Structure-based drug design: progress, results and challenges

Protein structure-based drug design is rapidly gaining momentum. The new opportunities, developments and results in this field are almost unbelievable compared with the situation less than a decade ago.

In the mid-eighties one of us wrote a review entitled "Protein crystallography and computer graphics — toward rational drug design" [1]. It listed about 10 projects, a major fraction of the number of protein structure-based drug design-related projects going on worldwide at that time. Yet at a meeting a few months ago, Alex Wlodawer showed a slide listing close to 200 structure determinations that have been performed worldwide on a single protein, HIV protease, complexed with a large variety of inhibitors; the number may have risen even further since then ...

These two facts dramatically illustrate the explosive growth in structure-based drug design in the last few years. The tremendous increase in detailed structural knowledge of medically relevant proteins is due to several factors. First, molecular biology techniques have made it possible to obtain large amounts of virtually any protein — although membrane proteins remain difficult to obtain in large quantities and with great purity [2]. Second, protein purification methods have been continuously improving, thanks in particular to more efficient chromatographic procedures [3]. Third, over-expression systems have facilitated the production of isotope-labeled proteins, which are the cornerstone of the heteronuclear multi-dimensional experiments used in NMR structural elucidations [4]. Ever higher field strengths have also increased the sensitivity and the information content of NMR spectra [5]. Fourth, data collection in protein crystallography has been revolutionized due to the widespread introduction of area-detectors [6], the availability of incredibly powerful synchrotron X-ray sources [7], and the development of cryo-cooling techniques [8]. These innovations make it possible to tackle weakly-diffracting and very radiation-sensitive crystals successfully. Finally, the introduction of workstations with ever-increasing computing and graphics capabilities has greatly facilitated the computational side of protein NMR and crystallography. All these developments have resulted in an exponential growth in the number of protein structures solved. Excluding mutants and complexes with small ligands, 226 structures were published in 1992 [9]. This is certainly an underestimate of the number

of new structures solved, since many protein structures are kept classified by pharmaceutical companies. The rate of structure determinations has doubled in the last two years, and this rate is still increasing [9].

The large number of structural investigations on medically relevant proteins reflects the general recognition that the structure of a potential drug target is very precious knowledge for a pharmaceutical company, not only for lead discovery and lead optimization but also in the later phases of drug development — stages where issues such as toxicity or bioavailability may crop up. At these late stages, knowledge of the binding mode of potential drug candidates to the target protein makes it easier to modify the compound in a rational manner.

One should never forget, however, that there is often a long road between the discovery or design of a tightly-binding inhibitor of a target protein and the commercial availability of a drug. A successfully developed inhibitor may be too toxic, teratogenic, too rapidly cleared, too quickly metabolized, unable to reach the target enzyme in sufficient concentration, unstable in solution, too difficult to synthesize in bulk or too costly to produce. The criteria for allowing a new compound to be administered to large populations need to be quite stringent, and this is the main reason for the failure of compounds to become useful drugs. To predict how a new compound will change the delicate balance of all metabolic, transport and signalling pathways in the human body is simply impossible, no matter how much pharmacological and toxicological know-how has been invested in tailoring of the compound for use in humans. Hence, many promising compounds will unfortunately have to be rejected when they are found to show unacceptable side effects in humans.

The emergence of structure-based drug design as a new technology is nevertheless a fascinating development of major, worldwide importance. The final verdict on the power of this method will not be clear for one or two decades, since it will take this long for enough cases to be studied to arrive at a statistically valid conclusion. At present, the field is exciting and full of surprising results, as we will show in this review.

Finding leads

Elucidating the three-dimensional structure of a target protein, no matter how challenging and demanding it is, is only the very first step in the structure-based design of new inhibitors. The next step is to find a lead — the term for a compound that binds to the protein of interest; it often exhibits weak affinity or is too toxic, too unstable or has other shortcomings, yet it forms a starting point to develop molecules with improved pharmacological properties. In the pharmaceutical industry an acceptable lead typically has a dissociation constant of 10 μM or better. Thus, only three to four orders of magnitude of affinity have to be gained in the lead optimization process before the low nanomolar range, characteristic for many successful drugs, is reached.

It is still a major challenge to design *de novo* leads on the basis of an unliganded protein structure alone, even though an increasing number of computational techniques for this process have become available (see below). Remarkable successes have been obtained with the program DOCK; 2–20 % of the suggested compounds found using this program proved to be micromolar inhibitors [10].

In practice, however, the fastest way to arrive at leads might still be to screen a corporate database of synthesized or naturally-occurring compounds in solution. Typically, a few hundred thousand compounds can be screened in a couple of months using cocktail screening methods and robotic technology. Screening strategies also include microbial broths and plant extracts; for billions of years nature has been in the business of developing small molecules which attach themselves tightly to essential protein molecules of attacking and competing organisms. Some of these molecules have such an unexpected complexity — take for instance taxol — that it is hard to believe that current drug design procedures would come up with such compounds.

To begin to rival the complexity provided by nature, several groups have turned to screening techniques aimed at discovering tightly-binding ligands from combinatorial libraries. For example, the phage display method is based on the display of a random sequence peptide on the surface of a phage. The phage library, typically including $10^6 - 10^8$ different peptides, is mixed with the target protein, which is immobilized on the surface of a plate. Non-binding phages are washed away while the bound ones can be used to decipher the sequence of the peptide bound to the target protein [11]. Alternatively, the affinity screening of synthetic peptide [12] and oligonucleotide [13] libraries offers the possibility of arriving at compounds which are not limited to the naturally-occurring amino acids or nucleotides. Peptoids, linear oligomeric N-substituted glycines, show great promise since they are metabolically stable against proteases [11]. Moving even further from peptide-based libraries, chemical libraries — cocktails of synthetic compounds where a central scaffold is decorated on various positions with a selection from a large range of different substituents — are also becoming increasingly popular (see, for example, [14]). An extensive review of these emerging techniques has appeared recently [15].

Optimizing leads

Screening procedures generally come up with leads which are far from perfect. These molecules then have to be optimized. At this point, the structure of the target protein in complex with the lead molecule can be extremely useful in suggesting ways to improve the affinity of the lead for the target. Some of the computational tools that assist in this are described in the next section. It is important to realize, however, that the basic scientific understanding of intermolecular interactions is still rather primitive. The *ab initio* prediction of the binding constant of a compound to a protein molecule is, in general, still far beyond current computational techniques and biophysical scientific precision, in large part because of the ubiquitous effects of water. Water molecules confound *ab initio* predictions in several ways; they compete with hydrogen bond donors and acceptors of both protein and ligand; they can mediate hydrogen bonds; they provide a large electrostatic screening effect; they are involved in van der Waals interactions with protein as well as inhibitor; and they are the root cause of hydrophobic effects, the fundamental nature of which is still under considerable debate. A second problem is the correct description of the charge distribution on protein and ligand. Not only is it difficult to arrive at 'true' partial atomic charges but, due to polarization effects, the charge distribution of a ligand is different in solution from that seen when it is bound to a protein molecule. In addition, there is the problem of flexibility — of both protein and ligand — so that the configuration space that is accessible to the molecular system is so large that it becomes difficult to generate a representative ensemble necessary for the calculation of free energies.

