Congeneric but Still Distinct: How Closely Related Trypsin Ligands Exhibit Different Thermodynamic and Structural Properties

Tobias Brandt[1]†, **Nicole Holzmann**[1]†, **Laveena Muley**[2], **Maan Khayat**[2], **Christof Wegscheid-Gerlach**[1], **Bernhard Baum**[1], **Andreas Heine**[1], **David Hangauer**[2]* and **Gerhard Klebe**[1]*

[1]*Department of Pharmaceutical Chemistry, Philipps University Marburg, Marbacher Weg 6, 35032 Marburg, Germany*
[2]*Department of Chemistry, University at Buffalo, The State University of New York, Buffalo, NY 14260, USA*

A congeneric series of benzamidine-type ligands with a central proline moiety and a terminal cycloalkyl group—linked by a secondary amine, ether, or methylene bridge—was synthesized as trypsin inhibitors. This series of inhibitors was investigated by isothermal titration calorimetry, crystal structure analysis in two crystal forms, and molecular dynamics simulations. Even though all of these congeneric ligands exhibited essentially the same affinity for trypsin, their binding profiles at the structural, dynamic, and thermodynamic levels are very distinct. The ligands display a pronounced enthalpy/entropy compensation that results in a nearly unchanged free energy of binding, even though individual enthalpy and entropy terms change significantly across the series. Crystal structures revealed that the secondary amine-linked analogs scatter over two distinct conformational families of binding modes that occupy either the inside or of the outside the protein's S3/S4 specificity pocket. In contrast, the ether-linked and methylene-linked ligands preferentially occupy the hydrophobic specificity pocket. This also explains why the latter ligands could only be crystallized in the conformationally restricting *closed* crystal form whereas the derivative with the highest residual mobility in the series escaped our attempts to crystallize it in the *closed* form; instead, a well-resolved structure could only be achieved in the *open* form with the ligand in disordered orientation. These distinct binding modes are supported by molecular dynamics simulations and correlate with the shifting enthalpic/entropic signatures of ligand binding. The examples demonstrate that, at the molecular level, binding modes and thermodynamic binding signatures can be very different even for closely related ligands. However, deviating binding profiles provide the opportunity to optimally address a given target.

Corresponding authors. E-mail addresses: hangauer@buffalo.edu; klebe@staff.uni-marburg.de.
† T.B. and N.H. contributed equally to this work.

Introduction

Successful drug research and development strongly depends on the discovery, accessibility, and validation of reliable and relevant targets for putative drug intervention.[1,2] Once a promising target has been selected, putative lead compounds that either inhibit, stimulate, prevent, or down-regulate the function of the target protein must be identified. Various high-throughput technologies have been developed for the discovery of small-molecule leads as candidates for further optimization into putative drugs.[3] However, it is often difficult to determine *a priori* which of the discovered leads should be taken into further steps of optimization and development. Because this step will allocate major resources, it has to be taken very carefully and needs reliable and relevant parameters to support the decision.[4]

Both ease of synthetic accessibility and the potential to improve interactions with the target protein are used to guide ligand design. The latter requires a meaningful structural characterization of target and lead, mostly performed by crystal structure determination and measurements of binding affinities. However, these measurements do not unravel details about the actual binding profile. Information on thermodynamic properties in terms of an enthalpy/entropy signature can instead be gained by characterizing ligand binding using isothermal titration calorimetry (ITC). Furthermore, limited data on the involved dynamics of protein ligand binding can be derived from crystal structure analysis. Conformational insight may be collected by computational methods such as molecular dynamics (MD) simulations. They can assist in the interpretation of ITC data by providing insights into the flexibility/dynamics of the protein–ligand complex.[5,6] Although of utmost importance, data on interaction kinetics—in terms of on-rates and off-rates, or the resulting residence time of a drug at its target—are usually not measured.[7] Possibly, our current view on drug binding is still too strongly biased towards the complementarity hypothesis in terms of shape, electrostatics, and interaction patterns, as influenced by the valuable but rather static image of Fischer's lock-and-key model.[8] Instead, a much more complex situation involving dynamics, induced-fit adaptations,[9] and the crucial influence of water results in a complex balance of enthalpy and entropy contributions.

Using trypsin as model protein, the present study tries to assemble relevant data to better characterize the binding profiles of a set of congeneric ligands. Trypsin belongs to the family of serine proteases. It is a digestive enzyme found in various human and animal tissues.[10,11] Physiologically, it is not a direct target for drug discovery; however, as a representative serine protease, it is a valuable model system that is applicable to other related serine proteases that are drug targets. Trypsin has fewer amino acids than most of the physiologically important proteases, and this renders trypsin more amendable to study by experimental and computational methods. Like thrombin, plasmin, factor Xa, or urokinase, trypsin hydrolyzes peptide sequences after a lysine or arginine residue. This selectivity is due to the presence of the amino acid residue aspartate 189 located at the bottom of the S1 pocket of the enzyme.[12]

The congeneric ligand series, displayed in Fig. 1, has been investigated for its binding to trypsin. The basic skeleton is conserved across all studied ligands: They all possess a benzamidine anchor group, which binds into the S1 pocket, and a proline moiety, which occupies the S2 pocket. The parts of the ligands addressing the S3/S4 pocket exhibit structural differences. They vary in the ring size of the terminal cycloalkyl moiety (from five-membered to eight-membered rings) and in the atom that connects the ring portion to the central ligand skeleton (secondary amine, ether, or methylene bridge).

In order to determine the binding affinities of the ligands for trypsin, we carried out kinetic assays. Subsequently, ITC was used to elucidate the enthalpic contribution ΔH^0 to binding and to characterize the thermodynamic signatures for each of the trypsin–ligand complexes. In order to gain more insights into the ligand binding modes, we obtained the crystal structures of all of the trypsin–ligand complexes and, in several cases, we obtained the structures in two different polymorphic forms. MD simulations were performed at physiological temperature to rationalize the thermodynamic profiles in greater detail. These simulations also elucidated the flexibility of the different ligands in the binding pocket. For comparative reasons, an MD simulation of uncomplexed trypsin was also carried out.

Results

Kinetic inhibition assay

The inhibitory efficacy of the eight congeneric compounds against trypsin was determined (Fig. 1). The ligands **1a–1d**, linked by a secondary amine functionality, inhibit trypsin slightly more strongly than the ether-linked (**2a** and **2b**) and methylene-linked (**3a** and **3b**) ligands.

Thermodynamic studies

As a second step, we carried out a thermodynamic analysis of the protein–ligand complexes. We used ITC to measure the dissociation constant K_d of the
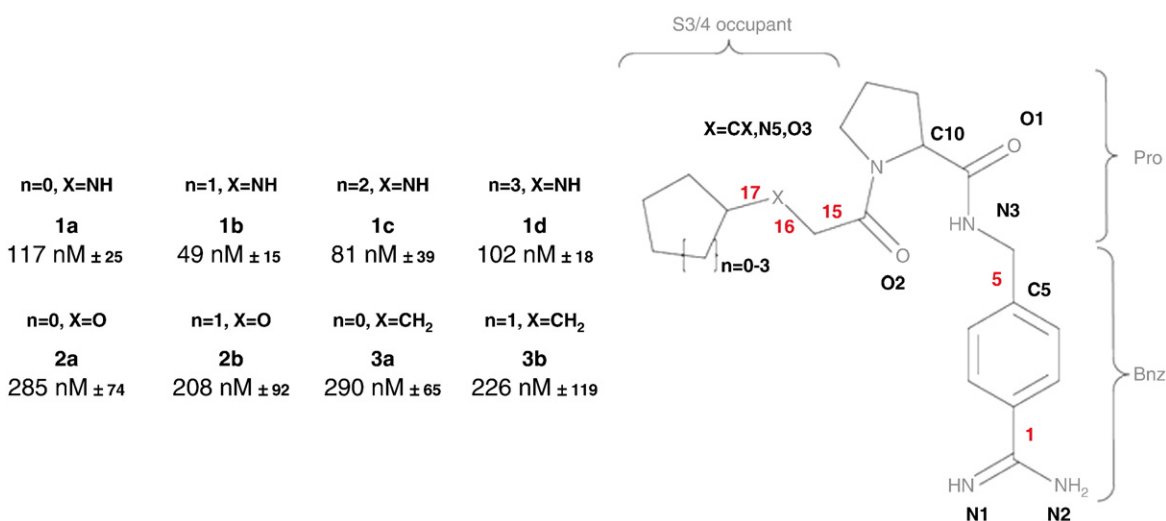
Fig. 1. Chemical formulas and abbreviations of the inhibitors used for this study. The kinetic inhibition constant K_i and the standard deviation are given below each inhibitor. Each measurement was performed at least in triplicate. For *B*-factor analysis, the ligands were divided into three parts, as indicated by brackets on the ligand scaffold. Pro=proline; Bnz=benzamidine. The atom numbers used are given in black, whereas torsion angles are indicates by numbers in red.

protein–ligand complexes, as well as the evolved heat, which comprises the ΔH_{bind}^0 of the exothermic binding process. The dissociation constant K_d is a more direct indicator than K_i for the binding affinity of a ligand for a protein because no competing binding of a labelled substrate is involved. Additionally, the thermodynamic parameters ΔG^0 and $T\Delta S_{bind}^0$ can be calculated using the equation: $RT\ln K_d = \Delta G^0 = \Delta H_{bind}^0 - T\Delta S_{bind}^0$.

We carried out ITC experiments for ligands **1a**–**1d**, **2a**, **2b**, **3a**, and **3b**. All experiments were repeated three times, and the results are summarized in Table 1 and Fig. 2. The measured K_d differences determined by ITC and the photometrically characterized K_i values are virtually equiv-

alent, considering the error margins of the methods. These results confirm the trend observed in the kinetic inhibition assay. The measured affinities (based on K_d) fall within a range between 239 nM (**1b**) and 562 nM (**3a**). The empiric rule that an additional methylene group adds 3–4 kJ/mol in affinity for a protein–ligand complex does not agree with the properties of the ligands studied here.[13,14] The enlargement of the terminal aliphatic ring moiety by up to three methylene groups (**1a**–**1d**) does not enhance binding affinity at all.

