Folding of a synthetic parallel β-roll protein

Hauke Lilie[a,*], Wolfgang Haehnel[b], Rainer Rudolph[a], Ulrich Baumann[c]

[a] *Institut für Biotechnologie der Martin-Luther-Universität Halle, Kurt-Mothes Strasse 3, D-06120 Halle/Saale, Germany*
[b] *Institut für Biologie II/Biochemie, Albert-Ludwigs-Universität Freiburg, Schänzlestrasse 1, D-79104 Freiburg, Germany*
[c] *Departement für Chemie und Biochemie der Universität Bern, Freiestrasse 3, CH-3012 Bern, Switzerland*

Received 10 January 2000; received in revised form 23 February 2000

Edited by Matti Saraste

Abstract Recently, the design of β-sheet proteins and concomitant folding studies have attracted increasing attention. A unique natural all-β domain occurs in a family of cytolytic bacterial toxins, the so-called RTX toxins. This domain consists of a variable number (about 6–45) of tandem repeats of a glycine-rich nine-residue motif with the consensus sequence GGXG-XDX(L/I/F)X. The analysis of the three-dimensional structure of alkaline protease from *Pseudomonas aeruginosa* which possesses six of these repeats revealed that they fold into a novel 'parallel β-roll' where calcium is bound within the turns connecting the β-strands. A 75-mer peptide of the sequence NH_2-WLS-$[GGSGNDNLS]_8$-COOH was chemically synthesised. Circular dichroism spectroscopy showed that this polypeptide folds in the presence of Ca^{2+} and polyethylene glycol into a β-structure which is presumably identical with the parallel β-roll. This synthetic β-roll behaves similarly to the isolated β-roll domains from *Escherichia coli* haemolysin or *Bordetella pertussis* cyclolysin in terms of calcium binding and polymerisation behaviour.

© 2000 Federation of European Biochemical Societies.

Key words: β-Sheet; Folding; Parallel β-helix; RTX toxin; Design

1. Introduction

RTX toxins (RTX = repeats in toxins) are a family of mostly cytolytic toxins which are secreted by Gram-negative bacteria into the medium [1,2]. The secretion pathway does not involve an N-terminal signal peptide but rather a C-terminal secretion signal as was shown for the first time for *Escherichia coli* haemolysin [3]. The name RTX stems from a glycine-rich sequence motif GGXGXDX(L/F/I)X which precedes the secretion signal. This sequence motif is tandemly repeated and the number of repeats varies between different proteins. It is found in all proteins which are secreted by the haemolysin pathway, so for example also in the alkaline protease from *Pseudomonas aeruginosa* [4] or a lipase from *Pseudomonas fluorescens* [5]. The exact function of these repeats remains somewhat obscure, they have been suggested to be receptor binding domains [6], enhancers of secretion [7] and internal chaperones [8]. In the three-dimensional structure of the alkaline protease from *P. aeruginosa* [8] it was found that these sequence repeats constitute a new calcium binding structure, called a parallel β-helix or parallel β-roll (Fig. 1), where the first six residues of each motif form a turn which binds calcium and the remaining three residues build a short β-strand. The consecutive β-strands are connected in such a way that a right-handed helix of parallel β-strands is formed. One turn of this helix consists of two consecutive nine-residue motifs (Fig. 1). The requirements for the particular amino acids at the individual positions in the nona-sequence can be rationalised easily: G2 and G4 have backbone dihedral angles which are unfavourable for non-glycine residues, a C^b on G1 would cause steric clashes. The side chain of position 8 forms the hydrophobic core of the β sandwich and is ideally leucine which nicely interdigitates with the corresponding residue of the next repeat. The aspartic acid on position 6 bridges two calciums and neutralises the positive charge of these buried metal ions. Residues 3 and 5 are small and mostly hydrophilic. It can be seen from Fig. 1 that the parallel β-roll is a highly regular super-secondary structure with calcium ions as an integral part. The calcium cannot be removed by EGTA or similar chelators under native conditions with the exception of those located at the edges of the roll which are solvent-exposed. This parallel β-roll is different from the parallel β-helix found in the pectate lyase [9] or the P22 tailspike protein [10]. This latter parallel β-helix does not consist of a repetitive sequence and involves three β-sheets rather than two.

