VOLUME 73, NUMBER 12

JOC Perspective

June 20, 2008

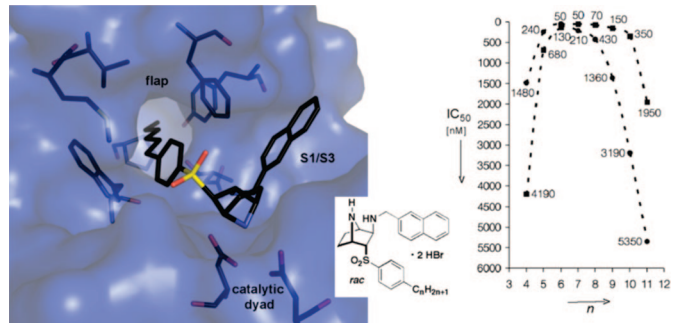
Structure-Based Drug Design: Exploring the Proper Filling of Apolar Pockets at Enzyme Active Sites

Martina Zürcher and François Diederich*

Department of Chemistry and Applied Biosciences, Laboratorium für Organische Chemie, ETH Zürich, HCI G 313, 8093 Zürich, Switzerland

diederich@org.chem.ethz.ch

Received March 7, 2008



The proper filling of apolar pockets at enzyme active sites is central for increasing binding activity and selectivity of hits and leads in medicinal chemistry. In our structure-based design approach toward the generation of potent enzyme inhibitors, we encountered a variety of challenges in gaining suitable binding affinity from the occupation of such pockets. We summarize them here for the first time. A fluorine scan of tricyclic thrombin inhibitors led to the discovery of favorable orthogonal dipolar C−F···C═O interactions. Efficient cation−π interactions were established in the S4 pocket of factor Xa, another serine protease from the blood coagulation cascade. Changing from mono- to bisubstrate inhibitors of catechol *O*-methyltransferase, a target in the L-Dopa-based treatment of Parkinson's disease, enabled the full exploitation of a previously unexplored hydrophobic pocket. Conformational preorganization of a pocket at an enzyme active site is crucial for harvesting binding affinity. This is demonstrated for two enzymes from the nonmevalonate pathway of isoprenoid biosynthesis, IspE and IspF, which are pursued as antimalarial targets. Disrupting crystallographically defined water networks on the way into a pocket might cost all of the binding free enthalpy gained from its occupation, as revealed in studies with tRNA-guanine transglycosylase, a target against shigellosis. Investigations of the active site of plasmepsin II, another antimalarial target, showed that principles for proper apolar cavity filling, originally developed for synthetic host−guest systems, are also applicable to enzyme environments.

Introduction

Structure-based ligand design finds increasing application in lead generation and optimization, given the large number of target proteins for which structural information is available from X-ray crystallographic or NMR analysis.[1] The success of this strategy is intimately coupled to a profound understanding of molecular recognition processes at atomic resolution.[2] In the early 1990s, we expanded our research program on synthetic host−guest chemistry in aqueous solution[3,4] to include the investigation of complexes between enzymes and small, non-peptidic ligands originating from structure-based design.[5] To complete this multidimensional approach toward molecular recognition, we also started extensive database mining in the Cambridge Structural Database (CSD) and the Protein Data

10.1021/jo800527n CCC: $40.75 © 2008 American Chemical Society
Published on Web 05/30/2008

Bank (PDB).[6–8] The knowledge gained on the nature of individual bonding interactions and their contribution to the overall free enthalpy of complexation is widely applicable to drug discovery research, enabling in particular structure-based design approaches but equally also pharmacophore-based lead optimization.

The small-molecule ligands in our work do not arise from any screening approach. In quite a number of cases, the first-generation compounds resemble the substrates of the targeted enzymes. But in others, a true de novo design was applied, based on chemical structure intuition and molecular recognition principles. The computer modeling program MOLOC[9] was used to visualize the three-dimensional structure of the protein target, to energy minimize the prospective ligand docked manually into the enzyme active site, and to optimize its stereoelectronic complementarity. None of the commercially available modeling packages and force fields feature a perfect parametrization with respect to weak intermolecular interactions and conformational analysis. For successful modeling, it is therefore important to know and take into account the strengths and weaknesses of the software used. Whereas design, database mining, and synthesis in this work took place at the ETH, biological study and protein X-ray crystallography were performed in the laboratories of outstanding academic and industrial collaborators whose names are found in the various references. In the majority of projects, the first compounds designed showed an inhibitory activity (K_i (inhibition constant) or IC_{50} (concentration at which 50% maximal initial reaction rate is observed)) between 100 nM and 20 μM. Within one to two X-ray structure-based optimization cycles, lead compounds with single-digit nanomolar activities have been reached in all advanced projects. Consequently, structure-based design in our hands offers an intelligent, economic alternative to screening approaches for the identification of first actives and hits.

In this paper, we describe the proper filling of apolar subpockets at enzyme active sites in several case studies. Appropriate occupation of such pockets by ligand moieties is the key to success for gaining suitable binding efficacy in numerous lead optimization programs.[10] This topic had already been at the center of our earlier host−guest studies with cyclophane receptors in aqueous solution.[3,11] In that work, we showed that tight apolar complexation of aromatic or steroidal guests in cyclophane cavities is characterized by a strong enthalpic driving force (the "nonclassical hydrophobic effect"),[6,12] originating from gains in dispersion interactions and favorable changes in solvent cohesive interactions. We found that the enthalpic driving force is partially compensated by unfavorable entropic terms.[12] Enthalpy−entropy compensation is nearly universally observed for chemical and biological recognition processes,[6,13,14] As a rule of thumb, tight apolar binding is more enthalpically driven, whereas loose binding benefits from a larger entropic driving force.[15] In the latter case, entropic losses resulting from restricted dynamic mobility of the binding partners in the complex are small and do not reduce the gains in desolvation entropy (i.e., the hydrophobic effect). Guest size also matters: with increasing surface, the entropic term becomes more favorable because of the enhanced desolvation entropy.

A Fluorine Scan To Map the Fluorophilicity/Fluorophobicity of the Subpockets at the Thrombin Active Site

With its limited flexibility, the active site of thrombin, a serine protease from the blood coagulation cascade, provides an ideal environment for molecular recognition studies. We introduced two series of inhibitors featuring rigid, tricyclic lactam (e.g., (+)-**1**) and imide (e.g., (±)-**2**) scaffolds, respectively, orienting residues into the three distinct subpockets at the active site (Figures 1 and 2).[5,16] Thus, inhibitor (+)-**1** (K_i = 7 nM, 740-fold selectivity for thrombin over the related digestive protease trypsin) binds to the hydrophobic selectivity (S1) pocket with a phenylamidinium residue (as a surrogate of the Arg side chain in the natural substrate) forming a bidentate salt bridge with the side chain of Asp189 at the bottom of the pocket. The large hydrophobic distal (D) pocket is occupied by a piperonyl residue (Phe side chain in the natural substrate), while the narrow proximal (P) pocket is filled with an isopropyl residue (Val side chain in the substrate). Owing to the rigidity of both the enzyme active site and the tricyclic scaffold, differently decorated inhibitors form complexes of similar geometry, allowing direct correlation of measured changes in the complexation free enthalpy with the contributions of individual ligand substituents. Nearly identical binding modes were observed in a large number of X-ray crystal structures of cocomplexes between thrombin and the tricyclic ligands.[5,16,17]

The phenylamidinium ring in the S1 pocket stacks nicely against the flat, β-pleated sheet-like backbone of Glu217-Gly216-Trp215 (Figure 1c). Such stacking is a frequently encountered motif of protein backbone-aromatic ring interaction.[6] For nearly two decades, thrombin has been an important target for antithrombotic therapies, and correspondingly, numerous low-molecular weight ligands have been reported.[19] While the majority feature a basic residue (such as arginine, amidine, aminopyridine, or imidazole) to fill the S1 pocket, potent inhibitors have also been reported that lack such residue to interact with Asp189 but rather feature nonpolar Arg surrogates such as chlorophenyl,[20a] benzothiophenyl,[20b] or benzothiazolyl[20c] "needles". We spent considerable efforts in replacing the phenylamidinium group by less basic or neutral residues to increase the bioavailability of our ligand, but with little success.[21] Apparently, the rigid tricyclic scaffold does not allow a proper positioning of such substituents in the S1 pocket.

Understanding the intrinsic recognition properties of fluorine substituents in protein environments is of great interest for structure-based lead design and optimization.[22] Over the past six years, we therefore undertook a fluorine scan of our tricyclic thrombin inhibitors in order to map the fluorophilicity and fluorophobicity of an entire enzyme active site. As part of this investigation, we systematically exchanged the H atoms of the phenylamidinium needle in (±)-**2** with one or two F atoms (resulting in the fluorinated compounds (±)-**3**−(±)-**7**, Figure 3a) in order to lower its basicity and increase the bioavailability of the ligands.[23] According to the modeling, such H/F-replacements do not cause any steric clashes in the S1 pocket. Upon fluorine introduction, the pK_{a2} values for the amidinium residue decrease as expected and the lower basicity of the needle[24] translates into a desirable increase in the logarithmic distribution coefficient log *D*. The potency of the ligands against thrombin and trypsin, however, was decreased upon F-substitution. Linear free-energy relationships between the free enthalpy of complexation ($-\Delta G$) to both enzymes and the pK_{a2} values unexpectedly revealed that binding affinity against thrombin is much more affected by the decrease in pK_{a2} than the affinity against trypsin (Figure 3b). This means that the pK_{a2} value of the phenylamidinium residue not only affects binding affinity but also selectivity. Improvement of the pharmacokinetic

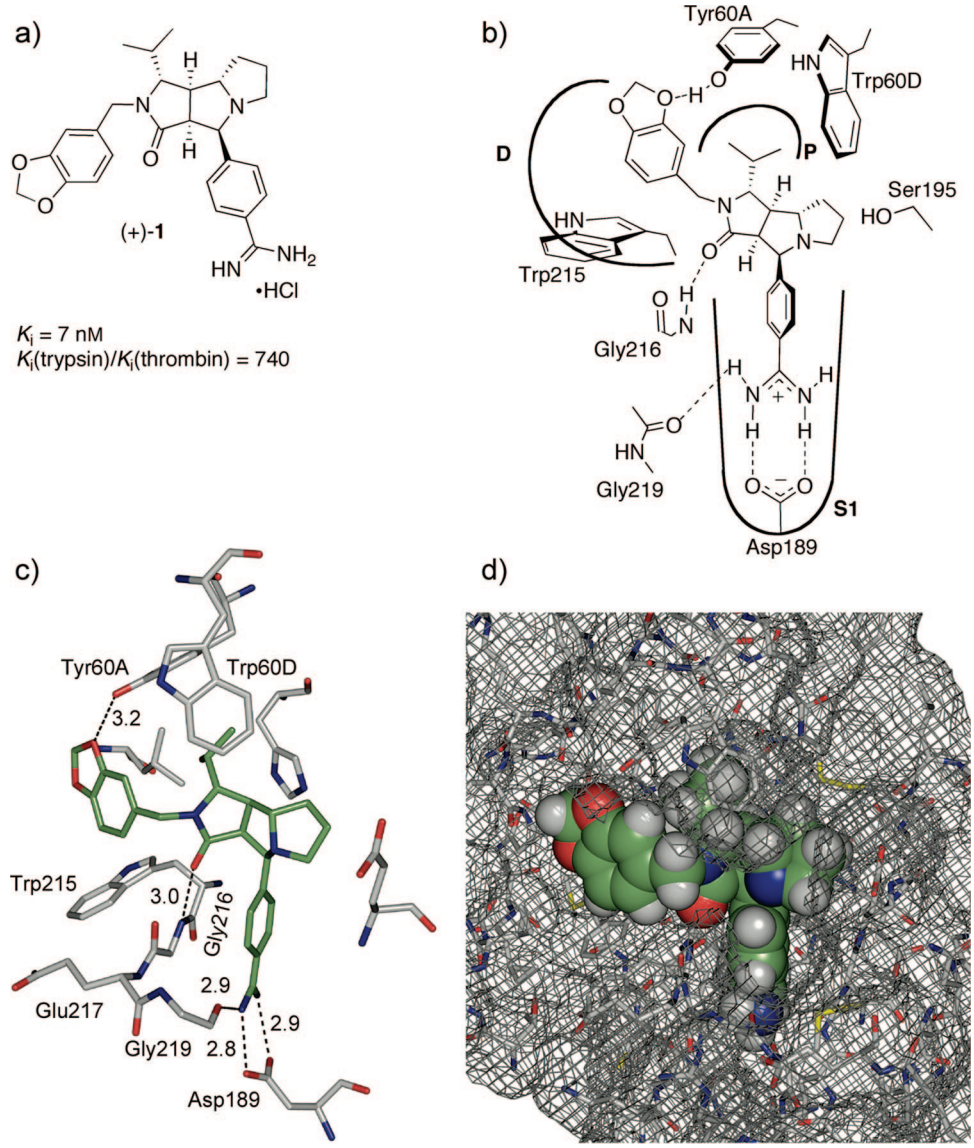


FIGURE 1. (a) Chemical structure, activity, and selectivity over trypsin of the optically pure tricyclic lactam inhibitor (+)-**1**. (b) Schematic representation of the binding mode of (+)-**1** at the active site of thrombin. (c) Active site region in the X-ray crystal structure of the complex between thrombin and (+)-**1** (resolution 1.93 Å).[5b] (d) Space-filling representation of (+)-**1** and solvent-accessible surface of thrombin. Color code: C atoms of inhibitor, green; C atoms of protein, gray; H atoms, white; O atoms, red; N atoms, blue; S atoms, yellow. H bonds are represented as dashed lines, and distances between heavy atoms are given in Å. The units for the indicated distances and the color code are maintained throughout the article, unless otherwise stated.

properties of thrombin inhibitors through F-substitution of the phenylamidinium needle seems therefore not to be a viable strategy despite the desirable, significant lowering of the pK_{a2} values. Note also the extremely low pK_{a1} values of the tertiary amine centers in the tricyclic scaffold, which can be largely rationalized by σ-inductive substituent effects.[23,24]

The D pocket of thrombin is spacious and has been shown to accommodate large residues such as dicyclohexylmethyl or diphenylmethyl.[25] Voluminous substituents departing from the tricyclic skeleton are also favorably directed into this pocket, generating highly active ligands such as (±)-**8** and (±)-**9** that feature some of the highest selectivities for thrombin over trypsin (Figure 4).[21b] Aromatic and aliphatic residues pointing into the D pocket benefit from interactions with the indole ring of Trp215 at the floor of the pocket. Aromatic rings such as the piperonyl or 4-fluorophenyl residues in (+)-**1** (Figure 1) and (±)-**2** (Figure 2) prefer edge-to-face interactions, also for conformational reasons, whereas aliphatic residues undergo CH$-\pi$ interactions.[6,26] In recent theoretical work, both interaction modes have been shown to be largely of a dispersive nature, with electrostatics being only a minor contributor.[27,28] This is also reflected in the complexation strength measured for ligands with similarly sized aromatic and alicyclic residues to fill the D pocket. Benzyl- and cyclohexylmethyl-substituted inhibitors (±)-**10** (K_i = 310 nM) and (±)-**11** (K_i = 350 nM) show similar binding affinities.[16]

