JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

COMMUNICATION

pubs.acs.org/JACS

Computational Design of the Sequence and Structure of a Protein-Binding Peptide

Deanne W. Sammond,[†,§,⊥] Dustin E. Bosch,[‡,⊥] Glenn L. Butterfoss,[†,∥] Carrie Purbeck,[†] Mischa Machius,[‡] David P. Siderovski,[‡] and Brian Kuhlman[†,*]

[†]Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, North Carolina 27599-7260, United States

[‡]Department of Pharmacology, University of North Carolina, Chapel Hill, North Carolina 27599-7365, United States

Supporting Information

ABSTRACT: The de novo design of protein-binding peptides is challenging because it requires the identification of both a sequence and a backbone conformation favorable for binding. We used a computational strategy that iterates between structure and sequence optimization to redesign the C-terminal portion of the RGS14 GoLoco motif peptide so that it adopts a new conformation when bound to $G\alpha_{i1}$. An X-ray crystal structure of the redesigned complex closely matches the computational model, with a backbone root-mean-square deviation of 1.1 Å.

Computational protein design tests our understanding of protein energetics and allows the creation of new functional proteins.[1−3] In the area of protein interface design, computer-based methods have been used to stabilize protein−protein interactions, redesign protein-binding specificities, and design new interactions from scratch.[4,5] In those studies, the designs were based on high-resolution structures and did not involve large perturbations to the backbone conformation of either interacting partner. However, in many naturally occurring interactions, large conformational changes accompany binding. Such changes particularly occur in peptide−protein interactions, where the peptide usually populates an ensemble of conformations in the unbound state but adopts a single conformation when bound to the target protein. To design a new protein-binding peptide, it is necessary to identify both an amino acid sequence and a backbone conformation that are favorable for binding.

A variety of approaches for coupling structure and sequence optimization have been described.[6−14] Here we iterated between sequence design and structure refinement with the Rosetta molecular modeling program[15,16] to design the bound conformation of a protein-binding peptide. Our model system was the GoLoco motif from the G-protein regulator and H-Ras effector[17] RGS14, which binds to the heterotrimeric G-protein α subunit $G\alpha_{i1}$ in its inactive, GDP-complexed state.[18] The 36-residue GoLoco peptide interacts with both the Ras-like and all-helical domains of $G\alpha_{i1}$·GDP (Figure S1 in the Supporting Information), burying 1900 Å^2 of surface area (PDB entry 2OM2).[19] In this study, we focused on the last 12 residues of the RGS14 GoLoco motif, which bind with an irregular secondary structure to a hydrophobic groove between the αA and αB helices of the $G\alpha_{i1}$ all-helical domain (Figure 1). Removing these residues from the GoLoco motif weakens the binding affinity for $G\alpha_{i1}$·GDP from 95 nM to ∼20 μM (Figure S2c). We used our flexible-backbone design protocol to replace the last 12 residues of the RGS14 GoLoco motif peptide with a new 16-residue sequence designed to adopt an α-helix when bound to $G\alpha_{i1}$.

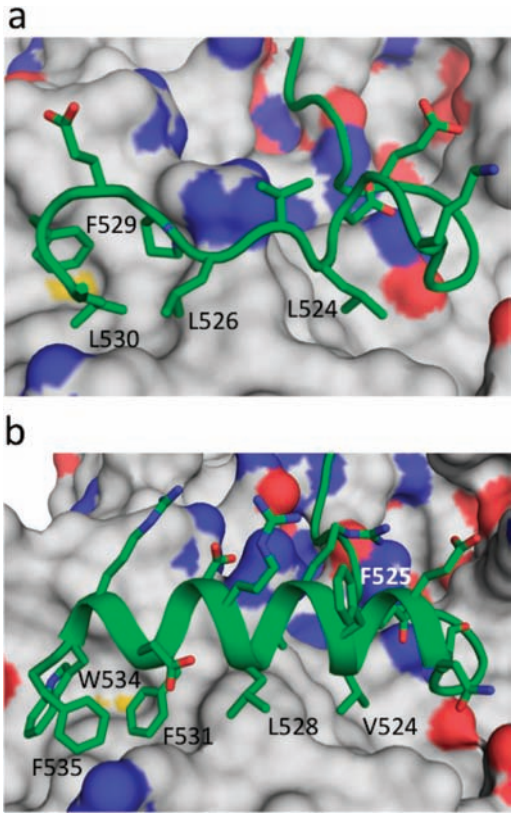

Figure 1. Redesigning the RGS14 GoLoco motif. (a) Crystal structure of the wild-type RGS14 GoLoco motif peptide bound to $G\alpha_{i1}$. (b) Model of the redesigned GoLoco motif, GL_{helix}-4, bound to $G\alpha_{i1}$. Both the wild-type GoLoco motif and GL_{helix}-4 are shown in cartoon representation with selected side chains displayed and labeled. $G\alpha_{i1}$ is shown in surface representation with side-chain nitrogens colored blue and side-chain oxygens colored red.

