Journal of Computer-Aided Molecular Design, 6 (1992) 61–78
ESCOM

J-CAMD 143

The computer program LUDI: A new method for the de novo design of enzyme inhibitors

Hans-Joachim Böhm

BASF AG, Central Research, D-6700 Ludwigshafen, Germany

Received 27 May 1991
Accepted 16 August 1991

Key words: Enzymes; Enzyme inhibitors; Molecular modeling; Drug design; De novo design

SUMMARY

A new computer program is described, which positions small molecules into clefts of protein structures (e.g. an active site of an enzyme) in such a way that hydrogen bonds can be formed with the enzyme and hydrophobic pockets are filled with hydrophobic groups. The program works in three steps. First it calculates interaction sites, which are discrete positions in space suitable to form hydrogen bonds or to fill a hydrophobic pocket. The interaction sites are derived from distributions of nonbonded contacts generated by a search through the Cambridge Structural Database. An alternative route to generate the interaction sites is the use of rules. The second step is the fit of molecular fragments onto the interaction sites. Currently we use a library of 600 fragments for the fitting. The final step in the present program is the connection of some or all of the fitted fragments to a single molecule. This is done by bridge fragments. Applications are presented for the crystal packing of benzoic acid and the enzymes dihydrofolate reductase and trypsin.

1. INTRODUCTION

The design of new potent and selective enzyme inhibitors is one of the most important applications in contemporary rational drug design. A considerable number of three-dimensional (3D) structures of enzymes are now available and the probability of having the 3D structure of the target enzyme at one's disposal is steadily increasing as more and more protein structures are deposited in the Brookhaven Protein Data Bank (PDB) [1]. The knowledge of the 3D structure of the target enzyme can be used in the design of new enzyme inhibitors. In the past, this was usually done by simply displaying the structure on a computer graphics system in order to identify possible binding sites for an inhibitor molecule. If, in addition, 3D structures of one or more enzyme–inhibitor complexes are known, one can also try to identify the essential interactions in those structures and search for new binding sites which are not used by the previously known structures.

Possible binding sites may be positions where hydrogen bonds ('H-bonds') can be formed with the enzyme, or hydrophobic pockets in the enzyme structure which can be occupied, e.g. by hydrocarbon groups.

The visual inspection of the 3D structure of the active site is of great help to the chemist in generating ideas for new potent inhibitors. However, the efficacy of the computer-aided drug design process could be greatly improved by additional tools, which propose new molecules as possible ligands by an automatic or semiautomatic procedure [2]. Indeed, over the past 5 years a number of different approaches towards the 'De Novo' design of drugs have been proposed [3–6]. There are at least two possible routes one can take to find automatically a compound that fits into the active site and binds to the enzyme. The first possibility is to search through databases of known 3D structures such as the Cambridge Structural Database (CSD) [7] and to identify those entries which fit into the active site. The program DOCK developed by DesJarlais et al. [3] is an important contribution to this approach. The advantages of this approach are that the molecules retrieved from the database do exist and that they are fitted into the active site in a low-energy conformation. There is, however, the problem of conformational flexibility. The crystal structure is not the only low-energy conformation and there are well-known cases where the conformation of an inhibitor bound to the enzyme is different from the crystal structure conformation [8]. Moreover, the number and variety of structures in the CSD is certainly too small to provide a comprehensive range of possible 3D structures. Both limitations could at least in principle be overcome by performing a conformational analysis of the compounds in the active site, for example with distance geometry methods [9] and/or by using larger datasets e.g. by converting 2D structures to 3D structures with the programs CONCORD [10] or WIZARD [11]. It is clear, however, that such a procedure will be very demanding both in CPU time and in disc space.

The second approach is based on molecular fragments. The idea is to position molecular fragments into the active site in such a way that hydrogen bonds can be formed with the enzyme and hydrophobic pockets are filled with hydrophobic fragments. These fragments are then connected by suitable spacer fragments to form a single molecule. There are several potential advantages of this approach. It is very fast and due to the large number of possible combinations of fragments to be connected the variety of molecules that can be generated is enormous. The computer program LUDI [12] described in the present paper adopts this fragment-based approach.

There have been previous attempts to provide related tools for inhibitor design [5,6]. Most notably the program GRID developed by Goodford [5] calculates binding energies for a given probe (e.g. carbonyl, amide, aromatic carbon) with a given receptor molecule. In our view this program is a very important first step towards the De Novo design of drugs and is therefore used frequently. Although a very helpful and easy to use tool, there are sometimes difficulties in interpreting the results from a GRID run. It is not always straightforward to translate the calculated binding energies and spatial positions of the putative interaction sites into chemical structures.

Our present approach attempts to go one step further than the GRID program by positioning molecular fragments into the active site at suitable positions. The advantage of such an approach is that the positioned molecular fragments immediately provide the drug designer with ideas, how putative binding sites offered by the protein might be saturated by chemical structures and how the structural fragments might be linked together.

The next section describes the technical details of the computer program. Applications are presented in Section 3 for the crystal packing of benzoic acid and the enzymes dihydrofolate reductase (DHFR) and trypsin.

2. DESCRIPTION OF THE ALGORITHM

For a description of the present computer program it is useful to define the three terms interaction site, molecular fragment and bridge.

An *interaction site* is a position in space, which is not occupied by the enzyme, where an atom out of a functional group of an inhibitor can have favorable interactions with the enzyme. The present program distinguishes between 4 different types of interaction sites: lipophilic-aliphatic, L_{ali}, lipophilic-aromatic, L_{aro}, hydrogen donor, D-X, and hydrogen acceptor, A-Y. The aliphatic and aromatic interaction sites denoted as L_{ali} and L_{aro} are suitable to form hydrophobic interactions. The H-donor and H-acceptor sites denoted as D-X and A-Y are suitable to form H-bonds with the enzyme. As the H-bond is strongly directional, H-donor and H-acceptor sites are stored as vectors so as to also retain information about the orientation. These vectors are represented by the atom pairs D-X and A-Y. Thus, if a hydrogen donor interaction site D is detected (a position where a polar hydrogen can form a hydrogen bond with the enzyme), the optimal position of the putative atom X bound to D 1.00 Å away from D is stored as well. If a hydrogen acceptor site A is detected (a position where e.g. an oxygen or nitrogen atom can form a hydrogen bond with the enzyme), then a second position Y atom 1.23 Å away from A is stored as well. The particular lengths of the vectors were chosen to simplify the subsequent fitting of molecules containing N-H or O-H groups (which are fitted onto the donor interaction site D-X) or carbonyl groups C=O (which will be fitted onto the acceptor interaction site A-Y). The use of atom pairs for H-donor and H-acceptor interaction sites makes subsequent fragment fitting straightforward although some fragments will not use both positions; e.g. if a fragment containing an ether moiety is fitted onto an H-acceptor site, the ether oxygen is fitted onto the position A and the adjacent position Y is not used.

