Structure of a secreted aspartic protease from *C. albicans* complexed with a potent inhibitor: Implications for the design of antifungal agents

CELE ABAD-ZAPATERO,¹ ROBERT GOLDMAN,² STEVEN W. MUCHMORE,¹
CHARLES HUTCHINS,³ KENT STEWART,³ JORGE NAVAZA,⁴
CANDIA D. PAYNE,⁵ AND THOMAS L. RAY⁵

¹ Laboratory of Protein Crystallography, D-46Y, AP-10, L-07, Abbott Laboratories, Abbott Park, Illinois 60064-3500

² Anti-infective Group, D-47M, AP-9A, Abbott Laboratories, Abbott Park, Illinois 60064-3500

³ Molecular Modeling Group, D-46Y, AP-10, Abbott Laboratories, Abbott Park, Illinois 60064-3500

⁴ Laboratoire de Physique (UPR 180, CNRS), Chatenay-Malabry, 92290, France

⁵ Department of Dermatology, University of Iowa College of Medicine, Iowa City, Iowa 52242

(RECEIVED November 14, 1995; ACCEPTED January 17, 1996)

Abstract

The three-dimensional structure of a secreted aspartic protease from *Candida albicans* complexed with a potent inhibitor reveals variations on the classical aspartic protease theme that dramatically alter the specificity of this class of enzymes. The structure presents: (1) an 8-residue insertion near the first disulfide (Cys 45–Cys 50, pepsin numbering) that results in a broad flap extending toward the active site; (2) a 7-residue deletion replacing helix h_{N2} (Ser 110–Tyr 114), which enlarges the S₃ pocket; (3) a short polar connection between the two rigid body domains that alters their relative orientation and provides certain specificity; and (4) an ordered 11-residue addition at the carboxy terminus. The inhibitor binds in an extended conformation and presents a branched structure at the P3 position. The implications of these findings for the design of potent antifungal agents are discussed.

Keywords: antifungal drug design; antifungal targets; aspartic proteinases; *C. albicans*; rational drug design

Aspartic proteinases are ubiquitous in nature and are involved in myriad commercial and biomedical processes of interest (Kostka, 1985; Davies, 1990). Although the HIV virus is the best known pathogen using an aspartic proteinase in a critical aspect of its life cycle, other infectious agents also use enzymes of the same class in different infectious processes.

Candida albicans is an opportunistic fungal pathogen associated with the rising incidence of serious, life-threatening fungal infections seen in immunocompromised and debilitated patients. Although the ability to cause disease is likely a complex process involving multiple interactions between *Candida* and the host, most studies have focused on the role of a secreted aspartyl (acid) protease (SAP), with unusually broad substrate specificity, as a virulence factor. Evidence in support of a role of SAP in virulence is as follows: (1) correlations exist between virulence and level of protease expression; (2) protease-deficient mutants are less virulent or avirulent; (3) both SAP and anti-

body to SAP are detected in infected hosts; and (4) SAP can degrade a broad range of substrates, including proteins related to immunological and structural defenses, such as IgG heavy chains, keratin, acidified collagen, and extracellular matrix proteins (Douglas, 1988; Ray et al., 1990; Cutler, 1991; Ruchel et al., 1992; White et al., 1995). Initially, *Candida* strains were believed to express a single SAP. Subsequently, the *C. albicans* genome has been shown to contain at least seven distinct genes (SAP1–7) constituting a gene family (Ray & Payne, 1990; Hube et al., 1991, 1994; Wright et al., 1992; Magee et al., 1993; White et al., 1993; Miyasaki et al., 1994; Monod et al., 1994), plus at least one other unrelated aspartyl proteinase (Lott et al., 1989). All contain pre-propeptide sequences with variable numbers of kex2 cleavage sites, an N-domain active site triplet DTG common to all aspartic proteinases, and a C-domain active site triplet DSG. The seven known *C. albicans* SAP genes fall into two sub-families represented by SAPs 1, 2, and 3, and SAPs 4, 5, and 6, with SAP 7 being the most divergent in sequence (Monod et al., 1994). The multiplicity and differential regulation of SAP genes suggests that this gene family plays an important role in the pathophysiology of *Candida*, with differential expressions during growth states, morphologic, and phenotypic transitions,

Reprint requests to: Cele Abad-Zapatero, Department D-46Y, AP-10, LL-07, 100 Abbott Park Rd., Abbott Laboratories, Abbott Park, Illinois, 60064; e-mail: abad@abbott.com or abad@mozart.pprd.abbott.com.

stages of infection, and tissue sites during infection. Regulation of SAPs by phenotypic switching (White et al., 1993; White & Agabian, 1995) and serum/hypha formation (Hube et al., 1994; White & Agabian, 1995) are particularly relevant to pathogenesis. Induction of SAP may involve a signal transduction event initiated by contact sensing of extracellular protein, with specific sequence requirements (Lerner & Goldman, 1993).

Most biochemical studies relate to SAP2, which is the major SAP produced in vitro during growth on protein as the sole nitrogen source. SAP2 is a single polypeptide chain (342 residues) with a deduced molecular mass of 35,880 Da (Wright et al., 1992). Although SAP2 has an optimal activity range of pH 2.2–4.5, it is active at neutral pH on appropriate substrates (Wagner T, Borg-v Zepelin M, and Ruchel R, 1995; R. Goldman et al., in prep.). SAP2 is inhibited by pepstatin, but not by inhibitors of other non-aspartyl proteases. Pepstatin can modulate the course of *C. albicans* infection in the mouse model (Ruchel et al., 1990; Zotter et al., 1990), further implicating SAP as a virulence determinant. Pepstatin, however, is not an ideal agent for therapy in view of its lack of selectivity, potency, and safety.

We previously developed a sensitive fluorogenic substrate assay for SAP2 and discovered a potent inhibitor, A-70450, with a K_i of 0.17 nM (Capobianco et al., 1992). In this report, we describe the three-dimensional structure of a close homologue of SAP2 (referred to as SAP2X) complexed with the potent inhibitor A-70450 refined at 2.5 Å resolution from an orthorhombic crystal form. In our determination of the initial structure of the complex, we used a triclinic crystal form with eight molecules in the asymmetric unit related by general noncrystallographic symmetry operations. We analyze the novel structural features of the SAP gene family and discuss the structure of analogues of A-70450 with differential binding to host aspartyl proteases. The implications for the design of inhibitors for use in the therapy and prophylaxis of fungal infections are also presented.

Results and discussion

Amino acid sequence

The refined structure of the SAP2X/A-70450 complex is consistent with a polypeptide chain consisting of 341 residues (numbered 1–342; residue 251 is not present). As expected, the amino acid sequence is highly homologous (~96% identity) to the gene SAP2 of *Candida* (Wright et al., 1992). Eight single-amino acid replacements are well-defined in the electron density maps of the refined structure and can be reassigned with confidence: Val 45 → Ile 45, near the first disulfide (Fig. 1); Lys 81 → Ser on the enzyme surface; Ala 164 → Ser near the interdomain hinge; Glu 278 → Asp and Leu 283 → Thr on the well-defined portions of the variable loops; Asp 302 → Lys, Leu 327 → Ile, and Ile 338 → Thr near the carboxy terminus. In addition, residues Ala 177, Lys 206, Asn 257, and Ala 280 of the SAP2 sequence have been refined as Gly due to the lack of side-chain density beyond the C_α carbon. In relation to the SAP2 gene product, the most striking difference is the deletion of residue Phe 251. Although this residue is strictly conserved in the SAP1–3 genes, it is substituted by Ala or Val in SAP4–6 and absent in SAP7, the latest *Candida* protease to be characterized (Monod et al., 1994). The definitive amino acid sequence cannot be assigned with confidence in the regions of the carboxy lobe where the electron density is very weak and the polypeptide chain appears to be disordered, namely Ser 245–Asn 250 and Gln 284–Tyr 291; in these areas, the amino acid sequence of SAP2 from strain ATCC 10261 (Wright et al., 1992) has been assumed (Fig. 2).

Variations on the aspartic protease theme

The overall architecture of the SAP from *C. albicans* conforms with the classical aspartic protease fold typified by pepsin (Sie-

Fig. 1. Stereo diagram of the electron density map corresponding to the insertion at the first disulfide bond (Cys 47–Cys 59) of the SAP2X structure. The figure comprises residues Ile 45–Lys 60 with the corresponding electron density ($2F_o - F_c$; $\sigma = 1$ level). It illustrates the quality of the electron density and the rigidity of the conformation. Residues are labeled sequentially from Ile 45 to Lys 60. See Figure 3 for the relative disposition of this loop in relation to the rest of the structure.

