Pharmacophore Discovery – Lessons Learned

John H. van Drie*

Vertex Pharmaceuticals, 130 Waverly St., Cambridge, MA 02139, USA

Abstract: Pharmacophore discovery is one of the major elements of molecular modeling in the absence of X-ray structural data. While pharmacophores initially made their debut as a means for lead discovery, more recent refinements have brought them into the domain of lead optimization, e.g. as a means to define the molecular alignment in 3D-QSAR. In this review, the experiences of over a decade of confronting and solving the challenges of pharmacophore discovery applied to actual drug discovery are summarized. Also, practical tips are described for using the author's methodology for pharmacophore discovery, DANTE..

INTRODUCTION

Nel mezzo del cammin di nostra vita	Midway upon the journey of our life
mi ritrovai per una selva oscura	I found myself within a forest dark,
ché la diritta via era smarrita.	For the straightforward pathway had been lost.
Ahi quanto a dir qual era è cosa dura	Ah me! How hard a thing it is to say
esta selva selvaggia e aspra e forte	What was this forest savage, rough, and stern,
che nel pensier rinova la paura!	Which in the very thought renews the fear.
Tant'è amara che poco è più morte;	So bitter is it, death is little more;
ma per trattar del ben ch'i' vi trovai,	But of the good to treat, which there I found,
dirò de l'altre cose ch'i' v'ho scorte.	Speak will I of the other things I saw there

Pharmacophore discovery is a way to bring structure to bear on the drug design problem, when no macromolecular structural data is available. When we first began to tackle the problem of pharmacophore discovery almost fifteen years ago, building upon the "active-analog approach" of G. Marshall and co-workers [1], no one ever imagined that after many years of effort and many advances, the precise, definitive solution to this problem would still elude us. The purpose of this manuscript is to summarize the lessons that have been learned about how to tackle pharmacophore discovery, based on the author's first-hand experiences accumulated over a decade and a half with many datasets in a variety of settings. Virtually all of these datasets with which the author has worked are proprietary; the focus of this manuscript is to describe the general strategies and lessons learned, illustrated by two specific applications to compounds acting against the 5-HT2a/c receptors, and to the oxazolidinone antibiotics, which target the ribosome. The first of these settings was at Abbott Labs, collaborating with Y. C. Martin, where we had developed ALADDIN [2,3], the first successful 3D database search methodology. The second setting was at BioCAD, a startup company in Silicon Valley that produced Catalyst, software which combined both the first robust commercial 3D database search system, and the first commercial pharmacophore discovery method, "Hypothesis Generation"; at BioCAD, the author was both developing the methodology and was applying it to a blizzard of potential applications that came from partners or customers. The third setting was at Upjohn (later to become Pharmacia & Upjohn, later still Pharmacia, now Pfizer), where the focus was on both fixing the methodologies and applying them in a prospective fashion working with drug discovery teams. The methodology that culminated these experiences has been published as DANTE [4-7]. This iterative cycling between developing computational methodology and applying it has been crucial to its evolution into an effective drug design technology. While most of the author's applications with both DANTE and Hypothesis Generation are proprietary, what can be published are the general rules for applying the methodology, based on the lessons learned from this wide-ranging experience. This manuscript documents these general rules and protocols for the first time, distilling the author's experiences of over 15 years in applying pharmacophore discovery to datasets of contemporary interest to pharmaceutical research.

This manuscript is not intended as a comprehensive review of pharmacophore discovery. Reviews have been written by Höltje [8], Bures [9], with a comprehensive and up-to-date review published recently by this author [10].

TERMINOLOGY AND CONCEPTS

Oh quanto è corto il dire e come fioco al mio concetto!	Oh, how faint and ineffective are the words to express my idea!

*Address correspondence to this author at the Vertex Pharmaceuticals, 130 Waverly St.,Cambridge, MA 02139, USA; E-mail: vandrie@mindspring.com

1381-6128/03 $41.00+.00

Definition of a Pharmacophore and Its Components

One may define the term 'pharmacophore' in two ways. One way is to tersely define a pharmacophore as an abstract description of the characteristics of a chemical structure that will confer a particular biological activity. Defined in this way, no particular reference is made to 3D properties. However, customarily to molecular modelers, pharmacophores include 3D properties; to medicinal chemists, the term pharmacophore typically only includes the "2D" properties (the atom types and their connectivity). Everywhere in this manuscript, the term 'pharmacophore' will refer to descriptions of 3D characteristics.

Alternatively, one can define a pharmacophore in a more intuitive, physically-motivated way. A pharmacophore may be defined as the distilled essence of what yields productive ligand-receptor interactions, a defined arrangement of individual interactions. Thus defined, it is explicitly three-dimensional. This physical picture should always be kept in mind when dealing with pharmacophores – they are *not* abstract mathematical entities. One of the reasons that medicinal chemists find pharmacophores so useful is that they explicitly represent fundamental physico-chemical aspects of ligand-receptor interactions.

With either definition of a pharmacophore, one can see that a pharmacophore encompasses a *set* of molecules. Pharmacophoric 3D database searching (Fig. (**1**)) is the process of using a pharmacophore as a search query for a database of conformations of existing molecules; those molecules which emerge has hits from the database search match the pharmacophore, i.e. they also possess the characteristics of chemical structure thought to confer biological activity. These hits can be tested for biological activity; the proportion of those hits with biological activity is a direct measure of the quality of the pharmacophore. Those which test positive may be novel options for new leads. This proportion of hits that test positive typically ranges from 0% - worthless pharmacophores -to 20% - *fantastic* pharmacophores. Lead discovery via pharmacophoric 3D database searching is a means for *prospectively* testing the value of a pharmacophore – the ultimate test. One should *not* judge pharmacophores by the degree to which they "fit" or "explain" the data; as we shall see shortly, many pharmacophores are generally consistent with the biological data.

Figure (**2a**) shows a simple pharmacophore for agonism of the dopamine D1 receptor. The basic amine must be 6.8-8.3 A from the indicated hydroxyl, and the hydroxyl must be 2.7-2.9 A from the center of the aromatic ring. Finally, the center of the aromatic ring must be 4.2-4.8 A from the basic amine. Amazingly enough, it was a simple pharmacophore similar to this that was used in an early application of ALADDIN, that led to the discovery of a constrained analog of dopamine, A-68930, a ligand highly-selective for the D1 receptor which had been synthesized earlier for a project targeting an adrenergic receptor [11].

This simple example of a pharmacophore in Fig. (**2a**) contains many of the parts that typically comprise a pharmacophore:

Features

Substructural elements that are defined purely by atom types and connectivity. The simple example contains 3 features: aromatic ring, hydroxyl, and a basic amine (an amine with a free lone pair, not tied up in conjugation as in an amide or an aniline).

Geometric Objects

3D quantities that are computed from the positions of the atoms in a Feature. In this simple example, the center of the aromatic ring is a Geometric Object not centered on an atom. The other Geometric Objects are atom-centered.

Constraints

All molecules which match the pharmacophore must be able to adopt a low energy conformation where the positions of the Geometric Objects satisfies the constraints. This simple example has 3 distance constraints. One of the key objectives of pharmacophore discovery is to determine the optimal values of these constraints based on the SAR (structure-activity relationship); it is *not* sufficient to merely measure the distances of a low-energy conformer of one molecule, and to add an arbitrary tolerance to these to define the values of these constraints.

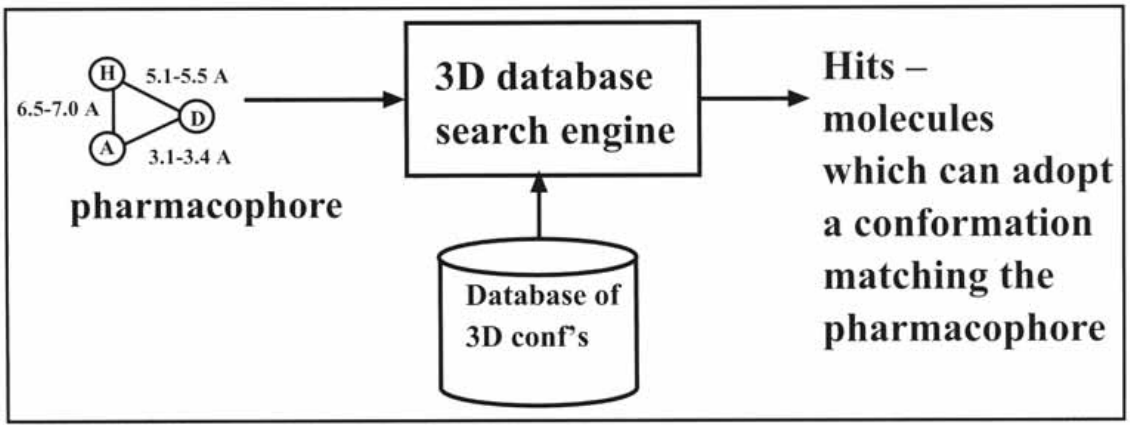

Fig. (1). Overall schematic of the process of 3D database searching.

Note that a pharmacophore, composed of Features, Geometric Objects, and Constraints, is an objective description. It is independent of the orientation of the molecule, independent of how the molecule is drawn, with the only ambiguity inherent in it being 'what constitutes a low-energy conformation?'.

Pharmacophores may be more complex than this simple example in Fig. (**2a**), or even simpler. The simplest pharmacophore aromatic ring 5 to 7 A from a basic amine is surprisingly effective in a 3D database search at retrieving all types of molecules active at GPCR's (G-protein coupled receptors). Pharmacophores can be more complex in many ways:

Number of Features

Pharmacophores typically have 3 Features, sometimes 4, but almost never more than 4.

Types of Geometric Objects

This simple example only shows points. One may also refer to oriented vectors, as shown in Fig. (**2b**). Rarely used, but also possible, are planes, as shown in Fig. (**2c**). The physical interpretation of vectors and planes is that these groups can make oriented interactions with the receptor, e.g. via hydrogen bonds or via pi-stacking interactions.

Types of Constraints

With oriented Geometric Objects, come a variety of constraints. Angle constraints between a point and an oriented Geometric Object are possible, and Torsion Angle constraints can be especially useful between two oriented Geometric Objects, as shown in Fig. (**2d**).

