FEBS Letters 550 (2003) 18–22 FEBS 27516

Structural studies on a twin-arginine signal sequence

Marc Kipping[a], Hauke Lilie[b], Ute Lindenstrauß[c], Jan R. Andreesen[c], Christian Griesinger[d], Teresa Carlomagno[d,*], Thomas Brüser[c,**]

[a] *Max Planck Research Unit for Enzymology of Protein Folding, Weinbergweg 22, D-06120 Halle, Germany*
[b] *Institute of Biotechnology, University of Halle, Kurt-Mothes-Str. 3, D-06120 Halle, Germany*
[c] *Institute of Microbiology, University of Halle, Kurt-Mothes-Str. 3, D-06120 Halle, Germany*
[d] *Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany*

Received 19 June 2003; revised 7 July 2003; accepted 9 July 2003

First published online 25 July 2003

Edited by Stuart Ferguson

Abstract Translocation of folded proteins across biological membranes can be mediated by the so-called 'twin-arginine translocation' (Tat) system. To be translocated, Tat substrates require N-terminal signal sequences which usually contain the eponymous twin-arginine motif. Here we report the first structural analysis of a twin-arginine signal sequence, the signal sequence of the high potential iron-sulfur protein from *Allochromatium vinosum*. Nuclear magnetic resonance (NMR) analyses of amide proton resonances did not indicate a signal sequence structure. Accordingly, data from H/D exchange matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry showed that the amide protons of the signal sequence exchange rapidly, indicating the absence of secondary structure in the signal sequence up to L29. We conclude that the conserved twin-arginine motif does not form a structure by itself or as a result of intramolecular interactions.

Key words: High potential iron-sulfur protein; Twin-arginine translocation; Signal sequence; Protein translocation; Nuclear magnetic resonance; H/D exchange matrix-assisted laser desorption/ionization-time of flight mass spectrometry

1. Introduction

The translocation of proteins across biological membranes usually requires the existence of N-terminal signal sequences [1]. Bacterial signal sequences are generally constituted by a positively charged N-terminal (n-) region, followed by an uncharged hydrophobic (h-) region and in most cases by a C-terminal cleavage site (c-) region. Although signal sequences of secreted proteins are similar in their overall structure, they have features which direct proteins to either the general secretory (Sec), or to the twin-arginine translocation (Tat) systems [2,3]. Signal sequences of Tat substrates contain usually a 'twin-arginine motif' in their n-region [4]. In addition, they are relatively long, less hydrophobic in their h-region and often charged in their c-region [5,6]. Tat substrates can be translocated in a folded state [7–9]. As incorrectly folded proteins are not efficiently translocated, it has been proposed that a further determinant of Tat substrates is their folded state [3,10,11].

Due to the fact that Tat substrates fold prior to translocation, we analyzed whether their long and motif-containing signal sequences also acquire a folded state, possibly including intramolecular interactions with the folded mature domain. Such a folded signal sequence could play a role in maturation and targeting processes. Further, the essential twin-arginine motif in the signal sequence is specifically recognized by the Tat system and therefore this pattern has to be considered to fold to a Tat substrate determinant structure.

So far, no Tat substrate signal peptide structure has been characterized. The only published structural study on a Tat substrate, the glucose:fructose oxidoreductase precursor (preGFOR) from *Zymomonas mobilis*, did not result in insight into the signal sequence structure [12]. The reported X-ray analyses of preGFOR crystals did not resolve the electron density corresponding to the signal sequence and therefore it could not be distinguished whether the signal sequence was unstructured or whether it was just flexible in the crystal.

