Protein Expression and Purification 76 (2011) 248–253

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

Heterologous expression, purification and characterization of heterodimeric monellin

Nilesh Aghera, Jayant B. Udgaonkar *

National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore 560065, India

ARTICLE INFO

Article history:
Received 15 September 2010
and in revised form 26 October 2010
Available online 6 November 2010

Keywords:
Monellin
pETDUET
Protein purification
Co-expression

ABSTRACT

Monellin is an intensely sweet-tasting protein present in the berry of *Dioscoreophyllum cumminsii*. Naturally occurring monellin (double chain monellin) is a heterodimer of two subunits commonly referred to as chain A and chain B. Monellin is a good model system for structural and dynamic studies of proteins. Single chain monellin has been generated by covalently linking the two subunits of naturally occurring double chain monellin, and has been used extensively for folding and unfolding studies, as well as for protein aggregation studies. There are, however, relatively few reports on such studies with double chain monellin. The primary difficulty associated with studies using double chain monellin appears to be the lack of a standard purification method. Here, a simple method for the purification of double chain monellin is presented. The genes encoding the two chains of monellin were cloned into a modified pETDUET vector under separate T7 promoters. The expression vector containing the genes of the two chains was expressed in *E. coli* BL21 Star (DE3). The expressed protein was purified using two steps of chromatography, ion exchange chromatography and gel filtration chromatography. This expression system consistently produced 40 mg of pure double chain monellin per litre of *E. coli* culture, in the correctly folded native state. The purity of the protein was confirmed by mass spectrometry and SDS–PAGE analysis. The purified protein was characterized using different spectroscopic methods, and the spectra obtained were in good agreement with the published spectra of naturally occurring double chain monellin.

Introduction

Monellin is a sweet protein found in the fruit of *Dioscoreophyllum cumminsii*. It is 3000 times sweeter than sucrose on the basis of weight [1], and 10^5 times sweeter on the basis of molarity [2]. Early on, monellin was believed to be a carbohydrate [3], but later two groups independently showed that the sweet taste of the berry of *D. cumminsii* arises from a protein [4,5]. Monellin is now known to be made up of two subunits, namely chain A and chain B. Chain A consists of 45 amino acid residues that form three antiparallel β-strands. Chain B consists of 50 amino acid residues that form two β-strands separated by an α-helix [6]. The two subunits are held together by non-covalent interactions. The five β-strands from both subunits form a single β sheet [7], and the protein adopts a β-grasp fold. This is among the ten most common protein folds, and the protein is a member of the cystatin protein family [8].

Monellin has been studied intensively by protein chemists because of its intense sweet taste. Several biophysical studies have been carried out to understand the mystery of the sweet taste of the monellin. It is clear that the native structure of the protein is essential for its sweet taste [9–11]. Several NMR studies have given insights into the "active site" of monellin which contributes to the sweet taste [10,12]. This appears to be located in a region containing residues 16–35 of chain A, which form a loop between the β-strands 4 and 5 [12]. Studies performed to understand the interaction between monellin and taste receptors, have indicated that the sweet sensation arises due to the interaction of monellin with the T1R2 and T1R3 receptors [13–15]. Monellin has also been used in gustatory studies with several animals to understand the neuronal basis of the taste sensation [16–19].

Protein chemists had intended to use this protein as a commercial low calorie sweetener [9], but its application was limited by its low temperature stability. The stability of monellin can be improved by covalently connecting its subunits. In the native state, the C-terminus of chain B is close to the N-terminus of the chain A. A single chain variant of monellin was generated by connecting the two ends of the chains. Two variants of single chain monellin have been made, one without a linker connecting chains A and B named as SCM [9], and the other with a Gly-Phe dipeptide linker named as MNEI [12]. The single chain variants of monellin remain native even at relatively high temperatures [9]. Structural studies have shown that single chain monellin has the same conformation and sweet taste as does double chain monellin [7,20].

* Corresponding author. Fax: +91 80 23636662.
E-mail address: jayant@ncbs.res.in (J.B. Udgaonkar).

1046-5928/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2010.11.002

Apart from its sweet taste, monellin is also a model system of choice for protein folding and aggregation studies because of its small size. Detailed studies have been done with single chain monellin to understand the folding and unfolding mechanisms [21–26], as well as structural dynamics [27,28]. A GroEL-assisted folding study of single chain monellin has given insight into the mechanism of chaperon-assisted folding of the protein [29]. Single chain monellin has also been used as a model system for understanding the protein aggregation process [30].

Double chain monellin is an attractive model system for the study of how chain complementation occurs during the folding of a heterodimeric protein, and for delineating the process of chain binding from folding. But there have been very few studies of the folding or aggregation of the double chain protein [31–35]. Such studies have been limited by the availability of the protein. So far, double chain monellin has been purified from the berry of *D. cumminsii* [1]. The commercially available protein is very expensive, and its purity is only 60–70%. It also invariably has two forms of chain A as well as of chain B. About 50% of the protein molecules contain a chain A lacking the phenylalanine residue at the N-terminus and 10% of the protein molecules contain a chain B lacking the glycine residue at the N-terminus [34,36].

