Acta Cryst. (1995). D**51**, 597–604

Structure-Based Analysis of Inhibitor Binding to Ht-d

BY ISTVAN BOTOS AND LEONARDO SCAPOZZA

Biographics Laboratory, Department of Biochemistry and Biophysics, Texas A & M University, College Station, TX 77843-2128, USA

JOHN D. SHANNON AND JAY W. FOX

Biomolecular Research Facility, University of Virginia Health Sciences Center, Charlottesville, VA 22908, USA

AND EDGAR F. MEYER*

Biographics Laboratory, Department of Biochemistry and Biophysics, Texas A & M University, College Station, TX 77843-2128, USA

(*Received* 20 *July* 1994; *accepted* 8 *February* 1995)

Abstract

A theoretical study was performed on the structure of both the native and inhibited metalloproteinase Ht-d (E.C. 3.4.24.42) solved at 2.0 Å resolution. The energy maps calculated by program *GRID* clearly showed the extended binding site of Ht-d and allowed localization and characterization of the pockets *S*1–*S*3 and *S*1′–*S*3′. The *GRID* energy contour maps point out the particular shape of the *S*1′ pocket in agreement with experimental density maps and inhibited Ht-d structures. Based on the high degree of sequence homology of the Ht-d active site to that of mammalian metalloproteinases, the characterization of active site pockets was extended to neutrophil collagenase, fibroblast collagenase, stromelysin 1 and 2. Thirty residues of the Ht-d propeptide were modeled and optimized with reference to the Ht-d structure, giving insight to the mechanism of natural inhibition in metalloproteinase proenzymes. Kinetic measurements of Ht-d inhibition by a series of synthetic peptides show, in agreement with our Ht-d propeptide model, the crucial role of cysteine and adjacent residues in the specificity of Ht-d propeptide. This study suggests the structural link between Ht-d and mammalian metalloproteinases, contributing to the understanding of the mechanism of natural and synthetic inhibitor binding to metalloproteinases. Therefore, Ht-d is a good model system for the design of novel inhibitors against these enzymes with enhanced potency and specificity.

Introduction

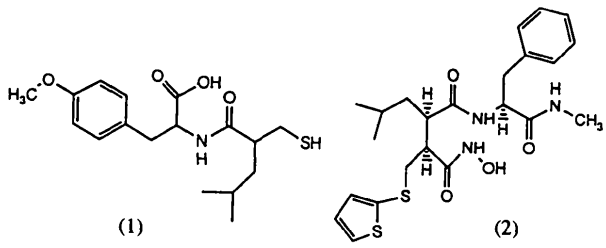
Matrix metalloproteinases (MMP's) are an important group of zinc enzymes which are responsible for degradation of the extracellular matrix components such as collagen and proteoglycan during normal connective tissue remodeling such as embryonic development and wound healing (Woessner, 1991; Birkedal-Hansen *et al.*, 1993). Type IV collagenases cleave *P*1—*P*1′ bonds, where *P*1 is typically Gly or Gln, and *P*1′ is a bulky hydrophobic residue (Phe, Leu, Ile) (Schechter & Berger, 1967; Seltzer *et al.*, 1989, 1990; Niedzwiecki, Teahan, Harrison & Stein, 1992). A series of pathological processes, including human breast, colon, hepatocellular, prostate, lung and thyroid cancers are correlated with abnormal MMP expression (Okada *et al.*, 1990). The direct role of MMP activity in tumor-cell invasion was demonstrated by endogenous inhibitors of MMP's (TIMP's) or synthetic inhibitors (DeClerck *et al.*, 1992; Davies, Brown, East, Crimmin & Balkwill, 1993). Considering the important physiological and pathological roles of the MMP's, efforts have been made to determine the inhibited structures of these enzymes, and recently the high-resolution crystal structures of several members of the MMP family were reported (Bode *et al.*, 1994; Borkakoti *et al.*, 1994; Gooley *et al.*, 1994; Stams *et al.*, 1994). The homology for the consensus sequence of the active-site region (Collier *et al.*, 1988; Jongeneel, Bouvier & Bairoch, 1989; Gomis-Rueth, Stöcker, Huber, Zwilling & Bode, 1993), found by previous studies were confirmed by crystallographic results.

Atrolysin C form d (Ht-d) is a haemorrhagic metalloproteinase from the venom of the Western diamondback rattlesnake *Crotalus atrox* (Bjarnason & Fox, 1994). Atrolysin C belongs to class P-I of venom proteinases (Hite, Jia, Bjarnason & Fox, 1994). Both Ht-d and the MMP's are involved in the hydrolysis of extracellular matrix components such as interstitial and basement-membrane collagens, fibronectin and laminin (Baramova, Shannon, Bjarnason & Fox, 1989).

The structure of the matrix metalloproteinase, Ht-d (E.C. 3.4.24.42), has been determined to 1.9 Å resolution (Zhang *et al.*, 1994). This proteolytic enzyme (202 amino acids, one Zn, one Ca) from rattlesnake venom has an active site which has a highly similar architecture to the

* To whom correspondence should be addressed.

mammalian collagenase from human neutrophils. Two synthetic and one natural inhibitors have been studied crystallographically. The natural inhibitor, pENW (pyroGlu-Asn-Trp), is present in the snake venom in m*M* concentration and has a K_i in the micromolar range; SCH (SCH 47890) is a synthetic inhibitor (1) that is more potent than pENW. BB-94 (2) has a nanomolar K_i and is a potent antitumor drug, demonstrating the pharmacological relevance of these studies, which point to the dominant role of van der Waals interactions in inhibitor binding (Botos *et al.*, 1995), together with Zn-atom ligation. The enzyme undergoes minimal displacement upon ligand binding, being a classical 'lock and key' enzyme (Fischer, 1894) and an ideal target for structure-based drug design.



