Journal of Computer-Aided Molecular Design, 6 (1992) 593–606
ESCOM

J-CAMD 183

LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads

Hans-Joachim Böhm

BASF AG, Central Research, 6700 Ludwigshafen, Germany

Received 10 July 1992
Accepted 17 August 1992

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

SUMMARY

Recent advances in a new method for the de novo design of enzyme inhibitors are reported. A new set of rules to define the possible nonbonded contacts between protein and ligand is presented. This method was derived from published statistical analyses of nonbonded contacts in crystal packings of organic molecules and has been implemented in the recently described computer program LUDI. Moreover, LUDI can now append a new substituent onto an already existing ligand. Applications are reported for the design of inhibitors of HIV protease and dihydrofolate reductase. The results demonstrate that LUDI is indeed capable of designing new ligands with improved binding when compared to the reference compound.

1. INTRODUCTION

The de novo design of protein ligands has recently gained increased attention [1–9]. Most effort so far has focused on the calculation of favorable binding sites [1–3] and on the docking of given ligands into the binding pocket of a protein [4,5]. A few groups have also reported on the automatic design of novel ligands [6–9].

Recently, I reported a new method for the de novo design of enzyme inhibitors, called LUDI [9]. This method is based on a statistical analysis of nonbonded contacts found in the Cambridge structural database (CSD) [10]. The first version of the program made direct use of the contact patterns retrieved from the CSD and utilized them to position small molecules or fragments in a cleft in a protein structure (e.g. an active site) in such a way that hydrogen bonds are formed with the protein and hydrophobic pockets are filled with suitable side chains of the ligand. In the first paper on LUDI [9] I presented a very simple set of rules to generate the positions of atoms on the basis of fragments found suitable to form favorable interactions with the protein. However, this first set of rules turned out to be too simplistic because it took into account only the most heavily populated hydrogen-bond geometries. The direct use of contact geometries from the CSD carries the danger that some potentially important contact patterns are not included because they have

0920-654X/$ 5.00

not yet shown up in the crystal structures of small molecules. One should keep in mind that despite the rather large number of structures (90 000) currently contained in the CSD (1991 version), the number of certain nonbonded contacts relevant for ligand protein interactions may be very small.

I therefore decided to develop a new set of rules for nonbonded contacts on the basis of the experimentally observed range of nonbonded contact geometries revealed by statistical analysis of the CSD [11–18]. This new set of rules is thought to have the advantage of covering the complete space of energetically favorable arrangements for hydrogen bonds and hydrophobic contacts. The analysis of the CSD is used to define the range of allowed angles and dihedrals (see Fig. 1 for definition of the angles and dihedrals) describing the nonbonded contact geometry. This space is then populated by discrete points (or vectors) that are equally spaced. The point density can be controlled by the user. Note that the data from the statistical analysis of the CSD are used merely to derive the allowed range of contact geometries. The rules derived from the CSD do not take into account the experimentally observed different populations of different contact geometries.

In addition, some other improvements to LUDI are reported concerning the positioning of fragments, the evaluation of positioned fragments and the possible prioritization of the structures found to fit the binding site of a protein. Another new functionality that has been added to LUDI is the ability to link a new fragment to an already existing ligand while forming hydrogen bonds with the protein and filling a hydrophobic pocket. This feature offers the important possibility to design new substituents for a given lead compound.

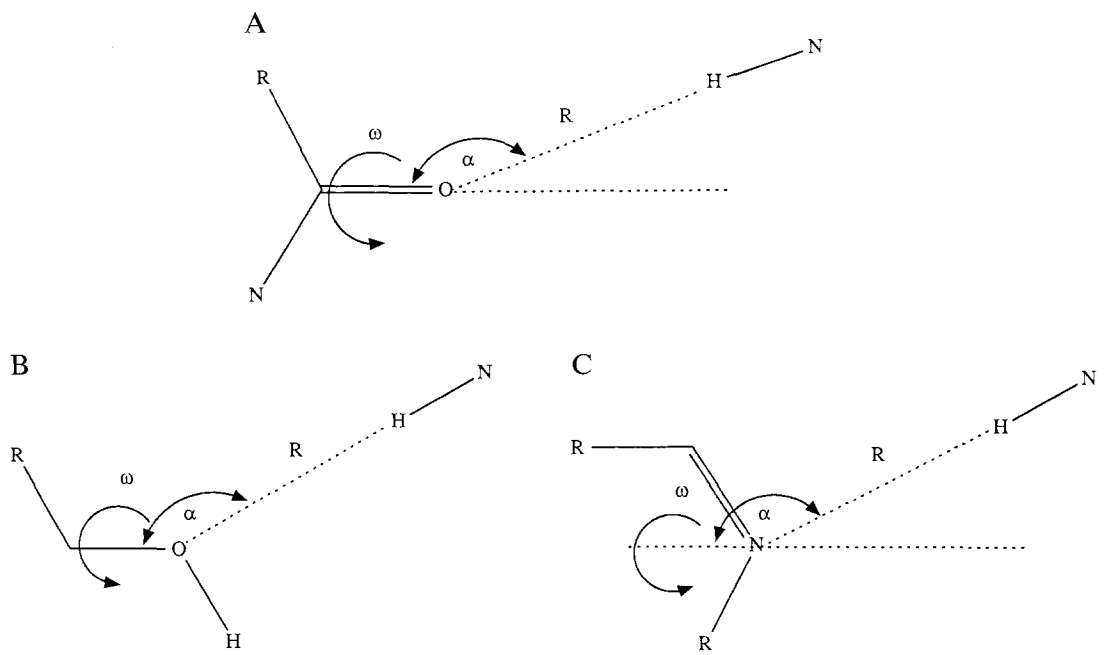


Fig. 1. Definition of the geometric parameters R, α and ω used in the rules for the allowed nonbonded contacts. A: definition for terminal groups; B: definition for -O-; C: definition for -N=. For -N= groups; α denotes the angle between the bisector of the angle C=N-R and the vector N..H.

