Thrombin inhibitors identified by computer-assisted multiparameter design

Daniel Riester*†, Frank Wirsching*, Gabriela Salinas*†, Martina Keller†, Michael Gebinoga†, Stefan Kamphausen†, Christian Merkwirth†, Ruediger Goetz†, Martin Wiesenfeldt‡, Jörg Stürzebecher§, Wolfram Bode¶, Rainer Friedrich¶, Marcel Thürk‡‖, and Andreas Schwienhorst*‖

*Abteilung für Molekulare Genetik und Präparative Molekularbiologie, Institut für Mikrobiologie und Genetik, Grisebachstrasse 8, 37077 Göttingen, Germany; †Novel Science International GmbH, Luisenstrasse 10-11, 30159 Hannover, Germany; ‡Matrix Advanced Solutions Ltd., 15 Berkley Street, London W1J 8DY, United Kingdom; §Zentrum für Vaskuläre Biologie und Medizin, Friedrich-Schiller-Universität Jena, Nordhäuser Strasse 78, 99089 Erfurt, Germany; and ¶Abteilung Strukturforschung, Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, 82152 Martinsried, Germany

Communicated by Manfred Eigen, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany, March 10, 2005 (received for review October 13, 2004)

Here, we present a series of thrombin inhibitors that were generated by using powerful computer-assisted multiparameter optimization process. The process was organized in design cycles, starting with a set of randomly chosen molecules. Each cycle combined combinatorial synthesis, multiparameter characterization of compounds in a variety of bioassays, and algorithmic processing of the data to devise a set of compounds to be synthesized in the next cycle. The identified lead compounds exhibited thrombin inhibitory constants in the lower nanomolar range. They are by far the most selective synthetic thrombin inhibitors, with selectivities of >100,000-fold toward other proteases such as Factor Xa, Factor XIIa, urokinase, plasmin, and Plasma kallikrein. Furthermore, these compounds exhibit a favorable profile, comprising nontoxicity, high metabolic stability, low serum protein binding, good solubility, high anticoagulant activity, and a slow and exclusively renal elimination from the circulation in a rat model. Finally, x-ray crystallographic analysis of a thrombin–inhibitor complex revealed a binding mode with a neutral moiety in the S1 pocket of thrombin.

crystallographic structure | drug design | early adsorption | toxicity | genetic algorithm

Thromboembolic diseases, such as myocardial infarction, stroke, and deep vein thrombosis, are a leading cause of death, particularly in the Western world. The past decade has seen major progress in the development of antithrombotic agents that are tailored to (*i*) exhibit antiplatelet activity, (*ii*) aid in the lysis of blood clots, or (*iii*) affect the activity and generation of thrombin. The latter thrombotic agent is a serine protease, which exerts its effects in the final steps of the blood coagulation cascade and in the activation of various cell types, including platelets. Despite great research efforts and the resulting impressive number of high-affinity thrombin inhibitors identified, there is still only limited clinical use of some parenterally available preparations of thrombin inhibitors. The reasons that render many inhibitors unattractive for clinical applications are manifold. A considerable number of thrombin inhibitors developed so far suffer from poor selectivity, inherent toxicity, high-plasma protein binding, poor metabolic stability, rapid elimination from the blood, low anticoagulant activity, or poor oral bioavailability, to name but a few problems. Because of the complexity of the problem, lead optimization is a tedious task. On the one hand, high-throughput assay systems only coarsely replace a living organism. This finding is particularly true for complex parameters such as bioavailability. On the other hand, the optimization problem is necessarily a multiparameter problem. The ideal functional compromise does not need to coincide with the optima of single parameters. Thus, very often, improvements gained for one parameter are only achieved at the expense of impairments of several others. Nevertheless, through laborious and costly medicinal chemistry, along with high-throughput screening programs, significant progress has been achieved in recent years. For some inhibitor candidates, efficacy in *in vitro* coagulation assays, and even a moderate oral availability, has now been reported (see ref. 1 for review). This finding is particularly true for compounds with basic group bioisosteres incorporated to bind to the acidic S1 pocket of thrombin. However, the ideal thrombin inhibitor has yet to be discovered.

We now present a series of lead molecules that has been obtained by an accelerated type of computer-assisted drug discovery (CADDIS) approach (2). The original procedure has now been further improved to serve as a powerful tool in the simultaneous optimization of a multitude of pharmaceutically relevant properties. One major advantage of our approach is that optimization does not use or depend on *a priori* information about the target, e.g., structural data on thrombin. Instead, it relies exclusively on data acquired within the cyclic process of optimization. Concerning the optimization of thrombin inhibitors, each cycle comprised synthesis of drug candidates, characterization of these compounds in a number of bioassays, and algorithmic processing of the data to generate or improve a predictive model that links compound structure and molecular properties, and which is used to design a new set of compounds. Starting from a set of randomly chosen compounds, a multiparameter optimization was carried out in eight design cycles. Synthesis and experimental fitness determination of a total of only ≈1,000 different compounds was necessary to yield lead compounds show extremely favorable properties concerning thrombin inhibition, selectivity, metabolic stability, (low) serum protein binding, (low) toxicity, activity in *in vitro* coagulation assays, and (slow) elimination from the bloodstream, making them attractive candidates for further development.

