The crystal structure of a major secreted aspartic proteinase from *Candida albicans* in complexes with two inhibitors

SM Cutfield[1], EJ Dodson[2], BF Anderson[3], PCE Moody[2], CJ Marshall[1], PA Sullivan[1] and JF Cutfield[1]*

[1]Biochemistry Department, University of Otago, PO Box 56, Dunedin, New Zealand, [2]Chemistry Department, University of York, York, YO1 5DD, UK, [3]Department of Chemistry and Biochemistry, Massey University, Private Bag 11222, Palmerston North, New Zealand

Background: Infections caused by *Candida albicans*, a common fungal pathogen of humans, are increasing in incidence, necessitating development of new therapeutic drugs. Secreted aspartic proteinase (SAP) activity is considered an important virulence factor in these infections and might offer a suitable target for drug design. Amongst the various SAP isozymes, the *SAP2* gene product is the major form expressed in a number of *C. albicans* strains.

Results: The three-dimensional structures of SAP2 complexed with the tight-binding inhibitor A70450 (a synthetic hexapeptide analogue) and with the general aspartic proteinase inhibitor pepstatin A (a microbial natural product) have been determined to 2.1 Å and 3.0 Å resolution, respectively. Although the protein structure retains the main features of a typical aspartic proteinase, it also shows some significant differences, due mainly to several sequence insertions and deletions (as revealed by homology modelling), that alter the shape of the binding cleft. There is also considerable variation in the C-terminal structural domain.

Conclusions: The differences in side chains, and in the conformations adopted by the two inhibitors, particularly at their P4, P3 and P'2 positions (using standard notation for protease-inhibitor residues), allows the A70450 structure to complement, more accurately, that of the substrate-binding site of SAP2. Some differences in the binding clefts of other SAP isoenzymes may be deduced from the SAP2 structure.

Structure 15 November 1995, **3**:1261–1271

Key words: aspartic proteinase, *Candida albicans*, crystal structure, inhibitors, isoenzyme

Introduction

Fungal infections are now widespread and pose a major threat in hospitals, a situation made more serious by the limitations of the anti-fungal drugs that are currently available [1]. *Candida* species, and *Candida albicans* in particular, are the main culprits, particularly in individuals whose immune systems have become compromised (e.g. patients with AIDS, cancer, organ transplants and drug addiction). Systemic infections in these patients are often life-threatening [2]. Several factors are believed to contribute to fungal pathogenicity. These include the means for adhesion to and subsequent invasion of host tissues, as well as the presentation of antigens [3,4]. Polymorphism, especially in the ability to form hyphae, and phenotypic switching in some strains, are particular properties of *C. albicans*, that may also contribute to virulence [5,6]. Of the enzymes implicated in host invasion, secreted aspartic proteinase (SAP) activity appears to have the strongest credentials as a virulence factor [1,7], and therefore offers a potential target for drug intervention in infections involving *Candida* species. The evidence for SAP involvement includes its presence in tissues and fluids of infected patients [7,8], correlation between extent of SAP secretion and pathogenicity [9], and the reduction in colonising ability shown by a SAP-deficient stable mutant of *C. albicans* [10]. The secretion of SAP by *Candida* sp. enables it to use a variety of exogenous proteins as a nitrogen source [11,12], the peptide cleavage sites being relatively non-specific, although a preference for hydrophobic residues immediately preceding the scissile bond is displayed [12].

Over the past three years it has been found that *C. albicans* possesses a family of at least eight *SAP* genes, of which *SAP2* is the major form expressed in a number of strains [13,14] under laboratory conditions. *SAP2* encodes a 398-residue preproprotein [15] that is processed to a 342-residue mature enzyme, a typical aspartic proteinase of pH optimum 3–4, displaying sensitivity to pepstatin A, a peptide-based inhibitor.

Members of the widespread aspartic-proteinase family are now well characterised structurally, being mainly β structures organised into two domains with an Asp–Thr–Gly (DTG)/Asp–Ser–Gly (DSG) motif in the centre of an extended cleft [16]. The two catalytic aspartic-acid residues are within hydrogen-bonding distance of each other and are believed to act in concert, via an intervening water molecule, to bring about the base-catalysed cleavage of the scissile peptide bond [16]. Sequence comparisons of SAP2 with aspartic proteinases of known three-dimensional (3D) structure show that SAP2 shares 22–27% identity with the fungal proteinases rhizopuspepsin, penicillopepsin and endothiapepsin and the

*Corresponding author.

mammalian proteinases pepsin, chymosin, renin and cathepsin D. Homology modelling [17] has indicated some significant structural differences, due to insertions and deletions, from these other enzymes.

The design of SAP-specific inhibitors depends crucially on a detailed understanding of the various SAP isozyme structures and how they differ from host aspartic proteinases. Tetragonal crystals of the SAP2 isoenzyme suitable for X-ray diffraction analysis have been grown in the presence of the strong inhibitor A70450 (Abbott Laboratories) [17,18]. We report the crystal structure of SAP2, the major secreted aspartic proteinase from *C. albicans* (strain ATCC 10261), complexed with the inhibitor A70450, to a resolution of 2.1 Å. In addition, we have solved the structure of an orthorhombic form of the SAP2–pepstatin complex to 3.0 Å resolution by using the SAP2–A70450 structure for molecular replacement.

Results and discussion

Structure determination

The conditions for successful crystallisation of SAP2 with A70450 took some time to establish [17] but once they were refined, tetragonal crystals of the complex could be grown in one to two days. The high salt conditions used were, however, not suitable for stable pepstatin binding as slow cleavage of SAP2 into two fragments occurred during crystallisation [17], the resulting crystals diffracting weakly and anisotropically. Consequently, quite different conditions were required for successful crystallisation of the SAP2–pepstatin complex. Using a rotating anode generator/image plate detector system, the orthorhombic crystals diffracted to just beyond 3.0 Å resolution, compared with 2.0 Å for the tetragonal SAP2–A70450 crystals.

The structure of SAP2–A70450 was solved first, by a combination of single isomorphous replacement with anomalous scattering, and molecular replacement (MR). Although SAP2 was a clearly identified member of the aspartic-proteinase family, the relatively low sequence homology with the other structurally-defined members, the anticipated variation in loop regions (as deduced from modelling studies [17]), and the uncertainty in orientation between the two rigid-body domains [19,20], suggested that the molecular replacement approach might not be straightforward. This turned out to be the case, and so it was decided early on in the analysis to obtain data for experimental phasing by isomorphous replacement. Interpretation of the single derivative-phased map was greatly aided by an MR solution based on the fungal aspartic proteinase rhizopuspepsin. The chain tracing was, however, a relatively slow process, largely due to weak density for several extensive loops (particularly in the C-terminal domain), that were also involved in intermolecular contacts. Eventually all 342 residues were placed although some uncertainty exists for parts of the large loops, as reflected in their high temperature factors. Assignments were in agreement with the published sequence [15]. A difference map revealed excellent electron density for the bound inhibitor, and all atoms were located.

The final model gave a crystallographic R factor of 0.195 for all data between 8 Å and 2.1 Å resolution. An example of the $2F_o–F_c$ electron-density map, calculated after the refinement was completed, is shown in Figure 1. The model, which includes 120 water molecules, shows good stereochemistry by PROCHECK analysis [21] and only two non-glycine residues, Asn160 and Gln11, fall outside favourable regions of the Ramachandran plot. Gln11 is part of a classical γ turn [22], with (ϕ,ψ) angles of (69°,–48°) and a hydrogen bond (2.8 Å) between the carbonyl oxygen of Glu10 and the amide nitrogen of Val12.

Using the SAP2 coordinates from the final model, it was relatively straightforward to solve the orthorhombic SAP2–pepstatin crystal structure by molecular replacement. Despite the limited data, a difference map showed good density for the bound pepstatin molecule.