The measured thermodynamic profiles are a remarkable example of the phenomenon of enthalpy/entropy compensation.[4,6,15] The binding affinities ΔG

Table 1. Thermodynamic data determined by ITC

Ligand	Buffer	ΔG^0 (kJ/mol)	K_d (nM)	n	ΔH^0 (kJ/mol)	$-T\Delta S^0$ (kJ/mol)
1a	Tris	−37.3±0.4	300±34 (26)	1.04±0.03	−22.4±2.4 (0.1)	−14.9±2.6
	Tricine	−37.2±0.7	316±60 (38)	0.99±0.02	−17.6±1.2 (0.1)	−19.5±1.3
	Hepes	−37.3±0.4	301±43 (40)	1.02±0.02	−15.1±0.3 (0.1)	−22.2±0.7
	PP_i	−37.8±0.5	245±38 (24)	1.04±0.02	−18.7±0.5 (0.1)	−19.1±1.0
1b	Tris	−37.8±0.4	239±26 (31)	0.92±0.00	−19.9±0.8 (0.2)	−17.9±0.4
1c	Tris	−37.4±0.6	287±48 (76)	0.92±0.03	−18.9±0.2 (0.5)	−18.5±0.6
1d	Tris	−37.5±0.8	276±76 (111)	1.04±0.06	−15.3±0.3 (0.4)	−22.2±1.1
2a	Tris	−36.6±0.7	402±78 (25)	0.95±0.04	−27.3±1.6 (0.2)	−9.3±1.6
2b	Tris	−36.6±0.4	386±49 (25)	0.91±0.04	−30.3±1.1 (0.2)	−6.3±1.5
3a	Tris	−35.7±0.4	562±66 (71)	0.99±0.04	−21.3±1.2 (0.2)	−14.4±1.3
	Tricine	−35.6±0.7	584±107 (56)	0.96±0.02	−19.3±1.4 (0.2)	−16.3±1.8
	Hepes	−35.9±0.2	506±34 (59)	1.10±0.02	−15.1±0.4 (0.1)	−20.8±0.5
	PP_i	−36.4±0.3	416±42 (38)	0.96±0.05	−17.3±0.6 (0.1)	−19.2±0.9
3b	Tris	−36.4±0.3	428±38 (59)	1.15±0.08	−17.1±0.2 (0.2)	−19.2±0.4

Values are given for ΔG^0 (calculated from K_d), the stoichiometry n of complex formation, ΔH^0, and $-T\Delta S^0$ (calculated from ΔG^0 and ΔH^0). Average experimental values (in triplicate)±standard deviations are given. The average fitting errors are stated in parentheses (if appropriate).
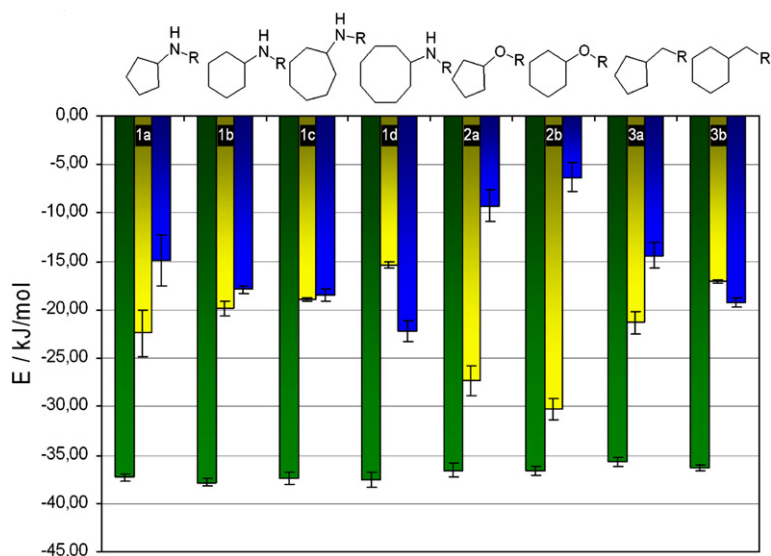
Fig. 2. Thermodynamic parameters determined by ITC for ligands **1a–1d**, **2a**, **2b**, **3a**, and **3b**.

of the amine ligands **1a–1d** are similar (mean: -37.5 ± 0.2 kJ/mol), whereas their individual thermodynamic profiles vary considerably with the size of the terminal aliphatic rings and the type of linker group (X). The most pronounced difference among the secondary amine-linked ligands occurs for the pair of the five-membered and eight-membered ring derivatives **1a** and **1d** with an absolute difference in $\Delta\Delta H^0_{bind}$ or $\Delta(-T\Delta S^0_{bind})$ of over 7 kJ/mol. The two ligands with a linking ether oxygen atom X (**2a** and **2b**) show considerably higher enthalpic contributions to the overall binding affinity as compared to **1a** and **1b** (4.9 and 10.4 kJ/mol, respectively). This gain in enthalpic binding results from a much lower entropic contribution. A slight increase in enthalpic contributions is registered when the terminal ring moiety is enlarged to a six-membered ring. This trend is inverted for the methylene-linked ligands **3a** and **3b**. The five-membered ring derivative **3a** shows a thermodynamic profile very similar to that of the secondary amine-linked ligand **1a**, but the six-membered ring derivative **3b** shows higher entropic contributions. Nevertheless, both ligands exhibit the same overall affinity for trypsin.

If the protein–ligand complex formation is overlaid by a net exchange of protons during the binding process, an additional heat of ionization arising from protonation and deprotonation steps will be experienced. Overall, in such a case, the enthalpy of binding ΔH^0_{bind} will depend on the buffers used for the experiments. Buffers exhibit different enthalpies of ionization ΔH^0_{ion}, and any release or pickup of protons is related to a different heat signal. To see whether any protonation step occurs in our studied system, we performed control experiments with **1a** and **3a** as representative ligands. Upon binding to trypsin, the secondary amine moiety of ligand **1a** might experience a shift of its pK_a value and, as a

consequence, of its protonation state. In contrast, **3a** has no such group that can potentially change its protonation state upon binding. Furthermore, several active-site residues of trypsin might change their protonation state. We determined the thermodynamic parameters of **1a** and **3a** in different buffers (Table 1) and recorded changes that maximally correspond to a release of 0.1 mol/L protons. This variation is within the error of the method; therefore, we assume that any superimposed protonation steps can be neglected.

The thermodynamic analysis revealed surprising differences in enthalpy/entropy partitioning. However, a complete explanation of the observed effects cannot be given solely based on these data. The distinctive enthalpy/entropy differences in thermodynamic profiles stimulated our interest to further study these protein–ligand complexes in more detail.

Crystallographic analysis

Trypsin can be crystallized in distinct space groups with different packings. We tried to crystallize **1a–1d**, **2a**, **2b**, **3a**, and **3b** in the orthorhombic space group $P2_12_12_1$. As the ligand binding site is freely accessible in this crystal form, we refer to it as the *open* form. We could easily cocrystallize the protein–ligand complexes of the secondary amine-linked ligands **1a–1d** at high resolution (1.40–1.94 Å); however, despite extensive screening for similar conditions, we did not succeed in growing trypsin crystals with **2a**, **2b**, **3a**, and **3b** in the *open* orthorhombic form. However, we could obtain well-diffracting crystals showing resolutions up to 1.51–1.98 Å for ligands **1a–1c**, **2a**, **2b**, **3a**, and **3b** in the trigonal space group $P3_121$. This trigonal form

shows a different packing. It restrains the available space next to the S3/S4 pocket due to close contacts with a crystal mate. This has an important consequence on the available space of the bound ligand and on the accessible binding modes, which we discuss below. All attempts to grow crystals in the trigonal space group with **1d** resulted in specimens of poor diffraction quality. Among the many crystals tested, we were only able to record a single data set. The initial difference electron density maps evaluated were similar to those of ligands **1a–1c** in this space group; however, only limited occupancy of the ligand was visible. Overall, we collected data sets of 11 well-resolved trypsin–ligand complexes (Tables 2 and 3).

Binding properties in the orthorhombic *open* form

The difference in electron densities for crystals in the *open* $P2_12_12_1$ form shows that at least two conformations of the ligands are populated (Fig. 3a–d). The superimposed ligands **1a–1d** (Fig. 3e) show a good agreement of the atom positions in the S1 pocket. Inside the S2 pocket, the atoms of the proline moiety show higher positional deviations indicative of a higher residual mobility. The aliphatic ring portions, as well as the linker atoms, preferentially adopt two orientations. In the first orientation "A" ("inside S3/S4"), the terminal aliphatic ring portion of the ligands occupies the hydrophobic S3/S4 pocket. Additionally, a second major orientation "B" ("outside S3/S4") can be observed. In this B form, the cycloalkyl portion is placed outside the S3/S4 pocket, but this orientation allows the formation of a hydrogen bond between the linking amino group of the ligand and Gly216O. Also, the distance between the neighboring ligand carbonyl oxygen atom O2 and Gly216N is shortened. This observation indicates that hydrophobic interactions experienced inside the S3/S4 pocket compete with the formation of hydrogen bonds and the placement of the ligand outside the pocket. Generally, the relative occupancy inside the S3/S4 pocket is higher for the ligands with

Table 2. Data collection and refinement statistics for four *open* trypsin complex structures determined in this study

Trypsin complex	*Open* trypsin complex			
	1a (PDB code 3LJJ)	1b (PDB code 3LJO)	1c (PDB code 2HZD)	1d (PDB code 2ZQ2)
Data collection				
Crystals used	1	1	1	1
Wavelength (Å)	1.54178	0.91841	1.54178	0.91841
Resolution range (Å)	50–1.55 (1.58–1.55)	50–1.50 (1.53–1.50)	50–1.94 (1.97–1.94)	30–1.40 (1.42–1.40)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters *a, b, c* (Å)	61.6, 63.7, 69.2	61.6, 63.6, 69.5	61.0, 63.6, 69.5	61.4, 63.9, 69.7
Unique reflections	39,672 (1715)	42,931 (2153)	18,886 (716)	54,593 (2701)
Redundancy	7.3 (3.4)	3.8 (3.4)	3.4 (2.3)	3.2 (3.2)
Completeness (%)	99.2 (88.0)	96.7 (98.4)	91.3 (70.4)	99.5 (100)
Mean I/σ	34.4 (3.1)	21.8 (2.9)	12.7 (2.6)	22.3 (2.9)
R_{sym} (%)	5.8 (36.4)	6.0 (53.8)	9.7 (36.3)	5.3 (39.0)
Solvent content (%)	56.1	56.2	57.4	58.0
Matthews coefficient (Å^3/Da)	2.8	2.8	2.9	2.9
Refinement				
Used reflections (work/free)	36,546 (1930)	38,342 (2025)	16,769 (859)	48,945 (2609)
Resolution range (Å)	10–1.55	10–1.50	10–1.94	10–1.40
R_{work} ($F_o > 4\sigma F_o$; F_o) (%)	15.9; 16.5	16.2; 17.7	15.9; 17.6	11.7; 12.5
R_{free} ($F_o > 4\sigma F_o$; F_o) (%)	20.2; 21.0	19.3; 20.8	22.4; 24.6	16.2; 17.2
Protein residues	223	223	223	223
Inhibitor atoms	27	28	29	30
Water molecules	233	213	151	269
Calcium ions	1	1	1	1
RMSD				
Bond lengths (Å)	0.013	0.013	0.008	0.012
Bond angles (°)	2.4	2.4	1.9	2.2
Mean *B*-factor				
Wilson *B*-factor (Å^2)	26.7	12.5	24.0	13.5
Protein (Å^2)	16.9	13.9	21.0	14.7
Main chain (Å^2)	15.1	12.0	18.8	13.0
Side chains (Å^2)	19.0	16.3	23.7	16.8
Inhibitor (Å^2)	21.3	18.5	25.9	23.1
Water molecules (Å^2)	34.2	32.5	30.2	30.7
Ramachandran plot				
Most favored (%)	87.2	87.2	88.3	87.8
Additionally allowed (%)	12.8	12.8	11.7	12.2
Generously allowed (%)	0.0	0.0	0.0	0.0

