Chemical Protein Synthesis

DOI: 10.1002/ange.201103237

Total Chemical Synthesis of Biologically Active Vascular Endothelial Growth Factor**

*Kalyaneswar Mandal and Stephen B. H. Kent**

Vascular endothelial growth factor (VEGF), also known as VEGF-A, is the principal endothelial-cell-specific mitogen that induces embryonic and somatic angiogenesis.[1–4] VEGF also plays an important role in tumor angiogenesis and in proliferative retinopathy.[5] In animal models of pathological conditions, it has been shown that over-expression of VEGF stimulates the generation of new blood vessels, and thus accelerates tumor growth and the progression of the neovascular form of age-related macular degeneration.[2,6,7]

VEGF is a homodimer and exists in multiple isoforms, each containing two identical polypeptide chains derived from the alternative splicing of exons in a single mRNA.[8] VEGF165 is the most abundant naturally occurring form.[9] All the isoforms of VEGF share a common N-terminal domain of approximately 115 residues;[10] it is this domain that interacts with the VEGFR1 and VEGFR2 receptors and causes the biological response. Targeting the N-terminal domain of VEGF with specific inhibitors has been shown to have broad therapeutic potential for a variety of medical conditions. Monoclonal antibody therapeutics that target VEGF are already approved for cancer and age-related macular degeneration.[11–13]

VEGF is a member of the cystine knot growth factor protein family, in which two disulfide bridges connect backbone segments of the protein to form a ring structure through which a third disulfide bond interpenetrates.[14] VEGF itself is a covalent homodimeric protein with three intramolecular disulfides in each polypeptide chain, and with two interchain disulfide bonds covalently joining the two identical monomers.[14] The resulting VEGF molecule has a twofold axis of symmetry and thus has two receptor-binding regions.[15,16] It has previously been shown that a truncated variant, VEGF (8–109), binds to the VEGFR2 receptor with wild-type affinity and stimulates a full biological response.[15] VEGF (8–109) has been the subject of extensive structural and biophysical studies including the determination of its X-ray structure.[14]

Robust, practical synthetic access to biologically active VEGF will enable novel approaches to the development of new anti-angiogenic therapeutics. Herein, we report a highly optimized total synthesis of human VEGF(8–109) by a one-pot, three-segment native chemical ligation (NCL) strategy.[17–19] Folding of the synthetic polypeptide chain with concomitant disulfide bond formation gave a VEGF protein molecule with full biological activity. The structure of the covalent homodimeric protein molecule was confirmed by high-resolution X-ray crystallography.

The amino acid sequence of the VEGF(8–109) target polypeptide chain contains eight cysteine residues and is shown in Figure 1a. In our initial synthetic design, we explored a potentially simple approach that made use of a

a) GQNHHEVVKF^{10} MDVYQRS***YC***H^{20} PIETLVDIFQ^{30} EYPDEIEYIF^{40}
KPSCVPLM***RC***^{50} GGCCNDEGLE^{60} CVPTEESNIT^{70} MQIMRIKPHQ^{80}
GQHIGEMSFL^{90} QHNKCECRPK^{100} KD^{102}

b)