The structure of SAP2

The tetragonal and orthorhombic crystal forms reveal the protein in almost identical conformations, despite the different crystal packing contacts and the different inhibitors that are bound. The root mean square (rms) difference between all 342 pairs of Cα atoms is 0.58 Å. SAP2 possesses an overall framework that, not unexpectedly, is

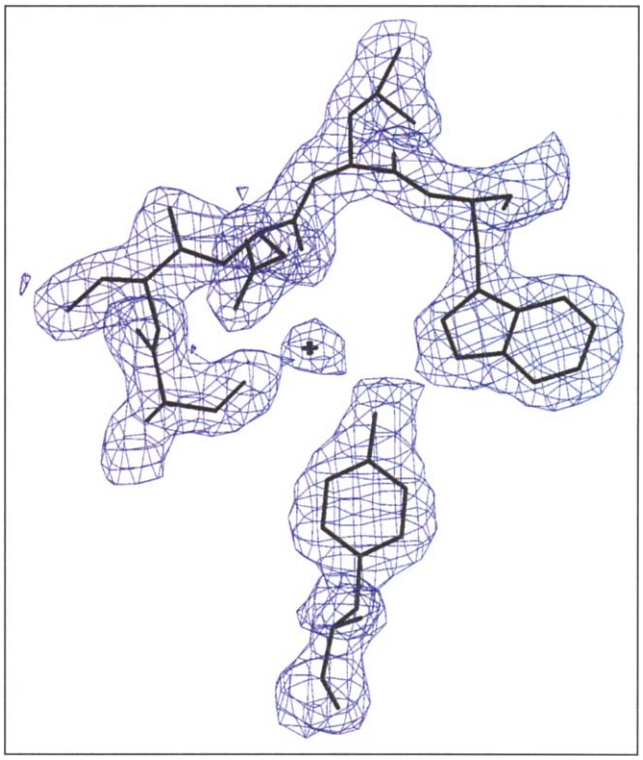

Fig. 1. Part of the final $2F_o–F_c$ electron-density map, at 2.1 Å resolution (contoured at 1σ). The map shows the interaction between Ser35, a well-ordered water molecule, Trp39 and Tyr84. (Figure drawn using O/OPLOT [34].)

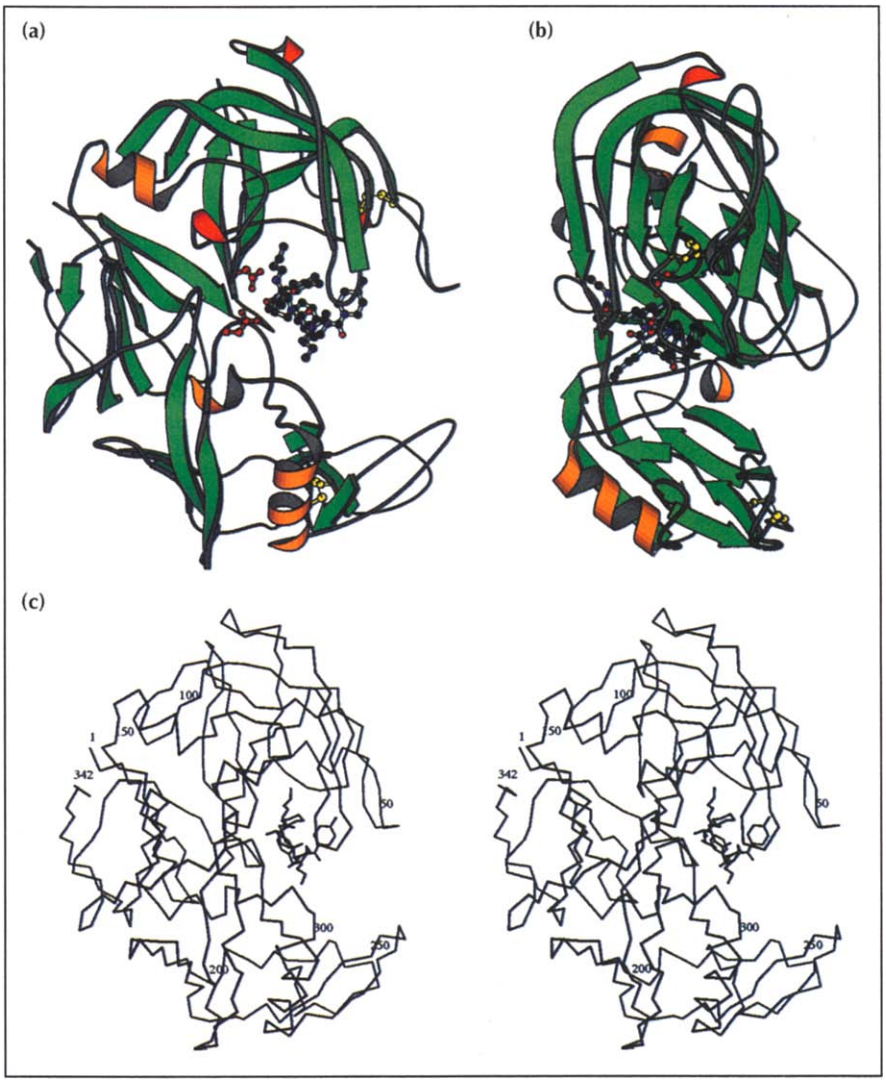

Fig. 2. Chain fold of SAP2 from *Candida albicans*. **(a),(b)** A ribbon presentation of SAP2–A70450 showing two orthogonal views. Active-site aspartates are highlighted in red; disulphides are shown in yellow. The C-terminal structural domain is identifiable in the lower third of (b). **(c)** Stereo Cα trace of SAP2–A70450. (Figure generated using MOLSCRIPT [47].)

typical of the monomeric aspartic-proteinase family. Such a fold can be described as a bi-lobed, mainly β structure, with a central binding cleft that can accommodate up to nine residues of peptide substrate. Figure 2 shows two representations of the SAP2 structure complexed to inhibitor A70450. Each domain provides one of the catalytic aspartate residues (Fig. 2a) that, in SAP2, are contained in DTG and DSG motifs, respectively. This mixed pairing is also seen in some other yeast and plant aspartic proteinases, whereas mammalian and other fungal enzymes have the DTG motif in both domains. The dimeric retroviral members possess either DSG or DTG, paired symmetrically. Within the C-terminal domain there is a structurally distinct subdomain of about 100 residues (see Fig. 2b), that in some other aspartic proteinases has been shown to behave as a separate rigid body [19,20]. This domain in SAP2 comprises residues 197–215 and 223–305, and contains one disulphide bond (between 256 and 294) which ties together a double loop (243–255, 282–293) of random structure (Fig. 2c). These C-terminal loops are a peripheral feature of aspartic proteinases, forming part of the wide entrance to the binding site. They are slightly different in the two crystal forms and are the most highly mobile regions of the molecule. Another loop, formed by the disulphide 47–59, is more well defined. It also flanks the binding site but is considerably closer (Fig. 2a,c).

The pattern of secondary structure elements in SAP2, in particular the organisation of the many β strands and the two major α-helical sections (140–146, 228–237), is similar to other known aspartic proteinases. However, some of the turns linking these elements are different in conformation. (Major differences are outlined in the next section.)

A highly conserved feature of the aspartic proteinases is the 'flap' region (a β hairpin loop) which interacts centrally with bound inhibitor/substrate, shielding the active site from bulk solvent. In SAP2 the flap comprises residues Lys81–Gln91, with the tyrosine at position 84 being highly conserved, as found in other aspartic proteinases (equivalent to Tyr75 in pepsin). There is a difference in conformation between the two SAP2 structures around Asp86–Gly87, due to the different inhibitors bound.

