Facile Synthesis of Homogeneous Artificial Proteins

Keith Rose

Contribution from the Department of Medical Biochemistry, University Medical Center, CH 1211 Geneva 4, Switzerland

Received August 27, 1993®

Abstract: A new approach, involving oxime formation, is described for the synthesis of stable macromolecules of defined structure and is exemplified with polypeptides. In a first step, purified polypeptide fragments, carrying either aldehyde or aminooxy groups but devoid of protecting groups, are prepared. In a second step, these fragments spontaneously self-assemble on an appropriate synthetic template under very mild conditions through formation of oxime bonds. The resulting oximes are stable in water at room temperature at pH 2–7. One of the compounds made, the covalent structure of which was defined by electrospray mass spectrometry [observed MW 19 916.61 ± 3.05 D; calculated MW 19 916.10 (average isotopic composition)], would seem to be the largest artificial protein ever made, in pure form in good yield, by controlled total synthesis. Possible applications of this method of synthesis of homogeneous macromolecules include vaccines, biosensors, and enhanced peptide libraries.

Introduction

There is considerable interest[1] in producing pure synthetic organic molecules in the molecular weight range 2000–50 000 D. This size range is accessible to "dendrimer" chemistry,[1] but the range of functional groups allowed is generally rather limited.

Polypeptide synthesis using solid-phase methods pioneered by Merrifield[2] is capable of producing homogeneous macromolecules up to about 10 000 D (80–100 amino acid residues).[3] The poor solubility (in organic solvents) of large protected fragments is the major problem facing extension of total synthesis of homogeneous material beyond the 10 000-D limit.[3] Recently, regiospecific condensation of unprotected polypeptide fragments has been achieved by chemoselective ligation in aqueous medium.[3,4] In this approach, pairs of groups not present elsewhere on the fragments to be joined, and having complementary reactivity (e.g. a thiol and a bromoacetyl function), are allowed to react. Chemoselective ligation has been used to prepare, via thiol chemistry, a multiple antigenic peptide,[5] but the product had poor HPLC characteristics and was not fully characterized. Thiol chemistry has also been used to prepare a totally synthetic, functioning HIV protease analog[3] and a template-assembled synthetic protein of MW 6650,[4] but the thioester products were not stable at neutral pH.[3]

I decided to use chemoselective ligation of unprotected polypeptides in an attempt to make totally synthetic macromolecules of controlled structure and of molecular weight in the protein range (greater than 10 000 D). To achieve spontaneous self-assembly involving up to nine unprotected fragments, I opted for oxime formation rather than the previously exploited thiol chemistry which is subject to unwanted disulfide bond formation[5] or involves labile thioester bonds.[3,4] The precursor *O*-alkylhydroxylamines are easily synthesized, and oxime bridges have been used successfully to attach a variety of substances to proteins,[6–9] often for *in vivo* applications.

