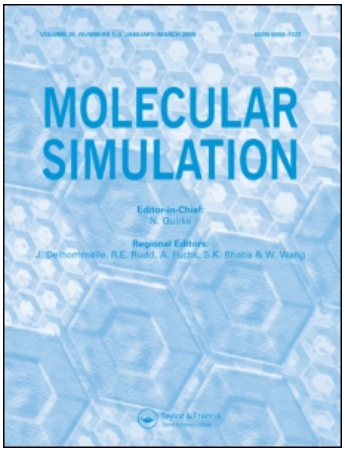


Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Computational Chemistry in Lead Identification, Library Design and Lead Optimisation

Andrew R. Leach[a]; Darren V. S. Green[a]
[a] Glaxo Wellcome Medicines Research Centre, Stevenage Hertfordshire, UK

To cite this Article Leach, Andrew R. and Green, Darren V. S.(2001) 'Computational Chemistry in Lead Identification, Library Design and Lead Optimisation', Molecular Simulation, 26: 1, 33 — 49
To link to this Article: DOI: 10.1080/08927020108024199
URL: http://dx.doi.org/10.1080/08927020108024199

Molecular Simulation, 2001, Vol. 26, pp. 33–49
Reprints available directly from the publisher
Photocopying permitted by license only

COMPUTATIONAL CHEMISTRY IN LEAD IDENTIFICATION, LIBRARY DESIGN AND LEAD OPTIMISATION

ANDREW R. LEACH* and DARREN V. S. GREEN

Glaxo Wellcome Medicines Research Centre, Gunnels Wood Road, Stevenage Hertfordshire SG1 2NY UK

(Received March 1999; accepted May 1999)

We describe a variety of the computational techniques which we use in the drug discovery and design process. Some of these computational methods are designed to support the new experimental technologies of high-throughput screening and combinatorial chemistry. We also consider some new approaches to problems of long-standing interest such as protein-ligand docking and the prediction of free energies of binding.

Keywords: Drug discovery; High-throughput screening; Combinatorial chemistry; Protein-ligand docking; Binding free energies

INTRODUCTION

A quick scan through recent issues of this journal reveals that many papers are concerned with molecular dynamics or Monte Carlo simulations of molecular systems. This comes as no surprise to us; in common with many practitioners of computational chemistry we tend to associate the word "simulation" specifically with these two techniques. However, in our day-to-day work in the research division of a large pharmaceutical company these two simulation techniques represent just two of the many computational methods available to us. In this short review we will try and give a flavour of some of the problems we are currently trying to tackle and the methods that are being used.

*Corresponding author.

Our research process can be conveniently divided into three phases. The first phase is often termed "hit discovery". A "hit" is a molecule which shows some reproducible activity in a biological assay. The second phase involves a careful examination of these hits (and possibly some exploratory chemical synthesis) to decide which can be considered as "leads". A lead molecule is one for which some relationship between the structure and the activity can be demonstrated and which shows selectivity for the target. It must also have the potential for being modified during the third phase, lead optimisation, during which the various properties of the molecule are modified using chemical synthesis with the aim of producing a compound for development. Of course there remain many hurdles still to be overcome before the medicine reaches the market, but we shall not consider those aspects here.

The techniques of high-throughput screening (HTS) and combinatorial chemistry have had a major impact on our research process. *In extremis* high-throughput screening enables one to screen every available sample – which could include hundreds of thousands if not millions of compounds. However, a more "rational" approach is often available, in which we use our knowledge of the target to identify that subset of molecules which are most likely to show the desired activity and which are most likely to pass successfully through the "hits-to-leads" hurdle to give candidates for lead optimisation. One of the most powerful ways to generate such focussed sets of compounds involves the use of 3D database searching.

3D DATABASE SEARCHING

Many chemists are familiar with chemical databases such as CAS-online and Beilstein. These databases can be searched in a variety of ways to identify molecules of interest. A particularly powerful search technique is the substructure search, which enables one to extract all molecules in the database that contain a particular structural template. For example, one could identify all molecules that contain the morphine skeleton using such an approach. Substructure searching makes use of the way in which the atoms and bonds in the molecule are connected (the *molecular graph*). A 3D database extends this to store information about the conformational properties of molecules. This conformational data may come from experimental sources such as X-ray crystallography. However, the X-ray structures of most compounds are not known and so computational methods are typically used to generate theoretical 3-dimensional models which are used as the starting point for an exploration of the conformational space [1].

A number of commercial 3D database systems are now available which can automatically perform these steps and store the conformational information in a manner that enables the database to be searched rapidly. Several review articles which describe the use of 3D searching in drug design are available [2].

In contrast to the substructure search that is often used for a "2D" database search, a 3D database is typically searched using a 3D pharmacophore query. The term pharmacophore refers to a set of features that is common to a series of active molecules. Typical features are hydrogen bond donors and acceptors, positively and negatively charged groups and hydrophobic regions. A 3D pharmacophore also includes information about the spatial relationships between such features. These spatial relationships are frequently expressed as distances or distance ranges as in the simple 3D pharmacophore for antihistamine activity shown in Figure 1. More complex 3D pharmacophore queries contain features such as exclusion volumes, angles between planes and the centroids of aromatic rings as illustrated in Figure 2. Typical 3D database systems are able to query the database to identify molecules that not only possess the features in the pharmacophore query but which can position these features in the appropriate spatial orientation in a low-energy conformation.

