COMPUTATIONAL REDESIGN OF PROTEIN-PROTEIN INTERACTIONS

by
Deanne Wallander Sammond

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry and Biophysics (Program in Molecular and Cellular Biophysics).

Chapel Hill
2008

Approved by

Advisor: Brian Kuhlman

Reader: Nikolay Dokholyan

Reader: Andrew Lee

Reader: David Siderovski

Chair: John Sondek

ABSTRACT

Deanne Wallander Sammond: Computational redesign of protein-protein interactions
(Under the direction of Dr. Brian Kuhlman)

Computational protein modeling predicts and manipulates the biophysical properties of proteins based on their amino acid sequences. Computational protein modeling has been applied to protein-protein and protein-peptide interactions in order to develop research tools as well as protein therapeutics. The intent of our work was to address three areas of protein interface design that are of special interest due to their potential applications. Affinity maturation has been used to improve biosensors as well as potential protein therapeutics. We developed a protocol to predict point mutations that will enhance the binding affinity of protein-protein interactions. Extending this work, we evaluated a protocol designed to increase protein-peptide binding specificity. Redesigning binding specificity can be used to isolate specific protein interactions within complicated signaling networks by limiting the interactions of redesigned proteins with their wild type counterparts and other natural binding partners. Finally, the de novo design of a peptide-protein interaction, or the design of a peptide that will bind a wild type protein, could enable the creation of biosensors or therapeutics from scratch. We take a step towards this goal by redesigning a portion of a peptide backbone in the context of its wild type binding partner.

ACKNOWLEDGEMENTS

I would like to thank Brian for the opportunity to pursue my doctoral dissertation under his direction. Brian's love of science and wonderful mentoring were invaluable to me throughout my dissertation. I would like to thank the members of my committee for their suggestions, advice and interest. I would especially like to thank David Siderovski, with whom it was a joy to collaborate. I would like to thank Barry Lentz for the opportunity to join the Molecular and Cellular Biophysics Training Program. I would also like to thank Alex Tropsha for the opportunity to join the Bioinformatics and Computational Biology Training Program. I would like to thank my peers and classmates, most notably Ziad Eletr and Gottfried Schroeder. We shared challenges and successes and grew as scientists and as people. I would like to thank Carrie Purbeck, whose help and friendship added so much to my educational experience. I would like to thank my parents, Ray and Carol, for believing in me. I would like to thank my children, Drew and Connor, for giving me the inspiration and courage to follow my ambitions. I would like to thank my husband, Doug, for his constant encouragement and advice. Finally, I would like to thank Pierre Morell and Arrel Toews for getting this whole journey started.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

6-IAF: 6-iodoacetamidofluorescein

β-ME: β-mercaptoethanol

BLIP: ß-lactamase-inhibitory protein

CaM: calmodulin

CASP: critical assessment of structure prediction

DNA: deoxyribonucleic acid

E6AP: E6-associated protein

GDP: guanosine diphosphate

GTP: guanosine triphosphate

GDI: guanine nucleotide dissociation inhibitors

HECT: homology to E6AP carboxy-terminus

ICAM-1: intercellular adehesion molecule

IPTG: Isopropyl β-D-1-thiogalactopyranoside

LFA-1: integrin lymphocyte function-associated antigen-1

PDA: Protein Design Automation

PDB: Protein Data Bank

RGS14: regulator of G-protein signaling subtype 14

smMLCK: smooth muscle myosin light chain kinase

TCEP: *tris*(2-carboxyethyl)phosphine

TEM1: tumor endothelial marker 1

UbcH7: ubiquitin-conjugating enzyme

CHAPTER I

INTRODUCTION TO COMPUTATIONAL PROTEIN-PROTEIN INTERFACE DESIGN

INTRODUCTION TO COMPUTATIONAL PROTEIN MODELING

The goal of computational protein modeling is to model the physical laws that dictate the form and function of proteins in order to predict or manipulate their biophysical properties. After it was demonstrated that the amino acid sequence of a protein determines its structure[1; 2] it became the goal of computational protein modeling to learn how to model the information in amino acid sequences to increase the efficiency of experimental research through the power of prediction. One of the greatest challenges in computational protein modeling is to predict the structure of a protein given only the amino acid sequence. The number of known proteins is fast outpacing the number of characterized structures.[3; 4] The critical assessment of structure prediction, or CASP, is a double-blind assessment of the computational algorithms designed to predict protein structures. The CASP experiments have highlighted significant progress that has been made in the field protein of structure prediction.[5; 6; 7; 8] Currently, however, high-resolution structure prediction, where the predicted structure is within approximately 1.5 angstroms of the x-ray crystal structure, seems to be limited to small proteins that are approximately 85 residues or less.[9] A major reason for this limitation seems to be the challenge of sufficiently modeling conformational space given current computer limitations.[9; 10]

The inverse to computational protein structure prediction is computational protein design. Computational protein design models the same physical laws, but starts with protein structures and makes predictions about the amino acid sequences that are compatible with those structures. There are numerous

research and industrial applications for protein design, including the creation of new enzymes[11; 12] and the development of biosensors.[13; 14] Many protein design applications implement fixed protein backbone design approximations, where the backbone torsion angles are those of the experimentally determined structure and only the side chains torsion angles are modeled. Bolon et al. stabilized thioredoxin using fixed-backbone protein design to search for sequence positions that could accommodate polar amino acids in order to alter the hydrogen bonding interactions.[15] Desjarlais et al. found that including backbone flexibility when modeling protein hydrophobic cores yielded comparable results to fixed-backbone designs.[16] Increasingly complex design challenges appear to require some type of backbone modeling. Joachimiak et al. designed a hydrogen bond network by modeling the colicin E6 DNase and Im7 immunity protein complex about an axis of rotation.[17] Ambroggio et al. designed a protein that can switch between two conformations[18] and Havrenak et al. engineered coiled-coil binding specificity.[19] Both of these accomplishments were done by modeling protein sequences against an ensemble of protein structures. The design of the switch protein required modeling two different protein backbone structures and designing a sequence that is compatible with both structures. The coiled-coil binding specificity was achieved by modeling the homodimer and heterodimer interactions and designing a sequence targeted to stabilize the heterodimer interaction and then designing a sequence to stabilize the desired interactions while destabilizing the undesired interactions. A new protein fold has been designed using a protocol closely related to those used in protein structure

prediction,[20] blurring the lines between protein structure prediction and protein design.