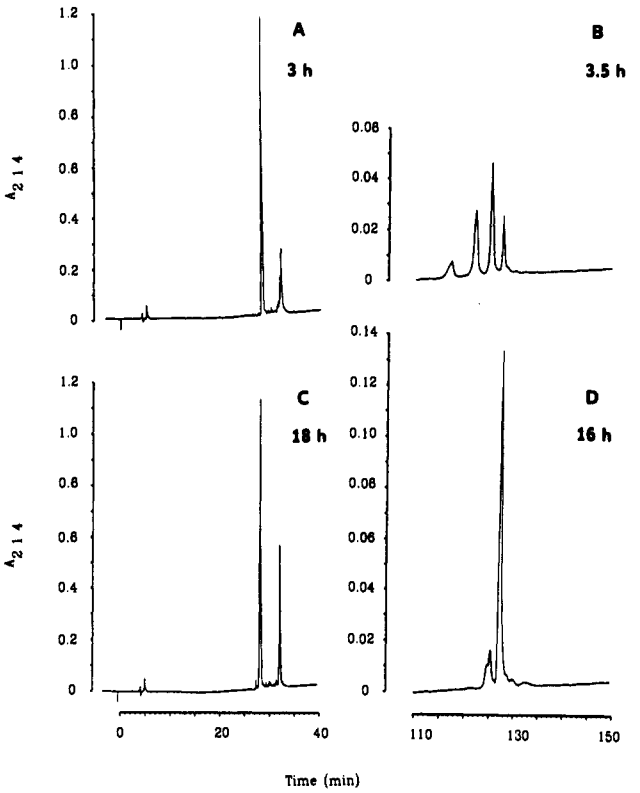

Figure 1. Analysis by reverse-phase HPLC of the self-assembly reaction between the peptide derivative NH_2OCH_2CO-KLEEQRPERVKG-OH and the template O=CH-CO-Gly_3-[Lys(COCHO)]$_5$-Gly-OH). Panel A (reaction time of 3 h; 5-min isocratic 100% A, then gradient of 2% B/min from 0% solvent B): the large peak at retention time of 28 min corresponds to excess peptide and the small group at 30–32 min to oxime formation. Panel B (reaction time of 3.5 h; 5-min isocratic 100% A, then gradient 0.2% B/min from 0% B) shows (left to right) tri-, tetra-, penta-, and hexaoxime products. Panel C (reaction time of 18 h, gradient 2%/min) shows further reaction progress. Panel D (reaction time of 16 h, gradient 0.2%/min) shows a high yield of hexaoxime product, retention time 127 min, and some pentaoxime isomers at about 125 min, confirmed by ESI-MS.

Results and Discussion

Synthetic polypeptides (NH_2OCH_2CO-KLEEQRPERVKG-OH and NH_2OCH_2CO-ELGGGPGAGSLQPLALEGSLQKR-OH) carrying an N-terminal aminooxyacetyl group were incu-

® Abstract published in *Advance ACS Abstracts*, December 15, 1993.

(1) Dagani, R. *Chem. Eng. News* **1993**, *12 April*, 26–28; Hodge, P., *Nature* **1993**, *362*, 18–19.
(2) Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149–2154.
(3) Schnölzer, M.; Kent, S. B. H. *Science* **1992**, *256*, 221–225.
(4) Dawson, P. E.; Kent, S. B. H. *J. Am. Chem. Soc.* **1993**, *115*, 7263–7266.
(5) Drijfhout, J. W.; Bloemhoff, W. *Int. J. Pept. Protein Res.* **1991**, *37*, 27–32.
(6) Kurth, M.; et al. *J. Med. Chem.* **1993**, *36*, 1255–1261.
(7) Ryser, J. E.; et al. *J. Nucl. Med.* **1992**, *33*, 1766–1773.
(8) Mikola, H.; Hänninen, E. *Bioconjugate Chem.* **1992**, *3*, 182–186.
(9) Webb, R. R.; Kaneko, T. *Bioconjugate Chem.* **1990**, *1*, 96–99.

