DOI: 10.1002/cbic.200700720

Glutamate–Glycine and Histidine–Glycine Co-oligopeptides: Batch Co-oligomerization versus Pulsed Addition of *N*-Carboxyanhydrides

Carole Lamy,[b] Jérôme Lemoine,[d] Denis Bouchu,[c] Peter Goekjian,*[b] and Peter Strazewski*[a]

Prebiotic α-amino acids are thought to have evolved into today's proteins because copolymers derived from these residues offer catalytically competent side chains within an evolvable dynamic scaffold. The manner in which peptides of main-chain lengths of above thirty residues, and that bear different side-chain functional groups might have formed spontaneously from chemically activated monomeric precursors is an issue that has attracted considerable attention. One possibility is the appearance of short peptides that are capable of catalyzing the formation and hydrolysis of amide bonds between peptides; these peptides thus bear, depending on the conditions, ligase or peptidase activities.[1] The ability to combine with themselves and other peptides might significantly enhance the variety and higher-order structural stability. An effective enrichment of stably folded peptides that bear a few catalytically competent side chains could result from the low expected solubility of the vast majority of the multimers formed from simple amino acids. The synthesis of libraries of random oligopeptides from a mixture of different amino acids under chemically simple conditions has thus become an important challenge.

The central difficulty in making such "random" oligopeptides lies in the fact that monomeric precursors of amino acids that bear potentially useful side-chain functional groups can polymerize at very different rates. Based on the principles of co-polymerization,[2] the amino acid distribution within a copolymer will not necessarily reflect the composition of the initial solution of the monomeric precursors. When rate differences between precursors reach several orders of magnitude, batch copolymerization of the mixture results in the exclusive formation of homopolymers rather than the desired copolymer. A further challenge lies in obtaining polymer chains of sufficient length in aqueous solutions. Whereas peptide chains above 30-mers have been achieved in nonaqueous solutions that contained mixtures of different *N*-carboxyanhydrides (NCAs = Leuch's anhydrides) that polymerized at similar rates,[3] and in homoplymers of simple amino acids such as oligoalanines,[4,5] the longest reported co-oligopeptide that has been synthesized in situ from aqueous mixtures of amino acid–NCAs are pentamers.[6] In order to generate copolymers that bear the requisite features for the potential evolution of catalytic activity, it is thus necessary to overcome the difficulties inherent to the large differences in polymerization rates of key amino acids under given aqueous conditions.

Amino acids such as histidine, glutamate, and serine are catalytically competent,[7] whereas glycine, alanine, valine, and similar amino acids can be considered to be good spacers and folders. Our preliminary studies focused on the preparation of co-oligomers of glycine (Gly, G) and glutamate (Glu, E), which are two partners that oligomerize at similar speeds. The addition of an unbuffered D_2O solution of Gly or Glu (120 mM) to dry 1,1′-carbonyldiimidazole (CDI, 2.5 equiv) led within a few minutes to the corresponding NCAs; these were readily identified by ^{1}H NMR spectroscopy and electrospray mass spectrometry. Homo-oligomerization of the individual solutions by initiation with free amino acid occurred within a few hours at room temperature to give oligomers Gly_n ($n = 4$–27) or Glu_m ($m = 2$–12; see the Supporting Information).[8] Batch co-oligomerization of an equimolar mixture of CDI-activated Gly and Glu (60 mM each, Figure 1) was followed by NMR spectroscopy. After the addition of initiating Gly, Gly_n–Glu_m co-oligomers were obtained that ranged from tetramers to 21-mers and contained evenly distributed amounts of both amino acids, with $n = 3$–11 and $m = 4$–10, as revealed by ESI$^+$ and MALDI-TOF$^+$ mass spectrometry.

We chose histidine and glycine as the first "unequal pair" of partners to be co-oligomerized and wish to report here our proof of concept in activating histidine (His, H, **1**), a highly catalytically competent residue, yet slow polymerizer, and in co-oligomerizing it with glycine (Gly, G), which is perhaps the fastest polymerizer among the proteinogenic amino acids. The process provides different populations of Gly–His co-oligomers where the peptide lengths and His contents can be varied in a controlled fashion.

