Chemical synthesis and X-ray structure of a heterochiral {D-protein antagonist *plus* vascular endothelial growth factor} protein complex by racemic crystallography

Kalyaneswar Mandal[a,1], Maruti Uppalapati[b,1], Dana Ault-Riché[c], John Kenney[d], Joshua Lowitz[d], Sachdev S. Sidhu[b,2], and Stephen B.H. Kent[a,2]

[a]Departments of Chemistry, Biochemistry and Molecular Biology, Institute for Biophysical Dynamics, University of Chicago, Chicago, IL 60637; [b]Banting and Best Department of Medical Research, University of Toronto, Toronto, ON, Canada M5S 3E1; [c]Reflexion Pharmaceuticals, San Francisco, CA 94104; and [d]Antibody Solutions, Sunnyvale, CA 94089

Edited by* James A. Wells, University of California, San Francisco, CA, and approved August 3, 2012 (received for review June 21, 2012)

Total chemical synthesis was used to prepare the mirror image (*D*-protein) form of the angiogenic protein vascular endothelial growth factor (VEGF-A). Phage display against *D*-VEGF-A was used to screen designed libraries based on a unique small protein scaffold in order to identify a high affinity ligand. Chemically synthesized *D*- and *L*- forms of the protein ligand showed reciprocal chiral specificity in surface plasmon resonance binding experiments: The *L*-protein ligand bound only to *D*-VEGF-A, whereas the *D*-protein ligand bound only to *L*-VEGF-A. The *D*-protein ligand, but not the *L*-protein ligand, inhibited the binding of natural $VEGF_{165}$ to the VEGFR1 receptor. Racemic protein crystallography was used to determine the high resolution X-ray structure of the heterochiral complex consisting of {*D*-protein antagonist + *L*-protein form of VEGF-A}. Crystallization of a racemic mixture of these synthetic proteins in appropriate stoichiometry gave a racemic protein complex of more than 73 kDa containing six synthetic protein molecules. The structure of the complex was determined to a resolution of 1.6 Å. Detailed analysis of the interaction between the *D*-protein antagonist and the VEGF-A protein molecule showed that the binding interface comprised a contact surface area of approximately 800 $Å^2$ in accord with our design objectives, and that the *D*-protein antagonist binds to the same region of VEGF-A that interacts with VEGFR1-domain 2.

One of the most remarkable aspects of the natural world is the homochirality of the protein macromolecules found in living systems (1). All protein molecules found in nature contain ribosomally translated polypeptide chains that are comprised exclusively of *L*-amino acids and the achiral amino acid glycine. Starting about 20 years ago, chemists and biochemists became interested in the properties of unnatural mirror image *D*-protein molecules; i.e., proteins with polypeptide chains of the same sequence as natural proteins, but made from *D*-amino acids and glycine. In pioneering work, Zawadzke and Berg showed that a racemic mixture of the enantiomeric forms of the small protein rubredoxin crystallized in a centrosymmetric space group and that the folded protein molecules were mirror images of one another (2). More recently, racemic crystallography of chemically synthesized protein enantiomers has been used to facilitate the crystallization of recalcitrant proteins and to solve protein structures that had not previously been determined by X-ray crystallography (3–7).

Chirality, the handedness of molecules (8), is fundamental to biological interactions. This aspect of chirality was first systematically demonstrated by Emil Fischer, who—based on his studies of sugar stereochemistry—formulated the 'lock and key' principle of enzyme action on chiral substrate molecules (9). Since then, it has been repeatedly demonstrated that natural proteins preferentially bind to only one enantiomer of chiral small molecule ligands. The principles of molecular chirality dictate that the inverse also be true—i.e., that the mirror image form of a natural protein molecule will preferentially bind to the opposite enantiomer of a chiral ligand (10). Chirality is believed to be of particular importance in protein-protein interactions—it is considered a self-evident conclusion of symmetry arguments that the enantiomer of a protein ligand will not bind to the same natural protein target, and that enantiomeric forms of a protein-(protein ligand) pair will have the same affinity for one another. Nonetheless, experimental demonstration of this principle may be instructive.

In 1996, Peter Kim and his colleagues at the Whitehead Institute at MIT used peptide phage display against a *D*-protein target as a way to systematically develop *D*-peptide ligands of natural proteins (11). As originally conceived, this 'mirror image (peptide) phage display' method involved the chemical synthesis of the mirror image form of a natural protein molecule, after which peptide phage libraries were screened to identify *L*-peptide ligands to the *D*-protein. Then, chemical synthesis was used to make the corresponding *D*-peptide ligand, which obligately bound with the same affinity to the natural protein target. Such *D*-peptide ligands would be resistant to proteolytic digestion in vivo, and for that reason they have excited a great deal of interest. Although mirror image phage display has been used in a number of academic studies (12–14), it has not yet led to the development of *D*-peptides as therapeutics.

We set out to apply the mirror image phage display approach to esigned libraries of a unique protein scaffold in order to develop high affinity *D*-protein antagonist(s) for biologically active protein targets. A properly engineered *D*-protein molecule would have near optimal properties as a therapeutic: A small *D*-protein can be produced by chemical manufacture, will resist proteolytic degradation, and is expected to be nonimmunogenic (15). Further, it should be possible to engineer a small *D*-protein to have affinity and specificity similar to that of antibodies for a therapeutic target molecule. Development of improved antagonists of growth factor activity is an important current objective in medicinal chemistry. The angiogenic factor vascular endothelial growth factor (VEGF-A) is the target of engineered monoclonal antibodies that inhibit angiogenesis by interfering with the interaction of VEGF-A with its

Author contributions: K.M., M.U., D.A.-R., S.S.S., and S.B.K. designed research; K.M., M.U., J.K., and J.L. performed research; K.M., M.U., D.A.-R., J.K., J.L., S.S.S., and S.B.K. analyzed data; and K.M., M.U., and S.B.K. wrote the paper.

Conflict of interest statement: This research has been carried out at the University of Chicago and the University of Toronto as part of a research program funded by the two universities under agreements with a start up company, Reflexion Pharmaceuticals, Incorporated. Both universities have minor equity positions in Reflexion. Ault-Riché, Kent, and Sidhu are founders of Reflexion. With the exception of Joshua Lowitz, all the authors of this paper own equity in Reflexion, and thus each of these authors declares a conflict of interest.

*This Direct Submission article had a prearranged editor.

Data deposition: Crystallography, atomic coordinates, and structure factors have been deposited in the Protein Data Bank, www.pdb.org [PDB ID codes 4GLU (D-VEGF-A), 4GLS (racemic complex in space group *P*21), and 4GLN (racemic complex in space group *P*21/*n*)].

[1]K.M. and M.U. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: sachdev.sidhu@utoronto.ca or skent@uchicago.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1210483109/-/DCSupplemental.

receptors (16), and several of these engineered antibodies have become important human therapeutics.

The objective of the work reported here was to develop a *D*-protein ligand that binds to VEGF-A and acts as an antagonist of VEGF-A binding to its receptor. Our target was the covalent homodimer VEGF[8–109] protein molecule; this form of VEGF-A retains full biological activity and was used in the development of the monoclonal antibodies bevacizumab and ranibizumab (16–18). Here we report the total chemical synthesis of the mirror image *D*-VEGF-A protein molecule, the use of mirror image protein phage display of designed libraries based on a unique protein scaffold, and the consequent generation of a specific *D*-protein ligand for VEGF-A that blocks binding to the VEGFR1 receptor. The high resolution X-ray structure of the {*D*-Protein Antagonist plusVEGF-A} protein complex was determined by racemic protein crystallography.

Results and Discussion

Chemical Synthesis of *D*-VEGF-A. The essential first step was to prepare the mirror image form of the VEGF[8–109] target protein. Recently we reported the total synthesis of the natural enantiomer of this form of the VEGF-A protein molecule with full biological activity (19). Here, we used total synthesis enabled by modern chemical ligation methods to prepare *D*-VEGF-A, the mirror image form of VEGF-A (Fig. 1). Three unprotected synthetic peptide segments were condensed by native chemical ligation (20) in a series of 'one pot' reactions; i.e., without purification of intermediate products (Fig. 1*A*). The purified synthetic 102 residue *D*-polypeptide chain (Fig. 1*B*) was folded in the presence of a redox couple, with concomitant formation of native disulfide bonds, to give the *D*-protein molecule (Fig. 1*C*). Synthetic *D*-VEGF-A was characterized by electrospray mass spectrometry, and had an observed mass (23‹849.3 ± 0.7 Da) in agreement with the expected mass (23,849.1 Da, average isotope composition) for the covalent homodimer containing eight disulfide bonds (19). Synthetic *D*-VEGF-A was crystallized, and X-ray diffraction data was acquired to a resolution of 1.9 Å. The structure was solved by molecular replacement, using inverted coordinates of the previously reported VEGF-A structure (PDB code: 3QTK) as a search model. A portion of the resulting 2Fo-Fc electron density map is shown in *SI Appendix*, Fig. S1*C*, and the structure of synthetic *D*-VEGF-A is shown in Fig. 1*D*. The chemically synthesized *D*-VEGF-A had a structure that was the mirror image of natural recombinant VEGF-A (17, 18), within experimental uncertainty.

A Protein Ligand for *D*-VEGF-A We used the B1 domain of streptococcal protein G (GB1) (21) as a scaffold for the development of small protein ligands. Natural protein GB1 is a well studied small protein that folds stably and reversibly and that binds to the Fc region of immunoglobulins with high affinity (22). GB1 consists of a polypeptide chain of 56 amino acids, small enough to be made by total chemical synthesis and yet large enough to provide a sufficient surface area for a strong binding interface between the *D*-protein antagonist and its target protein. While there are several miniprotein scaffolds that have been developed (23), they typically do not present enough solvent exposed surface area to match the affinity and specificity of antibodies. We displayed GB1 on M13 filamentous phage as a fusion with coat protein p3, and with an N-terminal FLAG tag. A library of GB1 mutants was constructed from 15 contiguous residues on the surface of the protein, within the region spanning residues 21–41 and with the insertion of one or two additional residues between positions 41 and 42 (Fig. 1*E*), in order to generate a binding interface with an area comparable to that of antibodies. The selected residues were randomized by a KHT codon, which allows the amino acids Y, A, D, S, F, and V at each position. This codon was chosen because one of us (Sidhu) has previously shown that minimalist libraries of aromatic amino acids and small residues are sufficient for generating high affinity antibodies (24, 25). This strategy allows better sampling of the potential sequence space in a library of approximately 10^{10} variants compared to randomizing with all 20 amino acids.

