provided helpful discussion. We are grateful to L. Flaks and the technical staff at the National Synchrotron Light Source at Brookhaven National Laboratory for their support at X8-C. D.L. was an Ontario Graduate Scholarship recipient. This work was supported by a NSERC grant to Z.J. and contract funding from Ontario Pork to C.W.F.

Correspondence should be addressed to Z.J. *email: jia@crystal.biochem.queensu.ca*

Received 23 June, 1999; accepted 6 December, 1999

1. Reddy, N.R., Sathe, S.K. & Salunkhe, D.K. *Adv. Food Res.* **28**, 1–92 (1982).
2. Graf, E. In *Phytic Acid Chemistry and Application* (ed. Graf, E.) 1–21 (Pilatus Press, Minneapolis; 1986).
3. Wodzinski, R.J. & Ullah, H.J. *Adv. Appl. Micro.* **42**, 263–302 (1996).
4. Greiner, R., Konietzny, U. & Jany, K.-D. *Arch. Biochem. Biophys.* **303**, 107–113 (1993).
5. Wyss, M. *et al. Appl. Env. Microbiol.* **65**, 367–373 (1999).
6. Chi, H. *et al. Genomics* **56**, 324–336 (1999).
7. Vincent, J.B., Crowder, M.W. & Averill, B.A. *Trends Biochem. Sci.* **17**, 105–110 (1992).
8. Van Etten, R.L. *Ann. N.Y. Acad. Sci.* **390**, 27–51 (1992).
9. Ostanin, K. *et al. J. Biol. Chem.* **267**, 22830–22836 (1992).
10. Ostanin, K. & Van Etten, R.L. *J. Biol. Chem.* **268**, 20778–20784 (1993).
11. Lindqvist, Y., Schneider, G. & Vihko, P. *Eur. J. Biochem.* **221**, 139–142 (1994).
12. Schneider, G., Lindqvist, V. & Vihko, P. *EMBO J.* **12**, 2609–2615 (1993).
13. LaCount, M.W., Handy, G. & Lebioda, L. *J. Biol. Chem.* **273**, 30406–30409 (1998).
14. Kostrewa, D. *et al. Nature Struct. Biol.* **4**, 185–190 (1997).
15. Kostrewa, D., Wyss, M., D'Arcy, A., van Loon, A.P.G.M. *J. Mol. Biol.* **288**, 965–974 (1999).
16. Cohen, G.H. *J. Appl. Crystallogr.* **30**, 1160–1161 (1997).
17. Jia, Z., Golovan, S., Ye, Q. & Forsberg, C.W. *Acta Crystallogr.* D**54**, 647–649 (1998).
18. Otwinowski, Z. In *Proceedings of the CCP4 Study Weekend: Data Collection and Processing* (ed Sawyer L., Issacs N.& Bailey S.) 56–62 (Daresbury Laboratory, Warrington; 1993).
19. Minor, W. *XdisplayF Program*. Purdue University, West Lafayette, USA (1993).
20. Collaborative Computational Project, Number 4. *Acta Crystallogr.* D**50**, 760–763 (1994).
21. Egloff, M.-P., Cohen, P. T. W., Reinemer, P. & Barford, D. *J. Mol. Biol.* **254**, 942–959 (1995).
22. de La Fortelle, E. & Bricogne, G. In *Methods in Enzymology, Macromolecular Crystallography* (eds. Sweet, R.M. & Carter, Jr. C.W.) **276**, 472–494 (Academic Press, New York; 1997).
23. Abrahams J.P. & Leslie A.G.W. *Acta Crystallogr.* D**52**, 30–42 (1996).
24. McRee, D.E. *J. Mol. Graphics* **10**, 44–47 (1992).
25. Brünger, A.T. *et al. Acta Crystallogr.* D**54**, 905–921 (1998).
26. Navaza, J. *Acta Crystallogr.* A**50**, 157–163 (1994).
27. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. *J. App. Crystallogr.* **26**, 283–291 (1993).
28. Engelen, A.J., Vanderheeft, F.C., Randsdorp, P.H.G., & Smit, E.L.C. *J. AOAC. Int.* 77: 760–764. (1994).
29. Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., & Klenk, D.C. *Anal. Biochem.* 150: 76–85. (1985).
30. Kraulis, P.J. *J. Appl. Crystallogr.* **24**, 946–950 (1992).
31. Merrit, E.A. & Murphy, M.E.P. *Acta Crystallogr*. D**50**, 869–873 (1994).
32. Nicholls, A., Sharp, K. & Honig, B. *Proteins* **11**, 281–296 (1991).

