Design, synthesis and solution structure of a helix–loop–helix dimer—a template for the rational design of catalytically active polypeptides

Susanne Olofsson,[a] Gunnar Johansson[b] and Lars Baltzer*[,a]
[a] *Department of Organic Chemistry, Göteborg University, 412 96 Göteborg, Sweden*
[b] *Department of Biochemistry, Biomedical Centre, University of Uppsala, Box 576, 751 23 Uppsala, Sweden*

A polypeptide with 42 amino acids, SA-42, has been designed to fold into a 'hairpin' helix–loop–helix motif. Its intended use is to serve as a template for the rational design of catalytically active polypeptides. SA-42 is made up of two amphiphilic helices, that are designed to interact through the hydrophobic effect and are connected by a short loop. Evidence is presented that it does adopt the intended motif and that it dimerises in an antiparallel mode at concentrations between 2 μmol dm^{-3} and 5.6 mmol dm^{-3}. The structural evidence was obtained from NMR and CD spectroscopy and the state of aggregation was determined with equilibrium sedimentation ultracentrifugation and CD spectroscopy. Special attention was paid to the design of SA-42 with respect to the expected NMR spectrum. The mean residue ellipticity at 222 nm of SA-42, −25 000 deg cm^2 $dmol^{-1}$, compares well with that of other published helix–loop–helices. Two polypeptides that were developed from SA-42 by the replacement of three and five residues, respectively, were recently shown to have catalytical activity.

De novo design and synthesis of polypeptides and small proteins with supersecondary and tertiary structures[1–3] shows great promise as a strategy for the development of complex molecules. Our own efforts at present, and those of others, focus on polypeptides with catalytic activity.[4–8] Other currently pursued goals are the design of artificial membrane ion channels,[9,10] immunogenic topologies[11] and metallobinding peptides.[12,13] So far the reported artificial polypeptides with supersecondary structure[4,6,8–17] have not attained the characteristics of native proteins, but this is not a prerequisite for catalysis.[5,7]

Only a small number of designed catalytically active polypeptides have been reported[4–8,18–20] and it appears that rational strategies are necessary. We have chosen a template approach, where the first task is to design a polypeptide that folds into a conformation with a well defined and well known structure. When that has been accomplished reaction sites can be introduced by minor modifications of the amino acid sequence. The sequence should contain a number of residues that can be replaced without affecting the supersecondary structure and it should be possible to study its structure and dynamics in solution. In the present investigation the helix–loop–helix motif was chosen because it can be made small enough to be synthesised by solid-phase peptide synthesis and large enough to form supersecondary structures. Polypeptides are well suited for studies of structure and dynamics by NMR and CD spectroscopy.

Sequences designed to form helix–loop–helix motifs have been reported previously.[11,21] CD spectroscopic studies have shown well developed helices in agreement with the formation of helix–loop–helix motifs and determinations of apparent molecular weight have demonstrated dimerisation.[21] In principle, the four different supersecondary structures that are shown in Fig. 1, are compatible with the reported results; the extended dimer, the parallel, the antiparallel and the interleaved 'hairpin' dimers. In addition, each motif has a 'front' and a 'back', which gives each hairpin a possibility to form a 'front-to-front', 'back-to-back', or 'front-to-back' dimer and the extended dimer can be in a 'head-to-tail' or a 'head-to-head' conformation. In order to develop further the helix–loop–helix motif more direct structural information must be accessible.

Structural information about polypeptides is difficult to obtain if they consist of more than a single helix. The methods used so far for structural studies, CD spectroscopy, equilibrium sedimentation ultracentrifugation[22] and small-angle X-ray scattering (SAXS),[11] can shed light on the degree of aggregation and on whether there is 'much' or 'little' helix. NMR spectroscopy has been used once[23] to demonstrate the supersecondary structure of a designed helix–loop–helix motif where a disulfide bridge was used as a conformational constraint and once to demonstrate the supersecondary structure of a designed helix–loop–helix motif.[24]

Detailed structural information can be obtained from NMR spectroscopy if the NMR spectrum can be assigned. NMR spectroscopy has not been used much in artificial polypeptide structure determination, mainly because of exchange broadening of the resonances, which in combination with the predominant use of repetitive sequences, the so-called minimalist approach,[25] make assignment of spectra difficult. To obtain more detailed structural information this problem must be addressed already in the design process, *e.g.* by extensive variation of the primary sequence and by incorporation of specific NMR probes and isotopically labelled amino acids. Peptide synthesis, unlike protein engineering, allows convenient introduction of artificial amino acids and specific isotopic labelling, which are of great use in the study of structure and dynamics. Exchange broadening of resonances can be reduced by small amounts of trifluoroethanol (TFE), not to induce structure, but to reduce overlap. The effect on polypeptide structure of short-chain alcohols such as TFE is to promote helix formation and, when added in large amounts, to disrupt tertiary structure.[26,27] In the present investigation the addition of small amounts of TFE was used to sharpen the NMR resonances to allow assignment of the ^1H NMR spectrum for extrapolation to aqueous solution. This provided a convenient assignment strategy as chemical shift changes as a function of fraction of cosolvent are usually approximately linear.

The subject of dynamics must also be addressed in the design of polypeptide catalysts because of its influence on the lifetimes of active conformations. The entropy contribution to the free energy of complexation and activation depends on the number of accessible conformations in the bound and transition states relative to those in the unbound and reactant states. Several processes can occur that will affect the lifetime of the motif. Each

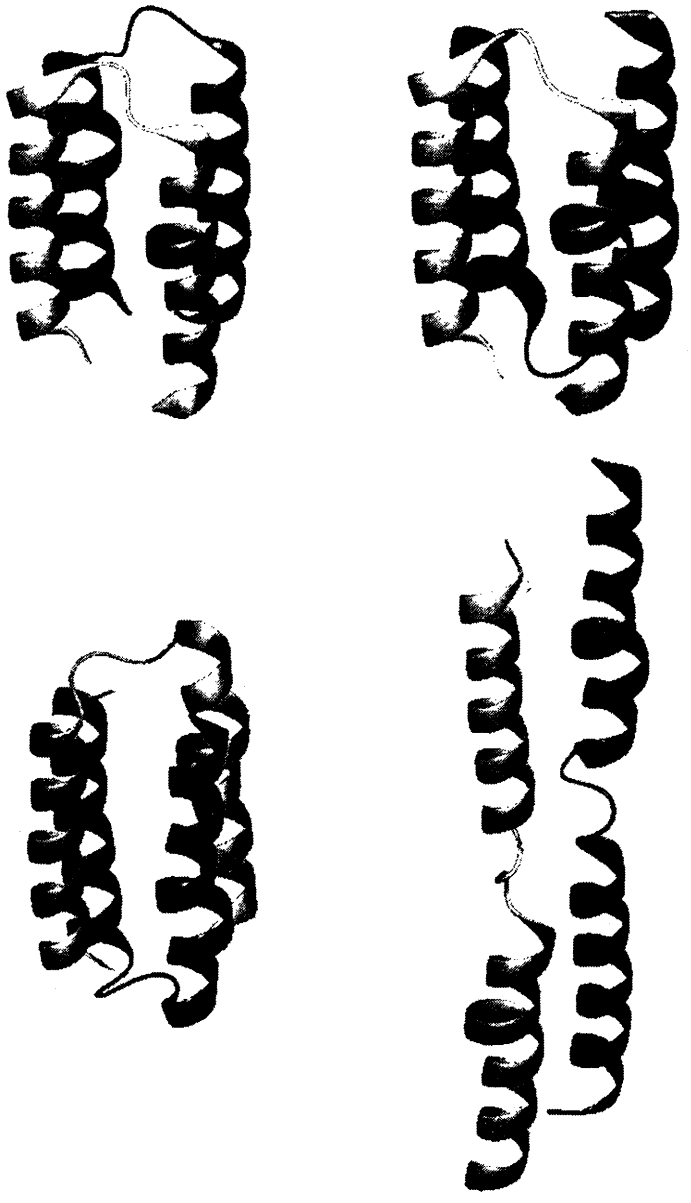

Fig. 1 Schematic representation of possible dimers of SA-42. With the parallel and antiparallel 'hairpin' dimers 'front-to-front', 'back-to-back' and 'back-to-front' dimers also have to be distinguished. The 'interleaved' dimer has two possible conformations and the extended dimer can be 'head-to-head' or 'head-to-tail'.

helix is inherently mobile due to imperfect hydrogen bonding in the backbone. In particular, the regions close to the loop and the peptide terminals can exhibit unordered structures, as helices are expected to 'fray' close to the termini. There is also a slow reorientation of helices relative to each other due to the ease with which the hydrophobic side chains can reorient. Without an understanding of the dynamics the design of binding sites becomes very uncertain. However, catalytically active polypeptides do not need to have protein-like stability. The catalytically active helical peptides[5] Oxaldie 1 and 2 clearly do not have the characteristics of a native protein.