In this study, we report the first structural analysis of a Tat signal sequence in solution. We have chosen the high potential iron-sulfur protein (HiPIP) from *Allochromatium vinosum* as our model precursor protein, because the structure of its mature domain has been elucidated with nuclear magnetic resonance (NMR) and X-ray techniques [13,14]. The 37 residues long signal sequence of HiPIP contains the typical twin-arginine motif and the other characteristics of Tat substrate signal sequences (Fig. 1). It has been demonstrated that HiPIP functions as a Tat substrate in *Escherichia coli*, and that the twin-arginine motif is essential for the translocation [15]. Secondary structure prediction programs suggest an α-helical structure for the twin-arginine motif and parts of the h-region of the HiPIP signal sequence (Fig. 1). In contrast, our experimental data show that the twin-arginine signal sequence has an unfolded conformation in the presence of a completely and correctly folded mature domain. We conclude that the n- and h-regions of the signal sequence, including the twin-arginine motif, do not form a structure by itself or by intramolecular interactions.

*Corresponding author on NMR-related questions.
**Corresponding author.
E-mail addresses: taco@mpibpc.mpg.de (T. Carlomagno), t.brueser@mikrobiologie.uni-halle.de (T. Brüser).

Abbreviations: HiPIP, high potential iron-sulfur protein; Tat, twin-arginine translocation; NMR, nuclear magnetic resonance; MALDI-TOF, matrix-assisted laser desorption/ionization-time of flight; preGFOR, glucose:fructose oxidoreductase precursor

0014-5793/03/$22.00 © 2003 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/S0014-5793(03)00804-4

2. Materials and methods

2.1. Preparation of folded HiPIP precursor

Inclusion bodies of HiPIP precursor were produced with pEXH5 in *E. coli* BL21DE3, and fully folded protein was obtained as described previously [15]. ^{15}N-labeled protein was obtained by expression in $^{15}NH_4Cl$ substituted M9 minimal medium with trace elements added (SL12; [16]). For NMR analyses, folded HiPIP precursor (preHoloHiPIP) was dialyzed against 50 mM sodium phosphate buffer, pH 6.8. For mass spectrometry, the buffer was changed to 50 mM ammonium acetate, pH 6.8. Samples were concentrated by ultrafiltration centrifugation (10 kDa cutoff, Millipore) and checked for purity by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) (Coomassie stain), and by ultraviolet-visible (UV-VIS) spectroscopy. The A_{283nm}/A_{388nm} ratio of the samples was at about 2.6, indicating homogeneity and complete cofactor reconstitution. The concentration of the samples was calculated from their absorbance at 283 nm ($\varepsilon = 41.3\ mM^{-1}\ cm^{-1}$; [17]).

2.2. NMR methods

^{1}H-^{15}N 2D correlations (HSQC) were acquired on a Bruker DRX-600 spectrometer for HiPIP dissolved in a H_2O (95%)/D_2O (5%) solution, containing 50 mM sodium phosphate buffer, pH 6.8. The protein concentration was 700 μM. Data acquisition was performed at 10°C, because concentrated HiPIP precursor tends to aggregate at higher temperature. The spectra were processed with the FELIX software (MSI, San Diego, CA, USA).

2.3. Hydrogen exchange experiments and mass spectrometry

Kinetics of deuterium incorporation by exchange of amide protons of peptide bonds to deuterons in preHoloHiPIP was monitored by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry using a Bruker REFLEX II mass spectrometer (Bruker Daltonik GmbH, Germany). The exchange with deuterium was carried out at 0°C and initiated by a dilution of 1 μl of 250 μM protein in 50 mM ammonium acetate (pH 6.8) with 9 μl D_2O containing the same buffer. At appropriate time points, the reaction was stopped by shifting the pH to 2.4 with 90 μl aqueous quenching buffer (25 mM succinic acid, 25 mM citric acid). Subsequently, a peptic digest of the deuterated protein was initiated by adding 1 μl of 250 μM pepsin (Roche Diagnostics, Mannheim, Germany). After 1 min at 0°C 30 μl methanol were added and the sample was frozen at −80°C. For MALDI analysis a special preparation protocol was used to minimize artificial H/D exchange during preparation and measurement [18]. Using a thin layer preparation with α-cyano-4-hydroxycinnamic acid at −20°C and fast evaporation, it is possible to reduce artificial exchange to values smaller than 10%. Back-exchange controls were performed using fully deuterated HiPIP (250 μM protein in 50 mM ammonium acetate; 3 weeks equilibrated). After correction for detected back-exchange and 10% residual H_2O, data of deuterium incorporation of peptic HiPIP peptides were fitted to the equation $D(t) = A^*(1-e^{-k_1*t})+B^*(1-e^{-k_2*t})$ with Sigma Plot 2000 (Jandel Scientific).

