Journal of Molecular Medicine

DOI 10.1007/s001090000084

Review

Recent developments in structure-based drug design

Gerhard Klebe(✉)

G. Klebe
Institute of Pharmaceutical Chemistry, University of Marburg, Marbacher Weg 6, 35032 Marburg, Germany

✉ E-mail: Klebe@mailer.uni-marburg.de

Received: 3 November 1999 / **Accepted:** 14 December 1999 / **Published online:** ▮ 2000

GERHARD KLEBE

received his Ph.D. in physical chemistry from the University of Frankfurt, Germany. After three years of postdoctoral work in crystallography and neutron diffraction, he jointed the BASF company in 1984 to found a group for drug design and crystallography. In 1991 he received his *Habilitation* degree in pharmaceutical chemistry from the University of Heidelberg. Since 1996 he has been Full Professor of Pharmaceutical Chemistry at the University of Marburg. His research interests are the crystallography of protein-ligand complexes, structure-based drug design, software development of drug design tools, database analyses, and thermodynamics. He is the coauthor of a textbook on drug design, which may serve as introduction to this area for less experienced readers.

Abstract. Structure-based design has emerged as a new tool in medicinal chemistry. A prerequisite for this new approach is an understanding of the principles of molecular recognition in protein-ligand complexes. If the three-dimensional structure of a given protein is known, this information can be directly exploited for the retrieval and design of new ligands. Structure-based ligand design is an iterative approach. First of all, it requires the crystal structure or a model derived from the crystal structure of a closely related homolog of the target protein, preferentially complexed with a ligand. This complex unravels the binding mode and conformation of a ligand under investigation and indicates the essential aspects determining its binding affinity. It is then used to generate new ideas about ways of improving an existing ligand or of developing new alternative bonding skeletons. Computational methods supplemented by molecular graphics are applied to assist this step of hypothesis generation. The features of the protein binding pocket can be translated into queries used for virtual computer screening of large compound libraries or to design novel ligands de novo. These initial proposals must be confirmed experimentally. Subsequently they are optimized toward higher affinity and better selectivity. The latter aspect is of utmost importance in defining and controlling the pharmacological profile of a ligand. A prerequisite to tailoring selectivity by rational design is a detailed understanding of molecular parameters determining selectivity. Taking examples from current drug development programs (HIV proteinase, t-RNA transglycosylase, thymidylate synthase, thrombin and, related serine proteinases), we describe recent advances in lead discovery via computer screening, iterative design, and understanding of selectivity discrimination.

Keywords. Drug design - Computer methods - X-ray crystallography - Three-dimensional QSAR - Comparative molecular field analysis

Abbreviations. *HIV*: Human immunodeficiency virus *TGT*: Guanosine transglycosylase

Introduction

The number of proteins of known three-dimensional structure has increased exponentially in recent years [*1*, *2*]. This is due mainly to significant improvements in gene technology, protein chemistry, and structure elucidation techniques. These methodological developments have given rise to the expectation that the structure determination or reliable homology modeling of a particular target protein can be achieved in a reasonably short time, that is, in a period that is sufficiently short for the structure to become available at an early stage of a drug design and development project. Furthermore, considerable effort is being spent in sequencing the genomes of various species. The human genome is expected to be fully characterized in the next few years. It is likely that structural genomics will carry on present efforts, and that substantial knowledge will be gathered in the near future about the sequence-structure-function relationship of proteins. This will discover new potential targets for disease therapy.

This will considerably increase the demand for new lead structures tailored to a particular target. Two strategies are presently used to discover new leads: experimental high through-put screening of large compound libraries and, alternatively or as a complement, computational methods for virtual screening and de novo design [*3*, *4*, *5*, *6*]. The latter approach requires the three-dimensional shape of the binding site of the target protein under consideration as prerequisite for the search of new ligands. Putative lead structures can be found via fast docking of candidate molecules from large databases into the binding site. As an alternative, ligands can be designed de novo, exploiting structural features at the binding site of the target protein. Novel molecular frameworks have been proposed, often assembled from smaller molecular portions that appear ideally suited to accommodate the binding site, and that operate as countergroups for the functional groups of active-site amino acid residues [*1*]. Initial leads require structure optimization to combine high-affinity binding with sufficient selectivity

and specificity for the target under consideration. Usually the optimization process is performed iteratively, involving a close link between design considerations, structure determination, and synthesis. Successful optimization requires a better understanding of the factors responsible for binding affinity and specificity. Structural evidence and analytical techniques must be developed to provide this understanding.

The present review demonstrates the potential of various techniques involved in state-of-the-art drug research, illustrated by case studies from the literature and our own work.

The design cycle

Structure-based drug design of protein ligands has emerged as a new tool in medicinal chemistry [*1*, *2*, *7*, *8*, *9*]. It is generally applied as an iterative approach. The process starts with a detailed analysis of the binding site of the target protein, which is preferably complexed with a ligand. All aspects thought to be responsible for binding affinity and selectivity are collected. This knowledge is then used to create new ideas on ways to improve existing ligands or to develop new alternative bonding skeletons. At this stage computational methods supplemented by molecular graphics are applied to assist the development of new hypotheses. Computer methods are used either to exploit information either from the binding site geometry of the given protein or to consider molecular shape similarity with other, already known ligands. They generate new molecules de novo from this knowledge, or they embark into elaborate computer searches of compound libraries to produce new suggestions. In the next step the experience of synthetic organic chemists is consulted to avoid proposals of compounds unfeasible for synthesis. The newly suggested compounds must then be produced and tested. Subsequently their binding properties are determined by biochemical, crystallographic, and spectroscopic methods. The next cycle of ligand design begins with careful analysis of the obtained results using the three-dimensional geometry of the new crystallographically characterized complexes.

Some historical remarks

The first example of structure-based design was reported by the group of Beddell and Goodford in 1973 at Wellcome Laboratories in the United Kingdom [10, *11*]. Hemoglobin was selected as a target, which at the time was the only example of pharmacological relevance with a known crystal structure. The goal of the studies was to develop a ligand that acts similarly to the natural allosteric effector diphosphoglycerate (**1**; Fig. 1). This endogenous ligand binds to hemoglobin and regulates its oxygen affinity. Taking this molecule as a reference, the Wellcome group designed dialdehyde derivatives (**2**) and related bisulfite adducts (**3**) which, as expected, modify the oxygen affinity to hemoglobin (Fig. 1). Several years later the antihypertensive captopril (**4**), inhibits the angiotensin-converting enzyme, was introduced onto the market; this was the first drug to be developed using structural information. Although a peptidic lead had already been known from a snake poison, the important breakthrough was achieved only in the mid-1970s by Cushman et al. [*12*] at Squibb after modeling the active site of the enzyme (Fig. 1). As a structural reference the binding site of carboxy peptidase was used. Shortly before this the crystal structure of this enzyme had been determined [*13*]. Similar to the angiotensin-converting enzyme carboxy peptidase, this is a zinc proteinase.

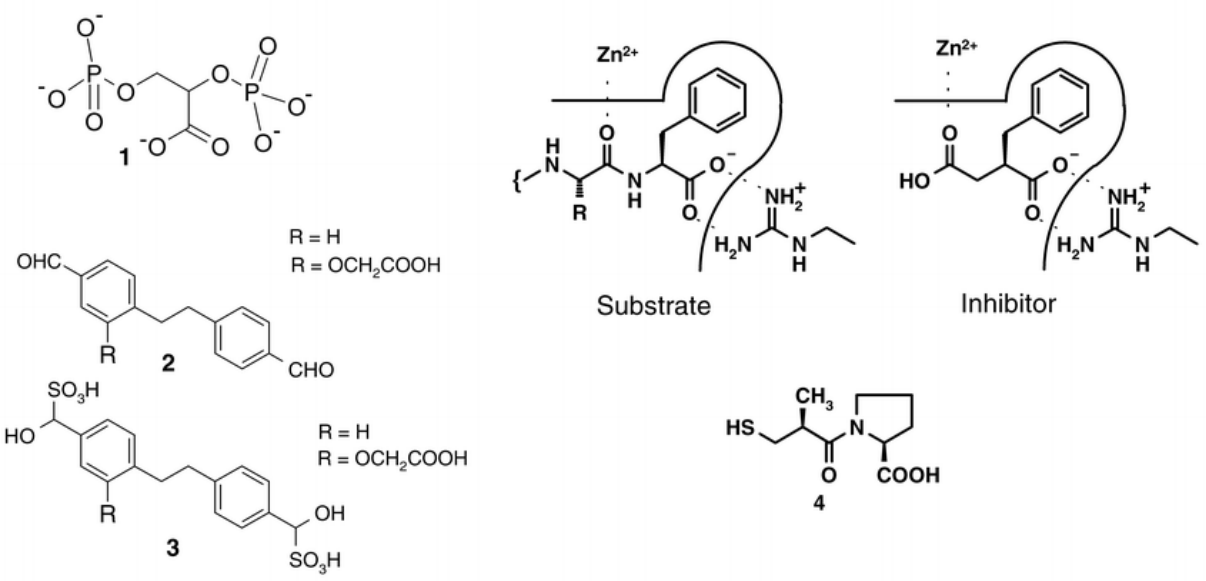


Fig. 1. *Left* The allosteric effector diphosphoglycerate (**1**) regulates oxygen affinity toward hemoglobin. In a design study dialdehyde derivatives (**2**) and bisulfite adducts (**3**) have been proposed to compete for the binding site in hemoglobin. *Right* Captopril (**4**) is a potent inhibitor of angiotensin-converting enzyme. Its design was guided by the crystal structure of an inhibitor complex with carboxy peptidase, a zinc proteinase whose structure had been determined shortly before

---

The past 20 years of drug design have witnessed the structural characterization of a tremendously number of therapeutically important targets. The increasing number of successful applications of drug design has led to the discovery of new therapeutics [*1*, *2*, *7*, *8*, *9*, *14*]. The recent development of human immunodeficiency virus (HIV) protease inhibitors has convincingly demonstrated the impact and the relevance of structure-based approaches to the development of new drugs.

