Journal of Computer-Aided Molecular Design, 6 (1992) 131–147
ESCOM

J-CAMD 152

In search of new lead compounds for trypanosomiasis drug design: A protein structure-based linked-fragment approach

Christophe L.M.J. Verlinde*, Gabrielle Rudenko and Wim G.J. Hol

BIOSON Research Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Received 27 August 1991
Accepted 14 October 1991

Key words: Modular structure-based inhibitor design; Chemical database; Triosephosphate isomerase

SUMMARY

A modular method for pursuing structure-based inhibitor design in the framework of a design cycle is presented. The approach entails four stages: (1) a design pathway is defined in the three-dimensional structure of a target protein; (2) this pathway is divided into subregions; (3) complementary building blocks, also called fragments, are designed in each subregion; complementarity is defined in terms of shape, hydrophobicity, hydrogen bond properties and electrostatics; and (4) fragments from different subregions are linked into potential lead compounds. Stages (3) and (4) are qualitatively guided by force-field calculations. In addition, the designed fragments serve as entries for retrieving existing compounds from chemical databases. This linked-fragment approach has been applied in the design of potentially selective inhibitors of triosephosphate isomerase from *Trypanosoma brucei*, the causative agent of sleeping sickness.

INTRODUCTION

In search of new lead compounds: from discovery towards design

Discovering new lead compounds from which to develop drugs has traditionally to a large extent been a matter of serendipity and mass-screening. Screening can be very tedious as exemplified by the quest for antimalarial drugs at the Walter Reed Army Institute of Research: it took 13 years and the testing of 300,000 compounds to discover seven effective drugs [1]. Alternatively, insight in a metabolic pathway may provide a new lead compound: e.g. the folic acid structure eventually led to the synthesis of pyrimethamine, a selective inhibitor of the dihydrofolate reductase of the malaria parasite [2]. However, if the structure of a molecular target is known down to atomic detail, it becomes possible to pursue structure-based drug design (see e.g. [3]). The first stage of this engineering exercise relies entirely on knowledge of the molecular target and should be coined

* To whom correspondence should be addressed.

'lead design and optimization'. The second stage consists of adapting the lead to the requirements of absorption, distribution, metabolism, excretion and toxicity; stage-one constraints apply also to this process.

Current protein structure-based methods to design new lead compounds fall into three types: (1) intuitive design, e.g. the invention of a nonpeptide inhibitor of HIV-1 protease by exploiting the structural information derived from a related peptide inhibitor–enzyme complex [4]; (2) automated site-directed inhibitor design: (a) by generating two- and three-dimensional molecular graphs subject to the constraints of a macromolecular environment [5,6], (b) by combining information about favourable binding sites for particular molecular probes [7,8] into molecules through peak parsing techniques [9]; (3) by screening, which is not a true design method, of three-dimensional databases such as the Cambridge Crystallographic Database [10] or a 2D-derived database (e.g. by using the program CONCORD [11]) using shape-matching algorithms [12] or pharmacophoric pattern-matching [13–15]. Like method (1), method (3) has proven its usefulness by identifying a butyrophenone derivative as a HIV-1 protease inhibitor [16]. Method (2) is clearly still under development.

The high-resolution structure of the macromolecular target which is required for all design methods is in almost all cases elucidated through X-ray crystallography. Recent developments in this field (see e.g. [17]) enable the rapid set-up of a database of the target macromolecule complexed with numerous existing compounds. Common binding mode patterns can then be identified and information gained about the adaptability of the protein's conformation. In this way a series of structures rather than a single structure forms the basis for the modelling of new lead compounds. We wondered if there was a design strategy which took greater advantage of biomacromolecular X-ray crystallography as the motor of a rational drug design cycle [3] and relied less on our imperfect understanding of drug–receptor interactions. This is particularly desirable when large inhibitors have to be designed.

The linked-fragment approach: a 'natural' component of the rational drug design cycle

An efficient design approach, particularly for larger inhibitors, would benefit greatly from a concept largely ignored in the aforementioned design strategies: the designed compound should be modular, i.e. it should consist of molecular building blocks. This has several advantages:

- through a combinatorial use of a few analogues of each building block a large variety of compounds can be designed;
- immediate experimental feedback from 3D structure determinations is possible, by simply carrying out soaking experiments with individual building blocks;
- each building block provides a logical entry point for analogue searches in chemical databases;
- synthesis of the many variants required for turning a lead compound into a drug should be facilitated.

Our efforts along these lines have resulted in a 'linked-fragment approach'. This strategy entails five steps:

(1) define the minimum molecular recognition environment required for the binding of a ligand (e.g. a selective competitive inhibitor); we call this macromolecular environment the 'design pathway';

(2) analyse the molecular surface of the design pathway in terms of shape, hydrophobicity, hydrogen bond properties and electrostatics;

(3) divide the design pathway into a number of subregions which each could accommodate a small molecular building block (approximately 4 to 11 atoms);
(4) evaluate the complementarity of a large number of building blocks with the pathway in each of the subregions using the properties revealed under (2); include the molecular strain of the building block conformation in this evaluation;
(5) link promising building blocks into potential lead compounds; avoid where possible introducing chiral centres as this may complicate synthesis.

We are currently trying this approach in a collaborative project aimed at developing new drugs for the treatment of sleeping sickness, also known as African trypanosomiasis.

Trypanosomiasis: glycolysis, a potential target for rational drug design

Sleeping sickness is considered by the World Health Organization as one of the major tropical parasitic diseases [18]. It is caused by the bloodstream form of the protozoon *Trypanosoma brucei*, and is always fatal if untreated. The few available drugs give rise to very serious side effects. In particular melarsoprol and analogues, the only trypanocidals effective in the late stages of the disease, induce a deadly encephalopathy in 10% of the treated patients [19]. Also the recently introduced compound 'eflornithine' (difluoromethyl ornithine, DFMO) has serious drawbacks: (i) it is less active against *Trypanosoma rhodesiense*, the more virulent form of sleeping sickness [20]; (ii) resistance has already been reported [21]; and (iii) treatment requires the administration of large quantities (the FDA recommendation is 400 mg/kg body weight i.v., daily for 2 weeks) under hospitalization of the patient [22]. Clearly, more effective and safer drugs are eagerly awaited.

Trypanosomes exhibit several unusual features which can be exploited for rational drug design [23]. One of them is the complete dependence of the bloodstream form on glycolysis to the stage of pyruvate as a sole source of energy supply [24]. Therefore, the glycolytic enzymes are attractive targets for designing selective inhibitors which should obviously exhibit minimal affinity for the equivalent enzymes in the human host.

In *Trypanosoma brucei* nine enzymes involved in glycolysis and glycerol metabolism are sequestered in a specialized organelle called the glycosome [24]. They have been purified and extensively characterized [25,26]. Thus far, the structures of two of these enzymes have been elucidated by X-ray crystallography, namely triosephosphate isomerase (EC 5.3.1.1) [27] and glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12) (F.M.D. Vellieux, personal communication; R.J. Read, personal communication). In this paper only the design of selective inhibitors of the glycosomal triosephosphate isomerase (gTIM) from *Trypanosoma brucei* is dealt with.