0002-7863/94/1516-0030$04.50/0

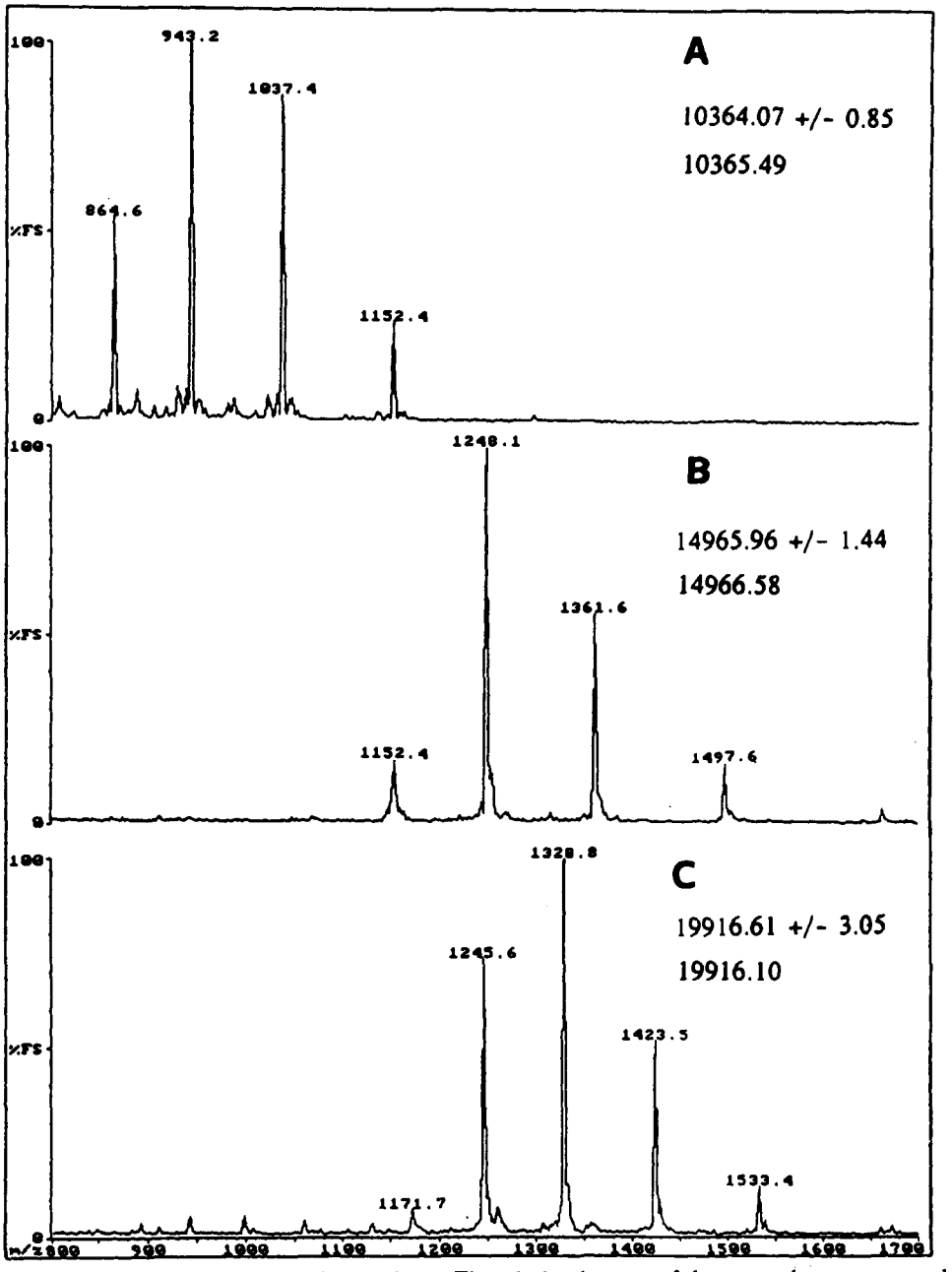

Figure 2. Electrospray ionization mass spectra of polyoxime products. The calculated masses of the expected structures are shown just below the measured values, and all are in excellent agreement. Shown are spectra of the hexaoximes formed between the template O=CH-CO-Gly_3-[Lys(COCHO)]$_5$-Gly-OH and the peptides NH_2OCH_2CO-KLEEQRPERVKG-OH (panel A) and NH_2OCH_2CO-ELGGGPGAGSLQPLALEGSLQKR-OH (panel B) and the spectrum of the octaoxime formed between the template O=CH-CO-Gly_3-[Lys(COCHO)]$_7$-Gly-OH and the peptide NH_2OCH_2CO-ELGGGPGAGSLQPLALEGSLQKR-OH (panel C).