PNO	1	7	17	27	32	37	45	PNO 184	193	201	206	210	215	223	233	
1PSA	IGDEPLE	NYLDTYEFGT	IGIGTFAQDF	TVIFDTGSSN	LWVPSVYCSS			1PSA	VSVVE.GXWQIT	LDSITMDG	ETIACSGGCQAI	VDTGTSLLTG	PTSAIANIQS			
SAP4	GPVAVKL	DNEIITYSAD	ITIGSNNQKL	SVIVDTGSSD	LWVPDSNAV	C		SAP4	ITSD	RTLSVG	LRSVNVMG	QNVNV..NAGVL	LDSGTTISYF	TPNIARSIIY		
SAP5	GPVAVTL	HNEAITTYAD	ITVGSNDQKL	NVIVDTGSSD	LWIPDSNVIC			SAP5	ITSE	KKLTVG	LRSVNVRG	RNVDA..NTNVL	LDSGTTISYF	TRSVIRNILY		
SAP6	GPVAVKL	DNEIITYSAD	ITVGSNNQKL	SVIVDTGSSD	LWIPDSKAI	C		SAP6	ITSD	RTLSVG	LRSVNVMG	RNVNV..NAGVL	LDSGTTISYF	TPSIARSIIY		
SNO	1	7	17	27	32	37	47	SNO 188	195	198	205	210	218	226	236	
SAP2X	QAVPVTI	HNEQVITYAD	ITVGSNNQKL	NVIVDTGSSD	LWVPDVNID	C		SAP2X	VTSD	RELRI	LGSVEVSG	GTINT.DNVDVL	LDSGTTITYL	QRDLADQIIK		
SAP1	QAIPTVL	NNELVSYAAD	ITIGSNNQKL	NVIVDTGSSD	LWVPDASVTC			SAP1	VTSD	RELRI	LNSLKAVG	KNIN..GNIDVL	LDSGTTITYL	QQDVAQDIID		
SAP2	QAVPVTI	HNEQVITYAD	ITVGSNNQKL	NVIVDTGSSD	LWVPDVNVDC	C		SAP2	VTSD	RELRI	LGSVEVSG	KTINT.DNVDVL	LDSGTTITYL	QQDLADQIIK		
SAP3	QTPVPKL	INEQVSYASD	ITVGSNNQKL	TVVIDTGSSD	LWVPDSQVSC			SAP3	VTSD	NELRIH	LNTVKVAG	QSINA.D.VDVL	LDSGTTITYL	QQGVADQVIS		
3APR	AGVGTVPMTD	YGNIDIEYYGQ	VTIGTFGKKF	NLDFDTGSSD	LWIASTLCTN			3APR	IDNSRGWVGIT	VDRATVGT	STVA	SS	FDGI	LDGTGTLIL	PNNIAASVAR	
PNO	48	50	58	68	77	87		PNO 234	241	249	260	270	279			
1PSA	LA.....	.CSDHNQFNP	DDSTFEATS	QELSITYG	T	GSMTGILGYD		1PSA	DIGASE..NS	DGEMVIS	C	SSIDSLPDI	FTINGVQYPL	SPSAIYLQDD		
SAP4	IPKWPGRGD	FKKNGSYSP	AASSTSKNLN	TPFEIKYADG	SVAQGNLYQD			SAP4	ALGGQVHYDS	SGNEAYVAD	C	KTSGTVDFQF	DRNLKISVPA	SEFLYQLYIT		
SAP5	IPKWRGDKGD	FKKSAGSYSP	ASSRTSQNLN	TRFDIKYGDG	SYAKGKLYKD			SAP5	AIGAQMKFDS	AGNKVYVAD	C	KTSGTIDFQF	GNNLKISVPA	SEFLFTQYIT		
SAP6	IPKWRGDCGD	FKKNGSYSP	AASSTSKNLN	TRFEIKYADG	SYAKGNLYQD			SAP6	ALGGQVHYDS	AGNKAYVAD	C	KTSGTVDFQF	DKNLKISVPA	SEFLYQLYIT		
SNO	48	57	59	67	77	87	97	SNO 237	246	256	266	276	286			
SAP2X	QVTYSQDQAD	FKQKGTYPD	SGSSASQDLN	TPFSIGYGDG	SSSQGTLYKD			SAP2X	AFGGKGTSDS	NGNS.YSVDC	C	NLSGDDVVFNF	SKNAKISVPA	SDFAASTQGD		
SAP1	DKPRPQGSAD	FKKGKGIYTP	KSTTSQNLG	SPFYIGYGDG	SSSQGTLYKD			SAP1	AFQAEKLKLDG	QGHTFYVTD	C	QTSQTVDFNF	DNNAKISVPA	SEFTAPLSYA		
SAP2	QVTYSQDQAD	FKQKGTYPD	SGSSASQDLN	TPFKIGYGDG	SSSQGTLYKD			SAP2	AFNGKLTQDS	NGNSFYVDC	C	NLSGDDVVFNF	SKNAKISVPA	SEFAASLQGD		
SAP3	QAG.QQQDPN	FKNEGTYSP	SSSSSSQNLN	SPFSIEYGDG	TTSQGTWYKD			SAP3	AFNGQETYDA	NGNLFYVDC	C	NLSGSDVDFAF	DKNAKISVPA	SEFTAPLYTE		
3APR		CGSGQTKYDP	NQSTYQADG	RTWSIYGDG	SSASGILAKD			3APR	AYGASDNGDG	TYTIS	C	DTSAFKPLVF	SINGASFQVS	.PDSL.V.F.E		
PNO	88	97	107	110	127	136		PNO	282	285	289	299	310	320	326	
1PSA	TVQVGGISDT	NQIFGLSETE	PGSFLYYAPF	DGILGLAYPS	ISA.SGATPV			1PSA	DS	CTS	GFEQMDVPTSSGELW	ILGDVFIRQY	YTVFDRANNK	VGLAPVA		
SAP4	TVGIGGVSVR	DQFLFANVRST	SAH.....	KGILGIGFQS	NEATRTPYDN			SAP4	NGEPYPKCEI	RVRES		EDN	ILGDNFMRSA	YIVYDLDDRK	ISMAQVK	
SAP5	TVGIGGVSVR	DQFLFANVWST	SAR.....	KGILGIGFQS	GEATEFDYDN			SAP5	SGKPFKCEV	RIRES		EDN	ILGDNFMRSA	YVYVNLDDKK	ISMAQVK	
SAP6	TVGIGGASVK	NQFLFANVWST	SAH.....	KGILGIGFQS	NEATRTPYDN			SAP6	NGKPYKCEI	RVRES		EDN	ILGDNFMRSA	YIVYDLDDKK	ISMAQVK	
SNO	98	107	117	121	130	140		SNO 287	294			304	314	324	331	
SAP2X	TVGFGGVSIK	NQVLADVST	SID.....	QGILGVGYKT	NEAGGSYDNN			SAP2X	DGQPYDKCQL	LFDVN		KAN	ILGDNFLRSA	YIVYDLDDNE	ISIAQVK	
SAP1	TVGFGGASIT	QKVFADITKT	SIP.....	QGILGIGYKT	NEAAGDYDNN			SAP1	NGQPYKPCQL	LLGIS		DAN	ILGDNFLRSA	YLVYDLDDDK	ISLAQVK	
SAP2	TVGFGGVSIK	NQVLADVST	SID.....	QGILGVGYKT	NEAGGSYDNN			SAP2	DGQPYDKCQL	LFDVN		DAN	ILGDNFLRSA	YIVYDLDDNE	ISLAQVK	
SAP3	TIGFGGISIT	KQGFADVST	SVD.....	QGILGIGYKT	HEAAGNYDNN			SAP3	DGQVYDQQL	LFQTS		DYN	ILGDNFLRSA	YIVYDLDDNE	ISLAQVK	
3APR	NVNLGGLLIK	GQTIELAKRE	AASFA.SGFN	DGLLGLGFTD	ITTVRGVKNP			3APR	EF	QQG	CIA	GFGYGNW	GFA	IIGDTFLKNN	YVFNQGVPE	VQIAPVAE
PNO	137	143	153	163	173	183		PNO				30%				
1PSA	FDNLWDQ	GLVSQDLFSV	YLSSNDSDGS	VVLGGIDSS	YTTGSLNMWVP			1PSA				30%				
SAP4	PITLKKQ	GIISKNAYSL	FLNSPEASSG	QIIFGGIDKA	KYSGSLVDLP			SAP4	YTSQSNIVGIN			59%				
SAP5	PISLRNQ	GIIGKAAYSL	YLNSAEASTG	QIIFGGIDKA	KYSGSLVDLP			SAP5	YTSSESDIVAIN			55%				
SAP6	PISLKKQ	GIITAKNAYSL	FLNSPEASSG	QIIFGGIDKA	KYSGSLVLEP			SAP6	YTSSESDIVAIN			60%				
SNO	141	147	157	167	177	187		SNO 332	342							
SAP2X	PVTLLKKQ	GVIKNAYSL	YLNSPDSATG	KIIFGGVDDG	KYSGSLIALP			SAP2X	YTSASSTSAIT			-				
SAP1	PVTLLKNQ	GVIKNAYSL	YLNSPNAATG	QIIFGGVDKA	KYSGSLIALP			SAP1	YTSASNIAALT			72%				
SAP2	PVTLLKKQ	GVIKNAYSL	YLNSPDAATG	QIIFGGVDNA	KYSGSLIALP			SAP2	YTSASSISALT			96%				
SAP3	PVTLLKNQ	GIISKNAYSL	YLNSRQATSG	QIIFGGVDNA	KYSGTLIALP			SAP3	YTSASNIAALT			70%				
3APR	MDNLISQ	GLISRPIFGV	YLGAKNNGG	EYIFGGYDST	KFKGSLTTFV			3APR				33%				

