J|A|C|S
ARTICLES
Published on Web 09/10/2002

Contribution of Multicentered Short Hydrogen Bond Arrays to Potency of Active Site-Directed Serine Protease Inhibitors

Bradley A. Katz,* Jeffrey R. Spencer, Kyle Elrod, Christine Luong, Richard L. Mackman,† Mark Rice, Paul A. Sprengeler, Darin Allen,‡ and James Janc

Contribution from Celera, 180 Kimball Way, South San Francisco, California 94080

Received January 15, 2002

Abstract: We describe and compare the pH dependencies of the potencies and of the bound structures of two inhibitor isosteres that form multicentered short hydrogen bond arrays at the active sites of trypsin, thrombin, and urokinase type plasminogen activator (urokinase or uPA) over certain ranges of pH. Depending on the pH, short hydrogen bond arrays at the active site are mediated by two waters, one in the oxyanion hole (H_2O_{oxy}) and one on the other (S2) side of the inhibitor (H_2O_{S2}), by one water (H_2O_{oxy}), or by no water. The dramatic variation in the length of the active site hydrogen bonds as a function of pH, of inhibitor, and of enzyme, along with the involvement or absence of ordered water, produces a large structural manifold of active site hydrogen bond motifs. Diverse examples of multicentered and two-centered short hydrogen bond arrays, both at and away from the active site, recently discovered in several protein crystal systems, suggest that short hydrogen bonds in proteins may be more common than has been recognized. The short hydrogen bond arrays resemble one another with respect to ionic nature, highly polar environment, multitude of associated ordinary hydrogen bonds, and disparate pK_a values of participating groups. Comparison of structures and K_i values of trypsin complexes at pH values where the multicentered short hydrogen bond arrays mediating inhibitor binding are present or absent indicate that these arrays have a minor effect on inhibitor potency. These features suggest little covalent nature within the short hydrogen bonds, despite their extraordinary shortness (as short as 2.0 Å).

Introduction

One controversy spanning the disciplines of chemistry and biology is the role of short (low-barrier) hydrogen bonds in enzymatic catalysis.[1−10] An important counterpart in drug design involves the contribution of short hydrogen bonds to inhibitor or ligand potency.[11−14] Inhibitor potency and enzyme catalysis can be intimately related, especially when short hydrogen bonds involving active site residues mediate both catalysis and inhibitor binding. Critical issues in both these circumstances are the relative strength and covalent character of short hydrogen bonds compared with their ordinary counterparts in the polar environment of the active site of an enzyme.

Recently we described a novel serine protease inhibition motif in which binding is mediated by very short hydrogen bonds involving the active site Ser195 hydroxyl ($O\gamma_{Ser195}$), a water co-bound in the oxyanion hole (H_2O_{oxy}) and an inhibitor phenolate oxygen (O_{phenol}).[14] In many complexes extension of the short hydrogen bond network through the benzimidazole or indole nitrogen produces an approximate tetrahedron of atoms, each one connected to the other within a cluster of six

* Address correspondence to this author. Phone: (650) 866-6270. Fax: (650) 866-6654. E-mail: brad.katz@celera.com.

† Current address: Gilead Sciences, 333 Lakeside Drive, Foster City, CA 94404.

‡ Current address: Sunesis Pharmaceuticals, Inc., 341 Oyster Point Blvd., South San Francisco, CA 94080.

(1) Cleland, W. W.; Kreevoy, M. M. Low-barrier hydrogen bonds and enzymic catalysis. *Science* **1994**, *264*, 1887−1890.
(2) Frey, P. A.; Whitt, S. A.; Tobin, J. B. A low barrier hydrogen bond in the catalytic triad of serine proteases. *Science* **1994**, *264*, 1927−1930.
(3) Cleland, W. W.; Frey, P. A.; Gerlt, J. A. The low-barrier hydrogen bond in enzyme catalysis. *J. Biol. Chem.* **1998**, *273*, 25529−25532.
(4) Schiøtt, B.; Iversen, B. B.; Madsen, G. K. H.; Larsen, F. K.; Bruice, T. C. On the electronic nature of low-barrier hydrogen bonds in enzymatic reactions. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12799−12802.
(5) Chen, J.; McAllister M. A.; Lee, J. K.; Houk, K. N. Short, strong hydrogen bonds in the gas phase and in solution: theoretical exploration of pK_a matching and environmental effects on the strengths of hydrogen bonds and their potential roles in enzymatic catalysis. *J. Org. Chem.* **1998**, *63*, 4611−4619.
(6) Warshel, A.; Papazyan, A.; Kollman, P. A. On low-barrier hydrogen bonds and enzyme catalysis. *Science* **1995**, *269*, 102−104.
(7) Warshel, A.; Papazyan, A. Energy considerations show that low-barrier hydrogen bonds do not offer a catalytic advantage over ordinary hydrogen bonds. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13665−13670.
(8) Guthrie, J. P. Short strong hydrogen bonds: can they explain enzymic catalysis? *Chem. Biol.* **1996**, *3*, 163−170.
(9) Ash, E.; Sudmeier, J. L.; De Fabo, E. C.; Bachovchin, W. W. A low-barrier hydrogen bond in the catalytic triad of serine proteases? Theory versus experiment. *Science* **1997**, *278*, 1128−1132.
(10) Warshel, A. Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J. Biol. Chem.* **1998**, *273*, 27035−27038.
(11) Usher, K. C.; Remington, J.; Martin, D. P.; Drueckhammer, D. G. A short hydrogen bond provides only moderate stabilization of an enzyme−inhibitor complex of citrate synthase. *Biochemistry* **1994**, *33*, 7753−7759.
(12) Schwartz, B.; Drueckhammer, D. G.; Usher, K. C.; Remington, S. J. α-fluoro acid and α-fluoro amide analogues of acetyl-CoA as inhibitors of citrate synthase: effect of pK_a matching on binding affinity and hydrogen bond length. *Biochemistry* **1995**, *34*, 15459−15466.
(13) Wang, Z.; Luecke, H.; Yao, N.; Quiocho, F. A. A low energy short hydrogen bond in very high-resolution structures of protein receptor-phosphate complexes. *Nature Structural Biology* **1997**, *4*, 519−522.
(14) Katz, B. A.; Elrod, K.; Luong, C.; Rice, M.; Mackman, R. L.; Sprengeler, P. A.; Spencer, J.; Hataye, J.; Janc, J.; Link, J.; Litvak, J.; Rai, R.; Rice, K.; Sideris, S.; Verner, E.; Young, W. A novel serine protease inhibition motif involving a multicentered short hydrogen bonding network at the active site. *J. Mol. Biol.* **2001**, *307*, 1451−1486.

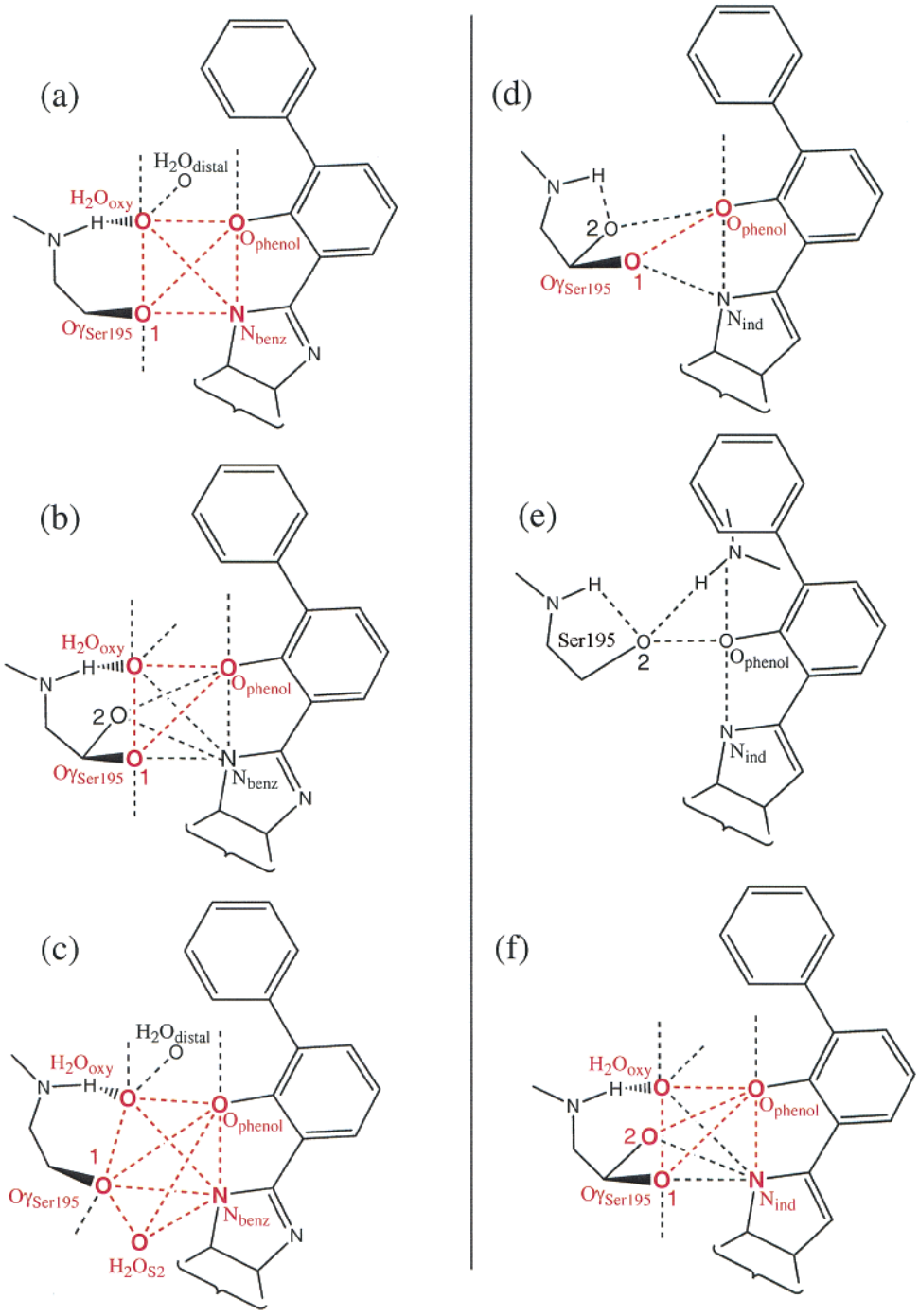

Figure 1. Multicentered and two-centered short and ordinary hydrogen bond arrays at the active site of protease-inhibitor complexes: (a) four-centered ($O\gamma_{Ser195}$, O_{oxy}, O_{phenol}, $N_{benzimidazole}$) short;[14] (b) three-centered ($O\gamma_{Ser195}$, O_{oxy}, O_{phenol}) short in equilibrium with two-centered ($O\gamma_{Ser195}-O_{phenol}$) short, as in trypsin-CRA-7806 (pH 8.2); (c) four-centered ($O\gamma_{Ser195}$, O_{oxy}, O_{phenol}, N_{indole}) short plus four-centered ($O\gamma_{Ser195}$, O_{S2}, O_{phenol}, N_{indole}) short, as in trypsin-CRA-8696 (pH 6.1); (d) two-centered ($O\gamma_{Ser195}-O_{phenol}$) short in equilibrium with two-centered ($O\gamma_{Ser195}-O_{phenol}$) normal (Ser195 discretely disordered), as in trypsin-CRA-8696 (pH 7.2); (e) two-centered normal, as in trypsin-CRA-8696 (pH 7.9); (f) three-centered ($O\gamma_{Ser195}$, O_{oxy}, O_{phenol}) short in equilibrium with two-centered ($O\gamma_{Ser195}-O_{phenol}$) short (Ser195 discretely disordered), as in trypsin-CRA-8696 (pH 9.0). Short hydrogen bonds and involved atoms are shown in red. The hydrogen bonds and hydrogen-bonded amidine-Asp189 salt bridge interactions at the S1 site of the trypsin and thrombin complexes are very similar at all pH values. The urokinase complexes are atypical at the S1 site, in that a bound water mediates interactions of the amidine of the bound inhibitor with Asp189, instead of direct hydrogen bonds.[14]

short hydrogen bonds (Figure 1a).[14] The shortness of the hydrogen bonds (as short as 2.0 Å) and their multicentered nature are both unprecedented.

