2315

Compass: Predicting Biological Activities from Molecular Surface Properties. Performance Comparisons on a Steroid Benchmark

Ajay N. Jain, Kimberle Koile,[†] and David Chapman[*]

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

We describe a new method, Compass, for predicting the biological activities of molecules based on the activities and three-dimensional structures of other molecules. The method improves on previous techniques by representing only the surface of molecules, by incorporating a nonlinear statistical method, and by automatically choosing conformations and alignments of molecules. We use a benchmark problem of steroid binding affinity prediction to compare the performance of the method with that of two previous systems: CoMFA and a molecular similarity method. Compass predicts steroid affinities substantially more accurately than the others, which represent the state of the art. We present experiments showing that the improved performance depends on each of the technical innovations.

Introduction

Drug discovery proceeds largely by trial and error. Typically, thousands of compounds are synthesized for each that finally becomes a drug. Each synthesis costs, on average, thousands of dollars and a few days to a few weeks of effort. This makes drug discovery tremendously expensive, and much slower than we would like. Current practice is unable to keep up either with the explosively increasing understanding of the biological processes underlying disease states, or with the number of potential targets for pharmaceutical intervention. Accurately predicting the biological activity of hypothetical molecules and understanding the basis of those predictions would make the process more productive.

Two approaches to prediction have been taken: *structure-based* methods and *QSAR* methods. Structure-based methods start with the three-dimensional structure of a protein target. This protein may be an enzyme or a receptor that has been implicated in a disease process. The goal is to inhibit the enzyme or to agonize or antagonize the receptor (that is, to induce or prevent its signaling function). The strategy, in all three cases, is to find a ligand molecule that will bind to the protein's active site. Among the factors that bind molecules together are van der Waals interactions, the hydrophobic effect, and electrostatics. The first two factors favor ligands whose shape is complementary to that of the target's active site, in the way a key fits a lock. The third favors ligands whose polar functionalities are complementary to those adjacent in the active site. If the overall shape of the target protein and the locations of its polar groups can be determined by X-ray crystallography or NMR, one can in principle design *de novo* a ligand intended to be complementary to the target. This process is as yet fraught with difficulties, is rarely quantitatively predictive, and is very unlikely to yield a drug on the first try. However, iterating the process of ligand design, synthesis, and co-crystal structure determination yields insight that makes the discovery process considerably more manageable. It has resulted in several drugs, and in some cases substantially decreased the amount of synthesis required to find them.[1] Structure-

based methods, however, require the active site structure, which is often unavailable; many proteins resist crystallization and are too large for NMR. Even when available, structures derived from crystallography and NMR may be misleading: they may represent an averaged ensemble of conformations or reflect a protein conformation that, due to side chain motion, is not that of the complex with the ideal ligand.

Quantitative structure–activity relationship (QSAR) methods predict the activity of hypothetical compounds based on the assayed activity of previously synthesized ones. QSAR methods require the identification of at least one lead molecule that binds to the active site, even if only weakly, and the synthesis and assay of several variants. The QSAR method then correlates properties computed from the structure of the ligands with their assayed activities. The resulting correlation can be used to predict the activities of hypothetical molecules and so help decide what to synthesize next, in the absence of information about the structure of the target. The form of the correlation may also be useful in understanding the structural features that contribute to activity, and so may help in designing improved ligands.

QSAR methods differ in the set of features of molecules that are correlated with activity and in the methods for determining the correlation. For the purposes of this paper, we can distinguish two broad classes of QSAR methods, "traditional" methods and "3D" methods. Traditional QSAR methods use some combination of three kinds of features.[2] The first type are bulk molecular properties, such as the computed or measured octanol/water partition coefficient (which predicts molecular hydrophobicity), computed or measured molar refractivity (thought to predict polarizability), and molar volume. The second type are topological and geometrical features that weakly correlate with molecular shape, such as the lengths of the principal axes, aspect ratios, number of aromatic bonds, and connectivity indices. The third are applicable only in cases in which all the molecules consist of substitutions on a common parent structure; they encode the identity or features of the substituent at each position. Traditional QSAR methods yield highly accurate predictions in some systems; in others they predict poorly. When accurate, they are useful in screening candidates for synthesis. However, they may provide little guidance for

---

design. If you find that a hypothetical molecule is predicted to have low activity due to high molar refractivity or low seventh-order valence chain index, it is not obvious how to improve it. Properties such as these are not directly grounded in the physics of binding, and when they do correlate with activity it is often unclear why.

3D QSAR methods use as features direct measurements of the three-dimensional shapes of molecules and the three-dimensional distribution of charges in and about them. These methods often yield better predictions than traditional QSAR, probably because the features more directly reflect the physical processes that govern ligand binding.[3] Further, 3D methods can relate predicted activity to specific portions a molecule, for example by explaining poor activity in terms of excessive steric bulk in a particular region or to a polar functionality of the wrong sign. This can directly guide the design of an improved ligand.

This paper describes a new method, Compass. Compass differs from previous methods in three ways: it computes physical properties only near the surfaces of molecules; it incorporates a nonlinear statistical methodology; and it automatically chooses conformations and alignments. We'll describe each of these here briefly.

The enthalpy of ligand–target binding depends on interactions across the ligand–target interface. We can think of the target active site "measuring" properties of the ligand; however, its measurements can occur only at the interface. The target cannot determine the internal structure of the ligand, it can only sense the field effects induced by this structure at the ligand–target interface. For example, the electrostatic field within a ligand is not relevant to binding; only the electrostatic force it exerts across the interface contributes to enthalpy. The same is even more obviously true of van der Waals interactions.

For this reason, Compass measures a ligand's properties only around its surface. By avoiding measurements of internal properties, Compass avoids spurious correlations that can lower predictivity. Further, the surface-only representation makes Compass more likely to predict across chemical classes. Structurally unrelated molecules may bind very similarly if they have similar surface characteristics, because they interact similarly with the active site, and models based on a surface-only representation should predict this. We have tested this hypothesis directly in other work[4] in which Compass was able to accurately extrapolate activity predictions from one structural class to another.

QSAR techniques have generally used linear statistical methods, such as linear regression or partial least squares, to find correlations. Nonlinear statistical methods can often provide better predictions than linear ones (though they must be used with care, to avoid overfitting). Such methods, particularly neural networks, have been used in traditional QSAR[5] and have produced better results in some cases. Compass's second innovation is the use of a new type of neural network, designed to make best use of the surface features and to support the conformation and alignment selection process described in the next paragraph.

We have spoken of "ligand shape" as though it were a determinate property. Molecules are flexible, however, and can adopt infinitely many conformations, each with a slightly different shape. Further, in comparing two ligands, one must first align them relative to each other to determine a correspondence of parts. Any two molecules

can be aligned in infinitely many ways. The conformation and alignment of a molecule relevant to predicting its biological activity are the "bioactive" ones—that is, the ones the molecule takes in binding to the active site of the target protein. The third innovation in Compass is a method for automatically selecting, for use in model construction, a conformation and relative alignment for each molecule. Ideally these should be the bioactive ones. Previous systems required the user to guess these, which was a significant difficulty and source of error. Typically, studies have used the vacuum minimum energy conformer derived from a conformational search or the small-molecule crystal conformation. The bioactive conformation may be quite different from these.

To determine whether these new methods actually improved performance, we benchmarked Compass on a steroid binding affinity problem to which CoMFA[6] and a molecular similarity method[7] had previously been applied. Compass's predictive performance is uniformly superior across a variety of conditions. In further experiments, we showed that each of the three advances contribute to predictive accuracy.

Methods

Overview. The Compass algorithm is outlined in Figure 1. This part of the Methods section surveys the algorithm at a high level only; subsequent parts supply additional details required to understand the system fully. The penultimate part of the Methods section describes the benchmark problem used in comparing Compass with other methods. Further algorithmic details appear in the Experimental Section at the end of the paper.

Compass operates in three phases. The first phase constructs a set of initial guesses as to the bioactive conformation and alignment of each molecule. We call a conformation in a particular alignment a *pose*, so these are guesses as to the bioactive pose. The second phase simultaneously chooses a bioactive pose for each molecule, starting from these guesses, and constructs a statistical model which explains quantitatively and predictively the relationship between the surface characteristics of the given molecules and their biological activity. The third phase predicts the activity and bioactive pose of a new molecule and can also graphically display the basis of the prediction in a way that aids molecular design.

Compass begins the first phase by conducting a standard conformational search to find low-energy conformers for each molecule. It will eventually choose, for each molecule, one of these conformers as the one most likely to be bioactive. We use a standard conformational search package.