Comparison with other aspartic proteinases

The pepsin-like monomeric aspartic-proteinase family is very well-characterised crystallographically and most of

these proteins have also been studied as complexes bound to a variety of transition-state analogues. Known structures include the fungal members rhizopuspepsin [23], endothiapepsin [24,25], penicillopepsin [26], mucorpepsin [27] and yeast proteinase A [28], and the mammalian representatives pepsin [29], chymosin [30], cathepsin D [31,32] and renin [33]. (References quoted relate to enzyme–inhibitor complexes, in most cases, and are representative, not comprehensive.) It is of considerable interest to compare the SAP2 structure with both of these groups, firstly to examine the diversity amongst aspartic proteinases in general and the fungal group in particular, and secondly to evaluate features exhibited in the substrate/inhibitor-binding site of SAP2 that might distinguish it from the binding sites of the human homologues.

The SAP2 structure most closely resembles rhizopuspepsin, with an rms difference between 295 pairs of Cα atoms of 1.54 Å; the next closest are pepsin and chymosin (about 1.7 Å), then endothiapepsin, penicillopepsin, and renin (~1.8 Å). Superpositions were carried out using the least-squares procedure in O [34]. Figure 3 shows backbone superpositions of SAP2 with three fungal and three human enzymes, the structures of which had been determined with a bound inhibitor present and for which atomic coordinates were available through the Brookhaven Protein Data Bank [35]. It is immediately apparent that the N-terminal domains align much better than the C-terminal domains. Figure 4 shows a sequence comparison of these proteinases, based largely on alignment of corresponding secondary structure elements. Seven regions of difference highlighted in this analysis involve additions or deletions, some of which influence the structure of the substrate/inhibitor-binding site.

Region I

A small deletion near the N terminus results in a foreshortening of the turn connecting the β-strand residues 3–9 and 15–20. The two strands are part of different sheet structures and are virtually superimposable amongst family members. A classical γ turn in SAP2 (involving residues 10–12) allows these two strands to maintain the same relative positions, while at the same time it protrudes less into the binding-site region (Fig. 3), which is particularly noticeable in comparison with the other fungal enzymes.

Region II

The disulphide loop enclosed by Cys47 and Cys59 is significantly longer than in the homologous proteins under comparison. Indeed, neither endothiapepsin nor penicillopepsin has an equivalent disulphide, whereas

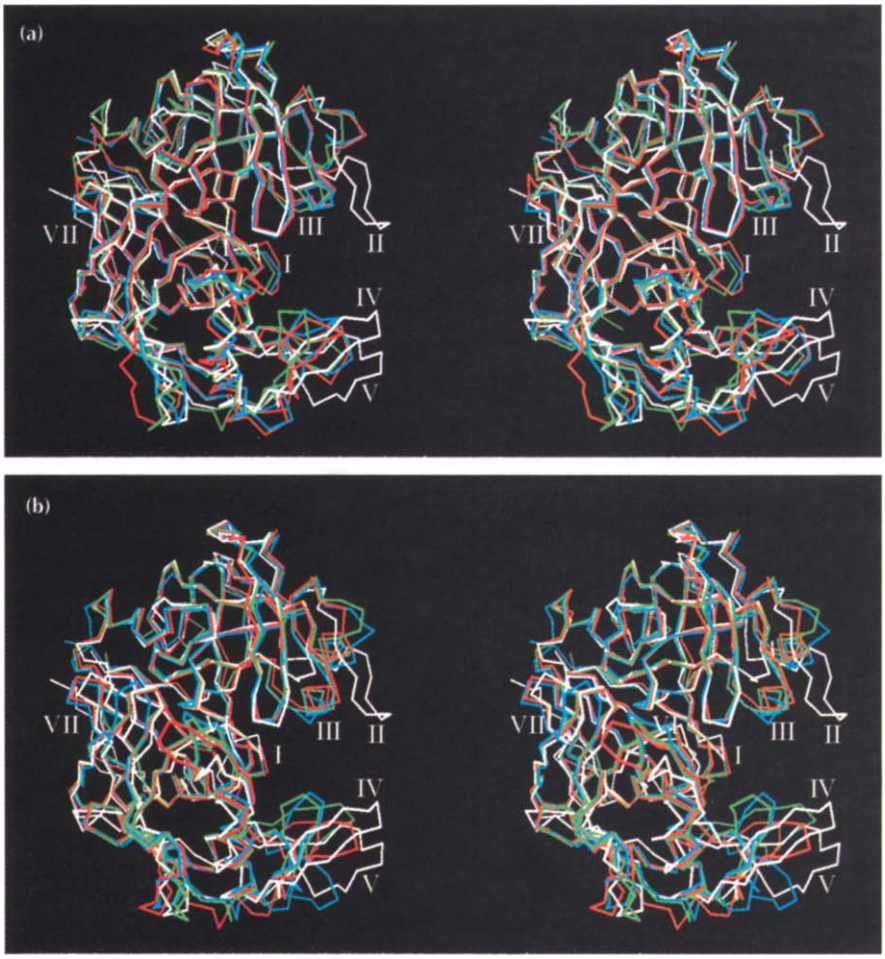

Fig. 3. Structural comparisons of aspartic proteinases. **(a)** Stereoview of the backbone superpositions of SAP2 (white) with three fungal aspartic proteinases: rhizopuspepsin (6APR;green), endothiapepsin (4ER2;red) and penicillopepsin (1APT;cyan). The roman numerals indicate regions where SAP2 differs from the other enzymes and these correspond to the sequences highlighted in Figure 4. **(b)** Similar comparisons of SAP2 (white) with three human aspartic proteinases: pepsin (1PSO;red), renin (1RNE;cyan) and cathepsin D (1LYB;green). (Figure generated using MOLSCRIPT [47].)

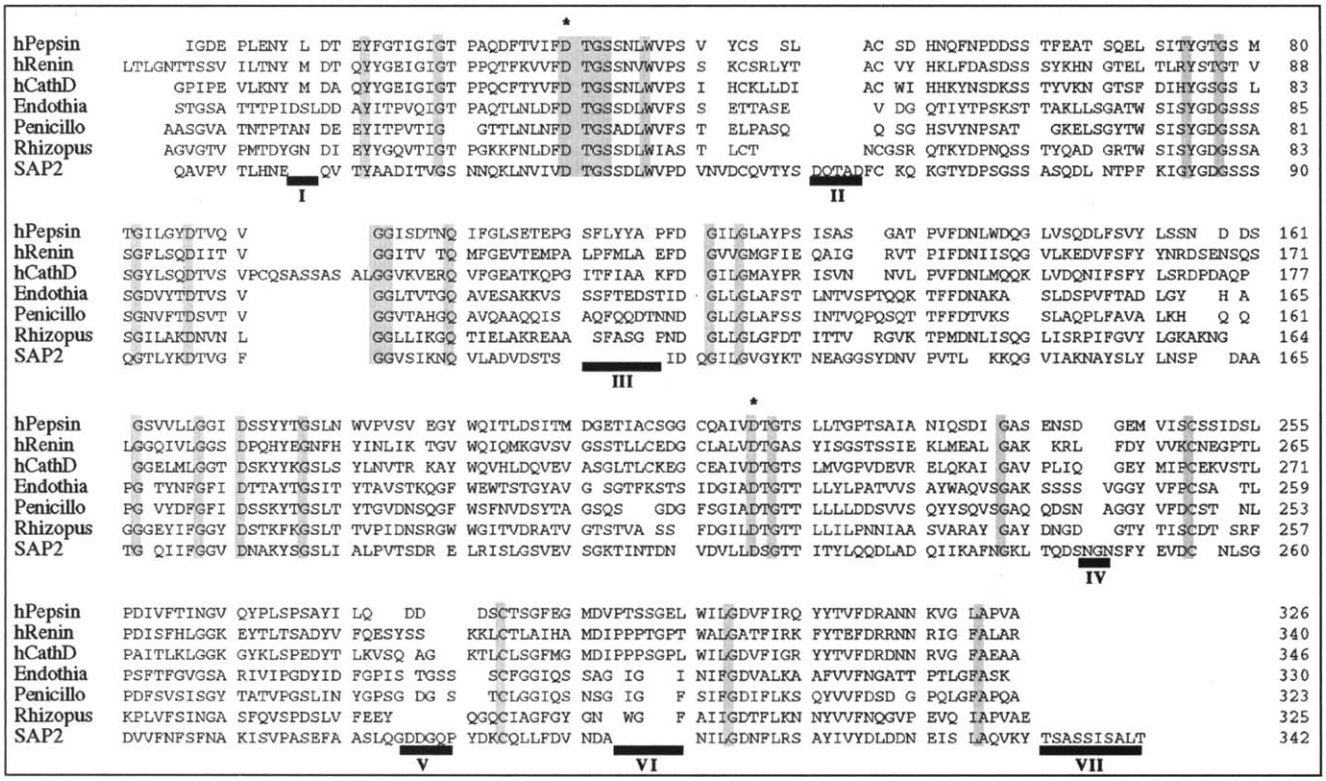

Fig. 4. Structure-based sequence alignment of SAP2 from *Candida albicans* 10261, rhizopuspepsin, penicillopepsin, endothiapepsin, human pepsin, human renin, and human liver cathepsin D. Regions indicated by a bar are highlighted in the tertiary structure comparisons of Figure 3. The two active-site aspartates, Asp32 and Asp218, are indicated by asterisks.