Investigating the occupancy of the D pocket of thrombin, Klebe and co-workers[26] recently obtained interesting new insight into how ligand mobility affects the thermodynamic characteristics of a complexation event.[26] They solved the crystal structures of ligands **12** and **13** bound to thrombin and found similar difference electron densities for both compounds, indicative of similar binding geometries, with one major exception (Figure 5). While the cyclopentyl ring in **12** in the D pocket is well-defined by the electron density, the cyclohexyl

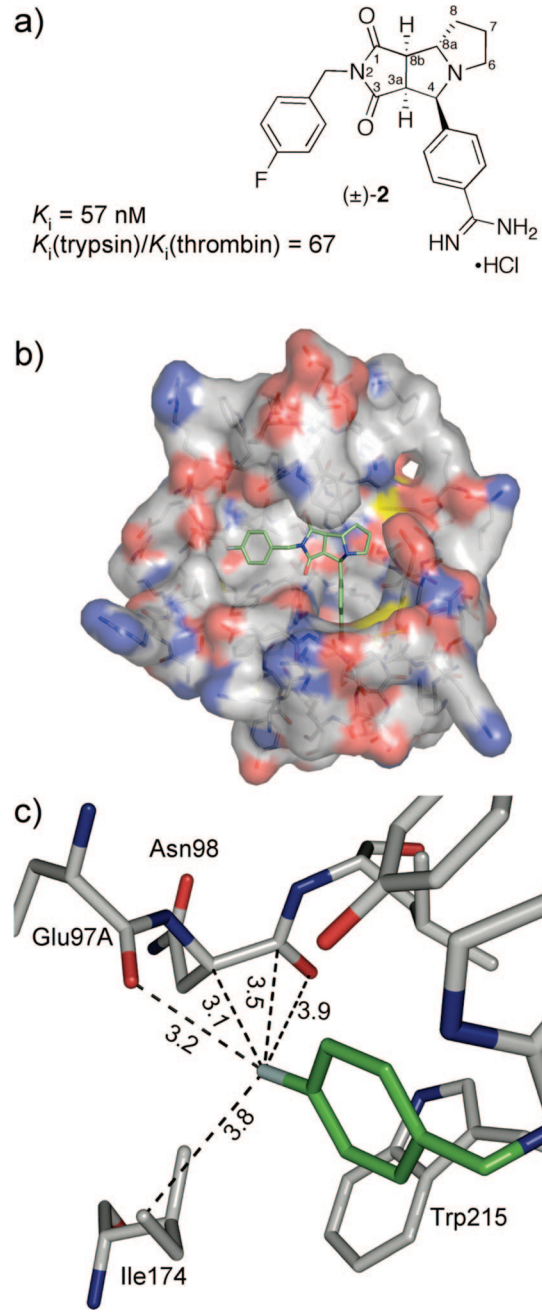
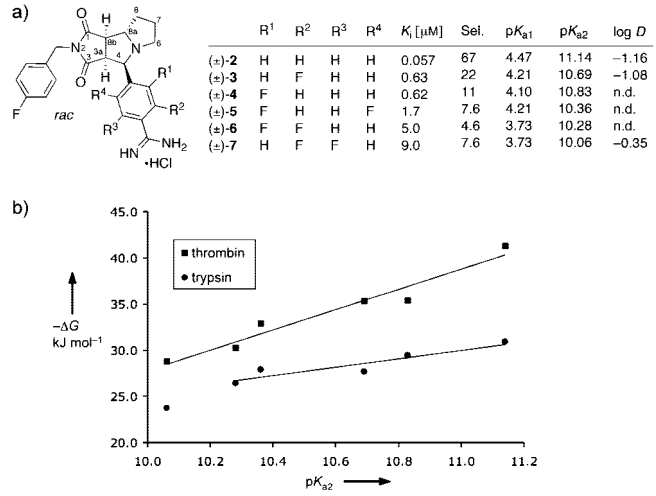


FIGURE 2. (a) Chemical structure, activity, and selectivity over trypsin of the tricyclic imide inhibitor (±)-**2**. (b) Ligand (±)-**2** bound to thrombin (PDB code 1OYT, resolution 1.67 Å; only the (3a*S*,4*R*,8a*S*,8b*R*)-configured enantiomer is bound).[18] The Connelly surface of the protein active site is shown. (c) Binding mode of the 4-fluorobenzyl moiety of imide (±)-**2** in the D pocket of thrombin.



	R^1	R^2	R^3	R^4	K_i [μM]	Sel.	pK_{a1}	pK_{a2}	log *D*
(±)-**2**	H	H	H	H	0.057	67	4.47	11.14	−1.16
(±)-**3**	H	F	H	H	0.63	22	4.21	10.69	−1.08
(±)-**4**	F	H	H	H	0.62	11	4.10	10.83	n.d.
(±)-**5**	F	H	H	F	1.7	7.6	4.21	10.36	n.d.
(±)-**6**	F	F	H	H	5.0	4.6	3.73	10.28	n.d.
(±)-**7**	H	F	F	H	9.0	7.6	3.73	10.06	−0.35

FIGURE 3. F scan in the S1 pocket. (a) Chemical structure, biological activities, and physicochemical properties of inhibitors (±)-**2**−(±)-**7** with fluorinated phenylamidinium needles. Uncertainty in K_i values (thrombin): ±20%. Sel.: K_i(trypsin)/K_i(thrombin). pK_{a1}: tertiary amine in the tricyclic skeleton. pK_{a2}: phenylamidinium. Reproducibility of pK_a values: ± 0.05 units. n.d.: not determined. (b) Plots of the binding free enthalpies $-\Delta G$ for the complexes of the fluorinated tricyclic inhibitors with thrombin and trypsin against the pK_{a2} values.

ring in the D pocket of **13** is ill-defined. The authors explained the poorly defined difference electron density with enhanced mobility of the cyclohexyl ring, possibly adopting multiple orientations and presumably undergoing rapid chair inversion. Subsequent molecular dynamics simulations provided support for this explanation, with distance fluctuations along the time trajectory being more substantial in the case of the cyclohexyl than the cyclopentyl ring (Figure 5b). Isothermal calorimetry data clearly reflect the differences in dynamic mobility. Both ligands bind with very similar association strength (K_b, ΔG^o in Figure 5a) based, however, on very different enthalpic and entropic contributions. Complexation of the apparently more rigidly binding cyclopentane derivative **12** is driven by similarly favorable enthalpic and entropic contributions. In contrast, binding of cyclohexane derivative **13** has a much weaker enthalpic driving force, since less dispersion interactions are gained in the apparently more dynamic complex. However, the mobility of the cyclohexyl residue and the surrounding protein environment seems also less reduced and therefore the entropic term is much more favorable, fully compensating for the loss in enthalpy. These results are quite remarkable but also not understood in full detail, in particular because the D pocket is very large (see the high potency of (±)-**8** and (±)-**9**, Figure 4) and should easily accommodate both cycloalkyl rings.

In an extension of our F scan, we replaced the benzyl ring in (±)-**10** (Figure 4) to occupy the D pocket by six mono-, di-, and pentafluorinated derivatives.[18] The 4-fluorobenzyl derivative (±)-**2** showed 5-fold enhanced activity ($\Delta\Delta G$ = 4.2 kJ mol^{-1}) relative to (±)-**10**. The X-ray crystal structure of the complex with (±)-**2** (PDB code 1OYT) was solved to reveal that the benzylic F atom is in close contact with the enzyme (Figure 2c).[18] It is only 3.1 Å away from the α-C atom of Asn98, forming a dipolar C−F···H−C contact. It is also only 3.5 Å away from the carbonyl C atom of the same residue, with the F atom pointing nearly orthogonally onto the backbone carbonyl of Asn98 (α(F···C═O) = 96°). A subsequent search in the CSD revealed a large number of sub van der Waals contacts (d(F···C═O) ≤ 3.0 Å) between organic F atoms and carbonyl C atoms, with the F atoms clearly preferring a position close to the pseudotrigonal axis of the carbonyl unit. At short distances, the F···C═O angle (α_2) tends toward 90° and the C−F···C(O) angle (α_1) approaches 180°, whereas at longer separation, the angle dependence is weaker (Figure 6). We proposed that the C−F bond undergoes an attractive dipolar interaction with the C═O bond, with the preference for orthogonal dipolar orientation at short F···C distances largely arising from steric reasons. No rehybridization of the carbonyl C═O was observed. A search in the PDB revealed similar short orthogonal C−F···C═O

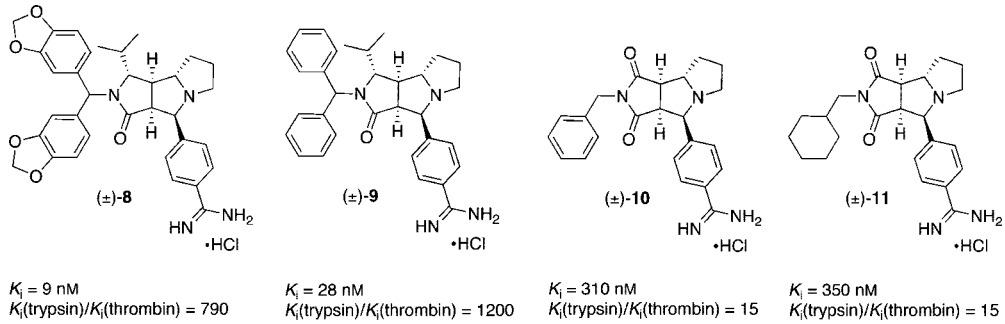
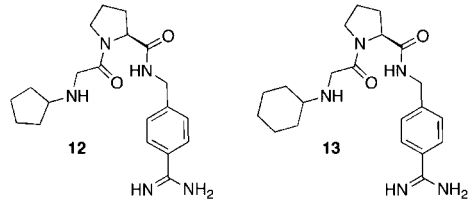


FIGURE 4. Activities of inhibitors (±)-**8**−(±)-**11** that target the D pocket of thrombin with different residues.

a)



	K_b [10^6 L mol^{-1}]	$\Delta G°$ [kJ mol^{-1}]	$\Delta H°$ [kJ mol^{-1}]	$-T\Delta S°$ [kJ mol^{-1}]
12	1.54 ± 0.41	−35.40 ± 0.55	−16.88 ± 0.58	−18.52 ± 0.56
13	2.18 ±0.62	−36.16 ± 0.58	−10.50 ± 0.28	−25.66 ± 0.43

b)

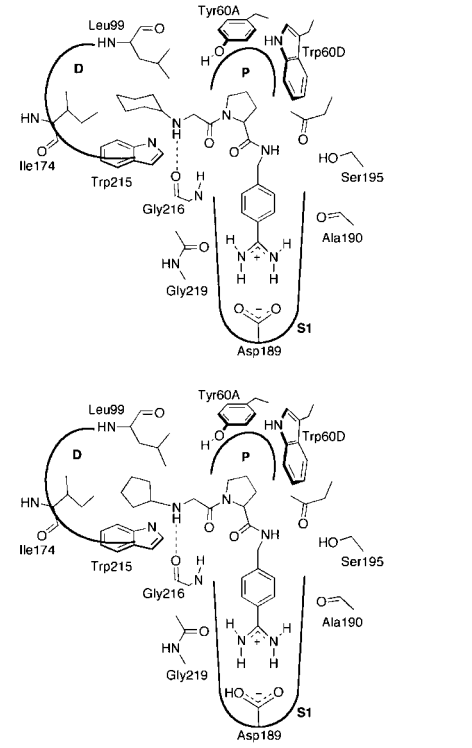


FIGURE 5. (a) Binding constants K_b and thermodynamic quantities determined by isothermal calorimetry at 298 K for the complexes of thrombin with **12** and **13**.[26] (b) Schematic drawing of the cyclopentyl derivative **12** (top) and the cyclohexyl derivative **13** (below) bound to thrombin.

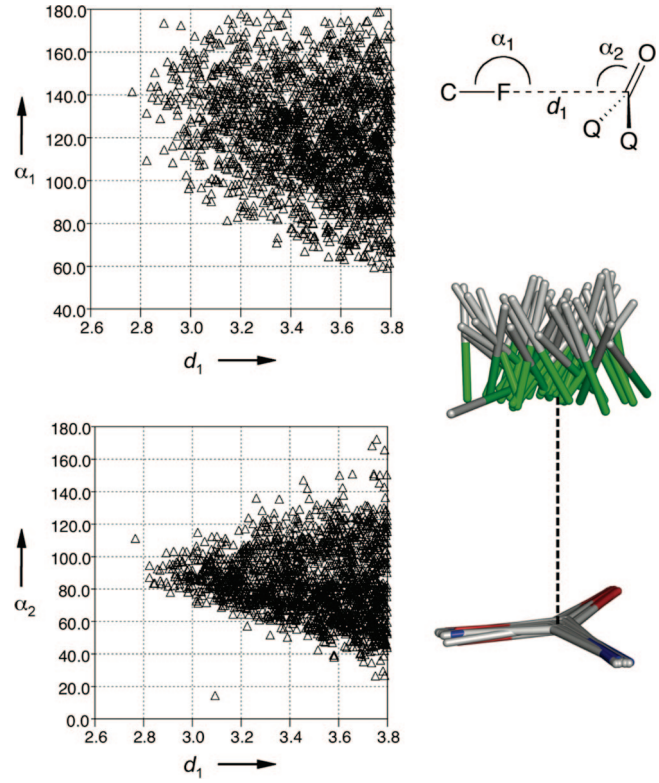
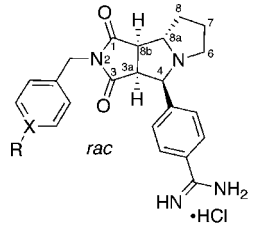


FIGURE 6. Scatter plots of α_1 versus d_1 (upper left) and α_2 versus d_1 (lower left) of 1848 occurrences of close C−F···C═O contacts with 2.3 Å < d_1 < 3.8 Å obtained from 1087 hits of the CSD search of fluorine-containing organic carbonyl derivatives (Q = C, N, O).[30] Right: Superposition of 50 intermolecular C−F···C═O contacts with 2.77 Å < d_1 < 3.0 Å extracted from the crystal structures and superimposed on the carbonyl unit (O atom pointing backwards). The vertical dashed line marks the pseudotrigonal axis of the carbonyl system. F atoms are colored green.

contacts in other protein−ligand complexes, which however had largely been unrecognized.[29]

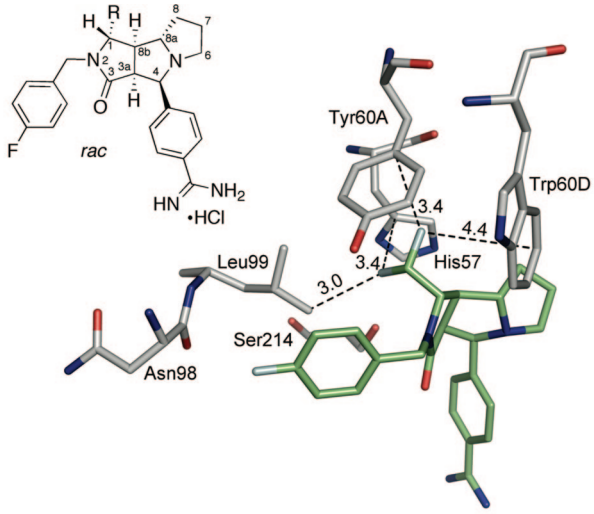
Upon variation of the substituent in position 4 of the benzyl ring in the series (+)-**2**, (±)-**10**, and (±)-**14**−(±)-**18**, the inhibitory constants of the imide ligands vary greatly, with K_i values increasing in the sequence F < Cl < H < OMe < OH < N_{pyr} ≪ Br (Figure 7).[31] This sequence can be explained by both steric fit and the occurrence of orthogonal interactions with the backbone C═O of Asn98. Interestingly, the latter interactions are absent in the lactam series because filling of the P pocket widens the D pocket slightly by shifting the peptide loop segment Asn98-Leu99 (Figure 8) away from the bound inhibitor, as revealed by crystal structure analyses.