The user then must identify either a pharmacophore or a substructure common to all molecules in the data set. This serves as a qualitative hypothesis about the ligand's shared binding mode. Compass uses this binding mode hypothesis to assign an approximate, initial alignment to each conformer on the basis of RMS fit of the pharmacophoric groups or substructure.[8] We will see that Compass later improves these crude alignments. As an example of basing alignment on a pharmacophore, in an unpublished study we aligned m1 muscarinic receptor ligands on the basis of the well-known pharmacophore, consisting of a protonated amine and either one or two acceptor functionalities.[9] In the case of the steroid experiments reported here, we aligned atoms in the steroid
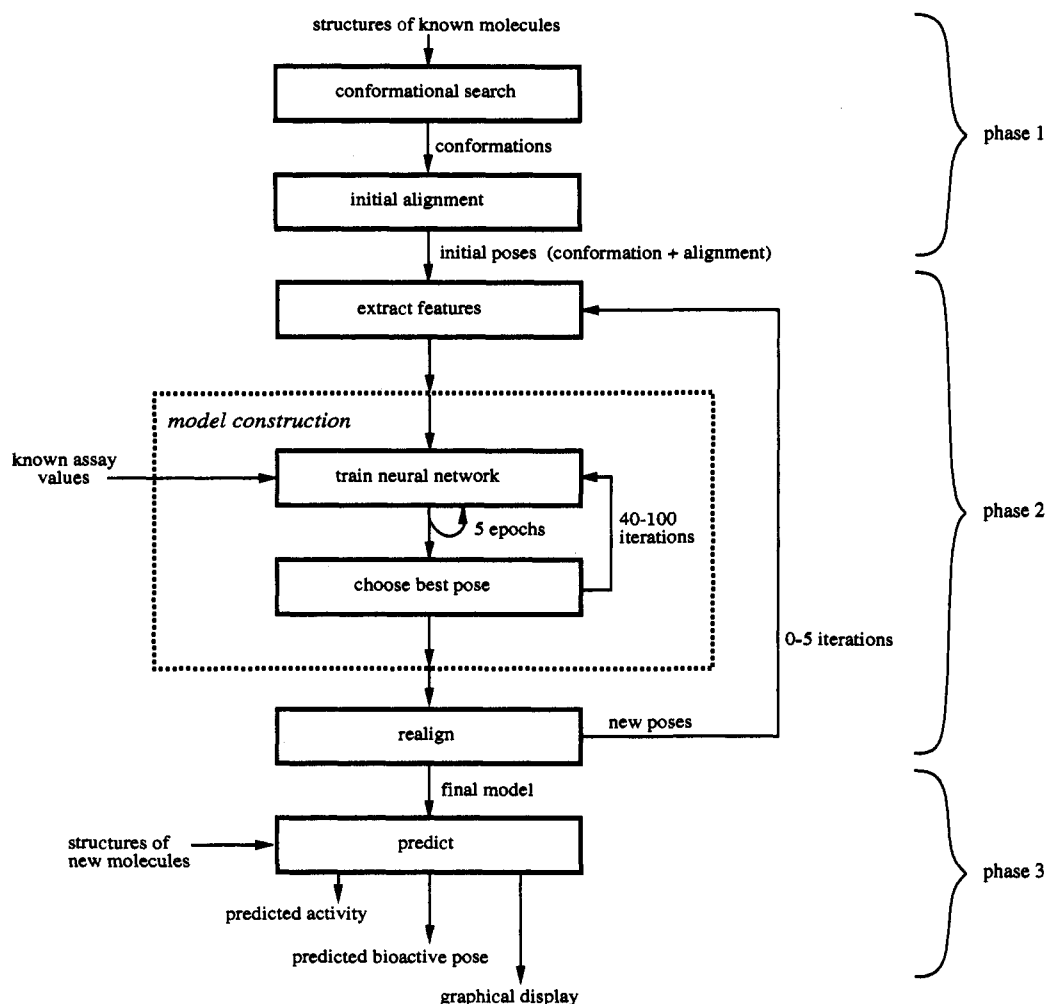
Figure 1. Flow diagram for Compass.

nuclei, which formed a substructure common to all the molecules. In cases in which the correct initial alignment is unclear, several can be supplied, and Compass will choose among them, using the pose selection algorithm of the second phase.

This second phase, discussed in greater detail in the Automatic Pose Selection and Nonlinear Statistical Method sections, proceeds iteratively, with three steps repeated in a loop.

1. The first step computes a set of real-valued features from each pose. Each feature measures the surface shape or the polar functionality of the pose in the vicinity of a particular point in space.

2. The second step constructs a statistical model relating structure to predicted activity. We use a neural network, which is a nonlinear function of the feature values, as the model. The neural network maps feature sets to activities and can be used to predict the activity of hypothetical molecules. This model can only be as good as the initial poses, however.

3. The third step uses the model to realign the molecules, thereby deriving better poses. Since the model measures local determinants of binding affinity, it is indirectly a model of the binding site, and the realignment process can be thought of as allowing a molecule to rotate, translate, and alter its conformation to achieve the best fit to a binding site. This is analogous to the physical process that results in molecules taking up the bioactive pose when interacting with the target protein. To the extent that

the model is accurate, realignment will find for each molecule a pose closer to the bioactive ones than those previously considered.

Given these improved poses, Compass can build an improved model that more accurately reflects the structure—activity relationship. This alternating process of model building and reposing iterates to convergence and yields, for each molecule, final predictions of activity and bioactive pose. Convergence completes the second phase.

The third phase predicts the activity and bioactive pose of a new molecule. Compass carries out a conformational search on each molecule and places them in initial alignments using the same common substructure or pharmacophore as was used in applying the first phase to the known molecules. These poses are realigned relative to the neural network model generated in the second phase; the likely bioactive pose is chosen using the selection algorithm of the second phase, and the model is used to predict activity.

The next several parts of this Methods section describe in more detail the three aspects of Compass that differ significantly from previous systems: the molecular representation, the use of a neural network as a nonlinear statistical methodology, and automatic pose selection.

Molecular Representation. Because molecular shape depends on conformation, Compass cannot represent a molecule with a single set of features. In fact, Compass's shape representation depends on alignment as well as conformation. Thus, each pose of a molecule is represented
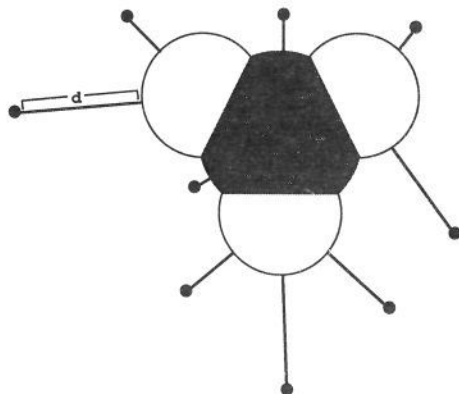
Figure 2. Steric feature computation illustrated for methane, using an arbitrary set of sampling points. Each point measures the distance to the van der Waals surface of the molecule.

by a different set of feature values. Each value corresponds to a *sampling point* in an invariant, common reference frame. The sampling points, scattered near the surface of the molecules, allow one to measure steric and polar characteristics in their vicinity. A single set of sampling points is used to derive feature values for all poses of all molecules.

Three types of features are used: steric features and hydrogen bond donor and acceptor features. Steric features consist simply of the distance from a sampling point to the van der Waals surface of a molecule in a particular pose (Figure 2).

Donor and acceptor features, similarly, measure the distance from a sampling point to the nearest H-bond donor or acceptor group. For example, a particular donor feature, in sampling a particular pose, might find the nearest amine, skipping over nearer hydrocarbon groups and a carbonyl. For the present study, it was sufficient to define the distance to a polar group as the distance to the one non-hydrogen atom in the group. In the benchmark data set, the only acceptor functionalities are oxygens and a fluorine; acceptor features measure the distance to them. The only donor functionalities were hydroxyls; in this case, the distance is measured to the oxygen, despite the fact that the hydrogen has the positive charge in order to suppress effects of the arbitrary hydrogen orientation chosen in conformational search. We originally used a polar feature type based on atomic partial charges and Coulomb's law but found this method inaccurate, presumably due to simplifying assumptions such as the nontreatment of polarization effects. Bohacek *et al.*[10] report similar conclusions and independently derived a similar H-bond representation. Adding hydrogen bond angle features as well as distance features may improve performance, but we have not tested this.