A *molecular fragment* is a small molecule or a functional group with ideal geometry. Examples of molecular fragments used by the present program are benzene or acetic acid. Of course, one can also use larger molecules such as peptides or substituted aromatic compounds as molecular fragments.

A *bridge* is very similar to a molecular fragment. It is also a small molecule with ideal geometry. The bridges are fitted between the molecular fragments in order to merge them into a single molecule.

The program consists of 3 modules BSITES, FRAGMT and BRIDGE. BSITES generates the potential interaction sites. FRAGMT selects molecular fragments and fits them onto the potential interaction sites. BRIDGE chooses bridge fragments in order to merge the molecular fragments into one molecule. A schematic flow diagram of the program is shown in Fig. 1. The complete program is written in standard Fortran 77 and comprises about 4000 lines of code. It was developed on a DEC VAX 11/785 but runs on Convex C210 and Silicon Graphics 4D computers without any changes. The program expects 5 types of input information:

(1) the cartesian coordinates of the enzyme in PDB format (it is necessary to provide also positions of the hydrogen atoms);
(2) a library of nonbonded contact distributions (see Section 2.1.2 below);
(3) a library of molecular fragments (PDB format);
(4) a library of bridge fragments (PDB format);
(5) the position of the active site (the cartesian coordinates of a single point in the active site and a cutoff radius).

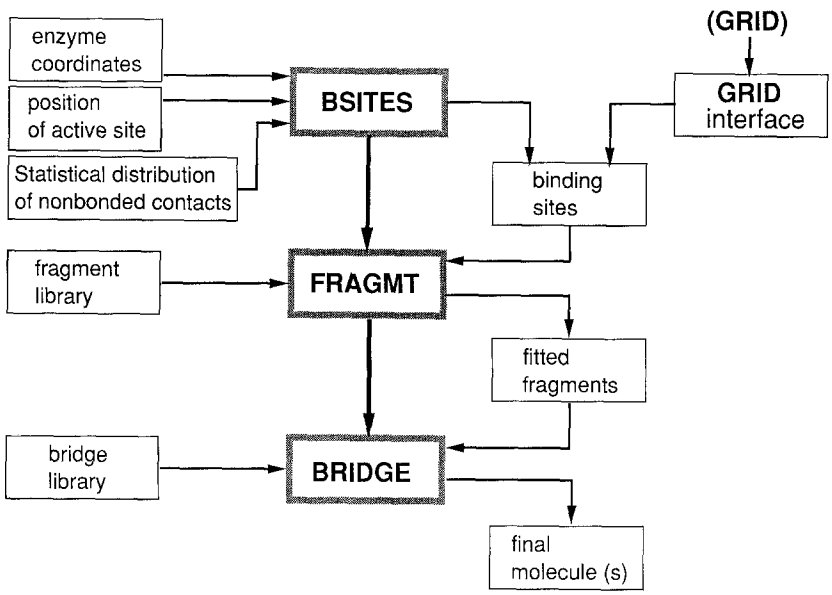

Fig. 1. Schematic flow diagram of the computer program LUDI. The module BSITES generates the interaction sites. FRAGMT fits molecular fragments onto the interaction sites. Finally, the module BRIDGE attempts to connect the fitted fragments to a single molecule.

As LUDI also requires the positions of the polar hydrogen atoms which usually cannot be determined by X-ray crystal structure analysis of proteins, they have to be calculated, e.g. in idealized geometry by one of the generally available molecular modelling packages. One also has to define the charge states of the amino acids Asp, Glu, Lys, Arg and His beforehand. Water molecules can be included in the input file, if they are thought to be important, e.g., in mediating an H-bond between the enzyme and the ligand.

We now describe the 3 program modules in more detail.

2.1. Calculation of the potential interaction sites (BSITES)

The program works in 3 steps. In the first step it calculates the potential interaction sites, as follows. The program classifies the protein atoms on the basis of their atom type and their connectivity into the following 4 categories:

(1) H-donor: all hydrogens, which are bound to N or O;
(2) H-acceptor: all nitrogen atoms which are not bound to hydrogen and all oxygen atoms;
(3) aromatic: carbon atoms occurring in the aromatic rings of Phe, Tyr, His and Trp;
(4) aliphatic: all aliphatic carbon atoms.

LUDI offers the possibility to use 3 different approaches to generate the interaction sites:

(1) rule based (see below);
(2) use of statistical hydrogen bond distributions and nonbonded contact patterns derived from the CSD [7];
(3) use of the output from the program GRID [5].

2.1.1. Rule-based generation of the interaction sites

The potential interaction sites are generated by the following rules:

(1) If the enzyme atom is an H-donor then 4 acceptor interaction sites A-Y are positioned with the following geometry $R_{H..A} = 1.9$ Å, $R_{A-Y} = 1.23$ Å, $\angle_{O/N-H..A} = 180°$, $\angle_{H..A-Y} = 120°$ and $\angle_{O/N-H..A-Y} = 0°,90°,180°,270°$.

(2) If the enzyme atom is an H-acceptor (oxygen) then two donor interaction sites D-X are positioned with $R_{O..D} = 1.9$ Å, $R_{D-X} = 1.0$ Å, $\angle_{O..D-X} = 180°$ and $\angle_{C-O..D} = 120°$. The D-X group is assumed to be in the plane of the R-C-O group: $\angle_{R-C-O..D} = 0°,180°$.

(3) If the enzyme atom is an unprotonated His side-chain nitrogen, then a donor interaction site D-X group with $R_{N..D} = 1.9$ Å and $\angle_{N..D-X} = 180°$ is generated on the bisector of the angle C-N-C.

(4) If the enzyme atom is an aliphatic carbon, then aliphatic interaction sites L_{ali} are generated with $R_{C..L} = 4$ Å. A user defined number (default: 14) of interaction sites are positioned, roughly equally spaced, on a sphere around each aliphatic carbon atom. The subroutine GENUN from Connolly's molecular surface program [13] is used for this purpose.

(5) If the enzyme atom is an aromatic carbon, then aromatic binding sites L_{aro} are generated with $R_{C..L} = 4$ Å using the same algorithm as for aliphatic carbon. Additional binding sites are generated with $R_{C..L} = 6$ Å above and below the plane of the aromatic rings. This was found to be necessary to allow for perpendicular arrangements of aromatic rings.

(6) If the enzyme atom belongs to an amide group, then aromatic interaction sites L_{aro} are positioned 4 Å above and below the plane of the planar amide group.

(7) If the enzyme atom is sulfur, then aromatic interaction sites L_{aro} are generated with $R_{S..L} = 4.8$ Å using the same algorithm as for rules 4 and 5.