(1) (2)

Recently, several efforts have been made to understand the mechanism of natural inhibition of metalloproteinase zymogens (Stetler-Stevenson, Krutzsch, Wacher, Margulies & Liotta, 1989; Okada *et al.*, 1990). The venom metalloproteinases are synthesized with highly conserved pro-sequences (Hite *et al.*, 1994). Approximately 23 residues into the amino terminal end of the start of the Ht-d sequence there is a conserved region containing a cysteinyl residue which may serve a function analogous to that of the nine-residue consensus sequence involved in the mechanism of proenzyme latency in the MMP's (Vallee & Auld, 1990; Van Wart & Birkedal-Hansen, 1990) variously referred to as the 'Velcro patch' or the 'cysteine switch'. The toxin sequence of this region, PKMCGVTQ is similar to the MMP consensus sequence KPRCGVPD. In the MMP's it is postulated that the cysteinyl residue binds to the active-site zinc, preventing substrate or water access to the active site and inhibiting proteolytic events. A similar mechanism is suggested for venom metalloproteinases (Hite *et al.*, 1994). Synthetic peptides which present this sequence from the putative latency site are shown to be weakly inhibitory to the MMP's.

Experimental

Peptide synthesis and kinetic measurements

Peptides (from Table 1) were synthesized using Fmoc synthesis protocol on a Biosearch 9600 automated peptide synthesizer. Synthetic peptides were purified by reverse-phase (C18) high-pressure liquid chromatography and then analyzed by MALDI-TOF MS on a Finnigan Mat Lasermat mass spectrometer. Immediately prior to kinetic analysis, peptides containing cysteine were assayed for free sulfhydryl groups using Ellman's reagent (Ellman, 1959). The inhibition of Ht-d hydrolysis of the fluorogenic substrate Abz-Ala-GlyLeu-Ala-Nba by the synthetic peptides was determined as previously described (Fox, Campbell, Beggerly & Bjarnason, 1986) using a Perkin–Elmer MPF-44A fluorometer.

Table 1. *Inhibition of Ht-d proteolytic activity by cysteine-switch peptide homologs*

Peptide		IC_{50} (m*M*)
I	Ac-EKEDEAPKMCGVTQNWES-NH_2	5.0 ± 0.6
II	Ac-EAPKMCGVTQ-NH_2	4.6 ± 0.3
III	Ac-KMCGVTQ-NH_2	1.7 ± 0.3
IV	Ac-EAPKMCGATQ-NH_2	3.2 ± 0.5
V	Ac-EAPKMAGVTQ-NH_2	9500
VI	Ac-MCPGMSEALELYSK-NH_2	147
VII	Ac-EGRNPGFYVEANPMPTFK-NH_2	NI*
VIII	Cysteine	85 ± 11

* NI = no inhibition.

GRID input preparation

Atomic coordinates for Ht-d were from the X-ray structure of Ht-d at 2.0 Å resolution complexed with pENW (Zhang *et al.*, 1994). The inhibitor pENW and all structurally non-significant crystallographic water molecules were removed. The resulting coordinates were submitted to the subroutine *GRIN* from the *GRID* program (Goodford, 1985). The standard *GRUB*-parameter set was used for the evaluation of the Lennard–Jones, electrostatic, hydrogen-bond functions and empirical energy. The resulting output of *GRIN* was used as input for the main routine of *GRID*. A regular array of grid points separated by 0.5 Å was established throughout and around the whole protein in order to enclose the entire binding site as well as the relevant surface area for the modeling of the 30-residue-long propeptide (NH_2-EAPKMCGVTQNWESYEPIKKASDLNLDPDQ). The potential energy was calculated for a neutral H_2O probe and a neutral methyl probe with van de Waals radii of 1.7 and 1.95 Å, respectively. Other probe parameters were from standard probe files distributed with the *GRID* program.

Modeling of the propeptide in proHt-d

With the overall aim to model the propeptide we calculated the energy contour map for other probes like amide, amine, carboxyl O atoms, hydroxyl O atoms and CH aromatic using standard parameters from *GRID*. The 20-amino-acid sequence (I11–I30) connecting the N terminus of the protein with the ten amino acids (I1–I10) in the extended binding site was modeled in an extended conformation according to the favorable energy contour map of the different probes calculated by *GRID*.

The obtained proHt-d model (Ht-d + 30–amino-acid propeptide) was optimized through 500 cycles of

conjugate-gradient minimization with *X-PLOR* (Brünger, Kuriyan & Karplus, 1987). The *X-PLOR* parameter set used was that of Engh & Huber (1991) derived from the Cambridge Structural Database. For the minimization the following energy terms were included: covalent bond, angle, dihedral angle, improper dihedral angle, intramolecular van der Waals, intramolecular electrostatic, explicit hydrogen bond. Besides the energy terms, non-bonded interactions were considered and crystallographic water included, not obstructing the position of the propeptide. All calculations were run on an SGI Extreme 2. Graphic display and analysis of the three-dimensional *GRID* energy contour maps were achieved using the program *PRONTO* (Swanson, S. & Laczkowski, A. personal communication) on an ESV-5 workstation.