COMPUTATIONAL PROTEIN-PROTEIN INTERFACE DESIGN

Another great challenge in the field of computational protein modeling is the prediction and design of protein-protein interactions. Protein-protein interactions form complex cellular networks that regulate most cellular functions. Alteration of individual proteins in these complex networks can lead to disease. The signaling protein p53 has been implicated in 50% of human cancers,[21] highlighting the need to better understand cellular networks and be able to modulate macromolecular complexes. The three dimensional atomic coordinates of a protein give considerable information as to the function of a protein. There is more information needed to thoroughly understand the role of a protein in the cell, however, including which other proteins it interacts with. Experimental methods such as yeast two-hybrid assays[22] and mass spectrometry[23] have been used to map out networks of protein-protein interactions. Computational prediction of protein-protein interactions, or computational protein docking, aims to increase the efficiency of this work. As with protein structure prediction, a major challenge in computational protein docking is the adequate sampling of conformational space.[10]

The inverse approach to protein docking prediction is protein-protein interface design. Interface design can benefit from the fact that many proteins do not have significant backbone movement upon binding,[24] greatly decreasing the degrees of freedom that need to be modeled. This allows many protein-protein design protocols to implement fixed-backbone design where the backbone

coordinates of the complex are taken from the experimentally determined structure and are held fixed. Cases where a only single sequence position is modeled seem to work well with fixed backbone approximation. Computational alanine scanning has been accomplished with fixed-backbone design.[25] Our work predicting affinity enhancing point mutations indicated that incorporating modest backbone flexibility gave comparable results to a much simpler fixed backbone protocol.[26] Several groups have reengineered the binding specificity of calmodulin using the fixed backbone approximation.[13; 27; 28] Computational protein-protein interface design seeks to look beyond the protein-protein complex by providing tools to investigate characteristics of protein-protein interactions such as phenotype and cellular location, or to enable us to create protein-protein interactions from scratch for the development of protein therapeutics and biosensors. As with monomeric protein design, the protein-protein interface design algorithms are increasingly incorporating backbone movement. This can include perturbation of backbone torsion angles, rigid-body docking between two protein chains, or the design of entirely new backbone structures. The most impressive specificity redesign results to date incorporated rigid-body backbone movement[17] where the two protein chains were rotated about each other on an axis of rotation. Huang et al. created a de novo designed protein-protein interaction using a docking algorithm.[29] Sood et al. enhanced protein-peptide binding affinity by designing extensions to peptides at interfaces in order to increase the buried surface areas.[30] The design of peptide extensions required

modeling new backbone coordinates followed by the search for complementary sequences.

PREDICTING AFFINITY ENHANCING POINT MUTATIONS

The work discussed here addresses three fundamental applications of protein-protein interface design. We developed a protocol to predict point mutations that will enhance protein-protein or protein-peptide binding affinity. This work allowed us to evaluate the predictive power of the Rosetta energy function. We then automated a protocol designed to increase the binding specificity of a protein complex. This protocol was first presented by Kortemme et al.[31] where they applied it to a protein-protein complex. We evaluated its use with a protein-peptide complex. Lastly we develop a protocol to design a new backbone and sequence of a peptide in the context of one a wild type binding partner. This protocol borrows from methods used for protein structure prediction[32] by building a peptide backbone from fragments taken from existing structures in the protein data bank[33] followed by iterations of sequence design.[34]

First, we developed a protocol for the prediction of point mutations that will enhance protein-protein binding affinity. Protein-protein binding affinity has practical applications for both research and protein therapeutics. Antibodies used in immunoassays can have improved performance with enhanced binding affinity,[35] or antibodies specific for predetermined targets that are isolated from libraries often require affinity maturation in order to achieve the desired therapeutic results.[36] One approach to enhance protein-protein binding affinity is to increase the electrostatic interactions between two proteins. This is typically done by modeling the residues on the periphery of the interface. This approach

has been successfully used to increase the affinity between TEM1 β-lactamase and its protein inhibitor BLIP by over 250-fold,[37] and the affinity between Ral and Ral guanine nucleotide dissociation stimulator by over 25-fold.[38] This approach works best when there are large like-charged regions on a protein chain that are not paired with the charges on the other side of the interface. An alternative approach is to model the residues that are buried at the protein-protein interface to improve hydrophobic packing, hydrogen bonding or desolvation energies. The basis for our protocol is that increasing buried hydrophobic surface area across the interface can stabilize the protein-protein interaction provided destabilizing effects such as steric clashes, unsatisfied hydrogen bonds or destabilization of individual protein chains does not occur. Mutational studies of buried hydrophobic residues highlight the importance of a tightly packed core in the stability of monomeric proteins.[39; 40] We apply these finding to protein-protein interactions, using the protein design software, RosettaDesign, to model point mutations at the interface and predict the resulting $\Delta\Delta G_{binding}$.

REDESIGNING PROTEIN-PEPTIDE BINDING SPECIFICITY

The second goal we address deals with the rational manipulation of protein-peptide binding specificity. Computational redesigns of protein-protein binding specificity have used positive protein design, where protein sequences are designed to stabilize the desired interaction, and combinations of positive and negative design. Negative protein design designs a sequence to destabilize undesired interactions. Protein-protein binding specificity for calmodulin was successfully reengineered using positive protein design by redesigning the interface of the desired interaction only.[27] Both positive and negative protein design were used to design binding specificity of a coiled-coil interaction using positive and negative design simultaneously. Sequence positions were deigned based on an amino acid's ability to stabilize the desired interaction while destabilize the undesired interactions.[19] Our protocol searched for point mutations at a protein-peptide interface that would disrupt the wild-type interaction. Sequence positions that neighbor the destabilizing mutation were then redesigned in an effort to regain wild-type binding affinity. The goal was for the resulting redesigned protein complex to bind as the wild-type complex does but not to cross-interact with its wild-type counterparts. The idea was that not only will the redesigned proteins not interact with their wild type counterparts, but that this specificity will extend to the homologous family members of the wild type counterparts. In this way the redesigned protein-protein interaction could be isolated from the complex cellular network in which it participates. This type of

redesigned specificity could allow for the rewiring of cellular networks as well as the creation of biosensors that could be used for live in cell imaging. This automated computational protocol is based on the work by Kortemme et al., where they introduced this protocol to redesign the specificity of the colicin E6 DNase-Im7 immunity protein complex.[31] Here we automated the protocol and applied it to a protein-peptide complex.

de novo PROTEIN-PEPTIDE INTERFACE DESIGN

Our final aim was to develop a protocol that can design a peptide backbone and sequence that will bind a wild type protein. This project was a small step toward a full *de novo* protein-protein interface design. Modeling backbone flexibility is currently seen as one of the major challenges in computational protein design mainly because of the increase in degrees of freedom that need to be sampled.[10; 41] Ultimately the goals of protein design depend on algorithms that can balance computational limitations with need to more extensively sample sequence and structure space. Our final aim addressed the area of backbone design at protein-protein interfaces. We began with an x-ray crystal structure of a protein-peptide interface; we removed a portion of one binding partner and redesigned the backbone of that region to have an entirely different secondary structure and sequence within the context of the wild-type binding partner.