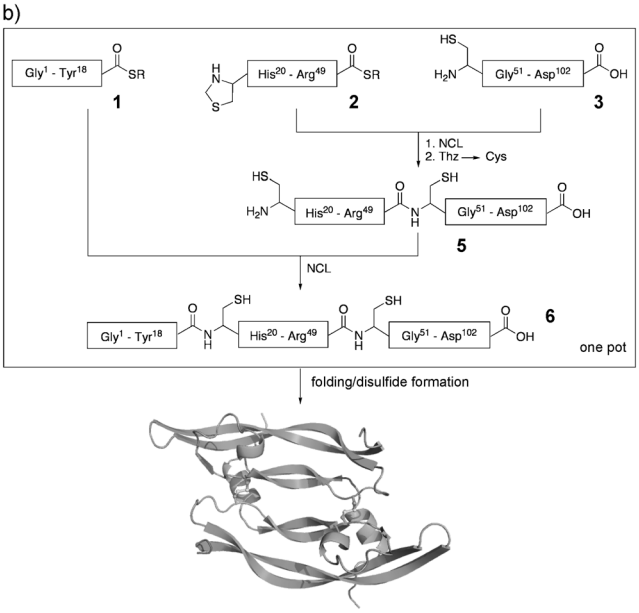

Figure 1. Total chemical synthesis of VEGF. a) Target amino acid sequence. b) Synthetic strategy for the total chemical synthesis of VEGF. One-pot sequential NCLs[18] of three peptide segments were used to prepare the 102 amino acid residue polypeptide chain; subsequent folding/disulfide bond formation gave the desired covalent dimer VEGF protein molecule. NCL = native chemical ligation, R = $CH_2CH_2SO_3H$, Thz = 1,3-thiazolidine-4-R-carboxo.

[*] Dr. K. Mandal, Prof. Dr. S. B. H. Kent
Department of Biochemistry and Molecular Biology
Institute for Biophysical Dynamics and
Department of Chemistry, The University of Chicago
Chicago, IL 60637 (USA)
E-mail: skent@uchicago.edu

[**] This research was supported by the Office of Science (BER), U.S. Department of Energy (grant no. DE-FG02 07ER64501 to S.B.H.K.) and by the National Institutes of Health (grant no. R01 GM075993 to S.B.H.K.). Use of NE-CAT beamline 24-ID at the Advanced Photon Source is supported by award RR-15301 from the National Center for Research Resources at the National Institutes of Health. Use of the Advanced Photon Source is supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under contract no. DE-AC02-06CH11357. We thank Dr. Kay Perry, Dr. Kanagalaghatta Rajashankar, and Somnath Mukherjee for useful discussions.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201103237.

single NCL[17] (at Arg49–Cys50) of two unprotected peptide segments of almost equal size for the assembly of the 102-residue polypeptide chain. However, unsatisfactory purity of the 1–49 thioester segment prompted us to use the three-segment NCL strategy shown in Figure 1b for the total synthesis of VEGF. In the optimized synthetic design, the N-terminal half of the full-length 102 amino acid polypeptide chain was divided into two synthetic peptide segments. The N-terminal cysteine residue of the middle segment was introduced as a 1,3-thiazolidine-4-R-carboxo moiety in order to prevent the reaction of the N-terminal Cys with the thioester in the same peptide.[18] To avoid excessive product losses from multiple intermediate purification steps, we used one-pot sequential NCLs, together with a single final purification of the full-length 102-residue ligation product.[18] The purified full-length polypeptide was folded with the concomitant formation of disulfides to give the homodimeric synthetic VEGF protein molecule.

The three synthetic peptides (peptide-thioester **1**, thiazolidine (Thz)-peptide-thioester **2**, and Cys-peptide **3**) were prepared by *tert*-butoxycarbonyl (Boc) solid-phase peptide synthesis (SPPS) using the in situ neutralization protocol[20] as detailed in the Supporting Information. Based on the synthetic strategy shown in Figure 1b, the C-terminal peptide segment Cys50–Asp102 (**3**, 1.7 mM) was reacted with Thz19–Arg49-αCOSR (**2**, 1.5 mM) under standard NCL conditions (aqueous guanidine hydrochloride (Gu·HCl, 6 M), Na_2HPO_4 (0.1 M), at pH 7.0, 4-carboxymethylthiophenol (MPAA, 100 mM),[21] tris(2-carboxyethyl)phosphine (TCEP·HCl, 50 mM)) to give the ligation product Thz19–Asp102 (**4**). Without purification of **4**, Thz19 was then converted to Cys19 by addition of methoxylamine·HCl (60 mM) at pH 4.0.[22] After the conversion of Thz to Cys was complete (16 h, product **5**), the peptide-thioester segment (Gly1–Tyr18-αCOSR, **1**, 2 mM) was added to the same reaction mixture and the pH value was adjusted to 6.8; within 2 h, NCL gave near-quantitative conversion to the target polypeptide **6**. Analytical data for all steps in the synthesis of the full-length polypeptide are shown in Figure 2. After HPLC purification, the target polypeptide **6** was obtained in 48% overall yield based on the limiting peptide segment **2**. The reverse-phase HPLC profile of the 102-residue polypeptide is shown in Figure 3a. The unfolded monomeric full-length polypeptide **6** was characterized by HPLC electrospray ionization mass spectrometry (LC-MS; Figure 3b) and had the expected mass: (11932.2 ± 0.7) Da (mean of the four most abundant charge states); calculated mass: 11932.5 Da (average isotope composition).