Materials and Methods

Peptide Synthesis. Peptides were synthesized by using an automated peptide synthesizer (MultiPep, Intavis, Koln, Germany) following the Fmoc-/tBu standard peptide synthesis protocol suggested by the supplier.

Assays. Peptides were tested for their thrombin inhibitory activity by using a fluorogenic assay as described (3) with the use of human thrombin (Sigma-Aldrich, Deisenhofen, Germany). The trypsin inhibitory activity was obtained in a similar way by using Tos-Gly-Pro-Arg-MCA and bovine trypsin (Sigma-Aldrich) as substrate. Assays were carried out with the assistance of a robotic workstation (CyBi-screen-machine, CyBio, Jena, Germany) that included a Polarstar fluorescence reader (BMG Labtechnologies, Offenburg,

Abbreviations: CADDIS, computer-assisted drug discovery; NAPAP, Na-(β-naphthylsulfonyl-glycyl)-4-amidinophenylalanine piperidide; Cha, cyclohexylalanine.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 1XM1).

‖To whom correspondence may be addressed. E-mail: aschwie1@gwdg.de or thuerk@matrix-as.com (computation).

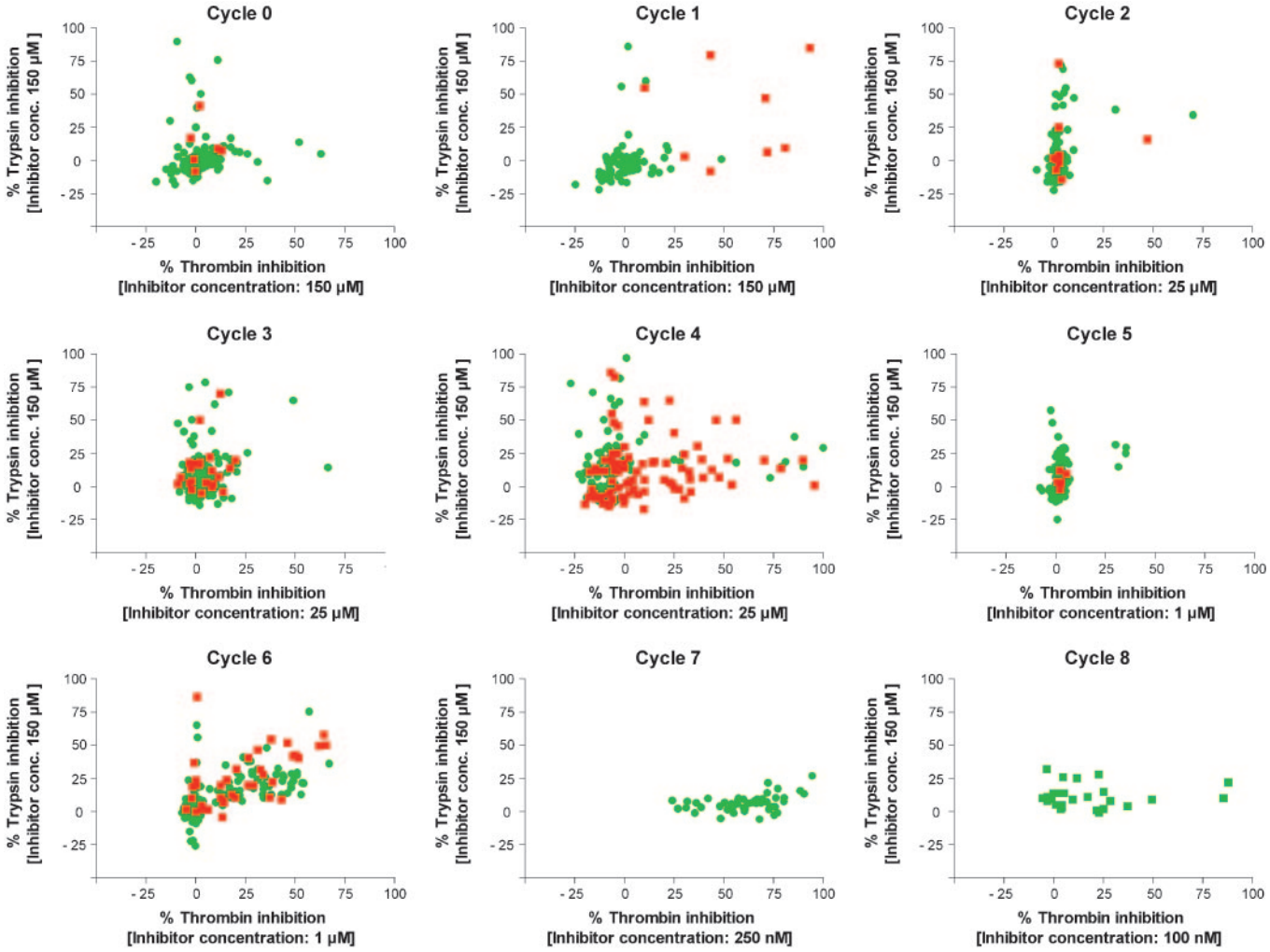

Fig. 1. Evolution of thrombin and trypsin inhibitory activities, toxicity, and hemolytic activity as obtained for compounds originating from design cycles 0–8. Compounds with toxicities or hemolytic activities of >40% at compound concentrations of 200 μM are red. Trypsin and thrombin inhibition were measured by using standard amidolytic assays. In trypsin assays, a final inhibitor concentration of 150 μM was used. Thrombin assays contained a final inhibitor concentration between 150 μM and 250 nM as indicated. Occasionally occurring negative values for inhibition refer to compounds enhancing enzyme activity in the assay.

