Indian Journal of Biochemistry & Biophysics
Vol. 60, September 2023, pp. 710-728
DOI: 10.56042/ijbb.v60i9.4061

Solution structures and thermodynamics of cis-trans X-Pro conformers of a novel single disulfide conopeptide

Aswani K Kancherla and Siddhartha P Sarma*

Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560 012, Karnataka India

Received 03 July 2023; *revised 08 August 2023*

The conopeptide Mo1853 (MW = 1853 Da) consists of 17 residues and a single disulfide bond. Structural studies using homonuclear solution NMR methods (2D ^{1}H,^{1}H DQF-COSY, TOCSY, NOESY and ROESY spectra) revealed that Mo1853 exists as two equally populated *cis* and *trans* X–Pro peptide bond conformers which are in slow exchange regime, compared to the chemical shift time scale. Temperature dependence of chemical shifts was measured and using coalescence temperature of two amide protons, the rate of exchange and the free energy of activation for the conformational exchange were determined to be 59 Hz and ≈ 67.2 kJ mol^{-1}, respectively, at 318 K. Additional evidence for this conformational equilibrium was also observed as exchange correlation peaks in the 2D-NOESY and ROESY spectra. Tertiary structures of both the *cis* (PDB ID 8K3N) and *trans* (PDB ID 8K3M) conformers were determined using distance restraints, backbone dihedral angle restraints, the disulfide bond restraint and the *cis* or *trans* conformation of the X–Pro peptide bond. The *trans* conformer of Mo1853 is stabilized by hydrogen bonds while the *cis* conformer seems to be stabilized predominantly by hydrophobic interactions. This was further corroborated by the fact that at lower temperatures, the hydrophobic interactions became weaker reducing the population of the *cis* conformer with respect to that of the *trans* conformer. The *cis* and *trans* X–Pro peptide bond conformational exchange could be another means to enhance the structural variability of the conopeptides and could have significance in the synergistic functional response caused by the cone snail venom peptides.

Keywords: cis peptide bond, Conopeptide, Conotoxin, Disulphide bond, Solution NMR

Cone snail venom consists of a diverse array of molecules which collectively assist the snail in capturing the prey and in deterring its predators. The primary components of the venom are conformationally constrained disulfide rich peptides called conotoxins (containing two or more disulfide bonds) and non-disulfide rich peptides (containing one or no disulfide bonds) called conopeptides. As evident from the information present in the ConoServer (http://www.conoserver.org/)[1,2], a comprehensive database of all the information on conotoxin research, most of the efforts have been directed at studying and characterizing the disulfide rich conotoxins. As of June 2023, ~ 90% of all the sequences (7646 out of 8523) deposited at the ConoServer and ~ 83% of the tertiary structures (204 out of 246) determined belong to the disulfide rich conotoxins. While these numbers clearly reflect the scientific interest generated by disulfide rich conotoxins for their high potency, specificity and stability, the numbers also indicate that the studies on other classes of molecules from the cone snail venom have been limited.

A search of ConoServer for wild-type conopeptides of the type "unclassified" returned about 50 sequences. About half of them are 10 to 20 amino acids in size and are listed in (Fig. 1). Five of these sequences have no cysteines, while two sequences have a single cysteine. In the case of two cysteine containing conopeptides, the number of residues in the inter-cysteine loop ranged from 1 to 9. Clearly, the sequences are diverse and no obvious trends in sequence conservation are observed. However, it is interesting to note that most sequences have a preponderance of hydrophobic and/or aromatic residues and contain at least one Proline residue. It is important to note that the structure and function of single disulfide peptides have not been examined in the same detail as the disulfide rich peptides[3].

Although there are only a handful of "unclassified" conopeptide sequences deposited at the ConoServer, it is noteworthy that more recent studies using an integrated analysis of transcriptome data and highly

*Correspondence:
Phone: +91 8022933454 (Mob)
Fax: +91 8023600585
E-mail: sidd@iisc.ac.in
Suppl. Data available on respective page of NOPR

ID	Sequence	Species	Diet
Mo1659[a]	FHGGSWYRFPWGY	*C. monile*	V
P05957	DVKSIGSWDFTVWHRV	*C. marmoreus*	M
P05518	TLQKLLNKTLLPNSATVL	*C. marmoreus*	M
P05519	TLTKAFEQTLLPNSATVL	*C. marmoreus*	M
P06672	TLQNASEQTLLPRLGIVLRV	*C. marmoreus*	M
P05515	STIPSLGSEWDDGW	*C. marmoreus*	M
P05972	GLVCTHVRPYHNAVWS	*C. eburneus*	P
P05971	GLVCAHVSPSQNSWWT	*C. virgo*	V
P04220	GPYRRYGNCYCPI#	*C. vitulinus*	V
P04233	GPYRRHGNCFCPS#	*C. vitulinus*	V
P05511	NEFLTHTFSWHPTWCPWC#	*C. marmoreus*	M
P06639	HPHNHFPFACPWC	*C. victoriae*	M
P05970	DVKCIGSCDSTVWHRV	*C. distans*	V
P05956	CIGSCDSTVWHRV	*C. marmoreus*	M
P05016	EIILPALRVQSFIACTMGWC	*C. ebraeus*	V
P05686	QYECPVGLWCD	*C. pulicarius*	V
P05688	QYGCPPGLWCH	*C. pulicarius*	V
P05684	QDECRIGLWCLVRIFN	*C. pulicarius*	V
P04915	DNSCTPKPSCFF	*C. californicus*	P
P05420	LVVGDQLCYRVLIKCLMNK	*C. marmoreus*	M
P05017	RSCSFPSNTWC	*C. ebraeus*	V
P05032	FCFFTITNMSGGCLV	*C. ebraeus*	V
Mo1853[b]	LVSGCNFVYVKPCRGGR	*C. monile*	V

Fig. 1 — Comparison of sequence of Mo1853 with those of the non-disulfide-rich, unclassified conopeptide sequences retrieved from ConoServer. The sequences are arranged in the increasing order of the number of cysteines they possess, followed by increasing order of the number of residues in the inter-cysteine loop. "ID" is the ConoServer ID of the protein. "#" denotes C-terminal amidation. Cysteine residues are highlighted in red. "M","V" and "P" denote the Molluscivorous, Vermivorous and Piscivorous dietary habits of the cone snails. [a]Details of Mo1659 are in the references Sudarslal *et al*[4] and Kumar *et al*[5] and are not present in the ConoServer. [b]Details of Mo1853 are from this study

sensitive mass spectrometric data revealed presence of unexpectedly large diversity in the peptide components of the cone snail venom[6,7]. These studies led not only to the discovery of a variety of new conopeptide sequences but also to the uncovering of molecular mechanisms that contribute to generating this diversity. The transcriptome studies indicated that "biological messiness"[7,8] present at the transcription level leads to several single amino acid variations, amino acid deletions, frame shifts and stop codon shifts resulting in toxin variants with alternate cleavage sites, interrupted or elongated cysteine frameworks. This diversity is further enhanced by variable post-translational processing. Mass spectrometric methods confirmed presence of the toxin variants at peptide level, albeit at low abundance. Although very little is known about the pharmacological properties and functional roles of these peptides, the observation that they are present in the venom of cone snails belonging to all the dietary habits (feeding on worms, fish or other molluscs) suggests that they are important for the overall efficacy of the cone snail venom and subsequently for the survival of the cone snails. In essence, a large fraction of cone snail venom components, such as conopeptides expressed at low abundance, continue to remain structurally and pharmacologically uncharacterized. It is important to study these peptides to understand their individual properties and also to uncover any common structural and thermodynamic principles underlying the stability and activity of these molecules. As part of this approach, we have focussed on the structure and function of these unusual single disulfide peptides that are found in cone snail venom[9].

One such low abundance conopeptide that contains a single disulfide bond was identified during the mass spectrometric analysis of venom from a vermivorous cone snail *Conus monile*. Further, using chemical modification and mass spectrometric fragmentation studies, the sequence of the peptide was determined to be LVSGCNFVYVKPCRGGR (Dr. Suman Thakur, personal communication). The peptide was named Mo1853 as it was identified from the venom of *Conus monile* and has an oxidized mass of 1853 Da. The primary structure of the peptide Mo1853 clearly indicated that Mo1853 belonged to neither the contryphan nor the conopressin class of conopeptides and hence was placed among the "unclassified" conopeptides. Since Mo1853 is present at low abundance in the cone snail venom, the sample for structural studies was obtained either by chemical synthesis or by recombinant expression using cytochrome-b_5 fusion protein host system.[10] The structural studies were carried out using solution NMR spectroscopy analysis of the unlabeled peptide samples using homonuclear and heteronuclear (natural abundance) spectra. NMR spectra unequivocally established that the peptide Mo1853 exists as 2 equally populated *cis* and *trans* X-Pro peptide bond conformers that are in a chemical exchange whose timescale is slow compared to the chemical shift time scale. Tertiary structures were determined for both the conformers using distance restraints obtained from NOESY/ROESY spectra and the rate of exchange was calculated using coalescence temperature of the amide protons obtained in a temperature study. The results from the solution NMR studies and the implications of the findings on the conopeptide function are presented below.

Experimental Procedures

Materials

Chemically synthesized peptide was purchased from Bioconcepts Laboratories, New Delhi. All the other chemicals were purchased from Sigma-Aldrich, India. NMR sample tubes (535-PP) were purchased from Wilmad Glass (Vineland, New Jersey, USA). Methanol for HPLC was purchased from Merck India Ltd and was purified by distillation prior to use.

Sample preparation

Chemically synthesized sample

The chemically synthesized sample of Mo1853 was obtained in reduced form at 50% purity. The peptide was dissolved in water to a final concentration of <100 µM, the pH was adjusted to be between 7.0 to 7.5 using 20 mM NH_4HCO_3 and was incubated at 37°C for 24 h to allow for the formation of disulfide bonds. After the oxidation step, the peptide was lyophilized to dryness and was purified by reverse phase HPLC using a C18 column.

Recombinantly expressed sample

Expression host *E. coli* BL21 (DE3) was transformed with expression vector pET21a(+) in which the gene coding for Mo1853 had been cloned[11] as a C-terminal fusion to Cytochrome-b_5 fusion protein system[10]. The initial stages of protein expression, lysis and purification up to the DEAE based anion exchange chromatography were carried out as described previously for cytochrome b5 fusion proteins. After cell lysis, the soluble and insoluble fractions of the recombinantly expressed protein were purified separately as described below. The details mentioned below correspond to a protein preparation starting from a culture volume of 2 L of LB medium.

Soluble fraction

After the DEAE–Sepharose based anion exchange chromatography purification step, the protein solution was lyophilized to dryness and was subjected to chemical cleavage by CNBr in 70% Trifluoroacetic acid (TFA). The CNBr cleavage reaction was terminated by carrying out rotary evaporation at ~50°C till all the solvent was removed. The reaction products formed a layer on the inner walls of round bottom flask. Reduced peptide was obtained by solvent extraction with stirring at room temperature using ~30 mL of 50% CH_3OH (*v/v*) in MilliQ water acidified with 0.1% TFA and containing 10 mM of Tris(2–carboxyethyl)phosphine (TCEP). In these conditions, the peptide Mo1853 came into the soluble fraction while majority of the cleaved fusion protein host Cytochrome-b_5 remained insoluble due to isoelectric precipitation (pI of Cytochrome-b_5 is ~5.2). The insoluble components were removed by centrifugation at 14,000 rpm for 30 min and the soluble fraction was rotary evaporated to remove methanol. Precipitants which appeared upon removal of methanol were pelleted by an additional round of centrifugation at 14000 rpm for 30 min. The soluble fraction containing the reduced peptide was dialyzed in MilliQ water acidified with 0.01% TFA using a dialysis membrane with a Molecular Weight Cut off (MWCO) of 1 kDa to remove excess TCEP. The peptide solution was lyophilized to dryness, redissolved in a smaller volume (1 to 2 mL) of MilliQ water acidified with 0.01% TFA and was purified using reverse phase High Pressure Liquid Chromatography (HPLC) using C18 column to remove any residual TCEP and fusion protein host. The fraction containing reduced Mo1853 was first concentrated by rotary evaporation and then oxidized by adjusting the pH of the peptide solution to 7.5 using 20 mM NH_4HCO_3 and incubating at 37°C for 24 h. The oxidized peptide was lyophilized to dryness and was finally purified by another round of reverse phase HPLC.