Despite the importance of β-sheet structures as regular secondary structure elements the principles underlying their formation and stability are not well understood. Until now, most information has been derived from mutagenesis experiments [11,12] and recently through the study of de novo designed β hairpins [13,14]. A major obstacle is the tendency of isolated β-sheets to aggregate, an example of this is the formation of amyloid fibrils in Alzheimer's disease.

The parallel β-roll $[GGXGXDXLX]_n$ found in RTX toxins seems to be an attractive system to study β-sheet formation: the size is amenable to chemical synthesis and the folding process can presumably be triggered by calcium ions. Studies on isolated β-roll domains from RTX toxins have been performed in the past on the *Bordetella pertussis* adenylate cyclase toxin (about 45 repeats [15]) and *E. coli* haemolysin (about 15 repeats [16]). In the former there are blocks of 8–12 repeats which are separated by intervening sequences. Both protein constructs show a large conformational rearrangement upon calcium binding.

We report here the design of a β-roll with eight identical sequence motifs NH_2-WLS-$[GGSGNDNLS]_8$-COOH and folding studies on this system.

*Corresponding author.
E-mail: lilie@biochemtech.uni-halle.de

Abbreviations: GdmCl, guanidinium hydrochloride; PEG, polyethylene glycol with a molecular mass $M_r = 8000$

0014-5793/00/$20.00 © 2000 Federation of European Biochemical Societies. All rights reserved.
PII: S0014-5793(00)01308-9

2. Materials and methods

2.1. Sequence design

The sequence motif was designed according to the rules outlined above. Positions 3, 5, 7 and 9 were occupied by Ser and Asn, with the intention of having polar residues on the outside of the sheet. These residues were also thought to be capable of forming an asparagine or serine hydrogen bond ladder similar to the ones observed in [9]. The model was constructed on an interactive graphics display using the program O [17] and energy-minimised using X-PLOR [18]. Sequence search for nine-residue motifs was done using the PATTERFIND server at ISREC, Lausanne, Switzerland.

2.2. Synthesis of the peptide sequence

The synthesis of the peptide with 75 amino acid residues began by solid-phase peptide synthesis using Fmoc-protected amino acids on a Ser-Pep Syn-KA resin in a Milligen Model 9050 peptide synthesiser. A four-fold excess of Fmoc L-amino acid was activated for 7 min with TBTU/DIEA and recirculated through the resin for 30 min. The side groups of Asn, Asp and Ser were protected by Trt, OBut, and But, respectively. Fmoc-protected amino acids and TBTU were purchased from Perseptive Biosystems. The Fmoc group was removed by washing the resin with 20% (v/v) piperidine in DMF for 7 min. After cleavage from the resin by trifluoroacetic acid (TFA)/anisole/dithiothreitol and precipitation with diethylether, the 75-mer peptide was purified by preparative HPLC on a Waters model 600E system equipped with a Waters DeltaPak C18 PrepPak column (40× 300 mm, 15 μm) with a linear gradient of water/acetonitrile containing 0.1% TFA at a flow rate of 10 ml/min. The correct sequence was verified by ESI mass spectrometry on a Finnigan model TSQ700 yielding 6818.2 Da (theoretical 6818.9 Da). Contaminating peptides with one or two Asp/Asn missing were lower than 10/3%, respectively. Other details were as described [19].

2.3. Spectroscopy

Analyses of secondary structure were performed by circular dichroism (CD) spectroscopy in the far-UV region using a spectrometer Aviv 62 D. Spectra were monitored at a protein concentration of 0.1 mg/ml. The samples were incubated at 20°C for at least 12 h. Kinetics of structure formation were analysed by diluting the unstructured protein (in water) in the respective buffer and following the change in the CD signal at 220 nm over time. For all measurements a 1-mm cuvette was used.

Based on the lag phase of structure formation upon folding the minimum protein concentration for structure formation was calculated according to:

$$y = a + b/(x-c) \quad (1)$$

where y is the time of the lag phase, x is the initial monomeric protein concentration, c is the minimum protein concentration and a and b the asymptotic boundary values.

The size of the nucleus of the polymerisation process was estimated from the slope of the lag phase using:

$$v = k*c^p \quad (2)$$

where v is the initial velocity, k is a constant independent of protein concentration and c is the initial concentration of monomeric protein [20]. The value of p represents the number of molecules involved in the formation of the nucleus.