The library of GB1 mutants was screened against the chemically synthesized *D*-VEGF-A and after four rounds of panning

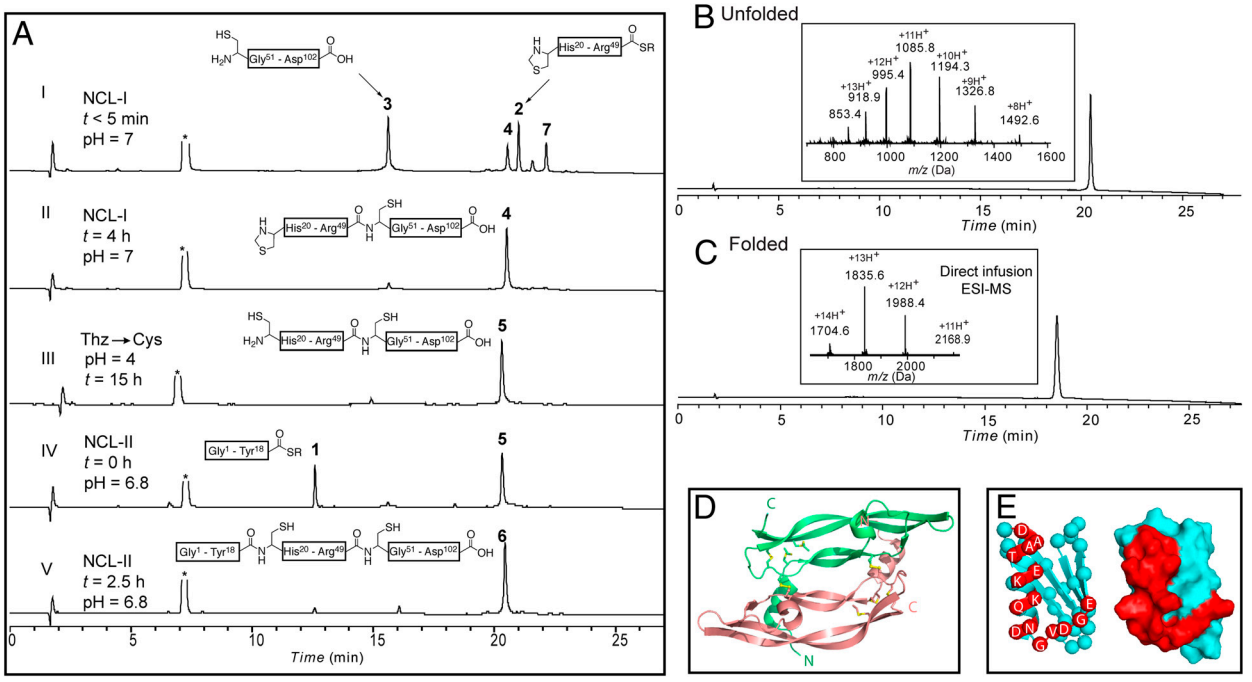
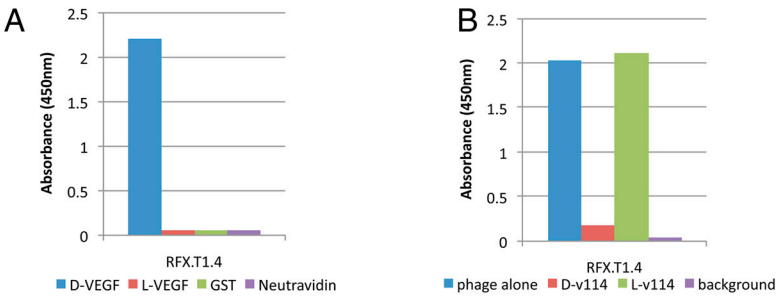


Fig. 1. Total chemical synthesis of *D*-VEGF-A. The amino acid sequence is given in (19). (*A*) Analytical HPLC data for the total chemical synthesis of the *D*-VEGF [8–109] polypeptide chain. (*B*) LCMS data for the synthetic *D*-VEGF[8–109] polypeptide chain [observed mass: 11,932.7 ± 0.3 Da; calculated mass 11,932.5 Da (average isotopes)]. (*C*) HPLC and direct infusion electrospray MS data for the folded, homodimeric synthetic *D*-VEGF-A protein [observed mass: 23,849.3 ± 0.7 Da; calculated mass 23,849.1 Da (average isotopes)]. (*D*) Cartoon representation of the 1.9 Å resolution X-ray structure of synthetic *D*-VEGF-A protein. (*E*) (left) Cartoon of the protein GB1 scaffold; (right) representation of the surface of GB1 with residues randomized in the designed library colored red.



C

	Sequence																								
Residue #	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	41a	42	43
RFX.T1.4	E	A	V	D	V	Y	D	A	F	D	V	F	F	V	Y	A	A	S	N	F	S	D	F	D	D	W	T
Library 1	E	A	V	X	X	X	X	A	F	D	V	F	F	V	Y	A	A	S	N	F	S	D	F	D	D	W	T
Library 2	E	A	V	D	V	Y	D	A	X	X	V	F	X	X	Y	A	A	S	N	F	S	D	F	D	D	W	T
Library 3	E	A	V	D	V	Y	D	A	F	D	V	F	F	V	Y	A	X	X	N	X	X	D	F	D	D	W	T
Library 4	E	A	V	D	V	Y	D	A	F	D	V	F	F	V	Y	A	A	S	N	F	S	X	X	X	X	W	T

X is a randomization site mutated with NNK codon to express all 20 amino acids

D

E

```
GB1       TYKLILNGKT LKGETTTEAV DAATAEKVFK QYANDNGVDG E-WTYDDATK TFTVTE

RFX001    TYKLILNGKT LKGETTTEAV DVFDAFDVFF VYAASNFSDF DDWTYDDATK TFTVTE
```

Fig. 2. Specificity and sequence diversity of selected phage-displayed protein ligands. (*A*) Phage-displayed protein ligand RFX.T1.4 specifically binds to only *D*-VEGF (blue) and does not bind to *L*-VEGF (red), GST (green) and Neutravidin (purple); (*B*) Competitive binding of phage-displayed protein ligand RFX.T1.4 in the presence of a saturating concentration (100 μM) of enantiomeric forms of the peptide competitor v114 (26): (red) *D*-v114 competitively inhibits binding; (green) *L*-v114 has no effect on binding. (*C*) Sequences of the best hit RFX.T1.4 (top) and the affinity maturation libraries used. (*D*) Sequence profile logos for selected clones (8–12 unique sequences); the size of the single letter code represents the frequency of occurrence of that amino acid at a given position. (*E*) Sequence of the optimized point mutant RFX001 (bottom); the parent sequence of GB1 (top). Underlined/bold residues are optimized in RFX001 compared to GB1; a residue was inserted between residues 41 and 42 as discussed in the text.

20-fold enrichment was observed by phage pool ELISA. Single clones were analyzed for binding-specificity (Fig. 2*A*), and then sequenced. *SI Appendix*, Table S1 shows the sequences of selected clones that bound to *D*-VEGF-A. *SI Appendix*, Fig. S2 shows the binding specificity of selected clones. Phage containing the clone RFX.T1.4 had the highest affinity for *D*-VEGF-A as determined by competitive ELISA. RFX.T1.4 phage bound specifically to *D*-VEGF-A (Fig. 2*A*), and the binding could be competed by the mirror image form of a known antagonistic peptide v114 (Fig. 2*B*) (26), suggesting that the GB1 mutants encoded by these clones would be likely to have antagonistic activity. Further affinity maturation was performed on clone RFX.T1.4, by relaxing codon constraints to allow all 20 amino acids at each of the randomized positions. Four libraries, each mutating a set of four residues (Fig. 2*C*), were constructed and three rounds of phage selections were performed. Twenty-four clones were selected and sequenced from each library, and from the unique sequences a binding profile was generated as shown in Fig. 2*D*. The binding profiles indicated that hydrophobic residues were preferred at positions 22, 23, and 40 and were conserved at positions 26, 30, and 37. Negatively charged residues were conserved in positions 27 and 42. Of the 16 randomized residues, these eight residues seemed to be important for binding to *D*-VEGF-A. From the binding profiles it appeared that the dominant mutations obtained from affinity maturation were Y23F and F40I. Preliminary experiments showed that only the Y23F mutation increased the affinity (twofold), while the F40I mutation did not improve affinity. Therefore a point mutant [Y23F]RFX.T1.4 was constructed and phage containing this point mutant had high affinity for *D*-VEGF-A. A GB1-derived protein molecule with this amino acid sequence (RFX001) was chosen for further study (Fig. 2*E*).

Chiral Specificity of VEGF-A Interactions with Mirror Image RFX001 Protein Ligands. The *D*-amino acid and *L*-amino acid forms of the RFX001 polypeptide chain were chemically synthesized and were each separately folded by dissolution in PBS at pH 7.4, and used to evaluate affinity for the enantiomeric forms of VEGF-A.

The mirror image RFX001 protein ligands showed reciprocal chiral specificity in surface plasmon resonance experiments as shown in Fig. 3*A*: *D*-RFX001 bound only to *L*-VEGF-A and did not bind to *D*-VEGF-A, while *L*-RFX001 bound only to *D*-VEGF-A and did not bind to *L*-VEGF-A. Using data from the two highest concentrations of protein ligand, the observed dissociation constant values K_d were 85 ± 12 nM for the D-RFX001 and 95 ± 8 nM for the L-RFX001 (*SI Appendix*, Table S2). The enantiomeric forms of the RFX001 protein molecule were also evaluated for their ability to inhibit the binding of $VEGF_{165}$, the most abundant form of VEGF-A found in vivo (16), to its receptor VEGFR1 (Flt1). Only *D*-RFX001 inhibited the binding of $VEGF_{165}$ to the VEGF receptor Flt1 (Fig. 3*B*).

Structure of the {VEGF-A + *D*-Protein Antagonist} Complex by Racemic Crystallography. We next set out to determine the molecular structure of the complex of VEGF-A with the protein antagonist *D*-RFX001. We attempted to crystallize a mixture of synthetic L-VEGF-A with two equivalents of synthetic D-RFX001 using 96 index conditions of the Quiagen Pro-complex suite; only two conditions produced micro crystals. Optimization of these two conditions did not lead to the formation of diffraction-quality crystals. Recently we showed that a racemic protein mixture crystallizes more readily than the natural *L*-protein alone*, and this has enabled us to use racemic crystallography to determine a number of

*Other potential advantages of racemic protein crystallography include: Facilitated crystallization to give well-ordered racemic crystals that diffract to high resolution; and, in the centrosymmetric space groups that can only be formed from a racemic mixture, phases of the reflections are quantized (e.g. for P1 ⟨bar⟩ or $P2_1/c$ it is 0 or π radians), which can simplify structure solution (2, 5, 32).

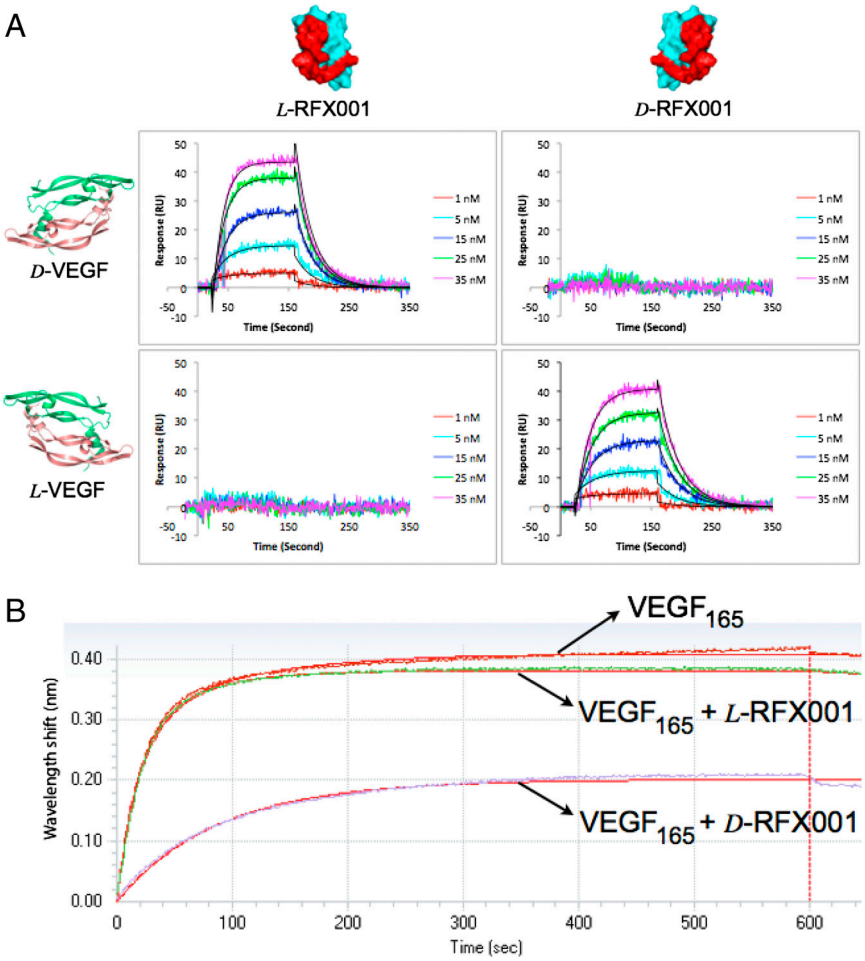


Fig. 3. Chiral specificity of protein ligand/VEGF-A interactions. (*A*) Surface plasmon resonance sensorgrams showing the interaction of enantiomeric forms of the protein ligand RFX001 with enantiomeric forms of VEGF-A: *L*-RFX001 bound to *D*-VEGF-A, but did not bind to *L*-VEGF-A; in contrast, *D*-RFX001 bound to *L*-VEGF-A, but did not bind to *D*-VEGF-A. (*B*) As measured by biolayer interferometry (ForteBio Octet) using the wavelength shift on binding, *D*-RFX001 is an antagonist of $VEGF_{165}$ binding to VEGFR1: *D*-RFX001 inhibited the binding of VEGF165 to VEGFR1 (lower curve), while added *L*-RFX001 had no effect even at 2 μM (upper curves).

