International Journal of Mass Spectrometry 302 (2011) 149–156

Contents lists available at ScienceDirect

International Journal of Mass Spectrometry

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

EX1 hydrogen–deuterium exchange in an all-helical protein and its cyclized derivative at neutral pH

Thitima Urathamakul[a], Neal K. Williams[b], Nicholas E. Dixon[a,b], Jennifer L. Beck[a,*]

[a] School of Chemistry, University of Wollongong, NSW 2522, Australia
[b] Research School of Chemistry, Australian National University, Canberra ACT 0200, Australia

ARTICLE INFO

Article history:
Received 31 May 2010
Received in revised form 5 September 2010
Accepted 9 September 2010
Available online 17 September 2010

Keywords:
Hydrogen–deuterium exchange
EX1 limit
Electrospray ionization mass spectrometry
ESI-MS

ABSTRACT

The all-helical protein, DnaBN, exhibited EX1 type hydrogen exchange at pH 7.2. Approximately 45 protons were exchanged relatively rapidly, while an additional ~50 protons exchanged more slowly. The rates of exchange for these slowly exchanging protons were the same, demonstrating that the slowest exchange events represent global unfolding. EX1 behavior is uncommon for native proteins. The protein was cyclized by joining the N- and C-termini through peptide linkers that were three, four, five or nine amino acids long. The corresponding "linear" proteins were extended by addition of the same amino acids to give proteins of identical amino acid composition as their cyclized versions but differing by the mass of a water molecule. All of the proteins unfolded approximately five times faster in 10 mM compared with 100 mM ammonium acetate. In all cases, the cyclized proteins showed slower rates of amide proton exchange related to global unfolding than their linear counterparts by a factor of approximately 7- to 12-fold. Interestingly, the rate of exchange for the slowly exchanging protons decreased for both the linear and cyclized proteins as linker length increased, and this correlated with predictions that the C-terminal helix of the protein would be extended by addition of these extra amino acids. This indicates that lengthening of this helix leads to a modest increase in stability of DnaBN.

1. Introduction

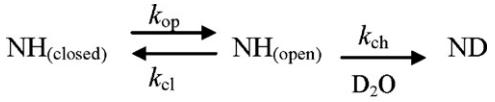
The familiar protein structures determined from NMR and X-ray crystallographic data provide images of proteins with elegant secondary structures (alpha helices and beta strands), folded into tertiary structures that are arranged through hydrophilic and hydrophobic interactions of the amino acid side chains and the protein backbone. This is an important but static view. It has become clear that proteins are dynamic and that during the course of their functions (*e.g.*, enzyme catalysis, transport and release of binding partners) there must be rearrangements including repositioning of water, salts and atoms of the protein itself that are involved in non-covalent bonds [1,2]. Furthermore, it is often not possible to obtain structural information about highly flexible regions of proteins that may have important roles in protein–protein interactions and molecular recognition in general [3]. That dynamic changes in protein structure are possible is a reflection of the ability of proteins in their native states to fold and unfold *via* parallel pathways and sample a range of different structural fluctuations from "breathing" of a few amino acid residues to substantial conformational changes and global unfolding [4].

Hydrogen–deuterium exchange (HDX) experiments have become increasingly applied to investigate the conformations and dynamics of proteins [5,6]. The exchangeable protons are the amide protons on the protein backbone and those attached to nitrogen, oxygen and sulfur atoms of the side chains and the amino- and carboxyl-termini. Comparisons of the extent of HDX provide information about the relative exposure to solvent and the involvement of protons in H-bonding. The amide backbone protons generally exchange more slowly than the side chain protons and are therefore the subject of most experimental investigations. In the first HDX experiments, the exchange was measured using tritium, but the development of NMR enabled determinations of the locations of exchanged protons to be made [7]. Since the development of matrix-assisted desorption ionization- and electrospray ionization mass spectrometry (MALDI- and ESI-MS), analysis of intact peptides and proteins using mass spectrometry has become routine and has enabled facile measurement of HDX [8,9]. The use of ESI-MS to measure HDX has some advantages over NMR, though it yields global rather than local information. These include that smaller amounts of protein can be used and that distinct protein populations representing different folded states can be readily observed. The ability to observe distinct populations

Abbreviations: HDX, hydrogen–deuterium exchange; NH_4OAc, ammonium acetate; ESI, electrospray ionization; NMR, nuclear magnetic resonance (spectrometry).

* Corresponding author. Tel.: +61 2 4221 4177; fax: +61 2 4221 4287.
E-mail address: jbeck@uow.edu.au (J.L. Beck).

1387-3806/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2010.09.011



$$k_{obs} = \frac{k_{op} \cdot k_{ch}}{k_{cl} + k_{ch}}$$

Scheme 1.

is important in understanding the kinetics of protein folding and unfolding.

In HDX experiments, a protein in buffer in 100% H_2O is treated with D_2O (or *vice versa*). This enables exchange to occur as described in Scheme 1 [10]:

$$NH_{(closed)} \underset{k_{c1}}{\overset{k_{op}}{\rightleftarrows}} NH_{(open)} \xrightarrow[D_2O]{k_{ch}} ND$$
$$k_{obs} = \frac{k_{op}.k_{ch}}{k_{c1} + k_{ch}} \quad (1)$$

In this Scheme, NH represents amide groups involved in peptide bonds, ND represents the amide groups that have undergone exchange, k_{op} is the rate constant for protein opening/unfolding (allowing exchange to occur), k_{cl} is the rate constant for protein closing and k_{ch} is the intrinsic rate of chemical exchange. It is only from the "open" protein that exchange can occur. This sets up two limiting conditions [11,12]. In the first exchange mechanism where $k_{cl} \gg k_{ch}$, there is a rapid opening event allowing just a few protons to exchange. The probability of exchange at each event is low. Each time such events occur, a few more protons exchange such that there is a gradual increase in deuterium levels in the protein. When ESI-MS is used to follow exchange this is demonstrated by a gradual increase in mass of the protein. Under these conditions, the overall rate of exchange, k_{obs}, equals $(k_{op} \cdot k_{ch})/k_{cl}$. Since k_{ch} can be calculated [13], the equilibrium constant for unfolding (k_{op}/k_{cl}) can be determined. This is called EX2 exchange.

EX1 exchange occurs when $k_{ch} \gg k_{cl}$. Under these conditions whenever a section of the protein is open, exchange will occur. Under EX1 conditions, mass spectrometry will reveal distinct populations of exchanged protein: one in which amide protons from a section of the protein that has opened have all exchanged and another population in which all of the labile amide protons have exchanged. This is a pseudo-first order process in which $k_{obs} = k_{op}$ [5,14]. Proteins under native conditions usually undergo EX2 exchange while denaturing conditions often favor EX1 exchange. In some experiments, the situation is more complex and it has been suggested that both mechanisms are in play. For example, Miranker et al. [15] initiated exchange in deuterated lysozyme by dilution into protonated buffer at pH 3.8 and 69 °C. While an EX2 mechanism was predominant, comparison of the peak shapes from ESI mass spectra with a simulation that assumed that only an EX2 mechanism showed that there were minor populations of protein molecules with higher or lower masses. These researchers suggested that this meant that there might be additional contributions to exchange from EX1 or other mechanisms. While EX1 exchange is uncommon for native proteins there have been a number of studies that report this mechanism [for example, [16,17]].

