

Isolation, characterization, sequencing and crystal structure of charybdin, a type 1 ribosome-inactivating protein from *Charybdis maritima* agg.

Eleftherios Touloupakis^{1,*}, Renate Gessmann^{2,*}, Kalliopi Kavelaki¹, Emmanuil Christofakis¹, Kyriacos Petratos² and Demetrios F. Ghanotakis¹

1 Department of Chemistry, University of Crete, Greece

2 Institute of Molecular Biology and Biotechnology (IMBB), FORTH, Heraklion, Crete, Greece

Keywords

active site; *Charybdis maritima* agg.; ribosome-inactivating protein; sequence; structure

Correspondence

D. F. Ghanotakis, Department of Chemistry, University of Crete, PO Box 1470, 71409, Heraklion, Crete, Greece
Tel: +30 2810545034
Fax: +30 2810393601
E-mail: ghanotakis@chemistry.uoc.gr

*These authors contributed equally to this work

Database

DNA sequence data from this article have been deposited with the GenBank data library under accession number DQ323742, protein sequence data with UniProt Knowledgebase under accession number P84786, and the crystal structure with the PDB database under accession code 2B7U

(Received 3 March 2006, revised 18 April 2006, accepted 19 April 2006)

doi:10.1111/j.1742-4658.2006.05287.x

Charybdis maritima agg. (previously *Urginea maritima* agg.) commonly known as squill, is a poisonous plant that belongs to the family of Liliaceae. It is a large, onion-like plant that grows wild on the coast around the Mediterranean Sea.

Both varieties of squill (red and white) have fibrous roots proceeding from the base of a large and tunicated

A novel, type 1 ribosome-inactivating protein designated charybdin was isolated from bulbs of *Charybdis maritima* agg. The protein, consisting of a single polypeptide chain with a molecular mass of 29 kDa, inhibited translation in rabbit reticulocytes with an IC₅₀ of 27.2 nM. Plant genomic DNA extracted from the bulb was amplified by PCR between primers based on the N-terminal and C-terminal sequence of the protein from dissolved crystals. The complete mature protein sequence was derived by partial DNA sequencing and terminal protein sequencing, and was confirmed by high-resolution crystal structure analysis. The protein contains Val at position 79 instead of the conserved Tyr residue of the ribosome-inactivating proteins known to date. To our knowledge, this is the first observation of a natural substitution of a catalytic residue at the active site of a natural ribosome-inactivating protein. This substitution in the active site may be responsible for the relatively low *in vitro* translation inhibitory effect compared with other ribosome-inactivating proteins. Single crystals were grown in the cold room from PEG6000 solutions. Diffraction data collected to 1.6 Å resolution were used to determine the protein structure by the molecular replacement method. The fold of the protein comprises two structural domains: an $\alpha + \beta$ N-terminal domain (residues 4–190) and a mainly α -helical C-terminal domain (residues 191–257). The active site is located in the interface between the two domains and comprises residues Val79, Tyr117, Glu167 and Arg170.

bulb. The bulb contains the pharmacologically active compounds of *Charybdis maritima* agg., which are bufadienolides and cardiac steroid glycosides. Squill has been used medicinally since ancient times. In human phytotherapy, the dried bulb of the white variety is used orally as a diuretic, emetic, expectorant and cardiotoxic [1].

Abbreviations

RIP, ribosome-inactivating protein.

Ribosome-inactivating proteins (RIPs) are a heterogeneous group of enzymes, identified in plants, bacteria and fungi. They are distributed throughout the plant kingdom and are active against ribosomes from different species, although the level of activity depends on the source of the RIP and of the ribosome. There are many reports that RIPs induce apoptosis [2,3]. The main application has been focused on the construction of chimeric molecules known as immunotoxins for cancer immunotherapy [4].

RIPs are RNA N-glycosidases that inactivate ribosomes by the selective cleavage of an adenine residue at a conserved site of the 28S rRNA, arresting protein synthesis. The nature of the enzymatic modification of ribosomes was discovered by Endo & Tsurugi [5]. Interest in RIPs has arisen from their potential medical and therapeutic applications, as several of these proteins have been found to be more toxic towards tumor cells than to normal cells [6].

RIPs have been classified into three types based on their primary structures [7]. Type 1 RIPs are single-chain proteins which contain the ribosome-inactivating entity, with a molecular mass of ≈ 30 kDa. Type 2 RIPs are two-chain proteins which consist of an A-chain, functionally equivalent to type 1, linked through a disulfide bond to a lectin-like B-chain which promotes uptake by the cell. Type 3 RIPs are composed of a single chain containing an extended C-terminal domain with unknown function. Although type 1 and type 2 RIPs are equally effective inhibitors of protein synthesis in cell extracts, the absence of the B-chain in type 1 does not allow the protein to bind and enter cells with high efficiency. Therefore they are considerably less cytotoxic [8].

In this study, we describe the purification, characterization and structural analysis of charybodin, a novel 29-kDa type 1 ribosome-inactivating protein, from bulbs of the white variety of *C. maritima* agg.

Results

Charybodin was purified from *C. maritima* agg. bulbs by using a combination of hydrophobic and ion-exchange chromatography (see Experimental procedures). It is interesting to note that the *C. maritima* agg. bulbs contain extremely high quantities of the charybodin protein. The initial extract contained mainly charybodin and very small amounts of other proteins, which were only observed when the gel was overloaded. The main impurities were pigments and other small hydrophobic molecules. The objective of the purification protocol was not only to remove traces of other proteins, but also smaller molecules, which

caused problems during the characterization and crystallization of charybodin. The yield of the purified protein was 150–200 mg protein per 100 g of bulbs. Charybodin appeared as a single band with a molecular mass of 29 kDa on SDS/PAGE (Fig. 1A). The pI was found by isoelectric focusing PAGE to be ≈ 7 (data not shown). The pI calculated from the derived sequence (see below) was 5.8.

