Pruned Receptor Surface Models and Pharmacophores for Three-Dimensional Database Searching

Jeffrey J. Sutherland,[†] Lee A. O'Brien,[‡] and Donald F. Weaver*[,§]

Departments of Chemistry and Pathology, Queen's University, Kingston, Ontario, K7L 3N6, Canada, and Departments of Medicine (Neurology) and Chemistry and School of Biomedical Engineering, Dalhousie University, Halifax, Nova Scotia, B3H 4J3, Canada

Received February 4, 2004

A pharmacophore represents the 3D arrangement of chemical features that are shared by molecules exhibiting activity at a protein receptor. Pharmacophores are routinely used in 3D database searching for identifying potential lead compounds. The lack of shape constraints causes the query to identify compounds that could not fit into the active site. In the absence of structural information, a receptor surface model (RSM) can be used to represent the active site. The RSM consists of a surface that envelops a set of known actives after these have been aligned using their common features. When used for database searching, a RSM is overconstraining as it restricts access to regions that could be occupied by ligands, such as the solvent−protein interface or unexplored pockets. We describe a protocol for developing pruned RSMs using information gleaned from 3D quantitative structure−activity relationship (QSAR) models. We examined the performance of queries that consist of pharmacophores used alone or with pruned or unpruned RSMs by performing searches on six databases containing known actives distributed among inactives. The pruned RSMs yield an average selectivity 1.8 times greater than that for pharmacophore queries, compared to 1.6 times for unpruned RSMs. However, the pruned RSMs retrieve on average 73% of the actives identified using the pharmacophores, compared to 40% for the unpruned RSMs. As such, pruned RSMs represent a useful compromise between the high sensitivity of pharmacophores and the high selectivity of unpruned RSMs.

Introduction

A major challenge in drug design is the identification of novel leads using information deduced from a set of compounds exhibiting activity at a protein receptor. When the compounds share common structural attributes and mechanisms of action, quantitative structure−activity relationships (QSAR) can be used to capture the effects of structural variation on the activity of a compound. However, applying a QSAR model for identifying compounds that are structurally divergent (i.e., contain a different scaffold) is generally unreliable; both 2D and 3D QSAR methods have limited accuracy when predictions are made for compounds outside the chemical space covered by the set used for developing the model.[1,2]

When the structure of the biological target is known, docking methods can be used to predict affinity for the target.[3] In the absence of structural information, the three-dimensional arrangement of atoms or functional groups that participate in key ligand−receptor interactions, or *pharmacophore*, can be deduced from the alignment of (diverse) compounds that bind to the same receptor.[4−6] The alignment or pharmacophore can be obtained manually or with an automated approach (see ref 7 for a review of automated methods). The pharmacophore can be used to search a database of compounds or to design "focused" combinatorial libraries.[8,9]

Despite their ability to identify active compounds with inherently different structural attributes, pharmacophores are known to be promiscuous when used as queries for database searches. In addition to actives (true positives), they identify many compounds that are inactive when tested (false positives). Such compounds may satisfy the pharmacophore, but the absence of shape constraints fails to eliminate those compounds that could not possibly fit into the active site. Many methods have been developed for representing the active site from the superposition of known ligands.[10−16] Pseudoreceptor models use an atomistic representation of the active site,[10] while others simply represent the receptor using a continuous surface enveloping the known ligands[11,12] (i.e., receptor surface models). Yet others use an intermediate representation, employing a genetic algorithm to place particles such as hydrogen bond donors and hydrophobes around the known ligands such that the particle−ligand interaction energies correlate with observed activities.[13,14] The use of surface-based constraints[12,17−19] and excluded-volume spheres[20] is most prevalent in database searching. In contrast to pharmacophore queries, receptor models constructed from a set of active ligands are sterically overconstrained. Compounds that satisfy the pharmacophore query and fit into the active site may be rejected by a model constructed from a set of ligands that do not fully explore the active site, giving a high rate of actives classified as inactives (false negatives). The fact that receptor models are little used in database searching is testimony to their practical limitations.

* To whom correspondence should be addressed. phone: (902) 494-7183. Fax: (902) 494-1310. E-mail: weaver@chem3.chem.dal.ca.
[†] Department of Chemistry, Queen's University.
[‡] Department of Pathology, Queen's University.
[§] Dalhousie University.

10.1021/jm049896z CCC: $27.50
Published on Web 06/17/2004

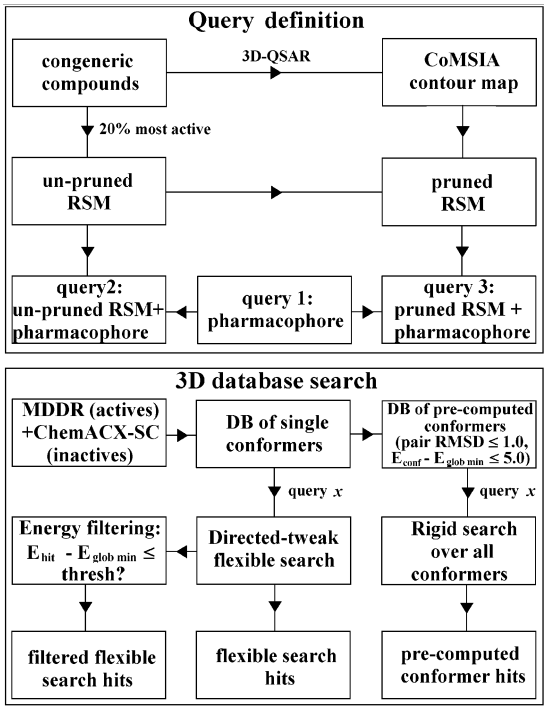

Figure 1. Flow chart indicating steps for query definition and 3D database searching.

To bridge the performance gap between pharmacophores and receptor surface models (RSMs) in database searching, we describe pruned RSMs where structure−activity information from a 3D-QSAR model is used to remove parts of the RSM that do not correspond to the physical limits of the active site. In particular, comparative molecular similarity indices analysis[21] (CoMSIA) is used to identify regions of the RSM for which binding affinity is influenced by steric contacts or electrostatic interactions; other regions are pruned away. Steric field contours are assumed to correspond to regions at which van der Waals receptor−ligand interactions may be occurring, while electrostatic field contours indicate regions at which hydrogen bond or electrostatic receptor−ligand interactions occur. In other words, protein residues are likely to be present near those CoMSIA contours. The parts of the RSM that are removed may correspond to pockets of the active site that are not explored by the ligands in the QSAR set or to the solvent−protein interface. This approach combines advantages from pharmacophore and 3D-QSAR models. A pharmacophore model can account for the key features required for activity, but deducing information concerning the shape of the active site from disparate compounds with possibly multiple mappings to the pharmacophore can be difficult. Conversely, the well-defined alignment of compounds in a congeneric series allows a 3D-QSAR contour map to be harnessed beyond its usual predictive power. We validate the method by examining the performance of pharmacophore queries used without shape constraints or combined with pruned and unpruned RSMs for retrieving known actives in database searches.

Methods

The development of pruned RSMs is outlined in Figure 1. For each of six QSAR data sets (section i), we develop a CoMSIA model (section ii). The most active 20% of QSAR compounds are selected for creating a RSM, which is subsequently pruned with reference to the CoMSIA contour map (section iii). Three types of database search queries are defined (section iv): (1) a pharmacophore model used alone, (2) a pharmacophore combined with the unpruned RSM, and (3) a pharmacophore with the pruned RSM. Databases for performing searches are assembled from MDDR compounds having activity at the same target as the QSAR compounds and from ChemACX-SC compounds assumed to be inactive (section v). Database searches are performed using two treatments of conformational flexibility (section vi).

(i) QSAR Data Sets. Six data sets were selected for developing QSAR and RSM models. The aligned structures for the BZR, COX-2, and ER sets are available in electronic format in the Supporting Information from our previous work;[22] the reader is referred to the "readme" file therein for details on the alignment procedure. The compounds used in this work are indicated in the Supporting Information.

(i.a) ACE. A set of 114 angiotensin converting enzyme (ACE) inhibitors has been studied with CoMFA[23] by Depriest et al.[24] Activities are spread over a wide range, with pIC_{50} values ranging from 2.1 to 9.9.