In parallel to these investigations, we conducted a comprehensive search on all types of dipolar interactions in chemistry (CSD) and structural biology (PDB) and found that, at short distances below van der Waals contact, orthogonal interactions are strongly preferred over other geometries such as the energetically more favorable antiparallel dipolar alignments.[7]



	X–R	K_i [μM]	Sel.
(±)-**2**	C–F	0.057	67
(±)-**14**	C–Cl	0.19	30
(±)-**10**	C–H	0.31	15
(±)-**15**	C–OMe	0.621	4.1
(±)-**16**	C–OH	0.89	9.7
(±)-**17**	N	1.54	15.4
(±)-**18**	C–Br	23.7	> 2.9

FIGURE 7. Chemical structure, K_i values (thrombin), and selectivity (Sel. = K_i(trypsin)/K_i(thrombin)) of compounds (±)-**2**, (±)-**10**, and (±)-**14**−(±)-**18**. Uncertainty in K_i values: ± 20%. For all compounds, only the (3a*S*,4*R*,8a*S*,8b*R*)-configured enantiomer is bound, as determined from the crystal structure analysis in refs 5, 18, 23, and 29.

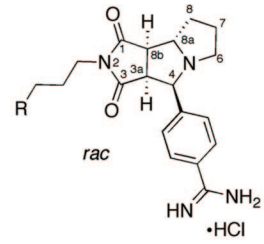


	R	K_i [μM]	Sel.	pK_{a1}	pK_{a2}	log *D*	*V*(R) [Å^3]
(±)-**19**	$CHMe_2$	0.005	413	6.52	n.d.	−1.08	60.0
(±)-**20**	CH_2Me	0.014	501	6.27	11.15	−0.79	43.3
(±)-**21**	$HC{=}CH_2$	0.008	835	6.02	11.15	−0.90	37.9
(±)-**22**	$(CH_2)_2Me$	0.078	70	6.16	n.d.	−0.59	60.0
(±)-**23**	$CH_2CH{=}CH_2$	0.036	156	6.19	n.d.	−0.91	54.7
(±)-**24**	CH_2F	0.028	407	5.95	10.98	−1.09	32.2
(±)-**25**	CHF_2	0.022	297	5.61	11.07	−1.16	38.0
(±)-**26**	CH_2CH_2F	0.012	395	6.08	10.96	−1.24	49.0
(±)-**27**	CH_2CHF_2	0.038	104	6.00	11.32	−1.01	54.8

FIGURE 8. Biological data of inhibitors (±)-**19**−(±)-**27** and binding mode of the difluoromethyl moiety of (±)-**25** in the P pocket of thrombin as determined by X-ray crystallography (PDB code 2CN0, resolution 1.30 Å, only the (3a*S*,4*R*,8a*S*,8b*R*)-configured enantiomer is bound).[35] F atoms are colored light blue. The volumes *V*(R) were calculated using Spartan '06[36] and van der Wals radii of Bondi.[37]

Using chemical double-mutant cycles[32] on a monomolecular system, a Wilcox molecular torsional balance,[33] we recently showed that the interaction between C_{sp2}−F and the C═O moiety of an amide group is energetically favorable, amounting to $\Delta\Delta G$ between −0.8 and −1.2 kJ mol^{-1} in apolar environments.[34] No experimental quantification of a dipolar interaction had previously been reported.

Whereas the reader is referred to a recent review for more details,[22] the major attractive interactions of organofluorine in protein environments can be summarized by the following: As small atoms of high electronegativity, fluorine substituents on ligands prefer to orient toward electropositive regions of receptor sites. Distinct fluorophilic environments in proteins are the ubiquitous peptide bonds, which undergo multipolar C−F···H−N, C−F···C═O, and C−F···H−C$_\alpha$ interactions,



		K_i [μM] thrombin	trypsin	factor Xa
(±)-**28**	R = $NMe_3^+Br^-$	7.76	10.5	0.28
(±)-**29**	R = CMe_3	0.13	7.1	29

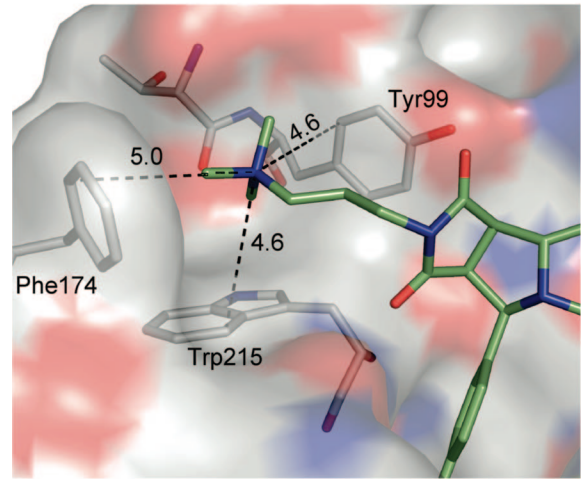


FIGURE 9. Inhibition constants of inhibitors (±)-**28** and (±)-**29** and binding mode of (±)-**28** in the S4 pocket of factor Xa as determined by X-ray crystallography (PDB code 2BOK, resolution 1.64 Å, only the (3a*S*,4*R*,8a*S*,8b*R*)-configured enantiomer is bound).[40]

as well as the side-chain amide residues of Asn and Gln, and the positively charged guanidinium side chain of Arg. Correspondingly, organofluorine introduction into regions of high negative charge density adversely affects binding affinity. At close intermolecular distances, fluorine thus avoids pointing directly onto the O atoms of C═O groups or into the π cloud of aromatic rings (see below).

The narrow hydrophobic P pocket at the thrombin active site is lined by Tyr60A, Trp60D, His57, and Leu99, whereas it is absent in trypsin. Optimal filling of this cavity therefore does not only give enhanced binding affinity but also greater selectivity against trypsin. We prepared a large series of tricyclic lactam inhibitors (±)-**19**−(±)-**27**, bearing similarly sized *exo*-alkyl, -alkenyl, and -fluoroalkyl groups (exo with respect to the bicyclic perhydropyrrolo[3,4-*c*]pyrrole scaffold) to explore possible advantages of organofluorine in such a π-electron-rich environment.[35] Figure 8 shows the crystal structure of the difluoromethyl-substituted inhibitor (±)-**25** at the active site of thrombin. While the overall binding geometry resembles that of the isopropyl derivative (±)-**1** (Figure 1), an overlay shows that the difluoromethyl residue in (±)-**25** orients further away from the π-electron-rich indole ring of Trp60D. In most cases, the biological results confirm the expected excellent selectivity against trypsin. Binding potency for both fluorinated and nonfluorinated derivatives is in the same range, varying between K_i = 0.005 and 0.078 μM. The weakest binding (0.078 μM) is actually observed for the *n*-propyl derivative, which is too large

to fit into the P pocket in its energetically more favorable *anti* conformation. According to the modeling, the C(ring)−CH_2−CH_2−CH_3 fragment adopts an energetically less favorable *gauche* conformation which could explain the observed increase in the K_i value. The study confirms that the introduction of fluoroalkyl residues into tight lipophilic pockets partially lined by aromatic rings neither increases nor decreases binding affinity substantially, when compared to similarly sized alkyl and alkenyl residues. However, taking into account the frequently advantageous effects on physicochemical properties, an overall benefit may result in many lead optimization projects from the decoration of ligands with fluoroalkyl residues to occupy apolar pockets.

Factor Xa: Cation−π Interactions in Aromatic Boxes

A true highlight in molecular recognition studies with synthetic cyclophane receptors has been the discovery of cation−π interactions by Dougherty and co-workers.[6,38,39] These interactions provide one of the strongest driving forces for biological complexation and, hence, are ubiquitous in Nature. In an attempt to establish cation−π interactions with the indole ring of Trp215 in the D pocket of thrombin, we prepared the tricylic imide ligand (±)-**28**, bearing a quaternary ammonium ion and, as a control, (±)-**29** with a *tert*-butyl residue (Figure 9). The biological assays showed that the *tert*-butyl-substituted inhibitor (±)-**29** formed the more stable complex ($\Delta\Delta G = 10.1$ kJ mol^{-1}) with thrombin as compared to (±)-**28**.[40] Presumably, the costs for desolvation of the cationic inhibitor are too large and cannot be compensated by the interactions with a single indole ring. On the other hand, factor Xa, another serine protease from the blood coagulation cascade[41] and hot target in the treatment of thrombotic diseases,[42] bound the onium ion ligand (±)-**28** better than the alkyl derivative (±)-**29** ($\Delta\Delta G = 11.5$ kJ mol^{-1}). The X-ray crystal structure of (±)-**28** complexed to factor Xa revealed that the onium ion is embedded in an aromatic box formed by the side chains of Phe174, Tyr99, and Trp215 in the S4 pocket of factor Xa. This aromatic box is a very effective onium binding site; contributions from ion pairing with carboxylate side chains of the protein are negligible because the closest ones (Glu217, Glu97) are nearly 10 Å away from the N^+ center. The large free enthalpy increment of $\Delta\Delta G =$ 11.5 kJ mol^{-1}, favoring (±)-**28** over (±)-**29**, shows that cation−π interactions are among the most potent bonding interactions in biology.

PDB searches subsequently revealed the more general occurrence of similar cation bonding sites in biological systems.[40] Ideally, the distances between the N^+ center and the C atoms of surrounding aromatic rings are between 4.0 and 5.0 Å and the cationic center is close to the normals passing through the centroids of the aromatic planes. Such aromatic boxes cannot be properly filled by quaternary ammonium ions only. Rather, because one side of the box is open, protonated tertiary amines can bind with their three alkyl residues in the box, while the N^+-H residue, requiring more efficient solvation, interacts directly or H_2O-mediated with H-bond acceptor residues of the protein. In our searches, we did not so far find protonated secondary ammonium ions embedded in such aromatic boxes: presumably the cost for desolvation of N^+-H residues is too high. Following these initial findings, a comprehensive investigation of cation−π interactions at the S4 pocket of factor Xa is now ongoing.

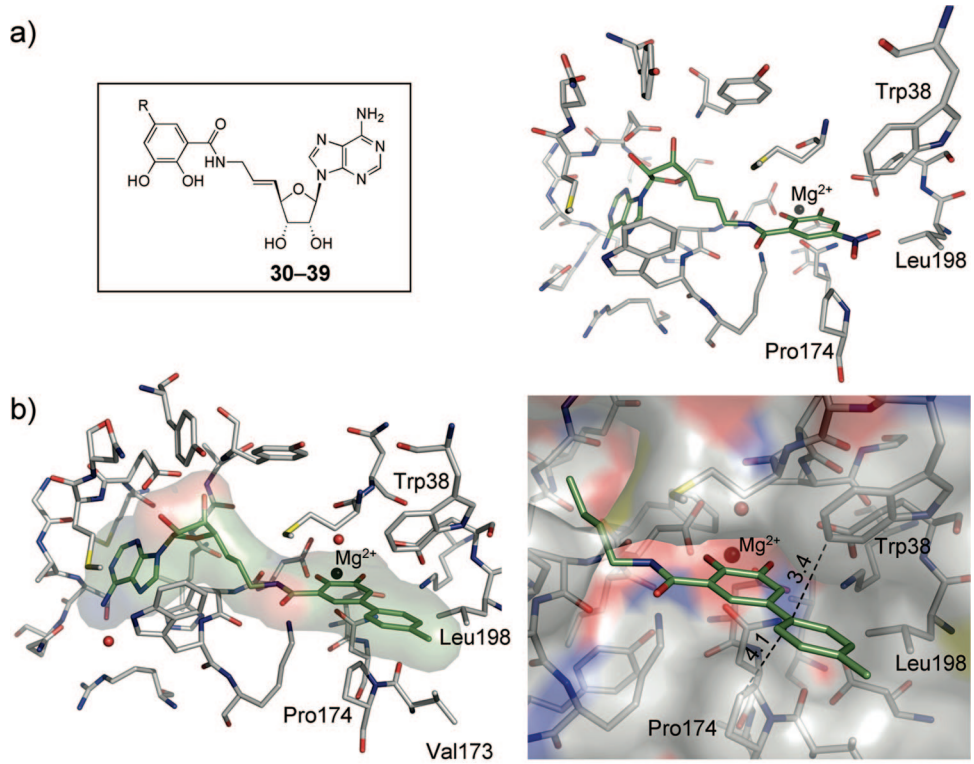
Catechol *O*-Methyltransferase (COMT): Moving from Mono- to Bisubstrate Inhibition Enables Full Exploitation of an Unexplored Hydrophobic Pocket

COMT catalyzes, in the presence of Mg^{2+} ions, the *O*-methylation of biologically active catechols, such as dopamine or its precursor L-Dopa, by *S*-adenosylmethionine (SAM).[43] Nitro-substituted catechols are potent COMT inhibitors and are therapeutic adjuncts to the L-Dopa-based treatment of Parkinson's disease, reducing the peripheral metabolic degradation of the dopamine precursor.[44] Nitro groups have for a long time been regarded as indispensable for tight and reversible binding to the substrate pocket of the active site: they enhance the binding of the catechol to the Mg^{2+} ion and reduce the nucleophilicity of the catechol HO groups, thereby greatly reducing substrate behavior.[45]

The latter issue became irrelevant when we developed potent bisubstrate inhibitors such as (−)-**30** (Figure 10).[46] Kinetic studies and X-ray analysis showed that this compound ($IC_{50} = 9$ nM, $K_i = 28$ nM) blocks both the SAM and catechol binding sites of COMT. Correct length and rigidity of the linker between adenosine and catechol moieties are essential for high binding efficacy.[47] Upon changing the linker from $-CH_2CH{=}CH-$ (in (−)-**30**) to $-CH_2CH_2CH_2-$ and $-CH_2CH_2OCH_2-$, the IC_{50} values increased from 9 nM, to 200 nM, and to 2 μM, respectively. A three-atom linker allows both efficient H bonding of adenine to the rigid protein backbone and catechol coordination to the Mg^{2+} ion. Furthermore, the introduction of the double bond is crucial to prevent hydrophobic collapse of the free inhibitor in water—with the catechol and nucleobase moieties approaching each other—which would reduce the binding free enthalpy.

A detailed examination of the protein revealed a previously unexplored hydrophobic pocket at the enzyme surface, located in proximity of the nitro group of (−)-**30** and shaped by the amino acid residues Trp38, Leu198, Val173, and Pro174 of COMT.[48,49] Computer modeling suggested that suitable substituents replacing the NO_2 group in (−)-**30** (position 5 of the catechol nucleus) could undergo favorable apolar interactions with this pocket (Figure 10b).

