The Construction of New Proteins*

II. Design, Synthesis and Conformational Studies of Folding Units with βαβ-Topology**

Manfred Mutter*, Karl-Heinz Altmann, and Thomas Vorherr
Institute of Organic Chemistry, University of Basel, St.-Johanns-Ring 19, CH-4056 Basel

Herrn Professor Dr. Ernst Bayer zum 60. Geburtstag gewidmet
Z. Naturforsch. 41b, 1315–1322 (1986); received April 18, 1986

Artificial Proteins, βαβ-Folding Units, Conformational Studies, Liquid-Phase Peptide Synthesis,
Tertiary Structure

The design, synthesis and preliminary conformational studies of two polypeptides exhibiting βαβ-type folding topologies are presented. In the design of the model peptides the general concept for the construction of new proteins developed in the preceding paper was applied. According to this strategy, amphiphilic helices and β-sheets are linked together via hydrophilic loops to attain three-dimensional structures of higher order (‘supersecondary structures’). Computer-assisted molecular modelling served as a valuable tool for minimizing conformational constraints within the molecules. The 38-residue peptide MI was synthesized using polyethylene glycol (PEG) as solubilizing polymeric support (‘Liquid-Phase synthesis’). Conformationally induced changes in the physico-chemical properties of the growing peptide chain stressed the significance of conformational effects in peptide synthesis reported earlier. Similar observations were made during the solid-phase synthesis of the 35-peptide MIL CD and IR spectroscopic studies revealed a high degree of secondary structure for both folding units. The present data strongly support the adoption of a three-dimensional structure for both models.

The construction of new proteins has become one of the most fascinating prospects in the field of peptide chemistry [1]. A milestone on the way to this goal was the development of the recombinant DNA technique [2], allowing – in principle – the synthesis of any arbitrary amino acid sequence. On the other hand, the question how a given primary sequence folds into its unique tertiary structure is open so far [3]. Furthermore the ability to attain a defined three-dimensional conformation seems to be restricted to a vanishingly small number of amino acid sequences [1]. Consequently, the critical hurdle on the way to proteins with unnatural (nonhomologous)

Abbreviations: Aib, α-aminoisobutyric acid; BOC, t-butyloxycarbonyl; Bu, t-butyl; Bzl, benzyl; CD, circular dichroism; CFm, 9-(4-carboxyfluorenyl)-methoxy; DVB, divinylbenzene; Fmoc, 9-fluorenylmethoxycarbonyl; HMPA, 4-hydroxymethyl-3-methoxy-phenoxyacetic acid; PEG, polyethylene glycol; PEGM, polyethylene glycol-monomethylether, M, = 5000; TFA, trifluoroacetic acid; TFE, 2,2,2-trifluoroethanol.

* The foundations of the present work have been the topic of a lecture given by M. M. on being awarded the Max Bergmann Prize, 3rd Max Bergmann Conference, October 7–9, 1982, Bolzano (Italy).

** For part I of this series see reference [1].

† Reprint requests to Prof. Dr. M. Mutter.

Verlag der Zeitschrift für Naturforschung, D-7400 Tübingen
0340–5087/86/1000–1315/S 01.00/0

The most fascinating prospects in the field of peptide chemistry [1]. A milestone on the way to this goal was the development of the recombinant DNA technique [2], allowing – in principle – the synthesis of any arbitrary amino acid sequence. On the other hand, the question how a given primary sequence folds into its unique tertiary structure is open so far [3]. Furthermore the ability to attain a defined three-dimensional conformation seems to be restricted to a vanishingly small number of amino acid sequences [1]. Consequently, the critical hurdle on the way to proteins with unnatural (nonhomologous) amino acid sequences is the elucidation of the relationship between primary and tertiary structure. Recent progress in the study of the topology and the folding mechanism of proteins has provided considerable insight into the rules governing the three-dimensional architecture of proteins [3–6]. For example, the increasing amount of X-ray crystallographic data revealed that only a small number of folding topologies exists, enabling a classification of proteins by patterns of similar tertiary structure [5–8]. Folding units with a specific spatial arrangement of secondary structure blocks (‘supersecondary structures’) seem to be a basic element of protein structure [9, 10]. Most notably, their topological features are not bound to a specific amino acid sequence (‘degenerate folding code’). They can be considered as independently foldable segments along the protein sequence and reflect the minimum chain length necessary for tertiary structure formation [1, 11]. As an example the general topology of βαβ-folding units is illustrated in Fig. 1.

