Supporting Information

Total chemical synthesis of biologically active vascular endothelial growth factor

Kalyaneswar Mandal and Stephen B. H. Kent*

Department of Chemistry, Department of Biochemistry & Molecular Biology, Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois, 60637, USA.

Table of Contents

1. General Methods

Reagents. 2-(1*H*-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) and Nα-Boc protected amino acids (Peptide Institute, Osaka) were obtained from Peptides International (Louisville, Kentucky). Side-chain protecting groups used were Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH_3Bzl), Glu(OcHex), His(Dnp), Lys(2Cl-Z), Ser(Bzl), Thr(Bzl), Tyr(2Br-Z). Boc-His(Dnp) was purchased from Chem-Impex. Aminomethyl-resin (1.0 mmol/gram) was prepared from Biobeads S-X1 (BioRad, California).[1] Boc-Ala-OCH_2-phenylacetic acid and Boc-Asp(OcHex)-OCH_2-phenylacetic acid were purchased from NeoMPS, Strasbourg, France. *N*,*N*-Diisopropylethylamine (DIEA) was obtained from Applied Biosystems. *N*,*N*-Dimethylformamide (DMF), dichloromethane, diethyl ether, HPLC-grade acetonitrile, and guanidine hydrochloride were purchased from Fisher. Trifluoroacetic acid (TFA) was obtained from Halocarbon Products (New Jersey). HF was purchased from Matheson. All other reagents were purchased from Sigma-Aldrich and were of the purest grade available.

Reverse phase HPLC and LC-MS analysis

Analytical reverse phase HPLC reported in this work was performed on an Agilent C-3 (3.5 μm, 300 Å) 4.6 x150 mm silica column at a flow rate of 1 mL/min or an in-house packed C-4 (Microsorb, 3 μm, 300 Å), 2.1 x 50 mm silica column at a flow rate of 0.5 mL/min using a linear gradient of 10-54% of buffer B in buffer A over 22 min or 5-65% of buffer B in buffer A over 60 min at 40 °C (buffer A= 0.1% TFA in H_2O; buffer B = 0.08% TFA in acetonitrile). The absorbance of the column eluate was monitored at 214 nm. The peptide masses were measured by on-line LC-MS using an Agilent 1100 LC/MSD ion trap. Calculated masses were based on average isotope composition, unless otherwise stated.

Preparative reverse phase HPLC of crude peptides was performed with an Agilent 1100 prep system on in-house packed C-18 or C-4 (10 μm, 300 Å), 10 x 250 mm columns at 40 ^oC using an appropriate shallow gradient of increasing concentration of buffer B in buffer A at a flow rate of 5 mL/min. Fractions containing the purified target peptide were identified by MALDI time-of-flight MS, and aliquots from each selected fraction were combined and checked by LC-MS. Selected fractions were then combined and lyophilized.

2. Peptide synthesis/purification/LCMS characterization. Peptide segments were synthesized using manual 'in situ neutralization' Boc chemistry protocols for stepwise SPPS.[2] The C-terminal peptide segment, [Cys50-Asp102]-COOH was synthesized on Boc-Asp(OcHex)-OCH_2-Pam-resin[1] at a 0.3

mmol scale. The peptide thioester segment [Gly1-Tyr18]-$^{\alpha}COSR_1$ was synthesized on trityl-SCH_2CH_2CO-Ala-OCH_2-Pam-resin[3] at a 0.3 mmol scale. The peptide thioester segment [Thz19-Arg49]-$^{\alpha}COSR_2$ was synthesized at a 0.4 mmol scale, on trityl-SCH_2CH_2CO-$(Arg)_4$-Ala-OCH_2-Pam-resin; the additional arginine residues were added to the thioester leaving group in order to improve the solubility and handling properties of this peptide. Prior to HF deprotection and cleavage, the Dnp groups of the residues $His^{79,83,92}$ were removed from the resin-bound C-terminal peptide segment by treatment with {20% v/v 2-mercaptoethanol plus 10% v/v DIEA in DMF} for 2 x 10 min. After removal of the N-$^{\alpha}$Boc group, peptides were cleaved from the resin and simultaneously deprotected by treatment at 0 °C for 1 h with anhydrous HF containing 5%v/v p-cresol as scavenger. After removal of HF by evaporation under reduced pressure, each crude peptide was precipitated and washed with diethyl ether, then dissolved in 50% aqueous acetonitrile containing 0.1% TFA and lyophilized. The analytical HPLC chromatograms of the crude synthetic peptide segments are shown in Figure S1.

