Disordered N-terminal residues affect the folding thermodynamics and kinetics of maltose binding protein

C. Ganesh[a], Antara Banerjee[a], Aseema Shah[a], Raghavan Varadarajan[a,b,*]

[a] *Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India*
[b] *Chemical Biology Unit, Jawaharlal Nehru Center for Advanced Scientific Research, Jakkur, P.O. Bangalore 560 004, India*

Received 16 March 1999; received in revised form 17 May 1999

Abstract Maltose binding protein (MBP) exhibits a slow phase of folding at pH 7.4, 298 K. The kinetics of this phase has been characterized as a function of denaturant concentration and temperature. Denaturant double-jump experiments and the activation energy for folding indicate that the slow phase involves processes other than proline isomerization. Although the first five N-terminal residues are disordered in the MBP crystal structure, mutations in this region slow down folding and destabilize the native structure. This is the first report showing that disordered N-terminal residues can affect folding kinetics and stability.
© 1999 Federation of European Biochemical Societies.

Key words: Folding; Kinetics; Stability; Maltose binding protein; Double jump; Mass spectrometry

1. Introduction

Maltose binding protein (MBP) is a large, two domain protein of 370 amino acid residues, present in the periplasm of *Escherichia coli*. The folding thermodynamics of MBP has been extensively characterized [1,2]. The kinetics of folding has also been characterized using manual mixing [3]. MBP serves as a good model system to study folding of large proteins but its folding is complex, involving several phases. The slow phase observed during the folding of several proteins has been attributed to isomerization about the X-Pro peptide bond in the polypeptide backbone. This typically occurs with a rate constant of about 10^{-3} s^{-1} [4]. There are 21 proline residues in MBP, all in the *trans*-conformation in the folded structure [5]. The slow phase of MBP folding is further retarded by the *E. coli* chaperone SecB, which is involved in the translocation of MBP across the bacterial inner membrane in vivo [6].

In the present study, the slow phase of MBP folding has been investigated by fluorescence spectroscopy under different folding conditions to assess the involvement of proline residues in the folding pathway. In order to obtain larger amounts of protein than those obtained through periplasmic expression, MBP was expressed in the cytoplasm using a derivative of the plasmid $pMalc_2$. The cytoplasmic (cmMBP) and periplasmic (pmMBP) proteins differ only at two N terminal positions which are disordered in the crystal structure of pmMBP. Surprisingly, cmMBP is destabilized and folds more slowly than pmMBP. This is the first report showing that disordered N-terminal residues can affect folding kinetics and stability.

*Corresponding author. Fax: (91) (80) 3348535, (91) (80) 3341683.
E-mail: varadar@mbu.iisc.ernet.in

Abbreviations: MBP, maltose binding protein; CGH, equimolar citrate-glycine-HEPES; MS, mass spectrometry; ESI, electrospray ionization; TOF, time of flight; MALDI, matrix assisted laser desorption ionization; cm, cytoplasmic mature protein; pm, periplasmic mature protein; k, rate constant; f, folding; u, unfolding; ΔG, free energy change; C_m, denaturant concentration at which half the molecules are unfolded; E_a, activation barrier; LA, lactalbumin

2. Materials and methods

2.1. Materials

Ultra-pure guanidinium hydrochloride, HEPES and urea were obtained from USB and ultra-pure Tris was from Gibco BRL. ANS, IPTG, maltose and PMSF were from Sigma. Either deionized, double distilled water or milliQ water was used for preparing all solutions. A stock solution of 0.1 M ANS was prepared in water and filtered just prior to use. ANS concentration was estimated using $\varepsilon_{350} = 5000$ M^{-1} cm^{-1} [7]. Buffers CGH10 and CGH1 were prepared as described earlier [2]. Pure acetonitrile was distilled and all solvents for HPLC were filtered and degassed just prior to use. All reverse-phase HPLC runs were carried out using a VYDAC C_{18} semi-preparative column. Pure cyanogen bromide (CNBr) was obtained locally and stored under cold, dry conditions. All other reagents were of analytical grade purity and obtained locally.