(i.b) AchE. A set of 111 acetylcholinesterase (AchE) inhibitors has been assembled from the work of Sugimoto et al., with pIC_{50} values ranging from 4.3 to 9.5. A subset of these compounds has been studied with CoMFA by Golbraikh et al.[25] We use their alignment rule (i.e., the pharmacophore described below), with other compounds flexibly fit onto E2020. However, the conformation of E2020 was not taken from the crystal structure of the complex (PDB 1eve) but was determined using the MCMM routine in Macromodel 7.2 (Schrodinger; Portland, OR) with the MMFF94S force field and GBSA implicit solvation model (500 Monte Carlo steps, other parameters default). The lowest-energy conformation has a heavy-atom root-mean-square deviation (rmsd) of 0.9 Å when compared to the conformation extracted from the crystal structure. We did not use the ligand-bound conformation deduced from the crystal structures for any QSAR modeling because the purpose of this work is to develop RSM models useful in the absence of such information.

(i.c) BZR. A set of 166 ligands for the benzodiazepine receptor (BZR) has been assembled from the work of Haefely et al.,[26] with pIC_{50} values ranging from 6.1 to 8.9. We use the syn conformation for esters such as Ro14-5974 instead of the anti conformation suggested by Cook et al.[27] (Figure 2). The syn conformer gives better overlap with other classes of BZR ligands and has a B3LYP/6-311+G(d,p)//HF/6-31G* energy only 2.1 kcal/mol higher than that of the anti conformer (the structures in the Supporting Information for ref 22 are in the anti conformation).

(i.d) COX-2. A set of 373 cyclooxygenase-2 (COX-2) inhibitors has been assembled from the work of Seibert et al. (see ref 22). pIC_{50} values range from 4.0 to 9.0.

(i.e) ER. A set of 87 estrogen receptor (ER) ligands has been assembled from two sources;[28,29] only compounds that were tested by Katzenellenbogen et al.

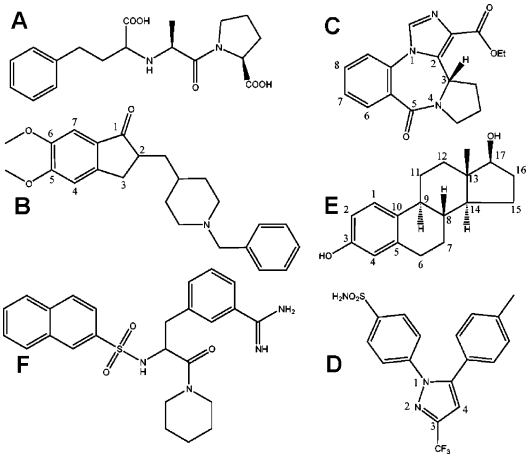

Figure 2. Representative compounds from each QSAR set: (A) enalaprat (ACE); (B) E2020 (AchE); (C) Ro14-5974 (BZR); (D) celecoxib (COX-2); (E) estradiol (ER); (F) naphtho derivative of 4-TAPAP (THR). For Ro14-5974, the ester is shown in the syn configuration. Ring labels correspond to those for diazepam.

using ER from lamb or rat uterine cytosol were retained from our previous compilation. Activities expressed as log(RBA) values range from −2.0 to 2.1.

(i.f) THR. A set of 88 thrombin inhibitors (THR) has been taken from the CoMSIA tutorial distributed with Sybyl; pK_i values range from 4.4 to 8.5. Representative compounds from each data set are shown in Figure 2.

(ii) CoMSIA. Comparative molecular similarity indices analysis[21] implemented in Sybyl 6.81 (Tripos Inc.; St. Louis, MO) was used for developing 3D-QSAR models for each data set. The method is conceptually similar to CoMFA,[23] in which steric and electrostatic fields are calculated at regularly spaced grid points of a lattice into which a series of aligned molecules are embedded. Partial least squares (PLS) is used to obtain a statistical model relating field values to observed activities. In CoMSIA, Coulomb and Lennard-Jones potentials are replaced with smooth Gaussian potentials having the form

$$A_{F,k}{}^{q}(j) = -\sum_i w_{\text{probe},k} w_{ik}\, e^{-\alpha r_{iq}{}^2}$$

in which the steric or electrostatic "field" k at grid point q for molecule j is obtaining by adding the probe atom potential for each atom i in the molecule. The coefficients w are the partial charge or atomic radius raised to the third power for the electrostatic and steric fields, respectively, α is a smoothing parameter set to 0.3, and r_{iq} is the distance (Å) between atom i and grid point q. The probe has a charge of +1 and radius of 1 Å. Net formal charges were determined by deprotonating carboxylic acids and phosphates and protonating non-aryl basic amines (except NH_2 groups that coordinate Zn in the ACE set), and scaled MNDO ESP-fit partial charges[30] were calculated with MOPAC 6.0 using atomic coordinates obtained by energy-minimizing the aligned molecules with the MMFF94S force field and MAXIMIN2 routine in Sybyl (200 steps; other parameters set to the default). A lattice with a 2 Å grid spacing and extending at least 4 Å in each direction beyond the aligned molecules was used for calculating fields. PLS analyses were performed after block-scaling the steric and electrostatic fields (CoMFA standard scaling). The statistical significance of models was evaluated by "leave-one-out" (LOO) cross-validation using the SAMPLS routine. The optimal number of components was determined by selecting the smallest s_{PRESS} value. The final PLS model (no cross-validation) was derived by setting the "minimum σ" standard deviation threshold to 1.0 kcal/mol. CoMSIA maps were obtained by contouring the "$\sigma \times$ coeff" field type at 25% and 75% contributions (i.e., the value of $\sigma \times$ coeff at which the sum of $|\sigma \times \text{coeff}|$ values reaches 25% or 75% of the sum for all grid points when sorted according to $\sigma \times$ coeff values). The contour thresholds were altered from their default values to give larger enclosed regions and to facilitate the subsequent RSM pruning steps.

(iii) RSMs. Receptor surface models were constructed by selecting the first 20% of QSAR set compounds when sorted by decreasing order of activity. These (aligned) compounds were merged into a single "supermolecule", for which a solvent-accessible surface was generated using the fast Connolly routine in the MOLCAD module of Sybyl (default parameters). For unpruned models, all atoms were used for generating the RSM. For pruned models, certain parts of the unpruned RSM were opened by excluding from the calculation atoms near regions that were deemed "unimportant" in ligand−receptor interactions (in some cases, it was more convenient to place dummy atoms or fragments such as benzene near parts of the surface that should be left open and excluding these atoms). "Important" parts of the RSM are selected using the following criteria:

1. Regions near grid points at which steric bulk causes a decrease in activity are indicative of ligand−receptor steric contacts; these are identified as yellow contours enclosing grid points with negative PLS coefficients.

2. Regions near grid points at which steric bulk causes an increase in activity suggest the presence of a receptor pocket that ligands may occupy; these are identified as green contours enclosing grid points with positive PLS coefficients. However, if no nearby grid points have negative coefficients, no molecule in the QSAR set is likely to be completely occupying the putative pocket. As such, parts of the RSM surface near such green contours are unimportant. We identify these regions by contouring the steric field below the value of $\sigma \times$ coeff that accounts for 90% of the *negative* signal; the other 10% is extremely diffuse and probably meaningless.

3. Some regions near grid points associated with electrostatic modulation of activity (blue or red contours) are selected as important. For electrostatic contributions, it is necessary to examine whether they represent positive or negative effects on activity. For example, grid points for which the corresponding PLS coefficients are positive (i.e., blue contours) may reflect a positive (enthalpic) contribution to binding if ligand atoms near that region bear positive charges or may reflect a negative (entropic) contribution when ligand atoms bear negative charges. This is readily determined by examining the charges on atoms of high-affinity compounds near the particular region. Only contoured regions that fall into the first category are considered important.