For the peptide–thioester segments [Gly1-Tyr18]-$^{\alpha}COSR_1$ and [Thz19-Arg49]-$^{\alpha}COSR_2$, the Dnp groups were removed from $His^{4,5,20}$ by treating the crude lyophilized peptide segments with 200 mM sodium 2-mercaptoethanesulfonate (MESNa) at pH 7 in 0.1 M Na_2HPO_4 containing 6 M Gu.HCl; this resulted in the corresponding transthioesterified (i.e. MESNa exchanged) products. Analytical HPLC traces of the Dnp group removal and transthioesterification are shown in Figure S2.

Preparative HPLC purification gave the following amounts of each peptide segment: 61 mg (10.1 μmol, 3.3% yield from 0.3 mmol peptide resin) of the C-terminal peptide [Cys50-Asp102]-COOH, observed mass. 6004.0 ± 0.6 Da, calc. 6004.85 Da (average isotope composition); 90 mg (38.1 μmol, 25.3% yield from 0.15 mmol starting peptide resin) of the peptide segment [Gly1-Tyr18]-$^{\alpha}COS\ CH_2CH_2SO_3H$,

observed mass 2361.7 ± 0.3 Da, calc. 2361.4 Da (av isotopes); and, 60 mg (15.5 μmol, 15.5% yield from 0.1 mmol starting peptide resin) of the pure thioester peptide segment [Thz19-Arg49]-$^{\alpha}$COS $CH_2CH_2SO_3H$, observed mass 3862.5 ± 0.3 Da, calc. 3862.3 Da (av isotopes). Analytical HPLC traces of the purified synthetic products are shown in Figure S3.

3. One-pot native chemical ligations/purification/LCMS characterization. Native chemical ligation was performed by dissolving peptide **3** [Cys50-Asp102]-COOH (22 mg, 3.66 μmol, 1.66 mM) and peptide **2** [Thz19-Arg49]-$^{\alpha}$COS $CH_2CH_2SO_3H$ (15 mg, 3.37 μmol, 1.53 mM) in ligation buffer (2.2 mL) containing 6 M guanidine hydrochloride, 100 mM Na_2HPO_4 and 20 mM TCEP hydrochloride at pH 7 to give ligation product **4** (Figure 2B). After conversion of the Thz-peptide product to the Cys-peptide **5** by addition of 61 mM methoxylamine hydrochloride (134.8 μmol) at pH 4.0 (Figure 2C), peptide **1** [Gly1-Tyr18]-$^{\alpha}$COS $CH_2CH_2SO_3H$ (10.4 mg, 4.4 μmol, 2 mM) was added to the same reaction mixture, and the solution adjusted to pH 6.8 by careful addition of 5 N NaOH. This final native chemical ligation reaction gave full-length polypeptide ligation product **6** (after purification: 19.3 mg, 1.6 μmol, 48% based on limiting peptide segment **2**), as a single peak eluting at 20.8 minutes on analytical LCMS, observed mass 11932.2±0.7 Da, calc. 11932.54 Da (av isotopes).

4. Folding/purification/HPLC and MS characterization. The purified reduced polypeptide [Gly1-Asp102]-COOH (18 mg, 1.5 μmol) was dissolved in degassed folding buffer (36 mL) containing 0.15 M guanidine hydrochloride, 97.5 mM tris-hydroxymethylaminomethane, 1.95 mM $[Glutathione]_{reduced}$ and 0.39 mM $[Glutathione]_{oxidized}$ at pH 8.4. The solution was incubated at room temperature without stirring. Folding was monitored by analytical HPLC (Figure S4) and was essentially complete in 5 days; the folded product was purified by reverse phase HPLC to give 8 mg (0.34 μmol, 23%) of the desired synthetic VEGF protein. Analytical HPLC showed a single sharp peak eluting at 18 minutes on HPLC analysis.