Results and discussion

The crystal structure of Ht-d was determined to a resolution of 1.9–2.0 Å with and without bound inhibitors (Zhang *et al.*, 1994). Extensive refinement led to structures with an *R* factor under 18%. The backbone consists of five α-helices and five strands which form a predominantly parallel β-sheet. There is one Zn^{II} ion and one Ca^{II} ion in the enzyme. The Zn^{II} atom is catalytic and is tetrahedrally coordinated with three histidines (His142, 146 and 152) and a water molecule. In the inhibited enzyme, the Zn^{II} atom is tetra- or pentacoordinate, being ligated also to the carboxyl O atom(s) of the inhibitor. The Ca^{II} ion coordination is octahedral, with coordination derived from OD1 of Asp93 and Asn200, OD2 of Asp93, OE1 of Glu9, O197 and water, the average Ca^{II}—O distance being 2.4 Å.

Extended binding site

The unusually deep (~10 Å) primary specificity pocket (*S1′*) (Schechter & Berger, 1967) can accommodate fairly large hydrophobic moieties such as the indole ring of pENW (pyroGlu-Asn-Trp), the *O*-methylphenol group of SCH and the ethylene thiophene group of BB-94. In this respect the *S1′* pocket of Ht-d more closely resembles human neutrophil collagenase, which is larger than the same pocket in fibroblast collagenase (Stams *et al.*, 1994). Sequence comparison (Fig. 1) lets us speculate about the size of the *S1′* pocket in different MMP's. The differences in size of the *S1′* pocket in Ht-d and human MMP's would suggest alternative physiological substrates for Ht-d. The *S1′* pocket is the largest (Fig. 2), defined by residues: 138–139 (top back wall), 142 (top left wall), 168 (top front wall), 170–171 (front middle and bottom), 163, 167 and 176 (bottom) (Table 2). Val138 from Ht-d is substituted for Arg in fibroblast collagenase and Leu in neutrophil collagenase, stromelysin 1 and 2 (Fig. 1). Comparing the above structures, fibroblast collagenase has the bulky Arg residue partially blocking access to the bottom of the *S1′* pocket. This would explain the fact that fibroblast collagenase prefers small hydrophobic chains at *P1′* while Ht-d can have Trp, or *O*-methyl tyrosine at this position. Arg167 from Ht-d is replaced (in mammalian MMP's) by a highly conserved, less bulky Tyr residue, which gives more 'opening' in the bottom wall of the pocket. The *S1′* pocket in Ht-d shows a clear bottom formed by the Leu163 (center), Tyr176 (right) and Arg167 (left) side chains (Fig. 2). For fibroblast collagenase as well as for neutrophil collagenase the change from Leu to Gly and from Tyr to Val, respectively, makes the bottom of the pocket more open. In the native fibroblast collagenase this structural feature forms a water tunnel with the surface of the protein, facilitating binding and release of substrate (Meyer, 1992). In stromelysins changing Leu to Glu and Tyr to Leu we have a compensating exchange, which should not affect the space-filling property of the bulky side chains. On the one hand the difference in the residue charges could be very important for the specificity of the pocket in relation to the design of

Table 2. *Residues defining individual subsites in the extended binding site of Ht-d*

Subsite	*S3*	*S2*	*S1*	Zn	*S1′*	*S2′*	*S3′*
Residues	Leu113	Leu113	Leu110		His142	Thr107	Leu108
	Glu151	His146	His152		Val138		Gly169
					Thr139		
					Leu163		
					Arg167		
					Pro168		
					Leu170		
					Thr171		
					Tyr176		

	106	113	137	152	163	176
Ht-d	ETLGLAPL		GVTMAHELGHNLGMEH		LCI MRPGLTKGRSY	
	162	169	196	211	216	229
FC	GNLAHAFQ		HRVAAHELGHSLGLSH		GALMYPSYTFSGDV	
NC	GI LAHAFQ		FL VAAHEFGHSLGLAH		GALMYPNYAFRETS	
S-1	NVLAHAYA		FL VAAHEI GHSLGLFH		EALMYPL YHSLTDL	
S-2	HSLAHAYP		FL VAAHELGHSLGLFH		EALMYPL YNSFTEL	

Ht-d, Atrolysin C form d; FC, human fibroblast collagenase; NC, human neutrophil collagenase; S-1, stromelysin 1; S-2, stromelysin 2

Fig. 1. Homology alignment of extended binding-site regions in MMP's.

specific inhibitors that are sufficiently long to reach this region. On the other hand the specificity of the pocket for polypeptidic substrates is dictated by the upper part of the $S1'$ pocket, that can accommodate Ile or Leu side chains (Fields, Van Wart & Birkedal-Hansen, 1987), and by the smaller pockets $S2'$, $S3'$ and $S1$–$S3$ which make up the extended binding site. Thr139 from Ht-d is replaced by a highly conserved Val residue in the structures compared. This substitution replaces a hydrophylic residue with a hydrophobic one, disfavoring hydrogen bonding. Pro168 is highly conserved in both Ht-d and the selected mammalian collagenases, forming the cornerstone of the 'lower wall' of the extended binding site. The $S1'$ pocket is lined by Leu170 and Thr171 in Ht-d, replaced by a well conserved Tyr in position 170 and various residues in 171. The Tyr residue is slightly bulkier than Leu, reducing the size of the $S1'$ pocket in the middle-front region in the mammalian enzymes, but at the same time offering the possibility of additional hydrogen-bond formation. His and Asn residues are present at position 171 in stromelysin 1 and 2, respectively; they are bulkier residues than Thr in Ht-d, reducing the size of the $S1'$ pocket. The bottom of the pocket is lined by residue Tyr176, replaced by less bulky Val in fibroblast collagenase, Ser in neutrophil collagenase and Leu in stromelysins. None of the inhibitors studied thus far form a hydrogen bond to Tyr176, and are important only in van der Waals interactions. Looking at the overall $S1'$ pocket-lining sequence, it is apparent that the neutrophil collagenase $S1'$ pocket and that of stromelysin is comparable in size with the Ht-d pocket but not with the fibroblast collagenase $S1'$ which is partially obstructed by Arg214.