ROSETTA PROTEIN DESIGN SOFTWARE

We used the computational prediction and design program, Rosetta, to redesign protein-peptide interactions.[32] Rosetta was created for protein structure prediction.[9] Rosetta has had tremendous success in the critical assessment of structure prediction experiments (CASP)[6; 8] and was subsequently applied to protein docking prediction[10; 42] and protein design.[34] Rosetta has been used to alter the folding path of a protein,[43] refine experimentally determined protein structures[44] and design a novel protein fold.[20] Rosetta uses a Monte Carlo simulated annealing search procedure by making random perturbations to backbone or side chain torsion angles. The Monte Carlo algorithm is used for both side-chain design as well as backbone movement. The search algorithm is directed by an all-atom energy function that focuses on short-range interactions. We used the all-atom mode of Rosetta to model the amino acids when designing protein sequences or predicting the free energy of binding and the stability of monomeric proteins. The energy function is comprised of a linear sum of molecular mechanics and knowledge-based terms. These include a Lennard-Jones potential, a distance and orientation dependent hydrogen bonding potential,[45] and the Lazaridis-Karplus implicit solvation model.[46] We primarily use a version of the energy function that was parameterized to best reproduce native sequences when redesigning whole proteins in fixed backbone simulations (command line option, -soft_rep_design , Rosetta v 2.1). This variation of the energy function significantly dampens repulsion energies to allow for small atom-atom clashes that may be accommodated by small changes in side chain and

backbone conformation. See supplementary material of Dantas et al. for a complete description of this version of the Rosetta energy function. It is referred to as Rosetta_DampRep.[47] We use this dampened repulsion model for most of the work presented here, including designs where the backbone torsion angles are fixed as well as design runs that include movement of the backbone. There is a reference energy that is meant to represent the energy of the amino acids in an unfolded state.[34] Amino acid side-chains are represented in discrete conformations taken from a rotamer library.[48] We expand the Dunbrack rotamer library by varying the chi 1 and chi 2 angles one standard deviation away from their most probable values in an effort to prevent the overestimation of destabilizing effects that can result from modeling with discrete side chain conformations on a fixed backbone scaffold. The backbone phi and psi angles are evaluated using Ramachandran torsion preferences.[49] In addition to Rosetta's atomic-level potential function we use a more course-grained, centroid-based energy function where amino acid side chains are represented by centroids located at the side-chain center of mass. The centroid mode is used when significant backbone modeling is performed. The centroid mode runs are followed by all-atom designs in order to obtain the high-resolution structures and energetic predictions needed for protein-peptide interface design. We also evaluate gradient based minimization of side chain and backbone torsion angles and rigid body docking using a quasi-Newton method.[32] This allows small side chain and backbone movements intended to relieve slight van der Waals

overlaps between amino acids as well as provide a way to improve hydrogen bond angles and distances.

MODEL SYSTEMS USED FOR EXPERIMENTAL VALIDATION

The model system that we use for the experimental validation of our computational protocols is the $G\alpha_{i1}$ protein from the heterotrimeric G-protein system bound to the GoLoco motif from the RGS14 multidomain regulatory protein.[50] Heterotrimeric Gα proteins have intrinsic GTPase activity, enabling them to switch from their active, guanosine triphosphate (GTP)-bound conformation to their inactive, guanosine diphosphate (GDP)-bound conformation by hydrolyzing the γ phosphate of the GTP. The GoLoco motif has been shown to selectively bind the GDP-bound Gα proteins from the adenylyl-cyclase-inhibitory subclass.[51] Upon binding, the GoLoco motif stabilizes the bound nucleotide in the Gα protein, thereby acting as a guanine nucleotide dissociation inhibitor (GDI).[52; 53; 54] The ability for the GoLoco motifs to bind the inactive, GDP-bound Gα proteins in place of the Gβγ heterodimer suggests a role for the GoLoco in G protein-coupled receptor signaling. Experimental evidence suggests, however, that the GoLoco motifs are involved in spindle-pole organization, chromosomal segregation, and asymmetric cell division.[55; 56] The exact biochemical roles of the GoLoco motifs is not currently understood. Questions remain as to the role the GoLoco motifs and the GoLoco-Gα i/o complexes might play in these processes, as well as the methods of regulation involved.[55] Rationally redesigned GoLoco-Gα complexes could provide novel investigative tools. A redesigned RGS14 GoLoco motif that binds a redesigned $G\alpha_{i1}$ protein, exhibiting biophysical behavior similar to that of the wild-type complex, could be used to investigate whether the RGS14 GoLoco-$G\alpha_{i1}$ complex

and the RGS14 GoLoco-$G\alpha_{i3}$ complex have distinguishable roles. GoLoco motifs with enhanced binding specificity could also be used as biosensors with improved signal to noise ratios.

In addition to a biochemical interest, the RGS14 GoLoco-$G\alpha_{i1}$ complex was selected because the interface has a number of structural features that make it a good target for computational protein design. The 36-residue GoLoco motif of RGS14 binds to the GDP-bound state of the $G\alpha_{i1}$ protein and interacts with both the Ras-like, guanine nucleotide binding domain and the all-helical domain of the $G\alpha_{i1}$ protein.[50] The $G\alpha_{i1}$–RGS14 GoLoco interface buries over 1900 $Å^2$ of surface area and has a low nanomolar binding affinity.[26; 50] We can investigate how our protocols work on the more flexible Ras-like $G\alpha_{i1}$ domain, which includes a flexible region termed switch II, which has been shown to undergo conformational change upon nucleotide exchange or hydrolysis. The all-helical domain of the $G\alpha_{i1}$ protein provides a less flexible domain which appears to undergo only a very slight backbone movement in helix αB as seen when comparing one of the unbound $G\alpha_{i1}$ structures (PDB ID: 1GIA)[57] with the GoLoco complex.[50]

Our development of a protocol to predict affinity enhancing point mutations is experimentally and computationally accessible enough that we were able to include a second protein model system for additional proof of principle. The second protein model system for our affinity enhancing protocol is the E2, ubiquitin-conjugating enzyme UbcH7, bound to the E3, E6-associated protein (E6AP), from the ubiquitin pathway.[58] This UbcH7-E6AP complex differs from

the GoLoco-Gαi1 complex in that the individual protein components undergo little conformational change upon binding. The structures of E6AP in the unbound state is virtually identical to the conformation that it adopts when forming a complex.[58] Likewise the UbcH7 fold observed in the UbcH7-E6AP complex is similar to that of numerous unbound structures of other ubiquitin conjugating enzymes.[59]

REFERENCES

1. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223-30.

2. Anfinsen, C. B., Haber, E., Sela, M. & White, F. H., Jr. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A* **47**, 1309-14.

