PRO_LIGAND: An Approach to *de Novo* Molecular Design. 2. Design of Novel Molecules from Molecular Field Analysis (MFA) Models and Pharmacophores

Bohdan Waszkowycz, David E. Clark, David Frenkel, Jin Li, Christopher W. Murray, Barry Robson,* and David R. Westhead

Proteus Molecular Design Ltd., Proteus House, Lyme Green Business Park, Macclesfield, Cheshire SK11 0JL, United Kingdom

Received May 18, 1994®

A computational approach for molecular design, PRO_LIGAND, has been developed within the PROMETHEUS molecular design and simulation system in order to provide a unified framework for the *de novo* generation of diverse molecules which are either similar or complementary to a specified target. In this instance, the target is a pharmacophore derived from a series of active structures either by a novel interpretation of molecular field analysis data or by a pharmacophore-mapping procedure based on clique detection. After a brief introduction to PRO_LIGAND, a detailed description is given of the two pharmacophore generation procedures and their abilities are demonstrated by the elucidation of pharmacophores for steroid binding and ACE inhibition, respectively. As a further indication of its efficacy in aiding the rational drug design process, PRO_LIGAND is then employed to build novel organic molecules to satisfy the physicochemical constraints implied by the pharmacophores.

Introduction

With the increasing importance of structure-based drug design in pharmaceutical research,[1-5] there has recently been much interest in the development of computer programs which automate the *de novo* design of molecular structures satisfying a set of steric and chemical constraints.[6-25] In a recent paper,[26] we have described our own approach for *de novo* molecular design called PRO_LIGAND and demonstrated its application to the design of organic molecules, with particular emphasis on the design of inhibitors of enzymes of known structure.

However, while the number of protein structures determined by X-ray crystallography and NMR is ever increasing and likely to grow rapidly within the foreseeable future,[27,28] there are many instances where the 3-D structure of the proposed biochemical target has not yet been elucidated. In such a situation, more indirect methods must be adopted. For instance, a series of analogues to a known lead may be synthesized and assayed for activity. In order to rationalize subsequent optimization of the lead molecule, it is necessary that the available structure–activity data are analyzed in order to define the structural and physicochemical properties which are required for biological activity. More specifically, we require that a pharmacophore is defined, which can be used as a template for the automated design of new molecules within PRO_LIGAND.

In this paper we describe two different approaches for objectively extracting a pharmacophore: molecular field analysis (MFA)[29] and pharmacophore mapping[30] and discuss the use of the PRO_LIGAND program in the generation of novel molecules which satisfy the pharmacophoric constraints derived from series of known active molecules.

***De Novo* Design with PRO_LIGAND.** As the previous paper in this series has detailed, PRO_LIGAND is a *de novo* design program which is an integral part of our in-house molecular design and simulation system, PROMETHEUS. The modules comprising PRO_LIGAND and their operation are described below. For fuller details, the reader is referred to ref 26.

The normal sequence of events in applying PRO_LIGAND to a molecular design study is as follows. The first stage is the definition of a *design base* from one or more input molecular structures. The design base represents the key structural features which will guide the design process and typically requires the extraction of the active site from a receptor or, as in the present example, the extraction of a pharmacophore from a set of active analogues.

Next, a *design model* is constructed, which is a 3-D template that describes the idealized steric and hydrogen-bonding features of the chemical structures to be designed. These features are represented by *interaction sites*.[12,13,31] Hydrogen bond acceptors and donors are represented by **A-Y** and **D-X** vectors, respectively, while lipophilic regions are characterized by L or R points according to whether the site is aliphatic or aromatic in nature. These sites are generated to be either complementary or similar to the design base atoms, depending on whether the object is to design a molecule to fit into a known receptor or to mimic a set of active analogues. The type and location of these sites are generated via a user-definable rule base.

The *structure generation* module produces molecular structures consistent with the design model by assembling small 3-D molecular fragments from preconstructed libraries. These library fragments are labeled to indicate the types of interaction sites they may match, and a rapid graph–theoretical algorithm is used to seek fits of the fragments on to the design model. The fitting procedure also corrects or eliminates any bad inter- or intramolecular van der Waals' clashes. A great variety of modes of fragment assembly are available to the user, including a continuous growth procedure and procedures for inter- and intrafragment bridging. In addition, the user also has full control over the structuring and ranking of the fragment libraries. Each gener-

* To whom all correspondence should be addressed.
® Abstract published in *Advance ACS Abstracts*, October 1, 1994.

0022-2623/94/1837-3994$04.50/0

ated solution is scored on the basis of the number of design model features it has succeeded in fulfilling and on certain structural characteristics, such as the number of rings or asymmetric carbon atoms.

Finally, the user may subject the built structures to a *structure refinement* stage, in which a genetic algorithm[32] approach is employed to breed further high-scoring structures from those produced by structure generation. This module will be described in detail in a future paper.

How information from MFA or pharmacophore mapping can be integrated into this strategy will be detailed in the following sections.

Construction of a Design Model by MFA. A PRO_LIGAND design model may be derived from a 3D-QSAR analysis of a set of molecules with known biological activities. The method we have used here is MFA, using software developed in-house which follows closely the approach of Cramer's comparative molecular field analysis (CoMFA). Below we offer a description of our method and its adaptation to pharmacophore identification.

The first step in MFA is the selection of a low-energy conformer of each molecule, followed by superimposition to a common reference frame by an appropriate alignment procedure. For each molecule, a molecular mechanics potential energy field is plotted on a cubic grid in 3-D space, using a suitable probe atom. The potential energy terms are generally separated into steric and electrostatic energies. The set of interaction energies derived for specific grid points is used as the set of 3D-QSAR parameters to describe each molecule in the subsequent regression analysis. Because the number of parameters (*i.e.*, grid energies) generally far exceeds the number of molecules, a component-extractive regression procedure such as partial least squares (PLS) is used.[33,34] This method extracts the orthogonal latent variables which optimally describe the variance within both the descriptor matrix and the activity matrix while ensuring maximal correlation between the descriptor and activity components.

In this paper we do not specifically address any deficiencies in the standard MFA method (such as the need for better procedures for structural alignment) but, rather, focus on the derivation of a pharmacophore. In order to define a pharmacophore from the MFA, we have developed a method of interpreting the PLS regression coefficients which to our knowledge has not been reported previously in the literature. Generally, a MFA is interpreted in terms of 3-D plots of significant PLS regression coefficients, each regression coefficient being associated with a particular grid point in 3-D space and a particular energy component (steric, electrostatic, etc.). Visual inspection of significant positive and negative coefficients allows the identification of regions of 3-D space in which steric interactions and electrostatic interactions are favored or disfavored. A problem with this approach is that the user tends to derive a qualitative rather than a quantitative judgement of which chemical features are correlated with activity. For example, a substituent which is too bulky may be disfavored in a certain region of space, but it may not be obvious what the optimal size of substituent should be. Similarly, a hydrogen bond donor in the ligand may be required to match a region of positive electrostatic coefficients, but it may be difficult to derive the optimal position and direction of the hydrogen bond donor.