It has long been recognized that His does not oligomerize under the same conditions as the NCAs of Gly, Glu and other amino acids.[9] The ^{1}H NMR spectroscopy data in Figure 2

[a] *Prof. Dr. P. Strazewski*
Laboratoire de Synthèse de Biomolécules, ICBMS UMR 5246
CNRS–Université Claude Bernard Lyon 1, Université de Lyon
Bâtiment Chevreul (5ième étage), 43 Boulevard du 11 Novembre 1918
69622 Villeurbanne Cedex (France)
Fax: (+33) 472431323
E-mail: strazewski@univ-lyon1.fr

[b] *C. Lamy, Prof. Dr. P. Goekjian*
Laboratoire de Chimie Organique II-Glycochimie, ICBMS UMR 5246
CNRS–Université Claude Bernard Lyon 1, Université de Lyon
C.P.E. Lyon–Bâtiment 308D, 43 Boulevard du 11 Novembre 1918
69622 Villeurbanne Cedex (France)
Fax: (+33) 478-898-914
E-mail: goekjian@univ-lyon1.fr

[c] *Dr. D. Bouchu*
Centre Commun de Spectrométrie de Masse, ICBMS UMR 5246
CNRS–Université Claude Bernard Lyon 1, Université de Lyon
C.P.E. Lyon–Bâtiment 308D, 43 Boulevard du 11 Novembre 1918
69622 Villeurbanne Cedex (France)

[d] *Prof. Dr. J. Lemoine*
Sciences Analytiques UMR 5180
CNRS–Université Claude Bernard Lyon 1–Université de Lyon
C.P.E. Lyon–Bâtiment 308D, 43 Boulevard du 11 Novembre 1918
69622 Villeurbanne Cedex (France)

Supporting information for this article is available on the WWW under http://www.chembiochem.org or from the author.

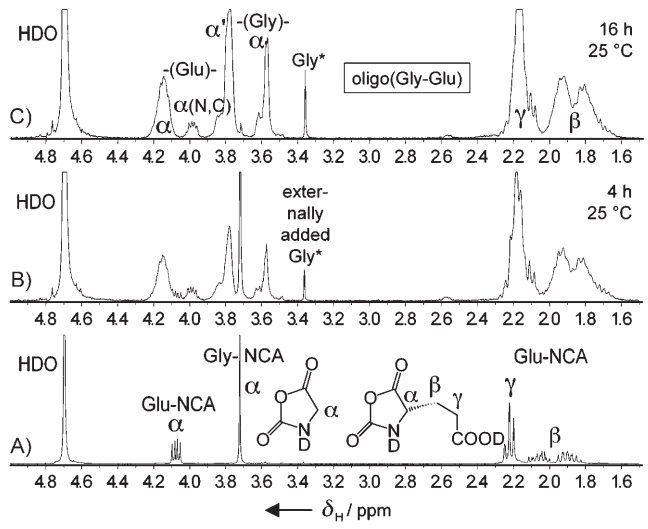

Figure 1. ^{1}H NMR spectra of an unbuffered D_2O solution of Gly and Glu (60 mM each) after the addition of CDI (2.5 equiv). A) Within the first 5 min at 25 °C the respective NCAs are formed. B) Oligomerization shown after 4 h, and C) after 16 h at 25 °C after the addition of Gly as initiator. See Supporting Information for the mass analysis of oligo(Gly–Glu).