There is an increasing evidence that secondary and supersecondary structures play a dominant role in the process of protein folding [5, 6, 12]. Thus, the accessibility of polypeptides having the propensity to fold into a supersecondary structure-like three-dimensional conformation would provide a powerful tool in the elucidation of the mechanism of protein folding.
In the present contribution the design, synthesis and conformational properties of \(\beta\alpha\beta\)-type folding units are described. For the first time our recently proposed strategy for the construction of new proteins [1] is applied.

**Design of Artificial \(\beta\alpha\beta\)-Folding Units**

So far, the design of model peptides with the potential for tertiary structure formation was mainly based on conformational probability parameters for single amino acids [13–16].

Alternatively to this approach, in our concept for the design of ‘artificial proteins’ some of the basic rules governing protein folding are taken into account: Extensive investigations by molecular modelling revealed a high degree of homology for the topological features of \(\beta\alpha\beta\)-units in a variety of proteins with very different size and functions [17]. Especially the geometry of the parallel \(\beta\)-strands and the overlying \(\alpha\)-helix is highly conserved, pointing to the stabilization of the structure by hydrophobic interactions between strand and helix residues forming the interior core of the folding unit.

According to these foundations, amphiphilic oligopeptide blocks adopting stable secondary structures in solution are assembled in a way to match the arrangement of secondary structure blocks in naturally occurring folding units. The amphiphilic nature of the helices and \(\beta\)-sheets acts as a major driving force for intramolecular folding, resulting in the formation of a hydrophobic core and a hydrophilic surface of the \(\beta\alpha\beta\)-unit.

Due to the lack of any stabilization by long-range interactions the intrinsic stability of the helical and \(\beta\)-sheet segments is of utmost importance for the folding into a stable tertiary structure. For this reason the detailed knowledge of the conformational properties of the single secondary structure blocks is a basic prerequisite for the construction of artificial folding units. These informations were obtained by numerous investigations dealing with the secondary structure forming potential of amphiphilic oligopeptides [18–20]; for the present approach the following results are relevant:

(i) Oligopeptides with alternating hydrophilic and hydrophobic amino acid residues are capable for \(\beta\)-structure formation in aqueous solution if the hydrophobic residues are strong \(\beta\)-formers, e.g. \((\text{Leu-Ser})_n\), \((\text{Val-Thr})_n\), \((\text{Ile-Thr})_n\), \((\text{Ile-Gln})_n\). The critical chain length for secondary structure formation is in the range as observed for hydrophobic homooligopeptides [18, 19]. As an example the CD spectrum of TFA · H − Ile − (Gln − Ile)$_2$ − Gly − PEGM in water is shown in Fig. 2. The spectrum exhibits the typical properties for peptides adopting a \(\beta\)-sheet
conformation (strong negative Cotton effect at 216 nm).

(ii) The construction of stable amphiphilic helices is achieved by incorporation of a sufficient number of the strongly helix-inducing Aib residue on the hydrophobic side of the helical cylinder [20].

The above considerations resulted in the design of a model sequence MI with \( \beta \)-sheet segments consisting of alternating Leu and Ser residues [18] and the stable amphiphilic helix Pro-Ala-Aib-[Glu-Ala-Ala-Aib]$_2$-Ala-Aib [20] as the central part of the structure*:

\[
\begin{align*}
\beta & : H - [L-S-L-S-L-S-A] - G \\
\beta & : G - [S-L-S-L-S-L] - G-OH
\end{align*}
\]

Ser was chosen as hydrophilic amino acid within the \( \beta \)-structure blocks in order to enhance the solubility in water. On the other hand, Leu was selected as hydrophobic residue as a result of studies by molecular modelling [17], suggesting a somewhat greater flexibility in intramolecular packing compared to Val or Ile.

The lengths of the secondary structure blocks correspond to the average chain lengths of these segments within \( \beta \alpha \beta \)-units in proteins (4—7 for \( \beta \)-sheets, 10—15 for \( \alpha \)-helices) [22].

