Prediction of Difficult Sequences in Solid-Phase Peptide Synthesis

R. C. de L. Milton,* Saskia C. F. Milton,* and Paul A. Adams[1]

Contribution from the MRC Regulatory Peptides Research Unit and Biomembrane Research Unit, Department of Chemical Pathology, University of Cape Town Medical School, Observatory 7925, South Africa. *Received September 26, 1989*

Abstract: The common mechanistic origin of the phenomena of segment insolubility in solution-phase and sequence-related incomplete aminoacylations in solid-phase peptide synthesis is used as the basis of a model for difficult sequences in which partial β-sheet hydrogen bonding by the pendant peptide chains of the peptidoresin is the dominant causative principle. A predictive method based on optimized Chou and Fasman type coil conformational parameters is proposed for the identification of such difficult sequences and is tested by its successful application to 101 previously performed solid-phase peptide syntheses (986 aminoacylation reactions). The use of optimized chemical tactics, secondary structure disrupting reagents and tertiary amide linkages between amino acid residues is discussed with a view to improving the solid-phase synthesis of peptides containing difficult sequences.

In the more than 25 years of its application the Merrifield solid-phase principle[2] has been responsible for the production of a great number and variety of synthetic peptides including polypeptides such as the 124-residue ribonuclease A[3] and the 140-residue interleukin-3.[4] In spite of the improved understanding of the chemical basis of the procedure, and the development of optimized approaches,[5-7] a class of peptide sequences has emerged that resists facile synthesis, the so-called "difficult sequences".[5,7,8] The difficulty arises out of the need for near-quantitative aminoacylation reactions at each cycle of a stepwise solid-phase synthesis to prevent the occurrence of peptide side products lacking one or more internal amino acids, but with properties similar to the target sequence.[5,7] Difficult sequences are characterized by reproducible stretches of repetitive incomplete aminoacylations and are more prevalent in some peptides than others (Table I). This phenomenon has been described as the most serious potential problem in stepwise solid-phase peptide synthesis.[5]

Results and Discussion

Sequence-Related Incomplete Aminoacylations. The rate of formation of any given peptide bond in solid-phase peptide synthesis may be affected by the nature of the support, the dispersing solvent, the acylating reagent, and the structure of the protected peptide chain up to that point.[11] The first three of these factors are amenable to optimization of the chemical tactics employed. A variety of supports and resin linkages[6] as well as more or less highly activated acylation intermediates[12] is available, and the use of more polar aprotic solvents such as dimethylformamide allows optimal solvation of the peptidoresin[13] to provide an immediate improvement in the acylation rate. However, sequence-related incomplete aminoacylations are more difficult to anticipate and eliminate.

Table I. Characteristics of Difficult Sequences[a]

1. reproducible, repetitive incomplete aminoacylations (0.5-15%)[5]
2. limited improvement by recoupling or capping[5,9]
3. occurs irrespective of resin type or strategy (orthogonal vs Merrifield)[5,10]
4. aggravated by high resin loadings[5,7,8,14]
5. aggravated by sterically hindered amino acids in the sequence[5]
6. weak or no correlation with the N-terminal residue of the peptidoresn[5,14]
7. sequence dependent[5,7,8,13,14,19]
8. mechanism: intermolecular aggregation of the protected peptide chains[5,7,8,13,14,16,19]

[a] Adapted from S. B. H. Kent.[5]

A study analyzing 723 aminoacylation reactions during the synthesis of 35 peptides by optimized manual stepwise solid-phase procedures revealed two distinct categories of sequence-related incomplete aminoacylations: random and nonrandom.[14] Random difficult aminoacylations may be ascribed to a sterically hindered reaction where the peptide bond is being formed between two amino acid residues with bulky side chains (or protecting groups) or N^{α}- or α-alkyl substituents.[15] Typically, residues such as isoleucine, valine, or threonine with their β-branched side chains fall into this category, but amino acid derivatives such as tosylhistidine and xanthenylasparagine or -glutamine may also cause difficulties by reason of limited solubility or the bulkiness of the protecting group. Any amino acid immediately subsequent to such a sterically hindered residue may also be affected and thus included in this category. The second category, the nonrandom difficulties, is the major source of incomplete aminoacylations[14] and comprises the previously described difficult sequences.

A Model for Difficult Sequences. High-resolution NMR studies of 1% divinylbenzene cross-linked polystyrene and polystyrene peptidoresins in $CHCl_3$ and dimethylformamide have revealed a high degree of segmental mobility in the polystyrene chains that is comparable to linear polystyrene in solution.[16] This would

(1) Biomembrane Research Unit, Department of Chemical Pathology, University of Cape Town Medical School, Observatory 7925, South Africa.
(2) (a) Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149-54. (b) Merrifield, R. B. *Science* **1986**, *232*, 341-7.
(3) Gutte, B.; Merrifield, R. B. *J. Biol. Chem.* **1971**, *246*, 1922-41.
(4) Clark-Lewis, I.; Aebersold, R.; Zittener, H.; Schrader, J. W.; Hood, L. E.; Kent, S. B. H. *Science* **1986**, *231*, 134-9.
(5) Kent, S. B. H. *Annu. Rev. Biochem.* **1988**, *57*, 959-89.
(6) Barany, G.; Kneib-Cordonier, N.; Mullen, D. G. *Int. J. Pept. Protein Res.* **1987**, *30*, 705-39.
(7) Kent, S. B. H.; Clark-Lewis, I. In *Synthetic Peptides in Biology and Medicine*; Alitalo, K., Partanen, P., Vaheri, A., Eds.; Elsevier Science Publishers B.V.: Amsterdam, 1985; pp 29-57.
(8) Kent, S. B. H. In *Peptides: Structure and Function. Proceedings of the 9th American Peptide Symposium*; Deber, C. M., Hruby, V. J., Kopple, K. D., Eds.; Pierce Chemical Co.: Rockford, IL, 1985; pp 407-14.
(9) Hancock, W. S.; Prescott, D. J.; Vagelos, P. R.; Marshall, R. *J. Org. Chem.* **1973**, *38*, 774-81.
(10) (a) Atherton, E.; Woolley, V.; Sheppard, R. C. *J. Chem. Soc., Chem. Commun.* **1980**, 970-1. (b) Atherton, E.; Sheppard, R. C. In *Peptides: Structure and Function. Proceedings of the 9th American Peptide Symposium*; Deber, C. M., Hruby, V. J., Kopple, K. D., Eds.; Pierce Chemical Co.: Rockford, IL, 1985; pp 415-8.
(11) Cameron, L. R.; Holder, J. L.; Meldal, M.; Sheppard, R. C. *J. Chem. Soc., Perkin Trans. 1* **1988**, 2895-901.
(12) Hudson, D. *J. Org. Chem.* **1988**, *53*, 617-24.
(13) Kent, S. B. H.; Hood, L. E.; Beilan, H.; Meister, S.; Geiser, T. *Peptides 1984: Proceedings of the 18th European Peptide Symposium*; Ragnarsson, U., Ed.; Almqvist and Wiksell: Stockholm, 1984; pp 185-8.
(14) Meister, S. M.; Kent, S. B. H. In *Peptides: Structure and Function. Proceedings of the 8th American Peptide Symposium*; Hruby, V. J., Rich, D. H., Eds.; Pierce Chemical Co.: Rockford, IL, 1984; pp 103-6.
(15) Stewart, J. M.; Young, J. D. *Solid Phase Peptide Synthesis*, 2nd ed.; Pierce Chemical Co.: Rockford, IL, 1984; p 46.

0002-7863/90/1512-6039$02.50/0

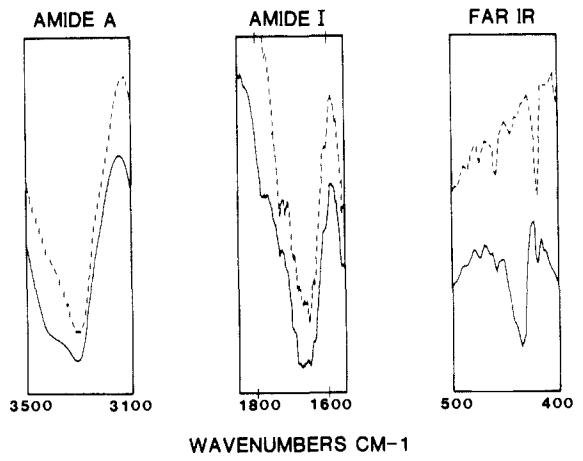

Figure 1. FTIR transmission spectra of peptidoresins from the synthesis of human gonadotropin-releasing hormone precursor peptide 14–36, residue 26 (KBr disks). Protonated peptidoresin (trifluoroacetic acid deprotected, CH_2Cl_2 washed) (—); neutralized peptidoresin (trifluoroacetic acid deprotected, diisopropylethylamine neutralized) (---).