Several reports have been published on the heterologous expression and purification of single chain monellin from different host systems. Work on the purification of single chain monellin has also been reviewed [37,38]. Surprisingly, there is not a single report on the heterologous expression and purification of double chain monellin. Here, we report a method for the heterologous co-expression of the two chains of monellin, in *E. coli*, and the purification of the double chain protein. The purification method is relatively simple, involving only two steps of chromatography. The only change in protein sequence is that an additional methionine gets introduced at the N-terminus of each chain. It shows absorbance, fluorescence and circular dichroism (CD) spectra similar to those reported for naturally occurring double chain monellin [4,34,39].

Materials and methods

All chemicals used for the purification were of analytical grade. Restriction enzymes were from Fermentas, and DNA ligase was from USB. DNA was synthesized by Genscript (USA).

Bacterial strain and plasmid

Escherichia coli DH5α was used as a host strain for cloning, while BL21 Star (DE3) was used as the host strain for protein expression. The pETDUET (Novagen) vector [40,41] was used to design the expression system.

Construction of pETDUET-DCMN

The genes encoding chain A and chain B were synthesized by Genscript in a pUC57 vector. The codons of the genes encoding chain A and chain B were optimized for *E. coli*. The synthesized DNA construct and pETDUET were digested by NcoI and BglII restriction enzymes (located in MCS1 and MCS2, respectively). The digested products were gel purified and ligated into pETDUET using T4 DNA ligase. The final construct was verified by DNA sequencing. The DNA sequences of the genes for chain A and chain B, as well as the protein sequences are given below.

Chain A: ATGCGTGAAATTAAAGGCTATGAATATCAGCTGTATGTG
TATGCGAGCGATAAACTGTTTCGTGCGGATATTAGCGAAGATTATAAA
ACCCGTGGCCGTAAACTGCTGCGTTTTAACGGCCCGGTGCCGCCGCCG

MREIKGYEYQLYVYASDKLFRADISEDYKTRGRKLLRFNGPVPPP

Chain B: ATGGGTGAATGGGAAATTATCGATATTGGTCCGTTTACA
CAGAACCTGGGCAAATTTGCTGTGGATGAAGAAAATAAAATCGGCCA
GTATGGTCGTCTGACGTTTAATAAAGTCATTCGTCCGTGTATGAAAAA
GACAATTTATGAAAATGAA

MGEWEIIDIGPFTQNLGKFAVDEENKIGQYGRLTFNKVIRPCMKKTIY
ENE

Protein expression and purification

BL21 Star (DE3) cells were freshly transformed with the pETDUET-DCMN construct. About 5–6 colonies were inoculated into 300 ml of LB starter culture containing 100 μg/ml ampicillin and the culture was grown at 37 °C until the OD_{600} reached 0.6–0.8. 10% inoculum was used from the starter culture to inoculate 3 L rich medium containing 100 μg/ml ampicillin. The culture was induced at an OD_{600} of 1, with IPTG[1] at a concentration of 250 μM, and then further grown for another 12 h. The bacterial cells were harvested by centrifugation at 6000*g* for 10 min at 4 °C, and then resuspended in 100 ml ice-cold 20 mM sodium acetate buffer containing 1 mM PMSF at pH 5.5. The cells, kept on ice, were then lysed for 25 min using a sonicator (Vibra cell from Sonics) at 60% amplitude setting, with 5 s on-pulses and 3 s off-pulses. Subsequently, the supernatant was collected after centrifugation at ⩾25,000*g* for 45 min at 4 °C. The supernatant was diluted to a final volume of 500 ml and was stored at 4 °C for 3 h. The diluted supernatant was centrifuged again at ⩾25,000*g* for 45 min at 4 °C. The supernatant was collected and loaded onto a 100 ml fast flow SP-Sepharose (Sigma) ion-exchange column at 4 °C. After loading the column, it was washed with at least 10 column volumes of the same buffer. The protein was eluted out using 20 mM sodium acetate buffer containing 100 mM NaCl at pH 5.5 (isocratic elution). The eluted protein was concentrated by ultra-filtration using a 3 kDa cut-off membrane. The concentrated protein was further purified using a Hi Load 16/60 Superdex 75 preparation grade column using an AKTA chromatography system from GE. Gel filtration was performed in 20 mM sodium acetate buffer containing 200 mM NaCl 1 mM DTT. The protein elution started at 60% of the column volume. The protein was desalted by ultra-filtration using a 3 kDa cut-off membrane, and was stored after lyophilization.

Separation and purification of the chains of monellin

The individual chains of monellin were purified using reverse phase chromatography (RPC). RPC separation was performed in the presence of 0.05% (v/v) TFA. Desalted protein was loaded onto a Resource 15 column from GE in solvent containing 20% (v/v) methanol at a flow rate of 2 ml/min. The chains were eluted using gradient of methanol from 20% to 100%. Chain A eluted at 60% methanol, and chain B eluted at 75% methanol. Methanol was removed by vacuum evaporation before lyophilization.

ESI-MS analysis

Desalted protein was dissolved in 40% acetonitrile/0.1% formic acid for mass spectrometric analysis. Mass spectra were acquired in ESI-MS mode on a Q-TOF machine from Waters.

SDS–PAGE

Tricine-SDS–PAGE was performed using a 16% resolving gel, as described previously [42,43]. Protein samples were loaded onto the gel without any heating in the presence of 2-mercaptoethanol.

