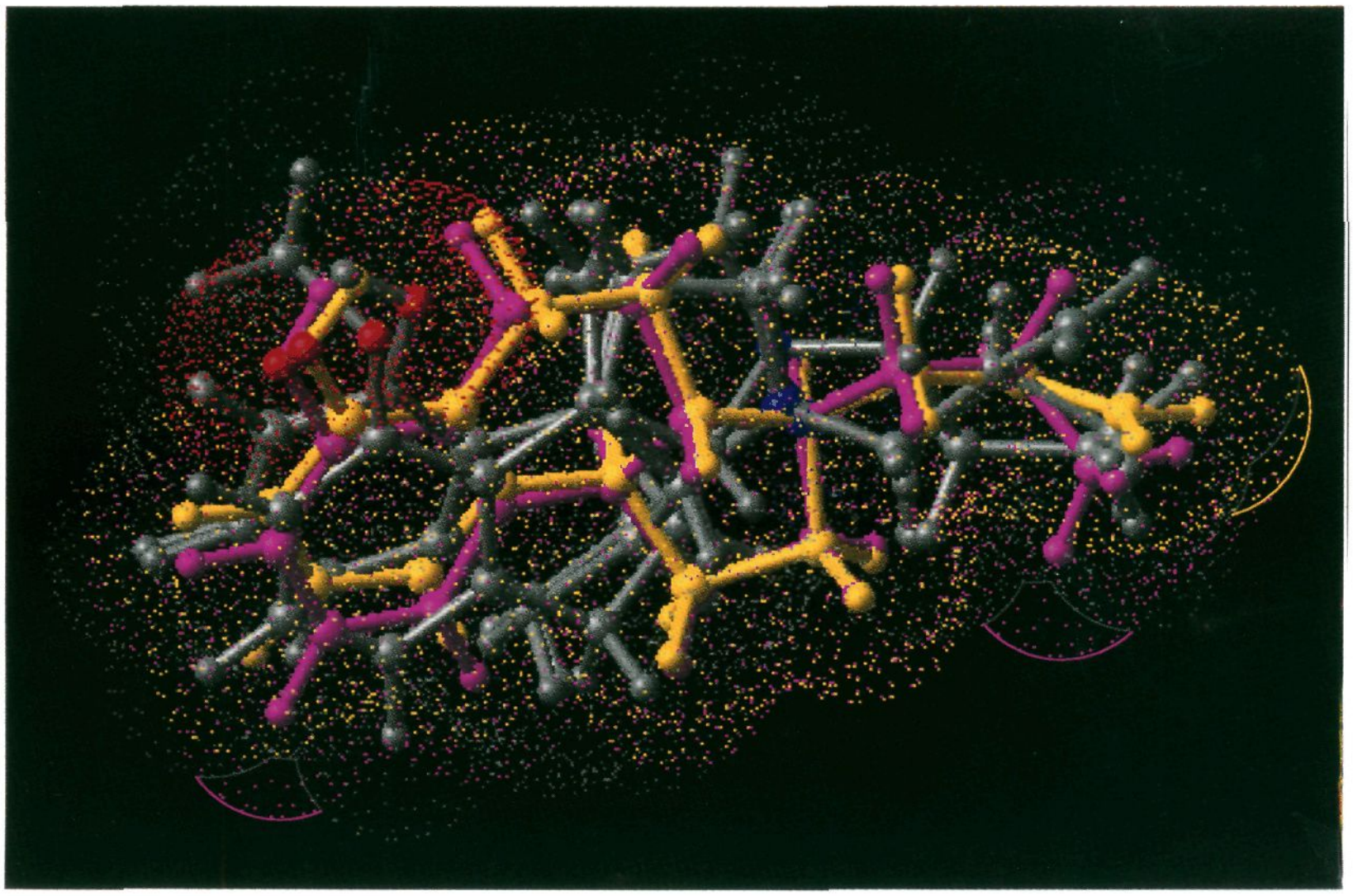

Figure 11. Propyl vs allyl comparison. Molecules **4a**, **2a**, and **16a** shown in gray; molecule **33b** shown in purple; molecule **32b** shown in yellow.