Compass's steric feature type is more precise than grid sampling methods,[6] which most closely resemble it. The van der Waals field of a molecule changes much more quickly than can be captured in a typical 2.0-Å or even 1.0-Å sampling. In effect, grid sampling of the steric field tells only whether the molecule occupies each point or not. By measuring distances to the surface, in contrast, each sampling point in our representation makes a steric measurement whose precision is limited only by the approximation of the molecular surface by van der Waals spheres.

A lesser advantage, relative to grid methods, is efficiency: instead of thousands of grid points, only 265

sampling points were used in the experiments reported here. The smaller number of features also reduces the problem of overfitting in model building.

Nonlinear Statistical Method. This method is used to construct the mathematical function that maps sets of feature values to activities.

The assayed activity of molecules will generally be highly nonlinear in the distance features defined above. An ideal distance to a sampling point places the sampled portion of the surface in contact with the hypothetical active site, contributing to binding. A shorter distance results in steric interference and disfavors binding. A greater distance results in lost favorable interactions. Since an optimal interaction involves an intermediate distance, with poor interactions resulting from either greater or lesser distances, any monotonic function of the distances (including linear functions) would make an inherently poor fit to the underlying physics.

We have instead aplied a nonlinear method, specifically a neural network.[11] The neural network is a nonlinear function that maps sets of feature values (distances) to predicted activities. This function constitutes the predictive model. In addition to the feature values, it takes as inputs a set of *parameters*. These parameters are set automatically, according to a method described in the next paragraph, to optimize the model's predictions of the compounds whose activities are known. The network consists of three "layers"; that is, it is a three-deep composition of simple nonlinear functions. The "input units", which are the functions applied directly to the feature values, have a form similar to that of a Gaussian, with a central peak and values that drop off toward zero on either side:

$$G_{\mu,\sigma,z}(x) = ze^{-(x-\mu)^2/2\sigma^2}$$

There is one G unit for each feature. The function G captures the physical intuition that there is an optimal distance, with activity dropping off as the distance goes either above or below this value. The argument x is the input feature value. The parameter μ is the optimal value of x, for which the output of G will be maximized. The parameter σ determines how quickly activity decreases as the feature value moves away from the optimum. This parameter might correspond to how flexible the corresponding portion of the receptor wall is, and so how tolerant it is of ligand variation, and thereby may capture in part the phenomenon of induced fit. The parameter z is a weighting term, which determines how much the feature contributes to overall binding. The outputs of the G units are fed through two layers of standard sigmoid units of the form

$$H_{b,w}(x) = \frac{w}{1 + e^{-(x+b)}}$$

In the benchmark experiments, we used 265 input G units (one for each feature), three H units in the intermediate layer, and one H output unit that produced the predicted activity.

A neural network—which is to say a continuous, differentiable nonlinear function with adjustable parameters—can be "trained" to approximate a given function using the "backpropagation" method.[12] This method is a gradient descent parameter fit; it repeatedly applies the

network to known values of the function to be approximated and selects parameter values that minimize the error. In our case, the known values of the function are the assayed activities of the known molecules, so the difference between the assayed activity and the predicted activity is used as the error term in training.

Backpropagation is an iterative process, which proceeds in steps called "epochs". During one epoch, each training pair, consisting in our case of the set of features derived from a molecule and its assayed activity, is presented to the neural network once, and the parameters are adjusted for each to decrease prediction error.

Automatic Pose Selection. The values of the features extracted from a molecule depend on its pose, since the pose determines the distance from sampling points to the surface. Since conformations and alignments can vary continuously, there is an infinite space of poses for each molecule, and so a corresponding infinite space of sets of feature values. Most previous systems require the user to choose, from this infinite space, a single pose for each molecule. Ideally this should be the bioactive pose, since it is the features of the bioactive pose that are relevant in determining activity. Predicting the bioactive pose is difficult, and incorrect predictions introduce a significant form of human bias into the model-building process. The need to supply a single, aligned conformation for each molecule has been a main hindrance in applying other systems.[13]

The second, main phase of Compass both builds a model predicting activity and selects, for each molecule, a pose predicted to be bioactive. The selected poses are used in constructing the model, and the model is used in selecting poses. Since each depends on the other, the problem is solved by an iterative refinement starting from the initial poses produced by the first phase. We will show, later in the paper, that Compass's choice of poses contributes more than any other single factor to its performance.[14]

To understand the automatic pose selection algorithm, we must first define the notion of the activity predicted for an individual pose. This corresponds to the activity that would be measured if the molecule were somehow locked into a conformation and could bind to the receptor in only one alignment—that is, in only one orientation and at only one position. It is computed by extracting the feature values for the pose and then applying the neural network model to produce an activity prediction.

If the neural network models the target protein's active site perfectly, the bioactive pose will be that which maximizes predicted activity, because ligands flex and move to bind the target proteins as tightly as possible. Pose selection exploits this fact; it chooses the pose that is predicted most active as the likely bioactive one.

At any given time, Compass has access to a finite set of previously generated poses. Initially, these are those produced in the first phase by conformational search and approximate alignment. In selecting a pose to treat as bioactive, Compass can choose among the existing poses or it can generate new poses out of the infinite space of possible poses. The second phase is organized as three nested loops (Figure 1), which perform different subparts of the process more and less often according to their cost.

Choice among the previously generated poses is cheap, so it is applied frequently during model building. On every fifth epoch of neural network training, the model is applied to the features extracted from each, and that with the highest predicted activity is chosen. (One might choose on each epoch, but doing so less frequently decreases run time without significantly affecting accuracy.) Only the features of the maximal pose of each molecule are used in the next five training epochs. Thus, only the characteristics of the pose currently predicted to be bioactive are used in improving the model, which reflects the physical reality that only the characteristics of the bioactive pose contribute to the measured activity.

Generating new poses is much more expensive than choosing between existing ones, so it is performed much less often, in the outer loop only. Compass generates new poses for consideration by using a continuous function optimization technique, such as gradient descent. To generate improved alignments, each conformation is rotated and translated to maximize activity. The features are continuous, piece-wise differentiable functions of the six alignment variables (three degrees of freedom in rotation and three degrees of freedom in translation). Since the neural network model is a continuous, differentiable function of the features, their composition is a piece-wise continuous, differentiable function mapping poses to activities. Gradient descent or other similar function optimization techniques can then find the alignment that maximizes activity.[15] It is a straightforward extension of the technique to allow the conformational variables, such as torsion angles, to vary continuously, as the alignment variables do, but we have not yet implemented this.

The outer, realignment loop is considered to have converged when the predicted activities of all molecules are close enough to the assayed values. The largest number of realignment cycles required for convergence was five in the experiments reported in this paper.

Automatic pose selection is one way Compass addresses the general phenomenon of induced fit, whereby a ligand and its target protein both change conformation upon binding. The use of the σ parameters in the neural network, which model the flexibility of regions of the receptor wall, is another. Finally, Compass can model some cases in which a protein side chain moves to either of two locations, resulting in distinct energetically advantageous interaction regions, by placing sampling points that measure proximity to each region. However, many cases of induced fit, such as those in which a whole protein loop shifts, are unlikely to be captured.

Benchmarking. We applied Compass to a steroid binding affinity prediction problem previously studied by Cramer, Patterson, and Bunce using CoMFA[6] and by Good, So, and Richards,[7] who applied a molecular similarity method.[16] The data set (Figure 3) consists of 21 steroids assayed for binding affinity to two transport proteins, corticosteroid binding globulin (CBG), and testosterone-binding-globulin (TBG)[17] and an additional 10 steroids assayed for CBG binding only.[18]