Where do 3D pharmacophores come from? If the protein structure of the biological target is available then it may be possible to deduce an appropriate pharmacophore (or set of pharmacophores) from an examination of the relevant binding site. This process can be assisted by knowledge of how inhibitors or substrates bind and also by computational techniques which probe the binding site to determine locations where particularly favourable binding would be expected. A particularly useful method in this regard is the GRID program [3] which calculates the energy of interaction between a

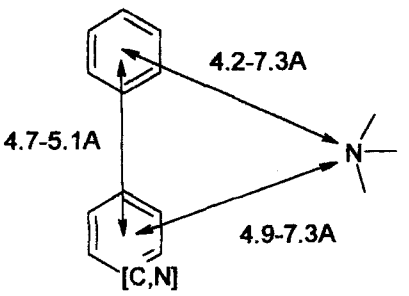


FIGURE 1 A simple antihistamine 3D pharmacophore.

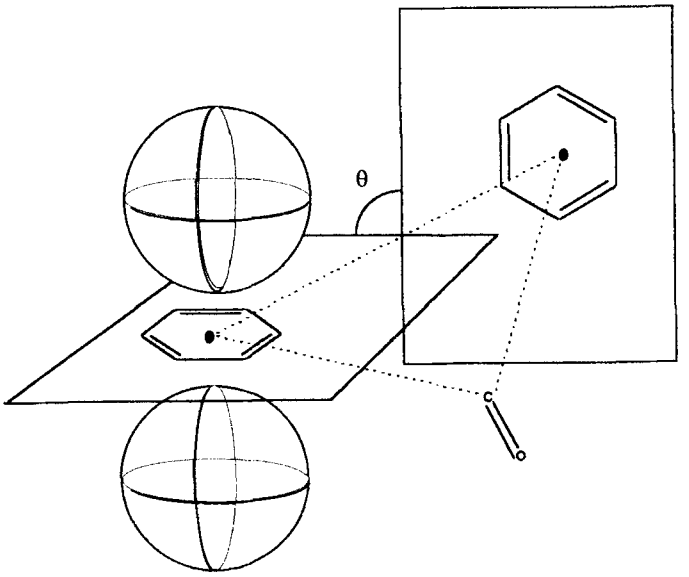


FIGURE 2 A 3D pharmacophore may include features such as the centroids of aromatic rings, best-fit planes and exclusion volumes.

series of probe functional groups and the protein over a regular grid. This interaction map can then be contoured and regions of lowest energy identified, from which the 3D pharmacophore(s) can be deduced.

Frequently the structure of the biological target is not available. The objective of a procedure called pharmacophore mapping is to derive one or more 3D pharmacophores from a series of active and (sometimes) inactive molecules. Simply stated, all of the active molecules should be able to fit the pharmacophore in at least one reasonable (*i.e.*, low-energy) conformation whilst none of the inactive molecules should be able to achieve this. This is illustrated in Figure 3 which shows a 3D pharmacophore derived for activity at the 5HT-3 receptor. Of the two molecules shown only one is able to match the pharmacophore. The other, inactive molecule, is unable to do so. A key assumption underlying pharmacophore mapping is that the active molecules all have a common binding mode – it is important to remember that this is not always the case.

We routinely use 3D database searching to identify molecules for focussed screening, in which a relatively small number of molecules (typically a few hundred or a few thousand) is identified from our corporate collection for testing in the biological assay. The alternative approach is to test every

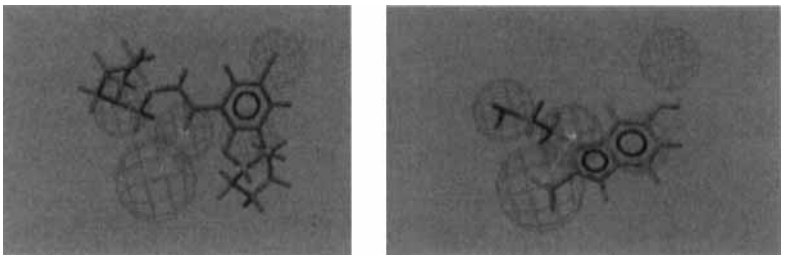


FIGURE 3 A 3D pharmacophore for activity at the 5HT-3 receptor derived using the Catalyst program [14]. Both active and inactive compounds were used in the generation of the pharmacophore. The red sphere corresponds to the location of a basic group; the green spheres correspond to a hydrogen bond acceptor and the blue spheres are regions where a hydrophobic region should be placed. On the left is shown an active molecule superimposed on the pharmacophore. The molecule on the right is inactive and cannot fit all features of the pharmacophore in a low-energy conformation. (See Color Plate III).