In the present investigation we report on the design, synthesis and structural characterisation of a novel helix–loop–helix motif, SA-42, Scheme 1, that was designed to be a template for the introduction of amino acids capable of binding and catalysis, with a primary sequence chosen to allow investigation by NMR spectroscopy. Two polypeptides based on the sequence of SA-42 recently exhibited catalytical activity.

K-A-Aib-Nle-H-E-A-L-A-K-I-A-A-E-Nle-D-A-Aib-N
19 1

G-P-V-D-
20 23

Aib-A-Q-Nle-A-E-Q-L-A-K-A-F-E-A-F-A-R-Aib-G
24 42

SA-42

Scheme 1

Design of a helix–loop–helix motif

The design process must lead to an amino acid sequence that will fold into two helical regions, connected by a short loop, and the helices must interact intramolecularly to form the 'hairpin' helix–loop–helix motif. We chose to use amphiphilic helices designed to interact through the hydrophobic effect. The starting sequence for each helix was poly-Ala because of the helix-stabilising properties of Ala.[28] The amino acid sequence was then modified by replacing Alas with amino acids that were chosen based on their structural propensities, according to the Chou–Fasmans rules, as developed by Richardson and Richardson.[29] The 'artificial' amino acids Aib (α-amino-isobutyric acid)[30] and norleucine[31,32] were used for extra helix stabilisation and for further variation of the amino acid composition. Aib is found in positions 2, 17, 24 and 41, and Nleu is found in positions 5, 16 and 27. The choice of amino acids was carefully optimised in the helical termini to provide helix dipole stabilisation (Asp-4, Lys-19, Asp-23 and Arg-40)[33] and N- and C-terminal capping (Asn-1 and Gly-42).[34] Intra-helical salt bridges were also included in the design to stabilise the structure (Glu-6, Lys-10, Glu-14 and Glu-29, Lys-33, Glu-36, Arg-40).[35] After optimisation of the helices, a short loop was introduced. The sequence Gly–Pro–Val–Asp was used to initiate the folding of the helix–loop–helix 'hairpin' motif. The role of Gly-20 and Pro-21 was to break the helix and the role of Asp-23 was to stabilise the positive end of the helical dipole. Val-22 was introduced to fill up the cavity that would otherwise arise within the loop.

Amphiphilicity was achieved by placing hydrophobic residues along one side of each helix and charged or polar residues on the opposite sides. The main difficulty was to estimate the number of hydrophobic residues needed to enforce the inter-helical interaction while retaining polypeptide solubility, a problem that could only be solved experimentally. Previously, several polypeptides were synthesised with systematic variations of the hydrophobic properties in order to probe the minimum requirements for structure formation. In SA-42 the hydrophobic amino acids were placed in positions 5, 9, 12 and 16 in helix I and in positions 27, 31, 35 and 38 in helix II.

The hydrophobic amino acids and their positions in SA-42 were chosen for optimisation of the packing of the hydrophobic core and to allow investigation by NMR spectroscopy. In SA-42, Phe-35 and Phe-38 are the only aromatic residues and they are placed next to each other, in helix II, at the lower part of the motif, when the sequence takes a helix–loop–helix conformation, Fig. 2(*a*). Ring current effects can be substantial and proximity of key residues to the aromatic ring of Phe is not only expected to give rise to nuclear Overhauser effects (NOEs) to the aromatic protons, but is also expected to cause shifts from random coil or near random coil values. Since the aromatic protons of Phe appear in a free region of the ^{1}H NMR spectrum,

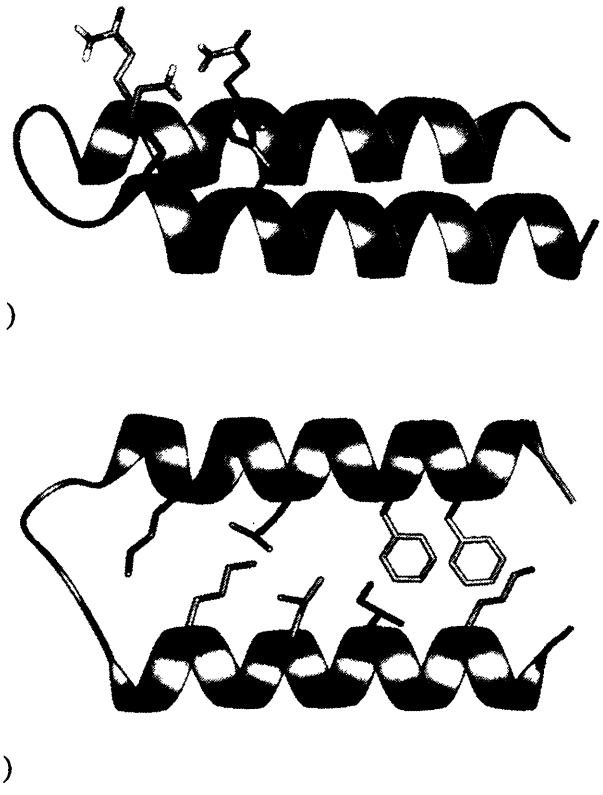

Fig. 2 (*a*) Modelled structure of SA-42 showing only backbone and side-chains of hydrophobic residues in the hydrophobic core. The design philosophy is clearly illustrated with two Phes at the 'lower' part of helix II. Proximity between Phe-35 and Leu-31 is obvious on helix formation, but the resulting NOEs are uninformative about super-secondary structure formation. Proximity between Phes and Ile-9 and Leu-12 gives rise to NOEs that provide conclusive evidence for inter-helical interactions. Observed proximity between Phes and Nle-16 and or Nle-27 is evidence for intermolecular interactions, due to the large distance from the lower part of helix II to the loop region, which is incompatible with intramolecular NOEs. (*b*) Modelled active site of SA-42 showing only backbone and reaction centre with sidechains of His-15, Lys-19, Gln-26 and Gln-30.

interactions with Phe are expected to be useful markers of structure. In the NOESY spectrum of SA-42, cross-peaks between the Phe aromatic protons and the hydrophobic methyl groups of Leu-31 are expected to appear because of their proximity in space on helix formation, but interaction with other methyl groups must be long-range in origin. Identification of these connectivities can then provide information about formation of the motif.

The overall design thus consists of two amphiphilic anti-parallel helices, connected by a short loop. The motif has a 'front' and a 'back' side free for the design of binding sites and catalytic functions. Although it was expected that the design of the active site on the 'front' side would have to be modified, attempts have been made already at this stage to introduce amino acid residues that would be complementary to meso-diesters of 3,3-disubstituted glutaric acids. Lys-19 and Gln-26 or Gln-30 were designed to interact with the carbonyl functions of the diester and the hydrophobic core of the peptide was intended to interact with bulky, hydrophobic groups in one of the 3-positions of the substrate. His-15 can then function as a general base or a nucleophilic catalyst for the ester hydrolysis, Fig. 2(*b*). The 'back' side of SA-42 was kept free from charged and bulky sidechains so that the 'back-to-back' dimer would be preferred in the case of dimerisation.