4. Parts of the RSM at which CoMSIA fields have little variance (identified by contouring the "σ field" at 4.0) are considered important. In literature data sets,

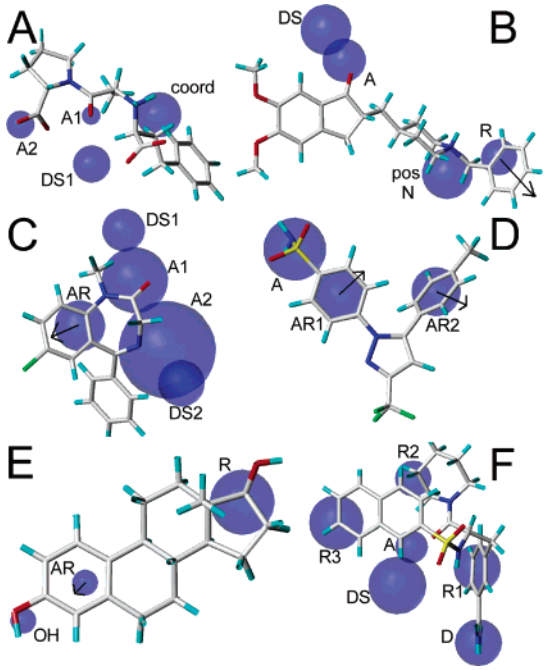

Figure 3. Pharmacophore queries for (A) ACE, (B) AchE, (C) BZR, (D) COX-2, (E) ER, and (F) THR. A is an acceptor atom, D a donor atom, DS a donor site, R a five- or six-member ring, and AR a five- or six-member aromatic ring. Arrows indicate plane normals. See text for other labels. The radii of the blue spheres depict the spatial constraints associated with each feature. Compounds shown are the same as in Figure 2 except (C) for which diazepam is shown.

a lack of variance at certain functional groups often reflects their necessary presence for maintaining activity (i.e., they define the pharmacophore). The benzamidine group of the thrombin inhibitors and the sulfonyl group of the COX-2 inhibitors are examples.

In summary, we prune regions of the RSM for which there are no nearby yellow contours, no blue or red contours deemed to have an enthalpic contribution to affinity, and regions near green contours for which there are no nearby grid points with (meaningful) negative PLS coefficients. We prune only if CoMSIA fields have sufficient variance. References to contour colors above assume the use of Sybyl defaults.

(iv) Query Definition. Queries for performing 3D database searches consist of a pharmacophore model used by itself or together with a RSM, encoded in Sybyl line notation (SLN). Pharmacophore models for each set were taken from the literature (Figure 3; see the Supporting Information for coordinates and constraints of features). Where available, we have used protein−ligand complex structures to help define pharmacophores because it was not the purpose of this work to test pharmacophore development methods. The ACE pharmacophore[31] consists of two carbonyl hydrogen bond acceptors, a zinc coordination atom, and a donor site on the carbonyl oxygen nearest the coordination atom. The coordination atom was restricted to either a carbonyl or phosphate oxygen atom, a non-aryl thiol or thioether sulfur atom, or a basic non-aryl nitrogen atom, all of which are present in the QSAR set. The pharmacophore tolerances were determined by superposing representative QSAR set molecules using the shared pharmacophore features. The AchE pharmacophore[25] consists of a ring, a positively charged nitrogen atom, and an acceptor atom−donor site pair (all references to rings indicate five- or six-member rings). Tolerances were deduced from the crystal structures of five AchE complexes (PDB codes 1eve, 1acj, 1dx6, 1gpn, 1h23). A BZR pharmacophore[27,32] consists of two donor sites, an acceptor site, and an aromatic ring. We have dropped the acceptor site that accounts for inverse-agonist activity because we make no distinction of activity types and have added loose constraints for the acceptor atoms that correspond to the donor sites because of the low selectivity of the query (see below). Feature tolerances were determined from the ligands diazadiindole, CGS-8216, RY-80, 6-PBC, CL-218872, zolpidem, BCCE, pyridodiindole, and diazepam (abbreviations are defined in ref 27). The COX-2 pharmacophore[33] consists of two aromatic rings and an oxygen acceptor atom. Tolerances were deduced from the crystal structures of three COX-2 complexes (PDB codes 1cx2, 3pgh, 4cox). The ER pharmacophore[28] consists of a hydroxy-substituted aromatic ring and a hydrophobic ring. Tolerances were deduced from the crystal structures of four ER complexes (PDB codes 1gwr, 1gwq, 3erd, 3ert). The thrombin pharmacophore[34] consists of three hydrophobic rings, a nitrogen donor atom, and an acceptor atom−donor site pair. Tolerances were deduced from crystal structures of seven thrombin complexes (PDB codes 1c4v, 1tom, 1d6w, 1d9i, 1dwd, 1fpc, 1d4p).

In all cases, tolerances were determined by superposing ligands using the shared pharmacophore features and the MULTIFIT routine in Sybyl. We did not vary the tolerances on pharmacophore features because they were adopted from published pharmacophores or deduced from multiple protein−ligand complexes.

RSMs were included in the query by adding the MOLCAD surface as a UNITY volume constraint using default parameters (expanding the surface by 1.0 Å outward and allowing penetration of the surface by up to 0.5 Å). UNITY automatically converts MOLCAD surfaces to excluded volume constraints by placing spheres in contact with the surface (outside for ligand-based surfaces) at regularly spaced points on it.

(v) Database Preparation. For the purpose of determining the sensitivity of queries, compounds having appropriate activity codes were selected from the MDDR database (MDL Information Systems, Inc.; San Leandro, CA) (activity codes: ACE 31410; AchE 09221; BZR 06210, 06211, 06212, 06213, 06216, 06214; COX-2 78454; ER 41300, 40210, 75711; THR 37110). For the AchE, BZR, COX-2, and ER MDDR sets having corresponding QSAR sets with a well-defined common scaffold, compounds containing the scaffold were removed. For the ACE and THR MDDR sets, compounds having a Tanimoto coefficient[35] $T_c \geq 0.85$ for any MDDR-QSAR set pair were removed. This allowed us to test the performance of queries for "extrapolating" beyond the QSAR set. Redundancy among the remaining MDDR compounds was reduced with a sphere-exclusion algorithm[36] implemented in Cerius2, version 4.8 (Accelrys, Inc.; San Diego, CA), using as threshold $1 - T_c = 0.15$. This gave sets for which all pairs of MDDR compounds have $T_c < 0.85$. All T_c values have been determined using 2D (structural) fingerprints in Cerius2. The threshold $T_c = 0.85$ is recommended for choosing diverse collections of compounds.[37]

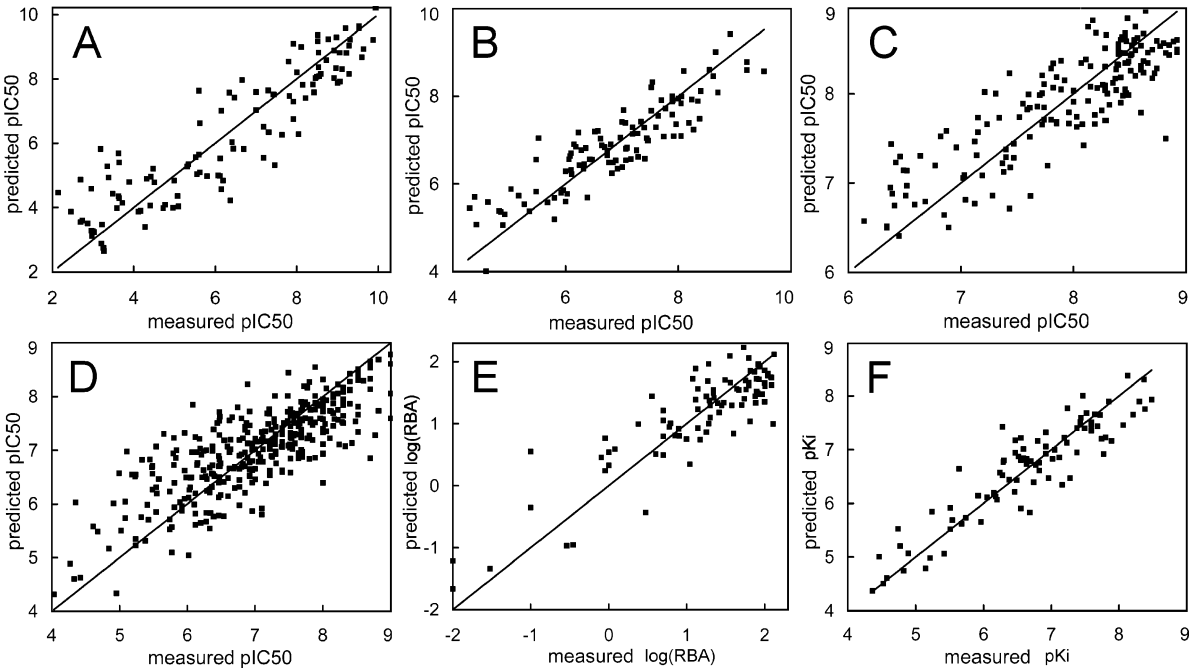

Figure 4. Fitted predictions vs measured activities from CoMSIA models for (A) ACE, (B) AchE, (C) BZR, (D) COX-2, (E) ER, and (F) THR.

