Published in final edited form as:

J Struct Biol. 2007 July ; 159(1): 111–121.

Human Topoisomerase I C-Terminal Domain Fragment Containing the Active Site Tyrosine is a Molten Globule: Implication for the Formation of Competent Productive Complex

Chandanamali Punchihewa¹, Jixun Dai¹, Megan Carver¹, and Danzhou Yang^{1,2,*}

¹College of Pharmacy, The University of Arizona, 1703 E. Mabel St, Tucson, AZ 85721

²Arizona Cancer Center, 1515 N. Campbell Avenue, Tucson, AZ 85724

Abstract

Human topoisomerase I (topo I) is an essential cellular enzyme that relaxes DNA supercoiling. The 6.3 kDa C-terminal domain of topo I contains the active site tyrosine (Tyr723) but lacks enzymatic activity by itself. Activity can be fully reconstituted when the C-terminal is associated with the 56 kDa core domain. Even though several crystal structures of topo I/DNA complexes are available, crystal structures of the free topo I protein or its individual domain fragments have been difficult to obtain. In this report we analyze the human topo I C-terminal domain structure using a variety of biophysical methods. Our results indicate that this fragment protein (topo6.3) appears to be in a molten globule state. It appears to have a native-like tertiary fold that contains a large population of α -helix secondary structure and extensive surface hydrophobic regions. Topo6.3 is known to be readily activated with the association of the topo I core domain, and the molten globule state of topo6.3 is likely to be an energy-favorable conformation for the free topo I C-terminal domain protein. The structural fluctuation and plasticity may represent an efficient mechanism in the topo I functional pathway, where the flexibility aids in the complementary association with the core domain and in the formation of a fully productive topo I complex.

Keywords

Human topoisomerase I C-terminal domain; molten globule; NMR; CD; fluorescence

Human topoisomerase I (topo I) is a pivotal cellular enzyme that catalyzes changes in topological forms of DNA and facilitates key cellular processes such as DNA replication, transcription, recombination and chromosome condensation (Wang 1996; Champoux 2001). Human topo I breaks one strand of duplex DNA via formation of a transient 3'-phosphotyrosine ester bond between the active site tyrosine (Tyr723) and the DNA scissile strand backbone phosphate. This is followed by the unwinding of DNA, and by a second transesterification reaction resulting in religation of the cleaved DNA and dissociation of the topo I-DNA complex. Topo I is the sole molecular target for clinical camptothecin anticancer drugs (Hsiang and Liu 1988; Jaxel et al. 1989).

* To whom correspondence should be addressed. Telephone: (520) 626-5969, Fax: (520) 626-6988, Email: yangd@pharmacy.arizona.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Based on proteolysis studies and X-ray structural analysis, the 91 kDa topo I is divided into four domains: the N-terminal domain (~24 kDa), the core domain (~56 kDa), the linker (~6-8 kDa) and the C-terminal domain (~6-8 kDa) (Stewart et al. 1996). The C-terminal and core domains are two highly conserved domains which are critical for topo I enzymatic activity (Champoux 2001). The 6.3 kDa C-terminal domain of topo I (topo6.3) contains the active site Tyr723, as well as a number of residues that are involved in interactions with DNA and camptothecin. Furthermore, several mutations in this domain make the topo I protein drug-resistant. The C-terminal region has also been shown to bind ATP and carry out splicing factor phosphorylating activity (Rossi et al. 1998). The C-terminal domain is conserved between human and vaccinia viral topo I, the two main members of the type IB topo I family (Stewart et al. 1996; Stewart et al. 1997) and therefore is significant on its own. However, the C-terminal domain by itself lacks enzymatic activity, which can only be reconstituted when it is mixed with the core domain (topo56) containing the remaining catalytically relevant residues (Stewart et al. 1996; Stewart et al. 1997). It has been suggested that both the C-terminal and the core domains are independently folded and are simply associated with each other to form an active enzyme without conformational changes (Stewart et al. 1997).

A number of crystal structures have been determined for the DNA or DNA/drug complexes of topo I (Redinbo et al. 1998; Stewart et al. 1998; Redinbo et al. 1999; Staker et al. 2002; Staker et al. 2005); however, crystal structures of the free topo I protein or its individual domain fragments have been difficult to obtain. Here we have carried out biophysical studies of the human topo I C-terminal domain using a variety of methods. Our results indicate that the topo I C-terminal domain appears to be in a molten globule state in solution, containing a substantial population of α -helix secondary structure and a native-like conformation with extensive surface hydrophobic regions. As topo6.3 has been shown to be readily activated with the association of the topo I core domain, this fluctuating molten globule conformation of topo6.3 may be an energy-favorable conformation that provides an efficient starting point for the formation of a complete productive topo I complex.

Materials and methods

Vector construction

The topo I C-terminal domain protein we studied consists of 53 residues from the C-terminus of topo I plus three additional residues, Gly-Ile-Pro, added to the N-terminus by the expression vector (Figure 1). A pGEX-3X vector containing the topo6.3 coding region (Stewart et al. 1997) was used as the template for construction of modified topo6.3 DNA fragments using a PCR reaction. The N-terminal modification used a forward primer coding for five extra lysines and a reverse primer coding for a new EcoRI site following the stop codon. The C-terminal modification used a forward primer complementary to the vector fusion domain including the 5' region of the wild-type topo6.3 coding region and a reverse primer coding for five additional lysine residues, a stop codon and a new EcoRI site. Following the PCR reaction, the DNA was purified using a PCR purification kit (Qiagen). The amplified fragments of correct length were purified using agarose gel electrophoresis followed by DNA extraction with a gel purification kit (GeneClean). Subsequently, the DNA fragments were BamHI/EcoRI-digested to create cohesive ends for directional cloning. The digested fragments were gel-purified again and cloned directly into double-digested and purified pGEX-3X vector. The plasmids were amplified in *E. coli* BL21 and the sequences were verified by DNA sequencing. Each of the vectors encodes GST followed by a Factor Xa cleavage site 5' to the topo6.3 start site. The first residue in the topo6.3 proteins produced is glycine, with the starting methionine being found in the GST tag which will be proteolytically removed following the affinity purification of recombinant proteins.