The $S2'$ site is lined by residue Thr107 in Ht-d, which defines a wall of the 'pocket'. In fibroblast collagenase this residue is replaced by Asn, permitting opportunity for the formation of more hydrogen bonds with the ligand. In neutrophil collagenase and stromelysins residue 107 is replaced with Ile and Val, respectively. Both residues are hydrophobic and suggest a preference for hydrophobic groups in $P2'$ for these enzymes. However the region containing residue 107 has a higher flexibility (temperature factor higher than 15 $Å^2$) suggesting a wider range of $P2'$ residues which can be accommodated in this region. It was shown (Gooley *et al.*, 1994) that in stromelysin 2 Arg can be accommodated in this region, and the Arg side chain up to the Cα atom interacts hydrophobically with the $S2'$ Val residue, then extending into the solution. This illustrates the dual character of the $S2'$ wall, accommodating short hydrophobic residues or extended hydrophilic residues.

The $S3'$ pocket is defined by residues Leu108 and Gly169 in Ht-d. The Leu is well conserved in mammalian collagenases, the Gly residue is replaced with hydrophylic residues in fibroblast and neutrophil collagenase and by Leu in stromelysins. This suggests a preference for hydrophylic residues in $P3'$ in fibroblast and neutrophil collagenase and preference for hydrophobic groups in Ht-d and stromelysins.

The $S1$–$S3$ pockets can be defined by examining the modeled fit of peptide III (Table 1) to the extended binding site (model not shown). The $P1$ group points towards the protein surface and is lined by Leu110 and His152. This $S1$ site is relatively small, being optimized for non-bulky residues as Gly but accepting Gln in stromelysin (Niedzwiecki *et al.*, 1992). $P2$ points towards the catalytic His146, which defines one side of the $S2$ pocket in Ht-d. The other side of the pocket is lined by residue Leu113. This pocket prefers hydrophobic residues (*e.g.* Val from the conserved Cys-Gly-Val repeat in the Ht-d propeptide). Residue Leu113 partly defines the $S3$ pocket, which is bordered by Glu151, forming a 'gate' through which the Ht-d propeptide passes.

The $S1$ site is similarly defined in fibroblast collagenase, neutrophil collagenase and stromelysins with Leu110 changed to His. $S2$ is defined by the well conserved Zn-binding His146 and Leu113 changed for Glu in fibroblast and neutrophil collagenase, Ala in stromelysin 1 and Pro in stromelysin 2. This suggests preference for hydrophilic residues in $P2$ for fibroblast and neutrophil collagenase and bulkier residues for stromelysins.

The $S3$ site is defined partly by Leu113 and Glu151 in Ht-d, the latter replaced by Ser in fibroblast collagenase, Ala in neutrophil collagenase and Phe in stromelysins. In this manner the 'gate' is better defined in fibroblast collagenase, given the nature of Gln–Ser interactions, while in neutrophil collagenase the $S3$ site is more open, being lined by a Gln or Ala residues. In stromelysins this 'gate' is higher, defined by the bulkier Phe residue paired with Ala and Pro in stromelysin 1 and 2, respectively. These sites have to be able to accommodate their propeptide strand optimally in the zymogen form.

None of the $S1$–$S3$ pockets is as deep and well defined as $S1'$, and thus plays a secondary role in small-molecule binding, but is more important in peptide binding, when the $S1'$ primary specificity pocket is not filled with bulky residues. When small molecules bind, and $S1'$ is filled to a greater extent by a hydrophobic group, van der Waals interactions become preponderant, being responsible for the tighter binding (Botos *et al.*, 1995).

The regions with greatest mobility are apparent upon examination of the temperature factors per residue (Fig. 3). Excluding the protein termini, the loop regions 102–105 and 153–157 near the extended binding site are very flexible, having an average temperature factor greater than 25 $Å^2$. However, these loops are not in close contact with the inhibitors. A less flexible region is defined by the 106–109 and 158–172 loops, which form the upper and lower walls of the $S1'$ pocket. The extended binding site helix (132–147) and the residues

defining the 'bottom' of the $S1'$ specificity pocket (163, 167, 176) are more rigid, suggesting a higher specificity for part of the inhibitor structure extending deep into the pocket. There is some mobility of the residues lining the $S2'$ region, suggesting a variety of $P2'$ residues which can be accommodated in this region. The 63–86 helix is located between two flexible loops and is itself fairly flexible. These mobile regions are topologically similar to those in fibroblast collagenase (Borkakoti *et al.*, 1994), especially the loops defining the extended binding site.