Fig. 2. Structural alignment of pepsin, rhizopuspepsin, and the SAPs from *C. albicans*. Identical colors are chosen to indicate the grouping of identical or similar residues in the two separate families SAP1–3 versus SAP4–6. PNO denotes the pepsin numbering system (residues renumbered 1–326 [Abad-Zapatero et al., 1990; Sielecki et al., 1990]). SNO denotes the sequential numbering system of the SAPs based on SAP2 (342 residues number 1–342; SAP2X lacks residue 251). Percentage numbers indicate percent identity with SAP2X.

lecki et al., 1990) and first described in detail by James and co-workers in the fungal enzyme penicillopepsin (Hsu et al., 1977), and by Blundell and collaborators for endothiapepsin from *Endothia parasitica* (Tang et al., 1978). The superposition of the three-dimensional structure of the liganded SAP2X from *C. albicans* with liganded pepsin (PDB entry 1PSA [Chen et al., 1992]) and the liganded structures of submaxillary gland renin (1SMR [Dealwis et al., 1994]), *E. parasitica* pepsin (1EED), rhizopuspepsin (3APR), and penicillopepsin (3APP) revealed that the closest structural homologue among the fungal enzymes is rhizopuspepsin (292 pairs, 1.48 Å RMS). Porcine pepsin is the closest structural relative within the mammalian proteases, with 283 C α pairs and an RMS of 1.70 Å (Table 1). As is the case with many other aspartic proteinases, the amino terminal lobe superimposes better (176 C α pairs, RMS 1.34 Å) than the carboxy lobe (116 pairs, RMS 1.60 Å after overall superposition). Although the RMS value is similar to the one obtained with pepsin in the amino-terminal domain (176 C α pairs, RMS 1.36 Å), the corresponding one for the carboxy domain (116 pairs, RMS 1.88 Å) is larger. The position of the C α carbons corresponding to the fully conserved disulfide bond among all monomeric aspartic proteinases (Cys 249–Cys 282, pepsin numbering in bold) is analogous in rhizopuspepsin and SAP2X. After over-

all superposition, the distance between Cys 256 (SAP2X) and Cys 253 (3APR) was 0.7 Å, and 1.7 Å between the structurally equivalent residues Cys 294 and Cys 285. The corresponding values when compared with pepsin were 1.7 and 4.1 Å respectively. However, the SAP2X structure presents several unique features in both the amino and carboxy lobes that put it into a different class among the monomeric proteinases.

Pro 23 is found in the *cis*-conformation in most of the aspartic proteinase structures refined thus far and was assumed to be an important element of the fold in the second β -hairpin of the structure. Although none of the members of the SAP family in *Candida* have proline at that position, the β -turn presents the same number of residues and the SAP2X structure follows very closely the classical β -hairpin (Figs. 2, 3A).

The classic pepsin fold presents three different disulfide bonds along the polypeptide chain. The first is in the amino terminal lobe between residues Cys 45 and Cys 50. This first disulfide is more common among the mammalian aspartic proteinases (renin, chymosin, and cathepsin D) than among the fungal enzymes, where neither endothiapepsin nor penicillopepsin exhibit it. The polypeptide chain contained within this first disulfide normally consists of a short helical or turn-like segment. In relation to pepsin, the amino acid sequence of the SAP2X from

Table 1. Overall superposition between several aspartic proteinase/inhibitor complexes^a

	ISMR	IPSA	SAP2X	3APR	IEED
ISMR	—	1.43 (317)	1.97 (279)	1.76 (301)	1.53 (244)
IPSA		—	1.68 (278)	1.55 (296)	1.56 (261)
SAP2X			—	1.49 (292)	1.73 (287)
3APR				—	1.32 (301)
IEED					—

^a Each entry of the table is the RMS deviation in Å between the structures denoted by the corresponding row and column of the matrix. The integer number in parenthesis specifies the number of C_α pairs resulting in superposition as described in the Materials and methods. Upper diagonal lists the results obtained from the overall superposition. IPSA, pepsin/A62095 complex (Chen et al., 1992); ISMR, mouse submaxillary gland renin (Dealwis et al., 1994); SAP2X, secreted aspartic protease from clinical isolate (this work); IEED, *Endothia parasitica*/inhibitor complex (Cooper et al., 1987); 3APR, *Rhizopus*/inhibitor complex (Gilliland et al., 1990).

C. albicans has a long insertion between Cys 47 and Cys 59 (Fig. 2) and the structure of this loop is completely novel among the known aspartic proteinases (Fig. 1 and Kinemage 1). The segment between Cys 47 and Cys 59 forms a well-defined, broad flap that extends toward the active site and defines a very wide entrance to the active site from the nonprime side (Fig. 3A,B). The aromatic ring of Phe 58 is in a hydrophobic environment next to the disulfide bond and is fully conserved among SAP1–6 (Fig. 2). In addition, the side chain of Gln 54 interacts with the main chain of residues Phe 51 (Asn 54 OE1...NPhe 51, 2.85 Å) and Thr 50 (Asn 54 OE1...NThr 50, 2.53 Å) to stabilize the final turn (Fig. 1). Therefore, Gln 54 may be critical for the protein to achieve this conformation. Gln 54 is not conserved in SAP4–6 (replaced by Asp, Fig. 2; see below), and this may indicate that the conformation of this loop is different for various members of the *Candida*-secreted aspartic protease gene family. The presence of this extended second flap, in addition to the well-recognized flap projecting over the active site (hairpin loop comprising residues 72–85 in pepsin), is probably also characteristic of other fungal aspartic proteinases having an insertion between the cysteine residues topologically equivalent to pepsin Cys 45–Cys 50. Among these are a variant of *C. tropicalis* (Monod et al., 1994) (SAPT1) and the product of gene BAR1, an aspartic proteinase with 24 amino acid residues between the above cysteines (MacKay et al., 1990).

A significant portion of the specificity displayed at the P3 position by several members of the aspartic protease family resides in the residues extending from helix h_{N2} (pepsin Ser 110–Tyr 114) (Dealwis et al., 1994). In the structure of the SAP2X from *Candida*, a 7-residue deletion in this region dramatically alters the enzyme pocket S3 in all the members of this secreted gene family (Monod et al., 1994) (Figs. 2, 3A,B). The predominant effect appears to be expansion of the physical size of this pocket, giving rise to a funnel-like active site cavity with the wide opening in the nonprime side of the enzyme (Kinemage 1).

Most members of the monomeric family of aspartic proteases have approximately 330 residues. Without exception, the carboxy termini of the different polypeptide chains ends in the proximity of the last Ala 326 residue in pepsin. The amino acid sequence of all the members of the SAP(1–6) gene family have 342 amino acid residues and the sequence alignment predicted

a carboxy-terminal extension 11 residues longer than pepsin that was difficult to build by homology modeling (Fig. 2; see the Materials and methods). In SAP2X, these additional residues wrap around the 176–182 region that connects the amino and carboxy domains, and is stabilized by hydrogen bonds in a β-sheet like conformation (Fig. 3A and Kinemage 1). The quality of the electron density, the corresponding temperature factors, as well as the regularity of the hydrogen bonds, suggest that this additional feature is rigid, at least in the two crystal forms that are the object of this study. The role that this extended tail plays in vivo is unknown at the present. It is conceivable that, because it is on the opposite side of the active site, it could possibly serve as an attachment point that will permit the enzyme to recognize certain surfaces or tissues, and still allow the enzyme unrestricted enzymatic activity.

The refined structure presents a well-ordered turn that wraps the carboxy terminus of the structure around the ridge of residues Gln 329–Ser 336 and later forms an additional antiparallel β-strand next to Gly 172–Asp 175, making hydrogen bonds with Ser 339 and Leu 341 of the SAP2X carboxy end. The structurally homologous residue to Pro 323 in pepsin is Gln 329 in SAP2X. The side chain of this residue makes hydrogen bonds with the main-chain atoms (329GlnOE1...N, 3.06 Å; 329GlnNE2...O, 3.05 Å) of Ser 273 and the hydroxyl group of Tyr 332 interacts with the carbonyl oxygen of Arg 312. Another contributing factor to the rigidity of the carboxy-terminal extension might be the presence of the strictly conserved positive charge of the residue at position 152 (Lys, Fig. 2), which is bracketed between Asp 322 (OD1...NE, 2.7 Å) and the carboxy terminus (Thr342OT...NE, 3.9 Å).