protein structures that had not previously been determined by X-ray crystallography (3, 4, 7). However, to date racemic protein crystallography has not been used to determine the structure of a heterochiral protein complex; i.e., a complex that consists of two different protein molecules of opposite chirality that bind to one another in solution. Here, we obtained crystals from a solution of chemically synthesized protein molecules that contained one equivalent each of *L*-VEGF-A and *D*-VEGF-A, and two equivalents each of the enantiomeric forms of the protein antagonist; i.e., *D*-RFX001 and *L*-RFX001[†]. Crystals were readily obtained, and synchrotron X-ray diffraction data were collected to 1.6 Å resolution. Data reduction statistics revealed that the protein complex had crystallized in the centrosymmetric space group $P2_1/n$. Matthews cell content analysis (27) suggested that there were one VEGF-A protein molecule and two RFX001 protein molecules in the $P2_1/n$ asymmetric unit. The structure was solved in $P2_1/n$ by molecular replacement using inverted and noninverted coordinates of the previously reported crystal structures of VEGF (PDB code: 3QTK) and GB1 (PDB code: 2QMT), as shown in Fig. 4.

The Binding Interface. Each *L*-VEGF-A protein molecule bound two *D*-RFX001 protein molecules. The structure of the VEGF protein molecule in the {VEGF-A + *D*-RFX001} complex is closely similar to that found in previously reported structures (17, 18). The structure of RFX001 in comparison to wild-type GB1 (PDB code: 2QMT) revealed that the major differences in the structure occur in the loop 2 region which has one amino acid inserted between residues 41 and 42 of the wild-type sequence. The Cα atoms in the rest of the two proteins align with a rmsd of 0.75 Å. Note that Phe37 and Phe40 in loop 2 form part of the RFX001 protein core and help order loop 2 to present a suitable surface for VEGF-A binding (*SI Appendix*, Fig. S5). Apart from the mutated residues, the side chains of the residues in the RFX001 molecule align well with those of GB1, except for

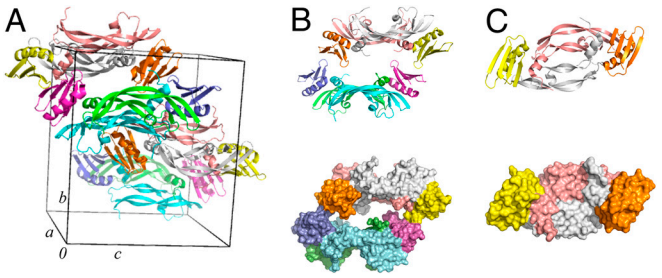


Fig. 4. X-ray structure of the {VEGF-A + *D*-RFX001} complex. Racemic crystallization gave crystals that diffracted to a resolution of 1.6 Å, and the structure was solved in $P2_1$ and $P2_1/n$ (21). (*A*) Unit cell representation of the {VEGF-A + *D*-RFX001} complex crystallized as a racemate. The unit cell contains a total of 12 synthetic protein molecules {2 × *L*-VEGF-A, 2 × *D*-VEGF-A, 4 × *D*-RFX001, 4 × *L*-RFX001}. (*B*) (upper): Asymmetric unit in the space group $P2_1$; (lower): Surface representation of the asymmetric unit in the space group $P2_1$. (*C*) (upper): Asymmetric unit in the space group $P2_1/n$; (lower): Surface representation of the asymmetric unit in space group $P2_1/n$. Color coding: green and cyan indicate the two chains of the homodimeric *D*-VEGF-A protein molecule; gray and salmon indicate the two chains of the homodimeric *L*-VEGF-A protein; yellow and orange indicate the two *D*-protein antagonist molecules; purple and mauve indicate the two *L*-protein antagonist molecules.

[†]There is a two-fold axis of symmetry in the homodimeric VEGF-A protein molecule (17,18); hence, one molecule of VEGF-A was expected to bind two molecules of the D-protein antagonist.

Trp42 that shows a shift in position most likely caused by the insertions of Phe37 and Phe40 into the core.

It is of particular note that *D*-RFX001 binds to the same region of VEGF-A that interacts with VEGFR1-domain 2 (17), consistent with the ability of *D*-RFX001 to act as an antagonist of the VEGF-A/VEGFR1 interaction. The structural basis for this antagonism is vividly illustrated in Fig. 5 (top). Despite its modest size, the protein *D*-RFX001 exhibits a contact surface area of approximately 800 Å^2 at the binding interface with VEGF-A and covers much of the contact surface that VEGF-A uses to interact with domain 2 of the VEGFR1 receptor molecule (17).

The interaction between the two protein molecules is dominated by a total of ten aromatic residues (six from *D*-RFX001 and four from the VEGF-A molecule) (Fig. 5, bottom, left). In the *D*-RFX001 ligand the helix (α1), the third β-strand (β3), and the loop spanning residues 37 to 41 all participate in contacts with VEGF-A. Each residue in the loop region (residue 37 to 41) of the *D*-RFX001 molecule makes direct polar contact with VEGF-A at the binding interface: Residue Ser38 makes a direct backbone-backbone H-bond contact with Gln82 of VEGF-A; and, residue 40 in *D*-RFX001makes two additional direct H-bonds involving the main chain amide bonds of residues 82 and 84 of VEGF-A. Similarly, the backbone amide oxygen atom of residue 23 of the *D*-RFX001 molecule makes a direct polar contact with the side chain of Asn55 residue of VEGF. In addition to the direct H-bonding network, another feature of the *D*-RFX001–VEGF-A interaction is the presence of a salt bridge at the binding interface between

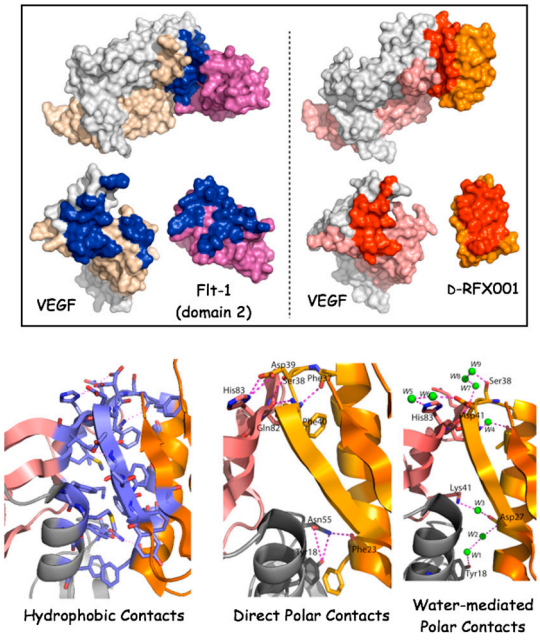


Fig. 5. Interactions at the binding interface between VEGF-A and *D*-RFX001. (Upper) Comparison of binding surfaces. (Left box) Recombinant VEGF-A interacts with VEGFR1-domain 2 (light magenta) with buried surface area of 1,300 Å^2. The interaction interface is shown in dark blue. Taken from Wiesmann, et al. (17). (Right box) Interaction interface of synthetic VEGF-A and *D*-RFX001 (orange). The buried surface area (shown in red) between VEGF and *D*-protein antagonist is ~800 Å^2. (Bottom) Contact residues in the *D*-RFX001–VEGF-A complex. (Left) A distinctive feature of the *D*-protein ligand-binding surface is the presence of numerous aromatic side chains originating from the N-terminal alpha-helix of VEGF and the helix/third β-strand of the *D*-protein ligand. Residues that are involved in the interactions are colored in light blue. (Middle) Direct hydrogen bonding interactions involving the side chains of the residues of VEGF and *D*-protein ligand are shown as dotted lines. (Right) Water-mediated hydrogen bonding networks at the interface. Water molecules are shown as green spheres. In all the three boxes the *D*-protein ligand is colored in orange, and the interacting VEGF chains are colored in silver-black and salmon.

the side chain of Asp39 of *D*-RFX001 and the side chain of His83 of VEGF-A (Fig. 5, bottom, middle). There are several other *D*-RFX001–VEGF-A hydrogen bonding networks mediated by well defined water molecules as shown in Fig. 5, bottom, right.

Comparison of Protein Enantiomers. In order to compare the enantiomeric forms of the proteins present in the unit cell, we also solved the structure in the space group $P2_1$[‡]. This procedure allows the enantiomeric protein molecules present in the complex to vary independently of one another in optimizing the fit of the molecular structures to the experimental electron densities. The $P2_1$ asymmetric unit contained a total of six synthetic protein molecules: Two *D*-RFX001 protein molecules bound to the opposite poles of one *L*-VEGF-A molecule; and, two *L*-RFX001 protein molecules bound to the opposite poles of one *D*-VEGF-A molecule (Fig. 4*B*). Because the complex crystallized in the centrosymmetric monoclinic space group P21/n, it possesses an inversion center. Comparison of the coordinates of main chain nonhydrogen atoms between enantiomeric VEGF-A molecules which are related by inversion gave an rmsd value of 0.2 Å. Comparisons between enantiomers for the four RFX001 protein molecules found in the $P2_1$ asymmetric unit are more complicated: Two enantiomeric pairs are related by inversion (L1-D2 or L2-D1 in Fig. 6), and two enantiomeric pairs are NOT related by inversion (L1-D1 or L2-D2 in Fig. 6). Comparison of pairs of enantiomeric RFX001 protein molecules that are related by inversion gave a main chain rmsd value of just 0.1 Å. Comparison of pairs of enantiomers that are NOT related by inversion gave a main chain rmsd value of 0.5 Å; this larger value is not unexpected, because in this case the enantiomers are in distinct environments within the crystal.

Conclusions

Total chemical synthesis of the mirror image form of VEGF-A has enabled the identification of an engineered protein ligand with high affinity and specificity for the *D*-VEGF-A protein molecule. The small protein GB1 was used as a unique scaffold to display a designed diversity library of mutants. Screening against *D*-VEGF-A, followed by a round of affinity maturation, identified mutants with good affinity and enabled the design of a point mutant with high affinity for *D*-VEGF-A. Chemical synthesis of the mirror image form of the point mutant gave a *D*-protein ligand that bound to *L*-VEGF-A but as expected did not bind to *D*-VEGF-A. The *D*-protein ligand acted as an antagonist of natural VEGF binding to its cognate receptor Flt1. The high resolution X-ray structure of the heterochiral {*L*-VEGF-A + *D*-protein antagonist} was determined by racemic protein crystallography. Detailed analysis of the interaction between the *D*-protein antagonist and VEGF-A showed that the binding interface comprised a contact surface area of approximately 800 Å^2 in accord with our design objectives, and that the *D*-protein antagonist binds to the same region of VEGF-A that interacts with VEGFR1-domain 2.