Steric Constraints

It is rare that one hears of pharmacophores with sterically-forbidden regions included (shape constraints), though this is

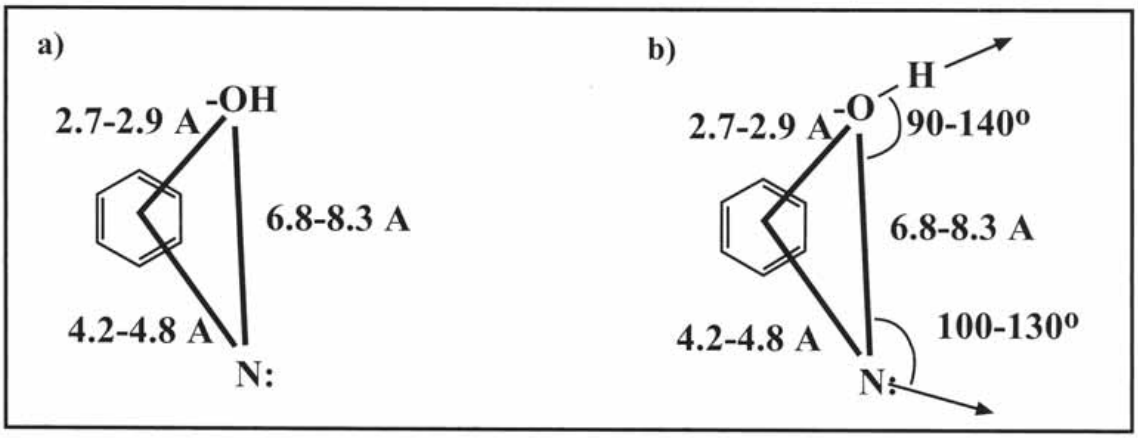

Fig. (2a). Simple D1 pharmacophore.

Fig. (2b). Simple D1 pharmacophore with vector relationships encoded with angles.

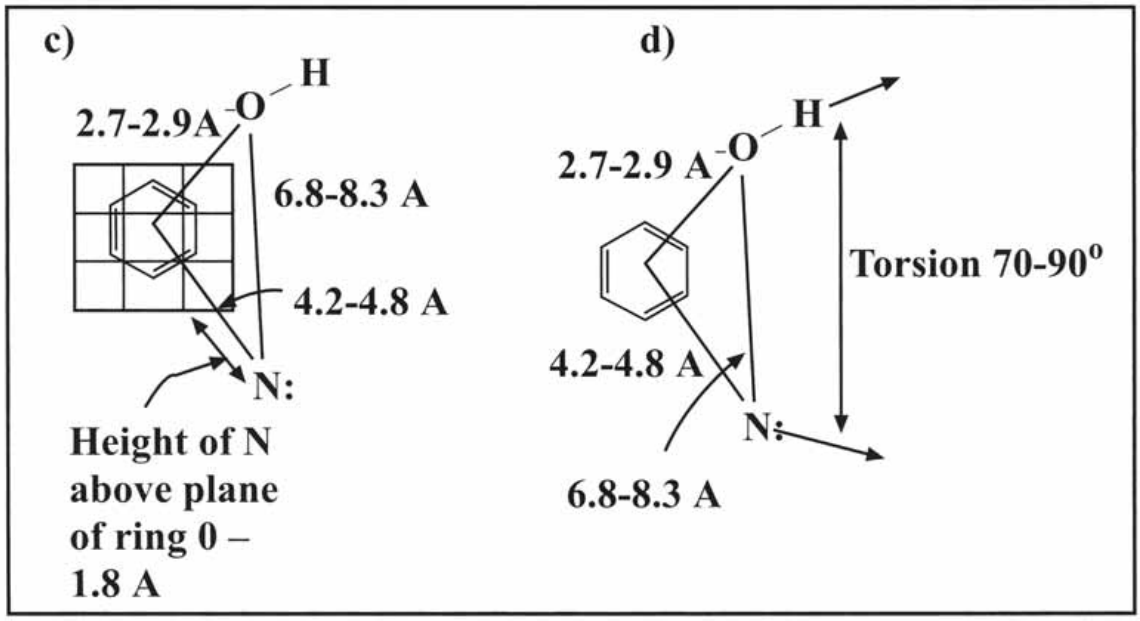

Fig. (2c). Simple D1 pharmacophore with a constraint between the basic amine and the plane of the ring.

Fig. (2d). Simple D1 pharmacophore with torsional relationship encoded.

less a reflection of the physical reality of ligand-receptor interactions and more a reflection of the poor methodology to support the definition of such steric constraints. A dopamine D2 pharmacophore similar to the one we used with ALADDIN is shown in Fig. (**2e**); it contains one steric constraint, a sphere of radius 3 A positioned at specific distances from the Geometric Objects defined on the Features. It was an egregious mistake to omit steric constraints from the automatically-generated pharmacophores of Catalyst; typically the guidance given the user in such cases is to omit molecules from the dataset, when the poor activity and steric size are suggestive of the need for a steric constraint. By contrast, DANTE pharmacophore discovery introduced a novel method for defining and using steric constraints, the "shrink-wrap algorithm".

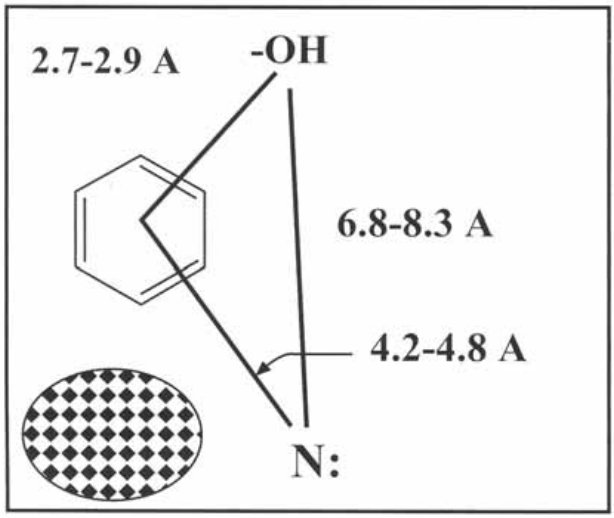

Fig. (2e). Sterically-forbidden region constraining space of active analogs. This region can be defined using either relative coordinates (distances from the center of the region to the three features), or absolute coordinates (positioning the pharmacophore features at specific Cartesian coordinates, and then positioning the region in that same coordinate frame).

Pharmacophores as Used in Lead Discovery

All medicinal chemistry efforts must begin with a *lead*, usually a molecule with modest biological activity against the chosen target, devoid of ineradicable liabilities, which is amenable to synthetic exploration. *Virtual screening* is the process of discovering leads by sifting through an electronic database of existing compounds, picking a subset which have defined properties, and submitting the hits from this database search to biological screening. Virtual screening may be done using either protein structures [12], or using a pharmacophore [2,3,11], in either case by electronically screening a *3D database*, a collection of conformation(s) of the molecules in the database. This process is depicted schematically in Fig. (**1**). Virtual screening had its biggest impact in the pharmaceutical industry in the days before high-throughput screening became commonplace, but it is still widely-used in special circumstances, e.g. where the biological assay is not amenable to high-throughput, or in academic environments, where massive compound libraries are not usually available.

Pharmacophores as Used in Lead Optimization

The core process of any drug discovery process is the iterative exploration of analogs around a lead, *lead optimization*. The essential question of the medicinal chemist in lead optimization mode is 'what molecule should I make next?'. The simple pharmacophores shown in Fig. (**2**) don't provide much assistance in answering this question, because usually every molecule in the series the chemist is exploring matches the pharmacophore. But pharmacophores can assist in answering the fundamental *cri de coeur* of the medicinal chemist, in multiple ways:

1) 3D-QSAR models may be used to predict the biological activity of proposed molecules [13], using methods like CoMFA [14]. Until recently, it has been little appreciated that a prerequisite for the construction of such models is a proper overlay of the conformations of the molecules in the dataset. Pharmacophores may be used to define the rules for overlaying molecules. Most published 3D-QSAR studies spent much time describing the statistical analyses, etc., but little time describing how the molecules were overlaid. Fortunately, current publications devote more effort to describing how the overlays were performed, either explicitly using a pharmacophore discovery method, or implicitly via the protocol used in performing the overlay (for a review, see [10]).

2) Shape-enhanced pharmacophores, of the type generated by DANTE, describe both the geometric arrangement of features and the steric boundaries of the binding site, as can be inferred from the dataset. While in theory the steric and electronic fields of CoMFA provide greater sophistication in predicting activity, the steric boundaries as derived by DANTE have demonstrated to be surprisingly useful in prospective applications, by defining the 'limits of the playing field', i.e. constraining the space of possible molecules a chemist should consider. Furthermore, unlike 3D-QSAR models, DANTE's shape-enhanced pharmacophore can easily be used as 3D database search queries, to screen databases composed of combinatorial libraries constructed around the lead (Fig. (**3**)) [7]. A greater utility of DANTE's shape-enhanced pharmacophores comes from their use in driving exploration of *terra incognita*. This is a very important but underappreciated concept for molecular design in lead optimization, as medicinal chemists need to discover novel compounds, and need to explore regions of space hitherto unexplored. In DANTE, regions of the binding surface are marked either as 'sterically forbidden' (those molecules in the dataset which are active lie within that boundary, while inactive molecules in the dataset protrude beyond it), or '*terra incognita*', i.e. active molecules lie within that region, and define the extent of that surface, but no molecules in the dataset protrude beyond that region. One can use a DANTE shape-enhanced pharmacophore to explicitly look for molecules to extend into these undefined regions, to probe new regions of chemical space, to see if the properties improve (Fig. (**4**)). In this example, a search was performed of a database of novel ring systems attached to a phenyl-oxazolidinone; those ring systems which lie within the sterically-forbidden regions, but

