*Pure & Appl. Chem.*, Vol. 66, No. 8, pp. 1589–1596, 1994.
Printed in Great Britain.


# Computer-aided drug design

W. Graham Richards

Physical Chemistry Laboratory, Oxford University

South Parks Road, Oxford OX1 3QZ, UK

**Abstract:** Two distinct approaches are possible in the area of computer-aided drug design. If the molecular structure of the target macromolecule is known the methods are obvious and direct and have achieved a high level of sophistication. That area may be extended by using computational techniques to predict protein structure as illustrated here by the interleukin-4 receptor.
When the only lead is a set of known active compounds or knowledge of a biochemical transformation which is to be interrupted, then the path is less direct. Currently favoured tactics include the use of molecular similarity methods and the employment of neural networks. Recent advances include the prediction of the relative potency of different chiral forms of drugs.

## INTRODUCTION

Although no single drug has been designed solely by computer techniques, the contribution of these methods to drug discovery is no longer a matter of dispute. All the world's major pharmaceutical and biotechnology companies use computational design tools. At their lowest level the contributions represent the replacement of crude mechanical models by displays of structure which are a much more accurate reflection of molecular reality, capable of demonstrating motion and solvent effects. Beyond this, theoretical calculations permit the computation of binding free energies and other relevant molecular properties. The theoretical tools include empirical molecular mechanics, quantum mechanics and, more recently, statistical mechanics. This latest advance has permitted explicit solvent effects to be incorporated. Underpinning all this work is the availability of high quality computer graphics, largely supported on workstations.

Two distinct categories of research are clearly distinguishable:

a) A detailed molecular structure of the target macromolecule, the drug receptor, is known from x-ray crystallography, nmr or homology modelling.

b) The target receptor binding site has properties which can only be inferred from a knowledge of the variable activity of otherwise similar molecules.

Both these types of approach will now be considered and illustrated with some recent examples.

## DNA AS TARGET

The sequencing of the human genome represents one of the major scientific endeavours of this century. A major aspect of the utilization of this information will be the provision of small molecules which will recognize selected sequences, perhaps with the goal of switching off particular genes as in cancer chemotherapy.

For some time antibiotics such as netropsin have been known to bind preferentially to sequences rich in A-T pairs. A variant based on this research has been to try to design a bioreductive ligand based upon netropsin (1). The idea of bioreductive anti-cancer agents starts with the fact that tumours receive less blood and hence less oxygen than normal tissue. Thus it becomes possible, at least in principle, to contemplate having a ligand which can exist in two forms, oxidized and reduced, and if the redox potential is appropriate to be in the oxidized form in normal tissue but reduced in tumours. If only the reduced form will bind to the macromolecular target and cause cell death, then differentiation in action between cells which it is desirable to destroy and normal cells is achievable, with concomitant reduction in side-effects.

A second starting point for sequence selective ligands is an organometallic molecule with chiral properties. The propeller- like ruthenium tris-phenanthroline complexes do show differential binding between A-T and G-C sequences (2) and moreover may exhibit a preference for purine 3′, 5′ pyrimidine sites in DNA (3).

Perhaps the most intriguing starting point for a molecule upon which to build nucleic acid selectivity is the ubiquitous spermine. It has been proposed (4) that spermine can bind to DNA in a cross-groove manner, with relatively non-specific interactions between the positive nitrogens of the spermine and the negatively-charged phosphate backbone. In addition there is a highly specific binding down the major groove of poly(dG-dC) involving hydrogen bonds to N7 and O6 of guanine. This latter interaction could be an aspect of the property of spermine in inducing conformational changes in DNA including the B-Z transition (5). It is then possible to speculate that the role of spermine as a messenger may be to control which portions of a DNA sequence are readable by altering the coiling of the nucleic acid.

## PROTEIN AS TARGET

If an enzyme structure is known then designing inhibitors which will block activity in the test-tube should be a relatively straightforward problem. More spice to such a challenge is added if we at the same time attempt to make the ligand bioreductive as outlined above (6). The published work has taken dihydrofolate reductase as the target enzyme, but current activity is being focussed on thymidylate synthetase.

The binding free energy of the inhibitor to the enzyme is a crucial quantity: strong binding is essential.

Computationally it is possible to calculate differences in the binding free energies of two ligands A and B to an enzyme E, using the cycle

$$\begin{array}{ccc} & \Delta G_1 & \\ E + A & \rightarrow & EA \\ \downarrow \Delta G_3 & & \downarrow \Delta G_4 \\ & \Delta G_2 & \\ E + B & \rightarrow & EB \end{array}$$

The desired ($\Delta G_1 - \Delta G_2$) is equated to ($\Delta G_3 - \Delta G_4$). These latter free energy charges are non-physical but may be computed using free energy perturbation techniques (7-9).