[1] *Abbreviations used:* IPTG, isopropyl-β-D-thiogalactopyranoside; TFA, trifluoroacetic acid; PMSF, phenylmethylsulfonyl fluoride; DTT, dithiothreitol; DCMN, double chain monellin.

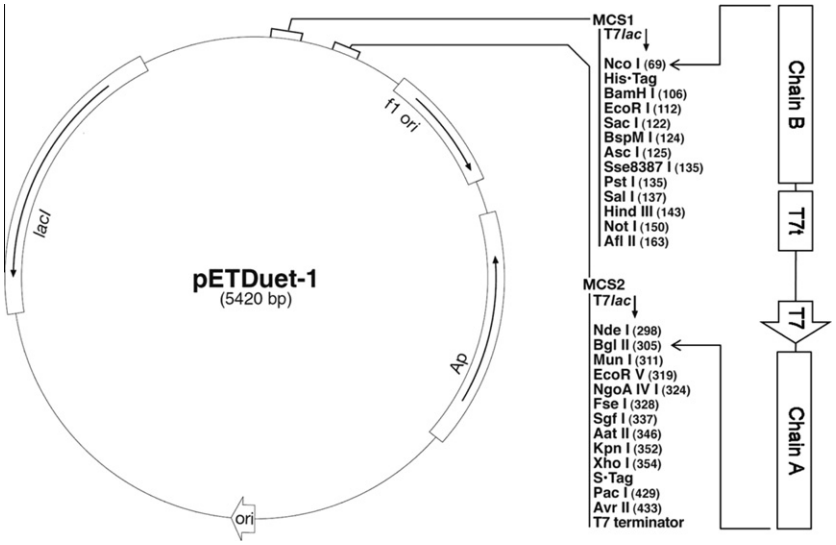

Fig. 1. Map of the plasmid constructed for the co-expression of chain A and chain B in *E. coli* BL21 Star (DE3). A synthesized DNA segment containing the genes of chain A and chain B, was inserted at the NcoI and BglII restriction sites located in MCS1 and MCS2, respectively. T7 represents the T7 promoter, and T7t represents the T7 termination site.

Electrophoresis was carried out at room temperature (25 °C) at constant 50 V for 12 h (Bio-Rad). The gel was stained with Coomassie brilliant blue R-250 and destained with 10% acetic acid.

Absorbance spectra

All spectroscopic studies of monellin and its chains were performed at pH 7. The absorbance spectra of monellin and chain B were acquired in 10 mM phosphate buffer containing 0.25 mM EDTA and 1 mM DTT. The absorbance spectrum of chain A was acquired in water because of the poor solubility of chain A in the presence of salt. Absorbance spectra were acquired using an Ultrospec 4000 UV/Vis spectrophotometer from Pharmacia. All absorbance spectra were acquired at 40 μM protein concentration using a 10 mm path length cuvette.

Fluorescence spectra

Fluorescence spectra were acquired using a Spex Fluoromax-3 spectrofluorometer. Spectra were acquired at 10 μM protein concentration in 50 mM phosphate buffer in the presence of 0.25 mM EDTA and 1 mM DTT. The protein was excited at 280 nm and the fluorescence emission was collected from 300 to 400 nm. The excitation bandwidth was 0.6 nm and the emission band width was 10 nm. Each spectrum was collected with a 2 s integration time.

CD spectra

CD spectra were acquired using a Jasco J-815 spectropolarimeter. Spectra were collected in 50 mM phosphate buffer in the presence of 0.25 mM EDTA and 1 mM DTT. Near-UV CD spectra were collected at 100 μM protein concentration using 1 cm path length cuvette. Far-UV CD spectra were collected at 10 μM protein concentration using a 0.2 cm path length cuvette. All CD spectra were acquired at a scan speed 50 nm/min, and a response time of 2 s. The band width was set to 1 nm.

Results

pETDEUT-DCMN

The monellin expression system was designed using a pETDUET vector and was expressed in *E. coli* strain BL21 Star (DE3). The DNA sequence containing the genes of chain A and chain B was synthesized. The synthesized DNA sequence contained the transcription termination sequence located downstream of the chain B gene, and the transcription initiation site with the ribosome binding site upstream of the chain A gene. The transcription termination sequence of chain B and the transcription initiation sequence of chain A were separated by a short stretch of DNA. The purpose of having two transcription termination sequences, one after the gene for each chain, was to ensure that both chains expressed to similar extents. Both chains of monellin are known to be unstructured [34], and it was thought that biased expression of either one of the two chains might lead to the formation of inclusion bodies rather than to the formation of native monellin. It was, however, not determined whether in the absence of the transcription termination sequence after the gene for the B chain, the protein yield is affected or not.

The synthesized DNA sequence was ligated into the pETDUET vector in between the NcoI and BglII sites. The final construct is shown in Fig. 1. The original pETDUET vector was modified for the expression of monellin. In the final construct, pETDUET-DCMN contains the transcription termination sequence after MCS1 (chain B), whereas the original pETDUET vector does not contain it. The final construct contains the two genes encoding chain B and chain A, with independent transcription initiation sites upstream and transcription termination sites downstream. After ligation, positive clones were selected on an LB agar plate containing 100 μg/ml ampicillin. Selected colonies were further screened by restriction digestion. The clone was confirmed by DNA sequencing.