Table 3. Data collection and refinement statistics for seven *closed* trypsin complex structures determined in this study

	Closed trypsin complex						
Trypsin complex	1a (PDB code 2ZFT)	1b (PDB code 2ZQ1)	1c (PDB code 2ZFS)	2a (PDB code 2ZDL)	2b (PDB code 2ZDM)	3a (PDB code 2ZDN)	3b (PDB code 2ZDK)
Data collection							
Crystals used	1	1	1	1	1	1	1
Wavelength (Å)	1.54178	0.91841	1.54178	1.00001	1.54178	1.54178	1.54178
Resolution range (Å)	50–1.76 (1.79–1.76)	50–1.68 (1.71–1.68)	50–1.51 (1.54–1.51)	50–1.80 (1.83–1.80)	50–1.93 (1.96–1.93)	50–1.98 (2.01–1.98)	50–1.67 (1.70–1.67)
Space group	$P3_121$	$P3_121$	$P3_121$	$P3_121$	$P3_121$	$P3_121$	$P3_121$
Unit cell parameters a, b, c (Å)	54.6, 54.6, 108.5	54.9, 54.9, 109.0	54.6, 54.6, 108.4	54.5, 54.5, 108.6	54.0, 54.0, 108.6	54.6, 54.6, 108.6	54.6, 54.6, 108.6
Unique reflections	19,216 (897)	21,968 (1481)	29,336 (1111)	16,745 (813)	14,347 (680)	13,582 (659)	22,193 (944)
Redundancy	5.8 (3.6)	1.8 (1.6)	5.8 (2.0)	4.0 (3.2)	5.5 (5.0)	4.1 (4.1)	6.2 (3.5)
Completeness (%)	99.6 (96.3)	94.5 (71.8)	97.3 (75.3)	93.1 (93.4)	99.6 (100)	99.6 (99.8)	98.7 (83.8)
Mean I/σ	21.6 (2.3)	19.7 (2.3)	29.1 (3.7)	15.4 (2.8)	17.3 (3.4)	14.3 (3.9)	31.7 (2.9)
R_{sym} (%)	7.7 (48.0)	4.1 (32.7)	4.6 (25.3)	8.5 (38.3)	10.0 (49.3)	9.8 (41.8)	5.9 (35.6)
Solvent content (%)	38.6	39.5	38.5	38.4	37.2	38.6	38.6
Matthews coefficient (Å^3/Da)	2.0	2.0	2.0	2.0	2.0	2.0	2.0
Refinement							
Used reflections (work/free)	18,657 (892)	19,802 (1001)	28,843 (1417)	13,298 (636)	11,569 (583)	12,979 (564)	21,559 (912)
Resolution range (Å)	10–1.76	10–1.68	10–1.51	10–1.80	10–1.93	10–1.98	10–1.67
R_{work} ($F_o > 4\sigma F_o$; F_o) (%)	15.2; 16.0	14.9; 16.2	13.2; 13.3	14.1; 15.3	15.1; 16.3	14.9; 16.0	14.9; 16.0
R_{free} ($F_o > 4\sigma F_o$; F_o) (%)	21.6; 23.1	21.8; 23.9	18.8; 19.1	20.8; 24.4	23.8; 25.6	22.7; 24.4	18.3; 19.3
Protein residues	223	223	223	223	223	223	223
Inhibitor atoms	27	28	29	27	28	27	28
Water molecules	172	175	199	229	164	150	166
Calcium ions	1	1	1	1	1	1	1
RMSD							
Bond lengths (Å)	0.008	0.009	0.011	0.008	0.007	0.007	0.010
Bond angles (°)	2.1	2.1	2.3	1.9	1.9	1.8	2.3
Mean *B*-factor							
Wilson *B*-factor (Å^2)	23.2	20.1	22.7	16.3	20.3	20.5	15.9
Protein (Å^2)	25.7	18.4	27.3	18.4	22.6	23.1	10.8
Main chain (Å^2)	23.7	16.5	25.6	16.4	20.6	21.1	9.3
Side chains (Å^2)	28.1	20.8	29.3	20.9	25.0	25.5	12.7
Inhibitor (Å^2)	25.9	20.6	26.5	20.9	23.5	24.4	11.9
Water molecules (Å^2)	33.4	28.7	37.0	29.5	29.7	30.3	25.9
Ramachandran plot							
Most favored (%)	88.3	88.8	87.2	88.3	85.1	86.7	83.3
Additionally allowed (%)	11.7	11.2	12.8	11.7	14.9	13.3	11.7
Generously allowed (%)	0.0	0.0	0.0	0.0	0.4	0.0	0.4

larger aliphatic ring moieties. For ligands **1a** and **1b**, refinement converged to 40% and 35% occupancies of the hydrophobic pocket, whereas **1c** and **1d** showed higher levels of 70% and 61%, respectively.

In all cases, the electron density of the ligands in orientation B is well-defined for the linker atoms, including the secondary amine functional group, which is fixed by a hydrogen bond to Gly216O. In contrast, the aliphatic ring portion is not comparably well-defined. This suggests a higher residual mobility for the aliphatic ring systems outside the S3/S4 pocket. The reverse is observed for ligands in orientation A, which are accommodated inside the S3/S4 pocket. The electron density appears more consistent with the aliphatic ring portions, whereas the secondary amine functional group and the other linker atoms, all of which are not fixed via a hydrogen bond in this orientation, show a poorly

Fig. 3. Crystal structures of inhibitors bound to trypsin in the *open* form. Atoms are color-coded (carbon: green; oxygen: red; nitrogen: blue; sulfur: yellow), and the protein surface is shown in light gray. Water molecules are indicated as red spheres. (a–d) View on the active site of trypsin. Inhibitor atoms of ligands **1a** (a), **1b** (b), **1c** (c), and **1d** (d), and the $F_o - F_c$ difference electron density (2.5σ; dark blue) are shown. (e) Alignment of ligands **1a–1d**. The two alternative orientations, A (inside S3/S4 pocket) and B (outside S3/S4 pocket), are indicated. All amino acids within 3 Å of the ligands are shown as sticks.
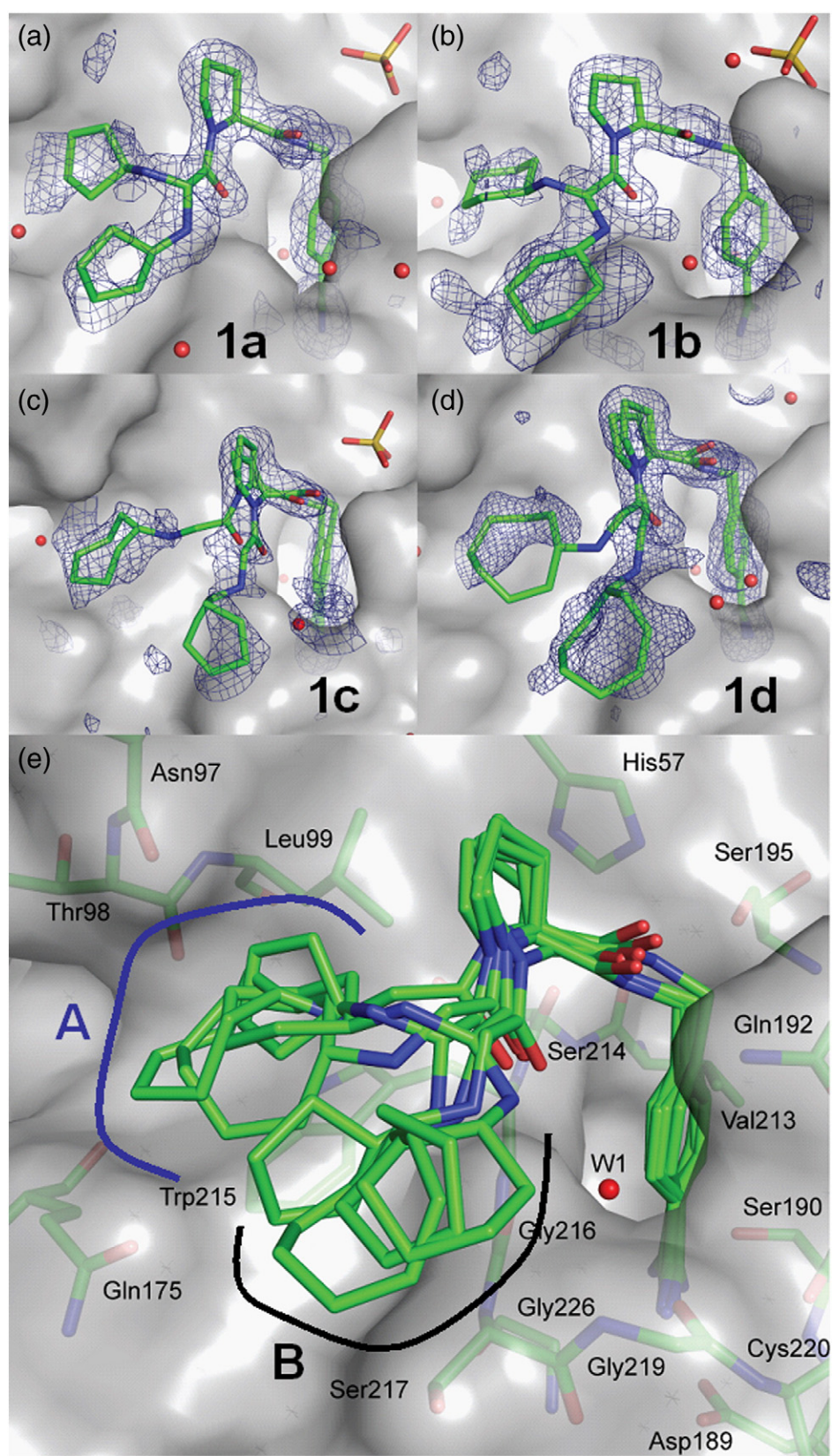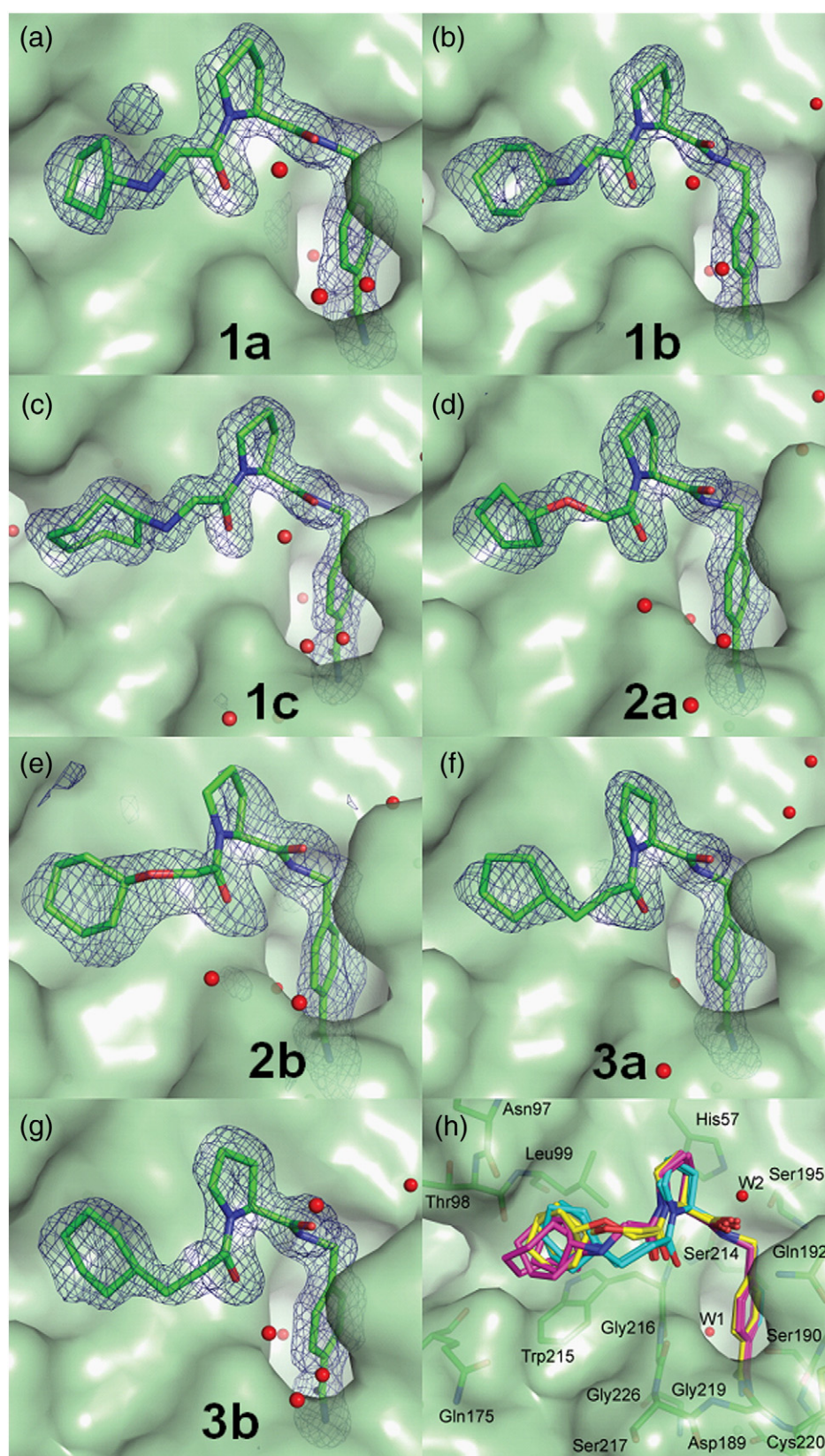
Fig. 3 (*legend on previous page*)