DATA AND METHODS

The triosephosphate isomerase structural database

Triosephosphate isomerase was the first of the glycosomal enzymes from *T. brucei* for which a crystal structure was obtained [27]. The structure has been refined against data between 6.0 and 1.83 Å resolution to an R-factor of 18.3% for 38,819 reflections [28]. The model of the dimeric enzyme includes 310 water molecules and a sulphate ion ($K_i = 4.5$ mM [29]) bound to the active site of subunit two (sequence numbers go from 2 to 250 for subunit one, and from 302 to 550 for subunit two). No sulphate ion is bound to the active site of subunit one because crystal packing effects prevent the required closing of a flexible loop (residues 168 to 180) in this case [27]. This mobile

loop is essential for binding the enzymatic reaction intermediate in a conformation that disfavours the phosphate elimination side reaction [30].

In addition to the native structure, crystal structures of the trypanosomal enzyme in a complex with several nonselective competitive inhibitors have been obtained (Table 1). This led to the following observations:

- with any of these inhibitors the flexible loop folds in an identical fashion further over the active site than with sulphate bound; in this way the inhibitors are sealed off from the solvent;
- the phosphate/phosphonate moiety adopts an identical position (within 0.2 Å) in all complexes, which is about 1.0 Å further inside the active-site pocket than the sulphate position;
- for the phosphonate compounds, the carbon atom bound to the phosphor atom remains partially accessible to the solvent;
- the catalytic residue Glu^{467} adopts a different conformation in each of the complexes;
- the water structure of the active site of subunit two is different in all six structures listed in Table 1.

The identical positioning of the phosphate/phosphonate irrespective of the nature of the attached group provides a particularly good starting point for the modelling of new inhibitors. Such a moiety could be used as a molecular anchor in a designed compound.

Computational tools for the linked-fragment approach

All modelling was carried out using the molecular modelling package BIOGRAF 2.10 [34] on an Evans and Sutherland PS390 interfaced to a MicroVAX II host computer. The protein structure used for design purposes was that observed in the complex between gTIM and glycerol-3-phosphate [32], with all water molecules and the inhibitor molecule deleted. Analysis of the properties of the design pathway was facilitated by calculating and displaying a molecular surface (probe radius = 1.4 Å) [35].

Only a static shell of protein atoms within 9 Å about the design pathway was included in the calculations. Standard united-atom types from the Dreiding-I force field were assigned to the

TABLE 1
CRYSTALLOGRAPHICALLY DETERMINED gTIM–INHIBITOR COMPLEXES

Inhibitor	K_i[a] (mM)	Resolution (Å)	R-factor[b] (%)	Ref.
Sulphate	4.5	1.8	18.3	[28]
Phosphate	4.5	2.4	15.0	[31]
Glycerol-3-phosphate	0.6	2.2	13.7	[32]
3-Phosphoglycerate	1.3	2.2	14.0	[32]
3-Phosphonopropionate	2.7	2.6	12.5	[32]
2-Phosphoglycerate	6.9	2.4	14.9	[33]
2-(*N*-formyl-*N*-hydroxy-amino)-ethyl-phosphonic acid	0.5	2.5	15.3	–[c]

[a] Inhibition constants from: [29]; F.R. Opperdoes and M. Callens, personal communication.

[b] R-factor $= \Sigma|(F_{obs,h} - F_{calc,h})|/\Sigma|F_{obs,h}|$.

[c] Noble, M.E.M., unpublished results.

atoms of both the protein and the modelled ligands; heteroatoms were supplemented where necessary with hydrogen atoms which are used in an explicitly geometrical hydrogen-bond potential term. All energy calculations were carried out 'in vacuo' with the electrostatic term of the potential energy function turned off. A long-range cutoff distance of 9 Å was chosen for the nonbonded interactions. The energy minimization convergence criterion was set to 0.4 kJ mol^{-1} $Å^{-1}$; conjugate gradient methods were used.

Evaluation of molecular interactions

Our aim is not to predict the ΔG of ligand binding, especially in view of the limited accuracy of calculating merely ΔΔG values at present (see e.g. [36]). Instead, we want to perform an educated guess whether a docked ligand fits 'reasonably' in a given protein environment. Such a fit is considered acceptable if (a) little strain is induced in the ligand, and (b) nonbonded interactions are favourable. The induced strain can be evaluated by a direct comparison with the strain calculated for the experimentally observed closest low-energy conformation. Such experimental conformations are readily available from template libraries in the modelling package, or can be retrieved from the Cambridge Structural Database [10].

Evaluating the nonbonded interactions is more difficult because of solvent effects. Four types of interactions should be considered: electrostatic, hydrogen bonding, van der Waals and hydrophobic. As mentioned above, our force-field calculations are devoid of an electrostatic energy term. However, the most important effect of electrostatics, namely the formation of salt bridges and hydrogen bonds is adequately taken care of by explicit geometrical hydrogen bond potentials of the Lennard–Jones 12-10 type. As discussed earlier [37], such an approach avoids the difficulties of calculating and calibrating charges for each new ligand molecule, and of treating dielectric effects in detail. Each hydrogen bond formed by the designed ligand has to be associated with a strong interaction energy term as otherwise, solvent hydrogen-bonding competition would render the formation of such a bond highly improbable. In this respect, satisfying all free hydrogen-bonding 'valencies' has to be considered essential. Van der Waals interactions are calculated through a Lennard–Jones 12-6 potential and allow an easy check of steric clashes. Shape complementarity is evaluated through visual inspection of the interface of the molecular surfaces of the protein and the designed ligand. Hydrophobic interactions are implicitly incorporated in our design strategy and are computationally not evaluated.

Searching through chemical databases

Analogues of the designed fragments are useful in two ways: (1) they can be used for crystallographic soaking experiments; and (2) they provide new entry points for further design. We will now describe a four-step strategy which identifies such compounds in 2D chemical databases without the necessity of transforming the complete databases into 3D. Firstly, the designed fragment is analysed in the protein environment for all possible substitution patterns which do not disrupt the complementarity. Secondly, the fragment, together with the allowed substitutions, is translated into a 2D connectivity representation. Thirdly, this 2D topological screen is used for searching a chemical database. Finally, the fourth step in our strategy consists of evaluating all retrieved compounds for complementarity in the design pathway using the aforementioned force-field calculations and visual inspection at the PS390 workstation. For this purpose, the selected 2D structures are converted into 3D using the 'fast builder' of the BIOGRAF package [34]. Subse-

quently, they are oriented and translated in the protein environment by superposition techniques, and eventually submitted to energy minimization. In case of chiral structures, both enantiomers are examined.

In principle, screening of databases of commercially available compounds provides the fastest feedback to new crystallographic soaking experiments. For this paper we performed a substructure search of the Janssen Chimica Database (about 12,500 compounds) using the OSAC program [38]. It should be noted that this program equally allows searches using similarity screens, thereby suggesting sometimes completely different frameworks for fragments (e.g. different sizes of ring systems). Besides using this small database, we also went through the Chemical Abstracts (about 10,000,000 compounds) using on-line services.