In its folded form, VEGF(8–109) exists as a covalent homodimer of two identical 102-residue polypeptide chains, with six intrachain disulfide bonds and two interchain disulfides connecting the two monomer polypeptides.[14] These disulfide bridges are crucial for the correct folding and structural integrity of functional VEGF.[23] Reduced polypeptide **6** was folded at pH 8.4 in aqueous Gu·HCl (0.15 M) containing a glutathione-reduced (2 mM)/glutathione-oxidized (0.4 mM) redox couple.[24] The completion of the folding was revealed by the near-quantitative formation (after 5 days) of a product that eluted earlier than the reduced (i.e., unfolded) polypeptide monomer in analytical reverse-phase HPLC (Figure 3c). Such a shift to an earlier retention time is typically observed upon folding of disulfide-containing globular proteins, because of the burial of the hydrophobic residues in the core of the folded protein molecule. The formation of eight disulfide bonds in the synthetic VEGF protein molecule was confirmed by measurement of the mass of this early-eluting peak (see next paragraph). Reverse-phase HPLC purification gave synthetic VEGF in 45% yield based on the amount of **6** that was used.

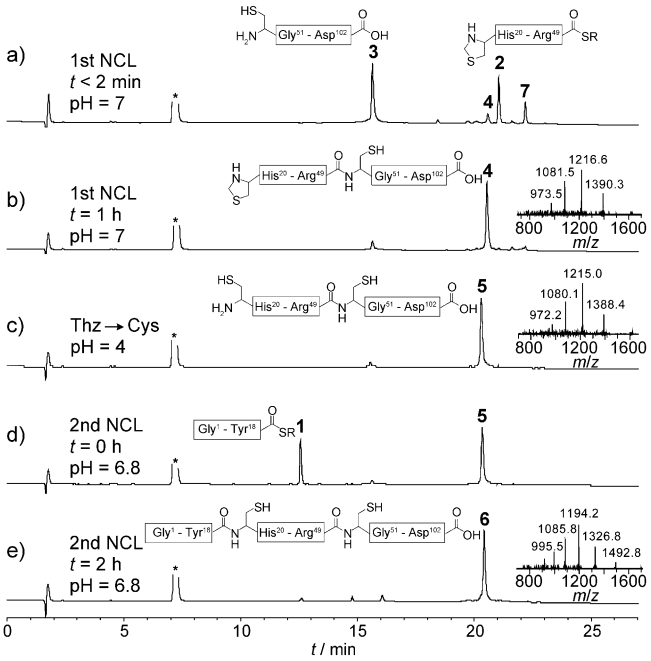

Figure 2. Analytical LC–MS data for the total chemical synthesis of VEGF. Analytical HPLC profiles ($\lambda = 214$ nm) are shown, together with online ESI MS data (inset) corresponding to each major product. a) NCL between Thz19–Arg49-αCOSR (**2**, 15 mg, 3.37 μmol, 1.53 mM) and Cys50–Asp102-COOH (**3**, 22 mg, 3.66 μmol, 1.66 mM) at $t < 2$ min. **7** is the MPAA-exchanged thioester intermediate of **2**. b) NCL reaction after 1 h. Thz19–Asp102-COOH (**4**) is the ligation product. c) Crude reaction mixture after Thz19 to Cys19 conversion using methoxylamine·HCl (60 mM) at pH 4. Formation of the desired product Cys19–Asp102-COOH (**5**) was confirmed by a mass decrease of 12 Da. d) One-pot NCL of Gly1–Tyr18-αCOSR (**1**, 10.4 mg, 4.39 μmol, 2 mM) and **5** at $t = 0$ min. e) Crude reaction mixture after 2 h at pH 6.8. Gly1–Asp102-COOH (**6**) is the ligation product. The overall yield of **6** after purification was 19.3 mg, 1.6 μmol, 48%, based on the limiting peptide segment **2**. Analytical HPLC was performed using a linear gradient (10–54%) of buffer B in buffer A over 22 min (buffer A = 0.1% trifluoroacetic acid (TFA) in water; buffer B = 0.08% TFA in acetonitrile) on a C-3 column (4.6 × 150 mm, Agilent) at 40 °C (flow rate = 1 mL min^{-1}). The asterisk (*) corresponds to MPAA.