Relative subdomain orientation

The approximate relative orientation of the segment of residues between 190 and 301 (pepsin numbering) is an important element in the structural variability observed within the carboxy domain of the monomeric aspartic proteinases (Sali et al., 1989, 1992; Abad-Zapatero et al., 1990; Sielecki et al., 1990). After the rigid body domain 1 (RB1, 1–189 and 301–326) has been superimposed between a certain pair, the relative orientation of rigid body domain 2 can range widely among the different enzymes. Using mouse renin as a reference, the lowest values are found among the mammalian enzymes, ranging from 1.9° for the mouse–human renin pair to 7.2° for the mouse renin–chymosin comparison. Using the same reference, the largest values are obtained when comparing with the fungal enzymes: 11.3° with penicillopepsin, 15.0° in relation to rhizopuspepsin, and 19.5° against endothiapepsin (Dealwis, 1993). After superposing RB1 of SAP2X with the corresponding rigid body domain of its closest structural-homolog (rhizopuspepsin), the relative orientation of the remaining rigid body domains 2 differs by approximately 4.2° (Fig. 3). This value is comparable to the relative rigid body rotation found between porcine pepsin and mouse renin (5.6°), and is similar to the values observed when comparing several fungal enzymes (Dealwis, 1993; Newman et al., 1993).

Two main factors have been suggested to contribute to the relative disposition of the two different rigid body domains: the conformation of the chain segment 184–192, and the residues making up the interface (Newman et al., 1993; Dealwis et al., 1994). An additional contributing factor in SAP2X is probably the shorter loop connecting Phe 298 to Ile 305 in SAP2X when

Fig. 3. Stereo diagram of the polypeptide backbone fold for the SAP2X from *C. albicans* (green) superimposed to the RMS given in Table 1 with porcine pepsin (blue) (1PSA) and rhizopuspepsin (red) (3APR). **A:** View emphasizing the major differences between the three enzymes. **B:** View chosen to illustrate the position of A-70450 in relation to the active site and the "specificity ridge." N- and C-termini for SAP2X have been identified and residues numbers have been intercalated along the polypeptide chain.

compared with rhizopuspepsin (Figs. 2, 3B, 4). Even though the fungal protease rhizopuspepsin is the closest structural homologue, the conformation of the polypeptide chain from Phe 298 to Ile 305 is very different between SAP2X and 3APR. In pepsin, renin, and cathepsin, this loop structure presents an insertion that includes the proline-rich "flap" in renin, which is responsible for a large portion of the specificity toward angiotensinogen (Dealwis et al., 1994). In the SAP2X structure, this region forms a unique bend, probably due to the unusual con-

formation of the strictly conserved Asn 304 residue. The carbonyl oxygens of residues Ala 303 and Asn 304 interact with the guanidium group of Arg 195 ($\text{NH1} \cdots 303\text{O}$, 2.8 Å; $\text{NH2} \cdots 304\text{O}$, 2.9 Å), probably forcing a left-helical conformation ($\phi = 48^\circ$, $\psi = 74^\circ$) for Asn 304. In addition, Arg 195NE forms a hydrogen bond with Glu 193OE1 (3.0 Å). These two residues are conserved in the SAP1–3 sequences, but not in the SAP4–6, suggesting that the conformation of the loop may be different in the two separate gene families.

Fig. 4. Stereo view of the proximity of the domain interface in the structure of the SAP2X of *C. albicans* (thick red) compared to the corresponding area in rhizopuspepsin (thin blue). The diagram is centered around the interactions of Arg 195, Glu 193, and Asn 304, and also illustrates the hydrophobic character of them in rhizopuspepsin when compared with SAP2X. Amino acids discussed in the text have been labeled by the one-letter code followed by the corresponding residue number color coded as above.

In the interdomain connection via the chain segment **Val 184-Ile 192** in pepsin, the mammalian proteases have a 1-residue deletion when compared to the fungals (penicillopepsin, rhizopus, and endothia); yet, SAP2X has the same number of residues as the mammalian proteins (Fig. 2). After superposition, the residues involved in the interdomain interface, topologically equivalent to Glu 193-Leu 194-Arg 195 and Asn 304 in SAP2X, are Trp 194-Trp 195-Gly 196 and Ala 303 in rhizopuspepsin and **Tyr 189-Trp 190-Gln 191** and **Trp 299** in porcine pepsin (Hutchins & Greer, 1991). The completely different character of the interatomic contacts in this region for the two fungal enzymes is of note, changing from a network of hydrogen bonds and polar interactions in SAP2X to a hydrophobic contact surface on the other fungal proteases. In fact, the interface in rhizopuspepsin is dominated by a cluster of four hydrophobic residues: Trp 194, Trp 195, Phe 296, and Trp 294 (Fig. 4). This departure is more significant because Trp 190 is strictly conserved among all the monomeric aspartic proteases known thus far. Although the side chain of **Gln 191** in pepsin also points to the outside of the molecule, it interacts with **Ser 294** in the "proline-rich" flap and does not play a role in the interdomain interface.

The conformation of the stretch of residues between Phe 298 and Ile 305 in the SAP2X structure is very different from the corresponding one in rhizopuspepsin, mainly due to the deletion of two residues in the SAP family (Fig. 2) when compared to 3APR. This shorter connection between Phe 298 and Ile 305 in the SAP2X structure allows for some of the residues to be in close proximity to the active site and to interact with the bound A-70450 inhibitor. In particular, three consecutive odd-numbered residues, Asn 301, Ala 303, and Ile 305, form a short "specificity ridge" unique to this fungal enzyme. Variations in the amino acid sequence of these three residues in the different SAPs could account for some of the different specificities (Fig. 4, see below).

Implications for the structure of other SAPs from different fungi

Monod and collaborators have reported the existence of up to seven SAPs from *Candida* and presented evidence for the existence of analogous secreted proteinases in other *Candida* spe-

cies such as *C. tropicalis*, *C. parapsilosis*, and *C. guilliermondii* and other pathogenic fungi (Monod et al., 1994). Analysis of their amino acid sequence showed a wide range of sequence identity ranging from 20% between PRA and SAP1 to 68% between SAP1 and SAP3. In terms of their amino acid sequence, the different SAPs from *C. albicans* could be clustered into two separate families with three members each (SAP1-3 and SAP4-6), with the longest and most divergent, SAP7, being closer to the other *Candida* species. The structural alignment between the clinical isolate SAP2X and the more homologous SAP1-6 proteinases as shown in Figure 2 supports the initial sequence analysis.

Although similar, the sequences in the vicinity of the unique structural features of the SAPs described above are distinctly different between SAP1-3 and SAP4-6. For instance, the sequence between the first disulfide bond (Cys 47-Cys 59) retains Phe 58 as fully conserved, but the intervening amino acids clearly segregate into two groups. The SAP1-3 enzymes have Thr 50 and Gln 54 providing hydrogen bonding interactions to support the rigid conformation of the loop. However, in the other family (SAP4-6), the replacements Thr 50 → Lys and Gln 54 → Asp could suggest the role of an electrostatic interaction to retain a rigid loop conformation (Fig. 2). The deletion of helix h_{N2} will probably be observed in all the SAPs, although the flanking sequences clearly fall into two separate groups (Fig. 2).

From the viewpoint of specificity in the domain interface, a clear dichotomy also exists between the SAP1-3 and SAP4-6. Although Asn 304 and Ile 305 are strictly conserved among the six SAPs, the sequence of residues 301 and 303 segregates the family basically into two groups. Residue 303 is Ala in SAP1-2, Tyr in SAP3, but is replaced uniformly by Asp in SAP4-6; Asn 301 is also replaced by Ser in all SAP4-6 and for SAP1 and SAP3. More uniform and intriguing are the alterations at the 185-195 region, where SAP1-3 preserve the characteristic sequence Glu 193-Leu 194-Arg 195, probably critical to force the conformation of Asn 304, whereas the SAP4-6 present either Thr-Leu-Ser or Lys-Leu-Thr. Whether or not this latter sequence will also alter the conformation of Asn 304 is still unknown.

The carboxy-terminal extension (residues 327-342), thus far unique to this family of enzymes, will probably have the same three-dimensional conformation in the SAP1-6 enzymes due to

the high sequence homology of the region of the polypeptide chain comprising Gln 329–Ser 336 (QVKYTS) (Fig. 2), and also likely for the SAPs from *C. tropicalis* and *C. parapsilosis*, which also have an identical sequence at the carboxy end (Monod et al., 1994).

Inhibitor conformation

The electron density corresponding to the inhibitor A-70450 (compound 1 in Fig. 5) was readily interpretable as an extended conformation (Fig. 6 and Kinemage 1), similar to that found previously in many other inhibitors bound to a variety of liganded aspartic proteinases (Bailey & Cooper, 1994). However, as described below, this ligand has a unique branched structure at the P3 position.