Expression and purification

The protein was expressed in BL21 Star (DE3). All incubations were done at 37 °C. Different IPTG concentrations were tried out under the same incubation condition. IPTG (250 μM) gave the highest yield of protein, and no further increase in the yield of the protein was seen at higher IPTG concentrations. Sonication was done in a beaker kept in ice with 5 s on-pulses and 3 s off-pulses for 25 min. Sonication was stopped periodically to ensure that the lysate remained at a low temperature, to avoid protein denaturation. The expressed protein was present in the soluble fraction of the lysate. After sonication, the supernatant was separated by centrifugation. The supernatant was incubated for 3 h at 4 °C. During incubation, many unwanted proteins

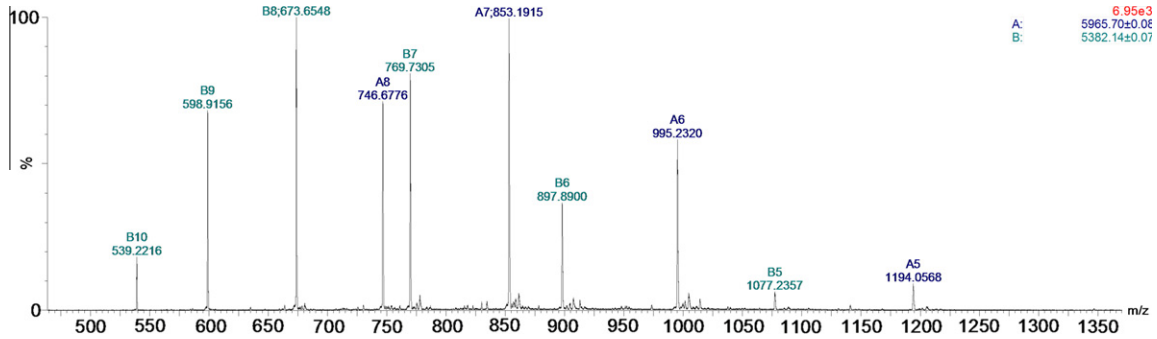

Fig. 2. Mass spectrum of double chain monellin acquired in the ESI-MS mode. The intact protein dissociated into chain A and chain B under the protein ionization conditions. The masses obtained for chain A and chain B are 5382.1 Da and 5965.7 Da, respectively, which are in good agreement with the calculated masses 5382.1 Da and 5965.8 Da.

precipitated out, but monellin did not undergo precipitation during this incubation. The supernatant was separated by centrifugation and loaded onto a SP-Sepharose ion exchange column. At pH 5.5, monellin has a net positive charge of 3, which allows weak binding between the protein and SP-Sepharose. Weakly bound protein can be eluted using a low salt concentration. Isocratic elution was done using 100 mM NaCl, and the eluted protein was 95% pure, as seen in the mass spectrum and by SDS–PAGE. The ratio of the absorbance at 280 to that at 260 nm was poor, in the range 1.2–1.4, which indicated some non-protein impurity. Gel filtration chromatography was then performed using a Superdex G75 column. After gel filtration, the ratio of absorbance at 280 to that at 260 improved to 1.8. The protein was desalted after gel filtration by ultra-filtration using a 3 kDa cut-off membrane. The protein was lyophilized and stored for future use. The purity of the protein after the final step was >98% as seen in the mass spectrum (Fig. 2) and SDS–PAGE (Fig. 3A).

Separation of the chains of monellin

The chains of monellin were separated using reverse phase chromatography. Methanol was used as the organic hydrophobic solvent, and TFA as the ion-pairing agent. Chromatographic separation was monitored using absorbance at 280 nm and at 295 nm. Chain A does not contain any tryptophan, and hence, does not absorb at 295 nm. Thus, the chromatogram itself indicates the quality of separation (Fig. 3B). The chains were separated by a shallow gradient of methanol. Chain A eluted first at around 60%, and chain B eluted later at around 75% methanol concentration. The result indicates that chain B is more hydrophobic than chain A in the presence of TFA. After the separation of the chains, methanol was removed by vacuum evaporation, and the chains were lyophilized.

ESI-MS analysis

ESI-MS analysis was performed with monellin, and the result is shown in Fig. 2. During mass spectrometric analysis, the chains of monellin dissociated from each other under the ionization conditions. Thus, the mass spectrum shows peaks that correspond to two different polypeptides. Analysis of the mass spectrum gave masses of 5382.1 Da and 5965.7 Da. The obtained masses are in good agreement with the theoretically calculated masses, which are 5382.1 Da for chain A and 5965.8 Da for chain B. The mass spectrometry result confirms the presence of the protein of interest, without any modification or adduct. The mass spectrum does not show any other peaks apart from those for chain A and chain B, indicating the very high purity of the protein. The mass spectrometric analysis of the separated chains was also done, and gave the expected masses, indicating that no modification of the chain occurred during the process of reverse phase chromatography (data not shown).