Finally, LUDI was used to design new inhibitors of the aspartic protease of the human immunodeficiency virus (HIV) and dihydrofolate reductase (DHFR).

2. METHODOLOGY

2.1. A new set of rules to generate the potential interaction sites

Interactions between a protein and its ligand are usually formed through favorable nonbonded contacts such as hydrogen bonds or hydrophobic interactions. These contacts may be divided into individual interactions between single atoms or functional groups of the protein and the ligand. Thus, for every atom or functional group of the protein that is involved in binding with the ligand, there exists a counterpart on the ligand. This counterpart is again an atom or a functional group. For example, the counterpart for a carbonyl group C=O of the protein may be an amino group N-H of the ligand. A suitable position for such a functional group or atom of the ligand is referred to as its 'interaction site'. A statistical analysis of hydrogen-bond geometries in crystal packings of small molecules [11–18] reveals that there is a rather broad distribution of hydrogen-bond patterns. Therefore, for every functional group of the protein there exists not only a single position but also a region in space suitable for favorable interactions with the protein. In LUDI, this distribution of possible contact patterns is taken into account by using an ensemble of interaction sites distributed over the whole region of possible contact patterns. This approach has the advantage that it is purely geometrical and therefore avoids costly calculations of potential functions.

The definition of an interaction site has been given previously [9]. LUDI distinguishes between four different types of interaction sites:

1. hydrogen-donor,
2. hydrogen-acceptor,
3. lipophilic-aliphatic,
4. lipophilic-aromatic.

In LUDI, the hydrogen-donor and hydrogen-acceptor interaction sites are described by vectors (atom pairs) to account for the strong directionality of hydrogen bonds. Hydrogen-donor sites are represented by D-X vectors ($R_{D\text{-}X} = 1$ Å) and hydrogen-acceptor sites are represented by A-Y vectors ($R_{A\text{-}Y} = 1.23$ Å). The particular lengths for the vectors were chosen to correspond roughly to the N-H/O-H and C=O bond lengths, respectively. A suitable type of interaction site is selected for each functional group or atom of the enzyme. Then a user-defined number of interaction sites is positioned. This positioning is guided by the rules.

The rules used to generate the hydrogen-donor and hydrogen-acceptor interaction sites will now be described. For the hydrophobic contacts the same rules are used as given in my previous paper [9]. The position of an interaction site is described by the distance R, angle α and dihedral ω as defined in Fig. 1. The available experimental data on nonbonded contact geometries in crystal packings of small organic molecules are used to define the allowed values for R, α, and ω. The region in space defined by the values is then populated by discrete interaction sites. The distance between the interaction sites is typically 0.2–0.3 Å. The rules are summarized in Table 1.

The hydrogen-bond geometry of carbonyl groups in the solid state has been investigated extensively [11,12,15]. The available data show a distribution of α from 110° to 180° with a preference for the lone-pair direction ($\alpha = 120°$, $\omega = 0°, 180°$). However, as this preference is not particularly pronounced and the other regions are also significantly populated, an even distribution of interac-

tion sites was used, with $R_{O..D} = 1.9$ Å, $\alpha = 110–180°$ and $\omega = 0–360°$. The optimal O..D-X hydrogen bond is assumed to be linear ($<_{O..D-X} = 180°$). This distribution is applied for the backbone carbonyl groups and those in the side chains of the amino acids Asn and Gln.

The distribution of hydrogen-acceptor atoms around a N-H group falls into a smaller region in space than that around a carbonyl group. The statistical analyses that have been published [12,14,15] all show a strong preference for a linear hydrogen bond with $<_{N-H..O/N} = 150–180°$. A very similar distribution has also been found around the N-H group in aromatic rings [13,15]. The available data indicate similar distributions for N-H and O-H. Therefore, identical rules for both groups were used to generate interaction sites with $R_{H..A} = 1.9$ Å, $\alpha = 150–180°$ and $\omega = 0–360°$. This distribution was used for the backbone N-H groups and for the hydrogen-donor groups in the side chains of the amino acids His, Gln, Asn, Ser, Thr and Tyr. For charged amino groups, a slightly shorter hydrogen-bond length of $R_{H..A} = 1.8$ Å was used. This shorter hydrogen-bond length for charged groups has also been observed experimentally [14].

A problem arises with the generation of the position of the second atom, Y, adjacent to the hydrogen-acceptor position A. The optimal position of this second atom is difficult to obtain from available experimental data. The position of the site Y was thus generated assuming $<_{N-H..A-Y} = 0°$, $<_{H..A-Y} = 110–180°$ and $R_{A-Y} = 1.23$ Å, although the particular choice of the dihedral is admittedly somewhat arbitrary.