Such short hydrogen-bonded complexes resemble the proposed transition state of serine protease catalysis in several aspects. In the inhibitor complexes the pK_a of His57, which makes ordinary hydrogen bonds to $O\gamma_{Ser195}$ and to H_2O_{oxy}, is significantly raised.[14] Likewise, during catalysis, elevation of the pK_a of His57 allows abstraction by $N\epsilon2_{His57}$ of the proton from $O\gamma_{Ser195}$ for it to attack the carbonyl carbon of the scissile bond.[15−17] While in the transition state of catalysis the oxyanion hole is occupied by the oxyanion resulting from attack of the scissile bond,[18] in the inhibitor complexes it is occupied by H_2O_{oxy}, also believed to bear a (partial) negative charge, required

for formation of short, ionic hydrogen bonds. Finally, like the oxyanion of a substrate, H_2O_{oxy} accepts hydrogen bonds from the oxyanion-stabilizing groups, N_{Gly193} and N_{Ser195}. Despite these intriguing similarities between multicentered short hydrogen-bonded inhibitor complexes of serine proteases and their catalytic transition states, the presumed involvement of a low-barrier hydrogen bond in the latter involves the $N\delta1_{His57}-O\delta2_{Asp102}$ interaction,[2,19−20] while in the former $O\gamma_{Ser195}$, H_2O_{oxy}, O_{phenol} (and often $N_{benzimidazole/indole}$) are the short hydrogen-bonding participants.[14]

Because of the similar architecture and chemical character of the active sites of members of the serine protease family, all are susceptible to short hydrogen-bond-mediated inhibition by 2-(2-phenol)indoles and 2-(2-phenol)benzimidazoles (Figure 1a). Thus, this multicentered short hydrogen bond inhibition motif has been harnessed in the design and development of potent, selective inhibitors of trypsin-like serine protease drug targets such as urokinase, thrombin, factor VIIa, and factor Xa.[14,21−24] Here we describe and compare the pH dependence of the inhibition constants (K_i values) and of the protease-bound structures of two inhibitor isosteres, a benzimidazole and an indole (Figure 1) whose active site binding is mediated by either short or normal hydrogen bonds, depending on pH. The K_i values and structures of the corresponding trypsin complexes, determined under conditions where short hydrogen bonds are present or absent, indicate that there is a minimal contribution to inhibitor affinity by short hydrogen bonds over ordinary ones. Other short hydrogen bond arrays recently discovered away from the active sites of trypsin and urokinase are also described, along with their similarities to the active site protease-inhibitor short hydrogen bond arrays.

Results

Diversity in Short Hydrogen Bonding Modes at the Active Site Involving Bound Inhibitors with an S1′-Directed Substituent. The structures of trypsin complexes of inhibitors such as the benzimidazole (CRA-7806) or indole (CRA-8696), bearing an S1′-directed group, usually exhibit an intricate pH dependence of active site hydrogen bonding that is a fingerprint of the particular complex. The inhibitors remain fully bound and well defined by density over a large range of pH, from pH 4.5 to 11.0. Figure 1 shows schematically some of the active site hydrogen bond arrays observed for these complexes as a function of pH, and Table 1 provides corresponding hydrogen bond lengths and angles.

The pH dependence of the active site structure of urokinase, thrombin, and trypsin complexes of the indole (CRA-8696) and benzimidazole (CRA-7806) isosteres is shown schematically in Figure 2, in which different active site hydrogen modes are color-coded and briefly described in the legend. An identification number (from 1 to 10) for each mode is assigned in the legend. The complete pH dependencies of the structures of the benzimidazole and indole complexes of trypsin, determined in a common crystal form (form *b*) between pH 4.5 and 11.0, are described and compared below. The crystal structures of these complexes over a more limited pH range are also determined for other crystal forms of trypsin (forms *a* and *c*).

Between pH 5.0 and 8.0 trypsin-CRA-7806 (Figure 2e) exhibits a four-centered centered short hydrogen bond array at the active site as described[14] (Figure 1a, Figure 3a). Each of the six hydrogen bonds connecting each of the six pairs of atoms of the ($O\gamma_{Ser195}$, O_{phenol}, O_{oxy}, $N_{benzimidazole}$) tetrahedron is short[14,25−28] (<2.50 Å for O−O and <2.55 Å for N−O; Table 1a). The average O−O and N−O hydrogen bond lengths in this short hydrogen bond array are 2.14 ± 0.07 Å and 2.42 ± 0.06 Å, respectively (Table 1a).

In trypsin-CRA-7806 between pH 8.0 and 8.5 there is an equilibrium between two alternate hydrogen-bonding modes (three-centered ($O\gamma_{Ser195}$, O_{oxy}, O_{phenol}) short and two-centered ($O\gamma_{Ser195}-O_{phenol}$) normal), involving discrete disorder of Ser195 between conformer 1 ($\chi1 = -76 \pm 9°$) and conformer 2 ($\chi1 = -24 \pm 8°$) (Figures 1b and 3b). In conformer 2, $O\gamma_{Ser195}$ is shifted by more than 1 Å into or toward the oxyanion hole, breaking the $N\epsilon2_{His57}-O\gamma_{Ser195}$ hydrogen bond and displacing H_2O_{oxy} (Figure 3b, Table 1b). At high pH lengthening of some of the active site hydrogen bonds in trypsin-CRA-7806 is associated with a subtle shift of the bound inhibitor away from the active site (Figure 3c).

Novel Multicentered, H_2O_{oxy}- and H_2O_{S2}-Mediated Short Hydrogen Bond Arrays at Low pH. An unanticipated, novel feature of trypsin-CRA-7806 (pH 4.7) and of trypsin-CRA-8696 (pH 5.7−6.3) is the participation in multicentered short hydrogen bonding not only of H_2O_{oxy} but also of another water, H_2O_{S2}, on the other (S2) side of the inhibitor (Figures 1c and 4a, Table 1c,d). In trypsin-CRA-8696 (pH 6.1) this motif can be viewed

(15) Finucane, M. D.; Malthouse, J. P. G. A study of the stabilization of tetrahedral adducts by trypsin and δ-chymotrypsin. *Biochem. J.* **1992**, *286*, 889−900.
(16) Liang, T.-C.; Abeles, R. H. Complex of α-chymotrypsin and *N*-acetyl-L-leucyl-L-phenylalanyl trifluoromethyl ketone: structural studies with NMR spectroscopy. *Biochemistry* **1987**, *26*, 7603−7608.
(17) O'Connell, T. P.; Malthouse, J. P. G. Determination of the ionization state of the active-site histidine in a subtilisin-(chloromethane inhibitor) derivative by ^{13}C NMR. *Biochem. J.* **1996**, *317*, 35−40.
(18) Wilmouth, R. C.; Edman, K.; Neutze, R.; Wright, P. A.; Clifton, I. J.; Schneider, T. R.; Schofield, C. J.; Hajdu, J. X-ray snapshots of serine protease catalysis reveal a tetrahedral intermediate. *Nature, Struct. Biol.* **2001**, *8*, 689−694.
(19) Kuhn, P.; Knapp, M.; Soltis, S. M.; Granshaw, G.; Thoene, M.; Bott, R. The 0.78 Å structure of a serine protease: *Bacillus lentus* subtilisin. *Biochemistry* **1998**, *37*, 13446−13452.
(20) Betzel, C.; Gourinath, S.; Kumar, P.; Kaur, P.; Perbandt, M.; Eschenburg, S.; Singh, T. P. Structure of a serine protease proteinase K from *Tritirachium album limber* at 0.98 Å resolution. *Biochemistry* **2001**, *40*, 3080−3088.
(21) Katz, B. A.; Sprengeler, P.; Luong, C.; Verner, E.; Elrod, K.; Kirtley, M.; Janc, J.; Spencer, J.; Breitenbucher, J. G.; Hui, H.; McGee, D.; Allen, D.; Martelli, A.; Mackman, R. L. Engineering inhibitors highly selective for the S1 sites of Ser190 trypsin-like serine protease drug targets. *Chem. Biol.* **2001**, *8*, 1107−1121.
(22) Verner, E.; Katz, B. A.; Spencer, J.; Allen, D.; Hataye, J.; Hruzewicz, W.; Hui, H.; Kolesnikov, A.; Li, Y.; Luong, C.; Martelli, A.; Radika, K.; Rai, R.; She, M.; Shrader, W.; Sprengeler, P.; Trapp, S.; Wang, J.; Young, W. B.; Mackman, R. L. Development of serine protease inhibitors displaying a multicentered short (<2.3 Å) hydrogen bond binding mode: Inhibitors of urokinase-type plasminogen activator and factor Xa. *J. Med. Chem.* **2001**, *44*, 2753−2771.
(23) Rai, R.; Kolesnikov, A.; Li, Y.; Young, W.; Leahy, E.; Sprengeler, P.; Verner, E.; Shrader, W.; Burgess-Henry, J.; Sangalang, J.; Allen, D.; Chen, X.; Katz, B. A.; Luong, C.; Elrod, K.; Cregar, L. Development of Potent and Selective Factor Xa Inhibitors. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1797−1800.
(24) Mackman, R. L.; Katz, B. A.; Breitenbucher, G.; Hui, H.; Verner, E.; Luong, C.; Liang, L.; Sprengeler, P. Exploiting subsite S1 of trypsin-like serine proteases for selectivity: potent and selective inhibitors of urokinase-type plasminogen activator. *J. Med. Chem.* **2001**, *44*(23), 3856−3871.
(25) Emsley, J. Very strong hydrogen bonding. *J Chem. Soc. Rev.* **1980**, *9*, 91−124.
(26) Hibbert, F.; Emsley, J. Hydrogen bonding and chemical reactivity. *Adv. Phys. Org. Chem.* **1990**, *26*, 255−379.
(27) Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*; Springer-Verlag: Berlin, 1991.
(28) Perrin, C. L.; Nielson, J. B. "Strong" hydrogen bonds in chemistry and biology. *Annu. Rev. Phys. Chem.* **1997**, *48*, 511−544.

