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

Hans-Joachim Böhm

BASF AG, Central Research, 6700 Ludwigshafen, Germany

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

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.
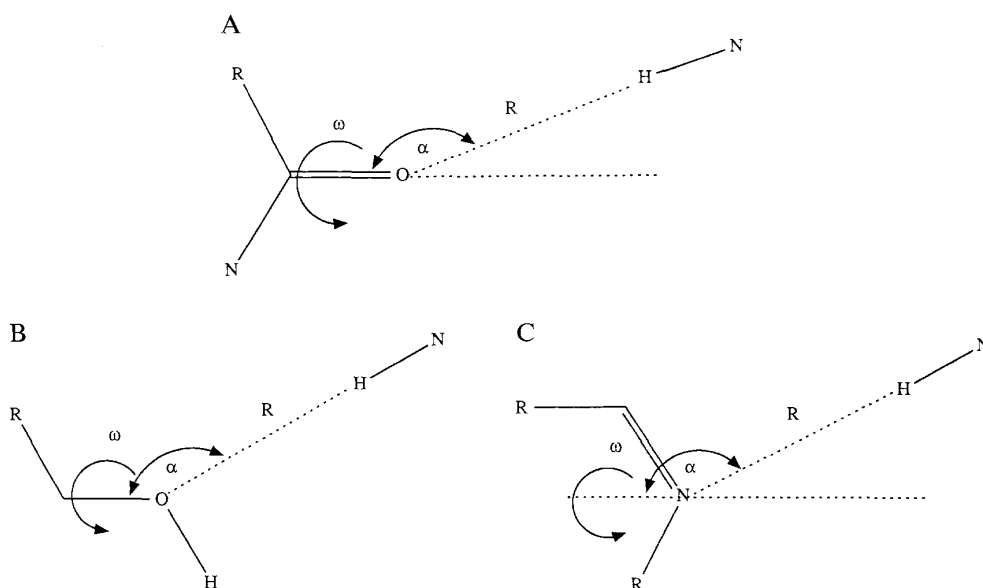

Fig. 1. Definition of the geometric parameters R, α and ω used in the rules for the allowed nonbonded contacts. A: definition for terminal groups; B: definition for -O-; C: definition for -N=. For -N= groups; α denotes the angle between the bisector of the angle C=N-R and the vector N..H.

Finally, LUDI was used to design new inhibitors of the aspartic protease of the human immu-
nodeficiency 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 pat-
terns. 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 distri-
bution 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 ($R_{D-X} = 1$ Å) and hydrogen-acceptor sites are represented by A-Y vec-
tors ($R_{A-Y} = 1.23$ 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 crys-
tal 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 exten-
sively [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 $<_{O-C-O..H}=0,180°$. 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 ($<_{C-O..H}=109\pm20°$). 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-

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.

fastcase
Smarter legal research.