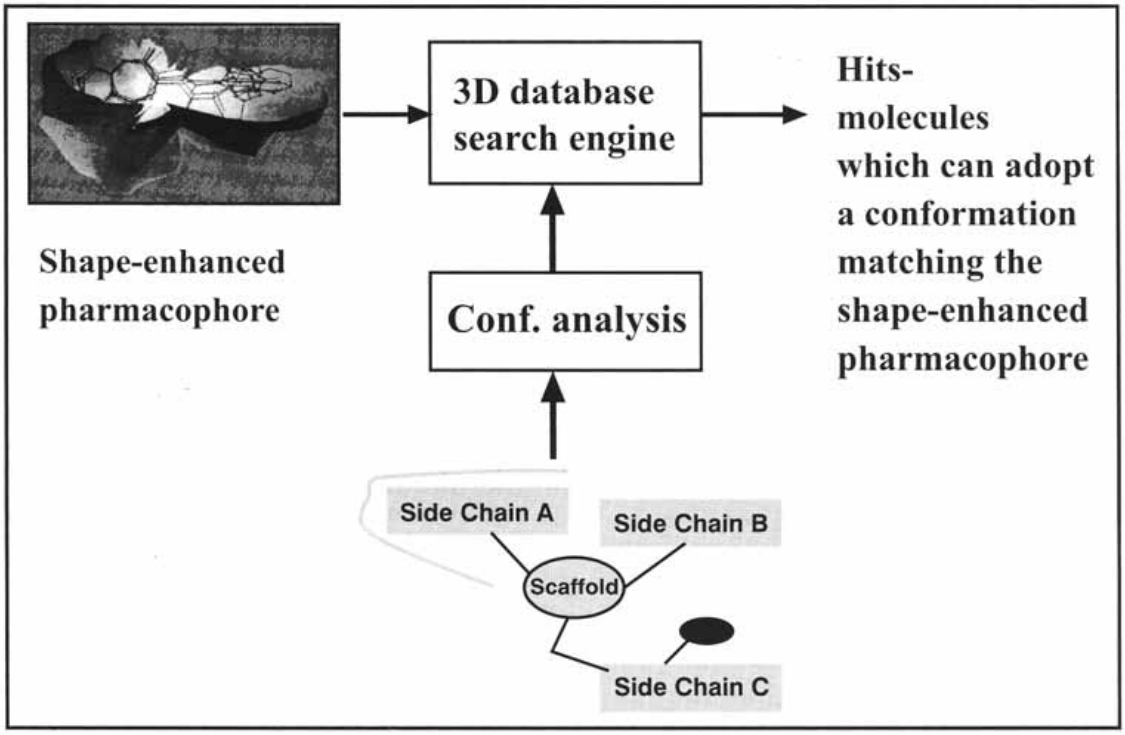

Fig. (3). Schematic flow of how to use shape constraints and terra incognita to virtually screen a combinatorial library. The shrink-wrap surface imposes constraints on reagents for A.

probe the *terra incognita*, were sought. In this way, a novel ring system was found. Note that 3D-QSAR methods attempt to assign electrostatic or steric field values at all regions of space, and have no way to indicate "don't know" for regions of space where the data in the dataset do not allow inferences to be made.

PROTOCOLS FOR PHARMACOPHORE DISCOVERY

Lume v'è Dato a Bene e a Malizia	You have been given reason, which can distinguish between bad and good.

The activities which one must undertake to discover and use a pharmacophore may be divided into 7 distinct stages:

1) Dataset preparation
2) Conformational analysis
3) Enumeration of candidate pharmacophores
4) Ranking of candidate pharmacophores
5) Overlaying molecules according to the best pharmacophore(s); determination of steric constraints/*terra incognita* or molecular fields for 3D-QSAR
6) Computational controls to evaluate robustness of the results
7) Modeling the activity of proposed new compounds; prospective application of the pharmacophore model

Our final goal is a high-quality pharmacophore, useful prospectively in the design of new bioactive molecules. Let us consider each of these steps along the path in more detail.

Dataset Preparation

At first glance, this step sounds trivial. Yet, the mere fact that it took multiple publications over a 5-year period until the "Cramer steroid dataset" was finally rid of all errors in chemical structure as well as all errors in biological data attests to how subtle and insidious this issue can be. The goals of dataset preparation should be (1) to verify the correctness of all chemical structures, especially stereochemistry, and, if the data source was a chemical database, discarding salts and other extraneous stuff (2) to ensure that all biological data is correct, and was gathered in a common way, (3) to ensure that the biological data is mechanistically homogeneous, in so far as that is possible to ascertain. Goal 2 warns against mixing data which comes from different assays or the same assay run by different labs; this is especially problematic when extracting data from publications. Most researchers in pharmaceutical companies rely on data retrieved from a corporate database, which generally can be relied upon for accurate structures and biological data acquired by one protocol. Goal 3 is always the most challenging goal to achieve. The notion of a pharmacophore assumes that there is a common pattern of ligand-receptor interactions among the molecules in the dataset. If that is not true, one is bound to fail. While some pharmacophore discovery methods, like DANTE, explicitly look to see if this assumption holds true, most do not, and it is up to the user to rely on biophysical data to provide guidance.

Because of Goal 3, some problems are simply not amenable to pharmacophore discovery. If, for example, your biological readout were %F, bioavailability, one would *never* want to derive a pharmacophore based on that, as there are

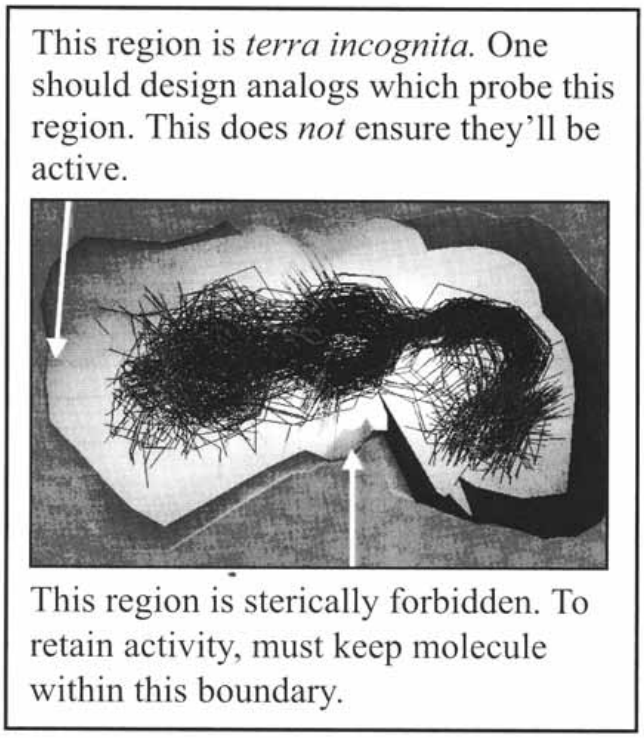

Fig. (4). How to use *terra incognita* as a design constraint.

many many mechanisms which contribute to the overall value, e.g. absorption, metabolism, clearance. If the biological readout were 'ability to be metabolized by cytochrome P450's (CYP's)', this would still be too heterogeneous, as multiple CYP's give rise to the overall value. If the biological readout were 'ability to bind to CYP 2D6', then potentially one may have a dataset that is mechanistically homogeneous.

Mechanistic inhomogeneity can still creep in, even if the biological readout results from binding a single, pure receptor. Multiple binding modes are difficult to detect in the absence of direct structural data, yet if they occur they will pose serious challenges for the pharmacophore discovery exercise (one of the best studied examples of multiple binding modes in a single series is the Roche thrombin inhibitors [15], where minor changes to the chemical structure leads to a new binding mode flipped 180 degrees about an axis – a phenomenon that has been called *precipitous binding modes*). The DANTE methodology is one of the few pharmacophore discovery methods that can detect this type of mechanistic inhomogeneity.

One thing which must be avoided in this step of dataset preparation is biasing the selection of compounds for the dataset to circumvent limitations in the pharmacophore discovery method being used. Users of Catalyst's "Hypothesis Generation" have long been encouraged to carefully select compounds to "teach" the computer properly. This introduces an enormous subjectivity into the process, and if the results are highly-dependent on the selection of compounds from a larger set of mechanistically-homogeneous compounds, then the methodology is weak. Good methodology leads to reproducible results – the hallmark of any scientific method.

Conformational Analysis

Some method of exhaustive conformational analysis must be applied to all the molecules of the dataset. It does not suffice to merely use the minimum energy conformation of all molecules in the dataset. By 'exhaustive conformational analysis', one refers to a process that systematically explores the entire conformational space of a molecule; this frequently includes exploring both stereochemistries for any chiral center specified as racemic. Many tools are available for this step. Below are a subset of those available tools, ones with which the author has had direct first-hand experience. The key tradeoffs regarding the protocol to follow are: (1) what energy threshold τ should be used (i.e. find all conformations within τ kcal/mol of the energy minimum), (2) how should duplicate conformations be detected and/or removed (or, alternatively, how should the results be clustered into families of conformations), and (3) how much time must be consumed. Conformational analysis is typically the slowest step in pharmacophore discovery.

Catalyst "Best" Conformational Analysis

This method [16] relies on a combination of algorithms for systematically exploring all conformational space, using a stripped-down version of the CHARMm force field to evaluate the energetics, with a novel term added to the energy function to ensure distinct regions of conformational space are explored. The default recommendation for the energy threshold τ is 20 kcal/mol, which is ridiculously large (if one takes that to be physically meaningful, one is including conformations whose probability is $\exp(-\tau/kT) = 3x10^{-15}$). In experiments with this conformational analysis method, this author has generally found a value of 8 kcal/mol to be acceptable; this was based on analyzing oxazolidinone antibiotics, comparing the active analogs vs. their ring-opened inactive analogs. Overall, one is left with the sense that the primary weakness of Catalyst's "Best" conformational analysis is the force field that is used; too frequently, odd conformations are chosen as the ones matching the pharmacophore. The high value of τ may be needed in part to overcome weaknesses of the force field. Also, this conformational analysis method appears to generate too few conformers, and it is difficult to adjust the variable parameters which control the clustering of conformers. This sparsity of conformers primarily manifests itself in poor coverage of the surface elements in the DANTE shrink-wrap surfaces. The speed of this method of conformational analysis is a limitation; one can easily spend 8 CPU-hours on a reasonably-sized dataset, and currently one is limited to Silicon Graphics computers, which are no longer the fastest. (Catalyst has another method for conformational analysis, the "Fast" method, which is appropriate only for the construction of 3D databases).