Germany). Assay data are mean values obtained from at least three independent measurements. Test compound concentrations were as indicated. Standard methods were applied to determine *in vitro* coagulation, serum protein binding, metabolic stability, and cytotoxicity.

Computational Methods. The CADDIS approach, which is based on a highly efficient type of genetic algorithm, has been described in great detail elsewhere (2). Previously, CADDIS has been applied only to single-parameter optimizations such as the generation of thrombin inhibitory peptides. In the latter case, "fitness" of a compound was defined by a single parameter. However, for the multiparameter optimization process described herein, the overall fitness was calculated as the geometrical mean of single fitness values f_i of all parameters i without different weighting. Values for each individual parameter varied between 0% and 100% for the desired function, e.g., 100% for a complete thrombin inhibition or no trypsin inhibition at a given compound concentration. To primarily obtain smaller molecules, the fitness function includes a penalty for large sequences as described (2).

Crystallography. Human α-thrombin was crystallized by using standard methods (4). Details of the structure determination will be described in an accompanying publication. The atomic coordinates have been deposited with Protein Data Bank (PDB), ID code 1XM1.

For detailed protocols and the list of building blocks, refer to *Supporting Materials and Methods* and Table 3, which are published as supporting information on the PNAS web site.

Results

Computer-Assisted Molecular Optimization. Previously, our CADDIS approach had proved to be well suited for the generation of thrombin inhibitory peptides in a single-parameter optimization process (2). Here, we apply the same concept to the problem of multiparameter optimization. In this case, the fitness function that was used to evaluate each molecule uses a geometric mean criterion, including the individual fitness of all parameters without different weighting.

The overall process was organized in design cycles. To start with, a set of 170 randomly chosen peptides with sequence lengths between 3 and 10 amino acids was generated by using solid-phase automated peptide synthesis. Rather than betting on a small set of "promising" building blocks, 66 building blocks that were as chemically diverse as possible were used. Synthesized compounds were characterized with respect to inhibition of thrombin and trypsin,

Table 1. Properties of selected thrombin inhibitors

Values	Selected inhibitors from cycles 7–8					Reference inhibitors		
	7-1	7-4	7-8	8-1	8-5	Argatroban	NAPAP	Melagatran
K_i [μM]								
Human thrombin	0.080	0.009	0.056	0.008	0.003	0.038	0.009	0.006
Bovine thrombin	0.150	0.021	0.073	0.013	0.0072	0.019	0.006	0.0036
Trypsin	>1,000	>1,000	>1,000	>1,000	>1,000	4.250	690	0.004
Factor Xa	>1,000	>1,000	>1,000	>1,000	>1,000	210	7.9	9.4
Factor XIIa	—*	>1,000	>1,000	>1,000	>1,000	>1,000	450	10.4
Urokinase	>1,000	>1,000	>1,000	>1,000	>1,000	>1,000	230	7.9
Plasmin	>1,000	>1,000	>1,000	>1,000	>1,000	600	30	1.4
Plasma kallikrein	—	>1,000	>1,000	>1,000	>1,000	>1,000	14	0.69
IC_{200}, μM								
Thrombin time	0.24	0.034	0.24	0.040	0.040	0.062	0.045	0.015
aPTT	6.0	1.0	2.8	0.95	0.60	0.42	0.50	0.24
Prothrombin time	14	2.0	—	2.0	1.45	0.66	1.0	0.37
Toxicity [%]								
HeLa (30 μM)	0	0	0	1.7	0	—	—	—
HeLa (200 μM)	17.1	—	0	5.0	2.9	—	13	12
Hemolysis	10	—	<2	<2	<2	—	—	—
ADME								
SPB, %	59	39	2	22	29	—	86	7
Metabolic Stab, %	88	80	92	70	76	—	97	93
Solubility, 200 μM	197.9	—	200	200	200	—	—	—

ADME, absorption, distribution, metabolism, elimination; SPB, serum protein binding; P-IC_{200}, inhibitor concentration doubling the respective clotting time; TT, thrombin time; aPTT, activated partial thromboplastin time.
*Not measured.

serum protein binding, metabolic stability, toxicity, hemolysis, and solubility (cycle 0). After algorithmic processing of the data, a set of 96 compounds was determined by the algorithm and subsequently synthesized. The compounds were again characterized in the aforementioned assays (cycle 1), and the data were used to devise a new set of molecules. This procedure was repeated afterward with 96–170 compounds synthesized and characterized per cycle. In the final cycle (cycle 8), only 26 compounds were synthesized and tested.

