

STRUCTURE-BASED DRUG DESIGN

edited by

PANDI

VEERAPANDIAN

Axiom Biotechnologies

San Diego, California

and

La Jolla Institute of Experimental Medicine

La Jolla, California



MARCEL DEKKER, INC.

NEW YORK • BASEL • HONG KONG

ISBN: 0-8247-9869-4

The publisher offers discounts on this book when ordered in bulk quantities. For more information, write to Special Sales/Professional Marketing at the address below.

This book is printed on acid-free paper.

Copyright © 1997 by MARCEL DEKKER, INC. All Rights Reserved.

Neither this book nor any part may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming, and recording, or by any information storage and retrieval system, without permission in writing from the publisher.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016
<http://www.dekker.com>

Current printing (last digit):
10 9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

21

Structure-Based Combinatorial Ligand Design

Amedeo Caflisch

University of Zürich, Zürich, Switzerland

Claus Ehrhardt

Novartis Pharma Inc., Basel, Switzerland*

I. INTRODUCTION

Structure-based ligand design is fascinating and challenging. Whenever it is possible to determine the three-dimensional structure of a pharmacologically relevant enzyme or receptor, computational approaches can be used to design novel high-affinity ligands. These methods can complement the broad screening efforts, which represent traditional lead discovery.

In this chapter we focus on our approach to computer-aided ligand design. It is based on the docking of a diverse set of molecular fragments into the active site of a macromolecular target and on the use of a combinatorial strategy to connect them to form candidate ligands. The methodology is illustrated by an application to human thrombin, a trypsin-like serine protease fulfilling a central role in both hemostasis and thrombosis. The selective inhibition of thrombin is expected to prevent thrombotic diseases.

Ligand-design programs are being developed at an ever increasing rate and some are related to various aspects of our ligand design approach. The LEGO software tool is based on the combination of multiple fragment docking, automatic connection by small linker units (one to four atom chains), and searching of three-dimensional databases for complementary molecules [1,2]. It has been implemented within the MOLOC molecular modeling system [3], which allows the visualization of the functionality maps and interactive model building of the growing ligands. Another related approach is that embodied in the program LUDI [4–8]. It makes use of statistical data from small-molecule

* Formerly Sandoz Pharma Ltd.

crystal structures to determine binding sites of molecular fragments, i.e., discrete positions on the binding site surface suitable to form hydrogen bonds and/or to fill hydrophobic sites of the receptor. Alternatively, it uses simple rules or the output of the program GRID [9–12] to generate the interaction sites. Finally, the fragments fitted in the interaction sites are connected by linker groups. Other fragment-based programs are GROUPBUILD [13]; GROW [14], HOOK [15], NEWLEAD [16], SPROUT [17], and TORSION [18]. These and other strategies for computer-aided structure-based ligand design have been reviewed by several contributors [13,19,20].

II. DOCKING MOLECULAR FRAGMENTS

A. Multiple Copy Simultaneous Search

The present approach for ligand design is based on the combinatorial selection of molecular fragments optimally docked on the protein binding site to form a population of diverse candidate ligands. The multiple copy simultaneous search (MCSS) procedure combines the advantages of random distribution and simultaneous minimization of a set of replicas of a chemical fragment to obtain maps of energetically favorable positions and orientations (local energy minima) [21,22]. These maps, which contain all possible low-energy minima of a fragment-protein complex, are called functionality maps. A plethora of structural and thermodynamic data on inhibitor-enzyme complexes [23–26] suggest that the burial of nonpolar groups of the ligand in hydrophobic pockets of the protein is important for binding affinity and that intermolecular electrostatic interactions determine selectivity. For this reason, and because most of the known enzymes' binding sites have both hydrophilic and hydrophobic character, very diverse functional groups are used in MCSS. Representative examples include charged (e.g., acetate, benzamidine, methylammonium, methylguanidinium, pyrrolidine); polar (e.g., methanol, 2-propanone, *N*-methylacetamide); aromatic (e.g., benzene, pyrrole, imidazole, phenol); and aliphatic (e.g., propane, isobutane, cyclopentane, cyclohexane) groups. Although most of these fragments are rigid, MCSS can also generate the functionality maps of flexible medium-size fragments, e.g., the amino acid side chains. Additional functional groups and more complex heterocyclic systems are currently being introduced to increase the diversity of the resulting ligands and to better characterize the specificity of the binding pockets (A. Cafilisch and C. Ehrhardt, unpublished results).

As shown by a flowchart in Figure 1, the method is fully automated, although certain critical parameters (e.g., number of replicas, radius of the sphere for random distribution, CHARMM parameters for the minimization) can be adjusted by the user to optimize it for specific applications. Several thousand replicas of a given group are randomly distributed inside a sphere

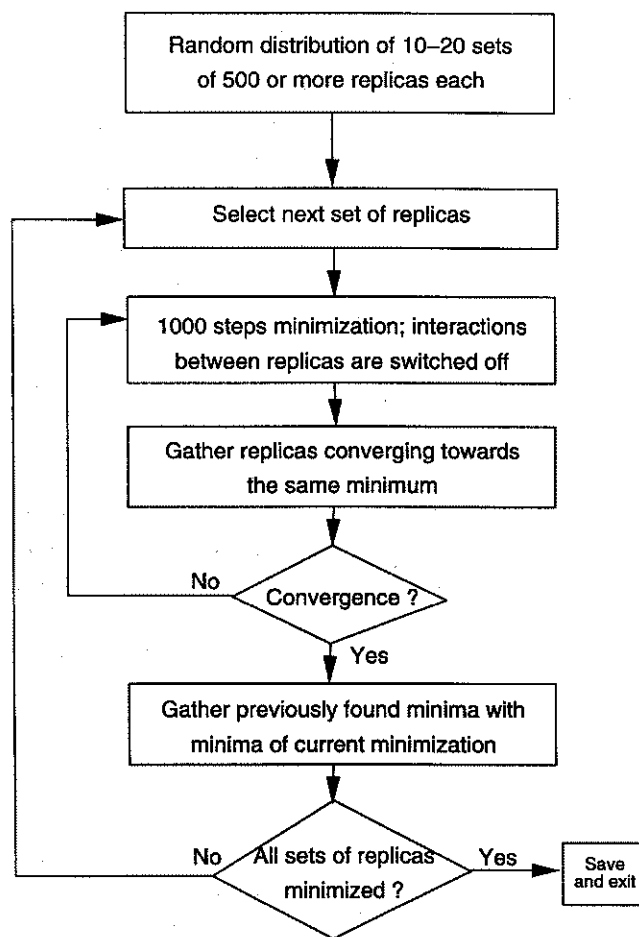


Figure 1 Schematic representation of the MCSS procedure. Conditional statements are enclosed by diamonds.

whose radius is chosen large enough to cover the entire region of interest. This can be a known binding site or the entire protein, if one wants to explore alternative binding pockets. The initial random distribution also can be performed inside a parallelepiped if the region of interest is elongated in one or two directions. A minimal distance can be given as input to avoid bad contacts between functional group atoms and protein atoms for the initial distribution.