Macromodel Monte-Carlo Multiple Minimum (MCMM)

This method [17] stochastically populates different regions of conformational space, and allows a variety of different force fields to be used to evaluate energetics; it uses an rms threshold for discarding similar conformations. For the types of molecules one presents to pharmacophore discovery, the Merck Molecular Force Field [18] appears to work best. Experimentation with steadily lower values of τ continue. The lowest values with which this author has had

success are 5 kcal/mol; this issue continues to be debated in the literature, with suggestions that values as low as 3 kcal/mol are acceptable [19]. For the quality of conformations, this author's experience suggests that MCMM is the best option for conformational analysis in pharmacophore discovery. The default setting for detection of duplicates (0.25 A rms deviation) ensures a high degree of coverage, even for mapping the shrink-wrap surface. The sole drawback is the execution speed, which can be measured in CPU-days for reasonably-sized datasets (though it should be noted that MCMM runs on most computers, and that this process is readily parallelizable, so multiple CPU-days can convert to acceptable elapsed times).

OMEGA

This unpublished method [20] appears to mainly rely on torsion-driving using a defined set of allowed torsions for each rotatable bond type, and does not have methods for assessing energetics or duplication. Its main advantage is speed (minutes for most datasets). It is recommended that post-processing be done to evaluate energetics, e.g. using Macromodel. The conformational coverage appears to be too coarse for routine use in pharmacophore discovery, though this is dataset-dependent.

CONFORT

This unpublished method [21] uses a sophisticated method for adaptively determining the optimum angles for driving torsions, among other things. Like OMEGA, it needs post-processing to evaluate energetics. Its speed lies between OMEGA's and Catalyst's Best method; the improved sampling of torsional angles comes at a computational cost.

Final Comments on Conformational Analysis

The list of methods above is not intended to represent a comprehensive list. It should also be noted that some pharmacophore discovery methods integrate in one program the operations of conformational analysis with those of pharmacophore discovery. The difficulty with these integrated approaches is that, when the wrong answer emerges from the software, it must be difficult to isolate the problems to the conformational analysis method, or the pharmacophore discovery method. Also, various workers augment their protocol with steps which cluster and/or minimize the conformations (for example, [22]); in contrast, the approach this author has generally followed using DANTE is to create more conformations than may be necessary, thousands per molecule if needed, and to let the natural clustering of the pharmacophore discovery method weed out unneeded conformers. Pharmacophore discovery is not the slow step, so an extra conformer is not a disadvantage, while a missing conformer may degrade the quality of the final result. (This was not always true – DANTE originally ran on a SGI Personal Iris with 32 Mb RAM – but with cheap powerful machines commonplace today, clock speeds > 1 GHz containing multiple Gb of RAM, this is not an issue).

Enumeration of Candidate Pharmacophores

Most datasets are compatible with many possible pharmacophores. Failure to appreciate this is the greatest source of subjectivity and irreproducibility in pharmacophore discovery; method A homes in on one pharmacophore, while method B homes in on a different one. The proper approach to pharmacophore discovery is to enumerate *all* pharmacophores consistent with a dataset, what may be termed *candidate pharmacophores*. The possibility also exists that there are no pharmacophores consistent with the dataset; this procedure of enumerating all candidate pharmacophores must also reckon with that possibility. There are some model building methods, e.g. COMPASS [23], which make an implicit assumption that a pharmacophore exists, and it is unique, and it is merely an algorithmic challenge to find it.

The procedure used in DANTE to enumerate all pharmacophores follows these three steps:

1) For all molecules in the dataset, for all conformations of each molecule, for each feature-type from a standard library of feature-types, identify all features on that conformation. The standard library of feature-types is the set we introduced with the Catalyst software: A,D,N,P, H,R (hydrogen bond acceptor, hydrogen bond donor, a group negatively charged at physiological pH, a group positively charged at physiological pH, a hydrophobic group, and an aromatic ring).

2) If triad pharmacophores are sought (pharmacophores containing 3 features), for all molecules in the dataset, for all conformations of each molecule, tabulate all inter-feature constraint values for all possible triads on each conformation. Each triad of inter-feature constraint values is referred to as an "MRS point" (point in Molecular Recognition Space). An MRS point may contain only 3 distances, if the 3 features are ones without orientation (N,P,H), or it may contain 3 distances and angles and torsions, if one or more features are oriented (A,D,R).

3) The tightest cluster of MRS points is identified, following the idea of Mayer, Naylor, Motoc, and Marshall [24], hereafter referred to as the MNMM algorithm (Fig. (**5a**)). This cluster should contain all the molecules in the dataset. The constraints for the pharmacophore are those that enclose this cluster, i.e. the minimum and maximum for each distance, angle, or torsion among all MRS points in the tightest cluster. Note that the MNMM algorithm works for any combination of distances, angles or torsions.

This final clustering step is the heart of the DANTE pharmacophore discovery method. Note that it is here where one can detect *if* a pharmacophore exists in the dataset – if no tight clusters emerge, then one is likely faced with that situation. Here one can also detect if the dataset is best partitioned into multiple pharmacophores: if the array of MRS points is best described by two tight clusters, each cluster containing a distinct subset of the dataset, then one may have a heterogeneous dataset (Fig. (**5b**)). Outliers are frequently evident in this clustering step, i.e. a tight cluster of MRS points exists for all molecules in the dataset but one (Fig. (**5c**)).

Evaluation and Ranking of Candidate Pharmacophores

When one applies the procedure described in the previous section to a variety of datasets, one is immediately struck by the problem that typically many pharmacophores can be inferred from the data. In fact, it is a hallmark of "easy"

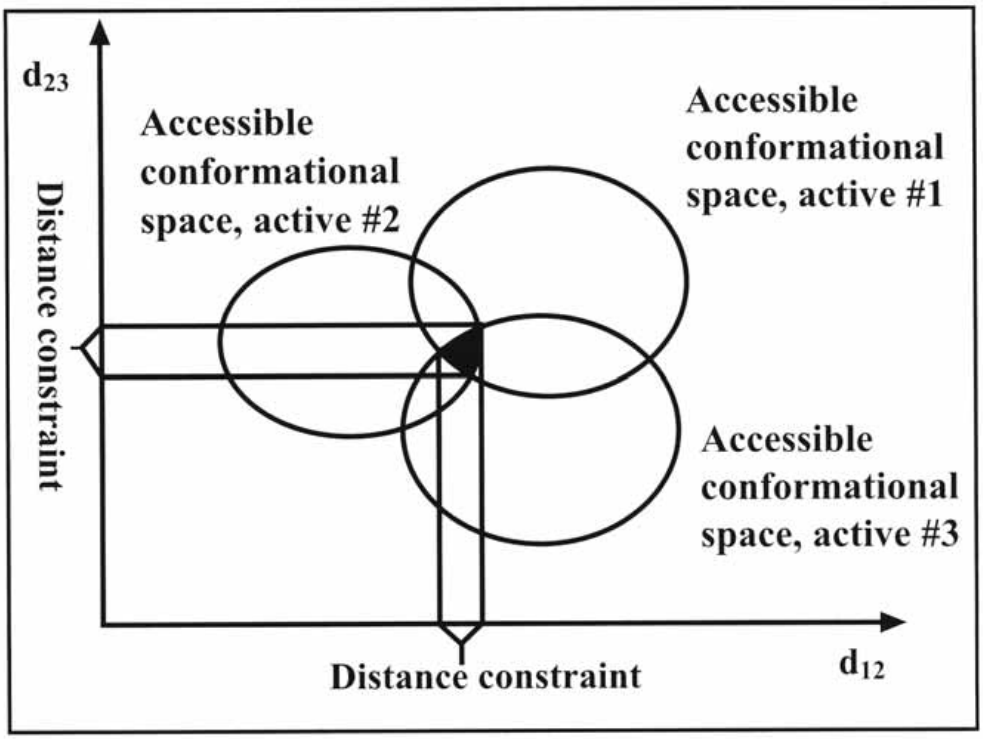

Fig. (5a). Algorithm of Mayer, Naylor, Motoc and Marshall.