Insoluble fraction

After centrifugation of the cell lysate, the insoluble fraction of the fusion protein was present as inclusion bodies in the pellet along with the other insoluble cellular components. The pellet was resuspended in ~40 mL of TEG buffer (25 mM Tris-HCl, 10 mM EDTA, 50 mM Glucose, pH 8.0). The membrane components were solubilized by addition of 0.04% Triton-X 100 followed by thorough vortexing and sonication for 2 min (2 sec ON; 5 sec OFF; 38% Amplitude). The suspension was centrifuged at 14000 rpm for 45 min and the supernatant was discarded. The inclusion bodies were retained as pellet and were solubilized by addition of ~30 mL of 8 M urea containing 40 mM β−mercaptoethanol and stirring at room temperature for ~1 h. Any remaining insoluble matter was removed by centrifugation at 14000 rpm for 30 min. The solubilized protein was dialyzed extensively in 20 mM NH_4HCO_3, pH 7.5 to remove the denaturant. Further purification was achieved using protocols described above for the soluble protein fraction.

Reverse Phase HPLC

Analytical and semi-preparative Reverse phase HPLC purification was carried out using Varian Pursuit XRs 5 C18 columns (bead size 5 μM, pore size 100 Å) or Dr. Maisch GmbH column (bead size 5 μM, pore size 70 Å) connected to a Shimadzu LC-20 AD Prominence Liquid Chromatography System equipped with SPD-20A UV-Visible detector capable of dual wavelength detection. Purification of the Mo1853 was achieved by application of a linear gradient of a binary solvent system consisting of solvent A (water: methanol: TFA in the ratio 90:10:0.1) and solvent B (water: methanol: TFA in the ratio 10:90:0.1). HPLC fractions containing the peptide were first concentrated by rotary evaporation to remove the methanol and were then lyophilized to dryness.

Cysteine modification reactions

Alkylation and Reduction / Alkylation reactions

Solutions of peptides (30 μM concentration) in 50 mM ammonium bicarbonate buffer, pH 7.5 and 4 M urea were used for all reactions. Alkylation of Mo1853 was carried out by adding excess of iodoacetamide (150 equivalents of the total cysteine content) to the peptide solutions. The reaction was allowed to proceed for 45 min at room temperature. For the reactions where reduction was followed by alkylation, the peptides were first incubated with 60 equivalents of DTT (based on total cysteine content) for 12 h at 4°C and then alkylated as described above. The reaction products were analyzed either by MALDI–MS or LC–ESI–MS.

Mass Spectrometry

Matrix Assisted Laser Desorption Ionization (MALDI) Mass Spectrometry

MALDI mass spectra were acquired either on Ultraflex TOF/TOF or Ultrafle Xtreme of Bruker Daltonics, Germany. One micro litre of (0.1 to 1 mg/ml) sample was mixed with an equal volume of saturated solution of either Sinapinic Acid (SA) or α-cyano-4-hydroxycinnamic acid (CCA) (in 50% Acetonitrile in water with 0.1% TFA) before spotting on the MALDI plate. The data were acquired in positive ion mode with detector either in linear or reflector mode. Five hundred to 1000 spectra were co-added. The data were processed and analyzed using the data analysis software Flex Analysis 3.1.

Electro Spray Ionization (ESI) Mass Spectrometry

ESI mass spectra were recorded on a HCT Ultra ETD II ion trap mass spectrometer (PTM Discovery System, Bruker Daltonics, Germany) connected to an Agilent 1100 series HPLC system. The samples were infused into the mass spectrometer either by direct injection or through a reverse phase HPLC column and eluted using a binary gradient of water (0.1% TFA): acetonitrile (0.1% TFA) at a flow rate of 0.2 mL/min. Data were acquired in positive ion mode over a m/z range of 300 - 2800. The data were processed and analyzed using the Bruker Compass Data Analysis 4.0 software.

Circular Dichroism

The far-UV CD spectra of Mo1853 in oxidized and reduced states were acquired at a final peptide concentration of 50 μM. Circular Dichroism (CD) spectra were acquired on a JASCO-715 spectropolarimeter over a wavelength range of 190 to 250 nm for the far-UV CD spectra and 250 to 350 nm for the near-UV CD spectra. Path length of the cuvette used was 0.2 cm. The data were acquired with a scan rate of 100 nm/min and a data pitch of 2 per nm. The final spectra were averaged over three scans, baseline corrected by subtracting the spectrum of appropriate blank solution and smoothed using a running average protocol available with the JASCO spectra analysis software. For reduction, the final concentration of TCEP used was 5 mM.

NMR Spectroscopy

Sample preparation

The NMR samples of Mo1853 were prepared by dissolving the oxidized, HPLC purified, lyophilized peptide in sterile MilliQ water with 10% ^2H_2O for the lock-signal. Whenever desired, the sample was micellized by addition of Dodecylphosphorylcholine-d_{38} (DPC) to a final concentration of 5 mM. The concentration of Mo1853 in the NMR samples ranged from 0.5 to 1 mM. Sample for "H/D" exchange studies was prepared by addition of appropriate amount of ^2H_2O to a lyophilized sample of Mo1853.

Data acquisition

All NMR data were acquired on Bruker Avance 700 MHz spectrometer having a cryogenically cooled triple resonance probe equipped with a z-axis pulsed field gradient accessory. The data for sequence specific assignments and structure determination were acquired at 30°C in samples of Mo1853 micellized with 5 mM DPC-d_{36} while the data for temperature study was acquired on samples of Mo1853 prepared in MilliQ water. The homonuclear ^{1}H, ^{1}H–TOCSY, NOESY and ROESY spectra acquired with mixing

times of 60 ms, 300 ms and 200 ms, respectively, were used for sequential assignments and for generation of distance restraints. The details of data acquisition for homonuclear spectra and heteronuclear spectra acquired using natural abundance of the heteronucleus are presented in (Suppl. Table S1).

Data Processing and analysis

All NMR data were processed on an Intel PC work station running SuSe Linux version 11.4 using NMRPipe/NMRDraw processing software[12]. Time domain data in the directly acquired and indirectly detected dimensions was apodized with a squared sine bell weighting function phase-shifted by 90° and zero-filled once prior to Fourier transform and base line correction. Linear prediction was applied in the indirectly detected dimensions to improve the resolution wherever applicable. Processed spectra were analyzed using CcpNmr Analysis versions 2.1 to 2.4[13].

Determination of exchange rate constant and activation energy barrier

A series of ^{1}H–1D and ^{1}H,^{1}H–TOCSY spectra were acquired by varying the temperature from 278 K to 332 K at an interval of 5 K. At every temperature, the sample was allowed to reach thermal equilibrium by waiting for15 min after which the shims and gain were adjusted prior to the data acquisition. The spectra were acquired using 700 MHz NMR spectrometer with a spectral width of 8012.8 Hz in both the F_1 and F_2 dimensions. A total of number of 4096 and 512 real points were acquired in t1 and t2 dimensions, respectively. The spectra were processed by applying a squared sine-bell apodization function phase shifted by 90°C. Prior to Fourier transformation, the data sets were zero filled to increase the number of points to the next nearest power of two, yielding a final data matrix of 8192×1024 real points. This results in a digital resolution of ~16 Hz (0.03 ppm) 2 Hz (0.003 ppm) in F_1 and F_2 dimensions, respectively.

Coalescence temperature

The H^N–H^α region of the ^{1}H,^{1}H–TOCSY spectra were overlaid and the pairs H^N–H^α correlations belonging to the *cis* and *trans* conformers were examined for coalescence in the amide proton dimension. The coalescence temperature was determined for whichever pairs exhibited coalescence.

Exchange rate at coalescence temperature

The exchange rate at coalescence temperature was calculated using the expression[14] given below (Eq. 1):

$$k_c = \frac{\pi \Delta \nu}{\sqrt{2}} \quad \ldots (1)$$

where:

Δv is the separation in Hz between the two amide protons in the absence of exchange; and

k_C is the exchange rate constant at coalescence temperature.

Δv is determined experimentally from spectra recorded at temperatures which are as far below the coalescence temperature as possible. The expression described above for determination of exchange rate (given by equation 1) is valid provided that the following conditions are satisfied:

- the dynamic process occurring is of first-order
- the two conformers are equally populated
- the exchanging nuclei are not coupled to each other

However, even if these conditions are not fulfilled exactly, equation 1 is a good approximation for estimation of exchange rate at coalescence temperature[14].

Free Energy of Activation

The relationship between exchange rate constant and free energy of activation is given by Eyring equation (equation 2) given below[14]:

$$k = x\frac{k_B T}{h} e^{-\frac{\Delta G^{\ddagger}}{RT}} \quad \ldots (2)$$

where:

k_B = Boltzmann Constant = $1.3805 \times 10^{-23} JK^{-1}$

x = transmission coefficient (usually assumed to be 1)

h = Planck constant = $6.6256 \times 10^{-34} Js$

The free energy of activation at coalescence temperature can be obtained by rearranging the Eyring equation (equation 2) and using the coalescence rate constant and coalescence temperature as shown below in equation 3:

$$\Delta G_C^{\ddagger} = 19.14 T_C \left(10.32 + log\frac{T_C}{k_C}\right) Jmol^{-1} \quad \ldots (3)$$

where:

$\Delta G^{\ddagger}$ = Free energy of activation at the coalescence temperature

T_C = Coalescence temperature in Kelvin

k_C = Exchange rate constant at coalescence temperature

Estimating the limits of error

The possible sources of error are inaccuracies in measuring the frequency separation ($\Delta\nu$), the absolute temperatures (T), and the coalescence temperature (T_C). A realistic estimate for the order of magnitude of the errors are given below[14]:

k_C: ± 25%

$\Delta G^{\ddagger}$: ± 0.8 kJ mol^{-1}

Generation of structural restraints

Distance restraints

Distance restraints for structure calculation were generated from unambiguously assigned correlations in the 2D–NOESY spectrum (mixing time = 300 ms) using "Make Distance Restraints" tool present in the CcpNmr Analysis. The default options in the tool were used. Using these options, the target inter-proton distance was calculated as the peak height raised to the power of -(1/6) multiplied by a scaling factor. The scaling factor was defined such that the reference peak intensity (average peak height of all the assigned peaks in the peak list) exactly corresponds to the reference distance (default value of 3.2 Å). An absolute maximum of 6.0 Å was set for upper distance limit. Lower distance limits were set at 1.8 Å for all inter proton distance restraints.