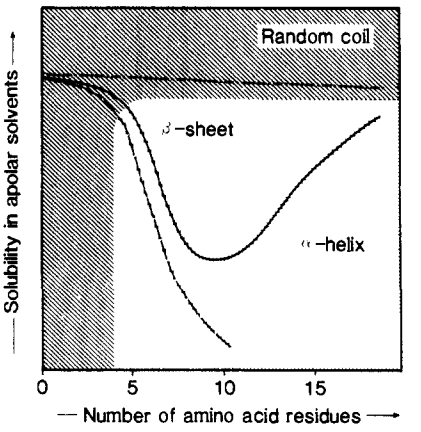

Figure 2. Phase diagram illustrating the relationship between solubility and chain length for homooligopeptides that adopt different secondary structures. Peptides with a predisposition to a random coil configuration (---); peptides with a predisposition to helical structure (--); peptides with a predisposition to β-sheet formation (——). Adapted from Pillai and Mutter.[31]

indicate that the pendant peptides are uniformly well solvated under conditions of solid-phase synthesis in a manner analogous to the free molecules in solution.[16] Thus, the macroscopic insolubility of the polystyrene support, which allows mechanical manipulations such as filtering, is complemented by solvation and diffusion kinetics comparable to the solution phase. This is clearly seen in the swelling of the peptidoresin in solvents such as CH_2Cl_2 and dimethylformamide (which would dissolve linear polystyrene) as opposed to methanol. The swelling of a cross-linked polymer is analogous to the process of dissolving a linear polymer,[17] in that both dissolution and swelling are driven by the normal changes in entropy and enthalpy associated with the mixing of solvent and solute molecules as well as the change in configurational entropy due to dilution of the flexible chain molecules.[18] However, in the cross-linked polymer this tendency toward dispersion is counterbalanced by an elastic restraining force exerted by the cross-linking groups. Peptidoresin swelling may therefore be viewed as an equilibrium between the free energy of solvation of the polymer with its pendant peptides and the elastic restraint of its cross-linker.[18]

NMR studies of difficult sequence peptidoresins (acyl carrier protein decapeptide and a human growth hormone releasing factor analogue) revealed a marked decrease in the segmental mobility of the polystyrene backbone.[19,20] Also, during the synthesis of gonadotropin-releasing hormone precursor peptide fragments, we have observed a reproducible decrease in CH_2Cl_2 and dimethylformamide resin swelling at the initiation of a series of incomplete aminoacylations while swelling in trifluoroacetic acid/CH_2Cl_2 remained normal.[21] These observations are consistent with a model in which a sudden, but reversible, increase in the cross-linking of the polymer[19,22] could occur. This could be accomplished by noncovalent interactions, which would not involve either the polymer support or the fully protected side chains of the pendant peptides, but rather an aggregation of some of the peptide chains by amide hydrogen bonding.[14,19] FTIR spectroscopy of a peptidoresin (Figure 1) from the synthesis of a difficult sequence (human gonadotropin-releasing hormone precursor peptide sequence 14–36, residue 26, KBr disks) supports this viewpoint in that the amide A (3500–3100 cm^{-1}) band of the deprotected protonated sample (trifluoroacetic acid) is broadened with a strong shoulder band at 3400 cm^{-1} compared to the deprotected neutralized sample. This is indicative of more non-hydrogen-bonded NH in the protonated peptidoresin[23] compared to the neutralized sample, which has a sharper band at 3300 cm^{-1}. In the amide I region (1800–1600 cm^{-1}), there is a stronger shoulder band at 1630 cm^{-1} and a shift of the C–O stretching absorbance from 1670 cm^{-1} in the protonated sample to 1650 cm^{-1} in the neutralized sample. This is suggestive of a degree of β-sheet hydrogen bonding.[24] Intermolecular β-sheet formation has been proposed as the underlying mechanism that results in the aggregation and incomplete aminoacylations associated with difficult sequences[7,8] and a partial development of β-structure, while sufficient to increase effective resin cross-linking with a concommitant effect on resin swelling, would accord with the observation that difficult sequence aminoacylations are predominantly incomplete in the 0.5–15% range.[5,14]

A serious problem affecting the fragment condensation method of solution-phase peptide synthesis is the decreasing solubility of the protected peptide segments in organic solvents with increasing chain length.[25,26] Here, too, the chemical mechanism has been demonstrated to be an intermolecular aggregation by β-sheet hydrogen bonding, which leads to precipitation of the reactants.[25] The phenomena of segmental insolubility and difficult sequences may therefore be regarded as fundamentally analogous on a mechanistic level,[8] the most important difference being the extent of the aggregation, which is limited by the improved solvation of the pendant peptide chains by the polystyrene support in the solid-phase approach.[8,18] Both phenomena are sequence related[8,27] and both are improved by the use of more polar solvents,[8,14,19,28] especially those with good proton-donating or -accepting properties.[29]

(16) Live, D.; Kent, S. B. H. In *Elastomers and Rubber Elasticity*; Mark, J. E., Lal, J., Eds.; ACS Symposium Series 193; American Chemical Society: Washington, DC, 1982; pp 501–15.
(17) Flory, P. J. *Principles of Polymer Chemistry*; Cornell University Press: Ithaca, NY, 1953.
(18) Sarin, V. K.; Kent, S. B. H.; Merrifield, R. B. *J. Am. Chem. Soc.* **1980**, *102*, 5463–70.
(19) Live, D. H.; Kent, S. B. H. In *Peptides: Structure and Function. Proceedings of the 8th American Peptide Symposium*; Hruby, V. J., Rich, D. H., Eds.; Pierce Chemical Co.: Rockford, IL, 1984; pp 65–8.
(20) Deber, C. M.; Lutek, M. K.; Heimer, E. P.; Felix, A. M. *Pept. Res.* **1989**, *2*, 184–8.
(21) Milton, R. C. de L.; Wormald, P. J.; Brandt, W.; Millar, R. P. *J. Biol. Chem.* **1986**, *261*, 16990–7.
(22) Ford, W. T.; Balakrishnan, T. *Macromolecules* **1981**, *14*, 284–92.
(23) Narita, M.; Doi, M.; Sugasawa, H.; Ishikawa, K. *Bull. Chem. Soc. Jpn.* **1985**, *58*, 1473–9.
(24) (a) Narita, M.; Ogura, T.; Sato, K.; Honda, S. *Bull. Chem. Soc. Jpn.* **1986**, *59*, 2433–8. (b) Miyazawa, T.; Blout, E. J. *J. Am. Chem. Soc.* **1961**, *83*, 712–9.
(25) Narita, M.; Fukunaga, T.; Wakabayashi, A.; Ishikawa, K.; Nakano, H. *Int. J. Pept. Protein Res.* **1984**, *23*, 306–14.
(26) Bodansky, M. *Principles of Peptide Synthesis*; Springer-Verlag: Berlin, 1984; p 209.
(27) Narita, M.; Ishikawa, K.; Chen, J.-Y.; Kim, Y. *Int. J. Pept. Protein Res.* **1984**, *24*, 580–7.
(28) Narita, M.; Ohkawa, N.; Nagasawa, S.; Isokawa, S. *Int. J. Pept. Protein Res.* **1984**, *24*, 129–34.
(29) Narita, M.; Honda, S.; Umeyama, H.; Obana, S. *Bull. Chem. Soc. Jpn.* **1988**, *61*, 281–4.