TABLE 1
GEOMETRIC PARAMETERS DESCRIBING THE ALLOWED RANGE OF NONBONDED CONTACT GEOMETRIES USED IN LUDI

Enzyme functional group	Interaction site	Geometric parameters	Reference
C=O	D-X	$R_{O..D} = 1.9$ Å $\alpha = 110–180°$ $\omega = 0–360°$	11,12,15
N-H,O-H	A-Y	$R_{H..A} = 1.9$ Å $\alpha = 150–180°$ $\omega = 0–360°$	12,14,15
N-H(charged)	A-Y	$R_{H..A} = 1.8$ Å $\alpha = 150–180°$ $\omega = 0–360°$	12,14,15
COO^-	D-X	$R_{O..D} = 1.8$ Å $\alpha = 100–140°$ $\omega - 50–50°, 130–230°$	16
=N-	D-X	$R_{N..D} = 1.9$ Å $\alpha = 150–180°$ $\omega = 0–360°$	13,15
R-O-R (sp^2)	D-X	$R_{O..D} = 1.9$ Å $\alpha = 100–140°$ $\omega = -60–60°$	13,15
R-O-R (sp^3)	D-X	$R_{O..D} = 1.9$ Å $\alpha = 90–130°$ $\omega = -70–70°$	12,15,18

The hydrogen-bond contact patterns around carboxylic acids have been studied by Görbitz and Etter [16]. The data indicate a preference for $<_{C=O..H}=120°$ and $<_{O\text{-}C\text{-}O..H}=0,180°$. These authors found no indication that *syn* hydrogen bonds are inherently more favorable than *anti* hydrogen bonds. Their data were translated into the following rules to generate the interaction sites around a carboxylic acid: $R_{O..D}=1.8$ Å, $\alpha=100$–$140°$, $\omega=-50..50°$, 130–230°.

The distribution of hydrogen donors around an unprotonated nitrogen in aromatic rings has been investigated by Vedani and Dunitz [13]. The distribution of hydrogen donors is narrower than that around a carbonyl group. The following rule (which applies to the unprotonated nitrogen in the side chain of His) is derived from the results of Vedani and Dunitz: $R_{N..D}=1.9$ Å, $\alpha=150$–$180°$, $\omega=0$–$360°$.

Hydroxyl groups can act both as hydrogen donors and as hydrogen acceptors. Although a detailed analysis of high-resolution protein structures [17] shows that hydroxyl groups act more often as donors than as acceptors, the possibility that hydroxyl groups act as acceptors has to be taken into account. For sp^3-oxygen, the data of Kroon et al. [18] indicate a preference for the donor group to lie in the plane of the lone pairs ($<_{C\text{-}O..H}=109\pm20°$). However, no evidence has been obtained for any preference of the lone-pair direction within this plane. This contrasts with data obtained by Vedani and Dunitz [13] and by Klebe [15], who report a preferred orientation of hydrogen-donor groups in the direction of the lone pairs. Since the experimental data are used merely to establish the allowed hydrogen-bond patterns, hydrogen bonds not pointing in the direction of the lone pair were also allowed for: $R_{O..D}=1.9$ Å, $\alpha=90$–$130°$, $\omega=-70..70°$. For sp^2-oxygen, as found in the side chain of Tyr, there is a clear preference for the hydrogen-donor groups to lie in the plane of the aromatic ring. The data of Vedani and Dunitz [13], Klebe [15] and Baker and Hubbard [17] were used to derive the following rule: $R_{O..D}=1.9$ Å, $\alpha=100$–$140°$ and $\omega=-50..50°$.

As most publications on statistical analyses do not present a quantitative analysis of the data, there is a certain amount of ambiguity involved in the choice of the rules given above. A very restricted definition of the allowed hydrogen-bond geometries would strongly reduce the number of hits obtained in the subsequent fragment fitting, and carries the risk of eventually missing some of the promising hits. On the other hand, a very broad definition would result in a very large number of hits, with the difficulty of selecting the most interesting ones. Thus, the present choice of rules represents a compromise.

The generated interaction sites were finally checked for van der Waals overlap with the protein.

2.2. Fragment linking

In my previous paper I described the 'bridge' mode which allows one to connect positioned fragments by suitable spacers. This concept has now been generalized. LUDI is now able to fit fragments onto the interaction sites and simultaneously link them to an already existing ligand or part of a ligand. For this purpose, 'link sites', which are X-H atom pairs suitable for appending a substituent to the ligand, can be specified by the user. Alternatively, the program assumes that all hydrogen atoms of the positioned ligand within a given cut-off radius, together with the heavy atoms they are bound to, are link sites.

LUDI can perform a single link, generating a single bond between the newly fitted fragment and the already existing ligand. Additionally, it is also possible to do a multiple link. The double link will generate two bonds between the newly fitted fragment and the existing ligand. For exam-

ple, it is possible to fuse a second phenyl ring onto an existing one to form a naphthyl group. This double link also includes the 'bridge-mode' as described previously [9]. The options are shown in Fig. 2.

In order to carry out the calculations in the link mode, a second library was built specifically for this purpose. The link sites (the atoms which form a bond with the already existing ligand) are explicitly defined for each entry in this library. Some examples are shown in Fig. 3. This library currently consists of 1100 entries. This number is larger than the number of entries in the standard library because, for many of the structures, there are several possible ways to form the link.

The link mode of LUDI is similar to the approach implemented in the computer program GROW by Moon and Howe [7]. The purpose of GROW is to construct peptides by linking amino acids, whereas LUDI attempts to construct arbitrary organic molecules. GROW is based on force-field calculations and will therefore be considerably slower than LUDI, because LUDI is completely based on geometric operations.