Using the *D*-protein ligand designed by phage display and the corresponding L-protein target molecule, we have shown that the determination of the high resolution X-ray structures of even large protein complexes can be facilitated by racemic crystallography. Experimentally, this work involved the total chemical synthesis of the enantiomeric forms of protein molecules of 6.4 kDa and 23.8 kDa. Crystallization of a racemic mixture of these synthetic proteins in appropriate stoichiometry gave a racemic protein complex of more than 73 kDa containing six synthetic protein molecules. The structure shown in Fig. 4 is the first example of a *new class of protein complex*, one consisting of two

[‡]Solving a structure in the centrosymmetric space group $P2_1/n$ involves a mathematical inversion that averages the electron densities of the protein enantiomers, and thus obscures any potential differences that may exist.

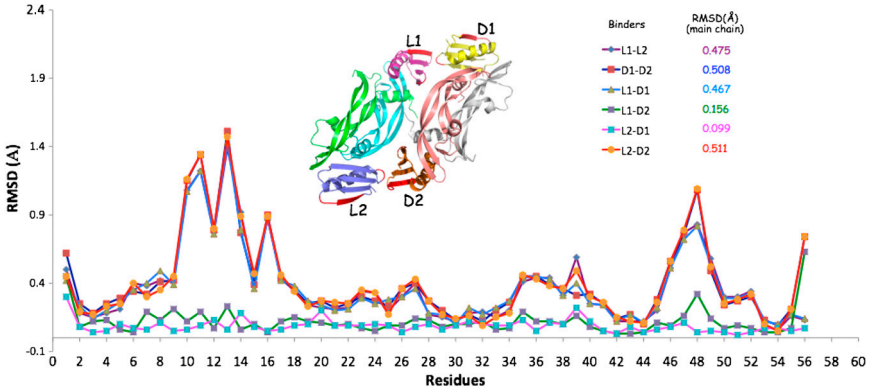


Fig. 6. Comparison of the structures of the four RFX001 protein molecules in the {VEGF-A+ *D*-RFX001} racemic complex solved in $P2_1$. The asymmetric unit (inset above) contained a total of six synthetic protein molecules. Backbone rmsd values for comparison of all possible pairs of RFX001 protein molecules are shown. The four RFX001 protein molecules are related as follows: Two enantiomeric pairs are related by inversion (L1-D2 or L2-D1), and two enantiomeric pairs are NOT related by inversion (L1-D1 or L2-D2). For the enantiomeric forms of RFX001 related by inversion, main chain rmsd value was just 0.1 Å. For the nonsymmetry related RFX001 pairs, the main chain rmsd value was 0.5 Å.

protein molecules of different amino acid sequence and with opposite chiralities that specifically interact with one another. Based on the principles of stereochemistry and associated nomenclature, we have termed this class a *'heterochiral protein complex'* (28).

We believe that this work illustrates the great potential of a systematic chemical protein synthesis plus protein phage display approach to the development of *D*-proteins as a previously unexplored class of molecules for antagonizing the action of natural protein molecules. Such *D*-protein antagonists may have significant advantages as human therapeutics.

Methods

Chemical Protein Synthesis. Total chemical synthesis of D-VEGF-A was performed as reported for the synthesis of biologically active VEGF-A (19). Final HPLC purification of the full length synthetic 102 residue polypeptide, folding with concomitant formation of disulfides gave synthetic D-VEGF protein with correct mass of 23,849.3 ± 0.7 Da by direct infusion electrospray ionization. Details are given in the *SI Appendix*.

Phage Display. The gene encoding the wild-type streptococcal protein GB1 domain sequence (22) was cloned into a display vector as N-terminal fusion to truncated protein 3 of M13 filamentous phage. A subset of 15 contiguous solvent exposed residues was chosen for randomization. Oligonucleotides with degenerate codon KHT (encoding Y, A, D, S, F, V) were used to construct a library of 8×10^9 transformants by previously described protocols (29, 30). Four rounds of selection against *D*-VEGFA were carried out following essentially the same protocols previously described (30). Because limited diversity (Y, A, D, S, F, V) was used in the initial library, we prepared affinity maturation libraries to allow all 20 amino acids to occur at each randomized position. A library of 1×10^9 transformants was obtained and selections were performed as described in the *SI Appendix*.

Racemic Protein Crystallography. The heterochiral protein complex was crystallized from the racemic mixture using 1:2 stoichiometry of protein:ligand. Diffraction data sets were collected to a resolution of 1.6 Å at the Advanced Photon Source, Argonne National Laboratory. The structures were solved by molecular replacement with the program PHASER (31) using the inverted and noninverted coordinates of previously reported X-ray structures of synthetic L-VEGF(8–109) (PDB code 3QTK) and GB1 (PDB code 2QMT) as search models. Full details are given in the *SI Appendix*.

ACKNOWLEDGMENTS. 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 Department of Energy, Office of Basic Energy Sciences, under contract no. DE-AC02-06CH11357. This work was supported by funds from the University of Chicago, the University of Toronto, and by Reflexion Pharmaceuticals.

1. Siegel JS (1998) Homochiral imperative of molecular evolution. *Chirality* 10:24–27.
2. Zawadzke LE, Berg JM (1993) The structure of a centrosymmetric protein crystal. *Proteins Struct Funct Genet* 16:301–305.
3. Pentelute BL, et al. (2008) X-ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers. *J Am Chem Soc* 130:9695–9701.
4. Mandal K, et al. (2009) Racemic crystallography of synthetic protein enantiomers used to determine the X-ray structure of plectasin by direct methods. *Protein Sci* 18:1146–1154.
5. Matthews BW (2009) Racemic crystallography—easy crystals and easy structures: What's not to like? *Protein Sci* 18:1135–1138.
6. Mortenson DE, Satyshur KA, Guzei IA, Forest KT, Gellman SH (2012) Quasiracemic crystallization as a tool to assess the accommodation of noncanonical residues in nativelike protein conformations. *J Am Chem Soc* 134:2473–2476.
7. Mandal K, et al. (2012) Design, total chemical synthesis, and X-ray structure of a protein having a novel polypeptide chain topology. *Angew Chem Int Ed* 51:1481–1486.
8. Pasteur L (1848) Mémoire sur la relation qui peut exister entre la forme cristalline et la composition chimique, et sur la cause de la polarisation rotatoire. *Compt Rend* 26:535–538.
9. Fischer E (1894) Einfluss der Konfiguration auf die Wirkung der Enzyme. *Ber Dtsch Chem Ges* 27:2985–2993.
10. de Lisle Milton RC, Milton SCF, Kent SBH (1992) Total chemical synthesis of a D-enzyme: The enantiomers of HIV-1 protease demonstrate reciprocal chiral substrate specificity. *Science* 256:1445–1448.
11. Schumacher TNM, et al. (1996) Identification of D-peptide ligands through mirror-image phage display. *Science* 271:1854–1857.
12. Welch BD, et al. (2010) Design of a potent D-peptide HIV-1 entry inhibitor with a strong barrier to resistance. *J Virol* 84:11235–11244.
13. Liu M, et al. (2010) D-peptide inhibitors of the p53-MDM2 interaction for targeted molecular therapy of malignant neoplasms. *Proc Nat Acad Sci USA* 107:14321–14326.
14. Funke SA, Willbold D (2009) Mirror image phage display—a method to generate D-peptide ligands for use in diagnostic or therapeutical applications. *Mol BioSyst* 5:783–786.
15. Dintzis HM, Symer DE, Dintzis RZ, Zawadzke LE, Berg JM (1993) A comparison of the Immunogenicity of a pair of enantiomeric proteins. *Proteins Struct Funct Genet* 16:306–308.
16. Ferrara N, Mass RD, Campa C, Kim R (2007) Targeting VEGF-A to treat cancer and age-related macular degeneration. *Annu Rev Med* 58:491–504.
17. Wiesmann C, et al. (1997) Crystal structure at 17 Å resolution of VEGF in complex with domain 2 of the Flt-1 receptor. *Cell* 91:695–704.
18. Wiesmann C, et al. (1998) Crystal structure of the complex between VEGF and a receptor-blocking peptide. *Biochemistry* 37:17765–17772.
19. Mandal K, Kent SBH (2011) Total chemical synthesis of biologically active vascular endothelial growth factor. *Angew Chem Int Ed* 50:8029–8033.
20. Dawson PE, Muir TW, Clark-Lewis I, Kent SBH (1994) Synthesis of proteins by native chemical ligation. *Science* 266:776–779.
21. Gallagher T, Alexander P, Bryan P, Gilliland GL (1994) Two crystal structures of the B1 immunoglobulin-binding domain of Streptococcal protein G and comparison with NMR. *Biochemistry* 33:4721–4729.
22. Gronenborn AM, et al. (1991) A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* 253:657–661.
23. Zoller F, Haberkorn U, Mier W (2011) Miniproteins as phage display-scaffolds for clinical applications. *Molecules* 16:2467–2485.
24. Fellouse FA, et al. (2005) Molecular recognition by a binary code. *J Mol Biol* 348:1153–1162.
25. Fellouse FA, et al. (2007) High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *J Mol Biol* 373:924–940.
26. Fairbrother WJ, et al. (1998) Novel peptides selected to bind vascular endothelial growth factor target the receptor-binding site. *Biochemistry* 37:17754–17764.
27. Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* 33:491–497.
28. Mislow K, Bickart P (1976/77) An epistemological note on chirality. *Isr J Chem* 15:1–6.
29. Kunkel TA, Roberts JD, Zakour RA (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol* 154:367–382.
30. Fellouse FA, Sidhu SD (2007) *Making and Using Antibodies*, eds GC Howard and MR Kaser (CRC Press, Boca Raton, FL), pp 157–180.
31. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658–674.
32. Mackay AL (1989) Crystal enigma. *Nature* 342:133.

Supporting Information

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-CH3Bzl), Glu(OcHex), His(Dnp), Lys(2Cl-Z), Ser(Bzl), Thr(Bzl), Tyr(2Br-Z). Boc-D-His(Dnp) was purchased from Bachem inc. Aminomethyl-resin (1.0 mmol/gram) was prepared from Biobeads S-X1(BioRad, California)[36]. Boc-L-Ala-OCH2-phenylacetic acid and Boc-L-Glu(OcHex)-OCH_2-phenylacetic acid were purchased from NeoMPS, Strasbourg, France. Boc-D-Asp(OcHex)-OCH_2-phenylacetic acid and Boc-D-Glu(OcHex)-OCH_2-phenylacetic acid were prepared following the literature procedure[37]. *N*,*N*-Diisopropylethylamine (DIEA) was obtained from Applied Biosystems. *N*,*N*-Dimethylformamide (DMF), dichloromethane, diethyl ether, guanidine hydrochloride and HPLCgrade acetonitrile, 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 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 or an Agilent C-3 (3.5 μm, 300 Å) 4.6 x150 mm silica column at a flow rate of 1 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 H2O; 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 °C 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.

1. Total chemical synthesis of D-VEGF-A

1.1 Peptide synthesis/purification/LCMS characterization. Peptide segments were synthesized using manual 'in situ neutralization' Boc chemistry protocols for stepwise SPPS [38]. After removal of the N-αBoc 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 C-terminal peptide segment and peptide thioester segments of the D-VEGF-A were prepared following the same procedure as described for the L-VEGF-A synthesis[19].

