

Structural insights into the substrate specificity and activity of ervatamins, the papain-like cysteine proteases from a tropical plant, *Ervatamia coronaria*

Raka Ghosh, Sibani Chakraborty, Chandana Chakrabarti, Jiban Kanti Dattagupta and Sampa Biswas

Crystallography and Molecular Biology Division, Saha Institute of Nuclear Physics, Kolkata, India

Keywords

3D structures; inhibitor complexes; multiple enzymes; plant cysteine proteases; proteolytic activity

Correspondence

J. K. Dattagupta, Crystallography and Molecular Biology Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700 064, India
Fax: +91 33 23374637
Tel: +91 33 23214986
Email: jibank.dattagupta@saha.ac.in

Database

The cDNA sequence of ervatamin-A has been deposited in the NCBI GenBank with accession number EF591130. The coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 3BCN and 2PRE for the two crystal structures ervatamin-A and ervatamin-C, both complexed with E-64

(Received 25 August 2007, revised 16 November 2007, accepted 27 November 2007)

doi:10.1111/j.1742-4658.2007.06211.x

Multiple proteases of the same family are quite often present in the same species in biological systems. These multiple proteases, despite having high homology in their primary and tertiary structures, show deviations in properties such as stability, activity, and specificity. It is of interest, therefore, to compare the structures of these multiple proteases in a single species to identify the structural changes, if any, that may be responsible for such deviations. Ervatamin-A, ervatamin-B and ervatamin-C are three such papain-like cysteine proteases found in the latex of the tropical plant *Ervatamia coronaria*, and are known not only for their high stability over a wide range of temperature and pH, but also for variations in activity and specificity among themselves and among other members of the family. Here we report the crystal structures of ervatamin-A and ervatamin-C, complexed with an irreversible inhibitor 1-[L-*N*-(trans-epoxysuccinyl)leucyl]amino-4-guanidinobutane (E-64), together with enzyme kinetics and molecular dynamic simulation studies. A comparison of these results with the earlier structures helps in a correlation of the structural features with the corresponding functional properties. The specificity constants ($k_{\text{cat}}/K_{\text{m}}$) for the ervatamins indicate that all of these enzymes have specificity for a branched hydrophobic residue at the P2 position of the peptide substrates, with different degrees of efficiency. A single amino acid change, as compared to ervatamin-C, in the S2 pocket of ervatamin-A (Ala67 → Tyr) results in a 57-fold increase in its $k_{\text{cat}}/K_{\text{m}}$ value for a substrate having a Val at the P2 position. Our studies indicate a higher enzymatic activity of ervatamin-A, which has been subsequently explained at the molecular level from the three-dimensional structure of the enzyme and in the context of its helix polarizability and active site plasticity.

The diverse roles of plant cysteine proteases in biological processes have already been established [1–3]. Some of them are involved in defense responses, such as papain in the latex of *Carica papaya*, which is triggered by invading pathogens [4]. Other papain-like proteases seem to be involved in the different signaling cascades

of plants [1]. These proteases belong to the C1 family, clan CA according to the classification in the MEROPS database (<http://merops.sanger.ac.uk>); this also contains mammalian intracellular proteases such as cathepsins (B, C, L, K, S, etc.) and proteases from pathogenic parasites, which act as drug targets in

Abbreviations

E-64, 1-[L-*N*-(trans-epoxysuccinyl)leucyl]amino-4-guanidinobutane; pNA, *p*-nitroanilide; β-ME, β-mercaptoethanol.

many diseases caused by uncontrolled proteolysis or parasite infection [5,6]. Plant proteases of this family have long been used in industry, owing to their high stability and broad specificity [7,8]. These proteases show high sequence similarity and they share a common fold with papain, the archetypal enzyme of the family, which has served as a model for mechanistic and structural studies. The papain-like fold consists of two domains with a V-shaped active site cleft at the interface of the domains, with a catalytic dyad comprising residues Cys and His situated at the opening of the cleft, one from each domain. The activity of the proteases is governed by the catalytic dyad that exist as a Cys[−]...His⁺ zwitterion – a prerequisite for enzyme catalysis [9]. The role of neighboring residues of the catalytic dyad, such as Asp158, Asn175, Glu50, and Gln19 (papain numbering), in catalysis and stabilization of the zwitterionic form of the dyad has been established from high-resolution X-ray structures and mutagenesis studies [10–12]. The active center chemistry and catalytic mechanism of this class of enzymes have been studied extensively by Brocklehurst *et al.* for a number of enzymes, using different reactive probes [13–18]. Other factors contributing to the stability of the Cys[−]...His⁺ ion pair, such as the dipole moment of the central helix, and interdomain interactions, have also been characterized [19–21]. Specificity subsites [22] for the members of this protease class have been identified from the crystal structures of the enzymes with substrate analog inhibitors [23–25]. The specific roles of individual amino acid residues in the subsites in substrate specificity have also been identified from mutagenesis studies [26].

In this family of papain-like cysteine proteases, it is not very uncommon to find, in a single species, multiple proteases that quite often differ in stability, activity and specificity in spite of their high homology in primary and tertiary structures [1,3,27]. Multiple lysosomal cysteine proteases of this family (cathepsins) from humans and their mammalian homologs have been widely studied [28]. Because these cathepsins are involved in the lysosomal proteolytic machinery, the uncontrolled regulation of their normal function leads to a number of pathological events in humans. These conditions may even arise when the regulatory protein inhibitors for these cathepsins, such as stefins or cystatins, are downregulated [28–30]. The cathepsins have been shown to be potential drug targets, having a relatively short and well-defined substrate-binding site [5]. In addition to the structural and biochemical studies on the individual cathepsins, comparisons of the subsite structures related to the functions of the proteases have also been made, and these studies serve as a

useful guide for drug or inhibitor design, which should be specific for a particular protease that is responsible for a particular pathological event in humans [5,28,31]. Studies on the plant multiple proteases, on the other hand, are limited. Structures of individual multiple proteases from the latex of *C. papaya* have been studied, and a few biochemical properties of some of these proteases have been compared [32,33], but elaborate studies relating their 3D structures with their properties, such as stability and activity, have not been reported. Such a study is particularly necessary because sometimes even a subtle change in the structure may cause variations in functions. Papain-like cysteine proteases from plants (papain, ficin, bromelain, etc.) have long been used in industry [34]. Stability and activity are two important parameters that practically determine the feasibility of the industrial application of an enzyme. Three-dimensional structures of multiple enzymes from the same species can thus be a useful source of natural variants for investigation of properties such as stability and activity, to judge the potential of the enzymes for industrial applications.

An attempt has therefore been made in this article by choosing one such species that has in its latex multiple enzymes sharing a common catalytic mechanism, but differing in properties such as stability, substrate specificity, and activity. The stability aspect has been extensively discussed in our earlier papers [27,35], and here we address the issue of substrate specificity and activity in greater detail. In the process, we have also studied the dynamic aspects of noncovalent interactions involved in substrate/inhibitor recognition and their effects on enzyme catalysis in this class of enzymes.

Three papain-like cysteine proteases, ervatamin-A, ervatamin-B and ervatamin-C, have been isolated and purified from the latex of a medicinal plant *Ervatamia coronaria* and biophysically/biochemically characterized [36–39]. The 3D crystal structures of ervatamins determined by us are used for investigation of the catalytic mechanism and substrate specificity and to understand the differences therein for this class of enzymes. In order to identify the subsites of the ervatamins and to understand substrate or inhibitor binding/recognition at the molecular level, we have crystallized ervatamin-A and ervatamin-C with a cysteine protease-specific inhibitor, 1-[L-N-(trans-epoxysuccinyl)leucyl]amino-4-guanidinobutane (E-64), which occupies the active site cleft from the S1 to the S3 subsites of the enzyme. Enzyme kinetic studies with chromogenic peptide substrates, along with the structural information from the enzyme–inhibitor

complexes, help us to understand the activity and substrate specificity at the molecular level. Molecular dynamics simulation studies at 300 K also help in revealing the dynamic behavior of amino acid side chains of the particular enzyme involved in substrate binding. It provides additional knowledge that complements the static conformation obtained from X-ray diffraction methods and helps us to understand the P2 specificity of ervatamins for branched hydrophobic residues, as compared to an aromatic residue in the case of papain. The high activity of ervatamin-A has been explained from the structural point of view, and it is seen that small differences in globally similar enzyme structures, even when the differences are at a remote position from the active center, may have considerable influence on specificity and activity. In this article, the domain plasticity has been correlated with the activity of the enzyme, which has the catalytic site at the interface of the two domains. This observation helps us to improve our understanding of the generic principle of catalysis for this family of proteases.

Results and Discussion

Fold of ervatamin-A

The overall structure and fold of ervatamin-A is similar to that of ervatamin-B, ervatamin-C [27,35] and other members of the papain family, and is made up of two domains, L and R, with a V-shaped active site cleft at the interface of the domains (Fig. 1).

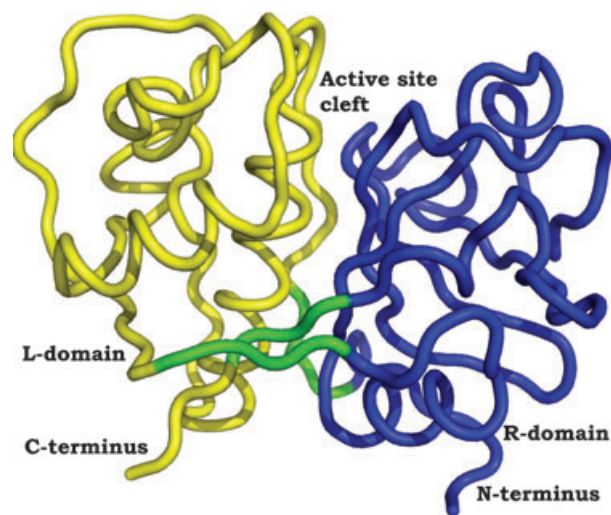


Fig. 1. Ribbon diagram of ervatamin-A; left domain, right domain and interdomain crossover are identified by yellow, blue and green color ribbons, respectively.