Both previous studies used a statistical technique called *cross-validation*[19] in measuring predictive ability. In a cross-validation experiment involving n molecules, a model is built from all but the first molecule, and this model is used to predict the activity of the first molecule. Then all but the second molecule are used to create a model that predicts the second molecule, and so on. In this way, each molecule is predicted, as though the system had never seen it before, on the basis of all the other molecules. (This is actually a special case of cross-validation, called "hold-
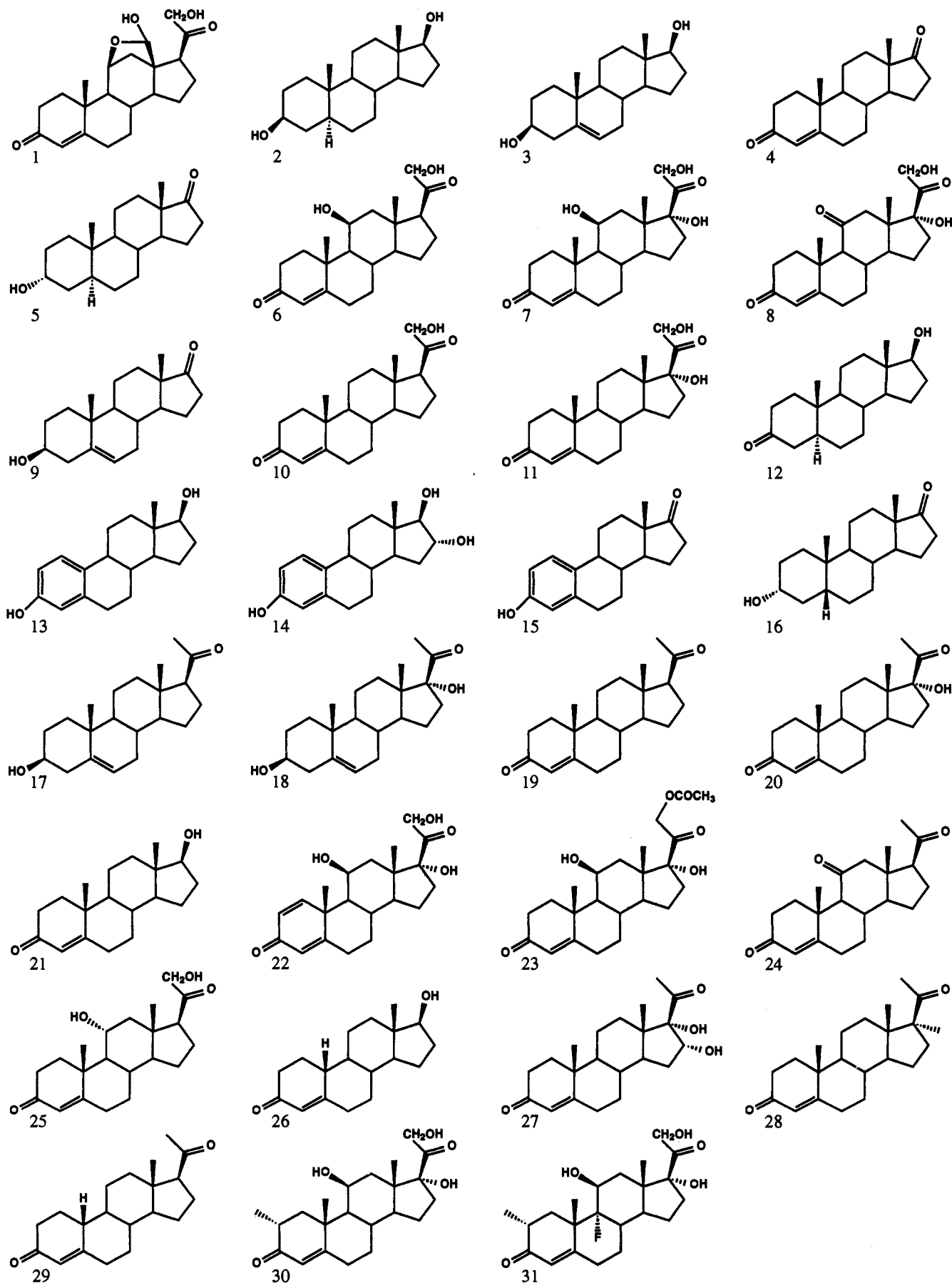
Figure 3. Structures of the 31 steroids in the benchmark data set.

one-out" cross-validation. Hold-one-out cross validation was the type used in all the studies described here.)

In the previous studies, systems were initially developed

using the 21-molecule set with a cross-validated experimental design. Only when development was complete were the 10 additional molecules predicted on the basis of a
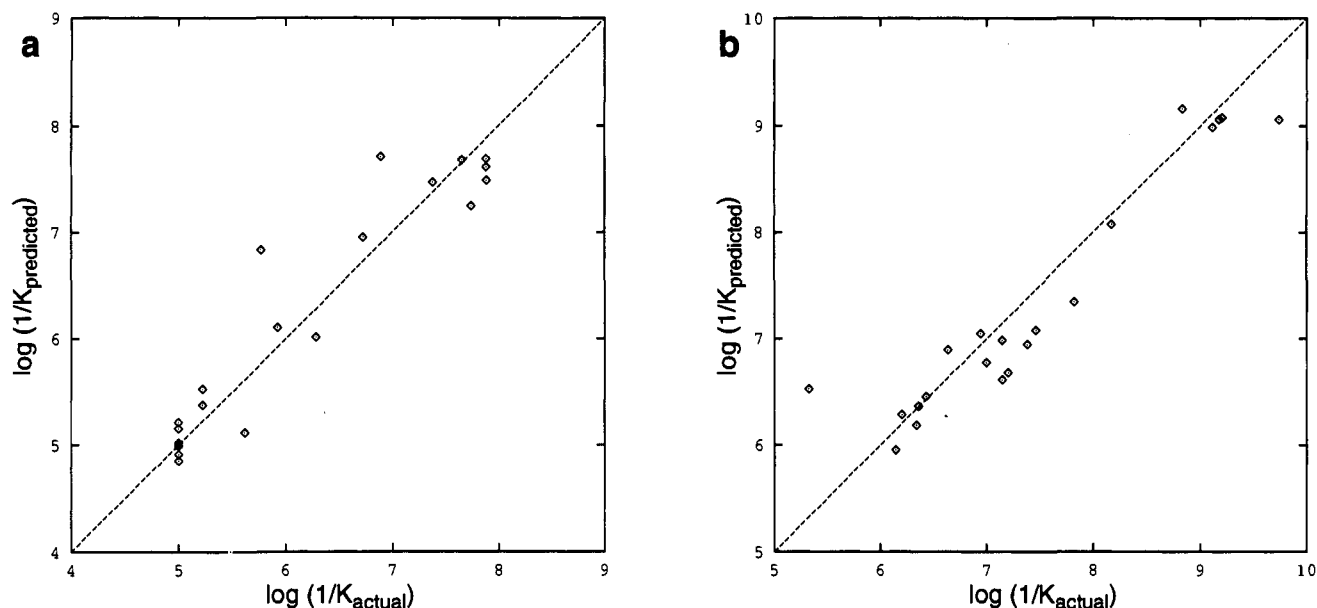
Figure 4. Scatterplots of predicted versus actual binding affinities for the 21 molecules in cross-validation. Panel a, CBG assay; b, TBG assay.

model derived from the 21. We used the same experimental design in applying Compass.

Good et al. applied both their molecular similarity method and CoMFA to generate three types of models, using steric features only, using polar features only, and using both steric and polar features. This allowed them to break down the contributions of steric and polar effects to binding. We generated the same three types of models using Compass.

We obtained the conformations used in the original CoMFA study from Tripos Associates[20] and took these as input to the conformational search. As initial alignments, we supplied the same alignments used in the previous studies, which optimize RMS fit of the 3, 5, 6, 13, 14, and 17 carbons of each conformation to the corresponding atoms of the standard conformation of deoxycortisol (11).

Performance Metrics. An activity prediction method has two functions: it should help choose between candidates for synthesis, and it should help design better candidates. It is hard to quantify the second role, but it is possible to quantify the first. An estimate of the predictive accuracy of a method helps its developers known whether they are making progress; it also helps medicinal chemists know how far to trust predictions.

The comparison papers report two measures of predictive performance: cross-validated r^2 is reported for cross-validations, and a standard error metric is reported for the 10-molecule prediction set. Cross-validated r^2 measures how well a model predicts data not used in model construction. An r^2 of 1.0 results from perfect prediction; $r^2 = 0.0$ corresponds to random prediction. Cross-validated r^2 should not be confused with the conventional r^2 metric, which measures only how well a model fits the data it was generated from. It is easy to develop models that achieve virtually perfect conventional r^2 but that have little or no predictive value. Accordingly, all r^2 values cited in this paper are cross-validated.

The "standard error" metric reported in previous papers is inapplicable to Compass because it is specific to the type of model the previous systems built (a partial least squares model). Therefore, in its place we report Kendall's τ, a measure of how well a system predicts the *ordering*

Table 1. Comparative Performance Results for Various Systems Applied to the Steroid Binding Problem[a]

	CoMFA	Similarity	Compass
	CBG		
both	**0.69**	**0.53**	**0.89**
steric	0.76	0.63	0.87
polar	0.64	0.50	0.81
	TBG		
both	**0.44**	**0.74**	**0.88**
steric	0.48	0.24	0.51
polar	0.60	0.73	0.67

[a] Numbers are cross-validated r^2 values. We report numbers for two different assays (CBG, TBG), for three different model-construction conditions (steric and polar features together, steric features only, polar features only), and for three methods (CoMFA, the molecular similarity method of Good et al., and Compass). The both-features condition is the significant result, and so is boldfaced; the other conditions are controls. Numbers for CoMFA and the similarity method are taken directly from Good et al.[7] who ran a more recent version of CoMFA and obtained somewhat better results than Cramer et al.[6] The results from Compass come from runs using uniform control settings throughout.

of data. A perfect τ (1.0) is obtained when molecules sorted by predicted assay value are in the same order as when they are sorted by actual assay value. Random prediction results in a zero τ; predicting the reverse order yields -1.0.