Table 1. Hydrogen Bond Lengths and Angles at the Active Sites of Trypsin, Thrombin, and Urokinase Complexes[a]

protease:	(a) trypsin;	(b) trypsin;	(c) trypsin;	(d) trypsin;	(e) trypsin;	(f) uPA; 7806,
inhibitor CRA#:	7806;	7806;	8696;	8696;	7806, 8696;	8696;
H-bond array type:[b]	3;	6;	2 (H_2O_{oxy});[d]	2 (H_2O_{S2});[d]	4;	4;
figures:	1a, 3a;	1b, 3b;	1c, 4a;	1c, 4a;	similar to 1a;	similar to 1a;
pH:	various[c]	8.2	5.8–6.1	5.8–6.1	various[e]	6.5
Hydrogen Bond Length (Å)						
$O\gamma_{Ser195(conf\ 1)}-O_{phenol}$	**2.10(02)**	**2.50**	**2.16(13)**	**2.16(14)**	**2.20(10)**	**2.35(05)**
$O\gamma_{Ser195(conf\ 2)}-O_{phenol}$		2.59				
$O\gamma_{Ser195(conf\ 1)}-O_{oxy}$ (or $-O_{S2}$)	**2.10(06)**	**2.20**	**2.16(14)**	**2.07(04)**	**2.25(04)**	**2.37(17)**
$O_{oxy}-O_{phenol}$	**2.22(06)**	**2.43**	**2.14(14)**	**2.07(06)**	**2.32(07)**	**2.44(08)**
$N_{ind(or\ benz)}-O_{oxy}$ (or $-O_{S2}$)	**2.45(06)**	3.16	**2.44**[d]	**2.41(08)**	2.61(10)	2.74(06)
$N_{ind(or\ benz)}-O_{phenol}$	**2.37(06)**	2.56	**2.53(01)**	**2.55(03)**	**2.52(07)**	2.58(01)
$N_{ind(or\ benz)}-O\gamma_{Ser195(conf\ 1)}$	**2.37(06)**	2.78	**2.52(01)**	**2.52(01)**	**2.42(04)**	2.60(11)
$N_{ind(or\ benz)}-O\gamma_{Ser195(conf\ 2)}$		3.04				
$N\epsilon2_{His57}-O\gamma_{Ser195(conf\ 1)}$	3.19(15)	3.21	3.23(04)	3.23(04)	3.16(08)	*3.30(09)*
$N\epsilon2_{His57}-O_{phenol}$	2.92(02)	**2.53**	3.06(02)	3.07(04)	2.88(07)	2.75(10)
$N_{Ser195}-O_{oxy}$	3.11(07)	3.04	3.17(01)		3.19(04)	3.03(05)
$N_{Gly193}-O_{oxy}$	3.00(06)	2.82	2.96(19)		2.87(01)	2.65(06)
$N_{Ser195}-O\gamma_{Ser195(conf\ 2)}$		**2.41**				
$O_{oxy}-O_{distal}$	2.99(07)		2.72(06)		2.69(09)	2.46
$O_{distal}-O_{Phe/Val41}$	2.75(07)		2.75(12)		2.61(08)	2.90
$N\epsilon2_{His57}-O_{S2}$				**2.49(07)**		
Hydrogen Bond Angle (deg)						
$O\gamma_{Ser195(conf\ 1)}-H\gamma_{Ser195(conf\ 1)}-O_{phenol}$	**171(5)**	**175**	**164(7)**	**168(2)**	**164(2)**	**153(15)**
$O\gamma_{Ser195(conf\ 2)}-H\gamma_{Ser195(conf\ 2)}-O_{phenol}$		**179**				
$O_{oxy}-H1_{oxy}-O\gamma_{Ser195(conf\ 1)}$	141(2)	147	145(4)	146(4)	140(2)	147(1)
$O_{oxy}-H2_{oxy}-O_{phenol}$	141(2)	147	145(5)	145(3)	140(2)	147(1)
$N_{ind(or\ benz)}-H_{ind(or\ benz)}-O\gamma_{Ser195(conf\ 1)}$	**156(2)**	149	**157(1)**	**157(1)**	**156(1)**	**156(4)**
$N_{ind(or\ benz)}-H_{ind(or\ benz)}-O\gamma_{Ser195(conf\ 2)}$		139				
$N_{ind(or\ benz)}-H_{ind(or\ benz)}-O_{oxy}$ (or $-O_{S2}$)	121(1)	124	124(1)	107(4)	121(1)	120(4)
$N_{ind(or\ benz)}-H_{ind(or\ benz)}-O_{phenol}$	116(1)	127	119(1)	121(3)	117(3)	117(3)
$N\epsilon2_{His57}-H\epsilon2_{His57}-O_{S2}$				119(4)		

protease:	(g) trypsin;	(h) trypsin;	(i) trypsin;	(j) trypsin;	(k) thrombin;
inhibitor CRA#:	8696;	8696;	8696;	8696;	7806, 8696;
H-bond array type:[b]	8;	9;	10;	5;	3;
figures:	1d;	1e, 4b;	similar to 1e;	1f;	similar to 1a;
pH:	7.2	7.7	8.9[f]	9.0	5.2–11.4
Hydrogen Bond Length (Å)					
$O\gamma_{Ser195(conf\ 1)}-O_{phenol}$	**2.15**			**2.12(11)**	**2.09(11)**
$O\gamma_{Ser195(conf\ 2)}-O_{phenol}$	2.54	2.58	**2.23**	**2.30(04)**	
$O\gamma_{Ser195(conf\ 1)}-O_{oxy}$				**2.02(05)**	**2.28(11)**
$O_{oxy}-O_{phenol}$				**2.29(01)**	**2.23(07)**
$N_{ind(or\ benz)}-O_{oxy}$				2.90(02)	**2.36(08)**
$N_{ind(or\ benz)}-O_{phenol}$	2.63	2.70	2.57	2.59(01)	**2.48(04)**
$N_{ind(or\ benz)}-O\gamma_{Ser195\ (conf\ 1)}$	2.71			2.71(09)	**2.24(05)**
$N_{ind(or\ benz)}-O\gamma_{Ser195(conf\ 2)}$	*3.35*	3.51	3.25	2.93(26)	
$N\epsilon2_{His57}-O\gamma_{Ser195(conf\ 1)}$	*3.55*			2.90(28)	3.07(12)
$N\epsilon2_{His57}-O_{phenol}$	2.72	2.65	2.70	2.74(16)	2.73(07)
$N_{Ser195}-O_{oxy}$				2.95 (02)	2.95(09)
$N_{Gly193}-O_{oxy}$				3.02(11)	2.91(11)
$N_{Ser195}-O\gamma_{Ser195(conf\ 2)}$	2.51	2.63	2.68	2.54(02)	
$O_{oxy}-O_{distal}$					2.79(09)
$O_{distal}-O_{Phe/Val41}$					2.63(08)
Hydrogen Bond Angle (deg)					
$O\gamma_{Ser195(conf\ 1)}-H\gamma_{Ser195(conf\ 1)}-O_{phenol}$	**164**			**157(13)**	**175(4)**
$O\gamma_{Ser195(conf\ 2)}-H\gamma_{Ser195(conf\ 2)}-O_{phenol}$	**169**	**174**	**177**	**172(5)**	
$O_{oxy}-H1_{oxy}-O\gamma_{Ser195(conf\ 1)}$				141(4)	140(3)
$O_{oxy}-H2_{oxy}-O_{phenol}$				141(4)	140(3)
$N_{ind(or\ benz)}-H_{ind(or\ benz)}-O\gamma_{Ser195(conf\ 1)}$	141			**168(5)**	**151(4)**
$N_{ind(or\ benz)}-H_{ind(or\ benz)}-O\gamma_{Ser195(conf\ 2)}$			137	144(8)	
$N_{ind(or\ benz)}-H_{ind(or\ benz)}-O_{oxy}$				124(5)	114(6)
$N_{ind(or\ benz)}-H_{ind(or\ benz)}-O_{phenol}$	119	122	130	117(3)	119(2)

[a] Except where noted, all trypsin structures are in form *b* ($P2_12_12_1$, large cell). Short H-bonds and ideal or near-ideal short H-bond angles (>150°) are given in boldface type. Values in italics correpond to distances longer than H-bonds, or to unreasonable H-bond angles. H-bond angles unambiguously determined from heavy-atom positions are underlined. [b] Figure 2 legend. [c] Averages and standard deviations are derived from four structures (pH 5.7, 6.4, 7.3, 9.0). [d] Hydrogen bond distances involving H_2O_{oxy} are in one column, while the counterpart distances involving H_2O_{S2} are in the adjacent column. The $N_{indole\ (or\ benzimidazole)}$-$O_{oxy}$ distance is short at pH 6.1 only. At pH 5.8 it is 2.76 Å. [e] Averages and standard deviations are derived from trypsin-CRA-7806 (pH 9.1 and 9.3) and trypsin-CRA-8696 (pH 4.6 and 5.4). [f] Crystal form *c* ($P3_121$).

as two fused tetrahedra ($O\gamma_{Ser195}$, O_{phenol}, N_{indole}, H_2O_{oxy}) (Table 1c), and ($O\gamma_{Ser195}$, O_{phenol}, N_{indole}, H_2O_{S2}) (Table 1d), sharing a common face. Nine short hydrogen bonds connect all the atom pairs, except for the non-hydrogen-bonded $H_2O_{oxy}-H_2O_{S2}$ pair