RESULTS

Significant differences between trypanosomal and human TIM 'near' the active site

The design of selective inhibitors of gTIM requires the structures of both the human and the glycosomal enzyme. Unfortunately, no crystal structure is available for human TIM. However, the chicken TIM 3D structure [39] is probably a good model as the sequence identity is 89% [40, 41]. Using crystal-structure coordinates of the complex between gTIM and glycerol-3-phosphate [32] together with the sequence alignment [42], the solvent-accessible topographical differences near the active site between trypanosomal and human TIM were investigated within a radius of 15.0 Å from the phosphate/phosphonate binding site, hereafter referred to as the 'anchor position'. Out of the 48 residues contributing to the examined molecular surface, 9 residues differ between the two enzymes (Table 2). Of these, residues 400 and 401 look promising to confer selectivity to inhibitors. The trypanosomal dipeptide Ala^{400}-Tyr^{401} has a completely different His-Val dipeptide as a counterpart in human TIM. His-Val is also present in chicken TIM, which justifies its use as a model for the human TIM structure. Furthermore, the two residues forming 'the selec-

TABLE 2
TARGET RESIDUES FOR CONFERRING SELECTIVITY TO INHIBITORS OF gTIM: THEIR COUNTERPARTS IN HUMAN AND CHICKEN TIM, TOGETHER WITH THE DISTANCE TO THE P ATOM POSITION OF THE MOLECULAR ANCHOR ARE GIVEN

gTIM	Human TIM	Chicken TIM	Distance (Å)[a]
Ser^{71}	Thr	Pro	13.7
Ala^{400}	His	His	14.2
Tyr^{401}	Val	Val	11.3
Tyr^{402}	Phe	Phe	16.0
Thr^{430}	Lys	Lys	15.5
Val^{477}	Thr	Thr	9.1
Asn^{515}	Thr	Thr	5.6
Lys^{517}	Ala	Gly	13.7
Asn^{518}	Thr	Asn	9.4

[a] Closest distance from any solvent-accessible atom of the gTIM residue to the phosphate P atom position as observed in the complex between gTIM and glycerol-3-phosphate [32].

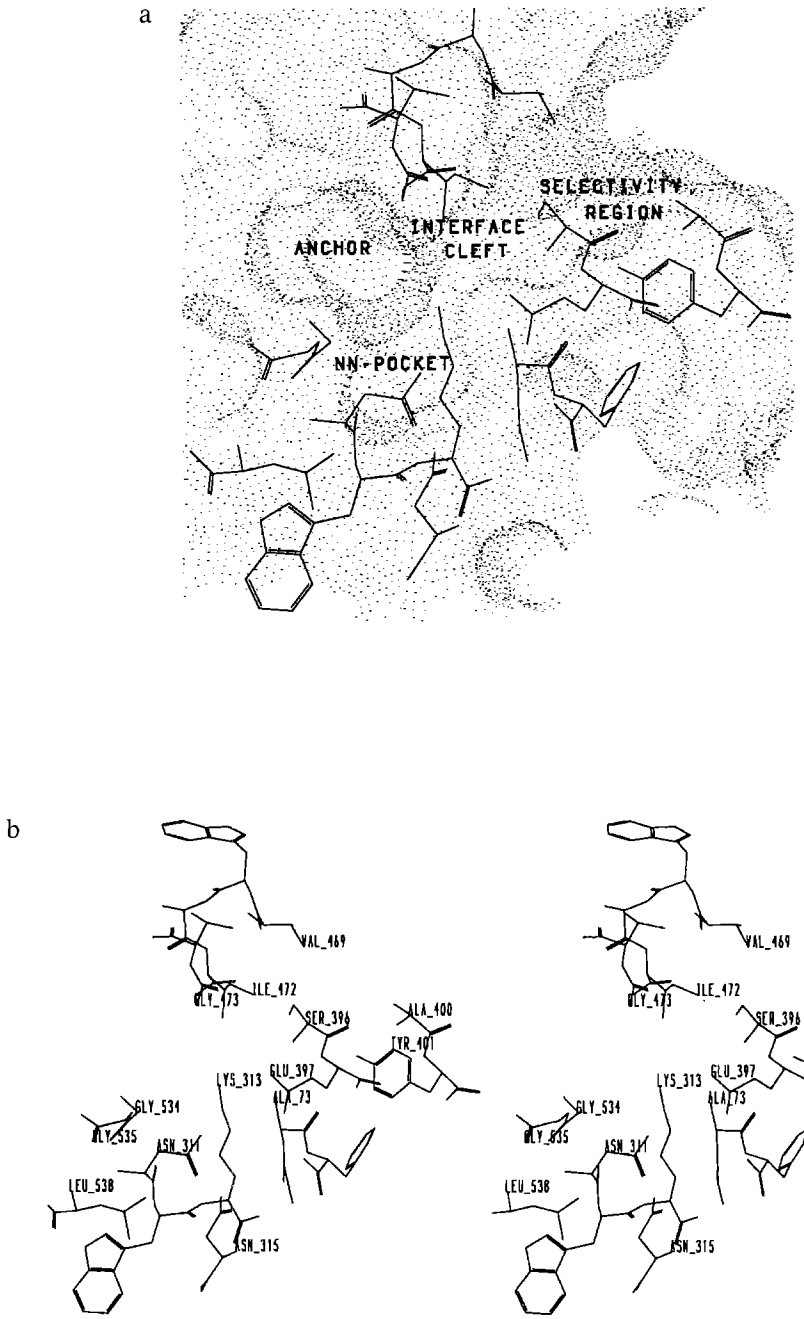

Fig. 1. (a) Molecular surface of the design pathway indicating the four subregions chosen for fragment design. The active site of gTIM is buried by the flexible loop, i.e. the residues immediately above the labels 'ANCHOR' and 'INTERFACE CLEFT'; (b) Stereoview of the design pathway; solvent-accessible residues are labelled. Notice the residues Ala^{400} and Tyr^{401} which are replaced in human TIM by His and Val respectively.

tivity region' can be reached from the anchor position along a groove on the surface of the protein; this groove spans about 15 Å (Fig. 1). Along this 'natural' design pathway two more regions can be discerned: the 'NN pocket', and the 'interface cleft'. Each of these four subregions of the design pathway will now be analysed in detail. This is greatly facilitated by using Table 3, which contains a complete list of all accessible hydrogen bond and hydrophobic partners in the design environment.

TABLE 3
CONTRIBUTORS TO THE 4 SUBREGIONS OF THE DESIGN PATHWAY

	Hydrogen bond		Hydrophobic	
	Donor	Acceptor	Atom[a]	Surface[b]
Anchor region				
Gly^{473}	N		C^{α}	21.2
Gly^{534}	N			
Gly^{535}	N		C^{α}	21.1
NN-pocket				
Ala^{73}			C^{α}	2.8
			C^{β}	24.6
Asn^{311}		O		
Lys^{313}			C^{δ}	9.2
			C^{ε}	7.5
Asn^{315}	$N^{\delta 2}$			
Gly^{535}			C	3.0
			C^{α}	21.1
Leu^{538}			C^{γ}	4.9
Interface cleft				
Ala^{73}			C^{β}	24.6
Lys^{313}[c]	N^{ξ}			
Glu^{397}[c]		O^{ε}		
Ile^{472}		O	$C^{\gamma 2}$	24.0
			$C^{\delta 1}$	4.1
Gly^{473}		O	C^{α}	21.2
Selectivity region				
Ser^{396}	O^{γ}	O^{γ}		
Glu^{397}[c]		O^{ε}	C^{γ}	10.3
Ala^{400}			C^{β}	28.7
Tyr^{401}	O^{η}	O^{η}	$C^{\delta 1}$	13.1
			$C^{\varepsilon 1}$	17.2
Val^{469}			$C^{\gamma 1}$	28.9

[a] Some atoms contribute to more than one subregion.
[b] The contribution to the molecular surface is given in $Å^2$.
[c] N^{ξ} of Lys^{313} and O^{δ} of Glu^{397} form a salt bridge.