Our imperfect understanding of intermolecular interactions coupled with the large number of degrees of freedom, makes it crucial that the protein structure-based drug design process be a tightly integrated, multidisciplinary activity where the intuition of the medicinal chemist is combined with the expertise of the protein structure specialist and the computational molecular modeler. The process must also be cyclic (Fig. 1); the predicted mode of interaction of a new inhibitor to the target protein must be experimentally verified to prevent the next cycles of the drug design process from being based on entirely wrong assumptions. There is a growing collection of examples where a modified inhibitor did indeed bind with higher affinity to the target protein than the parent compound — as predicted —

but where the binding mode was entirely different from what had been intended (see, for example, [16]).

Tools for structure-based drug design

Although quantitative *ab initio* prediction of binding constants remains a tremendous challenge [17,18], a number of qualitative rules for the design of high affin-ity ligands can be deduced from the many crystal structures of protein–ligand complexes:

(a) excellent steric and electronic complementarity to the target biomacromolecule is required;

(b) a fair amount of hydrophobic surface should be buried in the complex for tight binding;

(c) sufficient conformational rigidity is essential to ensure that the loss of entropy upon ligand binding is acceptable.
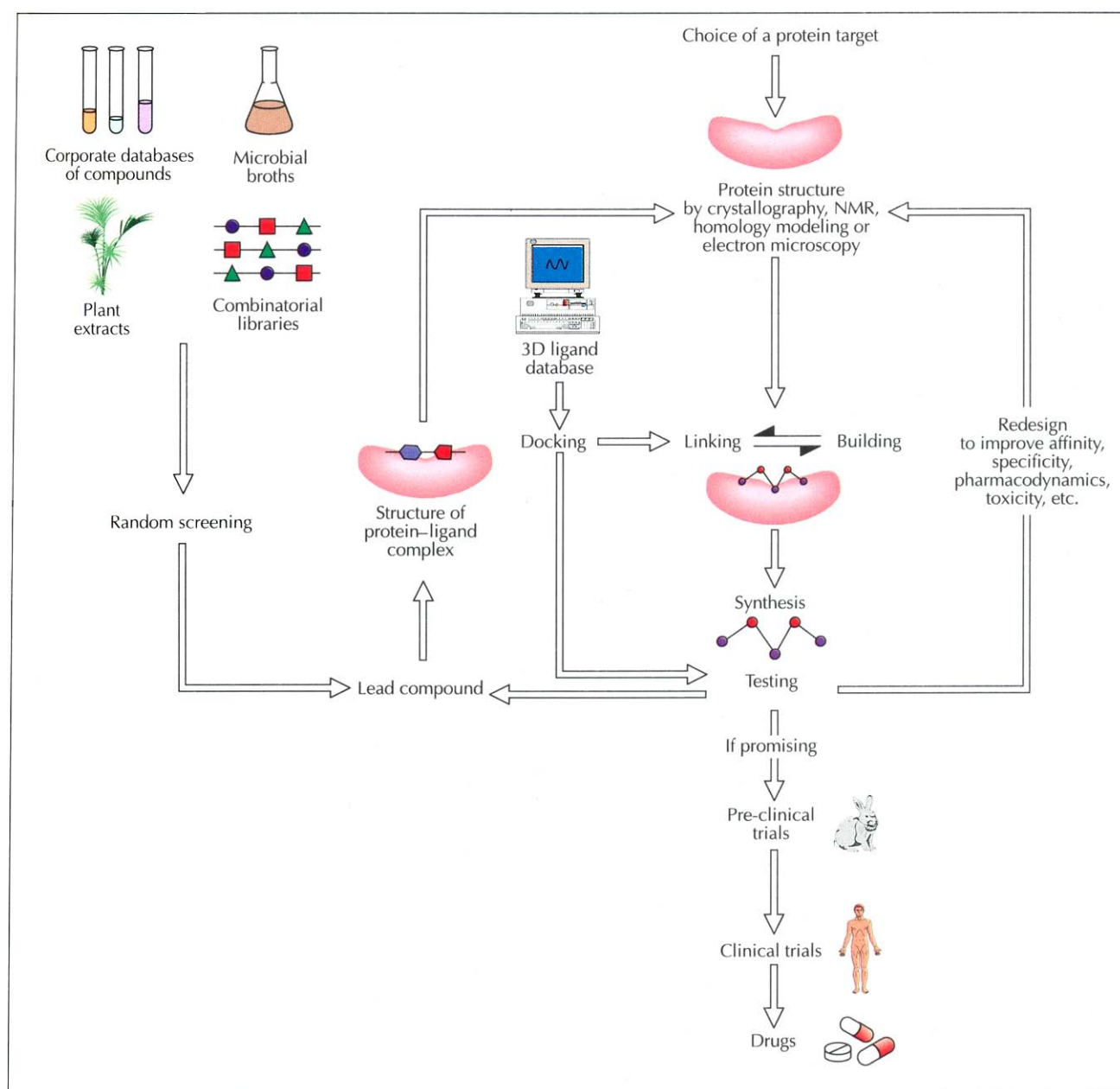

Fig. 1. Protein structure-based drug design cycle. Lead compounds originate from either random screening of a few hundred thousand compounds or from design. In the latter case, synthesis can be bypassed by using docking of compounds available commercially or in-house. Design is the result of docking, linking and building, or any combination of the three. Due to the imperfections of computer scoring, only about 2% of the designed compounds pass the first criterion to become a lead, namely having micromolar affinity. Verification of the structure of the protein–lead complex is essential. New rounds of structure-based design are then performed until a promising compound shows up for pre-clinical trials. At this stage the structure is still useful: knowledge of the essential protein–ligand interactions dictates where structural modifications to improve the pharmacodynamic properties should not be made. After successful clinical trials a new drug is born.

At least three additional criteria have to be taken into account in the inhibitor design cycle:

(d) chemical stability;

(e) sufficient solubility in water for inhibition tests and structural studies;

(f) ease of synthesis, including the avoidance of chiral centers and of 'dead-end leads' (i.e. compounds which are synthetically not easily amenable to many variations).

Some of these rules and criteria have been incorporated in a number of computer programs for automatic protein structure-based inhibitor design. The methods of these computer programs for protein structure-based ligand design can be grouped as: docking, linking or building of methods (Fig. 2). Only two of these programs came to light in the eighties, but at least ten new programs have been introduced in the last four years. A virtual explosion of methodologies is under way.

Docking algorithms

Three different strategies are currently in use for docking ligands on a target protein surface: optimal positioning of small chemical entities of molecules, the functional groups; searching for a subgraph isomorphism in a negative image description of the receptor, and Monte Carlo docking of complete molecules. The program GRID [19] is an example of the first strategy. It places functional groups, called probes, (e.g. amino, carboxylate, methyl moieties) at regularly-spaced lattice points in an active site and evaluates their interaction energy with the protein by means of empirical potential energy functions. Creating ligands from the favorable probe positions is left to the user's inspiration. The use of grids for energy calculations has since been incorporated in other design programs like AUTODOCK [20], LEGEND [21], and GroupBuild [22] (see below). Closely related to GRID is MCSS where thousands of copies of functional groups are simultaneously but independently positioned optimally on the protein surface by a molecular dynamics protocol [23].