Received: November 16, 2010
Published: March 09, 2011

   dx.doi.org/10.1021/ja110296z | J. Am. Chem. Soc. 2011, 133, 4190–4192

First, the backbone conformation of the GoLoco motif C-terminus was rebuilt using a fragment assembly protocol with helical fragments from the Protein Data Bank.[20] A low-resolution score function was used to prevent steric overlap with $G\alpha_{i1}$ and favor the burial of hydrophobic residues.[15] The initial sequence of the helix was set to have a hydrophobic face in order to favor docking into the hydrophobic groove between the αA and αB helices of $G\alpha_{i1}$. Independent fragment assembly trajectories were used to produce 2000 starting structures for high-resolution sequence design and backbone refinement. The new helix made contact with $G\alpha_{i1}$ in all of the starting structures, but the orientation and distance of the helix relative to $G\alpha_{i1}$ varied among the models (Figure S3).

Each of the starting structures was optimized with four rounds of sequence design followed by backbone and side-chain minimization with a high-resolution score function that evaluated van der Waals contacts, hydrogen bonding, desolvation energies, backbone and side-chain torsion potentials, and residue-type-based reference energies.[10] Sequence optimization was performed using a simulated annealing protocol with a backbone-dependent rotamer library.[21] Backbone and side-chain torsion angle minimization used a gradient-based quasi-Newtonian method.[15] During this process, the amino acid sequence of $G\alpha_{i1}$ was held fixed but side-chain and backbone torsion angles in the $G\alpha_{i1}$ αA and αB helices were allowed to vary. In general, after four rounds of sequence design followed by backbone refinement, the energies of the models stopped decreasing, suggesting that most of the sequence/structure combinations were trapped in local minima on the energy landscape.

The resulting models were filtered to remove designs that introduced unsatisfied, buried hydrogen-bonding partners. Of the remaining designs, those with the lowest Rosetta energy and highest-quality packing (as judged by the $SASA_{pack}$ score;[22] see Figure S4) were visually inspected. Several designs contained hydrophobic residues in positions that were primarily solvent-exposed. In most cases, these residues were replaced by performing fixed-backbone design, allowing only polar amino acids. In some of the best-scoring models, we also saw small cavities that could potentially be filled if neighboring GoLoco motif side chains were enlarged. However, when we forced a mutation to a larger amino acid, we introduced clashes that raised the overall energy of the model. This result suggested that the protein should be relaxed before the favorability of the mutation was evaluated. Our standard protocol performs mutation and backbone relaxation as separate steps and does not easily allow for coupled changes that may be needed to move from small to large side chains. We previously developed a strategy for identifying affinity-enhancing point mutations that cycles through every position at an interface and evaluates all possible point mutations.[19] Neighboring side-chain and backbone torsion angles are allowed to relax before evaluating the energy of the mutation. We used this protocol to scan our top-scoring GoLoco motif designs for mutations that could further improve packing at the GoLoco motif−$G\alpha_{i1}$ interface. In many cases no mutations were identified, but in our most successful design, GL_{helix}-4, an important mutation was identified: His 531 to Phe (Figure 1).

Four GoLoco motif peptide designs were selected for experimental characterization: GL_{helix}-1, GL_{helix}-2, GL_{helix}-3, and GL_{helix}-4 (Figure S2a and Table S1). To measure binding affinities for $G\alpha_{i1}\cdot$GDP, the GL_{helix} peptides were labeled with fluorescein, and the fluorescence polarization was monitored as a function of $G\alpha_{i1}\cdot$GDP concentration. Interestingly, two of the designs, GL_{helix}-1 and GL_{helix}-2, bound more weakly to $G\alpha_{i1}\cdot$GDP than the truncated GoLoco motif peptide with its C-terminal residues removed (Figure S2b,c). This result suggests that these two peptides may form intra- or intermolecular interactions that compete with binding to $G\alpha_{i1}$. Consistent with this interpretation, dye-labeled GL_{helix}-1 and GL_{helix}-2 had higher intrinsic fluorescence polarization values than GL_{helix}-3 and GL_{helix}-4. GL_{helix}-4 bound $G\alpha_{i1}\cdot$GDP more tightly (Kd = 810 nM) than the truncated GoLoco motif (Kd = ∼20 μM), indicating that the designed residues formed favorable interactions with $G\alpha_{i1}$ (Figure 1b and Figure S2b,c).