2.2. Cloning, protein purification and chemical cleavage

Mature MBP was directly purified from the *E. coli* periplasm as described earlier [2]. Mature MBP, lacking the leader sequence, was also separately cloned and overexpressed in *E. coli*, using the fusion protein vector $pMALc_2$ from New England Biolabs, Massachusetts, USA. $pMALc_2$ was double digested with *Bgl*II and *Hin*dIII and the resultant large fragment purified. Vector pBSKS$^+$malE was constructed by ligating the 1738 bp *Eco*RI-*Hin*dIII fragment of M13malE containing the wild-type MBP coding sequence into pBSIIKS$^+$. pBSKS$^+$malE was double digested with *Bgl*II and *Hin*dIII and the small fragment obtained at this step was ligated to the large fragment from $pMALc_2$. The resultant vector was pMALcmMBP and it was used for cytoplasmic expression of mature MBP. In this construct, amino acids 1–144 are from $pMALc_2$ and 145–370 are derived from M13malE.

Overexpressed MBP was obtained in a highly soluble form from *E. coli* cytosol using the above vector, and is referred to as cytoplasmic mature MBP (cmMBP). This nomenclature is used to differentiate it from the authentic wild-type periplasmic mature protein (pmMBP). cmMBP was purified from *E. coli* DH5α cells harboring the plasmid pMALcmMBP. cmMBP was extracted using a procedure described for another cytosolic protein [8] and purified by the same procedures used for pmMBP purification [2]. Concentrated stocks of proteins were stored at −70°C. MBP was chemically cleaved at its methionine residues by cyanogen bromide, as described [9]. Cyanogen bromide was used for analytical purposes only and not during any step of protein purification. Pure proteins and fragments were analyzed by SDS-PAGE, HPLC and MS.

2.3. Spectroscopic measurements and data analyses

All fluorescence measurements were made in a JASCO FP777 spectrofluorimeter fitted with a thermostatted cuvette holder. All temperatures reported here are that of the sample and have an accuracy of ±0.1°C. Three to six individual measurements were made for all experiments and the mean and standard deviation are reported. Excitation was at 280 nm and the emission was measured either at 341 nm

0014-5793/99/$20.00 © 1999 Federation of European Biochemical Societies. All rights reserved.
PII: S0014-5793(99)00826-1

(for folding kinetics) or scanned (for equilibrium experiments) at 100 nm min^{-1} with a time constant of 4 s. ANS binding experiments were performed using an excitation of 380 nm and emission scanned from 425 to 575 nm. Excitation and emission slit widths were 1.5 and 5 nm respectively in all cases. Refolding kinetics experiments were performed as described in [6]. Urea denaturation studies were performed at pH 7.4 in the presence of 0.15 M potassium chloride and all samples incubated until equilibria were established under the given set of conditions. For ANS binding experiments, 3 μM ANS was used. All protein concentration estimations, maltose binding titration and CD measurements were performed as described in [2].

Equilibrium unfolding was carried out as described previously [2]. The data were fitted to a two-state model by a non-linear least squares fitting procedure [10] to obtain ΔG_D, the free energy change upon unfolding as a function of denaturant concentration. The linear extrapolation model [10] was used to obtain $\Delta G°$ and the m value (the intercept and slope respectively of a plot of ΔG_D vs. denaturant concentration) for the protein. C_m (the denaturant concentration at which ΔG_D is zero) is given by the ratio $\Delta G°/m$. Refolding and unfolding kinetic traces were both fitted to a single exponential as $F(t)=a[1-\exp(-k\ t)]+b$ and $F(t)=a\ \exp(-k\ t)+b$ respectively. The dead time due to manual mixing was 6–8 s. In all the manual mixing studies reported here, the refolding kinetics were well described by a single exponential. All non-linear least squares fitting routines were unconstrained fits.

2.4. Analytical techniques

SDS-PAGE runs were carried out as described in [2]. PAGE under native conditions was performed at 7–10°C inside a cold room, using a buffer of pH 8.8. ESI-MS was performed on an HP1100 MSD and MALDI measurements made using a TOF KOMPACT SEQ machine from Kratos. The error in reported mass measurements by ESI-MS is about 1 Da in 10 kDa. Automated N-terminal sequencing was performed for both the proteins for several rounds.

3. Results

The refolding of both pm- and cmMBP is highly reversible, as judged from the final fluorescence yield and maltose binding characteristics of the refolded material (data not shown). The individual rates and amplitudes for folding and unfolding under different conditions are shown in Table 1. Examination of the data indicates that the two proteins differ only in their refolding rates while their unfolding rates are identical. The relative amplitudes of the fast and slow phases in both unfolding and refolding are identical in the two proteins. Both pm- and cmMBP exhibit similar affinity for binding maltose (data not shown).