Analysis of the design pathway

The anchor region. This region of the design pathway is the only one for which experimental information about ligand binding is available. Sulphate, monohydrogen phosphate, phosphate mono-esters or their phosphonate analogues are all bound by the same three Gly amide nitrogens: Gly^{534} and Gly^{535} at the N-terminus of a 3_{10}-helix, and Gly^{473}, which is near the tip of the flexible loop of the enzyme. Of the oxygen atoms of the dianion moiety of the ligands only the one bound to Gly^{535} is pointing towards the design pathway. Starting from this oxygen and going into the design pathway we encounter right next to the amide nitrogens the C^{α} atoms of Gly^{473} and Gly^{535} facing each other across a groove at a distance of 6.9 Å. The hydrophobic contribution of these two C^{α} atoms to the putative inhibitor binding region is quite substantial (Table 3). Next to this site, the design pathway splits into two directions, one leading into the NN-pocket and another one going into the interface cleft.

The NN-pocket. This pocket is named after the residues providing the hydrogen bonding valencies of this pocket, namely Asn^{311} and Asn^{315}. For our purposes, it is worth mentioning that the side chain of Asn^{315} is kept in place by forming a hydrogen bond to the main-chain carbonyl of Trp^{312}. Hydrophobic contributions to the NN-pocket come from Ala^{73}, Lys^{313}, Gly^{535} and Leu^{538}. The pocket shape of this region limits the size of complementary fragments to 5 or 6 nonhydrogen atoms. As is obvious from Fig. 1, it is not really necessary to occupy this part of the design pathway for obtaining selective gTIM inhibitors. However, tight binding at this subsite, which is a true pocket, may provide us with an alternative 'molecular anchor' which does not require the closing of the flexible loop of the enzyme for binding.

The interface cleft. This cleft is formed by loops stemming from different subunits of the enzyme: the loop between strand β3 and helix α3 of subunit one, and the flexible loop of subunit two. The main contribution from the loop of subunit one is the large hydrophobic surface of the C^{β} atom of Ala^{73}. At the bottom of the cleft there is a salt bridge between Lys^{313} and Glu^{397}. It is worth mentioning that the N^{ξ} of Lys^{313} also forms a water-mediated hydrogen bond to the dianion moiety of the inhibitors in all of the complexes determined crystallographically thus far. The flexible loop provides hydrophobic contributions through Ile^{472} and Gly^{473}. These residues also provide hydrogen bond acceptors at the edge of the cleft through their main-chain carbonyls. From Glu^{397} and Val^{469} onwards, the interface cleft continues into the selectivity region.

The selectivity region. The selectivity region derives its structure largely from the helix 395–402, also called 'the active-site helix' due to the presence of the catalytically essential His^{395}. Another, but much smaller, part of the region is a contribution from the flexible loop, namely Val^{469}. One of the main characteristics of the selectivity region is the presence of a substantial hydrophobic patch formed by the side chains of Val^{469}, Ala^{400} and Tyr^{401}. Such a continuous hydrophobic patch is absent in human TIM due to the presence of a His instead of Ala^{400}. Furthermore, two hydroxyl groups, one from Ser^{396} and one from Tyr^{401}, and one of the oxygens of the carboxylate of Glu^{397} point into this region. In the human enzyme an equivalent for the hydroxyl of Tyr^{401} is lacking as the corresponding residue is Val^{401}.

Complementary fragment design

The anchor region. Exploiting as much structural information as possible obtained through X-ray crystallography, all fragments were designed to contain phosphate as a molecular anchor. Phosphonates were not considered as good alternatives since the comparison between the K_i val-

ues of both types of compounds for the known inhibitors shows a 10^3-fold decrease in affinity [43]. The smallest ester providing hydrophobic partners for both C^{α}s in the anchor region is ethylphosphate (Fig. 2a). Its conformation has to be such that the ethyl group is antiperiplanar with respect to the oxygen hydrogen bonded to Gly^{534}. The terminal methyl group can then be substituted for linking this fragment to fragments in the NN-pocket and the interface cleft.

The NN-pocket. Due to the steric restrictions imposed by this narrow pocket only the design of a functionalized aliphatic chain is possible here. As an example we discuss here 4-hydroxy-2-but-