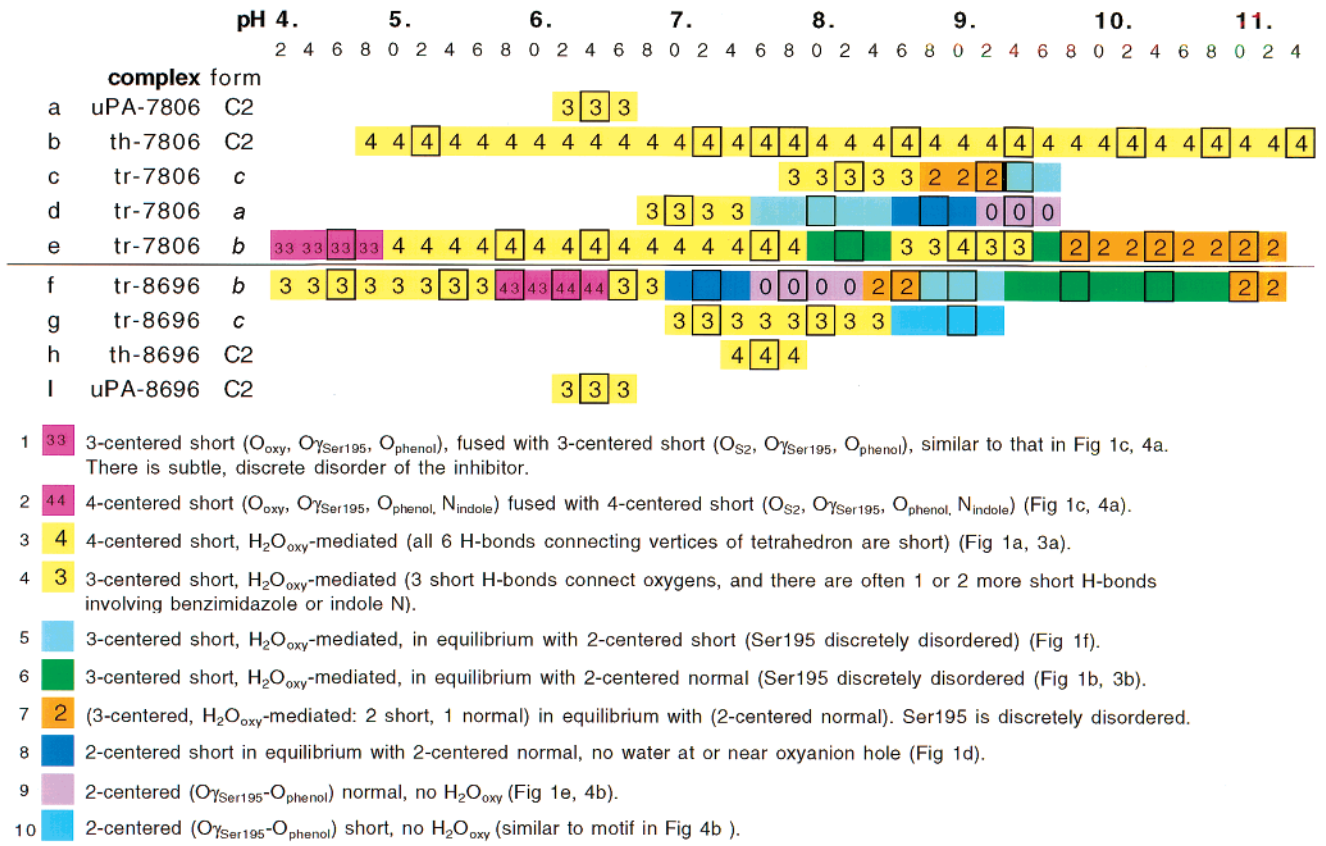

Figure 2. Color-coded representation of structure of active site hydrogen bond motifs as a function of pH. Actual structures are depicted with square boxes. Interpolations of transitions between binding modes are based on occupancies of H_2O_{oxy}, H_2O_{S2}, and the alternate conformers of Ser195 and/or on the structures of the complex at bracketing pH values. A bold vertical line represents a discontinuity in structures incurred by changing the soaking solution from $MgSO_4$ (for pH values less than about 9.0) to citrate (for pH values greater than about 9.0). Trypsin, thrombin, and urokinase are abbreviated as, tr, th, and uPA. See the Experimental Section for a description of crystal forms *a*, *b*, and *c* of trypsin.

(Figure 4a). One additional short hydrogen bond, $N\epsilon2_{His57}-O_{S2}$, extends outward from the periphery of the cluster.

Hydrogen Bond Array Transitions as a Function of pH, Inhibitor, and Protease. The pH dependence of the active site structure of trypsin-CRA-8696 (Figure 2f) is significantly different than that of the benzimidazole counterpart, trypsin-CRA-7806 (Figure 2e), especially over some pH ranges. Between pH 6.5 and 6.8, the hydrogen bonding is four-centered ($O\gamma_{Ser195}$, O_{phenol}, O_{oxy}, $N_{benzimidazole}$) short in the benzimidazole complex, while it is three-centered ($O\gamma_{Ser195}$, O_{phenol}, O_{oxy}) short in the indole complex (Table 1e). These two arrays are very similar to one another, except that in the three-centered short array the $N_{indole}-O_{oxy}$ hydrogen bond is somewhat longer (2.61 Å) (Table 1e). Between pH 6.9 and 7.4, H_2O_{oxy} participates in a four-centered short hydrogen bond array in the benzimidazole complex, while in the indole complex H_2O_{oxy} is completely absent. In the indole complex (pH 6.9−7.4) Ser195 is discretely disordered between two-centered ($O\gamma_{Ser195}-O_{phenol}$) short and two-centered ($O\gamma_{Ser195}-O_{phenol}$) normal hydrogen-bonding modes (Figure 1d, Table 1g). In the indole complex between pH 7.5 and 8.2, Ser195 adopts only one conformation ($\chi = -25°$), in which $O\gamma_{Ser195}$ occupies the oxyanion hole, receiving hydrogen bonds from N_{Gly193} and from N_{Ser195}, and donating only one ordinary hydrogen bond to O_{phenol} (Figures 1e and 4b, Table 1h). This hydrogen-bonding mode is significantly different from the multicentered short hydrogen bond array of the benzimidazole complex over the same pH range. The absence of multicentered short hydrogen bond arrays in the indole complex at near-neutral pH values may reflect a decreased ionizability of the indole compared with the benzimidazole.

The active site hydrogen-bonding motif of thrombin-CRA-7806 is unique in its pH independence (Figure 2b), with the four-centered, H_2O_{oxy}-mediated short hydrogen bond array persisting to pH values >11.4. The presumed donation by $N\epsilon2_{His57}$ of ionic hydrogen bonds both to O_{phenol} and to $O\gamma_{Ser195}$ (Table 1k) suggests a dramatic elevation in the pK_a of His57.[14] Lack of H_2O_{oxy} dissociation from the active site of thrombin at high pH may reflect the fact that H_2O_{oxy} is normally bound in apo-thrombin, but not in apo-trypsin, because of the wider oxyanion hole (by 0.5 Å) of thrombin.[14]

Comparison of pH Dependencies of Inhibition of Trypsin by CRA-7806 and CRA-8696. Figure 5 compares the pH dependencies of inhibition of trypsin and urokinase by the benzimidazole (CRA-7806) and indole (CRA-8696) isosteres. The parabolic shape and the minima (K_i(min) values) in the K_i versus pH profiles reflect ionization of the phenol, required for formation of the short, ionic hydrogen bond arrays.[14] The K_i-(min) values toward trypsin, urokinase, and thrombin for the indole are 6.7-, 5.0-, and 13-fold lower than for the benzimidazole (Table 2). Decreased potencies of benzimidazoles over indole isosteres have been observed in other comparisons.[14,21]

At the pH values where the potencies of CRA-7806 and CRA-8696 toward trypsin are maximal (pH 7.5 in both cases), the structures of trypsin-CRA-7806 and trypsin-CRA-8696 differ dramatically in terms of active site short hydrogen bonding. Although the indole is more than 6.7-fold more potent than the

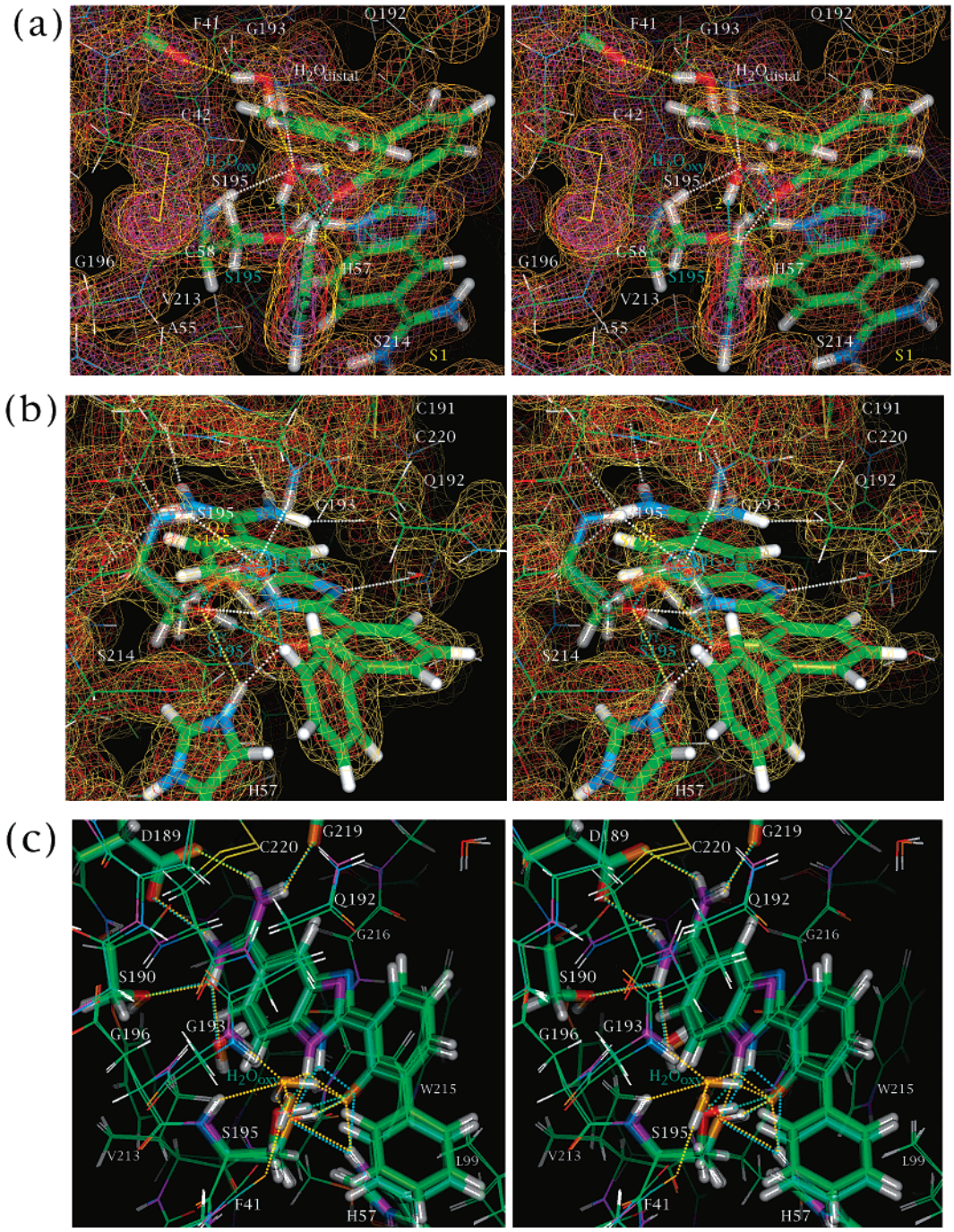

Figure 3. (a) Trypsin-CRA-7806 (pH 7.7) (1.50 Å resolution) superimposed on ($2|F_o| - |F_c|$), α_c map contoured at 1.0σ, 2.4σ, and 3.8σ. The standard atom color code is used (carbons, green; oxygens, red; nitrogens, blue). Short O−O and N−O hydrogen bonds are light blue and cyan, respectively, and the labels of their involved groups or atoms are shown in cyan. Protease-inhibitor and protease−H_2O_{oxy} hydrogen bonds are shown in white, while other protease−water and protease-protease hydrogen bonds are yellow. In these and subsequent figures some of the protons described in the Discussion are numbered in yellow. (b) Structure and ($2|F_o| - |F_c|$), α_c map for trypsin-CRA-7806 (pH 8.2) (1.39 Å resolution) superimposed on ($|F_o| - |F_c|$), α_c map, contoured at 2.5σ (blue) and 3.0σ (green) around H_2O_{oxy}. The hydroxyl group of Ser195 in conformation 1 is shown in red (labeled in cyan), while that of conformation 2 is orange (labeled in yellow). Short hydrogen bonds involving conformation 1 of Ser195 are cyan. (c) Superposition of trypsin-CRA-7806 (pH 5.7) (carbons, cyan; oxygens, orange; nitrogens, purple; hydrogen bonds, blue) onto trypsin-CRA-7806 (pH 11.0) (standard atom color code; hydrogen bonds, orange). Note that at pH 11.0 Ser195 is discretely disordered between two conformations, one of which is obscured by the similar conformation at pH 5.7. For clarity, H_2O_{oxy} (of partial occupancy at pH 11.0) is not shown in trypsin-CRA-7806 (pH 11.0).

benzimidazole, there are no short hydrogen bonds in the indole complex at pH 7.5, while there are six short hydrogen bonds in the benzimidazole complex at this pH.