The selectivity of each query was assessed using a database of >10000 organic compounds selected from the ChemACX-SC collection of >500000 screening compounds (Cambridgesoft Corp.; Cambridge, MA). The selection was not completely random but biased to reproduce the average profile of physicochemical properties of the six MDDR sets. Five such properties having relevance to "druglikeness" were used:[38] AlogP,[39] molecular weight, number of hydrogen bond donors and acceptors, and number of rotatable bonds. This biased selection was necessary for obtaining comparable estimates of the sensitivity and selectivity of queries because compounds from the ChemACX-SC collection are generally "leadlike", while those from MDDR are "druglike".[40] The set of ChemACX-SC compounds was further reduced to 10 000 compounds for which all pairs have $T_c < 0.85$ and all ChemACX-SC−MDDR pairs and ChemACX-SC−QSAR set pairs have $T_c < 0.85$. As such, it is reasonable to assume that the fraction of ChemACX-SC compounds showing activity at the various targets would be negligible if they were tested.

(vi) Database Searching. Databases searches were performed using the search3d expression generator in Sybyl. Two approaches were considered for treating the conformational flexibility of molecules. For each molecule, conformers were generated using (classic) distance geometry, followed by BFGS and truncated Newton minimization with the MM3 force field implemented in Tinker.[41] We have added to Tinker a parameter guessing routine[42] and a routine that reads Sybyl generated MM2/MM3 input files, allowing automatic atom-typing of molecules. This enables us to generate conformers for large databases in reasonable time by using distributed computing (ca. 8 h for 10 000 molecules with a 32 1.8 GHz processor PC cluster). The number of conformers was reduced by retaining only those having energies within 5.0 kcal/mol of the lowest-energy conformer and those having pairwise heavy-atom rmsd ≥ 1.0 Å after superposition.

Table 1. Summary Statistics for CoMSIA Analyses[a]

	ACE	AchE	BZR	COX-2	ER	THR
r^2	0.83	0.82	0.72	0.64	0.75	0.84
s	0.98	0.54	0.40	0.62	0.46	0.42
F	101.5	93.8	59.1	106.1	88.6	106.1
q^2	0.66	0.55	0.42	0.52	0.50	0.62
s_{PRESS}	1.35	0.85	0.57	0.71	0.65	0.64
components	5	5	7	6	3	4
fraction						
steric	0.47	0.32	0.30	0.21	0.21	0.52
electrostatic	0.53	0.68	0.70	0.79	0.79	0.48
n	114	111	166	373	91	88

[a] r^2 is the correlation coefficient; s is the standard error of prediction; F is Fischer's test of statistical significance; q^2 is the LOO cross-validated correlation coefficient; s_{PRESS} is the standard error from cross-validation; fraction indicates the relative importance of steric and electrostatic fields in the model; n is the number of compounds used for deriving models.

As an alternative to precomputed conformers, we considered conformational flexibility at search time using the directed-tweak algorithm[43] implemented in Sybyl/UNITY. As many as 10 conformers generated by torsional randomization were used as starting structures for directed-tweak optimization, and ring systems were treated flexibly; bump-checking was used to discard high-energy conformers. Other parameters were set to default values. We investigated the use of energy-filtering of hits returned by on-the-fly flexible searches using the MMFF94S force field in Sybyl. For pharmacophore queries, atoms not involved in defining pharmacophore elements were relaxed (maximum of 50 steps, 0 simplex steps). For RSM-containing queries, nonpharmacophore atoms were relaxed using a steep harmonic potential (500 kcal mol^{-1} Å^{-2} constant) applied when their coordinates differ from those returned by the flexible search by 0.5 Å or more. The energy of (relaxed) hits was compared to that of the lowest energy conformer obtained from distance geometry (described above, after reminimization of conformers with MMFF94S in Sybyl).

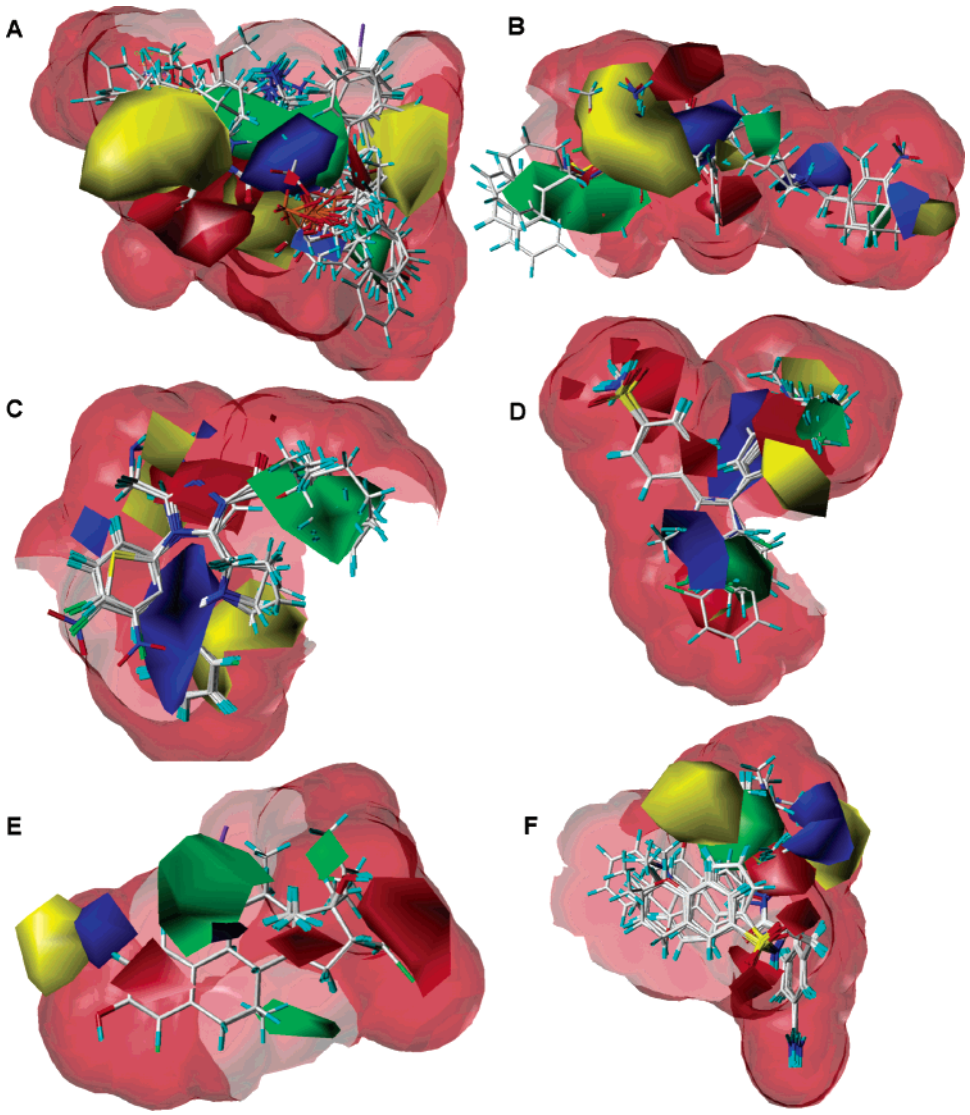

Figure 5. Pruned RSMs shown with CoMSIA model contours and the compounds used for defining the RSM: (A) ACE; (B) AchE; (C) BZR; (D) COX-2; (E) ER; (F) THR. Orientations of compounds are the same as in Figure 3.

Results

(i) CoMSIA Models. Statistically significant CoMSIA models were obtained for all six data sets (Table 1). Predicted vs actual activities are shown in Figure 4.