Absorbance spectra

Absorbance spectra are shown in panel A of Fig. 4. Spectra were acquired at 40 μM protein concentration at pH 7 in 10 mM

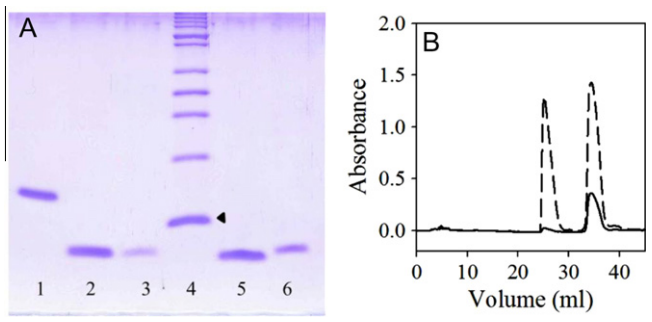

Fig. 3. Purity of monellin and separation of the chains A and B of monellin. (A) SDS–PAGE analysis of the purified protein. Lane 1 is single chain monellin, lane 2 and 3 are double chain monellin, 2 μg and 0.2 μg, respectively, lane 5 and 6 are chain A and chain B, respectively. The protein band in lane 4 marked with the arrow corresponds to a mass of 10 kDa. (B) Reverse phase chromatogram. Reverse phase chromatography was performed to separate the chains of monellin. In the chromatogram, the dashed line represents the absorbance monitored at 280 nm, and the solid line represents the absorbance monitored at 295 nm.

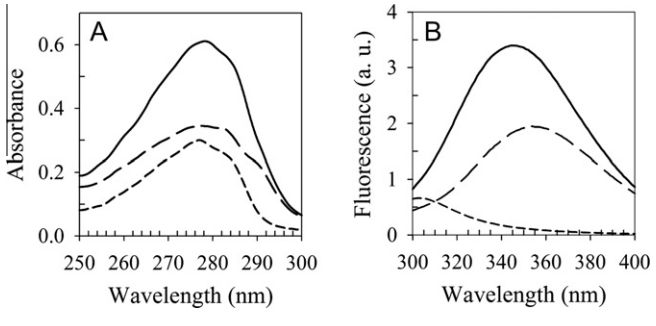

Fig. 4. Absorbance and fluorescence spectra of monellin and its individual chains. Panel A shows the ultraviolet absorbance spectra of double chain monellin and its purified chains. All absorbance spectra were acquired at 40 μM protein concentration. Double chain monellin and chain B were in 10 mM phosphate buffer, while chain A was prepared in milli-Q water. Panel B shows the fluorescence spectra of monellin and its chains. All fluorescence spectra were acquired at 10 μM protein concentration in 50 mM phosphate buffer, pH 7. In each panel, the short dashed line represents chain A, the long dashed line represents chain B, and the solid line represents double chain monellin.

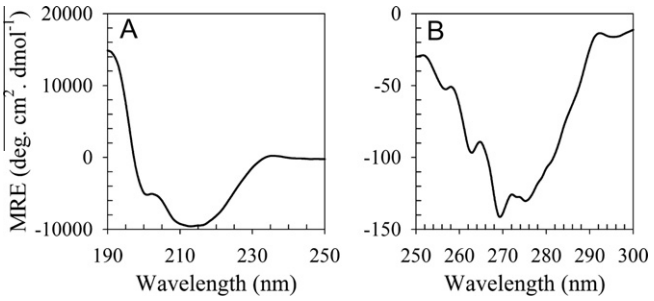

Fig. 5. CD spectra of double chain monellin. Mean residue ellipticity is plotted against wavelength. (A) Far-UV CD spectrum of native double chain monellin. (B) Near-UV CD spectrum of native double chain monellin.

phosphate buffer containing 0.25 mM EDTA and 1 mM DTT. The published values of the extinction coefficients were used to calculate the concentrations of chain A, chain B and monellin [4,34]. The published values were confirmed by estimating protein concentration using the BCA protein estimation method [44]. Each protein shows an absorbance maximum at 277 nm. Monellin and chain B show similar absorbance spectra with only a difference in their total absorbance. Chain A, which does not contain any tryptophan residue, does not show absorbance at wavelengths longer than 300 nm.

Fluorescence spectra

Fluorescence spectra of monellin and its chains are shown in panel B of Fig. 5. They were acquired by exciting monellin at 280 nm which is close to the wavelength of maximum absorbance of monellin. At 280 nm, tyrosine and tryptophan both absorb and contribute to fluorescence. The fluorescence spectrum was collected from 300 nm to 400 nm. It appears very symmetric, with the maximum of the spectrum at 346 nm. This maximum at 346 nm arises from the fluorescence of the single tryptophan residue, which is the fourth residue of chain B. Monellin contains seven tyrosine residues, but the protein does not show any fluorescence maximum in the region of 300–305 nm, where tyrosine emission peaks are typically seen. This is because the absorbance at 280 nm of tyrosine is much less than that of the tryptophan residue. Furthermore, the absorption spectrum of tryptophan overlaps with the emission spectrum of tyrosine. In native conditions, where the tyrosine and tryptophan residues are very close, tyrosine fluorescence is quenched by the tryptophan residue through the process of fluorescence resonance energy transfer (FRET). The fluorescence spectrum of chain A, which contains five tyrosine residues shows an emission maximum at 304 nm. Chain A does not show any fluorescence at any wavelength longer than 380 nm, which indicates that it is free of any contamination by chain B or any other tryptophan-containing protein. Chain B contains two tyrosine residues and one tryptophan residue. The fluorescence spectrum of chain B shows a fluorescence maximum at 355 nm, which indicates that the tryptophan is solvent-exposed. This observation suggests that chain B in isolation loses its native structure.