In Fig. 1, the results concerning thrombin and trypsin inhibitory activities, toxicity, and hemolysis are depicted for compounds originating from the design cycles 0–8. Each subsequent cycle revealed molecules with increased thrombin inhibitory activity. In cycle 0, the most active molecules exhibited thrombin inhibitory activities with K_i values $\gg 1$ μM (data not shown). However, in cycles 7 and 8, the best inhibitors showed K_i values in the lower nanomolar range and an excellent specificity, i.e., negligible inhibitory activities toward trypsin (see also Table 1). In cycle 0 and throughout most subsequent cycles, most compounds did not exhibit pronounced toxic or hemolytic activity. However, an exception was in cycle 4, where half of the compounds tested showed a significant toxic and/or hemolytic effect. Until then, toxic compounds appeared only scarcely; in other words, the training set for the parameter "toxicity" was strongly biased toward nontoxic compounds and thus incomplete! It was not surprising, then, that a small subpopulation of toxic (and related) compounds could arise transiently. However, in the last two cycles (7 and 8), all compounds proved to be nontoxic, probably also due to the more focused character of the corresponding libraries. Except for compounds from cycles 0 and 1, most molecules from subsequent cycles exhibited metabolic stabilities >60%, which correlates to the increased fraction of nonnatural amino acids as building blocks in late cycles. Solubility was never a problem throughout all cycles. This finding is possibly due to the chemical nature of the compounds, i.e., the polarity of the peptide backbone. This polarity could also explain why very strong serum protein binders were only occasionally identified throughout the optimization. From cycle 2 on, a fraction of 50–70% of the molecules displayed a serum protein binding of <60%.

The evolution of optimized molecules in the scope of the computer-assisted drug design process is reflected by changes in the sequence population of subsequent cycles. Peptides from cycles 0 and 1 generated essentially the same sequence distribution, i.e., the same mean pair distance as calculated for a random sample of the concomitant sequence space. In particular, no preferential sequence motif could be traced. In cycle 2, a considerably large fraction (10%) of peptides contained an N-terminal Ac-L-arginine. This fraction increased to 38% in cycle 3 and remained constant until cycle 5 before it decreased again. In cycle 3, for the first time, larger fractions of the population comprised common short sequence motifs such as Ac-L-Arg-L-Trp (12%) and Ac-L-Arg-D-cyclohexylalanine (Cha) (8%) at the N termini of the peptides. The most active molecules among the latter were two relatively small molecules, Ac-L-Arg-D-Cha-D-Phg-D-Tyr-L-Arg and a variant missing the C-terminal arginine. Interestingly, a two-error mutant of NSCI 521, the most active compound in previous single-parameter screenings (2), was also identified among the most active sequences of cycle 3. In cycles 4 and 5, the fraction of molecules containing the N-terminal motif Ac-L-Arg-D-Cha increased to 21% and 36%, respectively. Among the most active peptides were three small molecules, Ac-L-Arg-D-Cha-L-Pro-D-Tyr-L-Arg and two 2-error mutants thereof. In cycle 6, a fraction of 28% of the molecules comprised variants of the latter that contained one mutation, deletion, or insertion. The same sequence was further varied in cycle 7. Here, however, the motif D-Cha-L-Pro-D-Tyr-X (2%) was largely replaced by D-Cha-L-Aze-D-Tyr-X (35%). Twenty-nine percent of all sequences had an L-Arg at the X position, whereas 52% contained an L-homoArg at this position. Among the most active sequences were three pentapeptides, Ac-L-Ala-D-Cha-L-Aze-D-Tyr-L-homoArg (7-1), Ac-L-Arg-D-Cha-L-Aze-D-Tyr-L-homoArg (7-4), and NH_2-L-Arg-D-Cha-L-Aze-D-Tyr-L-homoArg (7-8). The final cycle, cycle 8, focused exclusively on one-error mutants of the latter sequences. The most successful compounds were Trx-D-Cha-L-Aze-D-Tyr-L-homoArg (8-1), and a variant thereof, which was guanylated at the N terminus (8-5). Compounds 7-8, 8-1, and 8-5 are depicted in Fig. 2.

Characterization of Optimized Molecules. The most active thrombin inhibitors from cycles 7 and 8 were thoroughly characterized, e.g.,

EVOLUTION

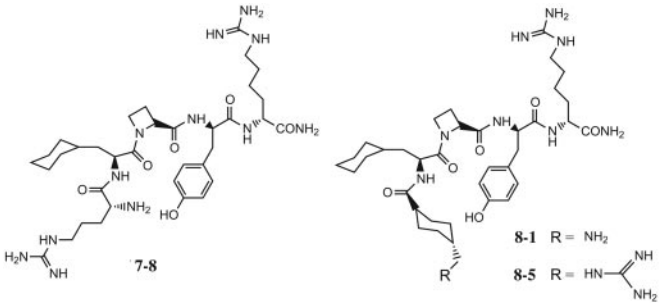

Fig. 2. Structures of thrombin inhibitors 7-8, 8-1, and 8-5 identified in design cycles 7 and 8 of the computer-assisted optimization procedure.

with respect to selectivity and anticoagulant activity (Table 1). Overall, these molecules exhibit excellent thrombin inhibitory activities with K_i values in the lower nanomolar range. For compounds 7-4, 8-1, and 8-5, the K_i values measured with human thrombin were 9, 8, and 3 nM, respectively; this finding is comparable with, or slightly better than, the K_i values of Na-(β-naphthylsulfonyl-glycyl-4-amidinophenylalanine piperidide (NAPAP) (9 nM) and melagatran (6 nM) measured in the same assay.