bated separately with a poly(aldehyde) template molecule (O=CH-CO-Gly_3-[Lys(COCHO)]$_5$-Gly-OH or O=CH-CO-Gly_3-[Lys(COCHO)]$_7$-Gly-OH) in aqueous solution at pH 4.6, and the oximation reactions were followed by reverse-phase HPLC. Groups of peaks corresponding to isomeric forms of di, tri, tetra, etc., oximes were soon visible (Figure 1A,B) and were characterized by electrospray ionization mass spectrometry (ESI-MS, see below). With time (16–18 h), the reactions proceeded surprisingly close to completion, e.g. about 90% in the case of the hexaoxime (Figure 1D), the remainder being pentaoxime. The final product was in each case characterized by mass spectrometry. ESI-MS provided experimentally determined molecular weights of 10364.07 ± 0.85 D for the hexaoxime produced by reaction between NH_2OCH_2CO-KLEEQRPERVKG-OH and O=CH-CO-Gly_3-[Lys(COCHO)]$_5$-Gly-OH (Figure 2A), 14965.96 ± 1.44 D for the hexaoxime formed between NH_2OCH_2CO-ELGGGPGAGSLQPLALEGSLQKR-OH and O=CH-CO-Gly_3-[Lys(COCHO)]$_5$-Gly-OH (Figure 2B), and 19916.61 ± 3.05 D for the octaoxime formed between NH_2OCH_2CO-ELGGGPGAGSLQPLALEGSLQKR-OH and O=CH-CO-Gly_3-[Lys(COCHO)]$_7$-Gly-OH (Figure 2C), all in excellent agreement with the calculated values of 10 365.49, 14 966.58, and 19 916.10, respectively. HPLC analysis showed the hexaoxime to be stable at room temperature at pH 2.1, 4.6, and 7.0 over at least 24 h (Figure 3): any hydrolysis of the oxime bonds would have given rise to earlier eluting forms (compare Figure 1B).

The polarity of the oxime bond was readily reversed by reacting an aldehydic form of the polypeptide (O=CH-CO-KLEEQRPERVKG-OH), formed by periodate oxidation of the seryl polypeptide according to Gaertner et al.,[10] with a hexa(aminooxy)-template molecule, NH_2OCH_2CO-Gly_3-[Lys($COCH_2ONH_2$)]$_5$-Gly-OH, in aqueous medium at pH 4.6. Oxime formation again

(10) Gaertner, H. F.; et al. *Bioconjugate Chem.* **1992**, *3*, 262–268.