available compound using advanced high-throughput robotic screening techniques. The relative merits of the two approaches have been a source of considerable debate; focussed screening has a higher chance of finding molecules of interest (*i.e.*, finding hits which would also be considered leads) but there remains the possibility that a molecule could be "missed" using this approach. Another aspect to the debate that should not be forgotten is that there is always a cost associated with screening compounds. The cost of testing a compound varies from assay to assay but even the cheapest screens require a considerable financial outlay when hundreds of thousands of compounds are to be tested. Nevertheless, both techniques will have a role in the foreseeable future, with 3D searching screening being used to prepare a list of "priority" compounds which should be screened first. One way in which we can refine this strategy is through retrospective analysis of high-throughput screening data. For example, in a recent screen a 3D pharmacophore containing 5 features was derived from a series of active molecules. This was used to select 2000 compounds which when screened gave 32 hits (1.6%). A high-throughput screen was subsequently run in which over 500,000 molecules were tested. Of these, 517 were deemed sufficiently active to have an inhibition constant measured (a hit rate of 0.1%). What is particularly interesting is that of these 517 molecules 237 were able to match the 5-point 3D pharmacophore. Moreover, 241 of the remainder were able to match at least one of the five 4-point pharmacophores that could be generated by sequentially removing one of the features. Only 39 compounds were unable to match either the original pharmacophore or one of the derived 4-point pharmacophores.

CHEMICAL LIBRARIES

In the short period since it was widely introduced in the pharmaceutical industry combinatorial chemistry has provided many challenging problems for the computational chemistry community. The early attraction of combinatorial chemistry was the potential to make many more compounds than was the case with conventional synthesis. To that end much emphasis was placed on the problem of designing "diverse" combinatorial libraries that were intended to cover a wide range of functionality. This of course raises the question: what do we mean by "chemical diversity", and how do we measure or calculate it? There is not enough space in this article to explore these issues; suffice to say that the early work made heavy use of properties derived from the molecular graph rather than any conformational or other 3D property. Some readers may be surprised to learn that many properties can be calculated solely from the computer equivalent of a chemical drawing on a piece of paper, often to an accuracy that rivals much more sophisticated approaches. A good example is the ClogP program [4] which is widely used to calculate octanol/water partition coefficients. Of particular significance so far as library design and compound selection is concerned is the fact that calculations based on the molecular graph can be performed extremely rapidly and so can be applied to collections of hundreds of thousands if not millions of molecules. It is often impractical to perform more sophisticated calculations (*e.g.*, involving conformationally-dependent or electronic properties) on this scale.

The earliest combinatorial libraries showed that the technology was certainly able to synthesise large numbers of compounds but it was often found that such libraries gave relatively few "hits" in biological assays. Some of the reasons for this lack of success became apparent when a variety of simple properties were calculated for these libraries and for sets of compounds known to possess biological activity. For example, the early libraries frequently contained very flexible molecules. When a flexible molecule binds to a protein to form a bimolecular complex it will have a larger entropic penalty than a less flexible molecules (all other factors being equal) due to the need to freeze out the conformational degrees of freedom. Based on these observations it was considered more desirable to synthesise libraries which had properties closer to sets of known active molecules. In other words, we would like to make our libraries more "drug-like". In Figure 4 we show the distribution of rotatable bonds for a number of combinatorial libraries as well as some collections of "drug-like" compounds. We also show in Figure 4 the results of a more complex analysis in which the number of 3D