Translation inhibition of rabbit reticulocytes by charybodin

The *in vitro* translation inhibitory effect of charybodin was analyzed. As shown in Fig. 1B, charybodin inhibits the rabbit reticulocyte translation system. The calcula-

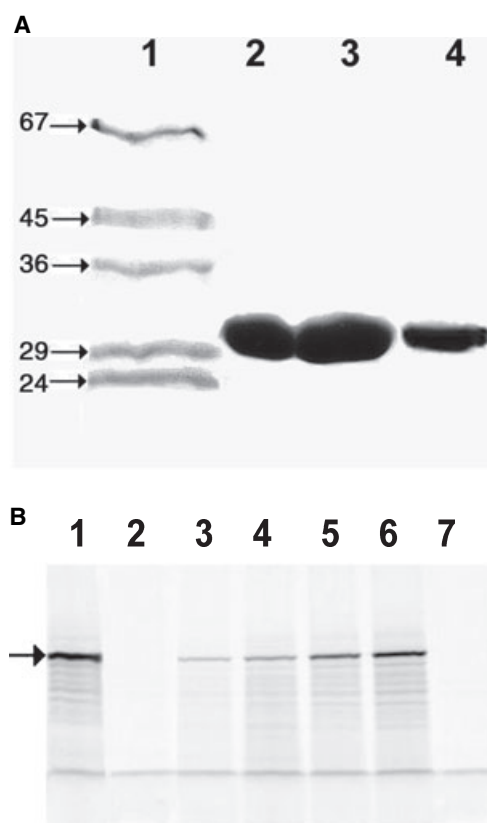


Fig. 1. Charybodin purification and biochemical properties. (A) (Lane 1) molecular mass markers (in kDa); (lane 2) crude extract containing charybodin; (lane 3) purified protein; (lane 4) protein crystal (SDS/12% polyacrylamide gel). (B) Inhibition of *in vitro* protein synthesis by charybodin. The rabbit reticulocytes were treated with different concentrations of charybodin (13.8–552 nM). The ^{35}S -labeled Met was used to label the product luciferase (arrow). Samples from the reactions were resolved by SDS/PAGE (12% gel) and analyzed by autoradiography. (lane 1) reticulocytes without charybodin; (lane 2) with 552 nM charybodin; (lane 3) with 138 nM charybodin; (lane 4) with 69 nM charybodin; (lane 5) with 34.5 nM charybodin; (lane 6) with 13.8 nM charybodin; (lane 7) with 13.3 nM saporin.

ted IC_{50} of 27.2 nM for charybdin is at least 100 times higher than the value (0.25 nM) reported for saporin L1 [9]. IC_{50} represents the concentration of charybdin that inhibited *in vitro* protein synthesis by 50%.

DNA sequence and derived amino-acid sequence

The DNA sequence and the derived amino-acid sequence are shown in Fig. 2. The amino-acid sequence shows homology to RIPs and exhibits identity of 46.7–37.7% with the musarmins [10], 36.6% with the RIP of *Hyacinthus orientalis* (UniprotKB/TrEmbl code Q677A1), 28.4% with pulchellin [11], which is highly homologous to abrin, and 25.3% with ricin. The sequence similarities were calculated using the program BLAST [12]. There are 15 identical residues among seven sequences (charybdin, musarmin I and III, *Hyacinthus*, *Iris holl*, pulchellin and ricin), which share high sequence similarity. Three of the four key residues of the active site, Tyr123, Glu177 and Arg180 (ricin numbering [13]), are among the identical residues. Thus, it is interesting to note that the fourth residue, which is an

invariant Tyr80 (ricin numbering) among more than 360 RIP sequences known to date, is replaced by Val in charybdin. To exclude the possibility of a local geographical mutation, DNA sequencing was also carried out on a plant collected from another region of Crete, and this residue substitution was confirmed. There are no N-glycosylation sites in the deduced sequence.

Quality of the model

The high quality of the collected diffraction data and the resulting refinement of the structure are shown in Table 1. A thin section of the structure with its electron-density map is shown in Fig. 3. A total of 232 out of 257 amino-acid residues fit very well in the electron-density map. Exceptions are certain regions on the surface of the molecule, which are quite flexible, as reflected in the higher thermal parameter values. These regions are the N-terminus and three turns comprising amino-acid residues 48–56, 96–102 and 183–188. Residues 1–3 and 99–101 are not included in the final refined model.