Compass was able to estimate the contribution of this substituent on the K_i quite well, ranging from very little (0.5 log unit) when presented on body type C to nearly 2 log units on body type A.

No Nitrogen Substituent. All of the compounds in the data set contained a nitrogen substituent of at least the size of an allyl group. From the discussion above, one sees that potency is highly dependent on the precise shape presented in that part of space. The constraints in that part of space are more subtle than the presence or absence of the acceptor functionality discussed above. Both **38** and **21** are predicted to be inactive (**21** is marked with a 2 in Figure 7). Surprisingly, **21** shows good potency. Figure 12 shows alignments of **21b** and **8b** to **4a** (all in their predicted active conformations). Panel A shows **21b** (purple) and **4a** (pink), panel B shows **8b** (yellow) and **4a**, and panel C shows all of them. Note that **21b** and **8b** show significantly different alignments. Since **21b** is unencumbered by constraints from an N substituent, it aligns its polar functionality directly with **4a**. Conversely, **8b** must accommodate its N substituent by shifting its alignment.

In accounting for the activity of the compounds, it appears that Compass has placed too much emphasis on the end of the aliphatic group and not enough on the middle region. This is consistent with the earlier discussion of Compass's predictions involving the allyl vs propyl effect. The molecule **21b** shows tight alignment to **4a** except for **4a**'s methyl proximal to position 2 of **21b**. Substitution at position 2 of **21b** may yield potent compounds of the linear type by taking advantage of both steric shape complementarity near the nitrogen and the polar contact made by the oxygen.

Other Variations on A–D. Cyclopropylmethyl substituents are slightly larger than the *n*-propyl substituents that are common in the series. For body types A and C, the cpm-substituted compounds (**13** and **28**) are predicted well. The constraints placed on the cleft into which the N substituents fit are easier to achieve with these body types (as evidenced by the forgoing discussion). But for body type B, the two compounds with cpm substituents as well as the benzyl-substituted compound are predicted significantly low (marked with a 3 in Figure 7). There is only one linear tricyclic compound in the modeled set (**8b**) with activity better than 20 nM, and it appears that the constraints learned to accommodate its *n*-propyl presentation are not sufficiently general (and are likely confounded with the allyl/propyl phenomenon). Note that these low predictions are not terribly serious since the compounds are not particularly active. The cpm-substituted body type D molecule was predicted well. All of the body type D molecules with an R_3 substituent are essentially dead and are predicted as such.

There were four compounds with position 8 (R_2 in Figure 5) substituents on the linear body types. They were are all predicted in the "unflipped" orientation, and all but **42** were predicted well. Molecule **42** (marked with a 4 in Figure 7) has three novel characteristics compared with the compounds used for model construction: body type D, 8-methoxy substituent, and a 2-methyl substituent. The methyl substituent affects its potency very dramatically, and this is not accounted for by the model. This type of error is reasonable, and it should be possible to flag such cases automatically by examining the space that is well-covered by the data that support a model. By explicitly identifying unexplored space, it would be possible to be more systematic in probing a binding cavity for favorable interactions.