Dihedral restraints

The backbone dihedral angle restraints φ and ψ were obtained using the chemical shift based prediction tools *viz.*, Talos+[15] and Dangle[16]. The 3J $H^N H^\alpha$ coupling constants were estimated from the H^N–H^α region of DQF–COSY spectrum which was processed without any apodization (or line-broadening) window function.

Disulfide restraints

Restraints for disulfide bond between two cysteines Cys^5 and Cys^{13} were included as distances S^γ-S^γ, C^β–S^γ and S^γ–C^β with both upper and lower limits at 2.05 Å, 3.05 Å, and 3.05Å, respectively. An additional C^β–C^β distance restraint of 4 Å was also included.

Structure calculation and refinement

The tertiary structures of *cis* and *trans* X-Pro conformers of Mo1853 were calculated by Cyana version 3.0[17-19] with the experimental distance restraints, dihedral angle restraints and disulfide bond restraints as input. The peptide bond between Lys^{11} and Pro^{12} was fixed in *cis* and *trans* conformation for the respective conformers. In each structure calculation run, 100 random conformers were taken as initial structures and were subjected to 20000 steps of simulated annealing by molecular dynamics simulation in torsion angle space. Twenty structures with the lowest target function were selected for further analysis. In the final round of structure calculation by Cyana, 200 initial structures were subjected to simulated annealing and the ensemble of best 50 structures from the run was written as output. These 50 structures were refined in explicit water using CNS version 1.3[20,21]. The format conversion from Cyana to CNS was done with the help of PDBStat[22] and the refinement was set up using the WaterRefCNS script (http://rtti7.uv.es/~roberto/Index.php?sec=water). The cut-off distance for automatic disulfide bond detection was decreased to 2.5 Å from 3.0 Å and the WateRefCNS script was executed using an option to include the *cis* Xaa–Pro peptide bond and with default values for all other options. All the structures which had no NOE violations greater than 0.2 Å, no dihedral violations greater than 5° and no van der Waal violations were curated and 20 structures with best covalent geometry were chosen for further analysis.

For the *trans* X-Pro conformer, after refining 50 cyana structures by CNS, 33 structures were without any violations. Of these, three structures have a positive value of chiSS. Remaining 30 have negative values of chiSS. Among the structures having negative chiSS values, the 20 models with lowest CNS refinement energy were chosen for submission to PDB.

Similarly, for the *cis* X-Pro conformer, after refining 50 cyana structures by CNS, 30 structures were without any violations. Of these, seven structures have a positive value of chiSS. Remaining 23 have negative values of chiSS. Among the structures having negative chiSS values, the 20 models with lowest CNS refinement energy were chosen for submission to PDB.

Results

Preparation of Mo1853 sample

The sample of Mo1853 was prepared either from chemically synthesized or recombinantly expressed protein. Chemically synthesized peptide was obtained as 50% pure reduced peptide, which was oxidized and purified by reverse phase HPLC. In the case of recombinant expression, about 50% of the fusion protein was found to be in the insoluble fraction as inclusion bodies. Further, majority of the fusion protein was also found to be present as a covalently linked dimer. Initial rounds of purification, CNBr cleavage and solvent extraction were carried out in reducing conditions to obtain linear (reduced)

Mo1853. The peptide was subjected to oxidation after first round of reverse phase HPLC and a pure sample of oxidized Mo1853 was obtained by a second (and final) round of reverse phase HPLC. Figure 2 shows gel pictures from SDS–PAGE analysis, HPLC chromatographs and mass spectra from various stages of sample preparation of Mo1853.

Preliminary characterization of Mo1853

The samples of reduced Mo1853 obtained by chemical synthesis and by recombinant expression were subjected to MALDI Post Source Decay (Suppl. Fig. S1). A comparison of the "y" and "b" ions observed in the MALDI LIFT mass spectra of both the samples, with those expected based on the sequence of Mo1853, ascertained that the Mo1853 samples prepared by chemical and recombinant methods are identical and that sequence of Mo1853 was as expected.

Establishing the formation of disulfide bond

The formation of disulfide bond in the oxidized samples of Mo1853 was confirmed using chemical modification of cysteines/cystines followed by mass spectrometry, and by Far-UV Circular dichroism of Mo1853 in native and reducing conditions. The

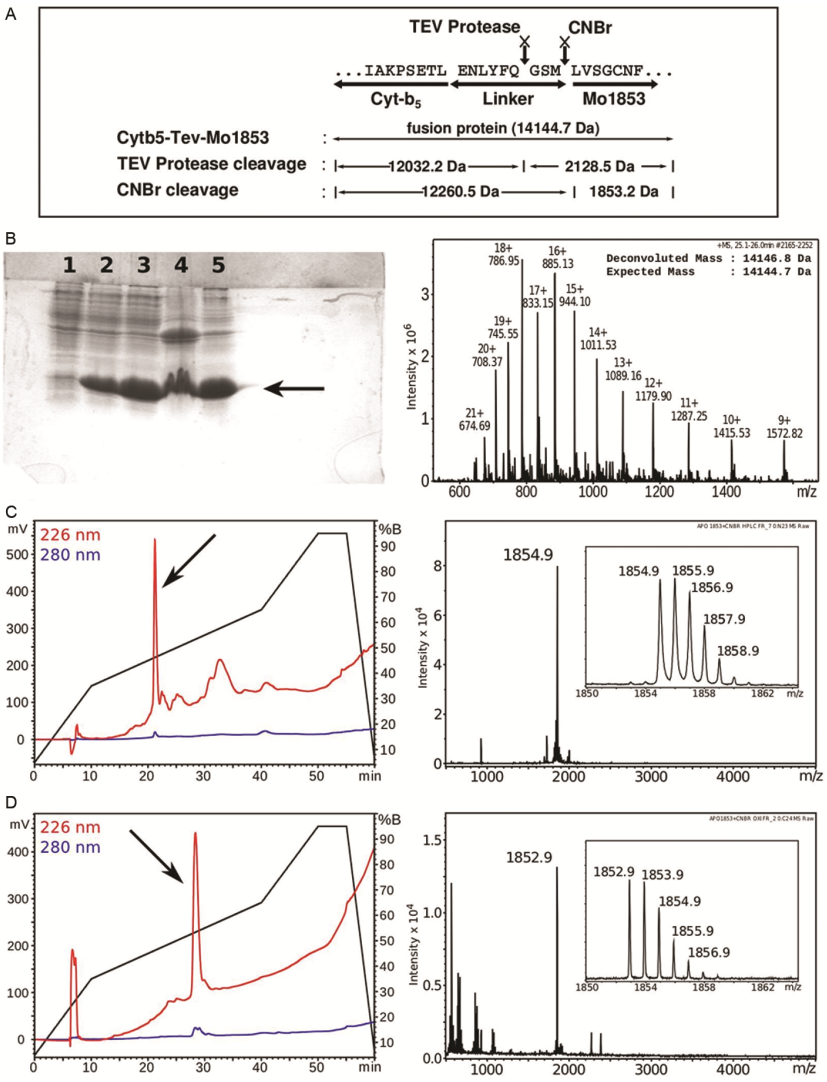

Fig. 2 — Recombinant expression and purification of Mo1853. (A) The fusion protein construct used for the expression of Mo1853. The linker region harbours recognition site for TEV protease and a methionine for cleavage by CNBr; (B): SDS–PAGE gel-picture analysis and LC–ESI–MS mass spectra for the fusion protein. Lanes 1 to 5 show the samples at the time of induction, 5 h after induction, soluble fraction of the cell lysate, insoluble fraction of the cell lysate and DEAE elution fraction, respectively. The arrow on the gel-picture shows the position of the fusion protein; and (C & D) The reverse-phase HPLC chromatograms and MALDI mass spectra of reduced and oxidized samples of Mo1853, respectively. The arrows on the HPLC chromatograms show the desired fractions containing Mo1853. The insets in the MALDI mass spectra contain a zoom-in of the peaks of interest to show the observed isotopic pattern

MALDI mass spectra of chemically modified samples of Mo1853 are shown in (Fig. 3A). No change in mass was observed upon subjecting the peptide to an alkylation reaction indicating that there were no free thiol groups in Mo1853. Further, an increase in mass by 252 Da (2×126 Da) upon alkylation (using NEM) of the chemically reduced Mo1853 clearly indicated that Mo1853 has two cysteines that are disulfide paired.

The far-UV CD spectrum of native Mo1853 exhibited a negative ellipticity with a minimum at ~205 nm (Fig. 3B). In the presence of reducing conditions *viz.,* 5 mM TCEP, far-UV CD spectrum exhibited a minimum at ~197 nm which is characteristic of unfolded peptides/proteins. These observations when taken together indicate that the Mo1853 is structured as a consequence of the disulfide bond.

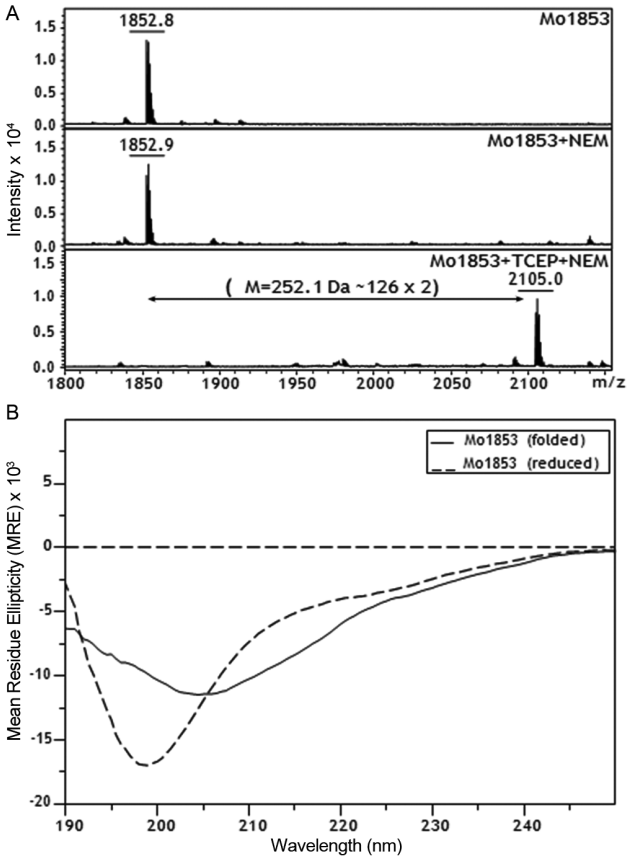

Fig. 3 — Establishing presence of disulfide bond within Mo1853. (A) MALDI mass spectra of Mo1853 after Reduction-Alkylation reactions indicates addition of 2 NEM moieties corresponding to mass increase of 252 Da suggesting that the 2 cysteines present in Mo1853 are linked by a disulfide bond; and (B): The far-UV CD spectra of Mo1853 in reducing and native conditions. The spectrum in the reducing conditions has a minimum at ~195 nm characteristic of unfolded proteins. However, the CD spectrum of Mo1853 in native conditions has a minimum at ~205 nm indicating presence of secondary structure that could be attributed to the presence of disulfide bond

Solution structural studies of Mo1853

Proton 1D spectrum

Proton 1D spectra of Mo1853 (Suppl. Fig. S2) prepared from chemical and recombinant expression methods were nearly identical reaffirming the reproducibility of sample preparation. The spectra have sharp lines indicating good solution properties of Mo1853 (absence of aggregation). However, it must be noted that the spectral dispersion is limited and the spectra lack up-field resonances below 0 ppm and down-field resonances above 10 ppm which are characteristic of ring-current shifted aliphatic protons and hydrogen-bonded amide protons, respectively.

