Structural determinants of trypsin affinity and specificity for cationic inhibitors

FABIO POLITICELLI,^{1,2} PAOLO ASCENZI,² MARTINO BOLOGNESI,³ AND BARRY HONIG¹

¹Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032

²Department of Biology, University of Rome TRE, 00146 Rome, Italy

³Advanced Biotechnology Center-IST, Department of Physics and INFN, University of Genova, 16132 Genova, Italy

(RECEIVED June 15, 1999; ACCEPTED September 21, 1999)

Abstract

The binding free energies of four inhibitors to bovine β -trypsin are calculated. The inhibitors use either ornithine, lysine, or arginine to bind to the S_1 specificity site. The electrostatic contribution to binding free energy is calculated by solving the finite difference Poisson–Boltzmann equation, the contribution of nonpolar interactions is calculated using a free energy-surface area relationship and the loss of conformational entropy is estimated both for trypsin and ligand side chains. Binding free energy values are of a reasonable magnitude and the relative affinity of the four inhibitors for trypsin is correctly predicted. Electrostatic interactions are found to oppose binding in all cases. However, in the case of ornithine- and lysine-based inhibitors, the salt bridge formed between their charged group and the partially buried carboxylate of Asp189 is found to stabilize the complex. Our analysis reveals how the molecular architecture of the trypsin binding site results in highly specific recognition of substrates and inhibitors. Specifically, partially burying Asp189 in the inhibitor-free enzyme decreases the penalty for desolvation of this group upon complexation. Water molecules trapped in the binding interface further stabilize the buried ion pair, resulting in a favorable electrostatic contribution of the ion pair formed with ornithine and lysine side chains. Moreover, all side chains that form the trypsin specificity site are partially buried, and hence, relatively immobile in the inhibitor-free state, thus reducing the entropic cost of complexation. The implications of the results for the general problem of recognition and binding are considered. A novel finding in this regard is that like charged molecules can have electrostatic contributions to binding that are more favorable than oppositely charged molecules due to enhanced interactions with the solvent in the highly charged complex that is formed.

Keywords: binding energy; ion pairs; molecular recognition; Poisson–Boltzmann

Understanding the physical basis of macromolecular recognition is a central problem in structural biology. The structures of many complexes formed by proteins are available but it has been difficult to establish simple rules that relate three-dimensional (3D) structure to thermodynamic properties. An essential step toward this goal is the development of computational methods based on simple but realistic physical models to dissect the energetics of binding. Free energy perturbation methods (see e.g., Hermans & Wang, 1997; Lamb & Jorgensen, 1997; Marrone et al., 1997) have encountered some success; however, molecular dynamics simulations are computationally time consuming and cannot, in general, be applied to protein–protein complexes. At the other extreme, phenomenological scoring functions are fast and can be quite useful (Andrews et al., 1984; Novotny et al., 1989; Williams et al., 1993; Bohm, 1994; Gilson et al., 1997b), but by their very nature

are not derived from a well-defined physical model. A possible intermediate approach is offered by continuum solvent models used in conjunction with an all-atom molecular force field (Honig & Nicholls, 1995). A number of binding free energy calculations based on this methodology have recently been reported (Jackson & Sternberg, 1995; Zhang & Koshland, 1996; Novotny et al., 1997) and a general approach to the partitioning of binding into distinct energetic contributions has been suggested (Froloff et al., 1997).

In the present work, we apply the method of Froloff et al. (1997) to study the structural determinants of the binding affinity of bovine β -trypsin for substrates and inhibitors, focusing our attention on the contribution of ion pairs formed by Asp189 with cationic ligands. Asp189 is located at the bottom of the S_1 specificity pocket of trypsin and is primarily responsible for the specificity of trypsin for substrates with arginine or lysine residues in the P_1 position (Bode & Huber, 1992; Perona & Craik, 1995). The buried ion pairs formed by Asp189 provide an apparent challenge to theoretical studies based on continuum models that indicate that buried salt bridges and hydrogen bonds are generally destabilizing (Honig &

Reprint requests to: Professor Barry Honig, Department of Biochemistry and Molecular Biophysics, Columbia University, 630 W. 168th Street, New York, New York 10032; e-mail: bh6@columbia.edu.

Hubbell, 1984; Hendsch & Tidor, 1994; Honig & Nicholls, 1995). It is, of course, possible that the role of Asp189 is to contribute to the specificity of binding and that the ion pairs it forms do not provide a stabilizing interaction.

To clarify the specific role of ion pairs formed by Asp189 and, more generally, of buried ion pairs to intermolecular interactions, we have carried out calculations of the binding free energies of the complexes formed by bovine β -trypsin with four different inhibitors: two synthetic inhibitors ($N\alpha$ -(N,N -dimethylcarbamoyl)- α -azaornithine p -nitrophenyl ester (Dmc-azaOrn-ONp) and $N\alpha$ -(N,N -dimethylcarbamoyl)- α -azalysine p -nitrophenyl ester (Dmc-azaLys-

ONp)) and two protein inhibitors (bovine pancreatic trypsin inhibitor (BPTI) and *Cucurbita maxima* trypsin inhibitor 1 (CMTI-1)). The rationale for the choice of these four complexes is that they differ in the nature of the ionizable side chain that binds in the trypsin specificity site; ornithine in Dmc-azaOrn-ONp, lysine in Dmc-azaLys-ONp, and BPTI and arginine in CMTI-1. In all complexes, water molecules are trapped at the protein-inhibitor binding interface. However, while in Dmc-azaOrn-ONp, Dmc-azaLys-ONp, and BPTI complexes one water molecule bridges through hydrogen bonds to the charged groups involved in ion pair formation, no bridging water is observed in CMTI-1 (Fig. 1). Dissection of the

Fig. 1. Chain trace representation of the complexes formed by bovine β -trypsin with (A) Dmc-azaOrn-ONp, (B) Dmc-azaLys-ONp, (C) BPTI, and (D) CMTI-1. Only the side chains involved in the buried ion pair and the water molecules trapped in the binding interface are shown. Carbon, nitrogen, and oxygen atoms are represented by black, blue, and red spheres, respectively. Hydrogen atoms have been omitted for clarity. Figure has been generated with the program MOLSCRIPT (Kraulis, 1991).