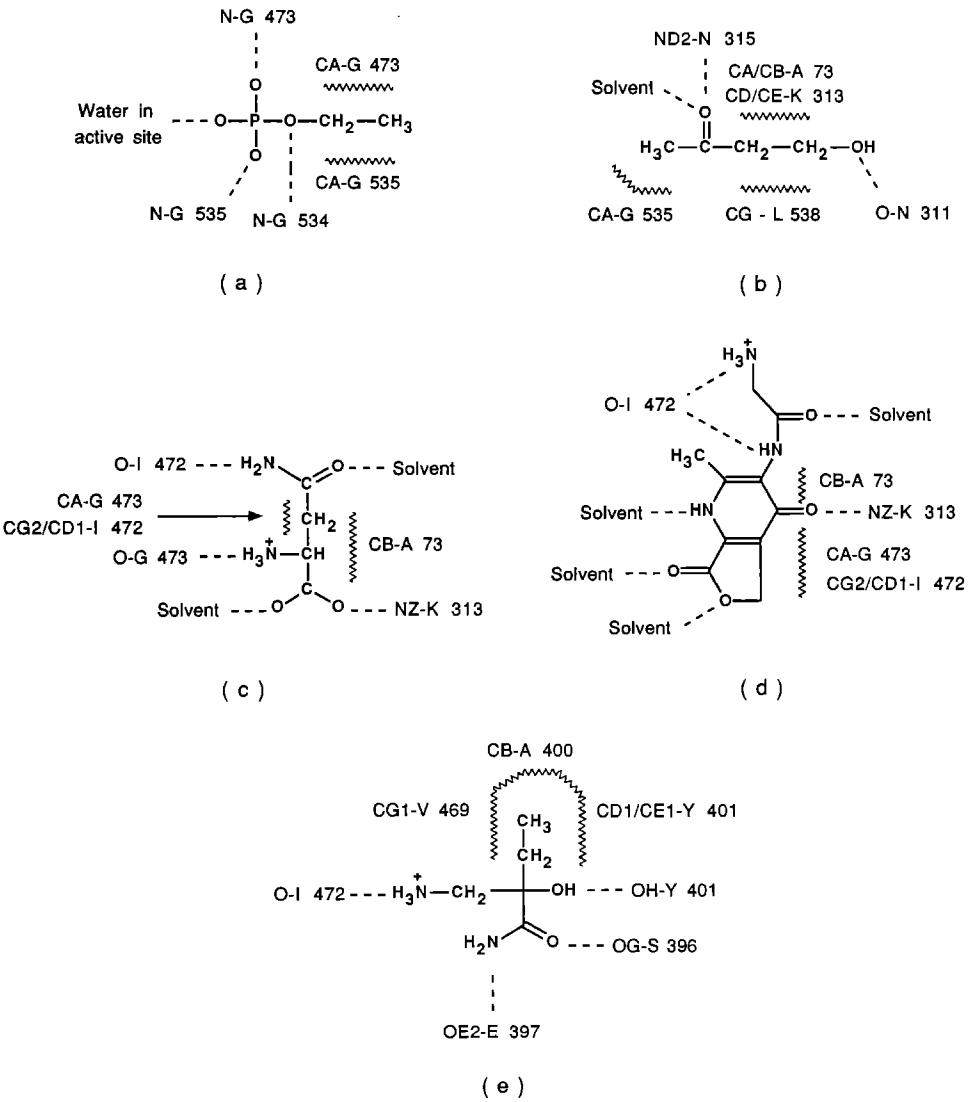

Fig. 2. Complementary fragment design (hydrogen bonds are indicated by dashed lines, hydrophobic interactions are interfaced by wavy lines): (a) by ethylphosphate in the anchor region; (b) by 4-hydroxy-2-butanone in the NN-pocket; (c) by D-asparagine and (d) by GMFP in the interface cleft; (e) by (*R*)-2-aminomethyl-2-hydroxy-butanamide in the selectivity region.

anone, which is in a strainless synclinal/antiperiplanar conformation for the dihedrals C1-C2- C3-C4 and C2-C3-C4-OH, respectively (Fig. 2b). The keto function of this compound provides Asn^{315} with a hydrogen bond partner while the primary alcohol can nicely interact with the main-chain carbonyl of Asn^{311}. It should be mentioned that the two oxygen atoms of the compound occupy the very same positions as two water molecules observed in the crystal structures of the gTIM complexes; together with a third bridging water molecule, these waters ought to be released upon binding of the designed compound. From the point of view of packing efficiency, the filling of the pocket by this molecule is nearly optimal.

The interface cleft. Two fragments complementary to this region, an acyclic and a cyclic one, will be presented, namely D-asparagine (Fig. 2c) and 3-glycinamido-2-methyl-1H,5H-furo[3,4-b]-pyridin-4,7-dione, abbreviated as GMFP (Fig. 2d). D-asparagine exploits all the protein hydrogen bond partners in the interface cleft and has two extra polar atoms pointing into the solvent. Acylation of the α-amino group might in further stages of the design be desirable in order to have a less charged compound. Modelling suggests that such a compound could still be accommodated by the protein. The cyclic compound GMFP does not exploit the main-chain carbonyl of Gly^{473} and has three polar atoms pointing into the solvent. A methyl group on the 2-position of the γ-pyridone ring should force the neighbouring amide substituent out of the plane of the ring, thereby improving the hydrogen bond of the amide to the main-chain carbonyl of Ile^{472}. Extensive hydrophobic interactions of the fused ring system with partners at both sides of the interface cleft are present.

The selectivity region. An example of a compound complementary to this region is shown in Fig. 2e. This compound, which has the (*R*)-configuration, is in fact derived from the modified amino acid C-ethyl-azaserine by replacing the amino group by an alcohol and the carboxylate by its amide. The ethyl group has extensive hydrophobic interactions with all hydrophobic partners in the selectivity region. It might be further extended into a big adjacent canyon which is accessible via a narrow groove between Val^{469} and Ala^{400}. This canyon has yet another exploitable difference between gTIM and human TIM at its surface, namely Thr^{430} which is a lysine in human TIM.

Linking the fragments

As might be expected, flexible fragments were the easiest ones to link into a potential lead compound for drug design. This can be explained by their inherent adaptability to small conformational changes without the introduction of substantial intramolecular strain. As an example we will discuss how the following four fragment molecules could be linked into a hypothetical molecule: (**1**) ethyl-phosphate, (**2**) 4-hydroxy-2-butanone, (**3**) D-asparagine, (**4**) (*R*)-2-amino-methyl-2-hydroxy-butanamide. We have given this molecule the code name DG_11.

In a first stage fragments (**1**) and (**3**) were merged. The distance between the terminal methyl group of (**1**) and the carboxylate oxygen exposed to the solvent in (**3**) is 2.5 Å. Therefore, we extended the ethyl-phosphate into propylamine-3-phosphate and carried out an in computro acylation of the amine group by D-asparagine. Subsequently it was found that the extra methylene unit added to compound (**1**) is at 1.6 Å of the terminal methyl group of compound (**2**), which is clearly within bonding distance. In this way fragments (**1**), (**2**) and (**3**) were linked. The addition of fragment (**4**) caused no problems since the primary amine functions of fragments (**3**) and (**4**) virtually coincide. Thus, the original amide of D-asparagine could be retained. After this linking process,

in which several intermediate energy minimizations were carried out, the entire DG_11 molecule was subjected to energy minimization in the protein environment. This resulted in a conformation with little strain and favourable interactions with the protein (Fig. 3) as exemplified by the intermolecular van der Waals energy of -198 kJ mol^{-1} and hydrogen bond energy of -252 kJ mol^{-1}. (As discussed in the Data and Methods section, we consider these energy values only as a check the quality of the complementarity.) Interestingly, the obtained conformation is also stabilized by an intramolecular hydrogen bond, namely between the alcohol function of the original fragment (**4**) and the amide carbonyl between fragments (**3**) and (**4**). With respect to stereochemistry, the linking process has created one extra chiral centre at the bridging methine between the fragments (**1**), (**2**) and (**3**): the absolute configuration is (*S*). Considering the complete DG_11 molecule the three chiral centres in the molecule are nicely separated in building blocks by the two amide bonds formed by the original building block (**3**).

Designed fragments as a link to database search methods

Compound DG_11 was analysed in the gTIM environment for allowed substitution patterns (Fig. 4). As the nature of possible substituents in a substructure search query hit is unknown in advance, only sterical considerations were taken into account. On five carbon atoms an extra substituent might be accommodated since then substituents are likely to point into the solution. Fur-