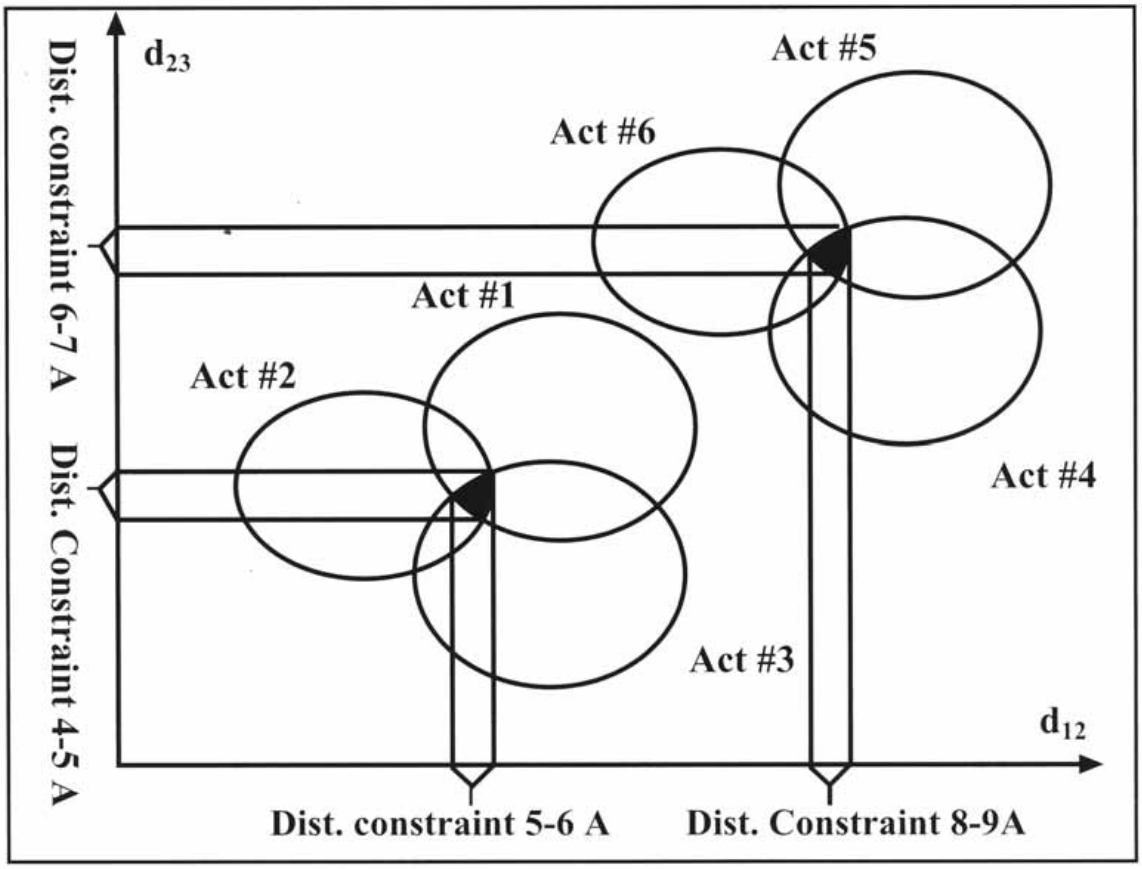

Fig. (5b). Algorithm of Mayer, Naylor, Motoc and Marshall in the presence of dataset heterogeneity. A pharmacophore with d12 constrained within 5-6 A and d23 constrained within 4-5A encompasses actives 1,2,3. A pharmacophore with d12 constrained within 8-9 A and d23 constrained within 6-7 A encompasses actives 4,5,6.

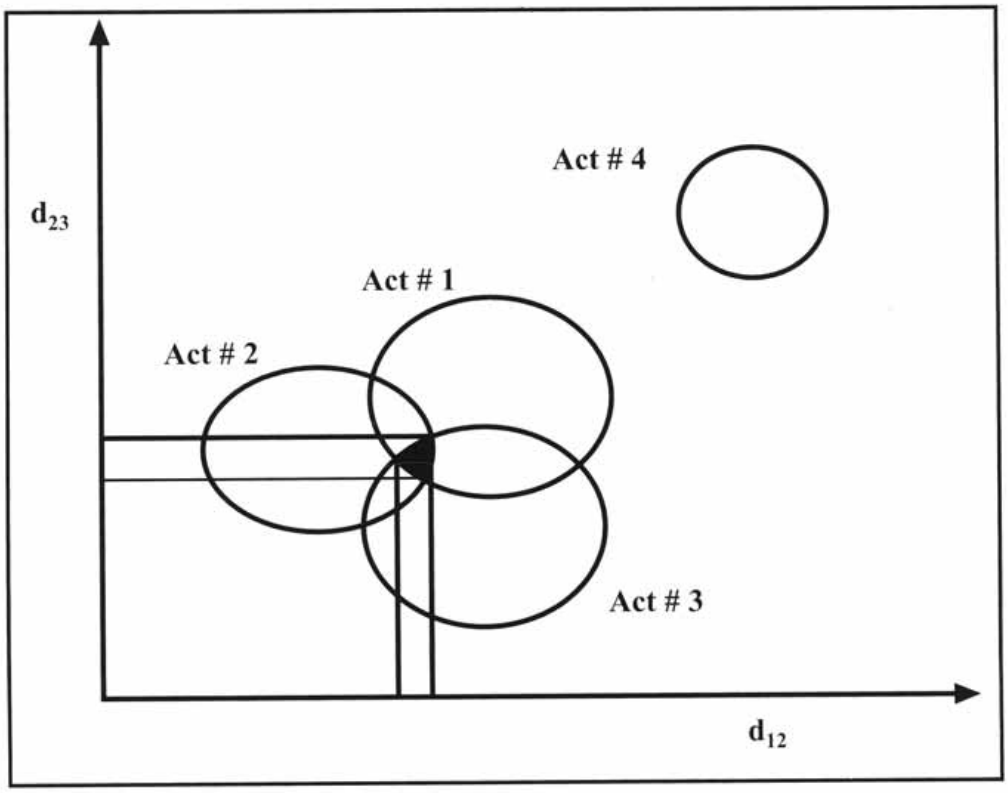

Fig. (5c). Algorithm of Mayer, Naylor, Motoc and Marshall in the presence of one outlier.

datasets that a single pharmacophore emerges from this procedure; the ACE dataset of Mayer *et al*. [24], is such an example. However, most datasets one encounters in the course of real drug discovery, and especially those which are most challenging to the intuitive approach of medicinal chemists, are ones for which multiple candidate pharmacophores emerge. How should one sift through these different possibilities, and identify the "best" one(s)?

The unique aspect of DANTE is that it uses the *principle of selectivity* to rank these candidate pharmacophores, with more selective pharmacophores preferred over less selective ones [5]. The selectivity index S is derived from notions of probability theory. For each candidate pharmacophore, one performs a 3D database search against a drug-like database, e.g. the database supplied by the National Cancer Institute [25]. Denote the proportion of the database returned as hits by that candidate pharmacophore as q. One can interpret q as the probability that that candidate pharmacophore will match a randomly-chosen druglike molecule. Hence, if our dataset contains N molecules, and the pharmacophore hits all N molecules, the likelihood that N random druglike molecules will all match the candidate pharmacophore is q^N, which is called S, the index of selectivity. The smaller the value of S, the less likely that candidate pharmacophore may have arisen by chance. Hence, by assessing q for all candidate pharmacophores, computing S for all candidate pharmacophores, and ranking all candidate pharmacophores by S, one can rank them according to the statistical likelihood that they do not represent a chance correlation.

If a candidate pharmacophore hits on a subset of M molecules in the dataset of N molecules, one must introduce a sum over the tail of the binomial distribution to properly evaluate this statistical likelihood:

$$S = \sum_{k=M}^{N} C_k^N q^k (1-q)^{N-k}$$

where C^N_k represents the number of ways k things may be selected from N things, $\frac{N!}{(N-k)!k!}$.

The only time-consuming step in this ranking procedure is the assessment of q for each candidate pharmacophore; with massively-parallel computing clusters, nowadays this is not an issue, though earlier one needed to rely on a mathematical trick to provide lookup tables for rapid estimation of q [4].

Because typical values of S are so small, it is more convenient to work with pS = -log(S). Values of pS smaller than 5 usually indicate a weak pharmacophore, i.e. if the most selective candidate pharmacophore only has a pS value of 5 ($S=1.0x10^{-5}$), this is a dataset that does not point clearly to a pharmacophore. Values of pS in the range 5 to 7 are moderately good pharmacophores; 7 to 10 are good pharmacophores; >10 are exceptional. Values of pS > 10 are characteristic of SAR's which have been well-worked-out, long after the drugs have already been discovered, like the ACE pharmacophore of Mayer *et al.* [24], or the D2 and beta-2 pharmacophores used as examples in the original DANTE publication [5].

The principle of selectivity is also useful in determining the degree to which a pharmacophore must be elaborated. Considering a dataset composed of 10 proprietary 5-HT2a/c

agonists [26], one can begin by discovering the simple dyad (two-feature) pharmacophore shown in Fig. (**6a**), a basic amine and an aromatic ring constrained to lie within a distance of 5.0 – 5.2 A. This selectivity values for this pharmacophore are q=0.15, pS=8.3. If one looks for triad pharmacophores containing this dyad, one discovers the pharmacophore shown in Fig. (**6b**), q=0.10, pS=5.1. This is a step *backwards*, and hence it is not productive with this dataset to advance from dyads to triads. Next, if one considers a dyad pharmacophore with angles and a torsion (taking the lone pair on the basic amine as providing a directionality, as is customarily done in modeling CNS compounds), one discovers the pharmacophore shown in Fig. (**6c**), q=0.04, pS=13.7. Note also that, due to the signed torsion angle, this pharmacophore is chiral, though it is only a dyad. Measured by the selectivity index alone, this looks like an excellent pharmacophore, though in fact it will hit agonists of many GPCR's. Additional selectivity comes from shape constraints.

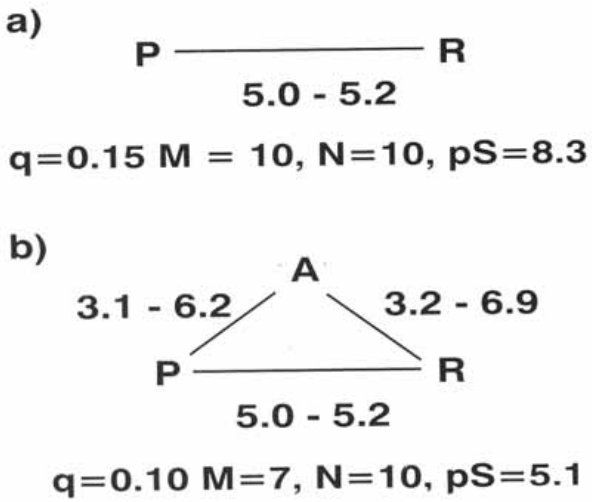

Fig. (6a). Simple dyad 5-HT2a pharmacophore.

Fig. (6b). Simple triad 5-HT2a pharmacophore.

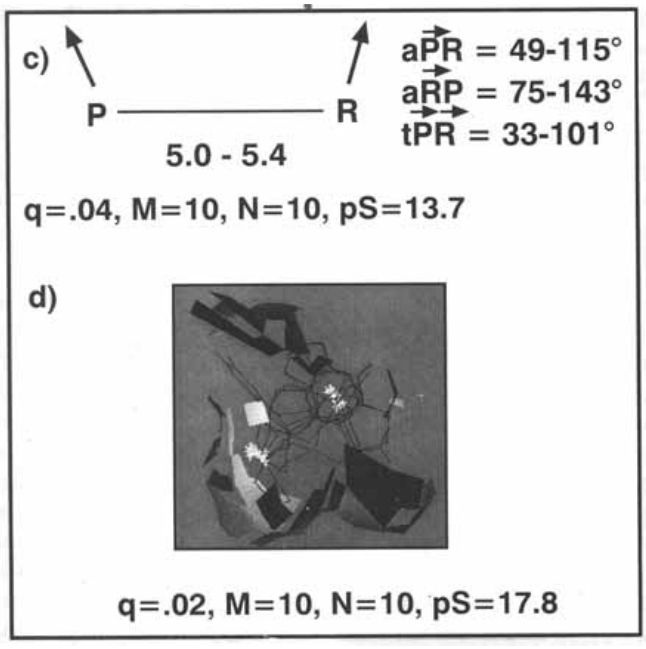

Fig. (6c). Simple dyad 5-HT2a pharmacophore with angle. torsion constraints.