Homonuclear Proton 2D spectra and optimization of data acquisition conditions

Structural studies on Mo1853 were initiated by acquiring homonuclear proton 2D spectra *viz.,* ^{1}H,^{1}H–DQF–COSY, ^{1}H,^{1}H–TOCSY and ^{1}H,^{1}H–NOESY. A total of ~30 correlations were observed in the signature H^{N}–H^{α} region of ^{1}H,^{1}H–TOCSY (Suppl. Fig. S3). The number of peaks observed is nearly double compared to the number of peaks expected in this region based on the amino acid sequence. Presence of impurities in the sample of Mo1853 was ruled out by HPLC and mass spectrometry. This observation hence indicated that Mo1853 might be undergoing chemical exchange which is in slow exchange regime in the chemical shift time scale. The spectra also lacked dispersion that is characteristic of well folded peptides. Peptides whose cognate partners are in membrane environment are known to undergo functionally relevant conformational changes in the membrane like environment created by micelles[23,24]. Keeping this in mind, the sample of Mo1853 was micellized by addition of DPC-d_{38} to a final concentration of 5 mM (CMC of DPC is 1.1 mM). However, the spectral changes observed upon micellization were only marginal as can be seen in (Suppl. Fig. S3).

Sequence specific assignments and restraints for structure determination were obtained by analyzing the spectra acquired on a sample of Mo1853 micellized with 5 mM DPC-d_{38}.

Sequence specific assignments of Mo1853

Sequence specific resonance assignments for Mo1853 were obtained by analysis of homonuclear

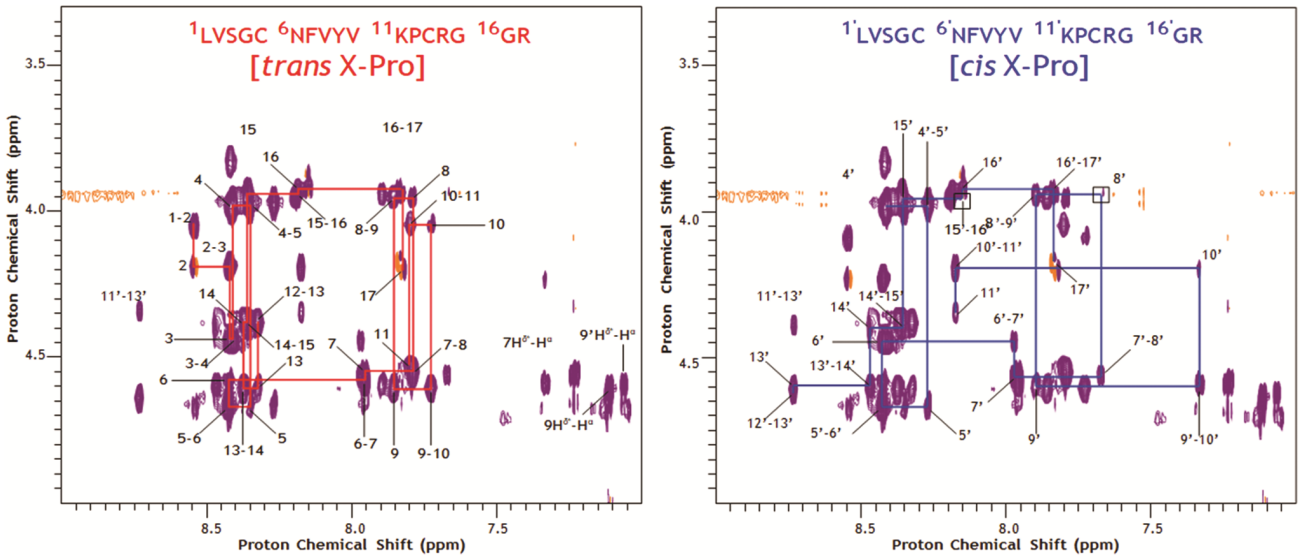

Fig. 4 — The signature H^N–H^α region of 2D–^{1}H,^{1}H–NOESY spectra of Mo1853 in 5 mM DPC showing sequential connectivity for the "*trans*" (left panel) and "*cis*" (right panel) X–Pro conformers of Mo1853. The horizontal lines connect the intra-residue H^Ni–H^α_i correlations with the sequential H^α_{i-1}–H^Ni correlations. Vertical lines connect the sequential H^Ni–H^α_{i-1} correlation with the intra- residue H^Ni–H^α_i correlations. Weak and missing correlations are shown by an open square. A few intra-residue correlations between $H^\alpha$and the side chain protons of aromatic residues are also seen in this H^N–H^α signature region of the NOESY spectra

2D– ^{1}H,^{1}H–DQF–COSY, ^{1}H,^{1}H–TOCSY and ^{1}H,^{1}H–NOESY spectra using protocols described by Wuthrich[25]. As described above, the number of correlations observed in the H^N–H^α signature region of the TOCSY spectrum correspond to approximately double the number of correlations expected for Mo1853. Further, during the assignment process, unambiguous sequential connectivity could be established separately for the two independent sets of correlations using the sequence of Mo1853 as shown in the H^N–H^α signature region of the ^{1}H,^{1}H–NOESY spectrum in (Fig. 4) Each set of correlations correspond to amino acid residues present in one of the two conformations of Mo1853. Except the three N-terminal residues Leu, Val and Ser, all the other residues of Mo1853 exist in both the conformations. The amino acid residues in the two conformations are numbered 1 to 17 and 4′ to 17′.

Despite the limited spectral dispersion, especially among the resonances belonging to the residues 1 to 17, all the expected sequential correlations between H^α_{i-1} and H^Ni were observed and assigned. The connectivity between the Lys^{11} and Pro^{12} residues was established based on the sequential correlations between Lys^{11}–H^α and the protons Pro^{12}–$H^{\delta a}$ and Pro^{12}–$H^{\delta b}$ (Fig. 5A) Similarly, the connectivity between the residues Lys$^{11'}$ and Pro$^{12'}$ was established using the sequential correlation between the protons Lys$^{11'}$–H^α and Pro$^{12'}$–H^α (Fig. 5A).

These correlations also unequivocally established that the peptide bond between Lys^{11} and Pro^{12} is in *trans* conformation while the peptide bond between the residues Lys$^{11'}$ and Pro$^{12'}$ is in *cis* conformation. This assignment was further validated by carbon–13 chemical shifts of side chain of the Proline residue obtained from the ^{1}H,^{13}C–HSQC spectrum acquired using the natural abundance of ^{13}C isotope. The difference in chemical shifts of $C^\beta$and C^γatoms *viz.,* ^{13}C $\Delta\delta_{\beta\gamma}$ in a Proline residue is an indicator of the conformation of the X–Pro peptide bond[26]. The values of $\Delta\delta_{\beta\gamma}$ are ~5 ppm and ~10 ppm for Proline residues in *trans* and *cis* conformation, respectively. In Mo1853, the values of $\Delta\delta_{\beta\gamma}$ for Pro^{12} and Pro$^{12'}$ are 4.4 ppm and 9.6 ppm, respectively (Fig. 5B). Hence the residues numbered 1 to 17 belong to the trans X–Pro conformer of Mo1853 while the residues numbered 4′ to 17′ belong to the cis X–Pro conformer of Mo1853.

The signature H^N–H^α region of the ^{1}H,^{1}H–TOCSY spectrum with sequence specific assignments is shown in the (Fig. 6). The residues Val^{10}, Lys^{11} and Cys^{13} show maximum variation in the amide proton chemical shift between the two conformers. It is noteworthy that these residues are sequentially adjacent to the Pro^{12}. This observation further corroborates assignment of NOESY correlations specific to the conformation of X–Pro peptide bond (Fig. 5A). As the two conformers of Mo1853 are attributed to *cis–trans* isomerization about the Lys^{11}–Pro^{12} peptide bond, maximum

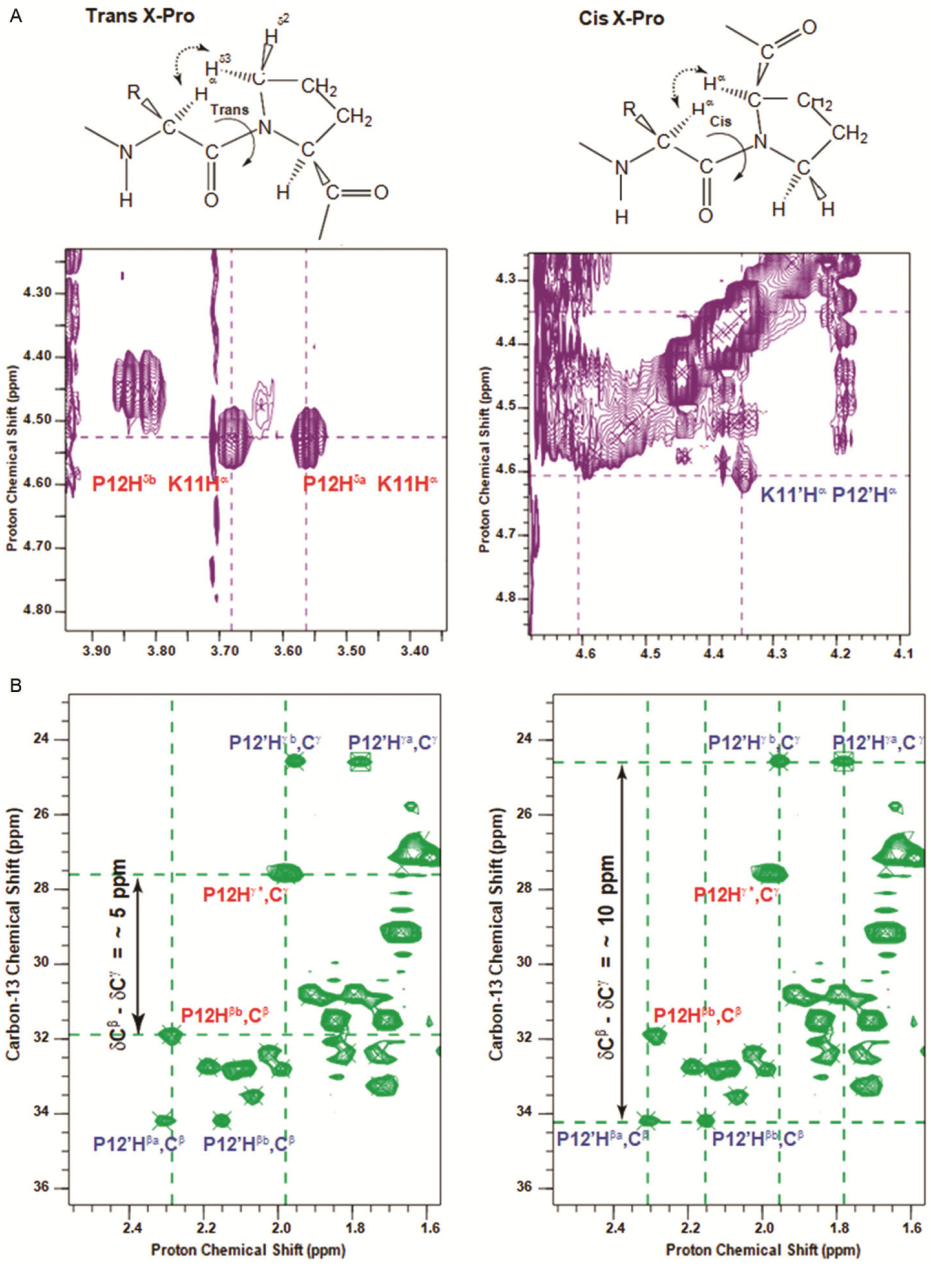

Fig. 5 — Identification of the *trans* and *cis* conformers of Mo1853 using NMR spectroscopy. (A) The characteristic sequential NOE correlations between Lys^{11}–H^{α}, and Pro^{12}–$H^{\delta a}$ and Pro^{12}–$H^{\delta b}$ indicate that Pro^{12} is in *trans* conformation. Similarly, the sequential NOE correlation between Lys$^{11'}$–H^{α} and Pro$^{12'}$–H^{α} indicate that Pro$^{12'}$ is in *cis* conformation; and (B) The chemical shift difference between C^{β} and C^{γ} atoms of the Proline side chains also indicate that Pro^{12} and Pro$^{12'}$ are in *trans* and *cis* conformations, respectively. See main text for details

perturbation in the conformation of peptide backbone (and hence in the chemical shift values) is expected to be proximal to the Proline residue.