3. Results and discussion

3.1. Structure formation of a synthetic β-roll

The synthetic β-roll polypeptide was completely soluble in a buffer of low ionic strength or even in water. The molecular mass determined by analytical ultracentrifugation was $M_r = 7420$, very close to the calculated value of 6818 Da (data not shown). However, as shown by a comparison of CD spectra of the protein in buffer and 6 M guanidinium hydrochloride (GdmCl) it did not possess any regular secondary structure (Fig. 2). Even the addition of millimolar concentrations of Ca^{2+}, which is a specific low-affinity ligand for this type of β-roll, did not lead to structure formation. Only in the presence of Ca^{2+} and polyethylene glycol (PEG) as an unspecific stabilising agent a β-sheet structure could be induced (Fig. 2). The absolute requirement of PEG for structure formation was a first indication of the low intrinsic stability of this structure.

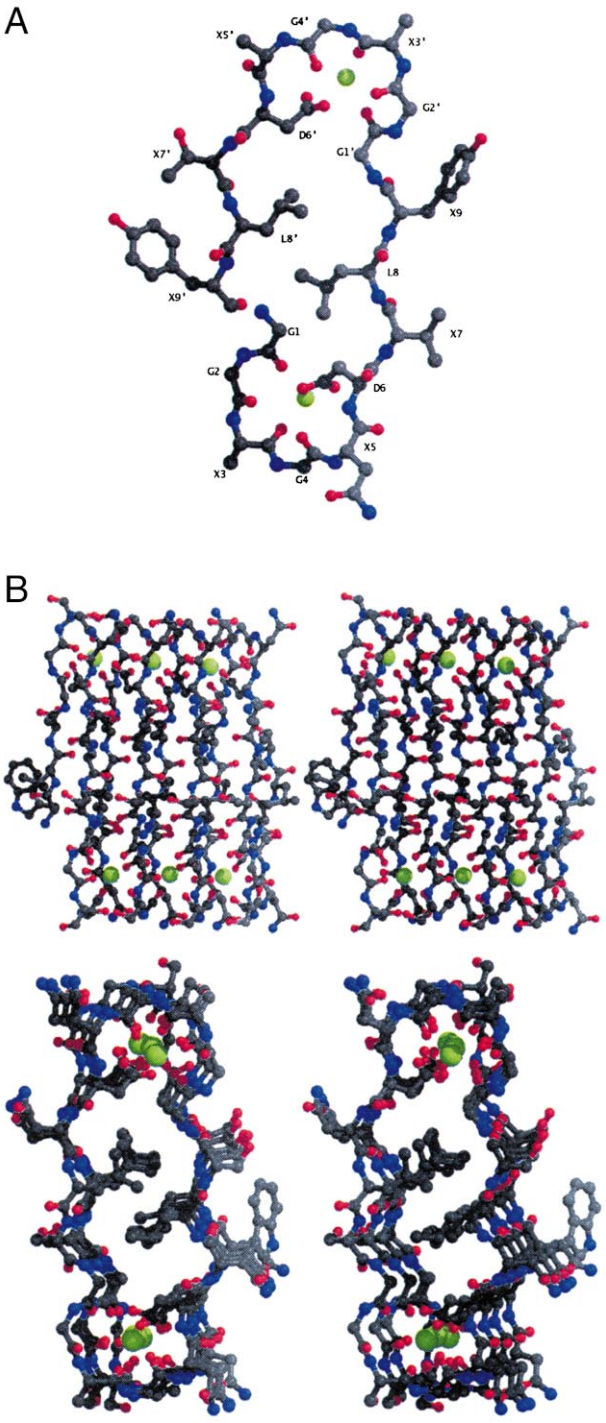

Fig. 1. The parallel β-roll. A: One turn of the β-roll. Residues of the first nine-residue motif are labelled G1, G2 etc. while the residues of the next motif have a prime, e.g. G1′, G2′. If G1 is at position n in the sequence and G2 at position $n+1$, then G1′ is at the position $n+9$ and G2′ at $n+10$ and so forth. Calciums are shown as green balls. B: Two orthogonal views of the β-roll found in alkaline protease from *P. aeruginosa*. Calciums are shown as green balls.