A typical representative of the second strategy for docking is the program LUDI [24]. First, it describes the protein site of interest in terms of a collection of complementary hydrogen bond donor and acceptor vectors, and lipophilic points, mathematically referred to as a graph. Subsequently, LUDI retrieves matching ligands from a database by a subgraph isomorphism algorithm. This means that the vectors describing the hydrogen bond donor and acceptors, and the lipophilic points of the ligands are fitted optimally onto the graph describing the protein. Other programs use a different description of the protein. CLIX [25] characterizes the site of interest by GRID potential energy maps. DOCK casts a negative image of the receptor as a set of spheres, and essentially therefore per-

forms shape matching [26]. A newer version of this program incorporates a GRID-like energy evaluation which requires charges to be assigned to all atoms of the database ligands [27]. It should be mentioned that the subgraph isomorphism techniques employed in the above programs are not new and are widely used in traditional pharmacophore matching programs like ALLADIN [28], FOUNDATION [29], MACCS-3D [30], ChemDBS-3D [31], and CATALYST [32] (for an excellent review see [33]).

A third method for docking is Metropolis Monte Carlo searching combined with a simulated annealing scheme. Energy evaluations from precalculated GRID-like potential energy maps speed up the procedure. The ligand is kept rigid in the program BOXSEARCH [34], while AUTODOCK [20] also alters the conformation of the ligand. Flexible ligand docking is computationally very demanding but may become more popular in the future as computing power continues to increase.

Three-dimensional ligand databases

For many of the computer programs developed for lead discovery or inhibitor optimization, large collections of three-dimensional structures of low molecular weight compounds are required as essential input. The basic source for experimentally-determined structures is the Cambridge Structural Database (CSD), containing over 110 000 organic molecules [35]. Its usefulness is limited, however, since most of its compounds are not readily available for carrying out inhibition tests. It is more practical to use databases of commercially available compounds such as fine chemicals (about 100 000 molecules), medicinal compounds and drugs (for a review, see [36]). All of these contain models of the compounds obtained by structure-generation programs [37], that convert two-dimensional connection tables into three-dimensional structures.

CONCORD [38] is the most popular of these programs, and has recently been used to convert 5 000 000 organic molecules of the Chemical Abstracts Service Registry file [39]. All currently available databases of three-dimensional structures of small molecules are essentially limited to one conformation per substance. It is to be expected that this limitation will be overcome to some extent in coming years — but the number of different conformations to be considered quickly becomes astronomical, as we will see below.

Linking recognition fragments

A useful strategy for obtaining powerful inhibitors is to incorporate different functional groups or small molecules bound to a target protein into a single, larger molecule. The larger molecule loses less entropy upon binding than the sum of the fragments and hence is
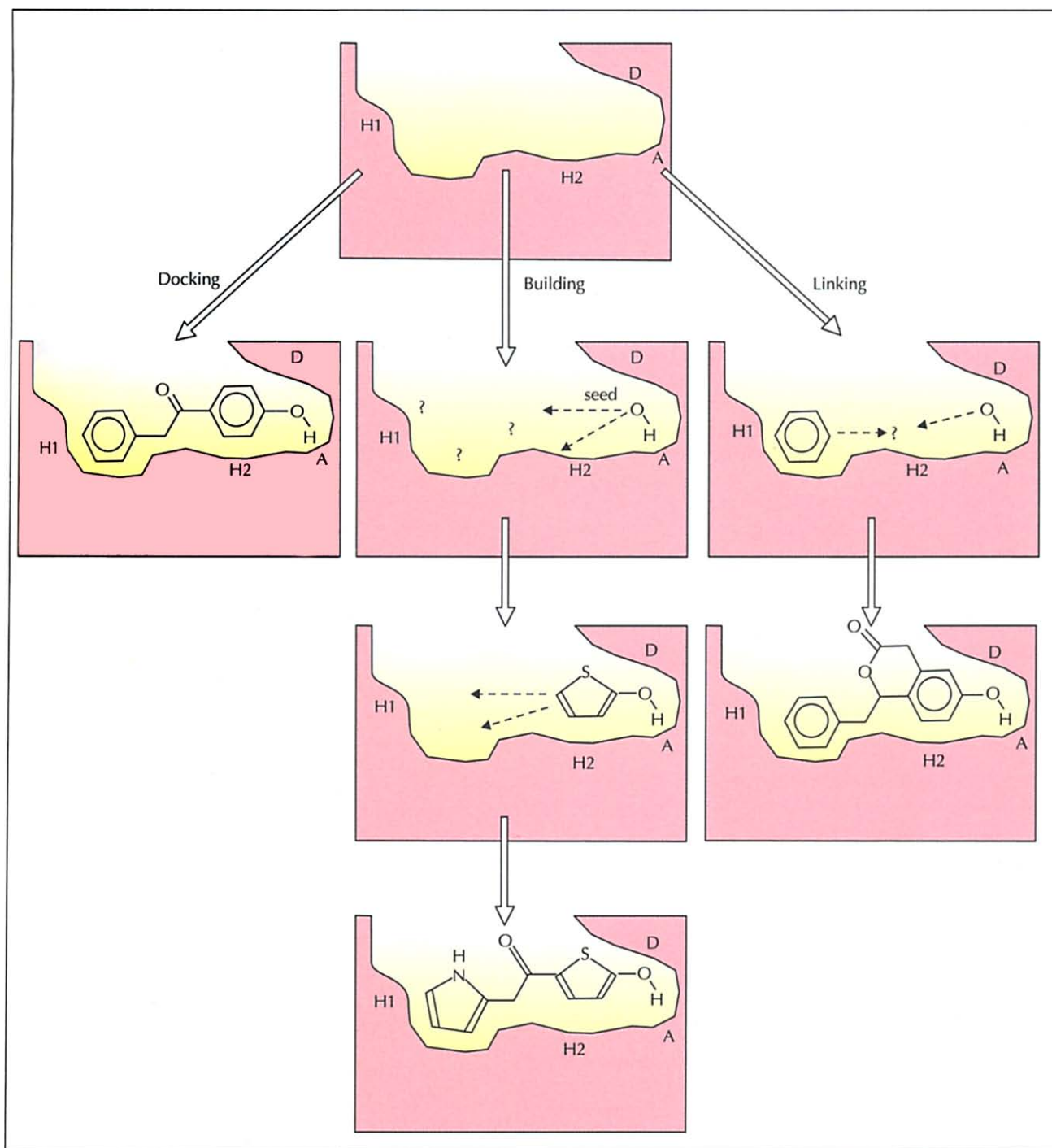
Fig. 2. Methods for protein structure-based inhibitor design. All methods first characterize the target site in terms of shape and the presence of specific surface properties, e.g. hydrophobic sites (H), hydrogen bond donors (D) and acceptors (A). Subsequently, docking, building or linking algorithms are applied. In docking, molecules are retrieved from a huge database and evaluated for complementarity to the target site. Usually, only one ligand conformation is tested because of computational intractability. Building starts from a highly complementary fragment, either found by docking or known from a previous protein–ligand structure. This fragment, called the 'seed', is appended with a myriad of different fragments, which each in turn can be substituted further. This combinatorial explosion, which is also the hallmark of the linking method, is contained by pruning techniques, of which the most recent ones are Monte Carlo methods and genetic algorithms. In linking, highly complementary fragments are linked into one molecule.