2.3. Prioritization of the fitted fragments

An important problem of every method based on searching through large numbers of structures is the prioritization of the hits. This problem is approached as follows:

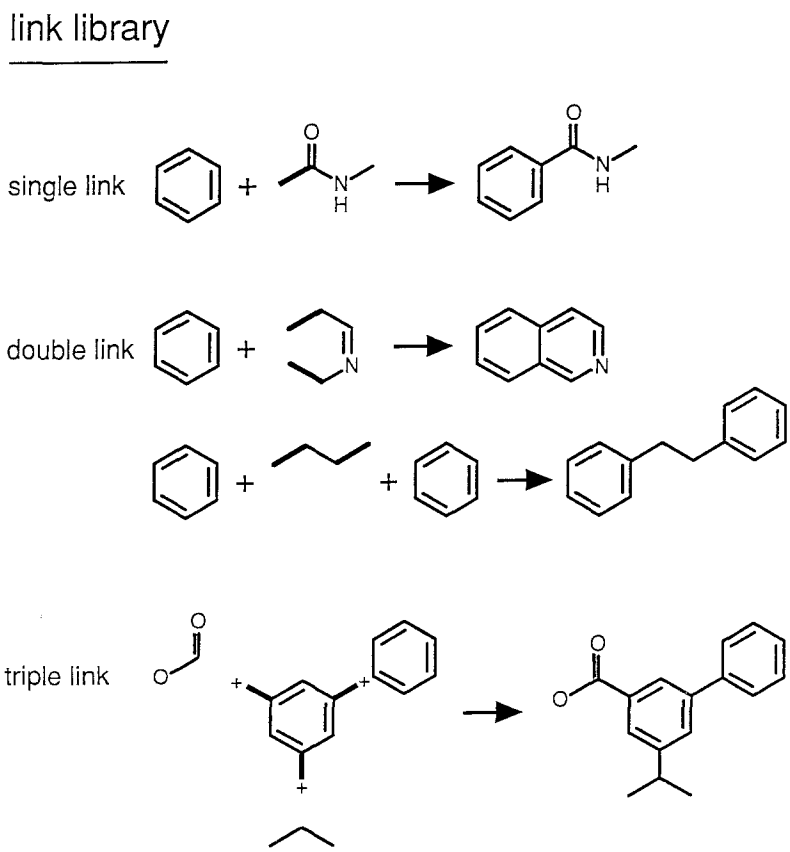


Fig. 2. Examples for a single, double and triple link as performed by LUDI in the link-mode.

Only those fragments with a root-mean-square (rms) deviation of the fit of the fragment onto the interaction sites below a certain threshold (typically 0.3–0.5 Å) are accepted. A further requirement for a successfully positioned fragment is that it does not overlap with the protein. LUDI also checks for electrostatic repulsion between protein and ligand: if a polar atom is closer to a protein atom of the same polarity than a threshold distance (typically 3.5 Å for O..O contacts),

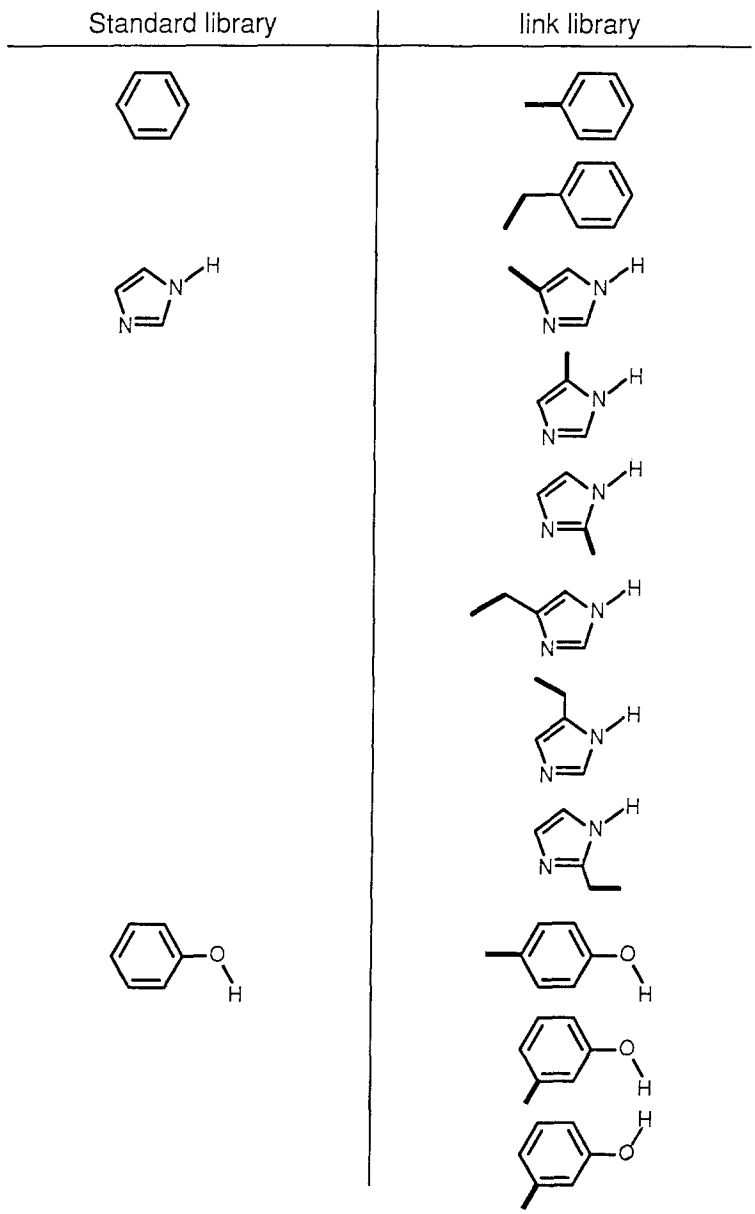


Fig. 3. Examples from the link library of LUDI. Each possible link that will be considered by LUDI has to be specified explicitly.

then the fit of the fragment is rejected. In the electrostatic repulsion check, only those protein atoms are taken into account that do not hydrogen bond with the ligand.