rhizopuspepsin has just two residues enclosed, pepsin four (between Cys45 and Cys50 using pepsin numbering), and renin and cathepsin D six each. The result is that this loop now forms a new boundary for the extended binding cleft of SAP2, as seen clearly in Figures 2 and 3.

Region III

Located to the inside of the disulphide loop (47–59) is a significant deletion centred about residue 118 which makes an even greater impact on the shape of the binding cleft. The absence of the 7–8 residues possessed by the other enzymes means the loss of a small helical segment (110–116 using pepsin numbering) that normally projects into the cleft and contributes to the S1 binding site (to use the standard notation [36]).

Region IV

A short insertion, corresponding to residues 247–249, is accommodated within an existing large loop, found well away from the active centre. This is a region of considerable flexibility, characterised by high B factors.

Region V

A similar insertion at positions 286–290 enlarges the adjacent loop, both loops being connected to the disulphide Cys256–Cys294. These two loops project out into the solvent further than in other aspartic proteinases and may be of antigenic significance, as may be the disulphide loop in the N-terminal domain.

Region VI

There is a significant deletion of 4–7 residues at SAP2 position 303, corresponding to the loss of a short looped segment that projects from the surface, markedly so in the case of the mammalian homologues (Fig. 3). This corresponds again to an alteration in the boundary of the extended binding site.

Region VII

A C-terminal extension of about ten residues is nicely stitched to the side of the N-terminal domain, residues 337–339 forming a small antiparallel β sheet structure with residues 174–176. The C terminus in SAP2 is in close proximity to the N terminus (a Cα–Cα separation of 8.4 Å) in contrast with the large separations seen in the other aspartic proteinases. The closest approach between the ends of the chain (4.8 Å) is made by residues 2 and 340.

Figure 4 also highlights the variability in aspartic proteinase sequences, not just in the external residues but also in the hydrophobic core. Conserved residues include those at the catalytic centre, the C-terminal domain cystine, the tyrosine in the flap, 14 glycines, and one tryptophan. SAP2 possesses only this one tryptophan residue (see Trp39 in Fig. 1). The active site of SAP2 contains the contributing sequences DTGS (32–35) and DSGT (218–221) with the serine at position 219 being unusual amongst monomeric

aspartic proteinase structures that are currently deposited in the Brookhaven Protein Data Bank [35]. However this substitution does not influence the geometry of the catalytic centre.

Inhibitor binding

Pepstatin is the archetypal aspartic-proteinase inhibitor and has been studied crystallographically in complex with rhizopuspepsin [26], cathepsin D [31,32], endothiapepsin [24] and pepsin (among the monomeric aspartic proteinases). It is an extended hexapeptide which is thought to act as a tetrahedral transition-state mimic by virtue of a central statine residue containing the non-scissile bond -(CHOH-CH_2)-. The hydroxyl group forms hydrogen bonds with the two catalytic aspartate side chains (Fig. 5). The insertion of these two extra carbon atoms in the backbone results in the P′1 side chain (alanyl) being displaced towards the S′2 pocket in the enzyme (using standard nomenclature). Inhibitor A70450, a renin-based inhibitor, is a pseudo-hexapeptide (Fig. 5) in which the peptide bond to be hydrolysed is replaced by the hydroxyethylene isostere, a common strategy in the design of aspartic proteinase inhibitors. Other noteworthy features of A70450 include a lactam ring that severely restricts the conformation around P3, the benzyl side chain of P3 having the (*R*)-configuration, a piperazine substituent at P4, and an *n*-butyl side chain at P2 [18]. The electron density for A70450 was well defined at 2.1 Å resolution (Fig. 6a). Although the SAP2–pepstatin complex was determined only at medium resolution (3 Å) and should therefore not be over-interpreted, the density for the pepstatin molecule was quite clear (Fig. 6b) and the final model is of good quality.

The pepstatin molecule binds in an open conformation, with an inverse γ turn between the CO and NH of the two statine residues, similar to that observed in other pepstatin-protein complexes. It also forms a very similar set of hydrogen bonds to the protein (Fig. 5). The involvement of active-site residues Asp32 and Gly34 of one loop and Asp218, Gly220 and Thr222 in the other, together with flap residues Gly85 and Asp86, is a common theme as these residues are highly conserved, especially amongst the fungal members. Where differences exist in the mammalian enzymes (serine for Thr222, serine or threonine for Asp86) corresponding hydrogen bonds are still made. Outside these eleven 'core' hydrogen bonds there are many van der Waals contacts, involving the residues listed in Table 1. Again, most of these residues or their equivalents, are found in the binding sites of other aspartic proteinases. Thus, it would appear that SAP2, despite its many differences from other related enzymes, binds pepstatin A in a conservative way, albeit less tightly than does

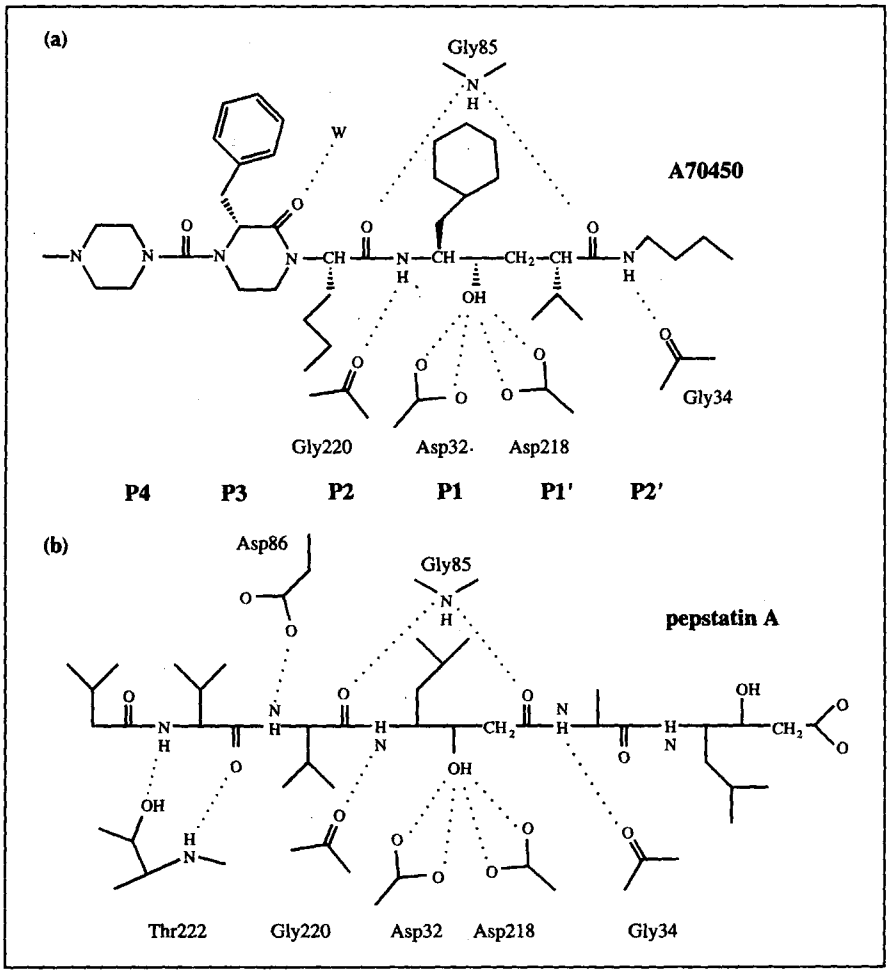

Fig. 5. Hydrogen-bonding diagrams. **(a)** A70450 bound to SAP2. **(b)** Pepstatin bound to SAP2. Distances of less than 3.5 Å, between electronegative atoms, are indicated.