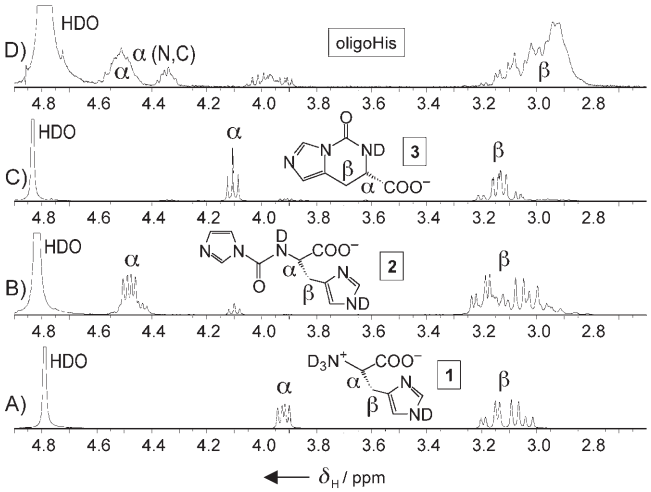

Figure 2. ^{1}H NMR spectra of the oligomerization of His (**1**) mediated by CDI and TMSCl. A) **1** in unbuffered D_2O. B) 5 min after addition of CDI (2.5 equiv) to **1** (120 mM in D_2O) at 25 °C, **2** accumulates as the main intermediate, which is identified as *N*-(imidazolylcarbonyl)histidine through electrospray MS–MS analyses of H_2O solutions under otherwise identical conditions (*m/z* 248 for the protic monoanion). C) After 1 h at 50 °C, **2** fully converts into the known bicyclic urea derivative **3** (*m/z* 180 for the protic monoanion), which oligomerizes within several hours when treated with either TMSCl or more slowly when acidified to pD 5.9 (measured pH* 5.5).[13] D) The ^{1}H NMR spectrum of oligoHis. See the Supporting Information also.

shows that the addition of a 120 mM aqueous solution of **1** to CDI results in the transient formation of the *N*-imidazolylcarbonyl derivative **2**.[10] **2** converts within a few hours into the thermodynamically stable, bicyclic urea derivative **3**,[9c] which does not oligomerize even after 24 h at room temperature, or after several hours at 50 °C. The addition of chlorotrimethylsilane (TMSCl) to an aqueous (D_2O or H_2O) solution of **3** led to the formation of His_4–His_{15} oligomers[9c] within a few hours at 50 °C, with a maximum abundance at His_6 and His_7 (Figure 2 and MALDI-TOF MS in the Supporting Information). Despite its high reactivity with water, TMSCl does not appear to act solely as a source of DCl, because polymerization at comparable pH in the absence of TMSCl is noticeably slower (Supporting Information). TMSCl thus makes **3** amenable to clean oligomerization under mild conditions.

We next turned our attention to the problem of producing His–Gly co-oligomers. The challenge arises from the fact that Gly oligomerizes rapidly at 50 °C, and batch co-oligomerization under these conditions will thus lead first to the formation of oligoGly, followed by the slower formation of oligoHis. Indeed, treatment of an equimolar mixture of urea **3** and Gly–NCA (60 mM each) with TMSCl at 50 °C gave primarily homo-oligomers, or at best, incorporation of a single Gly into oligoHis. The obvious solution was therefore to reduce the polymerization rate of the Gly species without limiting the molar ratio by limiting its concentration by the slow addition of activated Gly to oligomerizing His. Thus, treatment of an aqueous solution of His (120 mM) with CDI and warming to 50 °C for 1 hour provided **3**, which was treated with TMSCl. A separate fresh solution of the Gly–NCA was prepared by treatment of an aqueous solution of Gly (120 mM) with CDI at 0 °C. The pulsed addition of the Gly–NCA to the oligomerizing His urea over the course of 3 h at 50 °C, either 0.2 equivalents every 20 min (2 equiv total), 0.2 equivalents every 40 min (1 equiv total), or 0.4 equivalents every 40 min (2 equiv total), provided Gaussian-like distributions of His–Gly co-oligomers that showed extensive incorporation of Gly. The longest detected oligo(His–Gly) peptides were composed of, depending on the pulse addition mode, 24 to 29 amino acid residues (Figure 3).

