The structure of MP-4 from *Mucuna pruriens* at 2.22 Å resolution

Abha Jain,^a Amit Kumar,^b Meha Shikhi,^{a,c} Ashish Kumar,^a Deepak T. Nair^a and Dinakar M. Salunke^{b*}

^aRegional Centre for Biotechnology, NCR Biotech Science Cluster, 3rd Milestone, Faridabad–Gurgaon Expressway, Faridabad 121 001, India, ^bInternational Centre for Genetic Engineering and Biotechnology (ICGEB), Aruna Asaf Ali Marg, New Delhi 110 067, India, and ^cKalinga Institute of Industrial Technology, Bhubaneswar, Odisha 751 024, India.
*Correspondence e-mail: dinakar.salunke55@gmail.com

Received 21 November 2019

Accepted 8 January 2020

Edited by L. J. W. Shimon, Weizmann Institute of Science, Israel

Keywords: *Mucuna pruriens*; MP-4; high resolution; plant proteins; protease inhibitors.

PDB reference: MP-4, 6jbp

Supporting information: this article has supporting information at journals.iucr.org/f

The structure of the MP-4 protein was previously determined at a resolution of 2.8 Å. Owing to the unavailability of gene-sequence information at the time, the side-chain assignment was carried out on the basis of a partial sequence available through Edman degradation, sequence homology to orthologs and electron density. The structure of MP-4 has now been determined at a higher resolution (2.22 Å) in another space group and all of the structural inferences that were presented in the previous report of the structure were validated. In addition, the present data allowed an improved assignment of side chains and enabled further analysis of the MP-4 structure, and the accuracy of the assignment was confirmed by the recently available gene sequence. The study reinforces the traditional concept that conservative interpretations of relatively low-resolution structures remain correct even with the availability of high-resolution data.

1. Introduction

A large number of deaths occur owing to snakebites in many parts of the world (Swaroop & Grab, 1954; Juckett & Hancox, 2002; Chippaux, 1998; Hsu, 2015; Warrell, 2010; Kasturiratne *et al.*, 2008). Currently, the administration of preparations that contain antivenom antibodies is the most effective method for treating snakebites (Harrison *et al.*, 2011; Laloo & Theakston, 2003). However, the antivenom preparations are known to cause deleterious and unwanted side effects in a significant number of cases (Williams *et al.*, 2007; Cruz *et al.*, 2009). Also, the antivenom formulations involve the immunization of large animals with venom and therefore the mode of preparation raises ethical issues (Rojas *et al.*, 1994). Many traditional therapies exist in different parts of the world, including the oral ingestion of plant tissue extracts (Asuzu & Harvey, 2003; Selvanayagam *et al.*, 1994, 1995; Houghton & Osibogun, 1993; Meenatchisundaram & Michael, 2010). One such traditional formulation that has been found to be effective involves the seeds of the plant *Mucuna pruriens* (Meenatchisundaram & Michael, 2010; Tan *et al.*, 2009). The proteome of these seeds needs to be studied for the development of new therapeutic strategies against snake envenomation. The development of novel formulations is important as snake envenomation causes more than 100 000 deaths per year, with the vast majority of victims being poor farmers (Dey & De, 2012; World Health Organization, 2018).

It has been shown previously that the MP-4 protein provides protection against snake venom through an indirect

antibody-based mechanism (Kumar *et al.*, 2016). Sequence information suggested that this protein belongs to the Kunitz-type protease inhibitor (KPTI) family. However, biochemical experiments showed that MP-4 is a poor inhibitor of proteases such as trypsin. The structural data suggested that this might be owing to the presence of an isoleucine residue in the reactive-site loop (RSL). Thermodynamics and molecular-dynamics studies also supported this inference (Kumar *et al.*, 2018). MP-4 did not show a direct mechanism to neutralize venom; instead, it was observed that immunization with the MP-4 protein provided prophylactic protection against snake venom. Polyclonal antibodies raised against the MP-4 protein in mice showed cross-reactivity with the venom. Further, it was observed that immunization with MP-4 significantly enhanced survival in mice challenged with snake venom (Kumar *et al.*, 2016; Shudofsky, 2016).

The gene sequence for MP-4 was not available and the side-chain assignment in our previously published structure was based on the diffraction data, which were available to a maximal resolution of only 2.8 Å. At this resolution, the polypeptide backbone could be traced and it was observed that the MP-4 protein adopts a β -trefoil structure. The side-chain assignment was suboptimal, and the low resolution of the structure may raise questions regarding the analysis. In the current study, we have freshly purified, characterized and determined the high-resolution structure of the MP-4 protein. The higher resolution enabled an improved identification of the side chains. Recently, the genome sequence of *M. pruriens* became available and confirmed the present sequence derived using electron density and previous sequence information.