likely to have a higher affinity. In addition, the combination molecule will have a higher specificity than the separate functional groups. Linking different fragments together is not easy, however, since the optimal position and orientation of the fragments must be largely maintained while the linker must be chemically feasible. Moreover, the linker should be quite rigid, since the more rigid the linker is, the less rotational entropy is lost upon binding, giving higher affinity. Obviously, if the linker makes additional favorable interactions with

the target protein this further enhances the affinity as well as specificity. Hence, the problem of finding optimal linkers is by no means trivial. CAVEAT [40] tries to find a suitable cyclic linker from external databases, while NEWLEAD [41] builds a linker from an internal library of fragments. LUDI [24] can also be run in a link mode. The major benefit of these programs for automatic linking is that they show the medicinal chemist many alternatives for positioning the same key fragments.

tein site of interest. Completely novel molecules are also generated by so-called 'genetic algorithms' that cycle back and forth between making random modifications of a two-dimensional description of a ligand and evaluating the fit of the next generation of molecules at the three-dimensional level in the active site of the target protein (JM Blaney, D Weininger, and JS Dixon, personal communication). Such programs are idea-generating tools which show the ligand design team entirely novel ways to fill up active sites with small molecules.

Building or extending ligands

All ligand building methods rely heavily on one or more of the docking methods described above or on an experimentally-determined structure. They start from any docked chemical moiety or a part of a known inhibitor, usually called the 'core' or 'seed', and add atoms or fragments, one at a time, to build the inhibitor step by step. Because each step in this process generates thousands of possibilities, a method to contain the combinatorial explosion is needed. One possibility is to retain only the best solution at each step. But, in an excellent inhibitor not all fragments need to make optimal interactions with the protein — suboptimal arrangements of some fragments may simply allow other fragments to interact better with the target. Therefore, several tree-search methods have abandoned the best-first algorithm and choose at random a fragment from the top 25 % scoring ones. A second inherent problem of any building method is the possibility of generating chemically unstable or reactive compounds. Checking newly formed bonds against a table of disallowed connected atom pairs and triplets is the usual method to prevent the generation of groups such as peroxides or acetals.

Building ligands in a protein environment is carried out either atom-by-atom or fragment-by-fragment. The programs GenStar [42] and LEGEND [21] are examples of the first method. A possible drawback of GenStar is the limitation of its atom repertoire to sp^3 carbons. Typical representatives of fragment-wise builders are GROW [43] and GroupBuild [22]. GROW was originally developed to build peptides but has recently been equipped with a general template library [44]. It was successfully used to grow a micromolar inhibitor of renin [43]. A fragment-wise builder similar to GenStar is SPROUT [45]. Based on earlier ideas by Lewis and Dean [46,47] it first generates molecular skeletons by joining templates then carries out a chemical functionalization.

Building methods suffer from serious bias, as in each step of their tree-search algorithms a particular fragment is chosen, limiting the possibilities in the next step. A radically new program, CONCEPTS, eliminates this bias [48]. In this strategy, atom identities and their connectivities are subjected to Metropolis Monte Carlo; intermediate molecular dynamic runs explore the pro-

Flexible ligands and flexible proteins

The average organic molecule has eight rotatable bonds [37]. As a consequence, if we assume that 30° increments in dihedral angles define different conformations, then for just one average molecule about 430 million conformations have to be examined. Most of the docking algorithms for entire molecules make little attempt to address the issue of ligand flexibility. An exception is AUTODOCK, a Monte Carlo program, but testing such large numbers of conformations makes the CPU-time requirements for docking large databases unacceptable. Databases of conformationally restricted molecules may provide an avenue to expand and explore the power of Monte Carlo methods.

Instead of testing enormous numbers of conformations, one can investigate whether the conformation of a ligand can be altered to satisfy the constraints of a protein-binding site. Such algorithms are presently being tested on simple 3–5 point pharmacophores, instead of in a full three-dimensional protein context. From a comparison of five flexible pharmacophore docking methods it appears that genetic algorithms and directed 'tweak-search' methods are largely superior to distance-geometry, systematic-search, or random-search methods [49].

Not only are ligands flexible, so are proteins. A typical example is triosephosphate isomerase where the so-called flexible loop of the enzyme is closed in the presence of inhibitors and open otherwise. Furthermore, the conformation of the catalytic glutamate varies depending on which inhibitor is bound, as does the water structure [50]. Recently, inhibitor binding to the open loop conformation has also been observed [51]. The possibility of unexpected conformational changes of the protein upon ligand binding is one of the reasons that experimental verification of predicted binding modes of new ligands is a key step in a cyclic structure-based inhibitor design process.

An intermediate approach to incorporate protein flexibility is to allow for flexible side chains while keeping the backbone fixed. The 'Dead End Elimination' and 'A*' algorithms have recently been shown to be useful for that purpose [52].

It is of course by no means necessary that all possible conformations of the ligands in the database be

checked *versus* all possible conformations of the protein molecule. The key issue is whether a sufficient number of ligand conformations from a sufficiently large reservoir of small compounds are tested for their fit *versus* a sufficiently large number of conformations of the protein. Promising results have been obtained with DOCK, even though only a tiny fraction of all possible conformations of protein and ligands were investigated. This holds great promise for the future, as new algorithms and the continuous increase in computer power will allow the testing of more conformations.

Scoring — a serious problem

Because of the very large number of potential ligands generated by docking, building or linking strategies, it is essential to be able to estimate the free energy change of the protein–ligand interaction. For this, an efficient scoring algorithm is required. Three steps can be distinguished in the process of a ligand binding to a protein in solution:

(a) both protein and ligand have to be brought from their conformation free in solution to the conformation they will adopt in the complex;

(b) the surfaces of both protein and ligand that will be buried in the complex have to be desolvated;

(c) the ligand must be properly oriented and translated to interact with the protein and to form a complex.

For the energy evaluation of step (a) the flexibility of the protein can often be ignored; in other words, all ligands are designed to bind to the same static protein conformation. A full conformational analysis for the ligand in solvent is required, however. Storing a representative ensemble of low energy conformers together with their energies allows the calculation of enthalpy and entropy changes, in other words, the strain and loss of conformational entropy upon adopting the bound conformation. The loss of internal rotational entropy is in practice estimated from the number of rotatable bonds that are frozen out upon complex formation. Step (b) is generally approximated by empirically calculating the solvent-accessible surface contributions of various functional groups buried in the complex and translating this buried accessible surface into changes of free energy of hydration. The loss of overall translational and rotational entropy in step (c) is very similar for all ligands. The enthalpy change of forming the protein–ligand complex in step (c) can be estimated from a force-field calculation.

None of the described inhibitor design programs calculates all components of the free energy of binding outlined above. All of them estimate the enthalpy change of step (c), however, sometimes by considering only some of the various contributions (van der Waals interactions, hydrogen bonds, and so on). Loss of internal rotational entropy is only taken into account by LUDI

and SPROUT. Desolvation is scored by DOCK, Group-Build and GROW. In GROW, the enthalpy change of step (a) is also evaluated. Scoring is receiving considerable attention. For example, DOCK initially used shape fitting, but now has a considerably more sophisticated scoring function [27]. It is to be expected that in the next years considerable progress in developing reasonably reliable scoring algorithms will be made.