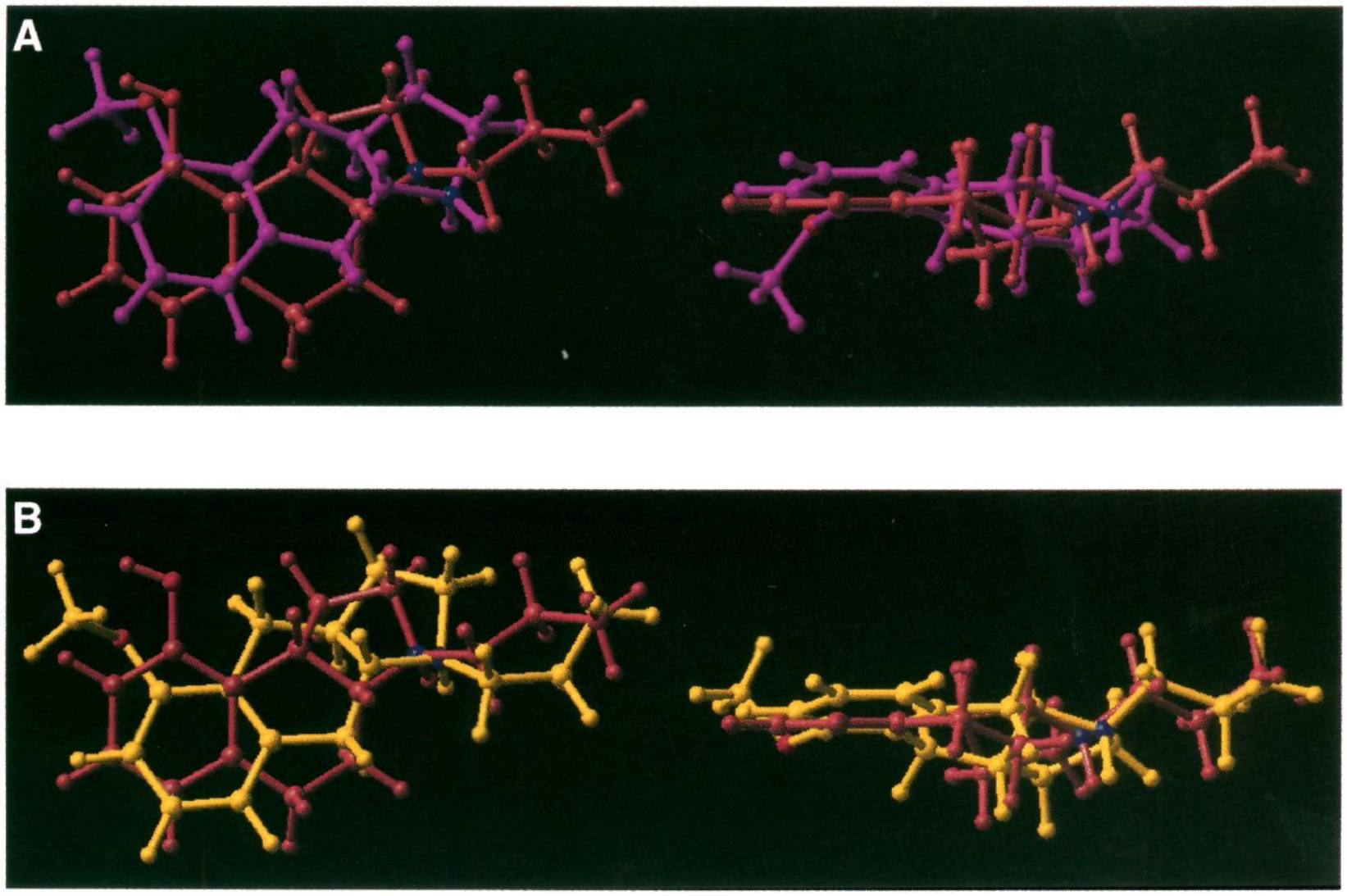

Figure 12. Effect of nitrogen substituent on body type B. Molecule **21b** (29 nM) shown in purple (panels A and C), molecule **8b** (17 nM) shown in yellow (panels B and C), and molecule **4a** (0.1 nM) shown in pink (all panels).

Conclusion

Pharmacophoric models of the 5-HT_{1A} receptor developed previously have been based on extensive conformational and crystallographic studies of ligands to gain insight into their potential bioactive states and interactions with the receptor.[3,6,7] Analysis of this type yields some quantitative data (e.g., specific distance constraints from a proton donor to a ring) along with a wealth of qualitative insight (e.g., "a group larger than an *n*-propyl will not fit"). With relatively sparse data and without requiring crystallography, Compass can construct accurate quantitative binding site models that replace qualitative insight with the quantitative constraints of ligand–receptor interactions. The full process, from entering chemical structures to generating the model, took approximately 1 week, with most of the time spent in conformational search of the molecules. New molecules can be screened in hours. With the ability to generalize to novel scaffolds, display the structural requirements for binding, and show the geometrical relationship between chemotypes, a tool such as Compass can speed the drug discovery process.

