Supramolecular Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713649759

Monte Carlo algorithms for docking to proteins

R. J. Read[a]; T. N. Hart[a]; M. D. Cummings[a]; S. R. Ness[a]
[a] Dept. of Medical Microbiology and Infectious Diseases, 1-41 Medical Sciences Building, University of Alberta, Edmonton, Alberta, Canada

To cite this Article Read, R. J. , Hart, T. N. , Cummings, M. D. and Ness, S. R.(1995) 'Monte Carlo algorithms for docking to proteins', Supramolecular Chemistry, 6: 1, 135 — 140
To link to this Article: DOI: 10.1080/10610279508032529
URL: http://dx.doi.org/10.1080/10610279508032529

SUPRAMOLECULAR CHEMISTRY, Vol. 6, pp. 135–140
Reprints available directly from the publisher
Photocopying permitted by license only

Monte Carlo algorithms for docking to proteins

R. J. READ*, T. N. HART, M. D. CUMMINGS and S. R. NESS

Dept. of Medical Microbiology and Infectious Diseases, 1-41 Medical Sciences Building, University of Alberta, Edmonton, Alberta, Canada T6G 2H7

(Received August 5, 1994)

The goal of docking is to predict binding interactions between molecules. We are primarily interested in docking as a tool for the structure-based design of new ligands that could serve as lead compounds for drug development. The program BOXSEARCH uses a Monte Carlo algorithm to explore the relative orientation and position of two molecules. Multiple runs are carried out from different random starting positions and orientations, and the temperature of the system is gradually reduced. An unbiased sampling of low energy states is the result. BOXSEARCH has been tested on a number of known complexes, involving both protein and small molecule ligands. Although a better treatment of solvent effects and of flexibility would improve the ranking of results, the complexes can be reconstructed successfully, even using uncomplexed conformations of the molecules. We are currently implementing two major enhancements. First, the code is being rewritten in a more general and adaptable form, using the object-oriented programming language C++. Object-oriented programming allows us to reuse code very easily and also lets us use a higher level of abstraction. In practical terms, this makes it much easier to program and test new ideas for molecular simulations, including better treatments of solvent and flexibility. Second, genetic algorithms are being implemented as a more general and powerful optimization tool. We envision simulations in which molecules "evolve" on the computer, by mutation and recombination in the binding site.

INTRODUCTION

The goal of docking is to predict binding interactions between molecules. In our work, we are interested in docking calculations where the target molecule is a protein (or other biological macromolecule) of known three-dimensional (3D) structure, determined either by crystallographic or nuclear magnetic resonance techniques. Depending on the nature of the probe being docked to the target, docking calculations can be put to a number of uses. If the probe is a second protein, one can predict protein-protein interactions, such as between enzymes and inhibitors, antibodies and antigens, or members of a multienzyme complex. If the probe is a small molecule, one can predict enzyme-substrate interactions, or choose potential ligands (which can serve as lead compounds for drug design) from a 3D database. Our initial interest in docking was in the possibility of docking molecular fragments into the active site of a potential drug target, in order to build up novel ligands in an approach to *ab initio* structure-based drug design.

In the fragment assembly approach to drug design, one would dock members of a library of molecular fragments to the target. Since each fragment of a tightly bound ligand need not be in its global energy minimum binding site, it would be best to find a set of minima for each fragment. Fragment docking, as a result, requires a search for multiple minima over (at least) the six dimensions of relative rotation and translation of the probe and target. If molecular flexibility is taken into account, the number of dimensions increases. The Monte Carlo method has been demonstrated to be very effective in a wide variety of high-dimension, multiple minima problems, so we decided to investigate its use in the problem of docking molecules and fragments to protein targets.

A number of other approaches to structure-based drug design are being developed. The DOCK method of Kuntz and coworkers[1–3] is probably the most widely used at present. In this method, members of a 3D database are fit by steric complementarity into the binding pocket, using a distance geometry algorithm, then scored by various measures of chemical complementarity. The program LUDI[4,5] fits fragments into the binding pocket, using rules for functional group interactions. These approaches, and others, have been reviewed recently[6–8].

*To whom correspondence should be addressed.