Amino acid sequence of ervatamin-A and comparison with other ervatamins

The first 10 N-terminal amino acids of mature ervatamin-A were already known to us from protein sequencing [40], and on the basis of this, the forward primer for cDNA amplification was designed. Sequences from residues 11–195 were derived from the partial cDNA sequence of ervatamin-A, which is described in Experimental procedures. The backbone of region 196–209 was traced from the electron density map. Within region 196–209, the side chain of the last residue could not be located from the electron density map; the rest could be fitted in the map, mainly guided by the sequence conservation in the family of papain-like plant cysteine proteases.

Ervatamin-A shows 90% sequence identity with ervatamin-C [including Cys114 and Cys193, forming the extra (fourth) disulfide bond], with no insertions or deletions. One would therefore expect ervatamin-A, like ervatamin-C, to have a fourth disulfide bridge at the equivalent position. However, the electron density map of ervatamin-A at various levels clearly indicates that Cys114 and Cys193 adopt rotamer conformations that are unfavorable for the formation of a disulfide bond, and remain in a reduced form. In addition, ervatamin-A was found to have a free Cys (108) apart from the active site Cys, which is not very common in the papain family. In comparison, ervatamin-B differs from ervatamin-A (65% identity), with insertions/deletions in its amino acid sequence.

The structure of ervatamin-A is reported for the first time; hence, model building and structural features have been described. On the other hand, as mentioned above, the 3D structure of ervatamin-C has been published previously [27], and therefore only its binding interactions with E-64 will be discussed here.

Modes of binding of E-64 with ervatamin-A and ervatamin-C

The modes of binding of E-64 with ervatamin-A and ervatamin-C have been analyzed and compared with the structures of other complexes of the same family. E-64 binds to these two ervatamins in the same manner as that found in other structures of complexes of papain-like cysteine proteases, and here also the binding is in the reversed orientation [5,23,24]. The C2 atom of the epoxy ring of E-64 is covalently bound to the active site Cys (Cys25) S γ atom, with an average bond distance of 1.8 Å (Fig. 2A). One of the E-64 carboxyl oxygen atoms in each of the ervatamins is

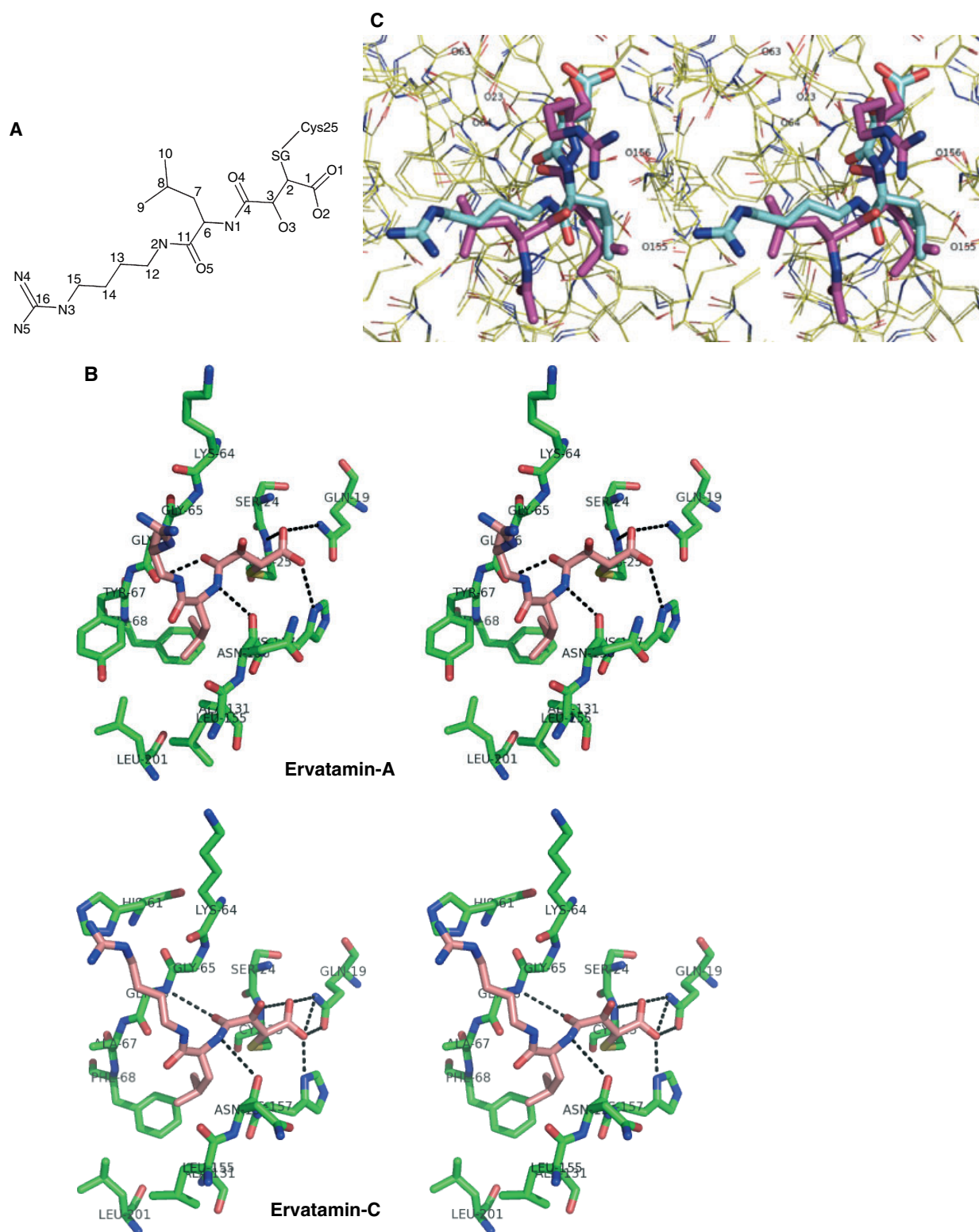


Fig. 2. (A) Schematic representation of E-64 covalently bound to the active site Cys (S γ) of an enzyme. (B) Ervatamin-E-64 interactions (shown in stereo view) for one of the two molecules in the asymmetric unit of ervatamin-A and ervatamin-C. Hydrogen bonds are marked by dashed lines. Molecules are represented by stick models, and E-64 carbon atoms are colored pink. (C) Superposition of ervatamin-C, complexed with leupeptin (carbon atoms in magenta) and E-64 (carbon atoms in cyan) at the active site region (stereo view).

stabilized in the oxyanion hole formed by Gln19 side chain N ϵ 2 and Cys25 main chain nitrogen atoms, while the second oxygen atom interacts with the active site HisN δ 1 (supplementary Table S1). The backbone nitrogen atom of residue 24 is also involved in stabilizing one of the carboxyl oxygen atoms of E-64. Other hydrogen bonds or electrostatic short contacts involving backbone atoms of E-64 and ervatamin-A or ervatamin-C are also listed in supplementary Table S1. In contrast to the inhibitor interactions at the active site, the subsite interactions (S2–P2 and S3–P3) are different in ervatamin-A and ervatamin-C.

S2–P2 interactions

The Leu side chain of E-64 at the P2 position is buried inside the S2 pocket of ervatamin-A, formed by the side chains of Tyr67, Phe68, Ala131, Leu155 and Leu201. This Leu side chain forms a van der Waals contact mainly with the side chain of Tyr67 and to some extent with Phe68, Ala131 and Leu155 (Fig. 2B). The side chain of Tyr67 in ervatamin-A points towards the S2 cleft, providing the maximum hydrophobic environment for S2–P2 stabilization.

The S2 pocket of ervatamin-C is formed by the side chains of Ala67, Ala131, Phe68, Leu155 and Leu201 (Fig. 2B). Tyr67 in ervatamin-A is replaced by Ala67 in ervatamin-C, and due to this replacement, the S2 cavity of ervatamin-C has a wider opening than that of ervatamin-A and lacks the proper environment in which to bind and fix the Leu side chain (P2) of E-64 tightly. In fact, S2–P2 binding is not compact in ervatamin-C, and free space is observed beyond this Leu side chain at the S2 cavity (Fig. 2B). Two molecules of the enzyme complex in the crystallographic asymmetric unit provide individual snapshots which reveals the dynamic nature of the binding of E-64 to ervatamin-C. The main van der Waals interactions are provided by Leu155, Phe68 and Ala131 in both molecules of the asymmetric unit. This observation also corroborates the IC₅₀ value of 225.0 nM for ervatamin-C as compared to 76.25 nM for ervatamin-A.

The mode of interaction of the Leu (P2) residue of E-64 is similar to that of another substrate analog inhibitor leupeptin, as revealed from our previous docking studies with ervatamin-C, although the direction of the peptide-binding mode is opposite in the two cases (Fig. 2C).