Table II. Amino Acid Residues in the Helix, β-Sheet, and Coil Regions of 29 Proteins and the Conformational Parameters of All Residues

amino acid	no. of residues[a]	helix residues[a]	sheet residues[a]	coil residues[a]	P_α[b]	P_β[b]	P_c[b]	P_c*[c]
Ala	434	234	71	129	1.42	0.83	0.70	0.75
Arg	142	53	26	63	0.98	0.93	1.04	0.96
Asn	230	58	40	132	0.67	0.89	1.35	1.26
Asp	273	105	29	139	1.01	0.54	1.20	1.07
Cys	94	25	22	47	0.70	1.19	1.18	1.08
Gln	162	68	35	59	1.11	1.10	0.86	0.83
Glu	234	134	17	83	1.51	0.37	0.83	0.80
Gly	422	91	62	269	0.57	0.75	1.50	1.47
His	129	49	22	58	1.00	0.87	1.06	0.96
Ile	233	95	73	65	1.08	1.60	0.66	0.74
Leu	358	164	91	103	1.21	1.30	0.68	0.75
Lys	347	153	50	144	1.16	0.74	0.98	0.90
Met	73	40	15	18	1.45	1.05	0.58	0.70
Phe	170	73	46	51	1.13	1.38	0.71	0.76
Pro	176	38	19	119	0.57	0.55	1.59	1.64
Ser	367	107	54	206	0.77	0.75	1.32	1.21
Thr	278	87	65	126	0.83	1.19	1.07	0.98
Trp	78	32	21	25	1.08	1.37	0.75	0.79
Tyr	184	48	53	83	0.69	1.47	1.06	0.98
Val	357	144	119	94	1.06	1.70	0.62	0.73
totals	4741	1798	930	2013				

[a] Data cited from Chou and Fasman[32] and Narita et al.[27] [b] The frequencies, $f_{i,x} = n_{i,x}/n_i$, of residues in the coil, α-, or β-region for the 20 amino acids are divided by the average frequency, $\langle f_x \rangle = \sum n_{i,x}/\sum n_i$, in that region, where $n_{i,x}$ is the number of i residues in the region, and n_i is the number of i residues in the proteins. [c] P_c* obtained by linear regression of the function $1/P_c = 0.739P_\alpha + 0.345P_\beta$.

Accepting that partial β-sheet formation is responsible for that fraction of the pendant peptide chains on the peptidoresin that resist aminoacylation in difficult sequences, it is relevant to consider the likely conformation of those peptide chains that are readily acylated. For this purpose the liquid-phase method,[30] utilizing protected peptide–poly(oxyethene) copolymers,[31] may be regarded as a hybrid between the solution and solid phase and may be used as a model for the solid-phase situation.[7,31] The solubility of these copolymers in solvents such as trifluoroethanol and water allows CD and IR studies to be performed in parallel with solubility studies in more apolar solvents as chain length is increased.[31] In this way, by use of homooligopeptide– and protected natural sequence–copolymers, three distinct modes of conformational behavior have been described (Figure 2).[31] The first few amino acid residues ($n < 5$) added to the copolymer adopt a random coil conformation, which may persist throughout the synthesis if the sequence has a predisposition to a random coil structure.[31] However, with sequences predisposed to β-structure [poly(Ala)–, poly(Val)–, and poly(Ile)–copolymers], there is a conformational transition to β-sheet after the initial random coil region and this persists as far as the sequence can be continued.[31] The third mode, characterized by poly(Met)–copolymers, involves two conformational transitions: from the initial random coil to β-sheet, and then when $n > 8$ residues, a significant proportion of helical structure is seen.[31] The initial conformational transformation from random coil to β-sheet, even in peptides with a predicted[31-33] predisposition to helix formation, is understandable in terms of hydrogen bond structure. Optimally, the carbonyl and amide dipoles are not linearly aligned; rather the amide hydrogen is closer to one of the carbonyl oxygen's sp^2 lone electron pairs.[34] This steric arrangement cannot be achieved in α-helix hydrogen bonding where the polypeptide backbone geometry prevents optimal alignment.[35] On the other hand, both parallel and antiparallel β-sheet hydrogen bonds can assume the optimal configuration more favorably (the average N···O distances are 2.91 Å for β-sheet and 2.99 Å for α-helix hydrogen bonds), thus contributing more stability to the secondary structure per residue than with α-helices.[36] Also, in the relatively apolar conditions of solid-phase aminoacylation there would seem to be little entropic advantage in the disordering of the solvent molecules by α-helix formation. This may, however, be the case with the more polar solvents used in conformational studies. For example, the far-IR spectrum of the deprotected, protonated difficult sequence peptidoresin (human gonadotropin-releasing hormone precursor sequence 14–36, vide supra) shows a peak at 440 cm^{-1} that is absent from the neutralized sample (Figure 1). This is believed to be indicative of a degree of helical structure[23,37] induced by the polar trifluoroacetic acid/CH_2Cl_2 reagent used for deprotection.

Based on the presented arguments, a steric model, which envisages a random coil conformation for the peptide chains of the peptidoresin under the conditions of aminoacylation in solid-phase peptide synthesis, is proposed. This is the predominant conformation of all, or at least the majority, of the pendant peptide chains regardless of their length and it is the most suitable for rapid and complete aminoacylation reactions. As the chain length increases with the addition of more amino acid residues, a small fraction of the pendant peptide chains may adopt a helical structure if the peptide sequence has this predisposition, but this is not thought to affect the rate or extent of aminoacylation significantly. Some sequences (or internal segments of these sequences—i.e., difficult sequences), dependant on their amino acid composition, have a predisposition toward β-sheet structure. Under aminoacylation conditions in solid-phase synthesis, a fraction (0.5–15%) of the pendant peptide chains of these sequences will form a partial β-sheet structure, which makes those peptide chains involved sterically resistant to aminoacylation and leads to an incomplete reaction. The partial β-sheet structure also increases the effective cross-linking of the polystyrene copolymer, thus reducing resin swelling and hindering the entry of reagents into the resin matrix.

Prediction of Difficult Sequences. The common mechanistic origins of segment insolubility in solution-phase peptide chemistry and the incomplete aminoacylations associated with difficult sequences in solid-phase[8] suggest that approaches shown to be successful in the one area may have a more general application. Narita and co-workers have developed a predictive method, based on the conformational parameters of Chou and Fasman,[32] which identifies potentially soluble protected peptide intermediates for segment-condensation strategies in solution-phase synthesis.[27] On the basis of their observation that protected peptides in a random coil conformation are most soluble in solvents suited to amino-

(30) (a) Mutter, M.; Hagenmeier, H.; Bayer, E. *Angew. Chem., Int. Ed. Engl.* **1971**, *10*, 811–2. (b) Bayer, E.; Mutter, M. *Nature* **1972**, *237*, 512–3.
(31) Pillai, V. N. R.; Mutter, M. *Acc. Chem. Res.* **1981**, *14*, 122–30.
(32) Chou, P. Y.; Fasman, G. D. In *Advances in Enzymology*; Meister, A., Ed.; John Wiley and Sons: New York, 1978; Vol. 47, pp 45–148.
(33) Tanaka, S.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 142–59.
(34) Taylor, R.; Kennard, O.; Versichel, W. *J. Am. Chem. Soc.* **1983**, *105*, 5761–6.
(35) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97–179.
(36) Schultz, G. E. *Annu. Rev. Biophys. Chem.* **1988**, *17*, 1–21.
(37) Itoh, K.; Katabuchi, H. *Biopolymers* **1973**, *12*, 921–9.

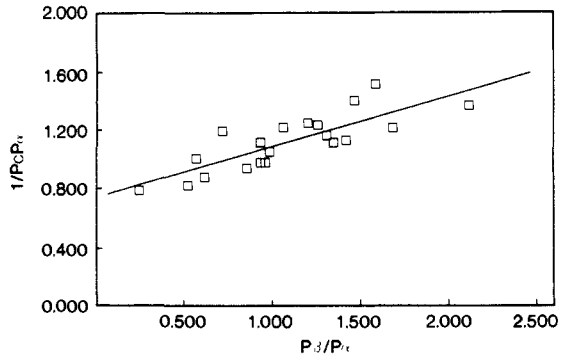

Figure 3. Linear regression of eq 2. $1/P_cP_\alpha = a + bP_\beta/P_\alpha$; $a = 0.739$; $b = 0.345$; $r = 0.87$. P_c^* values for each of the 20 amino acids were derived from the least-squares fit.

acylation conditions, they have derived the P_c or coil parameter (Table II), which, averaged for all the residues in a particular peptide segment, provides a reliable index of solubility for that peptide.[27] This principle may be extended to the stepwise solid-phase approach if the addition of each amino acid residue is regarded as creating a new segment. The average, $\langle P_c \rangle$, of each segment then reflects the inherent tendency of that sequence to assume a random coil conformation (as opposed to β-sheet or α-helix) and provides an index of the ease of incorporation of the next residue in the stepwise assembly of the target sequence.