This strategy led indeed to a new generation of highly active bisubstrate inhibitors that eliminate the need for NO_2 substitution of the catechol structural unit. In many cases, replacement of the NO_2 group afforded potent bisubstrate inhibitors with activities in the low double-digit nanomolar range. Compounds **31**−**39** were shown to be competitive inhibitors with respect to the cofactor binding site. There is no correlation between the free enthalpy of inhibition and pK_a values or Hammett parameters (Figure 10c). This finding is in sharp contrast to the interdependence of biological activities and catechol pK_a values in the case of monosubstrate inhibitors blocking only the catechol binding site.[45] Many of the inhibitors actually exhibit high pK_a values of the catechol HO group, but at the same time display high inhibitory activity. Examples are the 4-fluorophenyl- and 4-methylphenyl-substituted inhibitors (+)-**31** and (−)-**32** with pK_a values of 6.87 and 7.06, respectively, and K_i values of 55 nM. In the absence of cocrystal structure information, these data provide strong support for the gain in apolar interactions as a result of filling and desolvating the hydrophobic pocket at the enzyme surface (Figure 10b). Gains in binding free enthalpy do not simply result from more favorable partitioning (as estimated from *c* log *P* data); rather, proper



c)

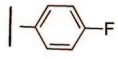
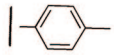
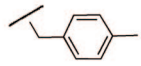
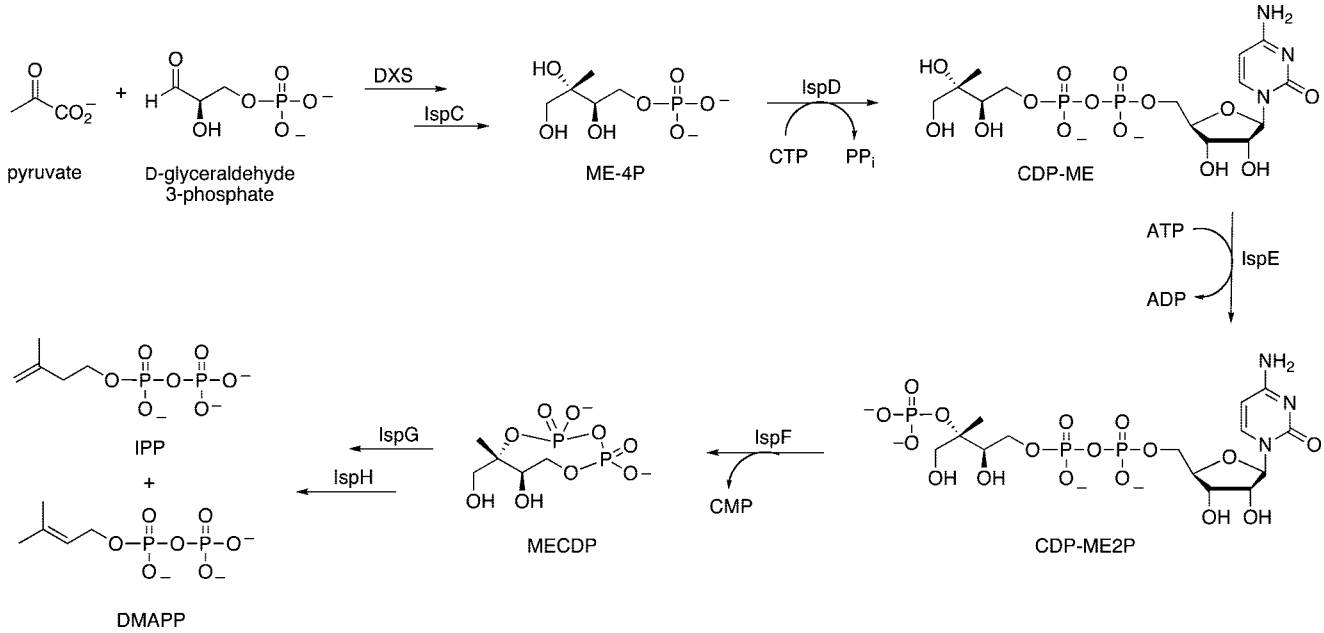
	R	IC_{50} [nM]	K_i [nM]	$-\Delta G_{inh}$ [kJ mol^{-1}]	pK_a	*c* log *P*	σ_p
(–)-**30**	NO_2	9	28 ± 2	44.8 ± 0.2	4.42	2.05	0.78
(+)-**31**		21	55 ± 5	43.1 ± 0.2	6.87	2.70	0.06
(–)-**32**		23	55 ± 3	43.1 ± 0.1	7.06	3.23	–0.03
(–)-**33**		23	55 ± 2	43.1 ± 0.1	6.89	1.38	0.23
(–)-**34**		27	60 ± 5	42.9 ± 0.2	6.15	3.12	–
(–)-**35**	Br	28	80 ± 5	42.1 ± 0.2	6.54	2.70	0.23
(–)-**36**		608	211 ± 4	39.6 ± 0.1	7.56	3.51	–0.09
(+)-**37**		1370	546 ± 9	37.6 ± 0.1	7.63	2.40	–0.15
(–)-**38**		2000	511 ± 36	37.3 ± 0.2	6.25	0.31	0.36
(–)-**39**	H	2600	1644 ±144	34.3 ± 0.2	7.37	1.06	0

FIGURE 10. (a) General structure of compounds **30**−**39** and X-ray crystal structure of (−)-**30** in complex with COMT (PDB code 1JR4, resolution 2.63 Å).[46b] (b) MOLOC model of bisubstrate inhibitor (−)-**32** complexed at the active site of COMT in the presence of a Mg^{2+} ion (left) and detailed view of the catechol binding site (right). The model was obtained by starting from the crystal structure shown in (a). The Connolly surface of the inhibitor (left) and the enzyme (right), respectively, is shown. Water molecules are represented as red balls. (c) Structure and properties of compounds **30**−**39**. σ_p = Hammett substituent parameter.

stereoelectronic complementarity to the newly exploited pocket is a prerequisite for such gains: 4-methylphenyl (in (−)-**32**) is a much better substituent than 4-methylbenzyl (in (−)-**36**). An isopropyl group (in (+)-**37**) is too small for efficient interactions and desolvation, whereas the dimethycarbamoyl group (in (−)-**38**) has the wrong polarity.

Thus, potent inhibition could for the first time be achieved, despite the lack of a NO_2 substituent on the catechol moiety, by exploiting favorable apolar interactions with a hydrophobic cleft-type subpocket near the enzyme surface. This result validates in an impressive way the bisubstrate inhibition approach on the way to new inhibitors of COMT.

SCHEME 1. Nonmevalonate Pathway for the Biosynthesis of the C_5 Precursors to Isoprenoids IPP and DMAPP Used by *P. falciparum*[a]



[a] Key: ADP = adenosine 5′-diphosphate; ATP = adenosine 5′-triphosphate; CDP-ME = 4-diphosphocytidyl-2*C*-methyl-D-erythritol; CDM-ME2P = 4-diphosphocytidyl-2*C*-methyl-D-erythritol-2-phosphate; DXS = 1-deoxy-D-xylulose 5-phosphate synthase (EC 2.2.1.7); IspC = 1-deoxy-D-xylulose-5-phosphate reductoisomerase (EC 2.7.7.60); IspG = 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate reductase (EC 1.17.4.3); IspH = 1-hydroxy-2-methyl-2-(*E*)-butenyl-4-diphosphate reductase (EC 1.17.1.2); MECDP = 2*C*-methyl-D-erythritol 2,4-cyclodiphosphate.

Preorganization of a Pocket Determines the Binding Free Enthalpy Gained From its Occupation: Enzymes of the Nonmevalonate Pathway for Isoprenoid Biosynthesis

In our program on new antimalarials, we have targeted two of the seven enzymes of the nonmevalonate pathway for isoprenoid biosynthesis, which was discovered in the early 1990s[50] and transforms pyruvate and glyceraldehyde 3-phosphate into the isoprenoid precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) (Scheme 1). This pathway is used exclusively by parasites such as *Plasmodium falciparum*, the causative agent of severe forms of malaria, and *Mycobacterium tuberculosis*. Mammals, on the other hand, only use the alternative mevalonate pathway. The pathway has been fully validated as antimalarial target by the inhibition of the second enzyme (IspC) with a small phosphonate, fosmidomycin, which was shown to cure malaria in rodents and has been tested in clinical trials.[51] When X-ray crystal structures became available, we applied our structure-based design approach to the development of inhibitors of the fourth and fifth enzymes in the pathway, IspE (4-diphosphocytidyl-2*C*-methyl-D-erythritol kinase, EC 2.7.1.148) and IspF (2*C*-methyl-D-erythritol 2,4-cyclodiphosphate synthase, EC 4.6.1.12).

X-ray crystal structures of IspF, which exists mostly as a symmetric homotrimer, show active sites located at each interface between adjacent subunits.[52,53] Each of the topologically equivalent active sites possesses two pockets (Figure 11). The rigid, well-conserved "Pocket III" of one monomer binds the cytidine moiety of the substrate CDP-ME2P. The larger, more flexible "Pocket II" in an adjacent monomer hosts the phosphate and erythritol moieties. In addition, "Pocket II" also contains one or two divalent metal ions (Mg^{2+}, Mn^{2+}, or Zn^{2+}), which participate in the binding of the diphosphate moiety of the substrate as well as in catalysis.

Our first inhibitors of *E. coli* IspF, (−)-**40** and (−)-**41**, closely resembled substrate CDP-ME2P, featuring a cytosine diphosphate moiety attached to a fluorescent anthranilate or dansylamide residue to reach into the hydrophobic region of "Pocket II".[54] These compounds were developed as fluorescent probes for assay development, and their binding mode (with K_d values in the lower micromolar and IC_{50} values in the low millimolar range) was elucidated by X-ray crystallography (PDB codes 2AMT and 2AO4).[54] In agreement with other X-ray crystal structures of IspF, the hydrophobic region of "Pocket II" is highly flexible and its conformation differs strongly from crystal to crystal. Correspondingly, its occupation by the fluorescent chromophores in (−)-**40** and (−)-**41** does not yield much better inhibitory power: the IC_{50} value of the dansyl derivative (−)-**41** (3.0 mM) is similar to that of CDP (7.3 mM), which does not occupy this pocket. Similarly, in another family of more druglike inhibitors, the binding potency gained from targeting "Pocket II" was disappointingly low.[55] The large flexibility of the hydrophobic part of "Pocket II", which is lined by Phe68′ and Leu76′ among others, limits the binding free enthalpy that can be gained through its occupancy. Upon occupation by a ligand fragment, good apolar interactions may well be established, however, at the (entropic) cost of freezing out one out of an entire ensemble of conformations of the protein. In the case of IspF, an appropriate residue for occupancy of the hydrophobic region of "Pocket II" still remains to be identified and the demonstrated conformational flexibility of this region of the protein makes IspF a "tough target".

A much more favorable case was presented by the kinase IspE from *E. coli*, where the discovery and appropriate occupancy of a small, highly preorganized hydrophobic pocket led to nanomolar activities (K_i values) for the first-generation ligands.[56] Figure 12a shows IspE in the ternary complex with CDP-ME and 5′-adenyl-β,γ-amidotriphosphate, a nonhydrolizable ATP analogue (PDB code 1OJ4).[57] The active site features three pockets: the adenosine, the cytidine, and the ME/phosphate pocket. To gain selectivity over other, in particular human,

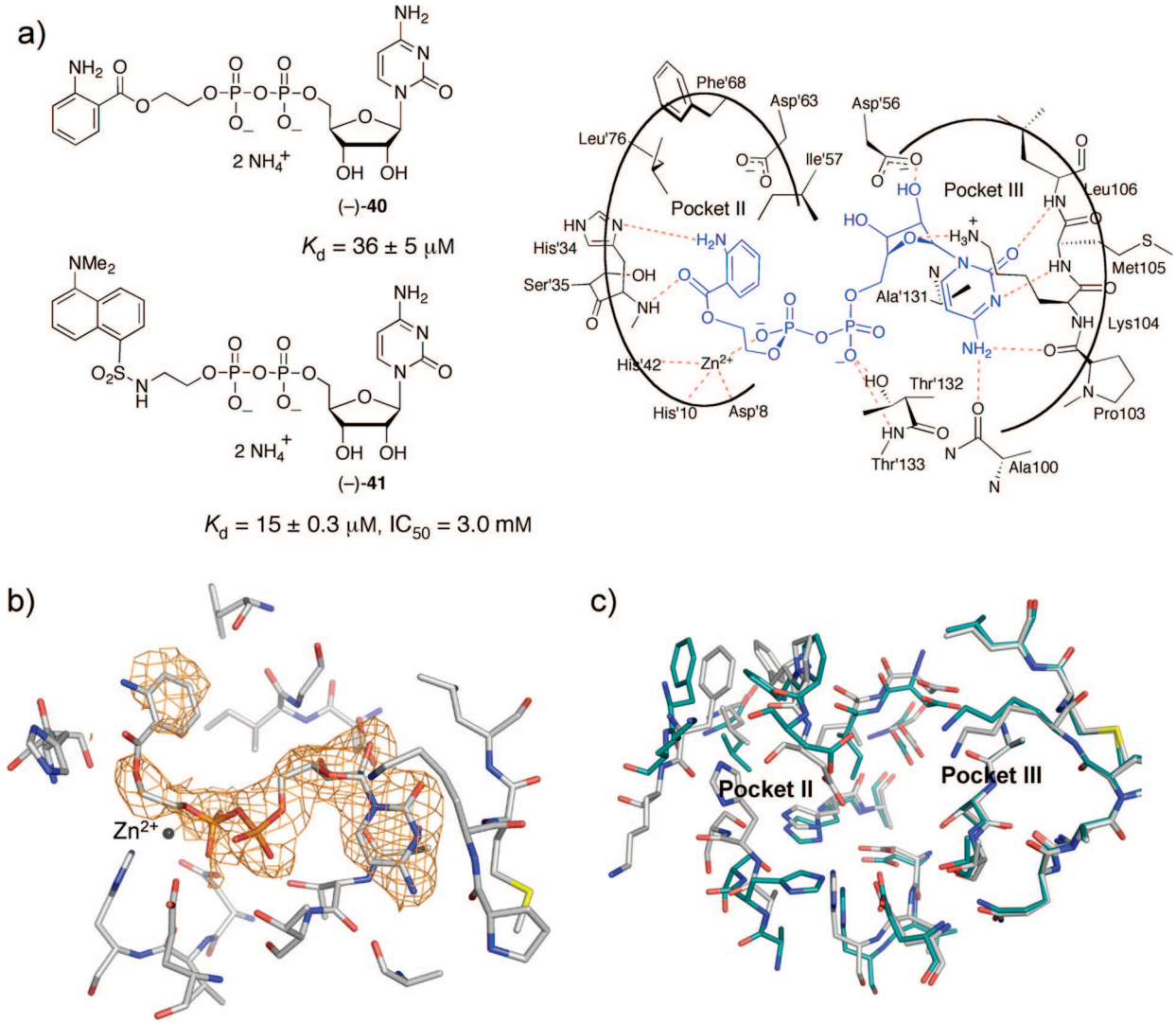


FIGURE 11. (a) First-generation substrate-like inhibitors for fluorescent assay development and schematic of the binding mode of the anthranilate derivative (−)-**40**. (b) Binding mode of the anthranilate derivative (−)-**40** in the ternary complex of IspF with Zn^{2+} as revealed by X-ray crystallography (PDB code 2AMT, resolution 2.30 Å).[54] The difference electron density of the compound is shown. Five of the six independent active sites of IspF show the depicted binding mode. (c) Overlay of two X-ray crystal structures of IspF (PDB codes 1GX1, resolution 1.80 Å (green)[53a] and 1JY8, resolution 2.50 Å (gray)[53b]) revealing the highly flexible "Pocket II" and the well-conserved "Pocket III".

kinases, we decided not to target the ATP, but rather the cytidine pocket. Molecular modeling revealed that the cytidine binding site features an additional, small hydrophobic subpocket, lined by Leu15, Leu28, and Phe185. The design of the first-generation ligands is illustrated in Figure 12b. Conserving a cytosine moiety as central scaffold, a tetrahydrothiophene ring was chosen (among others) as ribose substitute, sandwiching between Tyr25 and Pro182. According to the modeling, favorable S–aromatic interactions[6] with the phenolic ring of Tyr25 are established in both diastereoisomeric complexes formed by (±)-**42**, which are expected to be of similar stability. A propargylic linker attached to an alkyl sulfonamide in its favorable staggered conformation,[58] with the N-lone pair bisecting the OSO moiety, orientates an alkyl residue into the newly discovered hydrophobic pocket that is not occupied by the substrate.