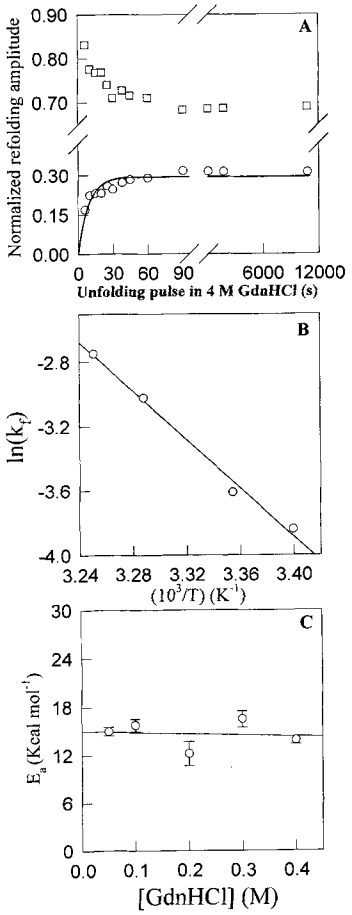

Fig. 1. A: Double-jump refolding experiment for pmMBP (4 M to 0.2 M GdnHCl jump). Shown are the changes in slow ( ) and fast phase (□) amplitudes as a function of increasing unfolding pulse. Also shown (solid line) is the single exponential fit for the rise in slow phase amplitude. B: Arrhenius plot for pmMBP refolding in 0.05 M GdnHCl. C: E_a values for the slow phase of pmMBP refolding at different denaturant concentrations. Experimental conditions similar to that mentioned under A.

Table 1
Kinetic parameters for folding and unfolding of MBP at 298 K, pH 7.4

[GdnHCl] (M)		a (normalized)	k (s^{-1})	b (normalized)
Unfolding				
1.50	pmMBP	0.700 (±0.005)	0.0038 (±0.001)	0.300 (±0.005)
	cmMBP	0.658 (±0.020)	0.0037 (±0.0001)	0.342 (±0.02)
1.75	pmMBP	0.710 (±0.003)	0.0126 (±0.002)	0.290 (±0.003)
	cmMBP	0.681 (±0.003)	0.0131 (±0.003)	0.319 (±0.003)
1.85	pmMBP	0.714 (±0.001)	0.0249 (±0.005)	0.286 (±0.001)
	cmMBP	0.685 (±0.01)	0.0194 (±0.006)	0.315 (±0.01)
2.00	pmMBP	0.722 (±0.015)	0.0495 (±0.001)	0.278 (±0.015)
	cmMBP	0.677 (±0.005)	0.0444 (±0.002)	0.323 (±0.005)
Folding				
0.1	pmMBP	0.382 (±0.003)	0.0317 (±0.003)	0.618 (±0.003)
	cmMBP	0.385 (±0.010)	0.0235 (±0.0013)	0.615 (±0.010)
0.2	pmMBP	0.394 (±0.004)	0.0265 (±0.002)	0.606 (±0.004)
	cmMBP	0.397 (±0.002)	0.0191 (±0.0007)	0.603 (±0.002)
0.3	cmMBP	0.416 (±0.007)	0.0141 (±0.0004)	0.584 (±0.007)
0.4	pmMBP	0.423 (±0.008)	0.0174 (±0.001)	0.577 (±0.008)
	cmMBP	0.423 (±0.011)	0.0119 (±0.0006)	0.577 (±0.011)
0.5	pmMBP	0.433 (±0.008)	0.012 (±0.0003)	0.567 (±0.008)

The slow phase of pmMBP folding was further investigated by the double-jump technique [4]. In these experiments the protein was unfolded in 4 M GdnHCl for varying amounts of time and then transferred to refolding conditions. The relative amplitudes of both the fast and slow phases of refolding as a function of incubation time in 4 M GdnHCl are shown in Fig. 1A. The build-up of the slow phase occurs with a rate of 0.11 s^{-1} in 0.2 M GdnHCl and this rate depends on the final GdnHCl concentration in the folding solution (data not shown). The activation energy (E_a) for the slow phase was estimated by examining the temperature dependence of the refolding rate in the range 20–37°C and fitting the data to the Arrhenius equation $k = A \exp(-E_a/RT)$ (Fig. 1B). Refolding experiments showed that E_a is relatively independent of denaturant concentration (Fig. 1C) and was estimated to be 15 ± 1.6 kcal mol^{-1}.