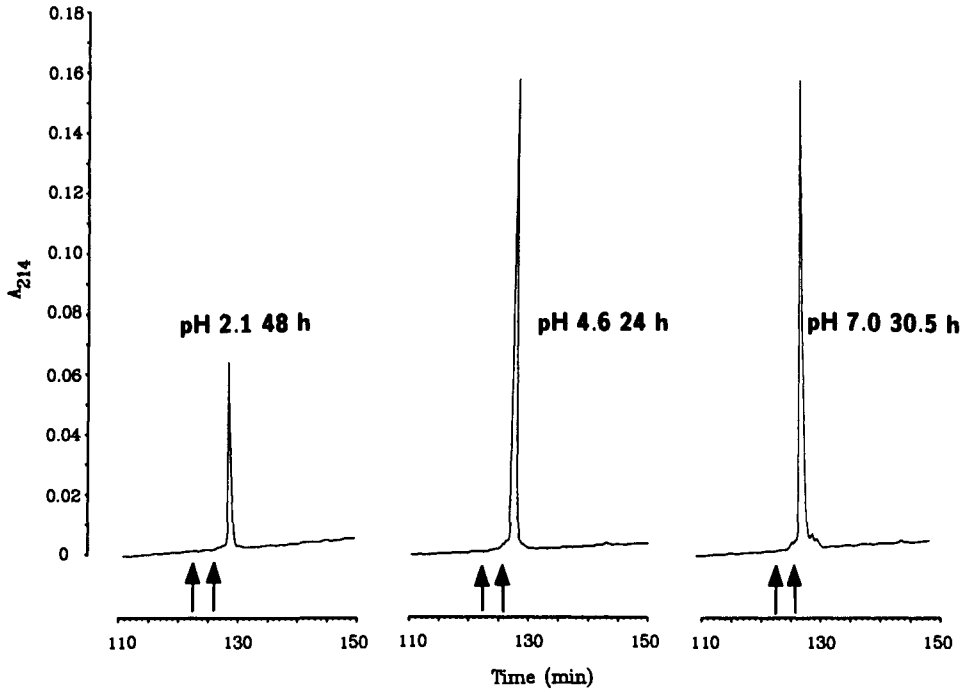

Figure 3. HPLC analysis of the purified hexaoxime described in the legend to Figure 1. Hexaoxime was incubated at room temperature at 0.1 mg/mL in 0.1% TFA (pH 2.1, 48 h), 0.1 M acetate (pH 4.6, 24 h), and phosphate buffered saline (pH 7.0, 30.5 h). HPLC conditions were essentially the same as those for Figure 1B,D. No evidence of hydrolysis of the hexaoxime was found; arrows indicate the elution positions of the tetra- and pentaoximes (compare Figure 1, parts B and D).

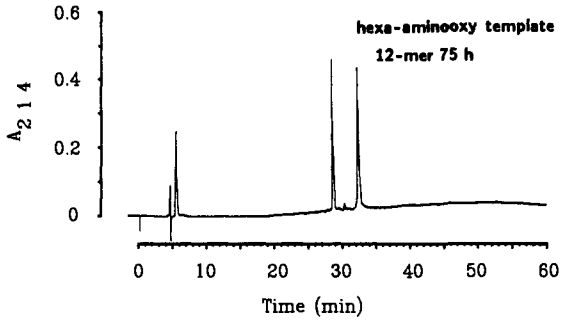

Figure 4. Analysis by HPLC of the self-assembly reaction between the peptide derivative O=CH-CO-KLEEQRPERVKG-OH and the template NH_2OCH_2CO-Gly_3-[$Lys(NH_2OCH_2CO-)]_5$-Gly-OH. HPLC conditions were the same as those for Figure 1A. Even after 75 h (shown here), there are no signs of side reactions. Excess peptide elutes after 28.5 min, and hexaoxime product, characterized by ESI-MS, after 32 min.

proceeded essentially to completion (Figure 4) to yield an isomer of the product whose synthesis is monitored in Figure 1. ESI-MS provided an experimentally determined molecular weight of 10 365.72 ± 0.88 D, in excellent agreement with the calculated value of 10 365.49.

The polyoximes possess multiple copies of a single oligopeptide and are thus comparable in structure with the branched multiple antigenic peptides previously described by Tam,[11] and there is some parallel with Mutter's TASPs (template-assembled synthetic proteins) and similar molecules.[12] However, these amide-bonded structures are exceedingly difficult (if not impossible) to obtain in pure form at high molecular weight for the reasons discussed in refs 3 and 4. In contrast, polyoximes containing up to 195 amino acid residues can be prepared easily and rapidly in pure form. We have prepared 11 mg of the hexaoxime whose synthesis is monitored in Figure 1. The compound is a white powder, freely soluble in water and in phosphate-buffered saline.