The rules are summarized in Table 1 and illustrated in Fig. 2. All generated interaction sites are checked for van der Waals overlap with the enzyme. H-donor (D-X) and H-acceptor (A-Y) interaction sites are rejected if they are closer than 1.5 Å to the nearest enzyme atom. This value is slightly less than the closest H..O/N contact found in salt bridges in crystal structures. Aliphatic and aromatic interaction sites L_{ali} and L_{aro} are rejected if they are closer than 3.0 Å to the enzyme. It is assumed that all hydrogen atoms are present in the PDB file for the enzyme. If only the polar hydrogens are present (if the 'united atom' approach as implemented for example in the program X-PLOR [14] for protein structure refinement is used), then the threshold value for the overlap check for aliphatic and aromatic interaction sites has to be slightly increased.

The rules outlined above require some further comments. The first 3 rules are in close accordance with the hydrogen bond geometries found in crystal structures of small molecules. Indeed a statistical analysis of the CSD [15–17] reveals a preference for $\angle_{H..O=C} \cong 120°$. The second rule is also applied to hydroxyl and carboxyl groups. E.g., for a -COOH group, 4 D-X interaction sites are initially generated, one of them along the O-H vector. However, this one will be rejected due to overlap with the polar hydrogen. In judging these first 3 rules, one should bear in mind that there is a considerable variation of the hydrogen bond geometry as indicated by the structures present in the CSD [15–17], which is not completely covered by these rules. E.g., no strongly preferred values can be derived from a statistical analysis of the CSD for the dihedral angle $\angle_{N-H..O=C}$. Thus the assumed values for this dihedral angle followed by the first rule are somewhat arbitrary. The 4th, 5th and 6th rule account for the hydrophobic interactions. The program distinguishes between aliphatic and aromatic interaction sites in order to allow for some flexibility

TABLE 1
SUMMARY OF THE RULES USED TO GENERATE THE INTERACTION SITES

Rule no.	Enzyme atom	Binding site	Rules
1	H-donor (X-H; X = N,O)	A-Y	$R_{H..A} = 1.9$ Å $R_{A-Y} = 1.23$ Å $\angle_{H..A-Y} = 120°$ $\angle_{O/N-H..A} = 180°$ $\angle_{O/N-H..A-Y} = 0°,90°,180°,270°$
2	Oxygen (C=O, ROR)	D-X	$R_{O..D} = 1.9$ Å $R_{D-X} = 1.0$ Å $\angle_{C-O..D} = 120°$ $\angle_{O..D-X} = 180°$ $\angle_{R-C-O..D} = 0°,180°$
3	Nitrogen (unprotonated)	D-X	$R_{N..D} = 1.9$ Å $R_{D-X} = 1.0$ Å $\angle_{N..D-X} = 180°$ N-H on bisector of $\angle_{CNC}$
4	$C_{aliphatic}$	L_{ali}	$R_{C..L} = 4$ Å[a]
5	$C_{aromatic}$	L_{aro}	$R_{C..L} = 4$ Å[a] $R_{C..L} = 6$ Å[b]
6	Amide	L_{aro}	$R_{X..L} = 4$ Å[c]
7	Sulfur	L_{aro}	$R_{S..L} = 4.8$ Å[a]

[a] Fourteen points on sphere around enzyme atom.
[b] Above and below the plane of the aromatic ring.
[c] Above and below the plane of the amide group.

in the subsequent fragment fitting. It has been argued by Burley and Petsko [18] that interactions between aromatic systems play a significant role in the packing of a protein. As benzene has a large quadrupole moment of $2.9*10^{-40}$ C m^2 [19], one might anticipate that quadrupole–quadrupole interactions contribute significantly to the binding of an inhibitor to its enzyme. It is clear, however, that aliphatic and aromatic side chains pack closely together to form the hydrophobic core of proteins. Thus, in order to account for hydrophobic pockets it may be advisable not to distinguish between aliphatic and aromatic interaction sites. The program allows the user to proceed in both ways. The amount of 14 interaction sites per atom for aliphatic and aromatic carbons was found to be sufficient for the fragment fitting. A large portion of these points will be rejected due to van der Waals overlap with the enzyme. In fact the number of lipophilic interaction sites to be generated per interaction site can be optionally set by the user. The final two rules are based on examinations of protein structures present in the Brookhaven PDB [1]. They may be rationalized by electrostatic interactions and/or interactions between π electron systems [20,21].

The binding positions are calculated as discrete points in cartesian space (or point pairs) and then written to disc in PDB format. Visual inspection can be performed easily by any graphics program, which accepts PDB format as input.

2.1.2. Use of statistical distributions derived from the CSD

The second approach to generate the potential interaction sites, that is implemented in LUDI,

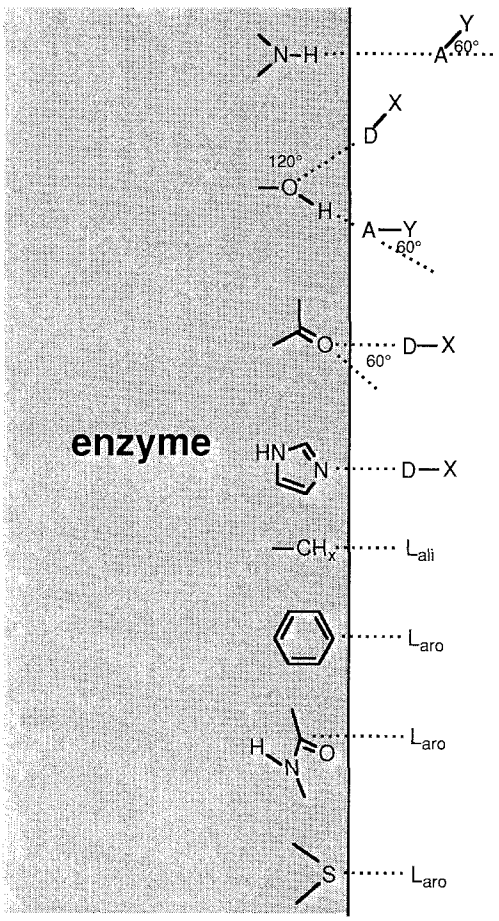

Fig. 2. Illustration of the various rules used by the LUDI module BSITES to generate the interaction sites. The program distinguishes between 4 different types of interaction sites: lipophilic-aliphatic, L_{ali}, lipophilic-aromatic, L_{aro}, hydrogen donor, D-X and hydrogen acceptor, A-Y.

is based on a search of nonbonded contacts in the CSD. We have carried out searches for the various functional groups present in proteins in order to cover the polar and nonpolar contacts typically present in enzyme–ligand interactions. The distributions are generated in a similar way as described in the study of Murray-Rust and Glusker [22] and then stored in a library using the PDB format. Since the present program describes H-donor and H-acceptor sites as atom pairs, we have tried to retrieve the positions of suitable atom pairs whenever possible. E.g., for the distributions of H-acceptor sites around a backbone nitrogen atom, all structures in the CSD were retrieved which possess nonbonded N..O=C contacts and for every hit the positions of both O and C were stored as the pair A-Y. As the H-donor interaction sites are described by the present program as a D-X pair, which corresponds to an N-H or O-H group, a problem arises with the positions of hydrogen atoms retrieved from the CSD, because these positions are not well defined due to the low X-ray scattering power of H. We have therefore decided to generate the position of hydrogen atoms (which is then used as the position of the interaction site D) at the idealized posi-

tions assuming a linear O/N...H-N arrangement with $R_{N\text{-}H} = 1$ Å. LUDI does not distinguish between interaction sites close to the maxima of the distribution and those being in the less populated regions. However, in the subsequent fragment fitting, the energetically more favorable regions will be statistically preferred simply because they are most densely populated by interaction sites.