The secondary CD of cmMBP is shown and compared to that of pmMBP in Fig. 2A. The secondary structural content of cmMBP appears to be very similar to that of pmMBP [2]. ANS binding to pm- and cmMBP were examined at both pH 2.8 and pH 7.4 (data not shown); it was found that both the proteins bind ANS only at the lower pH and not at 7.4. It has been shown previously that pmMBP forms a molten globule below pH 3 [2] and cmMBP also forms a molten globule under these conditions. Taken together, these results suggest

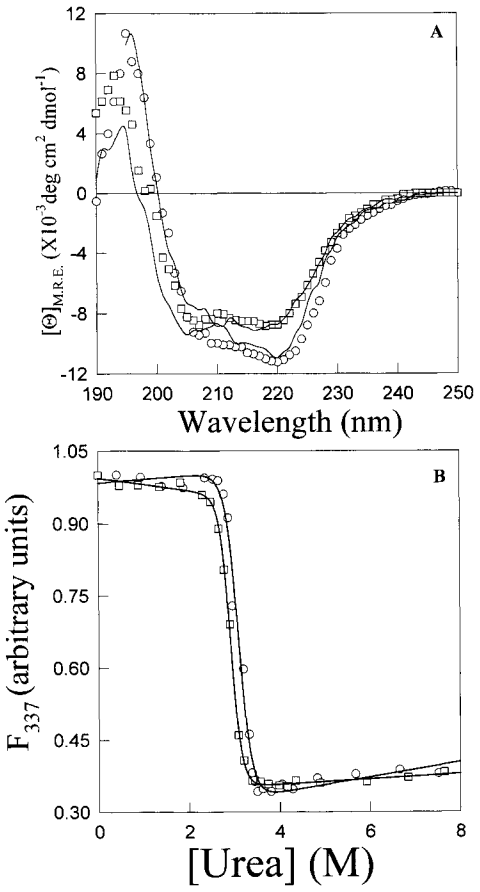

Fig. 2. A: Secondary CD of cmMBP at pH 7.4 and 2.8 at 21°C (○, □); continuous lines indicate CD of pmMBP under similar conditions. B: Urea denaturation; experimental data points for pm- (○), cmMBP (□) and (solid line) curves fitted for a two-state unfolding transition.

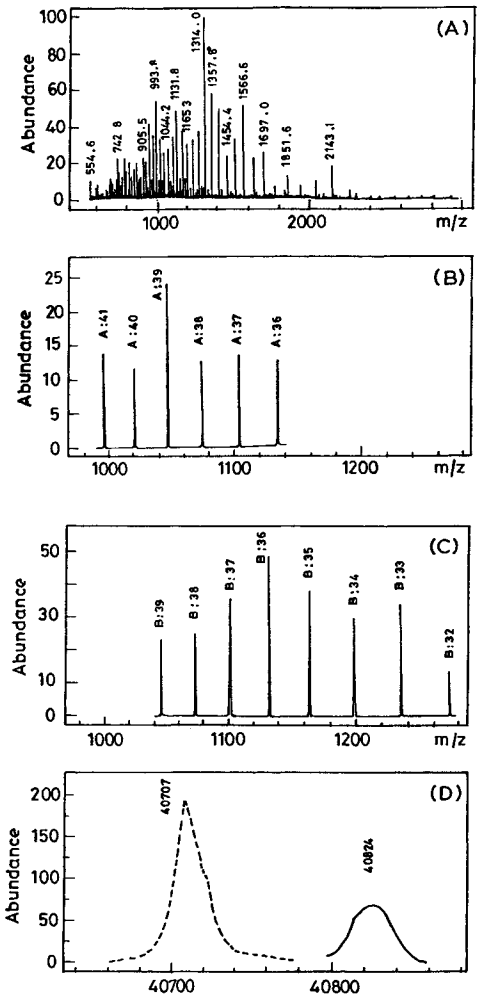

Fig. 3. ESI-MS of a 2:1 mixture of HPLC purified intact pm- and cmMBP. A: Actual charged states observed. B and C: Distribution of two distinct sets of charged states. D: Deconvoluted charge states yield only two unique masses, indicated in the figure.

that under native conditions, there may not be large-scale overall rearrangements in cmMBP compared to pmMBP but we do not rule out minor localized changes. The stability of both proteins was examined by performing equilibrium unfolding studies in urea. The values at pH 7.3 and 25°C for $\Delta G°$, m and C_m were 15.4 (±1.2) kcal mol^{-1}, 4.9 (±0.4) kcal mol^{-1} M^{-1} and 3.15 M for pmMBP and 11.9 (±0.9) kcal mol^{-1}, 4 (±0.3) kcal mol^{-1} M^{-1} and 2.96 M for cmMBP. Thus, cmMBP is destabilized at room temperature when compared to pmMBP.