Dependence of Active Site Hydrogen Bond Motif on Crystal Form. Long-range crystal packing interactions can have a profound effect on the type of active site hydrogen bond motif and on the length of the associated hydrogen bonds. For example, trypsin-CRA-7806 in crystal form *b* exhibits a three-centered ($O\gamma_{Ser195}$, O_{oxy}, O_{phenol}) short hydrogen bond network at pH 9.4 (Figure 2e, Table 1e), whereas in form *a* at the same

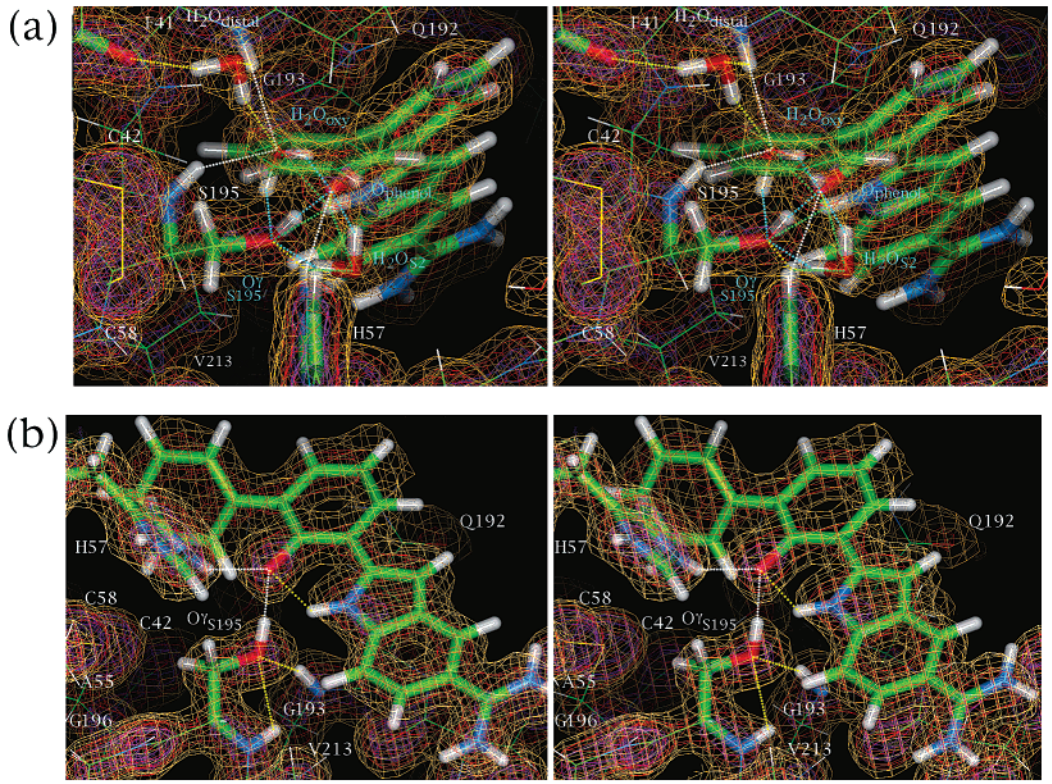

Figure 4. (a) Trypsin-CRA-8696 and $(2|F_o| - |F_c|)$, α_c map (pH 6.1) (1.50 Å resolution). The density and structure of the distal phenyl group of the inhibitor in the foreground is rendered transparent so as to not obscure the structure and interactions of H_2O_{oxy} behind it. (b) Structure of trypsin-CRA-8696 (pH 7.7) and $(2|F_o| - |F_c|)$, α_c map (1.39 Å resolution).

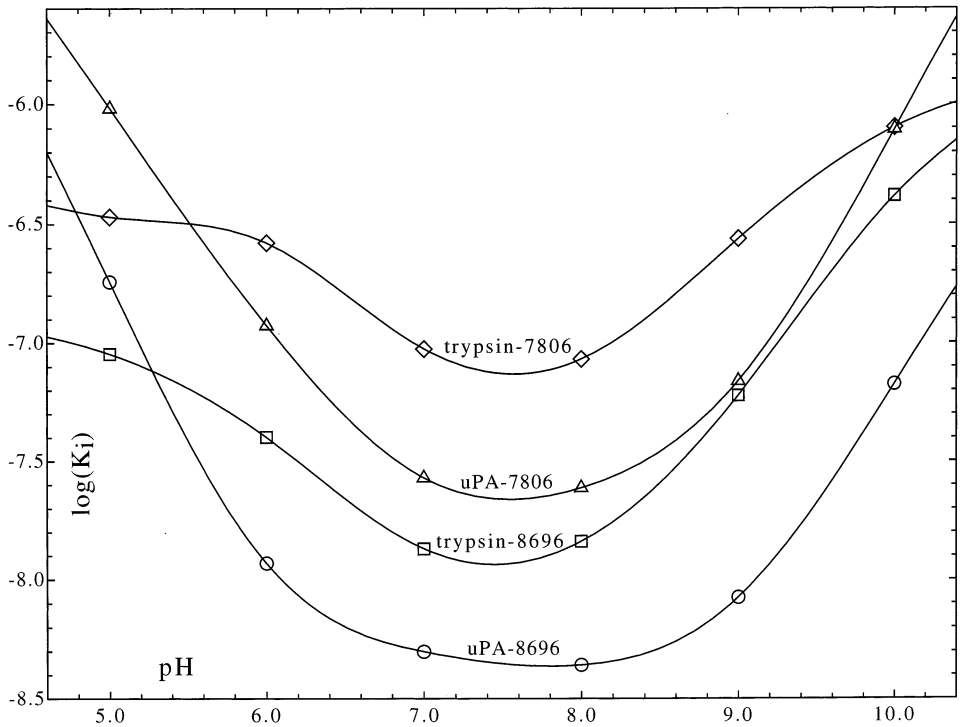

Figure 5. log K_i (toward trypsin and urokinase) versus pH for the indole, CRA-8696, and the benzimidazole, CRA-7806.

pH it exhibits a two-centered ($O\gamma_{Ser195}-O_{phenol}$) ordinary hydrogen bond (Figures 1e and 2d). In a third crystal form (form *c*) (Figure 2c), the structure adopts yet a third active site hydrogen bond motif at pH 9.4, involving an equilibrium between three-centered ($O\gamma_{Ser195}$, O_{oxy}, O_{phenol}) short and two-centered ($O\gamma_{Ser195}-O_{phenol}$) short, with Ser195 discretely disordered (Figure 1f). In addition to crystal packing interactions, synthetic mother liquor composition of the soaked crystals

Table 2. K_i(min) and pH(min) Values of Inhibitors toward Proteases

scaffold	CRA no.	protease	pH(min)	K_i(min) (μM)
benzimidazole	7806	trypsin	7.5	0.074
		urokinase	7.5	0.022
		thrombin	7.7	0.69
indole	8696	trypsin	7.5	0.011
		urokinase	7.9	0.0044
		thrombin	7.7	0.054

($MgSO_4$ versus sodium citrate) can also affect the type of active site hydrogen bond array, presumably due to ionic strength differences.

Other Short Hydrogen Bond Arrays Not Associated with Inhibitor Binding in Trypsin and Urokinase Crystals. In addition to a large manifold of short hydrogen bond arrays that mediate inhibitor binding (Figure 1), we have also discovered similar short hydrogen bond arrays that are integral components of trypsin and of urokinase crystal structures (Figure 6). In many trypsin complexes, there is a two-centered short hydrogen bond between $O\eta_{Tyr59}$ with $O\delta1_{Asp153}$ of a symmetry-related molecule (Figure 6a)[14] and between $O\eta_{Tyr228}$ and an ordered water molecule (Figures 6b and 7a). The lengths of these hydrogen bonds depend on pH, the space group, and the identity of the active site inhibitor. Hydrogen bond lengths and angles for the short hydrogen bond arrays in Figure 6 are given in Table 3.

In urokinase crystal structures, a four-oxygen centered short hydrogen bond array among two co-bound citrate molecules (Figures 6c and 7b,c) resembles the four-centered ($O\gamma_{Ser195}$, O_{oxy}, O_{phenol}, $N_{benzimidazole}$) short hydrogen bond array at the active site of trypsin-CRA-7806 (Figures 1a and 3a). The tetrahedral array of short hydrogen bonding groups involves one hydroxyl (O1) and one carboxylate oxygen (O2) in one citrate monomer and the chemically equivalent oxygens (O1′ and O2′) in the other (crystallographically independent) monomer (Figures 6c and 7b,c). There are six short hydrogen bonds connecting the six pairs of the four oxygens. However only one hydrogen bond is linear, and four of the remaining five hydrogen bond angles are poor (Table 3). Assumptions used for modeling the proton positions in the short hydrogen bond array are given in the caption of Figure 7b.

The oxygens of each citrate are hydrogen bond acceptors in an extensive array of ordinary hydrogen bonds from surface residues of two symmetry-related protein molecules. One protein provides the side chains of Arg36, Tyr67, Arg82, and Lys110A, while the symmetry-related molecule (colored differently in Figure 7b) provides the side chains of His165, Arg166, Arg230, and the main-chain NH groups of His165 and of Arg166. There are thus a total of 24 hydrogen bonds donated by protein or water groups to 13 of the 14 oxygens of the citrate dimer.