2. Materials and methods

2.1. Fractionation and purification

Protein purification was carried out as described previously (Kumar *et al.*, 2016). Briefly, seeds were purchased from M/S Shidh Seeds Sales Corporation, Dehradun District, India and a fine powder was prepared using an electric grinder followed by delipidification. The delipidified seed extract was homogenized in extraction buffer (50 mM sodium acetate buffer pH 5.0) and stirred in the dark at 4°C for 15 min. The supernatant was subjected to 0–95% ammonium sulfate fractionation post-centrifugation. The pellet obtained from the 60% fraction was resuspended in 50 mM phosphate buffer pH 7.2 and analyzed by SDS–PAGE. The protein was purified from this fraction by size-exclusion chromatography and the purity was checked by analytical gel-filtration and SDS–PAGE analyses. An S-100 GFC column was used to purify MP-4 from the 60% ammonium sulfate fraction. 50 mM phosphate pH 7.2 containing 140 mM NaCl was used to pre-equilibrate the GFC column and as a running buffer to elute the desired proteins at a flow rate of 0.5 ml min^{−1} in an ÄKTA pure system (GE Healthcare). Fractions corresponding to the second peak were collected and run on 15% SDS–PAGE to check their purity; they were then pooled together and concentrated to 10 mg ml^{−1} by ultrafiltration using Millipore Amicon filters

with a 10 kDa molecular-weight cutoff. The final protein concentration was estimated using a Micro BCA Protein Assay Kit (Thermo Scientific) with BSA (Thermo Scientific) as a control.

2.2. Protein identification and functional characterization

To confirm that the purified protein was indeed MP-4, mass spectrometry was carried out. MP-4 was diluted in 100% water containing 0.1% formic acid and loaded onto a C18 ZipTip matrix which was desalted using water containing 0.1% formic acid. 50% acetonitrile and 50% water (containing 0.1% formic acid) was used to elute the desalted protein, which was loaded into a MALDI-TOF/TOF 5800 (AB Sciex) in linear positive-ion mode for intact mass analysis. MALDI-TOF data were acquired and analyzed using the *BioAnalyst* software v.1.4.2.

All of the reagents used to estimate the protease inhibition activity of MP-4 were obtained from Sigma–Aldrich. The trypsin and chymotrypsin inhibitory activities of MP-4 were estimated by a slight modification of the protocol described by Erlanger *et al.* (1961). *N*- α -Benzoyl-DL-arginine-4-nitroanilide hydrochloride was dissolved in a minimum volume of DMSO until a clear yellow solution appeared and was further diluted in prewarmed buffer (100 mM Tris–HCl pH 8, 150 mM NaCl, 1 mM CaCl₂). All reactions were carried out under the same buffer conditions. 100 μ l trypsin at 0.074 mg ml^{−1} was incubated separately with 100 μ l soybean trypsin inhibitor (1.25 mg ml^{−1} to 0.305 μ g ml^{−1}) and 100 μ l MP-4 (10 mg ml^{−1} to 0.305 μ g ml^{−1}) for 10 min at room temperature. 50 μ l of 1.74 mg ml^{−1} BAPNA was added to each well; the 96-well plate was kept at 37°C for 45 min and the reaction was stopped by the addition of 50 μ l 30% (v/v) acetic acid. The optical density (OD) of the final reaction mixture was measured at 405 nm using a SpectraMax M2e microplate reader (Molecular Devices). A mixture of trypsin and BAPNA was used as a positive control.

The binding of MP-4 to trypsin and chymotrypsin was studied using an Octet RED96 system (Pall ForteBio Corporation). 100 μ g ml^{−1} MP-4 protein was biotinylated with a 3:1 molar coupling ratio of biotin reagent to ligand and immobilized on Streptavidin (SA) Biosensors (ForteBio) that had been pre-hydrated in 1× PBS + 0.2% Tween for 600 s to achieve an 0.8–1.2 nm shift. 20 μ M trypsin was diluted threefold using kinetic buffer (1× PBS) to 0.0823 μ M and used as an analyte. 40 μ M chymotrypsin was also diluted threefold to a final concentration of 0.1646 μ M for binding to MP-4. The SA biosensor was regenerated and neutralized by 2.5 M NaCl and kinetic buffer, respectively, after each binding cycle. A 1:1 global fit model was used to fit the binding curves of MP-4 with both trypsin and chymotrypsin. The data were acquired and analyzed using the *Octet Data Acquisition* (release 9.0) and *Octet Data Analysis HT* (release 9.0) software, respectively.