The number and quality of the hydrogen bonds between protein and ligand and the hydrophobic protein–ligand contact surface were then used to calculate a score. The relative weight of a hydrogen bond with respect to the hydrophobic interaction was derived from a value of 1.5 kcal/mol for the contribution of a hydrogen bond to the binding energy [19] and 25 cal/(mol Å^2) for the hydrophobic interaction [20]. Therefore, in the scoring function it is assumed that an unperturbed hydrogen bond has the same contribution to ligand binding as 60 Å^2 of hydrophobic contact surface. The following preliminary scoring function was used:

$$\text{Score} = \Sigma_{\text{hbonds}}\, 100 * f(\Delta R) * f(\Delta\alpha) + 5/3 * \text{NCONTACT}$$

$f(\Delta R) = 1, \Delta R \leq 0.2$ Å $\qquad f(\Delta\alpha) = 1, \Delta\alpha \leq 30°$

$f(\Delta R) = 1 - (\Delta R - 0.2)/0.4, \Delta R \leq 0.6$ Å $\qquad f(\Delta\alpha) = 1 - (\Delta\alpha - 30)/50, \Delta\alpha \leq 80°$

$f(\Delta R) = 0, \Delta R > 0.6$ Å $\qquad f(\Delta\alpha) = 0, \Delta\alpha > 80°$

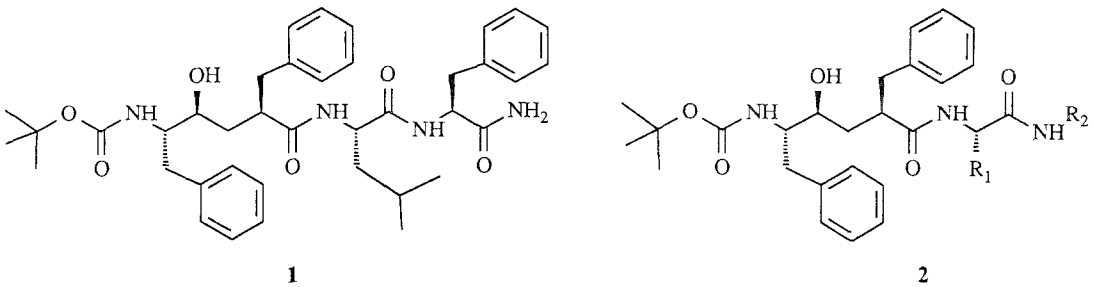
ΔR is the deviation of the H..O/N hydrogen-bond length from the ideal value 1.9 Å. $\Delta\alpha$ is the deviation of the hydrogen-bond angle $<_{\text{N/O-H..O/N}}$ from its ideal value 180°. NCONTACT represents the lipophilic contact area between protein and ligand in Å^2.

The scoring function was tested on the fit of fragments into the specificity pocket of trypsin and into the pteridine-binding site of dihydrofolate reductase. The fragments were taken from the standard LUDI library consisting of currently 800 fragments. For trypsin, the fragment with the highest score was benzamidine. In the case of DHFR, the highest score was found for the fragment 2,4-diamino-pteridine.

3. APPLICATIONS

3.1. Inhibitors of the HIV protease

As a first example, I report the application of LUDI to the design of inhibitors of the HIV-protease [21]. The 3D structure of the HIV-1 protease complexed with a peptidic inhibitor was recently solved by Wlodawer and coworkers [22] (entry 4HVP in the Brookhaven protein databank [23]). I used a recent publication by DeSolms et al. [24] on C-terminal variations of the HIV protease inhibitor L-682,679 (see **1** in Fig. 4) as a starting point for my calculations with LUDI. DeSolms et al. report binding data for 12 substituents at the P_2' position and for 18 substituents at the P_3' position. The 3D structure of the L-682,679–HIV protease is not available. For the calculations, I assumed that the Merck compound L-682,679 [24] binds to the HIV protease in the same manner as the compound MVT-101 that was used in the X-ray diffraction experiment by Wlodawer and coworkers. The validity of this assumption is supported by the further structural analysis of a HIV protease–inhibitor complex by Erickson et al [25], showing a binding mode very similar to that of MVT-101 [22]. The geometry of L-682,679 in the complex with the protease was generated as follows. First, the positions of the backbone atoms of the inhibitor were taken directly from the X-ray structure whenever possible and the side chains were added in a reasonable geometry. Hydrogen atoms were added using standard geometries with the molecular graphics program INSIGHT [26]. This structure was then optimized, including a critical buried water



1 2

Fig. 4. Chemical structure of the HIV-protease inhibitor L-682,679 [24] **1** and the reference compound **2** used in the present calculation. LUDI was used to search for suitable substituents R_1 and R_2.

molecule in the active site of the HIV protease, using the force-field CVFF [27]. The protein was kept fixed during the energy minimization. The amino acids Asp, Glu, Lys and Arg of the protease were assumed to be charged. A hundred steps of conjugate gradients energy minimization were carried out to remove unfavorable steric contacts between protein and ligand. The energy minimization caused a shift of the C-terminal nitrogen in compound **2** by 0.23 Å. The corresponding movement of the C_α atom at position P_2' was 0.43 Å. Therefore, with respect to the present calculation, the model structure of compound **2** with the protease is very close to the structure of the MVT-101 compound.

The purpose of the present calculations was to assess the ability of LUDI to design automatically analogs of L-682,679 with a modified C-terminus by comparing the results from LUDI with the data of DeSolms et al. [24]. Structure **2** (see Fig. 4) was used as a lead and calculations were performed with LUDI in the link mode. In this mode, LUDI attempts to append fragments to the already positioned inhibitor **2**. The results are summarized and compared with the data of DeSolms et al. [24] in Table 2.