Preparative HPLC purification gave the following amounts of each peptide segment: 70 mg (12 μmol, 6% yield from 0.2 mmol peptide resin) of the C-terminal peptide [Cys50-Asp102]-COOH, observed mass. 6004.5 ± 0.2 Da, calc. 6004.9 Da (average isotope composition); 125 mg (53 μmol, 27% yield from 0.2 mmol starting peptide resin) of the peptide segment [Gly1-Tyr18]-α$COSCH_2CH_2SO_3H$, observed mass 2361.5 ± 0.2 Da, calc. 2361.4 Da (av isotopes); and, 42 mg (11 μmol, 11% yield from 0.1 mmol starting peptide resin) of the pure thioester peptide segment [Thz19-Arg49]-α$COSCH_2CH_2SO_3H$, observed mass 3862.3 ± 0.1 Da, calc. 3862.3 Da (av isotopes).

1.2 One-pot native chemical ligations. The strategy for the chemical synthesis of D-VEGF-A is shown in Scheme S1. Native chemical ligation was performed by dissolving peptide **3** [Cys50-Asp102]-COOH (26 mg, 4.3 μmol, 1.72 mM) and peptide **2** [Thz19-Arg49]-α$COSCH_2CH_2SO_3H$ (15 mg, 3.88 μmol, 1.55 mM) in ligation buffer (2.5 mL) containing 6 M guanidine hydrochloride, 100 mM Na_2HPO_4, 100 mM MPAA and 50 mM TCEP hydrochloride at pH 7 to give ligation product **4** (Figure 1A-II). After conversion of the Thz-peptide product to the Cys-peptide **5** by addition of 60 mM methoxylamine hydrochloride at pH 4.0 (Figure 1A-III), peptide **1** [Gly1-Tyr18]-α$COSCH_2CH_2SO_3H$ (12 mg, 5.16 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** (Figure 1A-V) (after purification: 25 mg, 2.1 μmol, 54% based on limiting peptide segment **2**), observed mass 11,932.7±0.3 Da (average of the six most abundant charge states), calc. 11932.54 Da (av isotopes). LCMS data for the D-VEGF(8-109) polypeptide is shown in Figure 1B.

1.3 Folding/purification/HPLC and MS characterization. The purified reduced polypeptide [Gly1-Asp102]-COOH (25 mg, 2 μmol) was dissolved in degassed folding buffer (50 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 and was essentially complete in 3 days (Figure S1A); the folded product was purified by reverse phase HPLC to give 9 mg (0.38 μmol, 38%) of the desired synthetic D-VEGF protein. Analytical HPLC showed a single sharp peak eluting with 2.8 minutes earlier retention time compared with the unfolded protein 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. HPLC and direct infusion ESMS data for the folded, homodimeric synthetic *D*-VEGF-A protein is shown in Figure 1C; observed mass: 23849.3±0.7 (average of the two most abundant charge states); calculated mass: 23849.1 Da (average isotopes).

1.4 Circular Dichroism (CD). The circular dichroism spectrum of the synthetic D-VEGF-A protein was recorded on an AVIV-202 instrument. An aqueous solution of synthetic D-VEGF-A (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 is shown in Figure S1B.

2. Phage Display of GB1

2.1 Cloning

The gene encoding the wild-type *streptococcal* protein G B1 domain sequence[22] was ordered for synthesis from Genscript USA Inc. This sequence was synthesized with NcoI and XbaI restriction sites at 5' and 3' respectively and cloned into a display vector as N-terminal fusion to truncated protein 3 of M13 filamentous phage. The complete ORF of the display construct is shown below.

```
1         10         20         30         40         50
|         |          |          |          |          |
MKKNIAFLLASMFVFSIATNAYASMGDYKDDDDKGGSTYKLILNGKTLKG
ETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTEGGSDKTHT
CGRPSGSGDFDYEKMANANKGAMTENADENALQSDAKGKLDSVATDYGAA
IDGFIGDVSGLANGNGATGDFAGSNSQMAQVGDGDNSPLMNNFRQYLPSL
PQSVECRPFVFSAGKPYEFSIDCDKINLFRGVFAFLLYVATFMYVFSTFA
NILRNKES
Red-    StII secretion signal
Green-  FLAG tag
Blue-   GB1
Brown-  Hinge and dimerization sequence
Orange- truncated M13 protein3
```

The salient features include a N-terminal FLAG tag and C-terminal dimerization sequence derived for human IgG1. The fusion protein is expressed under the control of *ptac* promoter.

2.2 Library design and construction

The solvent accessible surface area (SASA) for each residue in the structure PDB 3GB1 was estimated using the GetArea tool. The tool also calculates the ratio of

SASA in structure compared to SASA in a random coil. An arbitrary cutoff ratio of 0.4 was used to differentiate the solvent accessible residues from buried residues. The solvent exposed residues are shown in bold and red below

TY**K**LI**L**N**G**K**TLKGE**T**T**T**E**A**V**DAAT**A**E**K**V**F**K**QY**A**ND**N**G**V**D**G**E**W**T**Y**DDATK**TFTV**T**E

A subset of 15 contiguous solvent exposed residues (based on GB1- Fc binding interface) was chosen for randomization. Preliminary experiments (unpublished results) demonstrate that GB1 tolerates insertions of one or two amino acids in the loop (residues 37-41). Therefore these insertions were allowed in the randomization scheme shown below.
Oligonucleotides with degenerate codon KHT (encoding Y,A,D,S,F,V) were ordered from Integrated DNA Technologies Inc.,USA. The sequences are shown below These oligonucleotides include a 15bp homology region (underlined) at 5' and 3'ends to facilitate site-directed mutagenesis. The libraries were constructed by previously described protocols[33,34] and library of 8 x 10^9 transformants was obtained.

TYKLILNGKTLKGETTTEAV**XXXX**A**XX**VF**XX**YA**XX**N**XXZXXX**WTYDDATKTFTVTE

Where X encodes a randomization of Y, A, D, S, F and V
Z stands for 0-2 insertions with a randomization of Y, A, D, S, F and V

5'-ACGACCGAAGCAGTG KHT KHT KHT KHT GCA KHT KHT GTT TTC KHT KHT TAC GCC KHT KHT AAT KHT KHT KHT KHT KHT TGGACCTACGATGAT-3' (no insertion)

5'-ACGACCGAAGCAGTG KHT KHT KHT KHT GCA KHT KHT GTT TTC KHT KHT TAC GCC KHT KHT AAT KHT KHT KHT KHT KHT KHT TGGACCTACGATGAT-3' (1aa insertion)

5'-ACGACCGAAGCAGTG KHT KHT KHT KHT GCA KHT KHT GTT TTC KHT KHT TAC GCC KHT KHT AAT KHT KHT KHT KHT KHT KHT KHT TGGACCTACGATGAT-3' (2aa insertion)

2.3 Phage panning

The selection procedure is essentially the same as mentioned in previous protocols [34]with some minor changes.

Briefly, *D*-VEGFA was adsorbed on 96 well Maxisorp plate (Thermo Fisher) at concentration of 5μg/ml. For Round 1 and Round 2 of panning, the library/phage pool was preincubated on Glutathione S-transferase (GST) coated Maxisorp plates (10μg/ml) to clear non-specific binders prior to binding to *D*-VEGFA. Bound phage was washed and eluted with 100mM HCl and neutralized with 1M Tris-HCl, pH 11. The eluted phage was amplified in *e.coli* XL1-blue cells with use of M13K07 helper phage (New England Biolabs). For Rounds 3 and 4, the phage pools were pre-incubated with 0.2 mg/ml GST in solution for 1 hr before transfer to selection wells. The enrichment of the phage was monitored by phage pool ELISA using anti-M13-HRP conjugate (GE Healthcare) antibody and the data is shown in Figure S2A.

Then, 48 single clones from 3rd round and 4th round selection pools were analyzed by phage ELISA as described previously[34]. The binding specificity is shown in Figure S2B.

2.4 Binding epitope mapping of RFX.T1.4

Fairbrother et. al.[26] developed antagonistic peptides that bind to VEGF-A at the same site as VEGFR1. The peptide v114 (VEPNCDIHVMWEWECFERL-NH_2) was ordered in both *D* and *L* forms from CSBio, Inc, USA. The following assay was used to determine whether the phage bound to the same epitope as v114.

D-VEGFA coated on Maxisorp plate was preblocked with saturating concentration of v114 (100μM) for 30min at room temperature. In a control well, *D*-VEGFA was left unblocked with an equal volume of buffer. RFX.T1.4 phage (~1 x10^{10} cfu/ml) was spiked into both the wells and allowed to bind for 1hr. The binding was estimated after washing using anti-M13-HRP conjugate antibody. Significant reduction of phage binding in the presence of peptide indicated that RFX.T1.4 binds to the same epitope as v114.

2.5 Affinity Maturation of RFX.T1.4

Because limited diversity (Y, A, D, S, F, V) was used in the initial library, we prepared affinity maturation libraries to allow all 20 aa to occur at each randomized position. The randomization scheme is shown in Figure 2C. The following oligonucleotides were used for site-directed mutagenesis.

Library 1
5′-ACGACCGAAGCAGTG NNK NNK NNK NNK GCA TTT GAT GTT TTC TTT GTT TAC GCC GCT TCT AAT TTT TCT GAT TTT GAT GAT TGGACCTACGATGAT-3′
Library 2
5′-ACGACCGAAGCAGTG GAT GTT TAT GAT GCA NNK NNK GTT TTC NNK NNK TAC GCC GCT TCT AAT TTT TCT GAT TTT GAT GAT TGGACCTACGATGAT-3′
Library 3
5′-ACGACCGAAGCAGTG GAT GTT TAT GAT GCA TTT GAT GTT TTC TTT GTT TAC GCC NNK NNK AAT NNK NNK GAT TTT GAT GAT TGGACCTACGATGAT-3′
Library 4
5′-ACGACCGAAGCAGTG GAT GTT TAT GAT GCA TTT GAT GTT TTC TTT GTT TAC GCC GCT TCT AAT TTT TCT NNK NNK NNK NNK TGGACCTACGATGAT-3′

A library of 1 x10^9 transformants was obtained and selections were performed as described in the previous sections.

3. Total chemical synthesis of L- and D-RFX001

Amino acid sequence of the binder RFX001: TYKLILNGKT LKGETTTEAV DVFDAFDVFF VYAASNFSDF DDWTYDDATK TFTVTE
L- and *D*-RFX001 peptides were synthesized using manual 'in situ neutralization' Boc chemistry protocols for stepwise SPPS on Boc-Glu(OcHex)-OCH_2-Pam-resin of

appropriate chirality. After removal of the N-αBoc 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 mM NH_4OAc (pH 7.6) and purified by revesed phase HPLC (Figure S3).

4. Surface Plasmon resonance. D- and L-VEGF were bound in separate flow cells of a GLM sensor chip in a ProteOn™ XPR36 Protein Interaction Array System (Bio-Rad) with 4000 RU of L-VEGF or 3500 RU of D-VEGF using an amine coupling kit. Sensorgrams were generated by injecting several concentrations of protein ligand over the chip (Figure S4). The interactions were analyzed by local and global kinetic analysis using nonlinear regression fits to a 1:1 Langmuir binding isotherm. The kinetic data are shown in Table S2. Considerable variation in the K_D values is observed between individual concentrations. Under the conditions used for this experiment the observed global dissociation constant values, K_D, were 53 nM for the *L*-RFX001 and 87 nM for the *D*-RFX001. The difference in observed global K_D values arises from experimental uncertainties in the determination of dissociation constants by SPR, which can be ± ~30%[39] ; here, taking the mean of the global K_D values for *L*-RFX001 and *D*-RFX001, gives an observed K_D of 70 ± 17 nM. This should be contrasted with K_D ~12 nM (Table S2.) found for recombinant RFX001 which protein molecule contains additional amino acid residues at the N- and C-terminals.