Protein production and purification

Protein production was carried out as described (Stewart et al. 1997) in *E. coli* BL21 cells containing the wild-type or modified topo6.3 coding sequence cloned into pGEX-3X plasmids. Cells were cultured in rich growth medium and gene expression was induced at an OD₅₉₅ of 0.7 by addition of 40mg/l IPTG. Following incubation for a further 3 hours, the cells were harvested by centrifugation at 5500xg_{max} for 10 minutes and frozen overnight at -20°C. Pelleted cells were then resuspended in Sonication Buffer (500 mM KCl, 10 mM Tris-HCl (pH 7.4), 1 mM DTT, 1 mM EDTA, 0.2% Triton-X100, 0.1 mM PMSF) and sonicated for 3 minutes. The sample was then centrifuged again at 17500xg for 50 minutes to remove the cell debris. The supernatant was incubated with glutathione-Sepharose 4B beads (Pharmacia) overnight at 4°C and then eluted by gravity using an Econo-Pac column (Pharmacia). The matrix was washed with one column volume of sonication buffer and 2 column volumes of Wash Buffer (500 mM KCl, 10 mM Tris-HCl (pH 7.5), 1 mM EDTA). At the end of the washes, more Wash Buffer, equal to the volume of beads, was added into the same column and the final CaCl₂ concentration was adjusted to 3 mM. The sample was incubated for 24 hrs with Factor Xa (New England Biolabs) to cleave the protein off the beads and the GST tag. Final samples of topo6.3 were eluted from the column and dialyzed into 2.5 mM KCl before lyophilization. The presence of topo6.3 protein in the resultant solution was confirmed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Final protein concentration was determined by Bio-Rad protein assay (Bio-Rad Laboratories). The solubility of the wild-type topo6.3 and the N-terminal modified topo6.3 (5K-topo6.3) proteins was determined to be about 0.06 mM; while the solubility of the C-terminal modified topo6.3 (topo6.3-5K) protein was about 0.9 mM.

The labeled topo6.3 protein production was carried out in M9 minimal medium containing 22.4 mM ¹⁵NH₄Cl, 42.3 mM Na₂HPO₄, 22 mM KH₂PO₄, 8.6 mM NaCl, 0.5% glucose, 2 mM MgSO₄, 10 mg/l biotin, 2 mg/l thiamine, 50 mg/l ampicillin, 0.02% yeast extract, 0.1 mM CaCl₂ and trace elements. The incubation after IPTG addition was longer compared to that for the protein production in rich medium due to the slower growth of cells in minimal medium. The same method of protein purification as described above for the unlabeled topo6.3 proteins was used for the ¹⁵N labeled proteins.

CD spectroscopy

CD spectra were recorded on a Jasco J-810 spectropolarimeter and CD measurements were performed in both the PBS solution (137 mM NaCl, 10 mM Phosphate, 2.7 mM KCl, pH 7.4) and the NMR solution (50 mM pH 6 phosphate buffer, 100 mM KCl). For far-UV CD, protein concentrations were between 2.5-50 μM and the measurements were performed in a 1 mm pathlength cuvette. For near-UV CD, the C-terminally modified topo 6.3 protein (topo6.3-5K) was used at a concentration of 0.6 mM and the measurements were performed in a 1 cm pathlength cuvette due to the low sensitivity in the range of near-UV CD. For thermal denaturation experiments, the temperature was varied from 20 °C to 90 °C and the CD at 222 nm was monitored. For all other experiments the temperature was kept constant at either 25 °C or 37 °C. A scan rate of 100 nm/min was used for each measurement. Each data point in a CD spectrum was the average of three individual scans. All spectra were blank-subtracted using a PBS solution of pH 7.4 or NMR solution of pH 6.

Fluorescence measurements

Fluorescence was measured using a Molecular Devices SpectraMAX GeminiXS 96w fluorescent plate reader. The sample solutions were mixed in a 96-well plate, which was subsequently used for fluorescence measurement. ANS dye was used to determine exposed hydrophobic patches in each protein sample. A 10 mM ANS stock solution in water was used to obtain a final concentration of 300 or 5 μM in each well. Each protein sample was at a

concentration of 4 μ M. The excitation wavelength was 379 nm and the spectra were recorded from 425 nm to 600 nm. The spectra of the samples with urea were blank-subtracted with a solution containing the relevant amount of urea. Each experiment was repeated three times. The intrinsic tryptophan fluorescence was measured using 5-10 μ M solutions of each protein. The excitation wavelength was set at 280 nm and the emission spectra were recorded in the range of 300-400 nm.

NMR sample preparation

In preparation of NMR samples, excess powdered proteins were dissolved into appropriate buffer solution to achieve the maximum protein concentration. Specifically, approximately 1 mg of each of the wild-type and N-terminally modified topo6.3 proteins and 5 mg of C-terminally modified protein were dissolved in 600 μ l of NMR solution (50 mM pH 6.0 phosphate buffer, 100 mM KCl, 15% D₂O). The undissolved protein was removed by centrifugation and the solution was transferred to a new tube. The protein concentration in each sample was determined using the Bio-Rad Protein Assay system. The wild-type and N-terminally modified topo6.3 were at an approximate concentration of 0.06 mM and the C-terminally modified topo6.3 was at a concentration of $\sim$ 0.9 mM.

NMR data collection

One- and two-dimensional NMR spectra were acquired at 5-45 $^{\circ}$ C using a Bruker Avance 600 spectrometer. Homonuclear NOESY spectra were collected using States-TPPI method with a sweep width of 12 ppm in two dimensions. The relaxation delay was 2 seconds and the mixing time was 200 ms. The Watergate method was employed to suppress solvent peaks. For the sample of urea-denatured protein, multiple solvent peak suppression was achieved by using a train of shaped pulses targeted at the solvent peaks during the relaxation delay to presaturate the solvent peaks. All homonuclear 2D spectra had 2048 complex data points in the direct acquisition dimension and 1024 FIDs collected in the F1 dimension. The 90 degree shifted square sinebell function was used as the window function. After zero-filling and Fourier transformation, the size of the final spectra was 2048 $\times$ 1024. For the ¹⁵N-labeled samples, the gradient-enhanced heteronuclear single-quantum coherence (HSQC) spectra were acquired with 2048 complex points in t2 and 512 points in t1. The sweep width in F1 and F2 dimensions was 40 ppm and 20 ppm, respectively. A 90 or 60 degree shifted square sinebell function was applied in both dimensions. Forward linear prediction in the F1 dimension and Fourier transformation was applied, the final spectra are with a size of 2048 $\times$ 1024 data points. All spectra were processed using Xwinnmr version 3.5 and analyzed with the SPARKY program.