2.3. Crystallization and structure refinement

Crystals were grown using the hanging-drop vapour-diffusion method in 50 mM Tris–HCl, 1.5 M ammonium sulfate pH 9.5 at 25°C as reported previously (Kumar *et al.*,

Table 1

Data-collection statistics.

Values in parentheses are for the outer shell.

Space group	$P3_121$
Unit-cell parameters	
a (Å)	75.46
b (Å)	75.46
c (Å)	94.23
Wavelength (Å)	1.5418
Unique reflections	15843 (2151)
Resolution range (Å)	53.70–2.21 (2.33–2.21)
Completeness (%)	99.2 (94.2)
Multiplicity	7.3 (7.0)
Mean $I/\sigma(I)$	23.8 (5.9)
R_{merge} (%)†	0.067 (0.322)
Average mosaicity (°)	0.62

† $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the i th measurement of the intensity of reflection hkl and $\langle I(hkl) \rangle$ is the mean intensity of reflection hkl .

2016). 33% glycerol was used as a cryoprotectant and the crystal was mounted in a nitrogen stream. Data were collected on a Mar μ X system (MAR Research) and were processed using *iMosflm* (Battye *et al.*, 2011) followed by scaling and merging using *SCALA* from the *CCP4* suite (Winn *et al.*, 2011). The quality of the data was analyzed using *xtriage* in *Phenix* (Liebschner *et al.*, 2019) and the data-collection statistics are provided in Table 1.

The previous low-resolution structure (PDB entry 5dss; Kumar *et al.*, 2016) was used as a model for molecular replacement. The 5dss structure without water molecules was used as a search model in *Phaser* (McCoy *et al.*, 2007). This

exercise yielded a single unambiguous solution and further refinement was carried out using *Phenix*. Subsequently, electron-density maps were viewed and certain side chains were reassigned based on the electron density. Iterative refinement using *Phenix* and side-chain reassignment using *Coot* were sequentially carried out until convergence of the R_{free} and R_{cryst} values. *PROCHECK* (Laskowski *et al.*, 1993) and *MolProbity* (Chen *et al.*, 2010) were used to evaluate the quality of the final model. The structure was analyzed and all figures were prepared using *Coot* (Emsley *et al.*, 2010) and *PyMOL* (DeLano, 2002; Schrödinger).

3. Results

3.1. Protein purification and characterization

The purification protocol standardized previously yielded pure MP-4 according to gel-filtration and SDS–PAGE analyses (Figs. 1*a* and 1*b*). Also, the intact mass of MP-4 was estimated to be 20.871 kDa, which was identical to that determined previously (Fig. 1*c*). Since MP-4 belongs to the protease inhibitor family, functional characterization was performed by analyzing its binding to trypsin or chymotrypsin. Results show that MP-4 binds to trypsin with a dissociation rate constant (K_d) of 2.114 μM , showing a weaker affinity than standard trypsin inhibitors such as soybean trypsin inhibitor (STI; $K_d = 0.048 \mu\text{M}$). Binding of MP-4 to chymotrypsin shows a similarly weak affinity, with a K_d value of 2.585 μM (Figs. 2*a* and 2*b*).

The IC_{50} value of MP-4 for trypsin inhibition was estimated as 3.75 μM , showing 50-fold weaker inhibition compared with

Figure 1

Purification and characterization of MP-4. (*a*) Gel-filtration profile for MP-4. (*b*) SDS–PAGE analysis of purified MP-4. (*c*) Mass spectrogram showing a molecular weight of 20.871 kDa, which is identical to that estimated previously.

Table 2

Refinement statistics.

Values in parentheses are for the outer shell.

Resolution range (Å)	38.22–2.217 (2.296–2.217)
No. of reflections, working set	15811
No. of reflections, test set	771
Final R_{free}	0.2203
Final R_{cryst}	0.1795
R.m.s. deviations	
Bonds (Å)	0.006
Angles (°)	0.773
Wilson B factor (Å ²)	26.9
Average B factor (Å ²)	30.0
Ramachandran plot (%)	
Most favoured	96
Outliers	0.57

STI ($\text{IC}_{50} = 0.07 \mu\text{M}$). Similarly, in the case of chymotrypsin inhibition MP-4 ($\text{IC}_{50} = 2.98 \mu\text{M}$) is about 20-fold weaker than chymostatin ($\text{IC}_{50} = 0.157 \mu\text{M}$) and 15 times weaker than chymotrypsin–trypsin inhibitor ($\text{IC}_{50} = 0.192 \mu\text{M}$) (Figs. 2c and 2d).