Fig. (6d). Pharmacophore in 6c with additional steric constraints.

Overlaying Molecules According to the Pharmacophore, Mapping Steric Boundaries, Identifying Terra Incognita

In the development of ALADDIN, this author spent a lot of time investigating the pro's and con's of using absolute coordinate systems (molecules positioned in Cartesian x,y,z coordinate systems) vs. relative coordinate systems (internal coordinates, distances, angles, etc.). Ultimately, the ALADDIN language provided either one to the user, but what emerged after much experimentation is that internal coordinates worked best for defining the types of pharmacophores that have been discussed up to this point, but that absolute coordinates work best for the description of shape, i.e. steric constraints. While Catalyst's "Hypothesis Generation" ignored these lessons learned, in DANTE pharmacophore discovery, what appears to work best is to discover the initial pharmacophore (features; distance, angle, and torsion constraints) in internal coordinates and then to use that pharmacophore to revert to absolute coordinates, and then to infer the shape constraints. One is left with two distinct problems: (1) how to convert pharmacophores composed of distances, angles, etc. to absolute coordinate frames, i.e. how to align all molecules in 3D space, and (2) how to infer the shape constraints.

To align molecules in 3D space, one selects out those conformations that match the initial pharmacophore. One can either arbitrarily pick one of these conformations as a reference (a step which introduces subjectivity), and overlay all other conformations to that using the Kabsch algorithm [27], or, more systematically, one can construct geometrically a set of (x,y,z)'s whose internal coordinates match the relationships of the centroid of all the MRS points which belong to the cluster from which the pharmacophore was formed, and overlay all conformations to those (x,y,z)'s (a hybrid approach is also possible – to choose the one conformation whose MRS point is closest to the centroid, and use that as a reference).

Once aligned, one has an ugly composition – an explosion of many molecules, possibly hundreds of conformations, all aligned in a way consistent with the pharmacophore. Inference and application of shape constraints suddenly turns it into a thing of beauty – reminiscent of what is in every chemist's and pharmacologist's mind of how these molecules may be binding to the receptor.

In DANTE, the procedure that is used is called 'shrink-wrapping' [6]; while it is sometimes called the 'shrink-wrap algorithm', properly the procedure used by DANTE combines (1) an algorithm, (2) a novel shape representation, and (3) a principle of minimum volume. In DANTE, the shape of the binding site is inferred by computing the surface enclosing the smallest volume that contains at least one conformer of each active molecule. This shape is represented by a series of polygons mapped onto a surface topologically equivalent to a sphere. The algorithm for performing the volume computations is described in detail elsewhere [6].

There is no guarantee that the surface of minimum volume should be physically identical to the binding site of the receptor. Empirically, this principle of minimum volume appears to work. Note that most SAR's do not allow one to define the entire surface; polygons are marked as 'sterically forbidden' when an inactive molecule must protrude though it, otherwise they are labeled as *terra incognita*. An example of the net result is shown in Fig. (**6d**), a patchwork of forbidden regions surrounded by transparent *terra incognita*; literature 5-HT2a compounds are shown mapped into this fully-fledged pharmacophore model [28]. Note that this approach is a refinement of the approach originally developed by Garland Marshall and co-workers, where the union volume of the actives is intersected with the union volume of the inactives to define sterically forbidden regions [1]. Fig. (**7**) shows a rare case of a binding site being essentially fully defined by the SAR, with almost no *terra incognita*: the binding site of the oxazolidinone antiobiotics [29], as inferred from the SAR of over 3,000 analogs, measuring the biological activity as the MIC against the bacterium *Haemophilus infuenza*. This is a cutaway view of that surface, cut away in a manner which intentionally obscures what little *terra incognita* exists.

Computational Controls

A colleague of mine likes to use the term "gonzo pharmacology" [30] to describe the process where a biologically active compound is applied to cells, a response is observed, and inferences are made based on that one experiment as to the molecular basis of how that response is generated. Part of the training of pharmacologists and cell-biologists includes engendering a deep respect for the importance of control experiments, to elucidate the molecular mechanisms of the behavior of a complex set of interrelated pathways, and to weed out incorrect hypotheses about what those mechanisms might be.

Too often, in molecular modeling, we are guilty of "gonzo modeling": running some data through a program, observing the results, and asserting "Eureka!", without running the types of additional computational experiments that ascertain whether one has discovered a significant relationship or not. Nowhere are computational controls more important than in the area of pharmacophore discovery.

The types of computational controls that this author has used to evaluate the quality and significance of both DANTE [5] and Hypothesis Generation are:

1) Adding noise to conformations. It is a simple matter to repeat the pharmacophore discovery operation, where after performing conformation generation, a small amount of noise is added to each coordinate of every conformation. One anticipates that adding small quantities of noise, e.g. 10^{-7} A, should leave the results unchanged. One anticipates that intermediate amounts, e.g. 0.001 A, should produce comparably-sized differences in the results. Surprisingly, the results of early versions of Hypothesis Generation were not robust relative to even

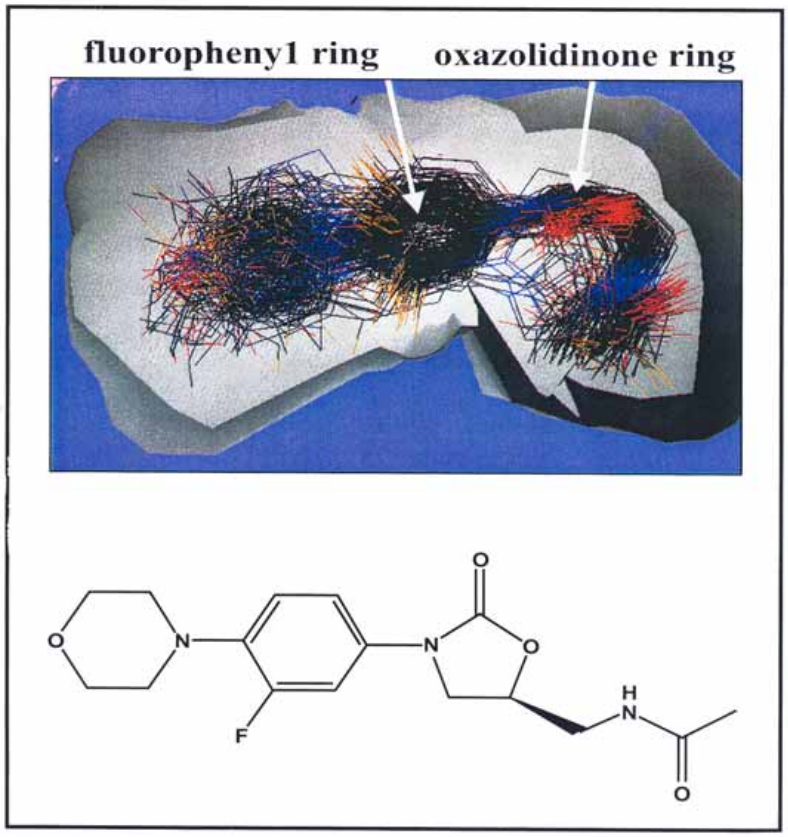

Fig. (7). Cutaway view of the shrink-wrap surface of oxazolidinone antibiotic binding site for *Haemophilus Influenza* activity. This steric boundary is defined by over 3,000 oxazolidinone analogs. The structure shown is that of linezolid, the first oxazolidinone antibiotic, approved for use in 2000.

the smallest quantities of noise: upon addition of noise of magnitude 10^{-7} the output was totally different – different features selected, changes in distances of the order of 1 A. DANTE results behave as anticipated [5].

2) Permuting atom indices. One anticipates that order by which atom numbers are assigned should have no effect whatsoever on any computation, e.g. that numbering the oxygens of dopamine 1,2, the nitrogen 3, and all the carbons 4-11 should give the same result as numbering the carbons 1-8, the nitrogen 9, the oxygens 10,11. This is another computational control that early versions of Hypothesis Generation flunked (non-programmers find this hard to believe, but it is actually non-trivial to write software that never makes the assumption "let's start by looking at atom 1…"). DANTE relies everywhere on a subgraph-isomorphism algorithm to refer to pieces of molecules, and is robust relative to this computational control [5].

3) Adding randomly-chosen molecules to the dataset. This is another challenge to most pharmacophore discovery algorithms, as they frequently begin with the assumption that a pharmacophore exists for all the molecules in the dataset. By contrast, DANTE, via the MNMM algorithm, *looks* for a pharmacophore – randomly added molecules will appear as outliers. DANTE is usually robust to this type of control [5].

4) Apply the entire pharmacophore discovery protocol, beginning to end, on a dataset of randomly-selected drug-like molecules. One anticipates some type of feedback from the software, suggesting that no common patterns are discernible. In DANTE, the selectivity index is a guide, in the published examples always yielding a value of the selectivity index $pS < 5.3$ [5]. At first glance, this computational control sounds foolhardy, yet one is frequently faced with the task of evaluating an SAR produced by a biological assay, where it later turns out that the biological readout was laden with artifacts [31].

These are only some examples of the types of computational controls one could apply. This should be an opportunity for some creative thinking among modelers, to better protect us from drawing conclusions from chance correlations.

Prospective Application, Design, Synthesize, Test, Iterate on Model Building

The *sine qua non* of any model-building exercise is the prospective application to molecular design. By prospective, we mean applying the model to molecules that have never been seen during the model-construction activity. This is distinct from the popular techniques like cross-validation, where a portion of the dataset is partitioned off, not used in the model-building, and against which the model is applied as the final step. The vicissitudes of these retrospective approaches are becoming steadily clearer [32,33]; in fact, Kubinyi has summarized his experience with the observation that those models that fit the data best tend to work prospectively the worst, an observation this author has dubbed the "Kubinyi Paradox".