Our approach for the interpretation of the PLS model is to calculate the contributions of individual atoms to the predicted activity of the molecule. It is also possible to decompose these atomic contributions into contributions from each energy term or principal component. We will use the acronym PACs to denote these *predicted activity contributions*. The advantage of this approach is that it allows the user to identify quantitatively those atoms which are responsible for the activity of the molecule and to explore the relative importance of individual energy terms and principal components.

It is worth noting that the derivation of PACs is simply a method of interpreting an MFA model: it is not capable of optimizing the parameters involved in the generation of the model. The quality of any MFA is dependent on many parameters, and of particular importance are the chosen conformations and alignments. All such parameters need to be optimized in terms of the predictive value of the resulting MFA model. Note also that the quality of the pharmacophore which is extracted by the analysis of PACs will always be limited by the nature of the training set: as with any MFA, one must ensure that the training set fully explores all areas of substituent space which are considered to be relevant.

As an example of the use of PACs in defining a pharmacophore, we have examined a typical QSAR study from the literature: a set of 35 steroids with reported binding affinities to the progestogen and androgen receptors.[35] In such an example, a typical design goal is to define a pharmacophore which represents the key chemical features required to bind to a particular receptor and then to use PRO_LIGAND to generate a set of novel molecules to fit the pharmacophore.

Construction of a Design Base by Pharmacophore Mapping. In some instances, the MFA technique may not be applicable in a straightforward manner. This is particularly so when the set of active structures shows considerable structural diversity and when the structures are conformationally flexible. In such a scenario, the choice of a suitable conformer for each structure and the subsequent alignment can be problematic. An alternative strategy to MFA in such circumstances is *pharmacophore mapping*—the elucidation of a set of structural features common to all the members of an active series which may be responsible for their observed activity.

Several strategies for pharmacophore mapping have been described in the literature including the well-established active analog approach[36-38] and ensemble distance geometry[39] as well as some more recent innovations.[40-42] For the purposes of PRO_LIGAND, we decided to adopt one of these newer techniques: the clique-detection approach described by Martin *et al.*[40]

The input to this algorithm is a set of low-energy conformations for each member of the active series. One member is then selected to be the reference compound, and the algorithm then takes one of its conformations and compares it with all the conformations of the other structures in the series. Each pairwise comparison uses the Bron–Kerbosch[43] clique-detection algorithm to determine the maximal common substructure (MCS) of the

two conformations. The MCSs from each pairwise comparison are stored, and when all the comparisons have been performed, an algorithm originally described by Brint and Willett[44] is used to extract the MCS that is common to all members of the series. If such a set of structural features exists, it constitutes a possible pharmacophore. If no MCS can be found for the whole series, the algorithm cycles to the next conformation of the reference compound and repeats the process until a solution is found or all the conformations of the reference compound have been investigated. A distance tolerance for the matching process can be specified by the user.

The clique-detection method has several advantages compared to existing strategies for pharmacophore mapping.[40] Firstly, there is the decoupling of the conformational analysis procedure from the pharmacophore elucidation step. This means that once the conformational analysis has been performed, any number of experiments can be performed upon the resulting conformers with minimal computational expense. The other techniques have the conformational analysis as an integral part of the pharmacophore elucidation. Secondly, the conformations involved in the clique-detection procedure are guaranteed to be low-energy, whereas in other techniques the solutions generated may be geometrically feasible but energetically unsatisfactory. Finally, and perhaps most significantly, the clique-detection method identifies both the bioactive conformations and the superposition rule simultaneously. In other words, the algorithm not only discovers those conformations capable of fitting the pharmacophore but also discovers the pharmacophore itself. This is in contrast to the other procedures which require the user to suggest a set of pharmacophoric points beforehand to which the algorithms seek to match the set of structures.

In PRO_LIGAND, each generated pharmacophore constitutes a viable design base. Interaction sites are then constructed to be either similar or complementary to the design base as desired, in order to form the design model required by the structure generation module.

Methods

Calculation of PACs. The first stage in the generation of a pharmacophore from MFA is the calculation of PACs. This follows the general procedure of deriving a predicted activity for a molecule on the basis of the PLS regression coefficients. A general algorithm is given in ref 45. In simple terms, for a one component model the predicted molecular activity, A_i, for any molecule is

$$A_i = \sum_{j=1}^{np} C_j^e E_{ij}^e + \sum_{j=1}^{np} C_j^s E_{ij}^s \quad (1)$$

where E_{ij}^e and E_{ij}^s are the electrostatic and steric energies at grid point j due to molecule i, C_j^e and C_j^s are the PLS regression coefficients for the two energy terms at grid point j, and np is the total number of grid points. For subsequent components, E_{ij} is replaced by the matrix of residuals. The basis of our novel approach to the interpretation of a PLS regression is to treat the predicted molecular activity, A_i, as the sum of atomic activities, a_{ik}, summed over na atoms

$$A_i = \sum_{k=1}^{na} a_{ik} \quad (2)$$

This is calculated by partitioning the energy at each grid point into the contributions due to each atom

$$a_{ik} = \sum_{j=1}^{np} C_j^e e_{ijk}^e + \sum_{j=1}^{np} C_j^s e_{ijk}^s \quad (3)$$

where e_{ijk} represents the grid energy at point j due to atom k in molecule i and

$$E_{ij} = \sum_{k=1}^{na} e_{ijk} \quad (4)$$

PAC is synonymous with a_{ik} when it represents total activity for a single atom, but note that PACs can also be derived easily to describe the activity of a functional group or be decomposed further into the contributions due to individual energy terms or principal components.

The sum of the PACs for all atoms in a molecule is equal to the predicted activity of the molecule (eq 2). For this to be true, there are several points which require attention. Most importantly, eq 4 is not true for all grid points in normal implementations of MFA since E_{ij}^s and E_{ij}^e are truncated when a clash occurs between a grid point and an atom. In our implementation, a clash is defined as a steric energy greater than E_{clash}^s (typically a value of 5 kcal/mol is used). When a clash occurs, E_{ij}^s is truncated to E_{clash}^s and then partitioned among all atoms in proportion to e_{ijk}^s (having first truncated e_{ijk}^s to E_{clash}^s). The electrostatic energy at that point, E_{ij}^e, is set to the average value of E_{ij}^e for all other molecules (because a mean value has no weight in the PLS regression) and is then partitioned equally among all atoms.

Equation 2 also assumes that suitable scaling of grid energies has taken place to take account of block-scaling, *i.e.*, when each set of energies has been scaled to unit variance, in order to weight equally the contribution of electrostatic and steric energies in the PLS regression.[46] In practice we have omitted block-scaling and instead have scaled the electrostatic contribution by suitable choice of probe charge.