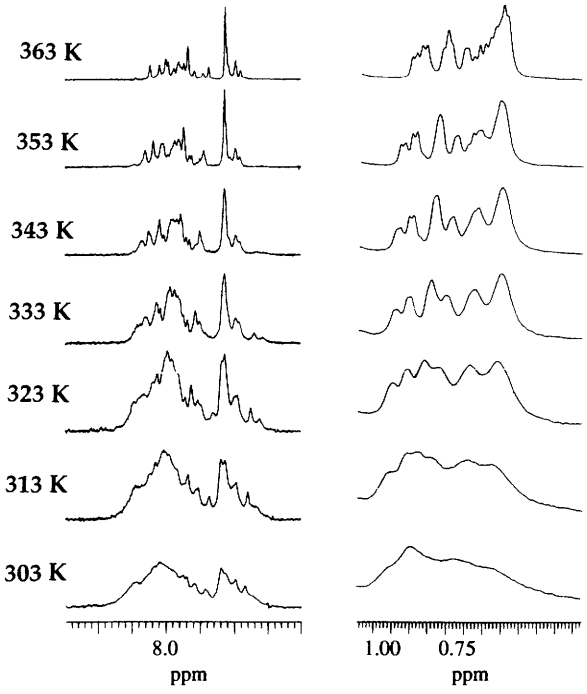

Fig. 3 Parts of the 500 MHz ^{1}H NMR spectrum of SA-42 in 90% H_2O–10% D_2O, pH 5.75, recorded at different temperatures from 303–363 K

Results

The *de novo* designed polypeptide SA-42, was successfully synthesised, purified and identified. Its identity was established from amino acid analysis and electro-spray mass spectrometry (ESMS). The amino acid composition, determined by amino acid analysis, was in good agreement with that expected for the amino acid sequence of SA-42. The molecular weight of 4289.5 g mol^{-1} determined by ESMS corresponds well to the theoretical value of 4289.9 g mol^{-1}. The purity was established from the appearance of the isocratic reversed-phase and ion-exchange HPLC chromatograms, where only single peaks were detected and from ESMS where only one distinguishable molecular ion could be observed in the transformed spectrum. The ^{1}H NMR spectrum of SA-42 in 90% H_2O–10% D_2O shows broadened resonances that arise from chemical exchange at an intermediate rate between several conformations. In order to optimise the resolution and increase the possibility of assigning the ^{1}H NMR spectrum and extract information about structure and dynamics, the temperature was varied from 303–363 K, Fig. 3. The best compromise between resolution and structural content was found to be at 323 K and the NMR studies were therefore conducted at that temperature.

The ^{1}H NMR spectrum of SA-42, Fig. 4, in 90% H_2O–10% D_2O at pH 5.75 and 323 K shows six resolved groups of methyl resonances. The NOESY spectrum in the same solvent shows cross peaks connecting the phenylalanine aromatic protons of Phe-35 and Phe-38 with at least four methyl groups from the hydrophobic core, Fig. 5. The chemical shift dispersion of the methyl region is probably due to the ring current effects of the phenylalanines. Complete assignment of the ^{1}H NMR spectrum is not possible in aqueous solution, but small additions of TFE greatly improve resolution. We have recorded the one dimensional ^{1}H NMR spectrum of SA-42 as well as the TOCSY and NOESY spectra in 0, 1.4, 2.8, 4.2, 5.6, 8.5, 12 and 14% of [$^2H_3$]TFE in 90% H_2O–10% D_2O (v/v). The sequential assignments were carried out in several of these mixtures. As an example, parts of the TOCSY and NOESY spectra of SA-42 in

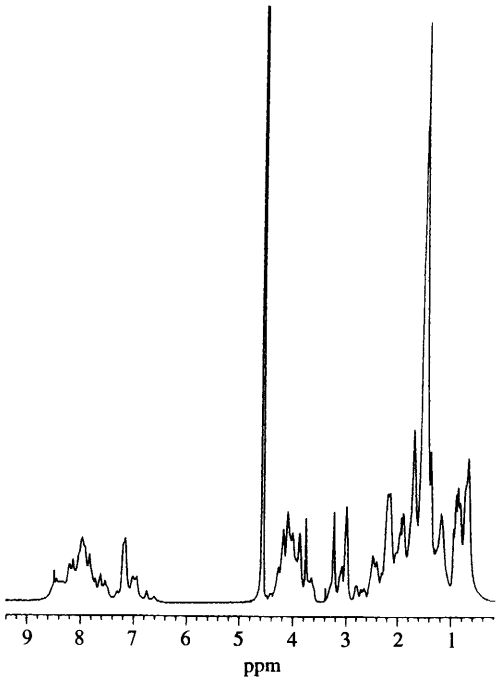

Fig. 4 500 MHz ^1H NMR spectrum of 2 mmol dm^{-3} SA-42 in 90% H_2O–10% D_2O pH 5.75 at 323 K

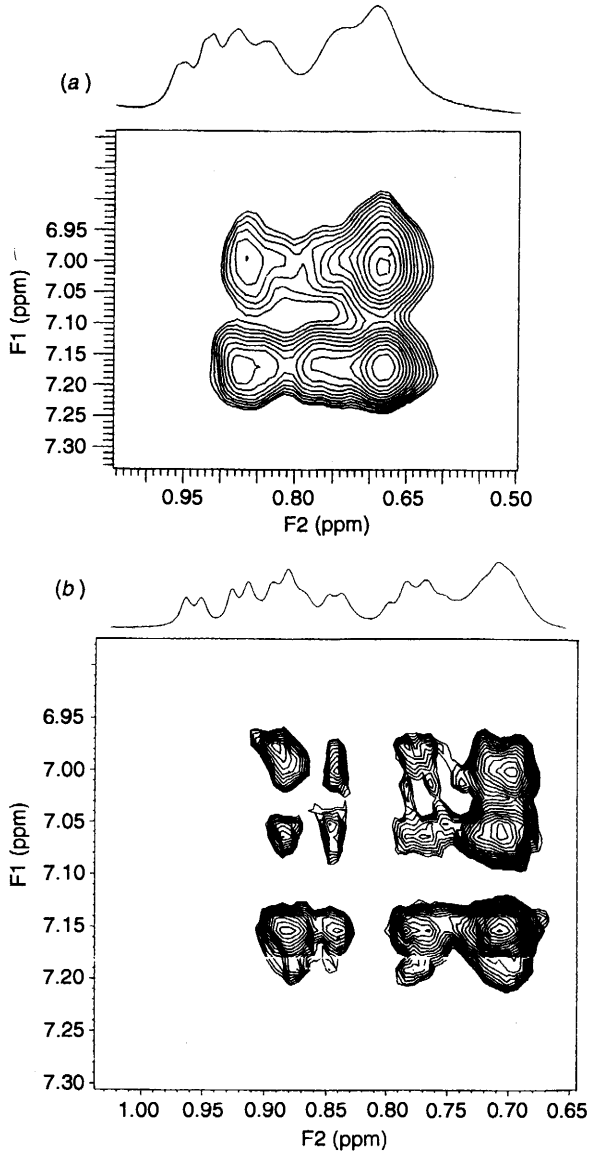

Fig. 5 (*a*) Part of the two-dimensional NOESY spectrum of SA-42 in 90% H_2O–10% D_2O at 323 K and pH 5.75, showing NOEs between aromatic Phe protons and methyl groups in the hydrophobic side chains. Cross-peaks are clearly visible between Phe aromatic protons and four different groups of resonances. (*b*) The corresponding part of the NOESY spectrum of SA-42 in 4.2% TFE in 90% H_2O–10% D_2O at 323 K and pH 5.75. The resolution is better and in particular the resonances of Nle-16 at 0.79 ppm and Nle-27 at 0.77 ppm are resolved providing evidence for intermolecular connectivity, see text. For assignment of methyl groups see Fig. 7.

14% [$^2H_3$]TFE are shown in Fig. 6. The spin systems are clearly observable in the TOCSY spectrum. Most of the sequential assignments were directly obtainable from the NOESY NH to NH crosspeaks. The assignment was simplified by the fact that SA-42 by design contains no repetitive sequences, *e.g.* there are only two phenylalanines in the sequence and identification of a segment of eight amino acids containing these two Phe and one Ala three residues away from one of them immediately provides the sequential assignment of all eight. Some readily identifiable amino acids (Arg, Ile, Pro, Val) occur only once and some occur in less crowded regions of the spectrum (Nle, Leu, Gly). Assignment of the methyl region was carried out in the series of solvents described above where the TFE content has been systematically varied; the results are shown in Fig. 7. The chemical shifts of the methyl groups of Leus, Iles and Nleus in aqueous solution were assigned from extrapolation. NOEs connect Phe aromatic protons and the methyl groups of Leu-12 and possibly also Ile-9. NOE contacts between the methyl groups of Nle-16 and or Nle-27 are clearly visible. NOEs to Leu-31 from the intra-helical neighbouring sidechain of Phe-35 are obviously strong. Potential NOEs to Nle-5 and the upfield methyl group of Ile-9 are therefore swamped by the strong NOE to Leu-31.