Molecular computer modeling

Molecular modeling calculations were performed using Insight II/Discover III software package (version 2000.1 Accelrys, CA). Multiple modules of Insight II were used for the calculation. The initial coordinates of topo6.3 protein was taken from the C-terminal fragment of the crystal structure of the ternary complex consisting of 70 kD human topoisomerase I, 22-base pair DNA and topotecan (pdb ID: 1K4T) (Staker et al. 2002). All hydrogen atoms were added at pH 6.0, and the missing side chains were added using the Insight II Biopolymer module. Bond orders of topo6.3 were corrected in Insight II. The starting model of the 5K-topo6.3 was generated by adding the random coil sequence G-I-P-K-K-K-K at the N-terminus of topo6.3. The starting model of the topo6.3-5K was generated by adding residues G-I-P at the N-terminus and five lysine residues at the C-terminus of topo6.3 fragment. Bad geometrical contacts from the crystal structure were first fixed by manually selecting the preferred rotamer of side chain conformation without changing the backbone conformation, followed by a series of minimizations to gradually relax the starting structure. A CFF force field in Insight II was used for all of the calculations (Maple et al. 1998), with the non-bond

cell multipole method. The energy minimization was carried out by first using the steepest descent method and then using the conjugate gradient method until a system convergence of 0.01 Kcal/mol/Å was reached. A distance-dependent dielectric constant ($\epsilon=4r_{ij}$) was used to simulate the aqueous solvation effect (Orozco et al. 1990; Garemyr and Elofsson 1999). The minimized structure was soaked in water solvent and subjected to further energy minimization to a gradient of less than 1 Kcal/mol/Å. The energy-minimized structure was then subjected to a molecular dynamics calculation with 20 ps equilibration and 80 ps simulations at 300 K. The dynamics time step was 1 fs and dynamic trajectories were recorded every 0.5 ps for analysis. The lowest energy structure from the dynamics simulation was subjected to further energy minimization to a convergence of 1 Kcal/mol/Å.

Results

Modifications of topo6.3 protein indicate that addition of multiple lysines to the C-terminus, but not the N-terminus, significantly increases the solubility of topo6.3

The C-terminal domain of the human topo I protein is an α -helical globular protein fragment, based on the crystal structures of DNA or DNA/drug complexes of topo I (Redinbo et al. 1998; Stewart et al. 1998; Redinbo et al. 1999; Staker et al. 2002) (Figure 1A). The protein we have studied consists of 53 residues from the C-terminus of topo I plus three additional residues, Gly-Ile-Pro, added to the N-terminus by the expression vector (Figure 1). Wild-type topo6.3 protein, whose pI is calculated to be 6.3, exhibits relatively low water solubility (0.06 mM) under physiological conditions. In an attempt to improve solubility properties, five lysine residues were added either to the N-terminus (5K-topo6.3) or to the C-terminus (topo6.3-5K) of the topo6.3 protein, a strategy that has been used in many systems (Park et al. 2003; Terpe 2003). Each expression vector was sequenced for verification of the presence of the correct coding sequence, and the presence of topo6.3 protein in the solution was verified by SDS-PAGE analysis at the end of protein purification as well as by mass spectroscopic analysis (Figure S1).

Addition of multiple lysines to the C-terminus increases the solubility of topo6.3-5K protein to 0.9 mM, more than ten-fold higher than that of the wild-type protein; however, the same modification on the N-terminus makes no obvious change in the solubility.

Far-UV CD analysis indicates that topo6.3 proteins contain a significant population of α -helix structure but the secondary structures are not completely ordered

All topo6.3 proteins in pH 7.4 PBS solution showed double minima at 205 and 222 nm, indicating the existence of a considerable population of α -helical secondary structure in topo6.3 proteins (Figure 2A-C) (Venyaninov and Yang 1996). However, while an all-helical protein like the native C-terminal domain of topo I is expected to give double minima of more or less similar intensities, the negative ellipticity at 222 nm was much less compared to the ellipticity at 205 nm for topo6.3 proteins. Similar observations have been reported in other molten globule proteins (Peng and Wagner 1994; Demarest et al. 1999; Song et al. 2001). While the helix content is estimated to be approximately 55% in the C-terminal domain of the crystal structures (Figure 1) (Redinbo et al. 1998; Stewart et al. 1998; Redinbo et al. 1999; Staker et al. 2002), the ellipticity at 222 nm only indicates about 35-40% helix population for topo6.3 proteins (Dyson et al. 1992; Venyaninov and Yang 1996), indicating that the secondary structures in topo6.3 proteins are not completely ordered. No change in the shape of spectrum was detected by increasing the sample temperature to 37°C, the physiological temperature. Far-UV CD measurement of denatured topo6.3 proteins in 8 M urea showed no detectable ellipticity, confirming that topo6.3 proteins in solution are in a state distinct from the unfolded state. Similar spectra were observed for the proteins in pH 6 phosphate buffer, the solvent used in the NMR experiments. Even though the C-terminally modified topo6.3-5K exhibits a ten-fold

higher water solubility, its CD spectrum is similar to that of the wild-type topo6.3 protein, suggesting a similar overall folding for both proteins.

The CD measurements were performed for different concentrations of the three proteins to verify that the observed folding state of topo6.3 proteins was not a result of protein aggregation. In the range of CD measurements (5-50 μ M), all three topo6.3 proteins appear to be monomeric, as no concentration dependence was detected (data not shown).

Thermal denaturation patterns measured by CD indicate the lack of cooperative folding in topo6.3 proteins

The thermal denaturation transition is an important measure in the study of protein folding. The ellipticity at 222 nm is usually correlated with the α -helical content of a protein and therefore is a parameter commonly monitored to observe folding and unfolding transitions of a helical protein. The thermal transition pattern obtained for C-terminally-modified topo6.3 protein by monitoring ellipticity at 222 nm, indicated a non-cooperative transition (Figure 3). This observation is consistent with the molten globule conformation of the topo6.3-5K protein. Non-cooperative transitions were also observed for the wild-type and the N-terminal-modified proteins (data not shown).