It should be emphasized that the unprotected polypeptide derivatives and templates described are easily prepared and purified, and the clean reaction of oxime formation permits the easy, rapid assembly of synthetic macromolecules which are devoid of protecting groups, soluble and themselves easily purified. Since it is possible to create, site-specifically on a recombinant-derived polypeptide, an aldehyde group,[10,13] there is no reason why polyoximation should not be successful in creating defined oligomers of recombinant-derived polypeptides: for such a purpose, the reactive groups on the template would probably have to be spaced somewhat further apart. Rather few steps are required to add to the template, prior to oximation, a reporter group (such as biotin or a chelator for a radiometal), a lipophilic anchor (such as *N*-palmitoyl-*S*-[2,3-bis(palmitoyloxy)-(2*RS*)-propyl]-[*R*]-cysteinyl[14,15]), or an orthogonal reactive group (such as bromoacetyl or a masked thiol such as *S*-acetylthioacetyl): we have successfully attached, separately, biocytin and *N*-acetylcysteine. It is then possible, respectively, to follow the polyoximes via the reporter group, to devise improved vaccines possessing both a lipophilic anchor and a defined molecular structure, or to attach the polyoximes to other macromolecules, to each other, or to surfaces via the orthogonal reactive group. Immediate possible applications include defined polypeptide vaccines and biosensors. In addition, polyoximes have much to offer the peptide library field (briefly reviewed in ref 15). If we consider a library of just the 400 L,L-dipeptide amides, all carrying the aminooxyacetyl group, and a template possessing three aldehyde groups, then neglecting symmetry considerations the number of possible trioximes obtained on mixing should be $400^3 = 6.4 \times 10^7$. Appropriately chosen templates should act as "the palm of a hand" with the peptide "fingers" ready to close around a target structure.

Given that aldehyde groups are terminal to reducing sugars (or can be created on sugars by mild chemical or enzymic

(11) Tam, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 5409–5413.
(12) Mutter, M.; Tuchscherer, G. G.; Miller, C.; Altmann, K.-H.; Carey, R. I.; Wyss, D. F.; Labhardt, A. M.; Rivier, J. E. *J. Am. Chem. Soc.* **1992**, *114*, 1463–1470. Hahn, K. W.; Klis, W. A.; Stewart, J. M. *Science* **1990**, *248*, 1544–1547.
(13) Geoghegan, K. F.; Stroh, J. G. *Bioconjugate Chem.* **1992**, *3*, 138–146.
(14) Nardelli, B.; et al. In *Solid Phase Synthesis*; Epton, R., Ed. Intercept: Andover, UK, 1992; pp 241–249.
(15) Jung, G.; et al. In *Solid Phase Synthesis*; Epton, R., Ed. Intercept: Andover, UK, pp 227–235.

oxidation) and can be used for site-specific ligation[16] and that aldehydes may be created by terminal oxidation of RNA and subsequently used for the attachment of dyes,[17] polyoxime chemistry may be applicable outside the area of polypeptides. It is particularly interesting that proteins to which various molecules have been linked by oxime bonds have been shown to be useful *in vivo* (refs 6–8 and references cited therein).

Conclusion

Homogeneous branched polypeptides have been readily synthesized containing up to 195 amino acid residues and with molecular weights up to 19 916 D. This is larger than the layer block dendritic copolymer (10 242 D) and 127-mer poly(phenylacetylene) (18 054 D) recently described[1] and would seem to be the largest artificial protein ever made, in pure form in good yield, by controlled synthesis. A new area of synthetic and semisynthetic chemistry is opened up now that it is possible, through polyoximation, to rapidly and easily produce macromolecular synthetic molecules of defined structure.

Experimental Section

Materials and Methods. HPLC was performed on Waters equipment using columns packed with Machery Nagel Nucleosil 300-Å 5-μm C_8 particles. The analytical column (250- × 4-mm i.d.) was operated at 0.6 mL/min, and the preparative column (250- × 21-mm i.d.) at 5 or 10 mL/min, monitoring at 214 nm. Solvent A was 0.1% TFA, and solvent B was 0.1% TFA in 90% acetonitrile. Mass spectra were obtained in positive ion mode on a Trio 2000 instrument (Fisons Analytical) equipped with a 3000-amu rf generator. Samples were infused at 2 μL/min in a solvent consisting of methanol/water/acetic acid (49.5/49.5/1 by volume). Spectra contain signals due to various protonation states of the samples, and the m/z information is used to measure the mass of the uncharged artificial protein. The standard deviations given are an indication of the precision of the measurements rather than the accuracy (accuracy, determined with standard proteins, was always better than 1.5 amu in 10 000).