(ii) RSM Pruning. The resulting contour maps for steric and electrostatic fields were used to prune the RSMs created using the most active 20% of compounds in each data set (Figure 5). For the ACE RSM, a large section was pruned for which there was significant field variance but no corresponding effect on activity. For the AchE RSM, the model was pruned at the green contour encompassing bulky substituents at ring positions 5 and 6 of E2020 because no nearby grid points have negative steric coefficients. The opening can be seen to correspond with that observed in the crystal structure of the complex (Figure 6). For the BZR RSM, the model was pruned for substituents at ring positions 3, 6, and 7 of diazepam for which there are no nearby contours and at the surface defined by one side of bulky R groups for ester-containing derivatives (e.g., Ro14-5974). The opposite side of the RSM near the same green contour was not pruned because of the presence of nearby grid points having negative steric coefficients. The COX-2 RSM was opened at position 4 of the pyrazole ring for which there are no contours and corresponds reasonably well with a poorly explored pocket in the active site (Figure 6). The ER RSM was pruned at the 7α and 11β positions having a nearby green contour but no nearby grid points with negative steric coefficients. The two opened regions in the resulting pruned RSM correspond to large pockets in the active site that occupy nearly 200 Å^3 [29] (Figure 6). For thrombin, the RSM was pruned near the naphtho group for which there are no nearby contours and corresponds well with the solvent-exposed face of the active site (Figure 6).

(iii) Database Searches. We examined the performance of three types of queries for database searching: (1) pharmacophores used alone, (2) pharmacophores with unpruned RSMs, and (3) pharmacophores with pruned RSMs. For each biological target, the fraction of hits among MDDR compounds was used to determine the sensitivity of queries, or true positive rate. The ChemACX-SC compounds were used to determine the selectivity of queries, or false positive rate. MDDR compounds are henceforth referred to as actives, and ChemACX-SC compounds are referred to as inactives.

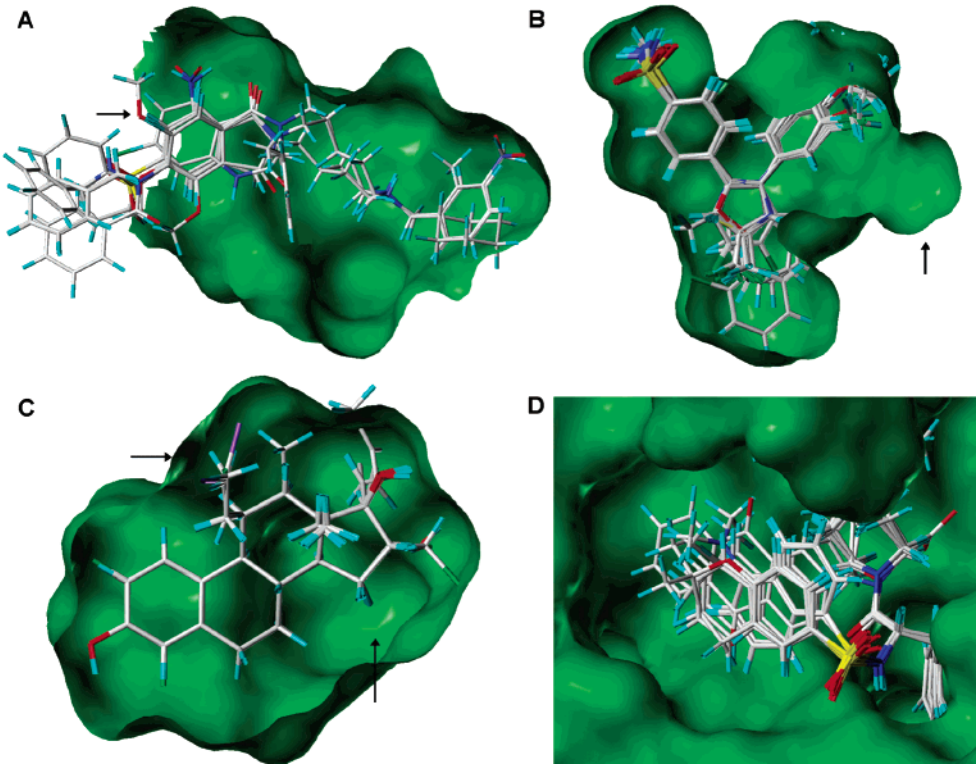

Figure 6. Crystal structure active site cavities obtained using the MOLCAD module of Sybyl: (A) AchE (PDB 1eve); (B) COX-2 (PDB 1cx2); (C) ER (PDB 1gwr); (D) THR (PDB 1ets). Orientations of compounds are the same as in Figures 3 and 5. Arrows indicate the pockets/solvent−protein interface that correspond to pruned regions in Figure 5. For THR, the ligands are visible through the solvent−protein interface.

Table 2. Comparison of Precomputed Conformers and On-the-Fly Flexibility for Database Searches Using Pharmacophore Queries

	ACE	ACHE	BZR	COX2	ER	THR
			Actives			
precomp conf[a] (%)	75.4	29.5	54.6	52.1	57.5	57.8
flexible[b] (%)	99.4	37.2	78.2	71.6	59.2	74.2
filter thresh[c] (kcal/mol)	22	14	7	7	15	22
% agree conf-flex[d]	64.9	93.2	72.2	87.4	96.6	75.9
rank (av ± σ)[e]	2.0 ± 1.8	1.8 ± 1.6	1.6 ± 1.5	1.8 ± 1.8	1.1 ± 0.5	1.9 ± 1.7
			Inactives			
precomp conf[a] (%)	0.88	10.1	41.5	28.7	2.11	1.7
flexible[b] (%)	3.65	17.4	75.2	57.1	3.89	3.55
filter thresh[c] (kcal/mol)	13	11	9	9	12	14
% agree conf-flex[d]	98.8	93.3	71.0	76.1	98.6	98.7
rank (av ± σ)[e]	2.4 ± 2.2	2.0 ± 2.0	2.3 ± 2.1	2.0 ± 1.9	2.1 ± 2.1	1.7 ± 1.4

[a] Percentage of actives or inactives retrieved using precomputed conformers. [b] Percentage of actives or inactives using flexible searching with no energy-filtering. [c] Energy threshold that gives the best agreement between hit rates from precomputed conformers and energy-filtered flexible searching. [d] As in footnote c but with the percentage of compounds having identical predictions of activity indicated. [e] Average and standard deviation of the number of flexible search hits that must be considered before one that satisfies the energy threshold for filtering is found.

The enrichment ratio, calculated as the fraction of actives contained in the screening hits divided by the fraction of actives in the entire database, is used to compare the performance of queries to random selection. Enrichment ratios greater than 1 indicate that a method is performing better than random selection.

(iii.a) Pharmacophore Query. All pharmacophore queries retrieve a larger fraction of actives than inactives (Table 2). The use of precomputed conformers gives greater enrichment ratios than does on-the-fly flexible searching with a simple bump-check (Table 4). The enrichment ratio for flexible searching can be improved by energy-filtering, in which the energy of query hits is compared to that of the lowest-energy conformer from distance geometry. We considered the conformation returned by the flexible search after relaxing parts of the molecule not used for defining the pharmacophore features. The energy threshold that gives the closest agreement with the hit rate obtained using precomputed conformers was identified (Table 2). For all targets, the selected thresholds give hit rates that differ by at most 3% of the hit rates for precomputed conformers (only integer values were considered for thresholds). There is good agreement between predictions from precomputed conformers and filtered flexible searching (Table 2). The greater range of thresholds for the actives compared to the inactives suggests that differences in data sets, not pharmacophores, are mostly responsible for the variation (the active compounds are substantially less diverse than the inactive compounds). The range 9−14 kcal/mol for the inactives might be further reduced and lie closer to the 5 kcal/mol threshold used for

precomputed conformers if pharmacophore elements were relaxed within their imposed tolerances (note that 9−14 kcal/mol refers to MMFF94S energies, while the 5 kcal/mol threshold for precomputed conformers refers to MM3 energies; after reminimization of precomputed conformers with MMFF94S in Sybyl, the 5 kcal/mol threshold is stretched to ∼10 kcal/mol). Because we prefer precomputed conformers for pharmacophore searches, we have not considered this possibility. Typical search times per molecule on a 1.8 GHz PC are 0.5 s for precomputed conformers, 0.7 s for flexible searching with no filtering, and 1.9 s when filtering with relaxation of hits. It is possible to use energy-filtering of flexible search hits without prior relaxation. However, the range of thresholds that gives closest agreement with searches using precomputed conformers is almost double that obtaining when relaxing hits. Dispensing with relaxation of hits results in only marginally faster searches (1.8 s/molecule).