ANS fluorescence spectroscopic study indicates the molten globule states of all three topo6.3 proteins

ANS (1-anilinonaphthalene-8-sulphonate) is a hydrophobic fluorescence probe that has been extensively used to monitor the molten globule states of proteins (Semisotnov et al. 1991; Li and Jing 2000; Vamvaca et al. 2004). ANS fluorescence studies were carried out with the different forms of topo6.3 protein. All three forms of topo6.3 protein bind ANS in water solution and give rise to enhanced fluorescence intensity with a characteristic blue shift of the emission maximum, indicating that the topo6.3 proteins are in a molten globule like state (Figure 4 A-C). Other molten globule proteins have been shown to exhibit similar increases in fluorescence upon ANS binding (Uversky et al. 1999; Li and Jing 2000; Vamvaca et al. 2004; Neyroz et al. 2006).

Very weak ANS binding was observed for the non-folded form of denatured topo6.3 proteins in 8 M urea. It is important to note that while topo6.3-5K exhibits much better solubility than the wild-type topo6.3, which is likely caused by the presence of a less extensive hydrophobic surface in topo6.3-5K (see below), topo6.3-5K exhibits similar ANS binding to that of the wild-type topo6.3. Thus the enhanced ANS binding of topo6.3 proteins are likely due to the molten globule states of conformation, rather than to the extensive hydrophobic surface of the topo6.3 proteins.

NMR studies of topo6.3 proteins indicate a fluctuating structure for all three topo6.3 proteins

The ^{15}N - ^1H HSQC spectra were collected for all three topo6.3 proteins; however, they all exhibit poor signal dispersion and substantial line broadening. The spectral line broadening appears to be independent of protein concentration as monitored by NMR. Because of the higher solubility of the topo6.3-5K protein (0.9 mM) compared to those of the topo6.3 and 5K-topo6.3 proteins (0.06 mM), the NMR study was focused on the topo6.3-5K protein. The HSQC NMR spectra of topo6.3-5K in pH 6 phosphate buffer and in 8M urea are shown in Figure 5. This HSQC spectrum is very similar to those of previously reported molten globule proteins (Schulman et al. 1997; Redfield et al. 1999; Song et al. 2001; Ramboarina and Redfield 2003; Wahlberg et al. 2003; Zhang et al. 2005). The line broadening of NMR resonances is characteristic for conformational fluctuations on a millisecond time scale. The existence of a dynamic equilibrium of multiple conformations can be clearly seen in the HSQC spectra. For example, the topo6.3-5K protein contains two asparagines and two glutamines (Figure 1),

whose characteristic side chain N-H crosspeaks all appear to be crosspeak clusters with broad linewidth (Figure 5A). Thus, the NMR data indicate that topo6.3 proteins are in a fluctuating molten globule state. In addition, we have carried out a hydrogen/deuterium exchange experiment on the topo6.3-5K sample. We found that the hydrogen/deuterium exchange rate is faster than that can be detected by NMR. When H₂O was added to the deuterated sample, the amide protons can be readily detected in a couple of minutes (data not shown). The HSQC spectrum of the denatured topo6.3-5K sample in 8 M urea is also shown in Figure 5B. There are a total of 61 residues in the topo6.3-5K protein, including three prolines, so 58 amide crosspeaks should be seen for the denatured topo6.3-5K sample. 56 amide peaks can be clearly seen in the HSQC spectrum, while the remaining two peaks are likely to be overlapping with other peaks. The purity of this sample is also supported by the SDS-PAGE gel and mass spectral analysis (Figure S1).

Because trifluoroethanol (TFE) has been known to be capable of stabilizing the α -helix secondary structure (Buck 1998), NMR spectra of topo6.3 proteins were also obtained in the presence of 10-20% TFE. The addition of TFE did not improve the HSQC NMR spectra, namely, the signal dispersion and line-broadening.

2D-NOESY spectra have also been collected for topo6.3 proteins. The C^αH protons of topo6.3 proteins are clearly up-field shifted (0.2-0.3 ppm) with respect to the denatured topo6.3 protein as observed in 2D-NOESY spectra (Figure S2A), suggesting the presence of α -helix secondary structure in topo6.3 proteins (Wishart et al. 1992). Poor signal dispersion and line broadening associated with structural fluctuations were also observed in the NOESY spectra (Figure S2B).

Tryptophan fluorescence spectra of topo6.3 proteins indicate the surface locations of tryptophans

The intrinsic fluorescence spectra can be used to monitor the hydrophobic environment of tryptophan residues and to provide indirect evidence of the locations of tryptophans in a protein. All three topo6.3 proteins give rise to similar fluorescence spectra. The tryptophan fluorescence spectra of the C-terminally modified topo6.3-5K protein in water and 8 M urea are shown in Figure 6A. The emission spectrum of the denatured topo6.3-5K protein in 8 M urea solution exhibits a maximum at 355 nm, which is typical for an unfolded protein where the indole chromophores of tryptophans are in full contact with free water molecules (Ladokhin 2000). Interestingly, a clear blue spectral shift to a lower wavelength (λ =340 nm) is observed for topo6.3-5K protein in water. This blue spectral shift is characteristic for the tryptophan residues at the surface of a folded protein, where their indole chromophores are in contact with bound water and other polar groups (Ladokhin 2000). Indeed, the topo6.3 protein contains three tryptophan residues, Trp 732, Trp 736, and Trp 754, which are all located at the surface of the C-terminal domain according to the crystal structures (Figure 1C).

Near-UV CD spectra of Topo6.3 proteins imply the lack of a rigid tertiary structure

The near-UV CD measurement provides information on the asymmetry of the immediate environment of aromatic groups, mainly tryptophan residues, of a protein, while the large negative near-UV CD signal is indicative of internal aromatic residues (Ptitsyn 1995). The topo6.3 protein has three tryptophan, three tyrosine, and two phenylalanine residues (Figure 1). The near-UV CD spectrum of the topo6.3-5K protein is shown in Figure 6B. No significant ellipticity was observed in the near-UV CD spectrum for topo6.3-5K. Denatured topo6.3-5K protein in 8 M urea gave a similar near-UV CD spectrum.