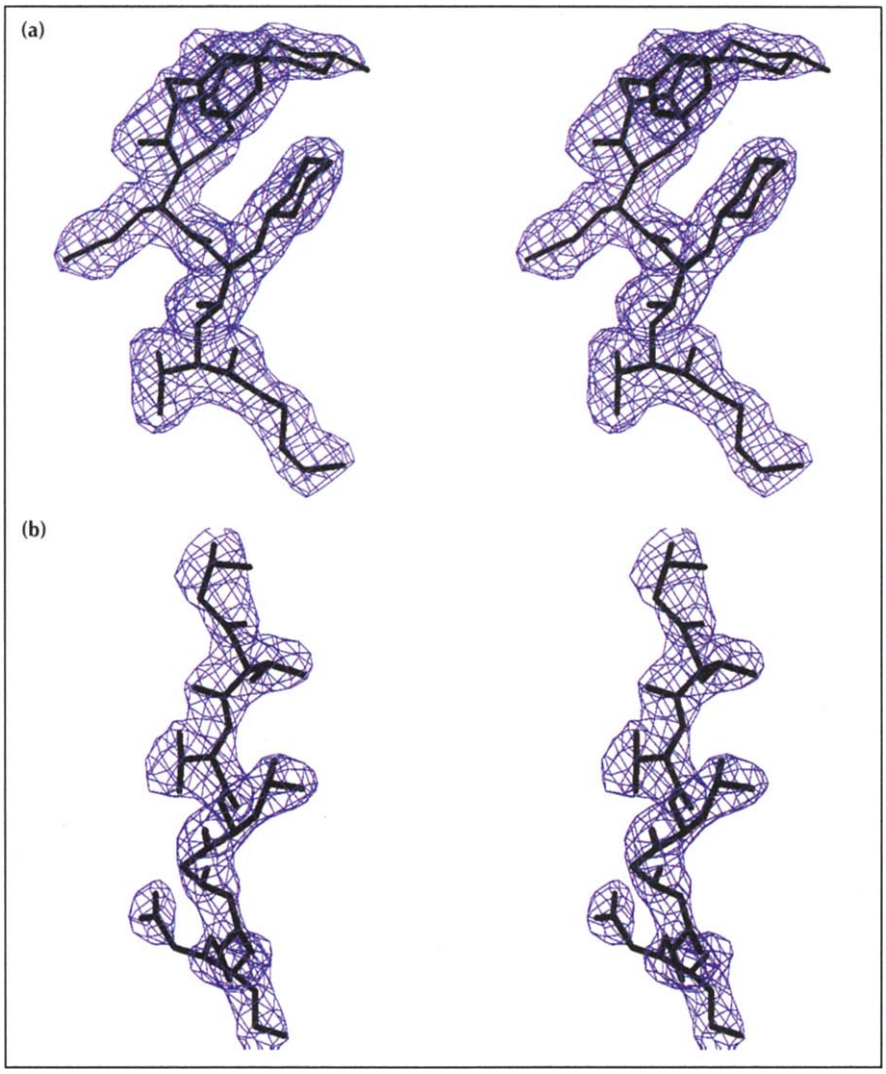

Fig. 6. Difference (F_o–F_c) electron density for inhibitors bound to SAP2. **(a)** A70450, using data between 20.0 Å and 2.1 Å. **(b)** Pepstatin, using data between 20.0 Å and 3.0 Å. The contour level is at 3σ in both cases. (Figure drawn using O/OPLOT [34].)

Table 1. SAP2 residues in contact (<4 Å) with inhibitors.

Subsite	Inhibitor	
	A70450	Pepstatin
S4	D86, S88, *S118*	*V12*, T222
S3	*V12*, *T13*, D86, S118, G220, T221, T222	*V12*, D86 G220, T221, T222
S2	Y84, *G85*, D86, T221 *N301*, *A303*, I305	Y84, *G85*, D86 G220, T221, Y225
S1	*I30*, D32, G34, Y84 *I119*, I123, D218, G220, T221	*I30*, D32, G34, Y84, *G85* I123
S'1	G34, Y84, *G85* *E193*, L216, D218	D218, G220, T221
S'2	G34, S35, I82, Y84 *N131*, A133	G34, S35, I82 *G83*, Y84
S'3		*E193*

Italicised residues are those which vary among the SAP 1–6 isozymes.

rhizopuspepsin (by a factor of 19), penicillopepsin (×19), endothiapepsin (×6) and considerably less so than pepsin (×63) or cathepsin D (×760), according to published K_i values [18,24,26,32]. There is no hydrogen bond to the carbonyl oxygen of P4, nor to the carbonyl oxygen of the alanine residue P'1 (P'2), as seen with cathepsin D [32]. In comparison with endothiapepsin there are fewer stabilising interactions to the C-terminal statine at P'2 (P'3) and to P3 valine [24]. The number of contacts made, or not made, because of sub-optimal filling of pockets, and the conformational determinants of these pockets, will all contribute to the strength of the enzyme–inhibitor interaction, but overall there seems to be no simple explanation for the observed differences in binding affinity.

Inhibitor A70450, although it binds to SAP2 with a strength around 20-fold greater than that of pepstatin, forms two fewer hydrogen bonds (Fig. 5) due to the lactam ring linking the main-chain nitrogen atoms of P3 and P2. One of the bonds formed involves a well-ordered water molecule that is close (2.7 Å) to the carbonyl oxygen of P3, a feature that has also been observed in complexes with other aspartic proteinases, especially endothiapepsin [25]. The other hydrogen bonds are

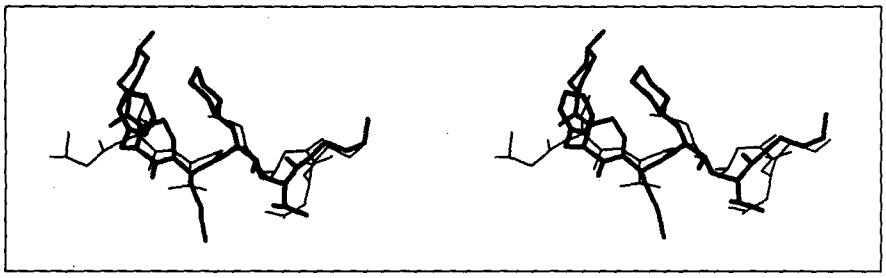

Fig. 7. Stereoview showing an overlay of the two inhibitors A70450 (thick line) and pepstatin (thin line) based on a superposition of the SAP2 structures as determined in each of the complexes. The orientation is consistent with the schematic representation in Figure 5.

highly conserved, with the shortest involving the interaction of an oxygen from each catalytic aspartate with the hydroxyl of the hydroxyethylene transition-state isostere. The number of contacts <4 Å from SAP2 to P′2 and to P3 is significantly greater than seen with pepstatin. Residue P3 possesses a side chain equivalent to that of D-phenylalanine and this aromatic ring forms close intermolecular interactions with the cyclohexyl group in P1 which folds back towards P3 and P4, resulting in a clustering of all four rings. The positioning of P4 and P3, that is, three of the rings, is controlled by rotation about the N-Cα bond of P2. As indicated in Table 1, the flap residues Ile82, Tyr84, Gly85, Asp86 and Ser88 make many contacts across the whole inhibitor. Indeed the tyrosine side chain slots between the P1 cyclohexyl ring and the *n*-butyl side chain at P′2.