It is also worth noting that it is usual in PLS to mean center the activity data. Thus in eq 2, A_i and a_{ik} represent *mean-centered* predicted activity contributions. The corrected (or activity-centered) predicted molecular activity, A_i^a, is simply

$$A_i^a = A_i + Y_{\text{mean}} \quad (5)$$

where Y_{mean} is the mean of the observed activity data. In order to derive activity-centered PACs, it is necessary to distribute Y_{mean} among individual atoms and energy components. There is no clear-cut way of doing this other than to divide Y_{mean} equally among all atoms. In practice we have found it more useful to deal with the mean-centered PACs as this allows one to define default PAC cutoffs (*i.e.*, for the purpose of defining interaction sites) which are less sensitive to the magnitude of the activity data. To avoid confusion, all PACs quoted as examples in this paper are mean-centered.

Extraction of a Design Model from PACs. The next stage in design model generation is to select from each molecule of the training set those functional groups which constitute the pharmacophore, *i.e.*, the hydrogen-bonding sites and the steric contacts which have been predicted by MFA to be most strongly correlated with activity. These groups are selected automatically within the program on the basis of chemistry and PACs, using default or user-defined PAC tolerances.

Hydrogen bond donor and acceptor sites are selected first and are generated as vectors as described earlier (*i.e.*, **D-X** and **A-Y** vectors to describe donor and acceptor sites, respectively). For example, a **D-X** vector will be generated from any HO or HN group with suitable electrostatic, steric, and total PACs. Typically one would desire a positive electrostatic PAC and a positive total PAC, but the user can make the criteria more or less strict as desired. (Note that a positive PAC implies a positive contribution to biological activity, which is assumed

to be the desired response.) Any atom not assigned as a donor or acceptor may be designated as a lipophilic site, L. Note that aliphatic and aromatic lipophilic sites are not presently differentiated.

Having generated sites from each molecule in the set, the algorithm then compresses sites of a similar type if they are within a defined distance cutoff: hence unnecessary sites are discarded and replaced by a site with the average Cartesian coordinates.

By suitable choice of PAC cutoff values, it is possible to generate a variety of pharmacophores for use in structure generation. Thus one may wish to select almost all atoms from the training set, in order to drive structure generation toward forming molecules which closely mimic the chemistry of the training set. Alternatively one may choose to form a very sparse pharmacophore which represents only a small number of key functional groups, which would result in a set of solutions with greater structural diversity.

A further use of PACs is for the generation of a receptor envelope. This can assist the visualization of the shape of the pharmacophore or the hypothetical binding site and can also be used for clash checking during structure generation to ensure that a designed molecule does not grow too far from the interaction sites and encroach into the putative receptor wall. The first step is to use PACs to define which atoms in the training set are *favorable*, *unfavorable*, or *indifferent* as sites for envelope generation. Then an envelope is generated at a suitable distance around all *favorable* atoms and allowed to clash with any atoms labeled as *unfavorable*, provided that it does not clash with any favorable or indifferent atoms in any molecule.

The envelope itself is generated from the original set of MFA grid points and is simply the subset of points which are in suitable van der Waals' contact with favorable atoms. The nature of the van der Waals' contact is determined by examining the steric energy associated with each grid point interacting with each atom (as this quantity, e^{s}_{ijk}, has already been calculated): a steric energy below a minimum cutoff (*e.g.*, −0.1 kcal/mol) defines a favorable contact, while an energy above a maximum cutoff (*e.g.*, 0.1 kcal/mol) defines a clash. In this way a shell of points is extracted from the original grid points which reflects a van der Waals's surface around the favorable pharmacophore sites.

The advantages of the use of PACs can be summarized as

(1) ease of interpretation of predicted activity in terms of individual atom or group contributions,

(2) ease of definition of pharmacophore atoms as points or vectors taken from coordinates of the original set of molecules,

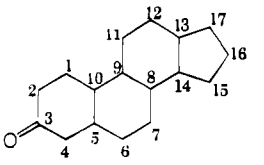
(3) a user-defined rule base to determine what PACs are considered favorable: hence the user may choose to create a very detailed or a very sparse pharmacophore or a receptor envelope which encloses a small or large volume, and

(4) easily adaptable to any type of MFA field (*e.g.*, hydrophobic potentials), and to take account of any additional chemical parameters which are not 3-D (*e.g.*, log *P* is sometimes appended to the 3-D parameter set). More generally, this approach could be applied to any 3D-QSAR method in which the chemical descriptors can be partitioned into contributions from individual atoms or groups.

In some cases, the elucidated pharmacophore may be very "sparse", *i.e.*, contain only a few features separated by large distances. In such instances, the design model generation module may be instructed to fill the volume defined by the pharmacophore features with a uniform density of L sites to help the growth of structures across the empty space.

Molecular Field Analysis of Progestogen and Androgen Receptor Binding. As an example of the generation of a pharmacophore from MFA, a 3D-QSAR analysis on a set of steroids was performed using an in-house implementation of MFA. The set of steroids (Table 1) was built and geometry-optimized within MOPAC 6.0,[47] using the AM1 Hamiltonian (with the PRECISE convergence criteria). Each steroid was then aligned to steroid 1 by RMS superimposition of the ring-junction carbon atoms (positions 5, 8–10, 13, 14). This is a reasonable starting point for aligning a set of such similar structures, given that our main purpose is to illustrate the features of *de novo* design.

Table 1. Structures and Relative Binding Affinities of the 35 Steroids Examined in this Study[a]



no.	Δ	C7α	C13	C17α	C17β	PR	AR
1	4		CH_3		OH	20	154
2	4,9		CH_3		OH	17	134
3	4,9,11		CH_3		OH	74	197
4	4		C_2H_5		OH	34	126
5	4,9		C_2H_5		OH	26	93
6	4,9,11		C_2H_5		OH	86	172
7	4		C_3H_7		OH	4.5	108
8	4,9		C_3H_7		OH	4.6	42
9	4,9,11		C_3H_7		OH	38	105
10	4		CH_3	CH_3	OH	100	146
11	4,9		CH_3	CH_3	OH	71	64
12	4,9,11		CH_3	CH_3	OH	208	204
13	4,9,11		C_2H_5	CH_3	OH	230	143
14	4	CH_3	CH_3	CH_3	OH	214	108
15	4,9	CH_3	CH_3	CH_3	OH	198	122
16	4,9,11	CH_3	CH_3	CH_3	OH	306	180
17	4,9,11	CH_3	C_2H_5	CH_3	OH	236	124
18	4		CH_3	C_2H	OH	156	43
19	4,9		CH_3	C_2H	OH	42	19
20	4,9,11		CH_3	C_2H	OH	63	70
21	4		C_2H_5	C_2H	OH	170	84
22	4,9		C_2H_5	C_2H	OH	68	41
23	4,9,11		C_2H_5	C_2H	OH	76	83
24	4		C_3H_7	C_2H	OH	73	44
25	4,9		C_3H_7	C_2H	OH	11	10
26	4,9,11		C_3H_7	C_2H	OH	61	66
27	4		CH_3			190	37
28	4,9		CH_3			218	29
29	4,9,11		CH_3			274	138
30	4		CH_3		$COCH_3$	230	6.4
31	4,9		CH_3		$COCH_3$	181	8.8
32	4,9,11		CH_3		$COCH_3$	230	16
33	4		CH_3	CH_3	$COCH_3$	317	5.5
34	4,9		CH_3	CH_3	$COCH_3$	230	1.1
35	4,9,11		CH_3	CH_3	$COCH_3$	230	1.9

[a] The relative binding affinities of progesterone and testosterone to the progestogen (PR) and androgen (AR) receptors, respectively, are arbitrarily taken to be 100.