S3–P3 interactions

The S3 subsite for papain-like cysteine proteases is not well defined like the S2 subsite; rather, it can be

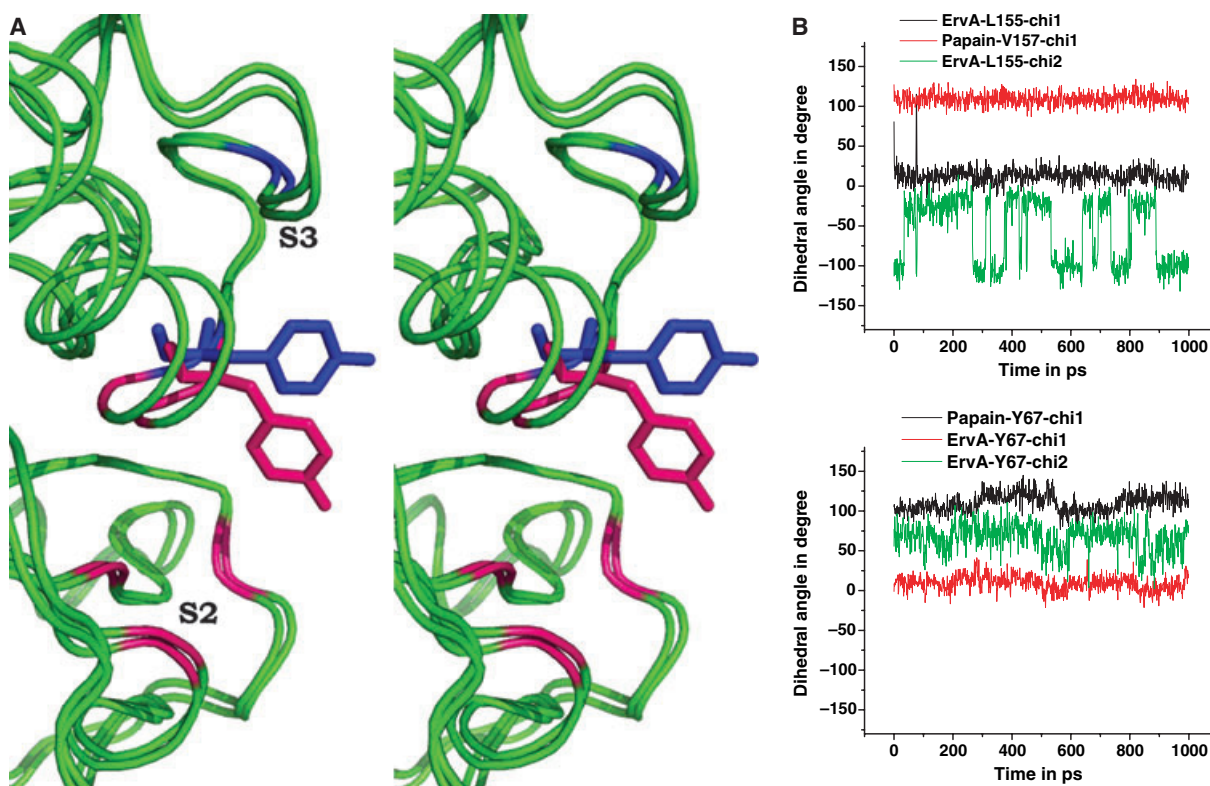
assigned to a region on the surface of domain L containing the active site Cys. The S3–P3 interactions are mainly governed by side chain interactions, and accordingly the amino-4-guanidinobutane (P3) moiety of E-64 orients differently in ervatamin-A and ervatamin-C. This difference in the orientation of the P3 moiety is further influenced by the different orientation pattern of the individual P2 Leu residues. In ervatamin-A, the P3 moiety of both the molecules of the asymmetric unit runs along the extended backbone of residues 65–64 and is partly exposed to the solvent (Fig. 2B). On the other hand, the P3 moiety of E-64 in each of the two ervatamin-C molecules in the asymmetric unit mainly interacts with His61 in both molecules (Fig. 2B).

The substrate specificity of ervatamins – a comparison from structural and kinetic studies

It is established that the specificity of papain-like cysteine proteases is primarily determined by S2–P2 interactions, as the S2 subsite is a deep pocket that makes a major contribution to the binding energy for enzyme–substrate interactions [5,31]. Our kinetic studies on the ervatamins using chromogenic peptides show a preference for Val at the P2 position in ervatamins (Table 1). Ervatamins show less activity towards benzoyl-Arg-*p*-nitroanilide, containing a phenyl ring at its P2 position, which is an ideal substrate for papain [26,41]. It is also shown that a substrate containing an aromatic ring (like Phe) at the P2 position is less preferred than Val or Leu by ervatamins. Of the five residues of the S2 subsite, a Leu (position 205 for ervatamin-A and ervatamin-C, and position 208 for ervatamin-B) at the bottom of the subsite is conserved in ervatamins. Compared to a Ser in the equivalent position of papain, the role of Leu here is to restrict the size and depth of the S2 subsite. The S2 subsites of ervatamin-A and ervatamin-C are similar, with only one substitution observed: Tyr67 in ervatamin-A is replaced by Ala67 in ervatamin-C. The same position is occupied by a Trp in ervatamin-B. In the case of ervatamin-A and ervatamin-B, this residue points towards the S2 cleft, whereas the Tyr at the equivalent position in papain moves towards the S3 subsite, resulting in less contribution to the S2–P2 interaction. This Tyr at the equivalent position of ervatamin-A and papain adopts different rotamer conformations (Fig. 3A), with a $\sim 90^\circ$ difference in χ_1 angle between the two. This difference is also observed to be maintained in the 1 ns dynamics trajectory of papain and ervatamin-A in solvated conditions without any inhibitor/substrate (Fig. 3B).

Table 1. Kinetic constants for peptidyl *p*-nitroanilides and IC₅₀ value for E-64. The P2 residues of the substrates are in bold.

Substrates	Ervatamin-A			Ervatamin-B			Ervatamin-C		
	<i>K_m</i> (mM)	<i>k_{cat}</i> (s ⁻¹)	<i>k_{cat}/K_m</i> (s ⁻¹ ·mM ⁻¹)	<i>K_m</i> (mM)	<i>k_{cat}</i> (s ⁻¹)	<i>k_{cat}/K_m</i> (s ⁻¹ ·mM ⁻¹)	<i>K_m</i> (mM)	<i>k_{cat}</i> (s ⁻¹)	<i>k_{cat}/K_m</i> (s ⁻¹ ·mM ⁻¹)
<i>N</i> -benzoyl-Phe- Val -Arg-pNA	0.071	35.330	497.605	0.057	0.285	5.0	1.063	9.312	8.760
D-Val- Leu -Lys-pNA	0.75	4.276	6.065	0.218	0.134	0.615	0.548	0.385	0.703
D-Ile- Phe -Lys-pNA	1.666	5.474	3.286	0.725	0.022	0.03	1.475	0.264	0.179
Ala-Ala- Val -Ala-pNA	0.501	0.333	0.665	0.013	0.002	0.154	0.355	0.192	0.541
D-Ile- Pro -Arg-pNA	0.683	0.017	0.025	0.364	0.241	0.662	0.716	0.015	0.021
<i>N</i> _α -benzoyl-Arg-pNA	0.550	0.07	0.127	No activity			0.776	0.007	0.009
D-Leu-Ser- Thr -Arg-pNA	2.36	0.198	0.084	Very low activity			1.683	0.037	0.022
<i>N</i> -succinyl-Ala- Ala -Ala-pNA	No activity			0.383	0.003	0.008	Very low activity		
<i>N</i> -benzoyl-Val- Gly -Arg-pNA	1.888	0.117	0.062	Very low activity			Very low activity		
<i>N</i> -acetyl-Leu-Glu- His -Asp-pNA	No activity			No activity			No activity		
<i>N</i> -acetyl-Val-Glu- Ile -Asp-pNA	No activity			No activity			No activity		
<i>N</i> -acetyl-Ile-Glu- Thr -Asp-pNA	No activity			No activity			No activity		
IC ₅₀ in nM for E-64	76.25			123.74			225.0		

**Fig. 3.** (A) Stereo view of superposition of the S2/S3 region of ervatamin-A on papain. The Tyr67 residues of ervatamin-A and papain are shown as stick models, colored magenta and blue, respectively. The remaining parts of the proteins are represented as ribbon diagrams, with S2 residues in magenta and S3 residues in blue. (B) Molecular dynamics trajectory of papain and ervatamin-A for residue 157 (papain numbering, 155 for ervatamin-A) and residue 67, respectively.

The 1 ns molecular dynamics trajectories at room temperature (300 K) for papain and ervatamin-A also reveal that side chain conformations of the S2 subsite residues are less flexible in papain than in ervatamin-A. The distribution of the chi2 dihedral angle of Tyr67 in

ervatamin-A with time (Fig. 3B) shows that the aromatic ring of the residue can move around the Cβ–Cγ bond, and may act as a lid that fixes the P2 side chain upon binding. A large degree of flexibility of the Leu155 chi2 dihedral angle is also observed in the

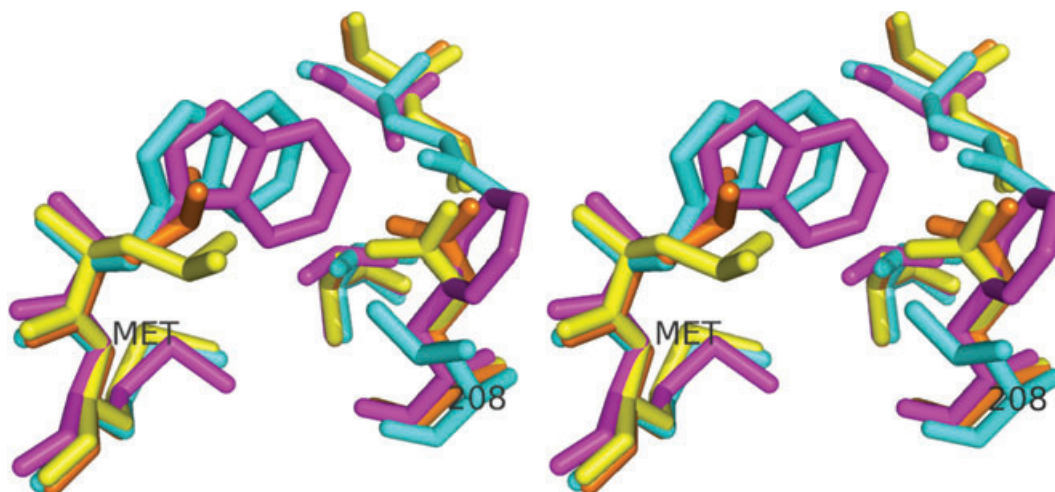


Fig. 4. Superposition of cysteine proteases specific for Pro (at P2): cyan, ervatamin-B; magenta, ginger protease II; orange, barley EP-B2; yellow, viganain; Protein Data Bank codes 1IWD, 1CQD, 1FO5 and 1S4V, respectively. The conserved Met residue is marked.

trajectory of ervatamin-A. This residue also provides hydrophobic interactions with the P2 leucyl side chain of E-64 in both the ervatamins (supplementary Table S1). Owing to its side chain flexibility, it can adopt a conformation suitable for binding a flexible P2 residue such as Val or Leu. On the other hand, side chains lining the papain S2 subsite show no conformational flexibility in the trajectory; this, and the larger volume of the S2 cleft, result in a preference for a bulky rigid aromatic side chain at the P2 position of the substrate in the case of papain [26,41].