Lead generation through extensive database searches: new nonpeptidic HIV protease inhibitors

The acquired immunodeficiency syndrome is caused by HIV, which codes for a proteinase required for its replication. The inhibition of HIV proteinase therefore offers the therapeutic possibility of interfering with replication of the virus [*15*]. The existence of the proteinase was postulated in 1985, its experimental characterization followed in 1988, and in 1989 its crystal structure was determined. The active site of this aspartyl proteinase defines the structural requirements for possible ligands (Fig. 2, left). Researchers in the drug design group of Dupont-Merck have transformed the arrangement of the active-site amino acid residues into a pharmacophore pattern required in putative ligands [*16*]. With a separation of 8.5-12 Å two lipophilic groups must be available and accompanied by additional hydrogen-bond donor or acceptor groups in a neighborhood of 3.5-6.5 Å distance. Furthermore, the replacement of a structural water molecule has been attempted, and consequently an appropriate functional group between the two lipophilic groups has been required (Fig. 2, right). Extensive search of the database on crystal structures of small organic molecules [*17*] (the Cambridge Structural Database) revealed a substituted phenol (**5**) as putative lead structure. Further computer experiments and extensive discussions with the synthetic chemists finally suggested a cyclic urea as a new central scaffold. This led to the development of the compound DMP-323 (**6**), which has now reached the stage of clinic trials. The crystal structure of a related compound has confirmed that the

postulated binding mode. The carbonyl oxygen replaces the structural water molecule, and the adjacent hydroxy groups bind to the two catalytic aspartates.

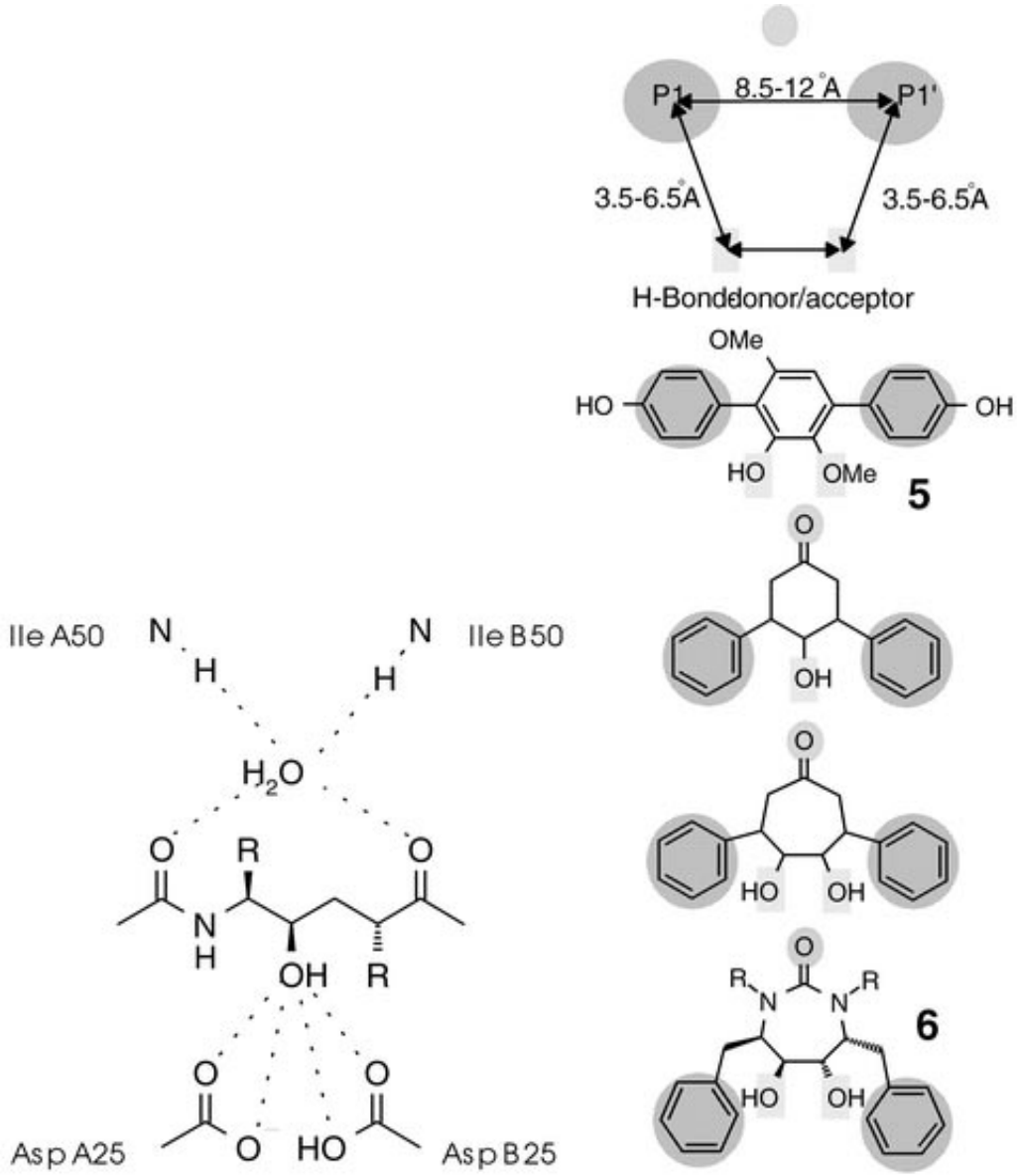


Fig. 2. *Left* Schematic binding mode of a substrate-analog inhibitor bound to the catalytic center of the HIV-aspartyl protease. The structural requirement determined by the contacts to the protein are used to define the geometric constraints for a database query to retrieve alternative molecular skeletons. *Right* Searches of the Cambridge Structural Database revealed a substituted phenol (**5**) that has been gradually optimized to the cyclic urea (**6**)

Lead generation by fast fragment docking: first putative inhibitors of t-RNA guanosine transglycosylase

The enzyme t-RNA guanosine transglycosylase (TGT) plays an important role in *Shigella* dysentery [*18, 19, 20*]. This disease is a frequent infection in the Third World and causes more than 500,000 infant deaths per year [*21*]. The administration of antibiotics provides a possible therapy; however, this results in a total loss of intestinal flora. Even worse is the fact that antibiotic resistance has developed rapidly against all the well established antibiotics. Therefore a new and selective drug is

greatly needed. The disease is caused by *Shigella* bacteria that are closely related to *Escherichia coli*. They cause rapid inflammation of the intestinal mucosa, acquiring their virulence via the transfer of pathogenicity coding genes. A strong reduction in virulence has been achieved by the reduction or loss of activity of TGT [*19*]. The enzyme is involved in queuine biosynthesis. Queuine is a modified guanine base that is introduced into t-RNA. This step is catalyzed by TGT. Queuine biosynthesis is not essential for *E. coli* or *Shigella*; however, inhibiting TGT may abolish the pathogenicity of the latter. This fact is important for the development of a selective antibiotic.

The crystal structure of TGT has been determined and the apo-form of the enzyme soaked successfully with $preQ_1$ (**7**; Fig. 3), a weak substrate-analog inhibitor [*22*]. To confirm the described therapeutic principle requires more potent with higher selectivity. Based on the crystal structure of the TGT-$preQ_1$ complex, computer screening for putative small molecule inhibitors has been started, applying the following strategy. The computer program used in the present study is LUDI [*23*, *24*], which initially analyses the protein binding site in terms of its amino acid composition. It first calculates interaction sites around the functional groups of the amino acid residues exposed to the binding site. These sites correspond to spatial positions at which appropriate functional groups of putative small molecule ligands should be placed [*25*]. Information about the spatial positions has been extracted from statistical evaluations of crystal data. The program accesses a database of several hundred thousand small molecules classified in terms of their favorable interaction properties, for example, whether they act as hydrogen-bond donors, acceptors, or favorable hydrophobic binding partners. In the next step the program tries to fit these small molecules onto the previously generated interaction sites in the protein. If the program finds a favorable placement for a particular candidate molecule, it is stored; otherwise the trial structure is discarded. In a final step the program tries to estimate the affinity of the found ligands. LUDI is a fast program; a database of the relevant size can be screened on a standard workstation overnight.

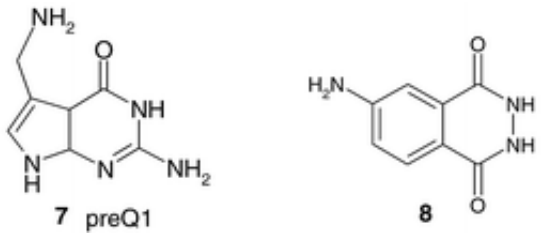


Fig. 3. The enzyme tRNA guanine transglycosylase (TGT), a t-RNA modifying enzyme, is cocrystallized with the substrate-analog inhibitor preQ$_1$ (**7**). Extensive computer searches with LUDI suggested the 4-amino phthalohydrazide (**8**) as an initial lead for inhibitor design. Its binding was confirmed by crystal structure analysis

---

In any computational approach such as this, suggested ligands must be ranked in terms of their expected potency. This potency is usually expressed in terms of the binding constant K_i, which is thermodynamically related to the Gibbs free enthalpy. This in turn is composed of enthalpic and entropic contributions. The enthalpic portion considers the energetic part of the protein-ligand interactions. These are determined principally by electrostatic and van der Waals interactions; however, they contribute only if they are stronger at the binding site than the aqueous solution from which the binding partners have assembled. The entropic portion measures the ordering of a system and the distribution of energetic parameters over the degrees of freedom of the entire system. The solvent plays an important role in this context since ubiquitously present water molecules are displaced, rearranged, or reordered at the protein-ligand interface upon complex formation. The major part of the entropic portion is generally correlated with the lipophilic surface, either of the ligand or the protein that becomes buried upon binding. Recent theoretical studies have explicitly considered the description of solvation properties [*26*, *27*].