Protein stability (maintenance of native structure) is an important consideration in the design of NMR experiments and in various industrial and pharmaceutical applications of proteins and enzymes. For example, some NMR experiments may require acquisitions to occur over many hours and at elevated temperatures. Enzymes used in industrial applications may need to function at high temperature or high or low pH [18], and proteins or peptides for use as drugs can be produced by grafting bioactive moieties onto stabilized peptide backbones [19]. One approach for producing stable proteins is to cyclize them by one of several methods. Backbone cyclization involves joining of their N- and C-termini by formation of a peptide bond, often *via* a flexible peptide linker. Cyclization can be made to occur *in vivo* in a process called protein splicing in which peptide sequences (inteins) are removed from the protein by reactions of particular amino acids that occur at the N- and C-terminal flanks of the intein sequence [20]. Previously, we cyclized the small protein, DnaBN ($M_r \sim 12{,}500$ Da), *in vivo* (over-expressed in *E. coli*) by the use of a split mini-intein from *Synechocystis* sp. PCC6803 (*Ssp*) DnaB [21]. The termini of the protein were connected by a nine amino acid linker, TRESGSIEF, and is called here 9-cz-DnaBN. To compare the stability of the cyclized protein with its non-cyclized (linear) counterpart, native DnaBN was modified so that the four amino acids, SIEF, were attached to its N-terminus and the five amino acids, TRESG, were attached to the C-terminus (9-lin-DnaBN) [22]. The difference in mass between 9-cz- and 9-lin-DnaBN is therefore the mass of a water molecule.

For applications where the native protein structure and function need to be retained when a protein is stabilized by cyclization, it is important that no strain is introduced into the molecule and this requires that the N- and C-termini are close in space. DnaBN satisfies this requirement. It is the N-terminal domain (residues 24–136) of the protein, DnaB. DnaB is the DNA helicase from *E. coli* that unwinds DNA ahead of the advancing replication fork during chromosomal replication. Its N-terminal domain, DnaBN, is important in protein–protein interactions during replication especially in interactions with primase, the RNA polymerase that synthesizes primers for replication [23]. Essentially identical structures of *E. coli* DnaBN have been determined by NMR [24] and X-ray crystallography [25]. It is an all-helical single domain protein (6 helices); the N- and C-termini protrude from the core. Furthermore, the NMR spectra of 9-lin-DnaBN and 9-cz-DnaBN were compared in our previous work and showed that cyclization did not substantially alter the structure of DnaBN, except for a three residue extension of the C-terminal helix to Ser136 and a minor reorientation of the solvent-exposed side chain of Tyr60 [21,24]. Therefore 9-lin- and 9-cz-DnaBN are suitable as models for comparing the thermodynamic effects of protein cyclization.

In our previous work, circular dichroism spectroscopy was used to follow denaturation of 9-lin- and 9-cz-DnaBN by urea. These experiments showed that the decrease in free energy achieved by cyclization ($\Delta\Delta G$) was 2.3 kcal/mol [22]. A significant contribution to this is the diminished conformational entropy that is expected for the cyclized protein in the unfolded state. A significant advantage of NMR is that the rates of exchange for individual protons can be assessed. In both the linear and cyclized proteins, the H-bonded amide protons of the helices showed the slowest exchange (first-order rate constants). Remarkably, the exchange rates (k_{obs}) of these most slowly exchanging protons for 9-cz-DnaBN were *all* about 0.010 min^{-1} slower than for the corresponding protons of 9-lin-DnaBN ($\sim$0.003 min^{-1} *cf.* $\sim$0.013 min^{-1}) and the rates for these protons in each of the linear and cyclized proteins were very similar, independent of amino acid residue type. This suggested that the slowest exchange rates were governed by global unfolding and suggested the possibility that EX1 exchange was occurring.

When ESI-MS was used to follow HDX, two distinct populations were observed, again supportive of an EX1 mechanism. Finally, ESI-MS was used to compare the first-order rate constants for exchange over the pH range of 6.8–7.8. At each pH, similar observations consistent with EX1 exchange were made. Since the chemistry of exchange (k_{ch}) is primarily base-catalyzed, if EX2 exchange is occurring it is expected that k_{obs} will increase by an order of magnitude for every pH unit. This was not observed. At pH 6.8, 7.2 and 7.8,

k_{obs} for 9-lin-DnaBN was 0.161, 0.085 and 0.056, respectively, and for 9-cz-DnaBN it was 0.0115, 0.0095 and 0.0044 min^{-1}, respectively. For both proteins, this represents a three-fold decrease with a ten-fold increase in concentration of OH^- [22]. This is again supportive of EX1 exchange; the increases in rate at lower pH may reflect an overall lower stability of the protein as pH is decreased, probably because of protonation of the buried His64 sidechain.

Taken together, the observations from the previous work strongly support an EX1 amide proton exchange mechanism for 9-lin- and 9-cz-DnaBN at near neutral pH. The ratio of the rates of exchange of the slowly exchanging protons in the linear and cyclized proteins ($k_{obs(lin)}/k_{obs(cz)}$) was ~10, supporting that cyclization stabilizes the protein against unfolding. The slower rate of unfolding for the cyclized protein supported that unfolding involves some separation of the N- and C-termini. While the ratio was the same whether the exchange rate was measured by NMR (in 10 mM phosphate buffer) or ESI-MS (in 10 mM ammonium acetate), the absolute values of the exchange rates were about 10-fold faster when measured by ESI-MS. It was suggested that this was because the protein is sensitive to different salts and salt concentrations [22].

To investigate in more detail the effect of cyclization on protein stability and to confirm the previous observation of EX1 exchange for 9-lin- and 9-cz-DnaBN, new protein constructs were prepared in which linkers used to cyclize the protein were only 3, 4 and 5 amino acids long. Linear versions of the proteins with the same amino acid compositions with extensions at the N- and C-terminal ends were prepared for comparison. It was hypothesized that shorter linkers would increase restriction of movement of the N- and C-termini and decrease rates of unfolding (decrease exchange rates of slowly exchanging protons). The results were consistent with EX1 exchange in all cases, but the exchange rates observed for the different linker lengths did not bear out the hypothesis. The rates of amide exchange were compared in 10 and 100 mM ammonium acetate, pH 7.2 at 10 °C, confirming that the proteins were less stable to unfolding at lower ionic strength.

2. Experimental

2.1. Reagents

Reagents were of the highest grade commercially available. MilliQ™ water (Millipore, Bedford, USA) was used in all experiments. Ammonium acetate, ammonia solution and acetic acid were from Ajax Finechem (Seven Hills, Australia). Deuterium oxide solution (D_2O) was from Cambridge Isotope Laboratories (Andover, USA).