The aspartic proteinase from *Saccharomyces cerevisiae* folds its own inhibitor into a helix

Mi Li[1–3], Lowri H. Phylip[1,4], Wendy E. Lees[4], Jakob R. Winther[5], Ben M. Dunn[6], Alexander Wlodawer[2], John Kay[4] and Alla Gustchina[2]

[1]These authors contributed equally to this work. [2]Macromolecular Crystallography Laboratory, Program in Structural Biology, National Cancer Institute-FCRDC, Frederick, Maryland 21702, USA. [3]Intramural Research Support Program, SAIC Frederick, National Cancer Institute-FCRDC, Frederick, Maryland 21702, USA. [4]School of Biosciences, Cardiff University, Cardiff CF1 3US, UK. [5]Department of Yeast Genetics, Carlsberg Laboratory, DK 2500 Copenhagen Valby, Denmark. [6]Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, Florida 32610, USA.

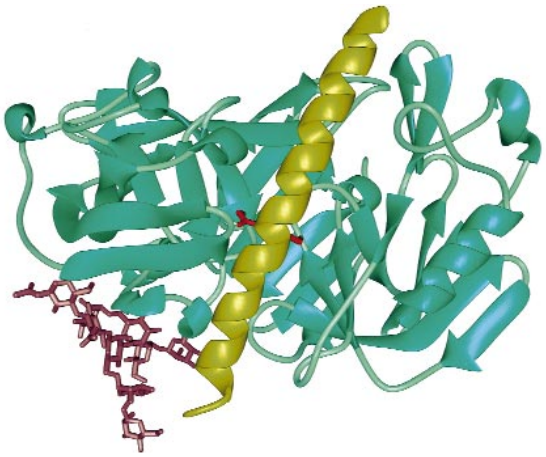

Fig. 1 Schematic diagram of the structure of proteinase A. Ribbon representation showing the tracing of the main chain of proteinase A (green) with the oligosaccharide attached to Asn 67 shown in mauve, together with the visible fragment of the inhibitor IA_3 (gold). Side chains of the active site residues Asp 32 and Asp 215 are shown in red.

Aspartic proteinase A from yeast is specifically and potently inhibited by a small protein called IA_3 from *Saccharomyces cerevisiae*. Although this inhibitor consists of 68 residues, we show that the inhibitory activity resides within the N-terminal half of the molecule. Structures solved at 2.2 and 1.8 Å, respectively, for complexes of proteinase A with full-length IA_3 and with a truncated form consisting only of residues 2–34, reveal an unprecedented mode of inhibitor–enzyme interactions. Neither form of the free inhibitor has detectable intrinsic secondary structure in solution. However, upon contact with the enzyme, residues 2–32 become ordered and adopt a near-perfect α-helical conformation. Thus, the proteinase acts as a folding template, stabilizing the helical conformation in the inhibitor, which results in the potent and specific blockage of the proteolytic activity.

Proteolytic enzymes are categorized into four main families[1]. Whereas structures have been solved for complexes of gene-encoded inhibitors (for example, serpins[2], cystatins[3] and TIMPs[4]) that target three of these protein families (serine, cysteine and metalloproteinases), protein inhibitors of the aspartic proteinase family are rare[5] and no mechanism of action has previously been elucidated for any of the few that are known, such as IA_3 (ref. 6). Human pathogens, including HIV, *Plasmodium falciparum* (malaria) and fungi such as *Candida*, the causative agent of thrush infections, are known to be crucially dependent on aspartic proteinases for their replication and survival[7–9]. Substantial progress has been made in the generation of chemically synthesized inhibitors to block the action of these enzymes[10,11], and several HIV-1 proteinase inhibitors are now licensed by the United States Food and Drug Administration as drugs for use in combating AIDS[10,12]. In contrast, only a few naturally occurring protein inhibitors of this proteinase family have been found, for example, in humans (renin binding protein[13]), in the parasitic nematode *Ascaris lumbricoides*[14],