Molecular modeling study shows that the 5-lysine extension adopts different conformations at the C-terminus and the N-terminus of topo6.3 protein

Even though both the N-terminus and the C-terminus of topo6.3 appear to be reasonable targets for the addition of charged hydrophilic tags to improve the water solubility, addition of multiple lysines to the C-terminus significantly increases the solubility of the topo6.3 protein, whereas the same modification on the N-terminus makes no change in the solubility. We thus carried out molecular modeling calculations on the two modified topo6.3 proteins. Our molecular dynamics calculation shows that the C-terminal 5-lysine extension adopts a stable fold-back conformation and covers one of the major hydrophobic surfaces, Ile728-Val730-Ala731-Trp732-Trp736 (Figure 1B), whereas the 5-lysine sequence at the N-terminus displays no such favorable interactions with the rest of the protein (Figure 7). As shown previously, despite the different extent of the hydrophobic surfaces of topo6.3-5K and wild-type topo6.3 proteins, the ANS bindings of the two proteins are very similar (Figure 4), suggesting that the increased ANS binding is a consequence of the molten globule states of conformation for both proteins.

Discussion

Topo6.3 protein is in a molten globule state

Although the topo I C-terminal domain does not exist on its own in cellular conditions, understanding its structure is important due to a number of reasons; (i) it contains the active site tyrosine residue and a number of mutations that confer camptothecin resistance to the protein, (ii) it is conserved between human and vaccinia viral topo I, the two main members of type IB topo I family (Stewart et al. 1996; Stewart et al. 1997), (iii) it has been designated as a separate domain and has been suggested to fold independent of the rest of the protein (Stewart et al. 1997), (iv) the structure of this domain or any other domains of human topoisomerase I protein, in the absence of DNA, is not yet known. According to our data presented above, the topo I C-terminal domain, the topo 6.3 protein, appears to be in a molten globule state. A molten globule is a compact but incompletely folded state in which a protein contains extensive secondary structure but lacks fixed tertiary interactions; this molecular state has been well established and can be readily identified by biophysical methods (Kuwajima 1989; Ptitsyn 1995). The molten globule states of topo6.3 proteins, including the C-terminally and N-terminally modified topo6.3 proteins, are clearly indicated by the non-cooperative thermal transitions of these proteins. The molten globule states of topo6.3 proteins are further supported by a variety of biophysical methods, including CD, fluorescence, and NMR spectroscopies. The dynamic structural fluctuation of the molten globule states of topo6.3 proteins is clearly indicated by the NMR spectra (Figures 5 and S2), and such fluctuations obscure the structure determination. Interestingly, despite the different solubility and the extent of hydrophobic surfaces of topo6.3-5K and wild-type topo6.3 proteins, the ANS bindings of the two proteins are very similar (Figure 4), suggesting that both proteins are in the molten globule like states. As observed in the crystal structures (Figure 1) (Redinbo et al. 1998; Stewart et al. 1998; Redinbo et al. 1999; Staker et al. 2002), substantial helical structure is clearly shown to be present in topo6.3 proteins by both the CD and NMR methods (Figures 2 and S2).

Topo6.3 protein contains extensive hydrophobic surface and thus low water solubility, while the hydrophobicity can be changed by addition of multiple lysine residues

The poor water solubility of the topo6.3 protein appears to be related with the extensive hydrophobic surface observed in the C-terminal domain (Figure 1B). These hydrophobic regions, which would otherwise be embedded in the interface with the core domain in the intact topo I enzyme according to the crystal structures (Figure 1D), are now located at the surface of the topo6.3 protein and can cause protein aggregation. Our results indicate that although the addition of charged lysine residues to topo6.3 can increase the solubility of the protein, the effect is largely dependent on the region modified. Since the addition of multiple lysines to the

C-terminus significantly increases the solubility of the topo6.3 protein while the same modification on the N-terminus makes no change, it is clear that merely having a favorable charge on a protein may not be sufficient to improve the solubility of the protein. This difference in solubility can be explained by our molecular modeling study, which shows that the C-terminal 5-lysine extension adopts a stable fold-back conformation and covers one of the major hydrophobic surfaces, whereas the 5-lysine sequence at the N-terminus displays no such favorable interactions with the rest of the protein (Figure 7).

It is important to note that the low solubility observed for the wild-type topo6.3 protein is likely attributed to the extensive hydrophobic surface rather than to the molten globule conformations, because while the C-terminally modified topo6.3-5K protein is clearly a molten globule, as indicated by the non-cooperative thermal transition, it shows much higher water solubility due to the reduction in its hydrophobic surface.

Implication in the formation of a productive topo I complex

As the topo I C-terminal domain can be readily activated with the association of the topo I core domain (Stewart et al. 1997), the fluctuating molten globule state of topo6.3 appears to be an energy-favorable conformation for the free topo I C-terminal domain protein, as shown in the cartoon presentation in Figure 8A. The association of topo6.3 and the core domain readily results in topo I enzymatic activity (Stewart et al. 1997) and thus forms a minimal complex for enzymatic function (Champoux 2001), most likely through rigidification of both the domain structures. The fluctuating molten globule state may provide an efficient starting point for the folding of the C-terminal domain into its rigid functional conformation with a minimal energetic penalty. Thus the molten globule state identified here for the topo I C-terminal may represent a folding intermediate of this region in the intact topo I or in its complex with the core domain. Estimates of backbone and side chain entropy (Doig and Sternberg 1995; Thompson et al. 2002) suggest that backbone entropy accounts for approximately 60% of the entropy lost in protein folding, thus reduction in the backbone conformational entropy in the molten globule could account for a significant fraction of the total entropy loss and could substantially affect the energetics of binding. While the native-like conformation of the topo6.3 can significantly reduce the entropic cost for association with the core domain, the dynamic fluctuation of the molten globule state may reduce the energy of the C-terminal domain and provide a means to reinforce the specific hydrophobic interactions between the C-terminal and the core domains. The entropic loss associated with rigidification of both the C-terminal and core domains is presumably paid by the extensive hydrophobic interactions between the two complementary domains (Figure 8B).