The transformation $\Delta G_3$ represents the molecule A being changed into molecule B in aqueous solution. This may be simulated either by using Monte Carlo techniques or by molecular dynamics. The other perturbation, $\Delta G_4$, is more readily followed by molecular dynamics since with protein involved, most Monte Carlo moves are not accepted. With care, however, relative binding free energies can be computed to an accuracy approaching 1 kcal $mol^{-1}$.

This same free energy perturbation approach can also yield the redox potentials which are fundamental to the bioreductive idea (10). In this case the thermodynamic cycle is a little more complex and it is also necessary to calculate a gas-phase energy difference between pairs of oxidized and reduced molecules. High quality *ab initio* molecular orbital methods including electron correlation effects provide those energy differences.

## DRUG TRANSPORT

Sceptics quite rightly point out that designing an enzyme inhibitor which will work in the test-tube is one thing; getting a compound which will work in a cell is another. Transport across the biological membrane is essential. Compounds must be soluble enough in the lipid to get into the membrane, but not so soluble that they remain there. Within the pharmaceutical industry the partition coefficient between water and n-octanol is used as a guide to membrane transport. The free energy perturbation technique just described can also be adapted to compute partition coefficients (7).

More excitingly, however, it is becoming possible to model biological membranes. Starting with crystal structures of membranes involving DMPC (1,2-dimyristoyl-sn-glycero-3-phosphoryl choline) a highly realistic simulation is possible, involving a hydrated lipid bilayer. After very long molecular dynamics simulations the resulting membrane model is in agreement with all the available experimental data; lead

group separation; order parameters and diffusion coefficients (J.W. Essex, D.Phil. thesis, Oxford, 1992). This model can be used as the 'solvent' in calculations of partition coefficients which should be considerably more realistic than experimental values in n-octanol. Furthermore it will be possible to introduce cholesterol and protein into the model membrane to produce a truer simulation of how a given drug is transported into a cell.

## PREDICTING PROTEIN STRUCTURE

One of the major contemporary scientific aims is to use the abundance of gene and hence protein sequences to predict the three-dimensional structure of proteins: going from primary to tertiary structure. Were this routinely possible the choice of drug target where the architecture of the binding site is known would increase from a handful of cases to many thousands. The currently favoured and only successful methods are all based upon finding similarities and homologies between the protein of known sequence but unknown topology and known structures from three-dimensional databases.

Generally sequences are compared with scoring matrices being used to ascertain just how similar a short length of polypeptide in the unknown is in comparison with a known case. One successful prediction (11), that of the important small protein big endothelin, was made using not the identities of amino acids in the sequence, but their properties, notably their hydrophobicities. The property profile is smoother than an identity specification where each amino acid can be one of twenty.

Where the similarity is low the use of colour graphics to permit the human eye to detect similarities has many advantages although it is inevitably subjective. This approach, using the computer program CAMELEON (12), has recently been used to predict the structure of the interleukin-4 receptor (13). It is believed that the folding topology of the beta sheets of IL4R is the same as that seen in the crystal structure of CD4, despite sequence identity being low. Each domain of the IL4R monomer was aligned with CD4 using single residue hydropathy properties. Loops were added from a database of immunoglobulins so as to connect the sheets; side-chains were added using a side-chain rotamer library and the unsolvated structure energy-minimized using molecular dynamics. The whole structure was thus placed in an 8Å shell of water and unconstrained molecular dynamics carried out for 60 ps. Finally the whole structure was minimized.

Assuming that the IL4 receptor acts in the same way as growth hormone receptor as a dimer, one molecule of IL4 was docked to a pair of receptor proteins. The docking shows which portion of IL4 binds to the receptor: in this prediction notably the D helix. On the basis of this it should be possible to design mimetics of the crucial parts of the IL4 D helix which would interfere with the biochemical consequence of the cytokine binding to its receptor, leading to antagonists with potential medicinal applications.

## TRANSITION STATE MIMETICS

Where no knowledge about the macromolecular target in atomic detail exists, then it is still possible to utilize computer-aided design techniques. A popular idealized approach would be to compute the energy profile of a biochemical transformation which it would be desirable to inhibit; locate the transition state or intermediate and then create a stable mimic of these unstable transients. Such a mimic should be recognised by the enzyme responsible for catalysing the reaction and would hence act as an inhibitor.