To study the binding mode of 8–5 within the active site, a number of derivatives of compound 8-5 were synthesized and characterized (Table 2). In general, all tested variants of 8-5 (including 7-1, 7-4, 7-8, and 8-1) showed an increase in their K_i values measured with human thrombin. Like many of the known small-molecule thrombin inhibitors directed against the active site, all compounds identified in the CADDIS process contained basic moieties that could very well occupy the S1 site of thrombin. However, neither of the basic moieties, tranexamic acid and homoarginine, found in the inhibitors, has so far been identified as a building block of selective inhibitors of thrombin or other closely related enzymes involved in the coagulation cascade. Only certain plasma kallikrein-selective inhibitors were reported to contain tranexamic acid (5). Because the tranexamic acid building blocks in 8-1 and 8-5 can be replaced not only by L-arginine (7-4 and 7-8) but also by L-alanine (7-1) without dramatically impairing inhibitory activity, we were inclined to assume that tranexamic acid is not a superior candidate for S1 site occupation. The second basic moiety in the newly generated inhibitors is L-homo-Arg. Here again, replacement by other amino acids including noncharged residues such as L-Met resulted in only moderate increases in the K_i value of 3 to 14-fold, indicating that the side chain of L-homo-Arg is also not the P1 moiety. The small hydrophobic S2 pocket of thrombin is typically filled with a small hydrophobic element such as proline. In the series of inhibitors, L-Aze is the best P2 candidate. Substitution by L-Pro resulted in a 26-fold increase in the K_i value. Larger moieties dramatically impair thrombin inhibitory activity (data not shown). The larger hydrophobic pocket usually accommodates bulkier hydrophobic residues such as D-Phe. The best candidate in the series of inhibitors is D-Cha. Support for this thesis comes from a derivative with substitution of D-Cha by D-Phe, which resulted in a moderate 19-fold increase in the K_i value. Further support comes from the fact that the D-Cha-L-Aze motif, which is common to this class of inhibitors, resembles the D-Chg-L-Aze motif of melagatran. In the crystallographic structure of the thrombin–melagatran complex, D-Cha and L-Aze occupy the S4 and S2 pockets, respectively (6). From this analogy, we assumed that D-Tyr could very well be the P1 moiety. However, only substitution of D-Tyr by the uncharged D-Phe, but not by typical basic P1 moieties such as D/L-Arg (as in Phe-Pro-Arg-chloromethylketone) or D-Phe-4-amidine (containing the benzamidine moiety of melagatran), largely preserves thrombin inhibitory activity. Thus, by adhering to the formal analogy between melagatran and the inhibitor series, one definitely had to assume (*i*) an uncharged moiety to occupy the S1 pocket, and

Table 2. Characterization of compound 8-5 derivatives

No.	Sequence*					Percent inhibition (trypsin)†	K_i, nM (thrombin)
8-5	G-Trx	D-Cha	L-Aze	D-Tyr	L-hArg	15	3
8-5A	G-Met					7	23
8-5B	G-Gly					0	133
8-5C		D-Phe				5	57
8-5D			L-Pro			6	80
8-5E				D-Phe		11	7
8-5F				D-Arg		14	405
8-5G				L-Tyr		18	488
8-5H				L-Arg		4	>1,000
8-5I				D-Ph1		18	483
8-5J				D-Ph2		20	183
8-5K				D-Ph3		26	>1,000
8-5L					Gly	5	42
8-5M					L-Met	0	13
8-5N					L-Lys	8	10
8-5O					AcLys	8	17

*G-Trx, guanylated tranexamic acid; G-Met, guanylated D-Met; G-Gly, guanylated Gly; D-Ph1, D-Phe-4-CN; D-Ph2, D-Phe-4-amidine; D-Ph3, D-Phe-4-hydroxy-amidine; AcLys, L-(ε-N-acetyl)-lysine.

†Percent trypsin inhibition at an inhibitor concentration of 100 μM.

(*ii*) a geometry of the inhibitors bound to the active site of thrombin that is significantly different from that of melagatran. To get things straight concerning this point, we solved the structure of the complex of thrombin with compound 8-5 by x-ray crystallography.

Human α-thrombin was crystallized in the presence of inhibitor benzamidine and inhibitory peptide hirugen 56-63 (hirugen) by using standard methods (4) in space group *C*2. The crystals diffracted to a resolution of 2.3 Å. With the exception of the 149-loop (chymotrypsinogen numbering of thrombin), the main chain of the thrombin molecule was fully defined by electron density, exhibiting a conformation virtually identical with the thrombin–hirugen complex (PDB ID code 1HGT) used for the solution of the structure.

The inhibitor could be almost fully traced, with the exception of a few atoms in side chains pointing toward the bulk solvent with only weak electron density. It was assumed that not all active sites were occupied by inhibitor molecules, and thus, the occupancy of the inhibitor atoms was set to 0.8. With the D-Cha and L-Aze moieties (positions 2 and 3), the inhibitor binds to the S4/S2 pockets, respectively (Fig. 3*A*). Whereas L-Aze moieties of both inhibitors almost ideally can be superimposed, the D-Cha moiety of 8-5 reaches, by 2.4 Å, deeper into the S4 pocket as compared with D-Chg in the melagatran structure. The D-Tyr moiety (position 4) indeed inserts in the S1 pocket of thrombin. Because of the kink in the Tyr side chain, the side chain necessarily slots in the S1 pocket other than an extended L-Arg or L-Lys. The OH group of D-Tyr nevertheless adapts a position nearly identical to one of the terminal nitrogens of either L-Arg in Phe-Pro-Arg-chloromethylketone or benzamidine in melagatran as P1 residue. It forms a rectangular bonding network, including a direct charged hydrogen bond to Asp-189, and to a bridging water molecule (water 38), which connects the D-Tyr OH with the second carboxyl oxygen of Asp-189, all at ideal hydrogen-bonding distances of ≈2.6 Å (Fig. 3*B*). This crystal structure is the first, to our knowledge, of an uncharged side chain forming a charged hydrogen bond with Asp-189 at the bottom of the pocket.