Attempts to determine the mass of the folded synthetic VEGF protein by LC–MS were unsuccessful. Under the conditions of direct infusion into the mass spectrometer from the analytical HPLC column, no ion current was observed for the UV-detected peak of the purified synthetic product. Poor ionization under electrospray-injection conditions is typical of tightly folded disulfide-cross-linked globular proteins dissolved in water/acetonitrile/trifluoroacetic acid (TFA,

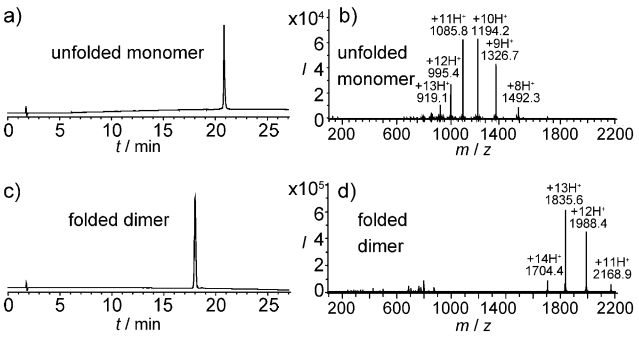

Figure 3. HPLC and MS characterization of synthetic VEGF. a) Analytical HPLC profile ($\lambda = 214$ nm) and b) LC–MS of the purified unfolded 102 amino acid polypeptide (Gly1–Asp102-COOH); observed mass: (11932.2 ± 0.7) Da (mean of the four most abundant charge states). c) Analytical HPLC profile ($\lambda = 214$ nm) and d) direct infusion ESI MS of the purified synthetic VEGF protein; observed mass: (23848.7 ± 1.2) Da (mean of the three most abundant charge states). The retention time of the folded VEGF was 2.8 min less than that of the reduced polypeptide. A decrease of 15.7 Da from twice the monomer mass confirmed the formation of eight disulfides. Note the narrow distribution of the charge states in the folded VEGF dimer compared with the unfolded monomer. Analytical HPLC was performed using a linear gradient (10–54%) of buffer B in buffer A over 22 min (buffer A = 0.1% TFA in water; buffer B = 0.08% TFA in acetonitrile) on a C-3 column (4.6 × 150 mm, Agilent) at 40 °C (flow rate = 1 mL min^{-1}).

0.1%).[25] The purified synthetic protein was dissolved in a solvent (1% acetic acid in methanol/water (80:20 v/v)) that is more conducive to electrospray ionization (ESI) of protein molecules,[26] and was characterized by direct infusion ESI mass spectrometry (Figure 3d). The synthetic protein had an observed mass of (23848.7 ± 1.2) Da (mean of the three most abundant charge states), calculated mass: 23849.1 Da (average isotope composition). The observed decrease of 15.7 Da, compared with twice the mass of the 102-residue polypeptide chain, was consistent with the formation of eight disulfide bonds and the consequent loss of 16 protons in the synthetic VEGF protein molecule.