In the design of a second model MII optimization of intramolecular packing was performed by molecular modelling [17]. As starting sequence for the optimization procedure a model was chosen containing an amphiphilic helical segment with three Leu residues on its hydrophobic side (see above) and \( \beta \)-strands consisting of (Gln—Ile)-units. Minimization of conformational strain for the hydrophobic residues in the interior of the folding unit suggested partial substitution of Ile by Val, resulting in the following optimized sequence MII:

\[
\begin{align*}
\beta & : H - [K-V-Q-I-Q-V-K] - G \\
\alpha & : S - [Q-L-Q-Q-L-Aib-E-Aib-L-Q-Q-Aib-A] - E \\
\beta & : G - [E-I-Q-V-Q-I] - G-OH
\end{align*}
\]

Additionally, charged residues were incorporated at positions enabling favourable interactions between pairs of opposite charges. It could be shown by molecular modelling that salt bridges are formed between Lys-1 and Glu-29 as well as Lys-7 and the C-terminal Gly when the molecule adopts a \( \beta \alpha \beta \)-like conformation. The stability of the amphiphilic helical block in the central part of the molecule was guaranteed by the insertion of three helix-inducing Aib residues on the hydrophobic side of the helical cylinder [20].

* One letter abbreviations are used for amino acids according to [21].

\( G = \text{Gly}; A = \text{Ala}; I = \text{Ile}; V = \text{Val}; L = \text{Leu}; P = \text{Pro}; S = \text{Ser}; Q = \text{Gln}; E = \text{Glu}; K = \text{Lys}. \)
For the loop-regions residues with high frequency of occurrence in the loops of proteins were used in both peptides (Gly, Pro, Ser) [16]. Furthermore the incorporation of hydrophilic loops (4–7 residues in length) as found in proteins) supports favourable interactions with the hydrophilic solvent and enhances the solubility of the peptide in water.

**Peptide Synthesis**

The synthesis of model MI was performed according to the general principles of the Liquid-Phase Method for peptide synthesis [23]. The acid-labile BOC-protecting group was used for Nα-protection and Bzl groups were employed for the protection of the side chains of Ser and Glu. The recently developed base-labile CFm-anchoring group [24] served as a reversible link between polymer and peptide chain. After esterification of the anchor with BOC–Gly anhydride further chain elongation was achieved mainly by a segment condensation strategy using activated dipeptides as coupling units. Besides the reduced number of coupling steps and the prevention of diketopiperazine formation, the simultaneous incorporation of hydrophilic and hydrophobic residues, e.g., BOC–Ser(Bzl)–Leu–OH, BOC–Glu(OBzl)–Ala–OH, Boc–Leu–Ser(Bzl)–OH, ensures the correct arrangement of hydrophilic and hydrophobic amino acids within the amphiphilic secondary structure blocks. The use of BOC–Ala–Aib–OH accounts for any difficulties arising from couplings to the sterically hindered Aib-residue [25].

After completion of stepwise synthesis, the peptide was cleaved from the polymeric support by treatment with piperidine/DMF; side chain protecting groups were removed by catalytic hydrogenation and the peptide was subsequently chromatographed on Sephadex LH 20 and on silica gel. Amino acid analysis after acid hydrolysis gave the expected ratios. The deprotected PEG-peptide (MI–PEGM) was obtained by treatment of the fully protected precursor with HBr/TFA.

Special attention was drawn to conformationally induced changes in the physico-chemical properties of the growing protected peptide chain throughout the synthesis. The close relationship between the physico-chemical properties of a peptide and its preferred conformation is well established [26–28] and has been shown to be of fundamental importance for the practice of peptide synthesis [26, 29, 30]. For example, the formation of β-structures in solution leads to a drastic decrease in the solubility of the peptide as well as a reduced reactivity in the coupling reaction. These observations were also confirmed during the synthesis of peptide MI, where β-structure formation resulted in a strong decrease in the solubility for the nonapeptide [Ser(Bzl)–Leu]–Gly–CFm–PEGM; at the same time a significantly reduced sensitivity of the ninhydrin reaction was detected (about 50% loss compared to Ser(Bzl)–Leu–Gly–CFm–PEGM), reflecting the reduced reactivity of the terminal NH2-groups. The latter result points to the importance of the conformational properties also for analytical control procedures commonly used in polymer-supported peptide synthesis.