3.2. Crystallization and data collection

The protein was crystallized in conditions that were effectively identical to those used previously and diffraction-quality crystals were obtained in two months. However, the new crystals belonged to space group $P3_121$, which differs from that of the previous structure of MP-4 ($P2_12_12$), and diffracted to a much higher resolution. This is not unusual. In fact, Manjula *et al.* (2018) reported crystals of two different morphologies and in different space groups from the same

crystallization drop. A single data set could be collected to a maximal resolution of 2.22 Å and the data were processed and scaled without difficulty (Table 1). Analysis of the data with *xtriage* in *Phenix* showed that the data were of good quality and free of defects such as twinning *etc.* It has been observed previously that the diffraction quality of crystals deteriorates rapidly during data collection. Therefore, a conservative data-collection strategy was employed to ensure that the crystals survived the entire duration of data collection at the home source, although the crystals could have diffracted to a higher resolution. Analysis has shown that more than 10% of the structures deposited in the PDB have an $I/\sigma(I)$ value of greater than 4 for the highest resolution shell.

3.3. Structure determination and crystallographic refinement

The structure was determined by the molecular-replacement method using the lower resolution structure of MP-4, PDB entry 5dss, as a starting model in *Phaser*. The initial model from *Phaser* showed RFZ and TFZ values of 11.3 and 33.5, respectively, with an LLG value of 1736. The packing arrangement was such that there were no symmetry-related clashes and no large voids between molecules. Fig. 3 illustrates a section of the electron-density map showing the excellent fit of the model.

Owing to the availability of the improved electron-density map, side chains could be assigned to the polypeptide backbone. The high-resolution data were particularly effective in identifying the correct amino acids in the C-terminal region of the MP-4 protein. Iterative rounds of model building and crystallographic refinement resulted in the identification of

Figure 2

Functional characterization of MP-4. (a) Binding kinetics of MP-4 and trypsin from biolayer interferometry (BLI). (b) Binding kinetics of MP-4 and chymotrypsin from BLI. (c) Inhibition of the protease activity of trypsin by MP-4 (STI; soybean trypsin inhibitor). (d) Inhibition of the protease activity of chymotrypsin by MP-4, chymostatin (Cst) and chymotrypsin–trypsin inhibitor (CtTI).

most of the residues of the entire chain with greater confidence. R_{free} and R_{work} converged to values of 0.2203 and 0.1795, respectively, in the final structure. The fully refined structure has only 0.56% of residues in disallowed regions of the Ramachandran plot and the $F_o - F_c$ correlation coefficient is 0.94. Pro31 is the only residue that was identified as an outlier in the Ramachandran plot. The presence of this residue in the disallowed regions is generally considered to be an acceptable outlier as its N atom is covalently locked within a ring with a constrained φ angle (Morgan & Rubenstein, 2013). Moreover, the backbone carbonyl of Pro31 interacts with the side chains of Arg157/165 and Glu51. The excellent stereochemical quality of the final structure can be gauged from the fact that it has an overall clashscore of only 3. The refinement statistics are tabulated in Table 2. The final refined coordinates and the MTZ file have been deposited in the Protein Data Bank with accession code 6jbp.

3.4. Overall structure

The overall structure of the protein (Fig. 4) closely resembles that described for the previous structure. It has the same β -trefoil fold as before, and this fold is known to be a signature of Kunitz-type protease inhibitors. A total of 179 residues of

MP-4 form 12 β -strands arranged in a β -hairpin-like conformation. The tertiary structure is further stabilized by two disulfide bonds: the first is formed between Cys45 and Cys90 and the second is formed between Cys144 and Cys152. The data set was collected at a wavelength of 1.54 Å and the corresponding anomalous Fourier map confirmed the positions of the disulfide bonds (Fig. 5). A total of 88 water molecules were present in the final structure.