Peptide Synthesis. A model 430A machine from Applied Biosystems Inc. (ABI) was used with the standard Fmoc protocol and 0.5-mmol starting resin. Sasrin resin (Bachem) was used for synthesis of the linear peptides and Gly-OCH_2-PAM polystyrene (ABI) for that of the branched templates. The aminooxyacetyl group was attached using Boc-aminooxyacetyl *N*-hydroxysuccinimide ester (0.1 M in dry DMSO, 2.5 equiv over each amino group, 50 mL for 0.69 g of resin, apparent pH 8–9 with *N*-methylmorpholine, room temperature, 2 h), which gave complete acylation. Ser residues were attached to templates using Boc-Ser(Bzl) *N*-hydroxysuccinimide ester in DMSO. Cleavage deprotection of linear peptides was achieved with a mixture of phenol (0.75 g), ethanedithiol (0.25 mL), thioanisole (0.5 mL), water (0.5 mL), and TFA (10 mL): after being stirred for 3 h, the mixture was filtered and the peptide precipitated with cold methyl *tert*-butyl ether, washed three times with ether, dried, taken up in water, and purified in portions by preparative HPLC. Cleavage deprotection of branched templates was achieved by treating with TFA, stirring for 30 min, then adding 1/10 volume trifluoromethanesulfonic acid; after agitation for 1 h, the peptide was precipitated with dry ether, washed three times with dry ether, and dried under vacuum. The resulting mixture of peptide and cleaved resin was taken up in water (the peptide is soluble and the resin is not) and filtered. After lyophilization, the filtrate was dissolved in water and purified in portions by preparative HPLC. All peptides gave a single peak on analytical HPLC and had the expected ESI mass spectrum. N-terminal Ser residues on linear peptides (and *N*-ε-Ser residues on branched templates) were oxidized to glyoxylyl functions according to Gaertner et al.[10] and repurified by HPLC.

Oxime Formation. Typical conditions were as follows. Reaction between the peptide derivative NH_2OCH_2CO-KLEEQRPERVKG-OH and the template O=CH-CO-Gly_3-[Lys(COCHO)]$_5$-Gly-OH was initiated by mixing 6.7 μL of template (10 mM in water) with 200 μL of peptide (10 mM in 0.1 M acetate buffer, pH 4.6, counterion sodium) at 22 °C. The peptide (2 μmol) was thus present in excess over the template (67 nmol, i.e. about 400 nmol of aldehyde groups). Reaction between the peptide derivative O=CH-CO-KLEEQRPERVKG-OH and the template NH_2OCH_2CO-Gly_3-[Lys(NH_2OCH_2CO)]$_5$-Gly-OH was initiated by adding the template (1 μL, 10 mM in water) to a mixture of 180 μL of buffer (0.1 M acetate, pH 4.6, counterion sodium) and the peptide (30 μL, 10 mM in water), mixing, and incubating at 22 °C. In both cases, the peptide was present in a 5-fold molar excess over each reactive group on the template, reaction progress was followed by HPLC (Figures 1 and 4), and the product was characterized by electrospray MS (Figure 2).

Acknowledgment. I thank Irène Rossitto-Borlat for expert technical assistance, P.-O. Regamey for mass spectra, and Lan Vu for a gift of protected peptidyl resin. I am grateful to the Swiss National Science Foundation, Schmidheiny Foundation, and Sandoz Stiftung for purchase of an automatic HPLC injector and the mass spectrometer.

(16) O'Shannessy, D. J.; Quarles, R. H. *J. Immunol. Methods* **1987**, *99*, 153–161.

(17) E.g.: Reines, S. A.; Cantor, C. R. *Nucleic Acids Res.* **1974**, *1*, 767–786.