Bacterial fermentation and isotope labelling optimized for amyloidogenic proteins

István Vida,1,2 Zsolt Fazekas,1,2 Gergő Gyulai,2 Dóra Nagy-Fazekas,1,2 Gyula Pálfi,1,4 Pál Strárner,4 Éva Kiss3 and András Perczel1,4

1 Laboratory of Structural Chemistry and Biology, Institute of Chemistry, Eötvös Loránd University, Pázmány P. stny. 1/A, Budapest, H-1117, Hungary.
2 Hevesy György PhD School of Chemistry, Eötvös Loránd University, Pázmány P. stny. 1/A, Budapest, H-1117, Hungary.
3 Laboratory of Interfaces and Nanostructures, Institute of Chemistry, Eötvös Loránd University, Pázmány P. stny. 1/A, Budapest, H-1117, Hungary.
4 MTA-ELTE Protein Modeling Research Group, Eötvös Loránd Research Network (ELKH), Institute of Chemistry, Eötvös Loránd University, Pázmány P. stny. 1/A, Budapest, H-1117, Hungary.

Summary

We developed a cost sensitive isotope labelling procedure using a fed-batch fermentation method and tested its efficiency producing the 15N-, 13C- and 15N/13C-labelled variants of an amyloidogenic miniprotein (E5: EEEAVRLYIQWLKEGGPSSGRPPPS). E5 is a surface active protein, which forms amyloids in solution. Here, we confirm, using both PM-IRRAS and AFM measurements, that the air-water interface triggers structural rearrangement and promotes the amyloid formation of E5, and thus it is a suitable test protein to work out efficient isotope labelling schemes even for such difficult sequences. E. coli cells expressing the recombinant, ubiquitin-fused miniprotein were grown in minimal media containing either unlabelled nutrients, or 15N-NH4Cl and/or 13C-D-Glc. The consumption rates of NH4Cl and D-Glc were quantitatively monitored during fermentation and their ratio was established to be 1:5 (for NH4Cl: D-Glc). One- and two-step feeding schemes were custom-optimized to enhance isotope incorporation expressing five different E5 miniprotein variants. With the currently optimized protocols we could achieve a 1.5- to 5-fold increase of yields of several miniproteins coupled to a similar magnitude of cost reduction as compared to flask labelling protocols.

Introduction

The 5-10 mg scale production of non-selective isotopic labelled polypeptide hormones (e.g. insulin, incretins; Baeshen et al., 2014) and miniproteins (e.g. Trp-cage, cyclotides, PPα-Tyr; Neidigh, 2002; Baker et al., 2017; Gould, 2017) is crucial for structure determination, binding studies, lead optimization etc. Solution or solid-state NMR structure determinations practically the only experimental method to provide atomic-level structural information concerning shorter or midsize proteins requires labelled materials for thorough investigations that reveal the structure and dynamics of such systems. Before the widespread use of recombinant expression methodologies, solid phase peptide synthesis (SPPS) was the only way to produce 15N, 13C labelled peptides but at extreme costs (Merrifield, 1964). The latter method requires N-protected isotopic labelled amino acid residues, largely increasing the production cost, especially for longer peptides or proteins (Strárner et al., 2016). In contrast, recombinant expression is a widely used environmentally friendly (green) and economical technique. Using a suitable strain of E. coli bacteria, high growth rates and fast protein production can be achieved, expressing directly the required gene or its constructs, such as the gene cloned side-by-side with a selected fusion protein. Host cells can be cultured and genetically manipulated in a water-based environment. Isotope labelling in bacteria is also cost-effective, as only two ingredients, NH4Cl and D-Glc, have to be 15N and 13C labelled to achieve uniform 15N-, 13C-incorporation.

The challenge in expressing medium-sized (20–40 residue-long) polypeptides or miniproteins is that their size makes them extremely vulnerable to proteolytic degradation. However, applying suitable fusion tag proteins (glutathione-S-transferase, GST; thioredoxin, TRX; ubiquitin; small ubiquitin-related modifier, SUMO;...
ketosteroid isomerase, KSI; Williamson et al., 2000; Bogomolovas et al., 2009; Panavas et al., 2009; Zorko, 2010) has been identified as a powerful strategy to overcome this difficulty. In this work, we used ubiquitin as a fusion tag (Bommarius et al., 2010). Ubiquitin fusion system has numerous advantages: the ubiquitin tag is relatively small (11 kDa), hence the target peptide has larger contribution to the total mass of the fusion protein and the yeast ubiquitin hydrolase (YUH) enzyme can cleave the fusion protein from the polypeptide without any remaining and unwanted residue at N-terminus (Kohno et al., 1998). One of the major drawbacks of bacterial expression systems is that in certain cases the expression levels can be too low (< 1–2 mg l−1). Eukaryotic proteins without post-translational modifications essential for their fold, toxic peptides, peptides or proteins with aggregation tendency and proteins lacking their complex disulfide bond pattern are typical systems inefficiently synthesized by bacteria. Changing the host cell to a higher organism (yeast, insect or mammalian cell) can be a solution to this problem; however, it extremely increases the isotope labelling costs. Another option is upscaling, however large volume apparatus (e.g. 30-50 l fermenter) are still uncommon in most academic laboratories.