VEGF contains 204 amino acid residues and has a total of 34 protonation sites (His, Lys, Arg side chains, plus 2 × N-terminal amino groups). Yet, the positive-ion ESI mass spectrum of the synthetic protein exhibited only a narrow distribution of charge states, centered around $+13H^+$ with protonation states ranging from 11^+ to 14^+ (Figure 3d). In contrast, the reduced and unfolded synthetic VEGF polypeptide chain with 102 amino acid residues contained a total of 17 protonation sites (His, Lys, Arg side chains, plus an N-terminal amino group) and its positive-ion ESI mass spectrum was dominated by highly charged ions with a maximum around $+11H^+$ with protonation states ranging from 8^+ to 13^+ in the LC–MS spectrum (Figure 3b). Such a dramatic effect on the charge state distribution in a folded versus an unfolded protein molecule has been reported previously and is still a matter of debate. Several different explanations have been offered by investigators.[25–31] The most plausible rationale for the low number of added protons in the folded protein molecule under ESI conditions was suggested by Fenn.[31] He noted that it is "the surface of the charged droplet that determines the nature and amount of the charge on a departing ion… in its compact configuration, a molecule has a smaller surface area in contact with the solution than when it is unfolded. Consequently, less work may be required to remove it from the droplet, so that it could lift off with fewer charges than when it is unfolded."

The circular dichroism spectrum of an aqueous solution of the folded VEGF revealed the presence of β-sheet and α-helix secondary structural elements, as shown in the Supporting Information (Figure S5).

The structure of the synthetic VEGF protein molecule was determined by X-ray crystallography. Synthetic VEGF was crystallized at a protein concentration of 2.5 mg mL^{-1} from NH_4OAc (0.2 M), 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)-1,3-propanediol (BIS-TRIS, 0.1 M, pH 5.5), and (±)-2-methyl-2,4-pentanediol (45% v/v). Synchrotron radiation diffraction data were collected to 1.85 Å resolution from a single crystal at the Advanced Photon Source, Argonne National Laboratory. The synthetic VEGF structure was solved in the monoclinic space group C2 by molecular replacement using the reported X-ray structure (protein data bank (PDB) accession code 2VPF) as a search model. The synthetic VEGF structure was refined to a crystallographic R-factor of 18.0% (R-free 22.3%) using the program Phenix.[32] X-ray data collection and the refinement statistics are summarized in Table S1 in the Supporting Information.

The X-ray structure of the chemically synthesized VEGF protein reported here is identical, within experimental uncertainty, to the previously reported X-ray structure of recombinant VEGF(8–109):[14] 96 Cα atoms of the 102-residue monomer unit can be superimposed with a root-mean-square deviation (rmsd) of only 0.7 Å (Figure 4c). The X-ray structure of the chemically synthesized VEGF protein consists of an N-terminal α helix that folds on top of the second monomer, followed by an antiparallel four-stranded β sheet forming the central part of the molecule. In addition, there is a short α-helical segment followed by a loop, and the second β strand is positioned between the first and the third β strands. Figure 4b shows a representative $2F_o - F_c$ electron density map contoured at 1 σ encompassing the interchain disulfide interface. The asymmetric unit contains three VEGF molecules: that is, it contains six crystallographically independent copies of the folded VEGF polypeptide chain in the form of three covalent homodimers. Figure 4a shows a cartoon representation of the synthetic protein structure. Superposition of the six independent copies of VEGF monomers reveals very similar structures in the core region as expected, but significant deviation is observed in one of the loop regions (Met71–Ser88) which is a part of the receptor-binding site of the VEGF protein molecule (see Figure S6 in the Supporting Information). Different copies of the molecule represent different snapshots of the loop movement, with a largest deviation of 6.7 Å at His79.

The synthetic VEGF protein had full mitogenic activity as demonstrated by the human umbilical vein endothelial cell (HUVEC) proliferation bioassay (Figure 5). The effective dose (ED_{50}) of 4.6 ng mL^{-1} is within the typical ED_{50} range observed for human VEGF165 (i.e. 1–6 ng mL^{-1}).