Compass's representation of molecules and treatment of the conformational problem is quite different from that employed by others who have modeled 5-HT_{1A} ligands.[3,6,7] The conformational issue is difficult. In cases of very rigid compounds, many types of analysis are sufficient. But when flexibility comes into play, relying on small-molecule crystallography or on detailed molecular mechanics computations to yield insight into bioactive conformations of ligands can be misleading. Ligand–receptor interactions may easily dominate weak torsional energetic barriers and result in quite different conformations from those observed crystallographically or in near-minimum computed energies. Compass's approach of deducing binding constraints through rigorous analysis of conformational space and structure–activity data results in predictive, intuitive models.

Pharmacophoric models have characterized the detailed position and presentation angle of functional groups relative to an aromatic ring.[3,6,7] Compass is able to produce binding constraints without directly modeling ring positions—an aromatic ring is simply a shape to Compass. Molecules with different aromatic structures from modeled molecules (e.g., the F series) are predicted well. Treatment of molecules based on surface shape facilitates this type of cross-chemotype extrapolation. Since it is possible to decompose the Compass model and paint the surface of molecules, visualizing the correspondence between chemotypes can be productive in design.

There are many areas for future research. Greater diversity in chemotypes than was considered in this study may necessitate explicit modeling of the effects of differing desolvation energies and relative differences in hydrophobicity. Other areas of focus will include refining the molecular representation to account for more types of interactions, developing automatic methods for identifying unexplored space in binding cavities, and improving the learning algorithms. Another clear direction will involve development of direct models of selectivity between receptors. For cases in which there are compounds that show marked selectivity, such as the 5-HT_{1A} versus D_2 receptor, one can get an idea about the basis for selectivity by building separate models and comparing them. However, in cases for which only weakly selective compounds have been generated, building direct computational models of selectivity that tie receptor preference to molecular shape will be required.

Experimental Section

Initial Alignment. The conformations of **4a** were aligned to each other using the aromatic rings so that the oxygens would correspond. One low energy conformation of each of the inversion isomers of **4a** was chosen (by eye) such that the *n*-propyl groups were in extended conformations. These were placed in the Cartesian frame such that their centroids were at the origin. This could be automated using a procedure that identifies canonical conformations based on clustering, but it was not necessary in this case. Each of the conformations of the remaining molecules was aligned to these two seed conformations, beginning from two alignments that superimposed the six-membered rings. Starting from multiple random starting orientations produces similar results without requiring the structure-based alignment but is more expensive computationally. For each conformation, the alignment with the best similarity score was retained.

In the similarity measure, the σ parameter was set to 2.5 (feature value quantities are measured in Angstroms). There are 42 reference points of each of the three types (van der Waals distance, acceptor surface distance, and donor surface distance) uniformly placed on a sphere of radius 6.0 Å centered at the origin of the Cartesian frame. There are also 42 reference points of each type placed on a sphere of radius 9.0 Å with the same center. This results in a total of 252 feature reference points. Neither the density nor the radii of the reference point spheres were chosen in a systematic way, and no experiments were performed to test the effects of variation. The atomic radii (Å) used in this study were hydrogen 1.2, carbon (aromatic) 1.7, carbon (nonaromatic) 1.6, nitrogen 1.5, and oxygen 1.4.

Generating initial alignment hypotheses for all conformations of all molecules using a gradient-based procedure can be expensive. Compass's procedure provides several levels of control to trade off time versus thoroughness. One can choose the number of molecules to be used as references (conformations of these are aligned to each other in descending potency rank with the best alignments retained), the number of conformations of each reference molecule that will be used, and the relative threshold for retaining alignment hypotheses for conformations. There are several methods for visualizing groups of matching molecular alignments that help to choose the necessary depth of the search.