On the basis of these assumptions, the P_c parameters were applied to predict incomplete aminoacylations in peptide sequences that had been previously synthesized by a manual solid-phase approach employing ninhydrin monitoring[38,39] of all acylation reactions. Disregarding the random incomplete aminoacylations[14] associated with sterically hindered amino acid derivatives, the expected correlation was obtained with most sequences, but misleading results were seen with certain sequences, especially those containing aspartic acid, methionine, serine, and valine residues. In view of the empirical derivation of the Chou and Fasman type conformational parameters,[27,32] and on the basis that secondary structure be either coil or α-helix/β-sheet in this system, an inverse relationship can be implied to exist between these parameters (Table II). The simplest relationship of this form would be expected to be bilinear:

$$1/P_c = aP_\alpha + bP_\beta \qquad (1)$$

which requires that a plot of $1/P_cP_\alpha$ vs P_β/P_α be a straight line (Figure 3). Linear least-squares analysis yields values of 0.739 and 0.345 for a and b, respectively, with $r = 0.87$. As the Chou and Fasman type parameters are obtained independently on the basis of the X-ray structures of 29 different kinds of globular proteins,[27,32] we consider that the P_c^* values (Table II) calculated from the least-squares analysis fit to eq 1 more closely approximate the values that would be obtained from an infinite data set. In addition, the use of refined P_c^* values calculated from eq 1 improved the accuracy of prediction of difficult sequences (in contrast to a range of other linear relationships that were tested), with the refined P_c^* values of the four amino acids (aspartic acid, methionine, serine, and valine) undergoing large changes relative to their original P_c values. These changes are in the direction of improved predictive accuracy.

The application of the P_c^* values to the prediction of incomplete aminoacylations in solid-phase peptide synthesis was found to require interpretation on the basis of assumptions derived from the preceding Model for Difficult Sequences and these could be formulated as a set of rules or guidelines for the predictive method.

Rule 1. The stepwise synthesis of a protected peptide chain is viewed as a succession of peptide segments increasing in size by one residue after each aminoacylation. Each segment has an individual inherent tendency to a random coil conformation, which can be expressed by the average of the P_c^* values of its constituent amino acid residues:

$$\langle P_c^* \rangle = \sum P_{c\,i-n}^*/n \qquad (2)$$

where n represents the number of amino acid residues in that segment and $P_{c\,i-n}^*$ represents the individual P_c^* value of each amino acid. The $\langle P_c^* \rangle$ of each segment is inversely proportional to the probability of incomplete aminoacylation of the subsequent residue.

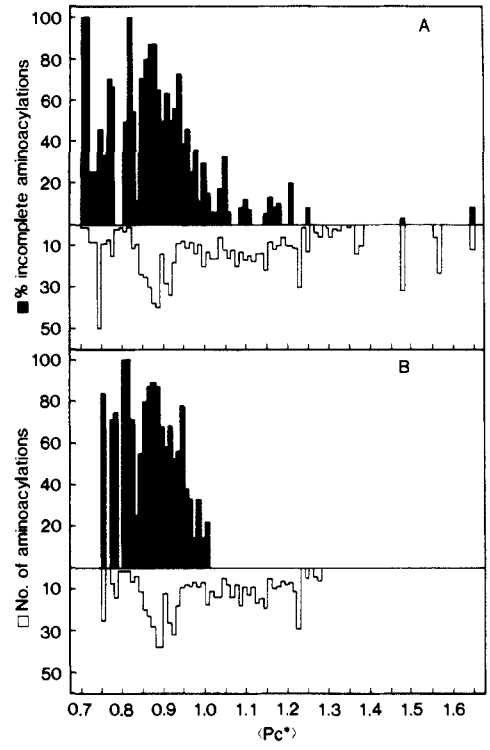

Figure 4. Cumulative $\langle P_c^* \rangle$ values plotted against percent incomplete aminoacylation reactions (■) and number of aminoacylations [lower histograms (□)] for 101 peptide syntheses to show the inverse relationship between $\langle P_c^* \rangle$ and probability of incomplete aminoacylations. (A) Application of rule 1 of the predictive method: $\langle P_c^* \rangle = \sum P_{c\,i-n}^*/n$ for 946 acylations. (B) Application of rules 2 and 3: random incomplete aminoacylations (33 reactions) and the first four amino acid residues of each synthesis (311 reactions) are disregarded.

In this manner a series of cumulative $\langle P_c^* \rangle$ values (C → N) is generated for each residue in the target sequence, with each value predicting the ease of aminoacylation for the next residue in the stepwise synthesis. Generally, we have found that $\langle P_c^* \rangle$ values greater than 1.0 are associated with near-quantitative incorporation (>99.8%) of the subsequent residue in a reasonable time, while values between 0.9 and 1.0 indicate that a longer reaction time or "recoupling" is required (Figure 4) (cf. ref 27). $\langle P_c^* \rangle$ values less than 0.9 were normally associated with persistent aminoacylation difficulties. These guidelines have been successfully applied irrespective of aminoacylation method [not even the (benzotriazol-1-yloxy)tris(dimethylamino)phosphonium hexafluorophosphate reagent[12,40] was able to improve the extent of aminoacylation following low $\langle P_c^* \rangle$ values] or the enantiomer of the amino acid derivative to be added or those already in the sequence.[41] A large increase in the molar excess (>5 equiv) of the incoming activated species may, however, improve the rate and extent of aminoacylation.

As expected, the cumulative $\langle P_c^* \rangle$ approaches the average of the 20 P_c^* values (0.97) with increasing peptide chain length in

(38) (a) Kaiser, E.; Colescott, R. L.; Bossinger, C. D.; Cook, P. I. *Anal. Biochem.* **1970**, *34*, 595–8. (b) Kaiser, E.; Bossinger, C. D.; Colescott, R. L.; Olsen, D. B. *Anal. Chim. Acta* **1980**, *118*, 149–51.

(39) Sarin, V. K.; Kent, S. B. H.; Tam, J. P.; Merrifield, R. B. *Anal. Biochem.* **1981**, *117*, 147–57.

(40) Castro, B.; Dormoy, J. R.; Evin, G.; Selve, C. *Tetrahedron Lett.* **1975**, 1219–22.

(41) This observation relates to sequences with a single D-amino acid substitution. Peptides with multiple D-amino acid insertions are likely to be substantially less difficult to assemble than the predictive method would imply.

Table III

(A) Standard Merrifield[42] Syntheses of $(Ala)_{12}$ and $[N^{\alpha}$-Me-Ala$^{4,9}](Ala)_{12}$[a]

	total acylation time (h)	Kaiser test	Ac_2O		total acylation time (h)	Kaiser test	Ac_2O
R-O-Ala				R-O-Ala			
Ala	1.5	-		Ala	1.5	-	
Ala[10]	1.5	-		Ala[10]	1.5	-	
Ala	1.5	-		N^{α}-Me-Ala	3.0	+	+
Ala	11.5	+	+	Ala	3.0		+
Ala	44.0	+	+	Ala	1.5	-	
Ala	20.0	+	+	Ala	1.5	-	
Ala[5]	11.0	+	+	Ala[5]	1.5	+	+
Ala	9.5	+	+	N^{α}-Me-Ala	3.0	+	+
Ala	11.0	-		Ala	3.0		+
Ala	12.0	+	+	Ala	1.5	+	+
Ala[1]	12.0	-		Ala[1]	1.5	-	