BOXSEARCH: MULTIPLE-START MONTE CARLO DOCKING

The multiple-start Monte Carlo algorithm used in our docking program, BOXSEARCH, has been described in detail elsewhere[6,9], so it will only be summarized briefly here. Monte Carlo methods are highly effective in finding energy minima in complicated multiple minimum problems. A series of random changes are applied to the state of the system. The change in energy is evaluated: if the energy drops, the new state is accepted; if the energy increases, a random choice is made to accept or reject the new state according to the Boltzmann probability of that increase in energy.

$$p(\Delta E) = \exp(-\frac{\Delta E}{kT})$$

Because there is a reasonable chance of traversing barriers between minima, if they are of the order of kT or less, the convergence radius is increased compared to gradient-driven minimization methods.

In the docking problem, the state of the system is changed by altering the mode of binding of the probe to the target. In BOXSEARCH, the target and probe are both considered to be rigid molecules, so that only their relative orientation and position are varied. As discussed below, some allowance for conformational flexibility in both molecules would be desirable, and is currently being implemented in a new version of the program.

There are a number of special features to our approach to Monte Carlo docking. We conduct a large number of short trials, instead of using the same amount of computing time to conduct a small number of long trials. There are two reasons for this choice. First, this tradeoff has often been found to be more efficient in other applications of the Monte Carlo method. Second, we do not expect each fragment of a good ligand to sit in its own global energy minimum, so we need to locate a number of promising minima for each fragment. Each trial is initiated by placing the probe in a random orientation and position within a box surrounding the binding site of interest. The search is therefore not biased by the choice of a "seed" orientation.

A "floating" function is used in the beginning of each Monte Carlo trial to enhance the sampling of deep binding pockets because, with rigid molecules, it could be quite difficult to find a path from the outside. Typically, the box surrounding the binding site will enclose a significant volume of protein, so in some of the trials the probe will start out buried inside the protein. To bring such buried probes to the nearest pocket or other surface region, a "floating" pseudo-energy function is calculated for each atom as the distance to the nearest surface point.

Once the probe has reached the surface, the energy evaluation switches to a conventional potential function with electrostatic and 6/12 van der Waals terms[9]. Since the calculation is performed without explicit solvent, it is necessary to compensate for solvent screening effects in the electrostatic calculations. The strategy chosen is to set net charges on charge groups to zero, thereby preserving short-range dipole interactions such as hydrogen bonds while eliminating the long-range electrostatic interactions that would otherwise be severely overestimated[9,10].

In each trial, a fixed number of steps is run according to a simulated annealing schedule. The maximum rotation and translation to be applied are specified, as is the temperature parameter that determines the height of energy barriers that can be traversed between local minima. The temperature is reduced through a trial, to ensure that at the end of the Monte Carlo run each probe is near a minimum of the potential function.

A typical run usually involves 5,000 to 20,000 or more trials. To simplify the analysis, trials above a cutoff energy value are discarded. The remaining trials are clustered, and only the lowest energy member of each cluster is saved for analysis.

BOXSEARCH TESTS

To validate the method, a number of docking experiments with known solutions have been run, with various sizes of probe: proteins, small molecule ligands, and molecular fragments. Given the approximations involved in the potential function, it would not be surprising if BOXSEARCH were only capable of generating the correct docking and ranking it among a reasonable number of possibilities. More sophisticated (and computationally expensive) calculations could then be performed on a small number of cases. It was gratifying to find that in some cases, particularly those involving protein probes, the lowest energy docking corresponds to reality, and is well separated from secondary minima. In the case of small molecule ligands, the correct answer ranks close to the top of the list, and even for molecular fragments the correct docking is generated among the minima. However, as one might anticipate, an analysis of the results shows that a number of improvements could be made. These will be discussed below.

SGPB:OMTKY3

Streptomyces griseus proteinase B (SGPB) is a bacterial serine proteinase. It is strongly inhibited by the third domain of the ovomucoid inhibitor from turkey eggs (OMTKY3), a protein proteinase inhibitor. The structure of their complex has been determined crystallographically at high resolution[11], as has the structure of native SGPB[12]. Docking tests in this system (as well as the

DHFR:MTX system discussed in the following section) have been described in detail[9].

As a first test, the crystallographic complex was reassembled using BOXSEARCH. The lowest energy docking reproduces the observed complex faithfully, as the rms difference between observed and docked atoms of the OMTKY3 inhibitor is only 0.4Å. In a more realistic test, OMTKY3 was docked to the native structure of SGPB. As one would expect, the results are not quite as good, but the lowest energy docking is still close to the observed structure, with an rms difference of 1.7Å. It should be noted that only a small segment of the inhibitor, the reactive site loop, is in direct contact with the enzyme, and this portion shows a considerably smaller rms difference (see Figure 1).