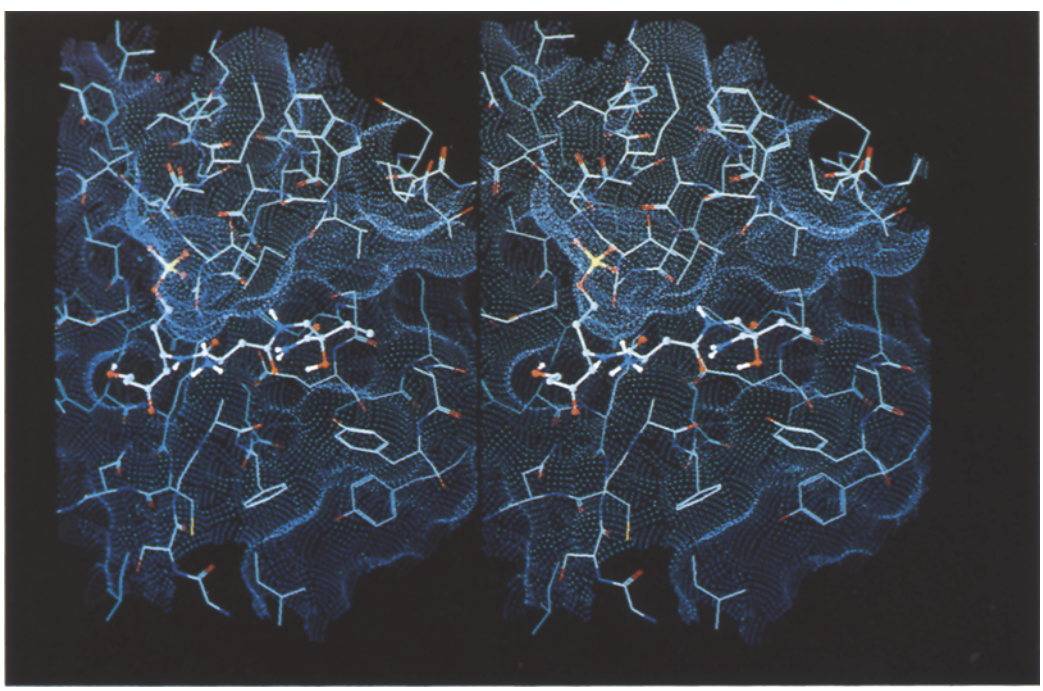

Fig. 3. Stereo colour picture of the designed DG_11 in the design pathway. The molecular surface of gTIM is represented by light blue dots, DG_11 by a ball-and-stick model, and the protein by vectors. The following half-bond colour code has been adopted: C, aqua (grey for DG_11); H, white; N, blue; O, red; and P, yellow.

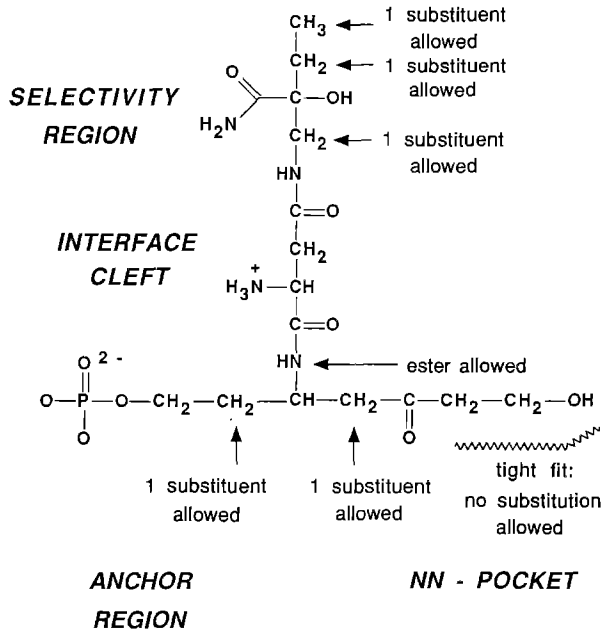

Fig. 4. Allowed substitution pattern of DG_11. The chiralities of the asymmetric carbon atoms are as follows: (*R*) in the upper part, (*R*) in the middle part, (*S*) in the lower part.

thermore, the amide linking fragments (**1**)-(**2**) with (**3**) is pointing into the solvent; it might possibly be replaced by an ester.

Within the limits imposed by the allowed substitution pattern, substructure searches and 'similar compound' searches (with a minimal similarity threshold of 80%) of fragments (**1**), (**2**), (**3**) and (**4**) were performed in the Janssen Chimica database. This revealed little more than the trivial commercial availability of D-asparagine and D-aspartic acid. In view of the limited success when using a small database, further substructure searches were carried out in the Chemical Abstracts database. No matching structures were found for fragment (**4**), the one designed in the selectivity region. However, a search for a fragment containing the anchor region fragment (**1**) linked to part of the interface cleft fragment (**3**) was successful. The different stages of this search are listed in Fig. 5.

After the straightforward substructure search, polymers and esters of common fatty acids (lauric acid, myristic acid, stearic acid, oleic acid, etc.) were eliminated as merely their size makes complementarity to the design pathway very unlikely. In the final stage, requiring human intervention with the aid of the modelling package, most of the 140 remaining molecules were eliminated on sterical grounds. Eventually, six compounds met all complementarity criteria without being constrained into unfavourable conformations, i.e. they match the protein surface in terms of shape, hydrophobicity, hydrogen bond properties and electrostatics.

Some complementarity details require further comment. The large exposure to solvent of the terminal methyl group of molecule (**a**) in Fig. 6 is probably not optimal. Its analogue, the urea-derivative (**b**), avoids this problem. Compound (**d**) is the ester analogue of amide (**a**), and has in addition an R_1 (Fig. 5) hydroxyl which points into the solvent. Molecule (**e**) in Fig. 6 is derived from

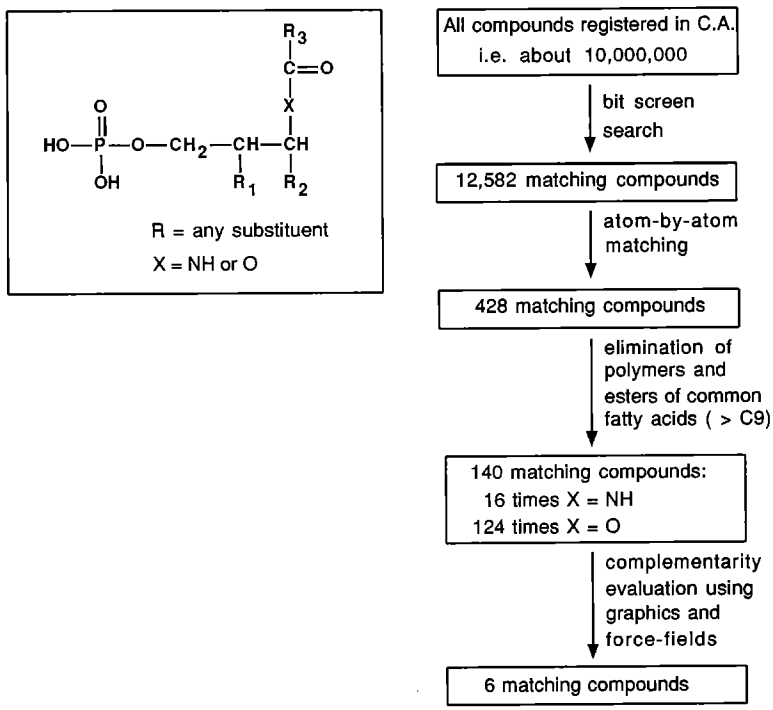

Fig. 5. The combined Chemical Abstracts–Molecular Modelling strategy. The substructure search fragment is enclosed within a box. Details of the search procedure are described in the text.

(**d**) through acetylation of the primary alcohol: it may make an extra hydrogen bond to $N^{\delta 2}$ of Asn^{515}. The glutamic acid part of molecule (**f**) reaches the O^{η} of Tyr^{401} with its amino function, and thus can form a hydrogen bond; the carboxylate group, however, is only pointing into the solvent. Potentially more promising is compound (**c**), which occupies a large part of the selectivity region with its benzyloxycarbonyl system: hydrophobic interactions of the aromatic ring with the C^{β} of Ala^{400} and with the $C^{\delta 1}$ and $C^{\varepsilon 1}$ of Tyr^{401} should be present in addition to a hydrogen bond between the oxy function and O^{η} of Tyr^{401}. The amide linking the oxycarbonyl to the search fragment makes a hydrogen bond to the O of Ile^{472}.