3. Results

3.1. Conformational analysis of preHoloHiPIP by NMR spectroscopy

The folded mature domain of HiPIP gives rise to well-characterized amide proton resonances in ^{15}N-^{1}H HSQC spectra [13]. So far, these analyses have been carried out with mature HiPIP only, but not with the precursor form of HiPIP. In order to obtain information about the signal sequence structure, we recorded the ^{15}N-^{1}H HSQC spectrum of folded HiPIP precursor (preHoloHiPIP) and tried to obtain information about the signal sequence structure. Notably, the signals previously reported for mature HiPIP were also produced by the precursor, indicating that the mature domain of HiPIP precursor is folded like in processed HiPIP (Fig. 2). Only small differences were detectable which can be attributed to the experimental conditions. Comparison of the precursor spectrum with the reported resonances from mature HiPIP allowed the identification of about 32 new signals which could be assigned to signal sequence resonances (indicated in Fig. 2 in black color). Note that two prolines are present in the signal sequence and that the N-terminal methionine is cleaved off (see below). Importantly, almost all ^{1}H signal shifts from signal sequence residues were found in the region where unstructured ^{1}H resonances are expected, i.e. within 8.5 and 7 ppm. Similarly, the corresponding ^{15}N shifts did not indicate any folded structure within the signal sequence.

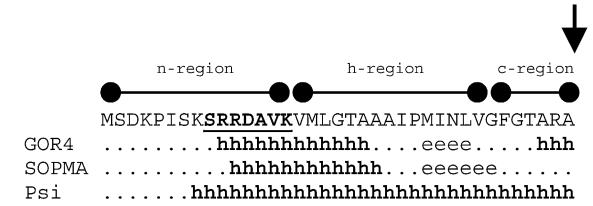

Fig. 1. The signal sequence of the HiPIP precursor. The twin-arginine motif (underlined bold) is located in the hydrophilic n-region adjacent to the h-region. The cleavage region (c-region; the arrow indicates the cleavage site) contains a charged residue (arginine), which is typical for Tat signal sequences. Secondary structure predictions with GOR IV [24], SOPMA [25], and PSIPRED [26] are indicated below the signal sequence. Symbols: h, α-helical; e, β-strand; dots, residues of random structures (coils).

3.2. Analysis of the signal sequence by hydrogen exchange mass spectrometry

To further analyze the structure formation of the HiPIP signal sequence, we applied H/D exchange mass spectrometry. This method is a valuable tool for the measurement of the exchange of amide protons within peptide bonds of defined protein regions. In a typical H/D exchange experiment, the protein of interest is subjected to D_2O. After desired incubation times, the exchange reaction is stopped by decreasing the deuterium concentration and switching the pH to 2.4. Subsequently, the protein is digested with the protease pepsin. Peptides of peptic digestion can be identified and their masses can be determined. One incorporated deuteron leads to an increase of one mass unit. The peptide mass therefore allows the calculation of the number of incorporated deuterons at a

Table 1
Results of fitting deuterium incorporation data (see Fig. 4) to the equation $D(t) = A^*(1-e^{-k_1*t})+B^*(1-e^{-k_2*t})$

Peptic peptide of HiPIP region	No. of amide peptide bonds (N)	Fast exchange rate constant (k_1)	No. of amide peptide bonds exchanging with $k_{ex} = k_1$ (A)	Moderate exchange rate constant (k_2)	No. of amide peptide bonds exchanging with $k_{ex} = k_2$ (B)	No. of amide peptide bonds exchanging very slow[a] (C)
S_2-L_{29}	25	$\geq 6.0\ min^{-1}$	22.0	$0.14\ min^{-1}$	3.0	0
A_{60}-E_{77}	14	$\geq 6.14\ min^{-1}$	5.0	$0.11\ min^{-1}$	2.6	6.4

[a]Number of peptide bonds which do not exchange within 120 min are calculated as difference $C = N-A-B$.