GRID modeling

As a third method for localizing and visualizing the position of the pocket in Ht-d we employed the *GRID* program which uses non-bonded potential functions (Lennard–Jones, Coulombic, hydrogen bonding) to calculate the interaction enthalpies between a chosen probe and target protein. *GRID* is a theoretical way of showing the real size of a binding pocket. Electron-density maps show the contour of the bound inhibitors but do not give insight to the unfilled regions of hydrophobic pockets. Connolly surfaces are able to depict the contours of binding pockets, but *GRID* maps in addition show preferred sites for certain chemical groups. This additional feature is optimal for the design of compounds with enhanced binding affinity and analogous to known inhibitors. This analysis was performed using a neutral H_2O probe and a neutral methyl probe with van der Waals radii of 1.7 and 1.95 Å, respectively. The 0 kcal mol^{-1} map (Fig. 4*a*) which describes the limits of all favorable interaction zones between the H_2O probe and the protein, shows clearly which residues are involved in building the pocket for possible interactions with substrate. The grid contour map confirms that Gln151 is the major residue defining the $S3$ pocket from the side, and Leu113 outlines the other side of the same pocket. Looking at the $S2$ pocket (Fig. 4*a*) it is clear the His146 would define the bottom and Leu113 the left side of this subsite. The $S1$ pocket is hard to localize and is in part lined by the Leu residue at position 110 and His152 (Fig. 4*a*). This pocket is very shallow and can accommodate only very small residues such as Gly. This is in agreement with previous sequence specificity studies of collagenase with Gly as the major $P1$ residue (Fields *et al.*, 1987). The *GRID* results shown in Fig. 4(*a*) demonstrate that the $S1'$ site is lined by His142, Val138, Thr139, Leu163, Arg167, Pro168, Leu170, Thr171 and Tyr176 in the manner described above.

From the −2 kcal mol^{-1} favorable contour map of the neutral methyl probe shown by Fig. 4(*b*) it appears that the $S1'$ pocket can be divided into three parts. The upper part (above Leu170) is very large and can accommodate bulky groups. Then a middle part close to Leu170 has a narrow constriction and the bottom part is quite large and has a special ring-like shape. This is in agreement with the X-ray electron-density map for the BB-94 inhibitor (Botos *et al.*, 1995) indicating the full occupancy of the pocket by this inhibitor. From the above model and other metalloproteinase models (data not shown) it appears that the natural amino acids are accommodated only in the upper part of the $S1'$ pocket, indicating an unfilled region at the bottom of the subsite, making it even more interesting for drug design. For the design of inhibitors that have good interactions with the $S1'$ site it seems very important to find a linkage group that can be placed in the narrow middle part of the pocket and join the bulky upper substituent with a large group in the bottom region of the pocket (like a dumbbell).

The results shown in Fig. 4(*b*) give an idea that the direction and size that the $P2'$ and $P3'$ substituents of an inhibitor should have to allow favorable interaction with the hydrophobic part of $S2'$ (Thr107) and the $S3'$ hydrophobic residues (Leu108 and Gly169). The results of this short theoretical analysis of the Ht-d subsites suggest that the *GRID* maps at the contour level presented in Figs. 4(*a*) and 4(*b*) give a realistic view of the extended binding site of the Ht-d protein and that *GRID* energy contour maps may be useful for the design of new MMP inhibitors.

Ht-d propeptide

1. *Kinetic measurements.* As can be seen from the IC_{50} values for peptides I–III in Table 1 there appears to be a length-dependent inhibition of Ht-d by these peptides. Successively shorter peptides containing the critical cysteinyl residue have greater inhibitory capacity than longer peptides. This suggests that the sequence at the amino-terminal end of the cysteinyl residue may moderate liganding of the cysteinyl group with the active-site zinc. When a valinyl residue is substituted by an alanyl residue a small decrease in the IC_{50} is observed suggesting that the branched side chain of valine found in the native structure may be important in preventing excessively strong binding of the pro-region consensus sequence to the active site of the proteinase domain. This is confirmed by the Ht-d propeptide model, which shows a greater distortion of the 'gate' formed by residues Leu113 and Glu151, when at the $P3$ position Ala is substituted by the bulkier side chain Val. Substitution of an alanyl residue for the cysteinyl residue (peptide V *versus* IV) produces a drastic loss of inhibitory potential. Peptides VI and VII, which are random sequences, produced moderate and no inhibition, respectively. Peptide VI, which contains a cysteinyl residue, gave an IC_{50} somewhat greater than free cysteine suggesting that the peptide structure may, to a degree, obstruct the interaction of the peptide cysteinyl residue with the active-site zinc. Comparison of the IC_{50}'s for peptides I–IV to that of peptide VI suggests that the peptide structure adjacent to the cysteinyl residue (at the carboxyl end of the Cys) is important in optimizing the interaction of the cysteinyl residue with the active-site Zn atom.