The apparent molecular weight in solution was determined at 1.3 mmol dm^{-3} concentration (based on an extinction coefficient of 300 dm^3 mol^{-1} cm^{-1} at 254 nm) and room temperature to be 7400 ± 500 g mol^{-1}, which after compensation for systematic errors due to electrostatic repulsion in the salt free solution corresponds well to the molecular weight of the dimer, 8579 g mol^{-1}. The systematic errors lead to experimentally determined molecular weights that are too low by 5–10%.

Quantitative determination of the mean residue ellipticity[36] of SA-42, Fig. 8, in aqueous solution at pH 5.75 gave a value of −25 000 ± 1000 deg cm^2 $dmol^{-1}$, a value that compares well with those of other published peptides, Table 1. The spectrum was recorded at room temperature in a 0.1 mm cell at a peptide concentration of 0.93 mmol dm^{-3}. The spectral changes upon variation of the concentration between 2 μmol dm^{-3} and 5.6 mmol dm^{-3} were shown to be negligible.

The denaturation process is conveniently monitored by CD spectroscopy and the temperature dependence of the CD spectrum in water is shown in Fig. 9. At 293 K the ellipticity is −25 000 deg cm^2 $dmol^{-1}$ and at 323 K it is still as large as −21 150 deg cm^2 $dmol^{-1}$. This corresponds to a loss of helical content of less than 15%. Consequently, the NMR studies that were conducted at 323 K were carried out at a temperature where the secondary structure is largely conserved. The ellipticity at 358 K was −15 300 deg cm^2 $dmol^{-1}$. The peptide does not exhibit a sharp melting point, in agreement with what can be expected of a peptide that is not as well packed as a native protein. An isodichroic point can be observed at 203 nm, suggesting a two-state equilibrium, with a cooperative

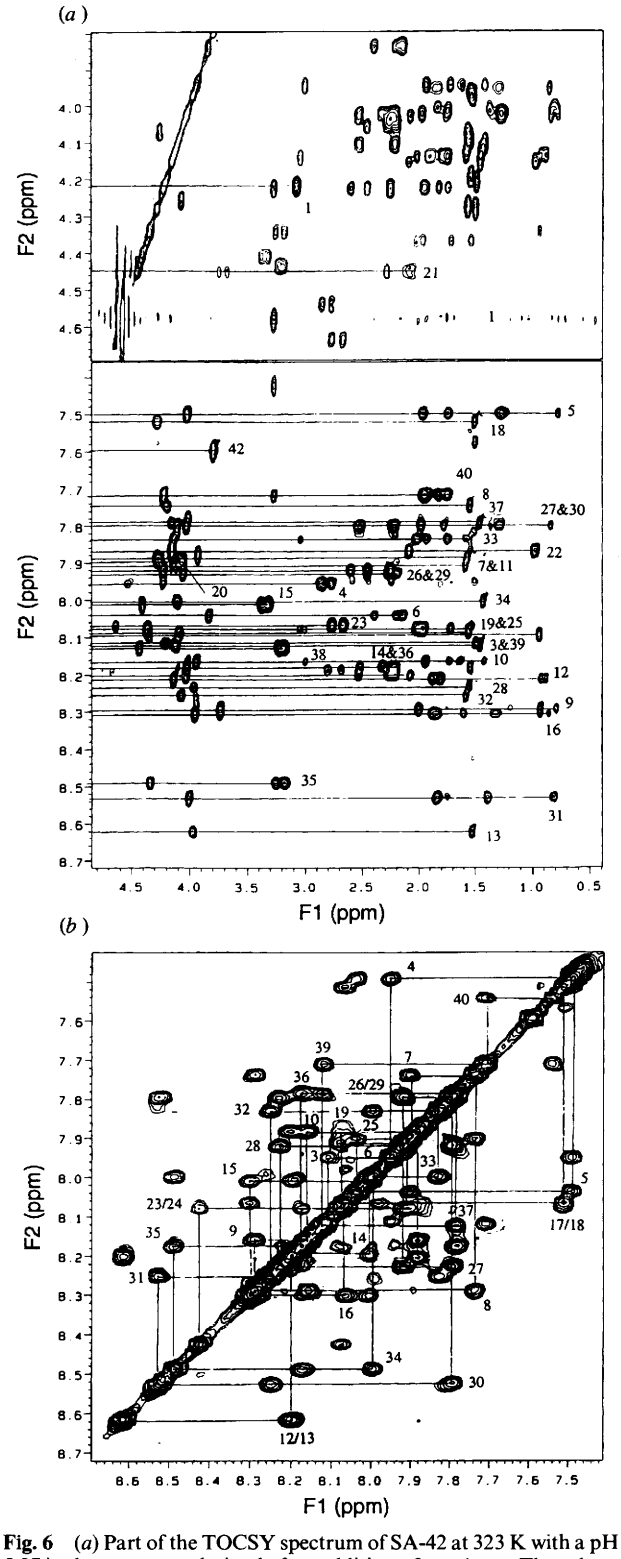

Fig. 6 (*a*) Part of the TOCSY spectrum of SA-42 at 323 K with a pH of 5.37 in the aqueous solution before addition of cosolvent. The solvent is 14% (v/v) of [$^2H_3$]TFE in 90% H_2O–10% D_2O, and the spectrum was recorded with a spin-lock time of 80 ms. Most of the spin systems are clearly visible in the lower spectrum, with the exception of the four Aibs that have no α-protons and some of the N- and C-terminal amino acids where the protons are not observable due to fast exchange. The spin systems are labelled with the number of the residue in the amino acid sequence. (*b*) Part of the NOESY spectrum of SA-42, recorded under the same conditions as in (*a*), using the same sample and with a mixing time of 250 ms. The amide part of the spectrum is well resolved and most of the sequential assignments of SA-42 can be carried out directly from the NH to NH NOEs. The cross peaks that connect pairs of amide protons are labelled with the number of the residue that has the lowest number in the sequence.

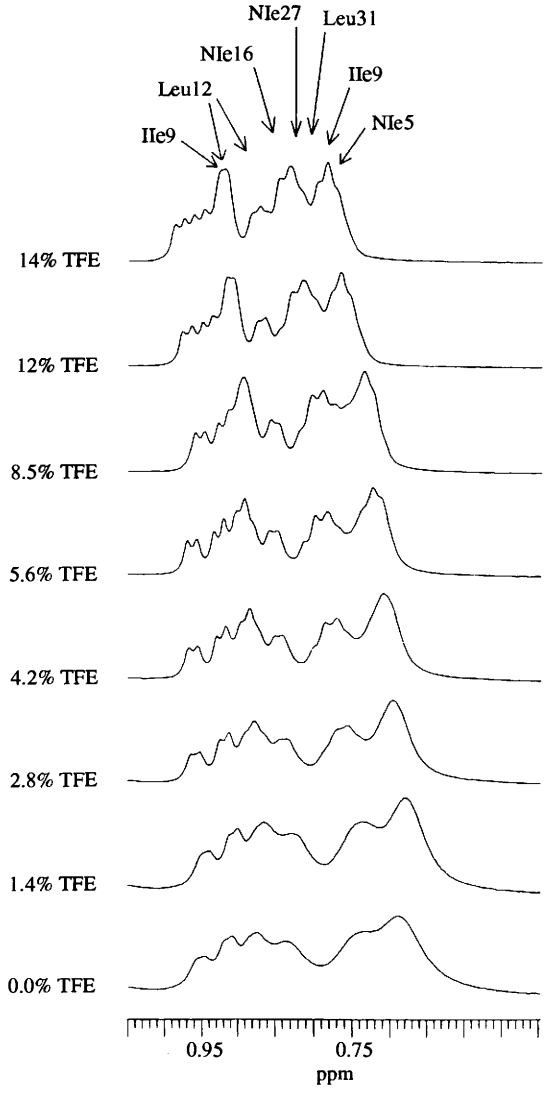

Fig. 7 Assignment of the methyl groups from the hydrophobic core in the ^1H NMR spectrum of SA-42 from TFE titration experiments at 323 K. [$^2H_3$]TFE was added to a solution of SA-42 in 90% H_2O–10% D_2O, with a pH of 6.3 after addition. Unequivocal assignments have been carried out in several TFE–water mixtures, and the chemical shifts vary roughly linearly with the fraction of TFE. The chemical shifts are shown in the top spectrum and the assignment in aqueous solution follows from extrapolation. The amounts of TFE that have been added are small; 14% of TFE by volume corresponds to 4.2 mol%. The effect of TFE is to reduce the linewidth but also chemical shift dispersion. At larger additions of TFE the latter effect will dominate, making spectral assignments impossible.