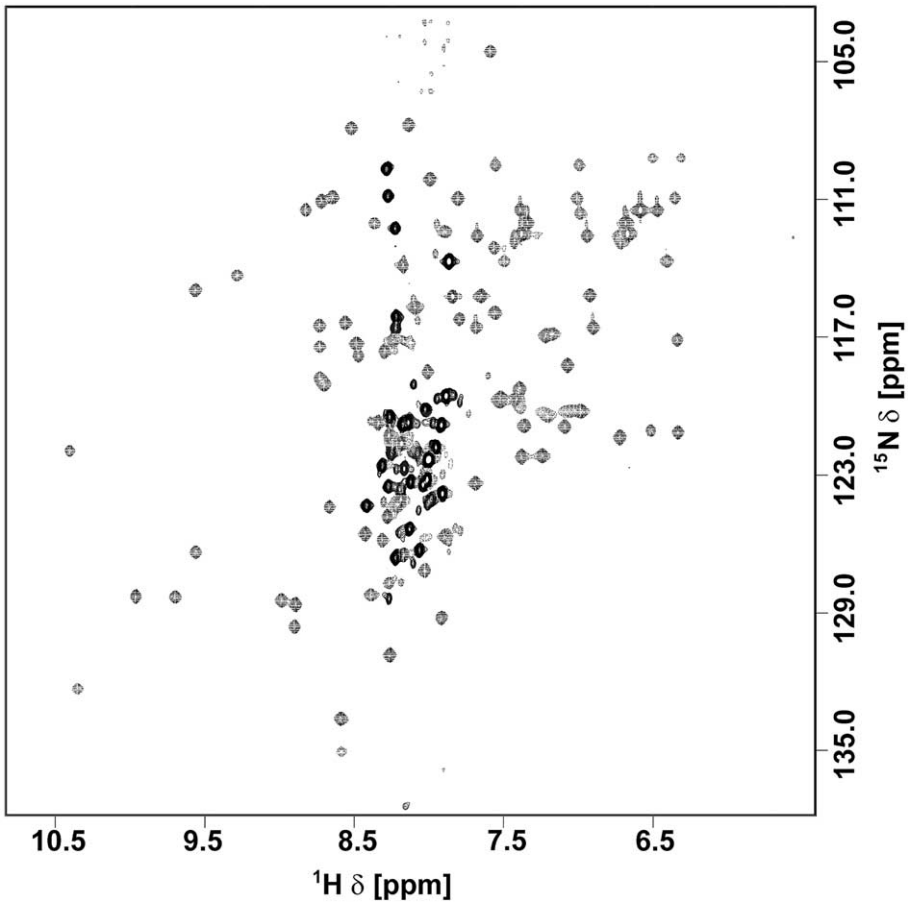

Fig. 2. ^{1}H-^{15}N HSQC spectrum of folded HiPIP precursor. The 32 resonances that are not present in the spectrum of mature HPIP and which can be attributed to the signal sequence are indicated in black.

given incubation time. As hydrogen bonding slows down the H/D exchange process, the kinetics of deuterium incorporation reflects directly and quantitatively the structure within the analyzed peptide [19,20].