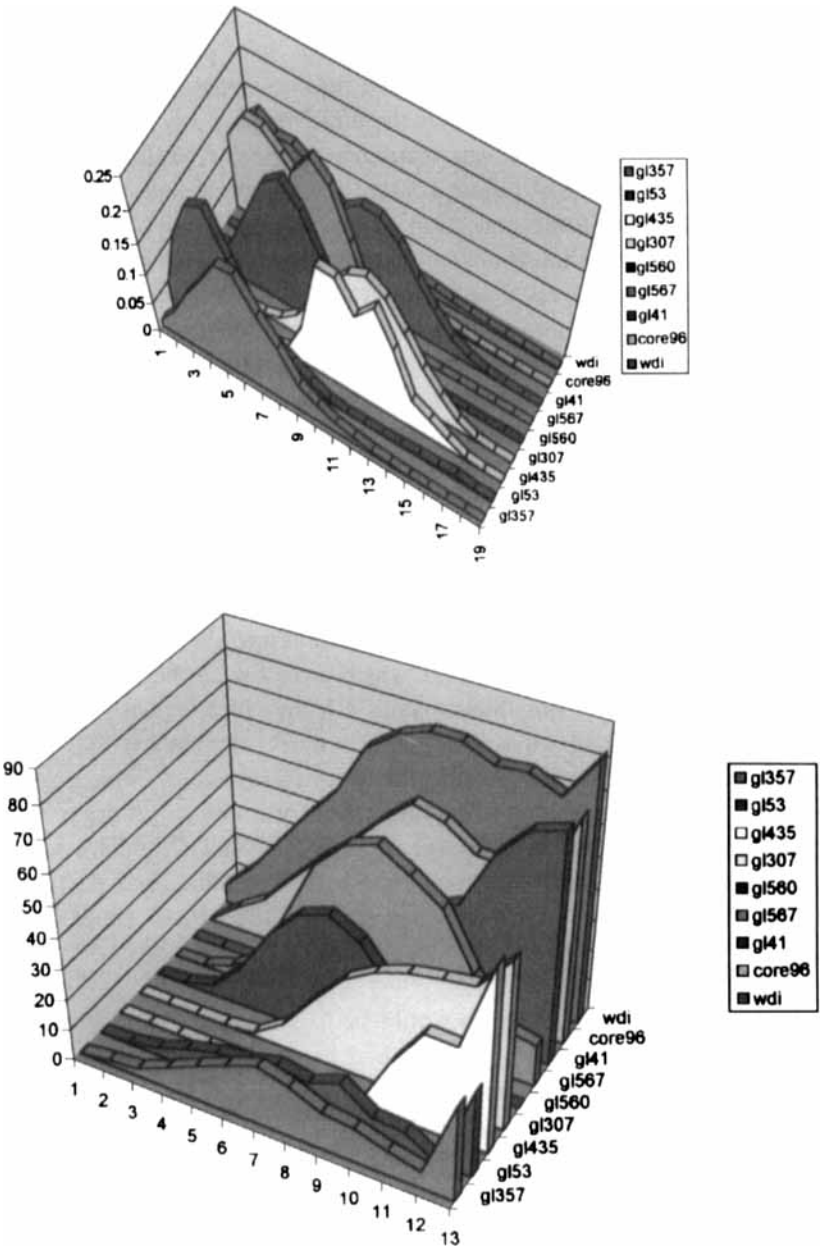


FIGURE 4 (Top). Distribution of rotatable bonds for a series of combinatorial libraries, the World Drug Index (WDI) and a representative sample of the Glaxo Wellcome compound collection (core96). (bottom) Graph showing how the number of 3D pharmacophores varies with the number of rotatable bonds for these sets of compounds. (See Color Plate IV).

pharmacophores displayed by a molecule varies with its flexibility. To calculate this data a conformational analysis was performed for every molecule in the various sets. During the conformational analysis we record how many 3D pharmacophores that the molecule displays. In this case the type of 3D pharmacophores was restricted to those containing just three features. The key feature of Figure 4 is that drug-like molecules such as those in our in-house corporate collection or the World Drug Index database (which contains more than 50,000 molecules known to have some biological activity) are able to express a significant number of 3D pharmacophores yet are not overly flexible. By contrast, many of the combinatorial libraries either contain very few 3D pharmacophores (*i.e.*, do not contain much functionality) or are only able to do so at the expense of very flexible molecules.

Another important aspect of the design of combinatorial libraries is the so-called subset selection problem. In a "true" combinatorial library of the form $A \times B \times C$ every molecule from the set of reagents A reacts with every molecule from B and every molecule from C to generate $n_A \times n_B \times n_C$ product structures where n_A, n_B, n_C are the numbers of reagent molecules A, B and C. Typically there will be many more possible reagents A, B, C available to us than we can actually incorporate into the library and so we need to select that subset of reagents which gives rise to the "best" library. Suppose the number of possible reagents A from which we wish to select is N_A *etc.* The number of ways of selecting n objects from N is ${}^{N}C_{n}$ and so the number of different combinatorial libraries of size $n_A \times n_B \times n_C$ that could be made for this three-component library is ${}^{N_A}C_{n_A} \times {}^{N_B}C_{n_B} \times {}^{N_C}C_{n_C}$. If we have 100 possible reagents for each of A, B and C and we wish to make a $10 \times 10 \times 10$ library then the number of possible libraries that we could make is approximately 10^{40}. Identifying the "best" single library from this extremely large number of possible libraries is clearly a "hard" problem.

One approach to this problem would be to select the "best" monomers from each of the monomer sets and then to combine these together, assuming that the "best" monomers give rise to the "best" library when they are combined. This is computationally a much less demanding approach. However, it has been shown [5] that whilst such an approach is better than selecting building blocks at random, it is generally less effective than selecting the reagents using the properties of the product molecules.

We have used genetic algorithms to tackle the library subset problem. Genetic algorithms are a powerful set of optimisation procedures which are able to generate near-optimal solutions to difficult problems in a reasonable

amount of time. There are many different flavours of genetic algorithm; in its simplest form such an algorithm uses the following steps:

1. Generate an initial population of possible solutions
2. Score each member of the population using a fitness function
3. Generate a new population using genetic operators
4. Score each new member of the population
5. Go to Step 3 until finished.

There are many variants on this simple scheme which are covered in detail elsewhere [6]. There are three key features of a genetic algorithm. The first of these is that a population of possible solutions is maintained during the procedure. Secondly, each of the members of this population is encoded as a *chromosome*, typically represented within the computer as a bit-string. It is necessary to have an encoding/decoding mechanism by which the chromosomal bit-string can be generated from a potential solution and *vice versa*. The third feature is that members of the new population (*i.e.*, the offspring) are generated from the current members of the population by applying a set of genetic operators. Many such operators have been developed but the two most widely used are mutation and crossover. In mutation a random change is made to the chromosome (*i.e.*, a random bit is flipped). In crossover two chromosomes are taken and a point is selected at random. The two halves of the two chromosomes are then switched to give two offspring (Fig. 5).