CD spectra

Monellin was characterized using far-UV and near-UV CD spectroscopy. Panel A of Fig. 5 shows the far-UV CD spectrum of monellin, whereas panel B shows the near-UV CD spectrum of monellin. The far-UV spectrum of monellin shows a single minimum at 214 nm, and there is no minimum observed at 208 nm or 222 nm. The result indicates that the protein is rich in β-sheet content and poor in α-helix content. The result agrees with the crystal structure of the protein and with published spectra [39]. The mean residue ellipticity (MRE) of the protein at 214 nm is ~9000 deg cm^2 mol^{-1}. The near-UV CD spectrum of monellin shows 4 local minima between 255 and 275 nm, and a global minimum at 269 nm. The near-UV CD spectrum suggests that the aromatic amino acid residues are in asymmetric environments.

Discussion

Different methods have been standardized for single chain monellin expression and purification using different host systems. Single chain monellin has been expressed in eukaryotic as well as in prokaryotic systems [45–47]. It has also been expressed in plants [48,49]. There are also reports of the solid phase synthesis of single chain monellin and chain B [35,50,51]. Surprisingly, there is not a single report on the heterologous expression and purification of wild type double chain monellin. In this study, we have developed a protocol for the heterologous expression and purification of double chain monellin.

Our protocol for the purification of double chain monellin is simple, and involves only two chromatographic steps. Our expression system does not use any protein tag for the expression and purification of protein. The purified protein is more than 98% pure as revealed by the mass spectrometry and SDS–PAGE. The yield of the protein is also good; up to 40 mg monellin per litre of *E. coli* culture are obtained. The protein purified using this method retains the first methionine residue on both subunits at their N-termini, without any detectable cleavage. This is commonly seen in the case of many recombinantly-expressed proteins. The expressed protein is suitable for all kinds of structural and biophysical studies, including mass spectrometric and aggregation studies.

Single chain monellin has a higher temperature stability than wild type monellin. This increases its applicability and for that reason considerable work has been done with the single chain protein. The availability of several protocols for the purification of single chain monellin has facilitated its use for studies of protein folding, unfolding and aggregation. The only purification method reported for wild type monellin is from its natural source, the fruit of *D. cumminsii* [1].The protein is therefore limited by the availability of the fruit around the world as a source of the protein. Currently, commercially available protein is very expensive and sold at very low purity. This report will be of help to researchers interested in studying double chain monellin.

Spectroscopic characterization was performed using different probes. All the spectroscopic data are in good agreement with published data [4,34,39] on the protein purified from its natural source. The fluorescence spectrum of native monellin shows a λ_{max} at 346 nm, which indicates that the tryptophan residue is buried. The observation is in accordance with the crystal structure [6]. The fluorescence spectrum of chain B shows a λ_{max} at 355 nm, which indicates that the tryptophan residue is exposed to the solvent. This observation suggests that chain B in isolation has non-native structure. The CD spectrum shows a minimum at 214 nm, which indicates a protein rich in β-sheet content. Again, the result is in accordance with the crystal structure of protein.

Very limited studies have been done so far with double chain monellin. Work with the commercially available wild type protein has given insight into the contribution of electrostatic interactions to the stability of the double chain protein [34,35]. Work on the aggregation of monellin has been reported but has not been continued [32,33]. We believe that the availability of the purification method will accelerate research on the folding and aggregation of double chain monellin.

In summary, we have reported a method that enables the expression and purification of double chain monellin. The purified

protein has an extra methionine residue at the N-terminus of each of the chains. The method allows purification of 40 mg of double chain monellin per litre of *E. coli* culture, and the protein obtained has more than 98% purity. The purified protein has spectroscopic properties that agree with the previously published results of studies on the commercially available protein from its natural source.

Acknowledgements

We thank Prof. R. Varadrajan for suggesting the use of the pETDUET vector for expressing the protein. The mass spectra were obtained in the Mass Spectrometry Facility of NCBS. This work was funded by the Tata Institute of Fundamental Research and by the Departments of Biotechnology and of Science and Technology, Government of India. JBU is the recipient of a J.C. Bose National Fellowship from the Government of India.