Docking tests were also performed with the dipeptide analogue *N*-isopropyl-isobutylamide, in several conformations selected to match fragments of the reactive site loop. For each fragment, a cluster close to the orientation in the whole inhibitor was found, though the rank in the list ranged from 2 to 361. This result supports the notion that fragment docking and assembly is a viable approach to structure-based drug design.

DHFR:MTX

Dihydrofolate reductase (DHFR) is an important drug target for both antibacterial and anticancer agents. A number of relevant crystal structures have been determined, including the complex of the *E. coli* enzyme with the anticancer drug methotrexate (MTX)[13]. Reconstructing this complex has become a standard test for docking methods. We tested docking of the whole molecule, and also of two fragments.

In contrast to the SGPB:OMTKY3 case, the correct dockings were more ambiguous. For the whole MTX, the correct answer was sixth in the list. In the case of the two fragments, the correct docking for the pteridine moiety ranked 18, while the fragment comprising the remainder of the inhibitor came at position 284. A number of the incorrect dockings were inserted into the binding site for the cofactor NADPH, which is not bound to DHFR in this crystal structure. On inspection, it can be seen that many of the high ranked incorrect dockings expose hydrophobic portions of the inhibitor to the solvent. An improved treatment of solvent effects would lower their ranking.

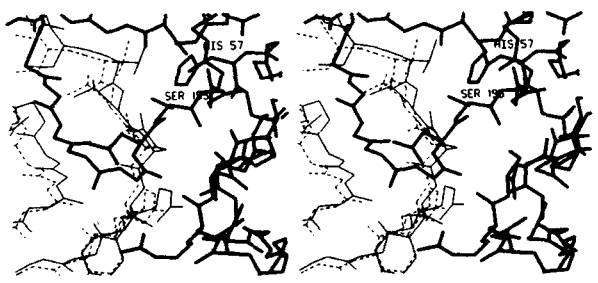

Figure 1 Closeup view of the docking[9] of the protein inhibitor OMTKY3 (dashed lines) to native SGPB (thick lines). The position of OMTKY3 obtained by superimposing the crystallographic complex is shown in thin lines. For clarity, only sidechains that are in contact between the inhibitor and the proteinase are shown.

CAT:CLM

Chloramphenicol acetyltransferase (CAT) acetylates and thereby inactivates the antibiotic chloramphenicol (CLM), rendering the bacteria that express CAT resistant to the antibiotic. A high resolution structure of the complex between *E. coli* CAT and CLM is available[14]. We chose this system to study the consequences of ligand flexibility[6], since CLM has seven rotatable bonds.

In a first test, BOXSEARCH was used to reassemble the crystallographic complex, using the bound conformation of CLM. This was quite straightforward, and the lowest energy docking was 0.4Å rms from the crystallographic structure. Next, a series of 18 low energy conformations of CLM were generated and used for identical docking runs. When the results from these runs were merged and analyzed, the correct binding mode (1Å rms) ranked 16 in the list. Reminiscent of the MTX case, the lowest energy docking exposed the phenyl ring of CLM to the solvent.

Diubiquitin

The protein ubiquitin is, as its name implies, ubiquitous among eukaryotic organisms due to the important role it plays in metabolic regulation. The covalent attachment of ubiquitin polymers acts as a signal for the regulated degradation of specific proteins. This conjugation occurs by the formation of an amide linkage (termed an isopeptide bond) between the carboxy-terminal residue of ubiquitin, Gly76, and a lysine side-chain on the target protein. In the formation of ubiquitin polymers, the major linkage is to Lys48. Crystal structures have been determined for the monomeric[15], dimeric[16] and tetrameric[17] forms of ubiquitin. We set out to reconstruct diubiquitin from its two halves, and to predict it by docking two copies of the monomeric ubiquitin.

A docking experiment with diubiquitin is a particular challenge because the covalent bond makes a major contribution to the stability of the dimer. Our docking method considers only the non-covalent interactions, which are quite weak in this case. The dissociation constant for non-covalent dimerization is estimated to have a lower limit of 10mM. Because the program will not make a covalent bond, there would be a serious van der Waals clash in the correct docking, so residue Gly76 was deleted in all trials.

Reassembly of diubiquitin from its two halves was quite straightforward. The lowest energy docking differed by 0.6Å rms from a reference structure obtained by minimizing the crystallographic result, and the highest

ranked incorrect docking was separated by about 27kcal/mole.