In a first calculation, LUDI was used to search for substituents R_1 at the P_2' position. LUDI predicts two substituents: $CH_2CH(CH_3)_2$ and $CH(CH_3)_2$. Both were also synthesized by DeSolms and indeed show improved binding by factors of 55 and 500, respectively. However, LUDI failed to retrieve the phenyl group [which shows the best binding of the compounds described by DeSolms (improved binding by a factor of 600)] as the substituent at R_1. The phenyl group was rejected by LUDI due to overlap with the protein structure. This calculation took only 45 s on a Silicon Graphics 4D35 workstation.

LUDI was then used to design ligands for R_2. The calculation took 105 s and yielded 10 possible substituents. Binding data from the paper of DeSolms et al. [24] are available for 3 of them: CH_2CH_2OH (Fig. 6), CH_2-2-pyridyl and CH_2-3-pyridyl (Fig. 5). In all cases, slightly improved binding, by factors of 5, 1.5 and 1.5, respectively, was observed experimentally. It is noteworthty that for several other suggestions of LUDI, experimental data of closely related compounds are given by DeSolms et al [24]. LUDI predicts both CH_2CH_2OH and $CH_2CH_2CH_2OH$ as substituents. Experimentally, the dihydroxy substituent $CH_2CH(OH)CH_2OH$ improved binding by a factor of 18–22. LUDI predicted *p*-hydroxy-phenyl as a substituent. Experimental information is available for *p*-amino-phenyl with a binding improvement of 11–17.

LUDI did not, however, find the methyl-benzimidazole compound, which shows the strongest

TABLE 2
COMPARISON OF THE SUGGESTED SUBSTITUENTS R_1 AND R_2 FOR THE HIV-PROTEASE INHIBITOR **2** WITH THE DATA OF DESOLMS ET AL. [24]

R_1	R_2	Experimentally observed binding improvement
$CH(CH_3)_2$	H	500[a]
$CH_2CH(CH_3)_2$	H	55[a]
$CH(CH_3)_2$	CH_2CH_2OH	5[b]
$CH(CH_3)_2$	CH_2-3-pyridyl	1.5[b]
$CH(CH_3)_2$	CH_2-2-pyridyl	1.5[b]
$CH(CH_3)_2$	$CH_2CH_2CH_2OH$	na
$CH(CH_3)_2$	$CH_2COOC_6H_5$	na
$CH(CH_3)_2$	$CH_2C_6H_4$-4'OH	na
$CH(CH_3)_2$	CH_2-1-imidazolyl	na
$CH(CH_3)_2$	CH_2-2-thiazolyl	na
$CH(CH_3)_2$	CH_2-2-furanyl	na
$CH(CH_3)_2$	CH_2-1-tetrahydroisochinolin	na

[a]As compared to R_1 = H (IC_{50} = 500 nM).
[b]As compared to R_2 = H (IC_{50} = 1 nM).
na = not available.

binding in the study of DeSolms et al. [24] (IC_{50}=0.06 nM), although the appropriate fragment is contained in the fragment library. It is tempting to speculate about the reason for this failure. The benzimidazole moiety can form two hydrogen bonds with the protein. The most likely partners in the protein to form these hydrogen bonds are the side-chain oxygen of Asp B29 and the backbone nitrogen of Gly B48. The distance between these atoms in the crystal structure 4HVP is 9.22 Å. The sum of two hydrogen-bond lengths (2*2.9 Å) and the intramolecular N-N distance (2.4 Å) in the benzimidazole moiety is, however, only 8.2 Å. This is 1 Å shorter than the distance in the X-ray structure. Therefore, it is likely that the conformation of the side chain of Asp B29 will change upon ligand binding to allow for two hydrogen bonds to be formed. In fact, when the side-chain conformation of Asp B29 was changed so that the distance Asp B29 OD–Gly B48 N was reduced to 8.0 Å, methyl-benzimidazol was retrieved by LUDI as a possible substituent at R_2.

3.2. Inhibitors of DHFR

The second example given is the design of new inhibitors of dihydrofolate reductase (DHFR). The 3D structure of DHFR complexed with the anticancer drug methotrexate **3** (MTX) was solved by Bolin et al. [28] (entry 4DFR in the Brookhaven protein databank [23]) (Fig. 7). This structure, without water molecules, was used in the present calculations. The purpose of the calculations was to use LUDI to design new substituents for the 2,4-diamino-pteridine moiety at position 6 on the ring system. Therefore, only the pteridine portion **4** of MTX was used from the X-ray structure and the substituent at position 6 was removed. Again, the hydrogen atoms were added using the program INSIGHT [26].

The results from LUDI on the design of substituents for 2,4-diamino-pteridine in position 6,

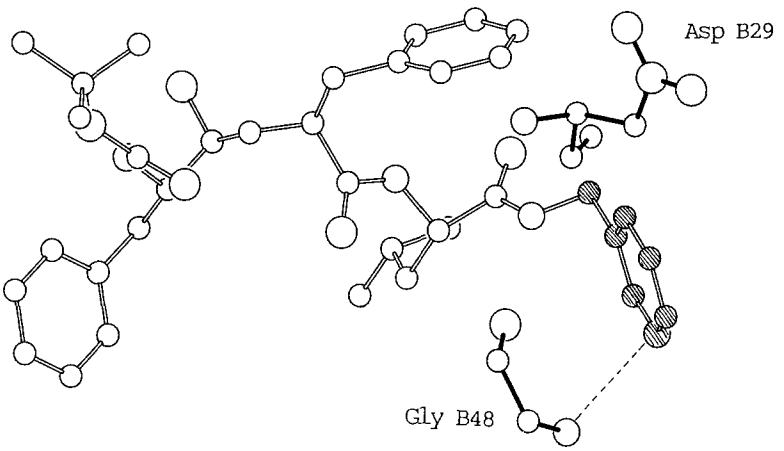


Fig. 5. Conformation of the CH_2-3-pyridyl substituent as R_2 of compound **2**. The substituent is shown with shaded atoms. LUDI suggests that the pyridine nitrogen forms a hydrogen bond with the backbone nitrogen of Gly B48 from the protein.

once again run in the link mode, are summarized in Table 3 and are compared with data from a compilation of experimental data prepared by Blaney et al. [29]. LUDI retrieved seven structures as possible substituents of R. Experimental data are available for two of them. The isobutyl substituent leads to a strong improvement in binding. The phenylethyl group leaves the binding unchanged. LUDI does not retrieve the substituent $CH_2N(CH_3)C_6H_4$-4′-CO-Glu (yielding MTX) because the link library does not contain such complex moieties.