2.2. Construction of a versatile circularly permuted intein expression vector

The bacteriophage λ promoter vector pNW1120 that contains a circularly permuted *Ssp dnaB* split intein gene [21] was modified to incorporate a new multiple cloning site with flanking *Bbs*I sites between the intein gene segments. Oligonucleotides 569°(5′-CACGCGTTCTAGAGAATTC*GAAGAC*CTGGCTGC ATCTCTGGTGATTCTCTG) and 570°(5′-CTTCGAATTCTCTAGAACGCGT*G AAGAC*ATAGAGTTGT GAACGATGATGTCG), overlapping at their 3′-ends, were used in conjunction with vector primers P10 and P9 [26], respectively, to amplify separately two DNA fragments by PCR, using pNW1120 as template. These fragments, containing the Int_C and Int_N gene segments, respectively, were isolated, mixed and PCR-amplified using P9 and P10. The resulting single fragment was digested with *Bam*HI and *Nco*I and inserted between the corresponding sites of pCL476 [27] to generate pNW1230

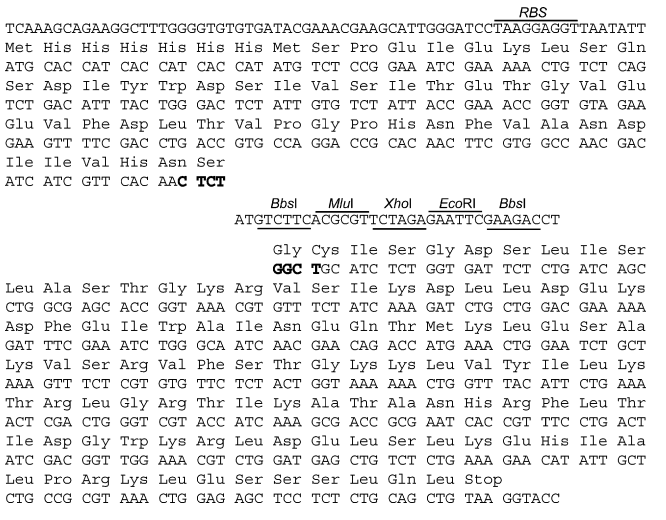


Fig. 1. Sequence of *Ssp dnaB* intein fragments and cloning region of the versatile circularly permuted intein expression vector pNW1230. The sequences of 5′-overhangs generated by *Bbs*I-digestion are shown in bold.

(Fig. 1). Digestion of pNW1230 with *Bbs*I, which cleaves outside its recognition sequence, results in DNA with 5′-overhangs (sequence in bold in Fig. 1) and no extraneous nucleotides flanking the Int_C and Int_N gene segments. For directional cloning, PCR fragments with 5′ overhangs complementary to the linearized pNW1230 can then be generated for insertion by incorporating *Bbs*I sites at each end, such that cleavage is internal to the recognition sequence. If, however, a *Bbs*I site occurs within the fragment to be cloned, any other restriction enzyme can be used that cuts remotely from its recognition sequence and produces 4-base 5′ overhangs. Alternatively, complementary oligonucleotides with the required overhangs can be synthesized directly for cyclization of short peptides.

2.3. Construction of lin- and cz-DnaBN expression vectors

Oligonucleotide PCR primers were designed to construct a series of DnaBN open reading frames for expression as linear or cyclized proteins, differing in the number of amino acid residues and their termini or cyclization points: 839 (5′-TTTTT TGAAGACGACTCTAAAGTGCCTCCGCACTCG), 840 (5′-TTGAATTCTTAGCC CGTCGAGATCATCTCACGGA), 841 (5′-TTTTTTGAAGACATCTCTTTCAAAG TGCCTCCGCACTC), 842 (5′-TTTTTTTTCATATGTCTTTCAAAGTGCCTCCGC ACTC), 843 (5′-TTTTTTGAAGACAGAGCCACGCGTCGAGATCATCTCAC), 844 (5′-TTGAATTCTTAGCCACGCGTCGAGATCATCTCAC), 845 (5′-TTTTTT GAAGACCTAGCCCGTCGAGATCATCTCACGGA) and 846 (5′-TTTTTTTTCAT ATGTCTAAAGTGCCTCCGCACTCG). Using pNW1130 [22] as template, primer pairs 842/844, 841/843, 842/840, 841/845, 846/840 and 839/845 produced PCR-amplified fragments for expressing 5-lin-DnaBN, 5-cz-DnaBN, 4-lin-DnaBN, 4-cz-DnaBN, 3-lin-DnaBN and 3-cz-DnaBN, respectively. For linear expression constructs, PCR fragments were digested with *Nde*I and *Eco*RI and inserted between the corresponding sites of λ promoter vector pND704 [27]. For cyclization constructs, PCR fragments were digested with *Bbs*I and ligated with *Bbs*I-treated pNW1230. All plasmid constructs were confirmed by nucleotide sequence determination. Some inconsistent expression was observed among these λ promoter constructs so the linear and cyclization DnaBN expression cassettes were subcloned as *Nde*I/*Nco*I fragments between the corresponding sites of the T7 promoter vectors pETMCSI and pETMCSIII [28], respectively. The sequence of the linkers of the various expressed proteins are shown in Table 1.

Table 1
Amino acid sequences of the DnaBN proteins used in this study.

DnaBN derivative	Amino acid sequences[a] (N-terminus. . .. . .DnaBN. . .. . .C-terminus)
3-lin- and 3-cz-DnaBN	*S*-DnaBN(F102E)-*TG*
4-lin- and 4-cz-DnaBN	*SF*-DnaBN(F102E)-*TG*
5-lin- and 5-cz-DnaBN	*SF*-DnaBN(F102E)-*TRG*
9-lin and 9-cz-DnaBN	*SIEF*-DnaBN(F102E)-*TRESG*

[a] Amino acid sequences of the linkers are shown in italics. The sequences are shown as the linear version. Use of an N-terminal serine residue ensures efficient removal of the N-terminal methionine residue from the linear versions by methionine aminopeptidase *in vivo*. In the cyclized versions, the serine (S) of the linear version is joined through a peptide bond to the glycine (G) residue.

2.4. Protein purification and characterization

Proteins 9-lin and 9-cz-DnaBN (Table 1) were purified as described previously [22]. For production of the other linear and cyclized DnaB-N proteins, the T7 promoter plasmids described above were separately transformed into *E. coli* BL21(λDE3)/pLysS [29]. Resulting strains were grown in 1-l cultures (except 3 l for 3-cz-DnaBN) for 2 days at room temperature in autoinduction medium as described [30], then cells were harvested and stored frozen at −80 °C. Cyclization was confirmed in the appropriate cases by SDS-PAGE analysis, by the diagnostic appearance of product intein fragments together with DnaBN proteins that migrated with higher mobility than the corresponding linear versions [21]. The proteins were all prepared essentially as described previously [21,22]. Briefly, cells were resuspended in 15 ml/g of lysis buffer containing sucrose and spermidine, and lysed using a French press. Proteins in the soluble fraction were precipitated by addition of 0.4 g/ml of ammonium sulfate, harvested by centrifugation, resuspended and dialyzed, before being subjected to chromatography on columns of DEAE-Fractogel, Sephadex G-50, and MonoQ (GE Healthcare). Pure DnaBN proteins all eluted in a single peak in a 0–300 mM NaCl gradient from the last column at ∼250 mM NaCl. The amino acid compositions and cyclic/linear status of all DnaBN proteins were confirmed by ESI-MS following extensive dialysis into 0.1% formic acid, and by SDS-PAGE analysis [21,22]. In each case the proteins contained the F102E mutation. This change was made to prevent dimerization of the proteins as occurs with wild-type DnaBN [24], and was shown in earlier work to have no effect on the protein structure and little effect on stability [22]. Protein concentrations were determined from ultraviolet spectra using a calculated [31] value of ε_{280} = 9530 M^{-1} cm^{-1}. Proteins were stored at −80 °C in 50 mM Tris–HCl buffer, pH 7.6, 100 mM NaCl, 5 mM $MgCl_2$ and 15% (v/v) glycerol.