Table 1. Comparison between calculated^a and experimental binding free energy of complex formation of Dmc-azaLys-ONp and Dmc-azaOrn-ONp with trypsin^b

	ΔG_{coul}	ΔG_{solv}	ΔG_{el}	ΔG_{np}	$-T\Delta S_l$	$-T\Delta S_{sc}$	ΔG_b	ΔG_{exp}
Model A ^c								
Dmc-azaOrn-ONp	-8.4	+14.2	+5.8	-25.9	+6.4	0.0	-13.7	-2.3 ^d
Dmc-azaLys-ONp	-10.3	+15.7	+5.4	-26.5	+7.0	0.0	-14.1	-5.2 ^e
Model B ^f								
Dmc-azaOrn-ONp	-2.7	+11.2	+8.5	-25.9	+6.4	0.0	-11.0	
Dmc-azaLys-ONp	-5.4	+14.0	+8.6	-26.5	+7.0	0.0	-10.9	
Model C ^g								
Dmc-azaOrn-ONp	-2.1	+17.3	+15.2	-27.1	+6.4	0.0	-5.5	
Dmc-azaLys-ONp	-5.2	+19.9	+14.7	-27.7	+7.0	0.0	-6.0	

^a Average values obtained on four different structures sampled every 5 ps after 35 ps molecular dynamics equilibration in water (see Materials and methods).

^b All energies are expressed in kcal/mol.

^c Buried waters included in calculations.

^d Data obtained from Ascenzi et al. (1995).

^e Data obtained from Sartori et al. (1995).

^f Buried waters not included in calculations, $\epsilon_{cavity} = 80$ (see Results and discussion).

^g Buried waters not included in calculations, $\epsilon_{cavity} = 2$ (see Results and discussion).

free energy determinants of binding of the four inhibitors to trypsin provides a means for identifying the common features of trypsin ligands and for an improved understanding of the relationship between structure and specificity in the trypsin binding site.

Experimental association constants for noncovalent binding of the four inhibitors to trypsin are available from the literature (Vincent & Lazdunski, 1972; Otlewski & Zbyryt, 1994; Ascenzi et al., 1995; Sartori et al., 1995) and span a wide range of binding free energy values (see Tables 1, 2). This provides an additional interest in the application of binding free energy calculations to these complexes as they represent an ideal test for the ability of the continuum approach to correctly rank the affinities of different ligands for the same protein. In this respect, the two synthetic inhibitors represent a particular challenge in that, despite the great similarity of their molecular structure, their affinity for trypsin is quite different (Table 1).

Results and discussion

Synthetic inhibitors

Table 1 shows the various free energy terms that characterize the binding of Dmc-azaOrn-ONp and Dmc-azaLys-ONp to trypsin. The electrostatic component of the binding free energy is calculated using three different electrostatic models of the complex in which the water molecules trapped at the inhibitor binding interface are considered explicitly (Model A), implicitly (Model B), or to be absent (Model C). In Model A the water molecules are considered fixed and included explicitly in the calculations. In Model B, the water molecules are not included directly in the calculations but a dielectric constant of 80 is assigned to the region that they occupy. This model is used to test the assumption that the electrostatic contribution of buried water molecules can be de-

Table 2. Comparison between calculated and experimental binding free energy of complex formation between BPTI and CMTI-1 with trypsin^a

	ΔG_{coul}	ΔG_{solv}	ΔG_{el}	ΔG_{np}	$-T\Delta S_{sc}$	ΔG_b	ΔG_{exp}
Model A							
BPTI	+137.7	-108.8	+28.9	-71.6	+9.8	-32.9	-18.1 ^b
CMTI-1	-101.3	+148.3	+47.0	-85.8	+13.2	-25.6	-14.4 ^c
Model B							
BPTI	+149.7	-111.0	+38.6	-71.6	+9.8	-23.2	
CMTI-1	-99.7	+147.3	+47.6	-85.8	+13.2	-25.0	
Model C							
BPTI	+149.7	-101.9	+47.8	-73.3	+9.8	-15.7	
CMTI-1	-99.7	+151.3	+51.6	-86.6	+13.2	-21.8	

^a All energies are expressed in kcal/mol.

^b Data obtained from Vincent and Lazdunski (1972).

^c Data obtained from Otlewski and Zbyryt (1994).

scribed with the bulk dielectric constant of water. Model C is similar to Model B except that a dielectric constant of 2 is assigned to the region occupied by the water molecules. Model C, though physically unrealistic, allows us to estimate the energetic contribution of individual water molecules to binding (see below).

The total electrostatic free energy opposes binding of both inhibitors in all three models. This is due in large part to the solvation free energy contribution, which is always positive due to the cost of burying the polar groups in the binding interface, including the two charges that form the salt bridge, upon complexation. Part of this desolvation penalty is recovered from hydrogen bond formation and from Coulomb interactions between the ammonium group of the inhibitor and the carboxyl group of Asp189 and between these two groups and the water molecules buried in the complex. It is interesting to note that when the water molecules are considered fixed, the Coulomb term is negative (Model A); when the water molecules are considered mobile, the solvation term becomes less penalizing but the Coulomb energy increases thus yielding a significantly higher total electrostatic binding free energy in both complexes (Model B). The strong Coulomb interactions in Model A is due to the contribution of charge-dipole interactions involving bound water molecules in the stabilization of the complex. When water molecules are absent (Model C), the solvation energy increases and the total ΔG_{el} is several kcal/mol higher than in the other two cases. The two water molecules stabilize the complex by participating in hydrogen bonds with the ammonium group of the inhibitors and the side chain of Asp189.