Feature and Network Computations. The diagram in Figure 3 is slightly simplified from the actual computation. The polar feature strength is computed from three vectors normalized to unit length: the feature point vector, the preferred hydrogen-bond direction, and the vector pointing from the hydrogen-bonding atom to the feature point. The dot products of the feature direction vector with each of the other two are multiplied, resulting in a scalar denoted q. In the case of perfect coincidence, q is maximized at 1. As the vectors diverge, q gets smaller. The strength function is $1/(1 + e^{10.0(q-0.3)})$. As the vectors become coincident, the strength approaches 1, and as the degree of divergence approaches 70°, the strength drops to 0.5. No extensive experiments have been performed to tune the parameters of the strength function. The basis functions of the neural network have been augmented to include an additive sigmoidal component: $z/(1 + e^{(x-\mu)/\sigma})$. Parameters z, μ, and σ are tunable independent of the parameters for the Gaussian,[8] with the restriction that the μ of the sigmoid be less than that for the Gaussian.

Conformational Search. All molecules were conformationally searched systematically with three samples per rotatable torsion. The resulting conformations were minimized by gradient descent using an implementation of Macromodel's MM2 force-field.[19] All conformations within 5 kcal/mol of the minimum were retained. Note that for nitrogen inversion isomers, each set of conformations was subject to its own minimum. This allowed Compass complete choice as to orientation of the donor protons.

Pair Ranking Correlation Coefficient. The PRCC is a nonparametric metric that evaluates the quality of the ranking of predicted data without the squared rank penalty in the SRCC and with a more intuitive treatment of "ties" than is found in Kendall's τ.[20] The PRCC (Δ) measures the probability that two molecules whose assays differ by Δ (in log space) will be ranked correctly. It is evaluated under cross-validation by comparing predictions for all pairs of molecules whose actual assays differ by more than Δ, counting the number of correctly ranked molecules, and dividing by the total number of comparisons. In cases of no ties (all molecules differ by more than Δ), the measure is linearly related to Kendall's τ. The Δ factor allows for more realistic evaluation of performance given some amount of assay noise.

Acknowledgment. We thank David Chapman for pointing out the data set and for productive discussions about representation of polar features and mathematical characterization of binding sites, Tomas Lozano-Perez for efforts in conformational search and comments on the manuscript, Tom Jenkins and Teri Klein for comments on the manuscript, and Mike Ross for corporate support and guidance.

References

(1) Lin, C.; Haadsma-Svensson, S. R.; Lahti, R. A.; McCall, R. B.; Piercey, M. F.; Schreur, P. J. K. D.; Von Voigtlander, P. F.; Smith, M. W.; Chidester, C. G. Centrally Acting Serotonergic and Dopaminergic Agents. 1. Synthesis and Structure-Activity Relationships of 2,3,3a,4,5,9b-Hexahydro-1H-Benz[e]indole Derivatives. *J. Med. Chem.* **1990**, *36*, 1053–1068.

(2) Lin, C.; Haadsma-Svensson, S. R.; Phillips, G.; Lahti, R. A.; McCall, R. B.; Piercey, M. F.; Schreur, P. J. K. D.; Von Voigtlander, P. F.; Smith, M. W.; Chidester, C. G. Centrally Acting Serotonergic and Dopaminergic Agents. 2. Synthesis and Structure-Activity Relationships of 2,3,3a,4,9,9a-Hexahydro-1H-Benz[*f*]indole Derivatives. *J. Med. Chem.* **1993**, *36*, 1069–1083.