During scale up, the number of cells is augmented, achieved either by a volumetric increase of the cell culture or by concentration enhancement. In an everyday biochemical environment even a moderate scale increase, for example from 1 to 10 l, introduces unexpected challenges to work with. In addition, this strategy results in the vast loss of isotope labelled ingredients. Furthermore, if instead of using a single batch, multiple flasks are in use, then the differences in metabolic rates can result in severe timing issues (e.g. induction and/or harvesting) to be matched. On the other hand, if cell concentration is increased to reach a higher yield by using a fermenter for example, then normal size laboratory equipment as well as improved and tunable biotechnology protocols can be applied (Klopp et al., 2018). However, the close monitoring of the pH, air flux and the concentration of selected metabolites requires a disciplined and automation-oriented technology.

When conditions such as temperature, pH, nutrient concentration, oxygen concentration for aerobic strains, etc., are optimal, bacteria grow at an exponential rate. However, if conditions are inappropriate, growth becomes limited. In nutrient limited growth, the rate limiting molecule is that which the medium runs out of the first, introducing a lag-phase, followed by a slower-than-exponential phase that utilizes an alternative nutrient (diauxic growth). The change caused by the petering of the limiting nutrient is called the diauxic shift (Monod, 1942). During exponential growth, the dissolved oxygen concentration steadily decreases, as the volume of the biomass increases. However, during the lag-phase there is practically no bacterial growth, and thus the oxygen concentration increases, indicating a shortage of a limiting nutrient. Several fermentation protocols to produce recombinant proteins are known in the literature. Working with labelled nutrients, cost-effective batch and fed-batch fermentation techniques are usually applied along the entire process (Studts, 1999; Ross et al., 2004). Cai et al. (1998) used a fed-batch fermentation process for 15N- and 13C/15N-labelling, in which the cells were grown in the presence of unlabelled isotope in the initial phase. In a comprehensive study, several fermentation protocols were implemented with auto- or IPTG-inducing on minimal or rich media (Klopp et al., 2018).

Here we present an economical fermentation protocol to produce isotope labelled polypeptides and miniproteins. During the development of our protocol we used a self-aggregating sequence, E5 (EEEEVLIQWLDGGPSSGRPPP; Rovó et al., 2013) derived from Exendin-4 miniprotein as a test system, which forms amyloid near the physiological conditions (80 μM < cprot. < 800 μM, 10 mM < csalt < 100 mM, 4 < pH < 5, T = 37 °C) after a few hours (Taricska et al., 2020) and thus disturbs and hinders high yield protein expression. We examined the structural rearrangement of E5 at air–water surface and its dependence on storage time. By using both PM-IRRAS and AFM measurements early amyloid formation triggered by the air–water interface could be effectively monitored. We then successfully applied our new protocol for the expression of Exendin-4 itself and its further variants, namely E11, E19, Tc5b-D9N (Rovó et al., 2014; Scheme 1).

Results and discussion

Our fermentation method for non-selective isotope labelling (15N, 13C) is based on the fact that the isotope carrying (15N-NH4Cl and 13C-D-Glc) components are added to the cell culture at the time of the diauxic shift, t1 (Fig. 1A). Prior to this, unlabelled 14N-NH4Cl and 12C-D-Glc are the exclusive nitrogen and carbon sources used in the minimal medium (batch phase, between t0−t1; Fig. 1A). As both the isotope source and the timing were examined thoroughly, the synchronized induction led to optimal use of isotope sources: predominantly to produce the required protein. In practice, cytosolic accumulation of both the ingredients and the metabolites could result in incomplete labelling, therefore it is common to overuse the isotope source, adding them in two consecutive steps (first feeding phase, between t1a−t1b and second feeding phase, between t1b−t2; Fig. 1B; Cai et al., 1998). As the superiority of two-step feeding has never
Scheme 1. The primary sequence of Exendin-4, a drug used in the treatment of type II Diabetes Mellitus and its four variants: E5, E11, E19, Tc5b-D9N.