Fig. 4 (*legend on next page*)

defined electron density indicative of a high level of flexibility. This analysis is supported by considering the differences in the *B*-factors of the corresponding ligand portions.

The overall binding modes are summarized in Fig. 3e. The benzamidine groups form an almost symmetrical salt bridge to Asp189 (OD1: 2.7–3.0 Å; OD2: 2.7–3.0 Å). Further hydrogen bonds can be assigned between the benzamidine nitrogens and Ser190O (2.8–3.6 Å), Ser190OG (2.8–3.1 Å), Gly219O (2.6–3.1 Å), and a conserved water molecule W1 (2.7–3.4 Å). The amide nitrogen atom N3 forms a hydrogen bond to Ser214O (2.6–3.3 Å). All hydrogen distances vary by no more than 0.3 Å, excluding ligand **1c** in orientation B, which is slightly rotated out of the plane of the aromatic ring systems. Furthermore, a weak hydrogen bond to Gln192NE2 is observed, but distances vary due to a high level of flexibility and a poorly resolved electron density. Comparably strong hydrogen bonds between the linking amine nitrogen atoms N5 and Gly216O are formed only in orientation B (2.5–2.9 Å), whereas the contacts between these atoms are beyond H-bond distance (4.0–5.4 Å) in orientation A. A weaker hydrogen bond is observed between the carbonyl oxygen atom O2 and Gly216N (3.1–3.6 Å). If the oxygen atoms are refined for different orientations A and B (ligands **1c** and **1d**; Fig. 3e), the hydrogen-bonding distance is shorter for orientation B (3.1–3.3 Å) than for orientation A (3.5–3.6 Å). This indicates a cooperative behavior of this hydrogen bond in orientation B: It only forms once the N5-Gly216O contact has been equally established. The binding of the different ligands does not have a significant impact on protein geometry (data not shown).

We analyzed the crystallographically observed *B*-factors, which, with some care, correlate to a lower ligand occupancy or a higher residual mobility.[16] By comparing the *B*-factors of the benzamidine anchor inside the S1 pocket with the residues surrounding the binding site (within 4 Å), we found that the *B*-factors are generally smaller than or equal to 1 for ligand **1a** ($B_{bnz}/B_{bs} = 0.87$), ligand **1b** ($B_{bnz}/B_{bs} = 0.72$), and ligand **1c** ($B_{bnz}/B_{bs} = 1.03$), suggesting full ligand occupancy or close to full ligand occupancy. Ligand **1d** shows slightly higher *B*-factors, and the resulting ratio is 1.17. This could indicate an overall higher residual mobility or an incomplete occupancy. *B*-factors for atoms refined in two orientations were averaged. Additionally, the flexibility of the aliphatic ring portions of the ligands can be assessed by comparing their *B*-factors to those of the remaining ligand atoms. We found ratios of 1.43 (**1a**), 1.78 (**1b**), 1.79 (**1c**), and 1.88 (**1d**). These ratios indicate a higher residual mobility because a full occupancy has been assumed by the benzamidine portion.

Binding properties in the trigonal *closed* form

As mentioned, it was not possible to crystallize ligands **2a**, **2b**, **3a**, and **3b** in the *open* crystal form. Because the ether and methylene ligands do not possess a hydrogen-bond donor functionality in their linker, the open orientation (which places the hydrophobic cycloalkyl moiety outside the S3/S4 pocket) would not be favorable with respect to the formation of an additional H-bond. Instead, we obtained trypsin complex crystals for ligands **2a**, **2b**, **3a**, and **3b** with trigonal symmetry. Additionally, we crystallized ligands **1a**–**1c** in this space group for comparison. As mentioned, for **1d**, only crystals of poor diffraction quality could be obtained in this trigonal form. In comparison to the previously discussed orthorhombic complexes with **1a**–**1d**, the trigonal crystals show obvious differences in ligand binding modes. The electron densities (Fig. 4a–g) of all studied ligands show the aliphatic ring moieties accommodated inside the hydrophobic S3/S4 pocket, which corresponds to orientation A in the *open* crystal form. The accessible space occupied by ligand moieties in orientation B of the *open* form is obstructed by a crystal contact of a packing mate (Fig. S1, Supporting Information). We therefore named the trigonal crystals as the *closed* form. In the case of the amine ligands **1a**–**1c**, the aliphatic rings and the amine-linking portion are forced into the hydrophobic S3/S4 pocket; thus, they cannot form hydrogen bonds as in orientation B of the *open* form. The observation that no *open*-form crystals could be grown with either ether-linked (**2a** and **2b**) or methylene-linked (**3a** and **3b**) ligands suggests that they bind more favorably inside the S3/S4 pocket, particularly as they lack the capacity to form hydrogen bonds to Gly216O. We tried soaking crystals of uncomplexed trypsin in the orthorhombic form with ligand solutions; however, all attempts to retrieve stable crystals with a bound ligand failed. In all cases, the crystal either dissolved or became extremely fragile within a very short time.

Some of the trypsin–ligand complexes exhibit a noncontiguously connected electron density in this crystal form across the aliphatic ring portions (**1b**,

Fig. 4. Crystal structures of inhibitors bound to trypsin in the *closed* form. View of the active site of trypsin with bound inhibitors **1a** (a), **1b** (b), **1c** (c), **2a** (d), **2b** (e), **3a** (f), and **3b** (g). The surface of trypsin is shown in light green, and atoms are color-coded (carbon: green; oxygen: red; nitrogen: blue). The $F_o - F_c$ difference electron density is shown at 2.5σ in blue. Water molecules are indicated as red spheres. (h) Alignment of ligands with the carbons of amino ligands **1a**–**1c** (magenta), ether ligands **2a** and **2b** (yellow), and methylene ligands **3a** and **3b** (light blue). All amino acids within 3 Å are shown as sticks.

2a, **2b**, and **3a**). In **3a**, electron density is not well-defined for the linking methylene carbon atom. Nevertheless, we could refine the complete ligands and aliphatic ring portions as *envelope* conformation (five-membered rings) or *chair* conformation (six-membered rings). A *B*-factor analysis was performed similarly to that of the *open* form. The ratios of the *B*-factors of the benzamidine moiety compared to those of the binding site are approximately 1 (0.89–1.03). Thus, we can assume a full ligand occupancy. In most cases, the aliphatic ring portions of the ligands are far more flexible than their benzamidine moieties, exhibiting *B*-factor ratios well above 1 (1.57–2.18). Surprisingly, for ligands **1a** and **1c**, we found lower ratios of only 1.14 and 1.20, respectively.