1	AGY	CAR	TGY	AAR	GCN	ATG	ACG	GTG	AAG	TTC	ACA	GTA	GAG	CTG	GAT	45
1	S	Q	C	K	A	M	T	V	K	F	T	V	E	L	D	15
46	ATC	GAG	AGG	CTC	ACC	GGC	CAA	ACC	TAC	ACC	GAY	TTC	ATC	AAG	AAC	90
16	I	E	R	L	T	G	Q	T	Y	T	D	F	I	K	N	30
91	CTT	CGT	CGT	AGT	CTG	GCA	ACC	TGG	TAC	TTG	CAC	GGT	GTC	CCA	GTG	135
31	L	R	R	S	L	A	T	W	Y	L	H	G	V	P	V	45
136	CTG	CCC	CTC	TAC	AAC	CAR	GAA	GCA	GAC	CCT	CGT	GGC	TTC	GAC	CTC	180
46	L	P	L	Y	N	Q	E	A	D	P	R	G	F	D	L	60
181	AAG	CTC	ACG	TTC	CGA	GGG	CAA	GTG	ACC	ACC	GTG	AGG	ATC	CAT	CGC	225
61	K	L	T	F	R	G	Q	V	T	T	V	R	I	H	R	75
226	GAC	GAC	CTC	GTC	TTG	AGA	GGT	TAC	CAG	ATG	CAG	GGA	GCT	GGC	AAG	270
76	D	D	L	V	L	R	G	Y	Q	M	Q	G	A	G	K	90
271	TGG	TTG	GAG	CTC	GAG	AGG	CCV	AGC	ACC	CAA	ACT	GGC	CAC	CTG	ATC	315
91	W	L	E	L	E	R	P	S	T	Q	T	G	H	L	I	105
316	GAG	GGA	TCC	GAG	CTG	CTG	GAG	TTC	GGT	CCG	AGC	TAC	GAG	GAA	CTG	360
106	E	G	S	E	L	L	E	F	G	P	S	Y	E	E	L	120
361	GCA	GCC	GCT	GCA	CAG	CAG	GAT	ATA	CTT	GAC	ATA	TCC	TAC	AAC	AAG	405
121	A	A	A	A	Q	Q	D	I	L	D	I	S	Y	N	K	135
406	AAC	GCA	CTC	CAA	GAY	GCC	GTA	TCC	AAG	CTA	GCC	GTC	TCC	ACA	AAC	450
136	N	A	L	Q	D	A	V	S	K	L	A	V	S	T	N	150
451	ACC	CGG	GAC	AGA	GCT	AGA	TCT	CTC	ATC	GTC	GTG	TCC	CAG	ATG	TTC	495
151	T	R	D	R	A	R	S	L	I	V	V	S	Q	M	F	165
496	TGC	GAA	GCC	ACC	AGG	TTC	GTC	GAC	ATT	GCT	AAT	CAT	TTC	GCT	TTC	540
166	C	E	A	T	R	F	V	D	I	A	N	H	F	A	F	180
541	AAT	CTC	GAG	AGC	TCA	GAG	CCC	GTG	AAG	CTG	CCT	CAG	TGG	ATG	CAG	585
181	N	L	E	S	S	E	P	V	K	L	P	Q	W	M	Q	195
586	AAT	GAT	CTT	GAG	AAG	AAC	TGG	GTG	CGA	TTT	TCT	TTC	ATG	ATT	CTC	630
196	N	D	L	E	K	N	W	V	R	F	S	F	M	I	L	210
631	AAG	TCG	AAT	GCG	GAT	CCC	TGC	TAC	AAG	TTC	GAG	CCA	CAG	ACG	ATT	675
211	K	S	N	A	D	P	C	Y	K	F	E	P	Q	T	I	225
676	TAC	GGG	AAA	ATA	ATT	AAG	ACT	GCC	GAT	GAG	CTC	CTT	AAC	TTC	CTT	720
226	Y	G	K	I	I	K	T	A	D	E	L	L	N	F	L	240
721	GGC	ATC	GTA	GAR	CAR	CAY	CCN	GAY	ACN	CGN	TCT	CCN	CCN	TGY	GCN	765
241	G	I	V	E	Q	H	P	D	T	R	S	P	P	C	A	255
766	GCN	GGN	771													
256	A	G	257													

Fig. 2. Nucleotide sequence and derived amino-acid sequence (GenBank accession number DQ323742 and UniProt Knowledgebase accession number P84786). Y = T or C, R = A or G, N = A or C or G or T, W = A or T, V = G or A or C. Underlined sequences are the primers used for PCR on the genomic DNA. The N-terminal and C-terminal protein sequences were determined by N-terminal and C-terminal amino-acid sequencing; the parts of the DNA sequence outside the primers (coding for SQC and CAAG) were taken from the genetic code table.

Table 1. Data, refinement and geometry statistics. The values in parentheses refer to the highest resolution shell.

Resolution range data (Å)	49.6–1.60 (1.69–1.60)
Observations	162385 (13536)
Multiplicity	4.6 (3.2)
R_{merge} (%)	6.4 (18.6)
$\langle I \rangle / \langle \sigma \rangle$	18.1(5.9)
Resolution range refinement (Å)	20–1.60 (1.64–1.60)
Number of reflections	33815 (1971)
Completeness (%)	97.2 (77.6)
R_{cryst} (%)	18.1 (17.8)
R_{free} (%)	20.8 (20.8)
Number of non-H atoms	2312
Protein atoms	2047
Water molecules	253
Buffer atoms (Mes)	12
Average B factors (Å ²)	18.25
B factor from Wilson plot	18.86
Rms deviations from ideal values	
Bond lengths (Å)	0.012
Bond angles (°)	1.607
Chiral volumes (Å ³)	0.164

The geometry of the model was analyzed by PROCHECK [14]. In the Ramachandran plot [15], 91.5% of the residues (glycine and proline residues excluded) lie in the core region, and 7.1% lie in the additional allowed region. Three residues, Leu48, Glu52 and Arg96, lie in less favored regions. These residues belong to the above mentioned poorly defined turns of the structure.

Overall folding and the active-site region

The overall folding is similar to the known RIP structures. There are two structural domains, a large N-terminal domain (Ser1–Leu190) and a smaller C-terminal domain (Pro191–Gly257). The cleft

between the two domain forms the active-site pocket (Fig. 4). The N-terminal domain is composed of a six-stranded β -sheet, which in turn contains four anti-parallel central β -strands (4–7, Fig. 4) and two parallel outer β -strands (1 and 8, Fig. 4). The β -sheet is attached to five α -helices (A, C–F, Fig. 4). In most of the RIPs there are six helices in the first structural domain. In charybodin the second helix (B) is missing. According to a structural alignment of 13 solved RIPs with charybodin (Fig. 5), this helix is a less conserved structural element. Helix B is expected to be in the region of the gap between Ser98 and Gly102. In the N-terminal domain, there is also an additional two-stranded β -sheet (strands 2 and 3, Figs 4 and 5), which lies opposite the C-terminal domain. This β -sheet is not well conserved among the known RIPs and is missing in the numbering of the structural elements of ricin [13]. The C-terminal domain consists of two consecutive α -helices (G, H, Figs 4 and 5), a third helix (I, Figs 4 and 5), which is less conserved among the RIPs, and a two-stranded β -sheet 9 and 10 (Figs 4 and 5). In charybodin there exists an additional 3_{10} helix (J, Figs 4 and 5 close to the C-terminus of the protein). This is a unique feature of charybodin. The solved structures of this family do not exhibit a 3_{10} helix near the C-terminus.