3.5. Sequence

The improved electron-density map of MP-4 enabled the assignment of side chains with considerably higher confidence than the previous structure. The revised sequence information for MP-4 is displayed in Fig. 6(a). A comparison of the amino-acid sequence in PDB entry 6jbp and that from the Edman degradation including the last 25 ambiguous residues, as identified previously in PDB entry 5dss, showed 83% identity. However, if we do not consider the last 25 residues of the C-terminus the identity increases to approximately 95% for the Edman degradation and 80% for PDB entry 5dss (Figs. 6b and 6c). The sequences of PDB entries 6jbp and 5dss thus differ mainly because of the interpretation of the electron-density maps obtained at the two different resolutions. The

Figure 3
A section of the $2F_o - F_c$ map for PDB entry 6jbp (2.22 Å) is displayed in stereo at a contour level of 1.8σ .

Figure 4
The overall structure of PDB entry 6jbp displayed here in stereo shows 12 antiparallel strands connected by long loops.

majority of the differences are localized in the C-terminal region of the structure, as expected since the interpretation of this stretch of the protein was previously stated to be tentative.

In the present study, we showed an unambiguous, improved and high-resolution structure of MP-4 that provides better understanding in terms of the side-chain details. The r.m.s.d. between the new and the old structure was only 0.848 Å as calculated in *PyMOL*. All of the structural inferences made at lower resolution were confirmed in the present structure. The availability of the genome sequence further confirmed that the final sequence obtained here based on the improved electron density is accurate.

The *DALI* server was used to identify the closest structural neighbours using the new coordinate file of MP-4 (Holm & Rosenström, 2010). The results of the analysis regarding structural homology are equivalent to those obtained for the previous structure. The closest structural homolog for the high-resolution structure of MP-4 was the Kunitz-type inhi-

bitor from the seeds of *Delonix regia* (PDB entry 1r8n; Krauchenco *et al.*, 2003), followed by the previous low-resolution MP-4 structure (PDB entry 5dss), with r.m.s.d.s of 1.7 and 1.9 and Z-scores of 26.5 and 24.0, respectively (Fig. 7). The new structure was superimposed onto PDB entries 1r8n and 1tie (Onesti *et al.*, 1991), which are Kunitz-type inhibitors from seeds of *D. regia* and *Erythrina caffra*, respectively, and this superimposition also showed that the core region superimposed well and the majority of the variations in structure and sequence were localized to the loop regions (Fig. 7a).

3.6. Comparative analysis of PDB entries 5dss and 6jbp

The backbone structure of PDB entry 6jbp was superimposed on that of PDB entry 5dss (Fig. 7b). The core region of the structure matched well. The differences between the two structures were primarily in the loop regions. These differences were expected for two main reasons. Firstly, the

Figure 5

A $2F_o - F_c$ map is displayed at a contour level of 1.5σ to highlight the clear electron density for the disulfide bonds (a) between Cys45 and Cys90 and (b) between Cys145 and Cys152. The anomalous Fourier map is displayed at a contour level of 3.0σ and shows clear peaks for the S atoms that are part of the two disulfide bonds (c) between Cys45 and Cys90, and (d) between Cys145 and Cys152.

protein crystallized in two different space groups, and secondly loop regions are generally flexible and adopt to the local environment. They also exhibit relatively high *B* factors. Apart from this, electron density for a short loop (Ser115-Glu116-Asp117) could clearly be seen in PDB entry 6jbp compared with PDB entry 5dss. The first three residues (KND) do not show clear density in the new structure and therefore were not modelled in PDB entry 6jbp.

This class of inhibitors is known to have a reactive-site loop (RSL) which is involved in interaction with the catalytic site of the target protease. Based on previous studies, the RSL is formed of nine residues and is positioned between strands 4 and 5 of the β -sheet. Based on the nomenclature used previously (Schechter & Berger, 1967), residues 66–74 are denoted P4–P5' and the central part is referred to as P1–P1' (Bode & Huber, 2000; Otlewski *et al.*, 2001). The new structure validates the identification of the RSL made using the previous structure. Structural comparison also showed that the RSL backbone and residue information is nearly identical

(Fig. 8). One change was observed in the RSL, in which residue 68 was thought to be Glu based on the sequence homology to the closest ortholog but is actually Gly based on the electron density observed in the new structure and the gene sequence. Therefore, the sequence of the RSL, which was thought to be IREILPRTI, is actually IRGILPRTI. The equivalent residue is Glu in the structure with PDB code 1r8n, which was used as a homologous structure for sequence identification for PDB entry 5dss, whereas it is Gly in PDB entry 5xoz, the Kunitz-type trypsin inhibitor from *Cicer arietinum* (Bendre *et al.*, 2019). This position is occupied by Tyr, Leu, Ala and Pro in other known Kunitz-type protease-inhibitor structures. The new structure thus resolves the ambiguity regarding the residue present at the 68th position in MP-4.