2.5. Preparation of deuterium oxide (D_2O) solution

D_2O solution was prepared by adding 50 μl of 3.0 M ammonium acetate (NH_4OAc) in H_2O to deuterium oxide (99.9% D), giving a final volume of 15.0 ml. This produced 10 mM NH_4OAc in D_2O. The 'pH' was then adjusted to 7.2 (meter reading) by adding either solutions of 3% (v/v) ammonium hydroxide or 3% (v/v) acetic acid in D_2O. During the preparation, nitrogen gas was flushed over the solution. A thermostat-controlled water bath was used to keep the solution at 10 °C prior to addition to proteins.

2.6. Hydrogen/deuterium exchange and analysis by ESI-MS

Appropriate volumes of proteins were dialyzed against either 10 or 100 mM NH_4OAc, pH 7.2, at 4 °C and concentrated to ∼20 μl using a Millipore Biomax centrifugal filter (5000 MWCO), which had been pre-washed twice with 400 μl of 10 or 100 mM NH_4OAc, pH 7.2. Dialyzed, concentrated proteins (1 mM protein, 10 μl) and the D_2O solution (above) were equilibrated separately at 10 °C in a water bath for 10 min. A t = 0, an aliquot of the D_2O solution (990 μl, giving 98.5 ± 0.5% deuterium) was added to the protein, giving a final protein concentration of 10 μM. The protein samples were then stored at 10 °C.

To slow amide exchange substantially and to prevent back exchange, the exchange process was quenched by acidification to pH 2.1. At various time points, a 4 μl aliquot of the deuterated protein solution was taken and quenched by dilution with 36 μl of ice-cold quenching solution (water:methanol:formic acid, 90:9:1, pH 2.1). An aliquot (10 μl) of the quenched solution was then injected through a Rheodyne injector with a 10 μl sample loop into the mass spectrometer. The ice-cold quenching buffer was used as the mobile phase throughout the experiment and injected using a Harvard model 22 syringe pump (Natick, MA, USA) at a flow rate of 50 μl/min.

HDX was monitored in positive ion mode using a Waters Q-ToF Ultima™ ESI mass spectrometer (Manchester, UK) equipped with a Z-spray ionization source. The capillary, cone and RF lens 1 energies were set to 2500, 50 and 120 V, respectively. The transport and aperture were both set to 5 V. The source block and desolvation temperatures were 40 and 120 °C, respectively.

3. Results and discussion

In preliminary experiments, the importance of careful control of conditions for quenching HDX was confirmed. Lowering the pH to ≤2.5 and the temperature to 0 °C decreases the chemical rate of HDX (k_{ch}) by approximately five orders of magnitude [6]. When HDX is not quenched, the observed mass change is a contribution from both the amide protons and labile protons on the side chains, less any that have undergone back exchange (replacement of amide deuterons with protons). Since the D_2O was present at around 98.5% in these experiments, back exchange was minimized. When samples were injected into the mass spectrometer without quenching, as expected, the number of exchanged protons was greater. In addition, a single potassium adduct was present on each of the unfolded forms of the proteins (A and B, see below), making calculations of first-order rate constants based on the abundance of these forms more complicated. Careful investigation of salts and protein stocks suggested that the source of potassium was the stock D_2O. Since DnaBN has an isoelectric point of 4.6, it is susceptible to adduction by cations. Quenching of the exchange reaction by lowering the pH therefore served two functions: (i) arresting proton exchange and (ii) removing the potassium adducted to the proteins.

All of the proteins, 3-, 4-, 5-, and 9-lin-DnaBN and 3-, 4-, 5-, and 9-cz-DnaBN were subjected to HDX in 10 mM ammonium acetate, pH 7.2 at 10 °C. These were the same conditions as applied in our previous work [22]. ESI mass spectra were acquired at various time points. The spectra were similar to those previously reported for 9-lin-DnaBN and 9-lin-DnaBN where there was one predominant charge envelope in which the $[M+7H]^{7+}$ and $[M+8H]^{8+}$ ions were the most abundant with $[M+6H]^{6+}$ also present [32]. Ions to lower and higher values of mass/charge (m/z) were present in very low abundance (≤5% of the abundance of the most abundant ion in the spectrum). Fig. 2 shows the ESI mass spectra (transformed to a mass scale) obtained at various time points for 3-lin-DnaBN (a) and 3-cz-DnaBN (b). Two peaks (labeled A and B) are evident in the spectra. These peaks were present for each of the $[M+6H]^{6+}$ to $[M+8H]^{8+}$ ions (see Supplementary data, Fig. S1). The differences in mass (number of protons exchanged) of peaks A and B relative to the proteins prior to treatment with D_2O is plotted against time in Fig. 3. A rapid increase in mass was observed within the first minutes of exposure to D_2O. The maximum number of protons exchanged for peak A for 3-lin-DnaBN was about 45 amide

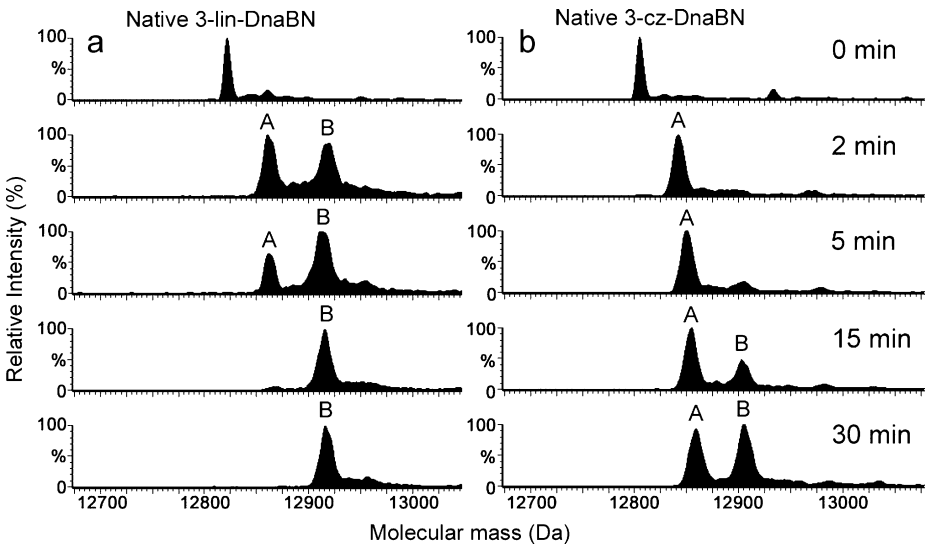


Fig. 2. Positive ion ESI mass spectra of linear and cyclized DnaBN with a three amino acid linker subjected to hydrogen/deuterium exchange at various time points. The proteins in 10 mM NH_4OAc, pH 7.2, were treated with 10 mM NH_4OAc, pH 7.2 in D_2O at 10 °C. (a) 3-lin-DnaBN and (b) 3-cz-DnaBN. The ESI mass spectra were transformed to a mass scale using MassLynx software™. Peak A corresponds to the amide proton exchange of the solvent-exposed surface of the folded protein; peak B corresponds to the amide proton exchange of the unfolded protein.