Results

We ran two classes of experiments: those aimed at evaluating Compass's predictive performance and those aimed at understanding the reasons for that performance. In the first class, the main result is that Compass makes substantially better predictions on the steroid binding benchmark than previous methods. This is true for both assays (CBG and TBG) and for both the 21-molecule data set and for predicting the 10 additional molecules on the basis of the structures and activies of the 21. Comparative performance is most clearly illustrated in Tables 1 and 2 and Figures 4 and 5. The results from the second class of experiments demonstrate that Compass's strong performance is in fact explicable in terms of the theoretical considerations that led to its design. These results are presented formally in Tables 1, 5, and 6, but are perhaps best understood by examining Figures 6–8.
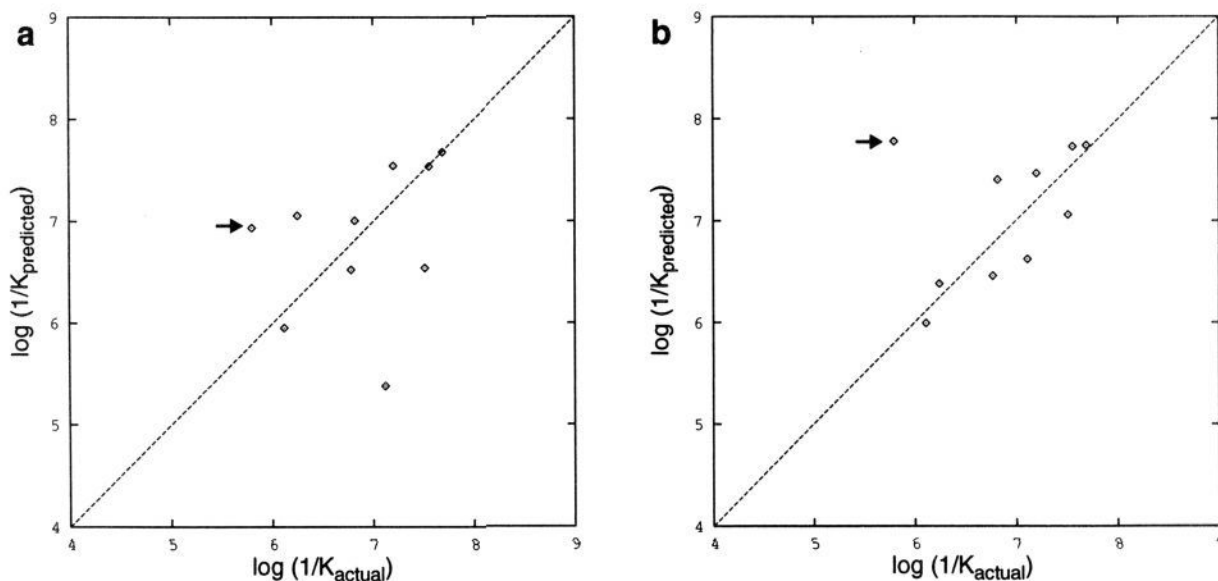
Figure 5. Scatterplots of predicted versus actual binding affinities for the 10-molecule set. Panel a, CoMFA; b, Compass. Molecule **31** is marked with an arrow in each plot. Good *et al.*[7] do not report raw data, so there is no corresponding scatterplot.
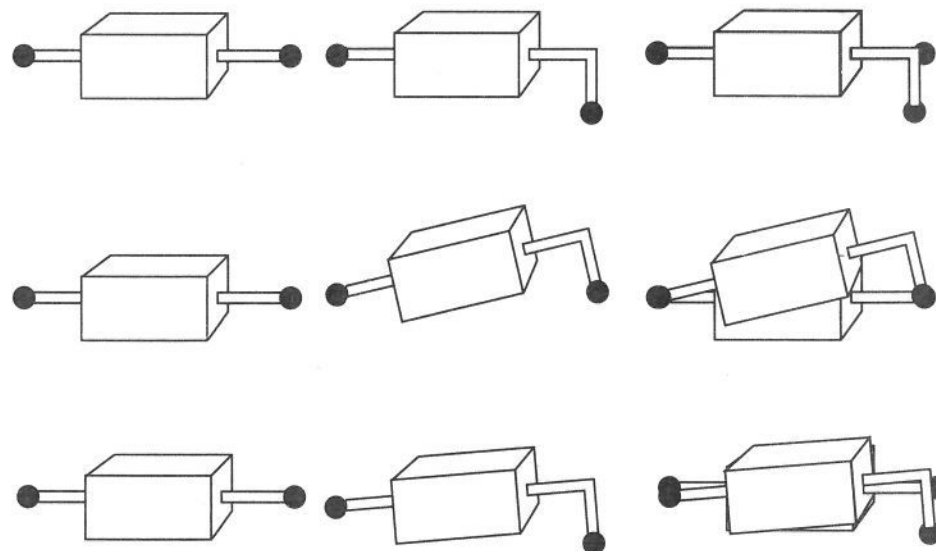
Figure 6. Effect of realignment in the polar-only model construction condition, illustrated in cartoon form for clarity. Top row, initial alignments of two molecules, each with two polar functionalities, one active and one inactive. The steroid nucleus, represented as the box shape, is aligned perfectly. The rightmost panel in this row shows the two molecules superimposed in their initial alignments. Middle row, realignment using only polar feature information results in the inactive molecule rotating to align its polar functionalities with those of the active one. Superimposition of the two, at the right, shows that they are almost indistinguishable in the positioning of polar functionalities, although sterically very different. Realignment hurts predictive performance in the polar-only condition because inactive molecules come to look like active ones. Bottom row, realignment using only steric features moves the second molecule only slightly, trading off slightly increased steric violations all over the molecule with decreased mismatch at the right end. In this case, the molecules are easily distinguished because the steric fit at this end is still poor.

Table 2. Kendall's τ Measure of Predictive Accuracy for Two Systems Applied to the 10 Held-Out Molecules[a]

prediction set	CoMFA	Compass
22–31	0.28	0.46
22–30	0.34	0.84

[a] Activities were predicted on the basis of a model built from the other 21 molecules and their assay values. Data for the similarity method are unavailable. We report results both for all 10 molecules and for the 9 molecules with **31** ignored.

Predictive performance of the four methods is shown in Table 1, which displays cross-validated r^2s for all three systems, both assays, and the three model construction conditions (steric features only, polar features only, and all features) described in the Benchmarking section above. (Numbers for CoMFA and the similarity method are taken from papers describing those methods.) In this table we see that Compass, in the standard model construction condition, achieved r^2 values of 0.88 and 0.89, whereas the best previous result, for any assay in any condition, was 0.76.

Figure 4 shows scatterplots of predicted versus actual activities for the 21 molecules in cross-validation tests. It shows plots for both assays, using seric and polar features combined. As one can see, the points cluster quite tightly about the 45° zero-error line. Similarly, Figure 5 shows scatterplots, for CoMFA and Compass, of actual versus

Table 3. Predicted and Assayed Activities for Molecules Held-Out in 21-fold Cross-Validation[a]

molecule	CBG assay	CBG prediction	TBG assay	TBG prediction
1	6.279	6.012	5.322	6.530
2	5.000	5.156	9.114	8.989
3	5.000	5.021	9.176	9.063
4	5.763	6.836	7.462	7.081
5	5.613	5.118	7.146	6.614
6	7.881	7.484	6.342	6.190
7	7.881	7.691	6.204	6.287
8	6.892	7.711	6.431	6.458
9	5.000	4.995	7.819	7.347
10	7.653	7.682	7.380	6.945
11	7.881	7.614	7.204	6.681
12	5.919	6.107	9.740	9.062
13	5.000	4.989	8.833	9.162
14	5.000	4.851	6.633	6.893
15	5.000	4.912	8.176	8.077
16	5.225	5.377	6.146	5.955
17	5.225	5.525	7.146	6.984
18	5.000	5.215	6.362	6.367
19	7.380	7.473	6.944	7.047
20	7.740	7.248	6.996	6.777
21	6.724	6.955	9.204	9.079

[a] Molecules are numbered according to the convention of Good et al.[7] Predictions are from models using steric and polar features combined. Values are log(1/K).