An intramolecular disulfide bridge (Cys217–Cys254) is formed.

The active site of the determined structure was found to be free of substrate. It is occupied by several well-ordered water molecules (Fig. 6). The four key residues for catalysis are well conserved among type 1 and type 2 RIPs [13]. In charybodin, Val79 unambiguously replaces the conserved Tyr. To our knowledge, this is the first observation of a natural substitution of a catalytic residue at the active site of an RIP.

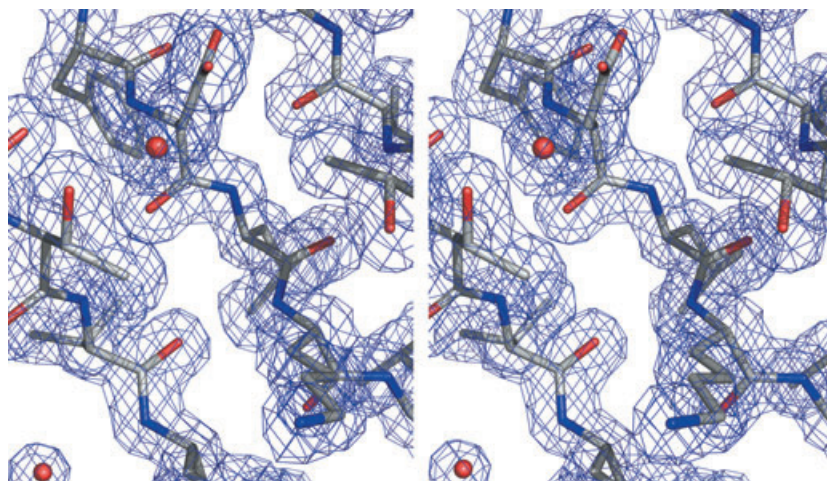


Fig. 3. Stereo view of a part of the final model in the 1.6-Å electron-density map. A section of the β -sheet in domain I is shown. The $2F_o - F_c$ map is contoured at 1σ .

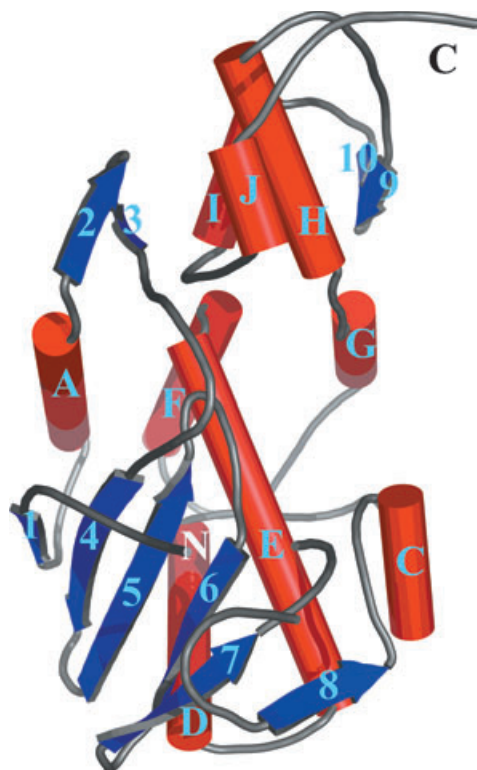


Fig. 4. Overall structure of charybdin. β -Strands are shown in blue and α -helices in red. The structural elements are labeled as follows: β -strands 1–10 and helices A–J. The N-termini and C-termini of the protein are marked. The molecule comprises two structural domains: domain I at the N-terminal part and domain II at the C-terminal end.

Discussion

In this work, we describe the purification, characterization and structural determination of charybdin, a novel 29-kDa protein from bulbs of *C. maritima* agg. Charybdin was characterized by biochemical methods and its structure determined by X-ray crystallography. The DNA sequence, and derived amino acid sequence, revealed significant homology with various RIPs. Although charybdin inhibited the rabbit reticulocyte translation system, the estimated IC_{50} of 27 nM indicates that it is not such a strong inhibitor of protein synthesis as other RIPs.

The active site of RIPs, which contains four key amino-acid residues, is highly conserved. Although three of the four key residues are present at the active site of charybdin, the fourth residue, which is an invariant Tyr80 (ricin numbering) among more than 360 RIP sequences known to date, is replaced by Val. This amino-acid change at position 79 of the active site of charybdin is a striking feature of the protein and possibly explains its low inhibitory activity compared

with other RIPs. In ricin A, the active-site residues were analyzed by site-directed mutagenesis to assess their role in the mechanism of action of the toxic enzyme [16,17]. It was found that replacement of Tyr (in ricin position 80) with Phe decreased activity by a factor of 15, and replacement with Ser decreased activity 170 times. It is expected that Val in this position would have an even more pronounced effect because the aliphatic side chain cannot form hydrogen bonds. Drastic attenuation of protein synthesis was also observed with two mutations in the Shiga-like toxin I A-chain [18]. Replacement of the active-site Tyr (position 77 in this case) with Phe resulted in 10–20-fold less activity, and replacement with Ser made the protein completely inactive.