The sequence specific assignments for the backbone H^N and H^{α} resonances of the *trans* (in red) and *cis* (in blue) conformers of Mo1853 are also shown in the H^N–H^{α} region of the ^{1}H,^{1}H–DQF–COSY spectrum in the (Suppl. Fig. S4). The $3J(H^NH^{\alpha})$coupling constants were estimated from the H^N–H^{α} region of DQF–COSY spectrum which was processed without any apodization (or line-broadening) window function. The $3J(H^NH^{\alpha})$ coupling constants for boththe *cis* and *trans* conformers of Mo1853 are shown in the (Suppl. Fig. S5). The correlations which were found to be weak in the ^{1}H,^{1}H–DQF–COSY spectrum were observed in the ^{1}H,^{1}H–TOCSY spectrum.

Ratio of the populations of the *trans* and *cis* conformers

The population ratio of the *trans* and *cis* conformers of Mo1853 was determined using the ratio of the peak volumes of the intra-residue H^N–H^α correlations of Val^{10} and Cys^{13} whose amide proton chemical shifts are well separated in the *trans* and *cis* conformations. Using these residues, the ratio of *trans* to *cis* conformers of Mo1853 was calculated to be approximately 60% to 40% at 303 K (exact values of *trans*:*cis* populations are 56%:44% and 61%:39% as calculated using the residues Val^{10} and Cys^{13}, respectively).

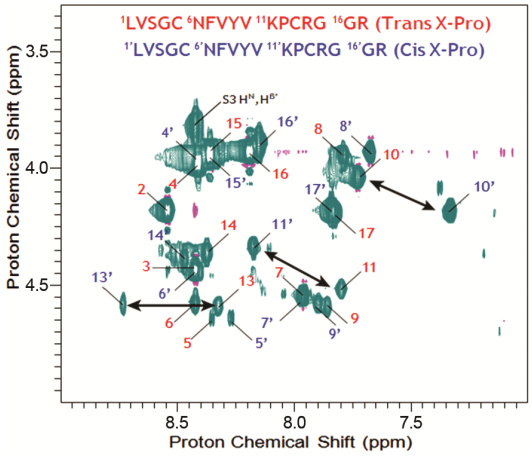

Fig. 6 — The signature H^N–H^α region of 2D–^{1}H,^{1}H–TOCSY spectrum of Mo1853 in 5 mM DPC showing sequence specific assignments for the *trans* (in red) and *cis* (in blue) Lys^{11}–Pro^{12} conformers of Mo1853. The residues whose amide protons show large variation in chemical shift between the two conformations are indicated by bold double headed arrows

Conformational exchange about the Lys^{11}– Pro^{12} peptide bond

The studies so far indicated that Mo1853 in solution exists as two different conformers that correspond to *trans* and *cis* conformations about the Lys^{11}–Pro^{12} peptide bond. The *trans* and *cis* conformers are in slow exchange regime with respect to chemical shift time scale and this manifested as doubling of resonances in the ^{1}H,^{1}H–TOCSY spectrum. Additional qualitative estimates for the rate of this exchange process came from the 2D–^{1}H,^{1}H–NOESY (mixing time of 300 ms) and the 2D–^{1}H,^{1}H–ROESY (mixing time of 200 ms) spectra.

Unambiguous exchange correlations *viz.,* Val^{10} H^N – Val$^{10'}$ H^N, Lys^{11} H^N–Lys$^{11'}$ H^N and Cys^{13} H^N – Cys$^{13'}$ H^N were observed in the H^N–H^α region of NOESY spectrum (Fig. 7, left panel). As expected for a peptide of this size, the exchange correlations are of the same sign as the NOE correlations due to cross relaxation. However, correlations due to chemical exchange and the NOE correlations due to cross relaxation could be discriminated using the ROESY spectrum. Two of the exchange correlations *viz.,* Val^{10} H^N – Val$^{10'}$ H^N and Cys^{13} H^N – Cys$^{13'}$ H^N were also observed in the H^N–H^N region of the ROESY spectrum (Fig. 7, right panel). Further, these exchange correlations are of the opposite sign compared to the correlations arising from the cross relaxation[27]. These observations conclusively showed that Mo1853 undergoes conformational two-site exchange involving *cis–trans* isomerization about the Lys^{11}–Pro^{12} peptide bond.

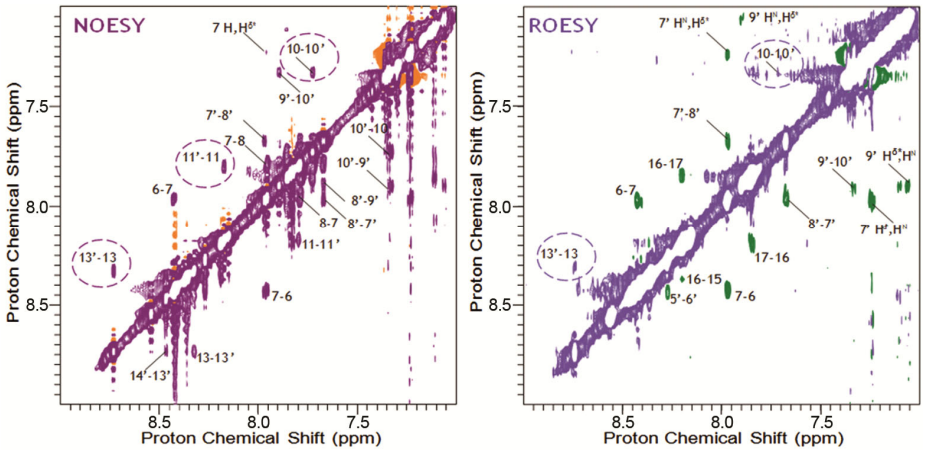

Fig. 7 — H^N–H^N region of 2D–^{1}H,^{1}H–NOESY(left) 2D–^{1}H,^{1}H–ROESY(right) spectra of Mo1853 in 5 mM DPC showing correlations due to chemical exchange as well as cross relaxation. The exchange correlations are highlighted by enclosing in dotted circles. In the ROESY spectrum, the cross peaks due to chemical exchange are of same sign as the diagonal, in contrast to the cross peaks due to through space interaction which are of the opposite sign as the diagonal. Thus ROESY provided additional evidence for the chemical exchange. See main text for details

Determination of the exchange rate and free energy barrier by temperature study

In order to determine the exchange rate, the temperature dependence of the chemical shifts was examined. The study was carried out by acquiring a series of 1D proton and 2D–^{1}H,^{1}H–TOCSY spectra at temperatures starting from 278 K up to 332 K at intervals of 5 K. Examination of the amide proton region of the 1D spectra (Suppl. Fig. S6) clearly showed temperature dependence of the signals *viz.,* coalescence of signals, appearance of new signals and sharpening of lines and overall simplification of the spectrum at higher temperatures, which are characteristic features of chemical exchange. However, the spectral overlap and absence of assignments for the signals in proton 1D spectrum precluded determination of coalescence temperature and subsequent extraction of any quantitative information directly from proton 1D spectra.

The temperature dependence of the H^N–H^α correlations in the TOCSY spectrum were examined to identify the coalescence temperature of intra-residue H^N–H^α correlations belonging to the *cis* and *trans* conformers. Figure 8 shows an overlay of H^N–H^α region of a representative set of TOCSY spectra acquired in the temperatures study. The correlations belonging to residues whose amide proton chemical shifts show large variation between the *cis* and *trans* conformers of Mo1853, *viz.,* Val^{10}–Val$^{10'}$, Lys^{11}–Lys$^{11'}$ and Cys^{13}–Cys$^{13'}$ continued to be well separated and did not coalesce indicating that they are in slow exchange regime even at temperatures as high as 332 K.

Coalescence was observed at ~318 K for correlations of two sets of residues *viz.,* Phe^{7}–Phe$^{7'}$ and Tyr^{9}–Tyr$^{9'}$. Using the expression given in equation 1, the exchange rate at coalescence temperature was calculated to be 59 Hz. The value of $\Delta\nu$, *i.e.,* the separation of the signals in the absence of exchange, was measured from the TOCSY spectrum acquired at 283 K. The Free Energy of activation $\Delta G^{\ddagger}$ at coalescence temperature (318 K) was determined to be 67.262 kJ mole^{-1} using the equation 3.

Additional information such as the temperature coefficients of amide proton chemical shifts (Suppl. Fig. S7) and effect of temperature on the populations of *cis* and *trans* conformers of Mo1853 (Fig. 10C) were also obtained from the temperature study. This information yielded further insight into the structure of the two conformers of Mo1853 and conformational exchange between them (see Discussion).

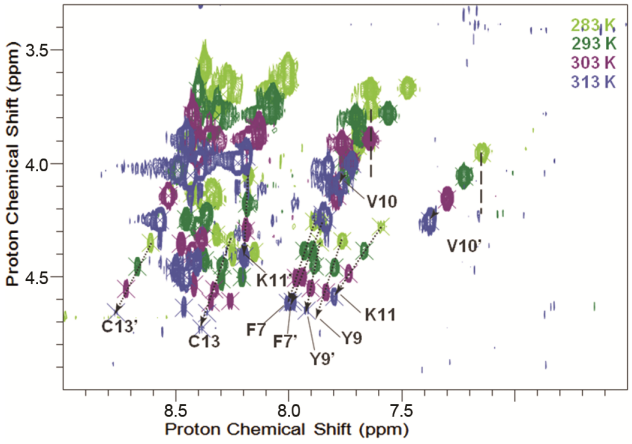

Fig. 8 — Overlay of H^N–H^α region of 2D–^{1}H,^{1}H–TOCSY spectra acquired in the temperature study of Mo1853. Coalescence of amide protons of cis and trans X-Pro conformers was observed for the residues Phe^{7} and Tyr^{9}. See text for details