The program selects for each functional group (e.g. backbone carbonyl, backbone amide, side-chain imidazole ring, etc.) the most suitable distribution from the library and superimposes the template of the distribution with the functional group of the enzyme. The generated positions are checked for van der Waals overlap with the enzyme and then stored in PDB format. Further details of this approach will be published elsewhere [23].

2.1.3. Interface to GRID

An alternative route to the rule-based approach or the use of statistical data from crystal structures of small molecules is the application of the program GRID [5]. Our program also includes a simple interface to GRID that reads the ASCII output file generated by GRID. It first finds the local minima in the grid and then selects a certain number (default: 3) of further low-energy points around the minimum suitable to fit the fragments upon. These points are then converted into vectors if necessary (for donor and acceptor interaction sites, the positions X and Y, respectively, are generated) and further treated as interaction sites.

2.2. Fit of the molecular fragments (FRAGMT)

The next step is the fitting of molecular fragments onto the interaction sites generated in the previous step and is performed as follows. The list of interaction sites is searched by distance criteria for suitable pairs, triplets or quadruples to match the molecular fragments. The criterion used for selecting suitable interaction sites is based upon the square of the distances between the interaction sites R_{ij}^2. If these values lie within a specified range (default range is calculated from the optimum value R_{opt}^2 as $R_{min}^2 = (R_{opt} - 2\Delta R_{max})^2$ and $R_{max}^2 = (R_{opt} + 2\Delta R_{max})^2$ with a typical value of 0.2–0.6 Å for ΔR_{max}), then a fit of the fragment is performed. For example, if 3 aromatic interaction sites i, j and k are found with suitable values for R_{ij}^2,R_{ik}^2 and R_{jk}^2, then 3 atoms of a lipophilic fragment, e.g. a phenyl group (e.g. the atoms 1, 3 and 5) are fitted onto these 3 points. The fit of the molecular fragments is done by a root–mean–square (RMS) superposition using the algorithm published by Kabsch [24]. A fit of a molecular fragment is accepted if the RMS value is less than a user-defined threshold (typically 0.2–0.6 Å) and no van der Waals overlap of the fitted fragment with the enzyme occurs. If the fit of the molecular fragment is accepted, then its coordinates are stored in PDB format.

The molecular fragments are read by the program from a fragment library. For each of these fragments, between 3 and 8 atoms are defined which are fitted onto the interaction sites. The fragment library contains structures of typically 5–30 atoms. Typical examples of fragments included in the library are given in Table 2. The fragment library was generated manually by carefully selecting suitable molecules in a low-energy conformation either obtained by force-field calculations using the force field CVFF [25] or by using experimentally determined crystal structures. Currently, the fragment library contains 600 entries. For the active site of the enzyme trypsin a run using all fragments takes about 20 min on a VAX 11/785.

The fragment library uses the PDB format. The program allows the user to define his own new

TABLE 2
TYPICAL MOLECULES CONTAINED IN THE FRAGMENT LIBRARY

benzene	naphthalene	cyclohexane
adamantene	acetic acid	indole
benzoic acid	nitrobenzene	aniline
phenole	imidazole	pyrrole
cyclohexanone	phenylamidine	*n*-propylguanidine
piperidine	piperazine	morpholine
N-acetyl-*N*'methyl-valineamide (10 conformers)		
N-acetyl-*N*'-methyl-D-phenylalanine amide (10 conformers)		
α-naphthol (2 conformers)		β-naphthol (2 conformers)

fragments simply by adding a corresponding PDB file to the library with a header card defining the name of the fragment and the number of atoms. This can be done without any changes to the program. For each new fragment to be added to the fragment library the user has to select the atoms actually used for matching the interaction sites and for each atom also the type of interaction site that it shall be fitted onto. The choice of the target atoms for the fitting procedure is rather straightforward for small molecules such as acetone: all heavy atoms are selected, the aliphatic carbons are fitted onto aliphatic interaction sites and the carbonyl group is fitted onto an H-acceptor interaction site. The situation is more complicated for larger fragments such as dipeptides or substituted heterocyclic compounds, because there are several atoms or functional groups which could be used as target atoms for the fragment fitting. One possible route to solve the problem is to use several independent sets of target atoms per fragment. Indeed, for some of the larger fragments in the current fragment library, the program makes several attempts to fit the molecule onto the interaction sites using different sets of target atoms. E.g., for the fragment 2,4-diamino-pteridine (which will be discussed in Section 3.2) LUDI uses 5 different sets of target atoms. Currently we are working on a procedure to select automatically the most promising target atoms for any given molecule.

All fragments are treated as rigid bodies. However, internal flexibility can be accounted for by including several conformers for a fragment in the library. One example of flexible molecules are dipeptides, where the conformers corresponding to the highly populated regions in the Ramachandran plot [26] are stored in the library. Another example are hydroxy-compounds, e.g. naphthol, where two rotamers of the hydroxy-group are stored in the fragment library.

There are clearly a large number of possibilities as to how to fit the fragments onto the interaction sites. We have found it useful to allow more than one fragment to be fitted per interaction site and then to select manually the appropriate fragments to be merged by visual inspection.

2.3. Fragment bridging (BRIDGE)

The final step in the present program involves the connection of some or all of the fitted fragments to a single molecule. This is done by bridge fragments, which are stored in the bridge library (PDB format). The algorithm used is somewhat similar to the program CAVEAT developed by Bartlett and co-workers [27]. The program first identifies the two fragments in closest contact,

then identifies those hydrogen atoms in both fragments which have the closest contact. These two hydrogen atoms, together with the adjacent heavy atoms, are used as link sites for selecting the bridge. The most suitable bridge is found by fitting bridge fragments onto both link sites in a similar manner as described in the previous section. The terminal atoms of the bridge are fitted onto the link sites. For some of the bridges the interaction sites are used for the fitting as well. E.g., the -CH2COO-bridge is fitted onto two suitable link sites and onto one H-acceptor site. Again the Kabsch algorithm [24] is used for the least-squares superposition. The program attempts to merge all fragments into one single molecule. The generated bridges are finally stored in PDB format.