Fig. 1. Comparison of the two batch-fed strategy. (A) During isotope labelled expression using one-step and (B) two-step feeding scheme, the dissolved oxygen (%) was monitored as function of the time. The relative consumption of D-Glc (green triangle) and NH₄Cl (orange triangle) are depicted schematically above using solid colours for ¹³C- or ¹⁵N-, while striped colours for natural abundant (¹²C- or ¹⁴N-) ingredients. Purple arrow points to the time of induction.

\( t_0 \): beginning of fermentation; \( t_1 \), time of feeding; \( t_{1a}, t_{1b} \), time of the first or second feeding; \( t_2 \), end of the fermentation.

Selected ¹H-NMR signals of backbone amid H(N) of Glu3 of ¹⁵N-labelled E5: using either (C) the one-step (red line) or (D) two-step feeding (blue line) fermentation protocol. The doublet stands for the ¹H-¹⁵N one-bond coupling non-existent in a ¹H-¹⁴N system resulting in a singlet. (E) Monitoring both D-Glc (green dot) and NH₄Cl (orange dot) concentration (g l⁻¹) changes with variation of the dissolved oxygen (%) (blue line) as function of the fermentation time (min).

© 2021 The Authors. Microbial Biotechnology published by John Wiley & Sons Ltd and Society for Applied Microbiology. Microbial Biotechnology, 14, 1107–1119
been unambiguously probed, we have carried out the following two experiments. Using a one-step feeding protocol, the time of induction and isotope addition was synchronized to the oxygen peak (Fig. 1A). During the two-step feeding approach, one quarter of the total amount of isotope is added in synchrony with the initial oxygen peak, followed by induction and addition of the remaining three-quarters when the second oxygen peak is detected (Fig. 1B). The extent of isotope incorporation achieved with each protocol was examined by NMR spectroscopy choosing one well separated signal. ¹H-NMR spectrum of ¹⁵N-E5 between 8.8 ppm and 9.1 ppm revealed a doublet, with ¹JHN-coupling of ~93.0 Hz with a singlet at the centre (8.96 ppm), belonging to the unlabelled derivative. Based on the ratio of integrals of the doublet and the singlet signals, we determined the ratio of the concentration of the labelled and unlabelled protein. This ratio was found to be 96% if the two-step feeding protocol was applied, suggesting a more complete isotope incorporation (Fig. 1D) than in case of the one-step feeding approach, when it was 91%. This might be due to presence of greater excess of cytosolic unlabelled metabolites in case of the latter. Based on these findings, we selected the two-step feeding protocol for further use.

It is important to distinguish whether the nitrogen or carbon source is exhausted at the time of the diauxic shift. Bacteria consume NH₄Cl and D-Glc at a different rate, however the ratio of the consumption rates remains constant over time. Therefore, it is possible to determine \textit{a priori} whether NH₄Cl or D-Glc will be the limiting nutrient, if the characteristic ratio of consumption rates is established. Thus, the concentration changes of the carbon and nitrogen source were measured as function of the time using initially 10 g of D-Glc and 1 g of NH₄Cl (Fig. 1E). The concentration of NH₄Cl was monitored by absorption spectroscopy detecting the indophenol derivative. Based on the ratio of integrals of the singlet at the centre (8.96 ppm), belonging to the unlabelled derivative, we determined the ratio of the labelled and unlabelled protein. This ratio was found to be 96% if the two-step feeding protocol was applied, suggesting a more complete isotope incorporation (Fig. 1D) than in case of the one-step feeding protocol (Fig. 1C) when it was 91%. This might be due to presence of greater excess of cytosolic unlabelled metabolites in case of the latter. Based on these findings, we selected the two-step feeding protocol for further use.