References

[1] J.A. Morris, R.H. Cagan, Purification of monellin, the sweet principle of *Dioscoreophyllum cumminsii*, Biochimica et Biophysica Acta 261 (1972) 114–122.
[2] R.H. Cagan, in: Chemostimulatory protein: a new type of taste stimulus, Science (New York) 181 (1973) 32–35.
[3] I. G. E, M.A.Y. Joann F, Serendipity Berries–Source of a New Intense Sweetener, Journal of Food Science, 1969, pp. 408–411.
[4] J.A. Morris, R. Martenson, G. Deibler, R.H. Cagan, Characterization of monellin, a protein that tastes sweet, The Journal of Biological Chemistry 248 (1973) 534–539.
[5] H. van der Wel, Isolation and characterization of the sweet principle from *Dioscoreophyllum cumminsii* (stapf) diels, FEBS Letters 21 (1972) 88–90.
[6] C. Ogata, M. Hatada, G. Tomlinson, W.C. Shin, S.H. Kim, Crystal structure of the intensely sweet protein monellin, Nature 328 (1987) 739–742.
[7] J.R. Somoza, F. Jiang, L. Tong, C.H. Kang, J.M. Cho, S.H. Kim, Two crystal structures of a potently sweet protein. Natural monellin at 2.75 A resolution and single-chain monellin at 1.7 A resolution, Journal of Molecular Biology 234 (1993) 390–404.
[8] A.G. Murzin, Sweet-tasting protein monellin is related to the cystatin family of thiol proteinase inhibitors, Journal of Molecular Biology 230 (1993) 689–694.
[9] S.H. Kim, C.H. Kang, R. Kim, J.M. Cho, Y.B. Lee, T.K. Lee, in: Redesigning a sweet protein: increased stability and renaturability, Protein Engineering 2 (1989) 571–575.
[10] M. Kohmura, N. Nio, Y. Ariyoshi, Highly probable active site of the sweet protein monellin, Bioscience, Biotechnology, and Biochemistry 56 (1992) 1937–1942.
[11] J.A. Morris, R.H. Cagan, Enzymatic proteolysis of monellin. Absence of sweet peptides, Journal of Agricultural and Food Chemistry 24 (1976) 1075–1077.
[12] T. Tancredi, H. Iijima, G. Saviano, P. Amodeo, P.A. Temussi, Structural determination of the active site of a sweet protein. A ^{1}H NMR investigation of pMNEI, FEBS Letters 310 (1992) 27–30.
[13] R. Spadaccini, F. Trabucco, G. Saviano, D. Picone, O. Crescenzi, T. Tancredi, P.A. Temussi, The mechanism of interaction of sweet proteins with the T1R2-T1R3 receptor: evidence from the solution structure of G16A-MNEI, Journal of Molecular Biology 328 (2003) 683–692.
[14] T. Tancredi, A. Pastore, S. Salvadori, V. Esposito, P.A. Temussi, Interaction of sweet proteins with their receptor. A conformational study of peptides corresponding to loops of brazzein, monellin and thaumatin, European Journal of Biochemistry/FEBS 271 (2004) 2231–2240.
[15] P.A. Temussi, Why are sweet proteins sweet? Interaction of brazzein, monellin and thaumatin with the T1R2-T1R3 receptor, FEBS Letters 526 (2002) 1–4.
[16] G. Hellekant, Y. Ninomiya, V. Danilova, Taste in chimpanzees II: single chorda tympani fibers, Physiology and Behavior 61 (1997) 829–841.
[17] C.R. Plata-Salaman, T.R. Scott, V.L. Smith-Swintosky, Gustatory neural coding in the monkey cortex: the quality of sweetness, Journal of Neurophysiology 69 (1993) 482–493.
[18] A. Schilling, V. Danilova, G. Hellekant, Behavioral study in the gray mouse lemur (Microcebusmurinus) using compounds considered sweet by humans, American Journal of Primatology 62 (2004) 43–48.
[19] K. Tonosaki, K. Miwa, F. Kanemura, Gustatory receptor cell responses to the sweeteners, monellin and thaumatin, Brain Research 748 (1997) 234–236.
[20] R. Spadaccini, O. Crescenzi, T. Tancredi, N. De Casamassimi, G. Saviano, R. Scognamiglio, A. Di Donato, P.A. Temussi, in: Solution structure of a sweet protein: NMR study of MNEI a single chain monellin, Journal of Molecular Biology 305 (2001) 505–514.
[21] S.K. Jha, D. Dhar, G. Krishnamoorthy, J.B. Udgaonkar, Continuous dissolution of structure during the unfolding of a small protein, Proceedings of the National Academy of Sciences of the United States of America 106 (2009) 11113–11118.
[22] S.K. Jha, J.B. Udgaonkar, Direct evidence for a dry molten globule intermediate during the unfolding of a small protein, Proceedings of the National Academy of Sciences of the United States of America 106 (2009) 12289–12294.
[23] T. Kimura, A. Maeda, S. Nishiguchi, K. Ishimori, I. Morishima, T. Konno, Y. Goto, S. Takahashi, Dehydration of main-chain amides in the final folding step of single-chain monellin revealed by time-resolved infrared spectroscopy, Proceedings of the National Academy of Sciences of the United States of America 105 (2008) 13391–13396.
[24] T. Kimura, T. Uzawa, K. Ishimori, I. Morishima, S. Takahashi, T. Konno, S. Akiyama, T. Fujisawa, Specific collapse followed by slow hydrogen-bond formation of beta-sheet in the folding of single-chain monellin, Proceedings of the National Academy of Sciences of the United States of America 102 (2005) 2748–2753.
[25] A.K. Patra, J.B. Udgaonkar, in: Characterization of the folding and unfolding reactions of single-chain monellin: evidence for multiple intermediates and competing pathways, Biochemistry 46 (2007) 11727–11743.