Table 4. Assayed and Predicted Activities for the 10 Molecules Held-Out during System Development[a]

molecule	CBG assay	prediction	molecule	CBG assay	prediction
22	7.512	7.062	27	6.247	6.383
23	7.553	7.729	28	7.120	6.625
24	6.779	6.462	29	6.817	7.403
25	7.200	7.466	30	7.688	7.741
26	6.114	5.994	31	5.797	7.779

[a] Assay values are available for CBG affinity only. Predictions are using both polar and steric features. Values are log(1/K).

predicted activities of the 10 molecules held out during system development. Here we see that only the Compass results approach linearity. (Raw data for the similarity method are unavailable, so we are unable to present a corresponding plot.) Tables 3 and 4 report the raw predictions corresponding to the figures; we present these data only so that others can perform alternative analyses of them.

Table 2 compares predictive performance on the 10 molecule set using the τ metric. We actually report results both for the full set of 10 and with molecule **31** dropped. Molecule **31** is the only one with a non-hydrogen 9-substituent. It is a fluorine, which is also the only non-oxygen functionality in the data set. It is otherwise identical to **30**, which is 100 times more active. All the methods substantially overpredicted the activity of **31**, and the comparison papers have reported results with and without it. As can be seen from the table and the scatterplots in Figure 5, Compass is substantially more predictive than alternatives, with or without **31**. In fact, CoMFA does not perform better than chance at $p = 0.1$. The Compass result with **31** is significant at $p = 0.03$, and at $p = 0.001$ without **31**. In fact, predictive performance on the nine molecules **22–30** is comparable to that on the 21 molecules from which the model was derived.

We conducted a series of experiments aimed at discovering the sources of the system's performance. These experiments removed from Compass, or added to CoMFA, various combinations of the three main technical advances incorporated in Compass. Table 5 displays the results of

Table 5. Cross-Validated Performance of Compass on the 21-Molecule Set with Automatic Pose Selection Partially or Completely Disabled[a]

assay	full posing	no realignment	single conformation	no posing
CBG	0.89	0.88	0.86	0.79
TBG	0.88	0.85	0.69	0.73

[a] Entries are cross-validated r^2s. No realignment, alignments held fixed. Single conformation, system supplied only with the single conformation used in the CoMFA study, but allowed to optimize alignment. No posing, system allowed to use only the original CoMFA conformations and alignments.

Table 6. Results of Experiments Substituting Compass Surface Features into CoMFA[a]

	assay					
	CBG			TBG		
poses	grid	surface	Compass	grid	surface	Compass
initial	0.76	0.73	0.79	0.60	0.72	0.73
selected	nd	0.85	0.89	nd	0.59	0.88

[a] First row, features derived from Tripos poses; second row, features derived from Compass-selected poses. For each assay, we report results under three conditions. The first, "grid", is the best result from all reported CoMFA runs, using whichever combination of polar and steric features gave the highest cross-validated r^2 value. The second, "surface", is the result of supplying the Compass surface features to PLS, using again the feature set that gave the best results. The third is the result from Compass (with pose selection disabled in the first row). Nd, experiment not done.

partially or completely disabling automatic pose selection. In this experiment, described later in more detail, we turned off the iterative realignment process or gave the system only the Tripos-supplied low-energy conformation used in the CoMFA study, or both, and measured predictive performance. Predictivity uniformly dropped. In fact, without pose selection, performance of Compass drops almost to the level of the best previous results. The effect of conformational selection was significantly greater than that of realignment.

Table 6 reports the results of two other experiments with a similar aim. In these experiments, we supplied Compass's surface features to partial least squares (PLS), the statistical component of CoMFA, in place of CoMFA's grid features. In one experiment, we supplied surface features for the CoMFA pose; in the other, we supplied the surface features of the pose Compass predicted to be bioactive. In the case of CBG, supplying the surface features of the bioactive pose resulted in better performance, and those of the initial pose in worse performance; in the case of TBG, the reverse effect held. The reason that supplying the surface features (of either set of poses) does not consistently improve performance is that the features are inherently nonlinear and may require a nonlinear statistical method to make use of them. PLS is a linear method and may not be able to make sense of nonlinear inputs. The CoMFA features, in contrast to the Compass ones, are in units of energy and so should combine linearly.

These results show that adding to CoMFA any one of the new capabilities incorporated in Compass may either improve or degrade performance. All three are required to get the best predictions because they act synergistically.

Discussion

Nonlinear Statistical Method. Compass's neural network model improves predictivity for two reasons: it

is able to combine multiple feature types synergistically and it is able to exploit the distance-based features.

Although the binding of steroids to CBG and TBG involves both steric and polar interactions, it was difficult for previous methods to combine information about these effects. Examination of Table 1 shows that in only one case (the similarity method applied to TBG) was a previous system able to produce a combined model that was as predictive as or more predictive than the single-feature-type models (and then not significantly so). Good et al. have suggested that the difficulty CoMFA and the similarity method had in combining feature types was an inherent problem with PLS: the greater dynamic range of the steric features leads to electrostatic features being eliminated in variable selection.

Compass, on the other hand, successfully combined feature types for both assays. (This despite the fact that its sterics-only and polar-only models were more predictive, on both assays, than any previous such model; in other words, Compass has the best r^2 value in each row. The one exception is the TBG assay modeled with polar features only, for which Compass did worse than the similarity method but better than CoMFA. Later in the paper we will suggest an explanation.)

This ability to combine information about sterics and polar functionalities is illustrated by Compass's performance in predicting molecule **16**'s TBG activity under the three different feature type conditions. Molecule **16**, like **9** and **15**, has a hydroxyl 3-substituent, a carbonyl 17-substituent, and no other heteroatoms. However, **9** and **15** are 100 times more potent. Molecule **16** has a fully saturated steroid nucleus, and its 3-hydroxyl substituent is α. In most but not all of its 11 energetically accessible conformational minima, the hydroxyl is placed well below the plane of the B, C, and D rings. However, under all three model construction conditions, the same conformation is chosen from among the 11, one of those that is relatively flat. In the steric-only model, it is oriented purely on the basis of steric fit. It is predicted to be about 1 log unit too low, because it is penalized for imperfect steric fit without being rewarded for good polar fit. Conversely, in the polar-only model, the conformation is oriented to maximize proper placement of the carbonyl and hydroxyl oxygens while ignoring sterics entirely. It is predicted to be 1.5 log units high in this case—very close to the activity of molecules **15** and **9**. In the combined feature type model, the conformation is oriented primarily on the basis of steric fit, but the prediction, incorporating polar factors as well, is very accurate (less than 0.2 log units low). Figure 8 shows molecules **15** and **16** in their chosen poses using Compass's graphical design tool. The saturation of the blue dots indicates how well the carbonyl is positioned; yellow dots indicate steric violations. Molecule **15** shows no steric violations and a good carbonyl fit; molecule **16**'s orientation is chosen to minimize steric violations (which cannot be entirely avoided) while picking up some benefit from a less-good carbonyl position. This model, then, is able to combine the information that the molecule has the appropriate polar groups with the fact that it is impossible to simultaneously place them correctly while respecting the steric requirements.

By comparing the "surface" columns of Table 6 with the "Compass" columns, one can see the effect of adding a nonlinear statistical method to the surface features. This improves performance in all cases, although much more

so in some than others. Our conclusion from the full series of PLS control experiments is that the three advances incorporated into Compass are synergistic in effect, and all three are required for best performance.

The importance of a nonlinear method is suggested also by a pilot experiment of Good et al.[7] replacing linear PLS with a quadratic model (GOLPE). This improved cross-validated r^2 for CBG, using both feature types, from 0.533 to 0.828. This number is still lower than that for Compass, however. Further, it appears that this experiment was not performed with a true cross-validation methodology, because the assay values of all molecules (including those held out for prediction) were used in variable selection prior to model building.

Automatic Pose Selection. Figure 7 illustrates the importance of automatic pose selection. It shows two conformations of molecule **28**, a member of the 10-molecule prediction set. The top conformation is the one used in the CoMFA study. The assayed CBG affinity of **28** is 7.12. On the basis of the top conformation, CoMFA predicts an affinity of 5.38, an error of 1.74 log units. This is more than half the range of the data (2.88 log units). By choosing the bottom conformation, unavailable to CoMFA, Compass predicts 6.63, an error of only 0.49 log units. When Compass is supplied only with the top conformation, it predicts 5.35, almost exactly the same value as CoMFA's prediction. Thus, in this case, the ability to choose among multiple poses is crucial to accurate prediction.