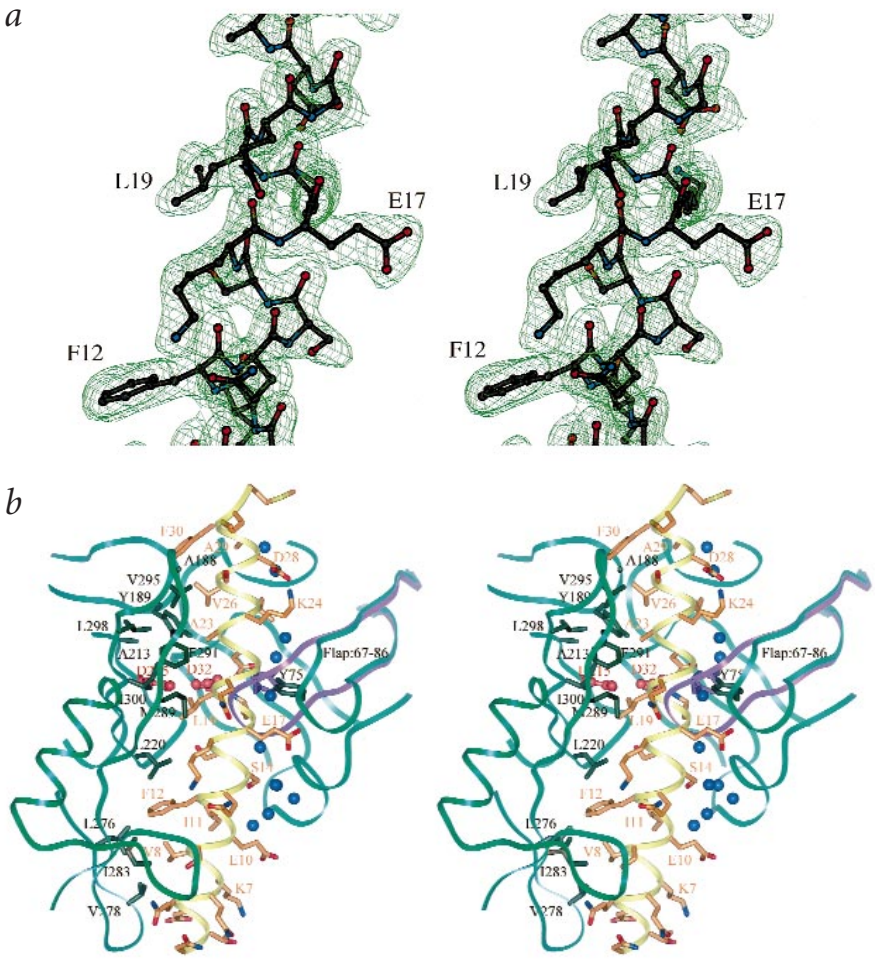

Fig. 2 Inhibitor IA_3 and its interactions with proteinase A. ***a***, Central part of the inhibitor helix covered by the $2F_o$ - F_c electron density map, contoured at 1 σ level. ***b***, Ribbon representation of the relevant fragments of the main chain of proteinase A is in green, and the main chain of IA_3 is shown in yellow. All side chains of IA_3 are shown in orange, and the catalytic aspartates of proteinase A are red, and selected hydrophobic side chains are dark green. The structure of the flap with Tyr 75 from the complex of proteinase A with a small-molecule inhibitor CP81,282 (ref. 17), superimposed on the IA_3 complex, are marked in purple. The amphipathic character of the IA_3 helix is due to the almost exclusive presence of hydrophobic residues (Val 8, Ile 11, Phe 12, Leu 19, Ala 23, Val 26 and Phe 30) on one face of the helix (left in this view). The only polar residue on this side, Ser 15, makes a hydrogen bond with the side chain of Thr 220 of the enzyme. The majority of the side chains on the other face of the helix are polar. Water molecules involved in essential polar interactions are blue.

in plants such as potato[15] and squash[5] and in the yeast *S. cerevisiae,* which encodes not only the IA_3 inhibitor but also its target enzyme, proteinase A[16].

The structure of the yeast vacuolar proteinase A has been previously solved and refined to 2.4 Å (ref. 17). This enzyme is the sole known target for its cognate protein inhibitor IA_3 (ref. 16), which does not inhibit any one of a large number of aspartic proteinases with similar sequence/structure[6]. Expression of the gene encoding IA_3 (ref. 18) in *Escherichia coli* enabled us to generate sufficient quantities of homogeneous inhibitor for structural studies. The recombinant wild type protein inhibited yeast proteinase A effectively at pH 3.1 (Table 1), and its potency increased even further at higher pH values. The efficiency of inhibition was not significantly affected in a mutant form of the inhibitor in which the ε-NH_3^+ group of two adjacent lysine residues was removed and replaced with the otherwise isosteric side chains of methionine (Met 31–Met 32 mutant, Table 1). A synthetic peptide spanning residues 2–34 of IA_3 also showed a potency comparable to that of the full-length protein inhibitor, as measured by K_i values (Table 1). This suggests that the initiator Met residue and residues 35–68 in the IA_3 sequence[19] are not important for inhibitory function. Peptides 2–15 and 16–34 did not inhibit the enzyme when added either alone or together. Indeed, separate incubation of each of these peptides with proteinase A at pH 4.7 resulted in cleavage at the Glu 10–Ile 11and Ala 29–Phe 30 bonds, respectively. In contrast, peptide 2–34 was not cleaved by proteinase A even upon prolonged incubation for 16 h. Thus, a contiguous stretch of residues from 2–34 not only prevents cleavage of this peptide by proteinase A but also acts an effective inhibitor of the enzyme.