Compound A-70450 was prepared originally as part of our renin inhibition program (US patent no. 5,164,388) and was later discovered to inhibit the SAP of *C. albicans* (US patent no. 5,120,718). The compound possesses a hydroxyethylene peptide bond isostere as a transition-state mimic with the hydroxylic carbon exhibiting the S-configuration. The hydroxy group is located between Asp 32 and Asp 218, equivalent to the pepsin active site residues **Asp 32** and **Asp 215**. The S-configurations found at the *n*-butyl (nor-leucine unit), cyclohexylmethyl, and isopropyl substituents (see Fig. 7) are in agreement with previous studies of optimal P2, P1, and P1' units in aspartyl proteinases inhibitors. The chemical structure of the P1, P1', and P2' units of A-70450 are identical to that of the corresponding units

of CGP38560, a potent inhibitor of renin (Bühlmayer et al., 1988). The conformation of the two inhibitors in their P1–P2' portions are similar when the structure reported here for A-70450 bound to SAP2X is compared with entry 1RNE for CGP38560 bound to renin (Rahuel et al., 1991).

The carbon in the ketopiperazine ring bearing the benzylic substituent is in the R-configuration. For typical acyclic linear aspartyl protease inhibitors, the optimal configuration of the P3 unit is the S-configuration. This switch in configuration is due to the cyclization of the P3 unit into a ketopiperazine ring, which forces the α -amino group of the P3 unit into a much different position than observed for acyclic inhibitors. The P3 benzyl group of A-70450 occupies an orientation traditionally associated with an S3 subsite of aspartyl proteases. An unusual feature of the conformation of A-70450 when compared with the conformation of other bound inhibitors is the conformation of the methylpiperazine ring of A-70450.

A priori, one might think that the terminal methylpiperazine ring of A-70450 would exist in a 6-membered ring chair conformation and serve as a typical P4 unit of an aspartyl protease inhibitor. Instead, this ring is in a boat conformation and is directed toward the S3 subsite in close contact with the Cys 47–Cys 59 loop, as described above. The S3 subsite of SAP2X is unusual in the aspartyl protease family due to its large size. We have operationally divided the S3 subsite into two sections (Figs. 4, 5, 7). We have designated S3a to be the site binding the P3 benzyl group (the traditional S3 subsite) and S3b to be the site binding the P3 methylpiperazine group of A-70450. Nota-

Cmpd.	X	R ₁	R ₂	Y	R ₃	<i>C. alb.</i> IC ₅₀ (nM)	Renin IC ₅₀ (nM)	Cat.D IC ₅₀ (nM)
1 (A-70450)	C=O	(R)-CH ₂ Ph	(S)-(CH ₂) ₃ CH ₃	C=O	-(CH ₂) ₃ CH ₃	1.4	7.1	770
2	C=O	(R)-CH ₂ Ph	(R)-(CH ₂) ₃ CH ₃	C=O	-(CH ₂) ₃ CH ₃	25	na	na
3	C=O	(S)-CH ₂ Ph	(S)-(CH ₂) ₃ CH ₃	C=O	-(CH ₂) ₃ CH ₃	2.0	>100*	na
4	C=O	(R)-CH ₂ Ph	(S)-(CH ₂) ₃ CH ₃	CH ₂	-(CH ₂) ₃ CH ₃	490	na	na
5	SO ₂	(R)-CH ₂ Ph	(S)-(CH ₂) ₃ CH ₃	C=O	-(CH ₂) ₃ CH ₃	2.7	1.5	150
6	C=O	(R)-CH ₂ Ph	(S)-(CH ₂) ₃ CH ₃	C=O	-(CH ₂) ₃ N(CH ₂ CH ₂) ₂ O	3.8	110	22000
7	C=O	(R)-CH ₂ Ph	(S)-(CH ₂) ₃ CH ₃	C=O	-(CH ₂) ₂ N(CH ₃) ₂	6.2	>100	58000
8	C=O	(R)-CH ₂ Ph	(S)-(CH ₂) ₃ CH ₃	C=O	-(CH ₂) ₃ N(CH ₃) ₂	3.8	140	16000

Fig. 5. Inhibitors of *C. albicans* protease. The figure shows the subsites P3–P2' of the inhibitors. *C. albicans*, renin, and cathepsin D protease assays carried out according to the published protocols (Capobianco et al. 1992, Mycek, 1970; Bolis et al., 1987). * 12% inhibition at 100 nM. The reference compound, pepstatin, gives IC₅₀ values of 27 and 21 nM in the *C. albicans* and cathepsin D enzyme assays, respectively. na, not available.

Fig. 6. View of the electron density corresponding to the inhibitor in the final $2F_o - F_c$ electron density map ($\sigma = 1$ level) with the corresponding model for A-70450. Notice the branching of the inhibitor structure into two subgroups, P3a and P3b.

bly, there is no portion of A-70450, which appears to occupy the S4 subsite of SAP2X.

This branching of the inhibitor A-70450 into both the S3a and S3b subsites is unprecedented in the structural studies of aspartic proteinases, and is believed to be critical for the high inhibition potency exhibited by this compound. The separation of the two adjacent sites at the P3 position is possible because of the deletion at the helical h_{N2} region. This creates a very large pocket that is dramatically different from the one observed in pepsin, renin, or even the other known fungal aspartic proteinases.

Inhibitor structure-activity relationships

The refined structure of the liganded *C. albicans* protease allows a retrospective interpretation of some of the available structure-activity relationships for analogues of compound **1** (A-70450) (US patent no. 5,120,718). Seven analogues of compound **1** are shown in Figure 5 along with their *C. albicans* protease, renin, and cathepsin D inhibitory potencies. Compounds **2** and **3** are epimeric to compound **1** at the P3a benzyl or P2 nor-leucine positions. Although there was a significant drop in potency for changing the configuration at the nor-leucine residue, interest-

ingly, the S-configuration of the P3a benzyl (compound **2**) retained the potency of the R-configuration in compound **1**. These changes in potencies may be reflective of the pocket sizes for the respective side-chains: the S2 site being small and restrictive, the S3 site being much more open (see discussion of S3 site above) and perhaps accommodating to isomeric structures. Compound **4** represents a reduced-bond analogue of compound **1** at the P1'-P2' linkage. The loss of the carbonyl group in compound **4** results in almost a 700-fold drop in potency relative to compound **1**. Interestingly, the structure of *C. albicans* protease complexed with compound **1** shows a specific hydrogen bond to this inhibitor amide linkage (H-bond donation from Gly 85 N-H to backbone C=O of compound **1**) that would be lost upon binding of compound **4**. The urea carbonyl of compound **1** makes no direct contact with the protein and may be replaced by a sulfonyl linkage, as in compound **5**, with only a twofold loss in potency. And finally, the P2' butyl group of compound **1** may be replaced with 3-morpholinopropyl, dimethylaminoethyl, or dimethylaminopropyl groups, compounds **6**, **7**, and **8**, respectively, with only a three- to fourfold drop in potency. The enzyme selectivity of these last three compounds is particularly interesting because each shows reduced activity against renin and sharply decreased affinity for cathepsin D.

Fig. 7. Schematic representation of the hydrogen bonding interactions of the A-70450 inhibitor with the protein atoms of the SAP2X and of the different enzyme pockets corresponding to the inhibitor subsites.

Table 2. Summary of crystallographic data collection and crystallographic refinement statistics

	Form A	Form C	Form B
A. Crystallographic data collection			
Space group	P1	P2 ₁ 2 ₁ 2 ₁	P2 ₁
<i>a</i> (Å)	122.2	98.3	80.1
<i>b</i> (Å)	49.2	49.3	280.0
<i>c</i> (Å)	123.0	63.6	49.2
α (°)	97.3	90	90
β (°)	99.5	90	101.6
γ (°)	98.7	90	90
Observations	88,229	30,743	—
Unique reflections	50,009	10,193	—
Completeness(%)	80	89.4	—
<i>R</i> _{merge} (%) ^a	15.2 ^b	9.90	—
No. mol./asym. unit	8	1	6
Resolution (Å)	10–2.9	8–2.5	2.5
B. Crystallographic refinement statistics			
Unique reflections	50,009	10,193	—
Completeness (%)	80	89.4	—
<i>R</i> -factor	0.267 ^c	0.145	—
No. of water molecules	189	189	—
RMSD bonds (Å)	0.009	0.009	—
RMSD dihedrals(°)	26.78	26.78	—
RMSD impropers(°)	1.40	1.41	—
RMSD angles (°)	1.66	1.66	—
Percentage of residues in Ramachandran allowed regions ^d	99	99	—

^a $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$, where *I* is the observed intensity and $\langle I \rangle$ is the average intensity obtained from multiple of observations possibly from different crystals.

^b Data merged from three different small crystals.

^c In view of the unfavorable variables/reflections ratio, the final model and solvent structure from Form C crystals were refined against Form A data with strict noncrystallographic restraints for the eight molecules in the asymmetric unit using all the available data with an *F*/ σ_F > 1.0. Only marginal differences were observed.