The affinity ranking in LUDI is evaluated by a crude empirical scoring function [*28*] which includes various contributions such as the number of formed hydrogen bonds, the lipophilic surface that becomes buried upon binding, and the number of molecular degrees of freedom immobilized during the binding process. More recent developments based on the structural information stored in databases on protein-ligand complexes. These approaches focus on the compiling of statistical pair potentials that are subsequently used to rank ligand-binding modes. These methods are very fast in computing time and show high reliability [*29*, *30*, *31*, *32*]. Computer screening with LUDI suggests several putative ligands for TGT, all scoring well in the range found for trypsin inhibitors of similar molecular weight that have been confirmed to bind to this serine protease. Furthermore, the LUDI score is calibrated on the basis of a protein-ligand complex of a substrate-analog inhibitor which has an affinity of 0.5 µM. Its computed affinity has been found to be about 20% better than the best scored hits from the LUDI search. Since part of the screened library (about 120,000 entries) have been derived from the Available Chemicals Directory (MDL Information Systems, San Leandro, Calif., USA), a collection of commercially available compounds, about ten of the retrieved hits were purchased and subsequently assayed. Several of these were unsuited due to poor solubility or limited chemical stability. However, for binding of three of the assayed LUDI hits was detected experimentally. Soaking with the enzyme has been attempted, and the compound (**8**) shown in Fig. 3 was successfully accommodated in the crystal structure of TGT. This ligand was the second best scored hit and proved to have a binding constant of 5.7 µM. Interestingly, experimental results confirm the computed binding mode of the probe ligand to be within the range of experimental error. This structure presently serves as a starting point for our next cycle of drug design. It is aimed at the development of a larger inhibitor using more of the exposed residues in the active site and thus yielding a higher affinity and specificity for the TGT binding site [33].

An unsatisfactory lead calls for the design of novel ligands: inhibitors of thymidylate synthese

The crystal structure of thymidylate synthese, a key enzyme in the biosynthesis of purine, was determined by Montfort et al. [*34*], together with two ligands: 5-fluorodesoxyuridylate (**9**) mimicking the substrate, and *N*10-propargyl-5,8-desazafolate (**10**; CB3717), mimicking the cofactor. Both compounds are potent inhibitors. Depending on the crystal structure, the inhibitor CB3717 binds with its quinazoline moiety via the protonated N3 to the carboxylate of Asp-169 (Fig. 4). The exocyclic NH_2 group forms a hydrogen bond to the backbone carbonyl of Ala-263. A second hydrogen bond is mediated via a water molecule (W402) to Asp-169. An additional water molecule (W430) binds to N1 while sustaining H-bonds to Ala-263 and Arg-21. The *p*-aminobenzyl moiety at the 6-position occupies a channel formed by aromatic residues and orients the terminal glutamate toward the entrance of the binding pocket.

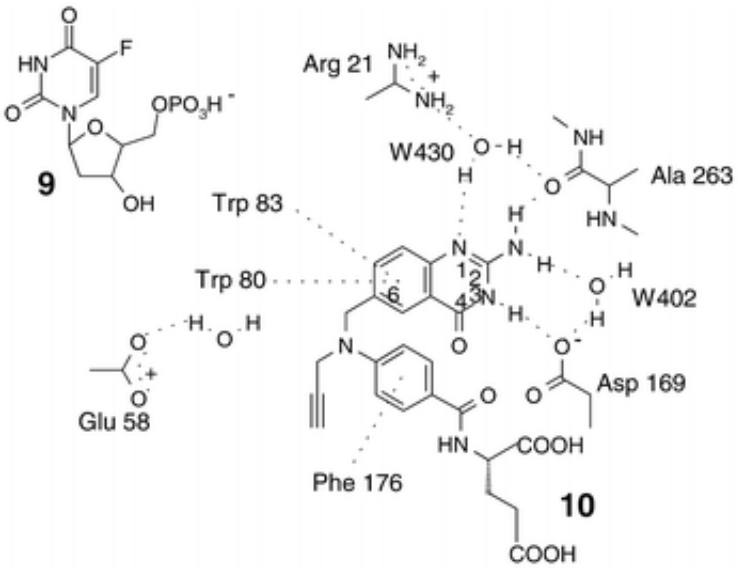


Fig. 4. The crystal structure of thymidylate synthase was determined together with the two ligands 5-fluorodesoxyuridylate (**9**) and N10-propargyl-5,8-desazafolate (**10**; CB3717). The binding mode of the latter ligand served as a starting point for the design of new inhibitors with alternative bonding skeletons and better physicochemical and metabolic properties

---

CB3717 showed unsatisfactory solubility and transportation properties, and an observed kidney toxicity terminated its clinical trials. Agouron Pharmaceuticals (San Diego, Calif., USA) began a drug design project for the development of alternative thymidylate inhibitors. In addition to an optimization and partial redesign of the CB3717 structure, the researches at Agouron accomplished a complete ligand de novo design [*35*, *36*, *37*, *38*].

The starting point was the crystallographically observed binding mode of CB3717. The inhibitor forms a hydrogen bond as a donor to the carboxylate of Asp-169. In turn, as an acceptor, it interacts with W430. A possible ligand therefore requires a structural building block with a donor and acceptor facility in close neighborhood, enclosing an angle of approximately 120°. Either an imidazole or *cis*-amide should satisfy these requirements (Fig. 5).

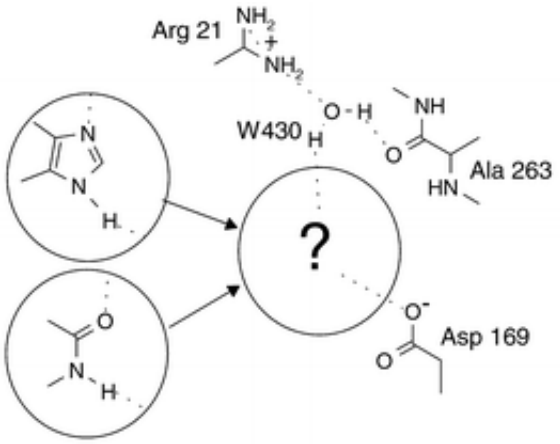


Fig. 5. The quinazoline moiety in CB3717 (**10**) operates as a hydrogen bond donor and acceptor to the protein. Possible fragments to replace this moiety are an imidazole or a *cis*-peptide group. Both fragments have been imbedded into a larger portion and serve as initial ideas for new lead structures

---

The next question was how to incorporate such structural elements into a larger molecular entity. The program GRID allows putative binding sites to be mapped using a variety of different probes [*39*]. A methyl probe has revealed a putative binding site close to the aromatic residues Trp-80, Trp-83, and Phe-176, and the primarily selected imidazole ring has therefore been embedded into an imidazo tetrahydroquinoline portion. The piperidine ring fills the lipophilic pocket, and its nitrogen serves as an achiral anchor to connect with a side chain. Based on previous experience, a phenylsulfonyl group with a terminal piperazine moiety has been selected for this purpose (Fig. 6). After synthesis and testing a micromolar inhibitor (**11**) was designed. The subsequently determined crystal structure with the enzyme confirmed the postulated binding mode and suggested the introduction of an additional hydrogen bond. The backbone carbonyl group of Ala-263 (Fig. 6) falls close to the 2-position of the imidazole ring. An additional NH_2 group at this position appeared to be ideally suited to enable another hydrogen bond to the protein. In fact, the corresponding 2-amino derivate shows a substantially increased affinity.

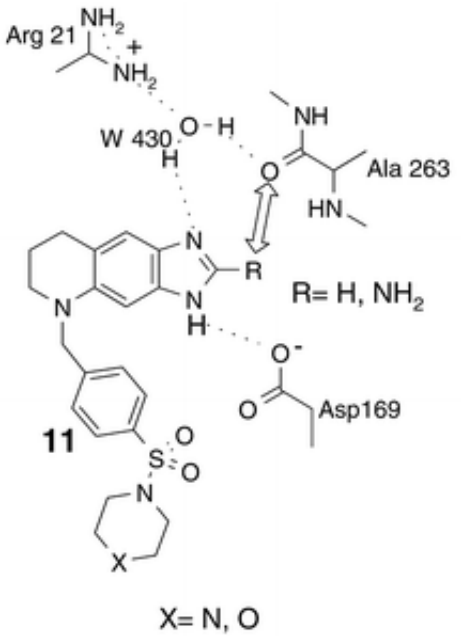


Fig. 6. Schematic binding mode of an imidazo tetrahydroquinoline (**11**) inhibiting thymidylate synthase. The crystal structure confirmed the design concept and suggested the introduction of an exocyclic NH_2 group in the 2-position of the imidazole ring

---

Its computed binding mode, based on the crystallographically observed one of the unsubstituted derivative, differs surprisingly from the subsequently determined crystal structure: the water W430, previously always involved in binding of inhibitors, is not observed in the structure. However, the predicted hydrogen bond to Ala-263 was detected. What causes the deviations from the computer predictions? Protons cannot be detected by protein crystallography. To obtain an idea about the protonation state of the ligand the pK_a value of the 2-amino derivate was determined. The measured value of 8.2 suggests that the ligand binds in its protonated state. Moving from the unsubstituted to the substituted derivative corresponds to a transfer from a nonprotnated to a protonated species. This has important consequences for binding. Nitrogen N1, thought to operate as an acceptor to W430, changes into a donor. In addition, protonation creates a charge on the imidazole moiety. Consequently the hydrogen bonds to Ala-263 and Asp-169 are transformed into much stronger but charge-assisted H-bonds. This effect is likely to be responsible for the strong increase in affinity upon NH_2 substitution.