5. Octet assay. Octet assay was measured on a ForteBio Octet instrument. All solutions were prepared in diluent containing 1X PBS w/ 0.1% BSA, 0.5% Proclin. Inhibitor-containing solutions were prepared by incubating 0-2,000 nM of inhibitor, with or without 5 nM hVEGF165 (Peprotech) for 1 hour at ambient temperature. A VEGR1 sensor was prepared by incubating ForteBio Octet anti-hIgGFc sensors for 700 seconds with 25 nM rhVEGFR1/hIgGFc (R & D Systems) followed by a 400 second incubation with 2000 nM human IgG (Jackson Immunoresearch). The VEGFR1 sensors were then incubated with diluent for 120 seconds to establish a baseline wavelength reading. Then, VEGFR1 sensors were incubated with the inhibitor-containing solutions for 600 seconds at 30 °C with mixing, and the wavelength shift was measured. Wavelength shifts for VEGF165 binding to the VEGFR1 sensor at each inhibitor concentration were normalized by subtracting the wavelength shift of the inhibitor alone at each respective concentration. Percent inhibition was calculated using the wavelength shift for VEGF165 binding in the absence of inhibitor and the wavelength shift for VEGF165 binding with a non-reactive sensor as 0% and 100% inhibition, respectively.

6. X-ray crystal structure determination.

6.1 General method

All crystals were grown at 19° C by hanging drop vapor diffusion technique using index screen from Hampton Research (HR2-144) or Qiagen Pro-complex suite. 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]. Diffraction data sets were 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, integrated, scaled and merged with HKL2000[40]. The structures were solved by molecular replacement with the program PHASER[35] and refined with PHENIX.REFINE[41] 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[42] using both 2Fo-Fc and Fo-Fc difference maps. At the later stages of refinement, the models were refined using TLS parameters in Phenix.

6.2 D-VEGF-A

Synthetic D-VEGF-A was crystallized using 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). A complete data set to 1.90 Å was collected from a single crystal at 100 K using 0.97 Å wavelength synchrotron radiation at the Argonne National Laboratory. The molecules crystallized in the C2 space group. Cell content analysis[30] revealed that there are three D-VEGF protein molecules in the asymmetric unit. The structure was solved by molecular replacement using the inverted coordinates of the previously reported X-ray structure of the synthetic L-VEGF(8-109) (PDB accession code 3QTK) as a search model. The resulting synthetic D-VEGF model was then refined with PHENIX.REFINE and after each refinement step the model was visually inspected in COOT using both 2Fo-Fc and Fo-Fc difference maps. The final model gave R-factor (R-free) of 17.6 % (21.4%). X-ray crystal diffraction data and refinement statistics are listed in Table S3.

6.3 Heterochiral complex of *L*-VEGF + *D*-RFX001

Synthetic *L*-VEGF-A (84 μM) and synthetic *D*-RFX001 (168 μM) were dissolved in 0.01 M HEPES, 50 mM NaCl, pH 8.4 and screened against 96 index conditions of Quiagen Pro-complex suite using 1 μL of protein solution and 1 μL of reservoir solution placed over 1 mL of reservoir solution and incubated at 19° C. However, only two conditions produced micro crystals. Optimization of these two conditions did not lead to the formation of diffraction-quality crystals.

In contrast, crystallization of the racemic mixture of the heterochiral complex, containing equal proportion of *L*-VEGF-A and *D*-VEGF-A and equal proportion of *D*-RFX001 and *L*-RFX001, using 1:2 stoichiometry of protein:ligand {i.e. *L*-VEGF-A (84 μM) + *D*-RFX001 (168 μM) + *D*-VEGF-A (84 μM) + *L*-RFX001 (168 μM)}, screened against 96 index conditions of Quiagen Pro-complex suite,

produced microcrystals in 7 out of 96 conditions. One of the seven conditions was optimized to produce a diffraction-quality single crystal. The crystal that diffracted to 1.6 Å was grown from 0.1 M MgCl2, 0.1 M HEPES, 12.5% PEG3350.

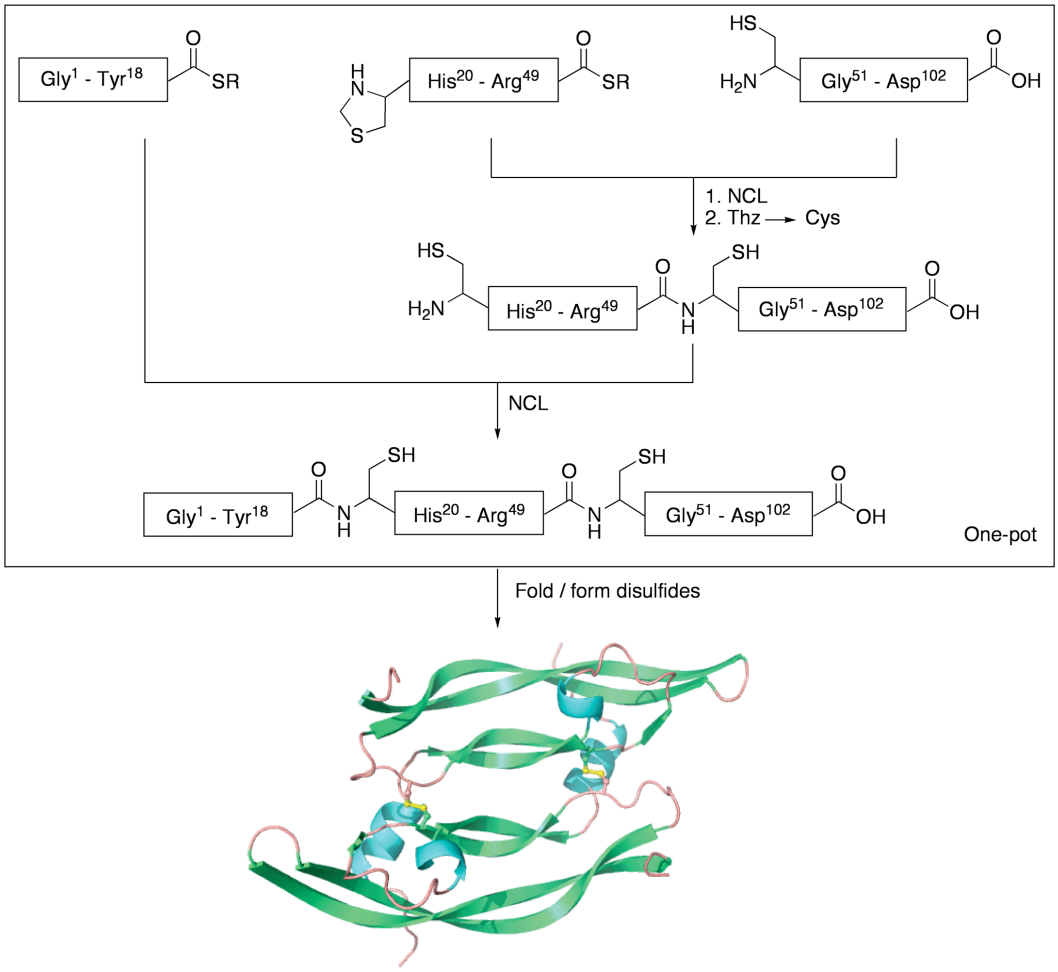
The diffraction data were then processed by HKL 2000 and the diffraction intensity statistics revealed that the protein complex had crystallized in the centrosymmetric space group $P2_1/n$. Matthews cell content analysis[30] suggested that there were one VEGF-A protein molecule and two RFX001 protein molecules in the $P2_1/n$ asymmetric unit. In order to treat protein enantiomers as independent molecules in the crystal rather than related by crystallographic symmetry operation, the structure was first solved in $P2_1$ by molecular replacement using inverted and non-inverted coordinates of the previously reported crystal structures of VEGF (PDB 3QTK) and GB1 (PDB ID 2QMT) as search models. The resulting racemic protein complex model was then refined by PHENIX.REFINE and after each refinement step the model was visually inspected in COOT using both 2Fo-Fc and Fo-Fc difference maps. The final model had R-factor (R-free) of 21.4% (24.9%).

The structure was also solved in $P2_1/n$ space group by molecular replacement using the coordinates of the complex of one D-VEGF molecule and two L-binders models obtained from the $P2_1$ solution. After refinement the final model in $P2_1/n$ space group had R-factor (R-free) of 26.3% (29.4%). X-ray crystal diffraction data and refinement statistics of the racemic protein complex are listed in Table S4.

A Sequence

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



Scheme S1. A. Target amino acid sequence (single letter code; lower case denotes D-amino acid); B. Synthetic strategy. Three unprotected synthetic peptides were condensed by native chemical ligation, and the 102 residue synthetic polypeptide was folded, to give the homodimeric D-VEGF-A protein molecule.

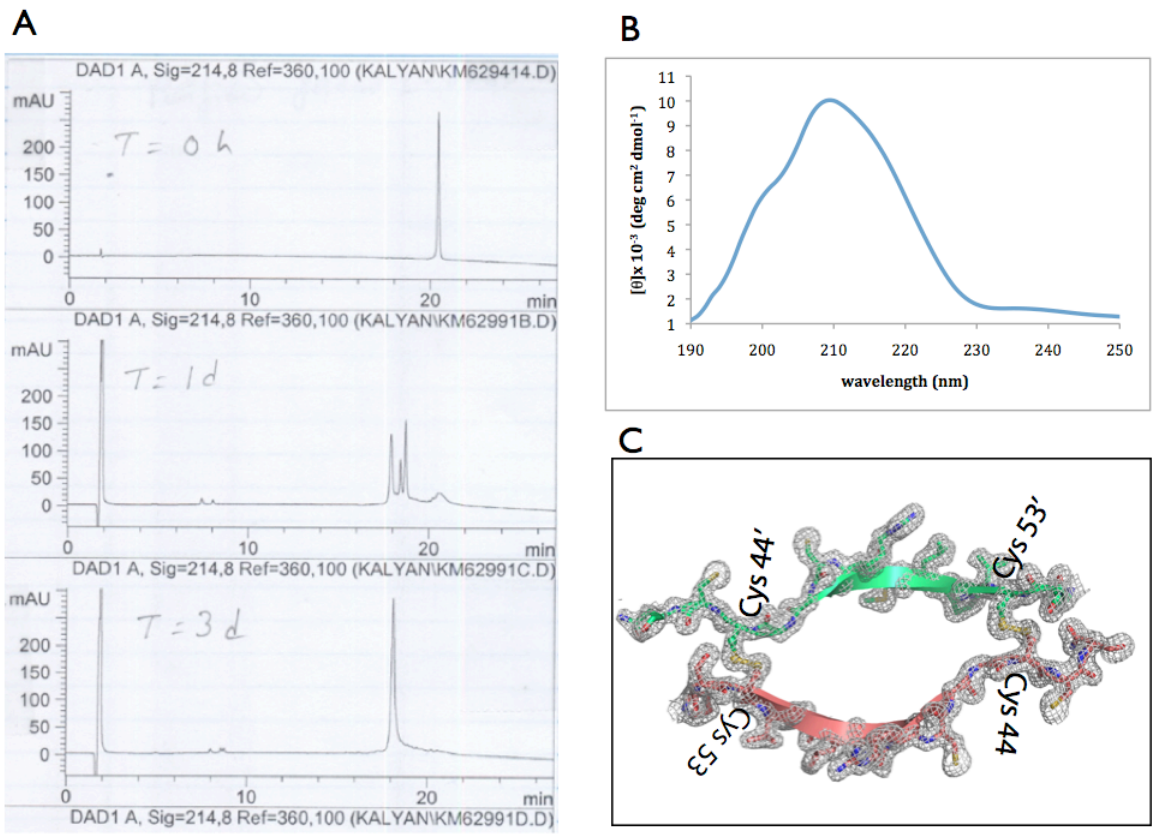


Figure S1 A) LC-monitoring of the oxidative folding of *D*-VEGF-A. Note the earlier retention time shift of the folded protein molecule compared to the unfolded polypeptide. B) Circular dichroism spectra of folded *D*-VEGF-A. CD spectrum was recorded at room temperature using 1.5 M (0.36 mg/mL) of the folded protein solution in water. CD cuvette path length = 0.1 cm. Total number of scans = 6. C) Sigma A-weighted 2Fo-Fc electron density map contoured at 1σ of the *D*-VEGF-A protein molecule in the region encompassing the disulfide bridges between the two identical monomers.