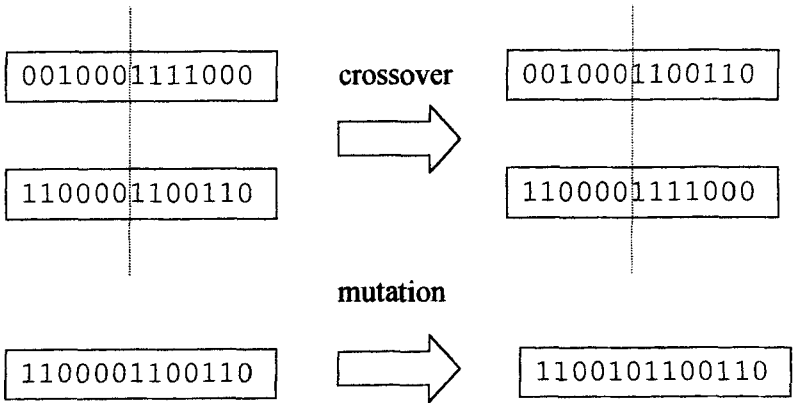


FIGURE 5 The two major operators used in genetic algorithms. The first is crossover, in which two new chromosomes are generated from two parents by first selecting (at random) a place to divide the bit-strings which are then combined as shown. In the mutation operator a random bit is changed.

Genetic algorithms are able to produce optimal or near-optimal library subsets in test cases where it is possible to determine exactly the optimal solution. A particularly attractive feature of the genetic algorithm approach is that a wide variety of different fitness functions can be used. As indicated above, the earliest libraries were designed to be as diverse as possible. More recently, we have incorporated a much wider range of properties into our library design. For example, we might wish to make a library that maximises the number of molecules that can fit a particular 3D pharmacophore whilst simultaneously matching a desired profile (*e.g.*, the distribution of rotatable bonds in the world drug index) and also being as diverse as possible. To incorporate 3D information into library design typically requires the virtual library to be enumerated; this corresponds to generating the structures of all

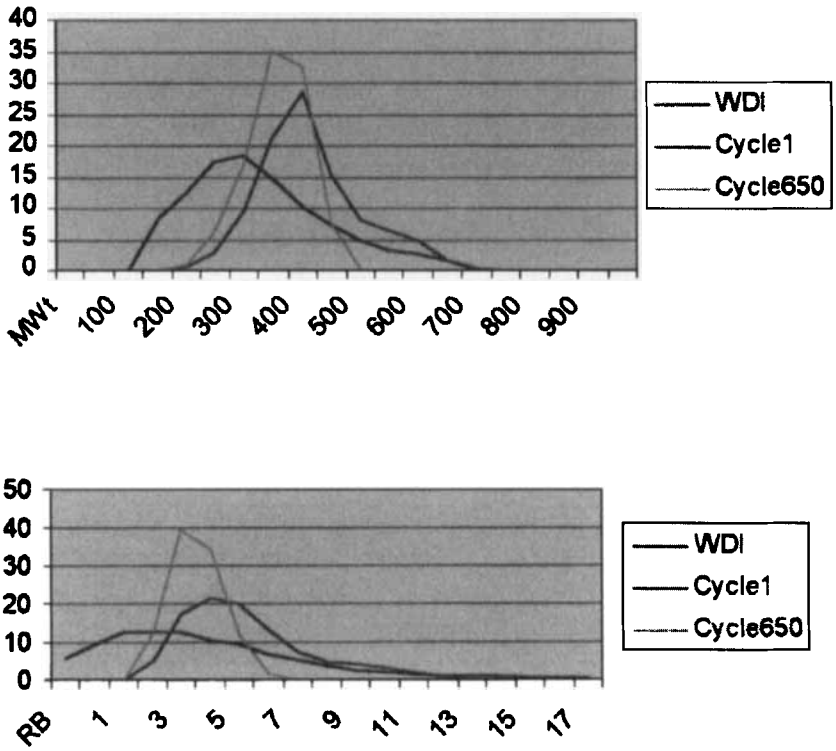
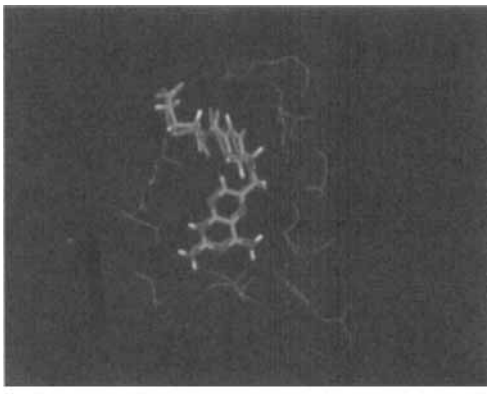


FIGURE 6 Graphs showing how the distribution of molecular weight and rotatable bonds varies during the optimisation of a library using a genetic algorithm. The aim is to produce a profile more closely matching that of the World Drug Index (labelled WDI). Initially (Cycle 1) both distributions have significant tails but by cycle 650 the distributions more closely match the WDI distribution, particularly in these tail regions. (See Color Plate V).