Despite the difference in binding modes observed in the region of the S3/S4 pocket, the remaining part of the ligands binds to the *closed* form of trypsin similarly as to the *open* conformation (Fig. 3e). The salt bridge between N1, N2, and Asp189 is symmetrical (2.6–2.9 Å). Further hydrogen bonds can be observed between the benzamidine nitrogen atoms and Ser190O (3.1–3.3 Å), Ser190OG (2.9–3.2 Å), Gly219O (2.8–2.9 Å), and a conserved water molecule W1 (2.8–3.0 Å). The amide nitrogen N3 of the ligand is within hydrogen-bonding distance of Ser214O (2.9–3.0 Å). The neighboring carbonyl oxygen atom O1 of the ligand can form hydrogen bonds to Gln192NE2 (2.8–3.4 Å). Due to poorly defined difference electron density for the solvent-exposed Gln side chain, refinement was not as accurate; therefore, in the case of the trypsin–**1b** complex, the side chain had to be refined as an alanine. Furthermore, due to the very highly resolved data sets, we were able to refine a water molecule within hydrogen-bonding distance of this carbonyl oxygen atom O1. In contrast to the amine ligands in the *open* form, distances between carbonyl oxygen atom O2 and Gly216N suggest very weak hydrogen bonds (3.8–4.4 Å). This could be due to the fact that the ligands are forced into the S3/S4 pocket. Distances between the linking amino nitrogen atom N5 and Gly216O are beyond 4.5 Å in this crystal form. An additional H-bond is formed between N5 and Cys128CO of a symmetry-related molecule in the packing. None of the ligands induces significant conformational changes in the protein (data not shown).

We observed a subtle but remarkable difference in the binding modes of the ligands inside the S3/S4 pocket by superimposing the aligned part of the ligand scaffolds (Fig. 4h). Intriguingly, the adopted positions of the linker atoms differ markedly between the ligands. The more hydrophilic ether oxygen atoms of ligands **2a** and **2b** are repelled from the surface of the hydrophobic pocket and find a more solvent-exposed orientation. We could not detect a hydrogen-bonding partner with a crystal

mate in the packing (data not shown). The methylene carbon atoms of ligands **3a** and **3b** are placed more than 1 Å closer than **2a** and **2b** to the hydrophobic protein surface, probably because they can form more favorable van der Waals interactions. The position of the nitrogen atoms of the secondary amine-linked ligands **1a**–**1c** lies in between those of the ether oxygen and methylene carbon atom. The positions of the atoms of the aliphatic ring system exhibiting the same linker functional group superimpose quite well. However, once similarly sized aliphatic ring systems with different linkers are compared, this superposition is no longer observed. The aliphatic ring portions of amine-linked ligands **1a**–**1c** penetrate deeper into the pocket, whereas the methylene-linked ligands **3a** and **3b** are slightly removed from the pocket. The latter ligands also show different atomic positions for the entire linker along with parts of the proline portion. Given the fact that we observed the same phenomenon in thrombin (unpublished data), it is unlikely to be solely a crystallographic artifact. In contrast, the atomic positions of the benzamidine moiety for all ligands superimpose nicely, highlighting the anchoring character of this portion of the ligand.

MD simulations

To gain a better insight into the properties of the protein–ligand complexes in solution, we used MD simulations. We performed MD simulations of trypsin in complex with ligands **1a**–**1d**, **2a**, **2b**, **3a**, and **3b**, as well as for uncomplexed trypsin, over 10 ns. Starting structures were based on our crystallographic data sets in the *closed* form [Protein Data Bank (PDB) code 1HJ9[17] in the case of uncomplexed trypsin]. The simulated system was first heated to 300 K before any data were analyzed. In the following, data analysis will focus on the ligand binding site, with particular emphasis on areas that show a higher level of flexibility in either protein or ligand.

We analyzed first the dynamics of the protein, particularly to detect possible fluctuations of residues in the interior or next to the binding site. A good indication for protein stability along the trajectory is the RMSD calculated for C^{α} carbons with respect to the starting structure. In the beginning of the simulation, the RMSD increases to about 1 Å in all cases. This reflects the relaxation of the proteins strained by crystal packing. Should the protein lose stability during the simulation, the recorded RMSDs would increase extensively; however, this was not observed for any simulation (Table 4). We analyzed the RMSDs of the ligand binding pocket in great detail. For uncomplexed trypsin, the flexibility of the binding pocket is rather high. This flexibility is demonstrated by RMSDs that

Table 4. RMSDs and standard deviations for C$^\alpha$ atoms and the binding sites of all simulated trypsin–ligand complexes

		Ligand								
		1a	1b	1c	1d	2a	2b	3a	3b	Free trypsin
Average	C$^\alpha$	1.38±0.13	1.26±0.16	1.13±0.10	1.02±0.12	1.26±0.21	1.32±0.18	1.35±0.26	1.46±0.39	1.35±0.21
RMSD(Å)	Binding site	1.12±0.14	1.09±0.14	1.12±0.19	1.24±0.20	1.09±0.16	1.24±0.21	1.08±0.17	1.14±0.19	1.57±0.17

The binding sites are defined by the residues His57, Leu99, Asp189, Ser190, Gln192, Ser195, Ser214, Trp215, Gly216, and Gly219.

oscillate about the overall average and by a two-dimensional RMSD plot that shows no clustering (Fig. S2a, Supporting Information). This flexibility is reduced once a ligand has been introduced in the binding pocket. In general, similar effects are observed when any ligand is bound (shown for ligand **1a** in Fig. S2c, Supporting Information); however, as a trend, we observe a slightly more pronounced clustering into distinctive conformational states for the ligands with larger aliphatic ring portions. The residues of the S1 subpocket (Asp189, Ser190, and Gly219) sacrifice much of their flexibility in the complexed state. This can be explained by the extensive network of hydrogen bonds that is formed upon ligand binding. Within the S2 pocket, His57, Ser195, and Ser214 are rather rigid, whereas Gln192 is highly mobile, corresponding to the findings in our crystal structures. Residues Leu99, Trp215, and Ser217 of the S3/S4 subpocket also lose their mobility upon ligand binding. Nevertheless, subsequently, they switch between a set of distinctive but similar conformational states. This correlates with several possible binding modes of the ligands in this area, again in accordance with our crystallographic observations.

We compared the motions of separate portions of the ligands addressing the subpockets of the active site. The mobility of the ligand inside the S1 pocket is highly constrained by the salt bridge of the ligand's benzamidine group and by Asp189OD1 and OD2. We analyzed the distances between N1 and AspOD1 and between N2 and Asp189OD2. The average values are 2.8–2.9 Å for all ligands, in agreement with the starting values in the crystal structures. Small fluctuations, which rarely exceed 3.5 Å, can be observed (Fig. S3a and b, Supporting Information).

Additionally, the aromatic portion of the benzamidine moiety fluctuates relative to the orientation observed in the crystal structure, although its movement is constrained in comparison to its properties in solution. Simulations of the free ligands in solution show that the benzamidine moiety adopts four main conformations (angle C$_{ar}$–C$_{ar}$–C$_{cat}$–N$_{pl3}$: −143°, −37°, 37°, 143°) and rotates randomly between them (data not shown). In complex with trypsin, this rotational movement is strongly constrained. Nevertheless, the phenyl ring keeps some flexibility and rotates around its axis by 40–50° (Fig. S4, Supporting Information). Overall, the ligands are bound tightly inside the S1 pocket and lose much of their flexibility exhibited in solution.

The S2 pocket of trypsin is flanked by residues His57, Gln192, Ser195, and Ser214. We showed that residue Gln192 is highly mobile and is therefore not appropriate for recording ligand movements. We analyzed six interatomic distances between Ser195OG, as well as Ser214O, and the ligand atoms C5, O1, and C10 in order to trace the possible ligand mobility of the central proline portions. In most cases, the ligands are positioned tightly inside the S2 pocket, and the distances show no major variation. Thus, only deviations from this pattern will be discussed. The simulations suggest that rotation of the amide bond between proline and benzamidine moiety within the S2 pocket is possible; however, this flexibility depends on the size of the terminal aliphatic ring system. Frequent rotation can be observed for the small five-membered ring analogs independent of the actual linker groups. As the aliphatic ring becomes larger, the rotation of the amide group appears to be restricted, and rotational movements occur rarely (shown for **1a**; Fig. S5, Supporting Information).

Finally, we analyzed the behavior of the ligands inside the hydrophobic S3/S4 pocket and concentrated on the distances between Gly216N and the carbonyl oxygen O2. The distance between the central atom of the linker group X (**1a–1d**: N5; **2a** and **2b**: O3; **3a** and **3b**: CX) and Gly216O was also recorded to detect whether the terminal aliphatic ring moieties are accommodated within the pocket. An analysis of the torsion angles 15, 16, and 17 gives information on the flexibility of the ligands inside the pocket.