^d According to all of the criteria used by PROCHECK to estimate the quality of the refined model (Laskowski et al., 1993), all the side-chain parameters in the refined structure were considered better than the ones observed for other refined proteins at similar resolution. This was also true for three of the six parameters used for the diagnosis of the main chain; the remaining three were within the standard tolerances. Only residue Gln 11 ($\phi = 65^\circ$, $\psi = -52^\circ$) falls outside the allowed regions of the Ramachandran plot in a classical γ -turn at the end of the first β -strand.

Substrate specificity

Although *C. albicans* protease has been reported to catalyze the cleavage of several protein substrates (Douglas, 1988), we know of only two reports of identification of the specific residues present in the substrate. Capobianco et al. (1992) report that *C. albicans* protease (SAP isolate whose structure is reported in this paper) catalyzes a hydrolytic cleavage at a His-Thr bond in a fluorogenic peptide substrate. The authors also observed an undocumented cleavage at a Lys-Thr bond (Goldman et al., 1996). A second, more extensive study of *C. albicans* protease (Strain C-74) substrate specificity with 30 chromogenic peptide substrates (Fusek et al., 1994) reports that cleavage can occur at a Phe-Phe (*p*-nitro) bond. Thus, observations of either a po-

lar residue, His or Lys, or a nonpolar residue, Phe, occupying the P1 position of a substrate have emerged. A discussion of how this dichotomy in substrate preference can exist has been published (James et al., 1985; Lowther et al., 1995) and we can reiterate these authors' suggestion of the role of the flap residue Asp 77 (SAP2X Asp 86) in playing a pivotal role in determining substrate specificity in the fungal proteases. Specifically, the S1 subsite is hydrophobic and prefers lipophilic P1 residues. However, for P1 lysine or histidine residues, ion-pairing with Asp 77 occurs, permitting hydrolysis of more hydrophilic or polar substrates.

Implications for the design of novel antifungal agents

There are specific residue replacements throughout the active site that differentiate the *Candida* fungal SAPs from mammalian aspartyl proteases such as renin, pepsin, and gastricsin. A focused chemical design effort centralizing on these specific changes might yield a clinically useful antifungal agent possessing the specificity required to be selective at impairing fungal virulence while not affecting mammalian function. More important, in our opinion, is the observation of a potentially much greater source of specificity for designing antifungal chemotherapy—the large S3 subsite of the SAP. Here, one can imagine entirely new structural units for the P3 portion of aspartyl protease inhibitors that would fill the S3a and S3b subsites with novel connectivities. For example, inhibitors could fill the S3a and S3b subsites as A-70450 does, i.e., separate branching from the inhibitor backbone. Alternatively, structural units of an inhibitor might fill the S3a subsite by branching from the S3b subsite (or vice versa). In addition, the S4 site of the SAP is unexplored by A-70450, suggesting that additional functional groups might bind in S4 and improve potency. With the solution of the structure of the SAP2X/A-70450 complex, we now have a spatial framework for rational and creative thought in this design effort.

Antifungal activity has been tested (systemic mouse model) for both pepstatin and A-70450 and the results were discouraging: either weak or no anti-*C. albicans* activity was detected in vivo. This lack of activity is possibly due to lack of potency, low specificity between fungal versus host aspartyl proteases, and/or the inability to inhibit all members of the SAP family, or at least those SAPs most important for virulence. The three-dimensional structure of the SAP2X/A-70450 complex can provide the basis for the design of novel anticandidal agents based on two different strategies: (1) broad specificity, focusing on residues in the neighborhood of the active site that are common to several SAPs; (2) narrow specificity, focusing on the homology modeling of a designated SAP from a specific pathogen. The prophylactic and therapeutic use of such agents would be a valuable addition to the limited repertoire of antifungal agents currently available.

Materials and methods

Protein isolation, purification, and characterization

SAP2X was purified from *C. albicans* strain val-1, a pathogenic clinical isolate from skin, using a minor modification of established methods (Ray & Payne, 1990). Briefly, culture super-

natant (7 L) from cells grown on bovine keratin (ICN Biochemicals, Cleveland, Ohio) as the sole source of nitrogen was filtered through glass wool, adjusted to pH 6.5 with NaOH, filter sterilized through 0.2- μ m filter units (Nalgene Labware Division, Nalgen Sybron Corp., Rochester, New York), flash evaporated to 500 mL, and further concentrated to 30–50 mL using a stirred cell ultrafiltration unit with a YM-10 membrane (Amicon Corp. Danvers, Massachusetts). The concentrate was applied to an S200 Sephacryl (Pharmacia, Uppsala, Sweden) gel filtration column (5.0 $\times$ 100 cm) equilibrated with 10 mM citrate buffer, pH 6.5. The remaining methods are as described (Ray & Payne, 1990). Pooled active fractions from the column were passed a second time over a smaller (2.5 $\times$ 100 cm) S200 Sephacryl column to further purify SAP. Purity was established by silver staining methods, the elicitation of mono-specific antisera in mice, and also by a single peak by capillary electrophoresis (BioRad). N-terminal amino acid sequence analysis was performed by Edman degradation and yielded the sequence: QAVP VTLHNEQVTYAADITVGSSN (Ray et al., 1991). This sequence agrees with the N-terminal sequence reported for SAP2 (Wright et al., 1992).

Amino acid sequence

Because the N-terminal amino acid sequence was identical to SAP2 (Wright et al., 1992), the amino acid sequence of the polypeptide chain was a priori assumed to be very closely homologous to SAP2. The sequence assignment for the individual residues was only changed when the electron density maps at different stages in the refinement differed clearly from the assumed sequence.

Crystallization

A twofold molar excess of A-70450 was added to the SAP recovered from S-200 chromatography. The sample was concentrated and buffer exchanged to 25 mM glycine, 25 mM NaCl buffer, pH 4.5, using Centricon 10 filtration units (Amicon Inc), while maintaining the inhibitor in a twofold molar excess. Samples in the range of 5–10 mg/mL protein in glycine/NaCl buffer were used for crystallization. Three different crystal forms were obtained, always in the presence of millimolar amounts of Zn^{2+} in the form of zinc acetate and various concentrations of PEG 8000. Attempts to grow crystals in the total absence of Zn^{2+} failed.

Form A crystals were obtained in the presence of 40 mM Cacodylate buffer, pH 6.5, 5–10 mM Zn^{2+} and approximately 22% PEG 8000. The crystals grew as thin plates in high concentrations of Zn^{2+} ($\sim$ 10 mM), but developed a more uniform prismatic habit at low concentrations ($\sim$ 5 mM) of the cation. Crystals appeared after a few days and reached full size in a few weeks. These crystals were characterized as triclinic, containing eight molecules of the complex in the unit cell (Table 2A).

Form B crystals were obtained in the presence of 40 mM imidazole/malate buffer, pH 6.5, 5 mM Zn^{2+} , and PEG 8000 concentrations ranging from 22 to 24%. The crystals grew consistently as prismatic in habit and took three to four weeks to grow to useable size. This form was characterized as monoclinic, containing six molecules of the complex in the asymmetric unit (Table 2A).

Form C crystals were only obtained in conditions similar to form B, but at low Zn^{2+} concentration ($\sim$ 1 mM). These crystals were prismatic in habit and took more than two months to grow. Unfortunately, these crystals were difficult to reproduce. They were orthorhombic and contained only one molecule in the asymmetric unit (Table 2A).

Data collection

Data were collected using a Rigaku RAXIS IIC image plate detector equipped with a graphite crystal monochromator mounted on a Rigaku RU-200 rotating anode X-ray generator, operating at 50 kV and 100 mA, equipped with a 0.3-mm focal cup. The crystal to detector distance ranged from 70 mm (Forms A and C) to 125 mm for Form B. Data were collected by the oscillation method using oscillation angles between 3° and 6°, depending upon the type and orientation of the crystal. The collected oscillation frames were processed and reduced using the RAXIS (Higashi, 1990) software, available with the instrument. Further data merging was done using the program PROTEIN (Steigemann, 1974).

Form A crystals were small (typical size: 0.050–0.100 mm in the largest dimensions) and diffracted initially to approximately 2.8 Å resolution. A data set was assembled using diffraction data from three different crystals with the PROTEIN software package, resulting in a total of 50,009 unique observations from a total of 88,229 observations. The overall R_{merge} was 15.2% based on reflections with an $F/\sigma F$ greater than 1.0. These data represent 80% of the possible reflections to a resolution of 2.9 Å, with a completion of 60% in the resolution shell between 3.0 and 2.9 Å. The large overall R_{merge} reflects the presence of a large number of weak reflections beyond 3 Å for the three different crystals (Table 2B).

A complete data set from the only small (0.1 $\times$ 0.1 $\times$ 0.2 mm³) and partially defective crystal of Form C was collected at 2.5 Å resolution (89% complete). It consisted of 10,193 unique reflections obtained from 30,743 observations with an R_{merge} of 9.9%. The data in the last resolution shell had an $(\langle F \rangle / \sigma \langle F \rangle) > 3.5$ and was 85% complete. In the course of the refinement, the crystallographic data from the Form C crystals were reprocessed with the HKL program suite (Gerwirth, 1995) ($R_{\text{merge}} = 10.1\%$, 11,013 unique reflections, 95% complete to 2.5 Å) and subsequent refinement proceeded using these data (Table 2B).