protons (from 12,799 to 12,844 Da) and for peak B was about 96 amide protons (12,895 Da; an additional 51 protons in excess of those that exchanged to give peak A). These differences were taken from the regions of the time courses where the number of protons exchanged was relatively stable. For example, for 3-lin-DnaBN the number of protons exchanged for the peak A was taken from the number of protons exchanged over the time period 10–15 min (3 values) and for peak B was 17.5–30 min (4 values). Similar behavior was observed for all proteins with small fluctuations (several protons) in the total number of protons exchanged over the time period where the number of exchanged protons remained relatively stable. This may be the result of localized unfolding involving a few residues in response to handling and/or the electrospray process.

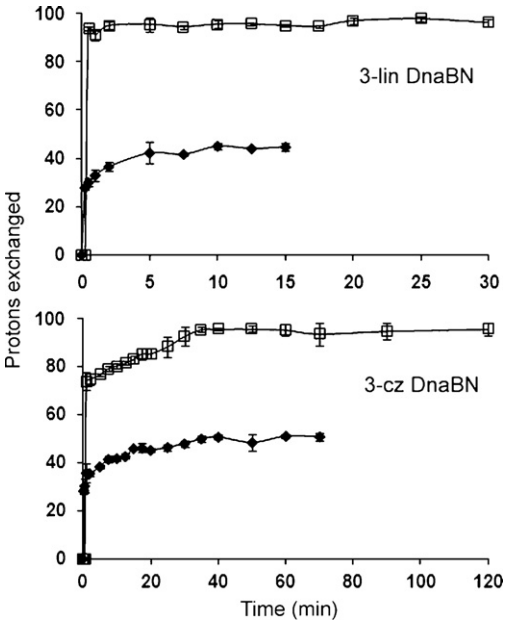


Fig. 3. The number of protons exchanged for 3-lin- and 3-cz-DnaBN at various times. Peak A (closed diamonds); peak B (squares). Each time point is the average of three experiments and the error bars show ±1 standard deviation. The number of peptide bonds in each of the proteins (maximum for 100% exchange) is: 3-lin (115), 3-cz (116), 4-lin (116), 4-cz (117), 5-lin (117), 5-cz (118), 9-lin (121) and 9-cz (122).

Peak A is the result of exchange of protons, not involved in H-bonding, on the solvent-exposed surface regions of the folded protein. The decay of peak A and the subsequent appearance of peak B is consistent with, and characteristic of, EX1 exchange. The loss of peak A and appearance of peak B was slower for the cyclized protein in agreement with our earlier observations of 9-lin- and 9-cz-DnaBN [22]. Similar results were observed for all of the proteins.

In both the linear and cyclized proteins, irrespective of linker length, the A and B forms were the same (or very similar) within experimental error, suggesting that while the cyclized proteins were more stable to unfolding, they unfolded by the same mechanism. The peak shapes of A and B were a little distorted, especially at the earlier time points (Fig. 2), suggesting that there were some additional local motions in small populations of protein molecules.

Previously, a 10-fold difference in the rates of exchange for the most slowly exchanging protons of 9-lin- and 9-cz-DnaBN was observed for measurements carried out by NMR compared with ESI-MS. This was attributed to the different buffers used in those experiments [22]. To compare the stabilities of the linear and cyclized proteins in solutions of different ionic strength, the HDX experiments were also carried out in 100 mM NH_4OAc, pH 7.2. In common with the results obtained in 10 mM NH_4OAc, two distinct populations were observed for all the proteins (peaks A and B) consistent with EX1 exchange. Fig. 4 shows the ESI mass spectra (transformed to a mass scale) for 3-lin- and 3-cz-DnaBN. There was a modest increase for all proteins in the number of amide protons on the solvent-exposed surface (peak A) at the higher ionic strength. Having established that the protein populations represented by peaks A and B (*cf.* Fig. 2) were similar in all cases, the pseudo-first order rate constants for the decay of peak A (appearance of peak B) were calculated. To determine the relative amounts of the protein forms A and B, the relative abundances in the spectra transformed to a mass scale were established by centering the data relative to peak heights and also by peak area. The relative amounts determined by the two methods were similar, but given the deviation from Gaussian distribution of the peak shape at early time points, peak areas were used to calculate all the rate constants presented here. It is unlikely that the peak shape involves a significant contribution by other conformational forms since this was also observed for individual ions (see Supplementary data, Fig. S1) and ions outside the $[M+8H]^{8+}$ to $[M+6H]^{6+}$ charge envelope were of very low

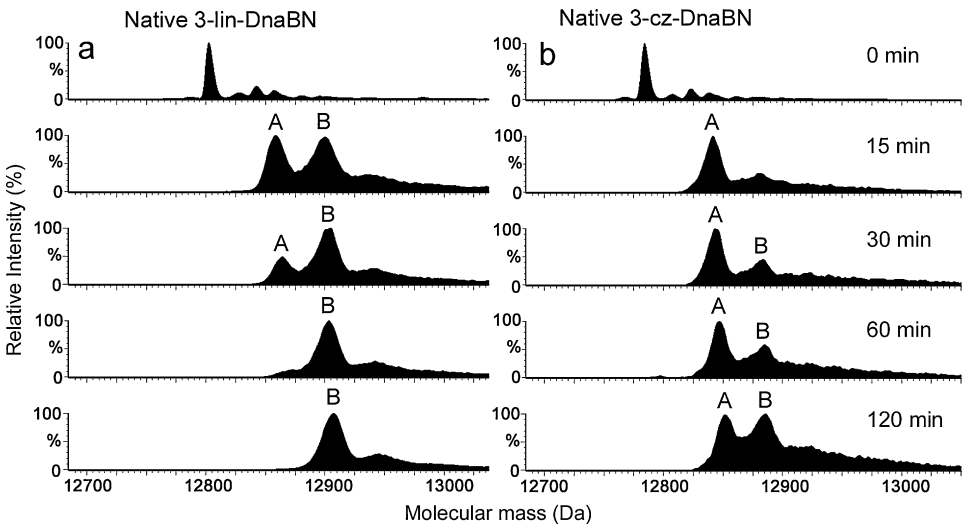


Fig. 4. Positive ion ESI mass spectra of linear and cyclized DnaBN with a three amino acid linker subjected to hydrogen/deuterium exchange at various time points. The proteins in 100 mM NH_4OAc, pH 7.2, were treated with 100 mM NH_4OAc, pH 7.2 in D_2O at 10 °C. (a) 3-lin-DnaBN and (b) 3-cz-DnaBN. The ESI mass spectra were transformed to a mass scale using MassLynx software™. Peak A corresponds to the amide proton exchange of the solvent-exposed surface of the folded protein; peak B corresponds to the amide proton exchange of the unfolded protein.