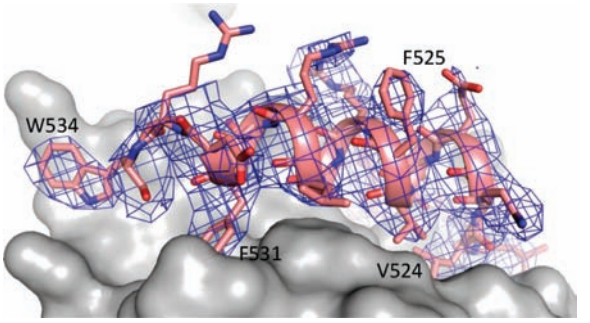

Figure 2. Crystal structure of GL_{helix}-4 bound to $G\alpha_{i1}\cdot$GDP. The unbiased $2F_o - F_c$ electron density (blue mesh, contoured to $\sigma = 1.5$; see Supplementary Methods in the Supporting Information) indicates a GoLoco motif peptide C-terminal α-helix (salmon) bound to the all-helical domain of $G\alpha_{i1}$ (gray surface). The well-defined aromatic side chains Phe 525, Phe 531, and Trp 534 establish the correct helical orientation and register.

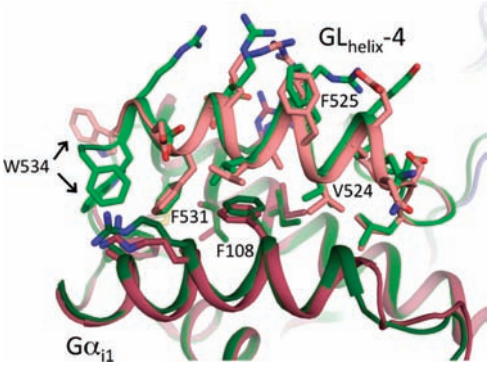

Figure 3. Model (light- and dark-green) and crystal structure (salmon and maroon) of GL_{helix}-4 bound to $G\alpha_{i1}$. The structures were superimposed by minimizing the rmsd between the backbone atoms in the all-helical domain of $G\alpha_{i1}$. The peptide side chains of Val 524 and Phe 531 pack on either side of Phe 108 from $G\alpha_{i1}$, as predicted. Only the orientation of Trp 534 and the absence of a definable residue 535 in the crystal structure deviate significantly from the predicted model.

To determine whether GL_{helix}-4 adopts the designed conformation when bound to $G\alpha_{i1}$, we elucidated the crystal structure of the complex (PDB entry 2XNS; Table S2). Our structure was determined using diffraction data to a resolution of 3.4 Å (R factor = 22.3, R_{free} = 24.4). The redesigned portion of the RGS14 GoLoco motif was well-defined in the electron density map and clearly adopts an α-helical backbone (Figure 2). Strong electron density was observed for the aromatic side chains Phe 525, Phe 531, and Trp 534 on the designed helix. Except for the last two residues of the peptide, the computational design model closely matches the crystal

dx.doi.org/10.1021/ja110296z | J. Am. Chem. Soc. 2011, 133, 4190–4192

structure (Figure 3). The backbone root-mean-square deviation (rmsd) between peptide residues 520−533 of the designed model and the equivalent residues in the crystal structure was 1.1 Å when the complexes were aligned with the helical domain of $G\alpha_{i1}$. As designed, the side chains of Val 524 and Phe 531 pack on either side of Phe 108 from $G\alpha_{i1}$.

The extreme C-terminus of the peptide is located close to a crystallographic symmetry axis, and thus, the corresponding electron density for Phe 535 overlaps with its symmetry-related counterpart. The deviation of the observed peptide structure from the computational model may be due in part to crystal-packing effects. In the crystal, the preceding Trp 534 packs against the neighboring GL_{helix}-4 peptide (Figure S5). Further evidence that Trp 534 does not form strong interactions with $G\alpha_{i1}$ was provided by mutagenesis studies. Mutating Trp 534 to an alanine had a modest effect on the binding affinity (K_d = 1.4 μM vs 810 nM). In contrast, mutating Phe 531 to an alanine weakened the binding over 10-fold (Figure S6).