As discussed above, the imidazole moiety is only one possible structural unit for satisfying the required hydrogen bonding pattern. A *cis*-amide bond may offer an alternative solution (Fig. 5). The hydrophobic nature of the remaining binding site was suggested by GRID, and the scientists at Agouron fused a *cis*-amide group with a naphthyl moiety to reveal a naphthostyril (Fig. 7). To connect with an appropriate side chain, an exocyclic nitrogen atom was introduced at position 5 (Fig. 7). An ethyl and a *p*-piperazinyl sulfonylbenzyl group have been selected to substitute this nitrogen.

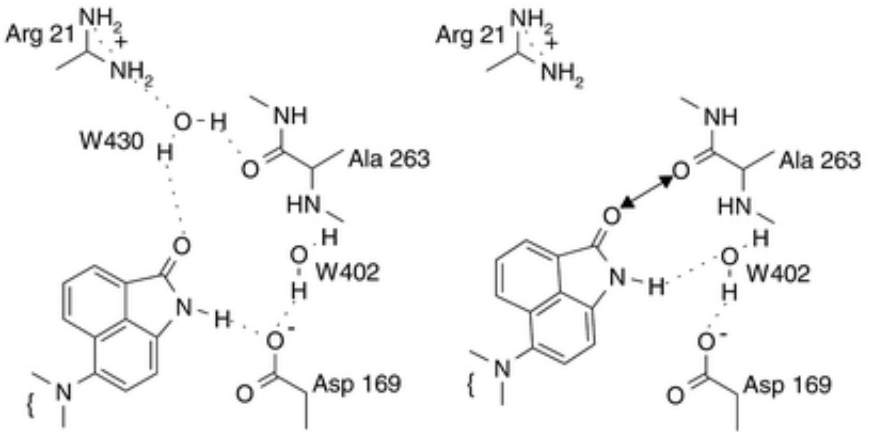


Fig. 7. Schematic binding mode of a naphthostyril moiety imbedding the suggested *cis*-amide bond. The subsequently performed crystal structure (*right*) showed that the anticipated hydrogen bond to the carbonyl group of the inhibitor, suggested by modeling (*left*), is not formed. To transform the presumed unfavorable interaction between the two neighboring oxygen atoms to an attractive H-bond, the lactame oxygen has been modified to an imino group (Fig. 8)

After synthesis the proposed binding mode was largely confirmed by crystal structure analysis. However, differences between predicted and observed binding modes arose around the lactame ring. The expected direct hydrogen bond to Asp-169 was not formed; instead a water molecule mediates an interaction between the ligand and this protein residue (Fig. 7). Furthermore, the water W430 is not observed in the crystal structure, and a surprisingly short repulsive contact is detected between the lactame carbonyl oxygen and the backbone carbonyl oxygen of Ala-263. To transform this presumably unfavorable interaction into an attractive H-bond the lactame oxygen has been modified into an imino group (**12**; Fig. 8). Synthesis of the imino lactame revealed an inhibitor of 20-fold increased affinity. Again, this major increase in affinity is due to changes in the protonation state. The pK_a value of the latter imino derivative is greater than 8, and therefore the compound should bind in its protonated state. Interactions to Asp-169 and Ala-263 are transformed into charge-assisted hydrogen bonds. The subsequently determined crystal structure of the imino derivate with the enzyme confirmed these postulations and showed that water W430 had returned to the complex (Fig. 8).

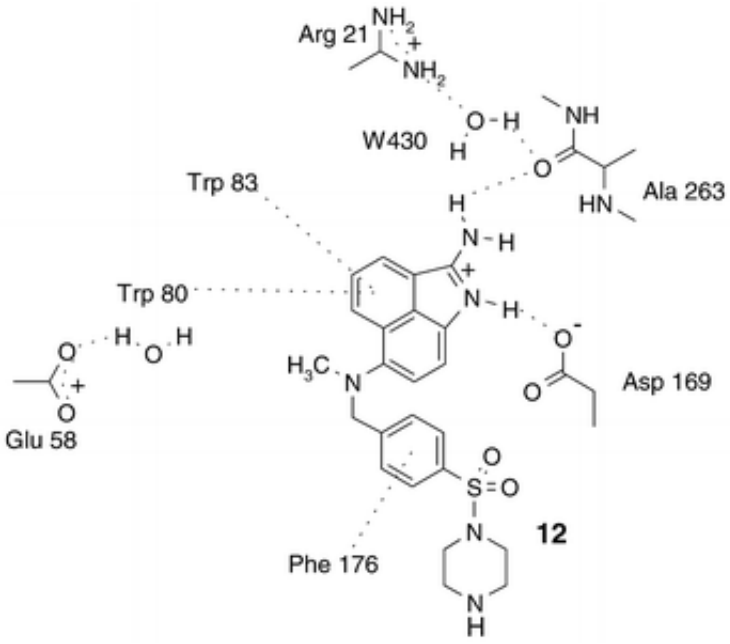


Fig. 8. Schematic binding mode of the newly designed inhibitor **12**, still with an imino lactame moiety. Increased binding affinity is observed, and the proposed binding mode was be confirmed by X-ray crystallography

---

Details from other inhibitor complexes prepare for a design from scratch: de novo design of nonpeptidic thrombin inhibitors

Thrombin is a trypsinlike serine protease that plays a major role in blood coagulation [*40*]. Selective inhibition of this enzyme would help to prevent life-threatening thrombotic diseases.

The binding site of thrombin is geometrically conserved in a large variety of ligand-protein complexes, and thus a fairly rigid structure can be presumed. In the early 1990s Obst et al. [*41*, *42*] investigated the available crystal structures of ligand-thrombin complexes to elucidate the common structural features in these examples. The enzyme provides a narrow specificity pocket comprising the carboxylate group of Asp-189 at the lower end along with a glycine backbone carbonyl group and a structural water molecule (Fig. 9). A positively charged benzamidinium group appears to be optimally suited as binding partner for this arrangement. A large hydrophobic pocket opens to the left of the catalytic center. This is structured in two parts, a smaller proximal (P) and a larger distal (D) pocket. The distal pocket is flanked by the indole moiety of Trp-215 and the aliphatic side chains of Ile-174 and Leu-99. Aromatic moieties such as phenyl or naphthyl rings are well accommodated in this so-called aryl binding pocket. The proximal pocket exhibits features specific to thrombin due to Tyr-60A and Trp-60D present in the extended so-called 60-loop. Aliphatic functional groups present at this position in a ligand ensure good binding properties. As found in all serine proteases, substrates and inhibitors must perform hydrogen bonds with the nonspecific peptide recognition site in thrombin with Gly-216. Figure 9 (top) shows the crystallographically confirmed binding mode of the potent inhibitor NAPAP (**13**).

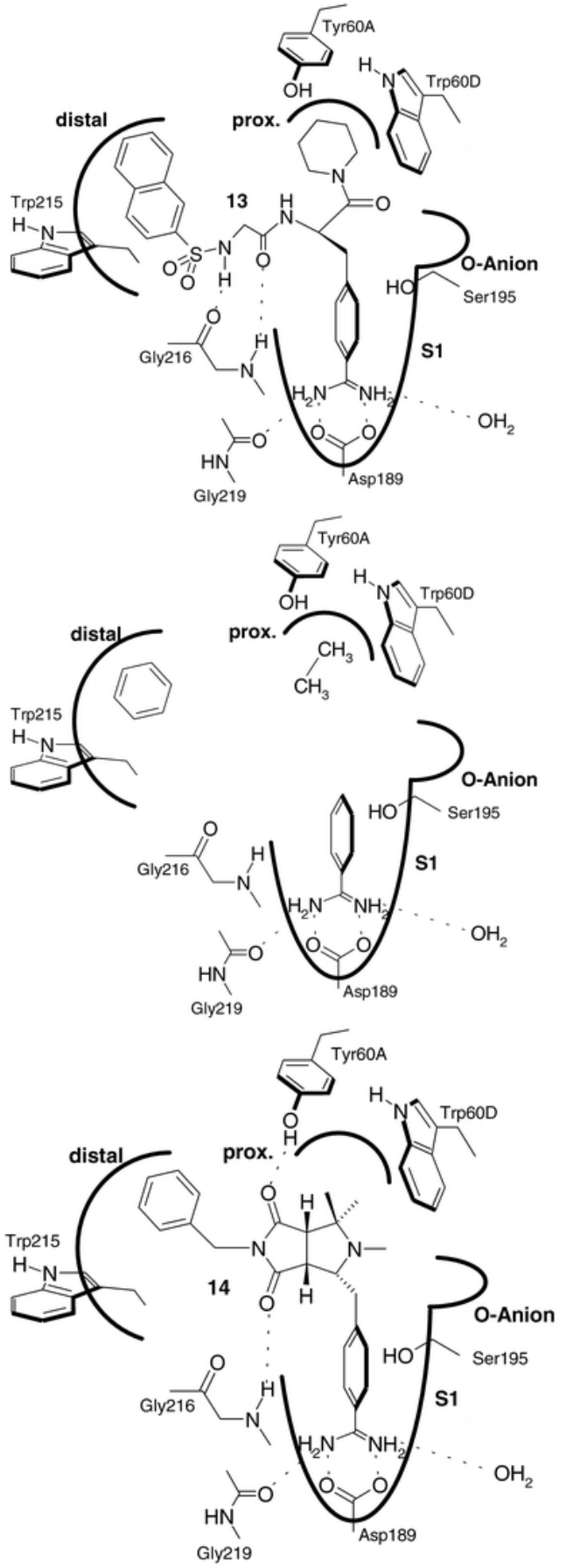

Tyr60A
Trp60D
distal
prox.
Trp215
13
O-Anion
Ser195
S1
Gly216
Asp189
Gly219
Tyr60A
Trp60D
distal
prox.
Trp215
O-Anion
Ser195
S1
Gly216
Asp189
Gly219
Tyr60A
Trp60D
distal
prox.
Trp215
14
O-Anion
Ser195
S1
Gly216
Asp189
Gly219