The molecular weights of both pm- and cmMBP were accurately determined by ESI-MS (Fig. 3). Both the proteins were cleaved by CNBr, the resultant fragments purified by HPLC and analyzed by both ESI-MS and MALDI-MS. The results are presented in Table 2. Automated protein sequencing indicated the presence of the following sequence at the N-termini: pmMBP, KIEEGKLV and cmMBP, MKTEEGKL. These are consistent with the known DNA sequences of wild-type MBP and $pMALc_2$ respectively. Native PAGE (data not shown) indicated that both pmMBP and cmMBP exhibit similar electrophoretic mobility under native conditions. All the available data suggest that pmMBP and cmMBP differ only at two amino acids in the amino terminus of the protein. The

Table 2
ESI-MS analyses of pmMBP, cmMBP and CNBr digestion products

Protein	Fragment	Position	Expected MW (Da)	Observed MW (Da)
pmMBP	Intact	1–370	40707	40707
	1	1–148	16471	16471
	2	149–204	6221	6221
	3	205–321	12474	12474
	4	225–321	10403	10403
	5	322–330	971	971
	6	331–336	651	651
	7	331–370	4390	4390
	8	337–370	3739	3743
cmMBP	Intact	1–371	40826	40824
	1	2–149	16459	16464
	2	150–205	6221	6221
	3	206–322	12474	12476
	4	226–322	10403	10404
	5	323–331	971	971
	6	332–337	651	651
	7	338–371	3739	3739

first five N-terminal residues containing these two mutations are disordered in the crystal structure of pmMBP. Despite this, the mutations affect the refolding kinetics and stability of MBP.

4. Discussion

The rate constants and amplitudes for the refolding kinetics reported here show good agreement to those previously reported [3,6,11]. The slow phase of folding of pmMBP has been further investigated in this study. The value of E_a for the slow phase is 15 ± 1.6 kcal mol^{-1}. This is significantly different from that expected (20 kcal mol^{-1}) for proline *cis-trans* isomerization [4]. The observed rate for slow phase amplitude build-up estimated from double-jump experiments was found to be about 100 times faster than that expected for the isomerization process. Thus, the slow phase of folding in this large protein must involve structural rearrangements other than proline isomerization. Since the N-terminal mutant, cmMBP, folds more slowly than pmMBP, residues close to the N-terminus are likely to be involved in the slow phase of folding. This is consistent with an earlier observation that the N-terminal signal peptide significantly slows folding of MBP [12]. It is also possible, though unlikely, that while the N-terminal residues are disordered in the protein crystal structure, the situation may be different in solution. Alternatively the N-terminus may be ordered in cmMBP but not in pmMBP.

The changes in physicochemical properties of a protein due to the presence of the extra unprocessed methionine at the N-terminus have recently been reported for two small recombinant proteins expressed heterologously in *E. coli*, goat α-lactalbumin [13] and bovine α-lactalbumin [14]. Table 3 compares the large protein MBP with these two small proteins and it is evident that destabilization of the native structure is a common theme. The additional substitution of Ile in pmMBP by Thr in cmMBP occurred because this mutation is already present in the pMALc$_2$ plasmid (sequence obtained from the web-site URL http://www.neb.com) and has not occurred during the cloning process. The observed MW of cmMBP from ESI-MS (Fig. 3) agrees well with that calculated using the mature pmMBP sequence [15] corrected for both the Ile to Thr mutation and the additional N-terminal methionine residue. Analyses of the CNBr fragments from both pm- and cmMBP by MS (Table 2) show that there are no other differences in sequence between the two proteins.

Peptide deformylase from *E. coli* is highly specific towards formyl group removal. The enzyme is active towards diverse peptide sequences bearing the fMet residue [16]. On the other hand, the action of methionyl aminopeptidase depends on the size of the side chain of the residue adjacent to terminal methionine and methionyl removal is negligible when the side chain length of the adjacent residue exceeds 4 Å [17]. The presence of an extra methionine and not formyl methionine residue at the N-terminus of purified cmMBP indicates that deformylation has indeed occurred but not demethionylation. In the case of cmMBP, the second residue is lysine, whose side chain length is 6.4 Å and thus the presence of uncleaved methionine in cmMBP may be accounted for.