Table 1 Hydrophobic and charged residues in designed helix–loop–helix motifs

Peptide and no. of residues	Hydrophobic residues	Charged residues	$\theta(H_2O)$/ deg cm^2 dmol^{-1}
ACL-28[a]	3	6	—
SA-29[b]	4	6	—
ACL-41[a]	8	10	−18 000
SA-42	9	10	−25 000
GTD-43[c]	11	9	−21 000
α-2B(PRR)-35[d]	12	18	−21 000
α_3-40[e]	17	14	−22 485

[a] A.-C. Lundh and L. Baltzer, unpublished. [b] S. Olofsson and L. Baltzer, unpublished. [c] G. Dolphin and L. Baltzer, unpublished. [d] Ref. 21. [e] Ref. 11.

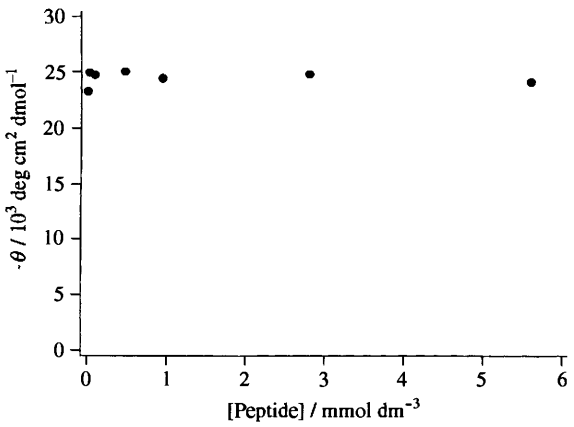

Fig. 8 The mean residue ellipticity at 222 nm of SA-42 as a function of concentration at pH 6.27 and ambient temperature. The ellipticity was determined at 5.6, 2.8, 0.94, 0.47, 0.093, 0.023 and 0.002 mmol dm^{-3} concentration of peptide. Only the value measured at 0.002 mmol dm^{-3} deviates measurably from the average. The mean residue ellipticity at 222 nm and 293 K is $-25\,000$ deg cm^2 dmol^{-1}.

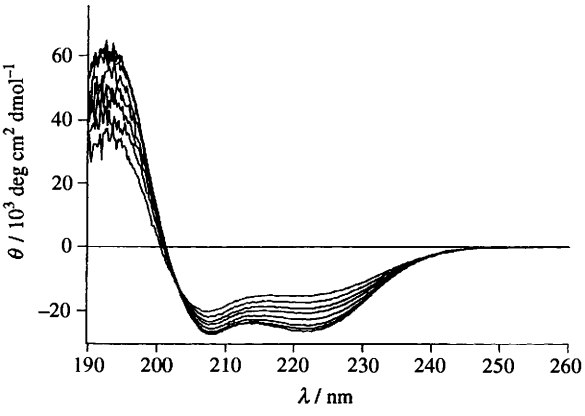

Fig. 9 The CD spectrum of SA-42 in aqueous solution at pH 5.75 as a function of temperature. From bottom to top at 222 nm the temperatures are 278, 288, 298, 313, 328, 338, 348 and 358 K. Note the isodichroic point at 203 nm, which is indicative of a two-state transition.

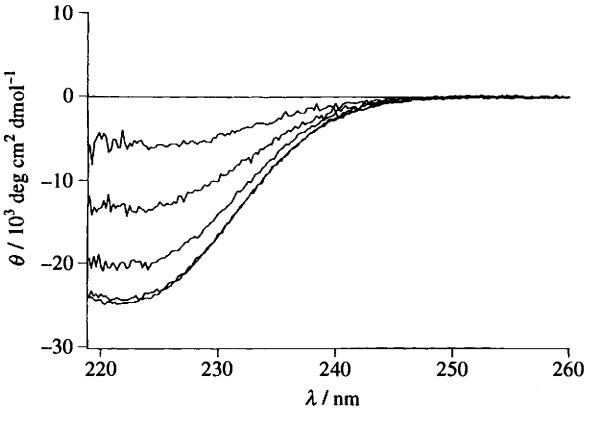

Fig. 10 The mean residue ellipticity of SA-42 in solutions of guanidine hydrochloride at ambient temperature. The concentrations of denaturant used were 0, 2.0, 3.0, 4.0 and 4.5 mol dm^{-3}.

transition between the unfolded (random coil) and folded states.[37]

Further investigations of stability were conducted by studying the denaturation in guanidine hydrochloride solution, Fig. 10. Under the assumption that the ellipticity of the denatured monomer is zero, the concentration needed to achieve half denaturation was close to 4 mol dm^{-3} at 60 μmol dm^{-3} peptide concentration, a guanidine hydrochloride concentration that was slightly lower than that reported by Ho and DeGrado[21] for α_2B.

The one-dimensional NMR spectrum of SA-42 was recorded at 0.05, 0.1, 0.5 and 3 mmol dm^{-3} concentration to investigate further possible aggregation, showing that the concentration dependence is very weak. Careful inspection shows a small change in the resolution of the methyl region above 0.5 mmol dm^{-3}, but more than one explanation can be invoked for this minor effect, see Discussion. To ascertain that the observed NOEs were not an artifact assignable to spin diffusion, the NOE build-up curves were recorded in two different mixtures of TFE and water. No signs of spin diffusion were detected with mixing times of 50–250 ms. There was no detectable dependence on ionic strength and pH in 0–0.2 mol dm^{-3} phosphate buffer and pH between 4.4 and 8.

Discussion

The development of polypeptides with supersecondary structure for binding and catalysis is of interest because of the ease of synthesis and the great variety of amino acids, natural and 'artificial', that can be introduced into the sequence. Systematic investigations of the effect of structure on reactivity are possible with a wide range of geometries and combinations of functional groups. The structure depends on the formation of non-covalent bonds and the main difficulty is the prediction of the three-dimensional structure from the amino acid sequence, the so-called folding problem. This problem should, however, be encountered only once if a template can be developed that folds into a supersecondary structure in solution and where a number of 'exchangeable' amino acids are not directly involved in the formation of structure. As long as the 'structural' amino acids are not replaced the polypeptide forms the same supersecondary structure with a variety of 'functional' amino acids.

We have chosen to use the helix–loop–helix motif as a template because it is large enough to allow design of many types of binding sites and because it is the simplest possible combination of secondary structures. In order to probe its usefulness spectroscopic investigations have been performed of its solution structure and of its thermodynamic and conformational stability.

The ^{1}H NMR spectrum of SA-42, Fig. 7, shows broadened resonances with much overlap which is typical of polypeptides with supersecondary structure that are not as well packed as native proteins but more restricted in mobility than small peptides.

The ^{1}H NMR spectra of SA-42 in mixtures of TFE and water have been assigned, Fig. 6. Thirty-seven spin systems were identified from the TOCSY spectrum and the sequential assignments were obtained from NH–NH, αH–NH and βH–NH connectivities from the NOESY spectrum, Figs. 6 and 11. The secondary structure was identified from short- and medium-range NOEs and from the chemical shifts of the α-protons, Fig. 11. Connectivities typical of helical structure were observed in the segments from residues 3 through 20 and from 23 through 41, whereas the segment from residue 21 to 22 is clearly not in a helical conformation, in agreement with the observed chemical shift deviations from the random coil values of the α-protons of the amino acid residues.

The identification of supersecondary structure comes from observation of NOEs between the aromatic protons of Phe-35 and Phe-38 in helix II and some of the methyl groups of Nle-5, Ile-9, Leu-12, Nle-16 and Nle-27, Fig. 5.