The striking difference in conformation between A70450 and pepstatin is the curved backbone between P4 and P1 in the former, due to the lack of flexibility imposed by the lactam ring (Fig. 7). The methylpiperazine group at P4 is able to occupy most of the space made available by the SAP2 helix deletion at position 118, a situation not exploited by pepstatin with its greater flexibility. Some of the residues present in regions unique to SAP2, outlined in the previous section, are involved in defining the inhibitor-binding pockets (Table 1). These include Val12 (in the S3 subsite), Ile119 and Ile123 (S1), and Asn301 and Ala303 (S2). The last pair help stabilise the *n*-butyl side chain at P2 of A70450 but do not interact with the shorter valyl side chain in pepstatin.

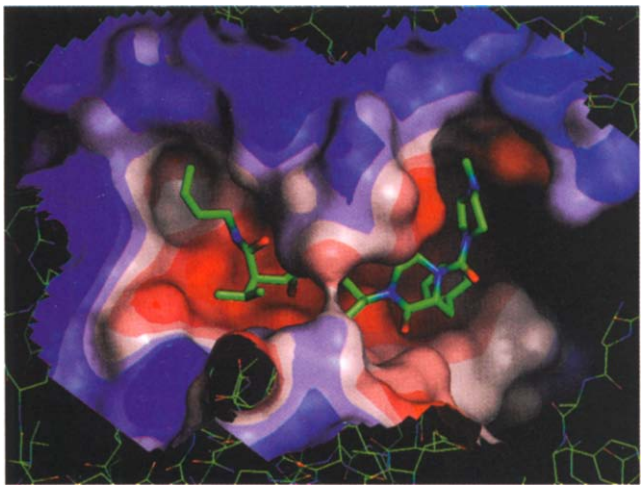

Fig. 8. Molecular surface of SAP2, with bound inhibitor A70450. The diagram is coloured to reflect electrostatic potential: negative charges in red, positive charges in blue. The inhibitor is oriented with P4 on the right and P′2 on the left. (Figure generated by MOLVIEWER [MJ Hartshorn, University of York].)

Modelling the A70450 structure, when bound to SAP2, into the renin structure results in a clash between the methylpiperazine (P4) and renin residue Pro111 of the helical region that is absent in SAP2 (region III). For A70450 to bind to renin in the same conformation, a cooperative movement of this helix and the nearby disulphide loop (region II) is required. Similar arguments apply to pepsin and cathepsin D.

Conclusions

The SAP2 structure displays some significant conformational as well as sequence differences from the other family members. It is too early to say if these differences are sufficient to allow a synthetic peptide-based inhibitor to select between *C. albicans* SAP2 (the putative virulence factor) and human host homologues such as pepsin, cathepsin D and renin. From the two crystallographic analyses presented here, there are some indications as to how one might design an inhibitor with higher specificity than A70450, for SAP2. Figure 8 shows a charged surface representation of the binding site. The topography suggests that an extension at P4 would result in a better fit to the hole around position 118 that is flanked by the disulphide loop (Fig. 8, right). The proximity of charged residues Asp86 (to the left of P4, as shown) and Asp120 (upper right) could also be considered in any new design. At the P′2 end there is a smaller hole that might be exploited, but clearly the structures of more complexes need to be analysed in order to properly assess conserved features of the binding site of SAP2.

It is worth noting that modelling studies indicate that a normal (flexible) peptide backbone would permit a putative inhibitor with an L-phenylalanine side chain at P3 to bind and interact with the enzyme at this site in a similar way to the D-phenylalanine of A70450, and still have P4 directed into the cavity.

Assuming that an inhibitor drug with the requisite pharmacological properties could be designed, there is still the problem of multiple isozyme expression by *C. albicans*, the control of which is not fully understood [14]. Would an inhibitor tailored to SAP2 be able to bind as well to the other isozymes? Of the eight *SAP* genes reported, one (*SAP7*) is probably not expressed [14] and

one (*SAP8*) has not yet been fully reported in the literature. Table 1 highlights the 12 subsite-defining residues (out of 27) that vary amongst the SAP 1–6 isozymes. Many of the changes involve conservative substitutions, and modelling indicates that these can be accommodated quite comfortably. Several substitutions that might cause some conformational rearrangement include: Ala303→Tyr (SAP3), Ala303→Asp (SAPs 4–6), Glu193→Lys (SAP5), Ile119→Ala (SAPs 4–6), Gly83→Glu (SAP3), Gly83→Lys (SAPs 4–6). Overall, SAP1 has the most similar binding cleft to SAP2 (only two differences); SAP3, which has the highest sequence identity (slightly greater than SAP1) with SAP2 shows a surprising eight changes, similar to the number seen for SAPs 4–6. SAP4 and SAP6, in fact, have identical binding sites, as defined in Table 1. It will be of interest to compare inhibitor-binding affinities for the different isozymes.

Biological implications

***Candida albicans* is the major fungal pathogen of humans, causing a variety of infections, both superficial (e.g. thrush) and systemic (that may be life-threatening in immunocompromised individuals). The increasing frequency of such infections, as well as the emergence of resistant strains, has prompted calls for new drug therapies. A possible drug target is the *Candida* sp. secreted aspartic proteinase (SAP) that enables the organism to use exogenous protein as a nitrogen source and that has long been cited as a putative virulence factor. Although it is now clear that there is a family of at least eight homologous *SAP* genes, one of these, *SAP2*, seems to be the major isoform expressed by *C. albicans*. Like many other aspartic proteinases, the SAP2 isoenzyme is not highly specific towards substrate and is inhibited by pepstatin, a microbial transition-state analogue. The synthetic hexapeptide analogue A70450 (Abbott Laboratories) is a stronger inhibitor, and the SAP2–A70450 complex can be readily crystallized. Under different conditions, the complex with pepstatin will also crystallise.**

The three-dimensional structures of the two enzyme–inhibitor complexes show similar protein conformations which are, in general organisation, typical for aspartic proteinases, that is, bi-lobed, mainly β sheet structures with a central groove. But several significant differences, especially in the region of the binding site, distinguish the SAP2 structure from structures of other aspartic proteinases. These changes include sequence insertions and deletions as well as sequence changes.

Whereas pepstatin binds in an extended conformation, A70450, by virtue of a lactam ring and an (*R*)-benzyl configuration at P3 (using standard notation), allows the P4 methylpiperazine to 'hook' into a pocket made by a sequence deletion.

Model building suggests that other SAP isozymes have similar binding sites. With the knowledge gained from these crystallographic analyses, an opportunity exists to design an inhibitor tailored to recognise the SAP isoenzymes in preference to host aspartic proteinases.

Materials and methods

SAP2–A70450 complex

The SAP2 isozyme from *C. albicans* (Strain ATCC 10261) was purified, and crystallized in the presence of the inhibitor A70450 (Abbott Laboratories) [18]. The crystals belong to the space group $P4_32_12$ (see below) with unit cell dimensions a=b=76.2 Å, c=126.1 Å, and one molecule in the asymmetric unit. The successful isomorphous derivative was prepared by soaking crystals of the complex in mother liquor containing 0.5 mM uranyl acetate for 24 h. Cell dimensions changed slightly as a result (a=75.7 Å, c=124.5 Å). Diffraction data were collected at room temperature using a Rigaku R-Axis IIC image plate detector in conjunction with a Rigaku RU-200HB rotating anode source, and processed using MOSFLM (AGW Leslie, LMB, Cambridge) and programs from the CCP4 suite [37]. Data to 2.05 Å and 2.45 Å resolution, respectively, were obtained for native and derivative crystals (see Table 2).