Determination of tertiary structure for *cis* and *trans* conformers of Mo1853

Determination of tertiary structures of the *cis* and *trans* conformations of Mo1853 was possible due to presence of structural restraints specific for each conformer. These restraints included experimentally derived distance restraints (from NOEs assigned in the 2D ^{1}H,^{1}H–NOESY), the conformation about the Lys^{11}–Pro^{12} peptide bond, the disulfide restraint (included as four distances between Cys$^{5}S^{\gamma}$–Cys$^{13}S^{\gamma}$, Cys$^{5}C^{\beta}$–Cys$^{13}S^{\gamma}$, Cys$^{5}S^{\gamma}$–Cys$^{13}C^{\beta}$ and Cys$^{5}C^{\beta}$–Cys$^{13}C^{\beta}$), 3 J(H^NH^α) coupling constants obtained from DQF-COSY, and dihedral angle restraints obtained using the programs Talos+ and Dangle. Initial rounds of structure calculation were performed using distance restraints alone. The dihedral restraints consistent with the initial structures were included in the later stages of structure calculation to improve the convergence of the structure solution. Table 1 lists all the restraints used for structure calculation in cyana and the refinement statistics for the best 20 structures obtained after refinement by CNS for both the *cis* and *trans* conformers of Mo1853. A superposition of the ensemble of best 20 structures of the *cis* and *trans* conformers of Mo1853 is shown in (Fig. 9). The "all backbone atom" and "all heavy atom" RMSD values and the Ramachandran Map analysis[28,29] for the ensemble of structures for both the *trans* and *cis* conformers indicate that the structures have been determined to good precision and stereochemical quality. The ensemble of structures for *trans* conformer has an "all backbone atom" and "all heavy atom" RMSD of 0.6 and 1.3 Å, respectively. Further,

Table 1 — NMR restraints and refinement statistics for *trans* and *cis* conformers of Mo1853. The statistics were prepared using PSVS server (https://montelionelab.chem.rpi.edu/PSVS/PSVS/)

NMR constraints	Conformer	
	Trans X-Pro	Cis X-Pro
Distance restraints		
Total NOE[a]	166	148
Intra-residue	104	93
Sequential (i-j) = 1	50	41
Medium range (i-j) <= 4	8	10
Long range (i-j) >= 5	4	4
Disulfide bond restraints[b]	4	4
Dihedral angle restraints		
φ	14	11
ψ	15	13
Structure statistics		
No. of NOE violations >= 0.5 Å	0	0
No. of dihedral violations > 5°	0	0
Deviation from idealized geometry		
Bond lengths (Å)	0.020	0.020
Bond Angles (°)	1.2	1.3
Average pairwise RMSD (Å)		
Backbone	0.6	0.7
Heavy atom	1.3	1.2

[a]Number of non-redundant and non-trivial NOE distance restraints
[b]Four distance restraints were included to constrain the disulfide bond. See methods section

~ 92% of the amino acid residues are in most favored regions of the Ramachandran Map with the remaining 8% in the additional allowed regions. Similarly, the ensemble of structures for *cis* conformer has an "all backbone atom" and "allheavy atom" RMSD of 0.7 and 1.2 Å, respectively. These structures have ~77%, 22% and 1% of the residues in most favored, additional allowed and generously allowed regions of the Ramachandran Map, respectively.

Description and comparison of the tertiary structures of *trans* and *cis* conformers of Mo1853

The disulfide bond between the residues $Cys^5$and Cys^{13} in Mo1853 causes cyclization of the pep- tide backbone through a side-chain to side-chain S–S covalent bond formation. Cyclization of the peptide backbone significantly reduces the number of conformations available for the molecule. Hence, both the *trans* and *cis* conformations of Mo1853 have a central macro-cyclic peptide structure (29 membered ring) formed by the nine amino acid residues from Cys^5 to Cys^{13}. Further, in both the conformers of Mo1853, the torsional angle about the disulfide bond *viz.,* χ^{SS}, is centered about 90°: the average values of χ^{SS} for the ensemble of *trans* and *cis* conformers of Mo1853 are -93° ± 6° and -78° ± 9°, respectively. Hence, the conformation of the macro-cyclic ring is primarily affected by the conformation about the Lys^{11}–Pro^{12} peptide bond, which is 180° in the *trans* conformer and 0° in the *cis* conformer. The *trans* conformation of the Lys^{11}–Pro^{12} peptide bond causes the *trans*–Mo1853 to have its central macro-cyclic ring in an extended and taut conformation with the amino acid side chains extending outward. In addition, the residues starting from Pro^{11} till the C-terminus exhibit a propensity to form a right handed *α*-helix. The change in conformation of the Lys^{11}–Pro^{12} peptide bond along with concomitant change in the backbone dihedral angles *φ* and *ψ* of a few other residues (*vide infra*) cause the central macro-cyclic ring of the *cis*–Mo1853 to be less extended and more slack in conformation. In addition, the C-terminal residues seem to have lost the propensity to be in helical conformation. The structures of macro-cyclic peptide ring, the orientation of the side chains and the interactions between the side chains are shown in (Fig. 10).

Comparison of the backbone dihedral angles of *trans* and *cis* X–Pro conformers of Mo1853

A residue-wise comparison of the backbone dihedral angles of the *trans* and *cis* conformers of Mo1853 helps to rationalize the structural features described in the previous section. Table 2 lists the average values of the dihedral angles *φ* and *ψ* for all the residues in both the *trans* and *cis* conformers of Mo1853.

Supplementary Figure S8 shows MolProbity Ramachandran analysis[31] for all the residues in the ensemble of structures. The following important observations can be made from the table of dihedral angles and the Ramachandran map:

- In the *trans* Mo1853, the residues Lys^{11} and Pro^{12} are present in the third quadrant, the right handed *α*-helical region of the Ramachandran map. However, both the residues move to the regions corresponding to extended conformation in the *cis* Mo1853. Hence in the *trans* conformer, Pro^{12} may be acting as a helix inducer, increasing the helical propensity of the C-terminal residues.
- All the residues of the *trans* conformer have negative values of *φ* and are in the second and third

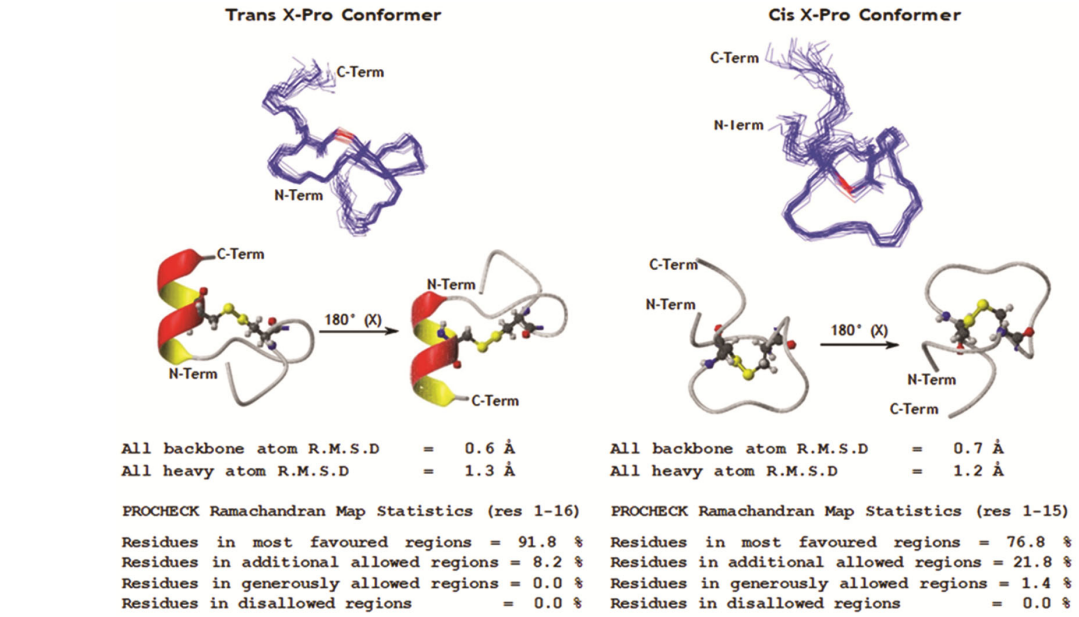

Fig. 9 — Tertiary structures of *trans* (PDB 8K3M)and *cis* (PDB 8K3N) conformers of Mo1853. Superposition of best 20 structures using backbone atoms is shown in blue (top panel). The disulfide bond is shown in red. Ribbon representation of a representative structure from the ensemble is shown below in two orientations (lower panel). The sulphur atoms and the disulfide bond are highlighted. Figures are prepared using molmol software[30]. Ramachandran Map statistics are obtained using PSVS server. The coordinates of the lowest energy model for both cis and trans conformer are included in the supplementary information

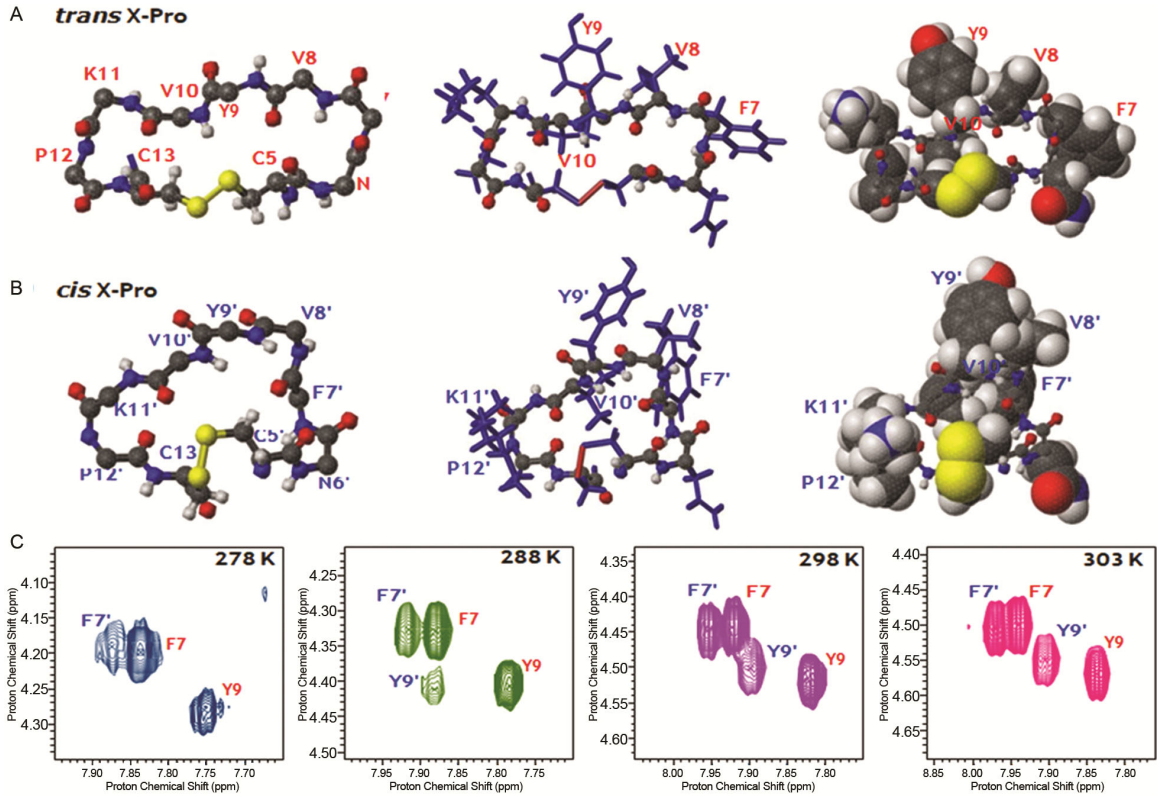

Fig. 10 — Interactions stabilizing the 29-membered macro-cyclic peptide ring in the *trans* (A) and *cis* (B) conformers of Mo1853. The H^N–H^α region of the $^1H,^1H$–TOCSY spectra acquired at four different temperatures showing the intra-residue H^N-H^α correlations for the residues Phe[7] and Tyr[9] in the *trans* and *cis* conformers of Mo1853 (C). See Discussion in the main text for details

Table 2 — The average backbone dihedral angles for the ensemble of structures of trans and cis conformers of Mo1853. The residues which show a large change in the backbone dihedral angles between the conformers are indicated by a right arrow in the middle column

Residue	Trans X-Pro			Cis X-Pro	
	φ (°)	ψ (°)		φ (°)	ψ (°)
Leu^{1}	-	145		-	128
Val^{2}	-70	33		-82	-26
Ser^{3}	-90	165		-138	138
Gly^{4}	-87	98	->	78	108
Cys^{5}	-70	-20	->	-93	121
Asn^{6}	-67	-37		-67	-44
Phe^{7}	-85	-40	->	-84	143
Val^{8}	-140	140		-69	134
Tyr^{9}	-77	95	->	-47	27
Val^{10}	-70	137		-87	155
Lys^{11}	-57	-48	->	-131	118
Pro^{12}	-66	-29	->	-71	-164
Cys^{13}	-80	-30	->	-78	53
Arg^{14}	-60	-32	->	-66	166
Gly^{15}	-68	-32	->	100	0
Gly^{16}	-77	-37		-109	28
Arg^{17}	-91	-		-131	-

quadrants of the Ramachandran Map having extended and right handed helical conformation, respectively. In *cis* conformer, three residues *viz.,* Gly^{4}, Tyr^{9} and Gly^{15} move to the first quadrant having with positive φ values and exhibit dihedral angles characteristic of left handed helical conformation.