The cyclopentyl ring of **1a** is only bound in the S3/S4 pocket at the beginning of the trajectory and moves out of this pocket for the remainder of the simulation. Accordingly, both recorded distances are indicative of hydrogen bonding (Fig. 5a; on average, 3.2 Å for Gly219N-O2 and 2.9 Å for Gly216O-N5), in accordance with the *open*-form crystal structure. Even though the two hydrogen bonds are formed, the ligand remains highly flexible. This is illustrated by several snapshots evenly collected along the trajectory (Fig. 5e).
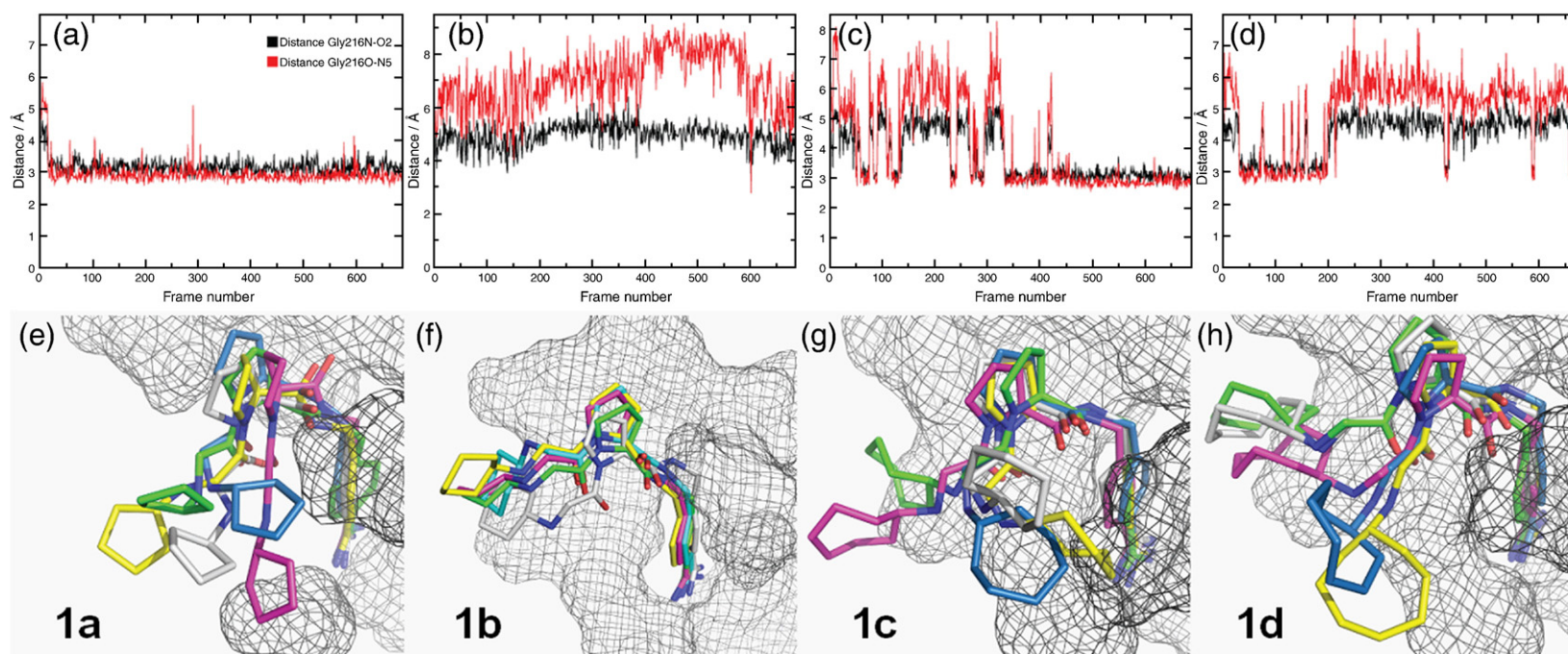
Fig. 5. MD simulations of amino ligands **1a**–**1d**. (a–d) Distance plots (every 30th frame shown) between atoms Gly216N-O2 (black) and Gly216O-N5 (red) for ligands **1a** (a), **1b** (b), **1c** (c), and **1d** (d). (e–h) Snapshots of representative conformations within the S3/S4 binding pocket of trypsin for ligands **1a** (e), **1b** (f), **1c** (g), and **1d** (h). Atoms are color-coded (carbon: green, gray, magenta, yellow, or light blue; oxygen: red; nitrogen: dark blue), and the protein surface of amino acids within 3 Å is shown as a mesh.

The secondary amine-linked ligand **1b** with a cyclohexyl ring behaves differently. The distances between the putatively hydrogen-bonded atoms remain, on average, well beyond 4 Å (Gly219N-O2: 4.9 Å; Gly219O-N5: 7.0 Å), indicating that no stable hydrogen bonds are formed (Fig. 5b). Intermediately, the amino nitrogen of the amide bond orients towards the solvent and increases the distance to Gly216O to >8 Å, allowing a water molecule to be inserted between the two atoms. A hydrogen bond between Gly219O and the linking nitrogen N5 is formed only for a very few frames. Consequently, the cyclohexyl portion remains accommodated inside the S3/S4 pocket along the whole trajectory (Fig. 5f).

Remarkably, the secondary amine-linked ligand **1c**, which bears a terminal cycloheptyl ring, is found both inside and outside the S3/S4 pocket at different stages during the simulation. The two hydrogen bonds analyzed (Fig. 5c) are formed cooperatively. In the first part of the simulation, the ligand resides more frequently in the S3/S4 pocket with average distances of >4.7 Å (Gly219N-O2) and 5.9 Å (Gly219O-N5), correlating well to the crystallographic data (4.7 and 5.9 Å, respectively). In the second part of the trajectory, the distances fall most of the time within the range of a hydrogen bond (3.1 and 3.0 Å). With this geometry, the cycloheptyl terminus remains outside the S3/S4 pocket, and the ring adopts many different orientations while still maintaining the hydrogen bonds. Snapshots illustrating these conformational changes are shown in Fig. 5g.

The distances between Gly216N/O2 and Gly216O/N5 for the amino-linked cyclooctyl derivative **1d** show a pattern similar to that for **1c** (Fig. 5d). Nevertheless, representative snapshots exemplify that the larger ring moiety allows this ligand to form an additional binding mode. Between frames 40 and 200, hydrogen bonds are formed (3.2 and 2.9 Å), with the cyclooctyl ring accommodated inside the S3/S4 pocket. This ring is large enough to fill the pocket without disrupting its hydrogen bonds to Gly216. Towards the end of the trajectory, the measured distances fall again into the range of hydrogen bonding; however, this time, the aliphatic ring portion remains outside the S3/S4 pocket. Apparently, an even larger conformational multiplicity is possible for this derivative (Fig. 5h), although binding inside the S3/S4 pocket constrains the mobility of the linker and the terminal ring.

The ether-linked ligands **2a** and **2b** and the methylene-bridged ligands **3a** and **3b** form hydrogen bonds between Gly216N and O2 (3.2, 3.1, 3.2, and 3.4 Å, respectively; Fig. 6a–d). This hydrogen bond measures about 4 Å in the crystal structures. The formation of the hydrogen bond is nevertheless remarkable, as we observed this hydrogen bond for the secondary amine-linked ligands **1a**–**1d** only once the second hydrogen bond between Gly216O and N5 has also been formed. A similar contact to the latter hydrogen bond cannot be formed by the ether (oxygen atom O3) and methylene (carbon atom CX) bridges in **2a**, **2b**, **3a**, and **3b**. The distances of O3 and CX to Gly216O differ considerably and have to be analyzed in detail.

The cyclopentyl ring in the ether-linked **2a** ligand always remains in the S3/S4 pocket during the simulation, in contrast to the secondary amine-linked derivative **1a**. This is illustrated by the changes in distance between the ether oxygen atom O3 and Gly216O, which fluctuates between 4 and 8 Å (Fig. 6a). Movements inside the S3/S4 pocket appear rather constrained in comparison to a simulation of the free ligand. These findings are confirmed by snapshots of representative frames along the trajectory (Fig. 6e). As already observed in the crystal structures, the oxygen atom of the bridging ether orients towards the solvent in all conformations.

Enhanced flexibility is observed for the ether-linked cyclohexyl derivative **2b**. The distance between Gly216O and O3 varies (on average, 5.1 Å) and drops towards the end of the simulation to 4.2 Å (Fig. 6f). Analyzing in detail the involved torsion angles suggests that the terminal cyclohexyl portion of the ligand remains inside the S3/S4 pocket, which restrains its rotation throughout most of the simulation. At the end of the simulation, it leaves the pocket (Fig. 6f), which induces strongly enhanced rotational motions.

The methylene-linked **3a** with a terminal cyclopentyl ring shows an average distance of 4.1 Å between its bridging carbon atom CX and Gly216O, which is considerably shorter than that for the ether derivative **2a** (Fig. 6c). Snapshots of conformations along the trajectory (Fig. 6g) show that the bridging carbon atom CX always points towards the S3/S4 pocket, as found in the crystal structures. This orientation enhances favorable van der Waals interactions. Along the trajectory, the cyclopentyl portion remains within the S3/S4 pocket.

Finally, we analyzed the movements of the methylene-linked cyclohexyl ligand **3b**. As mentioned, the hydrogen bond between Gly216N and O3 is, on average, slightly longer (3.4 Å) than that in the complexes with **2a**, **2b**, and **3a**. Most likely, this hydrogen bond is weaker and more easily broken, allowing for high fluctuations in the distance between both atoms (Fig. 6d). The distance between atoms CX and Gly216O is even less constrained, and no correlation with the latter distance can be observed. Snapshots of representative conformations along the trajectory (Fig. 6h) confirm that the terminal aliphatic ring and the linking group can move quite freely between conformations, which resemble those found in
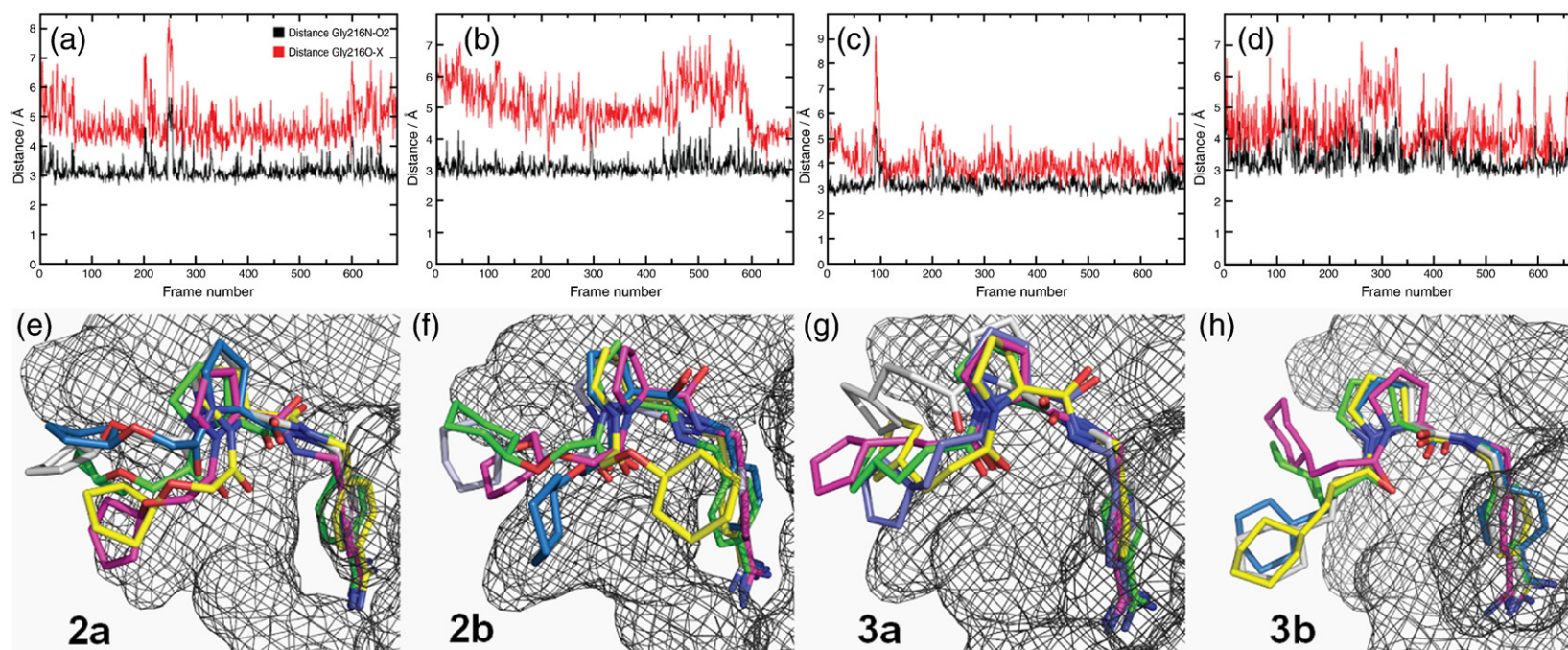
Fig. 6. MD simulations of ether ligands **2a** and **2b** and methylene ligands **3a** and **3b**. (a–d) Distance plots (every 30th frame shown) between atoms Gly216N-O2 (black) and Gly216O-X (red; X = O3, CX) for ligands **2a** (a), **2b** (b), **3a** (c), and **3b** (d). (e–h) Snapshots of representative conformations within the S3/S4 binding pocket of trypsin for ligands **2a** (e), **2b** (f), **3a** (g), and **3b** (h). Atoms are color-coded (carbon: green, gray, magenta, yellow, or light blue; oxygen: red; nitrogen: dark blue), and the protein surface of amino acids within 3 Å is shown as mesh.

the crystal structure inside and outside the S3/S4 pocket.