CONCLUDING REMARKS

The strategy of constructing DG_11 clearly illustrates that it is possible to design large inhibitors even in the absence of a known ligand in a macromolecular binding site. Characteristically for the method, the resulting DG_11 exhibits a very high degree of complementarity to the gTIM environment. This has been achieved through the systematic use of a detailed inventory of the design pathway (Table 3 in the case of gTIM). Great care has been taken not to deprive any hydrogen bond donor or acceptor of a partner upon complex formation. In addition, hydrophobic interactions have been engineered wherever possible. In this respect, it is worth mentioning that 213 $Å^2$ of hydrophobic surface, of both protein and DG_11, becomes buried. If one uses a recent estimate of the hydrophobic effect, about 197 J mol^{-1} $Å^{-2}$ [50], this would correspond to a contribution

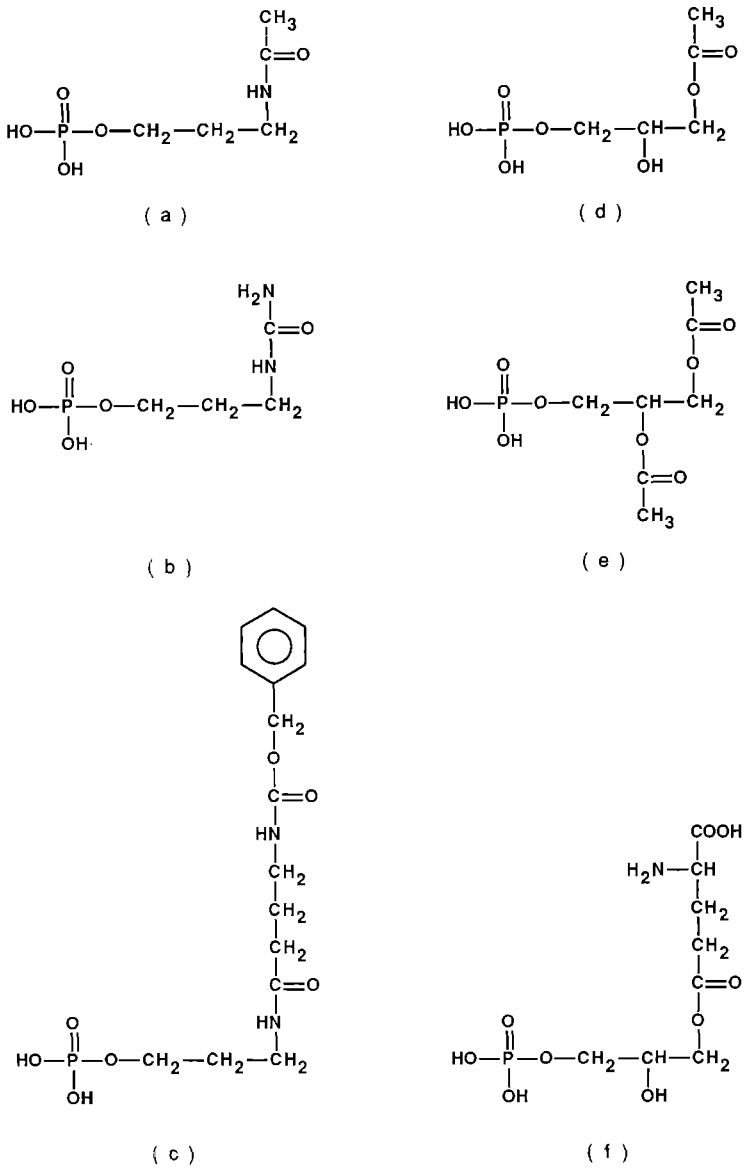

Fig. 6. Molecules matching the design pathway and retrieved through the search described in Fig. 5; the RN codes refer to the CAS-registry numbers of the compounds: (a) RN 114252-62-9 [44]; (b) RN 60855-11-0 [45]; (c) RN 75821-97-5 [46]; (d) RN 91573-82-2 [47]; (e) RN 65359-84-4 [48]; (f) RN 109735-22-0 [49].

of 41.9 kJ mol^{-1} to the binding energy. As still some 108 $Å^2$ of hydrophobic surface is not covered by DG_11 future designs may improve upon this. More important, however, is the need to restrict the conformational flexibility of the current design which possesses 18 torsional degrees of freedom. The ring compound fragment GMFP modelled in the interface cleft is a first step in this direction.

A limitation of the linked-fragment approach is the modelling in a fixed protein environment. Although this could be remedied by using sophisticated but computationally demanding molecular dynamics or Monte Carlo annealing strategies, we consider it more reliable to proceed to crystallographic soaking experiments with the designed compounds. At this stage the linked-fragment approach should especially prove its value. Instead of waiting for the synthesis of such a complicated molecule as DG_11 to be accomplished, one can initiate soaking experiments with its constituent fragments or fragment analogues revealed through the database searches. Synthesis of these less complicated molecules and soaking experiments are underway. We believe that our protein structure-based linked-fragment approach may be a general method for virtually ab initio inhibitor design studies.

ACKNOWLEDGEMENTS

The most stimulating discussions with Prof. K. Müller of Hoffmann-La Roche, Basle, and the many suggestions he made are gratefully acknowledged. It is a pleasure to thank Drs. R.K. Wierenga and M.E.M. Noble of the EMBL in Heidelberg for providing the coordinates of several of the gTIM complexes, Drs. F.R. Opperdoes and M. Callens of the ICP in Brussels for determining inhibition constants, and Mr. U. Kooystra and Mr. J. Kuindersma for assistance with database searches. This research project was supported by a grant of the WHO/UNDP/World Bank Special Programme for Research and Training in Tropical Diseases; by a fellowship from Hoffmann-La Roche, Basle (to C.L.M.J.V.); and by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organisation for Scientific Research (NWO). Use of the services and facilities of the Netherlands National NWO/SURF Expertise Center CAOS/CAMM, under grant numbers SON 326-052 and STW NCH99.1751 is gratefully acknowledged.