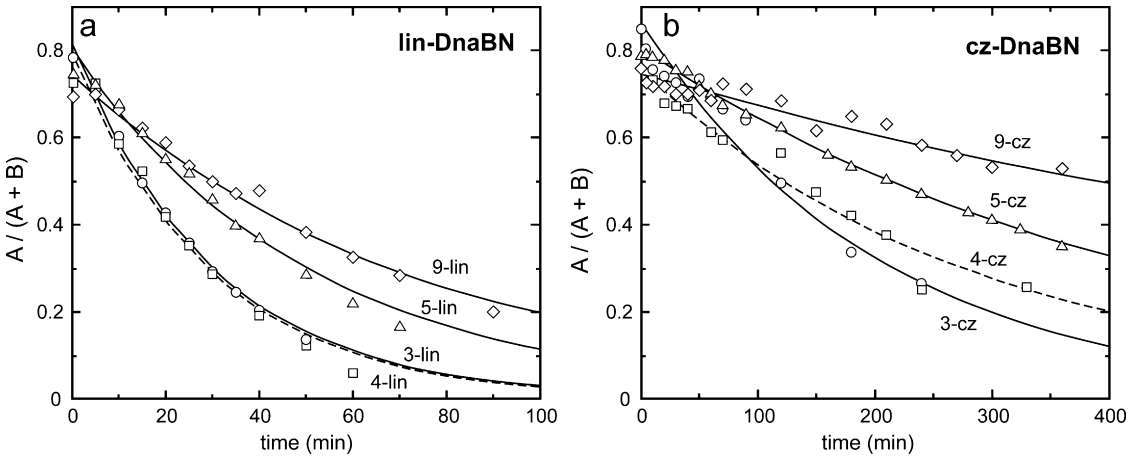


Fig. 5. Progress of hydrogen/deuterium exchange, expressed as the ratio of A ions to total of A and B ions (see Fig. 2) of linear (a) and cyclized (b) DnaBN containing linkers of 3 (circles), 4 (squares), 5 (triangles) and 9 (diamonds) amino acids in 100 mM NH_4OAc (98.5% D_2O), pH 7.2, 10 °C. Each data point was determined from the average of three separate experiments. The lines represent first-order fits using rate constants given in Table 3. Data collected at longer time intervals were used to obtain rate constants, but have been omitted for clarity. Note the different time scales in (a) and (b).

abundance (see above). The progress of HDX in 100 mM NH_4OAc, pH 7.2, expressed as the ratio of peak A to total of peaks A and B (see Fig. 4) is shown for each of the proteins in Fig. 5. Tables 2 and 3 show the rate constants (k_{obs}) calculated from the data for the HDX measured in 10 and 100 mM NH_4OAc, respectively. For each protein, the rate of exchange was approximately 5-fold greater when measured in 10 mM NH_4OAc. This is in agreement with our previous work where the stabilities of 9-lin- and 9-cz-DnaBN to denaturation by treatment with urea were enhanced in 50 mM Tris–HCl buffer, pH 8.0, 260 mM in NaCl compared to when the buffer was 10 mM sodium phosphate, pH 7.2 [22].

The ratios of the rate constants for unfolding of 9-lin- and 9-cz-DnaBN, (EX1 limit observed), $k_{obs(lin)}/k_{obs(cz)}$, in 10 mM NH_4OAc were in reasonable agreement with our earlier work (5.9 here and 8.9 previously [22]). Comparison of the results in Tables 2 and 3 shows that while all proteins unfolded at a slower rate in 100 mM

Table 2
First-order rate constants (k_{obs}) for unfolding of linear and cyclized DnaBN with different linker lengths in 10 mM NH_4OAc, pH 7.2, 10 °C.

Linker length (amino acids)	k_{obs} (min^{-1})[a]		$k_{obs(lin)}/k_{obs(cz)}$
	Linear	Cyclized	
3	0.21 ± 0.02	0.028 ± 0.001	7.4 ± 1.0
4	0.144 ± 0.008	0.021 ± 0.002	7.0 ± 0.9
5	0.109 ± 0.003	0.0147 ± 0.0009	7.4 ± 0.6
9	0.044 ± 0.002	0.0074 ± 0.0009	5.9 ± 1.0

[a] Rate constants were determined from the average of three sets of experiments; errors are ±1 standard deviation.

Table 3
First-order rate constants for unfolding of linear and cyclized DnaBN with different linker lengths in 100 mM NH_4OAc, pH 7.2, 10 °C.

Linker length (number of amino acids)	k_{obs} (min^{-1})[a]		$k_{obs(lin)}/k_{obs(cz)}$
	Linear	Cyclized	
3	0.034 ± 0.001	0.0049 ± 0.0005	6.8 ± 0.9
4	0.034 ± 0.002	0.0034 ± 0.0003	9.9 ± 1.4
5	0.0196 ± 0.0009	0.00225 ± 0.00003	8.7 ± 0.5
9	0.0134 ± 0.0005	0.00104 ± 0.00006	12.8 ± 1.2

[a] Rate constants were determined from the average of three sets of experiments; errors are ±1 standard deviation.

NH_4OAc, $k_{obs(lin)}/k_{obs(cz)}$ was approximately 10-fold for the proteins with the nine amino acid linker (12.8). This shows that the rate of global unfolding is decreased by about 10-fold when the distance between the N- and C-termini is restricted to a nine amino acid linker. In our previous experiments, the nine amino acid linker was chosen to join the N- and C-termini of DnaBN so that no strain was introduced to the molecule based upon calculations in which a polypeptide chain is represented as a Gaussian chain composed of statistical chain segments [21,22]. A statistical chain is a polypeptide segment in which the orientation of the first amino acid is uncorrelated with the last. The length of a statistical chain is about 35 Å or nine amino acids in random coil proteins [33,34]. NMR experiments showed that in both 9-lin- and 9-cz-DnaBN the nine amino acids in the linker (or extensions in the linear protein) were highly solvent exposed and mobile as expected for a random coil [22].