3. Rost, B. (1998). Marrying structure and genomics. *Structure* **6**, 259-63.

4. Berman, H. M., Bhat, T. N., Bourne, P. E., Feng, Z., Gilliland, G., Weissig, H. & Westbrook, J. (2000). The Protein Data Bank and the challenge of structural genomics. *Nat Struct Biol* **7 Suppl**, 957-9.

5. Bonneau, R., Tsai, J., Ruczinski, I., Chivian, D., Rohl, C., Strauss, C. E. & Baker, D. (2001). Rosetta in CASP4: progress in ab initio protein structure prediction. *Proteins* **Suppl 5**, 119-26.

6. Bradley, P., Chivian, D., Meiler, J., Misura, K. M., Rohl, C. A., Schief, W. R., Wedemeyer, W. J., Schueler-Furman, O., Murphy, P., Schonbrun, J., Strauss, C. E. & Baker, D. (2003). Rosetta predictions in CASP5: successes, failures, and prospects for complete automation. *Proteins* **53 Suppl 6**, 457-68.

7. Kryshtafovych, A., Venclovas, C., Fidelis, K. & Moult, J. (2005). Progress over the first decade of CASP experiments. *Proteins* **61 Suppl 7**, 225-36.

8. Lesk, A. M., Lo Conte, L. & Hubbard, T. J. (2001). Assessment of novel fold targets in CASP4: predictions of three-dimensional structures, secondary structures, and interresidue contacts. *Proteins* **Suppl 5**, 98-118.

9. Bradley, P., Misura, K. M. & Baker, D. (2005). Toward high-resolution de novo structure prediction for small proteins. *Science* **309**, 1868-71.

10. Schueler-Furman, O., Wang, C., Bradley, P., Misura, K. & Baker, D. (2005). Progress in modeling of protein structures and interactions. *Science* **310**, 638-42.

11. Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Rothlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Barbas, C. F., 3rd, Hilvert, D., Houk, K. N., Stoddard, B. L. & Baker, D. (2008). De novo computational design of retro-aldol enzymes. *Science* **319**, 1387-91.

12. Rothlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J. L., Althoff, E. A., Zanghellini, A., Dym, O., Albeck, S., Houk, K. N., Tawfik, D. S. & Baker, D. (2008). Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190-5.

13. Palmer, A. E., Giacomello, M., Kortemme, T., Hires, S. A., Lev-Ram, V., Baker, D. & Tsien, R. Y. (2006). Ca2+ indicators based on computationally redesigned calmodulin-peptide pairs. *Chem Biol* **13**, 521-30.

14. de Lorimier, R. M., Smith, J. J., Dwyer, M. A., Looger, L. L., Sali, K. M., Paavola, C. D., Rizk, S. S., Sadigov, S., Conrad, D. W., Loew, L. & Hellinga, H. W. (2002). Construction of a fluorescent biosensor family. *Protein Sci* **11**, 2655-75.

15. Bolon, D. N., Marcus, J. S., Ross, S. A. & Mayo, S. L. (2003). Prudent modeling of core polar residues in computational protein design. *J Mol Biol* **329**, 611-22.

16. Desjarlais, J. R. & Handel, T. M. (1999). Side-chain and backbone flexibility in protein core design. *J Mol Biol* **290**, 305-18.

17. Joachimiak, L. A., Kortemme, T., Stoddard, B. L. & Baker, D. (2006). Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. *J Mol Biol* **361**, 195-208.

18. Ambroggio, X. I. & Kuhlman, B. (2006). Computational design of a single amino acid sequence that can switch between two distinct protein folds. *J Am Chem Soc* **128**, 1154-61.

19. Havranek, J. J. & Harbury, P. B. (2003). Automated design of specificity in molecular recognition. *Nat Struct Biol* **10**, 45-52.

20. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L. & Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364-8.

21. Gasco, M., Shami, S. & Crook, T. (2002). The p53 pathway in breast cancer. *Breast Cancer Res* **4**, 70-6.

22. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S. & Rothberg, J. M. (2000). A

comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403**, 623-7.

23. Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M. A., Copley, R. R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G. & Superti-Furga, G. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-7.

24. Betts, M. J. & Sternberg, M. J. (1999). An analysis of conformational changes on protein-protein association: implications for predictive docking. *Protein Eng* **12**, 271-83.

25. Kortemme, T., Kim, D. E. & Baker, D. (2004). Computational alanine scanning of protein-protein interfaces. *Sci STKE* **2004**, pl2.

26. Sammond, D. W., Eletr, Z. M., Purbeck, C., Kimple, R. J., Siderovski, D. P. & Kuhlman, B. (2007). Structure-based protocol for identifying mutations that enhance protein-protein binding affinities. *J Mol Biol* **371**, 1392-404.

27. Shifman, J. M. & Mayo, S. L. (2002). Modulating calmodulin binding specificity through computational protein design. *J Mol Biol* **323**, 417-23.

28. Green, D. F., Dennis, A. T., Fam, P. S., Tidor, B. & Jasanoff, A. (2006). Rational design of new binding specificity by simultaneous mutagenesis of calmodulin and a target peptide. *Biochemistry* **45**, 12547-59.

29. Huang, P. S., Love, J. J. & Mayo, S. L. (2007). A de novo designed protein protein interface. *Protein Sci* **16**, 2770-4.

30. Sood, V. D. & Baker, D. (2006). Recapitulation and design of protein binding peptide structures and sequences. *J Mol Biol* **357**, 917-27.

31. Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L. & Baker, D. (2004). Computational redesign of protein-protein interaction specificity. *Nat Struct Mol Biol* **11**, 371-9.

32. Rohl, C. A., Strauss, C. E., Misura, K. M. & Baker, D. (2004). Protein structure prediction using Rosetta. *Methods Enzymol* **383**, 66-93.

33. Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D. & Zardecki, C. (2002). The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* **58**, 899-907.

34. Kuhlman, B. & Baker, D. (2000). Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A* **97**, 10383-8.

35. Siegel, R. W., Baugher, W., Rahn, T., Drengler, S. & Tyner, J. (2008). Affinity Maturation of Tacrolimus Antibody for Improved Immunoassay Performance. *Clin Chem*.

36. Sheedy, C., MacKenzie, C. R. & Hall, J. C. (2007). Isolation and affinity maturation of hapten-specific antibodies. *Biotechnol Adv* **25**, 333-52.

37. Selzer, T., Albeck, S. & Schreiber, G. (2000). Rational design of faster associating and tighter binding protein complexes. *Nat Struct Biol* **7**, 537-41.

38. Kiel, C., Selzer, T., Shaul, Y., Schreiber, G. & Herrmann, C. (2004). Electrostatically optimized Ras-binding Ral guanine dissociation stimulator mutants increase the rate of association by stabilizing the encounter complex. *Proc Natl Acad Sci U S A* **101**, 9223-8.

39. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A. & Sauer, R. T. (1990). Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science* **247**, 1306-10.

40. Matthews, B. W. (1995). Studies on protein stability with T4 lysozyme. *Adv Protein Chem* **46**, 249-78.

41. Kortemme, T. & Baker, D. (2004). Computational design of protein-protein interactions. *Curr Opin Chem Biol* **8**, 91-7.

42. Gray, J. J., Moughon, S. E., Kortemme, T., Schueler-Furman, O., Misura, K. M., Morozov, A. V. & Baker, D. (2003). Protein-protein docking predictions for the CAPRI experiment. *Proteins* **52**, 118-22.

43. Nauli, S., Kuhlman, B. & Baker, D. (2001). Computer-based redesign of a protein folding pathway. *Nat Struct Biol* **8**, 602-5.

44. Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A. J., Read, R. J. & Baker, D. (2007). High-resolution structure prediction and the crystallographic phase problem. *Nature* **450**, 259-64.

45. Kortemme, T., Morozov, A. V. & Baker, D. (2003). An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J Mol Biol* **326**, 1239-59.

46. Lazaridis, T. & Karplus, M. (1999). Effective energy function for proteins in solution. *Proteins* **35**, 133-52.

47. Dantas, G., Corrent, C., Reichow, S. L., Havranek, J. J., Eletr, Z. M., Isern, N. G., Kuhlman, B., Varani, G., Merritt, E. A. & Baker, D. (2006). High-resolution Structural and Thermodynamic Analysis of Extreme Stabilization of Human Procarboxypeptidase by Computational Protein Design. *J Mol Biol*.

48. Dunbrack, R. L., Jr. & Cohen, F. E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci* **6**, 1661-81.

49. Bowers, P. M., Strauss, C. E. & Baker, D. (2000). De novo protein structure determination using sparse NMR data. *J Biomol NMR* **18**, 311-8.

50. Kimple, R. J., Kimple, M. E., Betts, L., Sondek, J. & Siderovski, D. P. (2002). Structural determinants for GoLoco-induced inhibition of nucleotide release by Galpha subunits. *Nature* **416**, 878-81.

51. Siderovski, D. P., Diverse-Pierluissi, M. & De Vries, L. (1999). The GoLoco motif: a Galphai/o binding motif and potential guanine-nucleotide exchange factor. *Trends Biochem Sci* **24**, 340-1.

52. De Vries, L., Fischer, T., Tronchere, H., Brothers, G. M., Strockbine, B., Siderovski, D. P. & Farquhar, M. G. (2000). Activator of G protein signaling 3 is a guanine dissociation inhibitor for Galpha i subunits. *Proc Natl Acad Sci U S A* **97**, 14364-9.

53. Peterson, Y. K., Bernard, M. L., Ma, H., Hazard, S., 3rd, Graber, S. G. & Lanier, S. M. (2000). Stabilization of the GDP-bound conformation of Gialpha by a peptide derived from the G-protein regulatory motif of AGS3. *J Biol Chem* **275**, 33193-6.

54. Natochin, M., Gasimov, K. G. & Artemyev, N. O. (2001). Inhibition of GDP/GTP exchange on G alpha subunits by proteins containing G-protein regulatory motifs. *Biochemistry* **40**, 5322-8.

55. Willard, F. S., Kimple, R. J. & Siderovski, D. P. (2004). Return of the GDI: the GoLoco motif in cell division. *Annu Rev Biochem* **73**, 925-51.

56. Kimple, R. J., Willard, F. S. & Siderovski, D. P. (2002). The GoLoco motif: heralding a new tango between G protein signaling and cell division. *Mol Interv* **2**, 88-100.

57. Coleman, D. E., Berghuis, A. M., Lee, E., Linder, M. E., Gilman, A. G. & Sprang, S. R. (1994). Structures of active conformations of Gi alpha 1 and the mechanism of GTP hydrolysis. *Science* **265**, 1405-12.

58. Huang, L., Kinnucan, E., Wang, G., Beaudenon, S., Howley, P. M., Huibregtse, J. M. & Pavletich, N. P. (1999). Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science* **286**, 1321-6.

59. Pickart, C. M. (2001). Mechanisms underlying ubiquitination. *Annu Rev Biochem* **70**, 503-33.

CHAPTER II

STRUCTURE-BASED PROTOCOL FOR IDENTIFYING MUTATIONS THAT ENHANCE PROTEIN-PROTEIN BINDING AFFINITIES

Deanne W Sammond[1], Ziad M Eletr[1], Carrie Purbeck[1], Randall J Kimple[2], David P Siderovski[2], & Brian Kuhlman[1*]

[1]Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7260, USA

[2]Department of Pharmacology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7365, USA

*Corresponding Author

This work was published in the Journal of Molecular Biology (2007) 371(5):1392-1404.

ABSTRACT

The ability to manipulate protein binding affinities is important for the development of proteins as biosensors, industrial reagents, and therapeutics. We have developed a structure-based method to rationally predict single mutations at protein–protein interfaces that enhance binding affinities. The method is based on the premise that increasing buried hydrophobic surface area and/or reducing buried hydrophilic surface area will generally lead to enhanced affinity if large steric clashes are not introduced and buried polar groups are not left without a hydrogen bond partner. The procedure selects affinity enhancing point mutations at the protein–protein interface using three criteria: (1) the mutation must be from a polar amino acid to a non-polar amino acid or from a non-polar amino acid to a larger non-polar amino acid, (2) the free energy of binding as calculated with the Rosetta protein modeling program should be more favorable than the free energy of binding calculated for the wild-type complex and (3) the mutation should not be predicted to significantly destabilize the monomers. The performance of the computational protocol was experimentally tested on two separate protein complexes; $G\alpha_{i1}$ from the heterotrimeric G-protein system bound to the RGS14 GoLoco motif, and the E2, UbcH7, bound to the E3, E6AP from the ubiquitin pathway. Twelve single-site mutations that were predicted to be stabilizing were synthesized and characterized in the laboratory. Nine of the 12 mutations successfully increased binding affinity with five of these increasing binding by over 1.0 kcal/mol. To further assess our approach we searched the literature for point mutations that pass our criteria and have experimentally determined

binding affinities. Of the eight mutations identified, five were accurately predicted to increase binding affinity, further validating the method as a useful tool to increase protein–protein binding affinities.