Typical bridge fragments currently implemented are: $-CH_2-$, $-CH_2CH_2-$, $-CH_2CH_2CH_2-$, -CH=CH- (cis and trans), -O-, -CO-, -NH-, $-SO_2-$, $-SO_2NH-$, -CONH-, -COO-, -CH2COO-.

3. APPLICATIONS

3.1. Crystal packing of benzoic acid

As a very simple first test case we have applied our program to the crystal packing of benzoic acid (CSD Refcode BENZAC01 [28]). The structure was retrieved from the CSD and several unit cells were generated using the molecular modelling program package SYBYL [29]. A molecule in the center being fully coordinated by its nearest neighbours was then removed leaving a cavity in the crystal structure. The cavity with the nearest neighbouring molecules located in the plane of the removed molecule is shown in Fig. 3. This ensemble was used as the input to our program and the extracted benzoic acid molecule was added to the fragment library. The purpose of this first test was to verify that the program can fit the benzoic acid molecule back into the artificial pocket inside the crystal packing of benzoic acid.

The output generated by LUDI is shown in Figs. 4–6. The interaction sites generated by the program module BSITES are shown in Fig. 4. One H-donor, one H-acceptor and 14 aromatic interaction sites are generated. In the next step the module FRAGMT fits the benzoic acid molecule back into the cavity using the generated interaction sites. The fragment Benzoic acid is fitted onto 4 interaction sites: one H-donor, one H-acceptor and two aromatic interaction sites. The -COOH group and both carbon atoms in meta position are used as target atoms of the benzoic acid fragment. The fit is achieved with an RMS of 0.34 Å with respect to the interaction sites. The fitted benzoic acid molecule is shown with the interaction sites in Fig. 5. Figure 6 compares the position of the fitted molecule with the correct position in the crystal packing. The fitted molecule is rotated slightly by 8° with respect to its position in the crystal packing. The RMS deviation (all heavy atoms plus the polar hydrogen) of the fitted molecule from the corresponding molecule in the crystal packing is 0.23 Å. The slight misorientation is caused by the aromatic interaction sites which do not coincide exactly with the atomic positions in the crystal structure. A further slight discrepancy arises from the fact that the hydrogen bond in the crystal packing of benzoic acid is slightly shorter ($R_{O..H} = 1.74$ Å [28]) than the 'standard' hydrogen bond length of 1.9 Å, which is used by the present program.

3.2. DHFR–MTX complex

As a second test case we have applied our program to the enzyme dihydrofolate reductase

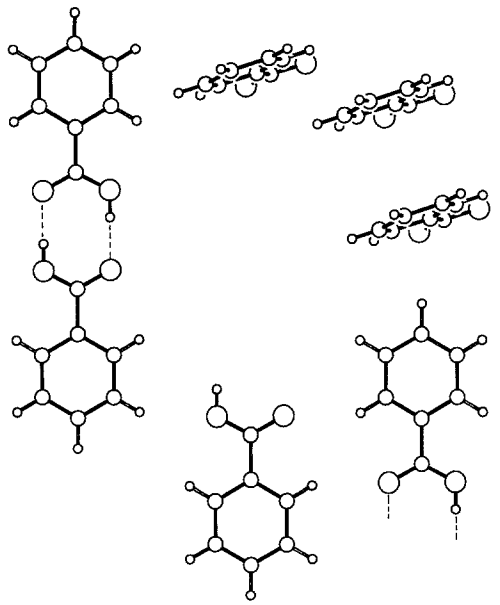

Fig. 3. Cavity in the crystal packing of benzoic acid generated by removing one molecule. Note that only the molecules lying in the plane of the removed molecule are shown.

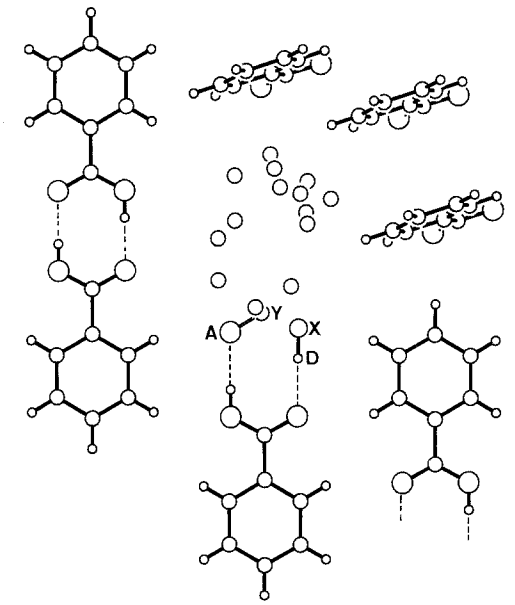

Fig. 4. Cavity in the crystal packing of benzoic acid with the interaction sites generated by BSITES using the rules described in Section 2.1.2. One H-donor, one H-acceptor and 14 aromatic interaction sites are generated.

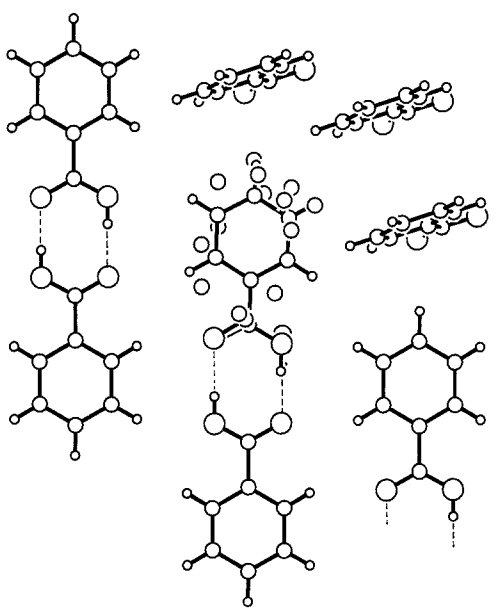

Fig. 5. Fit of a benzoic acid molecule (open bonds) onto the interaction sites. Hydrogen bonds are shown as dashed lines.

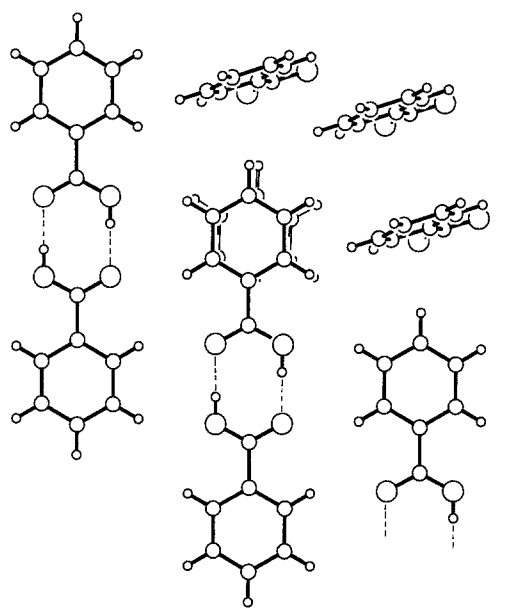

Fig. 6. Comparison of fitted benzoic acid molecule (open bonds) with the experimentally determined crystal packing of benzoic acid. The RMS deviation of the fitted molecule from the corresponding molecule in the crystal packing is 0.23 Å.