Two sets of experiments show quantitatively that automatic pose selection can improve performance of 3D prediction systems: experiments that remove pose selection from Compass and experiments that simulate adding it to a CoMFA-like system.

The first set of experiments is reported in Table 5, which displays the results of disabling realignment, conformer selection, or both. This substantially degrades performance. That supplying only the CoMFA conformation to Compass substantially degrades performance strongly suggests that the bioactive conformation is not the low-energy conformation and that optimal performance requires a 3D method to choose conformations itself. (Strictly, the CoMFA conformations are not necessarily the lowest-energy ones; they were derived in part from crystal structures and in part from conformational search.)

In this study disabling realignment (while allowing conformation selection) does not substantially degrade performance. This suggests that the initial alignments of the steroid nuclei are close to correct. In a previous study,[4] however, we found that realignment was crucial to performance; presumably our initial alignments of those molecules were less accurate.

Experiments with disabling realignment suggest an explanation for Compass's performance in the polar-features-only condition on the TBG assay, which was lower than that for the similarity method (but higher than that for CoMFA). Disabling realignment in this case actually improves performance, from 0.67 to 0.72 (comparable to the similarity method's 0.73). The reason is that the polar features alone do not sufficiently constrain realignment. Most of the molecules have polar functionalities only at the two extreme ends of the molecule, the 3 and 17 positions. Polar features are therefore insensitive to most of the molecule, and models based on them cannot constrain realignment well. In realignment, low-affinity molecules rotate drastically to align their polar function-
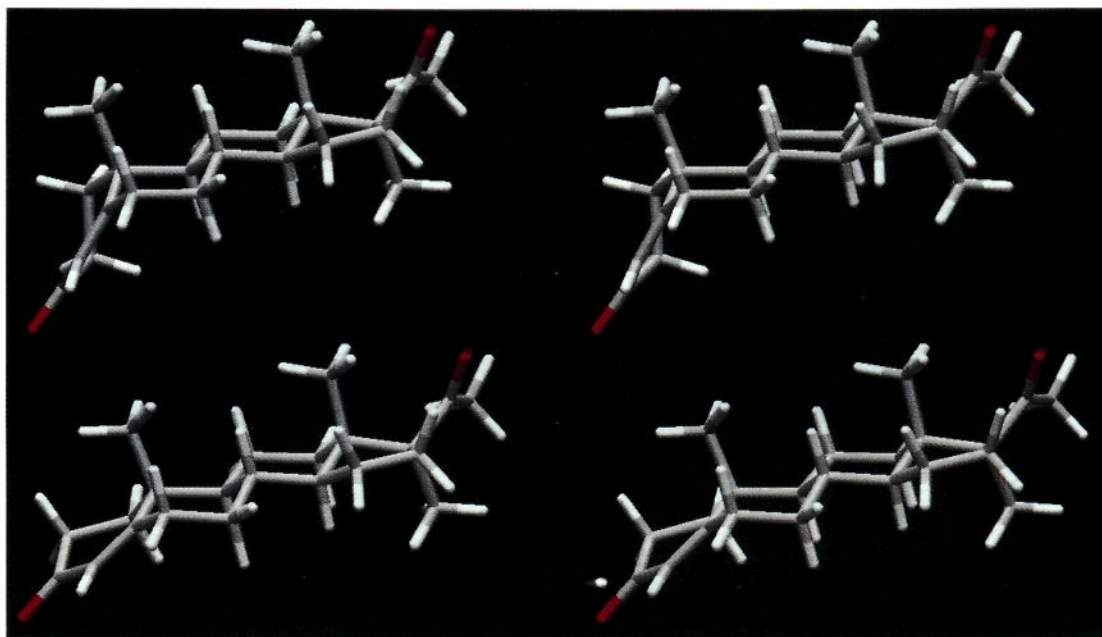
Figure 7. Conformations of molecule **28**. The top conformation is that used in the CoMFA study. The bottom conformation is that selected by Compass.

alities with those of the high-affinity molecules while ignoring unfavorable steric interactions, and their activities are overpredicted (Figure 6). Since the polar-only model is insensitive to the shape and alignment of the steroid nucleus, it is unable to prevent this. On the other hand, in the combined model, the steric features in effect constrain realignment by adding information about probable disallowed steric overlaps with the receptor. The combined model-building condition is, of course, the form we always use in practice.

The second set of experiments demonstrating the importance of automatic pose selection is reported in the "surface" columns of Table 6. Here we supplied either the CoMFA pose or the Compass-selected pose of each molecule to a system consisting of Compass surface features and a linear (PLS) model. Supplying Compass-selected poses simulates the effect of adding automatic pose selection to the linear system. Of course, although the Compass-selected poses are intended to be the bioactive ones, they are actually chosen to optimize Compass's neural network model and might not optimize PLS performance. Even if these poses are close to the bioactive ones, a linear model may not be able to exploit them. In fact, using the selected poses significantly improves performance on the CBG assay (0.73 to 0.85). On the other hand, it hurts performance on the TBG assay (0.72 to 0.59). The reason for this is that the conformations for several low-affinity molecules used in the CoMFA study are very dfferent from those chosen by Compass, which makes them easy to discriminate from high-affinity molecules. The Compass-selected conformations are much more similar to those of the high-affinity molecules because the pose selection process chooses the conformation with the highest predicted activity as the likely bioactive one. This conformation is likely to be the one most similar to conformations of high-affinity molecules, since that is what is required to fit the model. Had it by chance been high-affinity molecules that had the unusual conformations, pose selection would have helped PLS; as it was, pose selection hurts PLS in this case. (This illustrates the predictive

bias—accidental in this case—introduced by using only a single pose for each molecule.)

Conclusion

We have described a new method, Compass, for predicting unknown biological activities based on relating molecular structures to their known biological activities. The method incorporates several technical advances; theory predicts, and experiments confirm, that these result in substantially improved predictive accuracy. A benchmark study shows that Compass outperforms previous systems on the data set used in their development.

Many extensions and improvements to the system are planned or underway. First, we will add additional feature types to the system. For example, we plan to add the internal energy of conformations as a feature type. Surface hydrophobicity is another candidate feature.[21] A more detailed model of hydrogen bonding, including perhaps bond angles and the partial charges of the ligand donor or acceptor atoms, may improve predictivity, as may consideration of other polar phenomena such as aromatic interactions. Compass currently makes no reference to the entropic contribution to binding, which in some cases dominates the enthalpic contribution; we plan to add techniques that would estimate the entropy term.

Compass currently requires the user to supply a qualitative hypothesis about how ligands bind in the form of a common substructure or a pharmacophore. It should be possible to use standard maximum-common-subgraph algorithms to find common substructures and active analog techniques to find pharmacophores, thereby automating this step.

QSAR techniques currently give no estimate of the reliability of their predictions. A person, looking at ligand structures, can easily identify some molecules as outliers with, for example, a bulky substituent in a region where no other molecule protrudes. Little confidence should be placed in a prediction of the activity of such a molecule, because the data on which to base a prediction are simply

Figure 8. Comparison of molecules **15** (top) and **16** (bottom) in their chosen poses. Small white dots represent the Connolly surface. The saturation of the large blue dot indicates how close to optimal the distance from a polar sampling point to the carbonyl is. Large yellow dots indicate steric violations. Large brown dots indicate portions of the molecule that are close to optimal.

not available. We are developing means that will provide a confidence measure for predictions. In the later stages of the lead optimization process, one may choose to avoid synthesizing candidate molecules with low predictive confidence, preferring those that are predicted to be highly active with high confidence. In an earlier phase, one might deliberately synthesize molecules that are difficult to predict, in order to systematically explore the space of variation, and thereby to collect information about binding requirements as efficiently as possible, maximizing the information gained from each synthesis.

Although it would often be very useful to accurately predict binding affinities, in many drug discovery projects highly active compounds are well-known and the difficulty is selectivity instead. If one could predict activity against several targets sufficiently accurately, then that would suffice to predict selectivity; we believe that other methods can be developed that would predict selectivity more

accurately, in some cases, than activities against any of the individual targets could be predicted.

Ligand flexibility is a limitation on the applicability of the current implementation of Compass. Standard conformational search methods make it impractical to use the system to analyze ligands with much more than a thousand conformers. We are developing new search techniques that may make it practical to predict the activities of short peptides and other very flexible compounds.