Fig. 9. Starting with the known binding geometries (*O-Anion* the oxy-anion hole; see also Figs. 11, 12) of well established high-affinity thrombin inhibitors such as NAPAP (**13**), the minimal binding requirements were reduced to a benzamidine, a phenyl and hydrophobic alkyl portion (*center*). On the base of these molecular portions an entirely new skeleton composed of two *cis*-fused five-membered rings was designed and optimized to nanomolar inhibition

---

Benzamidine, phenyl, and ethyl portions have been selected as putative countergroups in designing a rigid template composed by two *cis*-fused five-membered rings. This skeleton correctly orients these groups in space (Fig. 9, bottom) along with the required stereochemistry. To allow some residual flexibility, the phenyl and benzamidine moieties were connected via CH_2 spacers. The "left" five-membered ring was accomplished by two carbonyl groups well placed for hydrogen bonding to the NH of Gly-216 and the OH of Tyr-60A. A methyl group was added to the "right-handed" ring to mimic the selected ethyl group in the proximal pocket. An additional methyl has been fixed to this ring position to reduce the number of stereogenic centers. The "apical" ring positions were replaced by nitrogens for the same reason and to ease synthetic access (Fig. 9). Actually the central bicycle can be synthesized via a 1,3-dipolar cyclo-addition between an azomethine ylid and a N-substituted maleinimide (Fig. 10).

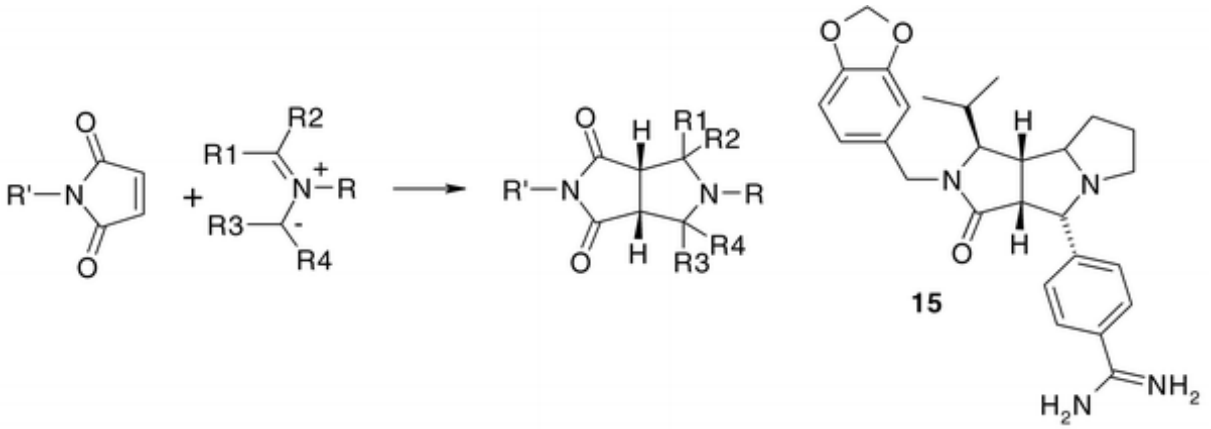


Fig. 10. The new type of thrombin inhibitors (**15**) are accessible via a 1,3-dipolar cyclo-addition between an azomethine ylid and a N-substituted maleinimide

---

Biological testing of the designed lead **14** revealed micromolar affinity to thrombin. However, selectivity with respect to trypsin inhibition was yet not satisfactory. Further optimization of the initial lead **14** considered the stereochemistry of the ligands and followed a thorough analysis of their binding geometries. This optimization led to the production of a 13-nM inhibitor (**15**) which is 760 times more selective for thrombin than trypsin.

Experimental and computational approaches to a better understanding of selectivity-determining factors

As discussed above, thrombin plays a crucial role in the blood-clotting cascade. It is located at the far end of this highly controlled system involving several sequential and feedback-regulated processes. Several enzymes, members of the class of serine proteinases, catalyze the bifurcated cascade of proteolytic reactions. The various serine proteinases involved are structurally related and, several of them have been selected as potential targets for drug development [*43*]. Inhibition at various levels of the cascade leads to alterations in the therapeutic profile. Controlled interference requires highly specific inhibitors being selective for the respective serine proteinases involved. As described in the example above, in addition to the enzymes of the clotting cascade other structurally related serine proteinases can be targeted, for example, trypsin and urokinase, particularly by nonspecific serine protease inhibitors [*44*, *45*].

Drug development requires selective inhibitors, and a thorough understanding of the molecular parameters controlling selectivity is therefore of utmost importance. We approach these important questions by both experimental and computational means. On the experimental side, have we selected trypsin as the parent structure for the described class of serine proteinases. Trypsin has been used as a plausible model in several studies to examine the binding modes of inhibitors optimized for the binding site of other, similar proteinases, such as thrombin and factor Xa [*45*, *46*]. The binding pockets of trypsin and factor Xa differ in two important loops. To better understand selectivity differences these two we have constructed a mutant variant of rat trypsin in the "99-loop," which forms part of the extended binding site (M.T. Stubbs, S. Reyda, F. Dullweber, M. Möller, G. Klebe, K. Stabe, D. Ullmann, H.D. Jakubke, D. Dorsch, W.W.K.R. Mederski, H. Wurziger, submitted). In bovine and rat trypsin the upper part of this pocket is composed of Tyr-172, Trp-215, and Leu-99, while in human factor Xa Leu-99 is replaced by Tyr-99. Furthermore, in factor Xa, a second loop deviates from the situation in trypsin: with Phe-174 replacing Gln-175 in trypsin the factor Xa binding

site is flanked by another aromatic residue, thus providing a ligand recognition site composed solely of aromatic groups. Both bovine trypsin and the X99rT mutant replacing the 99-loop of rat trypsin (Lys 97-Thr-98-Leu-99) by that of human factor Xa (Gln 97-Thr-98-Tyr 99) have been crystallized together with a potent factor Xa chloronaphthyl piperidinylpyridine inhibitor (**16**) first described by Faull et al. (Fig. 11; A.W. Faull, C.M. Mayo, J. Preston, A. Stocker, patent no. WO 9610022, 1996). In addition to showing a surprising pH-dependent binding mode to bovine trypsin, the ligand induces remarkable changes in the recognition properties at the binding site. Crystals grown at pH 7 show the inhibitor with its pyridine moiety binding toward Asp-189 in the specificity S1 pocket. Surprisingly, the hydrophobic S3 pocket (proximal and distal pocket in thrombin) remains unoccupied, and the inhibitor orients itself towards the S1 site (Fig. 11). Reversed binding is observed in a second crystal form obtained at pH 8. In this form the inhibitor binds to bovine trypsin with its hydrophobic chloronaphthyl group towards the S1 pocket which is restricted at its far end by Asp-189. Therefore a basic moiety is not - as usually assumed - an absolute requirement for the inhibition of trypsinlike proteinases possessing an acidic amino acid in the specificity pocket. The piperidinylpyridyl portion accommodates the hydrophobic pocket formed by Tyr-172, Trp-215, and Leu-99 (Fig. 11). Similarly, in the X99rT mutant the chloronaphthyl moiety is found in the S1 pocket, and the piperidinylpyridyl part is located in the aromatic pocket formed by Y172, W215, and Y99. In contrast to the human factor Xa, the "intermediate X99rT" mutant misses the proper aromatic boundary at one of the rims (represented by Phe-174 in human factor Xa). Interestingly, the active site in the rat mutant undergoes a rearrangement compared to the benzamidine-inhibited complex. The side chain of Tyr-217 is flipped toward this aromatic recognition pocket involving an additional reorientation of the 217-219 peptide bond in the mutant. As a result the new position of the side chain of Tyr-217 mimics nicely the position of Phe-174 in human factor Xa, thus almost isostructurally wrapping around the piperidinyl pyridine moiety of the well known potent factor Xa inhibitor (Fig. 12). The observed binding mode and adaptions of the binding site residues provide some ideas regarding the way in which molecular recognition and thus selectivity mechanisms operate in protein-ligand binding.