The structure was solved by a combination of single isomorphous replacement with anomalous scattering (SIRAS), and molecular replacement (MR). A high occupancy uranyl site was found by inspection of both the isomorphous and anomalous difference Patterson maps; three minor sites were located using VECSUM within the CCP4 package. The SIRAS phasing was based on these four uranyl sites, which were refined using the program MLPHARE (CCP4 package), then improved by solvent flattening and histogram matching using SQUASH [38,39]. It was noted that the solvent boundary was clearer when the phases were calculated for $P4_32_12$, as compared with $P4_12_12$. The final overall figure of merit for data in the range 20–2.5 Å was 0.58. Although the SIRAS/SQUASH phased electron density map looked promising, with elements of β sheet structure clearly visible, it was difficult to relate it to the known sequence of SAP2 due to discontinuities in the main chain. Eight other derivatives were screened in the hope of providing better phases but were not found to be useful.

Molecular replacement computations were performed using AMoRE [40] from the CCP4 implementation. Various structures of aspartic proteinases bound to inhibitors are available, and three of these were tried as search units: rhizopuspepsin (Brookhaven entry code 3APR), endothiapepsin (5ER2) and chymosin (4CMS). In addition, an homology model of SAP2 was constructed [17]. All four search models were aligned for ease of comparison. Rotation function solutions were not encouraging, with correlation coefficients all less than 10.0, and there was little consistency between searches done at different resolutions and between different models. However a systematic search for a translation function solution in both $P4_12_12$ and $P4_32_12$ was carried out using the top 40 rotation function solutions for each model. One solution (ninth highest peak) using the rhizopuspepsin model with 30–4 Å data in space group $P4_32_12$ gave an obviously 'best' solution with the highest correlation coefficient (23.3) and the lowest R factor (53.4%). These improved with rigid body fitting to a correlation coefficient of 26.7 and an R-factor of 52.0%. It is worth

Table 2. Data collection and phasing statistics.

	SAP2–A70450		SAP2–pepstatin
Derivative	Native	Uranyl acetate	Native
No. of unique reflections	24302	13453	6936
Redundancy	7.5	6.3	3.4
Completeness	98.0%	99.6%	99.9%
Resolution	2.05 Å	2.45 Å	3.0 Å
R_{merge}*	0.072	0.056	0.095
R_{iso}†, R_{ano}‡		0.32, 0.046	
Phasing power§			
acentric		1.2	
centric		0.9	
Cullis R#			
acentric		0.83	
centric		0.77	
anomalous		0.89	
Figure of merit			
acentric		0.36	
centric		0.48	

*$R_{merge}=\Sigma|I-\langle I\rangle|/\Sigma I$. †$R_{iso}$=fractional isomorphous difference, $\Sigma||F_{PH}|-|F_P||/\Sigma F_P$. ‡$R_{ano}$=fractional anomalous difference, $\Sigma|F_{PH}(+)-F_{PH}(-)|/\Sigma\langle F_{PH}\rangle$. §Phasing power=rms $\Sigma f_H/E$ where E is residual lack of closure error and f_H is the calculated heavy atom structure factor amplitude. #Cullis $R=\Sigma||F_{PH}\pm F_P|-f_H|/\Sigma|F_{PH}-F_P|$ and for anomalous data is calculated as anomalous lack of closure/anomalous difference.

noting that the correct rotation function solution was actually present in each set of the top 40 peaks.

The MR solution clearly matched the electron density of the 2.5 Å SIRAS/SQUASH phased map. Using the O program [34] the appropriate sequence changes were made and some reasonably straightforward rebuilding carried out. Partial model and experimental phases were combined using SIGMAA [42]. Some of the loop regions (47–63, 240–250, 280–290) were very poorly defined and initial refinement using the simulated annealing procedure in X-PLOR [43] did not reduce the R factor below 0.35 with R_{free} [44] above 0.50. Further rounds of model building into combined phase maps and restrained least squares refinement using PROLSQ [45] gradually improved the model, a breakthrough being the realisation that there was an out-of-register error involving the N-terminal region. When the R factor had dropped to 0.30, model-building continued using $2F_o-F_c$ maps with various ill-defined segments omitted, and most of the loop regions could then begin to be filled in. A difference map (F_o-F_c) clearly showed the density for the inhibitor, which was now included in the refinement (R=0.24 for all data between 8 Å and 2.1 Å). Water molecules were also located from difference maps and included if their density was above 3σ and if their positions were chemically sensible. The final model included all 342 residues (2560 non-hydrogen atoms), the inhibitor (53 atoms) and 120 water molecules, giving an R factor of 0.195 (R_{free}=0.268) for all data in the range 8.0–2.1Å (21780 reflections). The rms deviations from ideality (as per PROLSQ) are 0.013 Å (bond lengths), 0.032 Å (angle distances) and 0.033 Å (planarity). The average B value for protein atoms was 35 $Å^2$, consistent with the overall B estimated from a Wilson Plot [46], whereas for the inhibitor atoms the value was 27 $Å^2$. Several parts of the model are poorly defined with electron density less than 1σ in $2F_o-F_c$ maps, and with high B values. These are residues 246–249 and 283–287, both segments being contained in large loops within the C domain and also the side chain of Lys271.

SAP2–pepstatin complex

Orthorhombic crystals were grown by the hanging drop method from 0.1 M sodium cacodylate pH 6.5, 0.1 M zinc acetate, 16% PEG 8000. The space group was $P2_12_12_1$ with unit cell dimensions a=49.2 Å, b=67.1 Å, c=98.6 Å, and one molecule per asymmetric unit. Data collection and processing was as above except that DENZO [47] was used for indexing and integration of images. The crystals did not diffract much beyond 3 Å. Using the SAP2 coordinates, derived from the SAP2–A70450 structure as search unit, a molecular replacement solution using AMoRE was readily obtained. The highest peak in the rotation function was 9 times the rms deviation from the mean and more than twice the next peak, with a correlation coefficient of 25.3. When applied to the translation function a clear solution was obtained (correlation 58.7, R factor 0.42 which improved to 64.4 and 0.38 respectively after rigid-body refinement). Alternating model building (using O) into $2F_o-F_c$ OMIT maps with rounds of restrained least-squares refinement (PROLSQ) quickly reduced the R-factor. A difference map, calculated when the R factor was 0.23, revealed unambiguous density for the pepstatin molecule (47 atoms). Given the high ratio of parameters to data, B values were held constant at 30 $Å^2$ for most of the molecule and 50 $Å^2$ for the middle regions of the three largest loops. The final R factor was 0.196 for 8–3 Å data. No water molecules were included. Deviations from stereochemical ideality were 0.019 Å (bonds) and 0.046 Å (angle distances).

The coordinates of the SAP2–A70450 complex are being deposited with the Brookhaven Protein Data Bank.

Acknowledgements: This research was supported by the Health Research Council of New Zealand, the New Zealand Lottery Grants Board, and the Otago Medical School Bequest Funds. We are grateful to Ted Baker and Guy Dodson and members of their groups for assistance and use of equipment. The inhibitor A70450 (US Patent No. 5120718) was a gift from Abbott Laboratories, Illinois, USA.

Note added in proof

An independent and similar study of an inhibited SAP has been carried out by C Abad-Zapatero and colleagues (submitted for publication).

References

1. Sternberg, S. (1994). The emerging fungal threat. *Science* **266**, 1632–1634.
2. Odds, F.C. (1988). *Candida and candidosis: a review and bibliography.* (2nd edn), Cailliere and Tindall, London.
3. Cutler, J.E. (1991). Putative virulence factors of *Candida albicans. Annu. Rev. Microbiol.* **45**, 187–218.
4. Odds, F.C. (1994). Pathogenesis of Candida infections. *J. Am. Acad. Dermatology* **31**, (**suppl. 3**, part 2), S2–5.
5. Sobel, J.D., Muller, G. & Buckley, H.R. (1984). Critical role of germ-tube formation in the pathogenesis of candidal vaginitis. *Infect. Immun.* **44**, 576–580.
6. Soll, D.R., Morrow, B. & Srikantha, T. (1993). High frequency phenotypic switching in *Candida albicans. Trends Genet.* **9**, 61–65.
7. Ruchel, R., De Bernadis, F., Ray, T.L., Sullivan, P.A. & Cole, G.T. (1992). Candida acid proteinases. *J. Med. Vet. Mycol.* **30**, 123–132.
8. MacDonald, F. & Odds, F.C. (1980). Inducible proteinase of *Candida albicans* in diagnostic serology and in the pathogenesis of systemic candidosis. *J. Med. Microbiol.* **13**, 423–435.