With further chain elongation a conformational transition β-structure → helix/r.c. was observed, resulting in a continuous improvement of the solubility of the peptide. For example, the amide I region of the IR-spectrum for the 24-peptide in CH2Cl2 is dominated by a strong absorption around 1675 cm⁻¹, a much weaker band at 1630 cm⁻¹ indicating only small amounts of β-structure. This pattern of the IR spectrum in CH2Cl2 as well as the solubility properties of the peptide remained essentially unchanged until the end of the synthesis.

Contrary to this finding, a further decrease of the reaction rates was observed; reaction times of 24 h or more were necessary in the last steps of the synthesis to ensure for complete coupling reactions. A possible explanation for this remarkable behaviour may be the intramolecular aggregation of different chain segments, thereby reducing the accessibility of the terminal NH2-groups. Further work concerning these questions is in progress and a detailed study dealing with conformationally induced coupling problems in polymer-supported peptide synthesis will be published elsewhere [31].

**Peptide MII** was built up by a solid-phase synthesis on a polystyrene-1% DVB resin. An orthogonal protecting scheme was applied using the base-labile Fmoc-group [33] for Nα-protection and the acid-labile t-butyl group for the protection of alcoholic and carboxylic acid side chain functions. The acid-labile HMPA group [34] was used as anchor between peptide and resin.

Peptide assembly was performed by DCC/HOBt mediated couplings of Fmoc–Gly–OH, Fmoc–Ser(OBu')–OH, Fmoc–Leu–OH,
Fmoc-Glu(OBu')—OH, Fmoc—Lys(BOC)—OH and of the dipeptides Fmoc—Gln—Ile—OH, Fmoc—Gln—Val—OH, Fmoc—Leu—Aib—OH, Fmoc—Ala—Aib—OH and Fmoc—Glu(BOBu')—Aib—OH. Incorporation of a single Gln was achieved by coupling of Fmoc—Gln—ONp in the presence of HOBt. After completion of the synthesis the Fmoc protected peptide was cleaved from the resin by 50% TFA/CH₂Cl₂ in the presence of anisole with simultaneous removal of all side chain protecting groups. The peptide material obtained showed satisfactory amino acid analysis; it was used for the conformational studies without further purification.

**Conformational Studies**

In Fig. 3 the CD-spectrum of peptide MI is shown in aqueous phosphate buffer at pH 7.

It is characterized by a dominating negative Cotton effect at 219—220 nm (n—π* transition) and a second one at 206—207 nm (π—π* transition), indicating a highly ordered conformation of peptide MI in aqueous solution even at very low concentrations.

![Fig. 3. CD-spectra of peptide MI in aqueous buffer pH 7, c = 7·10⁻⁷ M (-----) and 7·10⁻⁵ M (-----).](image)

Fig. 4. CD-spectra of peptide MI in 5% TFE/95% aqueous buffer pH 7 (-----), 50% TFE/50% aqueous buffer pH 7 (-----) and 100% TFE (-----).

It should be emphasized that (Ser—Leu)₃ is not able to adopt an ordered conformation under the experimental conditions used here [18]; thus the observed spectrum points to an *intramolecular* stabilization of this N-terminal block. Interestingly the spectrum is very similar to those found for α/β-proteins [34].

The addition of 5% TFE to the aqueous solution of MI brings along a red shift of the n—π* transition to 224 nm, a slight blue shift for the π—π* transition and an increase in ellipticity for both minima (Fig. 4).
However, the curve shape remains essentially the same as in pure aqueous solution. Raising the TFE content to 50%, a drastic change in the spectral properties is induced, pointing to an increase in the helicity of the peptide. The helix promoting properties of TFE as well as its destabilizing effect on the tertiary structure of proteins are well known [35]. For this reason, the spectral changes could be rationalized by a transition from a more globular structure of MI, stabilized by intramolecular hydrophobic interactions, to a more extended helical conformation in going from H₂O to TFE solutions.

The conformational studies of peptide MI in solution were complemented by IR-spectroscopic measurements in the solid state. Furthermore the conformational properties of PEG-bound MI (MI−PEGM) were studied in detail by Fourier Transform-Attenuated Reflection (FT−ATR)−IR-spectroscopy in the solid state and in solution.