Atom-centered partial charges were derived from the AM1 electrostatic potential, with the default MOPAC ESP parameters. Steric parameters for van der Waals' energies were taken from the Robson–Platt force field.[48]

A cubic grid was created around the set of superimposed molecules at a spacing of 1.5 Å. The grid was then truncated such that, for each molecule, grid points were retained if they were no closer than 2.0 Å to any atom and within 5.0 Å of at least one atom; a grid point outside these limits with respect to one particular molecule would still be retained provided it fell within these limits for another molecule. This approach serves to reduce the number of grid points compared with using a full cubic grid (as in CoMFA) by eliminating points which are too close or too distant to contribute significantly to the model. The final grid consisted of 831 points.

The energy calculations were performed using a methyl probe with a charge of 0.5 and a distance-dependent dielectric. Hydrophobic fields were also evaluated, using the parameters of Viswanadhan *et al.*,[49] but omitted in the final model because they were not found to contribute significantly. A steric cutoff of 5 kcal/mol was used; thus, any steric energy greater than 5 kcal/mol was truncated to 5 kcal/mol, and the electrostatic energy associated with this point was set to the mean of the electrostatic energies of all other molecules at this point.

The resulting grid energies were correlated against log-

Table 2. Di- and Tripeptide Inhibitors of ACE

inhibitor	sequence	IC_{50} (nM)
1	Nle-Ala-Pro	700
2	Val-Trp	1700
3	Leu-Ala-Pro	2300
4	Ile-Tyr	3700
5	Phe-Ala-Pro	4200
6	Arg-Ala-Pro	16000
7	Phe-Pro-Pro	78000

(relative binding affinity) for both the progestogen and androgen receptors, using an in-house implementation of PLS regression. The final PLS models were chosen by a leave-one-out cross-validation method.[50]

In order to generate a pharmacophore, atoms were extracted from the original coordinates of the training set according to the following rule base, which was chosen in order to select only the most important interaction sites for this particular example:

PACs are summed for united functional groups (*e.g.*, carbonyls and methylenes each count as a single group).

D-X and **A-Y** sites are generated from, for example, NH and C=O groups, respectively, provided that the electrostatic PACs and total PACs are both greater than 0.05 (note PACs are expressed in the same units as the original activity data: here as log(relative binding affinity)).

L sites are generated from any atoms not defined as D or A provided that steric PAC is greater than 0.05.

Favorable sites for envelope generation are defined as any atom with a steric PAC greater than −0.05, with all other atoms defined as unfavorable sites (these cutoffs were chosen in order to generate an envelope which encompassed the whole volume of the steroids while clashing only with groups with the most negative PACs).

Sites of a similar type within 0.5 Å were compressed into a single site.

As a very sparsely populated pharmacophore was generated by the above criteria, the volume defined by the interaction sites was filled in by a wash of L interaction sites, at a minimum spacing of 0.5 Å.

Structure generation was performed only for the progestogen pharmacophore. This was because the training set does not explore a wide variety of substituent positions but concentrates mostly on D-ring substituents, which are generally disfavored for androgen receptor binding. Thus the androgen pharmacophore is very similar to the steroid skeleton, and it was decided that more varied designs would result from examining the progestogen pharmacophore.

The structure generation module used a continuous growth strategy, as this proved most efficient at growing a molecule which hit both of the hydrogen bond acceptor sites present in the progestogen pharmacophore. Structure generation accessed the full general organic library but was constrained to test fragments with an acceptor site before any other chemistries.

Pharmacophore Mapping—ACE Inhibitors. As a further test of PRO_LIGAND, we decided to use the data presented by Teig[51] for seven peptidic ACE inhibitors to generate a pharmacophore for ACE inhibition and then to build structures to conform to it. The seven peptides together with their activities are shown in Table 2.

The initial requirement is to generate a diverse set of low-energy conformers of each peptide. There are many methods of achieving this, and the one we have followed is to apply a novel molecular dynamics algorithm aimed at maintaining constant potential energy (*RUSH dynamics*), which has been shown to be effective at exploring conformational space efficiently.[52] Fifty conformations of each peptide were generated by taking snapshots from a RUSH dynamics simulation, followed by minimization of each snapshot. The molecular dynamics simulation was carried out *in vacuo* at 410 K; 15 000 steps were simulated, saving a snapshot every 300 steps. These conformations were minimized by the method of conjugate gradients to a gradient norm of less than 0.1 kcal/mol/Å. All calculations were carried out using the Robson–Platt force field.[48]

Table 3. Correlation coefficients (and in parentheses standard deviations) for the Principal Components Extracted for the Androgen (AR) and Progestogen (PR) PLS Models[a]

component	AR		PR	
	R^2_{conv} (SD)	R^2_{cross} (SD)	R^2_{conv} (SD)	R^2_{cross} (SD)
1	0.720 (0.311)	0.594 (0.375)	0.639 (0.308)	0.492 (0.365)
2	0.885 (0.202)	0.790 (0.274)	0.772 (0.248)	0.601 (0.328)
3	0.925 (0.167)	0.826 (0.254)	0.839 (0.212)	0.690 (0.294)
4	0.934 (0.159)	0.834 (0.252)	0.857 (0.203)	0.700 (0.294)
5	0.938 (0.157)	0.842 (0.249)	0.873 (0.195)	0.708 (0.295)
6		0.821 (0.271)		0.681 (0.314)
7		0.807 (0.286)		0.691 (0.315)
8		0.810 (0.289)		0.681 (0.326)

[a] Correlation coefficients (R^2) are quoted for conventional and cross-validated models. The final PLS model for each receptor consisted of five components.

The clique-detection algorithm was then used to deduce a pharmacophore from these input conformations. A solution was obtained using inhibitor **1** as the reference molecule and a distance tolerance of 1.5 Å. Only heavy atoms were considered in the matching process.

Results

Derivation of a Steroid Pharmacophore and Subsequent Design. A five-component PLS model was derived from MFA for both androgen and progestogen receptor binding. As can be seen from Table 3, the cross-validated correlation coefficient (R^2) is somewhat higher for the androgen receptor but for both receptors is comparable to the magnitude of values quoted in the literature for CoMFA studies on steroids binding to steroid receptors[53] and binding globulins.[54] Small modifications to the MFA parameters (*e.g.*, grid spacing, probe charge, addition of hydrophobic parameters) did not significantly improve the model.