Structure solution and refinement

Full details of the structure solution of the P1 form will be published elsewhere (C. Abad-Zapatero & J. Navaza, in prep.). The crystal structures for both Forms A and C were solved by molecular replacement methods (Rossmann, 1990) as implemented in the program suite AMoRe (Navaza, 1994). The structure of porcine pepsin complexed with Abbott inhibitor A-62095 (Chen et al., 1992) (set 1PSA) was used as an initial search model against data from Form A crystals. This cross-rotation function search readily found four molecules that satisfied most of the observed peaks of a previously calculated self-rotation function. However, attempts to place the remaining four molecules failed. A molecular model built by homology modeling (Hutchins & Greer, 1991) was used later as a probe and confirmed the ini-

tial orientation and placement of the four molecules in the asymmetric unit. Again, subsequent attempts to find the remaining four molecules failed and therefore the attention was directed toward the data set of the Form C crystals. The homology model probe suggested a solution in space group $P2_12_12_1$ and the corresponding model was partially refined. However, lack of density for the inhibitor in the active site suggested at that time that Form C crystals did not contain the inhibitor. The partially refined structure from Form C was then used as a probe to search for a complete solution for Form A crystals containing eight molecules in the asymmetric unit. Such a solution was found ($R = 39\%$, $CC = 51\%$, $15\text{--}4.5\text{ \AA}$). The phases of the solution of the P1 form were refined by noncrystallographic symmetry averaging and solvent flattening over the eight copies in the unit cell, using the software package PHASES (Furey & Swaminathan, 1990). The algorithm converged to a back-transform $R = 0.158$ and $CC = 0.98$ after 15 cycles and 2 molecular masks. The averaged map was of excellent quality and permitted the immediate revision of the model and showed satisfactory density for the inhibitor A-70450 at the active site.

This revised model was subject to two cycles of graphics revision with FRODO (Jones, 1985) and refinement by X-PLOR (Brünger et al., 1987) using strict noncrystallographic constraints. This model was then used to search the Form C crystal data, which confirmed the initial, preliminary solution for space group $P2_12_12_1$ at the orientation and position found previously; at this time, there was clear density for the inhibitor.

Refinement of the orthorhombic form proceeded rapidly by combining model revision using O (Jones et al., 1991) and simulated annealing refinement using X-PLOR (Brünger et al., 1987). The final complex model for the orthorhombic form consisted of 341 amino acid residues, 189 water molecules, 1 zinc atom in the crystal contacts (Asp 57–Glu 324), and one molecule of the inhibitor. The final crystallographic R -factor after individual temperature factor refinement was 0.145 using data between 8 and $2.5\text{ \AA}$ resolution. The final refinement statistics are listed in Table 2B and an example of the quality of the electron density is illustrated in Figure 1. The refined coordinates of the Form C crystal structure have been deposited in the Protein Data Bank (Bernstein et al., 1977) with accession code AZAP. Until their official release, requests for the refined coordinates from individual academic investigators should be sent to the corresponding author.

Structural superpositions

Structural superpositions were performed using the least-squares superpositions routines of the package O (Jones et al., 1991) using a cutoff limit for improvement (Lsq_improve) of $3.8\text{ \AA}$ and a consecutive stretch of three amino acid residues for the smallest fragment that can be aligned. All the superpositions were allowed to converge for these superposition parameters.

Homology modeling of the SAP structure

Examination of the sequence alignment of SAP2 (Wright et al., 1992) with several other members of the class showed immediately an 8-amino acid insertion in the *Candida* sequence after residue 52, and a 6-amino acid deletion after residue 115 in relation to that of rhizopuspepsin. The first region was on the ex-

terior of the protein near the P3, P4 position of a substrate or inhibitor. The second region is near an alpha helix-forming part of the P3 subsite. It was therefore intriguing to propose the *Candida* enzyme had deleted one group of residues (the alpha helix near 115) and had added a number of amino acids at a different location that would, in a structural sense, replace those deleted residues. An insertion at the carboxy end of the *Candida* structure was assumed to be disordered and was never incorporated into the model. The remainder of the protein appeared to be similar to the other fungal proteinases, although the loops were of different lengths.

The homology module of the INSIGHTII program (Biosym, 1992) was used for the homology modeling. The coordinates of the structurally conserved regions were taken from the rhizopuspepsin-inhibitor complex X-ray structure (Gilliland et al., 1990) (3APR). The loops were modeled to follow the same topology as those in other fungal proteinase structures. The helix of rhizopuspepsin was replaced in the *Candida* structure with a loop to connect the two structurally conserved β -strands. The residues of the insertion described above were modeled as an alpha helix and were placed in a position analogous to the position of the helix of rhizopuspepsin. The side-chain conformations for the amino acids of the modeled *Candida* protein were similar to those of the rhizopuspepsin structure. After the residues were in place, all side chains were examined for bad contacts and adjusted if necessary, preferably using one of the rotamers from a rotamer library (Ponder & Richards, 1987). The model was then subjected to 200 cycles of tethered molecular mechanics geometry optimization, using DISCOVER (Biosym, 1992), to improve the geometry of the connections and to relieve major stresses.

Acknowledgments

We thank Dr. J. Greer for critical reading of the manuscript and Dr. T. Perun for support during the course of this project. We appreciate the assistance of the personnel of the research computing department at Abbott for their support during the computations required for this work. The expert assistance of Mr. John Capobianco and Ms. Jean Severin is appreciated during the process of purification, characterization, and crystallization of the enzyme. C.A.-Z. and S.W.M. thank their colleagues Drs. C. Dealwis, Che-Fu Kuo, V. Giranda, and C. Park for discussions and insight. We acknowledge the dedication of the members of the renin project: Drs. W. Baker, H. Stein, and Mr. H.-S. Jac. Portions of this work were supported by NIH grant A124344 (T.L.R.).

Note added in press

While this work was being refereed, the three-dimensional structure of SAP2 from *C. albicans* (strain ATCC 10261) was published (Cutfield et al., 1995). Complexes with pepstatin A and the same A-70450 compound in a different crystal form were described. Although the shape of the active site, including the S3 subsite, is very similar to that described here, a more detailed comparison might reveal intriguing differences, particularly in the inhibitor conformation.

References

- Abad-Zapatero C, Rydel TJ, Erickson JW. 1990. Structure of porcine pepsin: Evidence for a flexible subdomain. *Proteins Struct Funct Genet* 8: 62–81.