Progress and successes

It is beyond the scope of a review like this to give a comprehensive list of published protein structures which are of relevance for the potential design of new drugs. Max Perutz, in his recent book [53], shows that there are numerous protein structures available to start a drug design process. And every week, if not every day, new structures are added to the list, ranging from small proteins like cytokines and growth factors to multi-enzyme complexes and viruses. The following selection of results is intended to give a sense of the range of projects currently in progress.

The structures of picornaviruses, like rhinovirus and poliovirus, form a basis for the development of tightly-binding ligands which prevent infection — presumably by blocking the uncoating process of the virus. For the rhinoviruses, responsible for the common cold, the great variability of the virus poses a tremendous challenge for the design of a broad-spectrum drug. This variability is seen not only in the antigenic regions but also in the ligand-binding pocket [54]. As well as the potential for designing a drug that inhibits viral replication, structure-based drug design offers the possibility of designing ligands that stabilize the virus coat, with potential application for poliovirus vaccines [55]. Full implementation of the polio vaccination program in third world countries is hampered by the thermolability of the virus used in the vaccine, which requires the vaccine to be kept at low temperatures, without a single interruption, from manufacturer to injection in the patient. A ligand which would stabilize the virus might be of tremendous value in allowing the vaccine to be exposed to higher temperatures and thus to be more easily used in tropical areas.

Dihydrofolate reductase (DHFR) was the first target protein solved in complex with a drug — even though the complex was a cancer drug, methotrexate, bound to a bacterial enzyme, it was an important first step [56]. Three-dimensional structures of dihydrofolate reductase have been the basis for the design of several improved inhibitors [57]. Both DHFR and thymidylate synthase (TS), another enzyme involved in folate metabolism, are excellent targets for drug design since they are crucial for pyrimidine and purine synthesis. In the search for new cancer chemotherapeutics a TS inhibitor with $K_i = 960\,nM$ was rationally changed into one with $K_i = 15\,nM$; the compound is now in clinical evaluation [58]. Recently, the three-dimensional struc-

ture of a bifunctional DHFR–thymidylate synthase from the protozoon parasite *Leishmania major* has been unravelled [59], providing an opportunity for a two-pronged attack for drug design for the treatment of leishmaniasis.

A stunning example of successful improvement of a lead compound has been published in the case of influenza virus neuraminidase. The investigators used GRID to find a favorable position for adding a positively charged substituent to the sialidase inhibitor 2-deoxy-2,3-didehydro-D-*N*-acetylneuraminic acid (Neu5Ac2en) [60]. This modified saccharide not only binds over three orders of magnitude better than the parent compound, but is also effective *in vitro* and *in vivo* against virus infection.

Structural studies on elastase have revealed the erratic mode of binding of closely related inhibitors. Three modifications of the same parent ligand molecule led to surprisingly different binding modes of the three daughter molecules [61].

Thrombin provides an example where natural compounds have been used as leads in a structure-based drug optimization process. The three-dimensional structure of the leech inhibitor hirudin has been the basis for the development of compounds such as 'hirulog', a potent inhibitor of thrombin [62]. Fragments of the natural target of thrombin, fibrinogen, have also been incorporated into high-affinity ligands [63]. Surprisingly, oligonucleotides can also be found that inhibit the activity of thrombin. The binding mode has been unravelled in a three-dimensional structure [64] and is likely to allow the development of yet another family of thrombin inhibitors.

Knowledge of the three-dimensional structure of HIV protease, a small but crucial protein from the AIDS virus, has led to structure-based design of nanomolar, bioavailable, non-peptide inhibitors [65]. Few proteins have been crystallographically studied more frequently than this protease — there are now very many known inhibitor–protease complexes. Interestingly, this goal-oriented research also provides an extremely precious database of known structures of protein–ligand complexes with associated binding constants. This is a gold-mine for the fundamental research of theoretical biophysicists, and will allow them to sharpen their theoretical and computational tools. Eventually this will of course be beneficial for improving the scoring tools in the ligand design process.

Other success stories for structure-based ligand design include ligands for such diverse proteins as streptavidin, purine nucleoside phophorylase, and various proteases. Although streptavidin is not a drug target, the discovery of a peptide ligand for streptavidin by library screening procedures is most illuminating, since the peptide binds in an entirely different mode from that of the natural ligand, biotin [66].

In vitro, acyclovir is a molecule with great potential for treating T-cell leukemias and autoimmune diseases, but it is limited *in vivo* by its inability to penetrate cells and by its chemical and enzymatic instability. Using the structure of purine nucleoside phosphorylase, it has proved possible to design a nanomolar membrane-permeable inhibitor [67].

In the case of essential proteases of malaria and schistosoma, the structures of the proteases of interest are not available. Since this class of proteins is so well understood, however, it was possible to construct protein models by 'homology modeling' and use these models for inhibitor design. Screening of 83 molecules from a DOCK run yielded two micromolar inhibitors [68].

Selective inhibitor design is yet another opportunity offered by structural knowledge. A search for new drugs for the treatment of sleeping sickness exploited the differences between the structures of human and trypanosomal glyceraldehyde-3-phosphate dehydrogenase. Starting from adenosine as a lead, a 45-fold increase in inhibition of the parasite enzyme was achieved while the new compound was a poorer inhibitor of the human enzyme than the lead (CLMJ Verlinde, *et al*, & WGJ Hol, unpublished data) (Fig. 3).

High-resolution NMR structures of medically important proteins are gradually appearing in the literature. An excellent example is the structure of FK506-binding protein (FKBP) complexed with the immunosuppressant ascomycin. The quality of such NMR structure determinations is considerable, as evidenced by a mere 0.8 Å root mean square deviation for all non-hydrogen atoms when compared with the FKBP complex with the highly analogous immunosuppressant FK506 [69].

Most of the examples we have given involve enzymes as targets. However, it is to be expected that the numerous structures of cytokines, growth factors and hormones and their interaction with receptors (see, for example, [70,71]) form new starting points for the design of low molecular weight compounds modulating signal transduction. Such compounds may become important tools in controlling inflammation, suppressing the immune system or controlling cell growth. The information from the large number of DNA-binding proteins which have been structurally characterized will one day be used to regulate gene expression by low molecular weight compounds. Also, many proteins of the immune system are being fully characterized (for example, MHC class I [72] and class II [73] molecules) and form yet another starting point for new drug discovery programs.