(B) Optimized[5,7] Synthesis of [D-Ala^{22}]pHGnRH 14–26 and [N^{α}-Me-D-Ala^{22}, N^{α}-Me-Leu^{18}]pHGnRH 14–26[b]

	total acylation time (h)	quant Sarin test (% coupled)	Ac_2O		total acylation time (h)	quant Sarin test (% coupled)	Ac_2O
R-Pam-Val				R-Pam-Val			
Ile[25]	3.0	99.74		Ile[25]	2.0	99.61	+
Glu	3.5	99.78		Glu	2.0	99.80	
Gln	3.5	99.75		Gln	2.0	99.80	
D-Ala	4.5	99.76	+	N^{α}-Me-D-Ala	9.5	99.80	
Ser	4.3	99.61	+	Ser	9.0		+
Asp[20]	5.0	99.70	+	Asp[20]	2.0	99.86	
Ile	12.5	96.54	+	Ile	2.0	99.84	
Leu	5.0	99.37	+	N^{α}-Me-Leu	13.0	99.89	
Asn	5.0	99.63	+	Asn	9.3		+
Glu	5.5	99.45	+	Glu	2.0	99.88	
Ala[15]	3.5	99.57	+	Ala[15]	2.0	99.92	
Asp	3.5	99.62	+	Asp	2.0	99.93	

[a] Aminoacylations: 1.5 molar excess dicyclohexylcarbodiimide and t-Boc-amino acid. [b] Aminoacylations: 3.0 molar excess dicyclohexylcarbodiimide + Boc-amino acid, preformed with HOBt.

most peptide sequences. This is in accordance with the observation that the maximum occurrence of consecutive low-yield aminoacylations is 5–15 residues from the resin,[5,7,14] where the cumulative $\langle P_c^* \rangle$ is most likely to fluctuate, and with the generally observed trend toward longer acylation times and "recoupling" with increasing length of the peptide chain[42] when the $\langle P_c^* \rangle$ approaches its average value.

Rule 2. Random incomplete aminoacylations[14] are ascribed to the steric hindrance (β-branched amino acids) or poor solubility of individual amino acid derivatives and occur irrespective of the cumulative $\langle P_c^* \rangle$ values. They are recognized by experience or by their isolated occurrence and are not predicted as part of a difficult sequence. If they occur within a difficult sequence ($\langle P_c^* \rangle$ values less than 1.0), then that specific aminoacylation and/or the one subsequent to it can be extraordinarily difficult.[5]

Rule 3. The first three cumulative $\langle P_c^* \rangle$ values for any synthesis are disregarded as there is insufficient sequence for secondary structure to form.[31] Incomplete aminoacylations occurring in the first four residues of a synthesis can invariably be ascribed to random-type difficulties.

Rule 3 was formulated on the basis of our observation that the earliest reproducible commencement of a series of incomplete aminoacylations in a peptide sequence was at residue 5 in the solid-phase synthesis of poly(alanine) peptides (Table IIIA). This was supported by the finding of Pillai and Mutter[31] that the first few amino acid residues ($n < 5$) of the peptide–poly(oxyethene) copolymers in their liquid-phase synthesis system adopted a random coil conformation prior to nucleation of secondary structure (Figure 2). We have, however, noted considerable variation in the initiation of difficult sequences in our database, with repetitive incomplete aminoacylations only commencing at residue 10 of one sequence in spite of preceding $\langle P_c^* \rangle$ values ranging from 0.78 to 0.97. The average number of residues at which predicted incomplete aminoacylations were first observed was found to be six. It was nevertheless decided to recommend overprediction in this region (i.e., n = 4–8) until the factors determining nucleation of secondary structure in peptidoresin sequences are better understood and can be incorporated into a predictive model. The worst consequence of such overprediction would be a relatively easy aminoacylation where an incomplete reaction was anticipated.

The predictive method was tested by applying it retrospectively to 101 solid-phase syntheses comprising 986 aminoacylation reactions (Figure 4) previously performed in this laboratory. These peptides were synthesized by standard Merrifield[43] or optimized[5,7] tactics, ranged in size from 3 to 41 residues, and included gonadotropin-releasing hormone and human gonadotropin-releasing hormone precursor fragments and analogues as well as poly(alanine) peptides with amino acid substitutions and miscellaneous sequences prepared for other research groups. Incomplete aminoacylations were defined in terms of yield, a positive Kaiser test[38] or a quantitative ninhydrin test[39] showing less than 99.5% acylation in 1.5 h.

Figure 4A demonstrates the correlation between ease of aminoacylation and the cumulative $\langle P_c^* \rangle$ value (rule 1), where the greater proportion of incomplete aminoacylations are associated with low $\langle P_c^* \rangle$ values. Figure 4B illustrates the application of rules 2 and 3 along with the first rule: 33 incomplete aminoacylations are disregarded as being random-type[14] under rule 2 and 311 aminoacylations representing the first four residues of each peptide are disregarded under rule 3. The resultant histogram now shows that all incomplete aminoacylations were associated with $\langle P_c^* \rangle$ values less than 1.0 and that difficulty of aminoacylation increased in a roughly linear fashion as the $\langle P_c^* \rangle$ value diminished. Both parts A and B of Figure 4 include a histogram of total aminoacylations at each $\langle P_c^* \rangle$ value from which the percentage incomplete aminoacylations was calculated. The database of aminoacylations utilized can be seen to be loaded

(42) Morrell, J. L.; Brown, J. H. *Int. J. Pept. Protein Res.* **1985**, *26*, 49–54.

(43) (a) Erickson, B. W.; Merrifield, R. B. In *The Proteins*, 3rd ed.; Neurath, H., Hill, R. L., Boeder, C. L., Eds.; Academic Press: New York, 1976; pp 255–590. (b) Barany, G.; Merrifield, R. B. In *The Peptides*; Gross, E., Meienhofer, J., Eds.; Academic Press: New York, 1980; Vol. 2, pp 1–284. (c) Stewart, J. M.; Young, J. D. *Solid Phase Peptide Synthesis*, 1st ed.; W. H. Freeman and Co.: San Francisco, CA, 1969.

toward incomplete aminoacylations, the reason being that syntheses where difficulty was anticipated were subjected to greater quality control. This is considered as an additional test of the predictive method.

Not all aminoacylations associated with low cumulative $\langle P_c^* \rangle$ values are necessarily incomplete in practice. We have found, on resynthesis of difficult sequence peptides, that reactions which proceed to completion in one synthesis may be incomplete in the subsequent attempt and vice versa. We therefore consider a low cumulative $\langle P_c^* \rangle$ value to be an indication of a potential incomplete aminoacylation. In view of the progressive increase in the probability of incomplete aminoacylation with diminishing $\langle P_c^* \rangle$, the database was divided into three parts in order to assess the predictive value of the method. At $\langle P_c^* \rangle$ values greater than 1.0, when all three rules are applied (Figure 4B), the predictive accuracy of the method is unity (successful predictions/number of aminoacylations). $\langle P_c^* \rangle$ values between 0.9 and 1.0 show a predictive accuracy of 0.50, and with values below 0.9 this increases to 0.76. This is to be expected of any predictive method of this type—at the extremes of the predictive range the accuracy would approach a 100% success rate with a crossover value of 50%. The overall accuracy of the predictive method with this database is 0.80. These correlations support the guidelines already given for interpretation of the cumulative $\langle P_c^* \rangle$ values.

The predictive method was also successfully applied to those reports from the literature where adequate synthetic details were given. These included the use of both Boc/Bzl[7,9,20,42,44] and orthogonal-Fmoc[45] tactics. Notably, the polypeptides bovine ribonuclease A[3] and interleukin-3[4] are both predicted as "possible" syntheses (cumulative $\langle P_c^* \rangle$ values greater than or of the order of 1.0).