Derivation of PACs allows an analysis of the PLS model in terms of atomic or group contributions. A typical breakdown for the progestogen receptor relative binding affinity is demonstrated for steroids **25** and **35** in Figure 1. It can be seen that most atoms have small negative or positive contributions, with the main contributions to activity coming from the C17 substituents. The observed distribution of PACs clearly demonstrates that the difference in predicted activity between steroids **25** and **35** is mostly due to the presence of the C13 propyl chain (which results in a steric clash) and the replacement of the favored C17 acetyl with the disfavored hydroxyl. It should be noted that this partition is entirely dependent on the quality of the MFA/PLS model: hence the C3 carbonyl, which is known to be essential for activity, is not seen to have a major contribution, simply because the training set does not feature molecules without this moiety. Note also that this carbonyl does not necessarily have a constant PAC throughout the training set because it does not occupy a constant position in space (as increasing the number of double bonds in the steroid skeleton produces a more planar skeleton and thus alters the position of the carbonyl).

The absolute values of PACs are less useful than the relative values across a series of molecules. Table 4 demonstrates the use of PACs to highlight differences in structural requirements between the two receptors. Thus androgen receptor binding favors the 17β hydroxyl

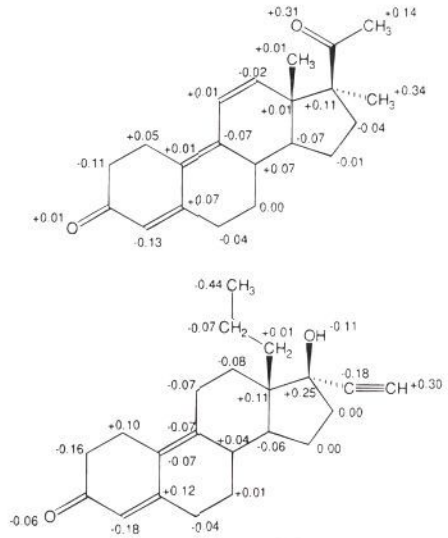


Figure 1. Example of PACs derived for progestogen receptor binding of steroids **35** (upper) and **25** (lower). The sum of the PACs within a molecule is equal to the mean-centered predicted log(relative binding affinity). In this case, this is equal to 0.65 for steroid **35** and −0.65 for steroid **25**. Note that the PACs have been summed for functional groups (*e.g.*, C=O).

Table 4. Examples of PACs for Certain 17β Substituents in Steroids **1**, **27**, and **30**[a]

	PR			AR		
substituent	E	S	tot	E	S	tot
OH	−0.10	−0.02	−0.12	0.22	−0.05	0.17
acetyl (C=O)	0.18	0.21	0.39	−0.24	−0.41	−0.65
acetyl(CH_3)	0.05	0.18	0.23	−0.08	−0.34	−0.42
lactone (−CO−)	0.08	0.22	0.30	−0.01	−0.30	−0.31
lactone (α-CH_2)	−0.02	0.19	0.17	0.11	0.00	0.11
lactone (β-CH_2)	0.07	0.23	0.30	−0.06	−0.10	−0.15

[a] PACs are quoted in terms of electrostatic (E) and steric (S) components, as well as their sum (tot). The PACs represent contributions to the mean-centered predicted log(relative binding affinity) for the progestogen (PR) and androgen (AR) receptors.

but disfavors the carbonyls of the acetyl and lactone groups. The PLS model interprets the poor activity of these latter groups as being in part an electrostatic effect and in part a steric effect. Conversely, bulky hydrogen bond-accepting groups are seen to be favored for progestogen receptor binding.

Figure 2 demonstrates the pharmacophores generated for the progestogen receptor and androgen receptor. The definition of interaction sites using the PACs detailed above results in very sparse pharmacophores, in which only the key steric and electrostatic interactions are present. These features are

The most active C3 carbonyl is for the most planar skeleton (*i.e.*, with Δ 4,9,11).

At C17, an H-bond donor is favored for androgen receptor binding (*cf.* the hydroxyl in steroid **1**), while two possible H-bond acceptor sites are favored for progestogen receptor binding (*cf.* the carbonyls in steroids **27** and **30**).

Several steric/lipophilic sites are retained, most notably those describing the bulky C17 substituents which favor progestogen receptor binding.

For the progestogen pharmacophore, the volume enclosed by these interaction sites was filled by a wash of lipophilic sites in order to facilitate structure building. In this way, PRO_LIGAND is encouraged to build

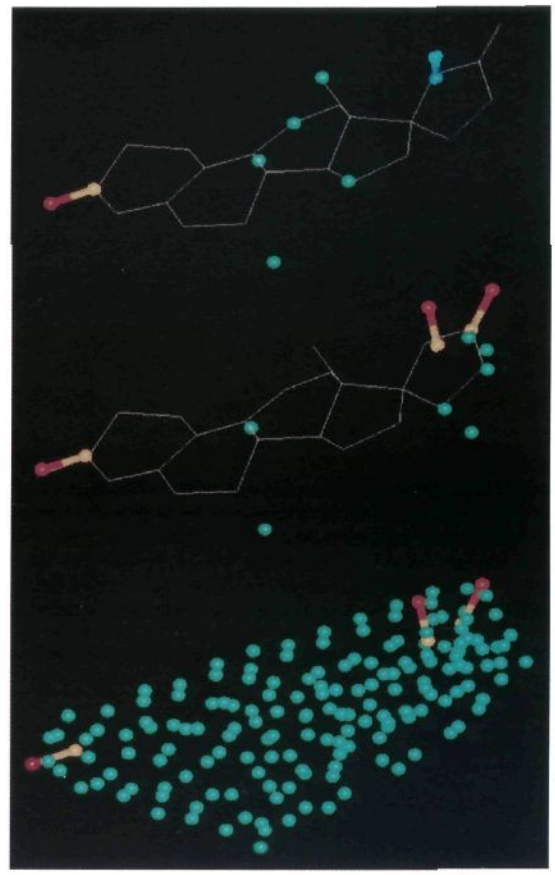


Figure 2. Pharmacophores generated for the androgen receptor (upper) and progestogen receptor (center), superimposed upon a steroid skeleton for reference. Hydrogen bond acceptor sites are colored red, hydrogen bond donor sites blue, and steric (lipophilic) sites green. Also shown is the design model (lower) from the sparse progestogen pharmacophore (center), generated by filling the enclosed volume with a wash of lipophilic sites.

designs which bridge the acceptor sites with a diversity of chemistries, *i.e.*, nonsteroidal ligands for the steroid receptor.