### Table 1. Summary of the initial and additional amount of D-Glc and NH₄Cl added to the samples making either ¹⁵N or ¹³C as well as ¹⁵N/¹³C labelled miniproteins.

|                      | Batch phase | First feeding phase | Second feeding phase |
|----------------------|-------------|---------------------|----------------------|
|                      | t₀ < t(min) < t₁a | t₁a < t(min) < t₁b | t₁b < t(min) < t₂   |
| N (g)                 | C (g)       | N (g)               | C (g)                |
| U                    | L           | U                   | L                   |
| ¹⁵N-miniprotein       |             | 1                   | 0                   |
| ¹³C-miniprotein       |             | 1                   | 0                   |
| ¹⁵N/¹³C-miniprotein   |             | 0                   | 1                   |

|                      | N          | C          |
|----------------------|------------|------------|
| ¹⁵N-NH₄Cl            | 10         | 0          |
| ¹³C-D-Glc            | 3          | 0          |
| ¹⁵N/¹³C-D-Glc        | 3          | 0          |

a. \( t₀, t₁a, t₁b, t₂ \) are the same codification as in Fig. 1A and B.

b. N = mass of NH₄Cl; C = mass of D-Glc.
c. U = natural abundance isotopic source.
d. L = isotopically enriched (¹³C and/or ¹⁵N) source.

© 2021 The Authors. *Microbial Biotechnology* published by John Wiley & Sons Ltd and Society for Applied Microbiology, *Microbial Biotechnology*, 14, 1107–1119.
the final yields. Following our first attempts, cell pellets were stored at −20 °C after fermentation. The cells which remained in frozen state at −20 °C for an extended time, did not provide reproducible yields, most likely due to aggregation processes leading to fibril formation to various extents. This is probably caused by the conformational heterogeneity of the systems that is sustained even at −20 °C (Yang et al., 2015). The formation of β amyloid fibril structure requires conformational change from the globular to elongated forms and the formation of numerous intermolecular contacts between the latter. If amyloid formation is triggered during the expression, the overall yield will decrease due to the precipitation of the fibrils.

Our isotope labelling (15N, 13C) protocol was developed using E5 known to be amyloidogenic sequence, Table 2. Yield and cost comparison of the flask expression and fermentation of Exendin-4 and four variants, namely Tc5b-D9N, E5, E11 and E19. Ni-IMAC chromatograms of raw materials were used to estimate the yields and the purity was checked by analytical HPLC (an example for E5 is shown in Supplementary material S2).

| Protein name and labelling type | Calculated yield (mg l⁻¹) | Estimated cost (EUR/10 mg) |
|--------------------------------|---------------------------|---------------------------|
|                                | Ferm. | Flask | Ferm. | Flask |
| E11, 15N                       | 64.8  | 31.2  | 18.7  | 38.8  |
| E19, 15N                       | 150.2 | 33.5  | 8.1   | 36.2  |
| E5, 15N                        | 99.6  | 19.4  | 12.2  | 62.4  |
| Exendin-4, 15N                 | 87.0  | 31.4  | 13.9  | 38.6  |
| Tc5b D9N, 15N                  | 19.1  | 12.1  | 63.5  | 100.1 |

a. Raw material yield based on the integral volume of the absorbance curve (see methods), determined after completing the first Ni-IMAC step.

![Fig. 2](image_url)

Fig. 2. (A) Time dependence of the surface tension (γ) of ovalbumin (OVA, red dots), E5 (black rectangles) and bovine serum albumin (BSA, blue triangles) at solution/air surface, OVA and BSA are included as a reference. (B) Representative PM-IRRAS spectrum of E5 recorded at the air–water interface at 15 °C. (C) Secondary structure composition (purple rectangles: β-sheets; red circles: α-helices and random coil structures; blue triangles: β-turn motives) of E5 interfacial films at 15 °C as a function of film age.

© 2021 The Authors. Microbial Biotechnology published by John Wiley & Sons Ltd and Society for Applied Microbiology.
which forms amyloid near physiological conditions after a few hours in solution (Yang et al., 2015). We followed the process by PM-IRRAS and AFM experiments. We established that E5 suffers a structural rearrangement at air–water surface causing reduced yields during protein expression and purification.

The PM-IRRAS technology allows the measurement of surface specific FTIR spectra of materials relying on the differences in the reflection of p- and s-polarized light from interfaces. This gives information of the chemical composition, phase transition, structure and orientation of the molecules in the examined sample. Polarization modulation results in the elimination of the background signals, such as water vapour and CO₂, thus it is not necessary to use protective gases or vacuum during the measurements. E5 is a highly surface active molecule, it accumulates promptly in air–water interfaces (Fig. 2A), similarly to other well-known globular proteins like bovine serum albumin (BSA) and ovalbumin (OVA). Thus, PM-IRRAS measurements were carried out at the air–water surface of spread E5 films (Fig. 2B). Multiple peaks could be identified in the amide I region indicating the presence of the protein molecules at the surface and allowing for the determination of the protein secondary structure (Shanmukh et al., 2005; Yang et al., 2015). The most notable feature is the presence of prominent peaks around 1630 cm⁻¹ that can be assigned to β-sheet structures. Since in bulk phase no β-structure is detectable (Taricska et al., 2019), the appearance of β-sheet and β-turn structures and the decreased presence of random coil and α-structures points to surface induced structural changes.