Previously we suggested, based on our observations for 9-lin- and 9-cz-DnaBN, that the ~10-fold difference in unfolding rates for the linear and cyclized proteins meant that global unfolding events involved separation of the N- and C-termini by a distance exceeding the length of the linker. Further, it was speculated that "unzipping" the N- and C-terminal helices was an event that triggered global unfolding. That the cyclized protein still underwent global unfolding showed that this process can also commence from different parts of the molecule, but less frequently [22]. These proposals were explored by preparing linear and cyclized proteins of the same amino acid composition but with different (shorter) linker lengths. Based on our previous proposals it might be expected that the presence of the amino acids extensions would have little effect on the rate of unfolding of the linear protein. This was not the case as the rate of unfolding (k_{op}), as judged by the rate of amide proton exchange for the slowly exchanging protons increased as the linker length was shortened (0.0134 min^{-1} for 9-lin-DnaBN *cf.* 0.034 min^{-1} for 3-lin-DnaBN; Table 3). Therefore, the longer linker presented a modest barrier to global unfolding. Given the previous NMR data that indicated the nine-residue linker to be unstructured [21,22], this seems unlikely to be due to some kind of steric hindrance or the nature of the amino acids themselves.

The source of this modest stabilization by increase of the N- and/or C-terminal extensions in both the lin- and cz-DnaBN proteins is, in fact, easily explained. In the original NMR structure of DnaBN(24–136) [24], residues Met134 to Ser136 following helix 6 were clearly observed to be poorly structured, yet in 9-cz-DnaBN helix 6 extended as far as Ser136 [21]. In the hexameric DnaB structures from other organisms [23,35,36], the C-terminal helix extends for several additional turns beyond Ser136, and is stabilized by an additional helix 7 that packs in an anti-parallel orientation. This arrangement of an identical helical core extended by a C-terminal helical hairpin structure is also seen in structures of the DnaB-binding domain of primases [23,37–39]. All of this evidence supports our prediction that the C-terminal helix 6 of the present DnaBN constructs are, in full-length DnaB, stabilized by its extension and its packing against a complementary helix 7.

Our expectation from this argument is that addition of amino acid sequences at the C-terminus of DnaBN(24–136) that have propensity to extend its C-terminal helix 6 (in either the linear or cyclized form) are likely to have a stabilizing effect on the protein as a whole. Our present HDX-MS data are consistent with this expectation. To demonstrate this in an objective way, we submitted sequences all of lin-DnaBN proteins (Table 1) for secondary structure prediction by the JPred3 server [40]. In the structurally defined region of DnaBN, JPred's predictions are consistent with the known structure of this region of *E. coli* DnaB, so our question reduces to its prediction of the effects of addition of a few additional residues at the N- and C-termini. In 9-lin-DnaBN JPred predicts helix 6 to extend to the Arg residue of the linker (Table 1). In 5-lin-DnaBN, it is predicted to be one residue shorter, while in 4-lin- and 3-lin-DnaBN, the prediction suggests it to be 2 residues shorter. Thus, the predicted length of the C-terminal helix is directly correlated with kinetic stability (Tables 2 and 3)—the longer the C-terminal helix 6, the slower is global EX1 amide exchange.

For the same reason, the rate of protein unfolding was also greater for the cyclized proteins as the linker lengths were shortened: global unfolding was facilitated by tethering of the N- and C-termini closer together. Interestingly, the ratio $k_{obs(lin)}/k_{obs(cz)}$ was similar for all amino acid linker/extension lengths, but tended towards smaller numbers as the linker length was reduced (12.8 for 9-lin-DnaBN *cf.* 6.8 for 3-lin-DnaBN; Table 3). This indicates that the rates of global unfolding of cz-DnaBN proteins with the shorter linkers are enhanced relative to predictions based solely on the reduced helical propensity of residues in helix 6. This is consistent with a mechanism by which drawing the termini together introduces some strain that moderately destabilizes the protein structure near the ends of the terminal helices.

4. Conclusions

Generally, the unfolding and folding of single domain, simple helix bundle proteins occur very rapidly under conditions that destabilize the native structure [41]. While DnaBN is a single domain protein consisting of six helices, it is not a simple helix bundle and its fold is rare. Its folding/unfolding behavior exhibited here and in our previous work [21,22] is unusual for an all-helical protein with a low contact order. Contact order describes the proportion of local to non-local contacts in a native protein [42]. The observation of EX1 exchange for a protein under native conditions at near neutral pH as is observed here for linear and cyclized versions of DnaBN is uncommon [5]. The experiments show that cyclization by joining the N- and C-termini of DnaBN substantially slows the rate of global unfolding and that this process is sensitive to ionic strength. The observed relationship between the unfolding rates and linker length was complex, but it could be explained by a combination of two effects: (1) lengthening of the C-terminal helix 6 in both the lin- and cz-DnaBNs by one or two residues in going from 3- or 4-DnaBN to 5- and 9-DnaBN proteins decreases their unfolding rates progressively by small factors (Tables 2 and 3); (2) use of the shorter linkers introduced some strain in the cz-DnaBN proteins that destabilizes them relative to predictions based on rates obtained with the lin-DnaBN series. This could be investigated in future experiments by altering the amino acids that are present in the linkers as well as investigating the effects of a greater range of linker lengths. Our arguments illustrate how measurements of rates of global HDX by ESI-MS can be used to probe very fine details of protein structure.

Acknowledgements

We thank the University of Wollongong, Australia, for supporting this work and Dr Stephen J. Watt for preliminary experiments. The award of a fee waiver scholarship (T. Urathamakul) from the University is also gratefully acknowledged. We also thank the Australian Research Council for research grants, including funding of the mass spectrometers used in this work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ijms.2010.09.011.

References

[1] A.A. Edwards, J.D. Tipton, M.D. Brenowitz, M.R. Emmett, A.G. Marshall, G.B. Evans, P.C. Tyler, V.L. Schramm, Conformational states of human purine nucleoside phosphorylase at rest, at work, and with transition state analogues, Biochemistry 49 (2010) 2058–2067.

[2] R.V Mauldin, A.L. Lee, Nuclear magnetic resonance study of the role of M42 in the solution dynamics of *Escherichia coli* dihydrofolate reductase, Biochemistry 49 (2010) 1606–1615.

[3] Z. Dosztányi, B. Mészáros, I. Simon, Bioinformatical approaches to characterize intrinsically disordered/unstructured proteins, Brief Bioinform. 11 (2010) 225–243.

[4] R. Jaenicke, Protein self-organization in vitro and in vivo. Partitioning between physical biochemistry and cell biology, Biol. Chem. 379 (1998) (1998) 237–243.

[5] I.A. Kaltashov, S.J. Eyles, Studies of biomolecular conformations and conformational dynamics by mass spectrometry, Mass Spectrom. Rev. 21 (2002) 37–71.

[6] J.R. Engen, Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS, Anal. Chem. 81 (2009) 7870–7875.

[7] C. Redfield, NMR studies of partially folded molten-globule states, Methods Mol. Biol. 278 (2004) 233–254.

[8] J.G. Mandell, A.M. Falick, E.A. Komives, Measurement of amide hydrogen exchange by MALDI-TOF mass spectrometry, Anal. Chem. 70 (1998) 3987–3995.

[9] V. Katta, B.T. Chait, Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry, Rapid Commun. Mass Spectrom. 5 (1991) 214–217.