The driving force for binding of both inhibitors to trypsin is provided by nonpolar interactions. In both cases more than 70% of the total accessible area of the inhibitors is buried in the complex resulting in a stabilization of 25.9 kcal/mol in the Dmc-azaOrn-ONp complex and 26.5 kcal/mol in the Dmc-azaLys-ONp complex (Table 1). Restriction of the torsional motion in the inhibitors is found to account for 6.4 kcal/mol in Dmc-azaOrn-ONp and 7.0 kcal/mol in Dmc-azaLys-ONp. More than 40% of the total accessible surface area of trypsin side chains forming the binding interface is buried in the isolated protein. Therefore, based on the Pickett and Sternberg approximation (Pickett & Sternberg, 1993), they are already restricted to one conformer and thus do not contribute to conformational entropy loss.

The free energy cost due to the loss of translational and rotational entropy upon complex formation is difficult to calculate, but estimates range from 7 to 15 kcal/mol (Froloff et al., 1997 and references therein). A free energy term that accounts for strain (including entropic effects) induced in the protein and the ligands upon binding (ΔG_{strain}) also makes a destabilizing contribution to binding. This term is not calculated in this work, and thus constitutes a potential source of error. However, the relevant effects are likely to be small, at least for the complexes considered here, since the root-mean-square (RMS) difference for heavy atoms positions between complexes analyzed in this paper and all other high resolution bovine trypsin structures is less than 0.8 Å, indicating no significant conformational rearrangements upon inhibitor binding. Since we do not know the exact value of these two terms that oppose binding, they have not been included in Tables 1 and 2. We expect, however, that the calculated binding free energies be approximately 10 kcal/mol more negative than the experimental ones (Froloff et al., 1997) as is in fact observed. Given the uncertainties in calculating absolute binding free energies, it is encouraging that the numbers are of a realistic magnitude. Moreover, even though results obtained are dependent on the choice of the MD snapshot

(deviation from the average ΔG_b values is ± 1 kcal/mol for both inhibitors), the order of affinities is correctly reproduced by model A in three of the four MD snapshots sampled.

Protein inhibitors

Table 2 shows the free energy terms that characterize the binding of BPTI and CMTI-1 to trypsin. As observed in the synthetic inhibitors, electrostatic interactions oppose binding but water molecules trapped at the binding interface reduce the penalty of burying the two charges that form the ion pair. In fact, when the water molecules are considered mobile (Model B) or absent from the binding interface (Model C), ΔG_{el} is several kcal/mol higher than when they are considered explicitly (Model A). This effect is more pronounced in the BPTI complex as compared to the CMTI-1 complex. In particular, the difference between the total electrostatic free energy calculated in Model A and that calculated in Model C is 18.9 kcal/mol for BPTI while it is just 4.6 kcal/mol for CMTI-1.

The relative magnitude of the Coulomb and solvation energy terms for formation of the trypsin-BPTI complex deserves further comment. As opposed to what is observed in the other three complexes, and more generally in binding free energy calculations using the continuum approach, the Coulomb energy is found to oppose binding ($\Delta G_{coul} > 0$, Table 2), while the solvation energy favors binding ($\Delta G_{sol} < 0$, Table 2). The latter result is particularly surprising since charges are always at least partially buried in the binding process. The source of the effect is in the fact that both trypsin and BPTI have large positive charges (net charge +8e and +6e, respectively). This accounts for the Coulomb repulsion but effect on the solvation free energy is less obvious. It is due to the fact that each protein induces a reaction field in the solvent that interacts favorably with its own charge distribution (the self-energy). But the same reaction field will interact favorably with another like charged object in the vicinity resulting in the net favorable solvation free energy change upon binding. In other words, the binding of trypsin and BPTI creates a complex with a net charge of +14. Since, according to the Born model, solvation goes as the square of the charge, the complex interacts with the solvent more favorably than either protein alone even though some individual charges are desolvated upon binding. This effect is counterbalanced by an increased electrostatic repulsion and hence a positive Coulomb energy so that the total ΔG_{el} opposes the binding of BPTI to trypsin, as observed in the other three complexes.

The same effect is not observed for the binding of trypsin to CMTI-1, which have an opposite net charge (+8e and -1e, respectively). In this case, as expected, the Coulomb term favors binding while the solvation term opposes binding. It is particularly interesting, however, that the change in electrostatic free energy upon binding is less unfavorable for BPTI than for CMTI-1. This suggests how like charged macromolecules can in principles bind to one another as effectively as oppositely charged species. What they lose in Coulomb free energy can be recovered through enhanced interactions with the solvent.

Nonpolar interactions represent the driving force for the binding process also in the case of protein inhibitors. The contact area between inhibitors and trypsin is 1,231 and 1,474 Å² for the BPTI and CMTI-1 complexes, respectively, yielding a stabilization of -71.6 kcal/mol in the former case and -85.8 kcal/mol in the latter. However, part of the free energy gained due to nonpolar interactions is reduced by conformational entropy loss. In both complexes only two trypsin side chains (Ser96 and Gln192) contribute

to conformational entropy loss. However, three side chains in BPTI (Lys15, Arg17, and Arg39) and seven in CMTI-1 (Arg1, Val2, Arg5, Leu7, Glu9, His25, and the C-terminal Gly29) are restricted to one rotamer upon complex formation. The resulting total conformational entropy loss is +9.8 kcal/mol for BPTI and +13.2 kcal/mol for CMTI-1, accounting for much of the difference in binding free energy between the two inhibitors.