The electrospray ionization mass spectrum of the synthetic VEGF protein was acquired by direct infusion on an Agilent 1100 LC/MSD ion trap using a 30 μM concentration of synthetic protein dissolved in 80% aqueous methanol containing 1% acetic acid (v/v). The solution was infused into the ion source at a flow rate of 2 μL/min using a syringe pump. Dry nitrogen was heated to 250 °C and introduced into the capillary region at a flow rate of 3 L/min. Compound stability and trap drive level parameters were adjusted to 25% and 70% respectively.

5. Circular Dichroism (CD). The circular dichroism spectrum of the synthetic VEGF protein was recorded on an AVIV-202 instrument. An aqueous solution of synthetic VEGF (15 μM, 358 μg/mL) was transferred to a CD cuvette of path length 0.1 cm. CD spectra were measured at room temperature over the range 190-250 nm using 1 nm step and averaging of 5 scans. CD data are shown in Figure S5.

6. X-ray crystal structure determination. Crystals were grown at 19°C using hanging drop vapor diffusion with 2 μL of aqueous protein solution (2.5 mg/mL) and 1 μL of reservoir solution placed over 1 mL of reservoir solution containing 0.2 M ammonium acetate, 0.1 M BIS-TRIS, pH 5.5, 45% v/v (±)-2-methyl-2,4-pentanediol (condition # 50, HR2-144, Hampton Research). Crystals appeared within 7 days.
For data collection, selected crystals were flash frozen in liquid nitrogen after a brief wash in the cryoprotectant [reservoir solution plus 20% (v/v) glycerol]. A complete data set to 1.85 Å was collected from a single crystal at 100 K using 0.97 Å wavelength synchrotron radiation at the Argonne National Laboratory (Advanced Photon Source, beamline 24-ID E, equipped with ADSC Q315 CCD detector). Crystal diffraction images were indexed in the C2 space group and then integrated, scaled and merged with HKL2000.[4]
Cell content analysis revealed that there are three VEGF protein molecules in the asymmetric unit. The structure was solved by molecular replacement with the program PHASER[5] using the previously reported X-ray structure of recombinant VEGF (8-109) (PDB accession code 2VPF) as a search model. The resulting synthetic VEGF model was refined with PHENIX.REFINE[6] using isotropic displacement

parameters and atomic positions against diffraction intensities using a maximum likelihood target function. After each refinement step the model was visually inspected in COOT[7] using both 2Fo-Fc and Fo-Fc difference maps. At the later stages of refinement, the model was refined using TLS parameters in Phenix with total 23 TLS groups refined for the six chains. Out of 102 residues, only residues 7-100 were well defined in all the folded polypeptide chains. In four chains, one additional N-terminal and one additional C-terminal residue were visible. In one chain two additional N-terminal residues were evident from the electron density map. The final model gave R_{factor} (R_{free}) of 18.0% (22.3%). 96.4% of all the residues had main chain torsion angle in the most favorable regions of the Ramachandran plot. X-ray crystal diffraction data and refinement statistics are listed in Table S1.

7. HUVEC assay. Human umbilical vein endothelial cell (HUVEC) proliferation assay was performed by R&D systems, Inc.(Minneapolis). Briefly, reconstituted standards and samples in PBS (plus 1 mg/mL BSA) were diluted to working concentration with assay medium containing medium-199 in 10 mM HEPES and supplemented with 10% fetal bovine serum and antibiotics (penicillin, 100 units/mL; streptomycin, 100 μg/mL) and added to a collagen-coated 96-well microtiter plate using serial dilution. HUVEC cells in assay medium were then added to all wells at 5000 cells/well. The tissue culture dishes were then incubated for 72 hours at 37 °C under 5% CO_2 in a humidified chamber. In the last 18-24 hours of incubation, 0.5 μCi of ^{3}H-thymidine was added per well. The cells were then harvested and ^{3}H-thymidine incorporation was determined by scintillation counting. The plate was read using a F-Max plate reader at Ex.544/Em.590 and the data was analyzed using a 4-parameter curve fit (Figure 5).