Ervatamin-B shows Pro specificity at the P2 position of the substrate. The S2 subsite of ervatamin-B is lined with Trp67, Met68, Thr132, Glu157 and Leu208 [35]. The Met residue at position 68 is conserved in other papain-like cysteine proteases with Pro specificity at the P2 position of the substrate (Fig. 4). These proteases specific for Pro (at P2) also contain a bulky residue at the equivalent position of 208 in ervatamin-B [42–44].

Kinetic studies indicate that the ervatamins have a preference for a long-chain positively charged residue such as Arg or Lys at the P1 position, and show no activity for substrates containing Asp at this position (Table 1). Our previous docking studies on ervatamin-B and ervatamin-C [27,35] with a substrate analog inhibitor, leupeptin, showed that an Arg at the P1 position of the inhibitor points away from the active site cleft towards the solvent. This Arg appears to have conformational flexibility, and in the case of ervatamin-C, only weak stabilizing interactions are provided by the enzyme through the backbone oxygen atom of residues 155 and 156 (Fig. 2C). We also observe from the model of the ervatamin-C complex with leupeptin (Fig. 2C) that if a smaller, negatively charged residue,

Asp, replaces Arg at the P1 position, it will lie in a region surrounded by an array of backbone oxygen atoms from residues 63, 64 and 23, and this region will provide an unfavorable electrostatic environment for Asp at the P1 position. This probably explains our observation that ervatamins show no activity towards substrates with Asp at the P1 position (Table 1).

High activity of ervatamin-A

The proteolytic mechanism of papain-like cysteine proteases have long been studied [10,22,33], and are known to be mediated by Cys and His forming a catalytic dyad. It has been established that the catalytic propensity of a His is highest, followed by a Cys, at the enzyme active site among the 20 naturally occurring amino acids [45,46]. In papain-like cysteine proteases, the presence of a zwitterionic form of the Cys[−]...His⁺ catalytic dyad was initially indicated experimentally [14] and later established theoretically [9,47,48], and was considered to be a prerequisite for catalysis. The most important contribution to stabilizing the zwitterions comes from the long central α -helix to which the catalytic Cys belongs [49]. The polarizing effect originating from the helix concerned facilitates the transfer of the proton from the catalytic Cys present at the N-terminus of the helix to the His of the dyad [19,50,51]. The first stage of catalysis is mediated by the highly active thiolate ion of the Cys. Biochemical studies on ervatamins in our laboratory and in the literature [42] show high activity of ervatamin-A among the ervatamins towards synthetic peptides and protein substrates. This phenomenon is difficult to explain from the structures, especially so when the

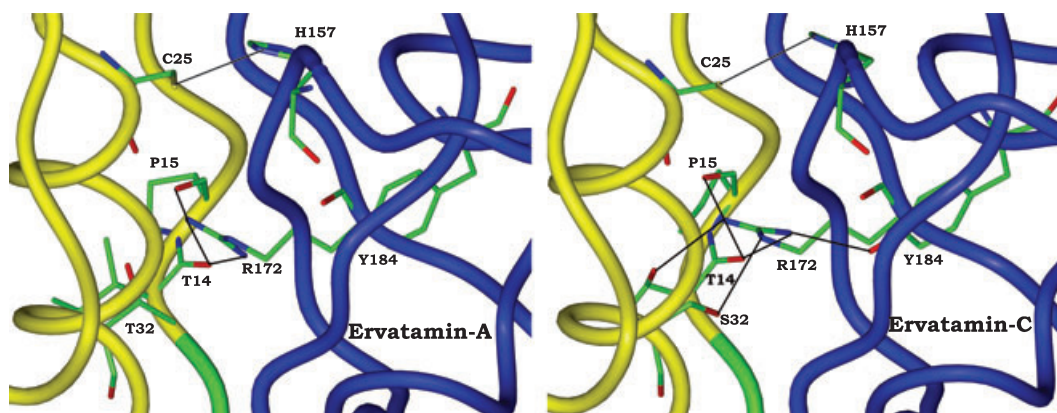


Fig. 5. The interdomain cleft of ervatamin-A and ervatamin-C in the same orientation. Ribbons colored yellow, blue and green represent domain L, domain R and the interdomain crossover, respectively. The catalytic dyad Cys25 and His157 is marked.

Table 2. The sequence and the dipole moments of the central helix were calculated by QUANTA (Accelrys Inc.). The residue at position 32 is in bold.

Enzyme	Sequence	Dipole moment (Debye)
Ervatamin-A (average from two molecules of the asymmetric unit)	CWAFSTV T TVESINQIRT	52.36
Ervatamin-C (average from two molecules of the asymmetric unit)	CWAFSTV S TVESINQIRT	49.39
Ervatamin-B	CWAFSAV A AVESINKIRT	51.06

sequence of ervatamin-A has 90% identity with that of ervatamin-C and has only one substitution at the S2 pocket within a 10 Å sphere around the catalytic Cys. This substituted residue at position 67 contributes to the S2–P2 interactions, as discussed above. The electrostatic surface calculated by GRASP [52] near the active site is also similar in the two enzymes. However, if we look carefully into the structure, we observe an important substitution in ervatamin-A that, although not near the active site, can influence the rate of catalysis in ervatamin-A. A Ser → Thr substitution at position 32 in the helix containing the catalytic Cys25 at its N-terminus has two effects in the structure of ervatamin-A as compared to ervatamin-C. The Thr32O γ in ervatamin-A points towards the helix and makes a hydrogen bond with the backbone oxygen atom of residue 28 of the same helix (Fig. 5), which is within the same turn of the helix starting from Cys25. Although residue 32 is not in the close vicinity of the catalytic dyad, it enhances the dipole moment of this particular helix in the case of ervatamin-A (Table 2), which may promote/stabilize the catalytic ion pair CysS γ^- HisIm $^+$, a prerequisite for catalysis in this class of enzymes. On the other hand, the Ser32O γ of ervata-

min-C pointing away from the helix forms a hydrogen bond with Arg172Ne from the other domain, and helps to form an intricate interdomain hydrogen bond network involving Ser32, Arg172, Tyr184, Thr14 and Pro15 (Fig. 5). A similar type of interdomain network mediated by the guanidinium group of the side chain of Arg172 (ervatamin-A numbering) is observed in ervatamin-B [35]. Owing to the presence of a bulkier residue, Thr, at position 32 of ervatamin-A, the guanidinium group of the corresponding Arg orients differently, and as a result, the interdomain hydrogen bond network is lost, and it consequently gains interdomain plasticity. The proteolytic activity of the papain-like cysteine proteases involves a number of steps to release the final product [53], and in these intermediate steps, conformational changes occur in the substrate. Conformational flexibility of the active site of the enzyme is thus required to allow/accommodate the conformational changes of the substrates, which are proteins or peptides of varying length and sequence. As the active site for this class of enzymes is at the interface of the two domains, interdomain plasticity plays a role in the activity of the enzyme. In the case of ervatamin-A, polarizability and active site plasticity can therefore be considered to be the primary factors responsible for its observed high activity.

Experimental procedures

Purification, enzyme–inhibitor complex formation, crystallization, and data collection

Protein purification from the latex of *Er. coronaria* was carried out as described previously [36–38]. Each of the three ervatamins (A, B and C) was reversibly inhibited by sodium tetrathionate during the purification. For enzyme–inhibitor complex formation, 1 mL of protein suspension

Table 3. Summary of diffraction data collection and refinement statistics. Values in parentheses are for the outer resolution shell.

	Ervatamin-A		Ervatamin-C
Crystal data			
Space group	Processed in P2 ₁	Processed in C222 ₁	P2 ₁ 2 ₁ 2 ₁
Unit cell parameters (Å, °)			
a	31.168	31.138	43.37
b	105.587	144.614	81.46
c	73.927	105.494	131.45
β	101.961		
No. molecules/asymmetric unit	2	1	2
Resolution (Å) (high resolution shell)	100–2.5 (2.56–2.5)	100–2.5 (2.56–2.5)	30–2.70 (2.79–2.7)
Data collection range (°)	115	–	82
No. observed reflections	34 362	34 609	41 609
No. unique reflections	14 392	8319	11 279
Completeness (%)	85.8 (81.4)	95.2 (92.8)	93.40 (94.60)
I/σ (I)	5.4 (3.0)	4.0 (2.5)	4.40 (1.90)
R _{merge} ^a (%)	7.43 (14.63)	10.46 (18.73)	10.51 (29.17)
Refinement statistics			
	With P2 ₁ dataset and using twinning options		
Resolution range (Å)	30–2.85		15–2.70
No. protein atoms	3208		3234
No. solvent molecules	40 (water)		166 (water), 6 (SO ₄ ²⁻)
No. ligand atoms	50 (E-64), 4 (β-ME)		50 (E-64)
R-factor ^b (%)	23.97		19.73
R _{free} ^b (%)	27.02		23.38
rmsd bond lengths (Å)	0.009		0.006
rmsd bond angles (°)	1.55		1.41
Ramachandran statistics			
Core region (%)	97.1		100.0
Generously allowed region (%)	2.3		0.00
Disallowed regions (%)	0.6		0.00

^a $R_{\text{merge}} = \sum |Ih - \langle Ih \rangle| / \sum (Ih)$, where $\langle Ih \rangle$ is the average intensity of reflection h and symmetry-related reflections.

^b $R = R_{\text{free}} = \sum ||F_o| - |F_c|| / \sum |F_o|$ calculated for reflections of the working set and test sets, respectively.