The results obtained with the first-generation ligands (±)-**42**−(±)-**47** are highly encouraging. The best ligands, featuring a cyclopropyl ((±)-**42**) or CH_2CF_3 ((±)-**43**) moiety to optimally fill the narrow hydrophobic subpocket, gave inhibitory constants K_i in the upper nanomolar range (Figure 12d). The proposed binding mode is strongly supported by SAR data and was recently largely confirmed by cocrystal structure analysis.[59] When the residue filling the subpocket becomes too small (e.g., Me in (±)-**45**), binding is weakened; when the sulfonamide substituent becomes too large, as in the case of phenyl in (±)-**47**, binding becomes further weakened. We assume that such large residues presumably turn away from the pocket, in the opposite direction toward solvent.[60]

The results obtained with IspE and IspF illustrate a general finding in drug discovery research: proper occupation of preorganized subpockets yields a large gain in binding efficacy, which is not the case when a pocket is conformationally highly flexible. Knowledge about the flexibility of a pocket-forming region of a protein is thus important for structure-based design, and it is worthwhile testing the flexibility of a pocket and its suitability for energy-gaining occupation. This can either be done by molecular dynamics simulations or, as in the cases reported here, by comparing through overlaying a larger set of X-ray crystal structures.

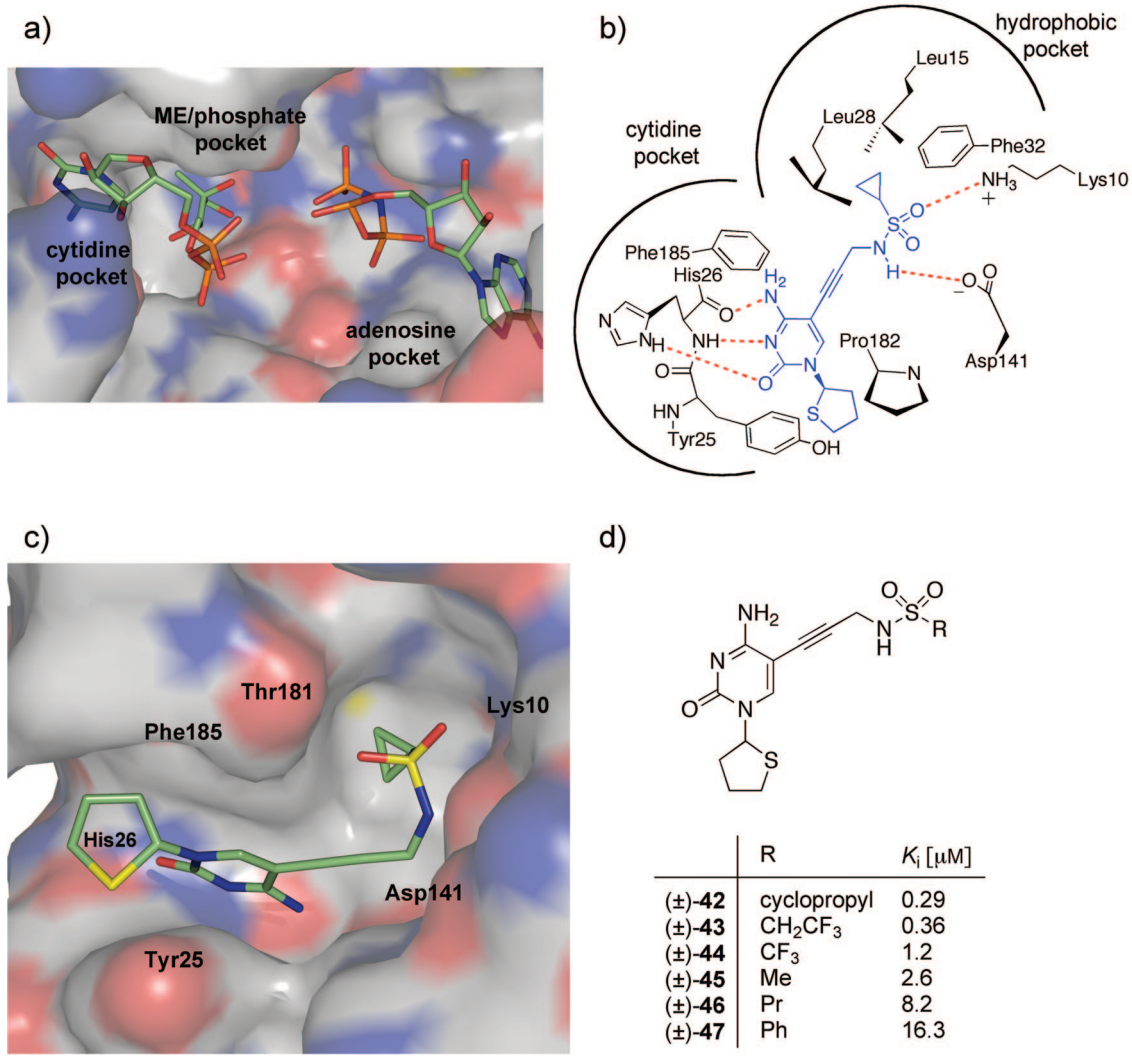


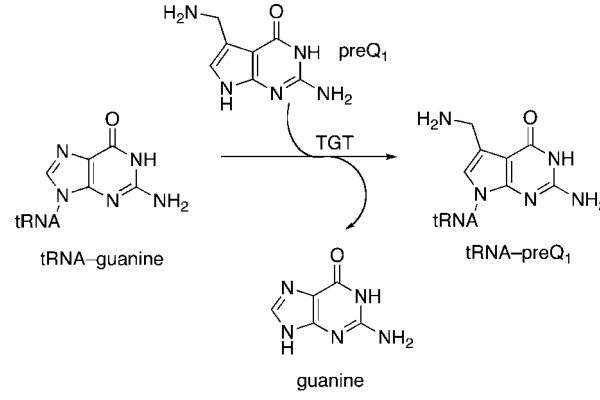
FIGURE 12. (a) Active site of *E. coli* IspE in the ternary complex with CDP-ME and an ATP analogue (PDB code 1OJ4, resolution 2.01 Å).[57b] (b) Design and proposed binding mode of the first-generation inhibitor (±)-**42** for IspE. (c) MOLOC model illustrating the proposed binding of the cyclopropyl moiety into the hydrophobic subpocket at the cytidine site of *E. coli* IspE. The Connelly surface of IspE is shown. (d) Activities of the new ligands with an SAR in agreement with the proposed binding of suitably sized residues R in the small subpocket.

tRNA-Guanine Transglycosylase: Penalty for Disrupting a Water Network on the Way into a Lipophilic Pocket

tRNA-guanine transglycosylase (TGT, EC 2.4.2.29) is a tRNA-modifying bacterial enzyme that catalyzes the exchange of guanine in position 34 by preQ$_1$ (Scheme 2).[61] As the enzyme is involved in the infection pathway of *Shigella*, the causative agent of dysentery (shigellosis), we have pursued the structure-based development of potent inhibitors for TGT.[62] Although the first X-ray crystal structure of the enzyme has been available for some time,[63] the detailed mechanism of action has only recently been elucidated in elegant structural work.[64] Two Asp side chains are involved in catalysis: Asp280 acts as a nucleophile and attacks tRNA, displacing guanine 34 under formation of a covalent intermediate. This intermediate is subsequently attacked by preQ$_1$, which is now incorporated at position 34 of the tRNA strand. All steps benefit from acid−base catalysis provided by Asp102. Figure 13 shows the X-ray crystal structure of TGT-bound tRNA after introduction of preQ$_1$.[64]

Our design approach involved the identification of a suitable nucleobase to bind into the guanine34/preQ$_1$ site, followed by the introduction of a vector that extends into the tRNA channel

SCHEME 2. Exchange of Guanine at Position 34 of Bacterial tRNA by preQ$_1$, Catalyzed by TGT



to competitively inhibit nucleic acid binding. Both 2,6-diaminoquinazolin-4(3*H*)-one (**48**)[65] and *lin*-benzoguanine (**49**)[66] were found to be suitable ligands to occupy the nucleobase site, with affinities (competitive inhibition constants K_i) in the micromolar range. The cocrystal structures of TGT (from *Zymomonas*

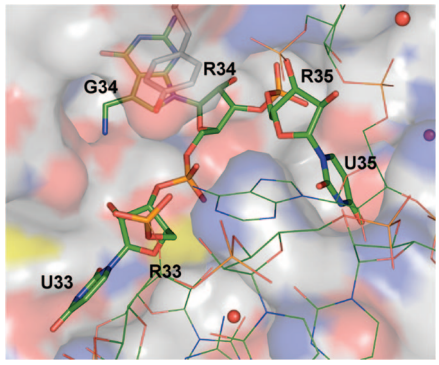


FIGURE 13. X-ray crystal structure of TGT-bound tRNA after introduction of preQ$_1$ (PDB code 1Q2S, resolution 3.20 Å).[64] Water molecules are represented as red balls. P atoms: orange.

mobiliz) bound to **48** and **49** (Figure 14a)[67] nicely reveal the sandwiching of the nucleobase between a Met and a Tyr side chain as well as the two catalytic Asp residues (Asp280 and Asp102) that are solvated by a crystallographically defined water network. We subsequently introduced substituents to fill a shallow hydrophobic pocket occupied by ribose34 and lined by Val282, Leu68, and Val45.[65] Figure 14b depicts two out of a larger series of crystal structures that confirmed that lipophilic residues, such as the propyl thioether moiety in **50** or the phenethyl moiety in **51**, were properly directed into this subpocket.[67] Biological assays revealed, however, that the binding efficacy of the inhibitors had not been increased and that the competitive inhibition constants K_i remained in the low micromolar range. Extensive crystallographic investigations showed that the lipophilic vector disrupted the highly conserved water network solvating the two catalytic Asp280 and Asp102 residues.[67a] Therefore, energetic gains from occupation of the hydrophobic pocket were entirely compensated by losses caused by the unfavorable disruption of this network and the violation of the high solvation requirements of the two catalytic Asp side chains.

We subsequently decided to direct a vector into the pocket of ribose33 (Figure 13), which could be elegantly accomplished by introducing side chains into position 2 of the *lin*-benzoguanine scaffold.[69] Gratifyingly, a series of inhibitors, such as **52**−**57**, with K_i values down to the single-digit nanomolar range were obtained. The crystal structure of the complexed naphthyl derivative **55** is shown in Figure 15, confirming the proper orientation of the substituent toward the ribose33 pocket. There are several reasons for the observed jump in activity:

(i) Binding of the new inhibitors leaves the solvation of the side chains of the catalytic Asp280 and Asp102 residues intact.

(ii) Complexation strength increases markedly upon introduction of a 2-amino group into the *lin*-benzoguanine core. The pK_a value of the resulting 2-aminoimidazole ring, in particular in the best ligand **57**, is increased, and we propose that the resulting 2-aminoimidazolium moiety participates in a charge-assisted hydrogen-bonding network with two ionic bonds ($d(O^-_{Glu235}\cdots HN_{Ala232}) = 2.9$ Å (distance between heavy atoms) and $d(C{=}O_{Leu231}\cdots HN^+_{aminoimidazole}) = 3.0$ Å, Figure 15c). A similar charge-assisted hydrogen-bonding network is seen in the complex of TGT with preQ$_1$, involving the basic aminomethyl residue of the nucleobase.[70] A minor energetic contribution could also result from a long hydrogen bond between the carbonyl group of Ala232 and the 2-amino group.

(iii) In the cocrystal structures of TGT with **54**−**57**, the 2-substituents reaching into the ribose33 pocket are not detected in the difference electron density maps (Figure 15b). Apparently, no strong directional interactions are established in the pocket. Nevertheless, binding free enthalpy should be gained by desolvation of the 2-substituents and the uracil33 binding pocket. Enthalpic gains in this loose binding might not be large; on the other hand, the favorable desolvation entropy should not be compensated from a substantial unfavorable entropic term resulting from the reduction in the mobility of the substituent and its environment. As one of the take-home messages of this investigation, modeling-based optimizing for the best host−guest fit with the tightest nonrepulsive van der Waals contacts might not be the only strategy to gain maximum binding efficacy from filling a hydrophobic pocket. As a result of enthalpy−entropy compensation, more loosely bound ligand fragments may provide similar complexation strength.

The question of how to deal with crystallographically defined water clusters is of general importance in structure-based drug lead design and optimization, and a careful examination of active site hydration patterns is strongly advised.[71] This study suggests that it is quite challenging to reach and occupy more remote pockets with a gain in binding free enthalpy, if high costs of desolvation are involved on the way to the pocket. The disruption of water clusters, such as the network solvating the two highly polar Asp side chains in TGT, can be energetically too costly. We are now exploring more polar linkers that could fully or partially restitute lost solvation when connecting the nucleobase scaffold to a residue for occupation of the small pocket near the ribose34 site, and hoping to harvest binding free enthalpy from apolar interactions in this pocket.