An NOE connecting Phe aromatic protons and the methyl

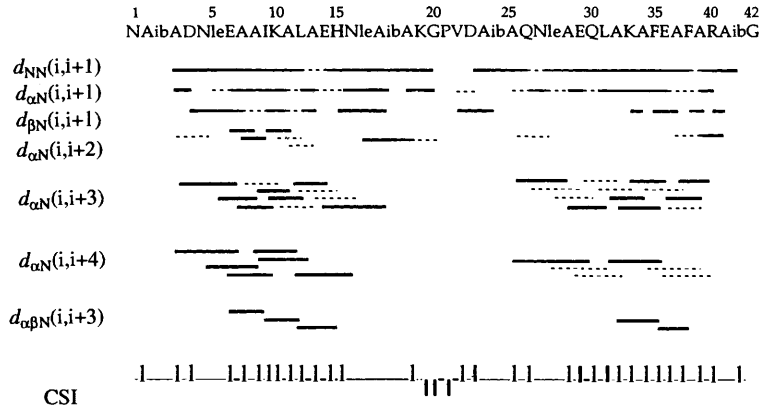

Fig. 11 Schematic representation of short- and medium-range NOEs[46] of SA-42 and of α-proton chemical shift deviations in SA-42 from those observed in random coil peptides.[47,48] The observed short-range NOEs connecting NH, αH and βH in successive amino acid residues, d_{NN}(i,i + 1), $d_{\alpha N}$(i,i + 1) and $d_{\beta N}$(i,i + 1), support the sequential assignments of the ^{1}H NMR spectrum and show that the loop region is not helical. The observed medium range (i,i + 2, i,i + 3 and i,i + 4) NOEs that connect αH and NH and αH and βH only appear in spectra of helices and show that SA-42 contains two predominantly helical regions. Chemical shift indexes are based on the observation that αH chemical shifts are shifted upfield in helices and downfield in β-structures and β-turns. Chemical shift deviations are not given for the artificial amino acids as Aib does not have α-protons and since reliable random coil chemical shifts have not been available for Nle. The size of the columns do not reflect the magnitude of the shifts.

groups of Leu-12 is observed at 0.83 ppm, and is unequivocal, whereas the crosspeak at 0.88 ppm can involve Leu-12 or Ile-9, or both. The chemical shift of the downfield methyl group of Ile-9 is close to that of Leu-12 and it is difficult to discriminate between the two. These crosspeaks provide unequivocal evidence for inter-helical connectivity. There is also at least one NOE to one of the methyl groups of Nle-16 or Nle-27, which overlap, but there may be NOEs to both. They resonate slightly downfield of Leu-31 at 0.74 ppm and the observable NOE demonstrates intermolecular interaction as the intramolecular distance is too large to give rise to an NOE, the modelled distance from Phe-35 to Nle-16 is more than 10 Å in the 'in register' conformation. An 'out of register' conformation will decrease the intramolecular distance from Phe-35 to Nle-16, but not from Phe-35 to Nle-27. In 4.2% TFE (v/v) in 90% H_2O–10% D_2O it is clearly demonstrated that NOE contacts exist between Phe aromatic protons and both Nle-16 and Nle-27, Fig. 5(*b*). The strong NOE at 0.68 ppm does not provide any long-range information, as it is probably due mainly to the interaction between Leu-31 and Phe-35 and it also prohibits the identification of possible interactions between Phe aromatic protons and the methyl groups of Ile-9 and Nle-5. No other intermolecular cross peaks were detected. The interface between the peptides contains mainly alanine residues that give rise to resonances within a narrow chemical shift range. Cross peaks that connect alanines therefore appear close to the diagonal and are difficult to observe.

The CD spectrum of SA-42 shows that it is largely helical. The measured ellipticity for SA-42, −25 000 deg cm^3 dmol^{-1} at 222 nm, can be used to estimate the fraction of secondary structure in solution. Using the equation of Chen *et al.*[41] and a mean residue ellipticity at 222 nm of −39 500 deg cm^2 dmol^{-1} for a completely developed and infinitely long helix[42] the estimated helical content of SA-42 in aqueous solution is more than 70%. This value is probably too low since the contribution from the loop region is expected to be positive and give a reduction in the apparent helical content. Taking into consideration the expected fraying of helix ends the observed ellipticity corresponds to that of well developed helices. Comparison with other published peptides, Table 1, shows that it compares well.

The apparent molecular weight of SA-42 shows that it exists as a dimer at 1.3 mmol dm^{-3} concentration. The concentration independence of the mean residue ellipticity at 222 nm between 2 μmol dm^{-3} and 5.6 mmol dm^{-3} shows that no further aggregation or dissociation takes place, and that SA-42 is a dimer over the concentration range studied here. An upper limit to the dissociation constant can be estimated from the fact that no decrease in ellipticity can be detected even at 10 μmol dm^{-3}, the first tendency towards a decrease (*ca.* 5%) was seen at 2 μmol dm^{-3} concentration although the deviation was within the limits of experimental error. Under the assumption that the monomer has less ellipticity than the dimer, one can conclude that the dimer is the dominant species at 2 μmol dm^{-3} concentration and that K_{diss} is smaller than 10^{-6} mol dm^{-3}.

Helix–loop–helix dimers have previously been reported twice and the evidence for dimerisation was based on CD spectroscopy, ultracentrifugation and SAXS data. The reported dissociation constants, K_{diss} for helix–loop–helix dimers are 5×10^{-10} and 2.4×10^{-6} mol dm^{-3}.[21,11]

The long-range NOEs show that the dominant conformation in solution is the hairpin helix–loop–helix that dimerises in an antiparallel mode. Inter-helical NOEs between Phes and Leu-12 and Ile-9 are compatible with this motif, but not with the formation of the extended 'head-to-head' dimer. NOEs between Phes and Nle-16 or Nle-27, or both, are compatible with the designed motif, but incompatible with formation of the extended 'head-to-tail' conformation which should not show NOE contacts between the end of helix II and the loop region. These NOEs also exclude the parallel 'hairpin' motif, for the same reason. The interleaved structure cannot be excluded on these grounds and may have to be considered a possibility. The short loop of SA-42, however, makes this structure unlikely because the helical regions close to the loop would have to be unwound to accommodate the larger inter-helical distance. The antiparallel dimer therefore remains the most probable motif that is consistent with the experimental results.

The thermodynamic stability of a template must be such that modifications of the amino acid sequence can be accommodated without affecting its thermodynamic stability. The 'exchangeable' positions in SA-42 are mainly occupied by alanines, the amino acid that is the most helix stabilising of all the natural ones. The replacement of Ala by any other amino will probably reduce the stability of SA-42 by at least a kcal mol^{-1}. SA-42 is therefore overdesigned to be able to accommodate several amino acid replacements. The guanidine

denaturation experiment can be used to calculate the free energy of unfolding in aqueous solution[43] and from the data in Fig. 10, it is estimated to be 5 ± 1 kcal mol^{-1}. This is the 'expendable' free energy that can be used to accommodate a number of amino acids that are less favourable in a helix than alanine.

The conformational stability is best probed with NMR and CD spectroscopy. The line broadening in the ^{1}H NMR spectrum and the poorly defined melting point of SA-42, as measured with NMR and CD spectroscopy, shows that it is not a single conformation peptide. The concentration independent CD spectrum shows that the fraction of monomers with less well developed helices are negligible and the dynamic processes are therefore probably restricted to rearrangements within the dimer or to exchange between dimers. The large helical content shows that there are no appreciable amounts of other secondary structural elements. Internal rearrangements of helices relative to each other are therefore the most likely cause of the line broadening. Aliphatic sidechains have low barriers to rotation around carbon–carbon bonds and several similar conformations are conceivable, but in terms of preorganisation of binding sites the movement of helices is probably less detrimental to binding than that of the sidechains of amino acids that make up the binding site.