Contribution of the buried salt bridge to binding energy

The fact that electrostatic interactions oppose binding of the inhibitors to trypsin is consistent with previous studies that have shown that the desolvation of polar groups is generally not completely compensated by the formation of hydrogen bonds in protein-ligand binding or protein folding processes (Froloff et al., 1997; Yang & Honig, 1995a, 1995b). The same observation is true for the formation of ion pairs (Hendsch & Tidor, 1994) yet, as discussed above, would appear inconsistent with the importance generally attributed to the ion pair formed between Asp189 and Arg/Lys side chains of the substrates in trypsin complexes (Bode & Huber, 1992; Perona & Craik, 1995). To study the specific contribution of the ion pair, we carried out calculations in which only the carboxylate group of Asp189, the cationic group of the inhibitors, and the interfacial water molecules, when these are considered explicitly in the calculations, are treated as charged. In this way, we are able to focus specifically on the role of salt-bridge formation by removing the contribution of hydrogen bond formation to the total electrostatic free energy.

As can be seen in Table 3, the salt bridge contribution depends on the nature of the ligand side chain forming the ion pair with Asp189. In particular, when water molecules trapped at the interface are considered explicitly in the calculations (Table 3, Model A), the ion pairs formed by ornithine and lysine side chains stabilize the complex while that formed by arginine has a destabilizing effect. However, when the less realistic models that ignore

structured waters are used, salt bridge contribution to complex stability is either irrelevant or destabilizing in all the four complexes (Table 3, Models B and C). This demonstrates the need to explicitly include buried waters in electrostatic calculations, a result that has been obtained in a number of previous applications (Yang et al., 1993; Sampogna & Honig, 1996).

The results presented in Table 3 indicate that the presence of water molecules in the binding interface plays a primary role in stabilizing all the complexes through the formation of hydrogen bonds with the ion pair. This contribution can be estimated by subtracting the electrostatic free energy values obtained from Model C from those obtained from Model A. The resulting difference is −2.5 kcal/mol for CMTI-1, −4.1 kcal/mol for Dmc-azaLys-ONp, −5.6 kcal/mol for Dmc-azaOrn-ONp, and −16.0 kcal/mol for BPTI. Differences observed between the complexes can be understood on the basis of the disposition of the bound water molecules with respect to each of the buried ion pairs (Fig. 1). In CMTI-1 each of the two water molecules trapped at the binding interface forms a single hydrogen bond with one of the two charges involved in the ion pair (Fig. 1D). In Dmc-azaOrn-ONp and Dmc-azaLys-ONp, one of the two water molecules bridges to both the charges forming the ion pair (Fig. 1A,B). In BPTI, three water molecules are found to form hydrogen bonds with the lysine side chain of the inhibitor and Asp189, one of which bridges to both charges forming the ion pair (Fig. 1C). Thus, the extent of ion pair stabilization is clearly related to the number of water molecules trapped at the binding interface and to the nature of their interaction with the ion pairs. We have not, however, calculated the energetic consequences of removing individual waters from the bulk, an effect which might be expected to penalize the complex with the largest numbers of bound waters, and hence reduce the calculated binding energy difference between BPTI and CMTI-1.

As mentioned above, that the buried salt bridge stabilizes trypsin/inhibitor complexes appears inconsistent with previous calculations on related systems. To illustrate this point in the context of

Table 3. Electrostatic free energy for formation of the salt bridge between Asp189 and inhibitors' side chains in the protein^a

	ΔG_{coul}	$-\Delta G_{\text{solv}}(\text{Asp})$	$-\Delta G_{\text{solv}}(\text{Inh})$	$\Delta G_{\text{solv}}(\text{Pair})$	ΔG_{el}
Model A					
Dmc-azaOrn-ONp ^b	−63.9	+26.9	+36.9	−2.6	−2.7
Dmc-azaLys-ONp ^b	−64.3	+28.8	+36.7	−2.6	−2.7
BPTI	−61.6	+24.1	+36.9	−4.1	−4.7
CMTI-1	−55.0	+26.9	+35.8	−3.4	+4.3
Model B					
Dmc-azaOrn-ONp	−57.5	+26.8	+36.9	−5.9	+0.3
Dmc-azaLys-ONp	−59.4	+28.8	+36.9	−6.3	−0.2
BPTI	−49.5	+28.8	+36.9	−11.2	+5.0
CMTI-1	−52.8	+27.8	+35.8	−6.0	+4.8
Model C					
Dmc-azaOrn-ONp	−57.5	+26.8	+36.9	−3.3	+2.9
Dmc-azaLys-ONp	−59.4	+28.8	+36.7	−4.2	+1.9
BPTI	−49.5	+28.8	+36.9	−5.0	+11.2
CMTI-1	−52.8	+27.8	+35.9	−4.0	+6.9

^a All energies are expressed in kcal/mol.

^b See Table 1.

Table 4. Electrostatic free energy for formation in a low dielectric medium^a of the salt bridge between Asp189 and inhibitors' side chains dissolved in water^b

	$\Delta G_{\text{coul}}(\text{Pair})$	$-\Delta G_{\text{solv}}(\text{Asp})$	$-\Delta G_{\text{solv}}(\text{Inh})$	ΔG_{el}
Dmc-azaOrn-ONp	-57.5	+41.1	+36.9	+20.5
Dmc-azaLys-ONp	-59.4	+42.2	+36.7	+19.5
BPTI	-49.5	+42.0	+36.9	+29.4
CMTI-1	-52.8	+42.1	+35.9	+25.2

^a($\epsilon = 2$)^bAll energies are expressed in kcal/mol.