- The *trans* conformation contains two Type-I turns between the residues Gly^{4}–Phe^{7} and Val^{10}–Cys^{13}. The *cis* conformation lacks these two turns but contains a Type-VIb like turn between the residues $Val^{10'}$–$Cys^{13'}$. The residue $Pro^{12'}$ which is in *cis* X-Pro conformation constitutes the characteristic third residue of the Type-VIb turn.

Interactions stabilizing the *trans* and *cis* conformations of Mo1853

The *trans* and *cis* conformers of Mo1853 are almost equally populated at 303 K, the temperature at which the solution structures were determined. Analysis of the ensemble of structures for intra-molecular interactions that stabilize these conformations (and present in at least half of the ensemble of structures) revealed the following: the ensemble of structures for the *trans* conformation has four hydrogen bonds, whereas the ensemble of structures for the *cis* conformation has only one hydrogen bond. The hydrogen bonds present in the *trans* conformer are $Phe^{7}H^{N}$→$Gly^{4}CO$, $Val^{8}H^{N}$→$Gly^{4}CO$, $Arg^{7}H^{N}$→$Lys^{11}CO$ and $Arg^{17}H^{N}$→$Cys^{13}CO$. The first hydrogen bond stabilizes the Type-I reverse turn between the residues Gly^{4}–Phe^{7} present in the *trans* Mo1853. The hydrogen bond present in the *cis* conformer is $Arg^{14'}H^{N}$→$Pro^{12'}CO$.

The observation that both the *trans* and *cis* conformers have equal populations (and hence equal free energy) despite the absence of hydrogen bond interactions in the *cis* conformer could be rationalized by the presence of hydrophobic interactions in the *cis* conformer. As mentioned in the previous section, the macro-cyclic peptide ring in the *cis* conformer of Mo1853 is less extended and has a more compact conformation when compared to that of the *trans* conformer. The compact conformation of *cis*–Mo1853 facilitates the interaction of the hydrophobic side chains. It can be seen in (Fig. 10) that the side chains of residues $Phe^{7'}$, $Val^{8'}$ and $Tyr^{9'}$ interact to form a well packed hydrophobic cluster. Similarly, the aliphatic side chain of $Lys^{11'}$ packs well with the side chain of $Pro^{12'}$. In *trans* conformer, side chains of these residues extend outward and do not interact with each other.

Further experimental evidence for the presence of hydrophobic interactions in the *cis* conformer of Mo1853 came from the temperature study. The hydrophobic effect is known to become weaker at lower temperatures and hence, if the *cis* conformer of Mo1853 is stabilized by hydrophobic interactions, it is expected to get destabilized and its population would decrease at lower temperatures. Indeed, the *trans*:*cis* population ratio was skewed towards *trans* conformer and the *cis* conformer is sparsely populated at lower temperatures (Fig. 10C) as measured using peak volumes/intensities of the correlations belonging to specific conformers. At 278 K, the intra-residue $Phe^{7'}H^{N}$–$H^{\alpha}$and $Tyr^{7'}H^{N}$–H^{α}TOCSY correlations belonging to the *cis* conformer of Mo1853 are very weak compared to the equivalent correlations from the *trans* conformer. As the temperature is increased, the intensity of these correlations increases. At 303 K, the intensities of correlations from the *cis* and *trans* conformers become nearly equal. All these observations strongly suggest that the *cis* conformer of Mo1853 is stabilized by hydrophobic interactions while the *trans* conformer is stabilized primarily by hydrogen bond interactions.

Discussion

Major focus of conotoxin research has been the characterization of disulfide rich conotoxins that are stabilized by two or more disulfide bonds. Of late, a large body of evidence has been accumulating to indicate the presence of a variety of non-disulfide-rich conopeptides (containing one or no disulfide bonds).[3] However, the structural studies of such peptides have been sparse[5,9]. In this study, we have structurally characterized a novel, single disulfide containing conopeptide Mo1853 from the cone snail species *Conus monile*. The samples of Mo1853 for structural studies were obtained either as chemically synthesized peptide or as recombinantly expressed and purified peptide. The recombinant expression was carried out using the Cytochrome b_5 fusion protein expression system[10]. In both the methods, Mo1853 was first purified in its reduced form and formation of disulfide bond was achieved by *in vitro* oxidative folding.

Utility of the Cytochrome b_5 fusion protein system for recombinant production of peptides

The cytochrome-b_5 fusion protein system[10] has been successfully employed in the past for enhancing expression and solubility of several disulfide rich peptides during recombinant expression[5,32,33]. In the case of Mo1853, though the fusion protein expressed well, it was primarily found as a covalent dimer and a significant part of the expressed protein went into the insoluble fraction as inclusion bodies. The possible reason for formation of inclusion bodies by the fusion protein could be the locally elevated intracellular concentrations of the fusion protein accentuated by low intra-cellular solubility due to basic pI of Mo1853 (theoretical pI = 9.5) though the theoretical pI of the full length fusion protein is 5.5.

The fusion protein from the inclusion bodies could be solubilized by using insoluble protein preparation protocols carried out in reducing conditions. The reducing conditions also helped to obtain monomeric fusion protein from the disulfide linked covalent dimer. The difference in pI values of Mo1853 and cleaved Cytochrome b_5 fusion host was used to advantage during the solvent extraction step carried out after CNBr cleavage of the fusion protein. The pH of the solvent (50% Methanol) used for extracting the peptide was adjusted to ~5.5. At this pH, the cleaved fusion host remained mostly in the insoluble fraction as a precipitate (pI of Cytochrome b_5 is close to 5.2) while most of the peptide Mo1853 was obtained in to the soluble fraction. At the end of the extraction process, Mo1853 was obtained as a reduced linear peptide. In essence, despite the need to make a few modifications to the purification protocols, the Cytochrome b_5 fusion protein host system greatly facilitated the recombinant expression and purification of Mo1853.

Role of Proline in folding as a turn inducing residue

The reduced Mo1853 could be oxidized in a facile manner by incubating in a buffer (preferably volatile buffer such as ammonium bicarbonate buffer) at mild alkaline pH (pH ~7.5) for 24 h at 37°C. The ease of oxidation could probably be attributed to the presence of turn inducing residue Pro^{12} between the two cysteines in the primary structure of the peptide. A turn inducing residue would bring the cysteines closer in space and hence enhance the rate of formation of intra-molecular disulfide bond. It is interesting to note that a majority of the disulfide poor conopeptides (Fig. 1) possess at least one Proline residue in the primary structure. Even in the case of disulfide rich conotoxins, the stability and activity of the folded toxin molecules is hinged on the formation of the disulfide bonds in the correct connectivity. It seems likely that Proline residues play an important role in nucleating the folding of conotoxins and driving the formation of the correct pairing among disulfide bonds.

Activity studies of Mo1853

The sequence of Mo1853 makes it an "unclassified" conopeptide. The lack of information about the signal sequence excludes the possibility of classifying it under any superfamily. Hence, prediction of the possible molecular targets and biological activity have been difficult. The online server iCTX-type (http://lin-group.cn/server/iCTX-Type)[34] for sequence based prediction conotoxin activity, predicts that Mo1853 is likely to act on a sodium channel with a probability of 0.7. Additional investigations will be carried out as a part of future studies to identify the exact molecular target(s) and mechanism of action of Mo1853.

Tertiary structure of Mo1853 and conformational equilibrium

Structural studies by homonuclear solution NMR methods clearly showed that Mo1853 exists in solution in two conformations that correspond to the *trans* and *cis* conformations of the Lys^{11} – Pro^{12} peptide bond. The most conspicuous difference between the two structures is the helical conformation of the C-terminal residues in the *trans* conformer of Mo1853 as seen in (Fig. 9). It is important to note that the secondary chemical shifts of H^{α} and C^{α} nuclei of

the C-terminal residues are not characteristic of α-helical conformation. Further, these residues did not show protection in the H/D exchange studies. Additionally, some of the NOE correlations characteristic of α- helical conformation *viz.,* $H^N i$ to H^N_{i+2}, H^α_i to H^N_{i+3} and H^α_i to H^N_{i+4} were not observed. However, the following observations are in support of the α-helical conformation of the C-terminal residues observed in *trans* Mo1853:

- A comparison of the temperature coefficients of the amide proton chemical shifts[35] of Mo1853 (Suppl. Fig. S7) indicated that the C-terminal residues in the *trans* conformer of Mo1853 are more structured than their counterparts in the *cis* conformer of Mo1853.
- Prediction of the protein backbone flexibility using the secondary chemical shifts[36] carried out by Talos+[15] also indicated that the C-terminal residues in the *trans* conformer of Mo1853 are "ordered" while those in the *cis* conformer are "disordered".

All these observations suggest that the C-terminal helix observed in the *trans* Mo1853 is dynamic (in time scale much faster than the *cis–trans* isomerization) and that the helix observed in the tertiary structure of *trans* Mo1853 is a representation of the ensemble and time averaged conformation of these residues[37].

The presence of exchange correlations in the NOESY spectrum gave a hint about the exchange rate of the conformational equilibrium. In order to observe such ZZ-exchange correlations, the k_{ex} must not be much less than longitudinal relaxation rate (R_1) for the exchanging sites; otherwise, the signals decay due to the relaxation faster than population transfer[38]. Presence of these correlations in ROESY spectrum with an opposite sign further corroborated that their origin is chemical exchange and not cross relaxation. The exchange rate was quantified by observing coalescence of signals from exchanging nuclei in a temperature study.

The coalescence temperature was determined to be 318 K. At this temperature, the *trans* and *cis* conformations have an exchange rate of 59 Hz and an activation barrier of ~67.2 kJ mole^{-1}. The activation barrier matches with the value expected for rotation about a peptide bond (barrier for rotation about peptide bond is ~80 kJ mol^{-1}). The exchange rate is higher than what is usually observed for Proline *cis – trans* isomerization but similar values have been reported in the past[39] especially in the case of cyclized molecules. The higher exchange rate also explains why the *cis* and *trans* conformers of Mo1853 could not be separated from each other using the reverse phase HPLC.