As charybdin is the main protein constituent of the bulb of *Charybdis*, one may speculate that protein translation inhibition is not its major (or only) function; it may, for example, act as a special storage protein [19].

Although charybdin was isolated by a series of purification steps, we cannot exclude the possibility that it exists in various isoforms (as is the case with other monocots such as *Muscari* sp., *Hyacinthus* and *Iris* [10]), some of them highly active and others inactive. If this is the case, the protein that we isolated and studied may be an inactive isoform, and the observed activity may be due to ‘impurities’ of another highly active isoform. A definitive answer to this question will be given by cDNA cloning, which is one of our objectives. We are also planning to carry out site-specific mutagenesis experiments to replace the active-site Val with Tyr and study the effects on the activity and structure of charybdin.

Experimental procedures

Fresh *C. maritima* agg. bulbs were collected from a hill near Agia Galini (N35.06′-E24.41′ Crete-Greece), and for DNA sequencing also from the hamlet of Samaria (N35.17′-E23.58′).

Preliminary sequencing experiments after tryptic digestion of the denatured protein provided small fragments and an 87-amino-acid sequence (F. Lottspeich, unpublished data). This allowed us to identify charybdin as a putative RIP.

Protein purification

Fresh bulbs of *C. maritima* agg. (100 g) were homogenized in a blender at 4 °C with 300 mL extraction buffer containing 60 mM sodium phosphate, pH 7.2, 100 mM NaCl, 5 mM EDTA, 5 mM dithiothreitol, 1 mM phenylmethanesulfonyl fluoride and 1.5% (w/v) polyvinylpyrrolidone. The homogenate was filtered through four layers of cheesecloth, and

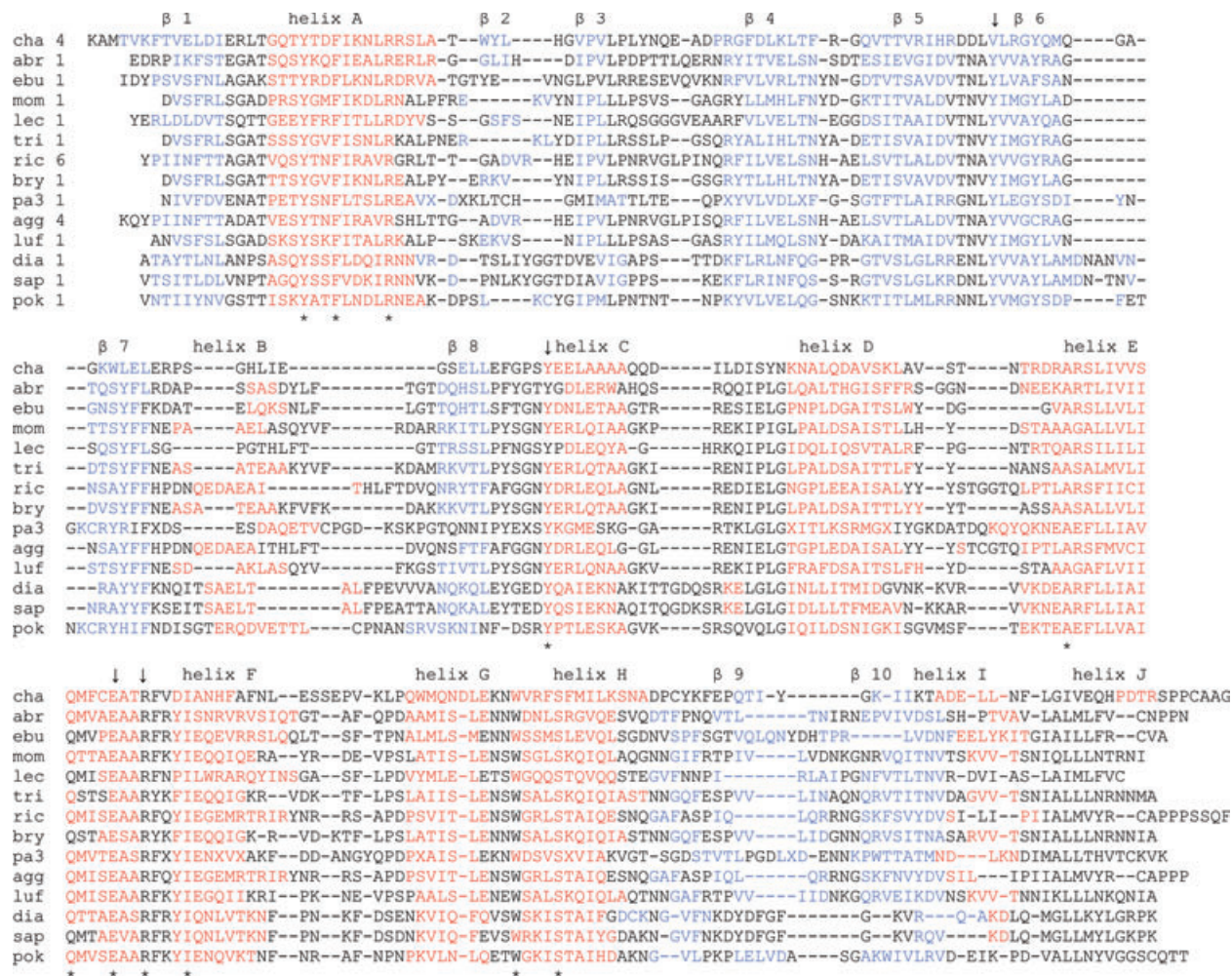


Fig. 5. Alignment of 14 crystal structures based on secondary-structure elements assigned by the program SPDBVIEW [23]. The structures are: cha, title compound (2B7U); abr, abrin (1ABR); ebu, ebulin (1HWM); mob, momordin (1MOM); lec, mistletoe lectin (1TFM); tri, trichosanthin (1MRJ); ric, ricin (1J1M); bry, bryodin (1BRY); pa3, pokeweed pap-III (1LLN); agg, agglutinin (1RZO); luf, luffin (1NIO); dia, dianthin (1LP8); sap, saporin (1QI7); pok, pokeweed antiviral protein (1QCG). Secondary-structural elements are colored as in Fig. 4. The key residues of the active site are marked with arrows; asterisks denote identical residues. The respective Protein Data Bank codes are given in parentheses.