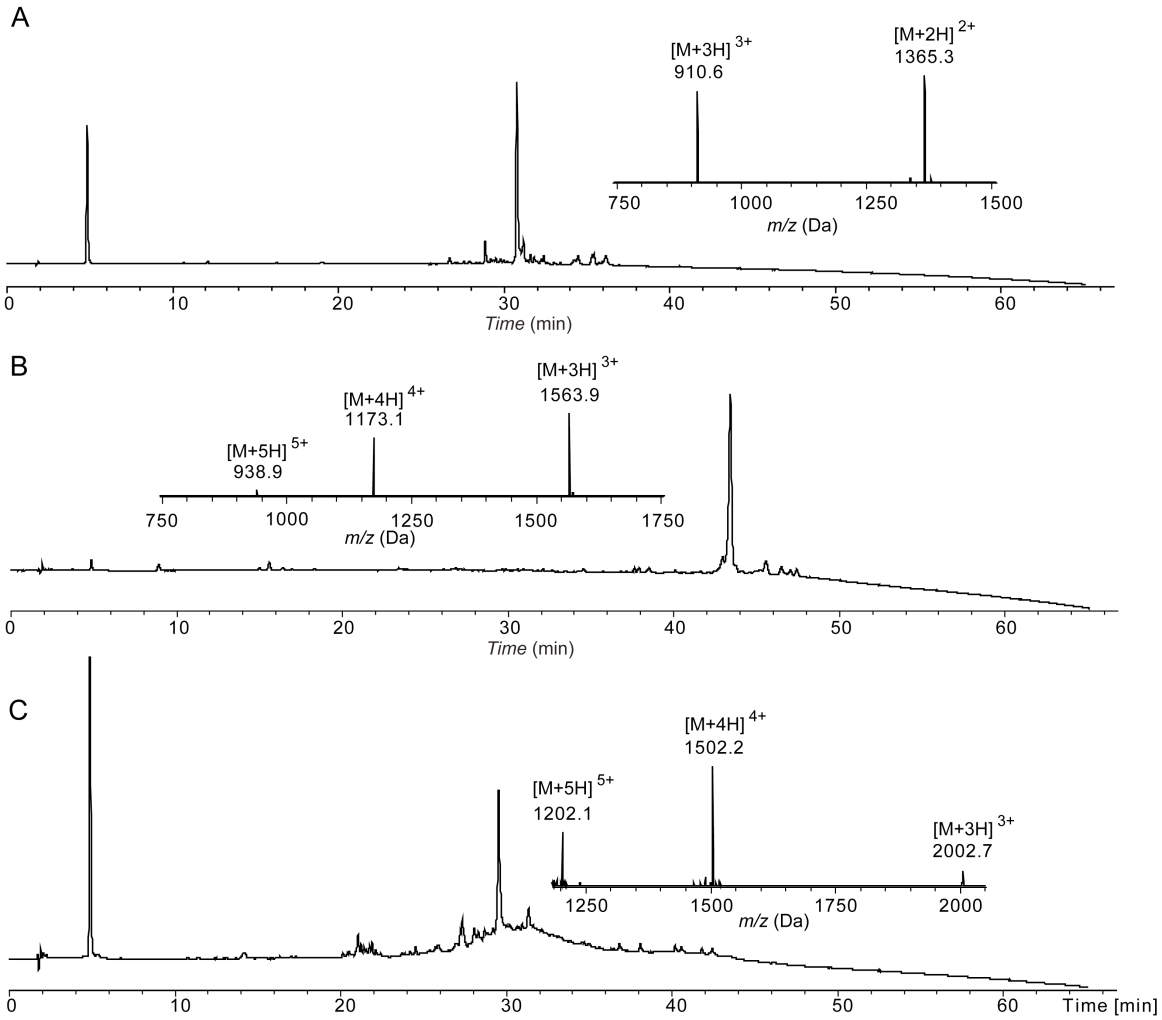

Figure S1. Analytical HPLC profiles (λ = 214 nm) for crude synthetic peptide segments together with LC-MS data (insets) corresponding to each major product. (A) Gly1-Tyr18 $-^{\alpha}COSR_1$; (B) Thz19-Arg49 $-^{\alpha}COSR_2$; and, (C) Cys50-Asp102 $-^{\alpha}COOH$. R_1 = -CH_2CH_2COAla-COOH and R_2 = -$CH_2CH_2CO(Arg)_4$Ala-COOH.

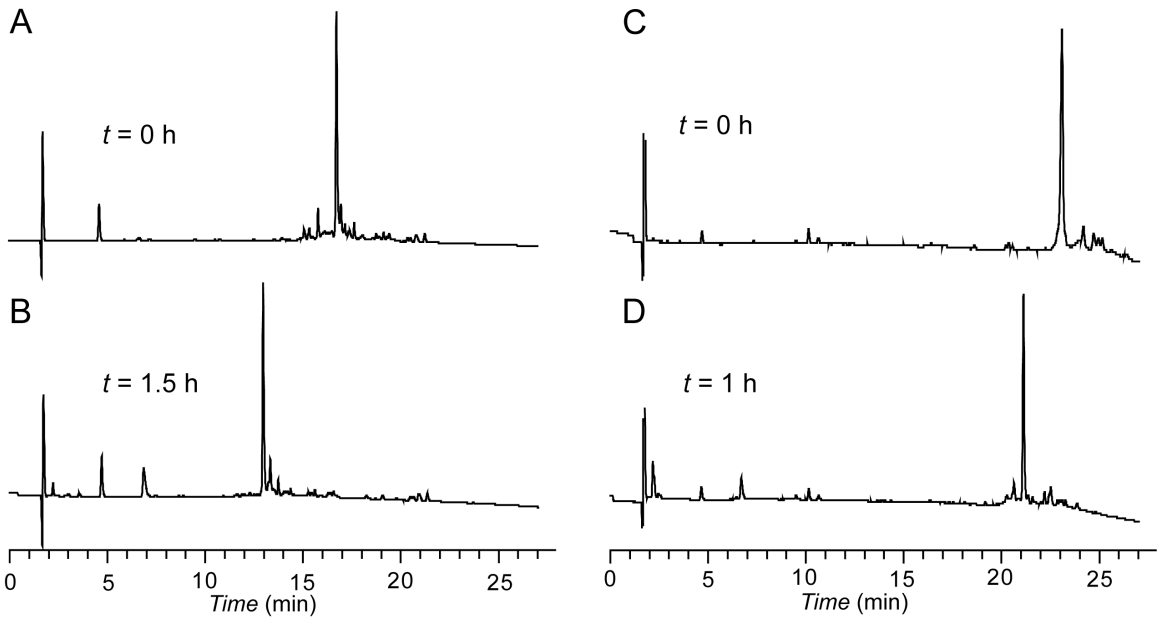

Figure S2. Dnp group removal and transthioesterification with 200 mM MESNa at pH 7. (A) Peptide Gly1-Tyr18 $-^{\alpha}COSR_1$ at T = 0 h; (B) Peptide Gly1-Tyr18 $-\alpha COSR_3$ after T = 1.5 h. Mass change of 367.3 Da from 2728.68 Da indicates the removal of two Dnp groups and thioester exchange with MESNa. (C) Peptide Thz19-Arg49 $-\alpha COSR_2$ at T = 0 h; (D) Peptide Thz19-Arg49 $-\alpha COSR_3$ after T = 1 h. Mass change of 826 Da from 4688.3 Da indicates the removal of one Dnp group as well as thioester exchange with MESNa. R_1 = $-CH_2CH_2COAla-COOH$, R_2 = $-CH_2CH_2CO(Arg)_4Ala-COOH$ and R_3 = $-CH_2CH_2SO3H$. For LC analysis the chromatographic separations were performed using a linear gradient of 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 (Agilent), 4.6 × 150 mm column at 40 °C (flow rate 1 mL/min).