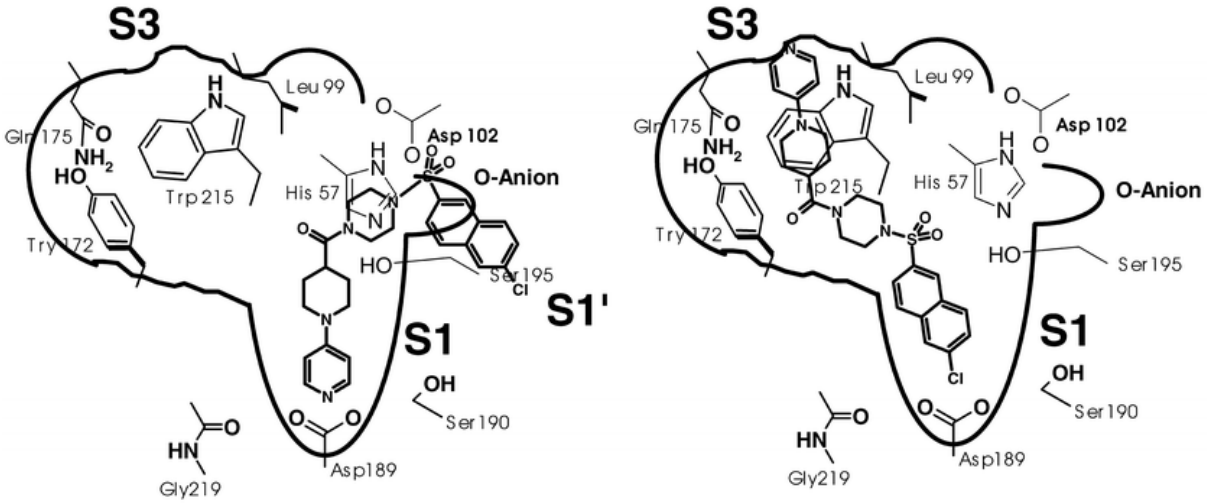


Fig. 11. Two crystal forms have been obtained for trypsin together with **16**, a potent inhibitor for human factor Xa. Surprisingly, the inhibitor adopts different binding modes. In one case (*left*) it binds with its amino pyridine moiety into the S1 pocket forming a hydrogen bond to Asp-189. The hydrophobic S3 binding pocket remains unoccupied and the inhibitor folds over into the S1 pocket. In the second crystal form (*right*) the ligand binds with its chloronaphthyl moiety into the S1 pocket and the piperidinylpyridyl part locates in the S3 pocket

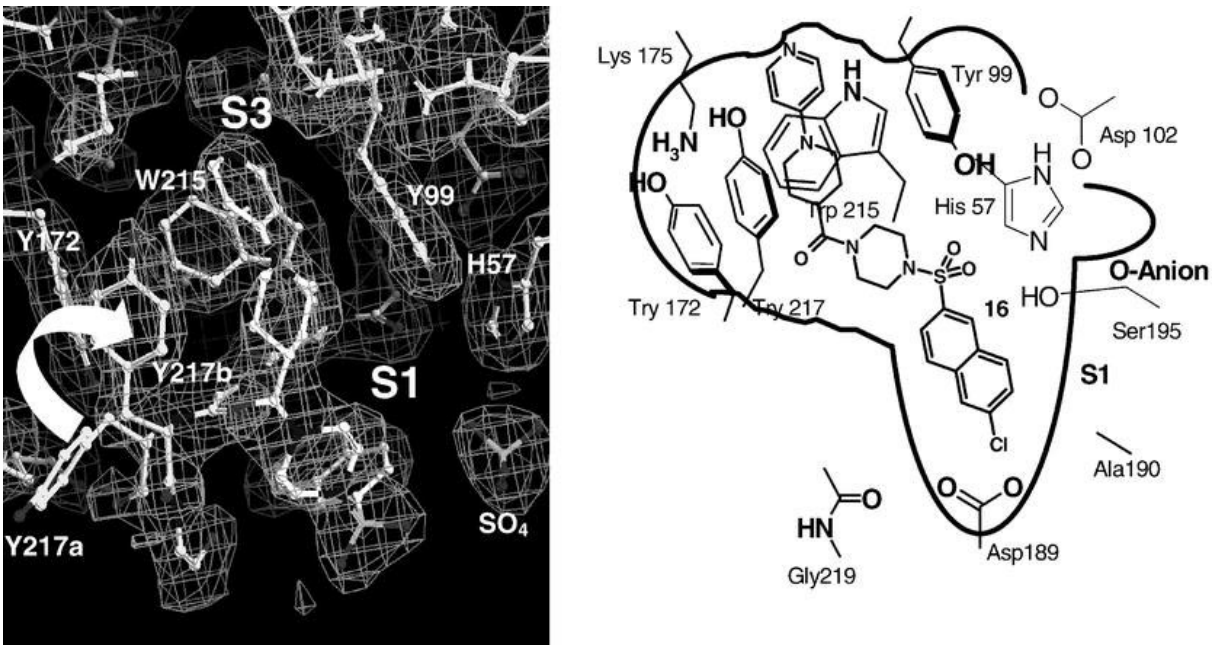


Fig. 12. The crystal structure of inhibitor 16 together with the X99rT mutant of rat trypsin shows again the chloronaphthyl portion to be located in the S1 specificity pocket (*right* schematic sketch of the binding mode; *left* experimentally determined electron density). The piperidinylpyridyl moiety accommodates the hydrophobic S3 pocket flanked by three aromatic residues. Compared to human factor Xa on the "left hand side" of the S3 pocket a proper aromatic boundary is not present in the mutant. However, the side chain of Tyr-217 undergoes an interesting conformational change (from position a adopted in the benzamidine-inhibited complex to position *b* in the present structure) and flips over toward the S3 pocket to complete the aromatic recognition pocket

---

From a computational point of view we try to extract selectivity-determining features by means of comparative molecular field analyses [*47*]. This three-dimensional QSAR method [*48*] quantifies and ranks the putative interaction properties that the various ligands can experience at different binding sites. The results of such an analysis can be mapped back onto the molecular structures, and graphic tools indicate the molecular properties that give rise to particular trends in the biological data [*49*].

A prerequisite for comparative molecular field analysis is a set of ligands binding with distinct affinities to a set for proteins. For the present case study we selected a set of 72 structurally modified 3-amidinophenylalanines (**17**) for which the binding constants (K_i) toward thrombin, trypsin, and factor Xa have been determined (Fig. 13) [*50*]. All ligands have been docked into the binding sites of the three related serine proteinases and a superposition model of the ligands derived. In each case the crystal structure of the three enzymes was taken as reference [*51*, 52, 53].

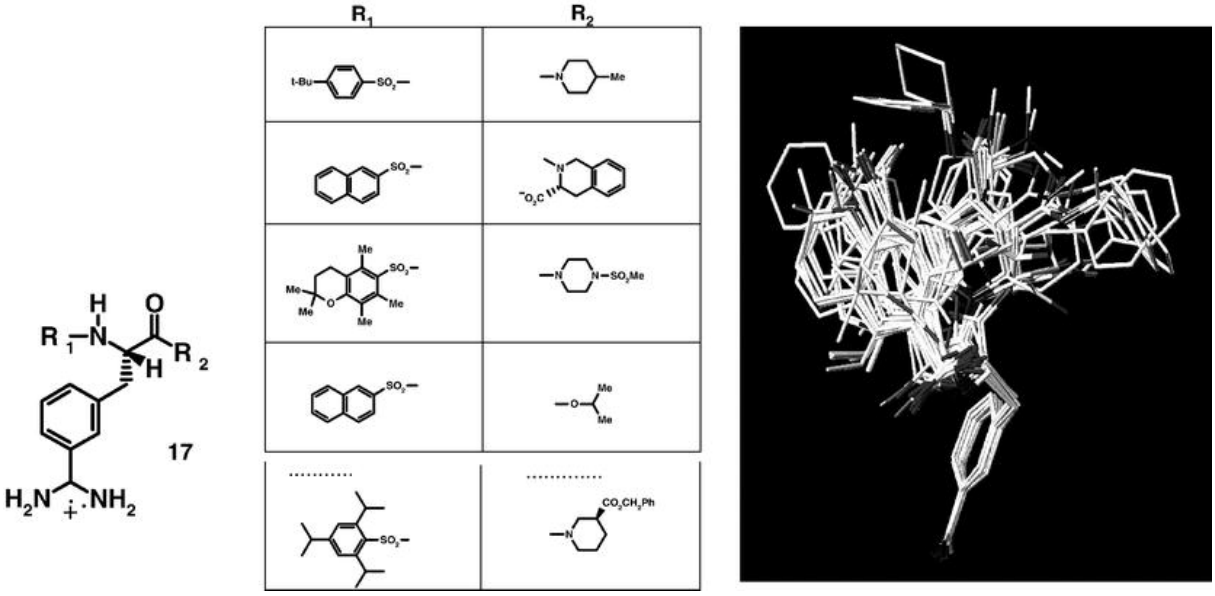


Fig. 13. Basic skeleton of differently substituted 3-amidinophenylalanines **17** for which the binding constants have been determined to thrombin, trypsin, and factor Xa and correlations calculated in a comparative molecular field analysis. The chemical formulas of the substituents R1 and R2 are given in the scheme for five of the considered inhibitors (for further details see [*50*]). *Right* the mutual superposition of all 72 inhibitors used in the analysis