the present study, we have calculated the electrostatic contribution to the formation in a low dielectric medium ($\epsilon = 2$) of a hypothetical salt bridge formed by an isolated Asp189 side chain and the inhibitors' cationic side chain dissolved in water in the uncomplexed state (Table 4). In this case, electrostatics is clearly destabilizing by about 20–30 kcal/mol for all inhibitors. By comparing the individual terms in Table 4 to those for Model C in Table 3 (which is appropriate since no buried waters are present in either case), an additional mechanism used by trypsin to stabilize the ion pair becomes clear. Specifically, the Coulomb contribution and the solvation contribution of the inhibitors are identical in both cases, but the desolvation free energy of the side chain of Asp189 in trypsin is less than that of the isolated side chain by approximately 13.0 to 18.0 kcal/mol in the different complexes. The reason for the difference is that 94% of the total accessible area of the carboxylate group of Asp189 is buried in the inhibitor free enzyme. Effectively, part of the energetic cost of desolvating this group is paid during the folding process of the enzyme. This reduces the free energy cost associated with salt bridge formation and in the case of ornithine and lysine ligands leads to a stabilizing electrostatic interaction when the water molecules trapped at the binding interface are explicitly taken into account.

Conclusions

As can be seen in Tables 1 and 2, binding free energy values calculated for the four complexes are more negative by about 10 kcal/mol than the experimental values. However, since factors that oppose binding such as translational/rotational entropy loss and strain energy have not been taken into account, the binding free energies obtained here are in a reasonable range. Moreover, the order of binding is correctly predicted including the approximately 10 kcal/mol free energy difference between the synthetic and protein inhibitors. The difference in the binding affinity of the two synthetic inhibitors appears due in large part to the difference in ionic bond distance in the two complexes (Table 1). For the protein inhibitors it appears due to a smaller number of side-chain rotamers that are frozen in the binding process and to a novel electrostatic effect, whereby like charged macromolecules can interact more favorably than oppositely charged macromolecules.

Overall, the extent of the agreement with experiment is good given the many uncertainties in the calculations. It is particularly significant in this regard that binding has been described with a well-defined physical model that yields a consistent physical picture in all cases. Specifically, nonpolar interactions provide much of the favorable driving force while electrostatic contribution op-

poses binding. However, due in part to the fact that Asp189 is almost completely buried in the inhibitor-free protein, the electrostatic penalty associated with binding is not large and, in three of the four cases, the salt bridge it forms actually favors binding. Water molecules trapped at the interface are found to play an important role in reducing the electrostatic penalty for binding.

An important result of this study is that ion pair formation, including the effects of bound water molecules, stabilizes complexes between trypsin and ornithine- or lysine-based inhibitors. To further establish the generality of this result, we analyzed the 29 high resolution crystal structures (resolution equal to/less than 2.0 Å) of bovine trypsin-inhibitor complexes available in the Protein Data Bank (PDB) (Bernstein et al., 1977; Sussman et al., 1998) in which solvent coordinates have been determined. In all the structures analyzed, one water molecule is present in the interface at hydrogen bonding distance with the Asp189 side chain; moreover, in 21 complexes a second buried water molecule is found at hydrogen bonding distance from the ligand. It is interesting to note that in all the four complexes involving lysine ligands at least one water molecule bridges to both the charges forming the ion pair while in all the four complexes formed by arginine ligands no bridging water molecule is observed. Thus, it seems likely that the ion pair stabilization observed in the present study and differences evidenced between different ligands are general features in the recognition of substrates and inhibitors by trypsin.

Another general feature of binding is a fairly rigid binding site in inhibitor-free trypsin. Not only is 94% of the accessible area of the carboxylate group of Asp189 buried, but in fact, all the residues forming the trypsin S_1 specificity subsite have most of their accessible area buried. This restricts their mobility in the free protein and results in a lower conformational entropy loss upon binding. Analysis of the crystal structure of human thrombin, another serine protease acting on charged substrates shows that, in this case as well, Asp189 is almost completely buried in the protein moiety and the mobility of the side chains forming the binding site is restricted. Thus, it appears that trypsin and related proteins use the dual strategy of partially burying charged groups and fixing side chains in the uncomplexed protein to achieve specificity and to increase binding affinity. It is, of course, likely that similar mechanisms are used by other proteins.

Materials and methods

Molecular structures

Molecular models of the complexes formed by trypsin with the two synthetic inhibitors were derived from the crystal structures of

the acyl enzyme adducts of trypsin with Dmc-azaOrn-ONp and Dmc-azaLys-ONp, which have been solved at 1.8 Å (Sartori et al., 1995) and 2.0 Å resolution, respectively (Ascenzi et al., 1995). Both inhibitors bind covalently to trypsin by acylating Ser195 with the concurrent release of the *p*-nitrophenyl group. The positively charged side chain of the inhibitors forms an ion pair with Asp189. In both complexes, two water molecules are trapped at the protein-inhibitor binding interface (HOH794 and HOH911 in Dmc-azaOrn/enzyme covalent complex and HOH794 and HOH772 in Dmc-azaLys/enzyme covalent complex), one of which (HOH794) bridges through hydrogen bonds to the charged groups involved in ion pair formation (Fig. 1A,B).

Binding free energy values calculated in this paper for the synthetic inhibitors refer to noncovalent binding of the native forms of the inhibitors (i.e., Dmc-azaOrn-ONp and Dmc-azaLys-ONp) to trypsin. Thus, to compare experimental and theoretical results, molecular structures of the noncovalent complexes were built by molecular modeling. Coordinates for noncovalent complexes were obtained by breaking the covalent bond between Ser195 and the inhibitors and adding the *p*-nitrophenolic group to the inhibitors' carbonyl carbon using the Biopolymer module of the program InsightII (Molecular Simulations Inc., Burlington, Massachusetts). The geometry of the *p*-nitrophenyl group relative to the azaaminoacid moiety was modeled so as to avoid steric clashes with trypsin side chains. The model-built structures were then refined by molecular dynamics simulations (see below).