($\sim 3 \text{ mg}\cdot\text{mL}^{-1}$) in 50 mM Tris/Cl and 1 mM EDTA was first activated at 22 °C for 15 min by 50 mM β-mercaptoethanol (β-ME), and the extra activator was removed by ultrafiltration using a YM-10 membrane. The protein was then mixed with an equal volume of 0.4 mg·mL⁻¹ E-64 solution in the same buffer, and incubated at 22 °C for 3 h. The extra inhibitor was removed, and the complex was concentrated by ultrafiltration on a YM-10 membrane. The total inhibition of the ervatamins was checked biochemically against azocasein as a substrate. Attempts were made to crystallize these protease-inhibitor complexes, and crystals could be grown by the hanging-drop vapor diffusion method at room temperature, using the same conditions as described previously [36,54], with 12% glycerol as cryoprotectant. However, for the ervatamin-C complex, diffraction-quality crystals of the complex were obtained only with ammonium sulfate as precipitant; the other condition, described by Chakrabarti *et al.* [54], using monobasic potassium phosphate as salt, did not produce any crystal. Diffraction-quality crystals of ervatamin-B-E-64 could not be grown. Diffraction data for the other two complexes were collected in-house with the MAR345dtb system and a BRUKER FR591 rotating anode generator equipped with

the Osmic Confocal Max-Flux optic system. The data for ervatamin-A-E-64 and ervatamin-C-E-64 were collected at 100 K with a crystal-to-detector distance of 200 mm. Both sets of data were processed with the AUTOMAR program suite (<http://www.marresearch.com/automar>), and the data statistics are listed in Table 3.

Indexing and scaling of the ervatamin-A-E-64 dataset was possible in primitive monoclinic P2 and C-centered orthorhombic C222 settings with acceptable data statistics for both cases (Table 3). Although molecular replacement with ervatamin-C (90% identity) as a search model did work in the higher space group (C222₁), *R*-factors continued to remain high during refinement, and poor electron density was observed in some parts of the model. The high *R*-factor in space group C222₁ and the relationship of $c \cos \beta = -a/2$ in the lower space group [55–57] led us to suspect that the crystals might be pseudo-merohedrally twinned, where a twinning operator acted as a symmetry operator leading to a pseudo-higher space group. Different twinning tests using the programs CNS [58] and DETWIN and SFCHECK in the CCP4 suite [59] confirmed that the data were twinned, and allowed us to calculate the twinning fraction (0.394) (Fig. 6A,B) and the twinning operator

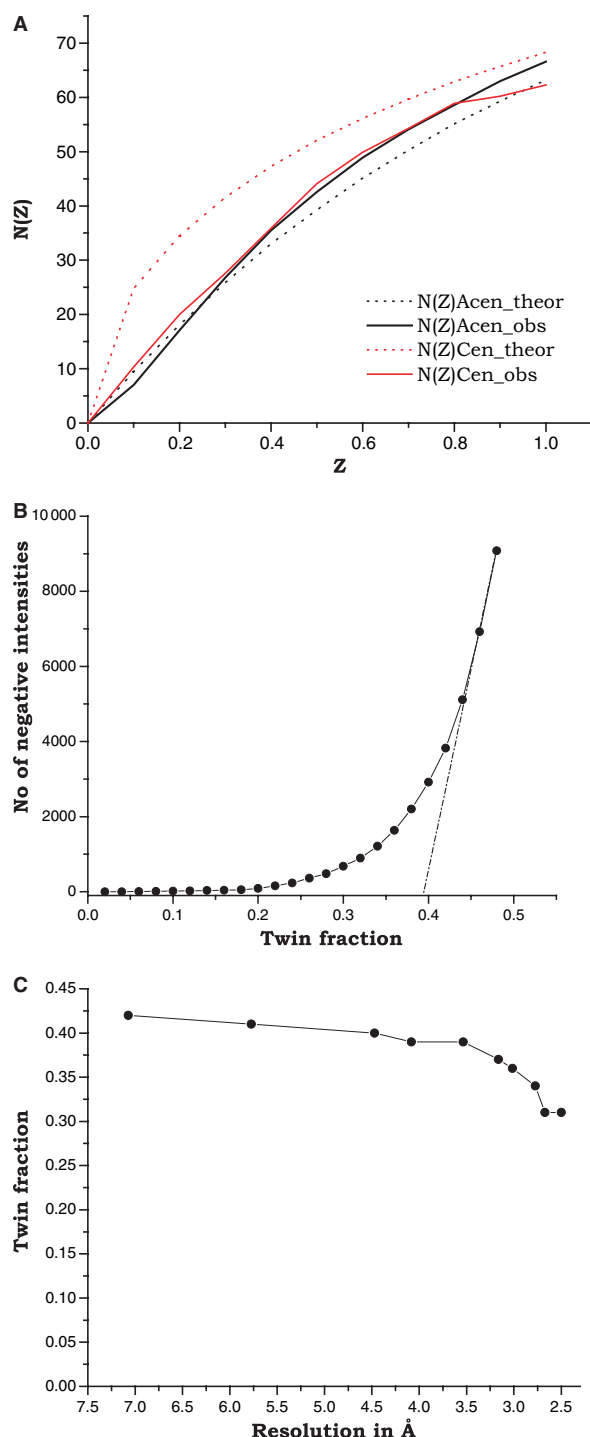


Fig. 6. (A) Cumulative intensity distribution of $Z = I/\langle I \rangle$, where I is the intensity for the centric (red lines) and acentric (black lines) reflections. The dashed lines and continuous lines show theoretical and observed distributions, respectively. The sigmoidal shape of the distribution of the acentric reflections (black continuous) indicates potential twinning. (B) Estimation of the twin fraction α from a Britton plot. The plot was calculated using the twinning operator $-h, -k, h + l$. (C) Variation of the twin fraction as a function of resolution.

$(-h, -k, h + l)$. The primitive monoclinic space group was thus identified as the true space group for ervatamin-A–E-64 complex data, and was used subsequently in structure solution and refinement.

Structure determination and refinement

The structure of ervatamin-A–E-64 was determined by the molecular replacement method using the program AMORE [60] implemented in the CCP4 suite [59], with the coordinates of ervatamin-C (Protein Data Bank ID: 2PNS) as the search model, keeping the mismatched residues as Ala, and using the diffraction data processed in the monoclinic space group. The Matthews coefficient ($2.5 \text{ \AA}^3 \cdot \text{Da}^{-1}$) suggested two molecules in the asymmetric unit. Molecular replacement was tried for space groups P2 and P2₁; a lower R -factor (39.9%), higher correlation coefficient (54.9%) and reasonable crystal packing confirmed the space group P2₁. Rigid body, positional and B -factor refinement using CNS [58] in the resolution range 30–2.9 Å, followed by electron density fitting using QUANTA (Accelrys Inc., San Diego, CA, USA), gave an R -factor of 32.3% and R_{free} of 34.7%. Refinement practically stalled at this stage, and from here we refined the structure by considering the twinning options in CNS [58]. A twinning factor of 0.4 and twinning operator $-h, -k, h + l$ were used during the course of refinement. A test set of reflections (6%) for cross-validation was chosen such that twin-related reflections were maintained together in either the test set or the working set. Although the data could be processed up to 2.5 Å, there were some practical problems in using the full dataset in the refinement – the twin fraction α varied from 0.42 to 0.31 with increasing resolution, and sharply decreased beyond 3 Å (Fig. 6C). CNS [58] does not have provision for incorporating a resolution-dependent twin fraction and refinement of the same. An average value of $\alpha = 0.4$ was used, which gave the lowest R -factor using data up to 2.85 Å. After a few cycles of positional refinement using the program CNS [58], the E-64 molecule was fitted in the $|F_o| - |F_c|$ electron density map (2.5σ). A covalently linked β -ME molecule, possibly trapped during complex formation, could be located near the free Cys108 in one molecule of the asymmetric unit, which was fitted in the map and refined subsequently. Finally, incorporation of 40 water molecules using the option X-SOLVE of QUANTA (Accelrys Inc.) yielded an R -factor of 23.97% and an R_{free} of 27.02%.

As the crystals of ervatamin-C–E-64 were isomorphous with the thiosulfate-inactivated enzyme (Protein Data Bank ID: 2PNS), the coordinates of the native molecule without the solvent molecules and the thiosulfate moiety were used for rigid body refinement using the program CNS [58]. In total, 5% of reflection data were set aside for R_{free} calculations. The $|F_o| - |F_c|$ electron density map (2.5σ) clearly indicated the presence of the E-64 molecule at the enzyme active site of both molecules of the asymmetric unit, and

the inhibitor was fitted in the electron density. As ammonium sulfate was used during crystallization, six SO_4^{2-} were located here, instead of the PO_4^{2-} reported in the thiosulfate-activated ervatamin-C structure (Protein Data Bank ID: 2PNS), which was crystallized in the presence of monobasic potassium phosphate as salt [27]. A few rounds of positional refinement, fitting, and introduction of water molecules and SO_4^{2-} , followed by manual fitting of the model by QUANTA (Accelrys Inc.), led to an R -factor of 19.78% and R_{free} of 23.42%. The final structure, after 20 cycles of group B -factor refinement using the same program, converged to an R -factor of 19.73% and R_{free} of 23.38%. Data and refinement statistics are given in Table 3. The stereochemistries of the final models of ervatamin-A and ervatamin-C in the complexes were checked by PROCHECK [61]. At the final stage of refinement, omit-maps covering the E-64 region in both the ervatamins were calculated (supplementary Fig. S1). For ervatamin-C–E-64, low-resolution data were excluded to improve map quality (supplementary Fig. S2). Molecular images were generated by using the programs INSIGHTII (MSI Inc.), QUANTA (Accelrys Inc.) and PYMOL (<http://pymol.sourceforge.net>). The coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 3BCN and 2PRE for the two crystal structures ervatamin-A and ervatamin-C, both complexed with E-64.