Drug resistance in infectious diseases and cancer

For infectious diseases the development of a single therapeutically useful compound is by no means the end of the story. Such a compound, when used on a large scale, will almost always lead to the occurrence
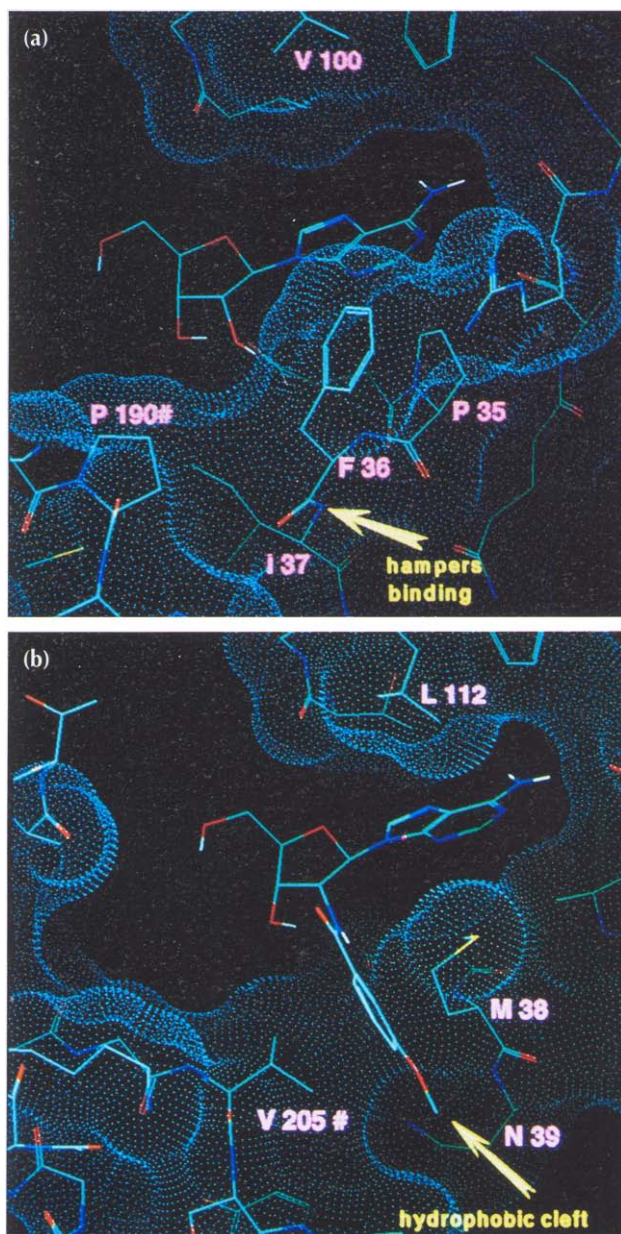
Fig. 3. Selective inhibitor design for blocking trypanosomal glycolysis. (a) Adenosine part of the nicotinamide adenine dinucleotide-binding region of human glyceraldehyde-3-phosphate dehydrogenase (GAPDH). (b) Equivalent view of glycosomal GAPDH from *Trypanasoma brucei*. Here, modeling shows how 2'-deoxy-2'-(3-methoxybenzamido)adenosine fills up a hydrophobic cleft, which is largely inaccessible in the human enzyme. The new inhibitor is 45 times more potent than adenosine and has hardly any effect on the human enzyme (CLMJ Verlinde, *et al.*, & WGJ Hol, unpublished data).

of pathogens which have cleverly developed one or more methods to avoid the harmful effects of the drug. Resistance is also a formidable problem in the case of cancer. The processes used by resistant organisms and cancer cells to avoid the harmful effects of a drug vary widely [74]. For example, the pathogen may mutate amino acids in the target protein. A special and discomforting example is HIV reverse transcriptase mutants which are simultaneously resistant to AZT, dideoxyi-

nosine and nevirapine [75]. Genes that affect sensitivity to a drug may be amplified, for example to increase the rate at which drugs are pumped out of the cell, leading to multi-drug resistance, a common phenomenon in cancer therapy [76]. In chloroquine-resistant malaria strains the gene for such a pump, related to that of cancer cells, is amplified [77]. Alternatively, the pathogen may recruit an entirely different enzyme to perform the same task as the original target enzyme. A typical example is found in some trimethoprim-resistant *Klebsiella* strains that possess a plasmid-encoded DHFR [78]. Drug metabolism can also be affected; for example, the β-lactamases destroy penicillins and cephalosporins. Recently, their mechanism has been unravelled on the basis of a crystal structure [79]. In the extreme, the pathogen may reorganize an entire enzymatic pathway, so that the drug target is simply no longer present. For example, enterococci became resistant to vancomycin by incorporating ester bonds rather than amide bonds in their cell wall, preventing the formation of a crucial hydrogen bond with a carbonyl of vancomycin [80].

The constant threat of resistance is a major source of concern. The disappointing battle against malaria is a tragic example of the recovery of a disease which once was thought to be on the road to disappearance. There exist tuberculosis strains that are simultaneously resistant to no less than nine different drugs — a truly frightening discovery [81,82].

The only way we will be able to combat resistant pathogens with new drugs with long-lasting utility is to develop cocktails containing multiple compounds acting in diverse ways. Thus, the chances of development of resistance can be decreased, in particular for diseases where the lack of compliance is greatest. Multi-drug therapy is very successful in the case of leprosy where three compounds are administered simultaneously [83]. This serves as an example for future approaches.

Conclusion and outlook

There are several potential drugs far advanced in clinical trials which are the result of structure-based design. In view of the large number of projects going on worldwide it is likely that many more will follow. Several may reach clinics and patients in the not too distant future.

As we discussed, more efficient drug design now requires the development of computer programs to cope with flexibility of ligands and proteins, and accurate ways of scoring interactions. Membrane proteins have so far been largely ignored in structure-based drug design processes since so few structures of membrane proteins are known. Recently, however, the structure of prostaglandin synthase, a membrane-associated protein which is the target of aspirin, has been reported [84]. High-resolution electron microscopy [85] offers hope for the future structural understanding of the important

membrane receptors of neurotransmission at atomic resolution.

In future we may also see further developments in computer programs which predict the effect of blocking an enzyme in a pathway on the flow of metabolites — or signals — through the pathway. Yet another area is the incorporation of toxicology data in knowledge-based systems. Eventually we may see the sequencing of entire genomes of pathogens, followed by the selection of target proteins which are either absent in the host or very different from analogous proteins in the host. Combining over-expression, structure determination and screening of libraries and databases with genome sequencing projects can give entirely new stimuli to the developments of new therapeutic agents.

So many activities are going on that it is hard to keep up with all developments on so many frontiers. We hope that this review has at least given a flavor of the excitement in the field, results obtained and challenges ahead.

References

1. Hol, W.G.J. (1986). Protein crystallography and computer graphics — toward rational drug design. *Angew. Chem.* 25, 767–778.
2. Schertler, G.F.X. (1992). Overproduction of membrane proteins. *Curr. Opin. Struct. Biol.* 2, 534–544.
3. Guide to protein purification (1990). *Methods Enzymol.* 182.
4. Pardi, A. (1992). Isotope labelling for NMR studies of biomolecules. *Curr. Opin. Struct. Biol.* 2, 832–835.
5. Boyd, J., Soffe, N. & Campbell, I. (1994). NMR at very high fields. *Structure* 2, 253–255.
6. Pflugrath, J.W. (1992). Developments in X-ray detectors. *Curr. Opin. Struct. Biol.* 2, 811–815.
7. Ealick, S.E. & Walter, R.L. (1993). Synchrotron beamlines for macromolecular crystallography. *Curr. Opin. Struct. Biol.* 3, 725–736.
8. Watenpaugh, K.D. (1991). Macromolecular crystallography at cryogenic temperatures. *Curr. Opin. Struct. Biol.* 1, 1012–1015.
9. Macromolecular Structures 1993 (1993). (Hendrickson, W.A. & Wüthrich, K., eds), Current Biology Ltd., London.
10. Kuntz, I.D. (1992). Structure-based strategies for drug design and discovery. *Science* 257, 1078–1082.
11. Hoess, R.H. (1993). Phage display of peptides and protein domains. *Curr. Opin. Struct. Biol.* 3, 572–579.
12. Zuckermann, R.N. (1993). The chemical synthesis of peptidomimetic libraries. *Curr. Opin. Struct. Biol.* 3, 580–584.
13. Famulok, M. & Szostak, J.W. (1992). *In vitro* selection of specific ligand-binding nucleic acids. *Angew. Chem.* 31, 979–988.
14. Bunin, B.A. & Ellman, J.A. (1992). A general and expedient method for the solid-phase synthesis of 1,4-benzodiazepine derivatives. *J. Am. Chem. Soc.* 114, 10997–10998.
15. Gallop, M.A., Barrett, R.W., Dower, W.J., Fodor, S.P.A. & Gordon, E.M. (1994). Applications of combinatorial technologies to drug discovery. 1. Background and peptide combinatorial libraries. *J. Med. Chem.* 37, 1233–1251.
16. Banner, D., *et al.*, & Wirtz, B. (1993). Serine proteases: 3D structures, mechanisms of action and inhibitors. In *Perspectives in Medicinal Chemistry. Proceedings of the XIIth International Symposium on Medicinal Chemistry.* (Testa, B., Kyburz, E., Fuhrer, W. & Giger, R., eds), pp. 27–43, Verlag Chimica Acta, Basel.