To characterize the time dependence of this structural rearrangement, measurements were carried out at 15 °C
Isotope labelling for amyloidogenic proteins

Protein expression in flask -> Ni IMAC

Protein expression in fermenter -> Ni IMAC

Further purifications

Raw material

Pure material

Analytic (isotope incorporation)

Mass spectrometry

1D $^{3}$H-NMR spectroscopy
subphase temperature as a function of film age. Already in the first measurement, started within 1 min after film deposition corresponding to a film age between 1 and 6 min, an excess in β-structures can be seen (Fig. 2C). Over time a further time dependent reorganization could be detected with an enrichment of β-structures. Due to

![Diagram](image-url)

Fig. 5. (A) 1H-NMR spectrum of 15N-E5 expressed in fermenter. (B) The zoom of the 1H-NMR spectrum shows two different signal groups in the H(N) region: one is a typical resonance of backbone peptide H(N) (right) with a doublet with J-coupling of 93.0 Hz (assigned to Glu), while the other belongs to the Trp indole side chain H(N) with a J-coupling of 99.0 Hz (left). (C) The zoom of the homonuclear 1H, 1H NOESY spectrum proves that the three signals of Trp side chain and the two signals of the backbone belong to the same H(N) due to the same crosspeak pattern.

© 2021 The Authors. Microbial Biotechnology published by John Wiley & Sons Ltd and Society for Applied Microbiology. Microbial Biotechnology, 14, 1107–1119
of 93.0 Hz (assigned to Glu), while the other belongs to the Trp indole side chain H(N) with a \(^1\)J-coupling of 99.0 Hz (Fig. 5A and B). The latter group is a doublet with a singlet in the centre, indicating only partial isotope labelling of Trp side chain. Similar effect was seen in the \(^1\)H,\(^1\)H-NOESY spectrum (Fig. 5C). In contrast, the signals belonging to backbone H(N)s clearly demonstrate a high isotope labelling ratio – in this region the intensity of the unlabelled signals in the centre are very low. The phenomenon is probably caused by the improper depletion of unlabelled indole, the metabolic precursor of Trp, which is present in increased amounts in the cytosol during the diauxic shift (Gaimster, 2015).

**Conclusion**

Lab-scale fermenters ensure greater control of bacterial culture processes and provide highly economical and scaled-up procedures to prepare isotope labelled miniproteins and peptides as compared to flask applications. Here we developed three protocols to produce \(^1\)C-, \(^1\)N- and \(^1\)C/\(^1\)N-labelled miniproteins. Our strategy consists of a batch phase, when cells grow in media containing mostly unlabelled isotopes, and a feeding phase, where the labelled isotopes were added in a two-step manner. We found that two-step feeding protocol provides greater extent of isotope incorporation than one-step feeding, and thus, could reach a 1.5- to 5-fold increase of yields and a similar magnitude of cost reduction compared to flask labelling methods. Our finding about improper isotope incorporation of the NH of the Trp indole ring suggests, that in some cases indole resonances might ‘underestimate’ the overall isotope incorporation efficacy.

The protocols are summarized here:

1. To produce \(^1\)C-labelled proteins, cells are grown in minimal media containing 1 g of unlabelled \(^1\)N-NH\(_4\)Cl and 3 g of unlabelled \(^1\)C-D-Glc. At the first oxygen peak 0.75 g of \(^1\)C-D-Glc is added, and at the second oxygen peak further 2.25 g of \(^1\)C-D-Glc is added to the media.

2. For making \(^1\)N-labelled proteins, cells are grown in minimal media containing 1 g of unlabelled \(^1\)N-NH\(_4\)Cl and 10 g of unlabelled \(^1\)C-D-Glc during the batch phase. At the first, oxygen peak 0.25 g of \(^1\)N-NH\(_4\)Cl is added, and at the second oxygen peak further 0.75 g of \(^1\)N-NH\(_4\)Cl is added to the media.