The mass of HiPIP precursor as derived from inclusion bodies could be determined to be 12 629 Da, indicating that the N-terminal methionine is cleaved off (Fig. 3). This cleavage of the N-terminal methionine has been previously observed with soluble HiPIP precursor purified from cytoplasmic fractions [5]. Subsequent analyses of pepsin-digested HiPIP resulted in the identification of a signal sequence-derived peptide, S2-L29, which is ideally suitable for H/D exchange analyses (Fig. 4A). This peptide covers the complete n-region of the signal sequence and most of its h-region, and therefore it includes all residues of the twin-arginine motif pattern. As a control, a peptide from the mature domain of HiPIP, A60-E77, has been analyzed in parallel (Fig. 4B). The results are summarized in Table 1. 22 of 25 amide protons of the analyzed signal sequence peptide exchange rapidly. They do not show retarded exchange kinetics, indicating that they are not involved in hydrogen bonding. The remaining three amide protons exchange with moderate rate constants and indicate weak hydrogen bonds. In contrast, six of 14 amide protons of the control peptide of the folded mature domain do not exchange even within 120 min, indicating very stable hydrogen bonds. Three amide protons of this peptide show moderate exchange rate constants and only five can exchange without a limitation of their exchange rate constant by intramolecular hydrogen bonding.

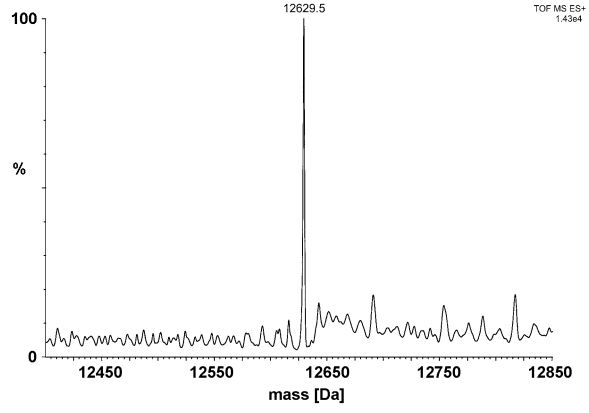

Fig. 3. Deconvoluted ESI Qq-TOF spectrum of preHoloHiPIP. Measurement was performed using a Micromass Q-TOF II instrument (Waters, Manchester, UK). Mass of 12 629.5 ± 1 Da was detected, the theoretical average mass of HiPIP S_2-G_{122} is 12 630.4 Da.

4. Discussion

This study reports the first structural analysis of a twin-arginine signal sequence. Based on NMR and mass spectrometry results, the signal sequence from soluble HiPIP precursor