Peptide 2–34 did not significantly inhibit any one of a number of aspartic proteinases from a wide variety of other species including human pepsin (data not shown). However, this peptide was readily cleaved by trypsin, leukocyte elastase and human pepsin, with pepsin cleaving the Glu 10–Ile 11 and Ala 29–Phe 30 bonds to generate the peptide fragments Asn 2–Glu 10, Ile 11–Ala 29 and Phe 30–Ala 34. None of these fragments alone or all three together were able to inhibit yeast proteinase A (Table 1). The aspartic proteinase pepsin has 40% sequence similarity to yeast proteinase A[20] and has a very similar fold and substrate specificity. The two pepsin cleavage sites in the intact peptide 2–34 are identical to those generated by proteinase A in peptides 2–15 and 16–34. Thus, it seems that the sequence that efficiently inhibits proteinase A is able to act as a good substrate for another, very similar enzyme. High-field NMR studies of IA_3 indicated that little intrinsic structure was present in the free inhibitor (A. Edison and R.A. Byrd, unpublished observations). Circular dichroism (CD) measurements were also performed on the peptide (residues 2–34) and protein forms of the free inhibitor in solution. Inspection of both spectra (data not shown) gave no indication of the presence of any ordered structure, be it helical or extended. Instead, the spectra were compatible with random coil. By contrast, CD spectra of proteinase A alone or complexed with IA_3 showed the expected mixture of different secondary structure elements, with the latter exhibiting slightly higher helical content (data not shown).

Complexes were made for the various forms of the inhibitor bound to proteinase A. The complex with the Met 31–Met 32 mutant of the inhibitor formed hexagonal crystals that diffracted to 2.2 Å on a synchrotron beam. The structure was solved by molecular replacement, and was refined to an R-factor of 0.188 (Table 2). The structure of proteinase A complexed to peptide 2–34 (the peptide form of the inhibitor consisting of only residues 2–34; Table 1) was solved and refined at 1.8 Å resolution (Table 2). Since the struc-

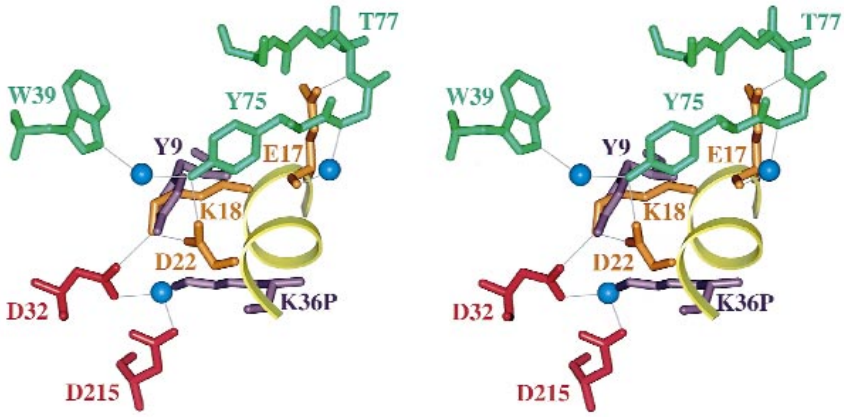

Fig. 3 Interactions in the vicinity of the active site of proteinase A. Proteinase A (green, with the active site aspartates red) complexed to IA_3 (yellow with orange side chains) is superimposed on progastricsin (purple). Water molecules in the proteinase A complex are blue, and hydrogen bonds are marked as thin lines.

tures bound to both forms of the inhibitor are very similar, only the Met 31–Met 32 mutant protein form is shown here and is used for numerical comparisons. In the resulting structures (Fig. 1), residues 2–32 of IA_3 become ordered and form a near-perfect amphipathic α-helix (Fig. 2). In contrast, the remaining 36 residues from the C-terminus of the inhibitor are unstructured. The visible parts of the inhibitor are not involved in any crystal contacts. The axis of the helix aligns well with the groove of the active site, and the side chains form numerous interactions with the enzyme along the whole length of the helix (Fig. 2*b*). Residues 2–30 of the peptide inhibitor superimpose on the corresponding residues of the Met 31–Met 32 mutant protein form of IA_3 to within a root mean square (r.m.s.) deviation of 0.21 Å. The structure of the enzyme in both complexes is similar to that of the reported unliganded enzyme and its complex with CP81,282, a small-molecule inhibitor[17]. When the structures of the enzyme–inhibitor complexes described here are compared with that of the CP81,282 complex (the only structure with available coordinates), the largest difference observed is up to 8.7 Å for the flap (a β-hairpin covering the active site, consisting of residues 72–82). Whereas the helical nature of IA_3 agrees in general terms with the helix-forming propensity of its primary structure, the N-terminal half of the molecule is predicted by the program PhD[21] to consist of three short helical stretches and not of a single, long helix. The C-terminal half of the protein, disordered both in solution and in crystals, is also predicted to be largely helical[21].