Subsets of between 500 and 3000 randomly distributed replicas of the same group are simultaneously minimized in the force field of the protein. The

interactions between the group replicas are omitted. The polar-hydrogen approximation (PARAM19) of the CHARMM force field is used [27]. In the application of the MCSS method to the sialic acid binding site of the influenza coat protein hemagglutinin [21], HIV-1 aspartic proteinase [22], and thrombin [20,28] the protein was kept fixed; hence, the forces on each replica consist of its internal forces and those due to the protein, which has unique conformation and, therefore, generates a unique field. The positions are compared every 1000 steps to eliminate replicas converging toward a common minimum. Further details concerning the methodology are given in References 21 and 22, while a critical assessment has been presented in Reference 20.

B. Simple Approximations of Solvation Effects

In previous applications of MCSS [21,22] the effects of the solvent were neglected, i.e., all protein-fragment interactions were calculated with a vacuum potential [27]. This choice was based on the principle that fast methods are necessary to perform effective searches of the binding site and that good candidate ligands subsequently can be ranked in terms of their binding free energy [20,28]. A possible difficulty with this approach is that minimized positions may be missed or misplaced due to the lack of a solvation correction during the MCSS minimization.

Electrostatic Shielding

In MCSS studies of thrombin [20,28], it was observed that minima of charged groups tend to cluster in the vicinity of charged side chains on the thrombin surface and in the S1 (basic groups) or S1' pocket (acidic groups). It is then necessary to estimate the electrostatic desolvation of both protein and fragment to obtain a realistic ranking of the minima [28]. As a simple test of the importance of electrostatic shielding, a distance-dependent dielectric function [29] was introduced instead of the unit dielectric constant in the vacuum potential. The overall shape of the acetate map did not change, but three more minima were found close to the Lys60F side chain in S1' [20]. It is difficult to find a physical meaning in favor of the distance-dependent dielectric function. Nevertheless, it is a simple and useful approximation, since it yields a smoother and more realistic potential surface than the vacuum Coulombic interaction.

Hydrophobic Effect

Aliphatic and aromatic fragments do not bear any partial charges in the polar-hydrogen approximation. Hence, MCSS determines their optimal position in the

protein binding site exclusively by van der Waals interactions. Minima of nonpolar fragments may be found in hydrophilic pockets because of the lack of an energy penalty for protein desolvation. A representative example of a cluster of propane minima in a mainly hydrophilic region of the HIV-1 aspartic proteinase binding site is shown in Reference 20. In a simple attempt to approximate desolvation of polar regions of the protein, the attractive contribution of the van der Waals interaction energy was switched off between atoms of nonpolar MCSS fragments and all protein polar hydrogen, nitrogen, oxygen, and sp^2 carbon atoms. In addition, the van der Waals radius of nitrogen and oxygen atoms was increased from the PARAM19 default value of 1.6 Å to 2.2 Å and the van der Waals radius of the aliphatic carbons was reduced by 0.1 Å to avoid the too large van der Waals distance between carbons often produced by PARAM19. The modified force field yields thrombin functionality maps of propane, cyclopentane, cyclohexane, and benzene in agreement with structural data of known inhibitors (see Section II. C). In addition, these nonpolar groups are prevented from occupying hydrophilic pockets.

As a further test of the modified force field, the MCSS procedure was used to generate the propane functionality map on the surface of the A peptide chain of the leucine zipper, i.e., residues 249–281 from the yeast transcriptional activator protein GCN4. Leucine zippers are composed of amphipathic α helices containing heptad repeats (*abcdefg*) in which hydrophobic residues are frequent at *a* and *d*. As the x-ray structure indicates [30,31], the two amphipathic α helices are held together by hydrophobic interactions between residues in the *a* and *d* positions (Figure 2). The B peptide chain was removed and MCSS was run separately with the original and the modified force field starting from the same random distribution of 5000 propane replicas around helix A. The sixty most favorable minima obtained with the modified force field are distributed in twelve clusters, seven of which match the Val and Leu side chains of the B helix involved in the interhelical interactions (Figure 2). Of the remaining five clusters, labeled A to E in Figure 2, B has minima in contact with Val and Ala side chains, D with Leu and Tyr side chains, while A, C, and E with both a Val or Leu side chain and the alkyl part of a Lys side chain. The sixty most favorable minima obtained with the original force field form sixteen clusters (not shown). Only four of these match Val and Leu side chains of the B helix, while eight clusters desolvate one (or more) polar group on the hydrophilic (exposed) surface of the A helix.

The functionality maps of polar groups generated with a distance-dependent dielectric and those of nonpolar groups obtained with the modified force field are closer to those obtained by using a continuum dielectric model [32–35] to postprocess the MCSS minima and estimate solvation effects [28]. Hence, simple modifications of the force field may result in more accurate projections of the binding free energy. Since these modifications are used during the