Kinetic measurements using chromogenic peptides

All the chromogenic substrates (Table 1) containing p -nitroanilide (pNA) were purchased from Sigma. Enzyme assays were performed in 50 mM Tris/Cl buffer (pH 8.0) containing 1 mM EDTA and 0.1% Brij35. Before addition of the substrate, proteases (0.1–0.5 μM) were preactivated for 15 min at 30 °C in the presence of 5 mM β -ME. Substrates were prepared in the same buffer as above. The time of activation, enzyme concentration and substrate concentration range for each enzyme–substrate combination were standardized. Reactions were initiated by addition of an equal volume of two-fold concentrated substrate to the active enzyme mixture, and incubation at room temperature was continued for the appropriate time. Liberated pNA was monitored continuously at 410 nm on a UV–visible spectrophotometer (Nicolet Evolution 100; Thermo Electron Corporation, Rockville, MD, USA). The range of substrate concentrations and the time of incubation were standardized for each substrate. An extinction coefficient of 8800 at 410 nm for pNA was used for the calculations. The software GRAPHPAD PRISM (<http://www.graphpad.com/prism>) was used to calculate the K_m and V_{max} values by nonlinear fitting of the Michaelis–Menten saturation curve. The k_{cat} values were determined by using the equation $k_{\text{cat}} = V_{\text{max}}/[\text{E}]_T$. $[\text{E}]_T$ is the total concentration of the active enzyme, the values of which were measured by active site titration with E-64 using appropriate substrates containing pNA.

Measurement of IC_{50} value of E-64

Enzymes were preactivated in the previously mentioned assay buffer at 37 °C for 5 min, using 5 mM β -ME. The optimum enzyme concentration was standardized to 0.25 μM for ervatamin-A and 0.5 μM for ervatamin-B and ervatamin-C. E-64 solution was added to the respective active protease solutions and incubated at 37 °C for 2–10 min. Residual activity was calculated with respect to the full activity of the enzyme ($\Delta A_{410 \text{ nm}} \cdot \text{min}^{-1}$) without any inhibitor under the same conditions described in the previous section, against the appropriate pNA peptide substrate for each of the ervatamins. A range of E-64 concentrations was used until the residual activity reached zero. Finally, the residual activity of the enzyme was plotted against the inhibitor concentration. From these plots, IC_{50} (the inhibitor concentration required for half-maximal inhibition) values of E-64 for the ervatamins were determined.

cDNA sequencing of ervatamin-A

Total RNA was extracted from the young leaves of *Er. coronaria* using the RNAqueous-4PCR Kit (Ambion, Austin, TX, USA) in accordance with the manufacturer's instructions, and quantized spectrophotometrically. Single-stranded cDNA was synthesized by RT using RevertAid M-MuLV reverse transcriptase with 3 μg of total RNA, 1 μg of oligo(dT)₁₈ primer, 1 mM dNTPs and 20 U of RNase inhibitor in a total volume of 20 μL . Then, the second cDNA strand was synthesized by PCR with Taq DNA polymerase. The protocol comprised a predenaturation step at 95 °C for 5 min, 35 cycles of two-step amplifications (the first five cycles comprised denaturation at 95 °C for 60 s, annealing at 50 °C for 90 s, and extension at 72 °C for 90 s, and the next 30 cycles comprised denaturation at 95 °C for 60 s, annealing at 60 °C for 90 s, and extension at 72 °C for 90 s) and a final extension step at 72 °C for 15 min. The forward primers for PCR were designed according to the N-terminal sequence of ervatamin-A [40]. The reverse primers were based on the conserved C-terminal sequence and guidelines from the electron density maps for the protein. Degeneracy of the primer sequences was fixed on the basis of frequency of occurrence of a particular DNA codon for an amino acid at a particular position for this family of plant cysteine proteases. The primers used were 5'-TTGCCTGAGCATGTTGATTGGAGAGCGAAAG-3' (forward) and 5'-GGGATAATAAGGTAATCTAGTGATTCCAC-3' (reverse). PCR-amplified products were purified from 1% agarose gel and ligated to the pTZ57R/T vector with the T/A cloning kit (Fermentas, Hanover, MD, USA). The ligation mixture was transformed into *Escherichia coli* XL1-Blue-competent cells. Recombinant clones carrying the insert were selected by blue–white screening. Plasmids containing DNA fragments were extracted with the QIAprep Spin Miniprep Kit

(Qiagen, Valencia, CA, USA). Isolated plasmids were verified by PCR and sequenced with forward and reverse M13 primers using the MEGABACE sequencing system (Amersham Biosciences, Piscataway, NJ, USA). The cDNA sequence and translated protein sequence were analyzed for similarity using BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) and CLUSTALW (<http://www.ebi.ac.uk/clustalw/>). The cDNA sequence of ervatamin-A has been deposited in the NCBI GenBank with the accession number EF591130.

Molecular dynamics simulation

The crystal structures of papain (Protein Data Bank ID: 9PAP) and ervatamin-A were used as starting models on which all the calculations were performed. The INSIGHT-II/DISCOVER package (MSI Inc., San Diego, CA, USA) was used for molecular dynamics study in a solvated condition of a 10 Å water layer along with the waters of crystallization associated with the biological unit, but omitting other ligands. The water molecules and generated hydrogen atom positions were refined in steps up to saturation. The properly optimized assembly was then subjected to simulation study. The temperature of the system was gradually increased to 300 K from 0 K in 120 ps with an increment of 50 K (for 20 ps) in each step. The system was then equilibrated for 1.4 ns at 300 K; the last 1 ns was considered as a product run for analysis. All the simulations were carried out with constant volume and temperature (NVT ensemble) through the velocity verlet integrator. A time-step of 2 fs for integration was used, and a bond constraint was applied by the RATTLE algorithm, a velocity version of SHAKE with a tolerance of $1e^{-5}$. Coordinates were saved at 1 ps intervals for the last 1 ns product run for analysis. All the simulations were carried out with the consistent valence force field, and the cell multipole method with a dielectric constant of 1 was used for nonbonded calculations. The 1 ns trajectory in each case was analyzed by the Analysis tool of the INSIGHTII/DISCOVER package (MSI Inc.).

Acknowledgements

This work was partially supported by the Department of Biotechnology and the Council of Scientific and Industrial Research, Government of India, with grants BT/PRO139/R&D/15/011/96 and 21/(0653)/06/EMR-II, respectively.

References

- Grudkowska M & Zagdanska B (2004) Multifunctional role of plant cysteine proteinases. *Acta Biochim Pol* **51**, 609–624.
- Schaller A (2004) A cut above the rest: the regulatory function of plant proteases. *Planta* **220**, 183–197.
- Beers EP, Jones AM & Dickerman AW (2004) The S8 serine, C1A cysteine and A1 aspartic protease families in Arabidopsis. *Phytochemistry* **65**, 43–58.
- Van der Hoorn RA & Jones JD (2004) The plant proteolytic machinery and its role in defence. *Curr Opin Plant Biol* **7**, 400–407.
- Turk D & Guncar G (2003) Lysosomal cysteine proteases (cathepsins): promising drug targets. *Acta Crystallogr D* **59**, 203–213.
- Steверding D, Caffrey CR & Sajid M (2006) Cysteine proteinase inhibitors as therapy for parasitic diseases: advances in inhibitor design. *Mini Rev Med Chem* **6**, 1025–1032.
- Kirk O, Borchert TV & Fuglsang CC (2002) Industrial enzyme applications. *Curr Opin Biotechnol* **13**, 345–351.
- Leisola M, Jokela J, Pastinen O, Turunen O & Schoemaker H (2002) Industrial use of enzymes. In *Encyclopedia of Life Support Systems (EOLSS)* (OOP Hänninen & M Atalay, eds.), pp. 1–25. EOLSS, Oxford, UK.
- Dardenne LE, Werneck AS, de Oliveira Neto M & Bisch PM (2003) Electrostatic properties in the catalytic site of papain: a possible regulatory mechanism for the reactivity of the ion pair. *Proteins* **52**, 236–253.
- Kamphuis IG, Drenth J & Baker EN (1985) Thiol proteases. Comparative studies based on the high-resolution structures of papain and actinidin, and on amino acid sequence information for cathepsins B and H, and stem bromelain. *J Mol Biol* **182**, 317–329.
- Carter CE, Marriage H & Goodenough PW (2000) Mutagenesis and kinetic studies of a plant cysteine proteinase with an unusual arrangement of acidic amino acids in and around the active site. *Biochemistry* **39**, 11005–11013.
- Ikeuchi Y, Katerelos NA & Goodenough PW (1998) The enhancing of a cysteine proteinase activity at acidic pH by protein engineering, the role of glutamic 50 in the enzyme mechanism of caricain. *FEBS Lett* **437**, 91–96.
- Gul S, Mellor GW, Thomas EW & Brocklehurst K (2006) Temperature-dependences of the kinetics of reactions of papain and actinidin with a series of reactivity probes differing in key molecular recognition features. *Biochem J* **396**, 17–21.
- Mellor GW, Thomas EW, Topham CM & Brocklehurst K (1993) Ionization characteristics of the Cys-25/His-159 interactive system and of the modulatory group of papain: resolution of ambiguity by electronic perturbation of the quasi-2-mercaptopyridine leaving group in a new pyrimidyl disulphide reactivity probe. *Biochem J* **290**, 286–296.
- Brocklehurst K, Brocklehurst SM, Kowlessur D, O'Driscoll M, Patel G, Salih E, Templeton W, Thomas