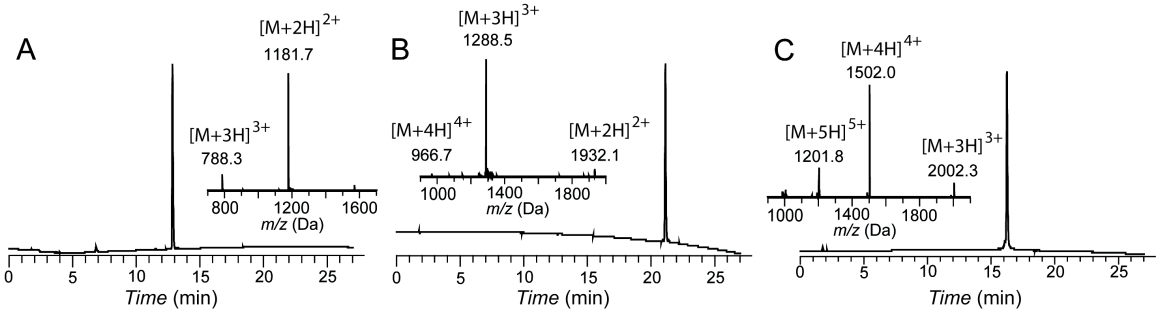

Figure S3. Analytical HPLC profiles (λ = 214 nm) together with LC-MS data (inset) corresponding to the purified peptide segments. (A) Gly1-Tyr18 –αCOSR (observed mass: 2361.7 ± 0.3 Da, calculated mass: 2361.4 Da (av isotopes), (B) Thz19-Arg49 –αCOSR (observed mass: 3862.5 ± 0.3, calculated mass: 3862.3 Da (av isotopes), (C) Cys50-Asp102 –αCOOH (observed mass: 6004.0 ± 0.6 Da, calculated mass: 6004.85 Da (av isotopes). R = $-CH_2CH_2SO_3H$.