As seen previously, the RSL forms a loop in PDB entry 6jbp and is critical for its inhibitory activity. In efficient KTIPIs the P1 position is usually occupied by Arg, Lys, Phe, Tyr, Leu or Met, but the electron-density map clearly shows that the residue at this position is Ile. The previous structure (PDB

Figure 6

Sequence comparison of PDB entry 6jbp with the old sequence. (a) The revised sequence of MP-4. (b) Sequence alignment of PDB entry 6jbp with the previously identified Edman sequence for MP-4. (c) Sequence comparison of PDB entry 6jbp with the old sequence from PDB entry 5dss.

entry 5dss) also has the branched aliphatic amino acid Ile at the P1 position and this was identified as an important factor responsible for its weak inhibitory activity (Fig. 8*b*), in contrast to Arg or Lys in strong inhibitors (Kumar *et al.*, 2018). The RSL shows the same residue profile and the loop does not exhibit any secondary structure or disulfide bonds in the high-resolution structure. In the RSL of PDB entry 5xoz, a Kunitz-type trypsin inhibitor from *C. arietinum*, the P1 position is occupied by Ile (IPGISPGII) and this is thought to be the reason for its non-substrate-like binding to trypsin (Bendre *et al.*, 2019; Fig. 8*c*).

4. Discussion

The determination of a protein structure at relatively low resolution in the absence of accurate sequence data is immensely challenging. It is definitely not possible to determine the structure in such cases using any of the automated methods currently in use. However, careful implementation of the molecular-replacement approach previously enabled the structure determination of MP-4 at 2.8 Å resolution. We were able to obtain diffraction intensity data for this protein at 2.22 Å resolution and refined the structure at higher resolution, confirming that all of the previous interpretations were indeed correct.

The space group for PDB entry 5dss is $P2_12_12$ and that for PDB entry 6jbp is $P3_121$. This is not unusual: there are numerous examples of proteins that have crystallized in two different symmetry environments from the same/similar crystallization conditions (Manjula *et al.*, 2018; Pohl *et al.*, 1998; Kurinov & Harrison, 1996). The difference in space group could be a result of local microenvironment differences at the nucleation stage. Closer inspection of the crystal packing shows compact packing in the case of PDB entry 5dss compared with PDB entry 6jbp. This could be owing to firstly the space group and secondly the solvent content. The solvent content for PDB entry 6jbp is higher compared with that for PDB entry 5dss, resulting in a lower number of molecules per unit cell.

The MP-4 protein was partially sequenced by Edman degradation (Kumar *et al.*, 2016). This partial sequence was used as a guide, but the electron-density map was primarily used to assign the side chains. It is important to emphasize that the critical residues highlighted in the Results and Discussion sections describing the earlier low-resolution structure were correctly identified. The improvement of the side-chain assignments at higher resolution compared with those at lower resolution is analogous to the modelling of a number of surface residues as alanines owing to a lack of clear electron density in many crystal structures. This is a consequence of the

Figure 7

(*a*) The superimposition of PDB entry 6jbp (cyan) with PDB entries 1r8n (green; r.m.s.d. of 1.7 Å) and 1tie (magenta; r.m.s.d. of 2.1 Å) is displayed in stereo and shows that majority of the variations are localized to the loop regions. (*b*) A superimposition of PDB entries 6jbp (cyan) and 5dss (red; r.m.s.d. of 0.848 Å) displayed in stereo showing that variations are mainly localized in loop regions.

high mobility of side chains such as those of lysine, glutamine, arginine and glutamic acid. It is well known that as the resolution of the structure decreases, the incidence of residues modelled as alanines increases (Karmali *et al.*, 2009). The sequence differences in PDB entries 6jbp and 5dss are owing to differences in the interpretation of the electron-density maps available at the two different resolutions. This new sequence should actually be compared with the published partial sequence of the MP-4 protein obtained from the N-terminal sequencing of peptide fragments. The sequence of PDB entry 6jbp shows 95% identity to the partial sequence of MP-4 obtained by Edman degradation. The differences observed between these two sequences are within the level of experimental error associated with the protein sequencer utilized for N-terminal sequencing and with side-chain assignment on the basis of electron density.