As shown by the average rank of hits that satisfy the energy thresholds (Table 2), it is necessary to consider multiple hits from the flexible search when using energy-filtering. When increasing the maximum number of hits returned from 10 to 20, average increases in hit rates are 2% for actives and 7% for inactives, despite a large increase in search time. For pharmacophore searching with precomputed conformers, retaining conformers having pairwise heavy-atom rmsd ≥ 0.5 Å instead of 1.0 Å gives hit rates that are at most 3% higher but requires on average 346 kB storage per molecule compared to 230 kB when using the larger threshold.

Many MDDR compounds having AchE activity are tacrine and huperzine derivatives that do not contain the pharmacophore element A; this is reflected by the lower sensitivity of the query. We suspect that the low enrichment rate for the BZR query (described in ref 32) is caused by the use of site points instead of donor/acceptor atoms; examination of a typical organic molecule reveals a proliferation of UNITY site points about the molecule.

(iii.b) Pharmacophore + RSM Queries. We use a two-pass approach for queries containing a RSM. In the first pass, compounds that satisfy the pharmacophore are identified, using precomputed conformers for treating conformational flexibility. In the second pass, the hits are evaluated further with the RSM-containing queries. We have found it necessary to use flexible searching for the second pass; using precomputed conformers retrieves 25−50% of actives identified using flexible searching with energy-filtering. Queries containing pruned RSMs retrieve a larger fraction of actives than unpruned RSMs (Table 3) while yielding enrichment ratios similar to those obtained using unpruned RSMs and greater than those obtained using pharmacophores alone (Table 4). In contrast to the behavior observed for pharmacophore searches, there are only small differences in enrichment ratios for flexible searching with or without energy-filtering. By use of precomputed conformers for the initial pharmacophore search, compounds that cannot satisfy the pharmacophore in a low-energy conformation have been excluded. The search without energy-filtering is substantially faster because it proceeds to the next compound as soon as one hit has been identified. When using energy-filtering, it is necessary to consider multiple hits from the flexible search. As for the filtered flexible pharmacophore search, increasing the maximum number of hits returned from 10 to 20 leads to marginal increases in the number of actives that satisfy the imposed energy threshold. Typical search times per compound for flexible searching are 50 s with energy-filtering and 20 s without with energy-filtering (excluding a 12 s setup time, which must be done once for the database in UNITY but for every compound using %search3d in Sybyl).

Nearly identical sensitivity and selectivity for the pruned and unpruned AchE RSMs are observed because the radius of gyration of compounds used for defining the RSM is greater than that for any of the hits.

In some cases, using precomputed conformers for the pharmacophore search may be impractical because of CPU and storage usage. Instead, pharmacophore hits can be identified with a flexible search and checked against the RSM-containing queries in a second pass (no energy-filtering for either search). The pruned RSMs retrieve a greater fraction of actives than unpruned RSMs (not shown), with enrichment rates for both pruned and unpruned RSMs exceeding those obtained using the pharmacophore alone (Table 4).

Table 3. Comparison of Database Searches Using RSM-Containing Queries for Refining Precomputed Conformer Pharmacophore Hits

	ACE	ACHE	BZR	COX2	ER	THR
			Actives			
unpruned filter[a] (%)	40.9	8.0	7.9	36.7	8.9	5.3
pruned filter (%)	45.0	8.3	30.3	39.1	49.3	16.7
unpruned no filter[b] (%)	63.2	9.4	9.5	37.7	8.9	7.7
pruned no filter (%)	69.0	9.4	34.1	40.5	51.4	23.2
			Inactives			
unpruned filter[a] (%)	0.22	1.6	4.6	5.9	0.35	0.06
pruned filter (%)	0.24	1.6	20.6	6.3	0.51	0.10
unpruned no filter[b] (%)	0.59	2.0	5.9	7.7	0.46	0.09
pruned no filter (%)	0.69	2.0	25.2	8.6	0.66	0.14

[a] Percentage of actives or inactives retrieved using flexible searches with energy-filtering (20 kcal/mol threshold). [b] Without energy-filtering.

Table 4. Enrichment Rates Obtained Using Pharmacophore and RSM-Containing Queries

	ACE	ACHE	BZR	COX2	ER	THR
no. of actives:	171	339	317	215	146	431
			Pharmacophore			
precomp conf	35.4	2.7	1.3	1.8	19.8	14.4
flexible	18.9	2.1	1.0	1.2	12.6	11.5
			Pharmacophore + RSM on Precomputed Conformer Pharmacophore Hits			
unpruned filter	45.3	4.3	1.7	5.7	18.8	19.2
pruned filter	45.3	4.5	1.5	5.6	40.7	21.3
unpruned no filter	38.5	4.3	1.6	4.5	15.3	19.0
pruned no filter	37.5	4.3	1.3	4.4	37.0	21.2
			Pharmacophore + RSM on Flexible Search Pharmacophore Hits (No Filter)			
unpruned no filter	24.8	3.3	1.1	3.2	10.1	18.6
pruned no filter	24.3	3.3	1.1	3.2	29.2	19.7

Discussion

We have described a method that allows users to develop pruned receptor surface models with tools that are available in widely used software packages. The

simple definition of the RSM allows for rapid identification of steric clashes within the putative receptor. When combined with a pharmacophore, the query can be used for efficient 3D database searching. The performance of pruned RSMs used in conjunction with pharmacophores was compared to the performance of the pharmacophores used alone or the pharmacophores with unpruned RSMs. Here, performance is quantified by examining the fraction of known actives retrieved by the query (true positives) and the fraction of compounds retrieved from a database of inactives (false positives). In virtual screening, one also makes reference to the enrichment rate, or the fraction of actives contained in the screening hits divided by the fraction of actives in the entire database.

The performance of the three approaches for database screening is generally consistent with our expectations based on intuition. The pharmacophore query retrieves the largest fraction of known actives but suffers from a high rate of false positives. The resulting enrichment rate is therefore the lowest of the three approaches. The use of a pharmacophore and unpruned RSM retrieves the smallest fraction of known actives but also retrieves fewer inactive compounds. Larger enrichment rates are observed at the expense of missing many actives. When a pruned RSM is used together with a pharmacophore, the enrichment rate is similar to that for the unpruned RSM, but the fraction of known actives retrieved is greater. As such, using the pruned RSM represents a reasonable compromise between the high specificity of the unpruned RSM and the high sensitivity of the pharmacophore query. All three approaches performed better than random screening.

While additional fields are sometimes used in 3D-QSAR modeling, we have chosen to use only steric- and electrostatic-type fields because we seek to identify regions in space at which receptor−ligand interactions occur. In other words, we wish to identify enthalpic contributions to binding. As for many approaches to QSAR, this assumes that the variation in binding free energy is largely governed by differences in binding enthalpies. The success of CoMFA[23] in deriving QSAR models for hundreds of data sets and the observation that there exists a good correlation between ΔH and $T\Delta S$ in some congeneric series of compounds[44] suggest that this is often a reasonable supposition.

Our use of CoMSIA contours for pruning RSMs assumes that the fields they enclose are causal variables in the SAR, not lurking variables simply correlated with those truly responsible for variation in activity. Some workers have cautioned against viewing the contours as a low-resolution representation of the active site.[23,45,46] There exist many published examples of 3D-QSAR models for which contours were found to be compatible with the positions of active site residues (e.g., refs 47−50). In the present work, the positions of most contoured regions were found to be compatible with active site residues when such a comparison was possible (all but ACE and BZR). The exceptions could be traced to two-level variables, for which fields fall near two distinct values; they are essentially sophisticated indicator variables and can be easily identified.