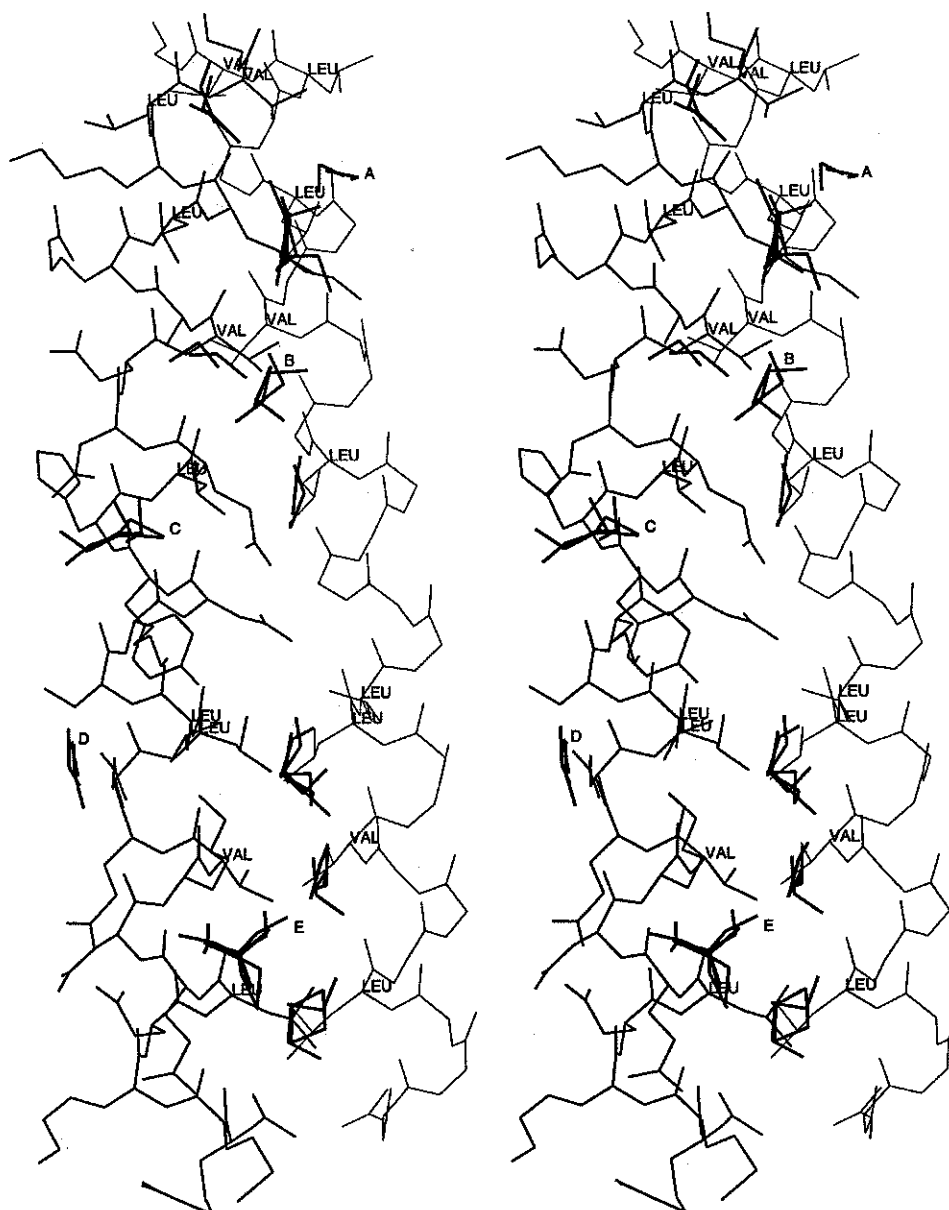


Figure 2 Stereo view of the sixty propane minima (thick lines) obtained with the modified force field (see text) on the surface of the A peptide chain (medium lines) of the GCN4 leucine zipper (PDB code 2ZTA). Although the B peptide chain was removed during the MCSS procedure, its backbone and hydrophobic side chains are also drawn (thin lines) to show how the propane minima match the aliphatic groups of chain B.

minimization phase, a more realistic distribution of functional group minima is generated.

C. Thrombin Functionality Maps

Human thrombin is one of the best characterized enzymes from a structural point of view (Figure 3). It binds a series of diverse inhibitors without major rearrangements in its conformation, as shown by a number of x-ray crystallography studies [26,36–39]. Its S3 and S2 precleavage subpockets have hydrophobic character, whereas at the bottom of the S1 or recognition pocket the carboxy group of Asp189 is a salt bridge partner for basic side chains (Figure 3). D-phenylalanyl-L-prolyl-L-arginine chloromethane, PPACK, (Figure 4a), and N α -((2-naphthylsulfonyl)glycyl)-DL-*p*-amidinophenylalanyl piperidine, NAPAP, (Figure 4b) are the archetypal active-site inhibitors of thrombin. The crystal structure of the thrombin–NAPAP complex is shown in Figure 3. The PPACK and NAPAP inhibitors bind to the thrombin active site by occupying the S3 and S2 pockets with their hydrophobic moieties and by positioning their basic group (guanidinium of PPACK, benzamidine of NAPAP) into S1 to form a salt bridge with Asp189.

In continuation of a project aiming at the structure-based design of low molecular weight, active-site directed inhibitors of human thrombin [40], MCSS was used to generate a series of functionality maps of the thrombin S3 to S2' pockets [20,28]. A detailed description of the thrombin functionality maps and the continuum approximation used to postprocess the MCSS minima is given in Reference [28]. From the analysis of the results for the nonpolar groups it is evident that hydrophobic moieties prefer to bind to the S3 and S2 pockets (Figure 5). The solvent exposed face of the Trp60D indole is another favorable site, though the intermolecular van der Waals interactions are much smaller. Binding to the S2' region is favored by interactions with the Leu40 side chain but implies a desolvation penalty because of the burial of part of the Arg73 guanidinium and/or the Gln151 side chain. The latter might be an artifact of the rigid protein structure used in the minimization, since the side chains of Arg73 and Gln151 are flexible enough to displace their polar groups towards a more exposed region. Binding to the neighboring Leu41 side chain in S1' is highly unfavorable because of the concomitant desolvation of Lys60F.

For polar groups with zero net charge there are several hydrophilic groups on the thrombin main chain that might be involved in strong hydrogen

Hydrophobic residues are labeled at their C α atom. Five clusters of propane minima that do not match the hydrophobic side chain of the B helix involved in the interhelical interactions are labeled from A (top right) to E (bottom center) and discussed in the text.