A set of structures generated by PRO_LIGAND is shown in Figure 3 as 2-D structure diagrams (**1**−**5**). It can be seen that they achieve suitable bridging of the C3 and C17 acceptor sites by a variety of ring structures. Although the design model encouraged the formation of nonsteroidal solutions, note also that one solution (**1**) is very steroid-like, having formed rings equivalent to rings A, C, and D. Not all favorable steric contacts have been achieved, but it is possible to take all or part of one of the solutions and use it as a seed from which to grow further fragments. In this way, it is possible for the drug designer to drive PRO_LIGAND toward areas of chemistry which are of particular interest. The results shown in Figure 3 can be thought of as the initial phase of design generation, when one is mostly concerned with brain storming for novel solutions which are capable of bridging across a particular pharmacophore. Development of any single idea into a feasible drug candidate clearly requires a greater concentration of effort to fine tune the physicochemical and structural features.

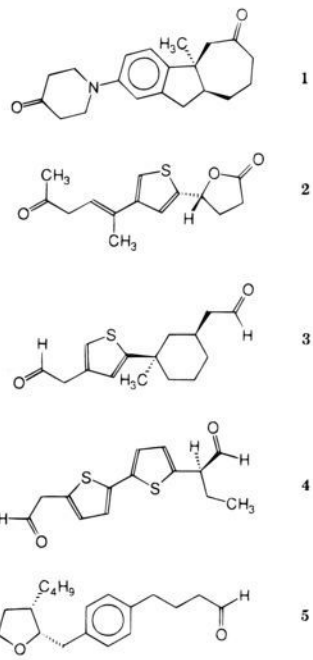
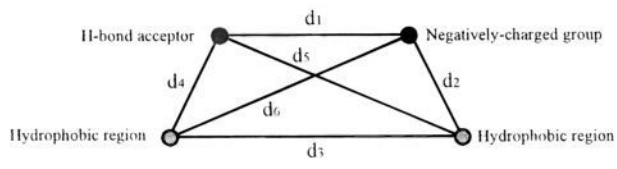


Figure 3. Some examples of PRO_LIGAND-generated designs for the progestogen pharmacophore.



d1 = 4.90 ± 1.5 d3 = 6.77 + 1.5 d5 = 6.59 + 1.5
d2 = 3.47 ± 1.5 d4 = 2.89 + 1.5 d6 = 3.78 + 1.5

Figure 4. Pharmacophore for ACE inhibition.

Pharmacophore Mapping—ACE Inhibitors. For the ACE inhibitor test case, a four-point pharmacophore was deduced consisting of two lipophilic regions, an acceptor group, and a negatively charged group as shown in Figure 4. The run in question took only 78 *CPU s on an R3000 SGI Indigo workstation. This model* is in agreement with that described by Teig[51] and also with the present consensus concerning the requirements for inhibitor binding to ACE, *viz.*,

1. an ionizable C-terminal carboxyl group capable of ionic binding to a positively charged enzyme residue (*e.g.*, Lys, Arg),

2. a carbonyl oxygen to accept a hydrogen bond from a donor XH group on the receptor, and

3. a zinc-binding functional group such as a carboxylate, hydroxamate, phosphonate, or thiolate.

It can be seen that our pharmacophore satisfies requirements 1 and 2; indeed, the carboxyl oxygen–carbonyl oxygen distance is in concurrence with the active-site geometry deduced by systematic/constrained search on a more structurally diverse training set[37,38] which included functional groups capable of zinc binding. This pharmacophore was then presented to PRO_LIGAND as the design base for a set of structure generation runs. The design model was generated to be similar to the design base, with the volume between the pharmacophore features filled with a wash of lipophilic sites in order to aid the growth of structures across empty space.

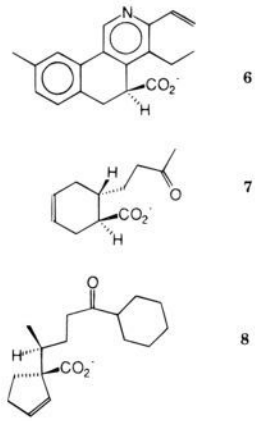


Figure 5. Some proposed ACE inhibitors designed by PRO_LIGAND.

The structure generation module was then invoked, and a variety of building strategies were employed to generate the three diverse structures (**6**–**8**) illustrated in Figure 5. All these structures satisfy the constraints of the pharmacophore both as built by PRO_LIGAND and after minimization with the COSMIC force field.[55,56] The detail of the building processes involved is given in what follows. All CPU times refer to an entry-level (R3000) SGI Indigo workstation.

The core of **6** was built by a grow-and-fuse strategy using the fragments shown in Figure 6a. In this strategy, molecular fragments are joined sequentially in a "build-up" approach to structure assembly; additionally the algorithm seeks to generate ring structures in the designed molecules. The generation of this core took about 120 CPU s. On examination of this structure, it was seen that while the requirement for an acceptor, the negative charge, and one lipophilic region (the ethyl side chain) has been fulfilled, another lipophilic region needed to be added. Accordingly, PRO_LIGAND was run in a ring-bracing mode which resulted in the joining of two *Z*-butadiene fragments on to the existing core to yield an intermediate which is likely to reduce spontaneously to yield the final structure (**6**). *The CPU time required for ring-bracing was* about 10 s.

To generate **7**, an alternative "outside-in" building mode was employed in which two fragments were first placed to satisfy the H-bonding requirements of the pharmacophore. Then, a lipophilic fragment was sought to bridge these two fragments. This is illustrated in Figure 6b. The CPU time required for this operation was about 4 s. Once again, it was observed that further augmentation of the lipophilic features was required, and thus a ring-bracing run was used to place the 2,3-butene fragment to form the cyclohexene ring as shown in the figure.

The third structure, **8**, was generated in a similar manner to **6**, *i.e.*, a grow-and-fuse strategy followed by a ring-bracing run as illustrated in Figure 6c. The CPU time required for these actions totaled 200 s.

Discussion

In the first paper of this series,[26] the capabilities of PRO_LIGAND were demonstrated in the context of "direct" drug design, *i.e.*, the search for novel ligands based upon a known target structure. In this paper, we have focused upon the alternative situation of "indirect" design where knowledge about the target