Optimal Filling of the Flap Pocket of the Antimalarial Target Plasmepsin II: An Example for the "55% Rule"

The plasmepsins (PMs), three aspartic proteases (PM I, II, IV), and a histo-aspartic protease (HAP) are used by *P. falciparum* in the stepwise degradation of human hemoglobin, ultimately providing the amino acids to fuel the exponential growth of the malaria parasite during the erythrocyte stage of its lifecycle.[72] Due to overlapping activity, all of the mentioned enzymes need to be targeted by a potential antimalarial drug.[73] Crystal structure information is available for PM II and PM IV: bound to peptidomimetics such as pepstatin A, the enzymes show a pepsin-like fold, which is also observed in all other published X-ray crystal structures.[74] However, three of the PM II structures (PDB codes 2BJU,[75a] 2IGX,[75b] 2IGY[75b]) feature a new cavity, the so-called flap pocket. This pocket is opened ("flap-open" conformation) and shaped by suitable nonpeptidic inhibitors, a process first observed for renin, another−human−aspartic protease.[76] The active site of PM II is schematically shown in Figure 16a. The potent inhibitors (±)-**58** and (±)-**59** (Figure 16b) address the catalytic Asp dyad with the protonated azanorbornane needle, while the naphthyl and bromophenyl residues occupy the S1/S3 site, and the hexyl chain opens and occupies the flap pocket.[77] The compounds are specific for the plasmepsins and bind only weakly to the human cathepsins (hCat) D and E, enzymes used to degrade disfunctional human hemoglobin. Preliminary SARs revealed that the occupation of the flap pocket by the alkyl chain is essential for high potency: if the *n*-hexyl chain in (±)-**58** is exchanged for a *p*-Br substituent, binding affinity drops dramatically to IC$_{50}$ values of 15 μM (PM II) and 17 μM (PM IV). The selectivity against

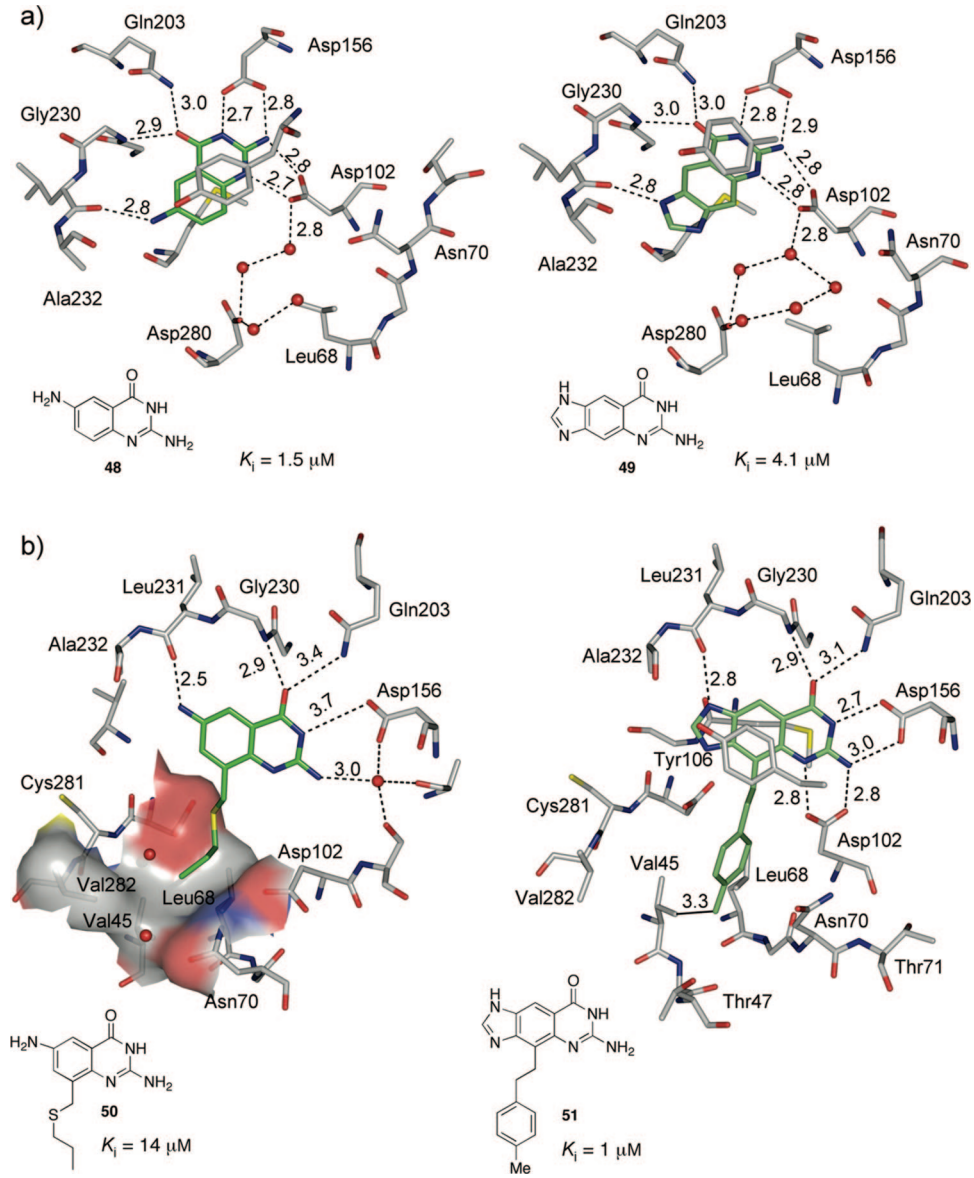
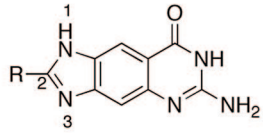


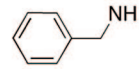
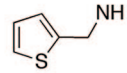
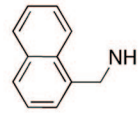
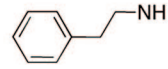
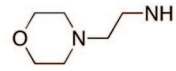
FIGURE 14. (a) Cocrystal structures of TGT (from *Zymomonas mobiliz*) bound to the unsubstituted nucleobase analogues **48** (left, PDB code 1R5Y, resolution 1.2 Å)[67a] and **49** (right, PDB code 2BBF, resolution 1.7 Å).[67b] (b) Cocrystal structures of TGT bound to inhibitors **50** (left, PDB code 1K4H, resolution 1.8 Å)[68] and **51** (right, PDB code 1Y5W, resolution 1.58 Å),[67b] featuring vectors to fill the shallow pocket lined by Val282, Leu68, and Val45. Water molecules are represented as red balls.

hCatD and hCatE also originates from the presence of the alkyl chain substituent: these human enzymes either do not have or do not easily reach a "flap-open" conformation for its accommodation. In view of these findings, we conducted a comprehensive study to optimize the occupancy of the flap pocket. While keeping the azanorbornane needle and the residue targeting the S1/S3 subsite constant, we systematically varied the length of the alkyl chain from *n*-butyl to *n*-undecyl and also introduced terminal cycloalkyl residues.[78] A striking SAR became evident for a series of inhibitors which target the flap pocket with a homologous series of *n*-alkyl chains (Figure 17a). Plotting the binding affinities for both PM II and PM IV against the chain length of the flap vector resulted in bell-shaped curves with broad maxima. Interestingly, the ideal chain length is different for PM II and PM IV. Not only does PM IV prefer shorter chains, the decrease in activity when exceeding optimal chain length is also much stronger for PM IV than for PM II. PM IV, for which no crystal structure in the "flap-open" conformation is known, seems to have a shorter flap pocket with a smaller binding volume compared to PM II.

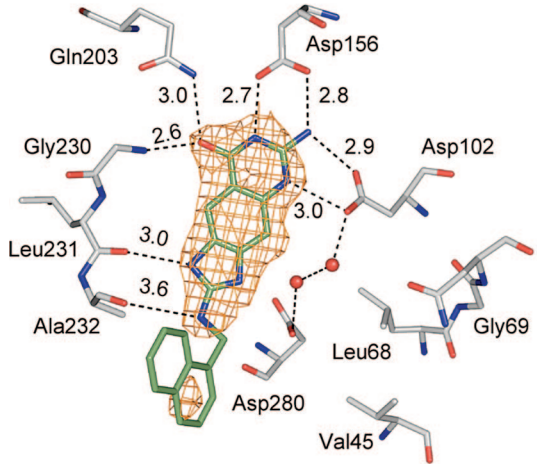
In this context, important concepts previously demonstrated in molecular recognition studies with synthetic receptors could be transferred and applied to an enzyme environment. Mecozzi and Rebek showed that inclusion complexes in confined cavities feature the highest stability when the guest occupies 55 ± 9% of the available space within a host.[79] This rule applies in particular to apolar binding processes.[80] If volume occupancy is lower than optimal, the number of van der Waals contacts is low and the enthalpic gain from dispersion interactions is reduced. If space occupancy is higher than optimal, a large number of van der Waals contacts is established resulting in a large enthalpic gain. At the same time, however, the dynamic mobility of both host and guest becomes reduced, which leads to an unfavorable complexation entropy that compensates the

a)



	R	K_i [nM]	pK_a	log *D*
49	H	4100±1000	5.2	–0.05
52	NH_2	77±12	6.3	–0.33
53		70±1	–	–
54		35±9	5.8	1.49
55		55±11	5.9	3.29
56		10±3	–	–
57		6±6	6.7	–0.31

b)



c)

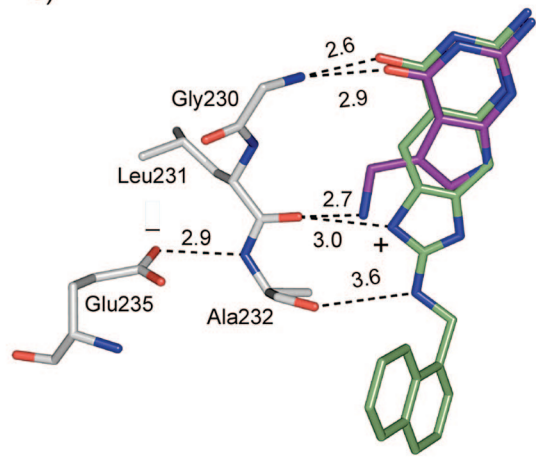


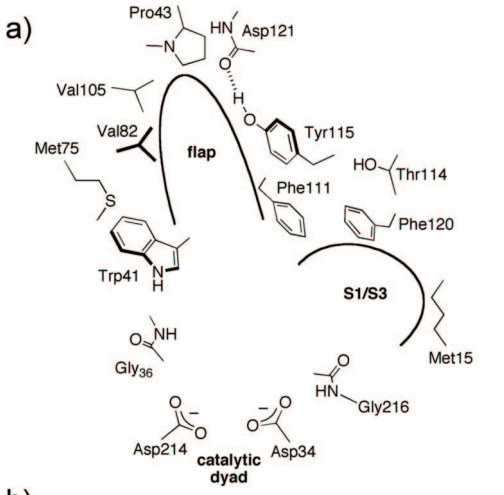
FIGURE 15. (a) Biological and physicochemical properties of TGT inhibitors **49**−**57**. (b) X-ray crystal structure of TGT-bound naphthyl derivative **55** (PDB code 2QZR, resolution 1.95 Å).[69] The difference electron density of the compound is shown together with the H-bonding contacts between the protein and the ligand. (c) Overlay of **55** (green) with preQ$_1$ (purple) bound at the active site of TGT showing the charge-assisted H-bonding interaction between protein and ligands. Water molecules are represented as red balls.

enthalpic gains. We determined the volume of the flap pocket by two methods, based on available crystal structures,[75] and found that this rule holds very well for the binding of alkyl and (cycloalkyl)alkyl substituents in the flap pocket of PM II (Figure 17b).

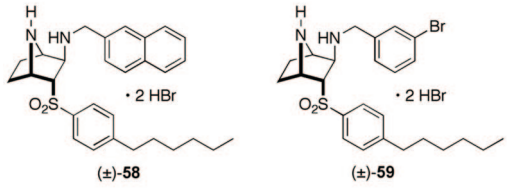
Furthermore, Rebek and co-workers observed in studies with synthetic receptors that alkyl chains, apparently too long in their fully extended conformation for a certain cavity, tend to adopt energetically less favorable conformations with a number of *gauche* dihedral angles in order to fit into the available space.[81] Driving forces for this guest contortion upon transferring an *n*-alkane from water into a confined apolar cavity are the establishment of chemical complementarity, better burial of hydrophobic surfaces, and proper filling of space in accord with the "55% rule".

In the case of the flap pocket of PM II, the *n*-heptyl substituent is the longest, according to computer modeling, that fits in an all-*anti* conformation (Figure 18). Ligands with longer *n*-octyl and *n*-nonyl chains, however, give similar IC_{50} values. According to the modeling, the longer alkyl chains have to reduce their length by folding to fit the flap pocket. We conclude that long *n*-alkyl chains in the flap pocket of PM II seem to become readily contorted, if this is required to fit inside the pocket and if volume occupation in the resulting complex is in the preferred range.

This investigation suggests that volume considerations provide valuable information whether a lipophilic enzyme or receptor pocket is optimally filled. Despite the often not as straightforward definition of the volume of an enzyme pocket, we foresee



b)



	IC_{50} [nM]			
	PM II	PM IV	hCat D	hCat E
(±)-**58**	130	50	2030	7050
(±)-**59**	210	30	2800	5200
(−)-**59**	45	10	n.d.	n.d.
(+)-**59**	3260	33900	n.d.	n.d.

c)

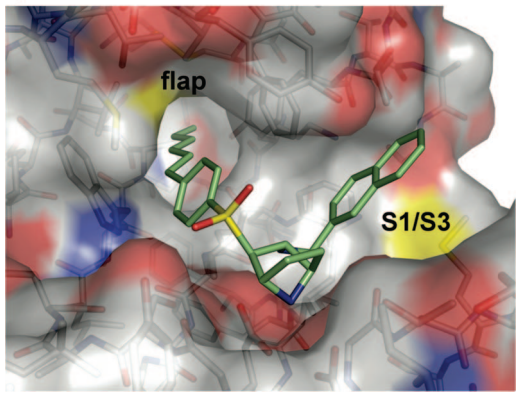
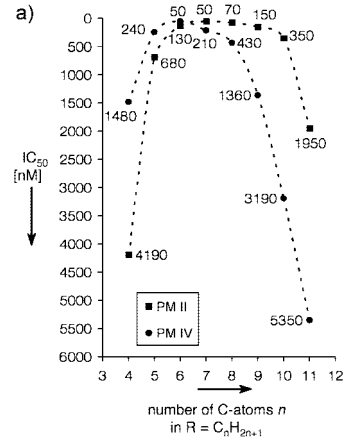
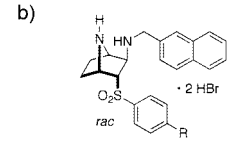


FIGURE 16. (a) Schematic representation of the active site of PM II in the "flap-open" conformation. (b) Biological activities of designed inhibitors for the plasmepsins. Based on the large difference in stability of diastereoisomeric complexes formed by (−)-**59** and (+)-**59**, the absolute configuration of the active enantiomer was assigned by molecular modeling. (c) MOLOC model of the active enantiomer of inhibitor (±)-**58** bound to the "flap-open" conformation at the active site of PM II. The crystal structure with the PDB code 2BJU (resolution 1.56 Å) was used for the docking experiment.[75a]



b)



R	PC [%]	IC_{50} [nM]
n-butyl	30	4190
n-pentyl	36	680
n-hexyl	43	130
n-heptyl	50	50
4-(isopropyl)butyl	50	30
3-cyclopentylpropyl	52	30
4-cyclobutylbutyl	53	170
n-octyl	57	70
4-cyclopentylbutyl	59	190
n-nonyl	63	150
4-cyclohexylbutyl	65	380
n-decyl	68	350
n-undecyl	75	1950

FIGURE 17. (a) Striking SAR becomes evident when plotting the IC_{50} values against the number of C atoms in inhibitors featuring an *n*-alkyl chain as flap vector R (cf. structure in (b)). (b) General structure of the discussed PM inhibitors, inhibitory constants, and packing coefficients (PC, PC = *V*(R)/*V*(flap pocket) × 100%).