Interestingly, different direct and indirect evidence suggest that structural fluctuations still exist to a certain extent after association of the C-terminal domain and the core domain, and even in the intact topo I proteins. For example, successful crystallizations of topo I proteins have only been obtained in the presence of a topo I-specific DNA (Redinbo et al. 1998; Stewart et al. 1998; Redinbo et al. 1999; Staker et al. 2002). The linker domain residues are typically disordered in the complex structures, even for the covalent complex of topo I and DNA (Stewart et al. 1998). In a different study, complexes of topo I and DNA have been crystallized in multiple non-isomorphous structures, and distinct regions of structural flexibility have been revealed in those structures (Redinbo et al. 1999). In addition, the structure flexibility of topo I protein is clearly shown in the molecular dynamics simulation study (Chillemi et al. 2001), which indicates that the DNA-protein contacts are made by a relatively flexible surface. In fact, the free topo I proteins, including the intact topo I protein and the complex protein of the C-terminal/core domain have all been difficult to crystallize, while crystallizations of topo I proteins can only be successful in the presence of a topo I-specific DNA, such as a specific sequence in *Tetrahymena* r-chromatin (Bonven et al. 1985), presumably due to the enhanced

rigidification of the overall tertiary structure of the topo I protein upon DNA binding (Figure 8 B-C). Indeed, the topo I protein becomes further ordered in a ternary complex with the presence of a topotecan drug compound, which traps the topo I protein in the covalent DNA-bound form (Staker et al. 2002).

During the catalytic mechanism, the intact topo I as well as the complex of core and C-terminal domains go through large conformational changes from an open conformation to a close conformation (DNA-bound) without any need for energy cofactors. Thus the conformational flexibility observed by us in the C-terminal domain, as well as similar possible flexibility in other domains, may be biologically important for facilitating such conformational changes with reduced energy requirements. The structural fluctuation and plasticity may represent an efficient mechanism in the topo I functional pathway, where the extent of the flexibility directs the formation of a fully productive topo I complex and provides an exquisite control of topo I function and specificity. Such flexibility transitions have been observed in many DNA binding proteins, where folding of locally disordered segments accompanies DNA binding of these proteins; moreover, the degree of folding and rigidification is enhanced when the protein binds to a specific, as opposed to less- or non-specific, DNA substrate (Spolar and Record 1994; Kalodimos et al. 2004; Pufall et al. 2005). A fluctuating structure can be more advantageous in many cellular processes, such as enzyme catalysis (Uversky et al. 1996; Vamvaca et al. 2004), allostery (Kern and Zuiderweg 2003), ribosomal complex assembly (Zurdo et al. 1997), and membrane transport (Bose et al. 1999). Facilitation of DNA binding appears to be a predominant molecular function of fluctuating proteins (Fink 2005). In a recently reported so-called affibody protein engineering (Wahlberg et al. 2003), the selected high affinity affibody-binding protein is a molten globule and can only achieve rigid tertiary structure upon binding of the target protein. The inherent flexibility of this molten globule state was suggested to enhance the high-affinity protein binding (Regan 2003), similar to the mechanism proposed here for the binding of topo6.3 protein to the topo I core domain.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was supported by the National Institutes of Health (1K01CA83886 and 1S10 RR16659) and Arizona Disease Control Research Commission (ADCRC). We are very grateful for Dr. James Champoux for his invaluable support and advices. We thank Drs. William R. Montfort and John J. Osterhout for their insightful discussions and comments.

References

- Bonven BJ, Gocke E, Westergaard O. A high affinity topoisomerase I binding sequence is clustered at DNAase I hypersensitive sites in Tetrahymena R-chromatin. *Cell* 1985;41(2):541–51. [PubMed: 2985282]
- Bose HS, Whittall RM, Baldwin MA, Miller WL. The active form of the steroidogenic acute regulatory protein, StAR, appears to be a molten globule. *Proceedings of the National Academy of Sciences of the United States of America* 1999;96(13):7250–7255. [PubMed: 10377400]
- Buck M. Trifluoroethanol and colleagues: cosolvents come of age. *Recent studies with peptides and proteins. Quarterly Reviews of Biophysics* 1998;31(3):297–355. [PubMed: 10384688]
- Champoux JJ. DNA topoisomerases: structure, function, and mechanism. *Annual Review of Biochemistry* 2001;70:369–413.
- Chillemi G, Castrignano T, Desideri A. Structure and hydration of the DNA-human topoisomerase I covalent complex. *Biophysical Journal* 2001;81(1):490–500. [PubMed: 11423431]
- Demarest SJ, Boice JA, Fairman R, Raleigh DP. Defining the core structure of the alpha-lactalbumin molten globule state. *Journal of Molecular Biology* 1999;294(1):213–21. [PubMed: 10556040]