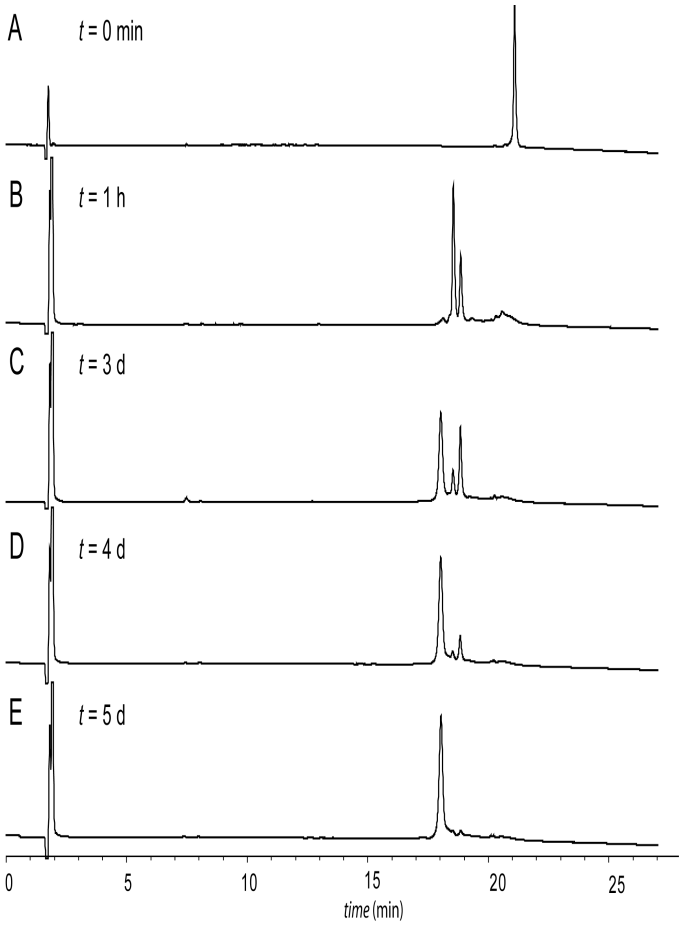

Figure S4. Folding/disulfide formation of the purified 102 AA VEGF polypeptide (18 mg, 1.5 μmol, 0.5 mg/mL) in presence of GSH/GSSG (1.95 M/ 0.39 M), 97.5 mM Tris, 0.15 M Gu.HCl, pH 8.4 at room temperature. (A) HPLC profile of the starting polypeptide before the addition of folding buffer; (B) Folding after 1 h; (C) Folding after 3 days; (D) Folding after 4 days; (E) Folding was essentially complete in 5 days. Folded protein had a 2.8 min earlier retention time compared to the reduced polypeptide. Preparative HPLC purification gave synthetic VEGF protein (8 mg, 0.34 μmol, 45%). 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 (Agilent), 4.6 × 150 mm column at 40 ^{o}C (flow rate = 1 mL/min).

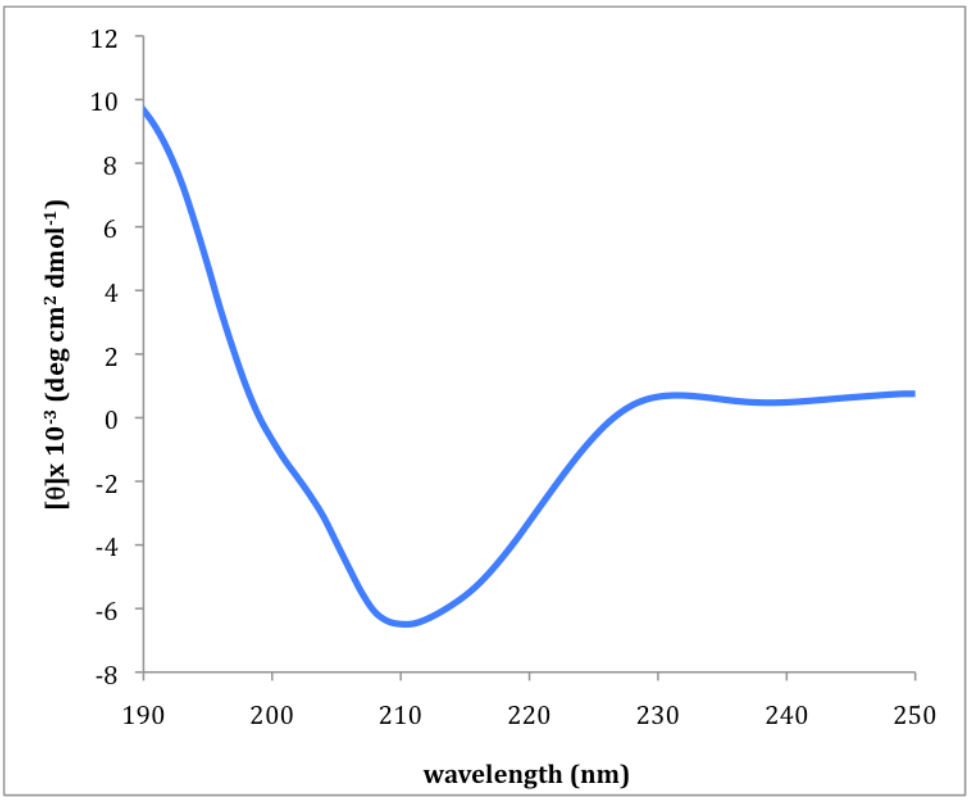

Figure S5. Circular dichroism spectra of synthetic VEGF protein. CD spectra were recorded at room temperature using 1.5 μM (0.36 mg/mL) of the synthetic protein dissolved in water. CD cuvette path length = 0.1 cm. Total number of scans = 6.