To assemble diubiquitin from two copies of monomeric ubiquitin required a bit more intervention. It turns out that a side-chain that is very flexible in mono-ubiquitin, Arg42, is buried in the dimer interface of diubiquitin, where it adopts a different conformation. In its mono-ubiquitin conformation, a serious van der Waals clash precludes the correct docking. In principle, the exposed nature of this side-chain and its high crystallographic B-factors would have allowed us to anticipate that it is flexible enough to adopt a different conformation in the dimer. One might deal with this by incorporating flexibility into the Monte Carlo calculation. Alternatively, for protein-protein docking one might reason that losing a small part of the interaction surface, by truncating the side-chain, is much less serious than risking a van der Waals clash. This simple approach is the one we chose, replacing Arg42 by an alanine residue.

The docking trials using Arg42->Ala mono-ubiquitin were fairly successful. The first and third clusters were recognizably similar to the crystallographic dimer, though rms deviations of 4.6Å and 3.3Å are worse than we have generally obtained in other tests. Clearly other minor conformational differences reduce the complementarity across the dimer interface. There were a number of incorrect dockings with similar energy scores, so it was necessary to apply additional filters to select the correct dockings. First, in a correct docking it must be possible to form the isopeptide bond. Unfortunately, the C-terminal four residues of mono-ubiquitin are rather flexible[15], so testing for the possibility of bond formation requires a systematic conformational search. This was carried out for cluster 2, the lowest energy incorrect docking, and ruled out the possibility of forming the required bond. Second, it has generally been found that homodimers exhibit pure two-fold symmetry; of the top clusters, only the correct ones satisfied this criterion.

Finally, we wished to test whether corrections for solvent effects, as judged by buried surface areas, would improve the ranking of results. The calculated interaction energies were adjusted by adding a term obtained from atomic solvation parameters[18] and the correct docking went from being barely discriminated to being separated by 13kcal/mole or more.

LESSONS FROM THE DOCKING TESTS

Knowing the approximations that have gone into the calculations, one can anticipate what might go wrong, but without running tests it would be difficult to know which effects are most important. We now have a clearer idea of the relative importance of a number of improvements that could be made to the method.

Solvent effects

In principle, with unlimited access to computing resources we could perform molecular dynamics simulations in the presence of water for each energy evaluation. In practice, the computer requirements of our method are already fairly intensive, so we need to use less expensive alternatives. There are several types of error that are introduced by the lack of solvent, and there are various approximations to deal with them.

One effect of solvent is to screen electrostatic interactions, because of the tendency for the polar water molecules to orient with the electrostatic field. While more sophisticated techniques exist, we feel that the use of neutral charge groups has been effective, and that other corrections deserve a higher priority.

Solvent-solute interactions are considerably more important, since the formation of a complex removes the interacting surfaces from contact with the solvent. Depending on the nature of the groups buried, burial of surface can be more or less favourable. We have seen in the tests that a number of high-ranked incorrect dockings expose hydrophobic surface to the solvent, implying that a correction for this effect is quite important. The atomic solvation parameters[18] used to adjust energies in the diubiquitin tests will be incorporated into the docking potential.

Finally, we have observed in incorrect dockings that hydrogen bond donors or acceptors are occasionally buried with their hydrogen bonding potential unsatisfied. While these will have less favourable binding energies than complexes where hydrogen bonds are formed, there should be an additional penalty reflecting the fact that these hydrogen bonds would have been satisfied in water.

Solvent-mediated interactions

Many protein-ligand interactions are mediated by solvent molecules that bridge between polar groups. In our docking calculations to date, all solvents are omitted. One way to deal with this would be to use an algorithm that automatically places water molecules in optimal positions at each step. Alternatively, one could use the experimentally determined solvent positions that frequently fill the active site of uncomplexed enzymes.

In the case of the CAT:CLM complex, a number of water molecules are observed to mediate interactions[14], and we found that the fit to the pocket in their absence was somewhat sloppy[6]. In the native CAT structure, several of these water molecules are found in the same positions (A.G.W. Leslie, personal communication). If the water molecules were included in the coordinates of the target, the energy evaluation could simply ignore the ones that clash with the position of the probe. Another possibility would be to vary their occupancy from 0 to 1 depending on the energy of interaction with the probe, weighting them by a Boltzmann factor. A similar ap-

proach has been used to weight different possible side-chain conformations (rotamers) in calculations of substrate binding energy[19].