The main uncertainty in the design of SA-42 was the prediction of the number of hydrophobic residues that would provide enough driving force for the random coil helix–loop–helix equilibrium towards the latter, without making the peptide insoluble. The packing of the hydrophobic core is probably also the most important factor in achieving thermodynamic and conformational stability. In our search for an adequate solution we designed helix–loop–helix motifs with 28 and 29 amino acids, with one or two hydrophobic residues per helix, that turned out to have little structure in water. DeGrado *et al.* designed a motif, α_2B(PRR), with 35 residues (12 hydrophobic) and Kaumaya *et al.* designed α_3 that had 40 residues (17 hydrophobic), whereas our present design has 42 residues, but only 9 hydrophobic ones. α_2B(PRR) has 18 charged residues to provide amphiphilicity and solubility, α_3 has 14 charged residues and our design has 11 charged residues including His-15, N- and C-terminals not counted. A comparison of ellipticities shows that roughly similar values are obtained, −21 000 deg cm^2 dmol^{-1} for α_2B(PRR), −22 485 deg cm^2 dmol^{-1} for α_3, and −25 000 deg cm^2 dmol^{-1} for SA-42. It is possible that with the shorter sequence, α_2B(PRR), a larger fraction of hydrophobic residues is needed to ensure structure formation, which then requires a relatively large number of charged residues for the sake of solubility. Even α_3, with 40 amino acids has a very large fraction of amino acids that are hydrophobic, and that may require several charged ones for solubility.

The *de novo* design of a helix–loop–helix motif was developed by DeGrado *et al.*, as an intermediate on the way towards a single-chain tetrahelix bundle. The same motif was also later designed by Kaumaya *et al.* with the objective of creating an immunogenic determinant. The present helix–loop–helix motif, SA-42, was designed to be a template for the construction and systematic variation of binding sites for the introduction and study of catalysis. The main difference between our approaches is that we already in the design have considered and introduced various probes for the study of structure and dynamics by NMR spectroscopy that represents the necessary step for further development of artificial polypeptides. We have also attempted to maximise the number of amino acids that do not have a purely structural role, to obtain maximum design flexibility. SA-42 has less than 50% of the total number of residues designated for structural purposes and both 'faces' free for other functions.

Conclusions

De novo design and synthesis of polypeptides amounts to addressing the folding problem. At the present level of understanding, well-packed, single conformation proteins are not to be expected, although it is clearly the ultimate goal in *de novo* design. However, protein-like stability is not necessary to obtain catalytic activity. The two peptides Oxaldie 1 and Oxaldie 2,[5] containing 14 amino acids each, show rate enhancements over simple amines with factors of 10^2–10^3 in the decarboxylation of oxaloacetate. Their respective helical content in aqueous solution is 18% and 33% at infinite dilution. Although these numbers depend on concentration and solvent they are far from single-conformation peptides under any circumstances. SA-42 folds into a helix–loop–helix motif that dimerises in aqueous solution and the measured mean residue ellipticity shows that the helical content is very large and that the population of monomers is negligible. The dynamic nature of SA-42 is therefore not likely to prevent the development of reaction sites. Two polypeptides based on the sequence of SA-42 have in fact shown catalytical activity. There are several positions on the solvent exposed surface of the motif where strongly interacting residues can be introduced without affecting the supersecondary structure. The conformational stability that was measured by guanidine denaturation is enough to accommodate the replacement of Alas for less helix stabilising amino acids. These modifications can also be assessed for structural effects and for effects on the overall stability by NMR and CD spectroscopy.

Experimental

SA-42 was designed, with the aid of computer graphics and molecular modelling software (Silicon Graphics personal Iris 4D-TG35 and the Quanta software package) to assume a helix–loop–helix motif with two anti-parallel, amphiphilic helices, connected by a short loop. Computer modelling included sequence generation, assignment of helical conformation for the separate helices, manual docking of the helices for the best fit possible in the hydrophobic core, introduction of the loop and minimisation of the potential energy of the complete structure before any dynamic simulations were performed. The result from the dynamic simulations, at elevated temperatures (300–500 K) in vacuum and in water (10 Å water shell), was interactively used in the design process. Special attention was paid to the packing of the hydrophobic core and the stability of the helical termini.

SA-42 was synthesised by solid-phase peptide synthesis on a Biosearch 9600 automated peptide synthesiser using *t*-BOC α-amino protection groups[44] on a Pam-linked resin, with a substitution level of 0.3 mmol Gly per gram of polymer. The following side-chain protection groups were used, benzyl-, (Asp, Glu), benzyloxymethyl-, (His), 2-Cl-CBZ-, (Lys), and tosyl-, (Arg). Polymer (2 g) was introduced into the reaction vessel and allowed to swell in DCM for several hours before the start of the synthesis. The α-amino groups were deprotected in 45 vol% TFA in DCM containing 1% anisole (v/v) as scavenger (two 1 min prewashes followed by a 30 min deprotection step), and neutralised in 10% DIPEA (v/v) in DCM. In the first five cycles the reaction vessel was removed and shaken manually since the 'stickiness' of the polymer gave rise to extensive formation of lumps.

Amino acid couplings were mediated by DIPCDI and HOBt in a mixture of 33% DMF and 67% DCM. A mixture of the amino acid, HOBt and DIPCDI was preactivated before addition to the polymer in the reaction vessel. Gly, Arg, Asn and Gln were used in the amounts of 2.4 mmol together with 0.6 mmol HOBt and 0.6 mmol DIPCDI. All other amino acids

were used in the amounts of 3.3 mmol with 0.3 mmol HOBt and 0.6 mmol DIPCDI. Gly, Arg, Asn, Gln, Glu-29, Ala-25 and all Aibs were preactivated for 10 min and all others were preactivated for 2 min.

The coupling reaction was monitored where possible with the Kaiser test.[45] A double coupling of 2 × 2 h was used for amino acids where the reaction could not be monitored and during nights. The couplings that could be monitored were allowed to continue until completion. No coupling reactions were allowed to continue for more than 2 h. The coupling reactions of Arg-40, Asp-23, Val-22, Gly-20, Nle-16-Ala-13 and Asn-1 could not be followed by the Kaiser test. The couplings of Aib-41, Ala-39-Glu-36, Gln-26-Aib-24, Lys-19-Aib-17, Lys-10-Ile-9, Glu-6 and Ala-3-Aib-2 could be followed by the Kaiser test. The remaining couplings were not tested.

Unreacted polypeptide chains were terminated by reaction with 0.3 mol dm^{-3} AcIm in DMF.

SA-42 was cleaved from the resin and deprotected by anhydrous HF according to the 'low-high' method, on a Teflon vacuum line (Peptide Institute, Inc.). Dimethyl sulfide (65 vol%) and *p*-cresol (10 vol%) were used as scavengers in the 'low' cleavage reaction (2 h at 273 K) and *p*-cresol [10% (v/v)] was used as scavenger in the 'high' cleavage reaction (1 h at 273 K). A total volume of *ca.* 6 cm^3 was used per gram of reacted polymer. Scavengers *etc.*, were removed by extraction with dry diethyl ether. The peptide was extracted from the polymer with 1 mol dm^{-3} acetic acid and lyophilised.

The peptide was purified by size-exclusion liquid chromatography (Sephadex G-25, aqueous solution) reversed-phase (Kromasil C-8, Eka Nobel, 42% propan-2-ol in 0.05% trifluoroacetic acid) and ion-exchange HPLC (POROS Q/H, PerSeptive Biosystems, 40% propan-2-ol in 0.04 mol dm^{-3} ammonium formate buffer, pH 8.2). SA-42 was identified by amino acid analysis (BMC, Uppsala, Sweden) and electrospray MS (VG Analytical AutoSpec, MW 4289.5 g mol^{-1}) and its purity was established by HPLC and ESMS. Single peaks were detected in the chromatograms and only one set of molecular ions were detected in the mass spectrum.

The apparent molecular weight was determined from the sedimentation equilibrium in a MSE Centriscan analytical ultracentrifuge using a photoelectric scanner at 254 nm. The partial specific volume was calculated from the amino acid composition.[46]

CD spectra were recorded on a Jasco J-720 spectropolarimeter from 300 to 190 nm. The instrument was routinely calibrated with (+)-camphor-10-sulfonic acid.[47] Samples were prepared by dilution of aqueous stock solutions by pipetting. The phosphate buffer and guanidine hydrochloride stock solutions were prepared gravimetrically. For quantitative measurements the peptide concentration was determined by quantitative amino acid analysis of the peptide stock solution, otherwise gravimetrically. The experiments were carried out in 0.1, 0.5, 1, 5 or 30 mm cells. For temperature studies, a 0.5 mm water-jacketed cuvette was used together with a circulating HETO constant temperature bath. The pH and ionic strength were adjusted by using phosphate buffers. Quantitative CD spectroscopic measurements and temperature denaturation were carried out with pure water as solvent.