- Doig AJ, Sternberg MJE. Side-Chain Conformational Entropy in Protein-Folding. *Protein Science* 1995;4 (11):2247–2251. [PubMed: 8563620]
- Dyson HJ, Merutka G, Waltho JP, Lerner RA, Wright PE. Folding of Peptide-Fragments Comprising the Complete Sequence of Proteins - Models for Initiation of Protein Folding .1. Myohemerythrin. *Journal of Molecular Biology* 1992;226(3):795–817. [PubMed: 1507227]
- Fink AL. Natively unfolded proteins. *Current Opinion in Structural Biology* 2005;15(1):35–41. [PubMed: 15718131]
- Garemyr R, Elofsson A. Study of the electrostatics treatment in molecular dynamics simulations. *Proteins-Structure Function and Genetics* 1999;37(3):417–428.
- Hsiang YH, Liu LF. Identification of mammalian DNA topoisomerase I as an intracellular target of the anticancer drug camptothecin. *Cancer Research* 1988;48(7):1722–6. [PubMed: 2832051]
- Jaxel C, Kohn KW, Wani MC, Wall ME, Pommier Y. Structure-Activity Study of Camptothecin Derivatives on Mammalian Topoisomerase I: Evidence for a Specific Receptor Site and a Relation to Antitumor Activity. *Cancer Research* 1989;49:5077–5082. [PubMed: 2548710]
- Kalodimos CG, Biris N, Bonvin A, Levandoski MM, Guennuegues M, Boelens R, Kaptein R. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* 2004;305 (5682):386–389. [PubMed: 15256668]
- Kern D, Zuiderweg ERP. The role of dynamics in allosteric regulation. *Current Opinion in Structural Biology* 2003;13(6):748–757. [PubMed: 14675554]
- Kuwajima K. The Molten Globule State as a Clue for Understanding the Folding and Cooperativity of Globular-Protein Structure. *Proteins-Structure Function and Genetics* 1989;6(2):87–103.
- Ladokhin, AS. Fluorescence spectroscopy in peptide and protein analysis. In: Meyers, RA., editor. *Encyclopedia of Analytical Chemistry*. Chichester: John Wiley & Sons Ltd; 2000. p. 5762-5779.
- Li YH, Jing GZ. Double point mutant F34W/W140F of staphylococcal nuclease is in a molten globule state but highly competent to fold into a functional conformation. *Journal of Biochemistry* 2000;128 (5):739–744. [PubMed: 11056385]
- Maple JR, Hwang MJ, Jalkanen KJ, Stockfisch TP, Hagler AT. Derivation of class II force fields: V. Quantum force field for amides, peptides, and related compounds. *Journal of Computational Chemistry* 1998;19(4):430–458.
- Neyroz P, Zambelli B, Ciurli S. Intrinsically disordered structure of *Bacillus pasteurii* UreG as revealed by steady-state and time-resolved fluorescence spectroscopy. *Biochemistry* 2006;45(29):8918–8930. [PubMed: 16846235]
- Orozco M, Laughton CA, Herzyk P, Neidle S. Molecular-Mechanics Modeling of Drug-DNA Structures - the Effects of Differing Dielectric Treatment on Helix Parameters and Comparison with a Fully Solvated Structural Model. *Journal of Biomolecular Structure & Dynamics* 1990;8(2):359–373. [PubMed: 2176506]
- Park SH, Mrse AA, Nevzorov AA, Mesleh MF, Oblatt-Montal M, Montal M, Opella SJ. Three-dimensional structure of the channel-forming trans-membrane domain of virus protein “u” (Vpu) from HIV-1. *Journal of Molecular Biology* 2003;333(2):409–424. [PubMed: 14529626]
- Peng, JW.; Wagner, G. Nuclear Magnetic Resonance, Pt C. 239. San Diego: Academic Press Inc; 1994. Investigation of Protein Motions Via Relaxation Measurements; p. 563-596.
- Ptitsyn OB. How the Molten Globule Became. *Trends in Biochemical Sciences* 1995;20(9):376–379. [PubMed: 7482708]
- Pufall MA, Lee GM, Nelson ML, Kang HS, Velyvis A, Kay LE, McIntosh LP, Graves BJ. Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science* 2005;309 (5731):142–145. [PubMed: 15994560]
- Ramboarina S, Redfield C. Structural characterisation of the human alpha-lactalbumin molten globule at high temperature. *Journal of Molecular Biology* 2003;330(5):1177–1188. [PubMed: 12860137]
- Redfield C, Schulman BA, Milhollen MA, Kim PS, Dobson CM. alpha-lactalbumin forms a compact molten globule in the absence of disulfide bonds. *Nature Structural Biology* 1999;6(10):948–952.
- Redinbo MR, Stewart L, Champoux JJ, Hol WGJ. Structural flexibility in human topoisomerase I revealed in multiple non-isomorphous crystal structures. *Journal of Molecular Biology* 1999;292(3):685–696. [PubMed: 10497031]

- Redinbo MR, Stewart L, Kuhn P, Champoux JJ, Hol WG. Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. *Science* 1998;279(5356):1504–13. [PubMed: 9488644]see comment
- Regan L. Molten globules move into action. *Proceedings of the National Academy of Sciences of the United States of America* 2003;100(7):3553–3554. [PubMed: 12657734]
- Rossi F, Labourier E, Gallouzi IE, Derancourt J, Allemand E, Divita G, Tazi J. The C-terminal domain but not the tyrosine 723 of human DNA topoisomerase I active site contributes to kinase activity. *Nucleic Acids Research* 1998;26(12):2963–70. [PubMed: 9611242]
- Schulman BA, Kim PS, Dobson CM, Redfield C. A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nature Structural Biology* 1997;4(8):630–634.
- Semisotnov GV, Rodionova NA, Razgulyaev OI, Uversky VN, Gripas AF, Gilmanshin RI. Study of the Molten Globule Intermediate State in Protein Folding by a Hydrophobic Fluorescent-Probe. *Biopolymers* 1991;31(1):119–128. [PubMed: 2025683]
- Song M, Shao H, Mujeeb A, James TL, Miller WL. Molten-globule structure and membrane binding of the N-terminal protease-resistant domain (63-193) of the steroidogenic acute regulatory protein (StAR). *Biochemical Journal* 2001;356(Pt 1):151–8. [PubMed: 11336647]
- Song MS, Shao HY, Mujeeb A, James ML, Miller WL. Molten-globule structure and membrane binding of the N-terminal protease-resistant domain (63-193) of the steroidogenic acute regulatory protein (StAR). *Biochemical Journal* 2001;356:151–158. [PubMed: 11336647]
- Spolar RS, Record MT. Coupling of Local Folding to Site-Specific Binding of Proteins to DNA. *Science* 1994;263(5148):777–784. [PubMed: 8303294]
- Staker BL, Feese MD, Cushman M, Pommier Y, Zembower D, Stewart L, Burgin AB. Structures of three classes of anticancer agents bound to the human topoisomerase I-DNA covalent complex. *Journal of Medicinal Chemistry* 2005;48(7):2336–45. [PubMed: 15801827]
- Staker BL, Hjerrild K, Feese MD, Behnke CA, Burgin AB Jr, Stewart L. The mechanism of topoisomerase I poisoning by a camptothecin analog. *Proceedings of the National Academy of Sciences of the United States of America* 2002;99(24):15387–92. [PubMed: 12426403]
- Stewart L, Ireton GC, Champoux JJ. The domain organization of human topoisomerase I. *Journal of Biological Chemistry* 1996;271(13):7602–8. [PubMed: 8631794]
- Stewart L, Ireton GC, Champoux JJ. Reconstitution of human topoisomerase I by fragment complementation. *Journal of Molecular Biology* 1997;269(3):355–72. [PubMed: 9199405]
- Stewart L, Ireton GC, Parker LH, Madden KR, Champoux JJ. Biochemical and biophysical analyses of recombinant forms of human topoisomerase I. *Journal of Biological Chemistry* 1996;271(13):7593–601. [PubMed: 8631793]
- Stewart L, Redinbo MR, Qiu X, Hol WG, Champoux JJ. A model for the mechanism of human topoisomerase I. *Science* 1998;279(5356):1534–41. [PubMed: 9488652]see comment
- Terpe K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology* 2003;60(5):523–533. [PubMed: 12536251]
- Thompson JB, Hansma HG, Hansma PK, Plaxco KW. The backbone conformational entropy of protein folding: Experimental measures from atomic force microscopy. *Journal of Molecular Biology* 2002;322(3):645–652. [PubMed: 12225756]
- Uversky VN, Gillespie JR, Millett IS, Khodyakova AV, Vasiliev AM, Chernovskaya TV, Vasilenko RN, Kozovskaya GD, Dolgikh DA, Fink AL, et al. Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH. *Biochemistry* 1999;38(45):15009–15016. [PubMed: 10555983]
- Uversky VN, Kutysenko VP, Protasova NY, Rogov VV, Vassilenko KS, Gudkov AT. Circularly permuted dihydrofolate reductase possesses all the properties of the molten globule state, but can resume functional tertiary structure by interaction with its ligands. *Protein Science* 1996;5(9):1844–1851. [PubMed: 8880908]
- Vamvaca K, Vogeli B, Kast P, Pervushin K, Hilvert D. An enzymatic molten globule: Efficient coupling of folding and catalysis. *Proceedings of the National Academy of Sciences of the United States of America* 2004;101(35):12860–12864. [PubMed: 15322276]