Probe flexibility

There are three ways to deal with the flexibility of the probe in docking studies. The first is to allow its conformation to change during the docking calculation[20,21]. The second is to model a collection of low energy conformations for the probe and dock all of them, as we have done in the CAT:CLM docking[6]. If the energies of these conformations differed significantly, a conformational energy term would have to be included. Finally, the probe can be docked as a collection of fragments that are either rigid or adopt a small number of possible conformations that are included in the library. This approach is also possible within the BOXSEARCH protocol.

Target flexibility

One must distinguish among different types of flexibility in the target. Small conformational changes, of the order of tenths of Ångstrom units, are probably best dealt with by softening the van der Waals potential. Variable side-chain conformations could be handled by choosing the optimal set of rotamers in each energy evaluation[22], using an efficient combinatorial algorithm such as that allowed by the dead-end elimination theorem[23,24]. Alternatively, Boltzmann weighting of rotamers could be applied[19], as discussed above.

Larger conformational changes, in the main-chain of the protein, occur fairly frequently during substrate binding. Ultimately, one would like to predict, and exploit, such conformational changes. However, dealing with relatively rigid targets is already sufficiently challenging, and there is much to be accomplished with such targets in the near future.

CURRENT PROGRESS

Object-oriented programming

Object-oriented programming (OOP) and design has emerged as the software design paradigm of the future. In this approach, emphasis is placed on the design of fundamental "objects", which constitute the basic elements appearing in the problem. Objects are specified by deciding how their internal data are stored, and how these data can be accessed by the other parts of the program. A primary advantage of this approach is that objects occur naturally in many programming problems, so the design of the software is more closely associated with how we might intuitively think about the problem. In addition, objects that are used to build one application can be reused in new applications with little or no redesign, because their use is generally independent of their internal construction.

The BOXSEARCH program has been extensively rewritten in the object-oriented programming language, C++. We have found that the Monte Carlo docking simulation is very amenable to the OOP approach. The basic objects that we use include atoms, molecules, potential functions, and molecular parameter libraries. In addition to the original functionality of BOXSEARCH, we have added routines to perform rigid-body minimization of the ligand with respect to the receptor, so that minimization can be run after the Monte Carlo search procedure. We find that with the design of these objects, the top level program is fairly simple in structure, and therefore can readily be altered to perform different types of simulations. Therefore, the effort in rewriting the code in C++ is already paying off, as new simulations are much easier to design than they were when the program was written in the C programming language.

We are currently enhancing the existing objects in a number of ways. First, flexibility is being introduced by including bond information in the molecule objects. Second, we are investigating the use of "multiple conformations" for molecules, particularly as a way of dealing with flexibility in proteins in docking procedures without having to introduce the prohibitive overhead of full flexibility. Third, we are in the preliminary stages of designing molecular fragments, which would be used in the *de novo* design of new ligands. Each fragment would represent a primitive functional group that could be altered, so the chemical structure of the molecule could be altered during the docking run. Genetic algorithms would be used to generate mutations and crossovers of the molecular species.

Genetic algorithms

Genetic Algorithms (GAs) are adaptive search mechanisms that model biological evolutionary processes in order to solve difficult optimization problems[24,25]. There are a variety of different models that are grouped under the heading of GAs, but they all share some common features. First, there is a population of individuals, each with its own genotype and phenotype. The genotype of an individual is typically a linear chromosome, represented by a string of values selected from an alphabet of two or more possible letters. This genotype is the recipe by which the phenotype is created, just as in biological organisms. The genotype is manipulated by biologically inspired operators such as crossover and mutation, each operating with a certain low probability. After a phenotype is produced, it is evaluated with respect to how close it comes to an optimal solution and is assigned a 'fitness'. The most fit individuals are preferentially chosen to form a new generation, and this cycle of mutation/crossover,

evaluation and selection is repeated until a desired level of overall population fitness is reached.

GAs have proven their worth in a wide variety of problems[24,25]. They are particularly useful when partial answers to a problem tend to give a better than random score, which is true of many complex optimization problems. Partial answers are generated randomly (by mutation) much more frequently than complete answers would be; they build up in the population, and then recombination of different partial answers builds up more complete answers. Recently, GAs have been applied to the protein folding problem, with encouraging results[26,27]. As well, the use of genetic algorithms has been proposed for structure-based drug design (see, for example[28]).