17. Hermans, J. (1993). An editorial comment: the limits of simulations. *Proteins* 17, ii.
18. Kollman, P.A. (1994). Theory of macromolecule–ligand interactions. *Curr. Opin. Struct. Biol.* 4, 240–245.
19. Goodford, P.J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules *J. Med. Chem.* 28, 849–857.
20. Goodsell, D.S. & Olson, A.J. (1990). Automated docking of substrates to proteins by simulated annealing. *Proteins* 8, 195–202.
21. Nishibata, Y. & Itai, A. (1993). Confirmation of usefulness of a structure construction program based on three-dimensional receptor structure for rational lead generation. *J. Med. Chem.* 36, 2921–2928.
22. Rotstein, S.H. & Murcko, M.A. (1993). GroupBuild: a fragment-based method for *de novo* drug design. *J. Med. Chem.* 36, 1700–1710.
23. Miranker, A. & Karplus, M. (1991). Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins* 11, 29–34.
24. Böhm, H.-J. (1992). The computer program LUDI: a new method for the *de novo* design of enzyme inhibitors. *J. Comput. Aided Mol. Des.* 6, 61–78.
25. Lawrence, M.C. & Davis, P.C. (1992). CLIX: a search algorithm for finding novel ligands capable of binding proteins of known three-dimensional structure. *Proteins* 12, 31–41.
26. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. & Ferrin, T.E. (1982). A geometric approach to macromolecule–ligand interactions. *J. Mol. Biol.* 161, 269–288.
27. Meng, E.C., Soichet, B.K. & Kuntz, I.D. (1992). Automatic docking with grid-based energy evaluation. *J. Comput. Chem.* 13, 505–524.
28. Van Drie, J.H., Weininger, D. & Martin, Y.C. (1989). ALLADIN: an integrated tool for computer-assisted molecular design and pharmacophore recognition from geometric, steric, and substructure searching of three-dimensional molecular structures. *J. Comput. Aided Mol. Des.* 3, 225–254.
29. Ho, C.M.W. & Marshall, G.R. (1993). FOUNDATION: a program to retrieve all possible structures containing a user-defined minimum number of matching query elements from three-dimensional databases. *J. Comput. Aided Mol. Des.* 7, 3–22.
30. Güner, O.F. & Dumont, L.M. (1990). 3D searching in computer-aided drug design. In *Pharmaceutical Manufacturing International 1991.* (Barber, M.S. & Barnacal, P.A., eds), pp. 65–68, Sterling Publications, Cambridge.
31. Murrall, N.W. & Davies, E.K. (1990). Conformational freedom in 3-D databases. 1. Techniques. *J. Chem. Inf. Comput. Sci.* 30, 312–316.
32. Sprague, P.W. (1991). CATALYST: a computer aided drug design system specifically designed for medicinal chemists. In *Recent Advances in Chemical Information - Proceedings of the 1991 Chemical Information Conference.* (Coller, H., ed), pp. 107–111, Royal Society of Chemistry, London.
33. Martin, Y.C. (1992). 3D database searching in drug design. *J. Med. Chem.* 35, 2145–2154.
34. Hart, T.N. & Read, R.J. (1992). A multiple-start Monte Carlo docking method. *Proteins* 13, 206–222.
35. Allen, F.H., *et al.*, & Watson, D.G. (1991). The development of versions 3 and 4 of the Cambridge Structural Database system. *J. Chem. Inf. Comput. Sci.* 31, 187–204.
36. Rusinko, A.III (1993). Commercially available databases of three-dimensional molecular structures. *Chem. Des. Auto. News* 8(9&10), 44–47.
37. Pearlman, R.S. (1993). Three-dimensional structures: how do we generate them and what can we do with them? *Chem. Des. Auto. News* 8(8), 3–15.
38. Pearlman, R.S. (1987). Rapid generation of high quality approximate 3-D molecular structures. *Chem. Des. Auto. News* 2(1), 5–6.
39. Fisanick, W., Cross, K.P., Forman, J.C. & Rusinko, A.III (1993). Experimental system for similarity and 3D searching of CAS registry substances. 1. 3D substructure searching. *J. Chem. Inf. Comput. Sci.* 33, 548–559.
40. Bartlett, P.A., Shea, G.T., Telfer, S.J. & Waterman, S. (1989). CAVEAT: a program to facilitate the structure-derived design of biologically active molecules. In *Molecular Recognition: Chemi-*

cal and Biological Problems. (Roberts, S.M., ed), pp. 182–196, Royal Society of Chemistry, Cambridge.

41. Tschinke, V. & Cohen, N.C. (1993). The NEWLEAD program: a new method for the design of candidate structures from pharmacophoric hypotheses. *J. Med. Chem.* 36, 3863–3870.

42. Rotstein, S.H. & Murcko, M.A. (1993). GenStar: a method for *de novo* drug design. *J. Comput. Aided Mol. Des.* 7, 23–43.

43. Moon, J.B. & Howe, W.J. (1991). Computer design of bioactive molecules: a method for receptor-based *de novo* ligand design. *Proteins* 11, 314–328.

44. Moon, J.B. & Howe, W.J. (1993). Automated receptor-based ligand design: development of the GROW program. In *Biomolecular Interactions: A Practical Workshop on Computational Approaches.* American Crystallographic Society, Buffalo, NY.

45. Gillet, V., Johnson, A.P., Mata, P., Sike, S. & Williams, P. (1993). SPROUT: a program for structure generation. *J. Comput. Aided Mol. Des.* 7, 127–153.

46. Lewis, R.A. & Dean, P.M. (1989). Automated site-directed drug design: the concept of spacer skeletons for primary structure generation. *Proc. R. Soc. Lond. [Biol.]* 236, 125–140.

47. Lewis, R.A. & Dean, P.M. (1989). Automated site-directed drug design: the formation of molecular templates in primary structure generation. *Proc. R. Soc. Lond. [Biol.]* 236, 141–162.

48. Pearlman, D.A. & Murcko, M. (1993). CONCEPTS: new dynamic algorithm for *de novo* drug suggestion. *J. Comput. Chem.* 14, 1184–1193.