An even distribution of Gly residues can be observed within each family. The overall distribution of His depends, not surprisingly, on the total amount of activated Gly that was added (compare yellow and blue vs. violet bars in Figure 3B). MS–MS analysis of the peaks that correspond to co-oligomers containing one or two His showed that the histidines are interspersed among the glycine residues (Supporting Information for MS–MS of Gly_8His_1 and Gly_6His_2 oligomers). The high complexity of MS–MS spectra from co-oligomers that contain more than two His residues showed the presence of complex mixtures; this precludes the formation of a single-block oligoGly–oligoHis co-oligomer. The choice of different addition programs with the same total Gly content (yellow vs. blue bars in Figure 3B) is expected to influence the distribution of His within the co-oligomer, that is, the distances between His residues within the main-chain. Additional studies are underway to address this issue. To the best of our knowledge, this is the first report of the successful CDI-mediated co-oligomerization that involves His.

In conclusion, mild conditions for efficient oligomerization of histidine have been established by chlorotrimethylsilane activation of the bicyclic urea **3**. Co-oligomers of glutamate–glycine and of histidine–glycine have been prepared through in situ activation of the amino acids by using batch co-oligomerization and pulsed addition protocols, respectively. The products

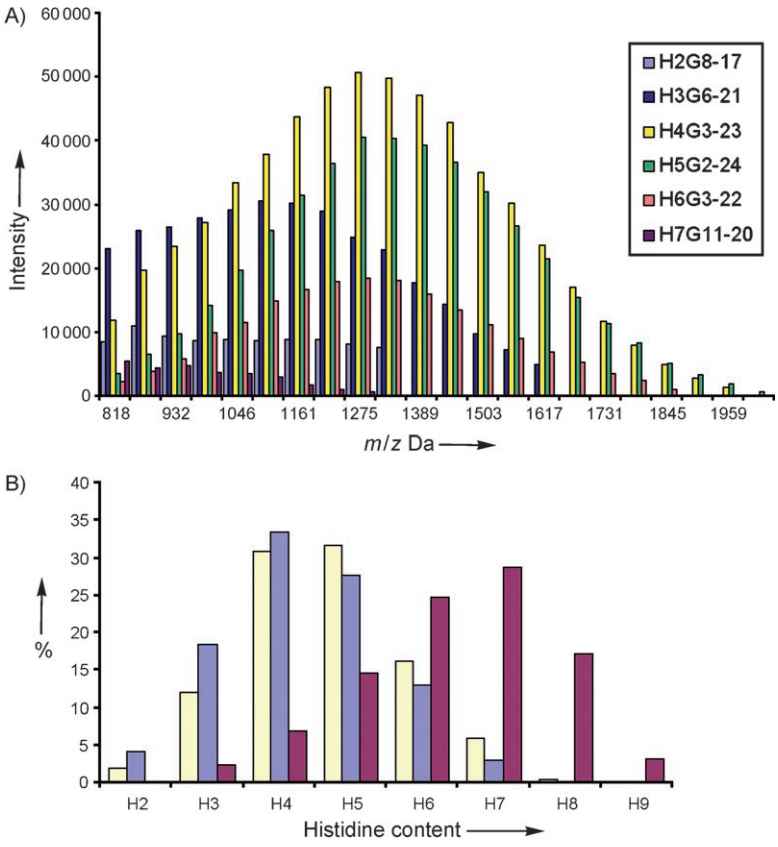

Figure 3. Analysis of a MALDI-TOF mass spectrum of the co-oligomerization of His and Gly. A) addition of 10×0.2 equiv Gly–NCA every 20 min (see SI for MALDI-TOF graphs for the remaining addition modes). B) His_n distribution (% His_2-His_9) within the His–Gly co-oligomers. Blue bars: addition mode as in (A); violet bars: addition of 5×0.2 equiv Gly–NCA every 40 min.; yellow bars: addition of 5×0.4 equiv Gly–NCA every 40 min.

showed even distributions of the amino acids and hitherto unprecedented main-chain lengths, that is up to 21 residues for oligo(Glu–Gly), and up to 29 residues for oligo(His–Gly). The pulsed addition mode allows for some control over the average relative amino acid composition and distribution, and allows, for instance, oligo(His–Gly) populations that bear maximal abundances either of $His_{4-5}Gly_{10-12}$, or of $His_{6-7}Gly_{6-8}$ to be generated. It is noteworthy that, contrary to the outcome under the batch co-oligomerization conditions, the His/Gly ratio that was observed in the most abundant co-oligomers closely reflect the ratio of total amino acids that were added to the reaction mixture.