During the course of computer-assisted optimization, inhibitor candidates were trained to be highly selective, i.e., not to inhibit trypsin as a representative of other serine proteases. Not surprisingly therefore, the majority of optimized molecules in cycles 7 and 8, including the inhibitors presented in Table 1, did not affect trypsin activity to a significant extent. Because none of the deriv-

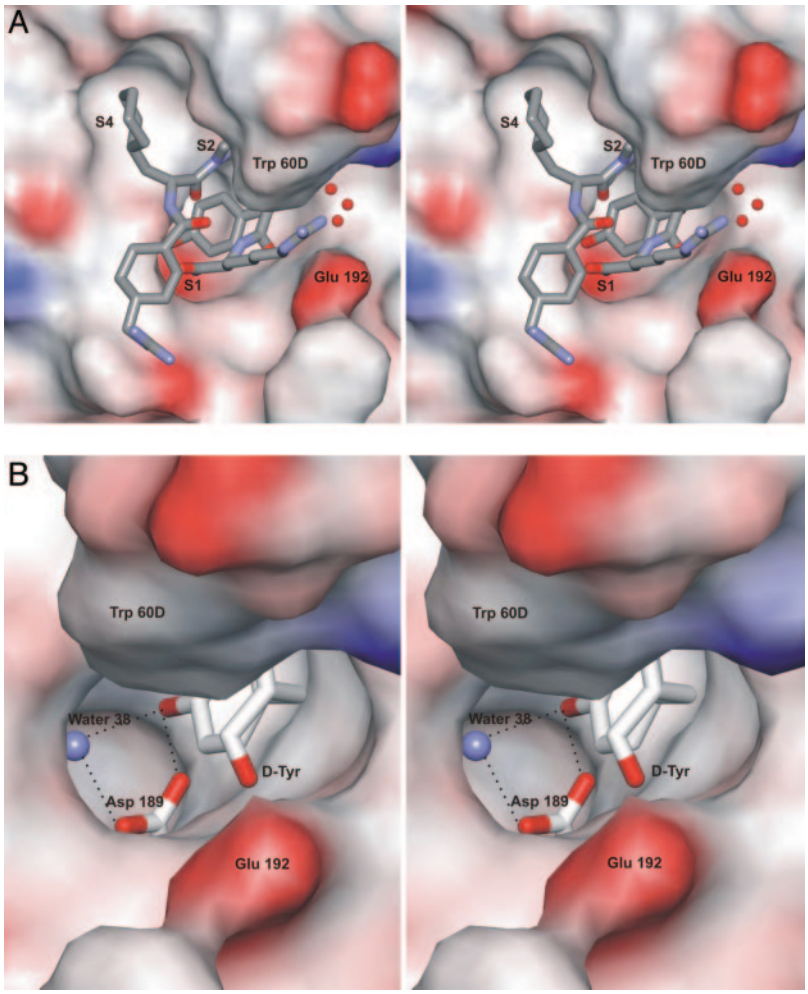

Fig. 3. Structure of the complex between human thrombin, hirugen, and compound 8-5. (*A*) Overall stereoview in standard orientation (active-site cleft running from left to right) slightly rotated around the *y* axis. The thrombin surface is shown with electrostatic coloring (red, negative potential; blue, positive potential). Compound 8-5 is represented as a stick model. Ordered water molecules are shown as green balls. (*B*) View of the S1 pocket. Asp-189 of thrombin and D-Tyr of 8-5 are shown as stick models. Water 38 is represented as a blue ball. The thrombin surface is shown with electrostatic coloring. Hydrogen bonds are represented as dotted lines.

atives of 8-5 showed a dramatic loss in selectivity against trypsin, selectivity presumably is not caused by a single specificity conferring element. Remarkably, selection against inhibition of the single "antitarget" trypsin also had an impact on the capabilities of inhibiting other proteases. Factor Xa, Factor XIIa, urokinase, plasmin, and Plasma kallikrein were not inhibited by the best inhibitors of cycles 7 and 8 (with a selectivity of >100,000-fold). Note that these inhibitors even discriminate between human and bovine thrombin. In contrast, argatroban, NAPAP, and melagatran exhibited only a moderate selectivity. In particular, melagatran, which nevertheless is believed to have sufficient selectivity toward the fibrinolytic enzymes at the necessary therapeutic concentrations, is well known for its inhibitory activity against other serine proteases such as trypsin (7).