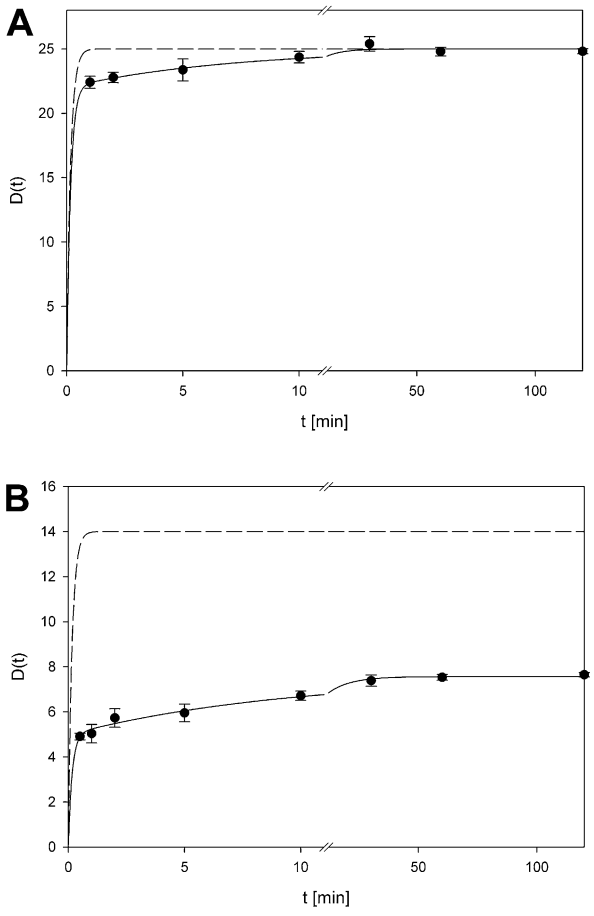

Fig. 4. Time course of deuterium incorporation into the amide peptide bonds of the regions. S_2-L_{29} (A) and A_{60}-E_{77} (B) within preHoloHiPIP. Data were corrected for 10% residual H_2O during deuterium incorporation and back-exchange of 10% (S_2-L_{29}) or 4% (A_{60}-E_{77}) during preparation and mass analysis. For results of data fitting to the equation $D(t) = A^*(1-e^{-k_1*t})+B^*(1-e^{-k_2*t})$ (solid line), see Table 1. The theoretical time course of unstructured ($k_{ex} = k_1 = k_2 = 6$ min^{-1}) is displayed by scattered lines.

has to be regarded as largely unstructured, whereas the mature region of in vitro folded HiPIP precursor is folded like native mature HiPIP, as in agreement with the reported electron paramagnetic resonance (EPR) studies on the cofactor in HiPIP precursor [15]. As these measurements were carried out with fully folded protein, it is apparent that the signal sequence does not fold by itself or as a result of intramolecular interactions with the mature domain. The observed amide proton resonances from the signal sequence are in areas where resonances from unstructured residues are expected (Fig. 2). Although unlikely, the NMR results did not completely exclude that the peptide is e.g. an α-helix, as α-helical shifts may overlap with shifts from unstructured regions. We therefore analyzed the signal sequence under native conditions by an independent and different approach which used H/D exchange mass spectrometry as a tool. The signal sequence up to residue 29 out of 37 does not build up any stable backbone hydrogen bonding pattern which would have resulted in a slowed down hydrogen exchange kinetics (Fig. 4). We conclude that at least up to L29 the signal sequence of soluble, pure HiPIP does not adopt a stable secondary structure. The unstructured region covers the complete n- and h-regions of the signal sequence, and therefore includes the twin-arginine motif. Importantly, these results indicate that in silico predictions of an α-helical structure in the twin-arginine signal sequence of HiPIP are incorrect (Fig. 1). Although we do not exclude that other twin-arginine signal sequences may form secondary structures, our results show that such a structure formation is not necessarily an intrinsic property of Tat signal sequences.

Could there be an interaction of side chains with the mature domain of the same polypeptide chain in a way which could be regarded as a structure? There is no evidence for this idea. The first analysis of the NMR data indicates that no resonances from the mature domain are significantly shifted in the presence of the signal sequence. This argues against a direct interaction of the signal sequence with the mature domain of HiPIP. The results from H/D exchange mass spectrometry further support this view. This method probes very sensitively structure formation within polypeptides [21] and any stable interaction between the signal sequence and the mature domain should be reflected in the deuterium incorporation kinetics, even in the absence of direct hydrogen bonding between the mature domain and an amide proton of the signal sequence. So far, we did not detect any interactions between the signal sequence and the mature domain by H/D exchange.

Does the signal sequence of HiPIP form a structure in vivo? The idea of a signal sequence structure is attractive, because this would have important physiological implications. The twin-arginine motif is part of the signal sequence and this motif appears to be specifically recognized by the translocation system. Even the conservative exchange of the two conserved arginines with lysines results in a complete block of translocation of so far tested Tat substrates [22]. HiPIP has also been tested and is no exception from this rule [15]. Moreover, it has been shown with the thylakoidal Tat system that the twin-arginine motif is required for a direct interaction with the Tat system components [23]. It must therefore be considered that the twin-arginine motif pattern forms a structure which is specifically recognized by the Tat system. However, we show in this study that a specific twin-arginine motif structure – if it exists – is not formed autonomously by the signal sequence or by intramolecular interactions of the signal sequence with the mature domain. The twin-arginine motif may therefore either form a recognizable structure by an interaction with the translocon (induced fit), or by an interaction with other components on the way to the translocon (substrate priming). Our studies are currently addressing these hypotheses.