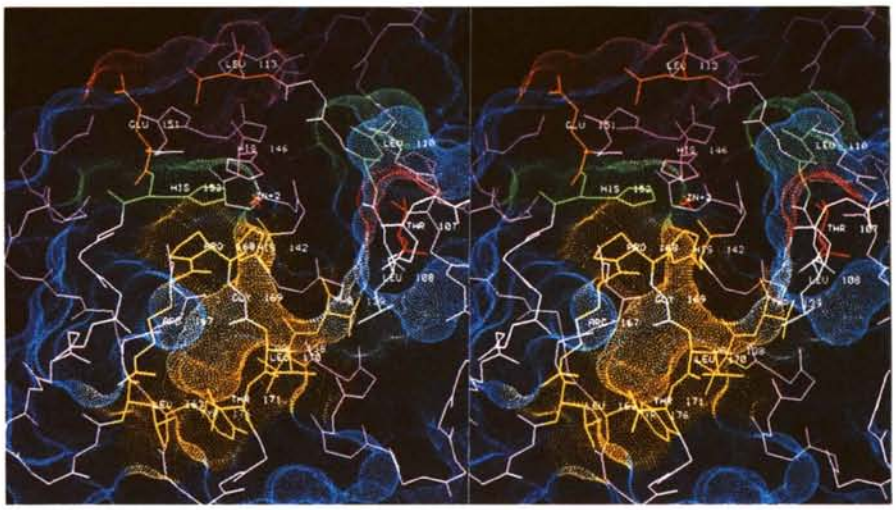


Fig. 2. Connolly surface (with a 1.4 Å probe) of the Ht-d extended binding site, generated with *Insight*II (Biosym Technologies, Inc., San Diego, USA). On the protein surface (blue) the individual subsites are color coded: *S*3 orange, *S*2 magenta, *S*1 green, *S*1′ yellow, *S*2′ red, *S*3′ grey. The backbone atoms are in pink. Residues defining individual binding sites are tabulated (Table 2).

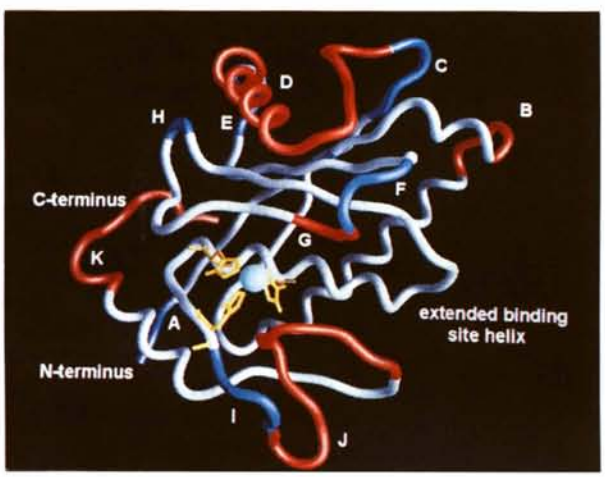
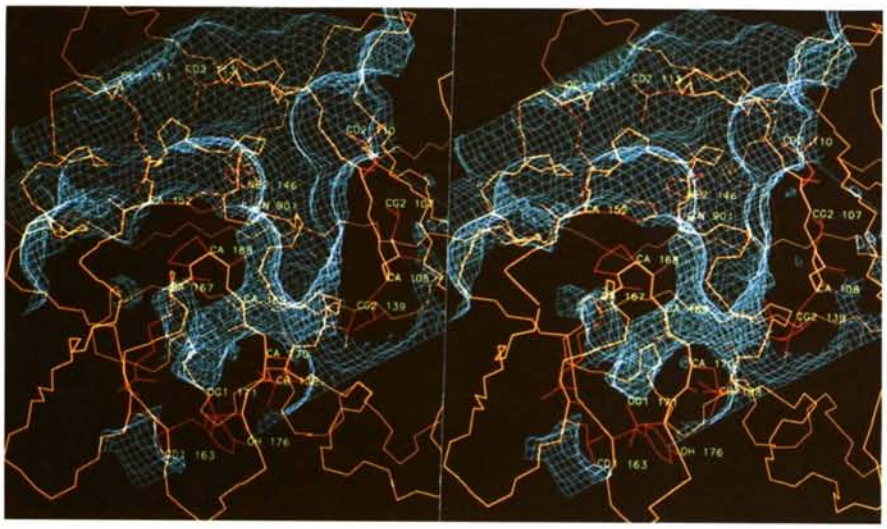


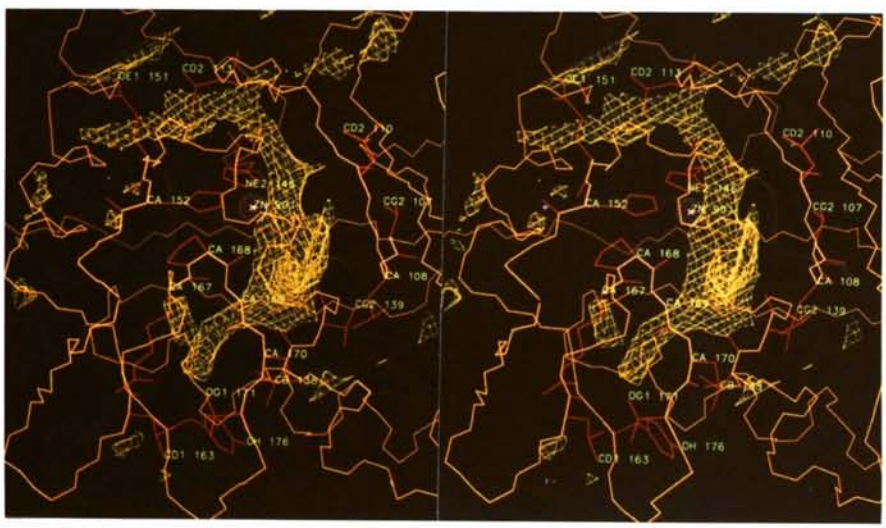
Fig. 3. Ht-d Cα backbone generated with *GRASP* (Nicholls, Sharp & Honig, 1991). The backbone is colored according to the average temperature factors; blue is greater than 25 Å^2 red is between 15 and 25 Å^2 and grey is less than 15 Å^2. The labelled segments are: A 1–3; B 23–26; C 57–63; D 63–86; E 87–91; F 102–105; G 106–109; H 118–120; I 153–157; J 158–172; K 191–202.