An analysis of the interactions between proteinase A and IA_3 revealed several important residues in the inhibitor that form specific interactions with the catalytic aspartates and residues from the flap (see below). The spatial juxtaposition of the two catalytic Asp residues in the complex is similar to that previously observed in a number of nonliganded mature aspartic proteinases[22]. The water molecule that is considered to be the catalytic nucleophile in reactions catalyzed by the aspartic proteinases occupies the same position between the two aspartates in the IA_3–proteinase A complex (Fig. 3). Such a water molecule has been frequently observed in the structures of a number of unliganded enzymes as well[22]. The presence of this water molecule is unprecedented; it has not been previously detected in inhibitor complexes of aspartic proteinases. The ε-NH_3^+ group of Lys 18 in the inhibitor is located within hydrogen bonding distance (2.75 Å) of the carboxyl oxygen of Asp 32. Precursor forms of aspartic proteinases (such as pepsinogen and progastricsin) also have a conserved lysine residue in the pro part of their sequences (Lys 36P), which interacts with the two catalytic aspartate residues (Fig. 3). This lysine has been shown to be important for the stability and folding of the precursor molecule[23]. However, the amino group of this lysine is equidistant from the two catalytic aspartates, occupying the position of the nucleophilic water molecule in the mature enzyme[23]. In the structure of the IA_3–proteinase A complex, the location of the NH_3^+ group of Lys 18 in IA_3 is different and is close to the position occupied by the hydroxyl of the conserved Tyr 9 in the structures of pepsinogen and progastricsin[24], which also forms a hydrogen bond with Asp 32. Another important residue in IA_3 is Asp 22. Both carboxyl oxygens of this residue are involved in the formation of two hydrogen bonds with Lys 18 in IA_3 and Tyr 75 in the flap region of proteinase A. Glu 17 in IA_3 also interacts with the flap region and forms a hydrogen bond with the main chain amide of Thr 77 in the enzyme (Fig. 3). Residues Ala 23, Val 25, Val 26, Phe 30 and Met 31 in IA_3, located on the hydrophobic side of the amphipathic helix, are involved in extensive hydrophobic interactions with residues along the length of the active site of the enzyme (Fig. 2). Residues Leu 19 and Phe 12 are oriented in the direction of the binding pockets S2 and S4, respectively, in the active site cleft of the proteinase.

Table 1 Interaction of yeast proteinase A at pH 3.1 with protein and peptide forms of the inhibitor IA_3

Designation	Sequence[1] (1 5 10 15 20 25 30)	K_i (nM)[2]
Wild type	M N T D Q Q K V S E I F Q S S K E K L Q G D A K V V S D A F K K M A~ +34 ~L E H_6	1.1 ± 0.4
M31–M32	M N T D Q Q K V S E I F Q S S K E K L Q G D A K V V S D A F M M M A~ +34 ~L E H_6	0.9 ± 0.3
2–34	N T D Q Q K V S E I F Q S S K E K L Q G D A K V V S D A F K K Z A	3 ± 0.6
2–15	N T D Q Q K V S E I F Q S S	N.I.
16–34	K E K L Q G D A K V V S D A F K K Z A	N.I.
2–10	N T D Q Q K V S E	N.I.
11–29	I F Q S S K E K L Q G D A K V V S D A N.I.	
30–34	F K K Z A	N.I.

[1]The sequence of residues 1–34 of IA_3 is given in full, with the C-terminal 34 residues of the protein depicted by +34 followed by the His tag encoded by the vector. Replacement of Lys 31–Lys 32 to generate the Met 31–Met 32 mutant was accomplished by overlapping PCR mutagenesis. Peptide forms consisted only of the residues indicated by the designated residue numbers; amino acids are indicated by the single letter code, Z = norleucine.
[2]N.I. = no inhibition at a concentration of 1 μM. K_i values (<0.1 nM) were also measured for the wild type and mutant protein and peptide 2–34 forms of the inhibitor at pH 4.7 and pH 6.0.

Table 2 Data collection and refinement statistics

	Met 31–Met 32 protein	Peptide 2–34
Data collection		
Resolution (Å)	2.2	1.8
Unique reflections	28,115	51,951
R_{merge} (%) (all/last shell)	6.9/33.2	7.3/62.9
I/I_σ (all/last shell)	15.8/6.0	25.2/2.5
Completeness (%) (all/last shell)	95.6/99.6	99.0/99.1
Redundancy	5.8	6.6
Structure refinement		
Resolution (Å)	30–2.2	30–1.8
R_{work} (%)	18.83	20.55
R_{free} (%)	23.49	22.16
R.m.s.d. bond lengths (Å)	0.011	0.017
R.m.s.d. bond angles (°)	1.65	1.88

Although residues 2–32 of IA_3 form an intact helix in the complex with proteinase A, cleavage of this sequence was readily accomplished by several other proteinases, as mentioned above. It is widely accepted that proteinases can only attack stretches of polypeptides with an extended and accessible conformation[25]. For aspartic proteinases in general, and pepsin in particular, chemically synthesized peptidomimetics adopt a β-strand conformation in the active site cleft of these enzymes[26]. Thus, IA_3 appears to bind pepsin as a β-strand and is therefore cut, whereas when it interacts with proteinase A it undergoes a random coil to α-helix transition. This spares it from degradation but at the same time ensures that it becomes lodged as a long helix in the active site of the enzyme. This arrangement guarantees that no peptide bond in the backbone of IA_3 is positioned any closer to the catalytic aspartate residues than 5 Å, which is not near enough to permit cleavage. Thus, by ensuring that its backbone is held away from the catalytic machinery, IA_3 inhibits its target proteinase.