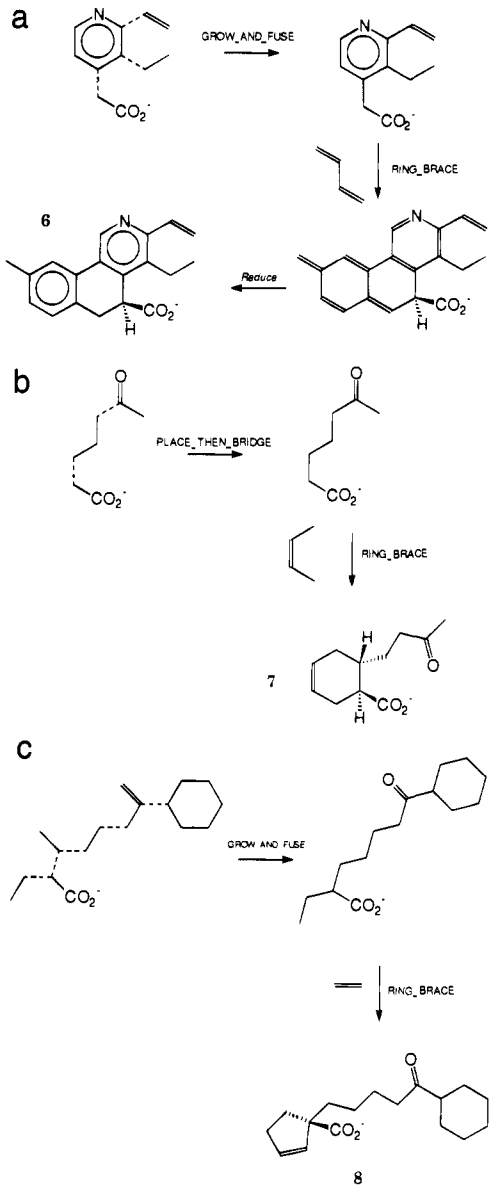


Figure 6. Construction processes for structures **6**–**8**. Dashed lines represent bonds that will be formed between molecular fragments.

molecule can only be obtained second-hand from compounds known to bind to it. Given that the structure of the majority of receptors and enzymes of therapeutic interest are yet to be determined, it is clear that the design of new pharmaceuticals will often be carried out indirectly. It is thus important that efficient, useful computational tools be developed to aid molecular design under these conditions.

Here, we have reported the integration of two such tools into our in-house *de novo* design methodology, PRO_LIGAND. Firstly, we have presented a novel method for the interpretation of MFA models by the derivation of atomic predicted activity contributions. Although useful in itself as a means of directly interpreting the PLS regression in terms of chemical structure and properties, the method also allows a straightforward approach to the generation of pharmacophores and receptor envelopes. Secondly, we have developed a pharmacophore-mapping procedure based on the proven technique of clique detection,[40,42] which is seen to be an efficient means of objectively deducing pharmacophores from multiple conformations of a series of molecules.

Pharmacophores extracted by MFA or clique detection can be used for searching 3-D structural databases.[57] However, while database searching is an efficient means of identifying structures which have already been synthesized and may even be immediately available, there remains a need to generate novel chemistries which satisfy the pharmacophore. Thus, as a complement to 3-D database searching, *de novo* design tools have recently become of great interest.

We have demonstrated that PRO_LIGAND is an intuitive and effective method for the *de novo* design of structures of considerable diversity while, it is hoped, retaining the activity of the original series. Hence PRO_LIGAND can be of considerable utility in the generation of novel lead molecules as part of a rational molecular design process. The application of PRO_LIGAND to the *de novo* design of peptides and DNA-binding drugs will be reported in future papers, as will the use of a genetic algorithm for structure refinement.

References

(1) Navia, M. A.; Murcko, M. A. Use of Structural Information in Drug Design. *Curr. Opin. Struct. Biol.* **1992**, *2*, 202–210.
(2) Kuntz, I. D. Structure-Based Strategies for Drug Design and Discovery. *Science* **1992**, *257*, 1078–1082.
(3) Navia, M. A.; Peattie, D. A. Structure-Based Drug Design: Applications in Immunopharmacology and Immunosuppression. *Trend. Pharmacol. Sci.* **1993**, *14*, 189–195.
(4) Reich, S. H.; Webber, S. E. Structure-Based Drug Design (SBDD): Every Structure Tells a Story. *Perspect. Drug Discovery Des.* **1993**, *1*, 371–390.
(5) Greer, J.; Erickson, J. W.; Baldwin, J. J.; Varney, M. D. Application of the Three-Dimensional Structures of Protein Target Molecules in Structure-Based Drug Design. *J. Med. Chem.* **1994**, *37*, 1035–1054.
(6) Moon, J. B.; Howe, W. J. Computer Design of Bioactive Molecules: A Method for Receptor-Based De Novo Ligand Design. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 314–328.
(7) Moon, J. B.; Howe, W. J. Recent Advances in De Novo Molecular Design. In *Trends in QSAR and Molecular Modelling 92*; Wermuth, C. G., Ed.; ESCOM: Leiden, 1993; pp 11–19.
(8) Miranker, A.; Karplus, M. Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 29–34.
(9) Caflisch, A.; Miranker, A.; Karplus, M. Multiple Copy Simultaneous Search and Construction of Ligands in Binding Sites: Application to Inhibitors of HIV-1 Aspartic Protease. *J. Med. Chem.* **1993**, *36*, 2142–2167.
(10) Nishibata, Y.; Itai, A. Automatic Creation of Drug Candidate Structures Based on Receptor Structure. Starting Point for Artificial Lead Generation. *Tetrahedron* **1991**, *47*, 8985–8990.
(11) Nishibata, Y.; Itai, A. Confirmation of Usefulness of a Structure Construction Program Based on Three-Dimensional Receptor Structure for Rational Lead Generation. *J. Med. Chem.* **1993**, *36*, 2921–2928.
(12) Böhm, H.-J. The Computer Program LUDI: A New Method for the De Novo Design of Enzyme Inhibitors. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 61–78.
(13) Böhm, H.-J. LUDI: Rule-Based Automatic Design of New Substituents for Enzyme Inhibitor Leads. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 593–606.
(14) Böhm, H.-J. Ligand Design. In *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM, Leiden, 1993; pp 386–405.
(15) Lewis, R. A.; Roe, D. C.; Huang, C.; Ferrin, T. E.; Langridge, R.; Kuntz, I. D. Automated Site-Directed Drug Design Using Molecular Lattices. *J. Mol. Graphics* **1992**, *10*, 66–78.
(16) Rotstein, S. H.; Murcko, M. A. GenStar: A Method for De Novo Drug Design. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 23–43.
(17) Rotstein, S. H.; Murcko, M. A. GroupBuild: A Fragment-Based Method for De Novo Drug Design. *J. Med. Chem.* **1993**, *36*, 1700–1710.
(18) Gillet, V.; Johnson, A. P.; Mata, P.; Sike, S.; Williams, P. SPROUT: A Program for Structure Generation. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 127–153.