Only two logical steps are necessary: find the transient structure and secondly design a stable mimic. The former task is probably best achieved by using a combination of quantum and molecular mechanics. A recent review (14) suggests that the combined potential method used by Bash et al (15) for the triosephosphate isomerase reaction is probably the technique likely to be followed in the future.

The second stage of the process invokes the introduction of the idea of molecular similarity, a quantitative measure of just how similar one molecule is to another. Perhaps the most important aspect of similarity is similarity of shape (16) and secondly similarity of molecular electrostatic potential, both of which can be represented by gaussian functions (17,18) which introduce major computational gains in the calculation of similarity indices, of which several different types may be defined (19).

Despite the simplicity of the logic this method of designing novel pharmaceutical products has not as yet had any major successes.

## SIMILARITY IN ACTIVITY RELATIONSHIPS

Much more striking has been the achievement of similarity measures in structure-activity relationships and in quantitative structure-activity relationships. Good et al (20) considered the series of steroids for which binding affinity data are available and which was the set studied in the earliest comparative molecular field three-dimensional structure-activity work. Every molecule in the series was compared in terms of shape and electrostatic potential similarity with every other member of the series yielding an n by n matrix for each property. The columns of these matrices were then used as input to a symmetrical neural network with numbers of nodes being n; n/3; 2; n/3 and n. The output was trained to be identical to the input and the values of the two central nodes (labelled x and y) noted. If one then plots x against y, clear structure-activity correlations emerge: strongly binding and weakly binding compounds cluster in different parts of the x/y plot.

The same matrices can be subjected to partial least squares analysis. The cross validated correlation coefficients obtained from the statistical analysis compare well with those obtained using the more commonly used matrices of similarities at grid points in the space surrounding the molecules which of course demand massive matrices of perhaps thousands of points. In addition there is no need for arbitrariness about the extent of molecular 'surface' or the size of the three-dimensional box into which

the molecules have to be placed. Although in its infancy molecular similarity matrices to seem to have a lot to offer in QSAR and in the optimisation of molecular structures for particular biological effects.

## MOLECULAR DISSIMILARITY

The molecular dissimilarity between a pair of molecules can be defined as (1 - Similarity). Similarity has a range of values 0 to 1 with unity representing identity. The interest in dissimilarity is in the comparison of chiral forms of the same molecule. The dissimilarity can be used as a 'chirality coefficient', a number which gives a range of values of chirality rather than this being an all-or-none property. Currently there is a great deal of research into producing pure chiral forms of compounds for use as pharmaceutical agents: the more active form being termed the eutomer and the less active the distomer, with their ratio being the eudismic ratio. For an homologous series of compounds we have shown that their is a direct correlation between the eudismic ratio and the chirality coefficient. An example is shown in Figure 1.

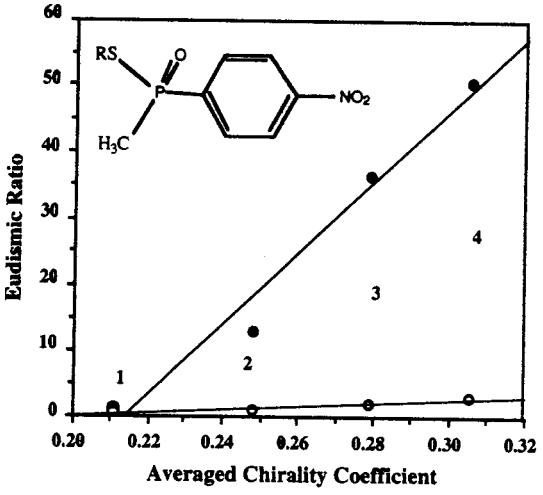


Fig. 1 The linear correlation between the eudismic ratios and chirality coefficients of muscarinic 1,3 dioxalone (R = (1) isopropyl; (2) ethyl; (3) methyl; (4) hydrogen)

This type of correlation should assist in the prediction of eudismic ratios prior to synthesis and provide a rational basis upon which to decide whether the production of a single optical form is important or not.

## CONCLUSION

Computer-aided drug design is no longer merely a promising technique. It is a practical and realistic way of helping the medicinal chemist. On its own it is unlikely to lead to pharmaceutical novelties but it has become a significant tool, an aid to thought and a guide to synthesis. Still drugs must be synthesised and tested by the computational techniques can contribute a clear molecular rationale and above all provide a spur to the imagination.

## ACKNOWLEDGEMENTS

Much of this work has been supported by a contract with the National Foundation for Cancer Research and by the Association for International Cancer Research.