INTRODUCTION

Engineered proteins are increasingly being used as therapeutics and as tools to probe cell biology.[1; 2; 3] Often, the effectiveness of these proteins depends in part on their affinity for their target ligand or protein. Directed evolution and combinatorial screening techniques such as phage display or ribosome display are successful and well accepted means of engineering protein complexes with enhanced affinity.[4; 5] These techniques, however, are labor intensive and are difficult to perform in cases where the target protein is difficult to express or is not stable under the conditions needed for binding selection. An alternative approach is to use structure-based modeling to predict affinity-increasing mutations. The limitation of this approach is that it requires a high resolution structure of the protein–protein interface that is being optimized. Benefits are that it has the potential to be a rapid way to identify stabilizing mutations and it can be used in cases where combinatorial screening is not feasible.

In general, two types of structure-based approaches have been used to enhance protein–protein binding affinities. One approach has focused on increasing the electrostatic attraction between proteins by mutating residues around the periphery of the binding interface.[6; 7; 8; 9] The rationale is that electrostatic interactions can work over long distances and therefore the additional charges do not need to be immediately adjacent to the partner protein.

The benefit of restricting mutations to the periphery is that the specific packing and hydrogen bonding that is often found at interfaces is not disrupted. This approach has been used with good success. Mutations have been identified that increase the affinity between TEM1 β-lactamase and its protein inhibitor BLIP by over 250-fold,[8] and the affinity between Ral and Ral guanine nucleotide dissociation stimulator by over 25-fold.[9] One limitation of this approach is that large increases in binding affinity are only possible in cases where there is an excess of like-charge on one side of the interface that is not already paired with the opposite charge on the other side of the interface.

An alternative approach is to use side-chain repacking algorithms to search for mutations that lead to better packing, hydrogen bonding, and desolvation energies at the interface. These protocols generally model protein energetics with a linear combination of terms that model van der Waals forces, steric repulsion, backbone and side-chain torsional energies, hydrogen bonding, desolvation energies and electrostatics. Often the protein backbone is held rigid and side-chains are restricted to the most commonly observed conformations in the Protein Data Bank (PDB), typically referred to as rotamers. This technology, often referred to as computational protein design, has been used to design new protein structures,[10] create new enzymes,[11; 12; 13] stabilize proteins[14; 15; 16; 17; 18] and perturb protein–protein binding specificities.[19; 20; 21; 22; 23] Recently, two studies made extensive use of a variety of protein design algorithms in an effort to enhance protein–protein binding affinities. Springer and colleagues used Protein Design Automation (PDA), Sequence Prediction Algorithm, and Rosetta

to search for mutations that would stabilize the interaction between integrin lymphocyte function-associated antigen-1 (LFA-1) and its ligand intercellular adhesion molecule-1 (ICAM-1).[24] Mixed results were obtained with the algorithms. Out of 24 single and double mutations that were selected for experimental characterization, four of the redesigns increased binding by over 1.5-fold. In a separate study, Clark and co-workers used a side-chain repacking algorithm to search for mutations that would increase the affinity of an antibody for its protein target.[25] Most of the point mutations studied did not significantly increase binding affinity, but by combining three mutations they were able to stabilize binding by 1.2 kcal/mol. Interestingly, in the Springer study three out of the four stabilizing mutations were from a polar amino acid to a non-polar amino acid, and in the Clark study two out of the three stabilizing mutations changed a polar amino acid to a non-polar amino acid. Although this is a small test set, this result suggests that side-chain repacking algorithms may be most successful at enhancing protein–protein binding affinities when limited to mutations that decrease the amount of buried hydrophilic groups and increase the number of buried hydrophobic groups.

The hydrophobic effect drives protein folding, and hydrophobic interactions often contribute significantly to protein–protein binding affinity. Removing a buried methylene or methyl group by site-directed mutagenesis almost always destabilizes a protein, often by more than 1 kcal/mol.[26; 27] Introducing new methylene groups can stabilize a protein, and side-chain repacking algorithms have been used to stabilize proteins by identifying residues

within a protein core that can accommodate larger hydrophobic amino acids.[15; 16; 17; 28] Burying hydrophilic groups has the opposite effect on protein stability. There is a large desolvation cost associated with placing a polar amino acid in the interior of a protein, to the point where buried polar groups are only found in positions where they can form hydrogen bonds with other polar groups in the protein. The balance between hydrogen bond energy and desolvation energy plays a large part in determining the favorability of a buried polar amino acid.[28; 29; 30] Accurately calculating this balance is difficult, in part because the calculation is sensitive to small structural perturbations given that hydrogen bonds have a sharp distance and orientation dependence.[30; 31] In addition, cooperative effects in hydrogen bond networks may change the average hydrogen bond energy within the network.[32]

Here, we test our ability to predict point mutations that will enhance protein–protein binding affinities. Because accurately calculating the favorability of a buried polar amino acid is difficult, we focus on mutations that do not rely on hydrogen bonding to overcome desolvation energies. We consider two types of mutations at protein interfaces: mutation from a hydrophobic amino acid to a larger hydrophobic amino acid, and mutation of partially buried polar groups, not involved in side-chain hydrogen bonding, to a non-polar amino acid. We model the mutations and calculate binding energies using Rosetta, protein modeling software that has been developed for protein structure prediction and design.[33] We use two separate model protein systems to experimentally validate our protocol; from the G protein system we use the $G\alpha_{i1}$ protein with one of its

binding partners, the GoLoco motif from the Gα regulator, RGS14,[34] and from the ubiquitin pathway, we use the ubiquitin-conjugating enzyme UbcH7 along with one of its binding partners, E6-associated protein (E6AP) (Figure 2.1).[35] We chose these interfaces as model systems because they are large (>2000 $Å^2$), which increases the probability of finding point mutations that meet our selection criteria. Additionally, these systems allow us to test our protocol on both a protein–peptide ($Gα_{i1}$–GoLoco) and a protein–protein interaction (UbcH7–E6AP). We do not explicitly model backbone conformational change in the unbound state, and therefore, our modeling procedure does not differentiate between a flexible peptide and a rigid protein. If mutations dramatically affect the unbound conformation of the GoLoco peptide, we may observe less accurate predictions for the GoLoco–$Gα_{i1}$ interface.

RESULTS

For both model systems, $G\alpha_{i1}$-GoLoco and UbcH7-E6AP, each residue at the protein–protein interface was sequentially mutated to the non-polar amino acids (V,I,L,M,F,Y,W) and changes to binding energy were predicted with the Rosetta energy function. In cases where a non-polar amino acid was being mutated to another non-polar, only mutations that increased the size of the side-chain were considered. Binding energies were calculated by subtracting the calculated energy of each unbound protein from the calculated energy of the complex. In the first round of calculations, the backbone and side-chain conformations of the unbound proteins were assumed to be identical to those in the bound state. The side-chain of the mutated residue was built by choosing the rotamer with the lowest energy when modeled in the context of the complex. Neighboring side-chains were held fixed in the positions observed in the crystal structure.