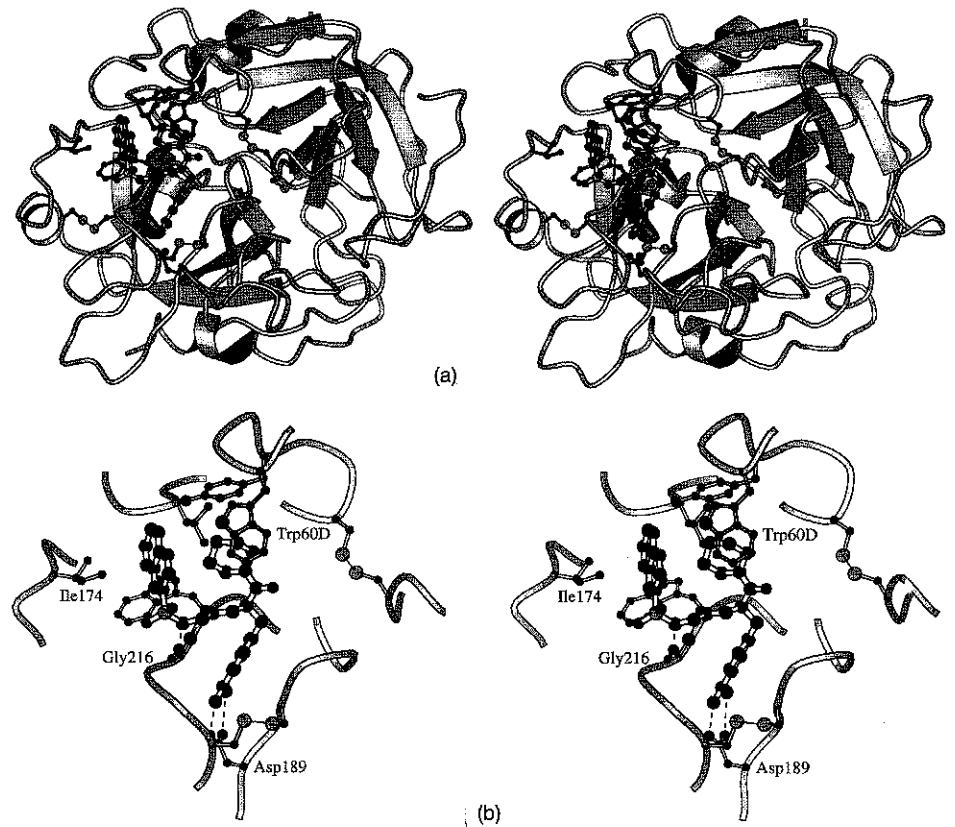


Figure 3 (a) Stereo view of the thrombin molecule in its complex with NAPAP. The side chains of thrombin involved in the binding of NAPAP and the disulfide bridges are shown in stick-and-ball representation with black sticks. NAPAP is shown in stick-and-ball representation with white sticks. (b) Zoom image of the active site region. Figures made with the program MOLSCRIPT [42].

bonds. These are 214CO, 216NH, 216CO, and 219NH in S1; 193NH and 195NH in the oxyanion hole; 41CO in S1'; 40CO in S2'; and 147NH and 148NH on the autolysis loop, whose exposure is dependent on crystallization conditions and inhibitor type.

Two main conclusions can be drawn from the analysis of the minimized positions of the charged functional groups. First, the minima with the lowest binding free energy have optimal hydrogen bonds with the Asp189 side chain in the S1 pocket. Representative examples are the lowest energy minimum of benzamidine (Figure 5) and the lowest energy minima of methylguanidinium and methylammonium (Figure 4 of Reference 28). Since the Asp189 side chain

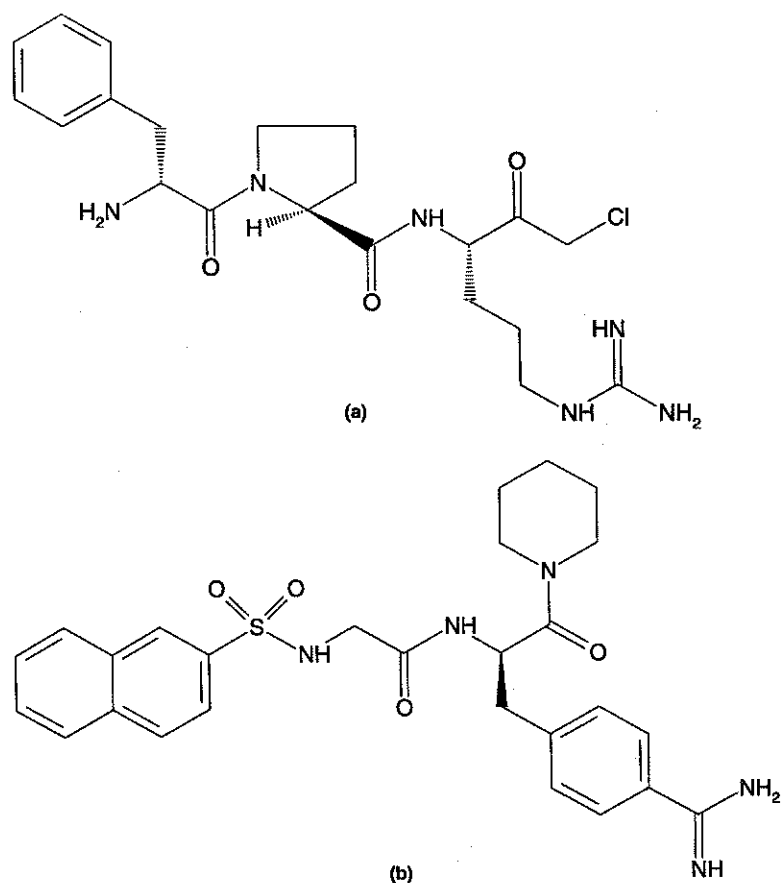


Figure 4 Chemical structure of PPACK (a) and NAPAP (b).

is more buried than the side chain of Lys60F, the minima of positively charged groups interacting with the former have a more favorable binding free energy than those of the negatively charged groups close to the latter. This is due to reduced shielding of the charge-charge interaction and the smaller desolvation of the carboxylate oxygens of Asp189 compared to the amino group in Lys60F [28]. Second, polar groups on the protein surface are not ideal partners for a charged functional group because the high desolvation penalty for these groups might not be completely compensated for by the favorable electrostatic interaction energy. This finding is analogous to the results for the polar functional groups, i.e., their binding to a partially exposed charged side chain of the protein may result in an unfavorable total binding free energy.