We are currently engaged in using object-oriented programming with C++ to write a flexible and easily extensible GA package designed to dock flexible and mutable ligands into the active sit of a given protein. Thanks to the object-oriented approach of C++, we have been able to easily incorporate the molecule and energy minimization classes designed in the new C++ version of BOXSEARCH. Object-oriented programming has also been very useful by making possible a higher level of abstraction in the problem, allowing us to change underlying implementation details of the GA without having to change the higher level code.

ACKNOWLEDGMENTS

This research was supported by the Medical Research Council of Canada, and in part by an International Research Scholars award to RJR from the Howard Hughes Medical Institute. RJR is a Scholar of the Alberta Heritage Foundation for Medical Research. MDC was the holder of a Natural Sciences and Engineering Research Council of Canada postgraduate scholarship while on leave from SynPhar Laboratories, Inc., Edmonton.

Since this manuscript was submitted, our work on ubiquitin docking has been published[29].

REFERENCES

1 Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269–288.
2 DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. *J. Med. Chem.* **1988**, *31*, 722–729.
3 Kuntz, I. D. *Science* **1992**, *257*, 1078–1082.
4 Böhm, H.-J. *J. Comput.-Aid. Mol. Des.* **1992**, *6*, 61–78.
5 Böhm, H.-J. *J. Comput.-Aid. Mol. Des.* **1992**, *6*, 593–606.
6 Hart, T. N.; Read, R. J. In *The Protein Folding Problem and Tertiary Structure Prediction*; Merz, K. M., Jr., LeGrand, S. M., Eds.; Birkhäuser: Boston, 1994; 71–108.
7 Rotstein, S. H.; Murcko, M. A. *J. Med. Chem.* **1993**, *36*, 1700–1710.
8 Verlinde, C. L. M. J.; Hol, W. G. J. *Structure* **1994**, *2*, 577–587.
9 Hart, T. N.; Read, R. J. *Proteins* **1992**, *13*, 206–222.
10 Aqvist, J.; van Gunsteren, W. F.; Leijonmarck, M.; Tapia, O. *J. Mol. Biol.* **1985**, *183*, 461–477.
11 Read, R. J.; Fujinaga, M.; Sielecki, A. R.; James, M. N. G. *Biochem.* **1983**, *22*, 4420–4433.
12 Delbaere, L. T. J.; Hutcheon, W. L. B.; James, M. N. G.; Thiessen, W. E. *Nature* **1975**, *257*, 758–763.
13 Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13650–13662.
14 Leslie, A. G. W. *J. Mol. Biol.* **1990**, *213*, 167–186.
15 Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531–544.
16 Cook, W. J.; Jeffrey, L. C.; Carson, M.; Chen, Z.; Pickart, C. M. *J. Biol. Chem.* **1992**, *267*, 16467–16471.
17 Cook, W. J.; Jeffrey, L. C.; Kasperek, E.; Pickart, C. M. *J. Mol. Biol.* **1994**, *236*, 601–609.
18 Eisenberg, D.; Wesson, M.; Yamashita, M. *Chemica Scripta* **1989**, *29A*, 217–221.
19 Wilson, C.; Mace, J. E.; Agard, D. A. *J. Mol. Biol.* **1991**, *220*, 495–506.
20 Goodsell, D. S.; Olson, A. J. *Proteins* **1990**, *8*, 195–202.
21 Caflisch, A.; Niederer, P.; Anliker, M. *Proteins* **1992**, *13*, 223–230.
22 Leach, A. R. *J. Mol. Biol.* **1994**, *235*, 345–356.
23 Desmet, J.; de Mayer, M.; Hazes, B.; Lasters, I. *Nature* **1992**, *356*, 539–542.
24 Holland, J. H. *Adaptation in natural and artificial systems*; MIT Press: Cambridge, Massachusetts, 1992.
25 Goldberg, D. E. *Genetic algorithms in search, optimization and machine learning*; Addison-Wesley: Reading, Massachusetts, 1989.
26 Unger, R.; Moult, J. *J. Mol. Biol.* **1993**, *231*, 75–81.
27 Dandekar, T.; Argos, P. *J. Mol. Biol.* **1994**, *236*, 844–861.
28 Robertus, J. *Nature Struct. Biol.* **1994**, *1*, 352–354.
29 Cummings, M. D.; Hart, T. N.; Read, R. J. *Protein Sci.*, **1995**, *4*, 885–899.