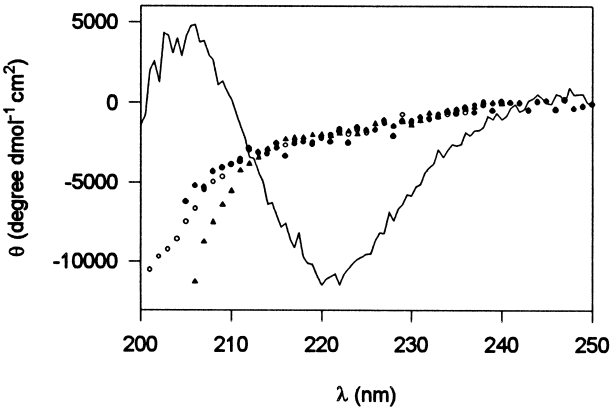

Fig. 2. Structure formation of the synthetic β-roll. CD spectra of the β-roll were measured after 12 h incubation at 20°C in 50 mM Tris, pH 7, supplemented with 100 mM $CaCl_2$ (○), 25% PEG 8000 (●), 6 M GdmCl (▲) and 100 mM $CaCl_2$, 25% (w/v) PEG 8000 (solid line), respectively. The protein concentration was 0.1 mg/ml, light path 1 mm.

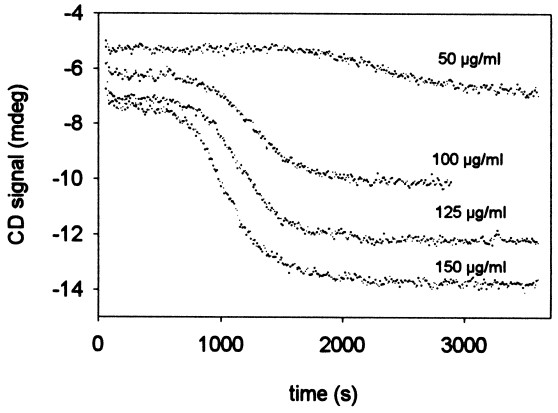

Fig. 4. Kinetics of structure formation. Unstructured peptide (1 mg/ml) in 50 mM Tris, pH 7 was diluted in Tris buffer with final concentrations of 100 mM $CaCl_2$, 25% PEG and protein concentrations of 50, 100, 125 and 150 μg/ml. The time course of structure formation was measured by CD at 220 nm. The different starting values of the kinetics represent the (not buffer-corrected) signal of denatured protein at the respective concentrations.

The induced β-helix seemed to be comparable to that of the respective parts of natural proteins from which it was deduced, since its ability to specifically bind Ca^{2+} was very similar to that of adenylate cyclase toxin or of haemolysin which bind Ca^{2+} with a dissociation constant in the millimolar range [15,16,21]. In the presence of 25% PEG the protein was half saturated with Ca^{2+} at about 3 mM (Fig. 3). It should be noticed that this binding cannot be described by a dissociation constant (the same holds for the data presented in the literature) since Ca^{2+} binding was not reversible. Addition of EDTA did not lead to a release of the ions (data not shown). Furthermore, the molecular basis of Ca^{2+} binding depends not only on binding but also on the induction of folding and polymerisation of the protein (see below). Thus, thermodynamic parameters characterising the interaction of Ca^{2+} ions with the β-roll protein could not be drawn from the data. Nevertheless, the properties of apparent cooperative Ca^{2+} binding at millimolar concentrations were very similar to that of haemolysin or adenylate cyclase toxin. Other ions such as Mg^{2+} or Tb^{2+} were not bound by the peptide and as a consequence did not induce the β-roll structure. Without being a final proof this specific Ca^{2+} effect at least suggests a similar topology of the induced β-sheet structure as compared to natural β-roll proteins.

In order to analyse the induction of structure in more detail we measured the kinetics of folding. Surprisingly, this process, monitored by far-UV CD, was characterised by a pronounced lag phase followed by cooperative structure formation. The overall kinetics showed a strong concentration dependence (Fig. 4). After completion of folding no monomeric or dimeric species could be observed by analytical ultracentrifugation (data not shown). Thus, structure formation quantitatively led to polymerisation of the protein. Variations of temperature and concentration of Ca^{2+} and PEG influenced the kinetics of folding. However, in all cases tested folding and polymerisation occurred simultaneously (data not shown).