Overall, the MD simulations provide valuable information on the dynamic properties of the ligands in complex with trypsin. Changes made at the terminal region of the inhibitors located inside the S3/S4 pocket have a remarkable impact on the flexibility experienced in the S2 pocket and even in the S1 pocket. Trends that depend on the size of the terminal aliphatic ring portions (such as a pronounced rotation of the amide group in the S2 pocket) are observed particularly for the smaller terminal rings. The linker group plays a crucial role in the formation of hydrogen bonds to Gly216. Cooperative hydrogen-bond formation is also detected for the amine-linked ligands.

Discussion

A combination of different biophysical techniques allowed us to study the protein–ligand interactions of a congeneric series of closely related ligands with trypsin. Across this data set, minor changes in the ligand scaffold lead to large differences in biophysical binding properties. A pronounced enthalpy/entropy compensation between the ligands results in an overall nearly unchanged free energy of binding. Our diverse data set, including MD simulations and crystal structures, provides a number of qualitative correlations with respect to the observed thermodynamic profiles. One has to keep in mind that several effects can compensate for each other, leading to the same thermodynamic profile despite differences in the binding modes.

For instance, there are possible energetic differences that result from the desolvation of the ligands while they are being transferred from the bulk water phase to the protein binding site. It is by no means obvious that the enthalpic and entropic contributions to this desolvation step are equivalent across the studied congeneric series, particularly when considering that a cyclopentyl ring has very different conformational properties compared to a six-membered, seven-membered, or eight-membered ring. We are in a comfortable situation to have data from two different crystal forms and MD simulations. This combination provides details about the preferred ligand binding modes also under biologically relevant conditions.

Particularly, the secondary amine-linked derivatives **1a**–**1d** scatter over two distinct families of binding modes. One ligand binding mode experiences hydrophobic interactions inside the S3/S4 pocket but lacks formation of hydrogen bonds. The other mode forms the hydrogen bonds but leaves the hydrophobic pocket unoccupied. Intriguingly, the respective populations for the two overall modes and their mutual interconversion result in similar thermodynamic profiles. In general, we observe a slight increase in entropic contributions for the ligands with larger terminal rings. The MD simulations help to rationalize this trend. For the small five-membered ring, derivative **1a** binding outside the S3/S4 pocket seems to be most favorable when two hydrogen bonds are formed. In contrast, the cyclohexyl moiety of **1b** remains more tightly bound in the S3/S4 pocket during the simulation and lacks the second hydrogen bond, yielding an entropically more beneficial thermodynamic profile. Derivatives **1c** and **1d** with the larger terminal cycloheptyl and cyclooctyl portions adopt more than one preferred orientation, which are easily interconverted. This might explain their increasing entropic advantage in the thermodynamic profile.

In contrast, the ether-linked and methylene-linked derivatives **2a**, **2b**, **3a**, and **3b** prefer to bind to the hydrophobic pocket with their terminal cycloalkyl moieties and do not form a permanent hydrogen bond between Gly216N and the carbonyl oxygen atom O2. Binding with less residual mobility possibly explains the more enthalpically favored binding profiles of **2a** and **2b**. The ether oxygen atom O3 remains solvent exposed in both crystal structures and MD simulations, in agreement with the expectation that this ether oxygen would be more difficult to desolvate than the methylene linker in **3a**/**3b**.

The methylene-linked and secondary amine-linked cyclopentyl ligands **3a** and **1a** show similar thermodynamic profiles, but for different reasons. The cyclopentyl portion of **3a** resides permanently inside the S3/S4 pocket along the MD trajectory, in contrast to **1a**. This suggests balanced contributions of hydrogen bonding and hydrophobic contacts. The methylene ligand **3b**, which contains a terminal cyclohexyl ring, binds with more entropically favored energetics due to fewer constraints in the rotational degrees of freedom of the terminal cyclohexyl group, as observed during the MD simulations.

Interestingly, the detailed analysis of the dynamic ligand properties provides an explanation on why we experienced problems in crystallizing the highly flexible **1d** in the conformationally restricting *closed* form. In contrast, **2a**, **2b**, **3a**, and **3b** show preference for one overall binding mode within the S3/S4 pocket; as a consequence, they do not crystallize in the *open* form, which accommodates the ligands with scattered orientations.

The secondary amine ligands exhibit a slightly higher affinity (~ 1 kJ/mol) for trypsin as compared to the homologous ether and methylene ligands, as confirmed in kinetic inhibition assays. Without knowledge of thermodynamic profiles, structures, and dynamics, one might anticipate superficially

that the secondary amine functional group forms an additional affinity-enhancing hydrogen bond with the protein, which cannot be formed by the other ligands. Our results strongly suggest that such a simple interpretation is clearly not valid and that much more subtle and complex differences in binding modes and residual mobilities are in operation.

Conclusions

The failure to determine crystal structures with potent ligands is a frustrating obstacle that is well-known to many crystallographers. The present study provides a possible explanation for this disappointing phenomenon. Furthermore, our work shows that even though congeneric ligands exhibit virtually identical affinities, their binding profiles at structural and thermodynamic levels can still be very distinct.

Optimizing subtle energetic differences between ligands is crucial to addressing a given target. Depending on the target's properties, quite different binding profiles will be required. Membrane-bound proteins in the nervous system demand properties different from, for instance, viral proteins. In one case, a drug candidate needs, for example, high isoform selectivity or promiscuous protein familywide binding, while in another situation, efficient tolerance against resistance mutations is desirable. This calls for very different ligand binding characteristics. In principle, binding events can be either enthalpy-driven or entropy-driven, for instance by rigid preorganized shape complementarity or pronounced residual mobility at the binding site.[4] These properties determine interaction kinetics that influence the on-rates/off-rates and residence times of drugs spent with a target. A detailed profile analysis, as described in this study, will be the prerequisite to subsequently tackle the problem to appropriately adjust ligand properties for a given target.

Materials and Methods

Synthesis

The synthesis of compounds **1a**, **1b**, **3a**, and **3b** has been described previously[18]. The synthesis of the remaining compounds is explained in detail in Supporting Information.

Kinetic inhibition assay

Kinetic inhibition of bovine trypsin (Sigma-Aldrich) was determined photometrically at 405 nm using the chromogenic substrate Pefachrom tPa (LoxoGmbH, Dossenheim, Germany) according to the protocols described by Sturzebecher *et al.* under the following conditions: 50 mM Tris/HCl (pH 7.8), 154 mM NaCl, 5% dimethyl sulfoxide, and 0.1% polyethylene glycol (PEG) 8000 at 25 °C using different concentrations of substrate and inhibitor.[19] K_i values were determined at least in triplicate, as described by Dixon.[20]

Isothermal titration calorimetry

ITC experiments were carried out using an MCS titration calorimeter[21] (Microcal, Inc., Northampton, MA). Inhibitor stock solutions were prepared by weighing of the corresponding hydrochlorides and dilution in the desired buffer to a final concentration of 0.75 mM. Protein solution was prepared freshly for each experiment. ITC measurements were routinely performed at 25 °C in 50 mM Tris/HCl (pH 7.8), 100 mM NaCl, and 0.1% PEG 8000. Inhibitor solutions were degassed (approximately 10 min) prior to use and titrated into a stirred cell (1.3513 ml) containing trypsin solution (0.075–0.15 mM) after a stable baseline had been achieved. The injection sequence consisted of an initial injection of 1.5 μl to prevent artifacts arising from the filling of the syringe (not used in data fitting), followed by injections of 6–10 μl every 5 min until a complete saturation of the enzyme binding site had been achieved. Raw data were collected, and the area under each peak was integrated, followed by correction for heats of dilution and mixing by subtracting the final baseline. Data were analyzed using the ORIGIN Software (Microcal, Inc.) by fitting to a single-site binding isotherm,[21] which yields ΔH^0 (enthalpy of binding) and K_d (dissociation constant). Measurements were performed at least in triplicate. The buffer dependence of ΔH^0 was tested for compounds **1a** and **3a** in Tris, Hepes, and N-(2-hydroxy-1,1-bis(hydroxymethyl)ethyl)glycine (Tricine) buffer (pH 7.8). The results were analyzed by linear regression (**1a**: $y = -0.096x - 15.881$; **3a**: $y = -0.111x - 15.290$).

Crystallization

Crystals of bovine trypsin in complex with ligands in *closed* form (inhibitors **1a**–**1c**, **2a**, **2b**, **3a**, and **3b**) and in *open* form (inhibitors **1a**–**1d**) were prepared using the sitting-drop vapor diffusion method.

For *closed*-form crystals, inhibitors (25 μg/μl) were dissolved in 1:1 dimethyl sulfoxide/H_2O, and bovine trypsin (20 mg/ml) was dissolved in 10 mM $CaCl_2$ and 1 mM HCl. Fifty microliters of trypsin solution, 10 μl of inhibitor solution, and 40 μl of water were mixed and incubated for 30 min at 16 °C. We used 100 mM imidazole (pH 7 or pH 8), 100 mM $(NH_4)_2SO_4$, PEG 8000 (25–35%), and 0.1% NaN_3 as crystallization buffer. Crystals were obtained within a week.

Trypsin–ligand complexes in *open* form were obtained by dissolving inhibitors (3 mM) in 50 mM Tris/HCl (pH 7.8), 100 mM NaCl, and 0.1% PEG 8000, and by dissolving bovine trypsin (60 mg/ml) in 10 mM $CaCl_2$ and 1 mM HCl. Both solutions were mixed 1:1 and incubated for 30 min. A 2:1 crystallization buffer [50 mM

4-morpholineethanesulfonic acid/NH₃ and 1.9 M ammonium sulfate (pH 6)]/H₂O mix was used for the initial 24 h and then replaced by pure crystallization buffer. After 24 h, precipitation was visible, and microseeding with *open* trypsin crystals (prepared with 10 mM benzamidine solution instead of ligands) was performed. Crystals suitable for X-ray diffraction grew very slowly (up to 4 months). No crystals were obtained for inhibitors **2a**, **2b**, **3a**, and **3b**.