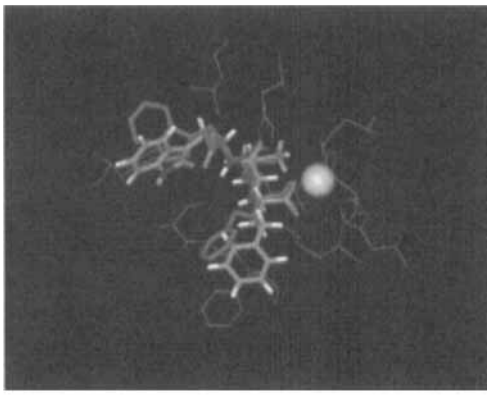
possible product molecules that could be made (*i.e.*, $N_A \times N_B \times N_C$). These are then converted into a 3D database which is searched using the 3D pharmacophore. We then aim to identify a set of reagents n_A, n_B, n_C which when combined give a combinatorial library with a maximal number of structures that hit the 3D pharmacophore whilst at the same time satisfying the profile and diversity requirements. We show in Figure 6 how the molecular weight and rotatable bond distributions change from the start of the optimisation to the final design.

PROTEIN-LIGAND DOCKING

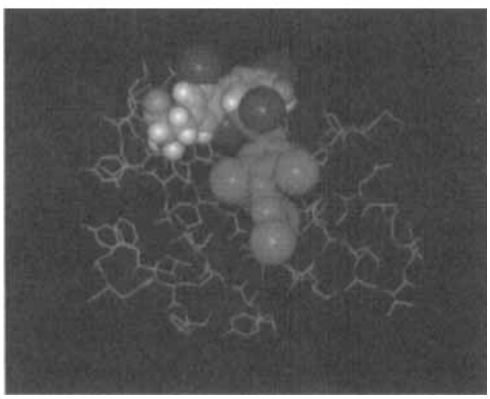
During the lead optimisation process one or more lead series is explored in order to try and identify candidates for the next phase of the process (*i.e.*, development). This involves the simultaneous optimisation of a large number of parameters ranging from the intrinsic activity (*i.e.*, how tightly does the molecule bind to the target macromolecule) to its pharmacokinetic properties (how does the molecule behave *in vivo*). One of the most useful pieces of information during the lead optimisation phase is knowledge about how the molecules bind to the target macromolecule. The most powerful way to determine such information involves the use of X-ray crystallography to determine the structure of the protein-ligand complex. However, it is not always possible to obtain X-ray data and so a theoretical approach is required. A large number of algorithms have been described over the years which attempt to predict the structure(s) of the intermolecular complex formed between a protein and an inhibitor molecule (the protein-ligand docking problem) [7]. This has proved a particularly attractive problem for a number of reasons, not least because it has yet to be completely solved. There are two main reasons why it has attracted so much attention. First, a very large number of degrees of freedom are involved. In addition to the six degrees of translational and rotational freedom of one body relative to another there are the conformational degrees of freedom of the ligand and also the conformational degrees of freedom of the protein. Secondly, one typically needs some means to discriminate between the many thousands (if not millions) of potential binding modes that a typical algorithm can generate. The most natural scoring function to use is of course the free energy of binding, but as we shall see later it can be very difficult and time-consuming to determine accurately such free energies. It is more common for a docking program to use a function that can be calculated more rapidly.



(a)



(b)



(c)

FIGURE 7 Results of three GOLD docking experiments. (a) shows methotrexate bound to dihydrofolate reductase; (b) shows N-carboxymethyl dipeptide inhibitor bound to thermolysin; (c) shows thyroxine bound to transthyretin. (See Color Plate VI).

In a LINK-sponsored project involving Glaxo Wellcome, the University of Sheffield and the Cambridge Crystallographic Data Centre we have tackled the protein-ligand docking problem using a genetic algorithm approach. In this case the chromosome codes for the conformation and orientation of the ligand within the binding site. A novel scoring function is used to assess each member of the population and the usual genetic operators are used to derive new members of the population. More information on this GOLD program can be found elsewhere [8]; one key feature of the method is that information derived from the Cambridge Structural Database of small-molecule crystal structures is used to direct the conformational analysis of the ligand to known low-energy regions.

The method has been evaluated on a large number of experimental structures (more than 130) obtained from the protein databank [9]. These structures were chosen to cover a wide range of types of system; not just those that one might anticipate were most suitable for the algorithm. In more than 70% of these test cases the top-ranked structure predicted by the program was the same as the experimentally determined structure. Two such structures predicted using GOLD are shown in Figures 7a and 7b superimposed on the experimentally determined structure. In Figure 7c we show an example of a system where the top-ranked solution predicted by GOLD was not in fact the same as the experimental result. In some cases it is possible to understand the reasons for these failures; usually they are due to the internal scoring scheme used to rank one orientation relative to another, rather than a lack of coverage in the exploration of the orientational or conformation space.

PREDICTING FREE ENERGIES OF BINDING

To many, the "holy grail" of computational drug design is a method that can accurately calculate the free energy of binding of a ligand molecule to a protein. In practice this remains an unsolved problem though some significant progress has been made, partly due to the improvements in computational power and partly due to methodological advances. The most appealing approach to this problem involves the determination of relative free energies of binding using the free energy perturbation method [10]. However, free energy perturbation usually requires significant computational resources and it is also limited to relatively small structural changes in the ligand.