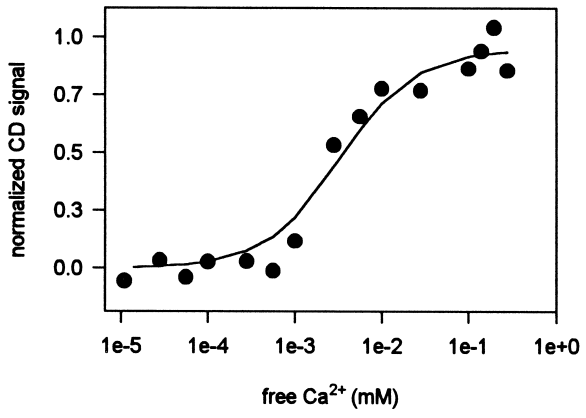

Fig. 3. Binding of Ca^{2+} to the β-roll peptide. The β-roll peptide at a concentration of 0.1 mg/ml in 50 mM Tris, pH 7, 25% PEG was incubated at different concentrations of $CaCl_2$ for 16 h at 20°C. The CD signal at 220 nm was monitored.

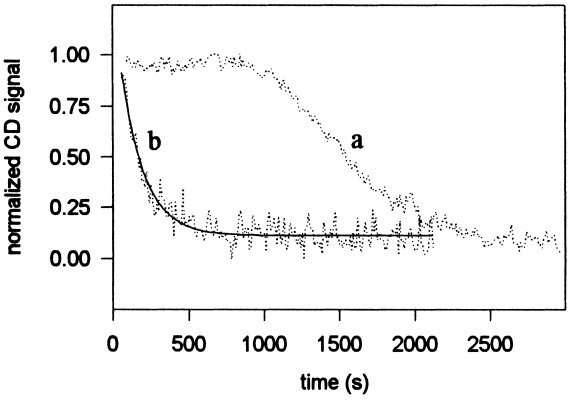

Fig. 5. Polymerisation of the β-roll peptide. Folding of the peptide at a concentration of 100 μg/ml was induced by diluting it in 50 mM Tris, pH 7, 100 mM $CaCl_2$, 25% PEG (curve a). After 50 min a further 100 μg/ml unstructured peptide was added and the kinetics of structure formation determined (curve b). The solid line represents a fit to a single first-order reaction with a rate constant of 4.5×10^{-3} s^{-1}. For comparison of the amplitude of both reactions the CD data at 220 nm were normalised.