Therefore, the structure of the complex of proteinase A with a protein inhibitor is very different from the structures of zymogens of aspartic proteinases, the only previously available examples of protein–protein interactions for these enzymes. The presence of a helical entity inhibiting the active site of an aspartic proteinase is unprecedented and suggests a possible new paradigm of inhibition, possibly leading to new directions in designing drugs that inhibit this important class of enzymes. This observation may even be applicable to the design of new inhibitors of retroviral proteinases, which have so far always been designed as extended molecules[10]. The structural transitions in the IA_3 inhibitor, from an unfolded structure in solution to a putative extended chain as a substrate and a helical structure as an inhibitor, are very significant and make this system particularly useful for future studies of protein folding and interactions.

Methods

Protein expression and purification. Proteinase A was purified to homogeneity and peptides were synthesized by solid-phase methods, as described[20]. Genomic DNA was extracted from *S. cerevisiae* and the gene encoding IA_3 (ref. 18) was amplified specifically by polymerase chain reaction (PCR) using G CAT ATG AAT ACA GAC CAA CAA AAA GTG and G CTC GAG CTC CTT CTT ATG CCC CGC as forward and reverse primers. Mutations were introduced into this wild type sequence, cloned into the pGEM.T vector (Promega) by overlapping PCR mutagenesis[27] using the mutagenesis primers 5'-TTT ATG ATG ATG GCC AGT CAA GAC AAG GAC GGC-3' and 5'-ACT GGC CAT CAT CAT AAA AGC GTC ACT CAC TAC CTT-3'. Wild type and mutant forms of IA_3 were subcloned into the *Nde*I-*Xho*I sites of pET-22b (Invitrogen), thus introducing a C-terminal Leu-Glu-His_6 tag, which was subsequently shown to have no effect on inhibitory potency. *E. coli* strain BL21 DE3 pLysS was transformed with wild type or mutant clones, then grown at 37 °C in LB medium to an A_{600} of approximately 0.6 before induction with 1 mM IPTG. The recombinant protein was soluble in the cell lysates and was purified to homogeneity by a two-step procedure that involved nickel-chelate chromatography (the sample was loaded in 0.05 M sodium phosphate buffer, pH 8.0, containing 0.3 M NaCl, the column was washed with the same buffer at pH 6.0 containing 10% glycerol, and IA_3 was eluted with the glycerol-containing buffer adjusted to pH 4.0), followed by heating twice at 100 °C for 5 min. Edman degradation of the wild type and Met 31–Met 32 mutant IA_3 proteins confirmed the cDNA predicted sequences and indicated that the initiator methionine residue had not been cleaved off by the *E. coli* enzymes. MALDI-TOF mass spectrometry analysis gave values of 8772.8 and 8778.2 (predicted 8772.2 and 8778.3) for the wild type and mutant inhibitors, respectively.

Kinetic and CD measurements. Kinetic measurements were made at pH 3.1 and pH 4.7 using Lys-Pro-Ile-Glu-Phe∗NitroPhe-Arg-Leu as the substrate[20]. Digestions were performed at 37 °C for 2–16 h, using IA_3:proteinase ratios of 40:1 for trypsin and proteinase A, 60:1 for leukocyte elastase, and 1,000:1 for human pepsin, respectively. Digests were separated by reverse-phase FPLC using a Pep-RP column (Pharmacia). Fractions were collected and subjected to amino acid analysis. CD measurements were made with a Jasco 720 spectropolarimeter at room temperature. The scanning started at a wavelength of 265 nm and ended at 195 nm, using a 0.2 mm cuvette. The signal was recorded at every 0.2 nm. The inhibitor sample was dissolved in 20 mM MES buffer, pH 6.0, containing 0.1 M NaCl.