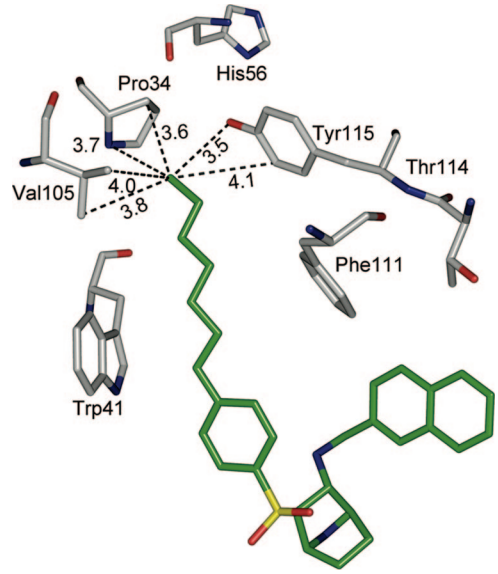


FIGURE 18. According to modeling, the longest *n*-alkyl chain that fits into the flap pocket in the all-*anti* conformation is *n*-heptyl. Distances between heavy atoms of the inhibitor and the amino acids are indicated as dashed lines.

growing importance of volume analysis as a tool in structure-based design.

Summary

This Perspective describes our experiences in identifying and properly filling apolar subpockets of enzyme active sites with the purpose of enhancing ligand binding affinity and selectivity. It is worthwhile to examine an enzyme active site from all possible angles during modeling, unbiased from the structure of the natural substrates, to identify suitable pockets to be occupied by ligand parts. This way, we discovered the small hydrophobic pocket near the cytidine binding site of IspE, which could be filled with large gains in binding free enthalpy. We presume that some of the preformed pockets at active sites, which are not occupied by the substrate, are used by downstream products in a pathway for feedback regulation purposes. It is also advisable to examine the degree of rigidity and preorganization of a pocket that correlates with the binding free enthalpy that can be gained. Molecular dynamics simulations or overlays of protein crystal structures, if available, help to evaluate the conformational preorganization of a pocket. Thus, the overlay of the ten currently available crystal structures of IspF shows that the hydrophobic region of "Pocket II" is highly flexible

and that a large ensemble of conformations can be populated. Correspondingly, the occupancy by a ligand part, chosen for concave-convex complementarity to one of the "pocket conformers" does not make a large contribution to the overall binding affinity.

Aromatic boxes such as the S4 pocket of factor Xa accommodate protonated tertiary or quaternary ammonium ions with a very large gain in binding affinity; in fact cation−π interactions are clearly among the strongest driving forces for complexation in biology. The fluorine scan conducted for a class of highly preorganized inhibitors of thrombin has helped identifying favorable interactions of organofluorine such as orthogonal dipolar interactions with backbone C═O residues. Thus, organofluorine may not only be introduced into ligands to modulate physicochemical and metabolic properties but also to gain binding affinity and selectivity. It is gratifying to note that important principles for estimating the optimum volume occupancy of constrained apolar cavities in synthetic host−guest chemistry, such as the "55%-rule", can also be applied to biological complexation as shown in this article for a pocket of PM II. We foresee growing importance of volume analysis as a tool in structure-based design.

The identification and occupation of a preorganized subpocket however is not sufficient to gain binding affinity. Linkers connecting the pocket-filling vector to the central ligand scaffold can disrupt defined water networks, such as those solvating the two catalytic Asp side chains in TGT, in an energetically costly way. More research is needed to learn how to properly substitute such networks without energetic losses. Finally, enthalpy−entropy compensation seems to be a principle that governs pocket filling to a large extent: loose binding leads to reduced dispersion interactions, while tight binding enhances these interactions but at the costs of reduced mobility of the binding partners and corresponding losses of entropy. Unfortunately, thermodynamic data, properly correlated with structural analysis, such as those discussed from the Klebe laboratory,[26] are missing for most protein−ligand interactions. Therefore, a number of broadly accepted concepts and hypotheses, which are also used in this article, need to be further validated. Deciphering of biological molecular recognition principles at atomic resolution remains a fascinating field of research that ultimately enables more efficient and more predictable small-molecule drug developments.

Acknowledgment. This work was generously supported by the ETH Research Council, F. Hoffmann-La Roche Ltd, Chugai Pharmaceuticals, and the Roche Research Foundation (Ph.D. fellowship to M.Z.).

References

(1) (a) Klebe, G. *J. Mol. Med.* **2000**, *78*, 269–281. (b) Craig, S. P., III; Eakin, A. E. *Vitam. Horm.* **2000**, *58*, 149–169. (c) Anderson, A. C. *Chem. Biol.* **2003**, *10*, 787–797. (d) Stahl, M.; Guba, W.; Kansy, M. *Drug Discovery Today* **2006**, *11*, 326–333.

(2) (a) Böhm, H.-J.; Klebe, G. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 2588–2614. (b) Babine, R. E.; Bender, S. L. *Chem. Rev.* **1997**, *97*, 1359–1472. (c) Hunter, C. A. *Angew. Chem., Int. Ed.* **2004**, *43*, 5310–5324. (d) Hobza, P.; Zahradník, R.; Müller-Dethlefs, K. *Collect. Czech. Chem. Commun.* **2006**, *71*, 443–531.

(3) (a) Diederich, F. *Angew. Chem., Int. Ed. Engl.* **1988**, *27*, 362–386. (b) Diederich, F.; Smithrud, D. B.; Sanford, E. M.; Wyman, T. B.; Ferguson, S. B.; Carcanague, D. R.; Chao, I.; Houk, K. N. *Acta Chem. Scand.* **1992**, *46*, 205–215.

(4) (a) Oshovsky, G. V.; Reinhoudt, D. N.; Verboom, W. *Angew. Chem., Int. Ed.* **2007**, *46*, 2366–2393. (b) Hooley, R. J.; Van Anda, H. J.; Rebek, J., Jr. *J. Am. Chem. Soc.* **2007**, *129*, 13464–13473.

(5) (a) Obst, U.; Gramlich, V.; Diederich, F.; Weber, L.; Banner, D. W. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 1739–1742. (b) Obst, U.; Banner, D. W.; Weber, L.; Diederich, F. *Chem. Biol.* **1997**, *4*, 287–295.

(6) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210–1250.

(7) Paulini, R.; Müller, K.; Diederich, F. *Angew. Chem., Int. Ed.* **2005**, *44*, 1788–1805.

(8) Hirsch, A. K. H.; Fischer, F. R.; Diederich, F. *Angew. Chem., Int. Ed.* **2007**, *46*, 338–352.

(9) Gerber, P. R.; Müller, K. *J. Comput.-Aided Mol. Design* **1995**, *9*, 251–268.

(10) For a recent article addressing proper cavity filling at the thrombin active site, see: Gerlach, C.; Münzel, M.; Baum, B.; Gerber, H.-D.; Craan, T.; Diederich, W. E.; Klebe, G. *Angew. Chem., Int. Ed.* **2007**, *46*, 9105–9109.

(11) (a) Diederich, F. *Cyclophanes*; The Royal Society of Chemistry: Cambridge, 1991. (b) *Comprehensive Supramolecular Chemistry*; Vögtle, F. Ed.; Elsevier-Pergamon: Oxford, 1996; Vol. 2. (c) Schneider, H.-J.; Yatsimirsky, A. *Principles and Methods in Supramolecular Chemistry*; Wiley: Chichester, 1999. (d) Diederich, F. In *Modern Cyclophane Chemistry*; Gleiter, R., Hopf, H. Eds; Wiley-VCH: Weinheim, Germany, 2004.

(12) (a) Ferguson, S. B.; Seward, E. M.; Diederich, F.; Sanford, E. M.; Chou, A.; Inocencio-Szweda, P.; Knobler, C. B. *J. Org. Chem.* **1988**, *53*, 5593–5595. (b) Smithrud, D. B.; Wyman, T. B.; Diederich, F. *J. Am. Chem. Soc.* **1991**, *113*, 5420–5426. (c) Peterson, B. R.; Wallimann, P.; Carcanague, D. R.; Diederich, F. *Tetrahedron* **1995**, *51*, 401–421.

(13) (a) Inoue, Y.; Liu, Y.; Tong, L.-H.; Shen, B.-J.; Jin, D.-S. *J. Am. Chem. Soc.* **1993**, *115*, 10637–10644. (b) Dunitz, J. D. *Chem. Biol.* **1995**, *2*, 709–712.

(14) (a) Benfield, A. P.; Teresk, M. G.; Plake, H. R.; DeLorbe, J. E.; Millspaugh, L. E.; Martin, S. F. *Angew. Chem., Int. Ed.* **2006**, *45*, 6830–6835. (b) Krishnamurthy, V. M.; Bohall, B. R.; Semetey, V.; Whitesides, G. M. *J. Am. Chem. Soc.* **2006**, *128*, 5802–5812.

(15) (a) Cromwell, W. C.; Byström, K.; Eftink, M. R. *J. Phys. Chem.* **1985**, *89*, 326–332. (b) For a detailed discussion, see ref 11a, Chapter 7.4.3, pp 254−257.

(16) Obst, U.; Betschmann, P.; Lerner, C.; Seiler, P.; Diederich, F.; Gramlich, V.; Weber, L.; Banner, D. W.; Schönholzer, P. *Helv. Chim. Acta* **2000**, *83*, 855–909.

(17) Fokkens, J.; Klebe, G. *Angew. Chem., Int. Ed.* **2006**, *45*, 985–989.

(18) Olsen, J. A.; Banner, D. W.; Seiler, P.; Obst Sander, U.; D'Arcy, A.; Stihle, M.; Müller, K.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 2507–2511.

(19) For reviews on thrombin inhibitors, see: (a) Rewinkel, J. B. M.; Adang, A. E. P. *Curr. Pharm. Design* **1999**, *5* (12), 1043–1075. (b) Srivastava, S.; Goswami, L. N.; Dikshit, D. K. *Med. Res. Rev.* **2005**, *25*, 66–92. (c) Di Nisio, M.; Middeldorp, S.; Büller, H. R. *N. Engl. J. Med.* **2005**, *353*, 1028–1040. (d) Schwienhorst, A. *Cell. Mol. Life Sci.* **2006**, *63*, 2773–2791.

(20) (a) Tucker, T. J.; Brady, S. F.; Lumma, W. C.; Lewis, S. D.; Gardell, S. J.; Naylor-Olsen, A. M.; Yan, Y.; Sisko, J. T.; Stauffer, K. J.; Lucas, B. J.; Lynch, J. J.; Cook, J. J.; Stranieri, M. T.; Holahan, M. A.; Lyle, E. A.; Baskin, E. P.; Chen, I-W.; Dancheck, K. B.; Krueger, J. A.; Cooper, C. M.; Vacca, J. P. *J. Med. Chem.* **1998**, *41*, 3210–3219. (b) Zhang, M.; Bailey, D. L.; Bastian, J. A.; Briggs, S. L.; Chirgadze, N. Y.; Clawson, D. K.; Denney, M. L.; Gifford-Moore, D. S.; Harper, R. W.; Johnson, L. M.; Klimkowski, V. J.; Kohn, T. J.; Lin, H.-S.; McCowan, J. R.; Richett, M. E.; Sall, D. J.; Smith, A. J.; Smith, G. F.; Snyder, D. W.; Takeuchi, K.; Utterback, B. G.; Yan, S.-C. B. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 775–780. (c) Baettig, U.; Brown, L.; Brundish, D.; Dell, C.; Furzer, A.; Garman, S.; Janus, D.; Kane, P. D.; Smith, G.; Walker, C. V.; Cockcroft, X.; Ambler, J.; Mitchelson, A.; Talbot, M. D.; Tweed, M.; Wills, N. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1563–1566.

(21) (a) Betschmann, P.; Lerner, C.; Sahli, S.; Obst, U.; Diederich, F. *Chimia* **2000**, *54*, 633–639. (b) Betschmann, P.; Sahli, S.; Diederich, F.; Obst, U.; Gramlich, V. *Helv. Chim. Acta* **2002**, *85*, 1210–1245. (c) Schweizer, E. ETH Dissertation No. 16531, Zürich, 2006.

(22) (a) Müller, K.; Faeh, C.; Diederich, F. *Science* **2007**, *317*, 1881–1886. (b) For another recent review on fluorine in medicinal chemistry, see Purser, S.; Moore, P. R.; Swallow, S.; Gouverneur, V. *Chem. Soc. Rev.* **2008**, *37*, 320–330.

(23) (a) Olsen, J.; Seiler, P.; Wagner, B.; Fischer, H.; Tschopp, T.; Obst-Sander, U.; Banner, D. W.; Kansy, M.; Müller, K.; Diederich, F. *Org. Biomol. Chem.* **2004**, *2*, 1339–1352. (b) Schweizer, E.; Hoffmann-Röder, A.; Schärer, K.; Olsen, J. A.; Fäh, C.; Seiler, P.; Obst-Sander, U.; Wagner, B.; Kansy, M.; Diederich, F. *ChemMedChem* **2006**, *1*, 611–621.

(24) For a review on tuning amine basicities, see: Morgenthaler, M.; Schweizer, E.; Hoffmann-Röder, A.; Benini, F.; Martin, R. E.; Jaeschke, G.; Wagner, B.; Fischer, H.; Bendels, S.; Zimmerli, D.; Schneider, J.; Diederich, F.; Kansy, M.; Müller, K. *ChemMedChem* **2007**, *2*, 1100–1115.

(25) Tucker, T. J.; Lumma, W. C.; Lewis, S. D.; Gardell, S. J.; Lucas, B. J.; Sisko, J. T.; Lynch, J. J.; Lyle, E. A.; Baskin, E. P.; Woltmann, R. F.; Appleby, S. D.; Chen, I-W.; Dancheck, K. B.; Naylor-Olsen, A. M.; Krueger, J. A.; Cooper, C. M.; Vacca, J. P. *J. Med. Chem.* **1997**, *40*, 3687–3693.

(26) Gerlach, C.; Smolinski, M.; Steuber, H.; Sotriffer, C. A.; Heine, A.; Hangauer, D. G.; Klebe, G. *Angew. Chem., Int. Ed.* **2007**, *46*, 8511–8514.

(27) (a) Sinnokrot, M. O.; Sherrill, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 7690–7697. (b) Lee, E. C.; Hong, B. H.; Lee, J. Y.; Kim, J. C.; Kim, D.; Kim, Y.; Tarakeshwar, P.; Kim, K. S. *J. Am. Chem. Soc.* **2005**, *127*, 4530–4537. (c) Ringer, A. L.; Sinnokrot, M. O.; Lively, R. P.; Sherrill, C. D. *Chem. Eur. J.* **2006**, *12*, 3821–3828.

(28) (a) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Fujii, A. *J. Phys. Chem. A* **2006**, *110*, 10163–10168. (b) Shibasaki, K.; Fujii, A.; Mikami, N.; Tsuzuki, S. *J. Phys. Chem. A* **2007**, *111*, 753–758.

(29) Olsen, J. A.; Banner, D. W.; Seiler, P.; Wagner, B.; Tschopp, T.; Obst-Sander, U.; Kansy, M.; Müller, K.; Diederich, F. *ChemBioChem* **2004**, *5*, 666–675.