It has been suggested that defining unpruned RSMs using all QSAR compounds rather than only the most active 20% may provide more tolerant shape constraints at the parts of the pruned RSMs that are open, achieving a similar result without the additional complications of pruning. The problem with this approach is that some low-activity compounds will have bulky groups directly overlapping the active site residues, loosening shape constraints at those parts where they are most necessary. For example, the large yellow contour in Figure 5F coincides with Tyr60A in the thrombin active site; it is not occupied by any of the most active 20% of compounds but is occupied by several low-activity compounds. Using all compounds for defining an unpruned thrombin RSM yields an enrichment rate of 16.1 compared to 14.4 for the pharmacophore used alone, 21.3 for the pruned RSM, and 19.2 for the unpruned RSM defined with the most active compounds (Table 4). For this reason, RSMs are usually developed with high-activity compounds only.[11]

There is a degree of subjectivity involved in choosing a CoMSIA contouring level, whether certain parts of the RSM are sufficiently far from contoured regions to warrant their removal, whether to prune the receptor in regions that show only positive steric contributions, and whether to open parts of the RSM where the data set shows no field variance. The smooth and continuous contours obtained from CoMSIA compared to "traditional" field-based 3D-QSAR methods (e.g., CoMFA,[23] GRID[45]) reduce difficulties in selecting which parts of the RSM to prune, although other approaches that improve the "interpretability" of contours could also be used.[51,52] If there is a lack of variation among the ligands that is not associated with pharmacophore features, it is sensible to synthesize and test a few additional derivatives that have a clear relationship to the existing compounds rather than attempting to obtain this information from disparate compounds identified by the database search. In many cases, examining the performance of different database queries in retrieving known actives can help one choose that most promising for identifying new leads.

While not the primary purpose of this work, some remarks can be made about conformational treatment in 3D database searching. We have considered two approaches: using precomputed conformers and treating conformational flexibility at search time. Smellie et al. have indicated that precomputed conformers retrieve a large fraction of true positives while minimizing false positives,[53] but Pearlman claims that flexibility should be considered at search time.[54] When only a pharmacophore is used, the directed tweak method[43] with a simple bump-check gives a higher rate of false positives than observed with precomputed conformers, yielding lower enrichment rates. As we have shown, the use of a force field energy filter gives results similar to those obtained with precomputed conformers. It is necessary to consider multiple hits from the directed-tweak search, and the use of energy-filtering requires that the global energy minimum be identified. Therefore, the use of precomputed conformers is more efficient. However, it may not always be necessary to use energy-filtering of flexible search hits. For comparing or designing combinatorial libraries, knowing the relative quantity of hits may be sufficient. Treating conformational flexibility at search time would be more efficient, both for the total

CPU time used (including the time required to generate conformers) and storage requirements.

The situation is different when the query includes a pruned or unpruned RSM. When using precomputed conformers, very few actives were identified. It was necessary to consider flexibility at search time, allowing for the search to "fold up" molecules into the RSM. In contrast, pharmacophores involve core parts of molecules, causing the search to be less sensitive to the particular conformation of substituents. These observations are consistent with the conclusions from a comparison of active site bound and low free energy conformations.[55] The energy threshold (20 kcal/mol) that we have adopted for filtering hits of the RSM pharmacophore query appears high in light of the <5 kcal/mol difference often observed between the lowest energy and active site bound conformers.[56] This discrepancy may be explained with the following arguments: we use the in vacuo conformer energy, not that in solution; allowing relaxation of the pharmacophore elements within the imposed constraints would cause a reduction in the energy; the threshold is not too distant from the range of values found to give good agreement with the results from precomputed conformers when using only a pharmacophore; a pruned RSM is at best only a crude model of the active site and likely remains overconstraining. This threshold is the same as that suggested by Hann et al.[17] in their use of unpruned RSMs for database searches. For greater efficiency, it is possible to dispense with energy-filtering when low-energy pharmacophore hits are subsequently filtered through the RSM, with only small decreases in enrichment rates.

We are not the first to recognize that receptor models are sterically overconstrained. Hahn et al.[11] have made provisions in their program allowing the user to prune the receptor surface model although it provides no assistance in determining if or where the model should be pruned. The genetic algorithms used in the semiatomistic model development methods of Walters et al.[13] and Vedani et al.[14] can be used to automatically develop active site models where regions are not delimited by particles; both methods are directed toward QSAR rather than 3D database searching applications. To our knowledge, neither approach has been validated by examining their ability to correctly identify such regions in the receptors of proteins having known structures. Recently, Van Drie has described the extension of his "shrink-wrap" surfaces[12] allowing compounds to protrude through parts of the surface that are not traversed by inactive compounds.[6] This approach is distinct from that presented here. Unfortunately, he provides no comparison of the modified shrink-wrap surfaces to queries lacking shape constraints or the unmodified surfaces for database searching. Among the methods that have been applied to virtual screening of databases, this issue has been largely neglected.

In summary, we have presented a method for developing pruned receptor surface models that are more useful than unpruned models in identifying potential lead compounds. Using screening sets for six biological targets, we have shown how pruned RSMs yield higher selectivity than pharmacophore models used alone, without the large decrease in sensitivity typical of unpruned RSMs. This methodology, which can be readily used with existing 3D database search methods, should prove to be useful in identifying new leads when structural information regarding the target is not available.

Acknowledgment. D.F.W. acknowledges support from the Natural Sciences and Engineering Research Council (NSERC), the Canadian Institutes of Health Research (CIHR), and the Canada Research Chairs Program. J.J.S. acknowledges support from a CIHR doctoral research award.

Supporting Information Available: Definitions of pharmacophore queries and lists of QSAR set compounds. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Stone, M.; Jonathan, P. Statistical thinking and technique for QSAR and related studies. 1. General theory. *J. Chemom.* **1993**, *7*, 455−475.
(2) Oprea, T. I.; Waller, C. L. Theoretical and practical aspects of three-dimensional quantitative structure−activity relationships. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 1997; pp 127−182.
(3) Muegge, I.; Rarey, M. Small molecule docking and scoring. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 2001; pp 1−60.
(4) Ghose, A. K.; Wendoloski, J. J. Pharmacophore modelling: Methods, experimental verification and applications. *Perspect. Drug Discovery Des.* **1998**, *9*−*11*, 253−271.
(5) Good, A. C.; Mason, J. S.; Pickett, S. D. Pharmacophore pattern application in virtual screening, library design and QSAR. In *Virtual Screening for Bioactive Molecules*; Bohm, H. J., Schneider, G., Eds.; Wiley-VCH: New York, 2000; pp 131−159.
(6) van Drie, J. H. Pharmacophore discovery—lessons learned. *Curr. Pharm. Des.* **2003**, *9*, 1649−1664.
(7) Lemmen, C.; Lengauer, T. Computational methods for the structural alignment of molecules. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 215−232.
(8) Mason, J. S.; Good, A. C.; Martin, E. J. 3-D pharmacophores in drug discovery. *Curr. Pharm. Des.* **2001**, *7*, 567−597.
(9) Hecker, E. A.; Duraiswami, C.; Andrea, T. A.; Diller, D. J. Use of catalyst pharmacophore models for screening of large combinatorial libraries. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1204−1211.
(10) Vedani, A.; Zbinden, P.; Snyder, J. P.; Greenidge, P. A. Pseudoreceptor modeling. The construction of 3-dimensional receptor surrogates. *J. Am. Chem. Soc.* **1995**, *117*, 4987−4994.
(11) Hahn, M. Receptor surface models. 1. Definition and construction. *J. Med. Chem.* **1995**, *38*, 2080−2090.
(12) Van Drie, J. H. "Shrink-wrap" surfaces: A new method for incorporating shape into pharmacophoric 3D database searching. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 38−42.
(13) Walters, D. E.; Hinds, R. M. Genetically evolved receptor models. A computational approach to construction of receptor models. *J. Med. Chem.* **1994**, *37*, 2527−2536.
(14) Vedani, A.; Dobler, M.; Zbinden, P. Quasi-atomistic receptor surface models: A bridge between 3-D QSAR and receptor modeling. *J. Am. Chem. Soc.* **1998**, *120*, 4471−4477.
(15) Jain, A. N.; Dietterich, T. G.; Lathrop, R. H.; Chapman, D.; Critchlow, R. E.; et al. Compass, a shape-based machine learning tool for drug design. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 635−652.
(16) Crippen, G. M. Validation of EGSITE2, a mixed integer program for deducing objective site models from experimental binding data. *J. Med. Chem.* **1997**, *40*, 3161−3172.
(17) Hahn, M. Three-dimensional shape-based searching of conformationally flexible compounds. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 80−86.
(18) Mount, J.; Ruppert, J.; Welch, W.; Jain, A. N. Icepick: A flexible surface-based system for molecular diversity. *J. Med. Chem.* **1999**, *42*, 60−66.
(19) Srinivasan, J.; Castellino, A.; Bradley, E. K.; Eksterowicz, J. E.; Grootenhuis, P. D. J.; et al. Evaluation of a novel shape-based computational filter for lead evolution: Application to thrombin inhibitors. *J. Med. Chem.* **2002**, *45*, 2494−2500.
(20) van Drie, J. H.; Weininger, D.; Martin, Y. C. Aladdin: An integrated tool for computer-assisted molecular design and pharmacophoric pattern recognition from geometric, steric and substructure searching of three-dimensional molecular structures. *J. Comput.-Aided Mol. Des.* **1989**, *3*, 225−251.