Prospective application is possible via collaborations with medicinal chemists, who can make molecules suggested by the model, or who will use the model to choose among many possible molecules they devise. The success of a pharmacophore is measured by frequency with which those new ideas are successful, compared to the frequency experienced by the medicinal chemist, relying on his/her intuition alone. Prospective validation may also be achieved by using the pharmacophore to virtually screen molecular databases; this is one of the advantages of pharmacophore models, that they can be directly coupled to a 3D database search.

SUMMARY OF LESSONS LEARNED

vuolsi così colà dove si puote	It is so willed there where is power to go
ciò che si vuole, e più non dimandare	That which is willed; and ask no further question

One may briefly summarize the lessons learned as follows:

1) Most datasets are consistent with many pharmacophores. The challenge in pharmacophore discovery is to enumerate *all* of these candidate pharmacophores, and to rank order them. The principle of selectivity - rank-ordering pharmacophores by their selectivity – is a useful method to identify the best pharmacophores

2) A facile link from pharmacophore discovery to 3D database searching is vital. This is useful both in performing prospective validation of the model, and in assessing different candidate pharmacophores, to determine which is most selective.

3) Inclusion of sterically-forbidden regions, or shape constraints, is vital to achieve good selectivity.

4) Identification of *terra incognita* is very important, and directing design towards uncovered territory is useful.

5) Computational controls are vital, to protect against chance correlations.

6) Some datasets are easy; some are hard. Medicinal chemists rarely need help with the former, and our methods are weakest on the latter. The challenge is to discover good pharmacophores for hard datasets; hence the need for a careful adherence to the proper protocols. This author has been searching for years for a metric that could quantify the intuitive sense of a dataset being "hard" or "easy"; one idea has recently emerged from studies of non-additivity, which should appear in print soon.

7) Mechanistic inhomogeneity among the molecules of a dataset can present an insuperable challenge for any pharmacophore discovery method. This is the one condition where picking subsets of the dataset is justified; beyond that, a good pharmacophore discovery method should process all the data.

8) Success in prospective application is the sole measure of the quality of a pharmacophore.

KEY OPEN ISSUES AND CONTROVERSIES

Le leggi son, ma chi pon mano ad esse?	There are the laws, but who commands respect to them?

The following questions represent, in the author's opinion, the crucial unresolved issues in the area of pharmacophore discovery, and point the way towards future investigations:

1) *What value of* τ *is best?* Values for the energy threshold in conformational analysis frequently are chosen based on tradition, with some empirical evidence based on actual studies suggesting values of 3-5 kcal/mol are best. In theory, the idea depicted in Fig. (**8**) should be the ideal way of addressing this. Referring back to the basic idea of the MNMM method, if we let each molecule span an increasing larger set of MRS points with increasing values of τ, one should eventually reach a value of τ where the circles all begin to form an intersection region. Thus, one should be able to "dial-in" the value of τ that leads to the tightest intersection region. In practice, the author has never been able to get this to work to his satisfaction, but it is suggestive of future directions. Using a traditional or arbitrary value of τ is a key source of subjectivity in the process of pharmacophore discovery; it is important to find a way to determine the proper value of τ, to wring out every element of subjectivity from this process, to ensure that the results parallel the physical reality and hence are useful. The proper value of τ may be dataset-dependent; a method like that shown in Fig. (**8**) would capture that.

2) *How should features be identified?* The idea of relying on a standard library of features A,D,N,P,H,R seems antiquated, and likely exacerbates the problem of multiple candidate pharmacophores arising from a dataset. It is an embarrassment to this field that we continue to rely on such a primitive method. Some creativity is needed here to provide a better solution, ideally to identify features in a way that is dataset-dependent.

3) *What is the best way to handle the multiplicity of pharmacophores consistent with a dataset?* This author's experience suggests that the principle of selectivity is indeed a useful guide to picking the pharmacophore most useful prospectively, but it is not foolproof. An alternative approach is to embrace all these pharmacophores, and to create in effect a 3D similarity measure based on the presence/absence of different pharmacophores [34,35]; it is unclear what the physical significance of such models is.

4) *How should shape constraints be handled? How can other field quantities be included (e.g. electrostatics)?* The author's experience suggests that the surface of minimum volume, as implemented in the 'shrink-wrap' procedure, is useful for steric constraints. It remains an open problem to "color the shrink-wrap surface", i.e. identify regions of the inferred binding surface that complement regions on the ligand of high- and low-charge density. In light of the Kubinyi Paradox, this author does not take it as a given, that the appropriate solution to this is partial-least-squares (PLS).

5) *Is the iterative model-building process convergent?* As shown in Fig. (**9**), the process of constructing a pharmacophore model, and using it prospectively, forms a loop. The question that has been posed by D. Agrafiotis is: is this process convergent? Generally, it was seen with Hypothesis Generation that every addition of a new molecule led to a totally different pharmacophore model; by contrast, in DANTE each additional molecule tends to reinforce the existing model. This clearly is an interesting question for future study.

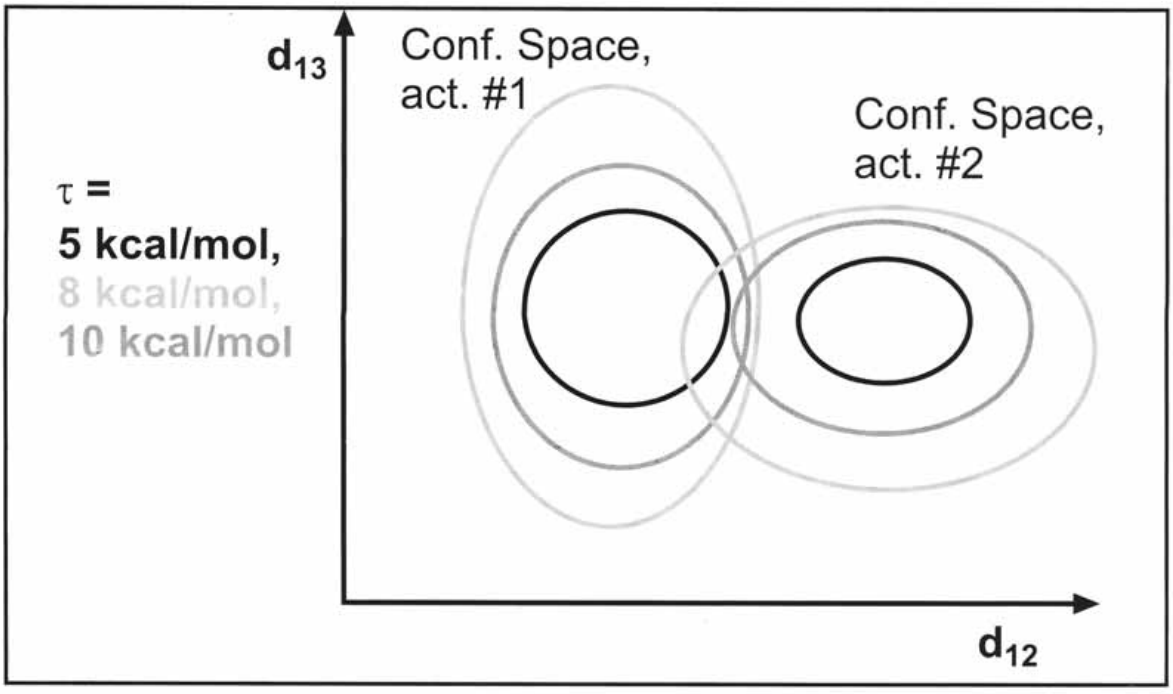

Fig. (8). Theoretical scheme for "dialing-in" the value of tau that leads to a pharmacophore. At tau=5 kcal/mol, there is no overlap, but at 8 kcal/mol, a very small overlap region may be found. At 10 kcal/mol, this overlap region grows.

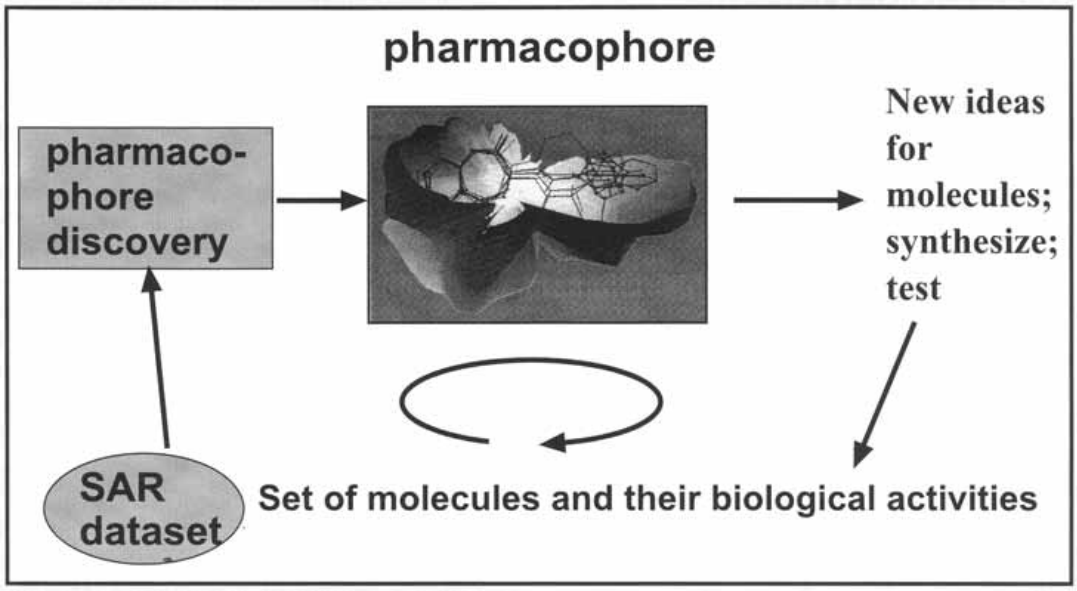

Fig. (9). Design using a pharmacophore is fundamentally iterative.