Discussion

Diversity of Short Hydrogen Bond Arrays both in Protease-Inhibitor Complexes and in Protein Structures Themselves. Alterations in the inhibitor scaffold or in the pH produce a rich diversity in short hydrogen-bond-mediated protease-inhibitor active site motifs (Figure 1). Our focus on the criteria, constraints, and energetics of short hydrogen bond formation for structure-based drug design also led to recognition of other short hydrogen bonds removed from the ligand binding sites (Figures 6 and 7), along with those at the active site, often in the same crystal structures. Proper crystallographic refinement of such short hydrogen-bonded structural motifs requires removal of appropriate van der Waals force field terms from conventional refinements. Thus, some short hydrogen bonds in protein crystal structures may go unnoticed, like that between $O\eta_{Tyr59}$ and $O\delta1_{Asp153'}$ (Figure 6a)[14] and that between $O\eta_{Tyr228}$ and H_2O3_{S1} (Figure 6b), recently observed in three well-studied crystal forms of trypsin.[29−31] The diverse examples discovered recently in these systems, both at and away from the active site, suggest that short hydrogen bonds in protein crystals may be more common than realized.

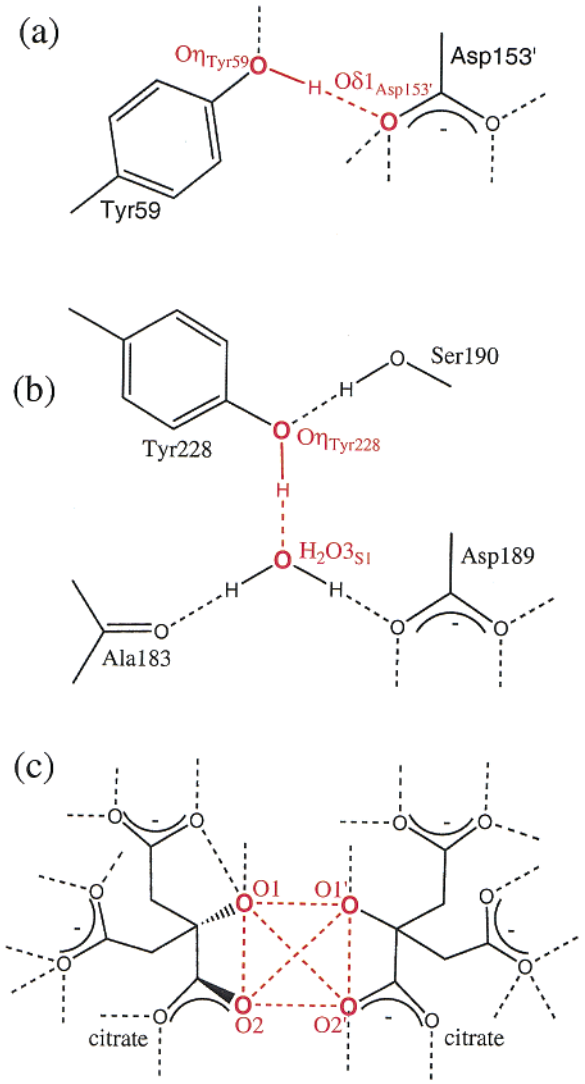

Figure 6. Short hydrogen bonds in trypsin and urokinase crystals away from the active site: (a) $O\eta_{Tyr59}-O\delta1_{Asp153'}$ hydrogen bond at a lattice interface in certain bovine trypsin complexes of crystal forms $P3_121$ and $P2_12_12_1$ (form *a*); (b) $O\eta_{Tyr228}-H_2O3_{S1}$ hydrogen bond at the S1 site of certain trypsin complexes; (c) multicentered short hydrogen bond array involving citrate dimer at the lattice interface of urokinase crystals.

(29) Krieger, M.; Kay, L. M.; Stroud, R. M. The structure and specific binding of trypsin: A comparison of inhibited derivatives and a model for substrate binding. *J. Mol. Biol.* **1974**, *83*, 209−230.
(30) Bode, W.; Schwager, P. The refined crystal structure of bovine beta-trypsin at 1.8 Å resolution. II. Crystallographic refinement, calcium binding site, benzamidine binding site and active site at pH 7.0. *J. Mol. Biol.* **1975**, *98*, 693−717.
(31) Mangel, W. F.; Singer, P. T.; Cyr, D. M.; Umland, T. C.; Toledo, D. L.; Stroud, R. M.; Pflugrath, J. W.; Sweet, R. M. Structure of an acyl-enzyme intermediate during catalysis: (guanidinobenzoyl)trypsin. *Biochemistry* **1990**, *29*, 8351−8357.

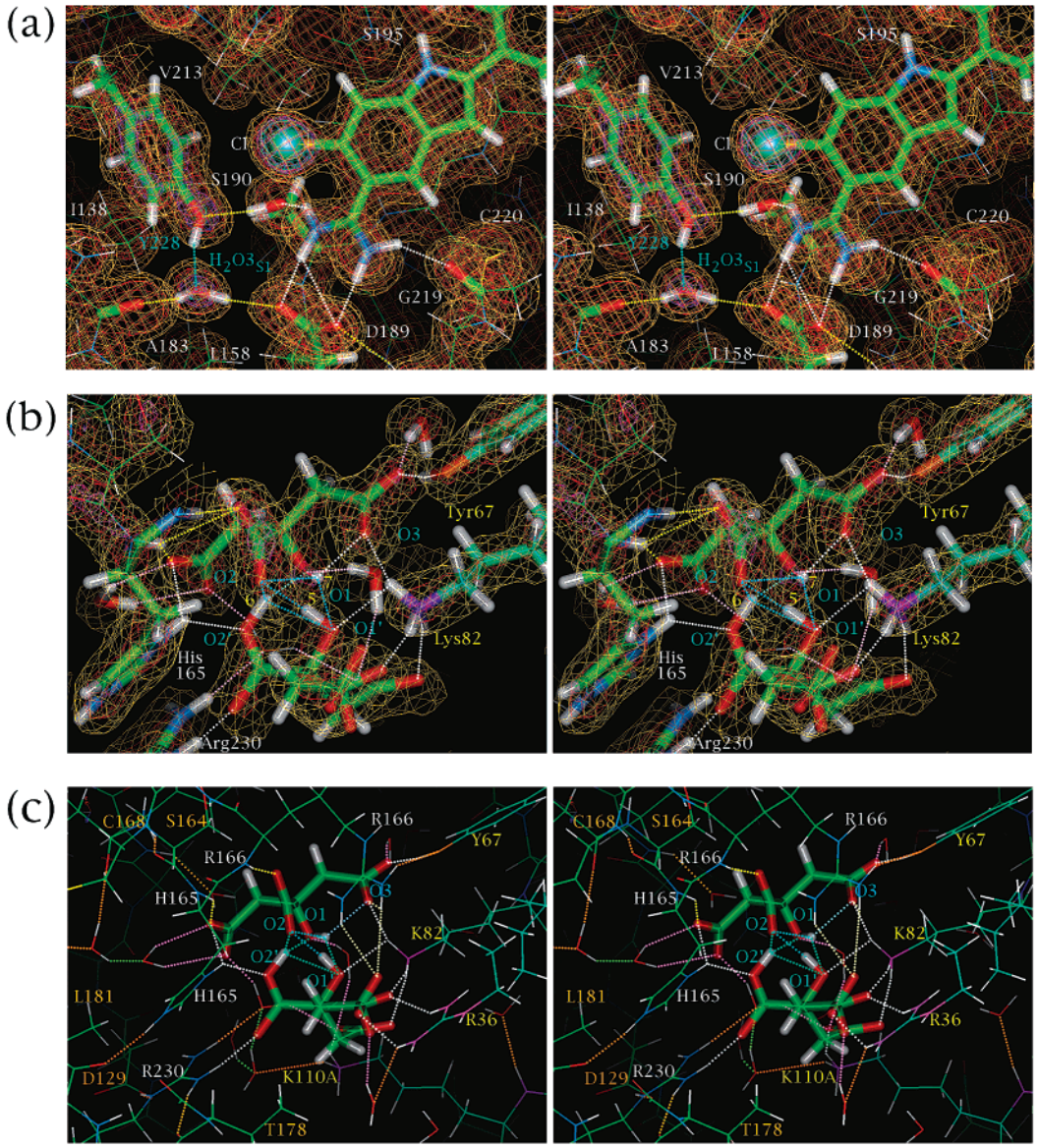

Figure 7. (a) Structure, environment, and $(2|F_o| - |F_c|)$, α_c map of the short $O\eta_{Tyr228}-H_2O3_{S1}$ hydrogen bond (2.42 Å) in trypsin-CRA-10991[21] (pH 8.3). (b) Structure and $(2|F_o| - |F_c|)$, α_c map for the four-centered short hydrogen bond array among the citrates co-bound in urokinase crystals. The hydrogen bond lengths and clarity of the motif depend on the crystal quality and identity of the inhibitor bound at the active site. Short hydrogen bonds are cyan. Hydrogen bonds to citrate from side chains, main chains, and waters are shown in white, yellow, and pink, respectively. It is assumed that one of the carboxylates of the two involved in the short hydrogen bond array is protonated, so that only one negative charge is distributed among the (O1, O2, O1′, O2′) hydrogen-bonding tetrahedron, and that the carboxylate proton is in the carboxylate plane. Protonation of O2′ rather than O2 is favored because it leads to a network in which O1 donates a short hydrogen bond to O2 and to O1′ as well as an ordinary hydrogen bond to O3. In alternate protonation schemes there are fewer hydrogen bonds, or hydrogen bonds with poorer angles. (c) Structure, environment, and extended hydrogen bond network of the citrate dimer.

Ionic, Noncovalent Nature of the Short Hydrogen Bond Arrays. Each of the multicentered short hydrogen bond arrays in Figures 1 and 6 is located in a highly polar environment that extends the core short hydrogen bond network with many additional ordinary hydrogen bonds that neutralize the negative charge of the core. Ionization of one or more constituent groups is a common feature of short hydrogen bonds.[25−28] In all but one of the short hydrogen bond arrays in Figures 1 and 6 a negative charge is expected to be either localized on the oxygen with the lowest pK_a or distributed, probably unevenly, among the constituent atoms. Figure 6b stands out as a case where neither the short hydrogen bond donor nor the acceptor is ionized.

Dissimilarity of pK_a Values of Short Hydrogen Bond Constituents. Another similarity among the short hydrogen bond motifs in Figures 1 and 6 is the large apparent difference in the pK_a values of the donor/acceptor groups. The pK_a values of the constituent groups of the protease-inhibitor hydrogen bond arrays (Figures 1, 3, and 4), a serine hydroxyl (∼14), a water (15.7),[32] a phenol hydroxyl (4.3−9.1, depending on the inhibitor),[14] and a benzimidazole nitrogen (∼4), are substantially different from one another in their isolated states. Although the interactions and environment of the bound inhibitor can alter

(32) Carey, F. A.; Sundberg, R. J. In *Advanced Organic Chemistry*, 4th ed.; Kluwer Academic/Plenum Publishers: New York, Boston, Dordrecht, London, Moscow, 2000; Part A (Structure and Mechanisms).