the filtrate was centrifuged at 34 000 *g* for 30 min at 4 °C. The supernatant was passed through filtration paper. The yellowish crude protein solution was first dialyzed against a solution containing 60 mM sodium phosphate, pH 7.2 and 0.75 M ammonium sulfate, and was subsequently loaded on to a column packed with a matrix substituted with hydrophobic ligands. The column (dimensions 1 × 10 cm) was packed with phenyl-Sepharose CL-4B (Pharmacia, Upsala, Sweden) and equilibrated with 10 column volumes of 60 mM sodium phosphate, and 0.75 M ammonium sulfate at 10 °C. The sample was applied to the column at a flow rate of 0.75 mL·min⁻¹. A fraction eluted with 60 mM sodium phosphate and 0.3 M ammonium sulfate contained the protein of interest. The eluted protein was dialyzed in 50 mM Hepes, pH 7.7, and then loaded on a Q-Sepharose anion-exchange

column pre-equilibrated with the same buffer. The purified protein was eluted with 0.3 M NaCl.

For crystallization experiments, the protein isolated by the chromatographic procedure described above, was further purified by an additional sucrose density gradient step. More specifically, a continuous sucrose density gradient (10–40% sucrose in 60 mM sodium phosphate buffer, pH 7.2) was used. Centrifuge tubes were put in a swing-out rotor and ultracentrifuged at 150 000 *g* for 22 h at 6 °C in a Sorvall Ultra 80 centrifuge. This sucrose density gradient step resulted in the removal of pigments, which were copurified with the protein, and it was necessary for the crystallization of charybdin. Protein concentration was determined by the method of Bradford, using BSA as standard.

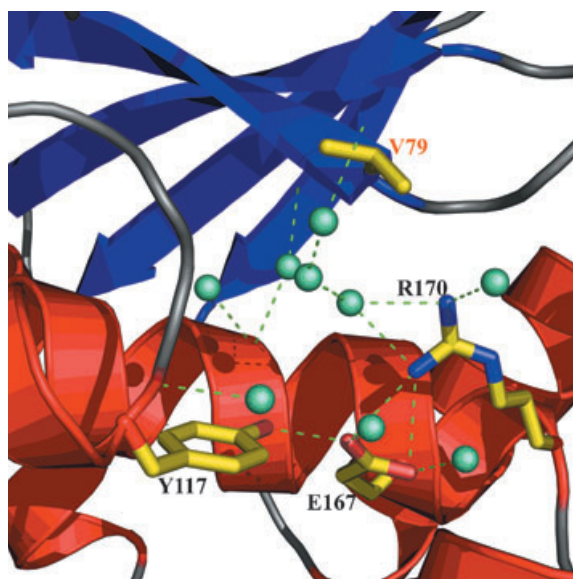


Fig. 6. The active-site region. The four key residues are shown as sticks, and water molecules which occupy the cleft are shown as spheres. Dashed lines indicate hydrogen bonds to main chain or side chain atoms. Secondary-structural elements are colored according to Figs 4 and 5.

Electrophoresis

Preparations were analyzed by SDS/PAGE by the method of Laemmli.

Translation inhibition of rabbit reticulocytes by charybdin

Charybdin was tested for *in vitro* protein synthesis inhibition activity by using a Flexi rabbit reticulocytes system (Promega, Madison, WI, USA). The translation was performed according to the manufacturer's protocol in the presence of [³⁵S]Met to label the products. Rabbit reticulocytes were incubated with increasing amounts of charybdin (13.8–552 nM) for 30 min at 30 °C before initiation of translation. Untreated rabbit reticulocytes were used as the negative control, while the RIP saporin (Fluka, Chemie Buchs, Switzerland) was used as the positive control. The reaction was initiated by adding luciferase control mRNA to the charybdin-treated reticulocytes. The reaction was carried out at 30 °C for 60 min and was terminated by centrifugation at 100 000 *g* for 15 min at 4 °C. The labeled products were analyzed by autoradiography. For autoradiography, the following instruments were used: Hypercassette™ (Amersham, Chalfont St Giles, UK) autoradiography cassettes, the Imaging Plate (Fujifilm, Tokyo, Japan) and the Storm 840 imaging system (Molecular Dynamics, Sunnyvale, CA, USA). ImageQuant software was used for quantification comparing the relative darkness

of the different bands on the film. Activity was expressed as a percentage of the control in which no charybdin was added. The IC₅₀ was calculated by linear regression analysis.

DNA sequencing

Total plant DNA was extracted from the bulb. Approximately 0.1 g of material cut from the inner part of the bulb was frozen and ground to powder in liquid nitrogen. Genomic DNA was further isolated by using the plant DNeasy Mini Kit (Qiagen, Hilden, Germany).

Crystals obtained as described below were dissolved in water, yielding 8 µg protein, which was used for N-terminal and C-terminal sequencing by the Protein Analysis Center at the Karolinska Institutet in Stockholm, Sweden. This was necessary in order to design primers suitable for the PCR experiments.

Based on the N-terminal sequence (SQXKAMTVKFTVELXI), the degenerate oligonucleotide primer (5'-AA RGCNATGACGGTGAAGTTCACAGTNGA-3'; where, R = A or G; N = A, C, G, T) was used as the upper primer. In this primer, several degenerate sites were converted into single nucleotides that were derived from the DNA sequences of homologous proteins.