Although not directly assessed in the course of the optimization, compounds 7-4, 8-1, and 8-5 also showed anticoagulant properties comparable with that of argatroban, NAPAP, and melagatran. As for the latter, the concentrations effective in doubling the respective plasma clotting times (IC_{200}) in the thrombin time, partial thromboplastin times, and prothrombin time assays *in vitro* have the typical ratio of ≈1:10:20, which is known for a variety of selective thrombin inhibitors with low or moderate serum protein binding (8).

For all selected inhibitors (Table 1), metabolic stability was high (> 70%), and solubility was excellent. In addition, the inhibitors showed low, or at least moderate, serum protein-binding capacities (< 60%).

Although we did not include a direct measure of elimination from the circulation as a parameter in the multiparameter optimization process, we nevertheless tested several of the thrombin inhibitors for their elimination from the bloodstream (Fig. 4) in a rat model system. Compounds 7-8, 8-1, and 8-5 revealed a delayed elimination from the bloodstream that was comparable with, or even slower than, that of melagatran. Only compound 7-4 showed a faster

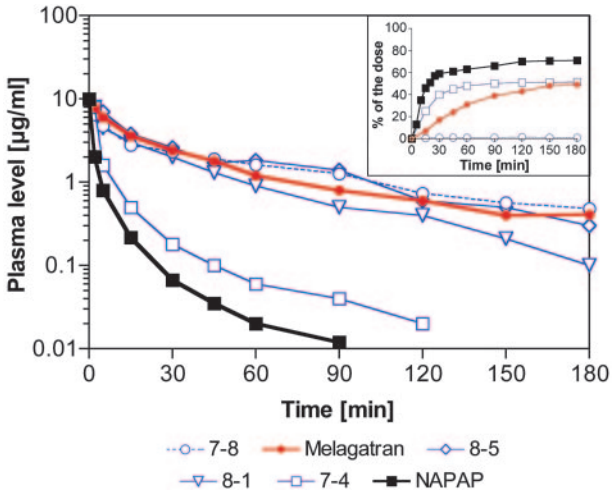

Fig. 4. Elimination of selected compounds after i.v. application of a 1 mg/kg (rat model). Elimination from the blood stream. (*Inset*) Excretion through the bile.

elimination that was intermediate between that of melagatran and NAPAP. Next, we tested two of the selected thrombin inhibitors for their biliary excretion, which usually is taken as an important mechanism for presystemic elimination, thereby limiting oral bioavailability. Whereas NAPAP, melagatran, and 7-4 underwent significant elimination through the bile (71%, 49%, and 51% of the dose within 6 h, respectively), no biliary excretion was detected for 7-8, which (like 8-1 and 8-5) was exclusively renally excreted. It is interesting to note that the only structural difference between 7-4 and 7-8 is an acetyl moiety at the N terminus of the peptide. Previous studies with NAPAP analogs (9, 10) had revealed that derivatives that contained more than one basic group showed lower systemic and biliary clearance compared with NAPAP. This finding is in line with our findings that a free, i.e., a positively charged N terminus as in compounds 7-8, 8-1, and 8-5 clearly increases the half-life in the bloodstream, compared with that of compounds with an acetylated N terminus such as 7-4.

Discussion

Previously, we have demonstrated that CADDIS, which is based on a recently developed type of algorithm (2), is well applicable to the optimization of macromolecules, such as RNAs and peptides, with respect to single-molecular properties, e.g., thrombin inhibition. Here, we show that the same concept is also valid as an approach to the problem of multiparameter optimization, and thereby, this opens up another perspective in early drug discovery. Whereas the definition of "fitness" is clear for a single-parameter optimization, different definitions may apply for a multiparameter optimization. Our studies revealed that the geometric mean of nonweighted fitness values of each, individual parameter provides a simple and appropriate fitness function. To primarily obtain small-molecule drug candidates, the fitness function includes a penalty for large sequences as described (2). To prove the feasibility of CADDIS in the context of multiparameter optimization, we sought to develop thrombin inhibitors with a property profile that makes them attractive candidates for further preclinical and clinical development.

Starting from a set of 170 randomly chosen compounds, eight design cycles with a total of $<$1,000 compounds turned out to be sufficient to identify a novel series of thrombin inhibitors. These compounds combined powerful thrombin inhibitory activity with an extraordinarily high selectivity, negligible toxicity, excellent solubility, moderate serum protein binding, high metabolic stability, and a slow systemic plasma clearance in a rat model. Mainly because of their unmatched selectivity along with their slow and exclusively renal clearance, certain advantages over other small-molecule thrombin inhibitors may be envisaged concerning, e.g., indications like arterial thrombosis, where usually higher concentrations of the drug are needed (11-13), and hence, the risk of side effects may be higher. Because the inhibitor series necessarily contains a neutral P1 moiety and largely preserves inhibitory activity when basic residues are substituted by noncharged residues, it has potential for further development into orally available compounds.