- E, Topham CM & Willenbrock F (1988) Supracrystallographic resolution of interactions contributing to enzyme catalysis by use of natural structural variants and reactivity-probe kinetics. *Biochem J* **256**, 543–558.
- 16 Brocklehurst K, Kowlessur D, O'Driscoll M, Patel G, Quenby S, Salih E, Templeton W, Thomas EW & Willenbrock F (1987) Substrate-derived two-protonic-state electrophiles as sensitive kinetic specificity probes for cysteine proteinases. Activation of 2-pyridyl disulphides by hydrogen-bonding. *Biochem J* **244**, 173–181.
 - 17 Willenbrock F & Brocklehurst K (1984) Natural structural variation in enzymes as a tool in the study of mechanism exemplified by a comparison of the catalytic-site structure and characteristics of cathepsin B and papain. pH-dependent kinetics of the reactions of cathepsin B from bovine spleen and from rat liver with a thiol-specific two-protonic-state probe (2,2'-dipyridyl disulphide) and with a specific synthetic substrate (*N*- α -benzyloxycarbonyl-L-arginyl-L-arginine 2-naphthylamide). *Biochem J* **222**, 805–814.
 - 18 Brocklehurst K, Mushiri SM, Patel G & Willenbrock F (1983) A marked gradation in active-centre properties in the cysteine proteinases revealed by neutral and anionic reactivity probes. Reactivity characteristics of the thiol groups of actinidin, ficin, papain and papaya peptidase A towards 4,4'-dipyridyl disulphide and 5,5'-dithiobis-(2-nitrobenzoate) dianion. *Biochem J* **209**, 873–879.
 - 19 Doran JD & Carey PR (1996) α -helix dipoles and catalysis: absorption and Raman spectroscopic studies of acyl cysteine proteases. *Biochemistry* **35**, 12495–12502.
 - 20 Menard R, Plouffe C, Khouri HE, Dupras R, Tessier DC, Vernet T, Thomas DY & Storer AC (1991) Removal of an inter-domain hydrogen bond through site-directed mutagenesis: role of serine 176 in the mechanism of papain. *Protein Eng* **4**, 307–311.
 - 21 Vernet T, Tessier DC, Khouri HE & Altschuh D (1992) Correlation of co-ordinated amino acid changes at the two-domain interface of cysteine proteases with protein stability. *J Mol Biol* **224**, 501–509.
 - 22 Schechter I & Berger A (1967) On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **27**, 157–162.
 - 23 Varughese KI, Ahmed FR, Carey PR, Hasnain S, Huber CP & Storer AC (1989) Crystal structure of a papain–E-64 complex. *Biochemistry* **28**, 1330–1332.
 - 24 Varughese KI, Su Y, Cromwell D, Hasnain S & Xuong NH (1992) Crystal structure of an actinidin–E-64 complex. *Biochemistry* **31**, 5172–5176.
 - 25 Schroder E, Phillips C, Garman E, Harlos K & Crawford C (1993) X-ray crystallographic structure of a papain–leupeptin complex. *FEBS Lett* **315**, 38–42.
 - 26 Khouri HE, Vernet T, Menard R, Parlati F, Laflamme P, Tessier DC, Gour-Salin B, Thomas DY & Storer AC (1991) Engineering of papain: selective alteration of substrate specificity by site-directed mutagenesis. *Biochemistry* **30**, 8929–8936.
 - 27 Guha Thakurta P, Biswas S, Chakrabarti C, Sundd M, Jagannadham MV & Dattagupta JK (2004) Structural basis of the unusual stability and substrate specificity of ervatamin C, a plant cysteine protease from *Ervatamia coronaria*. *Biochemistry* **43**, 1532–1540.
 - 28 Turk B, Turk V & Turk D (1997) Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors. *Biol Chem* **378**, 141–150.
 - 29 Turk V, Turk B & Turk D (2001) Lysosomal cysteine proteases: facts and opportunities. *EMBO J* **20**, 4629–4633.
 - 30 Turk B, Turk D & Turk V (2000) Lysosomal cysteine proteases: more than scavengers. *Biochim Biophys Acta* **1477**, 98–111.
 - 31 Turk D, Guncar G, Podobnik M & Turk B (1998) Revised definition of substrate binding sites of papain-like cysteine proteases. *Biol Chem* **379**, 137–147.
 - 32 Sumner IG, Harris GW, Taylor MA, Pickersgill RW, Owen AJ & Goodenough PW (1993) Factors effecting the thermostability of cysteine proteinases from *Carica papaya*. *Eur J Biochem* **214**, 129–134.
 - 33 Zucker S, Buttler DJ, Nicklin MJ & Barrett AJ (1985) The proteolytic activities of chymopapain, papain, and papaya proteinase III. *Biochim Biophys Acta* **828**, 196–204.
 - 34 Grzonka Z, Kasprzykowski F & Wiczak W (2007) Cysteine proteases. In *Industrial Enzymes: Structure, Function and Applications* (Polaina J & MacCabe AP, eds), pp. 181–195. Springer, Dordrecht, The Netherlands.
 - 35 Biswas S, Chakrabarti C, Kundu S, Jagannadham MV & Dattagupta JK (2003) Proposed amino acid sequence and the 1.63 Å X-ray crystal structure of a plant cysteine protease, ervatamin B: some insights into the structural basis of its stability and substrate specificity. *Proteins* **51**, 489–497.
 - 36 Chakraborty S, Biswas S, Chakrabarti C & Dattagupta JK (2005) Crystallization and preliminary X-ray diffraction studies of the cysteine protease ervatamin A from *Ervatamia coronaria*. *Acta Crystallogr F* **61**, 562–564.
 - 37 Kundu S, Sundd M & Jagannadham MV (2000) Purification and characterization of a stable cysteine protease ervatamin B, with two disulfide bridges from the latex of *Ervatamia coronaria*. *J Agric Food Chem* **48**, 171–179.
 - 38 Sundd M, Kundu S, Pal G & Jagannadham MV (1998) Purification and characterization of a highly stable cysteine protease from the latex of *Ervatamia coronaria*. *Biosci Biotechnol Biochem* **62**, 1947–1955.
 - 39 Ghosh R, Dattagupta JK & Biswas S (2007) A thermostable cysteine protease precursor from a tropical plant contains an unusual C-terminal propeptide: cDNA cloning, sequence comparison and molecular modeling studies. *Biochem Biophys Res Commun* **362**, 965–970.

- 40 Nallamsetty S, Kundu S & Jagannadham MV (2003) Purification and characterization of a highly active cysteine protease from the latex of *Ervatamia coronaria*. *J Protein Chem* **22**, 1–13.
- 41 Than ME, Helm M, Simpson DJ, Lottspeich F, Huber R & Gietl C (2004) The 2.0 Å crystal structure and substrate specificity of the KDEL-tailed cysteine endopeptidase functioning in programmed cell death of *Ricinus communis* endosperm. *J Mol Biol* **336**, 1103–1116.
- 42 Davy A, Sorensen MB, Svendsen I, Cameron-Mills V & Simpson DJ (2000) Prediction of protein cleavage sites by the barley cysteine endoproteases EP-A and EP-B based on the kinetics of synthetic peptide hydrolysis. *Plant Physiol* **122**, 137–145.
- 43 Bethune MT, Strop P, Tang Y, Sollid LM & Khosla C (2006) Heterologous expression, purification, refolding, and structural–functional characterization of EP-B2, a self-activating barley cysteine endoprotease. *Chem Biol* **13**, 637–647.
- 44 Choi KH, Laursen RA & Allen NN (1999) The 2.1 Å structure of a cysteine protease with proline specificity from ginger rhizome, *Zingiber officinale*. *Biochemistry* **38**, 11624–11633.
- 45 Bartlett GJ, Porter CT, Borkakoti N & Thornton JM (2002) Analysis of catalytic residues in enzyme active sites. *J Mol Biol* **324**, 105–121.
- 46 Gutteridge A & Thornton JM (2005) Understanding nature's catalytic toolkit. *Trends Biochem Sci* **30**, 622–629.
- 47 Dinakarandian D, Shenoy BC, Hilvert D, McRee DE, McTigue M & Carey PR (1999) Electric fields in active sites: substrate switching from null to strong fields in thiol- and selenol-subtilisins. *Biochemistry* **38**, 6659–6667.
- 48 Beveridge AJ (1996) A theoretical study of the active sites of papain and S195C rat trypsin: implications for the low reactivity of mutant serine proteinases. *Protein Sci* **5**, 1355–1365.
- 49 Costabel M, Vallejo DF & Grigera JR (2001) Electrostatic recognition between enzyme and inhibitor: interaction between papain and leupeptin. *Arch Biochem Biophys* **394**, 161–166.
- 50 Taylor MA, Baker KC, Connerton IF, Cummings NJ, Harris GW, Henderson IM, Jones ST, Pickersgill RW, Sumner IG & Warwicker J (1994) An unequivocal example of cysteine proteinase activity affected by multiple electrostatic interactions. *Protein Eng* **7**, 1267–1276.
- 51 Miranda JJ (2003) Position-dependent interactions between cysteine residues and the helix dipole. *Protein Sci* **12**, 73–81.
- 52 Nichols A, Sharp K & Honig B (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296.
- 53 Rullman JAC, Bellido MN & van Duijnen PT (1989) The active site of papain. *J Mol Biol* **206**, 101–118.
- 54 Chakrabarti C, Biswas S, Kundu S, Sundt M, Jagannadham MV & Dattagupta JK (1999) Crystallization and preliminary X-ray analysis of ervatamin B and C, two thiol proteases from *Ervatamia coronaria*. *Acta Crystallogr D* **55**, 1074–1075.
- 55 Rudolph MG, Wingren C, Crowley MP, Chien Y & Wilson IA (2004) Combined pseudo-merohedral twinning, non-crystallographic symmetry and pseudo-translation in a monoclinic crystal form of the γδ T-cell ligand T10. *Acta Crystallogr D* **60**, 656–664.
- 56 Lehtiö L, Fabrichniy I, Hansen T, Schönheit P & Goldman A (2005) Unusual twinning in an acetyl coenzyme A synthetase (ADP-forming) from *Pyrococcus furiosus*. *Acta Crystallogr D* **61**, 350–354.
- 57 Hakanpää J, Szilvay GR, Kaljunen H, Maksimainen M, Linder M & Rouvinen J (2006) Two crystal structures of *Trichoderma reesei* hydrophobin HFBI – the structure of a protein amphiphile with and without detergent interaction. *Protein Sci* **15**, 2129–2140.
- 58 Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS *et al.* (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D* **54**, 905–921.
- 59 Collaborative Computing Project (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D* **50**, 760–763.
- 60 Navaza J (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr A* **50**, 157–163.
- 61 Laskowski RA, MacArthur MW, Moss SD & Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26**, 283–291.