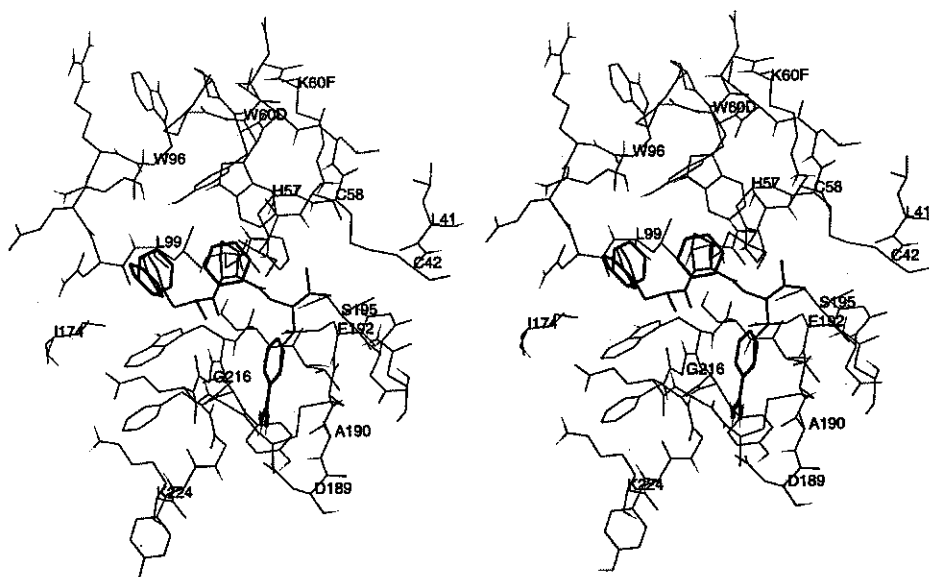


Figure 5 Stereo view of the three lowest energy minima of benzene obtained with the modified force field and the lowest energy minimum of benzamidine (thick lines for heavy atoms and thin lines for polar hydrogens) in the thrombin active site (thin lines). The inhibitor PPACK is also shown (medium lines), though it was removed during the MCSS procedure. Some C_{α} atoms of thrombin are labeled.

III. CONNECTING MOLECULAR FRAGMENTS

A. Computational Combinatorial Ligand Design

Overview

The recently developed program for computational combinatorial ligand design (CCLD) requires as input atomic coordinates and partial charges of the protein atoms, as well as the coordinates of the MCSS minima and the individual contributions to the free energy of binding [28]. An additional file contains a number of control parameters and, for each functional group used for MCSS, a list of atoms which can be used for connection (linkage atoms). The following procedures are performed during a regular execution of CCLD (Figure 6): The MCSS minima are first sorted according to their approximated binding free energies [28]; then a list of bonding fragment pairs and a list of overlapping fragment pairs are generated (see below). This is followed by the combinatorial generation of putative ligands (see below).

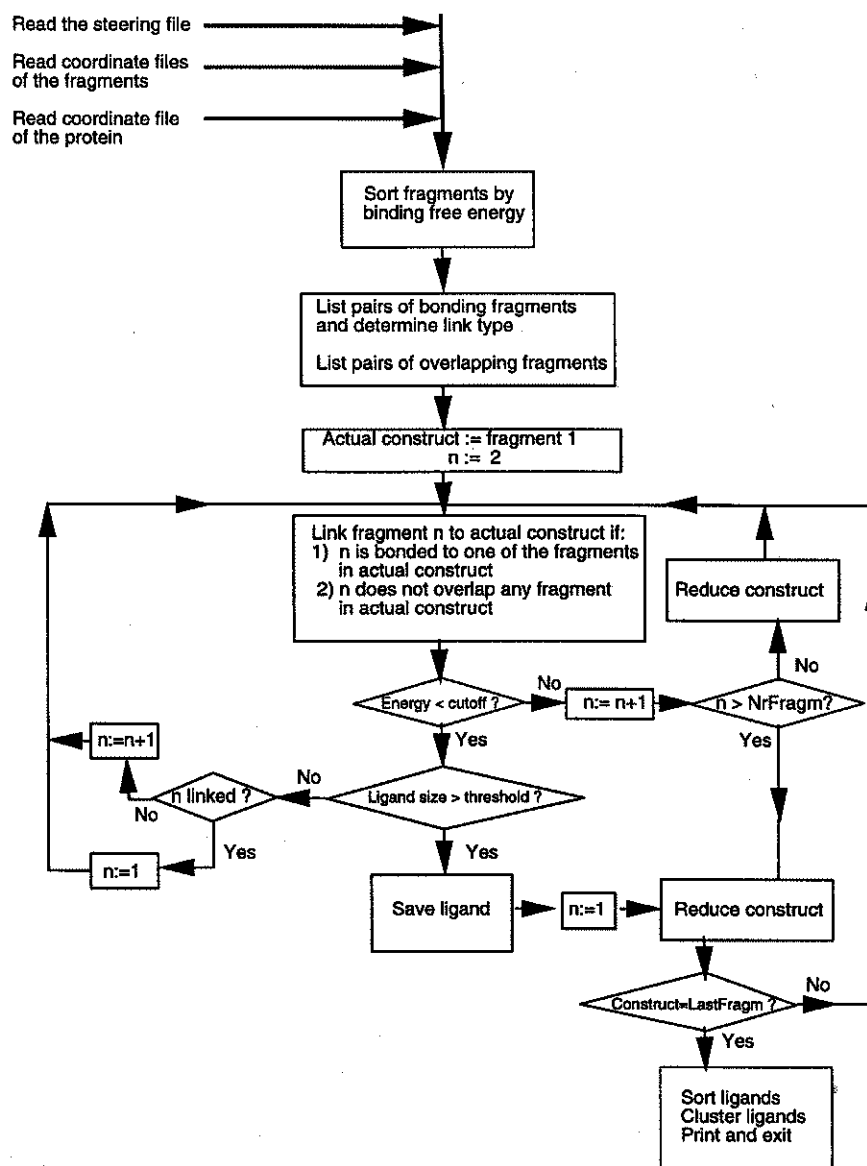


Figure 6 Schematic representation of the CCLD program. Variable assignments are symbolized by :=. Conditional statements are enclosed by diamonds. [From Reference 28.]