- Bailey D, Cooper JB. 1994. Structural comparison of 21 inhibitor complexes of the aspartic proteinase *Endothia parasitica*. *Protein Sci* 3:2129-2143.
- Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol* 112:532-543.
- Biosym. 1992. San Diego, California: Biosym Technologies.
- Bolis G, Fung AKL, Greer J, Kleinert HD, Marcotte PA, Perun TJ, Plattner JJ, Stein JH. 1987. Renin inhibitors. Dipeptide analogues of angiotensinogen incorporating transition-state, non-peptidic replacements at the scissile bond. *J Med Chem* 30:1729-1737.
- Brünger AT, Kuriyan J, Karplus M. 1987. Crystallographic *R* factor refinement by molecular dynamics. *Science* 235:458-460.
- Bühlmayer P, Caselli A, Fuhrer W, Göschke R, Rasetti V, Rüeger H, Stanton JL, Criscione L, Wodd JM. 1988. Synthesis and biological activity of some transition-state inhibitors of human renin. *J Med Chem* 31:1839-1846.
- Capobianco JO, Lerner CG, Goldman RC. 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.
- Chen L, Erickson JW, Rydel TJ, Park CH, Neidhart D, Luly J, 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 Crystallogr B* 48:476-488.
- Cooper JB, Foundling S, Hemmings A, Blundell TL, Jones M, Hallett A, Szelke M. 1987. The structure of a synthetic pepsin inhibitor complexed with endothiapepsin. *Eur J Biochem* 169:215-221.
- Cutfield SM, Dodson EJ, Anderson BF, Moody PCE, Marshall CJ, Sullivan PA, Cutfield JF. 1995. The crystal structure of a major secreted aspartic proteinase from *Candida albicans* in complexes with two inhibitors. *Structure* 3:1261-1271.
- Cutler E. 1991. Putative virulence factors of *Candida albicans*. *Annu Rev Microbiol* 45:187-218.
- Davies DR. 1990. The structure and function of the aspartic proteinases. *Annu Rev Biophys Chem* 19:189-215.
- Dealwis C. 1993. X-ray crystallographic analysis of aspartic proteinases and their inhibitor complexes. [dissertation]. London: University of London, Department of Crystallography, Birbeck College.
- Dealwis CG, Frazao C, Badasso M, Cooper JB, Tickle IJ, Driessen H, Blundell TL, Murakami H, Sueiras-Diaz J, Jones DM, et al. 1994. X-ray analysis at 2.0 Å resolution of mouse submaxillary renin complexed with a decapeptide inhibitor CH-667, based on the 4-16 fragment of rat angiotensinogen. *J Mol Biol* 236:342-360.
- Douglas LJ. 1988. *Candida* proteinases and candidosis. *Crit Rev Biotechnol* 8:121-129.
- Furey W, Swaminathan S. 1990. PHASES: A program package for the processing and analysis of diffraction data from macromolecules. *American Crystallographic Association Meeting Abstracts*, PA33 18:73.
- Fusek M, Smith EA, Monod M, Dunn BM, Foundling SI. 1994. Extracellular aspartic proteinases from *Candida albicans*, *Candida tropicalis* and *Candida parapsilosis* differ substantially in their specificities. *Biochemistry* 33:9791-9799.
- Gewirth D. 1995. *The HKL Manual*. 4th ed. Yale University, New Haven, Connecticut.
- Gilliland GL, Winborne EL, Nachman J, Wlodawer A. 1990. The three-dimensional structure of recombinant bovine chymosin at 2.3 Å resolution. *Protein Struct Funct Genet* 8:82-101.
- Goldman RC, Frost DJ, Capobianco JO, Kadam S, Rasmussen RR, Abad-Zapatero C. 1995. Antifungal drug targets: *Candida* secreted aspartyl protease and fungal wall β -glucan synthesis. *Infectious Agents and Disease* 4:228-247.
- Higashi T. 1990. Rigaku Corporation.
- Hsu I, Delbaere LTJ, James MNG, Hoffmann T. 1977. Penicillopepsin form *Penicillium janthinellum* crystal structure at 2.8 Å and sequence homology with porcine pepsin. *Nature* 266:140-145.
- Hube B, Monod M, Schofield DA, Brown AJ, Gow NA. 1994. Expression of seven members of the gene family encoding secretory aspartyl proteinases in *Candida albicans*. *Mol Microbiol* 14:87-99.
- Hube B, Turver CJ, Odds FC, Eiffert H, Boulnois GJ, Kochel H, Ruchel R. 1991. Sequence of the *Candida albicans* gene encoding the secretory aspartate proteinase. *J Med Vet Mycol* 29:129-132.
- Hutchins C, Greer J. 1991. Comparative modeling of proteins in the design of novel renin inhibitors. *Crit Rev Biochem Mol Biol* 26:77-127.
- James MGN, Sielecki AR, Hofmann T. 1985. X-ray diffraction studies on penicillopepsin and its complexes: The hydrolytic mechanism. In: Kostka V, ed. *Aspartic proteinases and their inhibitors*. Berlin: Walter de Gruyter. pp 163-177.
- Jones TA. 1985. Interactive computer graphics: FRODO. *Methods Enzymol* 115:157-171.
- Jones TA, Zou JY, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 47:110-119.
- Kostka V, ed. 1985. *Aspartic proteinases and their inhibitors*. Berlin: Walter de Gruyter.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK - A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283-291.
- Lerner CG, Goldman RC. 1993. Stimuli that induce production of *Candida albicans* extracellular aspartyl proteinase. *J Gen Microbiol* 139:1643-1651.
- Lott TJ, Page LS, Boiron P, Benson J, Reiss E. 1989. Nucleotide sequence of *Candida albicans* aspartyl proteinase gene. *Nucleic Acids Res* 17:1779-1784.
- Lowther WT, Majer P, Dunn BM. 1995. Engineering the substrate specificity of rhizopuspepsin: The role of Asp 77 of fungal aspartic proteinases in facilitating the cleavage of oligopeptide substrates with lysine in P1. *Protein Sci* 4:689-702.
- MacKay VL, Armstrong J, Yip C, Welsh S, Walker K, Osborn S, Sheppard P, Forstrom J. 1990. Characterization of the Bar proteinase, an extracellular enzyme from the yeast *Saccharomyces cerevisiae*. In: Dunn BM, ed. *Aspartic proteinase conference on structure and function of the aspartic proteinases: Genetics, structure and mechanisms*. New York/London: Plenum Press. pp 161-172.
- Magee BB, Hube B, Wright RJ, Sullivan PJ, Magee PT. 1993. The genes encoding the secreted aspartyl proteinases of *Candida albicans* constitute a family with at least three members. *Infect Immunol* 61:3240-3243.
- Miyasaka SH, White TC, Agabian N. 1994. A fourth secreted aspartyl proteinase gene (SAP4) and a CARE2 repetitive element are located upstream of the SAP1 gene in *Candida albicans*. *J Bacteriol* 176:1702-1710.
- Monod M, Togni G, Hube B, Sanglard D. 1994. Multiplicity of genes encoding secreted aspartic proteinases in *Candida* species. *Mol Microbiol* 13:357-368.
- Mycek MJ. 1970. Cathepsins. *Methods Enzymology* 19:285-315.
- Navaza J. 1994. AMoRe: An automated package for molecular replacement. *Acta Crystallogr A* 50:157-163.
- Newman M, Watson F, Roychowdhury H, Jones H, Badasso M, Cleasby A, Wood SP, Tickle IJ, Blundell TL. 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.
- Ponder JW, Richards FM. 1987. Tertiary templates for proteins: Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol* 193:775-791.
- Rahuel J, Priestle J, Gruetter MG. 1991. The crystal structure of recombinant glycosylated human renin alone and in complex with a transition state analog inhibitor. *J Struct Biol* 107:227-236.
- Ray TL, Payne CD. 1990. Comparative production and rapid purification of *Candida* acid proteinase from protein-supplemented cultures. *Infect Immunol* 58:508-514.
- Ray TL, Payne CD, Berhold A. 1991. The N-terminal amino acid sequence of *Candida* acid protease from *Candida albicans* strain WO-1. *Clin Res* 39:719A.
- Ray TL, Payne CD, Morrow BJ. 1990. *Candida albicans* acid proteinase: Characterization and role in candidiasis. In: Dunn BM, ed. *Aspartic proteinase conference on structure and function of the aspartic proteinases: Genetics, structures, and mechanisms*. Sonoma County, California: Plenum Press. pp 173-183.
- Rossmann MG. 1990. The molecular replacement method. *Acta Crystallogr A* 46:73-82.
- Ruchel R, de Bernardis F, Ray TL, Sullivan PA, Cole GT. 1992. *Candida* acid proteinases. *J Med Vet Mycol* 30:123-132.
- Ruchel R, Ritter B, Schaffrinski M. 1990. Modulation of experimental systemic murine candidosis by intravenous pepstatin. *Int J Med Microbiol* 273:391-403.
- Sali A, Veerapandian B, Cooper JB, Foundling SI, Hoover DJ, Blundell TL. 1989. High resolution X-ray diffraction study of the complex between endothiapepsin and an oligopeptide inhibitor: The analysis of inhibitor binding and description of the rigid body shift in the enzyme. *EMBO J* 8:2179-2188.
- Sali A, Veerapandian B, Cooper JB, Moss DS, Hofmann T, Blundell TL. 1992. Domain flexibility in aspartic proteinases. *Protein Struct Funct Genet* 12:158-170.
- Sielecki AR, Fedorov AA, Boodhoo A, Andreeva NS, James MNG. 1990. The molecular and crystal structure of monoclinic porcine pepsin refined at 1.8 Å resolution. *J Mol Biol* 214:143-170.
- Steigemann W. 1974. Dissertation. München: Technische Universität.
- Tang J, James MNG, Hsu IN, Jenkins JA, Blundell TL. 1978. Structural

- evidence for gene duplication in the evolution of the acid proteases. *Nature* 271:618–621.
- Wagner T, Borg-v Zepelin M, Ruchel R. 1995. pH-dependent denaturation of extracellular aspartic proteinases from *Candida* species. *J Med Vet Mycol* 284:72–94.
- White TC, Agabian N. 1995. *Candida albicans* secreted aspartyl proteinases: Isoenzyme pattern is determined by cell type, and levels are determined by environmental factors. *J Bacteriol* 177:5215–5221.
- White TC, Kohler GA, Miyasaki SH, Agabian N. 1995. Expression of virulence factors in *Candida albicans*. *Can J Botany*. Forthcoming.
- White TC, Miyasaki SH, Agabian N. 1993. Three distinct secreted aspartyl proteinases in *Candida albicans*. *J Bacteriol* 175:6126–6133.
- Wright RJ, Carne A, Hieber AD, Lamont IL, Emerson GW, Sullivan PA. 1992. A second gene for a secreted aspartate proteinase in *Candida albicans*. *J Bacteriol* 174:7848–7853.
- Zotter C, Haustein UF, Schonborn C, Grimmecke HD, Wand H. 1990. Die Wirkung von Pepstatin A auf die *Candida albicans* – Infektion der Maus. *Dermatol Monatsschr* 176:189–198.