(21) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indexes in a comparative-analysis (CoMSIA) of drug molecules to correlate and predict their biological-activity. *J. Med. Chem.* **1994**, *37*, 4130−4146.
(22) Sutherland, J. J.; O'Brien, L. A.; Weaver, D. F. Spline-fitting with a genetic algorithm: A method for developing classification structure−activity relationships. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1906−1915.
(23) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steriods to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(24) Depriest, S. A.; Mayer, D.; Naylor, C. B.; Marshall, G. R. 3D-QSAR of angiotensin-converting enzyme and thermolysin inhibitors. A comparison of CoMFA models based on deduced and experimentally determined active-site geometries. *J. Am. Chem. Soc.* **1993**, *115*, 5372−5384.
(25) Golbraikh, A.; Bernard, P.; Chretien, J. R. Validation of protein-based alignment in 3D quantitative structure−activity relationships with CoMFA models. *Eur. J. Med. Chem.* **2000**, *35*, 123−136.
(26) Haefely, W.; Kyburz, E.; Gerecke, M.; Mohler, H. Recent advances in the molecular pharmacology of benzodiazepine receptors and in the structure−activity relationships of their agonists and antagonists. *Adv. Drug Res.* **1985**, *14*, 165−322.
(27) Huang, Q.; He, X. H.; Ma, C. R.; Liu, R. Y.; Yu, S.; et al. Pharmacophore/receptor models for GABA(A)/BZR subtypes ($\alpha1\beta3\gamma$ 2, $\alpha5\beta3\gamma2$, and $\alpha6\beta3\gamma2$) via a comprehensive ligand-mapping approach. *J. Med. Chem.* **2000**, *43*, 71−95.
(28) Anstead, G. M.; Carlson, K. E.; Katzenellenbogen, J. A. The estradiol pharmacophore: Ligand structure−estrogen receptor binding affinity relationships and a model for the receptor binding site. *Steroids* **1997**, *62*, 268−303.
(29) Gao, H.; Katzenellenbogen, J. A.; Garg, R.; Hansch, C. Comparative QSAR analysis of estrogen receptor ligands. *Chem. Rev.* **1999**, *99*, 723−744.
(30) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990**, *11*, 431−439.
(31) Mayer, D.; Naylor, C. B.; Motoc, I.; Marshall, G. R. A unique geometry of the active site of angiotensin-converting enzyme consistent with structure−activity studies. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 3−16.
(32) Zhang, W.; Koehler, K. F.; Zhang, P.; Cook, J. M. Development of a comprehensive pharmacophore model for the benzodiazepine receptor. *Drug Des. Discovery* **1995**, *12*, 193−248.
(33) Palomer, A.; Cabre, F.; Pascual, J.; Campos, J.; Trujillo, M. A.; et al. Identification of novel cyclooxygenase-2 selective inhibitors using pharmacophore models. *J. Med. Chem.* **2002**, *45*, 1402−1411.
(34) Patel, Y.; Gillet, V. J.; Bravi, G.; Leach, A. R. A comparison of the pharmacophore identification programs: Catalyst, DISCO and GASP. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 653−681.
(35) Willett, P.; Winterman, V. A comparison of some measures for the determination of intermolecular structural similarity measures of intermolecular structural similarity. *Quant. Struct.−Act. Relat.* **1986**, *5*, 18−25.
(36) Clark, R. D. Optisim: An extended dissimilarity selection method for finding diverse representative subsets. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1181−1188.
(37) Matter, H. Selecting optimally diverse compounds from structure databases: A validation study of two-dimensional and three-dimensional molecular descriptors. *J. Med. Chem.* **1997**, *40*, 1219−1229.
(38) Muegge, I. Selection criteria for drug-like compounds. *Med. Res. Rev.* **2003**, *23*, 302−321.
(39) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. Prediction of hydrophobic (lipophilic) properties of small organic molecules using fragmental methods: An analysis of AlogP and ClogP methods. *J. Phys. Chem. A* **1998**, *102*, 3762−3772.
(40) Oprea, T. I.; Davis, A. M.; Teague, S. J.; Leeson, P. D. Is there a difference between leads and drugs? A historical perspective. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1308−1315.
(41) Ponder, J. W. *Tinker: Software Tools for Molecular Design*, version 4.1; http://dasher.wustl.edu/tinker/.
(42) Schnur, D. M.; Grieshaber, M. V.; Bowen, J. P. Development of an internal searching algorithm for parameterization of the MM2/MM3 force-fields. *J. Comput. Chem.* **1991**, *12*, 844−849.
(43) Hurst, T. Flexible 3D searching−the directed tweak technique. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 190−196.
(44) Searle, M. S.; Williams, D. H. The cost of conformational order. Entropy changes in molecular associations. *J. Am. Chem. Soc.* **1992**, *114*, 10690−10697.
(45) Cruciani, G.; Watson, K. A. Comparative molecular-field analysis using GRID force-field and GOLPE variable selection methods in a study of inhibitors of glycogen-phosphorylase-b. *J. Med. Chem.* **1994**, *37*, 2589−2601.
(46) Greco, G.; Novellino, E.; Martin, Y. C. Approaches to three-dimensional quantitative structure−activity relationships. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 1997; pp 183−240.
(47) Oprea, T. I.; Waller, C. L.; Marshall, G. R. 3D-QSAR of human immunodeficiency virus (I) protease inhibitors. III. Interpretation of CoMFA results. *Drug Discovery Des.* **1994**, *12*, 29−40.
(48) Kroemer, R. T.; Ettmayer, P.; Hecht, P. 3D-quantitative structure−activity-relationships of human immunodeficiency virus type-1 proteinase inhibitors. Comparative molecular-field analysis of 2-heterosubstituted statine derivatives. Implications for the design of novel inhibitors. *J. Med. Chem.* **1995**, *38*, 4917−4928.
(49) Bohm, M.; Sturzebecher, J.; Klebe, G. Three-dimensional quantitative structure−activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. *J. Med. Chem.* **1999**, *42*, 458−477.
(50) Chavatte, P.; Yous, S.; Marot, C.; Baurin, N.; Lesieur, D. Three-dimensional quantitative structure−activity relationships of cyclo-oxygenase-2 (COX-2) inhibitors: A comparative molecular field analysis. *J. Med. Chem.* **2001**, *44*, 3223−3230.
(51) Cho, S. J.; Tropsha, A. Cross-validated r(2)-guided region selection for comparative molecular-field analysis. A simple method to achieve consistent results. *J. Med. Chem.* **1995**, *38*, 1060−1066.
(52) Pastor, M.; Cruciani, G.; Clementi, S. Smart region definition: A new way to improve the predictive ability and interpretability of three-dimensional quantitative structure−activity relationships. *J. Med. Chem.* **1997**, *40*, 1455−1464.
(53) Smellie, A.; Kahn, S. D.; Teig, S. L. Analysis of conformational coverage. 2. Application of conformational models. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 295−304.
(54) Pearlman, R. S. 3D molecular structures: Generation and use in 3D searching. In *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; pp 41−79.
(55) Vieth, M.; Hirst, J. D.; Brooks, C. L. Do active site conformations of small ligands correspond to low free-energy solution structures? *J. Comput.-Aided Mol. Des.* **1998**, *12*, 563−572.
(56) Bostrom, J.; Norrby, P. O.; Liljefors, T. Conformational energy penalties of protein-bound ligands. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 383−396.

JM049896Z