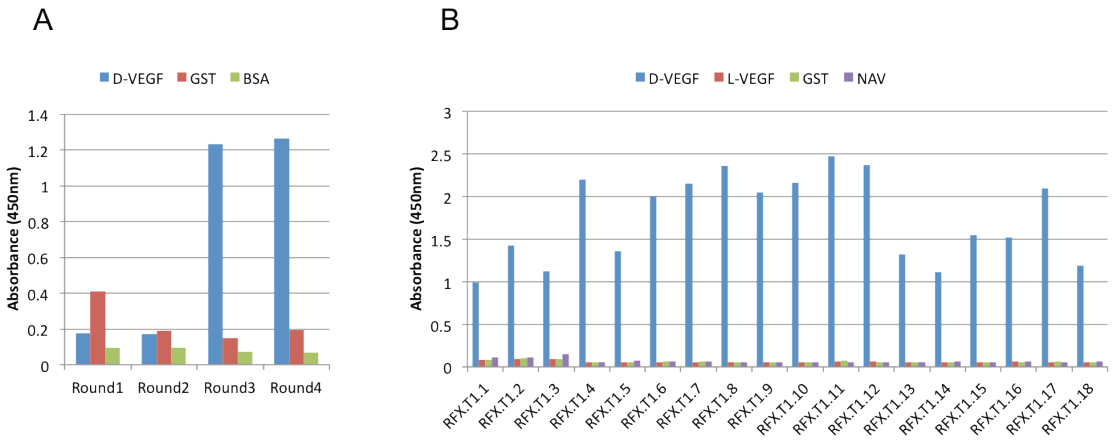


Figure S2. A) Phage pool ELISA showing enrichment with each round of panning. B) Positive clones bind specifically to *D*-VEGF and do not bind to *L*-VEGF, GST or Neutravidin (NAV).

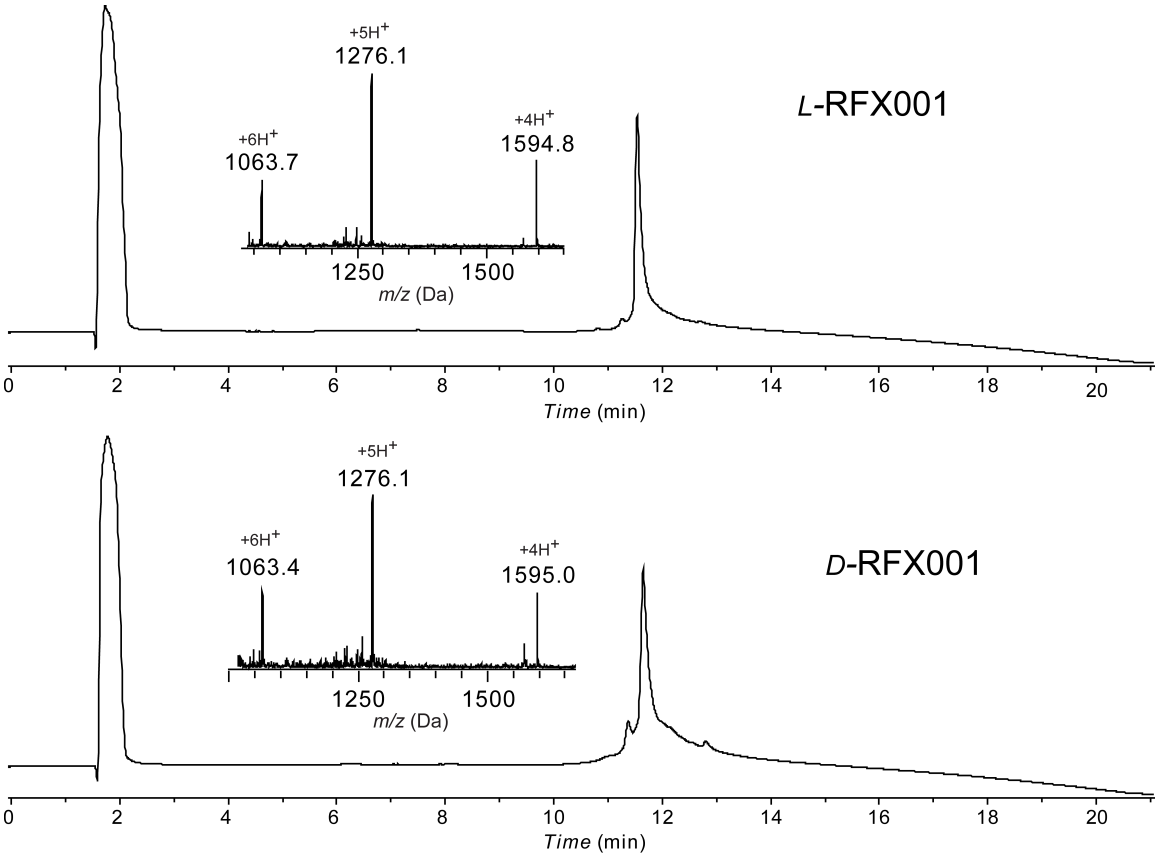


Figure S3 Analytical LC-MS data for the chemically synthesized *L*-RFX001 (top) and *D*-RFX001 (bottom). Analytical HPLC profiles (λ = 214 nm) are shown, together with online electrospray ionization mass spectrometry (LC-MS) data. Mass calculated: 6375.94 Da; Mass observed: 6375.4 ± 0.2 Da (*L*-RFX001) and 6375.3 ± 0.8 Da (*D*-RFX001). Analytical HPLC was performed using a linear gradient (25-85%) of buffer B in buffer A over 15 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 °C (flow rate = 1 mL/min). Amino acid sequence of RFX001: TYKLILNGKT LKGETTTEAV DVFDAFDVFF VYAASNFSDF DDWTYDDATK TFTVTE

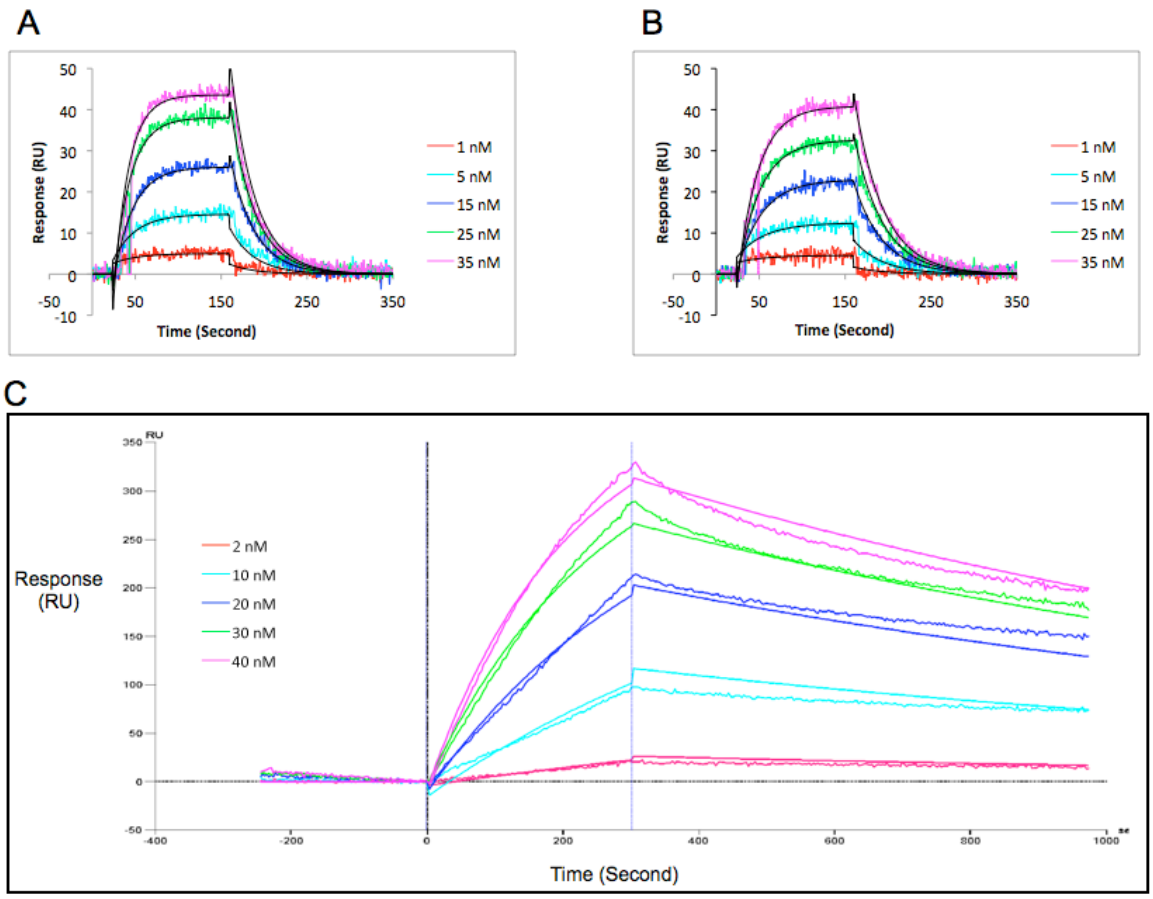


Figure S4 Surface plasmon resonance sensorgrams with best analytical fit of (A) *L*-RFX001 against *D*-VEGF-A, (B) *D*-RFX001against *L*-VEGF-A and (C) expressed *L*-RFX001 (containing His_{10}-tag) against *D*-VEGF-A. Sensorgrams were generated by injecting several concentrations (inset) of protein ligand over a chip coupled with 4000 RU of *L*-VEGF-A or 3500 RU of *L*-VEGF-A. Derived global K_D values are given in Table S2.

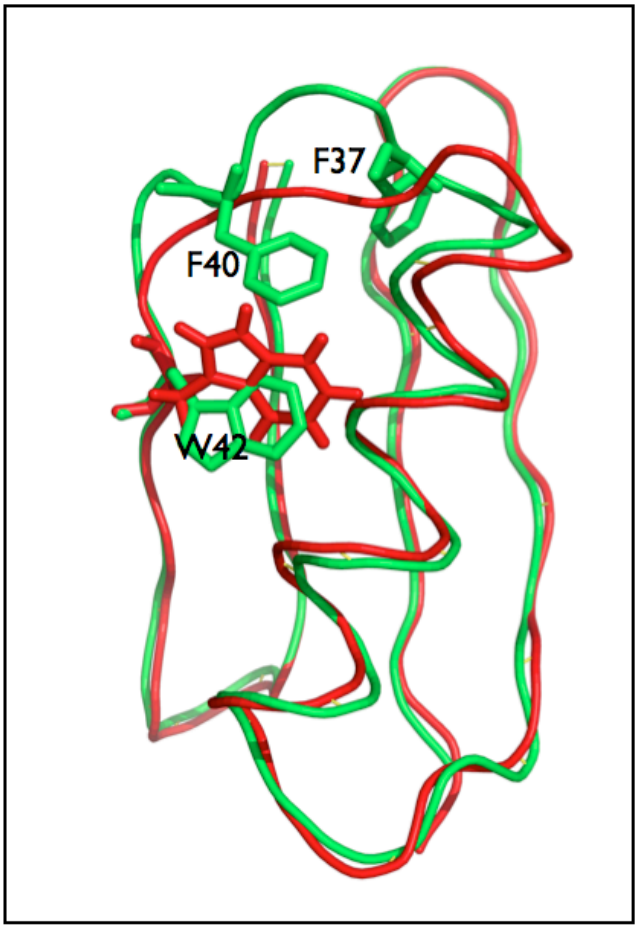


Figure S5 Comparison of RFX001 (red) and wild-type GB1 (green) structures. The major differences in the alignment of Cα atoms are seen in the loop2 region with the insertion mutation. Mutations Phe37 and Phe40 of RFX001 insert into the core causing a shift in Trp42 residue.