(DHFR) crystallized as a complex with the anticancer drug methotrexate (MTX) [30] (entry 4DFR in the Brookhaven PDB [1]). The MTX was removed and one DHFR molecule (molecule A) without water molecules was used as input file. In one run of LUDI a single water molecule (water A405) known to be important for the binding of MTX to DHFR [30] was added to the input file. The hydrogen atoms of this water molecule were positioned by a force–field calculation using the force field CVFF [26] in such a way that proper hydrogen bonds could be formed with the neighbouring amino acids of DHFR. The hydrogen atoms of the amino acids were added using the program MOLEDT [31]. The amino acids Asp, Glu, Lys and Arg were assumed to be charged. The interaction sites were generated using the rules as described in Section 2.1.1. We carried out 3 runs on the active site of DHFR using our program LUDI. The first and the second run used just 2,4-diamino-pteridine as molecular fragment to be fitted. The only difference between these two calculations was, that in the second run the water molecule A405 was added to the input file. The third run was limited to the use of the molecular fragments $CC(NH_2)_2^+$, CH_3COO^- and C_6H_6 only.

The result of the first run is shown in Fig. 7. The 2,4-diamino-pteridine molecule is fitted into the active site in the correct orientation. The RMS deviation of the fitted molecule from the corresponding MTX portion is 0.6 Å. The slight tilt of the fitted molecule as compared to MTX is a result of the position of the aromatic interaction sites, which are not located exactly in the plane of the pteridine moiety of MTX.

In the second run, a water molecule (# A405) known to form hydrogen bonds both to DHFR and to MTX was added to the input file. Again LUDI fits the 2,4-diamino-pteridine correctly into the active site of DHFR – this time, however, using a different set of target atoms for the fit. The

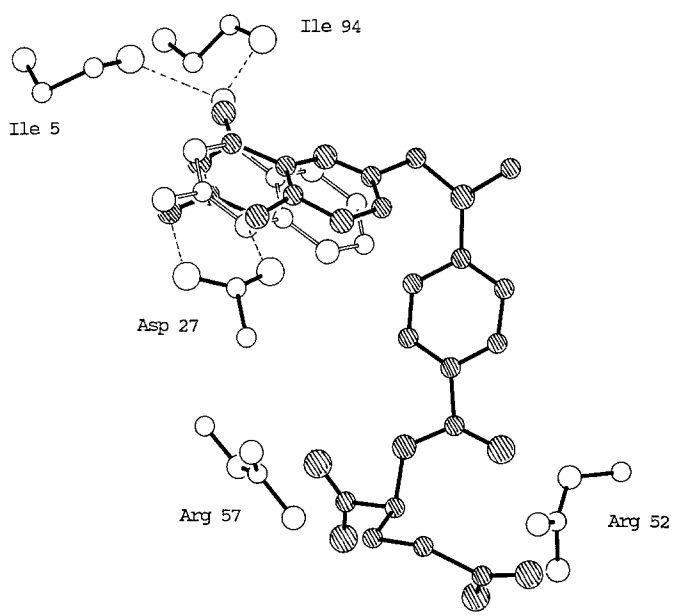

Fig. 7. Comparison of the fitted fragment 2,4-diamino-pteridine (open bonds) with the conformation of methothrexate (shaded atoms) as bound to DHFR. The hydrogen bonds between the fitted fragment and DHFR are shown as dashed lines.

RMS deviation of the fitted molecule with the corresponding MTX moiety is 0.47 Å. The tilt of the fitted fragment with respect to MTX is reduced as compared to the first run. Thus the inclusion of the water molecule slightly improves the quality of the fit. It is, however, interesting to note, that apparently the hydrogen bond with the water molecule is not required by LUDI to establish the correct orientation of the pteridine moiety.

The generated fragments of the third run are shown in Fig. 8. The amidino fragment #1 and the carboxyl fragment #2 are well positioned as compared to the pteridine moiety and the σ-COO group in the methotrexate molecule. These two fragments form the same hydrogen bonds with Asp A27 and Arg A57 of DHFR as does MTX. The RMS deviations of these fragments with respect to the corresponding MTX portions are 0.5 Å and 0.7 Å, respectively. A second carboxyl group #5 is positioned to form a chelate–type H-bond with Arg A52, whereas the terminal COO group of MTX forms only a single H-bond with Arg A52 and another H-bond with a water molecule. There are two phenyl fragments #3 and #4, which are positioned in close proximity to the aromatic rings of MTX. However, the relative and absolute orientation of the phenyl rings is significantly different from that in MTX. One may hope that slight misorientations of hydrophobic fragments can be corrected in a subsequent force–field calculation.

The fragments generated by the second run were used as input to the module BRIDGE. The bridge fragments are shown in Fig. 9. The amino fragment and the phenyl ring #3 can be directly linked together. LUDI suggests a methyl fragment as bridge between the two phenyl rings. This connection is one atom shorter as compared to MTX. A propyl fragment is calculated as bridge between the phenyl ring #4 and the carboxy group #2. This propyl bridge is one atom longer than the corresponding amide moiety of MTX. The reason for these discrepancies lies in the

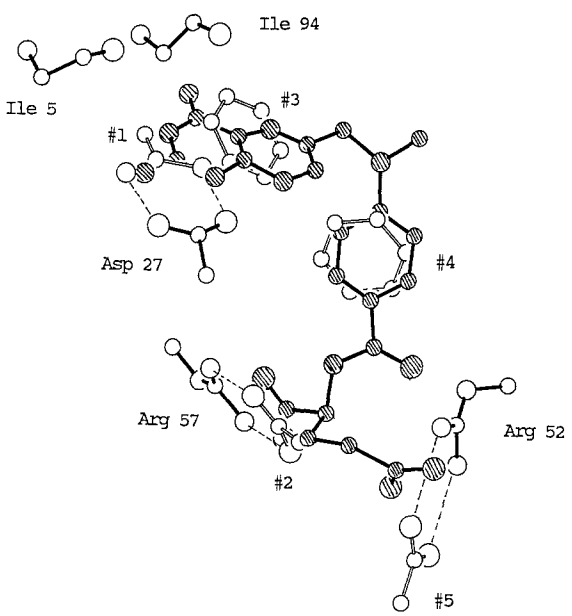

Fig. 8. Comparison of the conformation of methotrexate (shaded atoms) as bound to DHFR with 5 molecular fragments #1–#5 generated by the third run of LUDI.

slightly different orientations of the generated fragments. E.g. the carboxy fragment is positioned exactly in the plane of the guanidino group of Arg A57 as compared to the -COO moiety of MTX, which is slightly tilted out of the plane.