(3) Chidester, C. G.; Lin, C.; Lahti, R. A.; Haadsma-Svensson, S. R.; Smith, M. W. Comparison of 5-HT_{1A} and Dopamine Pharmacophores, X-ray Structures and Affinities of Conformationally Constrained Ligands. *J. Med. Chem.* **1990**, *36*, 1301–1315.
(4) Liljefors, T.; Wikström, H. A Molecular Mechanics Approach to the Understanding of Presynaptic Selectivity for Centrally Acting Dopamine Receptor Agonists of the Phenylpiperidine Family. *J. Med. Chem.* **1986**, *29*, 1896–1904.
(5) Liljefors, T.; Bøgesø, K. P.; Hyttel, J.; Wikström, H.; Svensson, K.; Carlsson, A. Pre- and Postsynaptic Dopaminergic Activities of Indolizidine and Quinolizidine Derivatives of 3-(3-Hydroxyphenyl)-*N*-(*n*-Propyl)Piperidine (3-PPP). Further Developments of a Dopamine Receptor Model. *J. Med. Chem.* **1990**, *33*, 1015–1022.
(6) Hibert, M. F.; Gittos, M. W.; Middlemiss, D. N.; Mir, A. K.; Fozard, J. R. Graphics Computer-Aided Receptor Mapping as a Predictive Tool for Drug Design: Development of Potent, Selective, and Stereospecific Ligands for the 5-HT_{1A} Receptor. *J. Med. Chem.* **1988**, *31*, 1087–1093.
(7) Mellin, C.; Vallgårda, J.; Nelson, D. L.; Björk, L.; Yu, H.; Andén, N.-E.; Csöregh, I.; Arvidsson, L.-E.; Hacksell, U. A 3-D Model for 5-HT_{1A} Receptor Agonists Based on Stereoselective Methyl-Substituted and Conformationally Restricted Analogs of 8-Hydroxy-2-(Dipropylamino)Tetralin. *J. Med. Chem.* **1991**, *34*, 497–510.
(8) Jain, A. N.; Koile, K.; Chapman, D. Compass: Predicting Biological Activities from Molecular Surface Properties. Performance Comparisons on a Steroid Benchmark. *J. Med. Chem.* **1994**, *37*, 2315–2327.
(9) Jain, A. N.; Dietterich, T. G.; Lathrop, R. H.; Chapman, D.; Critchlow, R. E.; Bauer, B. E.; Webster, T. A.; Lozano-Perez, T. Compass: A Shape-Based Machine Learning Tool for Drug Design. *J. Comput. Aided Mol. Des.* **1994**, 8.
(10) Sheridan, R. P.; Nilakantan, R.; Dixon, J. S.; Venkataraghavan, R. The Ensemble Approach to Distance Geometry: Application to the Nicotinic Pharmacophore. *J. Med. Chem.* **1986**, *29*, 899–906.
(11) Hopfinger, A. J. A QSAR Investigation of DHFR Inhibition by Bakers Triazines Based Upon Molecular Shape Analysis. *J. Am. Chem. Soc.* **1980**, *102*, 7196–7206.
(12) Marshall, G. R.; Barry, C. D.; Bosshard, H. E.; Dammkoehler, R. A.; Dunn, D. A. The Conformational Parameter in Drug-Design: The Active-Analog Approach. In *Computer-Assisted Drug Design*; Olsen, E. C., Christoffersen, R. C., Eds.; American Chemical Society: Washington, DC, 1979.
(13) BioCAD Corp. *Hypotheses in Catalyst*, 1992. BioCAD Corp. *Conformational Analysis in Catalyst*, 1992.
(14) Ghose, A. K.; Crippen, G. M. Use of Physicochemical Parameters in Distance Geometry and Related Three-Dimensional Quantitative Structure-Activity Relationships: A Demonstration Using *E. coli* Dihydrofolate Reductase Inhibitors. *J. Med. Chem.* **1985**, *28*, 333–346.
(15) Cramer, R. D. III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
(16) Good, A. G.; So, S.; Richards, W. G. Structure-Activity Relationships from Molecular Similarity Matrices. *J. Med. Chem.* **1993**, *36*, 433–438.
(17) Dietterich, T. G.; Jain, A. N.; Lathrop, R. L.; Lozano-Perez, T. A Comparison of Dynamic Reposing and Tangent Distance for Drug Activity Prediction. In *Advances in Neural Information Processing Systems 6*; Cowan, J. D., Tesauro, G., Alspector, J., Eds.; Morgan Kaufmann: San Francisco, CA, 1994; pp 216–223.
(18) Casy, A. F. Stereochemical Aspects of Parasympathomimetics and their Antagonists: Recent Developments. In *Progress in Medicinal Chemistry*; Ellis, G. P., West, G. B., Eds.; North-Holland Publishing Co.: Amsterdam, 1975; pp 1–65.
(19) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R. M. J; Lipton, M. A.; Caufield, C. E.; Change, G.; Hendrickson, T.; Still, W. C. MacroModel-an Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics. *J. Comput. Chem.* **1990**, *11*, 440–453.
(20) Snedecor, G. W; Cochran, W. G. *Statistical Methods*; Iowa State University Press: Ames, IA, 1989.

JM940644F