[26] Y.H. Sung, H.D. Hong, C. Cheong, J.H. Kim, J.M. Cho, Y.R. Kim, W. Lee, Folding and stability of sweet protein single-chain monellin. An insight to protein engineering, The Journal of Biological Chemistry 276 (2001) 44229–44238.
[27] J. Peon, S.K. Pal, A.H. Zewail, in: Hydration at the surface of the protein monellin: dynamics with femtosecond resolution, Proceedings of the National Academy of Sciences of the United States of America 99 (2002) 10964–10969.
[28] J. Xu, D. Toptygin, K.J. Graver, R.A. Albertini, R.S. Savtchenko, N.D. Meadow, S. Roseman, P.R. Callis, L. Brand, J.R. Knutson, Ultrafast fluorescence dynamics of tryptophan in the proteins monellin and IIAGlc, Journal of the American Chemical Society 128 (2006) 1214–1221.
[29] A.K. Patra, J.B. Udgaonkar, GroEL can unfold late intermediates populated on the folding pathways of monellin, Journal of Molecular Biology 389 (2009) 759–775.
[30] V. Esposito, F. Guglielmi, S.R. Martin, K. Pauwels, A. Pastore, R. Piccoli, P.A. Temussi, Aggregation mechanisms of cystatins: a comparative study of monellin and oryzacystatin, Biochemistry 49 (2010) 2805–2810.
[31] P. Fan, C. Bracken, J. Baum, Structural characterization of monellin in the alcohol-denatured state by NMR: evidence for beta-sheet to alpha-helix conversion, Biochemistry 32 (1993) 1573–1582.
[32] T. Konno, Multistep nucleus formation and a separate subunit contribution of the amyloid genesis of heat-denatured monellin, Protein Science 10 (2001) 2093–2101.
[33] T. Konno, K. Murata, K. Nagayama, in: Amyloid-like aggregates of a plant protein: a case of a sweet-tasting protein monellin, FEBS Letters 454 (1999) 122–126.
[34] W.F. Xue, J. Carey, S. Linse, Multi-method global analysis of thermodynamics and kinetics in reconstitution of monellin, Proteins 57 (2004) 586–595.
[35] W.F. Xue, O. Szczepankiewicz, M.C. Bauer, E. Thulin, S. Linse, in: Intra- versus intermolecular interactions in monellin: contribution of surface charges to protein assembly, Journal of Molecular Biology 358 (2006) 1244–1255.
[36] M. Kohmura, N. Nio, Y. Ariyoshi, Complete amino acid sequence of the sweet protein monellin, Agricultural and Biological Chemistry 54 (1990) 2219–2224.
[37] I. Faus, Recent developments in the characterization and biotechnological production of sweet-tasting proteins, Applied Microbiology and Biotechnology 53 (2000) 145–151.
[38] T. Masuda, N. Kitabatake, Developments in biotechnological production of sweet proteins, Journal of Bioscience and Bioengineering 102 (2006) 375–389.
[39] W.F. Xue, O. Szczepankiewicz, E. Thulin, S. Linse, J. Carey, Role of protein surface charge in monellin sweetness, Biochimica et Biophysica Acta 1794 (2009) 410–420.
[40] K. Ponniah, T.S. Loo, P.J. Edwards, S.M. Pascal, G.B. Jameson, G.E. Norris, The production of soluble and correctly folded recombinant bovine beta-lactoglobulin variants A and B in Escherichia coli for NMR studies, Protein Expression and Purification 70 (2010) 283–289.
[41] X. Wang, Z. Zheng, P. Dou, J. Qin, X. Wang, C. Ma, H. Tang, P. Xu, Cloning, expression, purification, and activity assay of proteins related to D-lactic acid formation in Lactobacillus rhamnosus, Applied Microbiology and Biotechnology 87 (2010) 2117–2123.
[42] H. Schagger, G. von Jagow, Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa, Analytical Biochemistry 166 (1987) 368–379.
[43] H. Schagger, Tricine-SDS-PAGE, Nature Protocols 1 (2006) 16–22.
[44] P.K. Smith, R.I. Krohn, G.T. Hermanson, A.K. Mallia, F.H. Gartner, M.D. Provenzano, E.K. Fujimoto, N.M. Goeke, B.J. Olson, D.C. Klenk, Measurement of protein using bicinchoninic acid, Analytical Biochemistry 150 (1985) 76–85.
[45] Z. Chen, H. Cai, F. Lu, L. Du, High-level expression of a synthetic gene encoding a sweet protein, monellin, in *Escherichia coli*, Biotechnology Letters 27 (2005) 1745–1749.
[46] I.H. Kim, K.J. Lim, Large-scale purification of recombinant monellin from yeast, Journal of Fermentation and Bioengineering 82 (1996) 180–182.
[47] K. Kondo, Y. Miura, H. Sone, K. Kobayashi, H. Iijima, High-level expression of a sweet protein, monellin, in the food yeast *Candida utilis*, Nature Biotechnology 15 (1997) 453–457.
[48] L. Penarrubia, R. Kim, J. Giovannoni, S.-H. Kim, R.L. Fischer, Production of the sweet protein monellin in transgenic plants, Nature Biotechnology 10 (1992) 561–564.
[49] K. Roh, K.-S. Shin, Y.-H. Lee, S.-C. Seo, H.-G. Park, H. Daniell, S.-B. Lee, Accumulation of sweet protein monellin is regulated by the psbA 5′UTR in tobacco chloroplasts, Journal of Plant Biology 49 (2006) 34–43.
[50] M. Kohmura, N. Nio, Y. Ariyoshi, Solid-phase synthesis of crystalline [Ser41] B-chain monellin, an analogue of the sweet protein monellin, Agricultural and Biological Chemistry 55 (1991) 1831–1838.
[51] M. Ota, Y. Ariyoshi, Solid-phase synthesis of single-chain monellin, a sweet protein, Bioscience, Biotechnology, and Biochemistry 62 (1998) 2043–2045.