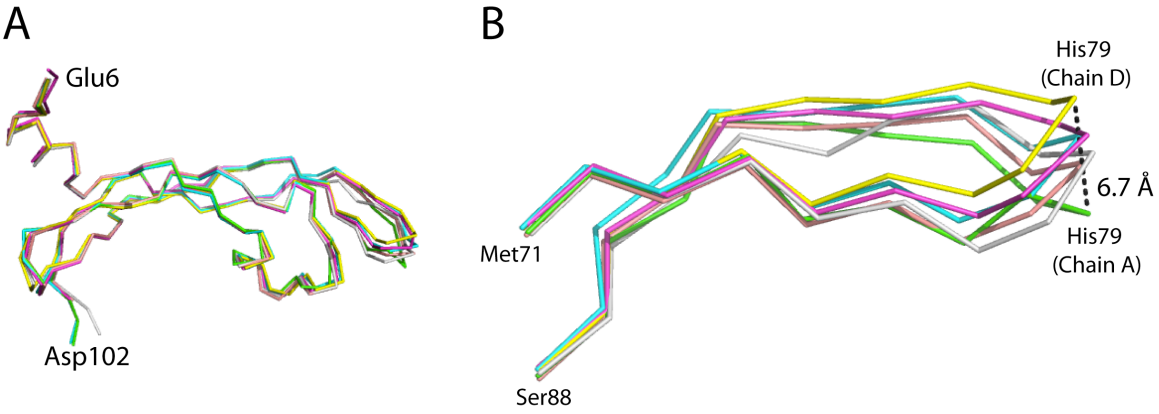

Figure S6. Superposition of the six crystallographically independent copies of the VEGF monomer at C-α. A) C-α superposition on secondary structure elements. B) C-α superposition showing the loop region (Met71-Ser88) representing the snapshots of a concerted loop movement.

Table S1. X-ray data collection and refinement statistics of VEGF.

Data collection*	
Space group	C_2
Wavelength (Å)	0.97919
Cell dimensions	
a, *b*, *c* (Å)	230.9, 44.0, 73.1
α, β, γ (°)	90.0, 99.9, 90.0
Mol/asymmetric unit	3 (dimer)
Mol/unit cell	12 (dimer)
Resolution (Å)	50.0 - 1.85 (1.92 – 1.85)
R_{merge}	0.074 (0.574)
$I/\sigma I$	18.0 (1.9)
Redundancy	3.1 (3.1)
Refinement*	
Resolution (Å)	32.5 - 1.85 (1.88 – 1.85)
No. reflections	61870
Completeness (%)	98.9 (94)
R_{work}/ R_{free}	0.180 / 0.223
No. atoms	
Non-solvent	4858
Solvent	422
Average B-factor (Å^2)	25.87
R.m.s deviations	
Bond lengths (Å)	0.007
Bond angles (º)	1.032

*Highest resolution shell is shown in parentheses.

Coordinates and structure factors have been deposited in the Protein Data Bank with accession code 3QTK.

References

[1] A. R. Mitchell, S. B. H. Kent, M. Engelhard, R. B. Merrifield *J. Org. Chem.* **1978**, *43*, 2845–2852.
[2] M. Schnölzer, P. Alewood, A. Jones, D. Alewood, S. B. H. Kent, *Int. J. Pept. Res. Ther.* **2007**, *13*, 31- 44.
[3] T. M. Hackeng, J. H. Griffin, P. E. Dawson *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 10068-10073.
[4] Z. Otwinowski, W. Minor, *Meth. Enzymol.* **1997**, *276*, 307–326.
[5] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Cryst.* **2007**, *40*, 658-674.
[6] P. V. Afonine , R. W. Grosse-Kunstleve, P. D. Adams PD (2005) PHENIX.REFINE. *CCP4 Newsl.* 42, contribution 8.
[7] P. Emsley, K. Cowtan, *Acta Cryst D* **2004**, *60*, 2126–2132.