3. To produce \(^1\)C/\(^1\)N-labelled proteins, cells grow in minimal media containing 1 g of \(^1\)N-NH\(_4\)Cl and 3 g of unlabelled \(^1\)C-D-Glc. At the first, oxygen peak 0.75 g of \(^1\)C-D-Glc is added, and at the second oxygen peak further 2.25 g of \(^1\)C-D-Glc is added to the media.
The protocols described here are straightforward, cost effective and can be adapted to produce any recombinant protein, especially for those miniproteins which are expressed with low yields due to their tendency of aggregation and amyloid formation. This was demonstrated for the case of E5, a miniprotein that is surface active and suffers an immediate structural rearrangement at air–water surface. We showed that the PM-IRRAS/AFM methodologies can provide detailed molecular information on aggregate formation. We propose that in case of E5 the process is initiated by the formation of metastable pre-aggregates of the globular form of the protein. Reaching the air–water interface triggers the restructuring and rearrangement of these pre-organized, high concentration clusters leading to β-amylloid formation. We concluded that extended storage of such sensitive samples in solution or cell pellet form leads to extensive material loss and thus should be avoided.

**Experimental procedures**

**Cultivation**

Fermentations were carried out in BIOSTAT A 2 L benchtop bioreactor (Sartorius). The inoculation procedure and fermentation protocols are presented in detail in the Supplementary material. Culturing was typically completed at 37°C, pH 7.0, 600 r.p.m. stirring, 1500 cm\(^{-2}\) min\(^{-1}\) aeration. Expression was induced with 1 mM IPTG. Following fermentation, cells were pelleted by centrifugation, followed by suspension in a Lysis buffer (50 mM NaPi, 300 mM NaCl, pH 8.0) and stored at −20°C. Flask expression controls were performed in 2YT media. For \(^{15}\)N labelling, 1 g of \(^{15}\)N-NH\(_4\)Cl and 4 g of unlabelled \(^{12}\)C-D-Glc was added to the media. For \(^{13}\)C/\(^{15}\)N labelling, the media were supplied with 1 g of \(^{15}\)N-NH\(_4\)Cl and 2 g of \(^{13}\)C-D-Glc.

**Purification of the recombinant miniproteins**

After thawing the suspension, cells were disrupted by sonication and cell debris was removed by centrifugation. The supernatant containing the protein of interest was purified with one of the two following methods depending on further uses. (i) If prompt information about protein yield was required, one-step Ni immobilized metal ion affinity chromatography (Ni-IMAC) was applied. The supernatant was loaded onto a HisTrap HP 5 ml column, and after washing with Wash buffer (Lysis buffer containing 20 mM imidazole), the protein of interest was eluted with Elution buffer (Lysis buffer containing 20 mM imidazole), the protein of interest was eluted with Elution buffer (Lysis buffer containing 20 mM imidazole). The resulting PM-IRRAS signal can be given by the following equation:

\[
S = \frac{\Delta R}{R} = \frac{(R_p - R_s)}{(R_p + R_s)}
\]

where \(R_p\) and \(R_s\) are the parallel (p) and perpendicular (s) polarized reflectances (Blaudez et al., 1993).
Normalized spectra a given as
\[
\Delta S = \frac{S - S_0}{S_0},
\]
where \(S_0\) is the signal measured on pure water surface at respective temperatures, while \(S\) is the signal from the protein covered surface. For each spectrum, an acquisition time of 5 min was used to record 3000 spectra in the 800–4000 cm\(^{-1}\). Spectra showed good reproducibility on separate films and the presented spectra are the average of at least 5 recordings. The normalized spectra were baseline corrected and the amide I band in the 1600–1700 cm\(^{-1}\) range was subjected to peak fitting using second-derivative peak identification and Gaussian curve-fitting with the FTYK software.

Atomic force microscopy (AFM)
The interfacial films of the E5 miniprotein was also studied applying atomic force microscopy. Following the PM-IRRAS measurements, the films were transferred onto a solid support using the Langmuir–Schaeffer method. Freshly cleaved mica was used as the flat substrate. Transferred films were dried in vacuum before imaging in air at room temperature. Sample morphology was studied with a Nanosurf FlexAFM atomic force microscope, operating in dynamic mode. Tap190GD-G cantilever (BudgetSensors) with a nominal tip radius of less than 10 nm was used for the measurements. Images were recorded over \(2 \times 2 \mu m^2\) window areas at 10 randomly selected locations with a resolution of 512 pixels/line.