2. *Modeling.* 30 residues (I1–I30) from the Ht-d propeptide were modeled using *GRID* maps calculated for the Ht-d surface and optimized with *X-PLOR* (Fig. 5). The propeptide lies in the extended binding site in the opposite orientation than the substrate peptide, then runs parallel to the C terminus, parallel to the loop formed by residues 89–92, takes a turn at the level of residue 57 and then runs antiparallel to the 51–57 strand connecting to the N terminus. Because of the opposite orientation of the propeptide strand the binding pockets defined by the propeptide model are in reverse order compared to the sites defined by the substrate; for clarity we keep the previous nomenclature, the subsites being defined by a substrate polypeptide. CysI6 from the conserved Cys-Gly-Val repeat is responsible for occupying the fourth ligation site of the tetrahedral Zn atom (Fig. 5), in agreement with the kinetic results, projecting the GlyI7 residue towards the surface of the protein, and shifting the whole sequence by one residue towards the Zn atom. GlyI7 is located at the surface of the *S*2 site, ValI8 in the *S*3 site. Because of the 'gate' in the *S*3 site the propeptide is very sensitive to amino-acid substitution in the corresponding position, shown by the kinetic results (Table 1).

As a result of O-atom ligation in the active enzyme, the backbone of a substrate polypeptide would lie closer to the subsites of the extended binding site than is the case of the propeptide strand. This is demonstrated by the model showing that MetI5 can interact only with the opening of the very deep *S*1′ pocket, the terminal methyl group being situated at the level of the Zn atom, leaving the pocket practically empty (*i.e.* hydrated). LysI4 would interact with the *S*2′ wall in a similar manner as an Arg residue described before. At *S*3′ we would have a conserved ProI3 residue in mammalian MMP's and the following AlaI2 and GluI1 residues would point towards the solvent. Because of S–Zn coordination, the propetide would not be able to participate in close peptide side-chain–pocket interactions, adopting a more distant docking position than a small-molecule inhibitor. This suggests an optimal specificity in the protein–propeptide interaction, crucial in the proenzyme activation mechanism. The fit of MetI5 into the *S*1′ pocket and the occupancy of the tetrahedral Zn ligation site by the S atom of CysI6 cause the peptide strand to be tilted 90° from the optimal parallel β-strand defined by 108–112 (Ht-d) and thus weaken the expected hydrogen-bond network between the parallel strands. The fit then would be dominated by Zn ligation and van der Waals interactions.



(*a*)



(*b*)

Fig. 4. Cα backbone trace with the side chain of the amino-acid residues forming the subsites *S3*–*S3′* of the extended binding site of Ht-d. A regular array of grid points separated by 0.5 Å is constructed to enclose the entire binding site. (*a*) In blue is shown the 0 kcal mol^{-1} energy contour maps generated by *GRID* using the H_2O probe. (*b*) In yellow is shown the –2 kcal mol^{-1} contour map indicating favorable interactions between the extended binding site and a neutral methyl probe. Both energy contour maps help with defining the extended binding site and visualizing its shape.

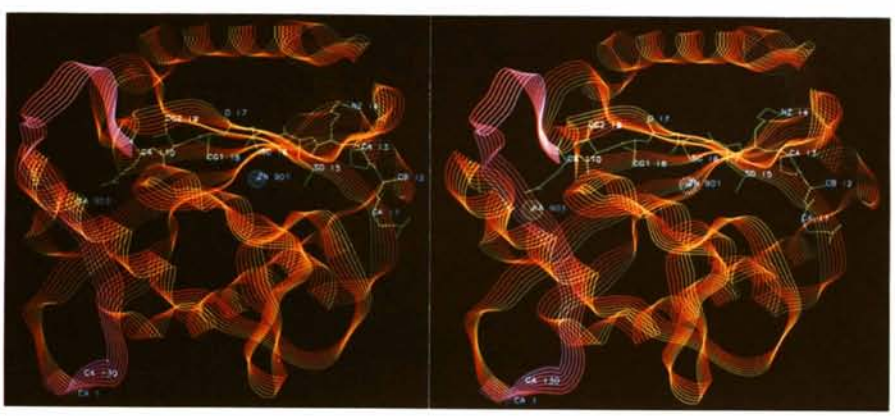


Fig. 5. ProHt-d model generated with *PRONTO* (S. Swanson & A. Laczkowski, personal communication). The Ht-d backbone is illustrated by a yellow–red ribbon, the I1–I10 region of the propeptide is entirely shown and the I11–I30 propeptide region is displayed as a violet–red ribbon. Active-site Zn atom is blue, the Ca atom is green.

Concluding remarks

This theoretical analysis of the extended binding site in Ht-d provides insight for the design of small synthetic (peptidomimetic) compounds as potential drugs. With the same carboxylate ligation, the tightness of inhibitor binding is dictated by the extent of van der Waals interactions between inhibitors and the *S1′* pocket. The primary specificity site (*S1′*) can contain large hydrophobic moieties up to a length (from Cα) of 9–10 Å with a specific shape. The *S2′–S3′* and *S1–S3* pockets can substantially contribute to the binding specificity of polypeptides but would play a secondary role in the binding of small molecules. The extended binding site is generally hydrophobic; local charges at specific subsites can be matched by the introduction of appropriately charged groups on the inhibitor. Kinetic measurements of Ht-d inhibition by a series of synthetic peptides as well as the proHt-d model point out the role of cysteine and adjacent residues in the specificity of Ht-d propeptide, giving insight to the mechanism of natural inhibition in metalloproteinase proenzymes. Because venom and mammalian enzymes show functional and structural active-site similarities, the venom enzyme Ht-d is a good model system for the development of potent drugs, to help control major pathological processes.*

We thank Mr Dachuan Zhang for valuable discussions. The support of the Robert A. Welch Foundation (EM: A-328) and the NIH (JWF: R01GM49042) is acknowledged.