Sequence comparison of Kunitz-type protease inhibitors provided further information. As many as 20 residues were identified as conserved residues in all of the Kunitz-type protease inhibitors, regardless of their associated strong or weak inhibitory nature. It was also observed that these

conserved residues are scattered throughout the sequence, with a few in secondary structure and others in loop regions. Multiple sequence alignment also helped to predict the consensus sequence for this class of protein. Surprisingly, 41 residues were identified as consensus residues (consensus >70%): Glu5, Asp9, Glu11 (Asp in PDB entry 6jbp), Gly12, Asn13, Gly18, Gly19, Ile23, Leu24 (Met in PDB entry 6jbp), Ile27, Gly32, Thr40, Asn42, Leu47 (Val in PDB entry 6jbp), Thr48, Val/Ile49, Asp/Glu55, Pro61, Ile74, Leu80, Ile82, Pro88, Val98, Asp/Glu101, Asp114, Glu116, Asp120, Gly126, Val143, Tyr/Phe144, Asp154, Ile155, Ile157 (Arg in PDB entry 6jbp), Asp/Glu160 (Asn in PDB entry 6jbp), Glu162, Arg165, Arg166, Val168 (Ile in PDB entry 6jbp), Pro174, Phe179 and Lys181 (Asn in PDB entry 6jbp) (numbering is as in PDB entry 6jbp; Fig. 9). Approximately 65% of consensus residues are in the loop region and 35% of them belonged to secondary structures. The consensus that is present in the β -sheets seems to be responsible for stabilizing the secondary structure, as evident in Supplementary Fig. S1.

Around the time that refinement of the structure with PDB code 6jbp was completed, the genome of *M. pruriens* was

Figure 8

(a) Stereoview of the active site displaying the loop formed by residues 66–74 denoted P4–P5'. (b) The $2F_o - F_c$ map is displayed at a contour level of 1.4σ and shows that the 69th residue in MP-4 is Ile. (c) Comparison of the RSLs of PDB entries 6jbp and 5xoz.

Figure 9

Sequence comparison of known protease-inhibitor sequences with that of MP-4. Conserved residues are shown on a red background. Consensus with a >70% cutoff is shown in small letters below the alignment. The cysteines that participate in the formation of disulfide bonds 1 and 2 are also marked in green.

released (Uniprot ID UNP A0A371E4L6). A comparison of the MP-4 sequence from the deposited genome with that from PDB entry 6jbp showed a difference in only one residue (Supplementary Fig. S2). This residue is assigned as glutamine in the genome sequence but is modelled as Ala171 in PDB entry 6jbp. No electron density was visible for the side chain beyond the C^β atom and this may be owing to high mobility of the Gln side chain.

To summarize, the accuracy of the protein sequence derived from crystal structures depends directly on the resolution of the X-ray diffraction data, and the importance of resolution is considerably amplified in the absence of a gene sequence. PDB entry 5dss was the best structural model possible for the corresponding data set, given the limited resolution and the

availability of only partial sequence information. The interpretations from the structure that were described in the corresponding manuscript were conservative and the new structure (PDB entry 6jbp) shows that all of these interpretations were correct. As mentioned earlier, the quality of a structure should be evaluated with respect to the information derived from the structure. As long as the interpretations are calibrated to the resolution of the data, all structures are useful.

Acknowledgements

We thank the RCB core facility for providing access to the X-ray diffraction equipment. We thank Dr Nipendra Singh for

help with mass spectrometry and Mr Ravindra Kumar for assistance in carrying out the biochemical experiments. We acknowledge financial support from the Departments of Science and Technology and Biotechnology, Government of India. Author contributions were as follows. DMS and DTN conceived and designed the work. AK, AK and MS performed purification, functional characterization and crystallization. AJ performed data collection, structure solution and refinement. AJ, DTN and DMS analysed the data and wrote the manuscript with input from all of the authors. All authors have read and approved the manuscript and declare no conflicts of interest.

Funding information

The following funding is acknowledged: Departments of Science and Technology and Biotechnology, Government of India.