[10] S.W. Englander, L. Mayne, Y. Bai, T.R. Sosnick, Hydrogen exchange: the modern legacy of Linderstrom-Lang, Protein Sci. 6 (1997) 1101–1109.

[11] A. Hvidt, S.O. Nielsen, H exchange in proteins, Adv. Protein Chem. 21 (1966) 287–386.

[12] D.M. Ferraro, A.D. Robertson, EX1 hydrogen exchange and protein folding, Biochemistry 43 (2004) 587–594.

[13] Y. Bai, J.S. Milne, L. Mayne, S.W. Englander, Primary structure effects on peptide group hydrogen exchange, Proteins 17 (1993) 75–86.

[14] T.E. Wales, J.R. Engen, Hydrogen exchange mass spectrometry for the analysis of protein dynamics, Mass Spectrom. Rev. 25 (2006) 158–170.

[15] A. Miranker, C.V. Robinson, S.E. Radford, R.T. Aplin, C.M. Dobson, Detection of transient protein folding populations by mass spectrometry, Science 262 (1993) 896–900.

[16] T.J.D. Jorgensen, L. Cheng, N.H.H. Heegaard, Mass spectrometric characterization of conformational preludes to β_2-microglobulin aggregation, Int. J. Mass Spectrom. 268 (2007) 207–216.

[17] J.P. Hodkinson, T.R. Jahn, S.E. Radford, A.E. Ashcroft, HDX-ESI-MS reveals enhanced conformational dynamics of the amyloidogenic protein β_2-microglobulin upon release from the MHC-1, J. Am. Soc. Mass Spectrom. 20 (2009) 278–286.

[18] J.R. Cherry, A.L. Fidantsef, Directed evolution of industrial enzymes: an update, Curr. Opin. Biotechnol. 14 (2003) 438–443.

[19] N.L. Daly, D.J. Craik, Design and therapeutic applications of cyclotides, Future Med. Chem. 1 (2009) 1613–1622.

[20] H. Paulus, Protein splicing and related forms of protein autoprocessing, Annu. Rev. Biochem. 69 (2000) 447–496.

[21] N.K. Williams, P. Prosselkov, E. Liepinsh, I. Line, A. Sharipo, D.R. Littler, P.M.G. Curmi, G. Otting, N.E. Dixon, *In vivo* protein cyclization promoted by a circularly permuted *Synechocystis sp.* PCC6803 DnaB mini-intein, J. Biol. Chem. 277 (2002) 7790–7798.

[22] N.K. Williams, E. Liepinsh, S.J. Watt, P. Prosselkov, J.M. Matthews, P. Attard, J.L. Beck, N.E. Dixon, G. Otting, Stabilization of native protein fold by intein-mediated covalent cyclization, J. Mol. Biol. 346 (2005) 1095–1108.

[23] S. Bailey, W.K. Eliason, T.A. Steitz, Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase, Science 318 (2007) 459–463.

[24] J. Weigelt, S.E. Brown, C.S. Miles, N.E. Dixon, G. Otting, NMR structure of the N-terminal domain of *E. coli* DnaB helicase: implications for structure rearrangements in the helicase hexamer, Structure 7 (1999) 681–690.

[25] D. Fass, C.E. Bogden, J.M. Berger, Crystal structure of the N-terminal domain of the DnaB hexameric helicase, Structure 7 (1999) 691–698.

[26] C.M. Elvin, P.R. Thompson, M.E. Argall, P. Hendry, N.P.J. Stamford, P.E. Lilley, N.E. Dixon, Modified bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*, Gene 87 (1990) 123–126.

[27] C.A. Love, P.E. Lilley, N.E. Dixon, Stable high-copy-number bacteriophage λ promoter vectors for overproduction of proteins in *Escherichia coli*, Gene 176 (1996) 49–53.

[28] C. Neylon, S.E. Brown, A.V. Kralicek, C.S. Miles, C.A. Love, N.E. Dixon, Interaction of the *Escherichia coli* replication terminator protein (Tus) with DNA: a model derived from DNA-binding studies of mutant proteins by surface plasmon resonance, Biochemistry 39 (2000) 11989–11999.

[29] F.W. Studier, A.H. Rosenberg, J.J. Dunn, J.W. Dubendorff, Use of T7 RNA polymerase to direct expression of cloned genes, Methods Enzymol. 185 (1990) 60–89.

[30] F.W. Studier, Protein production by auto-induction in high-density shaking cultures, Protein Express. Purif. 41 (2005) 207–234.

[31] S.C. Gill, P.H. von Hippel, Calculation of protein extinction coefficients from amino acid sequence data, Anal. Biochem. 182 (1989) 319–326.

[32] S.J. Watt, M.M. Sheil, J.L. Beck, P. Prosselkov, G. Otting, N.E. Dixon, Effect of protein stabilization on charge state distribution in positive- and negative-ion electrospray ionization mass spectra, J. Am. Soc. Mass Spectrom. 18 (2007) 1605–1611.

[33] W.R. Krigbaum, T.S. Hsu, Molecular conformation of bovine A1 basic protein, a coiling macromolecule in aqueous solution, Biochemistry 14 (1975) 2542–2546.

[34] M.S.Z Kellermayer, S.B. Smith, H.L. Granzier, C. Bustamante, Folding-unfolding transitions in single titin molecules characterized with laser tweezers, Science 276 (1997) 1112–1116.

[35] T. Biswas, O.V. Tsodikov, Hexameric ring structure of the N-terminal domain of *Mycobacterium tuberculosis* DnaB helicase, FEBS J. 275 (2008) 3064–3071.

[36] Y.-H. Lo, K.-L. Tsai, Y.-J. Sun, W.-T. Chen, C.-Y. Huang, C.-D. Hsiao, The crystal structure of a replicative hexameric helicase DnaC and its complex with single-stranded DNA, Nucleic Acids Res. 37 (2009) 804–814.

[37] A.J. Oakley, K.V. Loscha, P.M. Schaeffer, E. Liepinsh, G. Pintacuda, M.C.J. Wilce, G. Otting, N.E. Dixon, Crystal and solution structures of the helicase-binding domain of *Escherichia coli* primase, J. Biol. Chem. 280 (2005) 11495–11504.

[38] P. Soultanas, The bacterial helicase-primase interaction: a common structural/functional module, Structure 13 (2005) 839–844.

[39] X.-C. Su, P.M. Schaeffer, K.V. Loscha, P.H.P. Gan, N.E. Dixon, G. Otting, Monomeric solution structure of the helicase binding domain of *Escherichia coli* DnaG primase, FEBS J. 273 (2006) 4997–5009.

[40] C. Cole, J.D. Barber, G.J. Barton, The Jpred 3 secondary structure prediction server, Nucleic Acids Res. 36 (2008) W197–201.

[41] J.K. Myers, T.G. Oas, Preorganized secondary structure as an important determinant of fast protein folding, Nat. Struct. Biol. 8 (2001) 552–558.

[42] K.W. Plaxco, K.T. Simons, D. Baker, Contact order, transition state placement and the refolding rates of single domain proteins, J. Mol. Biol. 277 (1998) 985–994.