In conclusion, our results demonstrate that sequence optimization combined with backbone modeling can be used to rationally design the bound conformation of a protein-binding peptide. Such capabilities represent important steps toward the full de novo design of peptide−protein interactions. In this study, we designed an α-helical peptide to bind to a hydrophobic groove and therefore did not need to consider hydrogen bonds across the interface. One of the challenges for future designs is the creation of interfaces that are more polar in character and involve hydrogen-bonding networks with both side-chain and backbone groups of the peptide.

■ ASSOCIATED CONTENT

Supporting Information. Supporting figures, experimental procedures, and details of the X-ray analysis. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
bkuhlman@email.unc.edu

Present Addresses
[§]University of Colorado, Boulder, CO 80309.
[∥]New York University, New York, NY 10003.

Author Contributions
[⊥]These authors contributed equally.

■ ACKNOWLEDGMENT

Work in the Kuhlman lab was funded by NIH Grant R01 GM073960 and the W. M. Keck Foundation. Work in the Siderovski lab was funded by NIH Grants R01 GM074268 (to D.P.S.) and T32 GM008719 (to D.E.B.). We are grateful for the resources of the SER-CAT 22-ID Beamline at the Advanced Photon Source at Argonne National Laboratory (Argonne, IL).

■ REFERENCES

(1) Van der Sloot, A. M.; Kiel, C.; Serrano, L.; Stricher, F. *Protein Eng. Des. Sel.* **2009**, *22*, 537–542.
(2) Kang, S. G.; Saven, J. G. *Curr. Opin. Chem. Biol.* **2007**, *11*, 329–334.
(3) Lippow, S. M.; Tidor, B. *Curr. Opin. Biotechnol.* **2007**, *18*, 305–311.
(4) Karanicolas, J.; Kuhlman, B. *Curr. Opin. Struct. Biol.* **2009**, *19*, 458–463.
(5) Mandell, D. J.; Kortemme, T. *Nat. Chem. Biol.* **2009**, *5*, 797–807.
(6) Fu, X.; Apgar, J. R.; Keating, A. E. *J. Mol. Biol.* **2007**, *371*, 1099–1117.
(7) Humphris, E. L.; Kortemme, T. *Structure* **2008**, *16*, 1777–1788.
(8) Mandell, D. J.; Kortemme, T. *Curr. Opin. Biotechnol.* **2009**, *20*, 420–428.
(9) Harbury, P. B.; Plecs, J. J.; Tidor, B.; Alber, T.; Kim, P. S. *Science* **1998**, *282*, 1462–1467.
(10) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* **2003**, *302*, 1364–1368.
(11) Georgiev, I.; Donald, B. R. *Bioinformatics* **2007**, *23*, i185–i194.
(12) Georgiev, I.; Keedy, D.; Richardson, J. S.; Richardson, D. C.; Donald, B. R. *Bioinformatics* **2008**, *24*, i196–i204.
(13) Desjarlais, J. R.; Handel, T. M. *J. Mol. Biol.* **1999**, *290*, 305–318.
(14) Su, A.; Mayo, S. L. *Protein Sci.* **1997**, *6*, 1701–1707.
(15) Rohl, C. A.; Strauss, C. E.; Misura, K. M.; Baker, D. *Methods Enzymol.* **2004**, *383*, 66–93.
(16) Das, R.; Baker, D. *Annu. Rev. Biochem.* **2008**, *77*, 363–382.
(17) Willard, F. S.; Willard, M. D.; Kimple, A. J.; Soundararajan, M.; Oestreich, E. A.; Li, X.; Sowa, N. A.; Kimple, R. J.; Doyle, D. A.; Der, C. J.; Zylka, M. J.; Snider, W. D.; Siderovski, D. P. *PLoS One* **2009**, *4*, No. e4884.
(18) Kimple, R. J.; Kimple, M. E.; Betts, L.; Sondek, J.; Siderovski, D. P. *Nature* **2002**, *416*, 878–881.
(19) Sammond, D. W.; Eletr, Z. M.; Purbeck, C.; Kimple, R. J.; Siderovski, D. P.; Kuhlman, B. *J. Mol. Biol.* **2007**, *371*, 1392–1404.
(20) Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D. *J. Mol. Biol.* **1997**, *268*, 209–225.
(21) Kuhlman, B.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10383–10388.
(22) Sood, V. D.; Baker, D. *J. Mol. Biol.* **2006**, *357*, 917–927.