Crystallization and structure determination. Complexes were prepared for crystallization by mixing proteinase A and the appropriate form of the inhibitor in molar ratios of 1:5 in 20 mM MES buffer, pH 6.6. Each complex was separated from the excess of free inhibitor by gel filtration on Sephadex G-50 in the same MES buffer. The complexes were concentrated at 4 °C to approximately 5 mg ml^{-1} using Centriprep (Amicon) devices with a 3 kDa cutoff. Crystals of proteinase A bound to full-length protein inhibitor IA_3 were obtained by vapor diffusion. The well solution contained 28% PEG 1500, 0.2 M ammonium sulfate in 0.1 M MES buffer at pH 6.0. The complex sample was in 20 mM MES buffer, pH 6.6. The hanging drop was a 1:1 mixture of the sample and well solution. Crystals appeared after a few days and grew to 0.15 × 0.4 × 0.3 mm after two weeks. They belong to space group $P6_222$ with unit cell dimensions of a = b = 192.66 Å, and c = 52.09 Å, containing one complex per asymmetric unit. Fully isomorphous crystals of a complex with the peptide 2–34 (Table 1) were grown in a similar manner. Diffraction data (Table 2) were collected at 100 K with an ADSC 4K CCD detector on synchrotron beamline X9B at Brookhaven National Laboratory and were processed with HKL2000 (ref. 28). The structure of the complex with full-length inhibitor was solved by molecular replacement with AMoRe[29]. Proteinase A coordinates from the complex with CP81,282 (ref. 17) were used as the search model. A single solution was obtained with a correlation factor of 0.503 and R-factor of 0.385. The resulting model was rebuilt with the program O (ref. 30) and refined with CNS (ref. 31). After rigid-body refinement, discontinuous densities from the F_o - F_c map were observed in the inhibitor-binding site. The loops 73–80, 108–113, 158–164, 241–246 and 278–283 did not have good densities and were removed. After positional refinement and simulated annealing at 3,000 K, the inhibitor region had continuous electron density and showed obvious helical features. During the next five rounds of positional refinement and simulated annealing, 31 residues of the inhibitor (2–32) were built, including their side chains (Fig. 2*a*). In the final stage of the refinement, a polysaccharide chain of nine residues attached to Asn 67 was located and 391 water molecules were added. The structure of the complex with the truncated inhibitor was refined starting with the final model described above. The results of the refinements are summarized in Table 2.

Coordinates. Coordinates and structure factors for both complexes have been submitted to the Protein Data Bank (accession codes 1dp5 and 1dpj, respectively).

Acknowledgments
This paper is dedicated to the memory of H. Holzer, formerly of the University of Freiburg im Breisgau, Germany. We thank A. Chung at the Protein Chemistry Core, University of Florida, for synthesis of peptides; A. Edison and R.A. Byrd for their valuable contributions with the NMR; V. Dhanaraj and B. Brownsey for their extensive help with modeling and mass spectroscopy analysis, respectively; J. Collins for help with the preparation of Fig. 1; and A. Arthur for editorial comments. This work was supported in part by awards from the BBSRC (ROPA) and from Actelion, Allschwil, Switzerland (J.K.), by an NIH grant (B.M.D.) and by the National Cancer Institute, DHHS, under contract with ABL (A.W.).

Correspondence should be addressed to J.K. *email: smbjk@cardiff.ac.uk* or A.G. *email: alla@ncifcrf.gov*

Received 27 September, 1999; accepted 8 December, 1999.

1. *Handbook of proteolytic enzymes* (eds Barrett, A.J., Woessner, J.F. & Rawlings, N.D.) 1–1650 (Academic Press, London; 1998).
2. Schreuder H.A. *et al. Nature Struct. Biol*. **1**, 48–54 (1994).
3. Engh, R.A. *et al. J. Mol. Biol*. **234**, 1060–1069 (1993).
4. Gomis-Ruth, F.X. *et al. Nature*, **389**, 77– 81 (1997).
5. Christeller, J.T., Farley, P. E., Ramsay, R.J., Sullivan, P. A. & Laing, W.A. *Eur. J. Biochem.* **254**, 160–167 (1998).
6. Dreyer, T., Valler, M. J., Kay, J., Charlton, P. & Dunn, B.M. *Biochem. J.* **231**, 777–779 (1985).
7. Craig, C. *et al. AIDS* **12**, 1611–1618 (1998).
8. Dame, J.B. *et al. Mol. Biochem. Parasitol.* **64**, 177–190 (1994).
9. Monod, M., Togni, G., Hube, B. & Sanglard, D. *Mol. Microbiol*. **13**, 357–368 (1994).
10. Wlodawer, A. & Vondrasek, J. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 249–284 (1998).
11. Moon, R.P. *et al. Eur. J. Biochem*. **244**, 552–560 (1997).
12. Roberts, N.A., Craig, J.C. & Sheldon, J. *AIDS* **12**, 453–460 (1998).
13. Takahashi, S. *et al. J. Biochem*. **125**, 348 – 353 (1999) .
14. Kageyama, T. *Eur. J. Biochem*. **253**, 804 – 809 (1998).
15. Kreft, S. *et al. Phytochemistry* **44**, 1001–1006 (1997).
16. Saheki, T., Matsuda, Y. & Holzer, H. *Eur. J. Biochem*. **47**, 325–332 (1974).
17. Aguilar, C.F. *et al*. *J. Mol. Biol*. **267**, 899–915 (1997).
18. Schu, P. & Wolf, D.H. *FEBS Lett*. **283**, 78–84 (1991).
19. Biedermann, K., Montali, U., Martin, B., Svendsen, I. B. & Ottesen, M. *Carlsberg Res. Commun*. **45**, 225–235 (1980).
20. Kondo, H. *et al. J. Biochem.* **124**, 141–147 (1998).
21. Rost, B. & Sander, C. *J. Mol. Biol.* **232**, 584–599 (1993).
22. Davies, D.R. *Annu. Rev. Biophys. Biophys. Chem.* **19**, 189–215 (1990).
23. Richter, C., Tanaka, T., Koseki, T. & Yada, R.Y. *Eur. J. Biochem.* **261,** 746–752 (1999).
24. Khan, A.R. & James, M.N.G. *Protein Sci.* **7**, 815–836 (1998).
25. Wu, C.Y., Robertson, D.H.L., Hubbard, S.J., Gaskell, S. J. & Beynon, R. J. *J. Biol. Chem*. **274**, 1108–1115 (1999).
26. Cooper, J.B. *et al. Biochemistry* **28**, 8596–8603 (1989).
27. Tigue, N. J. & Kay, J. *J. Biol. Chem.* **273**, 26441–26446 (1998).
28. Otwinowski, Z. & Minor, W. *Methods Enzymol.* **276**, 307–326 (1997).
29. Navaza, J. *Acta Crystallogr. A* **50**, 157–163 (1994).
30. Jones, T. A. & Kieldgaard, M. *Methods Enzymol*. **277**, 173–208 (1997).
31. Brünger, A. T. *et al*. *Acta Crystallogr. D* **54**, 905–921 (1998).