Molecular structures of the complexes formed by trypsin and the protein inhibitors were obtained from the PDB (Bernstein et al., 1977; Sussman et al., 1998). In the BPTI-trypsin complex (PDB code: 2PTC) Lys15 residue of the inhibitor forms an ion pair with Asp189. Three water molecules are trapped in the binding interface (HOH402, HOH414, and HOH416). HOH 414 bridges the ammonium group of Lys15 to the carboxyl group of Asp189 (Fig. 1C). In the CMTI-1-trypsin complex (PDB code: 1PPE), the inhibitors' side chain Arg5 is involved in the ion pair with Asp189. In this case two water molecules are buried in the binding interface (HOH704 and HOH416) but each one is hydrogen bonded only to one of the two charged groups involved in the ion pair (Fig. 1D).

The structures of the four complexes were energy minimized using the program CHARMM (Brooks et al., 1983), and the CHARMM22 parameters, and force field (MacKerell et al., 1998). Hydrogen atoms were added to the structures and to the water molecules using the CHARMM routine HBUILD and then allowed to relax using a conjugate gradient algorithm energy minimization. Nonhydrogen atoms were kept fixed and the minimization was carried out until the maximum derivative of the total energy was less than 0.05 kcal/mol/Å. Water molecules buried at the protein-ligand interface were included in the minimization.

Molecular dynamics simulations

Structures of the noncovalent complexes formed between trypsin and Dmc-azLys-ONp and Dmc-azaOrn-ONp were obtained from a 50 ps molecular dynamics simulation using the CHARMM macromolecular mechanics package (Brooks et al., 1983) and the CHARMM22 parameters and force field (MacKerell et al., 1998). The three-site TIP3p model (Jorgensen et al., 1983) was used for water molecules. The minimized structures were placed in a truncated octahedron, constructed from a cubic volume of water molecules of dimension $77.758 \times 77.758 \times 77.758$ Å, and water molecules overlapping with protein-inhibitor atoms (cutoff = 2.8 Å)

were removed. Solvated structures, containing 6,166 solvent molecules, were energy minimized by applying a harmonic force of 10 kcal/mol to nonhydrogen atoms of the complexes to allow reorganization of the solvent. Minimized solvated structures were then subjected to a molecular dynamics simulation at 298 K in the microcanonical ensemble after a heating run of 10 ps during which the temperature was gradually increased from 0 to 298 K. The simulation time step was set to 0.002 ps. Equilibrium, as judged by convergence of the RMS deviation (RMSD) of nonhydrogen atoms position from the initial structures (RMSD difference between successive molecular dynamics snapshots, sampled every 5 ps, lower than 0.05 Å), was reached in both complexes after approximately 35 ps. Data relative to the synthetic inhibitors presented in Tables 1 and 3 represent the average of the results obtained on four molecular dynamics snapshots sampled every 5 ps from 35 to 50 ps.

Binding energy calculations

The total binding free energy ΔG_b for a protein-ligand interaction can be written in the form (Froloff et al., 1997):

$$\Delta G_b = \Delta G_{el} + \Delta G_{np} + \Delta G_{strain} - T\Delta S_{sc} - T\Delta S_l - T\Delta S_{t,r} \quad (1)$$

where ΔG_{el} is the electrostatic contribution to binding, ΔG_{np} is the contribution due to nonpolar interactions between the protein and the ligand, and ΔG_{strain} is a term that includes possible distortions of the protein and the ligand upon complexation. $T\Delta S_{sc}$ is the loss of conformational entropy of the protein side chains upon ligand binding, $T\Delta S_l$ is the corresponding term for the ligand, and $T\Delta S_{t,r}$ describes the loss of translational and rotational degrees of freedom. Contributions from ΔG_{strain} and $-T\Delta S_{t,r}$ to the total binding free energy are not considered in the present work. The first term is the most problematic in the context of the present study in that its magnitude can depend strongly on the nature of the ligand. $-T\Delta S_{t,r}$ is expected to remain approximately constant for a relatively similar set of molecules although since it depends in part on the volume swept out by the center of mass of the bound ligand (Gilson et al., 1997a) some small variation from ligand to ligand might be expected. Nevertheless the effect on *relative* binding free energies should be quite small. All of these terms do, however, weaken the overall computed binding affinities in that they are all essentially penalties the system must pay to achieve binding.

ΔG_{el} was calculated as described previously (Gilson & Honig, 1988; Smith & Honig, 1994) and is obtained from the sum of the Coulomb and solvation energies of the complex minus that of the isolated species.

$$\Delta G_{el}(\epsilon_i, \epsilon_o) = \Delta G_{coul}(\epsilon_i) + \Delta G_{solv}(\epsilon_i, \epsilon_o) \quad (2)$$

where $\Delta G_{el}(\epsilon_i, \epsilon_o)$ is the change in electrostatic free energy for two reactants with interior dielectric constant ϵ_i embedded in a medium of dielectric constant ϵ_o , $\Delta G_{coul}(\epsilon_i)$ is the Coulomb interaction between the two species embedded in a medium of dielectric constant ϵ_i and $\Delta G_{solv}(\epsilon_i, \epsilon_o)$ is the change in reaction field energy (solvation energy) upon transferring the reactants from a medium of dielectric ϵ_i to a medium of dielectric ϵ_o .