Table S1 Amino acid sequences of *D*-VEGF binding clones (residues 18-43, non-randomized residues not shown)

Clone	Sequence																											
Residue #	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	41a	41b	42	43
GB1-wt	E	A	V	D	A	A	T	A	E	K	V	F	K	Q	Y	A	N	D	N	G	V	D	G	E			W	T
RFX.T1.1	E	A	V	**D**	**V**	**F**	**Y**	A	**F**	**D**	V	F	**F**	**D**	Y	A	**S**	**D**	N	**F**	**A**	**D**	**V**	**A**	**D**		W	T
RFX.T1.2	E	A	V	**D**	**F**	**Y**	**S**	A	**F**	**D**	V	F	**F**	**Y**	Y	A	**S**	**S**	N	**F**	**Y**	**D**	**V**	**A**	**D**		W	T
RFX.T1.3	E	A	V	**D**	**V**	**Y**	**S**	A	**V**	**D**	V	F	**Y**	**D**	Y	A	**A**	**A**	N	**F**	**A**	**S**	**F**	**S**	**D**		W	T
RFX.T1.4	E	A	V	**D**	**V**	**Y**	**D**	A	**F**	**D**	V	F	**F**	**V**	Y	A	**A**	**S**	N	**F**	**S**	**D**	**F**	**D**	**D**		W	T
RFX.T1.5	E	A	V	**D**	**F**	**F**	**S**	A	**F**	**D**	V	F	**Y**	**A**	Y	A	**D**	**S**	N	**F**	**D**	**F**	**Y**	**D**	**D**		W	T
RFX.T1.6	E	A	V	**D**	**F**	**Y**	**A**	A	**F**	**S**	V	F	**F**	**D**	Y	A	**A**	**F**	N	**F**	**Y**	**D**	**V**	**D**	**D**		W	T
RFX.T1.7	E	A	V	**S**	**Y**	**Y**	**D**	A	**F**	**D**	V	F	**F**	**A**	Y	A	**S**	**S**	N	**F**	**D**	**F**	**F**	**D**	**D**		W	T
RFX.T1.8	E	A	V	**D**	**F**	**F**	**A**	A	**F**	**D**	V	F	**Y**	**S**	Y	A	**S**	**F**	N	**F**	**A**	**F**	**F**	**D**	**D**		W	T
RFX.T1.9	E	A	V	**D**	**V**	**Y**	**D**	A	**F**	**D**	V	F	**Y**	**Y**	Y	A	**A**	**A**	N	**Y**	**A**	**D**	**F**	**D**	**D**		W	T
RFX.T1.10	E	A	V	**D**	**F**	**F**	**D**	A	**F**	**D**	V	F	**F**	**D**	Y	A	**V**	**A**	N	**F**	**A**	**D**	**V**	**D**	**D**		W	T
RFX.T1.11	E	A	V	**D**	**V**	**Y**	**S**	A	**F**	**D**	V	F	**Y**	**D**	Y	A	**V**	**A**	N	**F**	**V**	**D**	**F**	**S**	**D**		W	T
RFX.T1.12	E	A	V	**D**	**F**	**F**	**A**	A	**F**	**A**	V	F	**D**	**S**	Y	A	**S**	**S**	N	**F**	**F**	**S**	**V**	**A**	**D**		W	T
RFX.T1.13	E	A	V	**D**	**F**	**Y**	**D**	A	**F**	**D**	V	F	**Y**	**D**	Y	A	**S**	**S**	N	**V**	**S**	**V**	**F**	**D**	**D**		W	T
RFX.T1.14	E	A	V	**D**	**V**	**F**	**D**	A	**F**	**D**	V	F	**A**	**V**	Y	A	**F**	**D**	N	**F**	**V**	**Y**	**V**	**D**	**D**		W	T
RFX.T1.15	E	A	V	**D**	**F**	**F**	**D**	A	**F**	**S**	V	F	**D**	**A**	Y	A	**S**	**S**	N	**Y**	**V**	**V**	**V**	**D**	**D**		W	T
RFX.T1.16	E	A	V	**D**	**V**	**F**	**D**	A	**F**	**D**	V	F	**F**	**Y**	Y	A	**A**	**A**	N	**F**	**D**	**V**	**Y**	**D**	**D**		W	T
RFX.T1.17	E	A	V	**D**	**F**	**Y**	**S**	A	**F**	**D**	V	F	**F**	**S**	Y	A	**A**	**S**	N	**F**	**F**	**V**	**F**	**D**	**D**		W	T
RFX.T1.18	E	A	V	**D**	**F**	**Y**	**D**	A	**F**	**D**	V	F	**F**	**V**	Y	A	**V**	**D**	N	**Y**	**F**	**F**	**D**	**V**	**D**	**D**	W	T

Table S2 Kinetic analysis of the VEGF-RFX001 interaction

Binder	Target protein	concentration	ka	kd	KD	Rmax	Chi2	kinetic model
		(M)	(1/Ms)	(1/s)	(M)	(RU)	(RU)	
L-RFX001	D-VEGF							
(Global fit)			7.39E+05	3.89E-02	5.26E-08	1.30E+02	2.13	Langmuir
(Local fit)		1.00E-09	1.16E+07	4.90E-02	4.22E-09	25.2		
		5.00E-09	3.61E+06	3.14E-02	8.70E-09	39.2		
		1.50E-08	5.32E+05	3.98E-02	7.48E-08	178		
		2.50E-08	4.77E+05	4.14E-02	8.67E-08	194		
		3.50E-08	3.97E+05	4.11E-02	1.03E-07	199		
D-RFX001	L-VEGF							
(Global fit)			3.80E+05	3.31E-02	8.69E-08	1.53E+02	1.63	Langmuir
(Local fit)		1.00E-09	1.13E+03	1.73E-01	1.53E-04	1.00E+06		
		5.00E-09	3.69E+06	4.33E-02	1.17E-08	37.7		
		1.50E-08	8.57E+05	3.24E-02	3.78E-08	78		
		2.50E-08	4.54E+05	3.29E-02	7.25E-08	130		
		3.50E-08	3.39E+05	3.27E-02	9.65E-08	167		
L-RFX001 (Expressed)	D-VEGF							
(Global fit)			1.05E+05	6.78E-04	6.46E-09	4.74E+02	70.3	Langmuir
(Local fit)		2.00E-09	35.2	3.99E-04		9.99E+05		
		1.00E-08	33.4	4.16E-04		1.00E+06		
		2.00E-08	37.2	5.29E-04		1.00E+06		
		3.00E-08	5.34E+04	6.95E-04	1.30E-08	792		
		4.00E-08	7.22E+04	7.76E-04	1.07E-08	595		

Table S3. Data collection and refinement statistics of D-VEGF-A.

Data collection*	
Synchrotron beam line	APS 24-ID-E
Space group	C_2
Wavelength (Å)	0.97916
Cell dimensions	
a, *b*, *c* (Å)	231.4, 44.0, 73.5
α, β, γ (°)	90.0, 100.3, 90.0
Mol/asymmetric unit	3 (dimer)
Mol/unit cell	12 (dimer)
Resolution (Å)	50.0 - 1.90 (1.93 – 1.90)
R_{merge}[a]	0.071 (0.625)
$I/\sigma I$	13.2 (2.3)
Completeness (%)	98.5 (99.4)
Redundancy	2.5 (2.5)
Refinement*	
Resolution (Å)	32.5 - 1.89 (1.92 – 1.89)
No. of reflections	57480
R_{work}(%)/ R_{free}(%)[b]	17.6 / 21.4
No. atoms	
Non-solvent	4849
Solvent	403
Average B-factor (Å^2)	25
R.m.s deviations	
Bond lengths (Å)	0.007
Bond angles (º)	1.032

*Highest resolution shell shown in parenthesis. [a] $R_{merge} = \sum | I-\langle I \rangle | / \sum I$. [b] R_{free} calculated using 5% of the data.

Table S4. Data collection and refinement statistics of the racemic {VEGF+RFX001} complex.

Data collection*		
Synchrotron beam line	APS 24-ID-E	
Wavelength (Å)	0.97916	
Cell dimensions		
a, *b*, *c* (Å)	57.2, 88.3, 77.9	
α, β, γ (°)	90.0, 100.0, 90.0	
Resolution (Å)	50.0 - 1.6 (1.66 – 1.60)	
R_{merge}[a]	0.046 (0.365)	
$I/\sigma I$	17.8 (1.8)	
Completeness (%)	97.8 (96.2)	
Redundancy	2.1 (2.0)	
Refinement*		
Space group	$P2_1$	$P2_1/n$
Mol/asymmetric unit	6 {1 × *L*-VEGF-A + 1 × *D*-VEGF-A + 2 × *L*-RFX001 + 2 × *D*-RFX001}	3 {1 × *L*-VEGF-A + 2 × *D*-RFX001}
Mol/unit cell	12	12
Resolution (Å)	34.6 -1.6 (1.62 – 1.60)	32.2 – 1.6 (1.62 – 1.60)
No. of reflections	97839	96772
R_{work}(%)/ R_{free}(%)[b]	21.4/24.9	26.3/29.4
No. of atoms		
Non-solvent	4953	2543
Solvent	1127	579
Average B-factor (Å^2)	15.6	48.7
R.m.s deviations		
Bond lengths (Å)	0.006	0.006
Bond angles (º)	0.986	0.976

*Highest resolution shell shown in parenthesis. [a] $R_{merge} = \sum | I-<I> | / \sum I$. [b] R_{free} calculated using 5% of the data.

References

36. Mitchell, A. R., Kent, S. B. H., Engelhard, M. & Merrifield, R. B. A new synthetic route to tert-butyloxycarbonylaminoacyl-4-(oxymethyl)phenylacetamidomethyl-resin, an improved support for solid-phase peptide synthesis. *J. Org. Chem.* **43**, 2845–2852 (1978).

37. Tam, J. P., Kent, S. B. H., Wong, T. W. & Merrifield, R. B. Improved synthesis of 4-(Boc-aminoacyloxymethyl)-phenylacetic acids for use in solid-phase peptide-synthesis. *Synthesis* 955–957 (1979).

38. Schnölzer, M., Alewood, P., Jones, A., Alewood, D. & Kent, S. B. H. In situ neutralization in Boc-chemistry solid phase peptide synthesis - rapid, high yield assembly of difficult sequences. *Int. J. Pept. Res. Ther.* **13**, 31-44 (2007).

39. O'Shannessy, D. J., Brigham-Burke, M., Sonenson, K. K., Hensley, P. & Brooks, I. Determination of rate and equilibrium binding constants for macromolecular interactions using surface plasmon resonance: use of nonlinear least squares analysis methods. *Anal. Biochem.* **212**, 457-468 (1993).

40. Otwinowski, Z. & Minor, W. in *Methods in Enzymology,* Charles W. Carter Jr., Ed.

    (Academic Press, New York, 1997), vol. 276, pp. 307–326.

41. Adams P. D. *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221 (2010).

42. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-2132 (2004).