3.3. Trypsin–benzamidine complex

The final example in the present paper involves the enzyme trypsin crystallized as a complex with the inhibitor benzamidine [32] (entry 3PTB in the Brookhaven PDB). The input coordinates of the enzyme were generated as described above for DHFR. Again, the benzamidine molecule and all water molecules were removed. The question was: If the benzamidine molecule is added to the fragment library (in fact this library contains molecules typically of the same size), can LUDI identify benzamidine among a large set of other molecules as a good candidate for enzyme inhibition?

The first run using 'rule–based' interaction sites failed. LUDI did not fit benzamidine into the specificity pocket of trypsin. This is due to the nonplanar arrangement of benzamidine with respect to Asp^{189} in the crystal structure. Since the 'rule-based' approach attempts to position the interaction sites in the plane of the carboxy group, it cannot cope with nonplanar arrangements.

The result of the second run using interaction sites generated from CSD distributions is shown in Fig. 10. The benzamidine fragment is positioned in close proximity to the experimentally determined structure. The RMS deviation of the fitted fragment from the benzamidine molecule found in the crystal structure is 0.39 Å. Using the full library of 600 fragments, LUDI positioned only

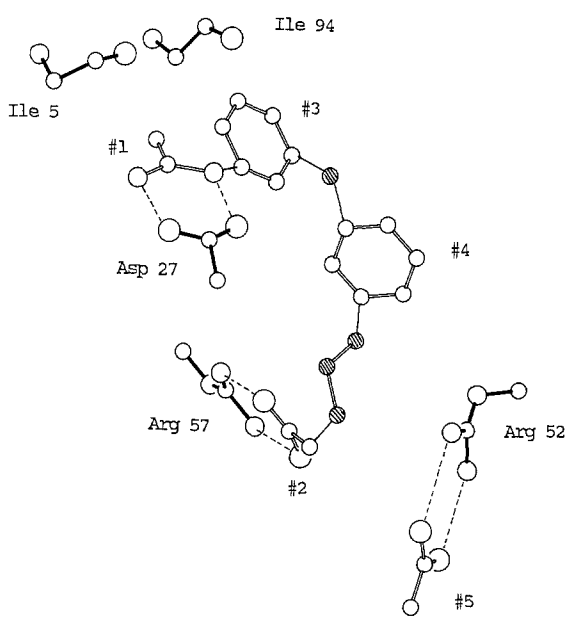

Fig. 9. Molecular fragments #1–#5 in the active site of DHFR (open bonds) together with bridge fragments (shaded atoms). A methyl bridge is inserted between the phenyl fragments #3 and #4. A propyl bridge is inserted between the fragments #2 and #4.

3 further fragments into this pocket, namely aniline, *n*-propylguanidine and *n*-butylamine. All 4 molecules form at least 2 hydrogen bonds with trypsin. It is interesting to note, that all these structures are known to bind into the specificity pockets of trypsin-like serine proteases [33,34].

4. DISCUSSION AND CONCLUSION

We have developed a new computer program LUDI for automated protein ligand design, e.g. the design of enzyme inhibitors. The program positions molecular fragments into the active site of an enzyme in such a way, that hydrogen bonds with the enzyme are formed and hydrophobic pockets are filled. These fragments are then linked together by suitable spacers ('bridges') to form a single molecule. The algorithm is very fast (e.g. the run on the active site of trypsin using 600 fragments takes less than 20 min on a VAX 11/785 and less than 2 min on a Silicon Graphics 4D25) and is very easy to use. It is also possible to define new fragments simply by adding a corresponding PDB file to the fragment library. This can be done without any changes of the program.

Several applications of LUDI have been presented. The program correctly positions a benzoic acid molecule back into a cavity in the crystal packing of benzoic acid. For the enzyme DHFR, two fragments are fitted very close to the positions where corresponding functional groups of MTX form strong hydrogen bonds with DHFR. Moreover, phenyl rings are positioned in close proximity to the aromatic portions of MTX. For trypsin the program fails to fit the benzamidine fragment into the specificity pocket of the enzyme, if rule-based interaction sites are used. Howev-

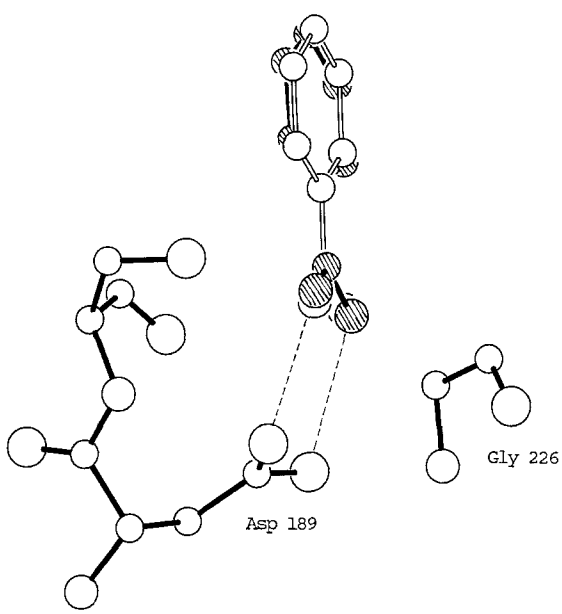

Fig. 10. Comparison of fitted fragment benzamidine (open bonds) with the crystal structure of trypsin (black bonds) complexed with benzamidine (shaded atoms). Note: Gly^{226} causes a nonplanar arrangement of the benzamidine molecule with respect to the carboxy group of Asp^{189}.

er, if the interaction sites are generated from a distribution of nonbonded contacts obtained through a statistical evaluation of crystal data in the CSD, then the benzamidine fragment is positioned correctly. Moreover, the program successfully identifies 3 further candidate molecules as trypsin inhibitors out of a library of 600 molecules.

The ultimate goal of the present approach is clearly a fully automated design of enzyme inhibitors. Although the results presented in this paper are promising, one should realize that we are still far away from a fully automatic procedure. In judging the results obtained up till now, one should bear in mind the underlying assumptions of the present algorithm:

(1) The enzyme is assumed to be rigid. Flexibility or partial flexibility of enzymes (e.g. in flap regions as indicated by high-temperature factors in the crystal structure determination) is not accounted for.
(2) The current algorithm does not distinguish between strong and weak hydrogen bonds.
(3) Entropic effects are only indirectly taken into account by placing hydrophobic fragments into the corresponding pockets.
(4) The molecules generated by the present version of the algorithm may not be in an energetically favorable conformation, because some of the bond lengths (bonds between fragments and bridges) may not be at their equilibrium values. The same also holds for bond and torsion angles.