The temperature study also revealed that the *cis* conformer of Mo1853 gets destabilized and its population decreases at lower temperatures. As described above. This observation could be rationalized based on the fact that the *cis* conformer is stabilized by hydrophobic interactions between the side chains of residues Phe[7], Val[8] and Tyr[9]. However, it was noted that none of the protons from the side chains of the interacting residues exhibited ring current shifts that are characteristic of aliphatic-aromatic interactions. One possible explanation could be that the aromatic interactions are dynamic in nature, in a much faster regime compared to the chemical shift time scale, and hence the chemical shifts are time averaged.

Another possible explanation for the stability of the *cis* conformer of Mo1853 in the absence of hydrophobic interactions could be its higher conformational entropy contributed by bond vector fluctuations in the "pico second-nano second" time scale[40,41]. A preferential reduction in the conformational entropy of the *cis*–isomer of Mo1853 at lower temperatures could explain its destabilization at lower temperatures.

In the absence of experimental evidence, it is difficult to ascertain if the presence of equally populated *cis* and *trans* X –Pro conformers and the conformational exchange in the millisecond time scale have any biological significance in the context of the Cone snail physiology. Although the phenomenon of conformational exchange could be an evolutionary side product of the primary structure of the peptide, having no real biological relevance, one could envisage situations in which conformational isomers could be of advantage to the cone snail. Two such examples are mentioned below:

- Presence of *cis–trans* conformational isomers adds to the existing sources of variability in the conotoxin repertoire of a cone snail, *viz.,* hyper variable mature sequences, "biological messiness" at transcription and translation level,[7,8] variable post-translational modifications[6] and disulfide connectivity isomers[42]. The different conformers could bind to different targets or different conformations of the same target.
- The milli second time scale of conformational exchange could have a more subtle role in

increasing the effectiveness of the venom in the following way: The cone snail venom components are known to act synergistically together in what are described as "toxin cabals"[43] to achieve a certain physiological end point in the target animal. The effectiveness of the venom depends on the proper orchestration of the action of the individual components of the venom. Since, the events involved in the ion-channel functioning are also in the milli-second time scale, the conformational exchange in the toxin molecule in a similar time scale might enable build up of the concentration of the right conformer at the correct point of time.

It will be interesting to design experiments and test the above hypotheses. For example, in the case of Mo1853, the first point mentioned above could be verified by carrying out the activity assays at lower temperatures where only the *trans* conformer is predominant. Experiments to validate the second point are more challenging as the scenario is described in the context of other toxin components in the venom.

Conclusion

In this study, the preparation of a novel, single disulfide containing, 17 residue conopeptide Mo1853 was optimized using chemical as well as recombinant expression methods. Structural studies using homonuclear solution NMR methods revealed that Mo1853 exists as two equally populated *cis* and *trans* X–Pro conformers which are in slow exchange regime, compared to the chemical shift time scale. The rate of exchange and the free energy of activation were determined at 318 K using temperature study. Tertiary structures of both the conformers include a 29-membered macro-cyclic ring formed by 9 amino acid residues which are cyclized by side chain to side chain disulfide bond. Analysis of the tertiary structures and the populations from the temperature study indicated that the *trans* conformer is stabilized by hydrogen bonds while the *cis* conformer is likely to be stabilized by hydrophobic interactions. The presence of *cis* and *trans* X-Pro peptide bond conformers will add to the existing diversity of the cone snail venom components.

Acknowledgement

A.K.K. thanks CSIR for a senior research fellowship. DBT and DST-FIST program funds to MBU are acknowledged for NMR and mass spectrometric facilities at the Indian Institute of Science.

Conflict of interest

All authors declare no competing interests

References

1 Kaas Q, Westermann JC & Craik DJ, Conopeptide characterization and classifications: an analysis using ConoServer. *Toxicon*, 55 (2010) 1491.

2 Kaas Q, Yu R, Jin AH, Dutertre S & Craik DJ, ConoServer: updated content, knowledge, and discovery tools in the conopeptide database. *Nucleic Acids Res*, 40 (2012) D325.

3 Lebbe EK & Tytgat J, In the picture: disulfide-poor conopeptides, a class of pharmacologically interesting compounds. *J Venom Anim Toxins Incl Trop Dis*, 22 (2016) 30.

4 Sudarslal S, Singaravadivelan G, Ramasamy P, Ananda K, Sarma SP, Sikdar SK, Krishnan K & Balaram P, A novel 13 residue acyclic peptide from the marine snail, *Conus monile*, targets potassium channels. *Biochemical and Biophysical Research Communications*, 317 (2004) 682.

5 Kumar GS, Ramasamy P, Sikdar SK & Sarma SP, Overexpression, purification, and pharmacological activity of a biosynthetically derived conopeptide. *Biochem Biophys Res Commun*, 335 (2005) 965.

6 Dutertre S, Jin Ah, Kaas Q, Jones A, Alewood PF & Lewis RJ, Deep venomics reveals the mechanism for expanded peptide diversity in cone snail venom. *Mol Cell Proteomics*, 12 (2013) 312.

7 Jin, Ah, Dutertre S, Kaas Q, Lavergne V, Kubala P, Lewis RJ & Alewood PF, Transcriptomic messiness in the venom duct of *Conus miles* contributes to conotoxin diversity. *Mol Cell Proteomics*, 12 (2013) 3824.

8 Tawfik DS, Messy biology and the origins of evolutionary innovations. *Nat Chem Biol*, 6 (2010) 692.

9 Mohan MK, Abraham N, Rajesh RP, Jayaseelan BF, Ragnarsson L, Lewis RJ & Sarma SP, Structure and allosteric activity of a single-disulfide conopeptide from Conus zonatus at human α3, β4 and α7 nicotinic acetylcholine receptors. *J Biol Chemy*, 295 (2020)7096.

10 Mitra A, Chakrabarti KS, Hameed MS, Srinivas KV, Kumar GS & Sarma SP, High level expression of peptides and proteins using cytochrome b_5 as a fusion host. *Protein Expr Purif*, 41 (2005) 84.

11 Kumar GS, Solution NMR studies of peptide toxins from cone snails and scorpion (PhD Thesis). Ph.D. thesis, Molecular Biophysics Unit, Indian Institute of Science, Bangalore, Indian Institute of Science, Bangalore, India, 2008.

12 Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J & Bax A, NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*, 6 (1995) 277.

13 Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J & Laue ED, The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, 59 (2005) 687.

14 Friebolin H, In *Basic One–and Two–Dimensional NMR Spectroscopy*, 3rd ed.; Bec- consall, J. K, Ed.; WILEY–VCH, (1998) 1.

15 Shen Y, Delaglio F, Cornilescu G & Bax A, TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR*, 44 (2009) 213.

16 Cheung M, Maguire ML, Stevens TJ & Broadhurst RW, DANGLE: a Bayesian inferential method for predicting

protein backbone dihedral angles and secondary structure. *J Magn Reson*, 202 (2010)223.
17 Guntert P, Structure calculation of biological macromolecules from NMR data. *Q Rev Biophys*, 31 (1998) 145.
18 Guntert P, Mumenthaler C & Wuthrich K, Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol*, 273 (1997) 283.
19 Herrmann T, Guntert P & Wuthrich K, Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol*, 319 (2002) 209.
20 Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS & Read RJ, Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*, 54 (1998) 905.
21 Brunger AT, Version 1.2 of the Crystallography and NMR system. *Nat Protoc*, 2 (2007) 2728.
22 Tejero R, Snyder D, Mao B, Aramini JM & Montelione GT, PDBStat: a universal restraint converter and restraint analysis software package for protein NMR. *J Biomol NMR*, 56 (2013) 337.
23 Wang G, Treleaven WD & Cushley RJ, Conformation of human serum apolipoprotein AI (166–185) in the presence of sodium dodecyl sulfate or dodecylphosphocholine by ^{1}H–NMR and CD. Evidence for specific peptide–SDS interactions. *Biochim Biophys Acta*, 1301 (1996)174.
24 Ahn HC, Juranic N, Macura S & Markley JL, Three–dimensional structure of the water–insoluble protein crambin in dodecylphosphocholine micelles and its minimal solvent–exposed surface. *J Am Chem Soc*, 128 (2006)4398.
25 Wuthrich, K, *NMR of Proteins and Nucleic Acids*; John Wiley & sons, (1986); pp 1–292.
26 Dorman DE & Bovey FA, Carbon–13 magnetic resonance spectroscopy. Spectrum of proline in oligopeptides. *J Org Chem*, 38 (1973) 2379.
27 Bothner-By AA, Stephens R, Lee J, Warren CD & Jeanloz R, Structure determination of a tetrasaccharide: transient nuclear Overhauser effects in the rotating frame. *J Am Chem Soc*, 106 (1984) 811.
28 Ramachandran G, Ramakrishnan C & Sasisekharan V, Stereochemistry of polypeptide chain configurations. *J Molr Biol*, 7 (1963)95.
29 Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R & Thornton JM, AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR*, 8 (1996) 477.
30 Koradi R, Billeter M & Wuthrich K, MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph*, 14 (1996) 51.
31 Lovell SC, Davis IW, Arendall WB, de Bakker PI, Word JM, Prisant MG, Richardson JS & Richardson DC, Structure validation by C^{α}geometry: φ, ψ and C^{β} deviation. *Proteins*, 50 (2003) 437.
32 Kumar GS, Upadhyay S, Mathew M & Sarma SP, Solution structure of BTK- 2, a novel hK$_v$1.1 inhibiting scorpion toxin, from the eastern Indian scorpion *Mesobuthus tamulus*. *Biochim Biophys Acta*, 1814 (2011) 459.
33 Kancherla AK, Meesala S, Jorwal P, Palanisamy R, Sikdar SK & Sarma SP, A disulfide stabilized β–sandwich defines the structure of a new cysteine framework M–superfamily conotoxin. *ACS Chem Biol*, 10 (2015) 1847.
34 Ding H, Deng EZ, Yuan LF, Liu L, Lin H, Chen W & Chou KC, iCTX- Type: A sequence-based predictor for identifying the types of conotoxins in targeting ion channels. *BioMed Res Int*, (2014) 1.
35 Baxter NJ & Williamson MP, Temperature dependence of ^{1}H chemical shifts in proteins. *J Biomol NMR*, 9 (1997) 359.
36 Berjanskii MV & Wishart DS, A simple method to predict protein flexibility us- ing secondary chemical shifts. *J Am Chem Soc*, 127 (2005)14970.
37 Cavanagh J, Fairbrother WJ, Palmer III AG & Skelton NJ, *Protein NMR Spectroscopy: Principles and Practice*, second edition ed.; Academic Press, (2007) 1.
38 Cavanagh J, Fairbrother WJ, Palmer III AG & Skelton NJ, Protein NMR Spectroscopy: Principles and Practice, second edition. Academic Press, (2007) 706.
39 Grathwohl C & Wuthrich K, NMR studies of the rates of proline cis–trans iso- merization in oligopeptides. *Biopolymers*, 20 (1981) 2623.
40 Yang D & Kay LE, Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J Mol Biol*, 263 (1996) 369.
41 Tzeng SR & Kalodimos CG, Protein activity regulation by conformational entropy. *Nature*, 488 (2012) 236.
42 Khoo KK, Gupta K, Green BR, Zhang MM, Watkins M, Olivera BM, Balaram P, Yoshikami D, Bulaj G & Norton RS, Distinct disulfide isomers of μ-conotoxins KIIIA and KIIIB block voltage-gated sodium channels. *Biochemistry*, 51 (2012)9826.
43 Terlau H & Olivera BM, *Conus* venoms: a rich source of novel ion channel- targeted peptides. *Physiol Rev*, 84 (2004) 41.