The commercially available pMALc2 vector has been widely used for high level overexpression of proteins in *E. coli* cytoplasm as MBP fusions [18]. Our results indicate that cmMBP is destabilized at room temperature. This factor should be considered, in addition to in vivo expression levels, in choosing whether pmMBP or cmMBP be used as a fusion partner. It is widely accepted that buried residues are impor-

Table 3
Effect of extra N-terminal methionine retained on overexpression in *E. coli*

Protein	N_{res}	Mobility in native gel	Stability	Kinetics		N-terminus first 4 residues
				Folding	Unfolding	
MBP	370	No change	Decreased	Slower	No change	Disordered
Goat α-LA	123	Greater	Decreased	No change	Faster	Ordered
Bovine α-LA	124	N.R.	Decreased	N.R.	N.R.	Ordered

N.R.: not reported.

tant determinants of protein stability [19]. However, the present work clearly shows that residues that are apparently in disordered regions of a protein may also affect folding kinetics and stability. A more complete explanation of this surprising observation must await structure determination of cmMBP.

Acknowledgements: We thank Prof. P. Balaram, S.S. Ray, S.K. Singh and S.R. Solomon at M.B.U., I.I.Sc. for help in use of the mass spectrometers. We are grateful to Dr. M. Hofnung, Pasteur Institut, Paris, France for providing pPD1 and M13malE. We also thank Kiran Bacchawat M.B.U., I.I.Sc. for help in use of $pMALc_2$ vector. R.V. acknowledges financial support from DBT and DST, Government of India.

References

[1] Novokhanty, V. and Ingham, K. (1997) Protein Sci. 6, 141–146.
[2] Ganesh, C., Shah, A.N., Swaminathan, C.P., Surolia, A. and Varadarajan, R. (1997) Biochemistry 36, 5020–5028.
[3] Liu, G., Topping, T.B., Cover, W. and Randall, L.L. (1988) J. Biol. Chem. 263, 14790–14793.
[4] Schmid, F.X. (1986) Methods Enzymol. 131, 70–82.
[5] Sharff, A.J., Rodseth, L.E., Spurlino, J.C. and Quiocho, F. (1992) Biochemistry 31, 10657–10663.
[6] Liu, G., Topping, T. and Randall, L.L. (1989) Proc. Natl. Acad. Sci. USA 86, 9213–9217.
[7] Stryer, L. (1965) J. Mol. Biol. 13, 482–495.
[8] Panse, V.G., Udgaonkar, J.B. and Varadarajan, R. (1998) Biochemistry 37, 14477–14483.
[9] Means, G.E. and Feeney, R.E. (1971) in: Chemical Modification of Proteins, Holden-Day, San Francisco, CA.
[10] Santoro, M.M. and Bolen, D.W. (1988) Biochemistry 27, 8063–8068.
[11] Raffy, S., Sassoon, N., Hofnung, M. and Betton, J.-M. (1998) Protein Sci. 7, 2136–2142.
[12] Park, S., Liu, G., Topping, T.B., Cover, W.H. and Randall, L.L. (1988) Science 239, 1033–1035.
[13] Chaudhuri, T.K., Horii, K., Yoda, T., Arai, M., Nagata, S., Terada, T.P., Uchiyama, H., Ikura, T., Tsumoto, K., Kataoka, H., Matsushima, M., Kuwajima, K. and Kumagai, I. (1999) J. Mol. Biol. 285, 1179–1194.
[14] Ishikawa, N., Chiba, T., Chen, L.T., Shimizu, A., Ikeguchi, M. and Sugai, S. (1998) Protein Eng. 11, 333–335.
[15] Duplay, P., Bedouelle, H., Fowler, A., Zabin, I., Saurin, W. and Hofnung, M. (1984) J. Biol. Chem. 259, 10606–10613.
[16] Hu, Y-J., Wei, Y., Zhou, Y., Ravi Rajagopalan, P.T. and Pei, D. (1999) Biochemistry 38, 643–650.
[17] Hirel, P.H., Schmitter, J-M., Dessen, P., Fayat, G. and Blanquet, S. (1989) Proc. Natl. Acad. Sci. USA 86, 8247–8251.
[18] Guan, C., Li, P., Riggs, P.D. and Inouye, H. (1987) Gene 67, 21–30.
[19] Lim, W.A. and Richards, F.M. (1993) Q. Rev. Biophys. 26, 423–498.