Supplementary material

The following supplementary material is available online:

Fig. S1. $F_o - F_c$ omit map, contoured at 1.5σ , in the E-64 region. (A) Ervatamin-A. (B) Ervatamin-C.

Fig. S2. Electron density ($2F_o - F_c$) map of the S2 subsite. (A) Ervatamin-A–E-64 contoured at 1.0σ . (B) Ervatamin-C–E-64 contoured at 1.2σ .

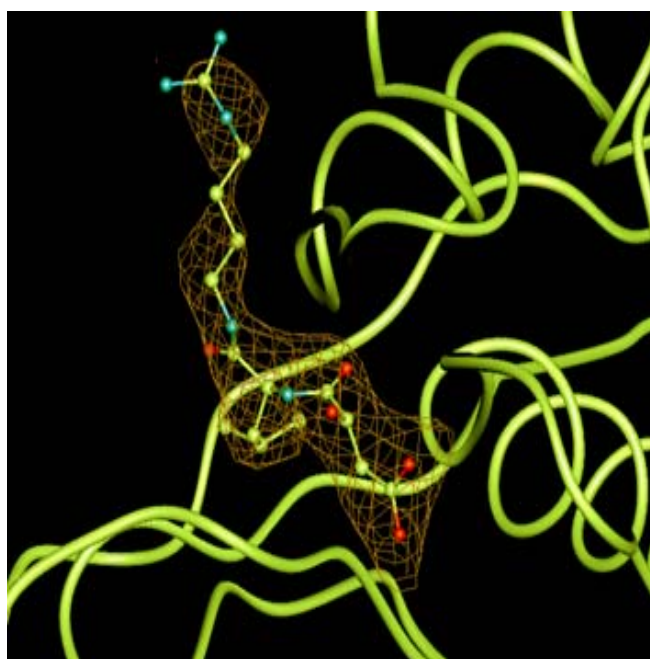
Table S1. Electrostatic and hydrophobic interactions of E-64 with ervatamins.

This material is available as part of the online article from <http://www.blackwell-synergy.com>

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

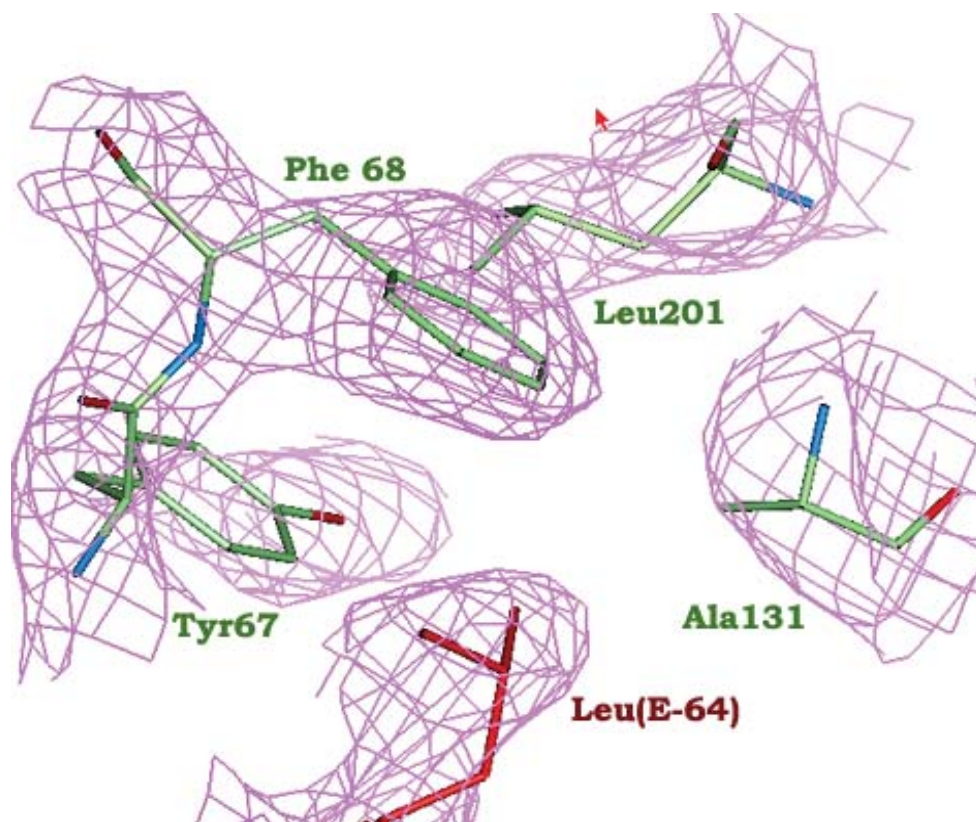


(a)

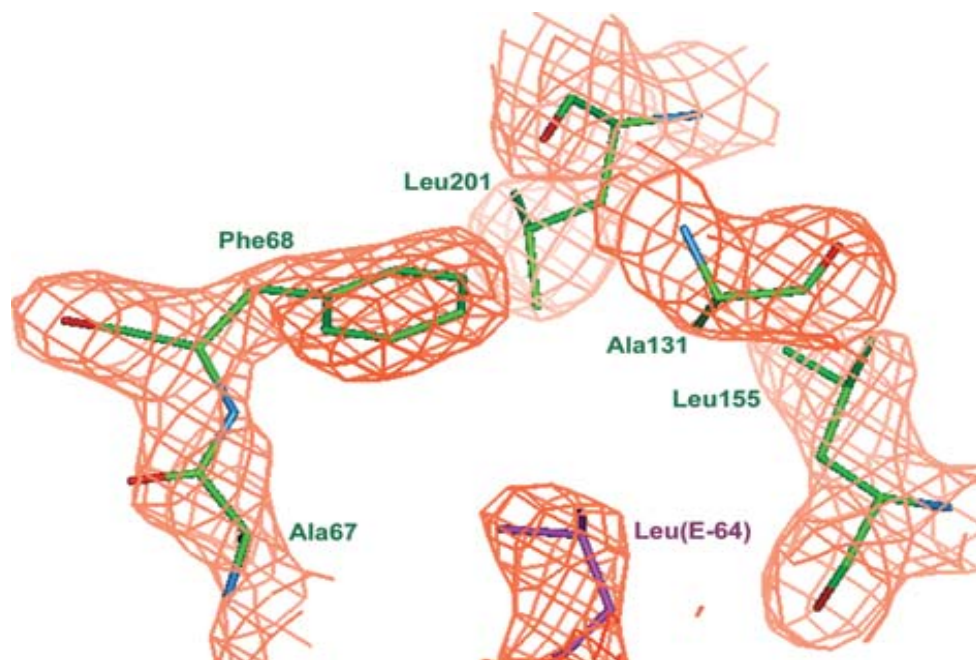


(b)

FIGURE S1: Fo-Fc omit map, contoured at 1.5σ , in the E-64 region. (a) Ervatamin-A, (b) Ervatamin-C.



(a)



(b)

FIGURE S2: Electron density (2Fo-Fc) map of the S2 subsite. (a) Ervatamin-A- E-64 contoured at 1.0σ and (b) Ervatamin-C- E-64 complex contoured at 1.2σ .

Table S1. Electrostatic and hydrophobic interactions of E-64 with Ervatamins

Type of interaction	Atom of the inhibitor	Residue of Ervatamin-A	Distance in Å	
			Molecule A	Molecule B
Hydrogen bond and electrostatic short contacts	O1	GLN 19 Nε2	3.42	3.07
	O1	SER 24 N	3.10	3.73
	O1	CYS 25 N	3.41	3.00
	O4	GLY 66 O	3.15	2.75
	O4	GLY 66 N	2.84	2.87
	N1	ASN 156 O	2.90	2.95
	O2	HIS 157 Nδ1	3.31	-
	N2	GLY66 O	3.79	-
Covalent bond	Sγ	C2	1.81	1.80
Hydrophobic interactions	C8,C9,C10 (P2)	TYR 67	< 5	<5
	C8,C9,C10 (P2)	PHE 68		
	C8,C9 (P2)	ALA 131		
	C9 (P2)	LEU 155		
	C12,C14 (P3)	GLY 65		

Type of interaction	Atom of the inhibitor	Residue of Ervatamin-C	Distance in Å	
			Molecule A	Molecule B
Hydrogen bond and electrostatic short contacts	O2	GLN 19 Nε2	3.20	3.50
	O1	GLN 19 Nε2	2.85	2.55
	O1	GLN 19 Oε1	3.48	-
	O1	SER 24 N	3.22	3.09
	O1	CYS 25 N	3.40	2.95
	O4	TRP 26 N	-	3.51
	N3	HIS 61 Nδ1	3.45	-
	N4	HIS 61 Nδ1	3.43	-
	N4	HIS 61 Nε2	2.92	-
	N5	HIS 61 Nε2	-	3.38
	N3	LYS 64 O	-	3.33
	O4	GLY 66 N	3.40	3.32
	N2	GLY 66 O	-	3.51
	N1	ASN 156 O	3.43	-
	O2	HIS 157 Nδ1	2.59	2.67
Covalent bond	Sγ	C2	1.82	1.82
Hydrophobic interactions	C10	TRP 26	< 5	< 5
	C7	TRP 26		
	C9	TRP 26		
	C8	PHE 68		
	C9	PHE 68		
	C10	PHE 68		
	C8 (P2)	ALA 131		
	C9	ALA 131		
	C10	ALA 131		
	C7	LEU 155		
	C10	LEU 155		
	C9	LEU 155		
	C7	ASN 156		
	C10	ASN 156		
	C14, C15, C16 (P3)	HIS 61		