Electrostatic potentials were calculated using the program DelPhi (Nicholls & Honig, 1991), which solves the finite difference Poisson-Boltzmann equation for molecules of complex shape by

mapping the dielectric and charge distribution of the molecules in a 3D grid. Atomic charges and radii were taken from the PARSE parameter set (Sitkoff et al., 1994). The ionic strength was set to 0.0 M in all calculations and dielectric constant values of 2 and 80 were assumed for solutes and solvent, respectively. The value of the molecular dielectric constant used in finite difference Poisson–Boltzmann calculations has been a subject of considerable discussion, with values as high as 20 reported in some studies. A value of two accounts explicitly for electronic polarizability and essentially assumes that there is no structural change upon binding (see e.g., Gilson et al., 1985). Larger values are often used to account for the effect of conformational changes, particularly on calculated pK_a s, when these are not treated explicitly in the calculations. As discussed previously (Froloff et al., 1997; Misra et al., 1998), we have preferred to incorporate the effects of conformational change in ΔG_{strain} so that the calculated electrostatic free energies are associated with a well-defined physical process: the binding of two molecules that are assumed to have the same conformation in the free and complexed state. The process of using the dielectric constant as a variable has some practical advantages in that it allows fitting parameters to experimental results, but suffers from the fact that the physical meaning of the dielectric constant is lost. Moreover, using a dielectric constant of 2 in conjunction with PARSE atomic charges and radii assures that the solvation model used in the calculations reproduces the solvation behavior of amino acid analogs (Sitkoff et al., 1994).

In our case a grid of $165 \times 165 \times 165$ points was used and to improve accuracy of the potentials, two focusing runs were carried out, yielding a final resolution of 3 grids/Å. To minimize numerical errors, the proteins and the ligands were placed in the same position in the grid both in the isolated and complexed state. Moreover, each ΔG_{solv} value was obtained as the average of seven different calculations in which the molecules were translated in the lattice in ± 1.0 Å increments along the x , y , and z axes. In each case, the deviation of ΔG_{solv} was less than ± 0.1 kcal/mol.

The nonpolar contribution to binding was calculated from the equation

$$\Delta G_{\text{np}} = \gamma_{\text{aw}} A_c \quad (3)$$

where γ_{aw} is the microscopic surface tension associated with the transfer of an alkane from liquid alkane to water and A_c is the curvature-corrected accessible area lost by the two reactants upon complexation (Nicholls et al., 1991; Sharp et al., 1991). γ_{aw} was set at $58.18 \text{ cal}/\text{\AA}^2$ (Froloff et al., 1997). A_c was calculated using the program SURFCV (Sridharan et al., 1992). The curvature correction of the accessible area accounts for the increased hydrophobicity of concave regions of the protein surface.

The procedure used here to calculate the nonpolar term has been a subject of some controversy. It has been suggested, for example, that molecular rather than accessible surface area should be used and that other values of γ_{aw} are preferable (Jackson & Sternberg, 1995 and references therein). We prefer the accessible area because it is related directly to the number of water molecules removed from a surface. The value of γ_{aw} is derived from a simple physical model (Nicholls et al., 1991), which gives a consistent physical picture when used in conjunction with a dielectric constant of 2 (Froloff et al., 1997). Nevertheless, the assumption that a single surface tension parameter [that may well account for packing effects as well as for hydrophobicity (Friedman & Honig, 1995; Yang & Honig, 1995a, 1995b)] can be used to characterize an entire interface is clearly quite crude.

Configurational entropy loss of the synthetic inhibitors was calculated from the equation

$$-T\Delta S_1 = \sum RT \ln(N_{\text{free}}/N_{\text{bound}}) \quad (4)$$

(Novotny et al., 1997) where N_{free} is the number of torsion angles available in the unbound state and N_{bound} is the corresponding number in the bound state. As more than 70% of the inhibitors' accessible area is buried in the complexes, the inhibitors were considered fixed in the complex. Three configurations per rotatable bond were assumed in the free state and only one in the bound state.

Conformational entropy loss $T\Delta S_{\text{sc}}$ for trypsin and protein inhibitors' side chains was obtained from the model of Pickett and Sternberg (1993) that assumes that side chains that are less than 40% buried rotate freely, while side chains that are more than 40% buried are restricted to one rotamer. An entropy scale is then obtained from the observed distributions of side-chain rotamers in 50 nonhomologous protein crystal structures (Pickett & Sternberg, 1993). Thus, $T\Delta S_{\text{sc}}$ is obtained by summing the entropy loss of all protein side chains that are less than 40% buried in the inhibitor-free protein and more than 40% buried in the complex. Percentage buried area was calculated with the program SURFCV (Sridharan et al., 1992) as the ratio between the accessible area of the amino acid side chain in the extended state and in the protein.

ΔG_{strain} and $T\Delta S_{t,r}$ are not calculated directly but their contribution is discussed in Results and discussion.

Acknowledgments

The authors would like to thank Dr. Felix Sheinerman for help in performing the molecular dynamics simulations. This work was supported by NIH grant GM30518 and by a long-term fellowship from Human Frontier Science Program to FP.

References

- Andrews PR, Craik DJ, Martin JL. 1984. Functional group contributions to drug-receptor interactions. *J Med Chem* 27:1648–1657.
- Ascenzi P, Balliano G, Milla P, Ferraccioli R, Sartori P, Djinnovic-Carugo K, Bolognesi M. 1995. Inhibition of bovine β -trypsin by the active site titrant $N\alpha$ -(N,N -dimethylcarbamoyl)- α -azaornithine p -nitrophenyl ester: A kinetic and X-ray crystallographic study. *Biochem Biophys Res Commun* 217:437–444.
- Bernstein FC, Koetzle TF, Williams GJ, Meyer EE Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol* 112:535–542.
- Bode W, Huber R. 1992. Natural protein proteinase inhibitors and their interaction with proteinases. *Eur J Biochem* 204:433–451.
- Bohm HJ. 1994. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Aided Mol Des* 8:243–256.
- Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM: A program for macromolecular energy minimization and dynamics calculation. *J Comput Chem* 4:187–217.
- Friedman RA, Honig B. 1995. A free energy analysis of nucleic acid base stacking in aqueous solution. *Biophys J* 69:1528–1535.
- Froloff N, Windemuth A, Honig B. 1997. On the calculation of binding free energies using continuum methods: Application to MHC class I protein-peptide interactions. *Protein Sci* 6:1293–1301.
- Gilson MK, Given JA, Bush BL, McCammon JA. 1997a. The statistical thermodynamic basis for computation of binding affinities. *Biophys J* 72:1047–1069.
- Gilson MK, Given JA, Head MS. 1997b. A new class of models for computing receptor-ligand binding affinities. *Chem Biol* 4:87–92.
- Gilson MK, Honig B. 1988. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies and conformational analysis. *Proteins* 4:7–18.