6) *Do the processes described above lead to models which are superior to or 3D-QSAR models like CoMFA [13,14], or those derived from ensembles of pharmacophores [34,35], or those derived from field-based overlay methods [36,37]?* 3D-QSAR models require a pharmacophore prior to calculating the fields, so these methods aren't really comparable – 3D-QSAR refines the results that emerge from pharmacophore discovery. Compared with the other classes of 3D model-building, prospective tests of these methods will eventually clarify the answer to this question, though a common set of computational controls applied to all methods with the same input data would be instructive.

CONCLUSION

Lo duca e io per quel cammino ascoso	The Guide and I into that hidden road
intrammo a ritornar nel chiaro mondo;	Now entered, to return to the bright world;
e sanza cura aver d'alcun riposo,	And without care of having any rest
salimmo sù, el primo e io secondo,	We mounted up, he first and I the second,
tanto ch'i' vidi de le cose belle	Till I beheld through a round aperture
che porta 'l ciel, per un pertugio tondo.	Some of the beauteous things that Heaven doth bear;
E quindi uscimmo a riveder le stelle.	Thence we came forth to rebehold the stars.

Many lessons have been learned in over a decade of use of modeling tools for pharmacophore discovery, but probably the most important one is: despite their immaturity, these tools are useful, and have their place at the side of any scientist engaged in drug design in the absence of direct crystallographic data on the biological target.

ACKNOWLEDGEMENTS

The author has benefited from interactions with many colleagues over the years on these matters. My former Upjohn colleague J.R. Blinn deserves special mention, as he was the the first user of DANTE, and the source of many ideas, including the suggestion of the name DANTE. The photograph of the author is © 2002, Laura H. van Drie, used with permission.

REFERENCES

References 38-40 are related articles recently published in Current Pharmaceutical Design.

The quotations beginning each section are from the Divine Comedy of Dante Alighieri. The English translations are taken from the Longfellow translation. The full text, in Italian and in translation, are now available on the World Wide Web, e.g. www.greatdante.net.

[1] Marshall GR, Barry CD, Bosshard HE, Dammkoehler RA, Dunn DA. The Conformational Parameter in Drug Design: The Active Analog Approach. In: Computer-Assisted Drug Design, Olson EC, Christofersen RE. Eds. Washington: Amer Chem Soc 1979; pp 205-226.

[2] Van Drie JH,Weininger D Martin YC. 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 Comp Aided Mol Des 1989; 3: 225-251.

[3] Van Drie JH. 3D Database Searching in Drug Discovery. Netsci 1995; www.netsci.org/Science/Cheminform/feature06.html.

[4] Van Drie JH. An inequality for 3D database searching and its use in evaluating the treatment of conformational flexibility. J Comp Aided Mol Des 1996; 10: 623-631.

[5] Van Drie JH. Strategies for the determination of pharmacophoric 3D database queries. J Comp Aided Mol Des 1997; 11: 39-52.

[6] Van Drie JH. "Shrink-Wrap" Surfaces: A New Method for Incorporating Shape into Pharmacophoric 3D Database Searching. J Chem Info Comp Sci 1997; 37: 38-42.

[7] Van Drie JH, Nugent RA. Addressing the challenges of combinatorial chemistry: 3D databases, pharmacophore recognition and beyond. SAR and QSAR in Env Res 1998; 9: 1-21.

[8] H-D. Höltje H-D. Pharmacophore identification and receptor mapping. In The Practice of Medicinal Chemistry, Wermuth CG, Ed, London: Academic Press 1996; pp. 437-458.

[9] Bures M. Recent Techniques and Applications in Pharmacophore Mapping. In Practical Applications of Computer-Aided Drug Design, Charifson P, Ed, New York: Dekker 1997; pp 39-72.

[10] Van Drie JH. Pharmacophore discovery – a critical review. In Computational Medicinal Chemistry and Drug Discovery, Tollenaere JP, Bultinck P, de Winter H. and Langenaeker W, Eds, New York: Dekker, in press.

[11] Martin YC. 3D database searching in drug design. J Med Chem 1992; 35: 2145-54.

[12] Shoichet BK, Stroud RM, Santi DV, Kuntz ID, Perry KM. Structure-based discovery of inhibitors of thymidylate synthase. Science 1993; 259: 1445-50.

[13] Kubinyi H. 3D-QSAR in Drug Design Volume 1: Theory Methods and Applications, Amsterdam: Kluwer Academic Publishers 1993.

[14] Cramer RD, Patterson DE, Bunce JD. Comparative Molecular Field Analysis (CoMFA). Effect of Shape on Binding of Steroids to Carrier Proteins. J Amer Chem Soc 1988; 1110: 5959-67.

[15] Hilpert K, Ackermann J, Banner DW, Gast A, Gubernator K, Hadvary P, et al. Design and synthesis of potent and highly selective thrombin inhibitors. J Med Chem 1994; 37: 3889-901.

[16] Smellie A, Teig SL, Towbin P. Poling: Promoting conformational variation. J Comp Chem 1995; 16, 171-87.

[17] Chang G, Guida W, Still W. An internal coordinate monte-carlo method for searching conformational space. J Amer Chem Soc 1989; 111: 4379-86. See also Macromodel 8.0 reference manual, available from Schrodinger, Inc. http://www.schrodinger.com/docs/mm 8.0/pdf/MM_minta_ref_manual/MM_Minta.pdf.

[18] Halgren TA. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization and Performance of MMFF94. J Comput Chem 1996; 17: 490-519, et seq.

[19] Gundertofte K, Liljefors T, Norrby PO, Pettersson I. A Comparison of Conformational Energies Calculated by Several Molecular Mechanics Methods. J Comput Chem 1996; 17: 429-49.

[20] Omega is software based on unpublished methodology, available from Open Eye Scientific Software, www. eyesopen.com.

[21] ONFORT is software based on unpublished methodology, available from R. S. Pearlman, Univ. of Texas, Austin, TX.

[22] Cavalli A, Poluzzi E, De Ponti F, Recanatini M. Toward a pharmacophore for drugs inducing the long QT syndrome: insights from a CoMFA study of HERG K(+) channel blockers. J Med Chem 2002; 45: 3844-53. While some aspects of their computational protocol may be disputed, these authors are to be commended for the level of detail they provide about their protocol.

[23] Jain AN, Koile K, Chapman D. Compass: predicting biological activities from molecular surface properties. Performance comparisons on a steroid benchmark. J Med Chem 1994; 37: 2315-27.

[24] Mayer D, Naylor CB, Motoc I, Marshall GR. A unique geometry of the active site of angiotensin-converting enzyme consistent with structure-activity studies. J Comp Aided Mol Des 1987; 1: 3-16.

[25] This database is available from the National Cancer Institute, and represents a collection of drug-like molecules submitted to the NCI for carcinogenicity evaluation. http://cactus.nci.nih.gov/ncidb2/download.html

[26] For example TenBrink RE. Oxazinocarbazoles for the treatment of CNS diseases. US Patent 6,448,397, Sept. 10, 2002; Nilsson, B. et al. Certain aryl-aliphatic and heteroaryl-aliphatic piperazinyl pyrazines and their use in the treatment of serotonin-related diseases. US Patent 6,465,467, Oct. 15, 2002.

[27] Kabsch W. A Solution for the Best Rotation to Relate Two Sets of Vectors. Acta Cryst. 1976; A32, 922-3; Kabsch, W. A discussion of the solution for the best rotation to relate two sets of vectors. Acta Cryst 1978; A34, 827.

[28] Monte AP, Waldman SR, Marona-Lewicka D, Wainscott DB, Nelson DL, Sanders-Bush E, et al. Dihydrobenzofuran analogues of hallucinogens. 4. Mescaline derivatives. J Med Chem 1997; 40: 2997-08; Bromidge SM, Dabbs S, Davies DT, Duckworth DM, Forbes IT, Ham P, et al. Novel and selective 5-HT2C/2B receptor antagonists as potential anxiolytic agents: synthesis, quantitative structure-activity relationships, and molecular modeling of substituted 1-(3-pyridylcarbamoyl)indolines. J Med Chem 1998; 41: 1598-1612.

[29] Brickner SJ, Hutchinson DK, Barbachyn MR, Manninen PR, Ulanowicz DA, Garmon SA, et al. Synthesis and antibacterial activity of U-100592 and U-100766, two oxazolidinone antibacterial agents for the potential treatment of multidrug-resistant gram-positive bacterial infections. J Med Chem 1996; 39: 673-9.

[30] The allusion is by V. E. Groppi to the "gonzo journalism" of Thompson HS, e.g. Fear & Loathing in Las Vegas, New York: Random House, 1973.

[31] McGovern SL, Caselli E, Grigorieff N, Shoichet BK. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. J Med Chem 2002; 45: 1712-22.

[32] Kubinyi H, Hamprecht FA, Mietzner T. Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices. J Med Chem 1998; 41: 2553-64.

[33] Golbraikh T. and Tropsha A. Beware of q2! J Mol Graph Model 2002; 20: 269-76.

[34] Mason JS, Good AC, Martin EJ. 3-D pharmacophores in drug discovery. Curr Pharm Des 2001; 7: 567-97.

[35] Bradley EK. Beroza P, Penzotti JE, Grootenhuis PD, Spellmeyer DC, Miller JL. A rapid computational method for lead evolution: description and application to alpha(1)-adrenergic antagonists. J Med Chem 2000; 43: 2770-4.

[36] Mestres J, Rohrer DC, Maggiora GM. A molecular field-based similarity approach to pharmacophoric pattern recognition. J Mol Graph Model 1997; 15: 114; 15, 103-6.

[37] Jones G, Willett P, Glen RC. A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. J Comp Aided Mol Des 1995; 9: 532-49.

[38] Langer T, Hoffmann RD. Virtual screening: an effective tool for lead structure discovery? Curr Pharm Des 2001; 7(7): 509-27.

[39] Li Y, Harte WE. A review of molecular modeling approaches to pharmacophore models and structure-activity relationships of ion channel modulators in CNS. Curr Pharm Des 2002; 8(2): 99-110.

[40] Wong MG, Tehan BG, Lloyd EJ. Molecular mapping in the CNS. Curr Pharm Des 2002; 8(17): 1547-70.