Åqvist described an alternative approach to the problem in 1994 [11]. Termed the linear interaction energy (LIE) method this requires just two

simulations, one of the inhibitor in water and the other of the inhibitor bound to the solvated protein. In the LIE method the following expression for the free energy is applied:

$$\Delta G = \beta\langle\Delta U_{\text{elec}}\rangle + \alpha\langle\Delta U_{\text{vdw}}\rangle$$

Here, ΔU_{elec} and ΔU_{vdw} are the differences in electrostatic and van der Waals energies for the inhibitor-solvent and inhibitor-protein systems. In his original work Åqvist used a value of 0.5 for β (based on the Born model for ion solvation) and the parameter α was optimised to give the best agreement between experiment and theory for a series of endothiapepsin ligands ($\alpha = 0.161$). Åqvist subsequently showed that the values of these parameters could be transferred to other systems. However, this has not always been the experience of other workers; the proposed value of α in particular does not lead to accurate predictions. Other equations have also been described; Jorgensen has described an alternative form which he found to be more predictive:

$$\Delta G = \beta\langle\Delta U_{\text{elec}}\rangle + \alpha\langle\Delta U_{\text{vdw}}\rangle + \gamma\langle\Delta\text{SASA}\rangle$$

Here, ΔSASA is the change in solvent accessible surface area.

In our work we have considered a series of ligands bound to the enzyme neuraminidase. Neuraminidase is a target for certain anti-influenza therapies now in clinical trials. The inhibitors considered (Fig. 8) show a wide range of inhibition constants [12] yet some at least have relatively small structural changes, making comparison with free energy perturbation possible. The protein is known to undergo conformational changes on binding of some inhibitors and the number of tightly bound water molecules can change. For these and other reasons this is a challenging system with which to investigate the LIE approach.

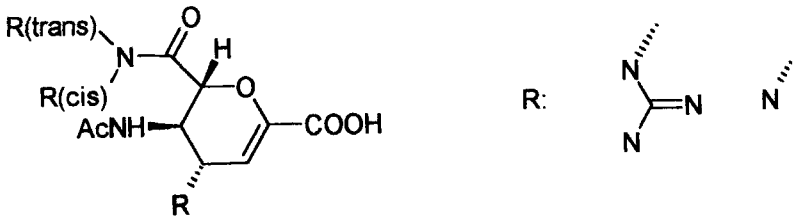


FIGURE 8 Generic structure of Neuraminidase inhibitors investigated using the Linear Interaction Energy method. The inhibition constants vary over 5 orders of magnitude depending on the substituents *R* (*trans*) and *R* (*cis*).

Monte Carlo simulation methods were used. Details of the calculations will be reported elsewhere [13]. A key result was that none of the sets of parameters described by other workers were found to give satisfactory agreement with experiment when applied to this data set. However, values of α and β could be found which did give a reasonable fit ($q^2 = 0.5$; cross-validated RMS = 1.7 kcal/mol). A more extensive statistical analysis was performed in which a wider set of parameters was considered, including terms in the intramolecular energies of ligand and protein and the changes in solvent accessible surface area for ligand and protein alone. Factor analysis was used to deduce the number of orthogonal components required to account for the variance in the free energy, and also to show which variables were most highly correlated with these factors. From this statistical analysis it was possible to construct a three-component model that gave good agreement between the observed and calculated free energies and was also of predictive value (Fig. 9).

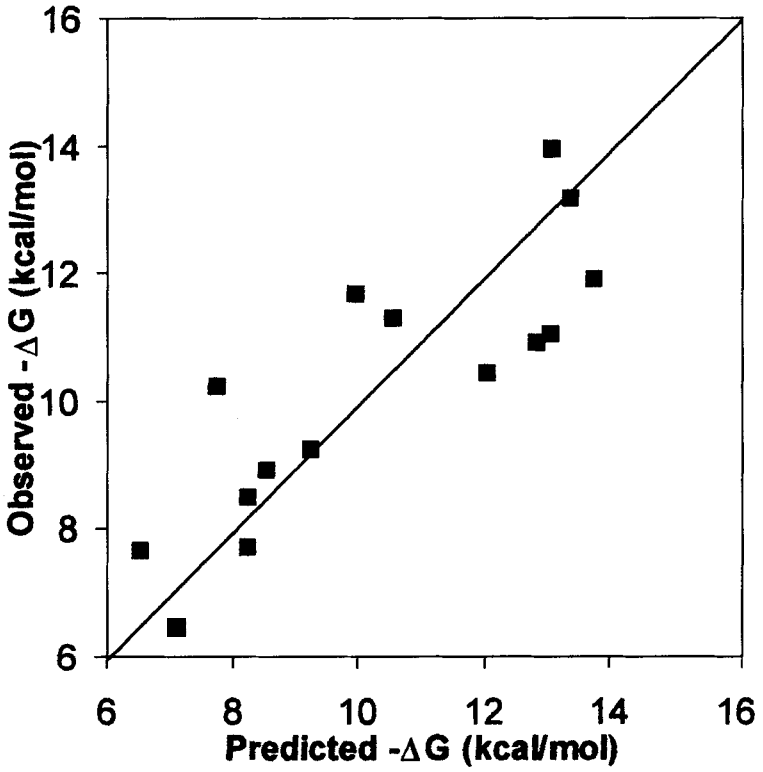


FIGURE 9 Comparison of observed and predicted ΔG values for 15 neuraminidase inhibitors using the Linear Interaction Energy method.