REFERENCES

1 Maugh, T.M., Science, 196 (1977) 413.
2 Falco, E.A., Goodwin, L.G., Hitchings, G.H., Rollo, I.M. and Russell, P.B., Brit. J. Pharmacol., 6 (1951) 185.
3 Hol, W.G.J., Angew. Chem. Int. Ed. Engl., 25 (1986) 767.
4 Erickson, J., Neidhart, D.J., VanDrie, J., Kempf, D.J., Wang, X.C., Norbeck, D.W., Plattner, J.J., Rittenhouse, J.W., Turon, M., Wideburg, M., Kohlbrenner, W.E., Simmer, R., Helfrich, R., Paul, D.A. and Knigge, M., Science, 249 (1990) 527.
5 Lewis, R.A. and Dean, P.M., Proc. R. Soc. Lond., B236 (1989) 125.
6 Lewis, R.A., J. Comput.-Aided Mol. Design, 4 (1990) 205.
7 Goodford, P.J., J. Med. Chem., 28 (1985) 849.
8 Boobbyer, D.N.A., Goodford, P.J., McWhinnie, P.M. and Wade, R.C., J. Med. Chem., 32 (1989) 1083.
9 Lesk, A.M., Acta Crystallogr., A42 (1986) 83.
10 Allen, F.H., Bellard, S., Brice, M.D., Cartwright, B.A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B.G., Kennard, O., Motherwell, W.D.S., Rodgers, J.R. and Watson, D.G., Acta Crystallogr., B35 (1979) 2331.
11 Rusinko III, A., Sheridan, R.P., Nilakantan, R., Haraki, K.S., Bauman, N. and Venkataraghavan, R., J. Chem. Inf. Comput. Sci., 29 (1989) 251.
12 DesJarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D. and Venkataraghavan, R., J. Med. Chem., 31 (1988) 722.
13 Sheridan, R.P., Nilakantan, R., Rusinko III, A., Bauman, N., Haraki, K.S. and Venkataraghavan, R., J. Chem. Inf. Comput. Sci., 29 (1989) 255.
14 Sheridan, R.P., Rusinko III, A., Nilakantan, R. and Venkataraghavan, R., Proc. Natl. Acad. Sci. U.S.A., 86 (1989) 8165.

15 VanDrie, J.H., Weiniger, D. and Martin, Y.C., J. Comput.-Aided Mol. Design, 3 (1989) 225.
16 DesJarlais, R.L., Seibel, G.L., Kuntz, I.D., Furth, P.S., Alvarez, J.C., Oritz de Montellano, P.R., De Camp, D.L., Babe, L.M. and Craik, C.S., Proc. Natl. Acad. Sci. U.S.A., 87 (1990) 6644.
17 Hol, W.G.J., In Blöcker, H., Collins, J., Schmid, R.D. and Schomburg, D. (Eds.) Advances in Protein Design, GBF Monographs, VCH Verlaggesellschaft, Weinheim, 1989, pp. 27–34.
18 WHO, TDR, Seventh Program Report, Tropical Disease Research, WHO, Geneva, 1985.
19 Haller, L., Adams, J.H., Merouse, F. and Dago, A., Am. J. Trop. Med. Hyg., 35 (1986) 94.
20 Tropical Drug Research News, 34 (1990) 1.
21 Bellofatto, V., Fairlamb, A., Henderson, G.B. and Cross, G.A.M., Mol. Biochem. Parasitol., 25 (1987) 227.
22 Tropical Drug Research News, 35 (1991) 7.
23 Hol, W.G.J., Wierenga, R.K., Groendijk, H., Read, R.J., Thunnissen, A.M.W.H., Noble, M.E.M., Kalk, K.H., Vellieux, F.M.D., Opperdoes, F.R. and Michels, P.A.M., In Roberts, S.M. (Ed.), Molecular Recognition: Chemical and Biochemical Problems, The Royal Society of Chemistry, 1989, pp. 84–93.
24 Opperdoes, F.R., Annu. Rev. Microbiol., 41 (1987) 127.
25 Misset, O., Bos, O.J.M. and Opperdoes, F.R., Eur. J. Biochem., 157 (1986) 441.
26 Opperdoes, F.R., Wierenga, R.K., Noble, M.E.M., Hol, W.G.J., Willson, M., Kuntz, D.A., Callens, M. and Perié, J., In Parasites: Molecular Biology, Drug and Vaccine Design, Wiley-Liss, Inc., 1990, pp. 233–246.
27 Wierenga, R.K., Kalk, K.H. and Hol, W.G.J., J. Mol. Biol., 198 (1987) 109.
28 Wierenga, R.K., Noble, M.E.M., Vriend, G., Nauche, S. and Hol, W.G.J., J. Mol. Biol., 220 (1991) 995.
29 Lambeir, A.M., Opperdoes, F.R. and Wierenga, R.K., Eur. J. Biochem., 168 (1987) 69.
30 Pompliano, D.L., Peyman, A. and Knowles, J.R., Biochemistry, 29 (1990) 3186.
31 Verlinde, C.L.M.J., Noble, M.E.M., Kalk, K.H., Groendijk, H., Wierenga, R.K. and Hol, W.G.J., Eur. J. Biochem., 198 (1991) 53.
32 Noble, M.E.M., Wierenga, R.K., Lambeir, A.M., Opperdoes, F.R., Thunnissen, A.M.W.H., Kalk, K.H., Groendijk, H. and Hol, W.G.J., Proteins, 10 (1991) 50.
33 Noble, M.E.M., Verlinde, C.L.M.J., Groendijk, H., Kalk, K.H., Wierenga, R.K. and Hol, W.G.J., J. Med. Chem., 34 (1991) 2709.
34 BIOGRAF 2.10: Molecular Simulations, Inc., 796 North Pastoria Ave, Sunnyvale, CA 94086, U.S.A.
35 Richards, F.M., Annu. Rev. Biophys. Bioeng., 6 (1977) 151.
36 Van Gunsteren, W.F. and Berendsen, H.J.C., Angew. Chem. Int. Ed. Engl., 29 (1990) 992.
37 Müller, K., Ammann, H.J., Doran, D.M., Gerber, P.M., Gubernator, K. and Schrepfer, G., In van der Goot, H., Domány, G., Pallos, L. and Timmerman, H. (Eds.) Trends in Medicinal Chemistry '88, Elsevier Science Publishers B.V., Amsterdam, 1989, pp. 1–12.
38 OSAC - Organic Structures Accessed by Computers: ORAC, Ltd., Leeds, UK, 1990.
39 Banner, B.W., Bloomer, A.C., Petsko, G.A., Philips, D.C., Pogson, C.I., Wilson, I.A., Corran, P.H., Furth, A.J., Milman, J.D., Offord, R.E., Priddle, J.D. and Waley, S.G., Nature, 255 (1975) 609.
40 Straus, D. and Gilbert, W., Proc. Natl. Acad. Sci. U.S.A., 82 (1985) 2014.
41 Maquat, L.E., Chilcote, R. and Ryan, P.M., J. Biol. Chem., 260 (1985) 3748.
42 Swinkels, B.W., Gibson, W.C., Osinga, K.A., Kramer, R., Veeneman, G.H., van Boom, J.H. and Borst, P., EMBO J., 5 (1986) 1291.
43 Waley, S.G., Biochem. J., 135 (1973) 165.
44 Cherbuliez, E. and Rabinowitz, J., Helv. Chim. Acta, 39 (1956) 1455.
45 Krueger, F. and Schmidt, G., Ger. Offen. DE 2453037 (1976).
46 Feuer, L., Furka, A., Hrcsel, J., Horvath, A. and Sebestyen, F., U.S. 4218404 (1980).
47 Moschidis, M.C. and Demopoulos, C.A., J. Chromatogr., 294 (1984) 403.
48 Sukhanov, V.A., Sergovskaya, N.L., Shvets, V.I. and Evstigneeva, R.P., Zh. Obshch. Khim., 47 (1977) 2130.
49 Laboratories le Brun, Fr. 1.072.327 (1954).
50 Sharp, K.A., Nicholls, A., Fine, R.F. and Honig, B., Science, 252 (1991) 106.