References

- Asuzu, I. & Harvey, A. (2003). *Toxicon*, **42**, 763–768.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Bendre, A. D., Suresh, C., Shanmugam, D. & Ramasamy, S. (2019). *J. Biomol. Struct. Dyn.* **37**, 2669–2677.
- Bode, W. & Huber, R. (2000). *Biochim. Biophys. Acta*, **1477**, 241–252.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Chippaux, J. P. (1998). *Bull. World Health Organ.* **76**, 515–524.
- Cruz, L. S., Vargas, R. & Lopes, A. A. (2009). *Ethn. Dis.* **19**, 42–48.
- DeLano, W. L. (2002). *PyMOL*. <http://www.pymol.org>.
- Dey, A. & De, J. N. (2012). *Afr. J. Tradit. Complement. Altern. Med.* **9**, 153–174.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Erlanger, B. F., Kokowsky, N. & Cohen, W. (1961). *Arch. Biochem. Biophys.* **95**, 271–278.
- Harrison, R. A., Cook, D. A., Renjifo, C., Casewell, N. R., Currier, R. B. & Wagstaff, S. C. (2011). *J. Proteomics*, **74**, 1768–1780.
- Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
- Houghton, P. J. & Osibogun, I. M. (1993). *J. Ethnopharmacol.* **39**, 1–29.
- Hsu, J. (2015). *Sci. Am.* **313**, 14–16.
- Juckett, G. & Hancox, J. G. (2002). *Am. Fam. Physician*, **65**, 1367–1378.
- Karmali, A. M., Blundell, T. L. & Furnham, N. (2009). *Acta Cryst.* **D65**, 121–127.
- Kasturiratne, A., Wickremasinghe, A. R., de Silva, N., Gunawardena, N. K., Pathmeswaran, A., Premaratna, R., Savioli, L., Laloo, D. G. & de Silva, H. J. (2008). *PLoS Med.* **5**, e218.
- Krauchenco, S., Pando, S. C., Marangoni, S. & Polikarpov, I. (2003). *Biochem. Biophys. Res. Commun.* **312**, 1303–1308.
- Kumar, A., Gupta, C., Nair, D. T. & Salunke, D. M. (2016). *J. Biol. Chem.* **291**, 11373–11384.
- Kumar, A., Kaur, H., Jain, A., Nair, D. T. & Salunke, D. M. (2018). *Sci. Rep.* **8**, 689.
- Kurinov, I. & Harrison, R. (1996). *Protein Sci.* **5**, 752–758.
- Laloo, D. G. & Theakston, R. D. G. (2003). *J. Toxicol. Clin. Toxicol.* **41**, 277–290.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* **D75**, 861–877.
- Manjula, R., Wright, G. S., Strange, R. W. & Padmanabhan, B. (2018). *FEBS Lett.* **592**, 1725–1737.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Meenatchisundaram, S. & Michael, A. (2010). *Int. J. PharmTech Res.* **2**, 870–874.
- Morgan, A. A. & Rubenstein, E. (2013). *PLoS One*, **8**, e53785.
- Onesti, S., Brick, P. & Blow, D. M. (1991). *J. Mol. Biol.* **217**, 153–176.
- Otlewski, J., Jaskólski, M., Buczek, O., Cierpicki, T., Czapinska, H., Krowarsch, D., Smalas, A. O., Stachowiak, D., Szpineta, A. & Dadlez, M. (2001). *Acta Biochim. Pol.* **48**, 419–428.
- Pohl, E., Holmes, R. K. & Hol, W. G. J. (1998). *J. Biol. Chem.* **273**, 22420–22427.
- Rojas, G., Jiménez, J. & Gutiérrez, J. (1994). *Toxicon*, **32**, 351–363.
- Schechter, T. & Berger, A. (1967). *Biochem. Biophys. Res. Commun.* **27**, 157–162.
- Selvanayagam, Z. E., Gnanavendhan, S., Balakrishna, K. & Rao, R. B. (1995). *J. Herbs Spices Med. Plants*, **2**, 45–100.
- Selvanayagam, Z. E., Gnanavendhan, S., Chandrasekharan, P., Balakrishna, K. & Bhima Rao, R. (1994). *Fitoterapia*, **65**, 99–111.
- Shudofsky, A. D. (2016). *Chem. Res. Toxicol.* **29**, 706.
- Swaroop, S. & Grab, B. (1954). *Bull. World Health Organ.* **10**, 35–76.
- Tan, N. H., Fung, S. Y., Sim, S. M., Marinello, E., Guerranti, R. & Aguiyi, J. C. (2009). *J. Ethnopharmacol.* **123**, 356–358.
- Warrell, D. A. (2010). *Lancet*, **375**, 77–88.
- Williams, D. J., Jensen, S. D., Nimorakiotakis, B., Müller, R. & Winkel, K. D. (2007). *Toxicon*, **49**, 780–792.
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* **D67**, 235–242.
- World Health Organization (2018). *Addressing the Burden of Snakebite Envenoming*. https://apps.who.int/gb/ebwha/pdf_files/EB142/B142_CONF1-en.pdf. Geneva: World Health Organization.