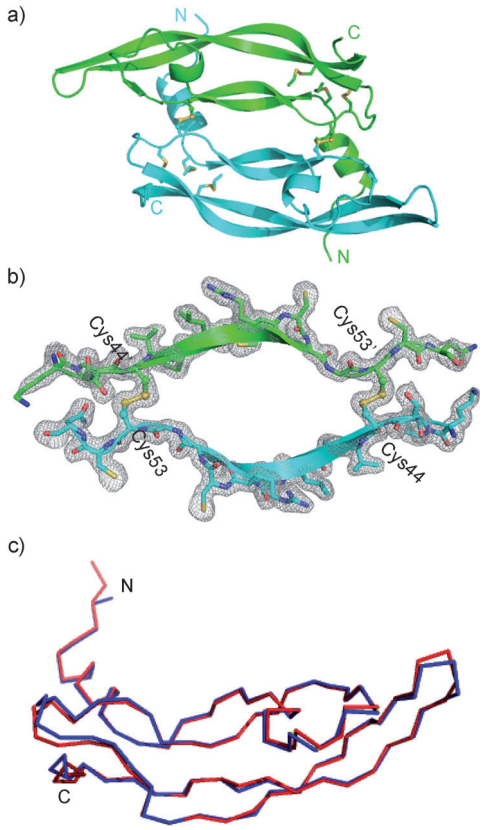

Figure 4. X-ray structure of chemically synthesized VEGF. a) Cartoon representation of the experimentally determined structure of the synthetic protein molecule. b) Sigma A-weighted $2F_o-F_c$ electron density map of VEGF contoured at 1σ encompassing the two disulfide bridges between the two monomer units. c) Superposition of the polypeptide chain backbones from the crystal structure of the chemically synthesized VEGF reported here (red) and the reported crystal structure (PDB 2VPF) of the recombinant VEGF (8–109) monomer (blue).

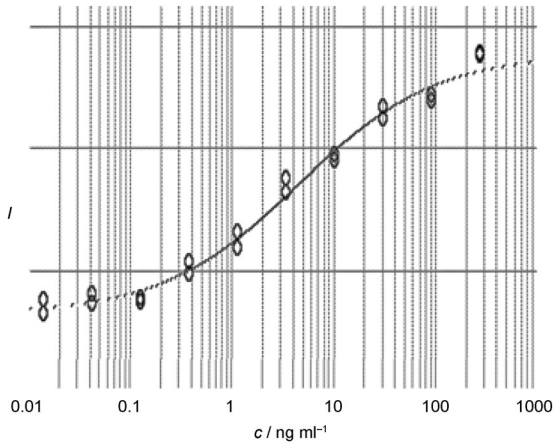

Figure 5. HUVEC proliferation assay: ED_{50} observed for the chemically synthesized VEGF (8–109) was 4.6 ng mL^{-1}; typical ED_{50} range for human VEGF165 is 1–6 ng mL^{-1}.

We have developed a robust and reproducible total chemical synthesis of the homodimeric VEGF protein molecule. The synthesis reported here is one of only a handful of reported total syntheses of protein molecules of this size.[33] With 204 amino acid residues, this is the largest protein molecule prepared by one-pot native chemical ligation; an overall yield of approximately 22 % was obtained based on the peptide building blocks used. The folding of proteins that contain large numbers of disulfides is not always straightforward; the near-quantitative formation of eight native disulfide bonds in the folded synthetic VEGF is notable. Precise mass measurement was consistent with the expected covalent structure of the synthetic protein. Furthermore, synthetic VEGF showed full biological activity and was characterized by high-resolution X-ray crystallography. Synthetic access to this important protein molecule, which was previously available only from natural and recombinant sources, is an essential first step in novel approaches to the development of improved anti-angiogenic agents.

Received: May 12, 2011
Published online: July 8, 2011

Keywords: chemical protein synthesis · growth factors · mass spectrometry · one-pot ligation · VEGF