- Venyaminov, SY.; Yang, JT. Determination of protein secondary structure. In: Fasman, GD., editor. Circular Dichroism and the conformational analysis of biomolecules. New York: Plenum Press; 1996. p. 69-107.
- Wahlberg E, Lendel C, Helgstrand M, Allard P, Dincbas-Renqvist V, Hedqvist A, Berglund H, Nygren PA, Hard T. An affibody in complex with a target protein: Structure and coupled folding. Proceedings of the National Academy of Sciences of the United States of America 2003;100(6):3185–3190. [PubMed: 12594333]
- Wang JC. DNA topoisomerases. Annual Review of Biochemistry 1996;65:635–92.
- Wishart DS, Sykes BD, Richards FM. The Chemical-Shift Index - a Fast and Simple Method for the Assignment of Protein Secondary Structure through Nmr-Spectroscopy. Biochemistry 1992;31(6): 1647–1651. [PubMed: 1737021]
- Zhang XC, Zhang JH, Li X, Xu JJ, Huang HD, Chen Q, Wu JH, Shi YS. Compact molten globule-like state of hUBF HMG Box1 at extremely low pH. Biochimica Et Biophysica Acta-Proteins and Proteomics 2005;1748(1):66–73.
- Zurdo J, Sanz JM, Gonzalez C, Rico M, Ballesta JPG. The exchangeable yeast ribosomal acidic protein YP2[beta] shows characteristics of a partly folded state under physiological conditions. Biochemistry 1997;36(31):9625–9635. [PubMed: 9236009]

Figure 1.

Models of the structure of topoisomerase I C-terminal domain (topo6.3 protein) extracted from the crystal structure (1K4T) of topo70-DNA-topotecan ternary complex (Staker et al. 2002). The amino acid sequence of the topo6.3 fragment is shown, including three additional residues G, I and P that are added to the N-terminus from the expression vector. The amino acids involved in helix conformation are underlined; the amino acids of the five potential hydrophobic regions are shown in red. (A) Schematic representation of the topo I C-terminal domain. (B) A ribbon structure of the topo I C-terminal domain with its surface represented by dotted display. Each residue is colored according to its hydrophobicity, as indicated by the color code bar with relative hydrophobicity value: red (negative number) represents high hydrophobicity and blue (positive number) represents high hydrophilicity. Five potential hydrophobic regions are indicated with numbers. (C) A view of the topo6.3-5K protein with tryptophan residues shown in yellow. (D) The interface between the topo I C-terminal domain and the core domain. The C-terminal domain is shown in solid surface representation color-coded by relative hydrophobicity as indicated in (B). The core domain is shown in white dotted surface representation. The extensive hydrophobic interactions between the two domains are clearly shown. The DNA is shown in green tubes.

Figure 2. Far-UV CD spectra of topo6.3 proteins, (A) wild-type topo6.3, (B) C-terminally modified topo6.3-5K, and (C) N-terminally modified 5K-topo6.3. Each protein was used at a concentration of 15-20 μ M in PBS.

Figure 3.

Thermal denaturation curve of C-terminally modified topo 6.3 protein (topo6.3-5K) obtained by measuring the ellipticity at 222 nm. 20 μ M protein was used in PBS. The temperature was changed from 20 °C to 90 °C in 1 °C steps.

Figure 4.

Fluorescence emission spectra of 1-anilinonaphthalene-8-sulfonic acid (ANS) in the presence (solid lines) and absence (dotted lines) of topo6.3 proteins: (A) wild-type topo6.3, (B) C-terminally modified topo6.3-5K, and (C) N-terminally modified 5K-topo6.3. Results shown are from one experiment of three performed. A 4 μ M protein solution in PBS was used for each experiment.

Figure 5. The ^{15}N - ^1H HSQC spectra of the C-terminally modified topo6.3-5K protein at a concentration of 0.9 mM in pH 6 phosphate buffer (A) and in 8 M urea solution (B). The side-chain crosspeaks of Gln/Asn residues (indicated by arrows) are only shown in (A), as this region of spectrum is noisy for the urea sample.

Figure 6.

(A) Tryptophan fluorescence spectra of the topo6.3-5K protein in the denatured form (solid line, in 8 M urea) and non-denatured form (dotted line, in water). The non-denatured form of the topo6.3-5K protein displays a λ_{max} shift of approximately 15 nm to a lower wavelength. (B) Near-UV CD spectra of the topo6.3-5K protein (0.6 mM) in phosphate buffer and in 8 M urea.

Figure 7. The molecular models of the C-terminally modified topo6.3 protein after molecular dynamics calculation. The hydrophobicity surface of the topo6.3-5K protein with (left) and without (right) the five C-terminal lysine residues. It is clearly shown that a significant hydrophobic surface area is covered by the five C-terminal lysine residues. The residue is colored according to its hydrophobicity, as indicated by the color code bar with relative hydrophobicity value. The blue represents high hydrophilicity, while the red represents high hydrophobicity.

Figure 8.

Cartoon presentation of the proposed pathway for the formation of a productive topo I complex. Upon complexation with the core domain, the flexible C-terminal domain (A) becomes better ordered and more rigid, which demonstrates enzymatic activity and DNA binding capability. However, structure fluctuation still exists to a certain extent in the topo I protein (B), which may explain the difficulty of crystallizing topo I proteins alone. Upon DNA binding, the rigidity of the overall tertiary structure of topo I protein is further enhanced (C), which leads to successful crystallizations. The structural fluctuation and plasticity may thus represent an efficient means to direct the formation of a fully productive topo I complex and to provide an exquisite control of topo I function and specificity.