Experimental Section

Computational Requirements. We performed our experiments on a Silicon Graphics Iris Indigo with an R4000 processor and 96 megabytes of main memory. The run time of the first phase, initial pose generation, is dominated by conformational search, which requires several hours per molecule. However, the results are independent of any subsequent variations in experi-

mental design and can be saved and reused in many experiments. The second phase, model building, takes only about 1 min per molecule, and the resulting model can be applied to predict a new molecule in a few seconds.

Sampling Points. The sampling points were scattered 2.0 Å outside the averaged van der Waals surface of all the initial poses. In addition, in areas where at least one pose extended an additional 2.0 Å beyond the 2.0-Å shell, points were added at the maximum excursion of any pose. This resulted in a total of 265 sampling points, of which 191 were steric and 74 polar. We have found that small changes in number and positions of sampling points make little difference in system performance.

Conformational Search. We used the BATCHMIN Monte Carlo search procedure[22] with the AMBER force field[23] to generate the set of conformers. We used the default parameter settings: energy minimization to less than 0.05 kJ/Å gradient, retaining all local minima within 50 kJ/mol of the global minimum, and performing 1000 search steps. The 1000-step default was sufficient to ensure oversampling for all molecules in the benchmark study, so the search probably found all low-energy conformers. The 50-kJ default energy window is probably wider than realistic; it is unlikely that the bioactive conformation of a relatively rigid molecule would be so far above the global minimum. However, the conformers Compass selected were almost always of much lower energy. We also conducted some experiments with a 20-kJ cutoff, and this did not significantly affect performance. We also believe that the force field used, and such details as solvated versus *in vacuo* computations will not significantly affect results in the case of relatively inflexible molecules. The reason is that all that is required is for the conformational search to turn up a conformation close in internal coordinates to the truly bioactive one. The energy assigned to this conformation is not used by Compass. Therefore it is not necessary to get the energies exactly correct, and indeed although the chose conformation must be fairly close to the global minimum in energy, it need not be at a local minimum. We have not, however, systematically experimented with alternative force fields.

Model Construction. All activities were normalized to the interval [0, 1] in order to accommodate the sigmoidal transfer function of the output unit. The learning rate (gradient descent step size) was −0.05. We did not use a momentum term, so the momentum coefficient was effectively 0.0. We ran 200–500 backpropagation epochs between reposings, with early termination on convergence. Learning is considered to have converged when the predicted activity of each molecule is within 0.05 of the assayed activity. The backpropagation code used in the experiments was based on that of Jain.[24]

In the study reported here, we used gradient descent to optimize alignments, with a step size of 0.03 and 20 descent steps per realignment. In other studies, we have applied more sophisticated optimization techniques, including simulated annealing and genetic algorithms but found gradient descent sufficient in this case.

Performance Metrics. Cross-validated r^2 is defined[6] as

$$r^2 = 1 - \frac{\sum (a_i - p_i)^2}{\sum (a_i - \bar{a})^2}$$

Here the a_i are the assayed activities of the molecules, $\bar{a}$ is the mean of the a_i, and the p_i are the predicted molecular activities. The numerator is the squared errors of the predictions, and the denominator is a measure of how much variation there is in the actual activities.

Kendall's τ metric is defined as follows.[25] Consider all $n(n-1)/2$ pairings of n distinct molecules. Each pair is considered to have been predicted correctly if the one with the greater assayed activity is given the greater predicted activity and incorrectly if the reverse is true. Then, in the absence of molecules with identical assayed or predicted values,

$$\tau = \frac{\text{correct} - \text{incorrect}}{n(n-1)/2}$$

We believe that τ is a better metric of activity prediction techniques than other alternatives (including r^2) because it measures directly how well a method answers the question "Which candidate should be synthesized next?". It also provides a linear penalty for error, instead of the quadratic penalty imposed by r^2. The quadratic penalty makes getting a single molecule very wrong, and all the others perfect, result in a bad r^2, even though operationally one may be happy with the results. The single error will typically not significantly slow the drug discovery process. Getting a single molecule very wrong, with all the others perfect, will give a good τ, however. We have found that the linear error penalty also makes τ significantly more stable under small system changes than r^2; this noise-insensitivity makes it easier to improve performance during development.

References

(1) Appelt, K.; *et al. J. Med. Chem.* **1991**, *34*, 1925–1934. Bugg, C. E.; Carson, W. M.; Montgomery, J. A. *Scientific Am.* **1993**, December, 92–98. Lam, P. Y. S.; *et al. Science* **1994**, *263*, 380–384.

(2) For reviews of these methods and the feature types used in them, see for instance: Blankley, C. J. In *Quantitative Structure-Activity Relationships of Drugs*; Topliss, J. G., Ed.; Academic Press: New York, 1983. Stuper, A. J.; Brügger, W. E.; Jurs, P. C. *Computer Assisted Studies of Chemical Structure and Biological Function*; John Wiley & Sons: New York, 1979.

(3) For a review, see: Marshall, G. R.; Cramer, R. D. *Trends Pharmacol. Sci.* **1988**, *9*, 285–289.

(4) Jain, A. J.; Dietterich, T. G.; Lathrop, R. H.; Chapman, D.; Critchlow, R. E.; Bauer, B. E.; Webster, T. A.; Lozano-Perez, T. Submitted.

(5) Andrea, T. A.; Kalayeh, H. *J. Med. Chem.* **1991**, *34*, 2824–2836. Aoyama, T.; Suzuki, Y.; Ichikawa, H. *J. Med. Chem.* **1990**, *33*, 2583–2590. Chastrette, M.; de Saint Laumer, J. Y. *Eur. J. Med. Chem.* **1991**, *26*, 829–833. Tetko, I. V.; Luik, A. I.; Poda, G. I. *J. Med. Chem.* **1993**, *36*, 811–814.

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

(7) Good, A. C.; So, S.; Richards, W. G. *J. Med. Chem.* **1993**, *36*, 433–438.

(8) Best RMS fit is computed using the algorithm of Ferro, D. R.; Hermans, J. *Acta Crystallogr., Sect. A* **1977**, *A33*, 345.

(9) Nordvall, G.; Hacksell, U. *J. Med. Chem.* **1993**, *36*, 967–976.

(10) Bohacek, R. S.; McMartin, C. *J. Med. Chem.* **1992**, *35*, 1671–1684.

(11) Rumelhart, D. E., McClelland, J. L., Eds. *Parallel Distributed Processing*; MIT Press: Cambridge, 1986.

(12) Rumelhart, D. E., Hinton, G. E.; Williams, R. J. In *Parallel Distributed Processing*; Rumelhart, D. E., McClelland, J. L., Eds.; MIT Press: Cambridge, 1986.

(13) Tripos Associates, Inc. *SYBYL Molecular Modeling Software, Version 5.4 Theory Manual*; 1991; p 2238.

(14) For further discussion, and comparison with alternative reposing algorithms, see: Dietterich, T. G.; Jain, A. N.; Lathrop, R. L.; Lozano-Perez, T. In *Advances in Neural Information Processing Systems 6*; Cowan, J. D.; Tesauro, G., Alspector, J., Eds.; Morgan Kaufmann: San Mateo, CA, 1994.

(15) For discussion of a similar technique, see: Dean, P. M.; Chau, P.-L. *J. Molecular Graphics* **1987**, *5*:3.

(16) Cramer *et al.* also compared CoMFA with traditional QSAR, using a variety of parameters including log P, molar refractivity, melting point, substituent constants, and the Hopfinger shape parameters V_0 and P_u. They found that CoMFA predicted more accurately than any of the traditional methods they applied.

(17) Dunn, J. F.; Nisula, B. C.; Rodbard, D. *J. Clin. Endocrin. Metab.* **1981**, *53*, 58–68.

(18) Mickelson, K. E.; Forsthoefel, J.; Westphal, U. *Biochemistry* **1981**, *20*, 6211–6218.

(19) Wold, S. *Technometrics* **1978**, *20*, 397–405.

(20) We thank David Patterson of Tripos Associates Inc. for these coordinates.

(21) Kellogg, G. E.; Semus, S. F.; Abraham, D. J. *J. Comput.-Aided Mol. Design* **1991**, *5*, 545–552.

(22) Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379–4386.

(23) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230–252.

(24) Jain, A. N. Ph.D. thesis, School of Computer Science, Carnegie Mellon University. Available as Tech. Rep. CMU-CS-91-208, 1991.

(25) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes in C*; Cambridge University Press: Cambridge, 1988.