From the crystallographic results, the C-terminal amino-acid sequence EQHPDTRSPPCAAG was found. C-Terminal sequencing of the protein confirmed the last four amino-acid residues. The seven underlined amino-acid residues were also deduced from sequencing after tryptic digestion. The highly degenerate primer (5'-GGNGGAGAN CGNGTRTCNGGRTGYTGYTC-3' where, Y = T or C) was used as the lower primer. As there are no homologous protein sequences for this part, the only assumption for lowering the degeneracy of the primer was made for Ser (genetic code assumed to be TCT) in analogy with the musarmin sequences, thus risking a maximum of three mismatches. Weak PCR-product bands with the expected molecular size of ≈ 800 nucleotides were obtained only with the 'Expand long template PCR system' (Roche, Basel, Switzerland) at an annealing temperature of 45 °C. The product was used as template for re-PCR (Deep Vent polymerase; New England Biolabs) after purification from a gel. Again the product of the re-PCR was purified from a gel and directly used for sequencing in an ABI-377 sequencer using the big determinant kit v.3.1 in the sequencing facility of IMBB. Sequencing was performed for both strands of DNA from two plants collected from different geographical environments in Crete, resulting in six sequences.

Crystallization

The protein was crystallized by the vapor-diffusion method. Crystals were grown during a several-day period by equilibrating a hanging drop of equal volumes of the protein

solution (5 mg·mL⁻¹ in 25 mM Hepes, pH 7.0) and reservoir solution (0.1 M Mes, pH 6.0, 16% PEG6000) at 10 °C. Crystals were first characterized 'in house' using as X-ray source a RU H3R rotating anode generator (Rigaku/MS, Woodlands, TX, USA) and a Mar300 imaging plate detector system (MarResearch, Hamburg, Germany). Before data collection, crystals were flash-frozen in liquid nitrogen in the presence of 25% glycerol as a cryoprotectant. The crystals belong to space group C2 with unit cell parameters, $a = 99.24 \text{ \AA}$, $b = 57.24 \text{ \AA}$, $c = 51.09 \text{ \AA}$ and $\beta = 104.08^\circ$. The Matthews ratio $V_M = 2.41 \text{ \AA}^3/\text{Da}$, which corresponds to 49% (v/v) solvent content. The asymmetric unit of the crystals contains one protein molecule. The crystals diffract synchrotron X-rays to 1.37 Å resolution.

Data collection, structure solution and refinement

The final diffraction data were collected using the ID14-1 beamline (ESRF, Grenoble, France) at 100 K on an ADSC detector. Data extending to 1.6 Å resolution were processed with MOSFLM 6.2.3 [20]. A high-resolution dataset with overloaded reflections was scaled together with a low-resolution dataset with limited overloaded reflections, using SCALA [21]. The structure was solved by the molecular replacement method by AMoRe [22], using 5452 reflections between 10 and 3 Å resolution. At the time of the structure solution, most of the protein sequence was unknown. This made necessary a careful inspection of the crystal structures of 12 RIPs in order to choose a suitable model for molecular replacement. In the N-terminal domain, a section of five strands of the β-sheet and one flanking α-helix was found to be relatively invariant on the basis of structural alignments using the program Swiss-PdbViewer [23]. This section was used as part A of the search model, whereby the residues were assumed to be alanine. Several flexible turns, e.g. not spatially conserved among the different RIPs, were omitted. The 87-amino-acid residue sequence deduced from a tryptic fragment was superimposed on the structures of the 12 RIPs, and a swiss model [23] was derived and used as part B of the search model. Both models were positioned on the consensus skeleton of the 12 RIPs by least-square fits. The molecular replacement search model comprised 271 atoms in 55 Ala residues (part A) and 722 atoms in 87 residues (part B), i.e. only 993 atoms, out of 2047 atoms (48.5%) of the final protein model. The rotation function with the correlation coefficient based on intensities and with the highest Patterson correlation coefficient was chosen to be the correct solution, in spite of the fact that the correlation coefficient based on F and R factor (56.2%) were not the best among the proposed solutions. The correctness of the solution was verified by building the symmetry related neighbors in the crystal lattice. No bad contacts were detected. Ninety two residues were built into electron density, which was derived from several runs of the program ARP/wARP 6.1 [24], whereby input parameters were varied. At

this stage, the protein consisted of five peptide fragments, the longest comprising 69 residues. Refinement was carried out using REFMAC v.5.2 [25] followed by manual modeling using XFIT [26]. TLS [27] refinement was also used for several cycles. One cocrystallized Mes molecule as well as all included water molecules were identified by manual model building. The final model comprises 251 out of 257 residues. The three N-terminal residues and residues 99–101 are not fitted in the final electron-density maps. The graphic illustrations of the protein were obtained using PyMOL [28].

Acknowledgements

We thank Dr F. Lottspeich for providing the sequence of various tryptic fragments, and Dr M. Aivaliotis and C. Karapidaki for their contributions during the isolation and characterization of the protein. RG would like to thank M. Providaki, A. Deli and L. Spanos for their contributions to the DNA sequencing. We thank the EMBL Grenoble Outstation, in particular, Dr Cusack and Dr Muziol, for providing support for measurements at the ESRF under the European Community – Access to Research Infrastructure Action FP6 program.