Lists of Bonding Fragment Pairs and Overlapping Fragment Pairs

The user has to specify for each functional group type which atoms are to be used for connection to other fragments. For each linkage atom CCLD generates a set of possible linkage points, i.e., points that will be used to determine the position and orientation of the link. All possible pairs of minimized positions are then analyzed and added to the list of bonding fragment pairs if they can be linked; otherwise, if two fragments have bad contacts they are added to the list of overlapping fragment pairs. A pair of bonding fragments may be connected by a linker unit, by a single covalent bond (1-bond), or by fusing two overlapping atoms belonging to different fragments (0-bond). The linker units are small since their function is to optimally connect two fragments without adding considerably to the molecular weight. The following linker elements have been implemented so far: Keto and methylene (2-bond), amide and ethylene (3-bond). The user is free to choose minimal and maximal values for the distance (d) between linkage atoms for each connection type. In the application to thrombin the following values in Ångstroms were used: $d < 0.43$, 0-bond; $1.2 < d < 1.8$, 1-bond; $2.2 < d < 2.7$, 2-bond; $3.6 < d < 4.0$, 3-bond. The list of bonding fragment pairs and the list of overlapping fragment pairs are created only once before entering the combinatorial search (Figure 6). The use of these lists results in a significant increase in the speed with which ligands are generated.

Combinatorial Ligand Generation

Starting from the MCSS minimum with the most favorable binding free energy the ligand generation algorithm proceeds in an iterative and exhaustive way by linking an additional fragment to the actual construct. Such an "elongation" step is very fast since it is sufficient to check that the new fragment may be connected to one of the fragments in the actual construct (by looking in the list of bonding fragment pairs) and that the new fragment does not overlap with any of the fragments in the actual construct (Figure 6). The combinatorial explosion problem is kept under control by pruning, which is performed according to the average value of the free energy of binding of the fragments. Whenever the addition of a fragment to the growing ligand results in an average value of the binding free energy higher than a user specified threshold, the construct is reduced by deletion of the latest added fragment (Figure 6). A ligand with energy below the threshold is saved if it is larger than a user specified minimal size and if it is not a substructure of a ligand found previously.

A CCLD run requires from 2–3 minutes (for 200 to 300 fragments) to less than one hour CPU time (for about 1000 fragments) on an SGI R4400 processor.

B. Candidate Ligands of Thrombin

To test if CCLD is able to reproduce the known thrombin inhibitors, the functional groups of PPACK and NAPAP were given as input molecular fragments in the orientation derived from the crystal structures of the complexes. The CCLD program generated a set of candidate ligands that not only contained the PPACK and NAPAP structures but also a number of interesting hybrid molecules consisting of fragments from both inhibitors. A representative example is shown in Figure 7. This putative ligand consists of the C-terminal part of NAPAP, whose piperidine ring is connected at the 3-position to the PPACK D-Phe by an amide linker. The latter has its carbonyl oxygen involved in a hydrogen bond with the Gly216 NH.

In another run, the MCSS minima of benzamidine (Figure 5), benzene (Figure 5), cyclopentane, and cyclohexane [28] were used as starting molecular fragments. In a few seconds of CPU time of an SGI Indigo2 (R4400 processor), CCLD generated a series of molecules showing the same interaction patterns as those of known thrombin inhibitors, i.e., hydrophobic moieties in S3 and S2, hydrogen bonds with the polar groups of Gly216, and benzamidine in S1. One of these putative ligands is shown in Figures 8 and 9. It is involved in the same interactions as in the NAPAP-thrombin complex except for the hydrogen bond with the CO of Gly216. Its cyclohexane ring in S3 is connected to the

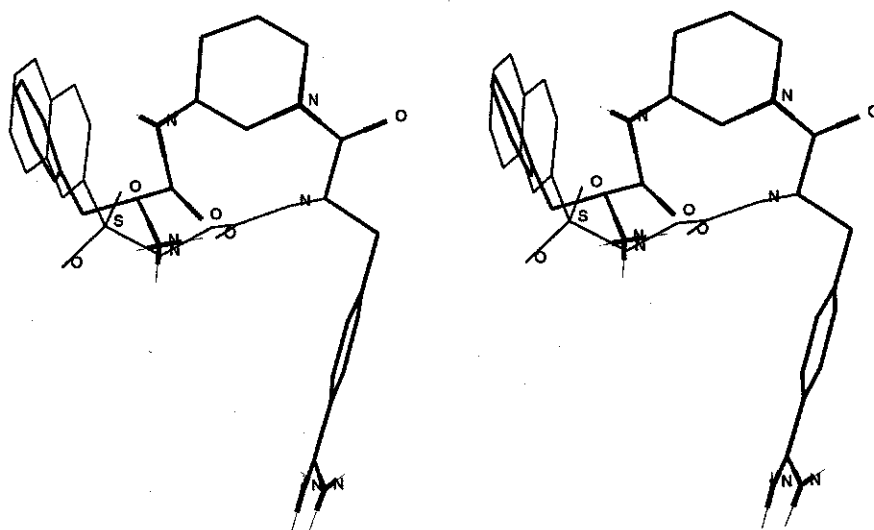


Figure 7 PPACK-NAPAP hybrid ligand (thick lines) generated by CCLD starting from the functional groups of PPACK and NAPAP (thin lines). The amide linker connecting the piperidine in 3-position to the D-Phe was created by CCLD.