9. Ray, T L. & Payne, C.D. (1990). Comparative production and rapid purification of *Candida* acid proteinase from protein-supplemented cultures. *Infect. Immun.* **58**, 508–514.
10. Ross, I.K., De Bernadis, F., Emerson, G.W., Cassone, A. & Sullivan, P.A. (1990). The secreted aspartate proteinase of *Candida albicans*: physiology of secretion and virulence of a proteinase-deficient mutant. *J. Gen. Microbiol.* **136**, 687–694.
11. Staib, F. (1965). Serum proteins as a nitrogen source for yeast-like fungi. *Sabouraudia* **4**, 187–193.
12. Negi, M., Tsuboi, R., Matsui, T. & Ogawa, H. (1984). Isolation and characterisation of proteinase from *Candida albicans*: substrate specificity. *J. Invest. Dermatol.* **83**, 32–36.
13. Monod, M., Togni, G., Hube, B. & Sanglard, D. (1994). Multiplicity of genes encoding secreted aspartic proteinase in *Candida* species. *Mol. Microbiol.* **13**, 357–368.
14. Hube, B., Monod, M., Schofield, D.A., Brown, A.J.P. & Gow, N.A.R. (1994). Expression of seven members of the gene family encoding secretory aspartyl proteinases in *Candida albicans*. *Mol. Microbiol.* **14**, 87–99.
15. Wright, R.J., Carne, A., Hieber, A.D., Lamont, I.L., Emerson, G.W. & Sullivan, P.A. (1992). A second gene for a secreted aspartate proteinase in *Candida albicans*. *J. Bacteriol.* **174**, 7848–7853.
16. Davies, D.R. (1990). The structure and function of the aspartic proteinases. *Annu. Rev. Biophys. Biophys. Chem.* **19**, 189–215.
17. Cutfield, S., Marshall, C., Moody, P., Sullivan, P. & Cutfield, J. (1993). Crystallization of inhibited aspartic proteinase from *Candida albicans*. *J. Mol. Biol.* **234**, 1266–1269.
18. Capobianco, J.O., Lerner, C.G. & Goldman, R.C. (1992). Application of a fluorogenic substrate in the assay of proteolytic activity and in the discovery of a potent inhibitor of *Candida albicans* aspartic proteinase. *Anal. Biochem.* **204**, 96–102.
19. Abad-Zapatero, C., Rydel, T.J. & Erickson, J. (1990). Revised 2.3 Å structure of porcine pepsin: evidence for a flexible sub-domain. *Proteins* **8**, 62–81.
20. Sali, A., Veerapandian, B., Cooper, J.B., Moss, D.S., Hofmann, T. & Blundell, T.L. (1992). Domain flexibility in aspartic proteinases. *Proteins* **12**, 158–170.
21. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291.
22. Nemethy, G. & Printz, M.P. (1972). The γ turn, a possible folded conformation of the polypeptide chain. Comparison with the β turn. *Macromolecules* **5**, 755–758.
23. Suguna, K., Padlan, E.A., Bott, R., Boger, J., Parris, K.D. & Davies, D.R. (1992). Structures of complexes of rhizopuspepsin with pepstatin and other statine-containing inhibitors. *Proteins* **13**, 195–205.
24. Bailey, D., *et al.*, & Szelke, M. (1993). X-ray crystallographic studies of complexes of pepstatin A and a statine-containing human renin inhibitor with endothiapepsin. *Biochem. J.* **289**, 363–371.
25. Bailey, D. & Cooper, J.B. (1994). A structural comparison of 21 inhibitor complexes of the aspartic proteinase from *Endothia parasitica*. *Protein Sci.* **3**, 2129–2143.
26. James, M.N.G., Sielecki, A., Salituro, F., Rich, D.H. & Hofmann, T. (1982). Conformational flexibility in the active sites of aspartyl proteinases revealed by a pepstatin fragment binding to penicillopepsin. *Proc. Natl. Acad. Sci. USA* **79**, 6137–6141.
27. Newman, M., *et al.*, & Blundell, T.L. (1993). X-ray analyses of aspartic proteinases. V. Structure and refinement at 2.0 Å resolution of the aspartic proteinase from *Mucor pusillus*. *J. Mol. Biol.* **230**, 260–283.
28. Dhanaraj, V., Dealwis, C., Bailey, D., Cooper, J.B. & Blundell, T.L. (1993). The three-dimensional structures of inhibitor complexes of monomeric aspartic proteinases. In *Innovations in Proteases and their Inhibitors*. (Avités, F.X., ed), pp. 141–159; Walter de Gruyter, Berlin.
29. Chen, L., *et al.*, & Abad-Zapatero, C. (1992). Structure of a pepsin/renin inhibitor complex reveals a novel crystal packing induced by minor chemical alterations in the inhibitor. *Acta Cryst. B* **48**, 476–487.
30. Gilliland, G.L., Winborne, E.L., Nachman, J. & Wlodawer, A. (1990). The three-dimensional structure of recombinant bovine chymosin at 2.3 Angstroms resolution. *Proteins* **8**, 82–101.
31. Metcalf, P. & Fusek, M. (1993). Two crystal structures for cathepsin D: the lysosomal targeting signal and active site. *EMBO J.* **12**, 1293–1302.
32. Baldwin, E.T., *et al.*, & Erickson, J.W. (1993). Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design. *Proc. Natl. Acad. Sci. USA* **90**, 6796–6800.
33. Dhanaraj, V., *et al.*, & Hoover, D.J. (1992). X-ray analyses of peptide-inhibitor complexes define the structural basis of specificity for human and mouse renins. *Nature* **357**, 466–472.
34. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building models in electron density maps and the location of errors in these models. *Acta Cryst. A.* **47**, 110–119.
35. Bernstein, F.C., *et al.*, & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.
36. Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157–162.
37. Collaborative Computational Project, Number 4. (1994). The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst. D* **50**, 760–763.
38. Zang, K.Y.J. & Main, P. (1990). The use of Sayre's equation with solvent flattening and histogram matching for phase extension and refinement of protein structures. *Acta Cryst. A* **47**, 110–119.
39. Cowtan, K.D. & Main, P. (1993). Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. *Acta Cryst. D* **49**, 148–157.
40. Navaza, J. (1992). AMoRE: a new package for molecular replacement. In *Molecular Replacement*. (Dodson, E.J., Gover, S. & Wolfe, W., eds), pp. 87–90, SERC Daresbury Laboratory, Warrington, UK.
41. Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst. A* **42**, 140–149.
42. Brünger, A.T. (1992). *X-PLOR Manual, Version 3.1*, Yale University, New Haven, CT, USA.
43. Brünger, A.T. (1993). Assessment of phase accuracy by cross validation: the free R value. Methods and application. *Acta Cryst. D* **49**, 24–36.
44. Konnert, J.H. & Hendrickson, W.A. (1980). A restrained-parameter thermal-factor refinement procedure. *Acta Cryst. A* **36**, 344–350.
45. Wilson, A.J.C. (1942). Determination of absolute from relative X-ray intensity data. *Nature* **150**, 151–152.
46. Otwinowski, Z. (1993). *DENZO: an oscillation data processing program for macromolecular crystallography*, Yale University, New Haven, CT.
47. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950.

Received: **18 Sep 1995**; revisions requested: **27 Sep 1995**; revisions received: **6 Oct 1995**. Accepted: **6 Oct 1995**.