Predicting affinity-enhancing mutations at the GoLoco–$G\alpha_{i1}$ interface

The 36-residue GoLoco motif peptide of RGS14 stretches across the $G\alpha_{i1}$ surface, interacting with both the Ras-like, guanine nucleotide binding domain and the all-helical domain of $G\alpha_{i1}$.[34] A total of 28 residues from the GoLoco motif and 59 residues from $G\alpha_{i1}$ interact at the protein–peptide interface. Each

interface position was sequentially mutated to a non-polar amino acid and the binding energies were calculated. From the 503 mutations that were considered, 55 have calculated binding energies that are more favorable by 0.5 kcal/mol or more than the wild-type protein. However, many of these mutations are predicted to create unfavorable intramolecular interactions ($\Delta\Delta G^{\circ}_{chain}$ in Table 2.1), which will indirectly weaken binding by disfavoring the backbone and side-chain conformations adopted in the bound state. For example, residue 203 in wild-type $G\alpha_{i1}$ is a glycine with a phi angle in a region of the Ramachandran plot that is disfavored for most amino acids. Because we are assuming that the conformation of the unbound state is the same as the bound state, when this glycine is mutated to a leucine, our binding energy calculation cancels out the repulsive energy between the leucine and its local backbone environment. To account for this weakness in our protocol, we calculate a folding energy for each chain of the complex (see METHODS). A total of 22 of the 55 mutations predicted to have favorable binding energies were removed from consideration because the energy of the individual chains increased by more than 1 kcal/mol.

Out of the 33 mutations that passed our filters, two mutations from $G\alpha_{i1}$ and four mutations from the GoLoco motif were selected for biophysical characterization. Binding affinities were measured by labeling the GoLoco motif peptide with the fluorescent dye fluorescein and monitoring fluorescence polarization as a function of $G\alpha_{i1}$ concentration. Four out of six mutations resulted in an increase in binding affinity, and in three cases the increase was greater than 1 kcal/mol (Table 2.2). The largest increase was from the mutation F529W,

which lowered the K_d from 95 nM to 5.9 nM (Figure 2.2). A model of one of the affinity-increasing mutations (E116L) is shown in Figure 2.3. In the wild-type structure, E116 is not involved in any hydrogen bonds with other protein atoms, and is partially shielded from water by leucine 518 on the GoLoco motif peptide. In the model of the mutant structure (E116L), the new leucine forms a close hydrophobic interaction with leucine 518, and the desolvation energy for complex formation is predicted by Rosetta to be −1.6 kcal/mol more favorable for the mutant than for the wild-type complex.

The biggest discrepancy between the predicted and experimentally derived binding energies was observed for the mutation, Q508L, in the GoLoco peptide. Glutamine 508 is buried in the middle of the interface, and its side-chain forms a hydrogen bond with a buried water molecule as well as the backbone nitrogen of residue 40 on $G\alpha_{i1}$. The water also makes hydrogen bonds with groups on $G\alpha_{i1}$. The water was not included in the Rosetta simulations and Rosetta predicted the mutation to leucine to be favorable because the side-chain on the glutamine did not satisfy its hydrogen bonding potential in the Rosetta simulation. Experimentally, Q508L destabilizes binding by more than tenfold. In the future, the calculations may be improved by explicitly modeling buried water molecules.

Predicting affinity-enhancing mutations at the UbcH7–E6AP interface

Over 30 residues interact across the UbcH7–E6AP protein–protein interface. The structure of E6AP in the unbound state is virtually identical to the conformation that it adopts when forming a complex.[35] Likewise the UbcH7 fold observed in the UbcH7-E6AP complex is similar to that of numerous unbound structures of other ubiquitin conjugating enzymes.[36] The interaction is dominated by a phenylalanine on UbcH7 (F63) that packs into a hydrophobic pocket on E6AP.[35] Mutation of the phenylalanine to an alanine destabilizes binding by over 3 kcal/mol.[37] As with the GoLoco motif/$G\alpha_{i1}$ system, each interface position was sequentially mutated to a non-polar amino acid and binding energies were calculated with Rosetta. Relatively few mutations are predicted to be stabilizing (Table 2.3). The residues that are packed around F63 are primarily hydrophobic and mutating them to larger non-polar amino acids leads to steric clashes. Thirteen mutations were predicted to stabilize binding by greater than 0.5 kcal/mol and not destabilize the individual chains by more than 1.5 kcal/mol. Most of these mutations are located on the periphery of the protein–protein interface.

Six mutations were selected for experimental characterization, D641Y, D641W, Q637W on E6AP and A98W, K64L, F63W on UbcH7. Again, a fluorescence polarization binding assay was employed to measure binding affinities but in this case a thiol-reactive bodipy fluorophore was used. Five of the

six mutations enhance binding. The A98W mutation lowers the K_d from 5 μM to 190 nM (Table 2.2). As with the mutations that enhanced affinity at the GoLoco motif–$G\alpha_{i1}$ interface, these changes create new hydrophobic interactions across the interface. A model of the A98W mutation on UbcH7 is shown in Figure 2.3. The new tryptophan is predicted to pack against D641, Y645 and M653 on E6AP.

Testing the protocol on other protein–protein complexes

To further test our protocol we searched the literature for point mutations that replace a polar amino acid with a non-polar amino acid or increase the size of a non-polar amino acid and have an experimentally determined $\Delta\Delta G^\circ_{binding}$. There are hundreds of point mutations characterized in the literature, but most of them involve mutation to alanine and therefore are not useful to our test. We identified 38 mutations from nine different protein complexes that fit our criteria, and Rosetta was used to predict the change in free energy of binding of each mutation. Eight of the mutations were predicted to increase binding affinity by more than 0.5 kcal/mol and not destabilize the individual proteins by more than 1.0 kcal/mol. Five of the eight did in fact enhance binding affinity. Of the 38 mutations that were examined, there were two mutations that enhanced binding by more than 0.5 kcal/mol that were not predicted to increase affinity by Rosetta.

Looking more closely at the mutations that were falsely predicted to enhance binding affinity, we noticed that in two cases the mutation removed a hydrogen bond across the protein interface. The hydrogen bonding term used in

the Rosetta energy function evaluates both the distance and the angle of the hydrogen bond, and the energy falls off rapidly with small changes in distance or angle. This method may be too stringent when used to predict energy from unminimized crystal structures. If Rosetta considers the hydrogen bond to be less than ideal, the energetic penalty for removing it can be insignificant compared to the other energy terms, resulting in a false prediction that the mutation is stabilizing to the complex. If we add a third filter to our scheme that requires $\Delta\Delta G^{\circ}_{binding}$ of the hydrogen bond energy term to be zero or less following a mutation, then two of the false positive predictions are removed from our results but all five of the correct predictions remain intact.