Data collection and processing

Crystals were prepared for data collection at 110 K using a cryoprotectant solution of 30% glycerol (wt/vol) in crystallization buffer. The data sets for **1b** (*closed* and *open*) and **1d** (open) were collected with synchrotron radiation at BESSY beamline 14.2 on a Mar CCD 165-mm detector. Diffraction data for **2a** (*closed*) were collected at DESY beamline X12 on a Mar CCD 225-mm detector. The remaining data sets were collected in-house using radiation from a rotating anode (Rigaku RU-H3R, 44–50 kV, 80–100 mA, 1.5418 Å Cu-Kα radiation; Rigaku, Austin, TX) and an image plate detector (R-AXIS IV; Rigaku). Data processing and scaling were performed using the HKL2000 package.[22]

Structure determination and refinement

The data set for trypsin in *closed* form in complex with **3b** was solved with molecular replacement using the program Phaser.[23] As starting structure, we used a trypsin structure of the same space group (PDB code 1GI1) at a resolution of 1.42 Å.[24] The coordinates of *open* bovine trypsin (PDB code 1K1P)[25] were used for the initial rigid-body refinement of the protein molecules (for data up to 3.5 Å), followed by repeated cycles of conjugate gradient energy minimization, simulated annealing, and *B*-factor refinement using the CNS program package.[26] Refinement at later stages was performed with the program SHELXL.[27] Here, at least 20 cycles of conjugate gradient minimization were performed with default restraints on bonding geometry and *B*-values. Five percent of all data were used for R_{free} calculations. Restraints were applied to bond lengths and angles, chiral volume, planarity of aromatic rings, and van der Waals contacts. Amino acid side chains were fitted into σ-weighted $2F_o−F_c$ and $F_o−F_c$ electron density maps using Coot.[28] After the first refinement cycle, water molecules and, subsequently, ions and ligand atoms were located in the electron density and added to the model. Ligand geometries were energy-minimized using SYBYL 8.0 (Tripos, Inc., St. Louis, MO), and restraints were calculated with the Monomer Library Sketcher software.[29] Multiple side-chain conformations were built in case an appropriate electron density was observed and maintained during the refinement, and if the minor populated side chain showed at least 10% occupancy. Side chains lacking appropriate electron density (*B*-factors >100) were refined as alanines. During the last refinement cycles, riding H-atoms were introduced without additional parameters. The final models were validated using PROCHECK.[30] Data collection, unit cell parameters,

and refinement statistics are given in Tables 2 and 3. Analysis of temperature factors was executed with Moleman,[31] and distances and angles were measured in PyMOL 0.99 (DeLano Scientific, Palo Alto, CA). Figures were prepared using Isis Draw (MDL, San Leandro, CA) and PyMOL 0.99. The accession codes are listed in Tables 2 and 3.

MD simulations

The protein–ligand complexes were placed in a box of approximately 7400 TIP3P water molecules. For the ligands, the Amber gaff force field was used; the benzamidine group was parameterized in accordance with McDowell *et al.*[32] The atomic charges of the ligands were calculated by fitting (program RESP) to the HF/6-31G* electrostatic potential (Gaussian 03) based on the crystallographically observed conformation. MD simulations were started by heating the solvent to 300 K over a period of 20 ps and by cooling to 100 K over a period of 5 ps, keeping the solute fixed. Then the system was gradually brought to 300 K over 25 ps. The simulation was carried on for 10 ns under constant temperature and pressure (NPT) and periodic boundary conditions (Amber 9.0). Solute coordinates were stored every 0.5 ps, and solvent coordinates were stored every 5 ps. All results presented refer to every 30th snapshot of 10-ns trajectories, excluding the first 100 ps required for temperature adjustment and equilibration (an equilibrated state with respect to the total potential energy in the system was reached after 100 ps). Analyses were carried out primarily using modules from the Amber program suite, while VMD[33] and PyMOL‡ were used for visualization purposes.

PDB accession codes

Coordinates and structure factors have been deposited in the PDB with the following accession numbers: *open* form: Try-**1a** (3LJJ), Try-**1b** (3LJO), Try-**1c** (2ZHD), and Try-**1d** (2ZQ2); *closed* form: Try-**1a** (2ZFT), Try-**1b** (2ZQ1), Try-**1c** (2ZFS), Try-**2a** (2ZDL), Try-**2b** (2ZDM), Try-**3a** (2ZDN), and Try-**3b** (2ZDK).

Supplementary materials related to this article can be found online at doi:10.1016/j.jmb.2010.11.038

‡ http://www.pymol.org

References

1. Bilello, J. A. (2005). The agony and ecstasy of "OMIC" technologies in drug development. *Curr. Mol. Med.* **5**, 39–52.

2. Vidalin, O., Muslmani, M., Estienne, C., Echchakir, H. & Abina, A. M. (2009). *In vivo* target validation using gene invalidation, RNA interference and protein functional knockout models: it is the time to combine. *Curr. Opin. Pharmacol.* **9**, 669–676.

3. Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. R. & Hofsteenge, J. (1989). The refined 1.9 Å crystal structure of human alpha-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J.* **8**, 3467–3475.

4. Ladbury, J. E., Klebe, G. & Freire, E. (2010). Adding calorimetric data to decision making in lead discovery: a hot tip. *Nat. Rev. Drug Discov.* **9**, 23–27.

5. Böhm, H. J., Klebe, G. & Kubinyi, H. (1996). *Wirkstoffdesign der Weg zum Arzneimittel.* Spektrum Akademischer Verlag, Heidelberg, Germany.

6. Krishnamurthy, V. M., Bohall, B. R., Semetey, V. & Whitesides, G. M. (2006). The paradoxical thermodynamic basis for the interaction of ethylene glycol, glycine, and sarcosine chains with bovine carbonic anhydrase II: an unexpected manifestation of enthalpy/entropy compensation. *J. Am. Chem. Soc.* **128**, 5802–5812.

7. Copeland, R. A., Pompliano, D. L. & Meek, T. D. (2006). Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discov.* **5**, 730–739.

8. Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Dtsch. Chem. Ges.* **27**, 2985–2993.

9. Koshland, D. E. (1958). Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl Acad. Sci. USA*, **44**, 98–104.

10. Bode, W. (1979). Activation, activity and inhibition of bovine trypsin. *Naturwissenschaften*, **66**, 251–258.

11. Wang, Y., Luo, W. & Reiser, G. (2008). Trypsin and trypsin-like proteases in the brain: proteolysis and cellular functions. *Cell. Mol. Life Sci.* **65**, 237–252.

12. Hedstrom, L. (2002). Serine protease mechanism and specificity. *Chem. Rev.* **102**, 4501–4524.

13. Andrews, P. R., Craik, D. J. & Martin, J. L. (1984). Functional group contributions to drug–receptor interactions. *J. Med. Chem.* **27**, 1648–1657.

14. Kuntz, I. D., Chen, K., Sharp, K. A. & Kollman, P. A. (1999). The maximal affinity of ligands. *Proc. Natl Acad. Sci. USA*, **96**, 9997–10002.

15. Dunitz, J. D. (1995). Win some, lose some: enthalpy–entropy compensation in weak intermolecular interactions. *Chem. Biol.* **2**, 709–712.

16. Baum, B., Muley, L., Smolinski, M., Heine, A., Hangauer, D. & Klebe, G. (2010). Non-additivity of functional group contributions in protein–ligand binding: a comprehensive study by crystallography and isothermal titration calorimetry. *J. Mol. Biol.* **397**, 1042–1054.

17. Leiros, H. K., McSweeney, S. M. & Smalas, A. O. (2001). Atomic resolution structures of trypsin provide insight into structural radiation damage. *Acta Crystallogr. Sect. D*, **57**, 488–497.

18. Gerlach, C., Smolinski, M., Steuber, H., Sotriffer, C. A., Heine, A., Hangauer, D. G. & Klebe, G. (2007). Thermodynamic inhibition profile of a cyclopentyl and a cyclohexyl derivative towards thrombin: the same but for different reasons. *Angew. Chem. Int. Ed. Engl.* **46**, 8511–8514.

19. Sturzebecher, J., Sturzebecher, U., Vieweg, H., Wagner, G., Hauptmann, J. & Markwardt, F. (1989). Synthetic inhibitors of bovine factor Xa and thrombin comparison of their anticoagulant efficiency. *Thromb. Res.* **54**, 245–252.

20. Dixon, M. (1972). The graphical determination of K_m and K_i. *Biochem. J.* **129**, 197–202.

21. Wiseman, T., Williston, S., Brandts, J. F. & Lin, L. N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179**, 131–137.

22. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.

23. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.

24. Katz, B. A., Elrod, K., Luong, C., Rice, M. J., Mackman, R. L., Sprengeler, P. A. *et al.* (2001). A novel serine protease inhibition motif involving a multi-centered short hydrogen bonding network at the active site. *J. Mol. Biol.* **307**, 1451–1486.

25. Dullweber, F., Stubbs, M. T., Musil, D., Sturzebecher, J. & Klebe, G. (2001). Factorising ligand affinity: a combined thermodynamic and crystallographic study of trypsin and thrombin inhibition. *J. Mol. Biol.* **313**, 593–614.

26. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D*, **54**, 905–921.

27. Sheldrick, G. M. & Schneider, T. R. (1997). SHELXL: high-resolution refinement. *Methods Enzymol.* **277**, 319–343.

28. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D*, **60**, 2126–2132.

29. Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D*, **50**, 760–763.

30. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.

31. Kleywegt, G. J., Zou, J. Y., Kjeldgaard, M. & Jones, T. A. (2001). Around O. In *International Tables for Crystallography: Vol. F. Crystallography of Biological Macromolecules* (Rossmann, M. G., & Arnold, E., eds), vol. F, pp. 353–356, 366–367, Kluwer Academic Publishers, Dordrecht, The Netherlands.

32. McDowell, L. M., McCarrick, M. A., Studelska, D. R., Guilford, W. J., Arnaiz, D., Dallas, J. L. *et al.* (1999). Conformations of trypsin-bound amidine inhibitors of blood coagulant factor Xa by double REDOR NMR and MD simulations. *J. Med. Chem.* **42**, 3910–3918.

33. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: Visual Molecular Dynamics. *J. Mol. Graphics*, **14**, 33–38, 27–28.