The NMR spectra were recorded on a Varian Unity 500 NMR spectrometer, equipped with an MHU 303 shim system from Resonance Research Inc. One- and two-dimensional NMR spectra were recorded at 323 K in 90% H_2O–10% D_2O (v/v) at pH 5–6 and in mixtures of this solvent with [$^2H_3$]TFE, at peptide concentrations of 1–2 mmol dm^{-3} with a sample volume of 600 μl (1 μl = 1 mm^3). The titration with [$^2H_3$]TFE was carried out in steps of 18 μl by withdrawing the corresponding volume with a syringe from the NMR sample and then adding the same volume of [$^2H_3$]TFE. The compositions reported in the present investigation were 0, 9, 18, 27 36, 54, 72 and 90 μl [$^2H_3$]TFE in water. The pH in the solution containing 90 μl TFE was 6.3. No adjustments of pH were carried out during the titration. For each composition the one dimensional ^1H NMR spectra were recorded as well as TOCSY and NOESY spectra. The 90° pulse was 10.4 μs and the acquisition times were 1, 0.2 and 0.2 s, respectively. The sweep width was 6500 Hz and solvent suppression was accomplished by transmitter presaturation for 1.5 s. The two dimensional experiments were carried out in the phase sensitive mode according to States *et al.*,[48] and for the titration experiments 2 × 350 increments were recorded with 16 transients in each. In the TOCSY experiment the spin lock pulse was 12.5 μs with a window function of 25 μs, the spin lock time was 80 ms. In the NOESY experiment the mixing time was 250 ms. For two TFE containing mixtures, NOE build-up curves were recorded with mixing times from 50 to 250 ms. One-dimensional NMR spectra were also recorded in D_2O as a function of peptide concentration at 0.05, 0.1, 0.5 and 3 mmol dm^{-3} with an acquisition time of 1 s.

Acknowledgements

We thank Professor Per Ahlberg for stimulating discussions and support. Financial support from the Swedish Natural Science Research Council and Magnus Bergvalls Stiftelse is gratefully acknowledged.

References

1 W. F. DeGrado, *Adv. Protein Chem.*, 1988, **39**, 51.
2 M. Mutter and S. Vuilleumier, *Angew. Chem., Int. Ed. Engl.*, 1989, **28**, 535.
3 J. S. Richardson and D. C. Richardson, *Trends Biochem. Sci.*, 1989, 304.
4 K. W. Hahn, W. A. Klis and J. M. Stewart, *Science*, 1990, **248**, 1544.
5 K. Johnsson, R. K. Allemann H. Widmer and S. A. Benner, *Nature*, 1993, **365**, 530.
6 H. Mihara, K. Tomizaki, N. Nishino and T. Fujimoto, *Chem. Lett.*, 1993, 1533.
7 L. Baltzer, A.-C. Lundh, P. Ahlberg, K. Broo and K. Ottosson, submitted for publication in *J. Chem. Soc., Perkin Trans. 2*.
8 T. Sasaki and E. T. J. Kaiser, *J. Am. Chem. Soc.*, 1989, **111**, 380.
9 W. F. DeGrado and J. D. Lear, *Biopolymers*, 1990, **29**, 205.
10 K. S. Åkerfeldt, R. M. Kim, D. Camac, J. T. Groves, J. D. Lear and W. F. DeGrado, *J. Am. Chem. Soc.*, 1992, **114**, 9656.
11 P. T. P. Kaumaya, K. D. Berndt, D. B. Heidorn, J. Trewhella, F. J. Kezdy and E. Goldberg, *Biochemistry*, 1990, **29**, 13.
12 T. Handel and W. F. DeGrado, *J. Am. Chem. Soc.*, 1990, **112**, 6710.
13 M. R. Ghadiri and M. A. Case, *Angew. Chem., Int. Ed. Engl.*, 1993, **32**, 1594.
14 K. Goraj, A. Renard and J. A. Martial, *Protein Eng.*, 1990, **3**, 259.
15 M. H. Hecht, J. S. Richardson, D. C. Richardson and R. C. Ogden, *Science*, 1990, **249**, 884.
16 D. P. Raleigh and W. F. DeGrado, *J. Am. Chem. Soc.*, 1992, **114**, 10079.
17 T. Handel, S. A. Williams and W. F. DeGrado, *Science*, 1993, **261**, 879.
18 P. K. Chakravarty, K. B. Mathur and M. M. Dhar, *Experientia*, 1973, **29**, 786.
19 B. Gutte, M. Daeumigen and E. Wittschieber, *Nature*, 1979, **281**, 650.
20 B. Robson and J. Garnier, *Introduction to Proteins and Protein Engineering*, Elsevier, Amsterdam, 1986.
21 P. S. Ho and W. F. DeGrado, *J. Am. Chem. Soc.*, 1987, **109**, 6751.
22 J. J. Osterhout Jr., T. Handel, G. Na, A. Toumadje, R. C. Long, P. J. Connolly, J. C. Hoch, W. C. Johnson Jr., D. Live and W. F. DeGrado, *J. Am. Chem. Soc.*, 1992, **114**, 331.
23 Y. Kuroda, T. Nakai and T. Ohkubo, *J. Mol. Biol.*, 1994, **236**, 862.
24 Y. Fezoui, D. L. Weaver and J. J. Osterhout, *Proc. Nat. Acad. Sci.*, 1994, **91**, 3675.
25 W. F. DeGrado, Z. Wasserman and J. D. Lear, *Science*, 1989, **243**, 622.

26 Y. M. Lau, A. K. Taneja and R. S. Hodges, *J. Biol. Chem.*, 1984, **259**, 13 253.
27 J. W. Nelson and N. R. Kallenbach, *Proteins*, 1986, **1**, 211.
28 S. Marqusee, V. H. Robbins and R. L. Baldwin, *Proc. Natl. Acad. Sci.*, 1989, **86**, 5286.
29 J. S. Richardson and D. C. Richardson, *Prediction of Protein Structure and the Principles of Protein Conformation*, ed. G. D. Fasman, Plenum Press, New York, 1989, ch. 1.
30 A. W. Burgess and S. J. Leach, *Biopolymers*, 1973, **12**, 2599.
31 S. Padmanabhan and R. L. Baldwin, *J. Mol. Biol.*, 1991, **219**, 135.
32 P. C. Lyu, J. C. Sherman and N. R. Kallenbach, *Proc. Nat. Acad. Sci.*, 1991, **88**, 5317.
33 W. G. J. Hol, P. T. van Duijnen and H. J. C. Berendsen, *Nature*, 1978, **273**, 443.
34 L. Serrano and A. R. Fersht, *Nature*, 1989, **342**, 296.
35 P. C. Lyu, P. J. Gans and N. R. Kallenbach, *J. Mol Biol.*, 1992, **223**, 343.
36 A. J. Adler, N. J. Greenfield and G. D. Fasman, *Methods Enzymol.*, 1973, **27**, 675.
37 N. Greenfield and G. D. Fasman, *Biochemistry*, 1969, **8**, 4108.
38 K. Wutrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1985.
39 N. J. Clayden and R. J. P. Williams, *J. Mag. Res.*, 1982, **49**, 383.
40 A. Pastore and V. Saudek, *J. Mag. Res.*, 1990, **90**, 165.
41 Y.-H. Chen, J. T. Yang and K. H. Chau, *Biochemistry*, 1974, **13**, 3350.
42 M. Engel, R. W. Wiliiams and B. W. Erickson, *Biochemistry*, 1991, **30**, 3161.
43 C. N. Pace, *Methods Enzymol.*, 1986, **131**, 266.
44 J. M. Stewart and J. D. Young, *Solid Phase Peptide Synthesis*, Pierce Chem. Co., Rockford, Illinois, 1984.
45 E. Kaiser, R. L. Colescott, C. D. Bossinger and P. I. Cook, *Anal. Biochem.*, 1970, **34**, 595.
46 T. L. McMeekin and K. Marshall, *Science*, 1952, **116**, 142.
47 W. C. Johnson Jr., *Proteins*, 1990, **7**, 205.
48 D. J. States, R. A. Haberkorn and D. J. Ruben, *J. Mag. Res.*, 1982, **48**, 286.

Paper 5/01920F
Received 27*th March* 1995
Accepted 8*th June* 1995