Table 3. Short Hydrogen Bond Lengths and Angles (Away from Active Site) in Trypsin and Urokinase Structures[a]

H-bond	distance (Å)	H-bond angle	angle (deg)
(a) $O\eta_{Tyr228}-O(H_2O3_{S1})$[b]	2.50(04)	$O\eta_{Tyr228}-H\eta_{Tyr228}-O(H_2O3_{S1})$	**153(4)**
$O(H_2O3_{S1})-O_{Ala183}$[b]	2.92(12)	$O(H_2O3_{S1})-H1(H_2O3_{S1})-O_{Ala183}$	164(1)
$O(H_2O3_{S1})-O\delta1_{Asp189}$[b]	2.66(06)	$O(H_2O3_{S1})-H2(H_2O3_{S1})-O\delta1_{Asp189}$	164(1)
(b) $O\eta_{Tyr59}-O\delta1_{Asp153'}$[c]	**2.14(06)**	$O\eta_{Tyr59}-H\eta_{Tyr59}-O\delta1_{Asp153'}$	**174(5)**
(c) O1−O2[d]	2.47(05)	O1−HO1−O2	113(3)
O1−O1′[d]	2.19(09)	O1−HO1−O1′	128(3)
O1−O2′[d]	2.30(07)	O2′−HO2′−O1	137(7)
O2−O1′[d]	**2.41(11)**	O1′−HO1′−O2	**172(8)**
O2−O2′[d]	2.28(18)	O2′−HO2′−O2	116(7)
O1′−O2′[d]	2.47(07)	O2′−HO2′−O1′	104(2)
O1−O3[d]	2.94(13)	O1−HO1−O3	113(3)

[a] Short hydrogen bonds with good hydrogen bond angles are in boldface type. [b] Trypsin-CRA-10991[21] (pH 4.7, 5.7, 6.6, 7.4, 8.3, 9.3, 10.1, 10.8). [c] Ten trypsin complexes with the shortest H-bond ($P2_12_12_1$ (form *a*) and $P3_121$).[14] [d] Citrate−citrate for eight urokinase complexes.[14,21,45]

its pK_a values, some inequivalence is most likely preserved in the complexes. Similarly, the $O\eta_{Tyr59}-O\delta1_{Asp153'}$ and $O\eta_{Tyr228}-H_2O3_{S1}$ short hydrogen bonds involve groups with significantly different pK_a values: 9.6 and 4.0 for the isolated Tyr and Asp side chains[33] and 9.6 and 15.7 for Tyr and H_2O. Finally, the pK_a values of the hydroxyl and carboxylate of the short hydrogen bond array of the citrate dimer in the urokinase crystals are also different from one another in solution (∼16[32] and 3.1,[34] respectively). Usher et al.[11] and Schwartz et al.[12] also found that equivalent pK_a values were not required for unusually short hydrogen bond formation, nor did the short hydrogen bonds significantly enhance inhibitor potency.[12]

In two-centered short hydrogen bonds there is often a low or absent vibrational barrier for transfer of a proton from one oxygen to another.[1−5,35] For such a "low-barrier" hydrogen bond the proton is equally or almost equally associated with the two oxygens, imparting covalency to the hydrogen bond.[35] However, short hydrogen bonds between groups with different pK_a values located in highly polar environments such as those described here are expected to be largely electrostatic with little covalent character.[6,7,10]

Minor Contribution of the Short Hydrogen Bonds to Inhibitor Affinity. Owing to the electrostatic, noncovalent nature of the short hydrogen bonds in Figure 1, they are not dramatically stronger than ordinary hydrogen bonds. For example, even though there are six short hydrogen bonds in the benzimidazole complex at pH 7.7 (Figures 1a and 3a, Table 1a), and none in the indole counterpart at the same pH (Figures 1e and 4b, Table 1h), the indole is still more potent by 1.1 kcal/mol. Under conditions of equivalent active site hydrogen bonding modes, it has been noted that indoles are more potent than benzimidazole isosteres because an unfavorable conformational change, corresponding to about 1.5 kcal/mol,[14] is

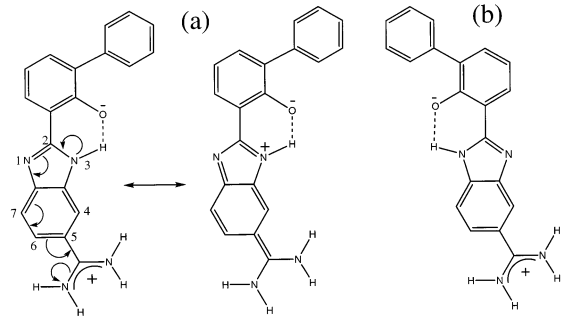

Figure 8. (a) Lowest energy (nonbinding) conformer of CRA-7806, in which the benzimidazole tautomer protonated at the 3-position is stabilized by resonance. (b) Binding conformation of CRA-7806.

required to convert the lowest energy benzimidazole tautomer (Figure 8a) to the binding tautomer (Figure 8b). However, the indole isostere, with only one possible internal hydrogen bond possible, is locked into the binding conformation. The moderately unfavorable conformational change required for binding of the benzimidazole appears to outweigh any binding enhancement imparted by the six short hydrogen bonds in the four-centered, H_2O_{oxy}-mediated array of the complex.

Another unfavorable change required for formation of H_2O_{oxy}-mediated multicentered short hydrogen bond arrays is the decrease in entropy associated with the co-binding and ordering of H_2O_{oxy}. In trypsin-CRA-7806 (pH 7.7) the fixing of the position and orientation of H_2O_{oxy} ($B = 18$ Å^2) (Figure 3a) is expected to correspond to a loss of up to 2.0 kcal/mol in binding energy,[36] whereas H_2O_{oxy} is completely absent in trypsin-CRA-8696 (pH 7.7) (Figure 4b).

Previously, the contribution to the potency of short hydrogen bonds over ordinary counterparts was assessed by comparison of the thermodynamics of binding of a short hydrogen-bond-forming inhibitor with that of an ordinary hydrogen-bond-forming isostere, in which the active site-directed phenol is replaced with a pyridone.[14] The four-centered short hydrogen bond array (Figure 1a) was estimated to contribute 1.7 kcal/mol more to binding than the same network of ordinary hydrogen bonds at pH 9.5.[14] Stabilization decreased to only 0.33 kcal/mol at pH 5.0. These values are in the lower range found for other short, single hydrogen bonds, 1.4−4.4 kcal/mol.[11−12,37−41]

Another reflection of the energetic similarity of the short, protease-inhibitor hydrogen bond arrays (Figure 1a) to ordinary hydrogen bonds (Figure 1e) is the dependence of hydrogen bond length on subtle factors such as long-range crystal-packing interactions. Because there are no direct crystal-packing contacts involving the bound inhibitor in any of the trypsin crystal forms, it is presumed that long-range electrostatic interactions affect the active site hydrogen-bonding motif, perhaps through per-

(33) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.* **1994**, *238*, 415−436.
(34) Bénézeth, P.; Palmer, D. A.; Wesolowski, D. J. Dissociation quotients for citric acid in aqueous sodium chloride media to 150° C. *J. Solution Chem.* **1997**, *26*, 63−84.
(35) Gilli, P.; Bertolasi, V.; Ferretti, V.; Gilli, G. Covalent nature of the strong homonuclear hydrogen bond. Study of the O−H- - -O system by crystal structure correlation methods. *J. Am. Chem. Soc.* **1994**, *116*, 909−915.
(36) Dunitz, J. D. The cost of bound water in crystals and biomolecules. *Science* **1994**, *264*, 670.
(37) Usher, K. C.; Remington, J.; Martin, D. P.; Drueckhammer, D. G. A short hydrogen bond provides only moderate stabilization of an enzyme−inhibitor complex of citrate synthase. *Biochemistry* **1994**, *33*, 7753−7759.
(38) Schwartz, B.; Drueckhammer. D. G. A simple method for determining the relative strengths of normal and low-barrier hydrogen bonds in solution: implications for enzyme catalysis. *J. Am. Chem. Soc.* **1995**, *117*, 11902−11905.
(39) Thorson, J. S.; Chapman, E.; Murphy, E. C.; Schultz, P. G.; Judice, J. K. Linear free energy analysis of hydrogen bonding in proteins. *J. Am. Chem. Soc.* **1995**, *117*, 1157−1158.
(40) Kato, Y.; Toledo, L. M.; Rebek, J., Jr. Energetics of a low barrier hydrogen bond in nonpolar solvents. *J. Am. Chem. Soc.* **1996**, *118*, 8575−8579.
(41) Shan, S.; Loh, S.; Herschlag, D. The energetics of hydrogen bonds in model systems. Implication for enzymatic catalysis. *Science* **1996**, *272*, 97−101.

turbation of the pK_a values of the involved groups. Such long-range crystal packing[42−44] or crystal mother liquor ionic strength[44] effects have been found to modulate pK_a values of protein residues, including that of His57 at the active site α-lytic protease[42] or of bound ligands.[43] Because the active site and ligand binding site of crystal form *b* of trypsin is the furthest removed (>10 Å) from the closest protein neighbor, this form is believed to reflect most closely the solution state.

Proton Locations in Hydrogen Bond Arrays. One intriguing aspect of the multicentered short hydrogen bond arrays in Figures 1 and 6 is the location of the involved protons. Although these proton positions have not been determined experimentally, there are many constraints and restraints on their locations. For example the benzimidazole or indole protons are unambiguously determined by the position of the bonded heavy atoms. Likewise, the positions of some of the other protons (HN_{Ser195}, HN_{Gly193}, and $H\epsilon2_{His57}$) involved in the accompanying ordinary hydrogen bonds are unambiguous. The further provisos that hydrogen bonds should be as linear as possible and that protons should avoid one another result in the relatively unique proton arrangements shown (Figures 3 and 4).