Surface tension measurement
The surface tension of aqueous protein solutions \((c = 0.1 \text{ g l}^{-1})\) was determined with an accuracy of 0.1 mN m\(^{-1}\) by the axisymmetric drop shape analysis using the OCA15 + instrument (Dataphysics, Filderstadt, Germany). Apart from the E5 miniprotein, the surface tension of bovine serum albumin (BSA) and ovalbumin (OVA) solutions were also measured as a reference. The profile of capillary surface required to determine the surface/interfacial tension is obtained by analysing the shape of the pendant drop using a CCD camera coupled to a video image profile digitizer board connected to a personal computer (Hill et al., 2008). Drop of 8 µl was formed and the surface tension was recorded as a function of the time for 20 min with a frequency of 10 frames per min.

Measurement of NH\(_4\)Cl concentration
The NH\(_4\)Cl concentration was measured by using the Spectroquant Ammonium Test kit (Merck, Darmstadt, Germany), using an eight-point calibration curve in the 0.0–140.0 mg l\(^{-1}\) range. This assay is based on the conversion of the NH\(_4^+\) to an indophenol derivative via the Berthelot reaction and the measurement of the absorption of the indophenol derivative (Searle, 1984).

Measurement of D-Glc concentration
The D-Glc concentration was measured using two different methods. At higher D-Glc concentrations, a commercial blood Glc meter (e.g. Méry Plusz TD-4255) was employed. The instrument was five-point calibrated using standard D-Glc solutions in the 2.0–10.0 g l\(^{-1}\) range. Lower D-Glc concentrations \((2.0 > c > 0.1 \text{ g l}^{-1})\) were measured using a D-Glc derivative, quantified by RP-HPLC as follows (Iqbal, 2009). Cells were removed by centrifugation and the diluted sample was filtered through an Amicon Ultra Ultracel 3K 0.5 ml ultrafilter, hence eliminating most macromolecular content. 200 µl DNPH solution \((10 \text{ mM } 2,4\text{-dinitrophenyl hydrazine, } 1\text{ V/V}\% \text{ H}_2\text{SO}_4, 4\text{ V/V}\% \text{ H}_2\text{O, } 95\text{ V/V}\% \text{EtOH})\) was added to 200 µl filtrate and the mixture was incubated at 65°C for 1 h. DNPH and D-Glc reacted in a quantitative manner and resulted in the UV active 2,4-dinitrophenyl hydrazone derivative. After derivatization, the sample was separated by RP-HPLC on a Phenomenex Gemini C18 5 µm column using the same water/acetonitrile gradient as mentioned above. The DNPH derivative’s peak area \((I_0 \approx 12 \text{ min})\) was calculated and compared to D-Glc concentrations: correlation was determined using a five-point calibration standard of D-Glc from the 0.1–0.4 g l\(^{-1}\) concentration range.

NMR spectroscopy
\(^{15}\)N-labelled miniproteins were dissolved in water \((c \sim 1 \text{ mM})\), and 10% D\(_2\)O with 1% DSS standard was added and pH was adjusted to 7.0. The NMR measurements were carried out at \(T = 25^\circ C\) on a Bruker Avance III 700 MHz spectrometer equipped with a 5 mm Prodigy TCI H&F–C/ND, z-gradient probehead operating at 700.05 MHz for \(^1\)H nucleus. Temperature calibration was performed using a standard methanol solution. All chemical shifts were referenced taking account the internal \(^1\)H-resonance of DSS standard. \(^1\)H-NMR and \(^1\)H,\(^1\)H-NOESY (mixing time: 150 ms) spectra were collected. All spectra were processed with the Bruker Topspin software.

Acknowledgements
NMR spectrometer measurement time (700 MHz Bruker) was courtesy of MedInProt Grant Facilitating Access to Instruments from the Hungarian Academy of Sciences. This research project was supported by the European Union and the State of Hungary and financed by the.
European Regional Development Fund (VEKOP-2.3.2-16-2017-00014) grant of the NKFIH of the Hungarian Academy of Sciences. The authors gratefully acknowledge Tünde Tóth and Dániel Fülöp for surface tension and PM-IRRAS measurements and Dóra Karancsiné Menyhárd for the fruitful discussions regarding the wording. This work was completed in the ELTE Thematic Excellence Programme supported by the Hungarian Ministry for Innovation and Technology. Project no. 2018-1.2.1-NKP-2018-00005 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the 2018-1.2.1-NKP funding scheme.

Conflict of interest

None declared.

References

Baeshen, N.A., Baeshen, M.N., Sheikh, A., Bora, R.S., Ahmed, M.M.M., Ramadan, H.A.I., et al. (2014) Cell factories for insulin production. Microbiol Cell Factories 13: 141.

Baker, E.G., Bartlett, G.J., Porter Goff, K.L., and Woolfson, I. Vida et al.