Structure of a photoactive rhodium complex intercalated into DNA

Clara L. Kielkopf[1,2], Kathryn E. Erkkila[3], Brian P. Hudson[3,4], Jacqueline K. Barton[3] and Douglas C. Rees[3–5]

[1]Division of Biology, California Institute of Technology, Pasadena, California 91125, USA. [2]Present address: Department of Molecular Biophysics, The Rockefeller University, New York, New York 10021, USA. [3]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA. [4]Present address: Department of Chemistry, Scripps Research Institute, La Jolla, California 92037, USA. [5]Howard Hughes Medical Institute, California Institute of Technology, Pasadena, California 91125, USA.

Intercalating complexes of rhodium(III) are strong photo-oxidants that promote DNA strand cleavage or electron transfer through the double helix. The 1.2 Å resolution crystal structure of a sequence-specific rhodium intercalator bound to a DNA helix provides a rationale for the sequence specificity of rhodium intercalators. It also explains how intercalation in the center of an oligonucleotide modifies DNA conformation. The rhodium complex intercalates via the major groove where specific contacts are formed with the edges of the bases at the target site. The phi ligand is deeply inserted into the DNA base pair stack. The primary conformational change of the DNA is a doubling of the rise per residue, with no change in sugar pucker from B-form DNA. Based upon the five crystallographically independent views of an intercalated DNA helix observed in this structure, the intercalator may be considered as an additional base pair with specific functional groups positioned in the major groove.

Before the elucidation of the Watson-Crick model for the DNA double helix, it was realized that many planar, aromatic compounds had mutagenic activity due to their interaction with DNA. Subsequent biochemical and fiber diffraction studies by Lerman in the early 1960s resulted in the intercalation model of binding[1]; the aromatic small molecules bound to otherwise B-form DNA by insertion of a planar ligand between the base pairs. Although this binding mode resulted in high affinities of the ligand for double-stranded DNA, the majority of the interactions were with the aromatic surfaces of the bases rather than with the distinctive functional groups in the DNA grooves. Consequently, most intercalators, such as ethidium and acridine, are relatively nonspecific DNA binding agents[2]. Despite the widespread use of intercalators in the fields of pharmacology and molecular biology, the lack of sequence-specific binding has limited high-resolution structural characterization of binding within a DNA helix[3].

In contrast with most naturally occurring intercalators, octahedral complexes of Rh(III) containing the phi ligand (where phi is 9,10-phenanthrenequinone diimine) constitute a class of intercalators with the ability to bind DNA specifically[4]. The octahedral geometry of these complexes provides a rigid framework that predictably orients the ancillary, nonintercalating ligands in the major groove, and these ancillary ligands can be synthetically modified to alter the sequence specificity and reactivity of the molecule. Furthermore, these metallointercalators are photoactive, and can carry out both short- and long-range photoinduced oxidation of DNA through reactions that are a sensitive function of their stacking within the DNA helix[4–7]. Metallointercalators have been exploited extensively for site-specific targeting of DNA[4], for the photochemical repair of DNA lesions[5], as photoluminescent reporters of DNA[8], in mismatch detection[9], as probes of DNA structure[10] and electron transfer chemistry[11], and to inhibit DNA binding proteins[12]. An atomic resolution structure of the ligand-bound DNA site would contribute to understanding the specificity and reactivity of metallointercalators, as well as address how intercalation modifies the DNA conformation.

The sequence specificity of the octahedral rhodium intercalator Δ-α-[Rh[(R,R)-Me_2trien]phi]$^{3+}$ (where (R,R)-Me_2trien is 2*R*,9*R*-diamino-4,7-diazadecane) was employed to obtain an atomic resolution structure of an intercalated DNA helix. This intercalator was successfully designed to recognize the major groove of the four-base pair sequence 5'-TG|CA-3' (where | indicates phi insertion)[13].