The combination of an exceptionally simple chemical reaction system (compared to in vitro random mRNA translation, for example) with the ability to select specific amino acids and different addition protocols opens the way to generating large populations of co-oligomers of two or more different amino acids. Because the amino acid composition, the average main-chain length, and the global distribution of catalytically competent side chains are tuneable, they have potential for novel medical applications,[11] and bear features that are well suited for various in vitro selection experiments, such as in catalysis, autocatalysis, molecular recognition (aptamer selection), or enzyme inhibition.[12] Future experiments with NCA-generated mixed co-oligopeptide populations will show if the autocatalytic emergence of prebiotic peptides or even proteins could have been a viable pathway for the evolution of early biotic catalytic activities.

Experimental Section

Oligohistidine: A solution of L-histidine (5 mL, 125 mM in D_2O) was added to solid 1,1′-carbonyldiimidazole (CDI; 247.2 mg, 1.52 mmol) at room temperature. The solution was stirred at room temperature overnight, then a sample (0.5 mL) was diluted in a 615 mM aqueous imidazole solution (0.5 mL D_2O) and heated to 50 °C. TMSCl was added (55 μL, 296 μmol) at room temperature, and the solution was heated at 50 °C. The reaction was monitored by ^{1}H NMR spectroscopy. Final conditions were 60 mM (activated) histidine, 580 mM imidazole (taking into account the contribution of the CDI), and 280 mM TMSCl; the pD at the end of the reaction was 6.7 (at 25 °C: pD = measured pH* + 0.43 ± 0.02).[13] The approximate observed half-conversion time of the urea **3** as a function of the quantity of TMSCl that was added under these conditions was as follows: < 30 min (75 μL); 7 h (55 μL); 17 h (25 μL); > 24 h (10 μL). These conditions were chosen to allow for comparative studies. A simplified procedure consisted of heating the histidine–CDI mixture (0.5 mL) at 50 °C for 1 h, followed by treatment with TMSCl (25 μL), and then by keeping the resulting solution at 50 °C overnight.

Glutamate–glycine co-oligomers: A solution of glycine (500 μL, 120 mM) and L-glutamic acid (500 μL, 120 mM) were added to solid CDI (50 mg, 0.3 mmol) at room temperature. After 5 min, more glycine solution (100 μL, 120 mM) was added, and the mixture was allowed to stand at room temperature. The oligomerization of Gly–Glu was essentially complete after 16 h.

Histidine–glycine co-oligomers: A 120 mM solution of L-histidine in H_2O (500 μL) was added to each of four eppendorf tubes that contained solid CDI (2.5 equiv). The solutions were heated at 50 °C for 1 h, and TMSCl (12 μL, 67 μmol) was added. A separate solution of glycine–NCA was prepared by treating solid CDI (2.5 equiv) with aqueous glycine (120 mM in H_2O), and was then kept at 0 °C. Five minutes after the addition of TMSCl to the activated histidine solution, the Gly–NCA solution was added to the oligomerizing histidine according to the following schedules over the course of 3 h: A) Gly–NCA (100 μL, 120 mM) every 20 min (10 additions, 1000 μL total); B) Gly–NCA (100 μL, 120 mM) every 40 min (5 additions, 500 μL total); C) Gly–NCA (200 μL, 120 mM) every 40 min (5 additions, 1000 μL total); D) no Gly–NCA added (control). The solutions were heated for an additional hour, then stored at 4 °C. Samples were prepared for ESI-MS by centrifugation at 10000 rpm (10 min), and then by diluting of some of the supernatant (20 μL) into MeOH/H_2O 1:1 (1 mL) that contained formic acid (0.5%, *v/v*).