On initial consideration it would seem unlikely that conformational parameters derived for proteins in aqueous solution should be applicable to protected amino acid derivatives and peptide chains in the relatively apolar conditions of peptide synthesis. However, closer examination reveals that none of the semipermanent protecting groups radically alters the electronic and inductive character of the amino acid side chains proximal to the protecting groups. Thus, the hydrophobic amino acids (phenylalanine, alanine, leucine, isoleucine, and valine) all have very similar P_c^* parameter values (Table II) while amino acid pairs, like threonine and serine or glutamic acid and aspartic acid (glutamine and asparagine), which have side chain functional groups that give them similar chemical properties, have differing P_c^* values and different electron distributions due to the inclusion of additional methylene or methyl groups. This conservation of the essential character in the protected amino acid side chains can be argued to provide a basis for the application of the Chou and Fasman type conformational parameters to the prediction of secondary structure (and hence segmental insolubility and incomplete aminoacylations) in peptide synthesis. Alternately, the predictive ability of the Chou and Fasman type parameters in organic media may be equated to their relevance in predicting secondary structures in the interior of globular proteins, where an environment of relatively low dielectric constant exists.[20]

In conclusion, the predictive method, proposed in this article, is considered a reliable index of potentially difficult sequences in peptides that are to be synthesized. It will facilitate the planning and logistics of a particular synthesis as well as allowing the choice of the most economical, yet effective chemistry for a particular sequence. It will also aid the preparation of more effective programs for automated syntheses.

Improvement of Sequence-Related Incomplete Aminoacylations. The recommended approach to the synthesis of peptides with difficult sequences is the adoption of optimized chemical tactics[5,7] to minimize the occurrence of side products unrelated to the incomplete aminoacylations. Also recommended is the use of low resin substitutions[8,46] and polar solvents such as dimethylformamide in the acylation reactions to disrupt hydrogen bond formation.[7] Although this approach does facilitate the synthesis of difficult sequences, incomplete aminoacylations still occur, typically at the 0.5–2.0% level,[5] and in the synthesis of longer peptides this can lead to unacceptably heterogenous crude products.

Most conveniently, the solution to the problem of the formation of secondary structure in peptidoresins would follow this approach: a solvent or additive is required that will be compatible with acylation conditions yet thoroughly effective in disrupting hydrogen bonding and in maintaining the pendant peptide chains of the peptidoresin in a random coil conformation. In this regard, solvent titration experiments have identified 1,1,1,3,3,3-hexafluoro-2-propanol as a particularly effective solvent (even when diluted with CH_2Cl_2) for protected peptide fragments that have a β-structure in the solid state.[29] Preliminary solid-phase aminoacylation trials using this reagent at approximately 10% dilution with preformed symmetric anhydrides of the CH_2Cl_2-soluble amino acid derivatives have shown resin swellings comparable to trifluoroacetic acid/CH_2Cl_2 (1:1) and reduced reaction times with difficult sequence aminoacylations. It is also helpful as the preliminary wash before the addition of preformed active esters and before recouplings, but is most useful when added to the acetic anhydride/CH_2Cl_2 reagent by allowing near-quantitative capping in reasonable times (30–60 min) with difficult sequences.[47] At this time, however, hexafluoro-2-propanol would seem to be an aid, rather than a cure, for the synthesis of difficult sequence peptides, but a better understanding of its mechanism of action and the possible side reactions associated with its use may allow for its more effective application. More economical alternatives to the perfluorinated alcohols are the chaotropic salts, $NaClO_4$ and KSCN, which have recently been reported to improve the solid-phase synthesis of difficult sequence peptides such as [Ala[13]-NH_2]RNAse 1–13.[48] They too would seem to function best in conjunction with preformed activated intermediates, as complex formation with side chain functional groups like Thr(Bzl) may be a problem.[48]

An alternative route to the disruption of amide hydrogen bonding in peptidoresins could be the provision of N^{α}-substituted blocking groups on the peptide bonds between amino acid residues. A limitation of this approach would be the difficulty associated with the incorporation of N^{α}-disubstituted amino acid derivatives into the peptide chain and the subsequent formation of a tertiary amide bond. This could, however, be alleviated by inserting these derivatives periodically along the difficult sequence. This principle, termed "peptide segment separation", has been successfully applied to the solubilization and synthesis of protected peptide segments for solution-phase chemistry.[25,49] The viability of this tactic in the solid-phase approach was tested by comparing the synthesis of two demonstrably difficult sequence peptides, an alanine dodecamer and an analogue of a human gonadotropin-releasing hormone precursor fragment ([D-Ala[22]]pHGnRH 14–26), with their analogues [N^{α}-Me-Ala[4,9]](Ala)$_{12}$ and [N^{α}-Me-Leu[18], N^{α}-Me-D-Ala[22]]pHGnRH 14–26. These comparisons employed standard Merrifield[42] and optimized tactics,[5,7] respectively, and in both cases, except for the N^{α}-methyl-substituted derivative and the subsequent residue which were treated as difficult aminoacylations, the remaining residues were incorporated more quickly and extensively in the tertiary amide bond containing analogues (Table III). Thus, the periodic incorporation of tertiary amide

(44) Wilkes, B. C.; Hruby, V. J.; Sherbrooke, W. C.; Castrucci, A. M. de L.; Hadley, M. E. *Biochem. Biophys. Res. Commun.* **1984**, *122*, 613–9.
(45) (a) Atherton, E.; Clive, D. I. J.; Sheppard, R. C. *J. Am. Chem. Soc.* **1975**, *97*, 6584–5. (b) Eberle, A. N.; Atherton, E.; Dryland, A.; Sheppard, R. C. *J. Chem. Soc., Perkin Trans. 1* **1986**, 361–7. (c) Cameron, L.; Meldal, M.; Sheppard, R. C. *J. Chem. Soc., Chem. Commun.* **1987**, 270–2. (d) Atherton, E.; Holder, J. L.; Meldal, M.; Sheppard, R. C.; Valerio, R. M. *J. Chem. Soc., Perkin Trans. 1* **1988**, 2887–94.
(46) Kent, S. B. H.; Merrifield, R. B. *Peptides 1980: Proceedings of the 16th European Peptide Symposium*; Brunfeldt, K., Ed.; Scriptor: Copenhagen, 1981; pp 328–33.
(47) Milton, S. C. F.; Milton, R. C. de L. *Int. J. Pept. Protein Res.* In press.
(48) Stewart, J. M.; Klis, W. A. In *Innovation and Perspectives in Solid Phase Synthesis*; Epton, R., Ed.; Wolverhampton Polytechnic: Wolverhampton, UK. In press.
(49) Narita, M.; Ishikawa, K.; Nakano, H.; Isokawa, S. *Int. J. Pept. Protein Res.* **1984**, *24*, 14–24.

linkages in difficult sequences could reduce incomplete aminoacylations to two out of every five (or more) residues, depending on the inherent predisposition of the particular sequence to the formation of secondary structure. A suitable N^{α} substituent for this purpose would be the 2,4-dimethoxybenzyl group,[50] which has been used by Narita and co-workers to improve the solubility of β-sheet-forming peptides in solution-phase chemistry.[25,27,49] It is relatively stable to trifluoroacetic acid yet labile under normal conditions of hydrogen fluoride deprotection and cleavage.[51]

Conclusion. The incomplete aminoacylations and resultant difficulties associated with the solid-phase synthesis of peptides containing difficult sequences arise from the inherent stereochemical properties of the constituent amino acid residues and are unlikely to admit facile solutions. Nevertheless, the preliminary identification and assessment of such sequences by a predictive method, such as the one presented here, should allow individual consideration of the potential problems associated with each target sequence. Judicious control over initial resin loadings and the choice of appropriate tactics should then ensure economical and effective syntheses in the majority of cases. Such tactics may be regarded as escalating from the use of optimized chemistries and secondary structure disrupting additives to the incorporation of semipermanent tertiary amide linkages in the peptide sequence.