## REFERENCES

1 A prototype bioreductive DNA groove binding ligand. I.S. Haworth, C. Burt, F. Gago, C.A. Reynolds and W.G. Richards. *Anti-Cancer Drug Design* 6, 59 (1991).

2 Sequence selective binding to the DNA major groove: tris (1,10 phenanthroline) metal complexes binding to poly(dG-dC) and poly (dA-dT). I.S. Haworth, A.H. Elcock, A. Rodger and W.G. Richards. *J. Biomol. Struct. and Dynamics* **9**, 23 (1991).

3 A binding mode of Λ-[tris(1,10-phenanthroline)ruthenium(II)]$^{2+}$ exhibiting preference for purine-3',5'-pyrimidine sites of DNA. I.S. Haworth, A.H. Elcock, A. Rodger, W.G. Richards. *J. Biomol. Struct. and Dynamics* **9**, 553-569 (1991).

4 A molecular mechanics study of spermine complexation to DNA: a new model for spermine-poly(dG-dC) binding. I.S. Haworth, A. Rodger and W.G. Richards. *Proc. Roy. Soc. Lond. B* **244**, 107 (1991).

5 A molecular dynamics simulation of a polyamine-induced conformational change of DNA. A possible mechanism for the B to Z transition. I.S. Haworth, A. Rodger and W.G. Richards. *J. Biomolec. Struc. Dynam.* **10**, 195 (1992).

6 Bioreductive anti-cancer drugs. W.G. Richards and C.A. Reynolds, in "Theoretical Biochemistry and Molecular Biophysics" (eds. D.L. Beveridge & R. Lavery), Adenine Press, Schenectady (1990).

7 Free energy calculations of pharmaceutically important properties. P.M. King, C.A. Reynolds, J.W. Essex, G.A. Worth and W.G. Richards. *Mol. Simulation* **5**, 265 (1990).

8 Free energy calculations in molecular biophysics. C.A. Reynolds, P.M. King and W.G. Richards. *Mol. Phys.* **76**, 251 (1992).

9 Errors in free-energy perturbation calculations due to neglecting the conformational variation of atomic charges. C.A. Reynolds, J.W. Essex and W.G. Richards. *Chem. Phys. Lett.* **199**, 257 (1992).

10 Theoretical calculation of electrode potentials: electron withdrawing compounds. S.G. Lister, C.A. Reynolds and W.G. Richards. *Int. J. Quantum. Chem.* **41**, 293 (1992).

11 A theoretical study of the structure of big endothelin. M.C. Menziani, M. Cocchi, P.G. De Benedetti, R.G. Gilbert, W.G. Richards, M. Zamai, V.R. Caiolfa. *J. Chim.Phys.* **88**, 2687 (1991).

12 CAMELEON, Oxford Molecular Ltd., Magdalen Centre, Oxford Science Park, Oxford OX4 4GA, UK

13 A computer model of the interleukin-4 receptor complex. P. Bamborough, G.H. Grant, C.J.R. Hedgecock, S.P. West and W.G. Richards. *Proteins* [in press].

14 Computer modelling of enzyme catalysed reaction mechanisms. A.J. Mulholland, G.H. Grant and W.G. Richards. *Protein Eng.* **6**, 133 (1993).

15 Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. Bash, P.A., Field, M.J., Davenport, R.C., Petsko, G.A. and Karplus M. *Biochemistry* **30**, 5826 (1991).

16 Similarity of molecular shape. A.Y. Meyer and W.G. Richards. *J. Comput. Aided Mol. Design* **5**, 427 (1991).

17 Rapid evaluation of shape similarity using gaussian functions. A.C. Good and W.G. Richards. *J. Chem. Inf. Comput. Sci.* **33**, 112 (1993).

18 Utilization of Gaussian functions for the rapid evaluation of molecular similarity. A.C. Good, E.E. Hodgkin and W.G. Richards. *J. Chem. Inf. Comput. Sci.* **32**, 188 (1992).

19 A linear molecular similarity index. C.A. Reynolds, C. Burt and W.G. Richards. *Quant. Struct. Act. Relat.* **11**, 34 (1992).

20 Structure-activity relationships from molecular similarity matrices. A.C. Good, Sung-Sau So and W.G. Richards. *J. Med. Chem.* **36**, 433 (1993).

21 Chiral drug potency: Pfeiffer's rule and computed chirality coefficients. A. Seri-Levy and W.G. Richards. *Tetrahedron Asymmetry* [in press].