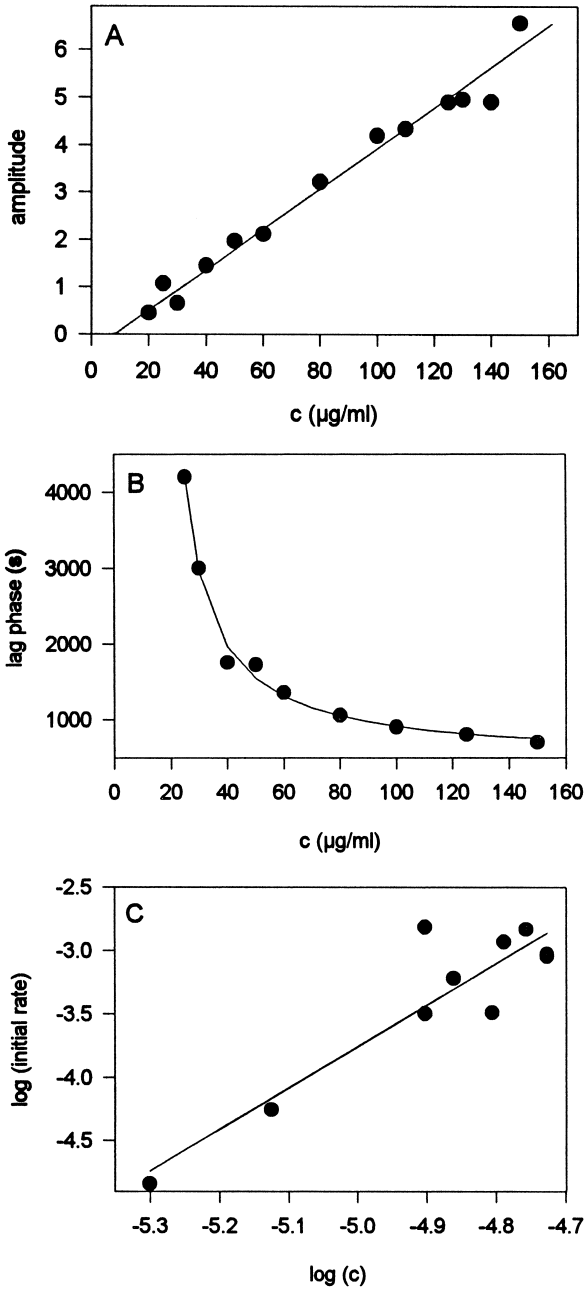

Fig. 6. Characterisation of the nucleation process. Structure formation of the β-roll peptide at different protein concentrations was induced as described in Fig. 4. A: The amplitude of the CD signal at 220 nm is plotted against the protein concentration. A linear extrapolation of the data indicate that below 9 μg/ml no amplitude and, therefore, no folding would be detected. B: The duration of the lag phase of folding was analysed with respect to protein concentration. A fit of the data according to Eq. 1 showed that below 15 μg/ml the time of the lag phase reaches infinity. C: The initial rate of folding was analysed dependent on the protein concentration. The slope p in a double logarithmic plot corresponds to the number of molecules in the nucleus. The fit yielded a value of $p = 3.27$.

3.2. Mechanism of polymerisation

The kinetics of structure formation characterised by a long lag phase and a subsequent cooperative folding reaction indicated a nucleated polymerisation process based on structured monomers with the nucleation as the rate-limiting step. This was confirmed by analysing the effect of preformed structured β-roll on the kinetics of folding of unstructured protein (Fig. 5). In the presence of already structured and polymerised protein the folding kinetics of unfolded protein occurred without a detectable lag phase. Instead, the folding kinetics obeyed an apparent first-order reaction. Thus, the preformed polymers acted as nucleation sites for the added denatured protein.

A nucleation reaction is characterised by (i) a minimum concentration necessary to start the reaction, (ii) the size of the nucleus and (iii) the polymerisation rate. The minimum concentration necessary to initiate the overall polymerisation reaction was determined in two different ways. First the dependence of the measured CD signal on the protein concentration was analysed (Fig. 6A). The extrapolation yielded a minimum protein concentration of ca. 9 μg/ml necessary for structure formation. In another approach the length of the lag phase was measured with respect to protein concentration. The lag phase was defined as the time in which only a linear change in the CD signal over time was observed. Extrapolation of this relation according to Eq. 1 showed that at 15 μg/ml the time of the lag phase became infinite (Fig. 6B) which means below a protein concentration of 15 μg/ml no folding and polymerisation would occur.

The lag phase of the time course of folding also contains information on the size of the nucleus of polymerisation. The initial rate of structure formation in the lag phase depends on the protein concentration. According to Eq. 2 a double logarithmic plot of the initial rate versus protein concentration should result in a straight line with a slope characterising the size of the nucleus. The analysis of the data is shown in Fig. 6C. From the slope of this curve the size of the nucleus was calculated to be three monomers.

These results indicate that the protein designed on the basis of the consensus sequence of several β-roll motifs is able to form a β-helix. In this first approach, however, intramolecular interactions are too weak to stabilise this structure in the monomeric state. In the presence of stabilising additives polymerisation is not only a dominant side reaction but instead the analysis of the nucleation reaction suggests that intermolecular interactions are essential to stabilise the β-roll. Recently, for another β-helix protein, the P22 tailspike protein, it was demonstrated that the polymerisation of folding intermediates leads to the formation of amyloidogenic fibrils [22]. Therefore, we investigated whether or not in our case the aggregates were amyloidogenic. Neither by Congo red staining nor by electron microscopy could specific fibrillar aggregates be observed (data not shown).