- Gilson MK, Rashin A, Fine R, Honig B. 1985. On the calculation of electrostatic interactions in proteins. *J Mol Biol* 184:503–516.
- Hendsch ZS, Tidor B. 1994. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci* 3:211–226.
- Hermans J, Wang L. 1997. Inclusion of loss of translational and rotational freedom in theoretical estimates of free energy of binding. Application to a complex of benzene and mutant T4 lysozyme. *J Am Chem Soc* 119:2707–2714.
- Honig BH, Hubbell WL. 1984. Stability of “salt bridges” in membrane proteins. *Proc Natl Acad Sci USA* 17:5412–5416.
- Honig B, Nicholls A. 1995. Classical electrostatics in biology and chemistry. *Science* 268:1144–1149.
- Jackson RM, Sternberg RJE. 1995. A continuum model for protein–protein interaction. *J Mol Biol* 250:258–275.
- Jorgensen WL, Chandrasekhar J, Madura J, Impley RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
- Kraulis J. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 24:946–950.
- Lamb ML, Jorgensen WL. 1997. Computational approaches to molecular recognition. *Curr Opin Chem Biol* 1:449–457.
- MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck J, Field MJ, Fischer S, Gao J, Guo H, Ha S, et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of protein. *J Phys Chem B* 102:3586–3616.
- Marrone TJ, Briggs JM, McCammon JA. 1997. Structure-based drug design: Computational advances. *Ann Rev Pharm Tox* 37:71–90.
- Misra VK, Hecht JL, Yang A, Honig B. 1998. Electrostatic contributions to the binding free energy of the λ cl repressor to DNA. *Biophys J* 75:2262–2273.
- Nicholls A, Honig B. 1991. A rapid finite difference algorithm utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J Comput Chem* 12:435–445.
- Nicholls A, Sharp KA, Honig B. 1991. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11:281–296.
- Novotny J, Bruccoleri RE, Davis M, Sharp KA. 1997. Empirical free energy calculations: A blind test and further improvements to the method. *J Mol Biol* 268:401–411.
- Novotny J, Bruccoleri RE, Saul FA. 1989. On the attribution of binding energy in antigen-antibody complexes McPC 603, DI.3, and HyHEL-5. *Biochemistry* 28:4735–4749.
- Otlewski J, Zbyryt T. 1994. Single peptide bond hydrolysis/resynthesis in squash inhibitors of serine proteinases. I. Kinetics and thermodynamics of the interaction between squash inhibitors and bovine β -trypsin. *Biochemistry* 33:200–207.
- Perona JJ, Craik CS. 1995. Structural basis of substrate specificity in the serine proteases. *Protein Sci* 4:337–360.
- Pickett SP, Sternberg MJE. 1993. Empirical scale of side-chain conformational entropy in protein folding. *J Mol Biol* 231:825–839.
- Sampogna RV, Honig B. 1996. Electrostatic coupling between retinal isomerization and the ionization state of Glu-204: A general mechanism for proton release in bacteriorhodopsin. *Biophys J* 71:1165–1171.
- Sartori P, Djinnovic-Carugo K, Ferraccioli R, Balliano G, Milla P, Ascenzi P, Bolognesi M. 1995. Active site titration of bovine beta-trypsin by $N\alpha$ -(N , N -dimethylcarbamoyl)- α -aza-lysine p -nitrophenyl ester: Kinetic and crystallographic analysis. *FEBS Lett* 358:53–56.
- Sharp KA, Nicholls A, Fine RF, Honig B. 1991. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science* 252:106–109.
- Sitkoff D, Sharp KA, Honig B. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J Phys Chem* 98:1978–1988.
- Smith KC, Honig B. 1994. Evaluation of the conformational free energy of loops in proteins. *Proteins* 18:119–132.
- Sridharan S, Nicholls A, Honig B. 1992. A new vertex algorithm to calculate solvent accessible surface areas. *Biophys J* 61:A174.
- Sussman JL, Lin D, Jiang J, Manning NO, Prilusky J, Ritter O, Abola EE. 1998. Protein Data Bank (PDB): Database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr D* 54:1078–1084.
- Vincent JP, Lazdunski M. 1972. Trypsin-pancreatic inhibitor association. Dynamics of the interaction and role of disulfide bridges. *Biochemistry* 11:2967–2977.
- Williams DH, Searle MS, Mackay JP, Gerhard U, Maplestone RA. 1993. Toward an estimation of binding constants in aqueous solution. *Proc Natl Acad Sci USA* 90:1172–1178.
- Yang AS, Gunner MR, Sampogna R, Sharp K, Honig B. 1993. On the calculation of pKas in proteins. *Proteins* 3:252–265.
- Yang AS, Honig B. 1995a. Free energy determinants of secondary structure formation: I. α -helices. *J Mol Biol* 252:351–365.
- Yang AS, Honig B. 1995b. Free energy determinants of secondary structure formation: II. Antiparallel β -sheets. *J Mol Biol* 252:366–376.
- Zhang T, Koshland DE. 1996. Computational method for relative binding energies of enzyme-substrate complexes. *Protein Sci* 5:348–356.