Acknowledgements

We are grateful to Prof. Pier Luigi Luisi, Dipartimento di Biologia, Università degli Studi di Roma Tre, Italy, for the inspiration and collaboration, Dr. Silvia Terenzi, at the time at EPFL Lausanne, Switzerland, for carrying out MALDI-TOF mass spectra of oligo-

Glu, oligo(Gly–Glu) and oligoHis. Financial support from the European Union (Contract No. LSHB-CT-2004-503467) and the COST Action D27 "Prebiotic Chemistry and Early Evolution" is gratefully acknowledged.

Keywords: chemical evolution · copolymerization · oligomerization · peptides · prebiotic chemistry

[1] P. L. Luisi, *Chem. Biodiversity* **2007**, *4*, 603–621.

[2] a) H. G. Elias, *An Introduction to Polymer Science*, VCH, Weinheim **1997**; b) P. L. Luisi, *The Emergence of Life. From Chemical Origins to Synthetic Biology*, Cambridge University Press, **2006**.

[3] More or less "ideal random copolymerizations", that is, compositionally faithful monomer incorporation into copolypeptides from nonaqueous solutions of previously *N*-carboxy-activated amino acids were reported by a) D. Teitelbaum, A. Meshorer, T. Hirshfeld, R. Arnon, M. Sela, *Eur. J. Immunol.* **1971**, *1*, 242–248; b) W. Sederel, S. Deshmane, T. Hayashi, J. M. Anderson, *Biopolymers* **1978**, *17*, 2835–2849; c) S. B. Mitra, N. K. Patel, J. M. Anderson, *Int. J. Biol. Macromol.* **1979**, *1*, 55–60; d) Y. Tajima, J. M. Anderson, P. H. Geil, *Int. J. Biol. Macromol.* **1980**, *2*, 186–192; e) H. R. Kricheldorf, W. E. Hull, D. Müller, *Macromolecules* **1985**, *18*, 2135–2140; f) H. R. Kricheldorf, D. Müller, W. E. Hull, *Int. J. Biol. Macromol.* **1986**, *8*, 20–26; g) Y. Iizuka, T. Endo, M. Oya, *Bull. Chem. Soc. Jpn.* **1991**, *64*, 1336–1341; h) Y. Iizuka, T. Endo, M. Oya, *Bull. Chem. Soc. Jpn.* **1993**, *66*, 1269–1272; i) Y. Iizuka, K. Wakamatsu, H. Mitomo, M. Oya, M. Iwatsuki, T. Hayashi, *Polym. J.* **1993**, *25*, 659–669; j) C. Uchida, Y. Iizuka, E. Ohta, K. Wakamatsu, M. Oya, *Bull. Chem. Soc. Jpn.* **1996**, *69*, 791–796; k) T. J. Deming, *Adv. Polym. Sci.* **2006**, *202*, 1–18; l) H. R. Kricheldorf in *Models of Biopolymers by Ring-Opening Polymerization* (Ed. S. Penczek), CRC, Boca Raton, **1990**.

[4] a) H. R. Kricheldorf, *Angew. Chem.* **2006**, *118*, 5884–5917; *Angew. Chem. Int. Ed.* **2006**, *45*, 5752–5784; b) H. R. Kricheldorf, C. von Lossow, G. Schwarz. *Macromol. Chem. Phys.* **2005**, *206*, 282–290; c) H. R. Kricheldorf, C. von Lossow, G. Schwarz, *Macromol. Chem. Phys.* **2004**, *205*, 918–924; d) J. P. Ferris, A. R. Hill, R. H. Liu, L. E. Orgel, *Nature* **1996**, *381*, 59–61.