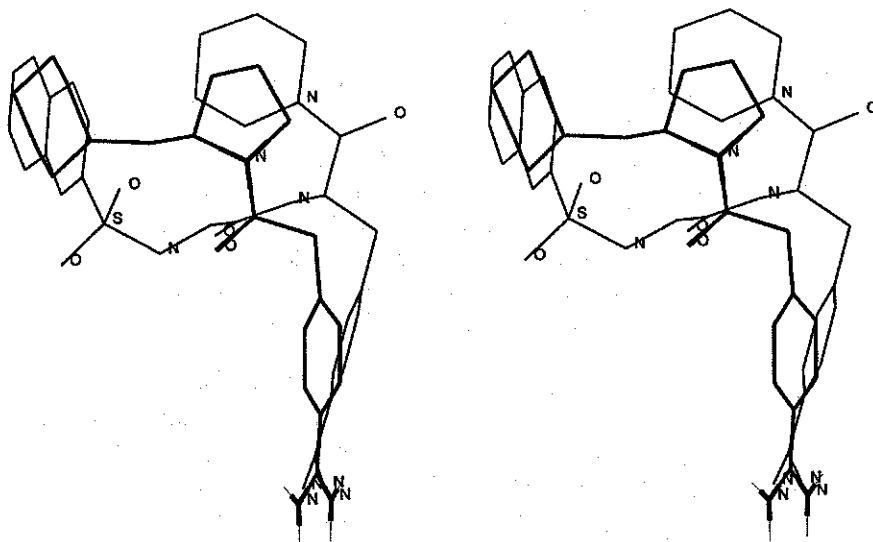


Figure 8 Minimized structure of a putative ligand suggested by CCLD (thick lines). The CCLD run used MCSS minima of benzamidine, benzene, cyclopentane, and cyclohexane. The putative ligand was minimized in the thrombin active site, whose residues within 8 Å of any atom of the ligand were allowed to move. The remaining residues of thrombin were kept rigid. The NAPAP structure is also shown (thin lines) as a basis of comparison.

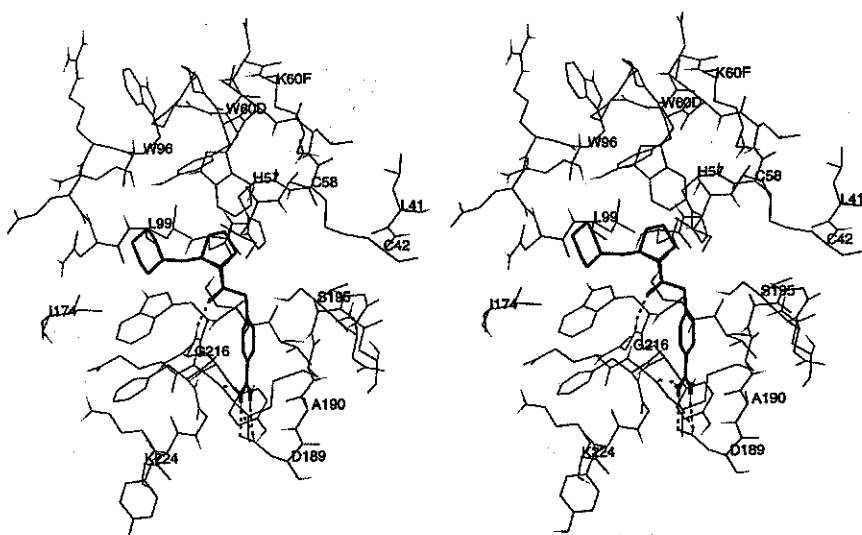


Figure 9 Stereo view of the minimized complex between the putative ligand shown in Figure 8 and thrombin (thin lines). Intermolecular hydrogen bonds are drawn by dashed lines.

N-acylpyrrolidine ring in S2 by a single methylene linker. This is a novel design and the candidate ligand appears to be more rigid than NAPAP, since it has a smaller number of rotatable bonds. Hence, the penalty paid for the loss in entropy upon binding should be smaller for this CCLD hit than for NAPAP.

In a study with a more diverse set of starting molecular fragments, ligands containing both a "core" similar to known inhibitors and additional intermolecular hydrogen bonds and/or van der Waals interactions were generated [28].

IV. CONCLUSION

A combinatorial approach for the computer-aided design of putative ligands of proteins or receptors of known three-dimensional structure has been presented. Diversity of these candidate ligands is provided by first docking a set of diverse molecular fragments. For aliphatic functional groups a modified force field (switching off the attractive part of the van der Waals interaction with polar atoms of the protein) was introduced to better approximate solvation effects, thereby avoiding the docking of apolar fragments into hydrophilic cavities of the macromolecular target. The second part of the present ligand design approach consists of a combinatorial strategy for the connection of optimally docked molecular fragments by small linkage elements having optimal interactions with the target molecule. An application was presented in which candidate inhibitors of thrombin were found.

It is important to note that the most difficult part of ligand design is the prediction of binding affinity. A method to estimate relative binding constants has recently been applied to a series of similar inhibitors of HIV-1 aspartic proteinase [41]. Our own efforts are concentrated on the development of an approach that will predict relative binding affinities and will be general enough to be used for any enzyme or receptor of known structure (A. Caffisch, D. Arosio, and C. Ehrhardt, work in progress).

One of the purposes of this chapter was to show that structure-based ligand design is a fascinating and progressing research field. It is fascinating not only for its ultimate goal, i.e., the discovery of ethical drugs, but also because it is based on, and thereby increases, our understanding of molecular interactions and recognition phenomena on an atomic level. Another reason is its multidisciplinary character, which requires skills in different branches of science, from theoretical physics and chemistry to computer science and statistics. That structure-based computer-aided ligand design is a progressing field is evident in many chapters of this book. This is mainly a consequence of the methodological developments and the ever-increasing performance of computers.

ACKNOWLEDGMENTS

We thank J. Apostolakis and Professor A. Plücker for helpful discussions. The calculations were performed on an SGI Indigo2 and an eight-processor SGI Challenge (R4400 processors). The CHARMM program within the version 4.0 of the QUANTA software package (Biosym-MSI Inc) was used for some of the minimization performed in this work. The CCLD program is available from A. Cafilisch.

This work was supported by the Swiss National Science Foundation (Schweizerischer Nationalfonds grant nr. 3100-043423.95) and by Novartis Pharma Inc.

REFERENCES

1. Gubernator K, Broger C, Bur D, Doran DM, Gerber PR, Müller K, Schaumann TM. Structure-based ligand design. In: Hermann EC, Frankle R, eds. *Computer-Aided Drug Design in Industrial Research* 1995:61-77.
2. Müller K. Paradigms of rational molecular design. In: Schwartz TW, Hjorth SA, Kastrup JS, eds. *Structure and Function of 7TM Receptors*, in press.
3. Gerber PR, Gubernator K, Müller K. Generic shapes for the conformational analysis of macrocyclic structures. *Helv Chim Acta* 1988; 71:1429-1441.
4. Böhm HJ. The computer program LUDI: a new method for de novo design of enzyme inhibitors. *J Comput-Aided Mol Design* 1992; 6:61-78.
5. Böhm HJ. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J Comput-Aided Mol Design* 1992; 6:593-606.
6. Böhm HJ. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput-Aided Mol Design* 1994; 8:243-256.
7. Böhm HJ. On the use of LUDI to search the fine chemicals directory for ligands of proteins of known three-dimensional structure. *J Comput-Aided Mol Design* 1994; 8:623-632.
8. Böhm HJ. Site-directed structure generation by fragment-joining. *Perspectives in Drug Discovery and Design* 1995; 3:21-33.
9. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* 1985; 28:849-857.
10. Bobbyer DNA, Goodford PJ, McWhinnie PM, Wade RC. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *J Med Chem* 1989; 32:1083-1094.
11. Wade RC, Clark KJ, Goodford PJ. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds. *J Med Chem* 1993; 36:140-147.
12. Wade RC, Goodford PJ. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known

- structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J Med Chem* 1993; 36:148–156.
13. Rotstein SH, Murcko MA. GroupBuild: A fragment-based method for de novo drug design. *J Med Chem* 1993; 36:1700–1710.
 14. Moon JB, Howe WJ. Computer design of bioactive molecules: a method for receptor-based de novo ligand design. *Proteins: Structure, Function and Genetics* 1991; 11:314–328.
 15. Eisen MB, Wiley DC, Karplus M, Hubbard RE. HOOK: A program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins: Structure, Function and Genetics* 1994; 19:199–221.
 16. Tschinke V, Cohen NC. The NEWLEAD program: a new method for the design of candidate structures from pharmacophoric hypotheses. *J Med Chem* 1993; 14:3863–3870.
 17. Gillet VJ, Myatt G, Zsoldos Z, Johnson AP. SPROUT, HIPPO and CAESA: Tools for de novo structure generation and estimation of synthetic accessibility. *Perspectives in Drug Discovery and Design* 1995; 3:34–50.
 18. Lewis RA, Roe DC, Huang C, Ferrin TE, Langridge R, Kuntz ID. Automated site-directed drug design using molecular lattices. *J Mol Graphics* 1992; 10:66–78.
 19. Kuntz ID. Structure-based strategies for drug design and discovery. *Science* 1992; 257:1078–1082.
 20. Caflisch A, Karplus M. Computational combinatorial chemistry for de novo ligand design: Review and assessment. *Perspectives in Drug Discovery and Design* 1995; 3:51–84.
 21. Miranker A, Karplus M. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins: Structure, Function, and Genetics* 1991; 11:29–34.
 22. Caflisch A, Miranker A, Karplus M. Multiple copy simultaneous search and construction of ligands in binding sites: application to inhibitors of HIV-1 aspartic proteinase. *J Med Chem* 1993; 36:2142–2167.
 23. Appelt K. Crystal structures of HIV-1 protease-inhibitor complexes. *Perspectives in Drug Discovery and Design* 1993; 1:23–48.
 24. Wlodawer A, Erickson JW. Structure-based inhibitors of HIV-1 protease. *Annu Rev Biochem* 1993; 62:543–585.
 25. Stubbs MT, Bode W. Crystal structures of thrombin and thrombin complexes as a framework for antithrombotic drug design. *Perspectives in Drug Discovery and Design* 1993; 1:431–452.
 26. Hilpert K, Ackermann J, Banner DW, Gast A, Gubernator K, Hadvary P, Labler L, Müller K, Schmid G, Tschopp T, van de Waterbeemd H. Design and synthesis of potent and highly selective thrombin inhibitors. *J Med Chem* 1994; 37:3889–3901.
 27. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 1983; 4:187–217.
 28. Caflisch A. Computational combinatorial ligand design: Application to human α -thrombin. *J Computer-Aided Molec Design* 1996; 10:372–396.
 29. Gelin BR, Karplus M. Sidechain torsional potentials and motion of amino acids in proteins: bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci USA* 1975; 72:2002–2006.

30. O'Shea EK, Rutkowski R, Kim PS. Evidence that the leucine zipper is a coiled coil. *Science* 1989; 243:538-542.
31. O'Shea EK, Klemm JD, Kim PS, Alber T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* 1991; 254:539-544.
32. Warwick J, Watson HC. Calculation of the electric potential in the active site cleft due to α -helix dipoles. *J Mol Biol* 1982; 157:671-679.
33. Gilson MK, Honig BH. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins: Structure, Function, and Genetics* 1988; 4:7-18.
34. Bashford D, Karplus M. pK_a s of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* 1990; 29:10219-10225.
35. Davis ME, Madura JD, Luty BA, McCammon JA. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program. *Comp Phys Comm* 1991; 62:187-197.
36. Bode W, Mayr I, Baumann U, Huber R, Stone SR, Hofsteenge J. The refined 1.9-Å crystal structure of human α -thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J* 1989; 8:3467-3475.
37. Banner DW, Hadvary P. Crystallographic analysis at 3.0-Å resolution of the binding to human thrombin of four active site-directed inhibitors. *J Biol Chem* 1991; 266:20085-20093.
38. Obst U, Gramlich V, Diederich F, Weber L, Banner DW. Design neuartiger, nichtpeptidischer Thrombin-Inhibitoren und Struktur eines Thrombin-Inhibitor-Komplexes. *Angew Chem* 1995; 107:1874-1877.
39. Håkansson K, Tulinsky A, Abelman MM, Miller TA, Vlasuk GP, Bergum PW, Lim-Wilby MSL, Brunck TK. Crystallographic structure of a peptidyl keto acid inhibitor and human α -thrombin. *Bioorganic and Medicinal Chemistry* 1995; 3:1009-1017.
40. Tapparelli C, Metternich R, Ehrhardt C, Cook NS. Synthetic low-molecular weight thrombin inhibitors: molecular design and pharmacological profile. *TIPS* 1993; 14:366-376.
41. Holloway MK, Wai JM, Halgren TA, Fitzgerald PMD, Vacca JP, Dorsey BD, Levi RB, Thompson WJ, Chen LJ, deSolms SJ, Gaffin N, Ghosh AK, Giuliani EA, Graham SL, Guare JP, Hungate RW, Lyle TA, Sanders WM, Tucker TJ, Wiggins M, Wiscount CM, Woltersdorf OW, Young SD, Darke PL, Zugay JA. A priori prediction of activity for HIV-1 protease inhibitors employing energy minimization in the active site. *J Med Chem* 1995; 38:305-317.
42. Kraulis P, Molscript, a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 1991; 24:946-950.