In summary, the synthetic parallel β-roll behaves in a similar manner as the isolated domains from *B. pertussis* adenylate cyclase toxins or *E. coli* haemolysin. The calcium binding in the millimolar range is of the same order. The synthetic protein is characterised by a strong tendency to polymerise in the presence of stabilising additives. The intermolecular interactions caused by polymerisation are essential to stabilise the β-roll structure in this designed peptide. A reason for the only marginal stability of the folded state of the synthetic β-roll could be a sub-optimal consensus sequence NH_2-WLS-[GGSGNDNLS]$_8$-COOH used in this study. Sequence alignment of 124 nine-residue motifs GGXGXDXXX revealed that the ideal consensus sequence would be GGAGNDILV. Especially in position 7 no asparagines are found at all in the data base but rather residues with higher β-sheet propensities [23],

namely Ile (16.1%) or Thr (14.5%), Phe (11.3%) and Tyr (10.5%). In position 9, Ser is present only in 3.2% of all cases, but Val (16.9%), Tyr (10.5%), aspartate (12.1%) and glutamate (11.3%) dominate. Also in position 3 Ser (7.3%) is disfavoured compared to Ala (21.0%). The other positions are optimally occupied in our current sequence.

Despite the sub-optimal sequence used in this study a β-roll structure could be induced in the polypeptide, similar to the structure of natural β-roll proteins. These results are quite encouraging to design a new polypeptide with an optimal consensus sequence.

Acknowledgements: We thank Rosemarie Loyal for synthesis of the polypeptide. Renate Nitsch is acknowledged for technical assistance. This work was supported by the Swiss National Science foundation, Grant 31-52398.97 to U.B.

References

[1] Coote, J.G. (1992) FEMS Microbiol. Rev. 8, 137–161.
[2] Welsh, R.A. (1991) Mol. Microbiol. 5, 521–528.
[3] Mackman, N., Nicaud, J.M., Gray, L. and Holland, I.B. (1986) Curr. Top. Microbiol. Immunol. 125, 159–181.
[4] Duong, F., Lasdunski, A., Cami, B. and Murgier, M. (1992) Gene 121, 47–54.
[5] Duong, F., Soscia, C., Lasdunski, A. and Murgier, M. (1994) Mol. Microbiol. 11, 1117–1126.
[6] Ludwig, A., Jarchau, T., Bens, R. and Goebel, W. (1988) Mol. Gen. Genet. 214, 553–561.
[7] Létoffé, S. and Wandersman, C. (1992) J. Bacteriol. 174, 4920–4927.
[8] Baumann, U., Wu, S., Flaherty, K.M. and McKay, D.B. (1993) EMBO J. 12, 3357–3364.
[9] Yoder, M.D., Keen, N.T. and Jurnak, F. (1993) Science 260, 1503–1507.
[10] Steinbacher, S., Seckler, R., Miller, S., Steipe, B., Huber, R. and Reinemer, P. (1994) Science 265, 9765–9771.
[11] Minor, D.L. and Kim, P.S. (1994) Nature 367, 660–663.
[12] Smith, C.K. (1995) Science 270, 980–982.
[13] de Alba, E., Jimenez, M.A., Rico, M. and Nieto, J.L. (1996) Fold. Des. 1, 133–144.
[14] Ramirez-Alvarado, M., Blanco, F.J. and Serrano, L. (1996) Nature Struct. Biol. 3, 604–612.
[15] Rose, T., Sebo, P., Bellalou, J. and Ladant, D. (1995) J. Biol. Chem. 270, 26370–26376.
[16] Ostolasa, H., Soloaga, A. and Goni, F.M. (1995) Eur. J. Biochem. 228, 39–44.
[17] Jones, T.A., Sou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Acta Crystallogr. A47, 110–116.
[18] Brünger, A.T. (1992) X-PLOR, A System for X-ray Crystallography and NMR, Yale University Press, New Haven, CT.
[19] Rau, H.K. and Haehnel, W. (1998) J. Am. Chem. Soc 120, 468–476.
[20] Kasai, M., Asakura, S. and Oosawa, F. (1962) Biochim. Biophys. Acta 57, 22–31.
[21] Bakas, L., Veiga, M.P., Soloaga, A., Ostolaza, H. and Goni, F.M. (1998) Biochim. Biophys Acta 1368, 225–234.
[22] Schuler, B., Rachel, R. and Seckler, R. (1999) J. Biol. Chem. 274, 18589–18596.
[23] Kim, A.C. and Berg, J.M. (1993) Nature 362, 267–270.