[1] N. Ferrara, *Breast Cancer Res. Treat.* **1995**, *36*, 127–137.
[2] J. Folkman, *Nat. Med.* **1995**, *1*, 27–31.
[3] H. F. Dvorak, L. F. Brown, M. Detmar, A. M. Dvorak, *Am. J. Pathol.* **1995**, *146*, 1029–1039.
[4] G. Neufeld, S. Tessler, H. Gitay-Goren, T. Cohen, B. Z. Levi, *Prog. Growth Factor Res.* **1994**, *5*, 89–97.
[5] L. P. Aiello et al., *N. Engl. J. Med.* **1994**, *331*, 1480–1487.
[6] P. F. Lopez, B. D. Sippy, H. M. Lambert, A. B. Thach, D. R. Hinton, *Invest. Ophthalmol. Visual Sci.* **1996**, *37*, 855–868.
[7] A. Kvanta, P. V. Algvere, L. Berglin, S. Seregard, *Invest. Ophthalmol. Visual Sci.* **1996**, *37*, 1929–1934.
[8] M. Kowanetz, N. Ferrara, *Clin. Cancer Res.* **2006**, *12*, 5018–5022.
[9] N. Vaisman, D. Gospodarowicz, G. Neufeld, *J. Biol. Chem.* **1990**, *265*, 19461–19466.
[10] K. A. Houck, N. Ferrara, J. Winer, G. Cachianes, B. Li, D. W. Leung, *Mol. Endocrinol.* **1991**, *5*, 1806–1814.
[11] L. G. Presta, H. Chen, S. J. O'Connor et al., *Cancer Res.* **1997**, *57*, 4593–4599.
[12] N. Ferrara, *Endocr. Rev.* **2004**, *25*, 581–611.
[13] N. Ferrara, L. Damico, N. Shams, H. Lowman, R. Kim, *Retina* **2006**, *26*, 859–870.
[14] Y. A. Muller, H. W. Christinger, B. A. Keyt, A. M. De Vos, *Structure* **1997**, *5*, 1325–1338.
[15] Y. A. Muller, B. Li, H. W. Christinger, J. A. Wells, B. C. Cunningham, A. M. De Vos AM, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 7192–7197.
[16] C. Wiesmann, G. Fuh, H. W. Christinger, C. Eigenbrot, J. A. Wells, A. M. De Vos, *Cell* **1997**, *91*, 695–704.
[17] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776–779.
[18] D. Bang, S. B. H. Kent, *Angew. Chem.* **2004**, *116*, 2588–2592; *Angew. Chem. Int. Ed.* **2004**, *43*, 2534–2538.
[19] S. B. H. Kent, *Chem. Soc. Rev.* **2009**, *38*, 338–351.
[20] M. Schnölzer, P. Alewood, A. Jones, D. Alewood, S. B. H. Kent, *Int. J. Pept. Res. Ther.* **2007**, *13*, 31–44.
[21] E. C. B. Johnson, S. B. H. Kent, *J. Am. Chem. Soc.* **2006**, *128*, 7140–7141.
[22] M. Villain, J. Vizzavona, K. Rose, *Chem. Biol.* **2001**, *8*, 673–679.
[23] Y. A. Muller, C. Heiring, R. Misselwitz, K. Welfle, H. Welfle, *J. Biol. Chem.* **2002**, *277*, 43410–43416.

[24] A. H. Zisch, M. P. Lutolf, M. Ehrbar, G. P. Raeber, S. C. Rizzi, N. Davies, H. Schmökel, D. Bezuidenhout, V. Djonov, P. Zilla, J. A. Hubbell, *FASEB J.* **2003**, *17*, 2260 – 2262.
[25] J. A. Loo, C. G. Edmonds, H. R. Udseth, R. D. Smith, *Anal. Chem.* **1990**, *62*, 693 – 698.
[26] S. K. Chowdhury, V. Katta, B. T. Chait, *J. Am. Chem. Soc.* **1990**, *112*, 9012 – 9013.
[27] V. Katta, B. T. Chait BT, *J. Am. Chem. Soc.* **1991**, *113*, 8534 – 8535.
[28] R. Grandori, *J. Mass Spectrom.* **2003**, *38*, 11 – 15.
[29] H. Prakash, S. Mazumdar, *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1409 – 1421.
[30] L. Konermann, D. A. Simmons, *Mass Spectrom. Rev.* **2003**, *22*, 1 – 26.
[31] J. B. Fenn, *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 524 – 535.
[32] P. D. Adams et al., *Acta Crystallogr. Sect. D* **2010**, *66*, 213 – 221.
[33] K. S. A. Kumar, S. N. Bavikar, L. Spasser, T. Moyal, S. Ohayon, A. Brik, *Angew. Chem.* **2011**, *123*, 6261 – 6265; *Angew. Chem. Int. Ed.* **2011**, *50*, 6137 – 6141.