Experimental Section

Materials. *t*-Boc amino acid derivatives were obtained from Peninsula Laboratories Europe (St. Helens, Merseyside, UK), Protein Research Foundation (Osaka, Japan), and Bachem Feinchemikalien AG (Bubendorf, Switzerland) with the following side chain protecting groups:[52] Arg(Tos), Asn(Xan), Asp(OBzl), Asp(OcHex), Cys(4-MeBzl), Gln(Xan), Glu(OBzl), Glu(OcHex), His(Bom), His(Dnp), His(Tos), Lys(Cl-Z), Ser(Bzl), Thr(Bzl), Trp(For), Tyr(Br-Z), and Tyr(Cl_2Bzl). *t*-Boc-N^{α}-methyl derivatives were prepared by the method of Olsen.[53] Trifluoroacetic acid, diisopropylcarbodiimide, diisopropylethylamine (over CaH_2 and ninhydrin), and triethylamine (over ninhydrin) were distilled before use. Dicyclohexylcarbodiimide and dimethylformamide were distilled under reduced pressure (the latter over P_2O_5 and ninhydrin). Dioxane was filtered through activated neutral aluminum oxide to remove peroxides. All other solvents were HPLC grade and all chemicals met ACS standards. Standard Merrifield[43] tactics employed *t*-Boc-aminoacyl resins prepared with chloromethylated 1% divinylbenzene–polystyrene (Bio-Rad, Richmond, CA) and anhydrous KF.[54] For optimized tactics,[5,7] 4-(bromomethyl)phenylacetic acid phenacyl ester was obtained from RSA Co. (Ardsley, NY) and recrystallized from CH_2Cl_2/petroleum ether (5:2) before esterification with the potassium[56] or cesium[57] salts of *t*-Boc amino acids. Subsequent reductive cleavage of the phenacyl group[55] and acylation with aminomethyl resin[55] using carbodiimide or preformed benzotriazolyl active esters produced the *t*-Boc-aminoacyl-Pam resins.[55,56] MBHA resin was prepared as previously described.[58]

Peptide Synthesis. Stepwise solid-phase syntheses were accomplished using standard Merrifield[43] or optimized tactics[5,7] with variations to suit individual sequences. General descriptions of these methods are provided with specific synthetic details for each peptide sequence listed in the supplementary material.

Standard Merrifield Tactics.[43] Syntheses were performed manually in a filter-frit reaction vessel or with a Beckman 990S synthesizer (Palo Alto, CA). *t*-Boc-aminoacyl or MBHA resins were subjected to the following schedule of operations for each cycle of aminoacylation: (i) two washings (5 and 20 min) with trifluoroacetic acid/CH_2Cl_2 (1:1) or 4 N HCl/dioxane containing 1% ethanedithiol as a carbocation scavenger; (ii) one wash with 2-propanol containing 1% ethanedithiol; (iii) neutralization with triethylamine/CH_2Cl_2 or $CHCl_3$ (1:9) for 2 min; (iv) one wash with methanol; (v) reneutralization; (vi) three washings with methanol; (vii) three washings with CH_2Cl_2; (viii) 90-min reaction with a 3 molar excess of (a) *t*-Boc-amino acid and carbodiimide (dicyclohexyl- or diisopropylcarbodiimide) in CH_2Cl_2 or dimethylformamide in situ or (b) *t*-Boc-amino acid, 1-hydroxybenzotriazole, and carbodiimide in dimethylformamide in situ; (ix) three washings with dimethylformamide; (x) three washings with methanol; (xi) three washings with CH_2Cl_2; (xii) Kaiser test,[38] quantitative ninhydrin assay,[39] or amino acid analysis; (xiii) "recouple" (steps vi–xii) or cap with acetic anhydride/CH_2Cl_2, dimethylformamide (1:5), or acetylimidazole in CH_2Cl_2. The N-terminal *t*-Boc protection was removed on completion of the target sequence. After being dried under high vacuum, the peptide was cleaved from the support and concomitantly deprotected in liquid HF at −20 °C for 30 min and 0 °C for 60 min ("high" HF)[59] in the presence of anisole (approximately 1 mL/g of peptidoresin) and/or ethyl methyl sulfide (approximately 0.3 mL/g of peptidoresin), triturated, and washed with anhydrous ether, and the crude peptide was extracted from the resin with aqueous acetic acid or ammonia solution prior to lyophilization.

Optimized Tactics.[5,7] Syntheses were accomplished manually in a filter-frit reaction vessel with Boc-aminoacyl-Pam resins.[55,56] The following schedule of operations was performed for each cycle of the stepwise syntheses: (i) two washings (5 and 20 min) with trifluoroacetic acid/CH_2Cl_2 (1:1) containing 1% ethanedithiol or 0.5% indole as carbocation scavengers; (ii) two washings with CH_2Cl_2; (iii) two washings with 2-propanol; (iv) neutralization with diisopropylethylamine/dimethylformamide (1:19) for 2 min; (v) one wash with methanol; (vi) reneutralization; (vii) three washings with methanol; (viii) three washings with CH_2Cl_2; (ix) three washings with dimethylformamide; (x) 30–90-min reaction with (a) a 3 molar excess of the preformed symmetric anhydride[60] in dimethyl formamide (preformed in CH_2Cl_2, evaporated and redissolved) or (b) a 3–5 molar excess of the preformed hydroxybenzotriazolyl active ester[61] in dimethylformamide (both in a double-coupling protocol[7]); (xi) three washings with dimethylformamide; (xii) three washings with methanol; (xiii) three washings with CH_2Cl_2; (xiv) a quantitative ninhydrin test;[39] (xv) when incomplete aminoacylations were encountered, the peptidoresin was reneutralized and "recoupled" in dimethylformamide by using the alternative preformed activated intermediate where possible [if a complete reaction was still not attained, other acylating reagents such as (benzotriazol-1-yloxy)tris(dimethylamino)-phosphonium hexafluorophosphate[12,40] or chaotropic agents such as trifluoroethanol[62] were employed]; (xvi) acetylation with acetic anhydride or acetylimidazole in dimethylformamide, as required, was followed by a quantitative ninhydrin test.[39] After removal of the *t*-Boc N-terminal protection, the peptidoresin was dried under high vacuum and cleaved from the resin support, with concomitant deprotection, by either the "high" HF[59] procedure or by the "low–high" HF technique[63] [with dimethyl sulfide and *p*-cresol as scavenger (2.5:6.5:1) for 2 h at 0 °C and, after evaporation of the initial HF and dimethyl sulfide under high vacuum, with HF/*p*-cresol (9:1) for 30 min at 0 °C]. Trituration and washing with anhydrous ether was followed by extraction with aqueous acetic acid or ammonia solution and lyophilization.

Purification of Synthetic Peptides. The crude peptide lyophilizates were dissolved and applied directly through the pump onto a cartridge packed with 15–20-μm C18 silica (The Separations Group, Hesperia, CA) fitting a Waters Associates Prep LC500 chromatograph[64] and eluted with a gradient of acetonitrile in 0.1 M ammonium acetate (pH 6.5) or 0.25 M triethylammonium phosphate (pH 2.25). The eluants were monitored by analytical C18 RP HPLC so that cuts could be made and often rerun until a pure product was obtained. Peptides purified by using the triethylammonium phosphate buffer required desalting with a gradient of acetonitrile in 0.1% trifluoroacetic acid prior to lyophilization. Homogeneities of >98% (the integrated area under the peak vs the total integrated areas recorded at 210 nm during analytical C18 RP HPLC) were obtained after lyophilization.

(50) (a) Weygand, F.; Steglich, W.; Bjarnason, J.; Akhtar, R.; Khan, N. M. *Tetrahedron Lett.* **1966**, 3483–7. (b) Weygand, F.; Steglich, W.; Bjarnason, J.; Akhtar, R.; Chytil, N. *Chem. Ber.* **1968**, *101*, 3623–41. (c) Weygand, F.; Steglich, W.; Bjarnason, J. *Chem. Ber.* **1968**, *101*, 3642–8.

(51) Pietta, P. G.; Biondi, P. A.; Brenna, O. *J. Org. Chem.* **1976**, *41*, 703–4.

(52) Abbreviations: Tos, tosyl; Xan, xanthenyl; OBzl, benzyl ester; OcHex, cyclohexyl ester; 4-MeBzl, *p*-methylbenzyl; Bom, benzyloxymethyl; Dnp, 2,4-dinitrophenol; Cl-Z, 2-chlorobenzyloxycarbonyl; Bzl, benzyl ether; For, formyl; Br-Z, 2-bromobenzyloxycarbonyl; Cl_2Bzl, 2,6-dichlorobenzyl ether; Pam, phenylacetamidomethyl; MBHA, *p*-methylbenzhydrylamine; RP HPLC, reverse-phase high-performance liquid chromatography.

(53) Olsen, R. K. *J. Org. Chem.* **1970**, *35*, 1912–5.

(54) Horiki, K.; Igano, K.; Inouye, K. *Chem. Lett.* **1978**, 165–8.

(55) Mitchell, A. R.; Kent, S. B. H.; Engelhard, M.; Merrifield, R. B. *J. Org. Chem.* **1978**, *43*, 2845–52.