The amide I region of the spectra in the solid state (KBr-pellet) is characterized by two distinct bands at 1661 cm⁻¹ and 1632 cm⁻¹ for MI as well as for MI−PEGM. This clearly demonstrates the coexistence of β-structure and α-helix within this model peptide [36, 37]. Thus, the polypeptide exhibits various types of ordered conformations in different parts of the chain.

The presence of different secondary structure types is also observed for MI−PEGM in aqueous solution. The amide I band of the spectrum in D₂O is centered around 1652 cm⁻¹, two distinct shoulders being present at 1671 cm⁻¹ and 1629 cm⁻¹. These spectral properties are in full agreement with a folded state of the molecule, the 1652 cm⁻¹ absorption being characteristic for the helical segment, the 1629 cm⁻¹ absorption for the β-structure and the shoulder at 1671 cm⁻¹ for the loop-regions.

Contrary to the behaviour in water, only one intensive absorption at 1662 cm⁻¹ is visible in TFE-d₃. This indicates an essentially helical conformation of the whole peptide under helix promoting conditions and confirms the results obtained by CD spectroscopy in TFE and TFE/H₂O.

Preliminary conformational studies of MII revealed valuable information concerning the tendency of this model peptide to adopt ordered conformations in aqueous solutions. The CD-spectrum in aqueous buffer at pH 8 is shown in Fig. 5.

It is characterized by a single minimum at 218 nm, reflecting a highly ordered conformation of the peptide. Most notably, a simple extended β-structure can be excluded due to the high amount of β-breaking amino acids [16] as well as the presence of three Aib residues [38]. Furthermore, the spectrum is insensitive to dilution (not shown), indicating intramolecular stabilization of the structure. The spectral changes induced by the addition of 50% TFE to an aqueous solution of MII are less pronounced than in the case of MI (Fig. 5). This points to a strong conservation of the conformation found in pure aqueous solution even in the presence of high amounts of TFE.

Interestingly the type of CD pattern observed for MII is also found for α+β- and even for α/β-proteins [34]. The somewhat different curve-shape compared to the spectrum of MI in pure aqueous solution may be attributed to structural rather than topological deviations of the two polypeptides.
Conclusions

The design and synthesis of two polypeptides with the potential to fold in a $\beta\alpha\beta$-type supersecondary structure has been achieved. The data of CD- and IR-spectroscopic investigations in aqueous solution are consistent with the adoption of a folded conformation, both molecules exhibiting high amounts of secondary structure. Most notably, the coexistence of $\alpha$-helix and $\beta$-structure has been confirmed for MI by IR spectroscopy. As the critical chain length for intermolecular $\beta$-sheet formation of the amphiphilic blocks is not attained under the experimental conditions, these findings can only be rationalized by a $\beta\alpha\beta$-type folding topology of the model peptide. Further support for the indicated structure comes from dilution experiments and conformation dependent properties like solubility or aggregation tendency.

The general strategy applied here for the first time — the construction of folding units by the assembly of amphiphilic segments with secondary structure forming potential — offers new perspectives for the design of polypeptides with tailor-made catalytic and immunological properties. Studies along these lines are in progress.

Materials and Methods

All amino acids are of the L-configuration. All solvents and reagents were of the highest purity available and, in the case of liquids, they were freshly distilled and stored over molecular sieves. Polyethylene glycolmonomethylether (HO—PEGM, $M_r = 5000$) was a product of Union Carbide (U.S.A.), chloromethylated polystyrene-1% DVB (1,34 mmol Cl/g) was purchased from FLUKA (Switzerland). The preparation of H$_2$N—PEGM (“amino-PEG”) and aminomethylated polystyrene was performed according to the literature [39, 40]. All coupling steps throughout the syntheses were tested for completion by ninhydrin and fluorescamine reactions [41].

Amino acid derivatives and dipeptides were synthesized by standard procedures [42—45] and checked for purity by means of tlc, IR, NMR and elemental analysis. CD spectra were recorded on a Yvon Jobin Mark V circular dichrograph using quartz cells of 0.5 and 1.0 mm pathlength. For conventional IR spectroscopy a Perkin Elmer model 781 infrared spectrophotometer was used, FT—ATR—IR-spectra were performed on a Bruker spectrometer model IFS 45 with a Ge crystal.

This work has been supported by the Swiss National Science Foundation, Hoffmann-La Roche Inc. and the Ciba Research Fund.