---

A comparative molecular field method developed by us was applied [*54*]. This determines molecular similarities among the ligands considering physicochemical properties in space. Using a common probe atom, molecular similarity indices are calculated for each molecule of the data set at regularly spaced grid points of a virtual lattice in which all molecules of the data set are embedded. The computed grid values are correlated and compared with respect to the differences in binding affinities. First, this approach reveals a relationship that allows the affinity of a particular ligand to be estimated from the actual grid values. This is very convenient in predicting the affinity of a newly designed inhibitor. However, more important with respect to molecular design, grid areas responsible for the trends in the affinity data can be highlighted graphically. This allows features to be specified and mapped that are directly responsible for affinity and selectivity differences between the superimposed ligands and with respect to the various proteins. Figure 14 presents the steric properties derived from the factor Xa affinity data [*50*]. The areas contoured in white correspond to regions in which steric occupancy with bulky groups increases affinity. Areas set off by magenta isopleths should be sterically avoided, otherwise reduced affinity can be expected. A favorable region is indicated in the distal pocket; the magenta contour on the right, next to the catalytic center, points to a sterically unfavorable area. The two molecules shown together with this map occupy the two regions differently. The less active 18 orients its methyl ester group into the disfavored region whereas the more active 19 fills the white-contoured area by its *p*-isopropyl substituent. The steric maps for the other two proteins, trypsin and thrombin, reveal deviating contouring. This finding shows that steric properties of the ligands contribute in distinctive ways to affinity in the respective enzymes. Other physicochemical features, such as electrostatic, hydrophobic, and hydrogen bonding properties can be evaluated and interpreted by similar contour maps. These demonstrate the properties that are of importance and at which locations in space. They furtherfore suggest ways in which to further optimize a particular lead structure.

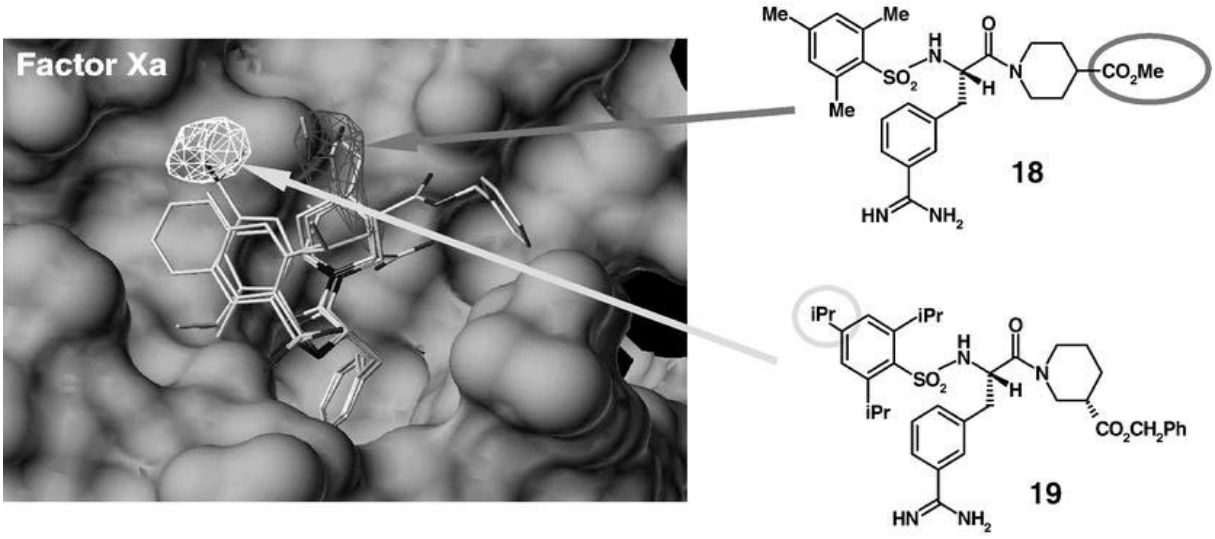
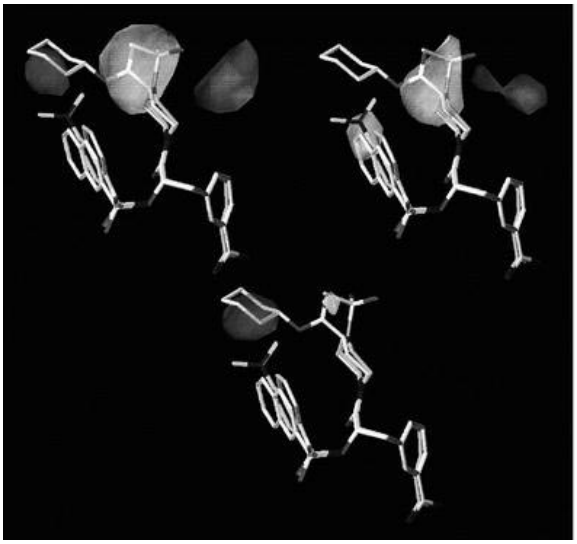
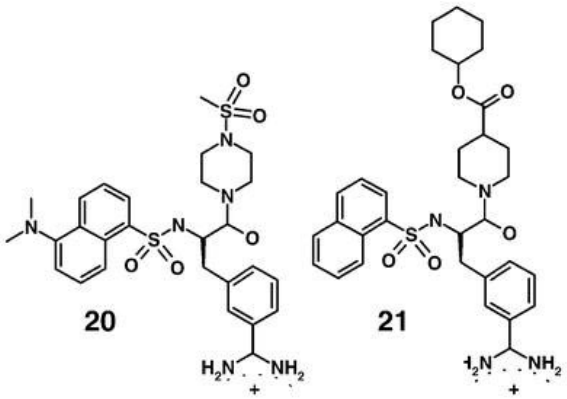


Fig. 14. Contribution map of steric properties for factor Xa. Steric occupancy of the *white-contoured region* increases affinity whereas the *gray-contoured area* should be sterically avoided. The weak binding inhibitor **18** (pK_i=4.1) places its COOMe group in the latter unfavorable region whereas the more potent **19** (pK_i=6.1) occupies the favorable area with its isopropyl group

---

A similar analysis was been performed to elucidate selectivity-discriminating features; however, instead of examining the binding affinities toward a single protein, the affinity differences in binding toward two proteins were evaluated, and the correlations between were calculated [*50*]. In the data set considered here, the binding constants of individual inhibitors differ by more than three orders of magnitude with respect to trypsin and thrombin inhibition. A statistically significant correlation model was obtained. The graphic interpretation, for example, of the steric selectivity map, shows one pronounced area. This clearly indicates that leaving this particular area sterically unoccupied enhances binding towards thrombin with respect to trypsin. Figure 15 shows two inhibitors together with this area. The inhibitor **20** possesses higher affinity toward thrombin and leaves the indicated area unoccupied. The inhibitor **21** with higher affinity toward trypsin places its terminal cyclohexyl moiety into this affinity-discriminating area. Consulting the related maps of the other physicochemical properties further helps to extract molecular parameters responsible for selectivity. It is interesting to note that the highlighted areas fall next to the 60-loop in thrombin. This loop occurs as a special characteristic in thrombin compared to other trypsinlike serine proteinases, and it is therefore reasonable that areas in which affinity is discriminated between thrombin and trypsin fall close to this 60-loop. The contour diagrams derived from comparative molecular field analysis help to elucidate molecular characteristics in space responsible for the affinity and selectivity determining features of small molecule ligands.





	20	21
trypsin:	6.77	7.10
thrombin:	8.38	5.58

Fig. 15. Steric contribution maps for thrombin (*upper left*), trypsin (*upper right*), and the selectivity discriminating map (*center below*). Steric occupancy of the *dark-gray contoured area* in the latter map indicates decreasing affinity towards thrombin. Inhibitor **21** with higher affinity (listed in this figure as pK_i=-log K_i) to trypsin than to thrombin places its terminal cyclohexyl moiety into this area. In contrast, **20** with stronger affinity to thrombin leaves this area unoccupied

Conclusions and summary

This review selects several examples from the literature and from our own research to demonstrate the way in which an iterative process of computer design, synthesis, testing, structure elucidation, and data analysis can aid in developing potent inhibitors. An important aspect is the discovery of a first lead that is structurally different from other, already known examples. Based on a binding site derived pharmacophore or a pattern of putative interaction sites elaborate computer screenings of structural databases can be performed to reveal initial hits for putative lead structures. However, if a large structural variety of ligands is already known, thorough analyses of structural binding site features complemented by a creative chemical and structural thinking can be used to design new ligands. Therapeutic strategies generally require inhibitors that are highly selective for a particular target. This objective can be achieved only by means of potent and highly selective ligands. However, the molecular features driving selectivity remain only little understood. Detailed structural studies of gradually modified ligands and binding sites provide a possibility of elucidating some of the controlling factors. Computational tools have been and are are continuing to be developed to extract molecular parameters from the large body of ligand binding data responsible for affinity discrimination toward structurally related proteins.

Acknowledgements. The author is grateful for a very fruitful and stimulating collaboration with all members of his research group and related groups in Marburg. In particular I thank Ulrich Grädler, Klaus Reuter, Ralf Ficner (TGT inhibitors), Milton Stubbs, Sabine Reyda (trypsin mutants), and Markus Böhm (field analysis). The financial support of the Deutsche Forschungsgemeinschaft, the German Ministry of Science and Technology, and Merck (Darmstadt, Germany) are gratefully acknowledged.