It is well known that ligand-induced conformational changes of the protein do occur upon substrate binding. They may be very small as in the trypsin–benzamidine complex [32,35], or too large to allow for any reasonable prediction of potential ligands such as in the case of the rhinovirus coat protein [36]. It should also be noted, that in some cases a predicted hydrogen bond between a putative ligand and the protein will not yield an improved free energy of binding as compared to a ligand, which cannot form that hydrogen bond. This can be explained through consideration of the differences in protein–ligand and solvent–ligand interactions [37]. The current algorithm considers only interactions of the ligand with its environment in the active site, but does not explicitly take into account solvation effects. Frequently, the generated molecules will be very flexible with a considerable number of rotatable bonds. Moreover, the complete molecules may be difficult to synthesize. Nevertheless, despite the limitations, in our view the program is a helpful new tool in the drug design process. It offers several interesting features to the drug designer:

(1) It provides an unbiased list of fragments that fit into the active site and may keep the drug designer from overlooking possibilities to make modifications to known inhibitors.
(2) The spacer groups give some estimate on bridging the gaps between the localized fragments in terms of chemical structures.
(3) The algorithm is very fast and easy to use. The usage of the PDB format for all input and output files makes interfacing to existing molecular modelling software simple.

The program can also be used to propose new substituents for an already known inhibitor, if the 3D structure of the protein–inhibitor complex is known. Our first experience using LUDI on the trypsin-benzamidine structure suggests, that the program is capable to predict substituents for benzamidine, which yield significantly improved binding [38].

People interested in using the program are requested to contact the author.

ACKNOWLEDGEMENTS

I would like to thank my colleagues Wolfgang Höffken, Gerhard Klebe, Hugo Kubinyi and Nigel Walker for many helpful discussions. I am grateful to Friedrich Diezinger for his help in generating the fragment library.

REFERENCES

1 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, T., J. Mol. Biol., 112 (1977) 535.
2 Cohen, N.C., Blaney, J.M., Humblet, C., Gund, P. and Barry, D.C., J. Med. Chem., 33 (1990) 883.
3 DesJarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D. and Venkataraghavan, R., J. Med. Chem., 31 (1988) 722.
4 Lewis, R.A., J. Comput.-Aided Mol. Design, 4 (1990) 205.
5 Goodford, P.J., J. Med. Chem., 28 (1985) 849.
6 Tomioka, N., Itai, A. and Iitaka, Y., J. Comput.-Aided Mol. Design, 1 (1987) 197.
7 a. Allen, F.H., Kennard, O. and Taylor, R., Acc. Chem. Res., 16 (1983) 146.
b. Allen, F.H., Bellard, S., Brice, M.D., Cartwright, B.A., Doubleday, A., Higgs, H., Hummelink-Peters, T., Kennard, O., Motherwell, W.D.S., Rodgers, J.R. and Watson, D.G., Acta Crystallogr., B35 (1979) 2331.
8 Hambley, T.W., Chan, H.K. and Gonda, I., J. Am. Chem. Soc., 108 (1986) 2103.
9 Crippen, G.M., In Bawden, D. (Ed.) Distance Geometry and Conformational Calculations, Research Studies Press (Wiley), New York, 1981, p. 1.
10 Rusinko, III, A., Skell, J.M., Balducci, R. and Pearlman, R.S., University of Texas at Austin, distributed by Tripos Ass., 1699 S. Hanley Road, St. Louis, MO 63144, 1987.
11 Leach, A.R., Prout, K. and Dolata, D.P., J. Comput. Chem., 11, (1990) 680.
12 Hermann Hesse, Das Glasperlenspiel.
13 Connolly, M.L., Science, 221, (1983) 709.
14 Brünger, A.T., Kuriyan, J. and Karplus, M., Science, 235 (1987) 458.
15 Taylor, R., Kennard, O. and Versichel, W., J. Am. Chem. Soc., 105 (1983) 5761.
16 Taylor, R., Kennard, O. and Versichel, W., J. Am. Chem. Soc., 106 (1984) 244.
17 Taylor, R., Kennard, O. and Versichel, W., Acta Crystallogr., B40 (1984) 280.
18 Burley, S.K. and Petsko, G.A., J. Am. Chem. Soc., 108 (1986) 7995.
19 Battaglia, M.R., Buckingham, A.D. and Williams, J.H., Chem. Phys. Lett., 78 (1981) 421.
20 Reid, K.S.C., Lindley, P.F. and Thornton, J.M., FEBS Lett., 190 (1985) 209.
21 Stoddard, B.L., Bruhnke, J., Porter, N., Ringe, D. and Petsko, G.A., Biochemistry, 29 (1990) 4871.
22 Murray-Rust, P. and Glusker, J.P., J. Am. Chem. Soc., 106 (1984) 1018.
23 Böhm, H.J. and Klebe, G., manuscript in preparation.
24 Kabsch, W., Acta Crystallogr., A32 (1976) 922, A34 (1978) 827.
25 Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V., J. Mol. Biol., 7 (1963) 95.
26 Dauber-Osguthorpe, P., Roberts, V.A., Osguthorpe, D.J., Wolff, J., Genest, M. and Hagler, A.T., Proteins, 4 (1988) 31.
27 Bartlett, P.A., Shea, G.T., Telfer, S.J. and Waterman, S., In Roberts, S.M. (Ed.) Molecular Recognition: Chemical and Biological Problems, Royal Society of London, 1989, pp. 182–196.
28 Bruno, G. and Randaccio, L., Acta Crystallogr., B36 (1980) 1711.
29 Program SYBYL, distributed by Tripos Ass., 1699 S. Hanley Road, St. Louis, MO 63144.
30 Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C. and Kraut, J., J. Biol. Chem., 257 (1982) 13650.
31 Program MOLEDT, distributed by BIOSYM Technologies, San Diego, CA 92121.
32 Marquart, M., Walter, J., Deisenhofer, J., Bode, W. and Huber, R., Acta Crystallogr., B39 (1985) 480.
33 Baker, B.R. and Erickson, E.H., J. Med. Chem., 10 (1967) 1123.
34 Stürzebecher, J., Horn, H., Markwardt, F., Wagner, G. and Walsmann, P., Pharmazie 36 (1981) 639.
35 Bartunik, H.D., Summers, L.J. and Bartsch, H.H., J. Mol. Biol., 210 (1989) 813.

36 Badger, J., Minor, I., Kremer, M.J., Oliveira, M.O., Smith, T.J., Griffith, J.P., Guerin, D.M.A., Krishnaswamy, S., Luo, M., Rossmann, M.G., McKinlay, M.A., Diana, G.D., Dutko, F.J., Fancher, M., Rueckert, R.R. and Heinz, B.A., Proc. Natl. Acad. Sci. U.S.A., 85 (1988) 3304.
37 Morgan, B.P., Scholtz, J.M., Ballinger, M.D., Zipkin, I.D. and Bartlett, P.A., J. Am. Chem. Soc., 113 (1991) 297.
38 Böhm, H.J., to be published.