[5] a) I. Rubinstein, R. Eliash, G. Bolbach, I. Weissbuch, M. Lahav, *Angew. Chem.* **2007**, *119*, 3784–3787; *Angew. Chem. Int. Ed.* **2007**, *46*, 3710–3713; b) I. Weissbuch, L. Leiserowitz, M. Lahav, *Top. Curr. Chem.* **2005**, *259*, 123–165; c) J. G. Nery, G. Bolbach, I. Weissbuch, M. Lahav, *Chem. Eur. J.* **2005**, *11*, 3039–3048; d) J. G. Nery, G. Bolbach, I. Weissbuch, M. Lahav, *Angew. Chem.* **2003**, *115*, 2207–2211; *Angew. Chem. Int. Ed.* **2003**, *42*, 2157–2161.

[6] a) G. Danger, L. Boiteau, H. Cottet, R. Pascal, *J. Am. Chem. Soc.* **2006**, *128*, 7412–7413; b) R. Pascal, L. Boiteau, A. Commeyras, *Top. Curr. Chem.* **2005**, *259*, 69–122; c) A. Commeyras, L. Boiteau, O. Vandenabeele-Trambouze, F. Selsis, *Lectures in Astrobiology Vol. 1, Part 2* (Eds.: M. Gargaud, B. Barbier, H. Martin, J. Reisse), Springer, Berlin, **2004**, p. 517–542; d) J. Taillades, H. Cottet, L. Garrel, I. Beuzelin, L. Boiteau, H. Choukroun, A. Commeyras, *J. Mol. Evol.* **1999**, *48*, 638.

[7] a) Y. Li, Y. Zhao, S. Hatfield, R. Wan, Q. Zhu, X. Li, M. McMills, Y. Ma, J. Li, K. L. Brown, C. He, F. Liu, X. Chen, *Bioorg. Med. Chem.* **2000**, *8*, 2675–2680; b) Q. Zen, Q. Yin, Y. Zhao, *Bioorg. Med. Chem.* **2005**, *13*, 2679–2689.

[8] a) K.-J. Wang, N. Yao, C. Li, *Origins Life Evol. Biospheres* **2005**, *35*, 313–332; b) M. Blocher, D. Liu, P. L. Luisi, *Macromolecules* **2000**, *33*, 5787–5796; c) A. R. Hill, L. E. Orgel, *Origins Life Evol. Biospheres* **1996**, *26*, 539–545.

[9] a) A. Patchornik, A. Berger, E. Katchalski, *J. Am. Chem. Soc.* **1957**, *79*, 5227–5230; b) R. S. Dewey, E. F. Schoenewaldt, H. Joshua, W. J. Palaveda, Jr., H. Schwam, H. Barkemeyer, B. H. Arison, D. F. Veber, R. G. Denkewalter, R. Hirschmann, *J. Am. Chem. Soc.* **1968**, *90*, 3254–3255; c) K. W. Ehler, E. Girard, L. Orgel, *Biochim. Biophys. Acta Protein Struct.* **1977**, *491*, 253–264.

[10] A. Brack, *Origins Life Evol. Biospheres* **1987**, *17*, 367–379.

[11] a) M. Sela, *Polym. Adv. Technol.* **2002**, *13*, 859–862; b) M. Fridkis-Hareli, E. F. Rosloniec, L. Fugger, J. L. Strominger, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12528–12531; c) T. J. Deming, *Prog. Polym. Sci.* **2007**, *32*, 858–875.

[12] For recent reviews on the application of combinatorial peptide libraries, see for example: a) J. D. Revell, H. Wennemers, *Curr. Opin. Chem. Biol.* **2007**, *11*, 269–278; b) K. Gazarian, *Front. Drug Des. Discovery* **2005**, *1*, 29–67; c) M. Bogyo, *Methods Enzymol.* **2005**, *399*, 609–622.

[13] a) A. K. Covington, M. Paabo, R. A. Robinson, R. G. Bates, *Anal. Chem.* **1968**, *40*, 700–706; b) A. Krezel, W. Bal, *J. Inorg. Biochem.* **2004**, *98*, 161–166.

Received: November 26, 2007
Published online on February 25, 2008