References

1Böhm H J, Klebe G (1996) What can we learn from molecular recognition in protein-ligand complexes for the design of new drugs? Angew Chem Int Ed Engl 35:2588-2614

2.Babine RE, Bender SL (1997) Molecular recognition of protein-ligand complexes: applications to drug design. Chem Rev 97:1359-1472

3.Müller K (1995) De novo design, vol 3. In: Anderson PS, Kenyon GL, Marshall GR (eds) Persp drug discovery and design. ESCOM, Leiden

4.Kubinyi H (1998) Structure-based design of enzyme inhibitors and receptor ligands. Curr Opin Drug Discov Dev 1:4-15

5.Van Drie JH, Lajiness MS (1998) Approaches to virtual library design. Drug Discov Today 3:274-283

6.Walters WP, Stahl MT, Murcko MA (1998) Virtual screening - an overview. Drug Discov Today 3:160-178

7.Böhm HJ, Klebe G, Kubinyi H (1996) Wirkstoffdesign. Spektrum, Heidelberg

8.Hol WGJ (1986) Protein crystallography and computer graphics - toward rational drug design. Angew Chem Int Ed Engl 25:767-778

9.Walkinshaw MD (1992) Protein targets for structure-based drug design. Med Res Rev 12:317-372

10.Bedell CR, Goodford PJ, Norrington FE (1976) Compounds designed to fit a site of known structure in human hemoglobin. Br J Pharmacol 57:201-209

11.Goodford P (1984) Drug design by the method of receptor fit. J Med Chem 27:557-564

12.Cushman DW, Cheung HS, Sabo EF, Ondetti MA (1977) Design of potent competitive inhibitors of angiotensin-converting enzyme. Carboxyalkanoyl and mercaptoalkanoyl amino acids. Biochemistry 16:5484-5491

13.Lipsbomb WN, Hartsuch JA, Reehe GN, Quiocho FA, Bethge PH, Ludwig ML, Steitz TA, Muirhead H, Coppola JC (1968) The structure of carboxypeptidase A. VII. The 2 Å resolution studies of the enzyme and of its complex with glycyltyrosine, and mechanistic deductions. Brookh Symp Biol 1:24-90

14.Navia MA, Murcko MA (1992) Use of structural information in drug design. Curr Opin Struct Biol 2:202-210

15.West ML, Fairlie DP (1995) A test for drug-design methodologies. Trends Pharmacol Sci 16:67-74

16.Lam PYS, Jadhav PK, Eyermann CJ, et al (1994) Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. Science 263:380-384

17.Allen FA, Kennard O, Taylor R (1983) Systematic analysis of structural data as a research technique in organic chemistry. Accounts Chem Res 16:146-153

18.Okada N, Nishimura S (1979) Isolation and characterization of a guanine insertion enzyme, a specific tRNA transglycosylase, from *Escherichia coli*. J Biol Chem 254:3061-3066

19.Durand JM, Okada N, Tobe T, Watarai M, Fukuda I, Suzuki T, Nakata N, Komatsu K, Yoshikawa M, Sasakawa C (1994) VacC, a virulence-associated chromosomal locus of *Shigella flexneri*, is homologous to tgt, a gene encoding tRNA-guanine transglycosylase (Tgt) of *Escherichia coli* K-12. J Bacteriol 176:4627-4634

20.Persson BC (1993) Modification of tRNA as a regulatory device. Mol Microbiol 8:1011-1016

21.Rohde JE (1984) Selective primary health care: strategies for control of disease in the developing world. XV. Acute diarrhea. Brookh Dis 6:840

22.Romier C, Reuter K, Suck D, Ficner R (1996) Crystal structure of tRNA-guanine transglycosylase from *Zymononas mobilis*: RNA modification by base exchange. EMBO J 15:2850-2857

23.Böhm HJ (1992) The computer program Ludi: a new method for the de novo design of enzyme inhibitors. J Comput Aided Mol Des 6:61-78

24.Böhm HJ (1992) Ludi: rule-based automatic design of new substituents for enzyme inhibitor leads. J Comput Aided Mol Des 6:593-606

25.Klebe G (1994) The use of composite crystal-field environments in molecular recognition and the de novo design of protein ligands. J Mol Biol 237:212-235

26.Majeux N, Scarsi M, Apostolakis J, Erhardt C, Caflish A (1999) Exhaustive docking of molecular fragments on protein binding sites with electrostatic solvation. Proteins Struct Funct Genetics 27:88-105

27.Zhou X, Sun Y, Kuntz ID (1999) Inclusion of solvation in ligand binding free energy calculations using the generalized-Born model. J Am Chem Soc 121:8033-8043

28.Böhm HJ (1994) 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 Des 8:243-256

29.Mügge I, Martin YC (1999) A general and fast scoring function for protein-ligand interactions: a simplified potential approach. J Med Chem 42:791-804

30.Mitchell JBO, Laskowski RA, Alex A, Thornton JM (1999) BLEEP-potential of mean force describing protein-ligand interactions. I. Generating potential. J Comp Chem 20:1156-1176

31.Mitchell JBO, Laskowski RA, Alex A, Forster MJ, Thornton JM (1999) BLEEP-potential of mean force describing protein-ligand interactions. II. Calculation of binding energies and comparison with experimental data. J Comp Chem 20:1177-1185

32.Gohlke H, Hendlich M, Klebe G (2000) Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol 295:337-356

33.Grädler U, Gerber HD, Goodenough-Lashua DM, Garcia GA, Ficner R, Reuter K, Stubbs MT, Klebe G (19[redacted]) A new target for shigellosis: rational design and crystallogarphic studies of inhibitors of tRNA-guanine transglycosylase [redacted]:[redacted]-[redacted]

34.Montfort WR, Perry KM, Fauman EB, Finer-Moore JS, Maley GF, Hardy F, Marley F, Stroud RM (1990) Structure, multiple site binding, and segmental accommodation in thymidylate synthase on binding UMP and anti-folate. Biochemistry 29:6964-6977

35.Reich SH, Fuhry MA, Nguyen D, et al (1992) Design and synthesis of novel 6,7-imidazotetrahydroquinoline inhibitors of thymidylate synthase using iterative protein crystal structure analysis. J Med Chem 35:847-858

36.Reich SH, Webber SE (1993) Structure-based drug design (SBDD): every structure tells a story. Perspect Drug Discov Des 1:371-390

37.Webber SE, Bleckman TM, Attard J, et al (1993) Design of thymidylate synthase inhibitors using protein crystal structures: the synthesis and biological evaluation of a novel class of 5-substituted quinazolinones. J Med Chem 36:733-746

38.Jones TR, Varney MD, Webber SE, et al (1996) Structure-based design of lipophilic quinazoline inhibitors of thymidylate synthase. J Med Chem 39:904-917

39.Goodford P (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J Med Chem 28:849-857

40.Davie EW, Fujikawa K, Kisiel W (1991) The coagulation cascade: initiation, maintenance and regulation. Biochemistry 30:10363-10370

41.Obst U, Banner DW Weber L, Diederich F (1997) Molecular recogniton at the thrombin active site: structure-based design and synthesis of potent and selective thrombin inhibitors and the X-ray crystal structure of two thrombin-inhibitor complexes. Chem Biol 4:287-295

42.Obst U, Gramlich V, Diederich F, Weber L, Banner DW (1995) Design of novel, nonpeptidic thrombin inhibitors and structure of a thrombin-inhibitor complex. Angew Chem Int Ed Engl 34:1739-1742

43.Bode W, Brandstetter H, Mather T, Stubbs MT (1997) Comparative analysis of haemostatic proteinases: structural aspects of thrombin, factor Xa, factor IXa and protein C. Thromb Haemost 78 501:511

44.Stubbs MT, Huber R, Bode W (1995) Crystal structures of factor Xa specific inhibitors in complex with trypsin: structural grounds for inhibition of factor Xa and selectivity against thrombin. FEBS Lett 375:103-107

45.Renatus M, Bode W, Huber R, Stürzebecher J, Stubbs MT (1998) Structural and functional analyses of benzamidine-based inhibitors in complex with trypsin: Implications for the inhibition od factor Xa, tPA and urokinase. J Med Chem 41 5445:5456

46.Stubbs MT (1996) Structural aspects of factor Xa inhibition. Curr Pharm Des 2:543-552

47.Cramer RD III, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110:5959-5967

48.Cramer RD III, DePriest SA, Patterson DE, Hecht P (1993) The development of comparative molecular filed analysis. In: Kubinyi H (ed) 3D QSAR. ESCOM, Leiden, pp 583-618

49.Klebe G (1998) Comparative molecular similarity indices analysis: CoMSIA. Perspect Drug Discov Des 12:87-104

50.Böhm M, Stürzebecher J, Klebe G (1999) Three-dimensional quantitative structure-activity relationship analysis using comparative molecular field analysis and comparative similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. J Med Chem 42:458-477

51.Turk D, Stürzebecher J, Bode W (1991) Geometry of binding of the Nα -tosylated piperidides of *m*-amidino-, *p*-amidino- and *p*-guanidino phenylalanine to thrombin and trypsin. X-ray crystal structures of their trypsin complexes and modeling of their thrombin complexes. FEBS Lett 287:133-138

52.Brandstetter H, Turk D, Höffken HW, Grosse D, Stürzebecher J, Martin PD, Edwards BF, Bode W (1992) Refined 2.3 Å X-ray crystal structure of bovine thrombin complexes formed with the benzamidine and arginine-based thrombin inhibitors NAPAP, 4-TAPAP and MQPA. A starting point for improving antithrombotics. J Mol Biol 226:10851099

53.Padmanabhan K, Padmanabhan KP, Tulinsky A, Park CH, Bode W, Huber R, Blankenship DT, Cardin AD, Kisiel W (1993) Structure of human des (1-45) factor Xa at 2.2 Å resolution. J Mol Biol 232:947-966

54.Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J Med Chem 37:4130-4146