However, computer-assisted multiparameter optimization is not restricted to the development of thrombin inhibitors but has also been successfully applied e.g., to the optimization of antibiotics (data not shown). Therefore, parallel optimization of a multitude of molecular properties can indeed generate valuable lead molecules in a significantly less costly and time-consuming process as compared with concepts based on high-throughput activity screening and subsequent optimization of primary hits with respect to additional properties. In addition, the risk of eliminating seemingly suboptimal candidate molecules that may nevertheless be superior in their capability for further development too early in the drug developmental process is significantly lower compared with a sequentially organized high-throughput screening approach that largely voluntarily retains a certain percentage of the initial hits for further development. However, as with any sampling strategy that is not testing all members of a given compound space, there is certainly no guarantee to identify "the fittest" molecule. Because within the overall process, molecules may improve in different ways, i.e., properties, chances are low that molecular solutions become trapped in a low-grade local optimum. Nevertheless, a prerequisite to successful optimization is the generation of high-quality data obtained in pharmacologically relevant assays. Therefore, assay data quality has to be monitored on the basis of standard statistical procedures (14, 15). A major advantage of the discovery approach described herein is that only medium throughput is mandatory. Therefore, complex and expensive, i.e., more predictive assay systems early in drug discovery, can now be afforded. However, even in this case, a multiparameter optimization process would not necessarily yield a satisfactory lead molecule. This finding is particularly true when the optimization comprises conflicting properties. For thrombin inhibitors, very often, basic groups such as guanidine moieties proved to be favorable for S1 binding, i.e., enzyme inhibition, and slow elimination, but they significantly hinder acceptable enteral absorption. In these cases, only prodrug approaches like the one that has been applied with melagatran to yield H 376/95 (16) may increase the structural and functional variability to meet all of the necessary demands. The main limitation of the current computer-assisted multiparameter optimization approach is its inability to perform on nonpolymeric small molecules. However, this restriction could be overcome, e.g., by using a fragment-based type of encoding (17, 18). Altogether, computer-assisted multiparameter optimization may add a valuable option to present drug discovery approaches and speed the development of new patentable drugs.

This work was supported in part by Ministry of Research of Lower Saxony Grant 203.32329-4/1-99 (12), Bundesministerium für Forschung und Technologie BioFuture Grant 0311852, and Novel Science International GmbH, European Commission Structural Proteomics in Europe Contract QLG2-CT-2002-00988 and the "Fonds der Chemischen Industrie."

1. Steinmetzer, T., Hauptmann, J. & Stürzebecher, J. (2001) *Exp. Opin. Invest. Drugs* **10,** 845–864.
2. Kamphausen, S., Höltge, N., Wirsching, F., Morys-Wortmann, C., Riester, D., Goetz, R., Thürk, M. & Schwienhorst, A. (2002) *J. Comput. Aided Mol. Des.* **16,** 551–567.
3. Mosmann, T. (1983) *J. Immunol. Methods* **65,** 55–63.
4. Skrzypczak-Jankum, E., Carperos, V., Ravichandran, K. G., Tulinsky, A., Westbrook, M. & Maraganore, J. (1991) *J. Mol. Biol.* **221,** 1379–1393.
5. Okada, Y., Tsuda, Y., Tada, M., Wanaka, K., Hijikata-Okunomiya, A., Okamoto, U. & Okamoto, S. (1999) *Biopolymers* **51,** 41–50.
6. Dullweber, F., Stubbs, M. T., Musil, G., Stürzebecher, J. & Klebe, G. (2001) *J. Mol. Biol.* **313,** 593–614.
7. Gustafsson, D., Antonsson, T. & Bylund, R. (1998) *Thromb. Haemostasis* **79,** 110–118.
8. Hauptmann, J. & Stürzebecher, J. (1999) *Thromb. Res.* **93,** 203–241.
9. Hauptmann, J., Steinmetzer, T., Vieweg, H., Wikstrom, P. & Stürzebecher, J. (2002) *Pharm. Res.* **19,** 1027–1033. 24
10. Steinmetzer, T., Schweinitz, A., Kunzel, S., Wikstrom, P., Hauptmann, J. & Stürzebecher, J. (2002) *J. Enzyme Inhib. Med. Chem.* **17,** 241–249.
11. Schumacher, W. A., Balasubramanian, N., St. Laurent, D. R. & Seiler, S. M. (1994) *Eur. J. Pharmacol.* **259,** 165–171.
12. Schumacher, W. A., Heran, C. H. & Steinbacher, T. E. (1996) *J. Cardiovasc. Pharmacol.* **28,** 19–25.
13. Lyle, E. M., Lewis, S. D., Lehman, E. D., Gardell, S. J., Motzel, S. L. & Lynch, J. J. (1998) *Thromb. Haemostasis* **79,** 656–662.
14. Taylor, P. B., Stewart, F. P., Dunnington, D. J., Quinn, S. T., Schulz, C. K., Vaidya, K. S., Kurali, E., Lane, T. R., Xiong, W. C., Sherrill, T. P., Snider, J. S., Terpstra, N. D. & Hertzberg, R. P. (2000) *J. Biomol. Screening* **5,** 213–225.
15. Zhang, J.-H., Chung, T. D. Y. & Oldenburg, K. R. (1999) *J. Biomol. Screen.* **4,** 67–73.
16. Gustafsson, D., Nyström, J. E. & Carlsson, S. (2001) *Thromb. Res.* **101,** 171–181.
17. Westhead, D. R., Clark, D. E., Frenkel, D., Li, J., Murray, C. W., Robson, B. & Waszkowycz, B. (1995) *J. Comput. Aided Mol. Design* **9,** 139–148.
18. Schneider, G., Lee, M.-L., Stahl, M. & Schneider, P. (2000) *J. Comput. Aided Mol. Design* **14,** 487–494.