References

- 1 Gemmill CL (1974) The pharmacology of squill. *New York Acad Med Bull* **50**, 747–750.
- 2 Narayanan S, Suroliya A & Karande AA (2004) Ribosome-inactivating protein and apoptosis: abrin causes cell death via mitochondrial pathway in Jurkat cells. *Biochem J* **377**, 233–240.
- 3 Bolognesi A, Tazzari PL, Olivieri F, Polito L, Falini B & Stirpe F (1996) Induction of apoptosis by ribosome-inactivating proteins and related immunotoxins. *Int J Cancer* **68**, 349–355.
- 4 Frankel AE, Neville DM, Bugge TA, Kreitman RJ & Leppla SH (2003) Immunotoxin therapy of hematologic malignancies. *Semin Oncol* **4**, 545–557.
- 5 Endo Y & Tsurugi K (1987) RNA N-glycosidase activity of ricin A-chain. Mechanism of action of the toxic lectin ricin on eukaryotic ribosomes. *J Biol Chem* **262**, 8128–8130.
- 6 Lin JY, Tserng KY, Chen CC, Lin LT & Tung TC (1970) Abrin and ricin: new anti-tumour substances. *Nature* **227**, 292–293.
- 7 Girbes T, Ferreras JM, Arias FJ & Stirpe F (2004) Description, distribution, activity and phylogenetic relationship of ribosome-inactivating proteins in plants, fungi and bacteria. *Mini Rev Med Chem* **4**, 461–476.
- 8 Barbieri L, Battelli MG & Stirpe F (1993) Ribosome-inactivating proteins from plants. *Biochim Biophys Acta* **1154**, 237–282.

- 9 Ferreras JM, Barbieri L, Girbes T, Battelli MG, Rojo MA, Arias FJ, Rocher MA, Soriano F, Mendez E & Stirpe F (1993) Distribution and properties of major ribosome-inactivating proteins (28S rRNA-N-glycosidases) of the plant *Saponaria officinalis* L. (Caryophyllaceae). *Biochim Biophys Acta* **1216**, 31–42.
- 10 Arias FJ, Antolin P, de Torre C, Barriuso B, Iglesias R, Rojo MA, Ferreras JM, Benvenuto E, Mendez E & Girbes T (2003) Musarmins: three single-chain ribosome inactivating protein isoforms from bulbs of *Muscari armeniacum* L. & Miller. *Int J Biochem Cell Biol* **35**, 61–78.
- 11 Silva AL, Goto LS, Dinarte AR, Hansen D, Moreira RA, Beltramini LM & Araujo APU (2005) Pulchellin, a highly toxic type 2 ribosome-inactivating protein from *Abrus pulchellus*. *FEBS J* **272**, 1201–1210.
- 12 Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
- 13 Robertus JD & Monzingo AF (2004) The structure of ribosome inactivating proteins. *Mini-Rev Med Chem* **4**, 483–492.
- 14 Laskowski RA, MacArthur MW, Moss D & Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26**, 283–291.
- 15 Ramachandran GN & Sasisekharan V (1968) Conformation of polypeptides and proteins. *Adv Protein Chem* **23**, 283–437.
- 16 Ready MP, Kim Y & Robertus JD (1991) Site-directed mutagenesis of ricin A-chain and implications for the mechanism of action. *Proteins* **10**, 270–278.
- 17 Kim YS & Robertus JD (1992) Analysis of several key active site residues of ricin A chain by mutagenesis and X-ray crystallography. *Protein Eng* **5**, 775–779.
- 18 Deresiewicz RL, Calderwood SB, Robertus JD & Collier RJ (1992) Mutations affecting the activity of the Shiga-like toxin I A-chain. *Biochemistry* **31**, 3272–3280.
- 19 Liu RS, Wei GG, Yang Q, He WJ & Liu WY (2002) Cinnamomin, a type II ribosome-inactivating protein, is a storage protein in the seed of the camphor tree (*Cinnamomum camphora*). *Biochem J* **362**, 659–663.
- 20 Leslie AGW (1992) Recent changes to the MOSFLM package for processing film and image plate data. In *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, No. 26.
- 21 Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760–763.
- 22 Navaza J (1994) AmoRe: an automated package for molecular replacement. *Acta Crystallogr A* **50**, 157–163.
- 23 Guex N & Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative modeling. *Electrophoresis* **18**, 2714–2723.
- 24 Lamzin VS, Perrakis A & Wilson KS (2001) The ARP/wARP suite for automated construction and refinement of protein models. In *International Tables for Crystallography*, Vol. F: Crystallography of Biological Macromolecules (Rossmann, MG & Arnold, E, eds), pp. 720–722. Kluwer Academic Publisher, Dordrecht, The Netherlands.
- 25 Murshudov GN, Vagin AA & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **5**, 240–255.
- 26 McRee DE (1999) XtalView/Xfit: a versatile program for manipulation atomic coordinates and electron density. *J Struct Biol* **125**, 156–165.
- 27 Winn MD, Isupov MN & Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D Biol Crystallogr* **57**, 122–133.
- 28 DeLano WL (2002) The PyMOL Molecular Graphics System, <http://www.pymol.org>.

Supplementary material

The following supplementary material is available online:

Fig. S1. Comparison of the derived amino-acid sequence of charybdin with other known RIPs. MusI (Q8L5M2), MusIII (Q8L5M4), *Hyacinthus* (Q677A1), *Iris holl* (O04356), pulchellin (Q5C8A3) and ricin (P02879). The putative signal peptide of musarmins are underlined; key residues of the active site are marked with arrows. Asterisks and double points denote identical and conserved residues, respectively. The respective Swiss/TrEMBL accession codes are given in parentheses.

Fig. S2. Possible DNA sequences of charybdin and homologous proteins coding for the N-terminal and C-terminal region of charybdin after alignment of the protein sequences. Underlined nucleotides denote different bases at the same position in different proteins; the colored sequence is the deduced primer. Other sequences: musarmin 1–4, *Iris holl* 1,2,3, *Iris holl* 4,5 (GenBank AF256085, AF256084).

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