49. Clark, D.E., Jones, G., Willett, P., Kenny, P.W. & Glen, R.C. (1994). *J. Chem. Inf. Comput. Sci.* 34, 197–206.

50. Noble, M.E.M., *et al.*, & Hol, W.G.J. (1991). The adaptability of the active site of trypanosomal triosephosphate isomerase as observed in the crystal structures of three different complexes. *Proteins* 10, 50–69.

51. Verlinde, C.L.M.J., *et al.*, & Opperdoes, F.R. (1992). Structure of the complex between trypanosomal triosephosphate isomerase and N-hydroxy-4-phosphono-butanamide: binding at the active site despite an 'open' flexible loop. *Protein Sci.* 1, 1578–1584.

52. Leach, A.R. (1994). Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* 235, 345–356.

53. Perutz, M. (1992). *Protein Structure. New Approaches to Disease and Therapy.* W.H. Freeman and Company, New York.

54. Kim, K.H., *et al.*, & Pevear, D.C. (1993). A comparison of the anti-rhinoviral drug binding pocket in HRV14 and HRV1A. *J. Mol. Biol.* 230, 206–227.

55. Caliguiri, L.A., McSharry, J.J. & Lawrence, G.W. (1980). Effect of arildone on modifications of poliovirus *in vitro. Virology* 105, 86–93.

56. Matthews, D.A., *et al.*, & Kraut, J. (1978). Dihydrofolate reductase from *Lactobacillus casei.* "X-ray structure of the enzyme methotrexate complex". *J. Biol. Chem.* 253, 6946–6954.

57. Kuyper, L.F., *et al.*, & Goodford, P.J. (1985). Receptor-bound design of dihydrofolate reductase inhibitors: comparison of crystallographically determined enzyme binding with enzyme affinity in a series of carboxy-substituted trimethoprim analogues. *J. Med. Chem.* 28, 303–311.

58. Webber, S.E., *et al.*, & Bartlett, C.A. (1993). Design of thymidylate synthase inhibitors using protein crystal structures: the synthesis and biological evaluation of a novel class of 5-substituted quinazolinones. *J. Med. Chem.* 36, 733–746.

59. Knighton, D.R., *et al.*, & Matthews, D.A. (1994). Structure and kinetic channelling in bifunctional dihydrofolate reductase–thymidylate synthase. *Nature Struct. Biol.* 1, 186–194.

60. von Itzstein, M., *et al.*, & Penn, C.R. (1993). Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* 363, 418–423.

61. Mattos, C., Rasmussen, B., Ding, X., Petsko, G.A. & Ringe, D. (1994). Analogous inhibitors of elastase do not always bind analogously. *Nature Struct. Biol.* 1, 55–58.

62. Qiu, X., *et al.*, & Fenton, J.W.II (1992). Structure of the hirulog 3–thrombin complex and nature of the S' subsites of substrates and inhibitors. *Biochemistry.* 31, 11689–11697.

63. Nakanishi, H., *et al.*, & Kahn, M. (1992). Peptide mimetics of the thrombin-bound structure of fibrinopeptide A. *Proc. Natl. Acad. Sci. USA* 89, 1705–1709.

64. Padmanabhan, K., Padmanabhan, K.P., Ferrara, J.D., Sadler, J.E. & Tulinsky, A. (1993). The structure of alpha-thrombin inhibited by a 15-mer single-stranded DNA aptamer. *J. Biol. Chem.* 268, 17651–17654.

65. Lam, P.Y.S., *et al.*, & Erickson-Viitanen, S. (1994). Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* 263, 380–384.

66. Webber, P.C., Pantoliano, M.W. & Thompson, L.D. (1992). Crystal structure and ligand binding studies of a screened peptide complexed with streptavidin. *Biochemistry* 31, 9350–9354.

67. Erion, M.D., *et al.*, & Montgomery, J.A. (1993). Structure-based design of inhibitors of purine nucleoside phosphorylase. 3. 9-Arylmethyl derivatives of 9-deazaguanine substituted on the methylene group. *J. Med. Chem.* 36, 3771–3783.

68. Ring, C.S., *et al.*, & Cohen, F.R. (1993). Structure-based design by using protein models for the development of antiparasitic agents. *Proc. Natl. Acad. Sci. USA* 90, 3583–3587.

69. Meadows, R.P., *et al.*, & Fesik, S.W. (1993). Three-dimensional structure of the FK506 binding protein/ascomycin complex in solution by hetronuclear two- and three-dimensional NMR. *Biochemistry* 32, 754–765.

70. De Vos, A.M., Ultoch, M. & Kossiakoff, A.A. (1992) Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* 255, 306–312.

71. Banner, D.W., *et al.*, & Lesslauer, W. (1993). Crystal structure of the soluble human 55 kD TNF receptor-human TNFβ complex: implications for TNF receptor activation. *Cell* 73, 431–435.

72. Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L. & Wiley, D.C. (1987). Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* 329, 506–512.

73. Brown, J.H., *et al.*, & Wiley, D.C. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364, 33–39.

74. Neu, H.C. (1992). The crisis in antibiotic resistance. *Science* 257, 1064–1072.

75. Larder, B.A., Kellam, P. & Kemp, S.D. (1993). Convergent combination therapy can select viable multi-drug resistant HIV-1 *in vitro. Nature* 365, 451–453.

76. Endicott, J.A. & Ling, V. (1989). The biochemistry of P-glycoprotein-mediated mutltidrug resistance. *Annu. Rev. Biochem.* 58, 137–171.

77. Wilson, C.M., Serrano, A.E., Wasley, A., Bogenschutz, M.P., Shankar, A.H. & Wirth, D.F. (1989). Amplification of a gene related to mammalian mdr genes in drug resistant *Plasmodium falciparum. Science* 244, 1184–1186.

78. Kraut, J. & Matthews, D.A. (1987). Dihydrofolate reductase. In *Biological Macromolecules & Assemblies. Volume 3: Active Sites of Enzymes.* (Jurnak, F.A. & McPherson, A., eds), pp. 1–71, John Wiley & Sons, New York.

79. Strynadka, N., *et al.*, & James, M.N.G. (1992). Molecular structure of the acyl-enzyme intermediate in β-lactam hydrolysis at 1.7Å resolution. *Nature* 359, 700–705.

80. Walsh, C.T. (1993). Vancomycin resistance: decoding the molecular logic. *Science* 261, 308–309.

81. Weiss, R. (1992). On the track of "killer" TB. *Science* 255, 148–150.

82. Bloom, B.R. & Murray, C.J.L. (1992). Tuberculosis: commentary on a reemergent killer. *Science* 257, 1055–1062.

83. Tropical diseases. Progress in Research 1989–1990 (1991). pp. 89–96, World Health Organization, Geneva.

84. Picot, D., Loll, P.J. & Garavito, R.M. (1994). The X-ray crystal structure of the membrane protein prostaglandin H synthase-1. *Nature* 367, 243–249.

85. DeRosier, D.J. (1993). Turn-of-the-century electron microscopy. *Curr. Biol.* 3, 690–692.

Christophe LMJ Verlinde and Wim GJ Hol, Department of Biological Structure, and Biomolecular Structure Program, SM-20, School of Medicine, University of Washington, Seattle, WA 98195, USA.