CONCLUSIONS

In this short article we have been able to provide no more than a superficial overview of just a few of the computational techniques that we use in our work to support the drug discovery process. Nevertheless, certain themes are emerging, such as the increasing use of the conformational properties of molecules when searching for new leads or designing combinatorial libraries – even though the number of molecules that need to be considered may exceed several million. Moreover, there is increased emphasis on approaches that can also take other properties into account; frequently these properties are related to a compound's *in vivo* activity and so increase the chances of a lead molecule possessing not only potency but also a good pharmacokinetic profile. In addition to the problems associated with these new experimental techniques there remain some problems of long-standing interest for which interesting new techniques are emerging. Computational methods such as genetic algorithms underpin much of the work in these areas; the use of such generic techniques means that a number of different problems can be tackled using the same underlying computational approach.

Acknowledgements

Much of the work described in this article would not have been possible without the contributions of our colleagues and collaborators. We are pleased to acknowledge our colleagues John Bradshaw, Mike Hann, Gianpaolo Bravi and Alfonso Pozzan and our collaborators Peter Willett, Val Gillet and Gareth Jones from Sheffield University, Robin Taylor from the Cambridge Crystallographic Data Centre, Jon Essex and Ian Wall from Southampton University and Martyn Ford and Dave Salt from Portsmouth University.

References

[1] Green, D. V. S., "Automated three-dimensional structure generation". In: "*Design of Bioactive Molecules*", Willett, P. and Martins, Y. C. (Eds.) American Chemical Society, Washington D.C. 1998, pp. 47–71.

[2] Martin, Y. C. (1992). "3D Database Searching in Drug Design". *J. Med. Chem.*, **35**(12), 2145–2154; Manallack, D. T. (1996). "Getting that hit, 3D database searching in drug discovery". *Drug Discovery Today*, **1**(6), 231–238; Charifson, P. S., Leach, A. R. and Rusinko, A. III, "The Generation and use of Large 3D Databases in Drug Discovery". http://www.netsci.org/Science/Cheminform/feature03.html.

[3] Goodford, P. J. (1985). "A Computational Procedure for Determining Energetically Favourable Binding Sites on Biologically Important macromolecules". *J. Med. Chem.*, **28**, 849–857.

[4] Leo, A. J. (1993). "Calculating log P_{oct} from Structures". *Chem. Rev.*, **93**, 1281–1306.
[5] Gillet, V. J., Willett, P. and Bradshaw, J. (1997). The Effectiveness of Reactant Pools for Generating Structurally-Diverse Combinatorial Libraries. *J. Chem. Inf. Comp. Sci.*, **37**(4), 731–740.
[6] Goldberg, D. E., "Genetic Algorithms in Search, Optimisation and Machine Learning", Addison-Wesley, Reading, M.A. 1989; Clark, D. E. and Westhead, D. R. (1996). "Evolutionary Algorithms in computer-aided molecular design". *J. Comp-Aided Mol. Des.*, **10**, 337–358.
[7] Blaney, J. M. and Dixon, J. S. (1993). "A good ligand is hard to find: Automated docking methods". *Pers. Drug Disc., And Des.*, **1**, 301–309; Shoichet, B. K., "Docking ligands to proteins". In: "*Protein Structure Prediction*". Sternberg, M. J. E. (Ed.), IRL Press, Oxford, U.K. 1996, pp. 263–290.
[8] Jones, G., Willett, P. and Glen, R. C. (1995). "Molecular recognition of a receptor sites using a genetic algorithm with a description of desolvation". *J. Mol. Biol.*, **245**(1), 43–53.
[9] Jones, G., Willet, P., Glen, R. C., Leach, A. R. and Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, **267**(3), 727–748.
[10] Kollman, P. A. (1996). "Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the Properties of Organic and Biological Molecules". *Acc. Chem. Res.*, **29**(10), 461–469.
[11] Åqvist, J., Medina, C. and Samuelsson, J. E. (1994). "A new method for predicting binding affinity in computer-aided drug design". *Pro. Eng.*, **7**(3), 385–391.
[12] Smith, P. W., Sollis, S. L., Howes, P. D., Cherry, P. C., Cobley, K. N., Taylor, H., Whittington, A. R., Scicinski, J., Bethell, R. C., Taylor, N., Skarzynski, T., Cleasby, A., Singh, O., Wonacott, A., Varghese, J. and Colman, P. (1996). "Novel inhibitors of Influenza Sialidases related to GG167". *Bioorg. Med. Chem. Lett.*, **6**(24), 2931–2936.
[13] Wall, I., Ford, M., Salt, D., Leach, A. and Essex, J., to be published.
[14] Catalyst is available from MSI, 9685 Scranton Road, San Diego, CA.