(56) Tam, J. P.; Kent, S. B. H.; Wong, T. W.; Merrifield, R. B. *Synthesis* **1979**, *12*, 955–7.

(57) Gisin, B. F. *Helv. Chim. Acta* **1973**, *56*, 142–3.

(58) Milton, R. C. de L.; Mayer, E.; Walsh, J. H.; Rivier, J. E.; Dykert, J.; Lee, T. D.; Shively, J. E.; Reeve, J. R. *Int. J. Pept. Protein Res.* **1988**, *32*, 141–52.

(59) Rivier, J.; Kaiser, R.; Galyean, R. *Biopolymers* **1978**, *17*, 1927–38.

(60) Stewart, J. M.; Young, J. D. *Solid Phase Peptide Synthesis*, 2nd ed.; Pierce Chemical Co.: Rockford, IL, 1984; p 80.

(61) Stewart, J. M.; Young, J. D. *Solid Phase Peptide Synthesis*, 2nd ed.; Pierce Chemical Co.: Rockford, IL, 1984; p 82.

(62) Yamashiro, D.; Blake, J.; Li, C. H. *Tetrahedron Lett.* **1976**, 1469–72.

(63) Tam, J. P.; Heath, W. F.; Merrifield, R. B. *J. Am. Chem. Soc.* **1983**, *105*, 6442–55.

(64) Rivier, J.; McClintock, R.; Galyean, R.; Anderson, H. *J. Chromatogr.* **1984**, *288*, 303–28.

Characterization of Synthetic Peptides. Purified peptide products were characterized by C18 RP HPLC for homogeneity, and by amino acid analysis following gas-phase hydrolysis in 5.7 N HCl.[65] In all cases a single peak was obtained and the amino acid compositions were consistent with the calculated molar ratios. Certain peptides were also subjected to sequence analysis, either by gas-phase methodology[66] with the phenylthiohydantoin derivatives quantitated by RP HPLC or, in the case of N- and C-terminally blocked peptides, by FAB mass spectrometry.[58,67]

FTIR Spectrometry. KBr disks were prepared with dried peptidoresin (100 mg of KBr:3 mg of peptidoresin) and scanned from 4000 to 400 cm^{-1} (resolution 1.0 cm^{-1}) with a Bruker IFS48 spectrometer (Rheinstetten, BRD).

Acknowledgment. We thank S. B. H. Kent for valuable advice and encouragement, R. Roeske for reading the manuscript and helpful suggestions, M. Narita for pioneering work in solution-phase peptide synthesis, which provided the inspiration, C. Tobler for computer programs, and L. Odes for assistance in manuscript preparation. This work was supported by the Medical Research Council of South Africa and the University of Cape Town.

Supplementary Material Available: Sequences, cumulative ($\langle P_c^* \rangle$) values for the predictive method, and synthetic details of the 101 syntheses (986 aminoacylations) used as the database for the histograms shown in Figure 4 (34 pages). Ordering information is given on any current masthead page.

(65) Bidlingmeyer, B. A.; Cohen, S. A.; Tarvin, T. L. *J. Chromatogr.* **1984**, *336*, 93-104.
(66) Brandt, W. F.; Alk, H.; Chauhan, M.; von Holt, C. *FEBS Lett.* **1984**, *174*, 228-32.
(67) Seki, S.; Kambara, H.; Naoki, H. *Org. Mass Spectrom.* **1985**, *20*, 18-24.

Multicyclic Polypeptide Model Compounds. 1. Synthesis of a Tricyclic Amphiphilic α-Helical Peptide Using an Oxime Resin, Segment-Condensation Approach[1]

George Ösapay* and John W. Taylor*

Contribution from the Laboratory of Bioorganic Chemistry and Biochemistry, The Rockefeller University, New York, New York 10021-6399. *Received November 29, 1989*

Abstract: An idealized model amphiphilic α-helical peptide, *cyclo*(3-7,10-14,17-21)H-[LysLeuLysGluLeuLysGlu]$_3$-OH (peptide **1-1-1**), comprising three repeats of a Lys^3-Glu^7 side-chain bridged heptapeptide, has been synthesized by a generally applicable segment-condensation approach that involves a novel solid-phase cyclization reaction. The linear heptapeptide, Boc-Lys-(2Cl-Z)LeuLys(Trt)Glu(OBzl)LeuLys(2Cl-Z)Glu(oxime resin)-OPac, was built on a *p*-nitrobenzophenone oxime derivatized polystyrene solid support by standard methods. After selective detritylation with TFA, the Lys^3 ϵ-amino group was liberated with DIEA, and then intrachain cyclization in the presence of AcOH released the protected cyclic heptapeptide precursor to peptide **1-1-1** into the solvent in 61% yield and high purity. Selective N$^\alpha$- and C$^\alpha$-group deprotection, followed by two solution-phase segment-condensation reactions and then complete deprotection with trimethylsilyl triflate, yielded peptide **1-1-1**. Circular dichroism spectra indicated that peptide **1-1-1** adopted mostly disordered conformations in aqueous solution, but a high α-helix content was induced in 50% TFE and upon adsorption of peptide **1-1-1** from aqueous solution onto siliconized quartz slides.

Introduction

Many biologically active peptides have flexible structures and exist in multiple disordered conformational states in aqueous solution. Studies of peptide models indicate that the conformations induced in these peptides by their functional environment, which is generally an interface, will often be a folded structure not present in solution and may include segments of amphiphilic secondary structure.[2] For several such peptides, including the serum apolipoprotein A-I, the bee venom peptide toxin melittin, and the peptide hormones β-endorphin and calcitonin, the functional requirement for amphiphilic α-helical structures has been demonstrated through the design and study of analogues incorporating minimally homologous models of these structures that are able to reproduce all of the functional properties of the native sequences.[2,3] However, similar studies of other peptide hormones, including glucagon,[4] calcitonin gene-related peptide,[5] and neuropeptide Y,[6] are less conclusive, because multiple substitutions in the amino acid sequence result in greatly reduced potencies. In these cases, which may represent the majority of intermediate-sized peptide hormones acting at cell-surface receptors, it is likely that functional requirements for both primary and secondary structural features are superimposed in the same peptide segments. For example, when highly conservative, helix-stabilizing substitutions in a potential amphiphilic α-helical segment of glucagon were limited to only three residue positions, enhanced receptor-binding potency was achieved.[7] In order to investigate structure-activity relationships in such peptides, the study of conformationally constrained analogues,[8] where primary structure is largely conserved, may be more appropriate than the use of minimally homologous model peptides. However, meaningful conclusions concerning potential folded conformations that may involve 10-20 residues or more will require the introduction of multiple conformational constraints, each of which is compatible with the conformation in question and limits the conformational freedom of a peptide segment several residues long. With this

(1) Dedicated to the memory of Professor Emil Thomas Kaiser.
(2) (a) Kaiser, E. T.; Kézdy, F. J. *Science* **1984**, *223*, 249-255. (b) Taylor, J. W.; Kaiser, E. T. *Pharmacol. Rev.* **1986**, *38*, 291-319.
(3) Kaiser, E. T.; Kézdy, F. J. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 1137-1143.
(4) (a) Musso, G. F.; Kaiser, E. T.; Kézdy, F. J.; Tager, H. S. In *Proceedings of the Eighth American Peptide Symposium*; Hruby, V. J., Rich, D. H., Eds. Pierce Chemical Co.: Rockford, IL, 1983; pp 365-368. (b) Musso, G. F.; Kaiser, E. T.; Kézdy, F. J.; Tager, H. S. *Biochem. Biophys. Res. Commun.* **1984**, *119*, 713-719.
(5) Lynch, B.; Kaiser, E. T. *Biochemistry* **1988**, *27*, 7600-7607.
(6) Minakata, H.; Taylor, J. W.; Walker, M. W.; Miller, R. J.; Kaiser, E. T. *J. Biol. Chem.* **1989**, *264*, 7907-7913.
(7) Krstenansky, J. L.; Zechel, C.; Trivedi, D.; Hruby, V. J. *Int. J. Pept. Protein Res.* **1988**, *32*, 468-475.
(8) Hruby, V. J. *Life Sci.* **1982**, *31*, 189-199.

0002-7863/90/1512-6046$02.50/0