4. DISCUSSION AND CONCLUSIONS

This paper describes recent advances in a new approach to the de novo design of protein ligands as implemented in the computer program LUDI [9]. A new set of rules to generate the interaction

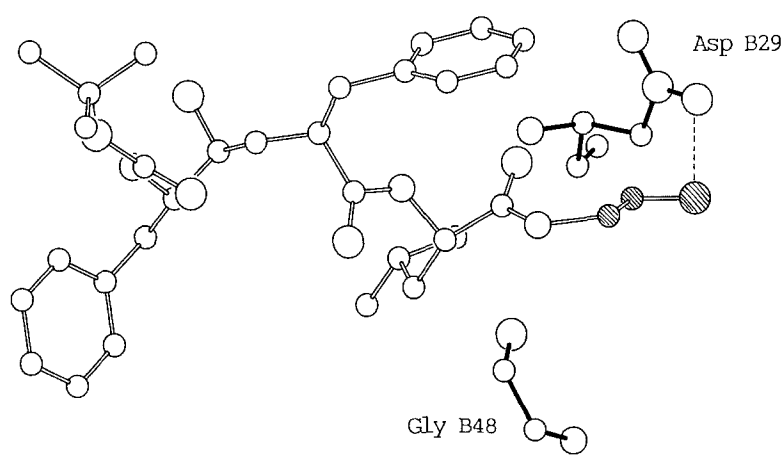
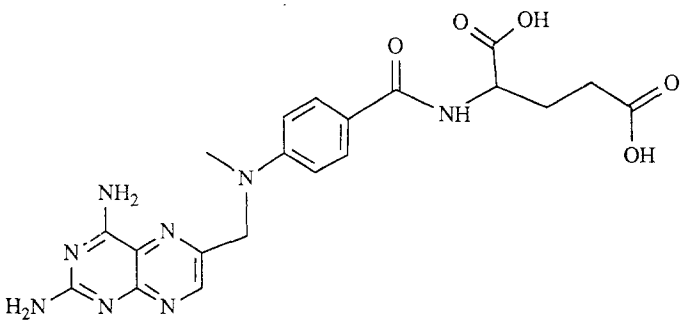
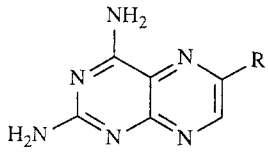


Fig. 6. Conformation of the CH_2CH_2OH substituent at R_2 of compound **2**. The substituent is shown with shaded atoms. LUDI suggests that the hydroxyl group forms a hydrogen bond with the side chain of Asp B29 from the protein.



3



4

Fig. 7. Chemical structure of methotrexate **3** and the reference compound **4** used in the present calculation. LUDI was used to search for suitable substituents R at position 6 of the pteridine ring.

sites is described. LUDI is now capable of designing new substituents for a given enzyme inhibitor lead. A scoring function for the fitted fragments was implemented that is based on the number and quality of the hydrogen bonds and the hydrophobic contact surface.

LUDI was successfully applied to the design of inhibitors for the enzymes HIV protease and dihydrofolate reductase. The first application of LUDI given in the present paper is the design of a new C-terminal substituent for an inhibitor of HIV protease. In this case, LUDI predicted two fragments as substituents for the P_2' site; both were found experimentally to yield substantially improved binding [24]. For the P_3' site, LUDI retrieved ten candidate structures. The available experimental data show improved binding for three of them. For DHFR, LUDI predicted seven fragments as possible substituents for 2,4-diamino-pteridine moiety at position 6. For one of them, the available experimental data indeed showed improved binding as compared to the unsubstituted lead compound. These results demonstrate that LUDI is indeed able to suggest active compounds.

The positioning of fragments by LUDI is done by a fit onto the interaction sites. This approach offers the advantage that only purely geometrical calculations are required, thereby avoiding the very CPU-intensive evaluation of energy functions and their derivatives. In comparing the present

TABLE 3
COMPARISON OF SUGGESTED SUBSTITUENTS R AT POSITION 6 OF 2,4-DIAMINO-PTERIDINE WITH EXPERIMENTAL DATA FROM THE SURVEY OF BLANEY ET AL. [29]

R	Experimentally observed binding improvement
$CH_2CH(CH_3)_2$	> 100
CH_2-1-naphthyl	na
$CH_2CH_2C_6H_5$	1
$CH(CH_3)_2$	na
$CH_2C_6H_3$-3′,5′$(CH_3)_2$	na
$CH_2C_6H_4$-4′CN	na
$CH_2C_6H_4$-4′OH	na

na = not available.

approach with the well-established method of positioning a putative ligand by force-field calculations, one should bear in mind that the traditional force-field approach will also encounter the multiple minima problem. Therefore, a considerable number of force-field calculations are required before the optimal position of the ligand can be specified unambiguously. Methods based on force-field calculations will therefore be much slower than LUDI.