Klopp, J., Winterhalter, A., Gébeux, R., Scherer-Becker, D., Ostermeier, C., and Gossert, A.D. (2018) Cost-effective large-scale expression of proteins for NMR studies. J Biomol NMR 71: 247–262.

Kohn, T., Kusunoki, H., Sato, K., and Wakamatsu, K. (1998) A new general method for the biosynthesis of stable isotope-enriched peptides using a decahistidine-tagged ubiquitin fusion system: An application to the production of mastoparan-X uniformly enriched with 15N and 15N/13C. J Biomol NMR 12: 109–121.

Merrifield, R.B. (1964) Solid-phase peptide synthesis. Ill. An improved synthesis of Bradykinin. Biochemistry 3: 1385–1390.

Monod, J. (1942) Recherches sur la croissance des cultures bactériennes. Paris, France: Hermann & cie.

Neidigh, J.W., and Andersen, N.H. (2002) Peptide conformational changes induced by tryptophan-phosphocholine interactions in a micelle. Biopolymers 65: 354–361.

Panavas, T., Sanders, C., and Butt, T.R. (2009) SUMO fusion technology for enhanced protein production in prokaryotic and eukaryotic expression systems. In SUMO Protocols. Methods in Molecular Biology. Ulrich, H.D. (ed). Totowa, NJ: Humana Press, pp. 303–317.

Ross, A., Kessler, W., Krumme, D., Menge, U., Wissing, J., van den Heuvel, J., and Flohé, L. (2004) Optimised fermentation strategy for 13C/15N recombinant protein labelling in Escherichia coli for NMR-structure analysis. J Biotechnol 108: 31–39.

Rovó, P., Farkas, V., Stráner, P., Szabó, M., Jermendy, Á., Hegyi, O., et al. (2014) Rational design of α-Helix-stabilized Exendin-4 analogues. Biochemistry 53: 3540–3552.

Rovó, P., Stráner, P., Láng, A., Bartha, I., Huszár, K., Nyitray, L., and Perczel, Á. (2013) Structural insights into the Trp-Cage folding intermediate formation. Chemistry 19: 2628–2640.

Searle, P.L. (1984) The berthelot or indophenol reaction and its use in the analytical chemistry of nitrogen. A review. Analyst 109: 549–568.

Shanmukh, S., Biswas, N., Waring, A.J., Walther, F.J., Wang, Z., Chang, Y., et al. (2005) Structure and properties of phospholipid–peptide monolayers containing monomeric SP-B1–25: II. Peptide conformation by infrared spectroscopy. Biophys Chem 113: 233–244.

Stráner, P., Taricska, N., Szabó, M., Tóth, K.G., and Perczel, A. (2016) Bacterial expression and/or solid phase peptide synthesis of 20–40 amino acid long polypeptides and miniproteins, the case study of Class B GPCR ligands. Curr Protein Pept Sci 17: 147–155.

Studs, J.M., and Fox, B.G. (1999) Application of fed-batch fermentation to the preparation of isotopically labeled or selenomethionyl-labeled proteins. Protein Expr Purif 16: 109–119.

Taricska, N., Bokor, M., Menyhárd, D.K., Tompa, K., and Perczel, A. (2019) Hydration shell differentiates folded and disordered states of a Trp-cage miniprotein, allowing characterization of structural heterogeneity by wide-line NMR measurements. Sci Rep 9: 2947.

© 2021 The Authors. Microbial Biotechnology published by John Wiley & Sons Ltd and Society for Applied Microbiology. Microbial Biotechnology, 14, 1107–1119
the amyloid state: exploring the potential energy surface of a drug-like miniprotein. *Chemistry* **26**: 1968–1978.
Williamson, P.T.F., Roth, J.F., Haddingham, T., and Watts, A. (2000) Expression and purification of recombinant neurotensin in *Escherichia coli*. *Protein Expr Purif* **19**: 271–275.
Yang, H., Yang, S., Kong, J., Dong, A., and Yu, S. (2015) Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. *Nat Protoc* **10**: 382–396.
Zorko, M., and Jerala, R. (2010) Production of recombinant antimicrobial peptides in bacteria. In *Antimicrobial Peptides. Methods in Molecular Biology*. Giuliani, A., and Rinaldi, A.C. (eds). Totowa, NJ: Humana Press, pp. 61–76.

**Supporting information**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Supplementary material S1.** Fermentation protocols.

**Supplementary material S2.** chromatogram and HPLC profile of $^{15}N$-E5 miniprotein.