(19) Gillet, V. J.; Newell, W.; Mata, P.; Myatt, G.; Sike, S.; Zsoldos, Z.; Johnson, A. P. SPROUT: Recent Developments in the De Novo Design of Molecules. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 207–217.
(20) Pearlman, D. A.; Murcko, M. A. CONCEPTS: New Dynamic Algorithm for De Novo Drug Design. *J. Comput. Chem.* **1993**, *14*, 1184–1193.
(21) Tschinke, V.; Cohen, N. C. The NEWLEAD Program: A New Method for the Design of Candidate Structures from Pharmacophoric Hypotheses. *J. Med. Chem.* **1993**, *36*, 3863–3870.
(22) Ho, C. W. M.; Marshall, G. R. SPLICE: A Program to Assemble Novel Partial Query Solutions from Three-Dimensional Database Searches into Novel Ligands. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 623–647.
(23) Leach, A. R.; Lewis, R. A. A Ring-Bracing Approach to Computer-Assisted Ligand Design. *J. Comput. Chem.* **1994**, *15*, 233–240.
(24) Leach, A. R.; Kilvington, S. R. Automated Molecular Design: A New Fragment-Joining Algorithm. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 283–298.
(25) Eisen, M. B.; Wiley, D. C.; Karplus, M.; Hubbard, R. R. HOOK: A Program for Finding Novel Molecular Architectures That Satisfy the Chemical and Steric Constraints of a Macromolecule Binding Site. *Proteins: Struct., Funct., Genet.* **1994**, *19*, 199–221.
(26) Clark, D. E.; Frenkel, A. D.; Levy, S. A.; Li, J.; Murray, C. W.; Robson, B.; Waszkowycz, B.; Westhead, D. R. PRO_LIGAND: An Approach to *De Novo* Molecular Design. 1. Application to the Design of Organic Molecules. *J. Comput.-Aided Mol. Des.*, in press.
(27) Walkinshaw, M. D. Protein Targets for Structure-Based Drug Design. *Med. Res. Rev.* **1992**, *12*, 317–372.
(28) Ealick, S. E.; Armstrong, S. R. Pharmacologically Relevant Proteins. *Curr. Opin. Struct. Biol.* **1993**, *3*, 861–867.
(29) Cramer, R. D.; DePriest, S. A.; Patterson, D. E.; Hecht, P. The Developing Practice of Comparative Molecular Field Analysis. In *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 443–485.
(30) Wermuth, C.-G.; Langer, T. Pharmacophore Identification. In *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 117–136.
(31) Klebe, G. The Use of Composite Crystal-field Environments in Molecular Recognition and the *de Novo* Design of Protein Ligands. *J. Mol. Biol.* **1994**, *237*, 212–235.
(32) Goldberg, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley: Reading, MA, 1989.
(33) Lindberg, W.; Persson, J.-A.; Wold, S. Partial Least-Squares Method for Spectrofluorimetric Analysis of Mixtures of Humic Acid and Ligninsulfonate. *Anal. Chem.* **1983**, *55*, 643–648.
(34) Höskuldsson, A. PLS Regression Methods. *J. Chemom.* **1988**, *2*, 211–228.
(35) Doré, J.-C.; Gilbert, J.; Ojasoo, T.; Raynaud, J.-P. Correspondence Analysis Applied to Steroid Receptor Binding. *J. Med. Chem.* **1986**, *29*, 54–60.
(36) Marshall, G. R.; Barry, C. D.; Bosshard, H. E.; Dammkoehler, R. A.; Dunn, D. A. The Conformational Parameter in Drug Design: The Active Analog Approach. In *Computer-Assisted Drug Design*; Olson, E. C., Christoffersen, R. E., Eds.; ACS Symposium Series 112; American Chemical Society: Washington, DC, 1979; pp 205–226.
(37) 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.
(38) Dammkoehler, R. A.; Karasek, S. F.; Shands, E. F. B.; Marshall, G. R. Constrained Search of Conformational Hyperspace. *J. Comput.-Aided Mol. Des.* **1989**, *3*, 3–21.
(39) Sheridan, R. P.; Nilakantan, R.; Dixon, J. S.; Venkataraghavan, R. The Ensemble Approach to Distance Geometry: Application to the Nicotinic Pharmacophore. *J. Med. Chem.* **1986**, *29*, 899–906.
(40) Martin, Y. C.; Bures, M. G.; Danaher, E. A.; DeLazzar, J.; Lico, I.; Pavlik, P. A. A Fast New Approach to Pharmacophore Mapping and its Application to Dopaminergic and Benzodiazepine Agonists. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 83–102.
(41) Hodgkin, E. E.; Miller, A.; Whittaker, M. A Monte Carlo Pharmacophore Generation Procedure: Application to the Human PAF receptor. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 515–534.
(42) Golender, V. E.; Vorpagel, E. R. Computer-Assisted Pharmacophore Identification. In *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 137–149.
(43) Bron, C.; Kerbosch, J. Algorithm 457: Finding All Cliques of an Undirected Graph. *Commun. ACM* **1973**, *16*, 575–577.
(44) Brint, A. T.; Willett, P. Algorithms for the Identification of Three-Dimensional Maximal Common Substructures. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 152–158.
(45) Geladi, P.; Kowalski, B. R. Partial Least-Squares Regression: A Tutorial. *Anal. Chim. Acta* **1986**, *185*, 1–17.
(46) Wold, S.; Johansson, E.; Cocchi, M. PLS: Partial Least-Squares Projections to Latent Structures. In *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 523–550.
(47) MOPAC: A General Molecular Orbital Package (Version 6.0); Stewart, J. J. P.; QCPE Program 455.
(48) Robson, B.; Platt, E. Refined Models for Computer Calculations in Protein Engineering: Calibration and Testing of Atomic Potential Functions Compatible with More Efficient Calculations. *J. Mol. Biol.* **1986**, *188*, 258–281.
(49) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic Physicochemical Parameters for Three-Dimensional Structure Directed Quantitative Structure-Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and their Application for an Automated Superposition of Certain Naturally Occurring Nucleoside Antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163–172.
(50) Cramer, R. D.; Bunce, J. D.; Patterson, D. E.; Frank, I. E. Crossvalidation, Bootstrapping and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18–25.
(51) Teig, S. L. The Development of Meaningful 3-D Search Queries for Drug Discovery. In *Recent Advances in Chemical Information II*: Collier, H. R., Ed.; Royal Society of Chemistry: Cambridge, 1993; pp 195–208.
(52) Byrne, D. G.; Li, J.; Platt, E.; Robson, B.; Weiner, P. Novel Algorithms for Searching Conformational Space. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 67–82.
(53) Loughney, D. A.; Schwender, C. A. A Comparison of Progestin and Androgen Receptor Binding using the CoMFA Technique. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 569–581.
(54) Norinder, U. Experimental Design Based 3-D QSAR Analysis of Steroid-Protein Interactions: Application to Human CBG Complexes. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 381–389.
(55) Morley, S. D.; Abraham, R. J.; Haworth, I. S.; Jackson, D. E.; Saunders, M. R.; Vinter, J. G. COSMIC(90): An Improved Molecular Mechanics Treatment of Hydrocarbons and Conjugated Systems. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 475–504.
(56) *Pimms*: Personal Interactive Molecular Modelling System; Oxford Molecular Ltd., The Magdalen Centre, Oxford Science Park, Sandford-on-Thames, Oxford OX4 4GA.
(57) Martin, Y. C. 3-D Database Searching in Drug Design. *J. Med. Chem.* **1992**, *35*, 2145–2154.