Effect of Hydrogen Bond Geometry on Short Hydrogen Bond Strengths. The modeled proton positions define a spectrum of hydrogen bond linearities that depend on various constraints, such as whether the proton is donated to one, two, or three acceptors. The two-centered short hydrogen bonds are quite linear (Tables 1i and 3a,b, Figure 7a). Within the multicentered short hydrogen bond arrays a constituent hydrogen bond involving one donor and one acceptor can also be essentially linear, such as for proton 1 (Figure 3a) ($O\gamma_{Ser195}-H\gamma_{Ser195}-O_{phenol} = 171 \pm 5°$) or for proton 5 (Figure 7b) (O1′−HO1′−O2 = 172 ± 8°). However, other hydrogen bonds involving protons donated to only one acceptor are constrained to be nonlinear in these arrays, such as for proton 2 ($O_{oxy}-H1_{oxy}-O\gamma_{Ser195} = 141 \pm 2°$) or for proton 3 ($O_{oxy}-H2_{oxy}-O_{phenol} = 141 \pm 2°$) (Figure 3a and Table 1a). When a proton is donated to two or to three acceptors (such as proton 4, Figure 3a), some of the hydrogen bonds involving the donated proton may be near-linear ($N_{benzimidazole}-H_{benzimidazole}-O\gamma_{Ser195} = 156 \pm 2°$), while others may be distinctly nonlinear ($N_{benzimidazole}-H_{benzimidazole}-O_{phenol} = 116 \pm 1°$ and $N_{benzimidazole}-H_{benzimidazole}-O_{oxy} = 121 \pm 1°$). In the case of the citrate dimer, all hydrogen bonds involving protons 6 and 7 (Figure 7b), each donated to three acceptors, have poor, nonlinear geometries (Figure 7b,c and Table 3c). For nonlinear hydrogen bonds donated to two or more acceptor atoms, the proton often lies above an edge formed by the two acceptor atoms or above a face formed by the three acceptor atoms. The calculated hydrogen bond strength of nonlinear hydrogen bonds is reduced to 90%, 60%, and <10%, for O−H−O angles of 165, 149, and 110°, respectively.[35] Because many hydrogen bond angles in the multicentered arrays are <150° (Tables 1 and 3), the strength of the associated hydrogen bonds is expected to be significantly attenuated, despite their shortness.

Conclusions

Diverse examples of short hydrogen bonds both at and away from the active site discovered recently in several protein crystal systems under investigation suggest that short hydrogen bonds may be more common than recognized. Engineered scaffolds that form short hydrogen bonds at serine protease active sites embody a unique and powerful suite of potent, selective small-molecule inhibitors.[14,21−24] The pK_a values of the involved short hydrogen bonding groups are significantly different from one another. Although some of the hydrogen bonding interactions within the multicentered short hydrogen bond arrays have nonoptimal angular components, in each of the arrays there is at least one near-linear short hydrogen bond and often two of them (Table 1). Nevertheless, the contribution to inhibitor binding over that of ordinary hydrogen bonds is modest. Because of the disparate pK_a values, polar environment, and modest strength over ordinary hydrogen bonds the multicentered short hydrogen bond arrays are considered as electrostatic, noncovalent interactions, similar to ordinary hydrogen bonds.

Experimental Section

Synthesis of Inhibitors. Syntheses and characterizations of a variety of 2-(2-phenol)indoles and 2-(2-phenol)benzimidazoles, including CRA-7806, CRA-8696, and CRA-10991 have been described.[22−24]

Crystallization and Preparation of Trypsin, Thrombin, and Urokinase Complexes. Trypsin-benzamidine was crystallized by vapor diffusion as described[14,29−31,45] in space group $P2_12_12_1$, small cell (form *a*, *a* = 54.8 Å, *b* = 58.7 Å, *c* = 67.6 Å), $P2_12_12_1$, large cell (form *b*, *a* = 63.7 Å, *b* = 63.5 Å, *c* = 69.3 Å), or $P3_121$ (form *c*, *a* = 54.9 Å, *c* = 109.2 Å). Benzamidine was removed by extensive soaking in inhibitor-free synthetic mother liquor. Trypsin-inhibitor crystals were prepared from apo-trypsin crystals by soaking for 7−10 days in solutions of 2.00 M $MgSO_4 \cdot 7\ H_2O$ (pH 3.5−9.0) or 85% saturated sodium citrate (pH >9.0). Solutions were buffered with 100 mM potassium acetate, MES, Tris, or 50 mM CAPS, contained 2.5% DMSO and 1.0 mM $CaCl_2$, and were saturated in inhibitor. The soaking solutions for form *b* ($P2_12_12_1$, large cell) also contained 10 mg/mL of trypsin.

Thrombin was purchased from Haematologic Technologies, Inc., and acetyl-hirudin from Bachem. Thrombin-acetyl-hirudin was prepared as described.[46] Thrombin (1.0 mg/mL in 50 mM HEPES, 50% glycerol, pH 7.0) was incubated with 1.0 mM acetyl-hirudin in the presence or absence of 1.0 mM inhibitor for 1 h. at 4° C. Glycerol was removed during concentration of the enzyme to ∼10 mg/mL with a Centricon 10 (Amicon). The concentration was quantitated with the Biorad protein assay kit using bovine serum albumin as the standard. Crystals of the CRA-7806 and CRA-8696 complexes of thrombin-acetyl-hirudin in space group *C*2 (*a* = 71.2 Å, *b* = 71.8 Å, *c* = 72.7 Å, β = 100.7°) were grown at pH 7.3 or 7.8 in hanging drops by vapor diffusion after streak seeding as described.[45] The drops were made from 3 μL complex and 3 μL reservoir solution (0.10 M HEPES, 0.30 M NaCl, 22% (by volume), PEG 5K monomethyl ether, pH 7.5 or 8.2). To change the pH of crystals, they were transferred to and soaked in a mother liquor solution composed of 3.0 mg/mL of thrombin, 0.71 mM acetyl-hirudin, 24% PEG 5K monomethyl ether, 0.42 M NaCl, buffered with 25−100 mM MES, Tris, or CAPS. Inhibitor concentrations were close to saturation. The pH of the soaking solutions used was monitored before, during, and after data collection.

(42) Smith, S. O.; Farr-Jones, S.; Griffin, R. G.; Bachovchin, W. W. Crystal versus solution structures of enzymes: NMR spectroscopy of a crystalline serine protease. *Science* **1989**, *244*, 961−964.
(43) Katz, B. A.; Cass, R. T. In crystals of complexes of streptavidin with peptide ligands containing the HPQ sequence the pK_a of the peptide histidine is less than 3.0. *J. Biol. Chem.* **1997**, *272*, 13220−13228.
(44) Katz, B. A. Binding of biotin to streptavidin stabilizes intersubunit salt bridges between Asp61 and His87 at low pH. *J. Mol. Biol.* **1997**, *274*, 776−800.
(45) Katz, B. A.; Mackman, R.; Luong, C.; Radika, K.; Martelli, A.; Sprengeler, P. A.; Wang, J.; Chan, H.; Wong, L. Structural Basis for Selectivity of a Small Molecule, S1-binding, Sub-micromolar Inhibitor of Urokinase Type Plasminogen Activator. *Chem. Biol.* **2000**, *7*, 299−312.
(46) Skrzpczak-Jankun, E.; Carperos, V. E.; Ravichandran, K. G.; Tulinsky, A.; Westbrook, M.; Maraganore, J. M. Structure of the hirugen and hirulog 1 complexes of alpha-thrombin. *J. Mol. Biol.* **1991**, *221*, 1379−1393.

Low molecular weight (LMW), mutagenically deglycosylated (Ala145) human urokinase[45] was concentrated to about 10 mg/mL in 50 mM HEPES, 5.0 mM NaCl, pH 7.4. Urokinase-inhibitor crystals were grown in space group $C2$ ($a = 82.1$ Å, $b = 49.8$ Å, $c = 66.7$ Å, $\beta = 113.3°$) by vapor diffusion in hanging drops as described.[45] Drops were composed of equal volumes of protein-inhibitor solution (0.28 mM LMW uPA/A145, 5.0 mM inhibitor) and well solution (20% 2-propanol, 20% PEG 4K, 100 mM sodium citrate, pH 6.5).

Crystallographic Data Collection. X-ray diffraction data sets were collected with an R-AXIS IV image plate (Molecular Structure Corp.) and processed as described.[14,45] X-rays (Cu Kα) were generated with a Rigaku RU200 generator operating at 50 kV and 100 mA. The completeness of each data set, before rejection of weak reflections ($F_o/\sigma < 0.5-2.0$) was >98.5% for trypsin crystals and >96% for the lower symmetry urokinase and thrombin crystals.

Determination and Refinement of Crystal Structures. Initial structures were determined and refined as described[14,45] with X-PLOR[47,48] and with difference Fourier analysis. All heavy and hydrogen atoms were included in refinements. Force field energy terms for atoms involved in short hydrogen bonds were removed in the refinements. The structures of uPA-CRA-7806 (1gjb), of uPA-CRA-8696 (1gjc), and of thrombin-CRA-7806 (1gj5) have been deposited into the RCSB database. The structures of all the complexes represented in Figure 2 are available upon request.

Superposition and Comparison of Structures. For each pairwise comparison, structures were initially superimposed based on sets of S1 and S1′ and active site main-chain atoms.[14] After inspection of the initially superimposed structures, residues were then excluded from the superpositions in order to maximize similarities and differences in the inhibitor-binding sites. Averages and standard deviations in hydrogen bond lengths and other parameters were calculated for multiply determined structures.

Enzyme Assays. Bovine trypsin, high-molecular-weight urokinase (HMW uPA), and thrombin were from Worthington Biochemicals, American Diagnostica, and Calbiochem, respectively. The substrate for trypsin and thrombin, tosyl-Gly-Pro-Lys-pNA, was from Sigma and that for uPA, Bz-Ala-Gly-Arg-pNA, from Centerchem, Inc. Inhibition constants were determined as described.[14,45] All inhibitor potency and enzyme activity measurements were performed at room temperature in 50 mM MES, Tris, or CAPS (pH 5−6, 7−9, and 10, respectively), 150 mM NaCl, 100 μM ethylenediaminetetraacetic acid (EDTA), and 2.5% dimethyl sulfoxide (DMSO) in 96-well microtiter plates (Falcon) using a Thermomax-Kinetic plate reader (Molecular Devices). The trypsin concentration was 50 nM (pH 5.0), 10 nM (pH 6.0−8.0), or 20 nM (pH 9.0−10.0). Thrombin was 150 nM (pH 5.0), 20 nM (pH 6.0), or 5.0 nM (pH 7.0−10.0), while uPA was 200 nM (pH 5.0), 100 nM (pH 6.0), or 30 nM (pH 7.0−10.0).

The enzyme kinetic parameters (K_m and k_{cat}) were determined at each pH by fitting the velocity data to the Michaelis−Menten equation. Inhibition determinations were performed at a substrate concentration close to the K_m value. Apparent inhibition constants, K_i' values, were calculated from the velocity data collected at various inhibitor concentrations with the software package BatchKi (developed and provided by Dr. Petr Kuzmic, BioKin Ltd., Pullman, WA), using methodology similar to that described for tight-binding inhibitors.[49] K_i' values were converted to K_i values by the formula $K_i = K_i'/(1 + S/K_m)$.

Acknowledgment. We thank Arnold Martelli for synthesis of CRA-8696.

JA020082M

(47) Brünger, A. T. *X-PLOR Manual, Version 3.1: A System for X-ray Crystallography and NMR*; Yale University: New Haven, CT, 1990; pp 187−206.
(48) Brünger, A. T. The free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **1992**, *355*, 472−474.
(49) Kuzmic, P.; Sideris, S.; Cregar, L. M.; Elrod, K. C.; Rice, K. D.; Janc, J. W. High-throughput screening of enzyme inhibitors: Automatic determination of tight-binding inhibition constants. *Anal. Biochem.* **2000**, *281*, 62−67.