(30) CSD search: CSD version 5.29, November 2007, inclusive updates January 2008, containing about 440,000 entries. All searches were performed for structures with R factor ≤0.1 (not distorted, no error, no partial disorder), excluding polymeric, powder, or organometallic compounds.

(31) Schweizer, E.; Hoffmann-Röder, A.; Olsen, J. A.; Seiler, P.; Obst-Sander, U.; Wagner, B.; Kansy, M.; Banner, D. W.; Diederich, F. *Org. Biomol. Chem.* **2006**, *4*, 2364–2375.

(32) Cockroft, S. L.; Hunter, C. A. *Chem. Soc. Rev.* **2007**, *36*, 172–188.

(33) Bhayana, B.; Wilcox, C. S. *Angew. Chem., Int. Ed.* **2007**, *46*, 6833–6836.

(34) (a) Hof, F.; Scofield, D. M.; Schweizer, W. B.; Diederich, F. *Angew. Chem., Int. Ed.* **2004**, *43*, 5056–5059. (b) Fischer, F. R.; Schweizer, W. B.; Diederich, F. *Angew. Chem., Int. Ed.* **2007**, *46*, 8270–8273.

(35) Hoffmann-Röder, A.; Schweizer, E.; Egger, J.; Seiler, P.; Obst-Sander, U.; Wagner, B.; Kansy, M.; Banner, D. W.; Diederich, F. *ChemMedChem* **2006**, *1*, 1205–1215.

(36) *Spartan'06*, Wavefunction, Inc.: Irvine, 2006.

(37) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441–451.

(38) Stauffer, D. A.; Dougherty, D. A. *Tetrahedron Lett.* **1988**, *29*, 6039–6042.

(39) (a) Dougherty, D. A.; Stauffer, D. A. *Science* **1990**, *250*, 1558–1560. (b) Dougherty, D. A. *Science* **1996**, *271*, 163–168. (c) Ma, J. C.; Dougherty, D. A. *Chem. Rev.* **1997**, *97*, 1303–1324. (d) Zacharias, N.; Dougherty, D. A. *Trends Pharmacol. Sci.* **2002**, *23*, 281–287. (e) Dougherty, D. A. *J. Nutr.* **2007**, *137*, 1504S–1508S. (f) Hughes, R. M.; Benshoff, M. L.; Waters, M. L. *Chem. Eur. J.* **2007**, *13*, 5753–5764.

(40) Schärer, K.; Morgenthaler, M.; Paulini, R.; Obst-Sander, U.; Banner, D. W.; Schlatter, D.; Benz, J.; Stihle, M.; Diederich, F. *Angew. Chem., Int. Ed.* **2005**, *44*, 4400–4404.

(41) (a) Davie, E. W.; Fujikawa, K.; Kisiel, W. *Biochemistry* **1991**, *30*, 10363–10370. (b) Maignan, S.; Mikol, V. *Curr. Top. Med. Chem.* **2001**, *1*, 161–174.

(42) (a) For Rivaroxaban (BAY 59-7939), an oral direct factor Xa inhibitor in phase III clinical trials, see: Agnelli, G.; Gallus, A.; Goldhaber, S. Z.; Haas, S.; Huisman, M. V.; Hull, R. D.; Kakkar, A. K.; Misselwitz, F.; Schellong, S. *Circulation* **2007**, *116*, 180–187. (b) For a general recent review on factor Xa inhibitors, see: Ieko, M.; Tarumi, T.; Nakabayashi, T.; Yoshida, M.; Naito, S.; Koike, T. *Front. Biosci.* **2006**, *11*, 232–248.

(43) (a) Molinoff, P. B.; Axelrod, J. *Annu. Rev. Biochem.* **1971**, *40*, 465–500. (b) Guldberg, H. C.; Marsden, C. A. *Pharmacol. Rev.* **1975**, *27*, 135–206.

(44) Keating, G. M.; Lyseng-Williamson, K. A. *CNS Drugs* **2005**, *19* (2), 165–184.

(45) Borgulya, J.; Bruderer, H.; Bernauer, K.; Zürcher, G.; Da Prada, M. *Helv. Chim. Acta* **1989**, *72*, 952–968.

(46) (a) Masjost, B.; Ballmer, P.; Borroni, E.; Zürcher, G.; Winkler, F. K.; Jakob-Roetne, R.; Diederich, F. *Chem. Eur. J.* **2000**, *6*, 971–982. (b) Lerner, C.; Ruf, A.; Gramlich, V.; Masjost, B.; Zürcher, G.; Jakob-Roetne, R.; Borroni, E.; Diederich, F. *Angew. Chem., Int. Ed.* **2001**, *40*, 4040–4042.

(47) Lerner, C.; Masjost, B.; Ruf, A.; Gramlich, V.; Jakob-Roetne, R.; Zürcher, G.; Borroni, E.; Diederich, F. *Org. Biomol. Chem.* **2003**, *1*, 42–49.

(48) (a) Paulini, R.; Lerner, C.; Jakob-Roetne, R.; Zürcher, G.; Borroni, E.; Diederich, F. *ChemBioChem* **2004**, *5*, 1270–1274. (b) Paulini, R.; Lerner, C.; Diederich, F.; Jakob-Roetne, R.; Zürcher, G.; Borroni, E. *Helv. Chim. Acta* **2006**, *89*, 1856–1887.

(49) (a) Bonifácio, M. J.; Archer, M.; Rodrigues, M. L.; Matias, P. M.; Learmonth, D. A.; Carrondo, M. A.; Soares-da-Silva, P. *Mol. Pharmacol.* **2002**, *62*, 795–805. (b) Tervo, A. J.; Nyrönen, T. H.; Rönkkö, T.; Poso, A. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 797–810.

(50) (a) Rohmer, M.; Knani, M.; Simonin, P.; Sutter, B.; Sahm, H. *Biochem. J.* **1993**, *295*, 517–524. (b) Schwarz, M. K. ETH Dissertation No. 10951, Zürich, 1994. (c) Eisenreich, W.; Schwarz, M.; Cartayrade, A.; Arigoni, D.; Zenk, M. H.; Bacher, A. *Chem. Biol.* **1998**, *5*, R221–R233.

(51) Jomaa, H.; Wiesner, J.; Sanderbrand, S.; Altincicek, B.; Weidemeyer, C.; Hintz, M.; Türbachova, I.; Eberl, M.; Zeidler, J.; Lichtenthaler, H. K.; Soldati, D.; Beck, E. *Science* **1999**, *285*, 1573–1576.

(52) Hunter, W. N. *J. Biol. Chem.* **2007**, *282*, 21573–21577.

(53) (a) Kemp, L. E.; Bond, C. S.; Hunter, W. N. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6591–6596. (b) Steinbacher, S.; Kaiser, J.; Wungsintaweekul, J.; Hecht, S.; Eisenreich, W.; Gerhardt, S.; Bacher, A.; Rohdich, F. *J. Mol. Biol.* **2002**, *316*, 79–88, To date, 10 crystal structures of IspF have been published.

(54) Crane, C. M.; Kaiser, J.; Ramsden, N. L.; Lauw, S.; Rohdich, F.; Eisenreich, W.; Hunter, W. N.; Bacher, A.; Diederich, F. *Angew. Chem., Int. Ed.* **2006**, *45*, 1069–1074.

(55) Baumgartner, C.; Eberle, C.; Diederich, F.; Lauw, S.; Rohdich, F.; Eisenreich, W.; Bacher, A. *Helv. Chim. Acta* **2007**, *90*, 1043–1068.

(56) Hirsch, A. K. H.; Lauw, S.; Gersbach, P.; Schweizer, W. B.; Rohdich, F.; Eisenreich, W.; Bacher, A.; Diederich, F. *ChemMedChem* **2007**, *2*, 806–810.

(57) (a) Wada, T.; Kuzuyama, T.; Satoh, S.; Kuramitsu, S.; Yokoyama, S.; Unzai, S.; Tame, J. R. H.; Park, S.-Y. *J. Biol. Chem.* **2003**, *278*, 30022–30027. (b) Miallau, L.; Alphey, M. S.; Kemp, L. E.; Leonard, G. A.; McSweeney, S. M.; Hecht, S.; Bacher, A.; Eisenreich, W.; Rohdich, F.; Hunter, W. N. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9173–9178.

(58) (a) Petrov, V.; Petrova, V.; Girichev, G. V.; Oberhammer, H.; Giricheva, N. I.; Ivanov, S. *J. Org. Chem.* **2006**, *71*, 2952–2956. (b) Senger, S.; Convery, M. A.; Chan, C.; Watson, N. S. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5731–5735, The staggered conformation is favored both for steric reasons and as a result of $n\rightarrow\sigma^*_{S-C}$ interactions.

(59) Hirsch, A. K. H.; Alphey, M. S.; Lauw, S.; Seet, M.; Barandun, L.; Rohdich, F.; Hunter, W. N.; Bacher, A.; Diederich, F. *Org. Biomol. Chem.* **2008**, . submitted. PDB code of the crystal structure: 2VF3.

(60) We observed the turning away of larger substituents in cocrystal structures: Crane, C. M.; Hirsch, A. K. H.; Alphey, M. S.; Sgraja, T.; Lauw, S.; Illarionova, V.; Rohdich, F.; Eisenreich, W.; Hunter, W. N.; Bacher, A.; Diederich, F. *ChemMedChem* **2008**, *3*, 91–101.

(61) Romier, C.; Meyer, J. E. W.; Suck, D. *FEBS Lett.* **1997**, *416*, 93–98.

(62) Okada, N.; Sasakawa, C.; Tobe, T.; Yamada, M.; Nagai, S.; Talukder, K. A.; Komatsu, K.; Kanegasaki, S.; Yoshikawa, M. *Mol. Microbiol.* **1991**, *5*, 187–195.

(63) Romier, C.; Reuter, K.; Suck, D.; Ficner, R. *EMBO J.* **1996**, *15*, 2850–2857.

(64) Xie, W.; Liu, X.; Huang, R. H. *Nat. Struct. Biol.* **2003**, *10*, 781–788.

(65) Meyer, E. A.; Furler, M.; Diederich, F.; Brenk, R.; Klebe, G. *Helv. Chim. Acta* **2004**, *87*, 1333–1356.

(66) Meyer, E. A.; Donati, N.; Guillot, M.; Schweizer, W. B.; Diederich, F.; Stengl, B.; Brenk, R.; Reuter, K.; Klebe, G. *Helv. Chim. Acta* **2006**, *89*, 573–597.

(67) (a) Brenk, R.; Meyer, E. A.; Reuter, K.; Stubbs, M. T.; Garcia, G. A.; Diederich, F.; Klebe, G. *J. Mol. Biol.* **2004**, *338*, 55–75. (b) Stengl, B.; Meyer, E. A.; Heine, A.; Brenk, R.; Diederich, F.; Klebe, G. *J. Mol. Biol.* **2007**, *370*, 492–511.

(68) Meyer, E. A.; Brenk, R.; Castellano, R. K.; Furler, M.; Klebe, G.; Diederich, F. *ChemBioChem* **2002**, *3*, 250–253.

(69) Hörtner, S. R.; Ritschel, T.; Stengl, B.; Kramer, C.; Schweizer, W. B.; Wagner, B.; Kansy, M.; Klebe, G.; Diederich, F. *Angew. Chem., Int. Ed.* **2007**, *46*, 8266–8269.

(70) Stengl, B.; Reuter, K.; Klebe, G. *ChemBioChem.* **2005**, *6*, 1926–1939.

(71) Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W. *J. Am. Chem. Soc.* **2007**, *129*, 2577–2587.

(72) (a) Silva, A. M.; Lee, A. Y.; Gulnik, S. V.; Majer, P.; Collins, J.; Bhat, T. N.; Collins, P. J.; Cachau, R. E.; Luker, K. E.; Gluzman, I. Y.; Francis, S. E.; Oksman, A.; Goldberg, D. E.; Erickson, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10034–10039. (b) Boss, C.; Richard-Bildstein, S.; Weller, T.; Fischli, W.; Meyer, S.; Binkert, C. *Curr. Med. Chem.* **2003**, *10*, 883–907.

(73) Liu, J.; Gluzman, I. Y.; Drew, M. E.; Goldberg, D. E. *J. Biol. Chem.* **2005**, *280*, 1432–1437.

(74) (a) Asojo, O. A.; Gulnik, S. V.; Afonina, E.; Yu, B.; Ellman, J. A.; Haque, T. S.; Silva, A. M. *J. Mol. Biol.* **2003**, *327*, 173–181. (b) Clemente, J. C.; Govindasamy, L.; Madabushi, A.; Fisher, S. Z.; Moose, R. E.; Yowell, C. A.; Hidaka, K.; Kimura, T.; Hayashi, Y.; Kiso, Y.; Agbandje-McKenna, M.; Dame, J. B.; Dunn, B. M.; McKenna, R. *Acta Crystallogr. Sect. D* **2006**, *62*, 246–252.

(75) (a) Prade, L.; Jones, A. F.; Boss, C.; Richard-Bildstein, S.; Meyer, S.; Binkert, C.; Bur, D. *J. Biol. Chem.* **2005**, *280*, 23837–23843. (b) Boss, C.; Corminboeuf, O.; Grisostomi, C.; Meyer, S.; Jones, A. F.; Prade, L.; Binkert, C.; Fischli, W.; Weller, T.; Bur, D. *ChemMedChem* **2006**, *1*, 1341–1345.

(76) Oefner, C.; Binggeli, A.; Breu, V.; Bur, D.; Clozel, J.-P.; D'Arcy, A.; Dorn, A.; Fischli, W.; Grüninger, F.; Güller, R.; Hirth, G.; Märki, H. P.; Mathews, S.; Müller, M.; Ridley, R. G.; Stadler, H.; Vieira, E.; Wilhelm, M.; Winkler, F. K.; Wostl, W. *Chem. Biol.* **1999**, *6*, 127–131.

(77) Hof, F.; Schütz, A.; Fäh, C.; Meyer, S.; Bur, D.; Liu, J.; Goldberg, D. E.; Diederich, F. *Angew. Chem., Int. Ed.* **2006**, *45*, 2138–2141.

(78) Zürcher, M.; Gottschalk, T.; Meyer, S.; Bur, D.; Diederich, F. *ChemMedChem* **2008**, *3*, 237–240.

(79) Mecozzi, S.; Rebek, J., Jr. *Chem. Eur. J.* **1998**, *4*, 1016–1022.

(80) (a) Hayashida, O.; Sebo, L.; Rebek, J., Jr. *J. Org. Chem.* **2002**, *67*, 8291–8298. (b) Scarso, A.; Trembleau, L.; Rebek., J., Jr. *Angew. Chem., Int. Ed.* **2003**, *42*, 5499–5502. (c) Scarso, A.; Trembleau, L.; Rebek, J., Jr. *J. Am. Chem. Soc.* **2004**, *126*, 13512–13518. (d) Gottschalk, T.; Jaun, B.; Diederich, F. *Angew. Chem., Int. Ed.* **2007**, *46*, 260–264.

(81) (a) Schramm, M. P.; Rebek, J., Jr. *Chem. Eur. J.* **2006**, *12*, 5924–5933. (b) Rebek, J., Jr. *Chem. Commun.* **2007**, 2777–2789.

JO800527N