Acknowledgements: This work was supported by the Deutsche Forschungsgemeinschaft with grant BR2285/1-1 to T.B.

References

[1] Blobel, G. and Dobberstein, B. (1975) J. Cell Biol. 67, 852–862.
[2] Pugsley, A.P. (1993) Microbiol. Rev. 57, 50–108.
[3] Brüser, T. and Sanders, C. (2003) Microbiol. Res. 158, 7–17.
[4] Stanley, N.R., Palmer, T. and Berks, B.C. (2000) J. Biol. Chem. 275, 11591–11596.
[5] Brüser, T., Deutzmann, R. and Dahl, C. (1998) FEMS Microbiol. Lett. 164, 329–336.
[6] Cristóbal, S., de Gier, J.-W., Nielsen, H. and von Heijne, G. (1999) EMBO J. 18, 2982–2990.
[7] Rodrigue, A., Chanal, A., Beck, K., Müller, M. and Wu, L.-F. (1999) J. Biol. Chem. 274, 13223–13228.

[8] Santini, C.-L., Bernadac, A., Zhang, A., Ize, B., Blanco, C. and Wu, L.-F. (2001) J. Biol. Chem. 276, 8159–8164.
[9] Thomas, J.D., Daniel, R.A., Errington, J. and Robinson, C. (2001) Mol. Microbiol. 39, 47–53.
[10] Halbig, D., Wiegert, T., Blaudeck, N., Freudl, R. and Sprenger, G.A. (1999) Eur. J. Biochem. 263, 543–551.
[11] Sanders, C., Wethkamp, N. and Lill, H. (2001) Mol. Microbiol. 41, 241–246.
[12] Nurizzo, D., Halbig, D., Sprenger, G.A. and Baker, E.N. (2001) Biochemistry 40, 13857–13867.
[13] Banci, L., Bertini, I., Dikiy, A., Kastrau, D.H.W., Luchinat, C. and Sompornpisut, P. (1995) Biochemistry 34, 206–219.
[14] Carter, C.W., Kraut, J., Freer, S.T., Xuong, N.-H., Alden, R.A. and Bartsch, R.G. (1974) J. Biol. Chem. 249, 4212–4225.
[15] Brüser, T., Yano, T., Brune, D.C. and Daldal, F. (2003) Eur. J. Biochem. 270, 1211–1221.
[16] Imhoff, J.F. (2003) in: The Prokaryotes, 3rd Edn., online release 3.13 (Dworkin, M., Ed.), Springer, New York.
[17] Bartsch (1971) Methods Enzymol. 23, 644–649.
[18] Kipping, M. and Schierhorn, A. (2003) J. Mass Spectrom. 38, 271–276.
[19] Zhang, Z. and Smith, D.L. (1993) Protein Sci. 2, 522–531.
[20] Englander, S.W. and Kallenbach, N.R. (1984) Q. Rev. Biophys. 16, 521–655.
[21] Kipping, M., Zarnt, T., Kiessig, S., Reimer, U., Fischer, G. and Bayer, P. (2001) Biochemistry 40, 7957–7963.
[22] Buchanan, G., Sargent, F., Berks, B.C. and Palmer, T. (2001) Arch. Microbiol. 177, 107–112.
[23] Cline, K. and Mori, H. (2001) J. Cell Biol. 154, 719–729.
[24] Garnier, J., Gibrat, J.-F. and Robson, B. (1996) Methods Enzymol. 266, 540–553.
[25] Geourjon, C. and Deleage, G. (1995) Comput. Appl. Biosci. 11, 681–684.
[26] McGuffin, L.J., Bryson, K. and Jones, D.T. (2000) Bioinformatics 16, 404–405.