* Atomic coordinates have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1HTD). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: GR0512).

References

Baramova, E. N., Shannon, J. D., Bjarnason, J. B. & Fox, J. W. (1989). *Arch. Biochem. Biophys.* **275**, 63–71.

Birkedal-Hansen, H., Moore, W. G., Bodden, M. K., Windsor, L. J., Birkedal-Hansen, B., Decarlo, A. & Engler, J. A. (1993). *Crit. Rev. Oral Biol. Med.* **4**, 197–250.

Bjarnason, J. B. & Fox, J. W. (1994). *Pharmacol. Ther. Dent.* **62**, 325–372.

Bode, W., Reinemer, P., Huber, R., Kleine, T., Schnierer, S. & Tschesche, H. (1994). *EMBO J.* **13**, 1263–1269.

Borkakoti, N., Winkler, F. K., Williams, D. H., D'arcy, A., Broadhurst, M. J., Brown, P. A., Johnson, W. H. & Murray, E. J. (1994). *Struct. Biol.* **1**, 106–110.

Botos, I., Scarpozza, L., Zhang, D., Liotta, L. A. & Meyer, E. F. (1995). *Proc. Natl Acad. Sci. USA*. Submitted.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.

Collier, I. E., Wilhelm, S. M., Eisen, A. Z., Marmer, B. L., Grant, G. A., Seltzer, J. L., Kronberger, A., He, C., Bauer, A. & Goldberg, G. I. (1988). *J. Biol. Chem.* **15**, 6579–6587.

Davies, B., Brown, P. D., East, N., Crimmin, M. J. & Balkwill, F. R. (1993). *Cancer Res.* **53**, 2087–2091.

DeClerck, Y. A., Perez, N., Shimada, H., Boone, T. C., Langley, K. E., & Taylor, S. M. (1992). *Cancer Res.* **52**, 701–708.

Ellman, G. L. (1959). *Arch. Biochem. Biophys.* **82**, 70.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Fields, G. B., Van Wart, H. E. & Birkedal-Hansen, H. (1987). *J. Biol. Chem.* **262**, 6221–6226.

Fischer, E. (1894). *Ber. Dtsch. Chem. Ges.* **27**, 2984–2993.

Fox, J. W., Campbell, R., Beggerly, L. & Bjarnason, J. B. (1986). *Eur. J. Biochem.* **156**, 65–72.

Gomis-Rueth, F. X., Stöcker, W., Huber, R., Zwilling, R. & Bode, W. (1993). *J. Mol. Biol.* **229**, 945–968.

Goodford, P. J. (1985). *J. Med. Chem.* **28**, 849–857.

Gooley, P. R, O'Connell, J. F., Marcy, A. I., Cuca, G. C., Salowe, S. P., Bush, B. L., Hermes, J. D., Esser, C. K., Hagmann, W. K., Springer, J. P & Johnson, B. A. (1994). *Struct. Biol.* **1**, 111–118.

Hite, L. A., Jia, L.-G., Bjarnason, J. B. & Fox, J. W. (1994). *Arch. Biochem. Biophys.* **308**, 182–191.

Jongeneel, C. V., Bouvier, J. & Bairoch, A. (1989). *FEBS Lett.* **242**, 211–214.

Meyer, E. (1992). *Protein Sci.* **1**, 1543–1562.

Nicholls, A., Sharp, K. A. & Honig, B. (1991). *Proteins Struct. Funct. Genet.* **11**, 281–296.

Niedzwiecki, L., Teahan, J., Harrison, R. K. & Stein, R. L. (1992). *Biochemistry*, **31**, 12618–12623.

Okada, Y., Morodomi, T., Enghild, J. J., Suzuki, K., Yasui, A., Nakanishi, I., Salvesen, G. & Nagase, H. (1990). *Eur. J. Biochem.* **194**, 721–730.

Schechter, I. & Berger, A. (1967). *Biochem. Biophys. Res. Commun.* **27**, 157–162.

Seltzer, J. L., Akers, K. T., Weingarten, H., Grant, G. A., McCourt, D. W. & Eisen, A. Z. (1990). *J. Biol. Chem.* **33**, 20409–20413.

Seltzer, J. L., Weingarten, H., Akers, K. T., Eschbach, M. L., Grant, G. A. & Eisen, A. Z. (1989). *J. Biol. Chem.* **33**, 19583–19586.

Stams, T., Spurlino, J. C., Smith, D. L., Wahl, R. C., Ho, T. F., Qoronfleh, M. W., Banks, T. M. & Rubin, B. (1994). *Struct. Biol.* **1**, 119–123.

Stetler-Stevenson, W. G., Krutzsch, H. C., Wacher, M. P., Margulies, I. M. K. & Liotta, L. A. (1989). *J. Biol. Chem.* **264**, 1353–1356.

Vallee, B. L. & Auld, D. S. (1990). *Biochemistry*, **29**, 5647–5659.

Van Wart, H. E. & Birkedal-Hansen, H. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 5578–5582.

Woessner, J. F. Jr (1991). *FASEB J.* **5**, 2145–2154.

Zhang, D., Botos, I., Gomis-Rueth, F.-X., Doll, R., Blood, C., Njoroge, G., Fox, J. W., Bode, W. & Meyer, E. F. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 8447–8451.