When comparing the accuracy of LUDI to the traditional force-field calculations, one must consider that the error introduced by using discrete positions or vectors is roughly of the order of the distance between the interaction sites. In the examples described in the present paper, the point density corresponds to distances between neighboring interaction sites of about 0.3 Å. This is roughly equal to the accuracy of the atomic positions in a high-resolution protein structure and well within the tolerance of most of the results of today's best force-field calculations. I therefore conclude that the use of discrete points does not introduce significant errors into the calculation.

LUDI does not distinguish between interaction sites at the optimal positions, e.g., for carboxylic groups those with $<_{O\text{-}C\text{-}O..H} = 0°$ and those slightly shifted off these positions. However, the position of a protein ligand is usually determined by several interactions that all occur simultaneously. This means that in most cases the hydrogen-bond geometries will not adopt their optimal values. The geometrical constraints involved in maximizing the number of hydrogen bonds will be more important than the electrostatic effects in determining the hydrogen-bond geometry [30]. Therefore, the present approach appears to be justified, as only geometries are considered. A detailed evaluation of a protein–ligand complex generated by LUDI can be made afterwards using a force-field calculation.

A very important advantage of the geometry-based approach adopted by LUDI is the possibility to combine the search for favorable nonbonded interactions with the search for a suitable bond for the fragment with an already existing ligand. This offers the possibility to design new protein ligands in a stepwise manner.

In conclusion, I have further developed a new algorithm for the de novo design of protein ligands. The method has been applied successfully to predict improved inhibitors for two enzymes (DHFR and HIV protease). The present results indicate that the approach may be useful for the rational design of drugs when the 3D structure of the target protein is known.

ACKNOWLEDGEMENT

I would like to thank my colleague Gerhard Klebe for many helpful discussions.

REFERENCES

1 Goodford, P.J., J. Med. Chem., 28 (1985) 849.
2 Boobbyer, D.N.A., Goodford, P.J., MCWhinnie, P.M. and Wade, R.C., J. Med. Chem., 32 (1989) 1083.
3 Miranker, A. and Karplus, M., Proteins 11 (1991) 29.
4 DesJarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D. and Venkataraghavan, R., J. Med. Chem. 31 (1988) 722.
5 Lawrence, M.C. and Davis, P.C., Proteins 12 (1992) 31.
6 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, London, 1989, pp. 182–196.
7 Moon, J.B. and Howe, W.J., Proteins 11 (1991) 314.

8 Lewis, R.A., J. Comput. Aided Mol. Design, 4 (1990) 205.
9 Böhm, H.-J., J. Comput. Aided Mol. Design, 6 (1992) 61.
10 a. Allen, F.H., Kennard, O. and Taylor, R., Acc. Chem. Res. 16 (1983) 146.
10 b. Allen, F.H., Bellard, S., Brice M.D., Cartwright, B.A., Doubleday, A., Higgs, H., Hummeling-Peters, T., Kennard, O., Motherwell, W.D.S., Rodgers, J.R. and Watson, D.G., Acta Crystallogr. B35 (1979) 2331.
11 Murray-Rust, P. and Glusker, J.P., J. Am. Chem. Soc. 106 (1984) 1018.
12 Taylor, R. and Kennard, O., Acc. Chem. Res., 17 (1984) 320.
13 Vedani, A. and Dunitz, J.D., J. Am. Chem. Soc. 107 (1985) 7653.
14 Görbitz, C.H., Acta Crystallorgr., B45 (1989) 390.
15 Klebe, G., Habilitationsschrift, University of Heidelberg, 1991.
16 Görbitz, C.H. and Etter, M.C., J. Am. Chem. Soc. 114 (1992) 627.
17 Baker, E.N. and Hubbard, R.E., Progr. Biophys. Mol. Biol. 44 (1984) 97.
18 Kroon, J., Kanters, J.A., Van Duijneveldt-Van de Rijdt, J.G.C.M., Van Duijneveldt, F.B. and Vliegenhardt, J.A., J. Mol. Struct., 24 (1975) 109.
19 Fersht, A.R., Shi, J.P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M.Y. and Winter, G., Nature 314 (1985) 235.
20 Richards, F.M., Ann. Rev. Biophys. Bioeng. 6 (1977) 151.
21 Huff, J.R., J. Med. Chem. 34 (1991) 1834.
22 Miller, M., Schneider, J., Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.H. and Wlodawer, S., Science 246 (1989) 1149.
23 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.
24 DeSolms, S.J., Giuliani, E.A., Guare, J.P., Vacca, J.P., Sanders, W.M., Graham, S.L., Wiggins, J.M., Darke, P.L., Sigal, I.S., Zugay, J.A., Emini, E.A., Schleif, W.A., Quintero, J.C., Anderson, P.S. and Huff, J.R., J. Med. Chem. 34 (1991) 2852.
25 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, N., Kohlbrenner, W.E., Simmer, R., Helfrich, R., Paul, D.A. and Knigge, M., Science 249 (1990) 527.
26 Program INSIGHT, distributed by BIOSYM Technologies, San Diego, CA 92121, U.S.A.
27 Dauber-Osguthorpe, P., Roberts, V.A., Osguthorpe, D.J., Wolff, J. Genest, M. and Hagler, A.T., Proteins, 4 (1988) 31.
28 Bolin, J.T., Filman, D.A., Matthews, D.A., Hamlin, R.C. and Kraut, J., J. Biol. Chem., 257 (1982) 13650.
29 Blaney, J.M., Hansch, C., Silipo, C. and Villon, A., Chem. Rev. 84 (1984) 333.
30 Mitchell, J.B.O. and Price, S.L., Chem. Phys. Lett., 154 (1989) 267.