Allowing backbone and side-chain relaxation in the energy calculations

All of the results presented so far were based on calculations in which only the side-chain of the residue being mutated was allowed to sample alternate conformations in search of lower energy structures. The neighboring residues were not allowed to relax to better accommodate the mutation and the unbound structures of the proteins were assumed to be identical to the bound structure. In reality, many proteins can relax to lower free energy conformation(s) in the unbound state, and binding energies are partially determined by how much free energy is needed to bring the protein into the bound conformation. To test if our protocol for identifying affinity enhancing mutations can be improved by allowing relaxation in the bound and/or unbound states, we performed a variety of

simulations in which different elements of structure were allowed to relax. In each case, the method being tested was used to predict the $\Delta\Delta G^{\circ}_{binding}$ of point mutations that replace a polar amino acid with a non-polar amino acid or increase the size of a non-polar amino acid. The test set included point mutations from our laboratory as well as from literature and included 57 mutations from 11 different protein complexes.

Seven protocols were tested in addition to the standard fixed backbone approach. We considered three levels of relaxation, and one additional energy term. The first flexibility option allows for the residues neighboring a mutated position to adopt alternate side-chain coordinates (repack neighbors). Low energy conformations are identified using Monte Carlo optimization of side-chain rotamers. The second flexibility option allows the interface side-chains of the unbound proteins to adopt alternate rotamers (relax unbound). This option is designed to more closely model the physical behavior of proteins by allowing the unbound proteins to adopt low energy side-chain conformations that may be incompatible with the bound state. The third flexibility option allows for backbone relaxation. Gradient-based minimization of backbone torsion angles, side-chain torsion angles and rigid-body orientation of the two proteins is used to identify small conformational changes that lower the energy of the complex or the unbound structures. Gradient-based minimization was performed with and without constraints derived from the wild-type structure. In both cases only small motions were observed; the backbone typically moved by less than 0.2 Å rmsd. The fourth option modified the energy function. Our standard energy function for

these studies has a Lennard-Jones potential with a damped repulsive term to compensate for the fixed backbone and rotamers.[17] In one set of simulations with gradient-based minimization we used a stiffer repulsive term to more accurately model the energy required to bring two atoms near each other. These four options were combined in a variety of ways to test the effects of including backbone and side-chain relaxation on predicting $\Delta\Delta G^{\circ}_{binding}$.

We compared the protocols by examining the percentage of correctly predicted affinity enhancing mutations (true positives) and the percentage of recovery of the known stabilizing mutations (20 out of the 57 mutations were experimentally determined to be stabilizing). As before, we only considered mutations that were predicted to stabilize binding by more than 0.5 kcal/mol and have chain energies that increase by less than 1 kcal/mol. In addition, to be considered stabilizing there could not be a decrease in hydrogen bonding energy across the interface. In general, including more flexibility did not improve the results. With the standard protocol (no flexibility) 83% of the mutations that were predicted to be stabilizing were indeed stabilizing and the recovery of experimentally determined stabilizing mutations by modeling was 50%. When the side-chains were allowed to relax to new rotamers in the bound and/or the unbound state there was a fall off in performance (Table 2.4). The percentage of true positives was lower and the recovery rate was decreased as well. It is not straight forward to determine why relaxing the side-chains did not improve the predictions, but overall the results suggest that some of the relaxed side-chains were placed incorrectly, which can have significant effects on the calculated

binding energies. In side-chain prediction benchmarks Rosetta performs similarly to most other side-chain repacking algorithms, on average 70% of the residues with chi 1 and chi 2 torsion angles are modeled with chi 1 and chi 2 within 30 degrees of the angles in the crystal structures. This indicates that we should expect some of the side-chains at the interface to be relaxed incorrectly when forcing mutations.

To further examine the predictions with flexible side-chains we looked in detail at two mutations whose binding energies were predicted correctly with fixed side-chains, but were predicted incorrectly when allowing the side-chains to relax (Figure 2.4). Mutating aspartic acid 641 to a tryptophan enhances binding affinity between E6AP and UbcH7 by 0.9 kcal/mol. The fixed side-chain simulations accurately predict that this mutation will favor tighter binding. When the side-chains are allowed to relax the new tryptophan and lysine 96 on UbcH7 adopt alternate rotamers that creates a new packing arrangement between the side-chains. This packing arrangement lowers the total calculated energy for the complex as a whole (in part because the internal energy of W641 and K96 drops by 1.9 kcal/mol), but it does not lower the calculated binding energy, or the energy across the interface of the complex. A new steric clash is introduced between the side-chain of W641 and the backbone oxygen of residue 95 that makes the predicted change in binding energy unfavorable. This result highlights the interplay between the various energy terms in Rosetta. A drop in one energy term can lead to an increase in another. Small errors in how the terms are

balanced can lead to possible error in side-chain prediction and hence the calculated binding energy.

Mutating threonine 662 to a phenylalanine destabilizes binding between E6AP and UbcH7 by 0.2 kcal/mol. The simulation with fixed side-chains predicts that this mutation will destabilize binding by 0.7 kcal/mol, in part because there is a steric clash between F662 and E60 on UbcH7. In the flexible side-chain simulations E60 adopts a new rotamer to make room for the phenylalanine, and binding is predicted to be stabilized by 3.0 kcal/mol. In the crystal structure of the wild-type protein E60 is making a strained hydrogen bond with its own backbone amide group. In the relaxed structure this hydrogen bond is broken and the amide is buried by the phenylalanine. This may be a scenario where Rosetta is not properly penalizing the burial of a polar group that has no hydrogen bond partners. Again, this result highlights the sensitivity of side-chain prediction and binding energy calculations to how the various energy terms are balanced.

Gradient-based minimization was used in a variety of ways: without repacking, with repacking, and with relaxation of the unbound state plus repacking. In most cases the rate of true positives was equal if not better than the fixed backbone simulation. However, the recovery rate was generally lower with minimization. In one case, all the mutations predicted to be stabilizing were stabilizing, but the rate of recovery fell to 25%. The rate of recovery was probably lower because the energies of the wild-type structures were more favorable, making many of the mutations appear less favorable.

DISCUSSION

Overall, the results suggest that increasing buried hydrophobic surface and/or reducing buried hydrophilic surface is an efficient approach for enhancing protein–protein binding affinities. In addition, it is important that the mutations do not create large clashes, do not destabilize the individual chains, and do not remove key hydrogen bonds across the interface. With this approach we selected 12 point mutations for experimental characterization and showed that nine of them increase protein binding affinity. In addition, mutations from the literature that pass our filters have a greater than 60% chance of being stabilizing. These are encouraging results considering that most randomly chosen mutations at a protein interface will weaken binding affinity.

Our energy filters do remove from consideration some mutations that increase protein–protein binding affinities. Our standard fixed-backbone protocol identified approximately half of the known stabilizing mutations from our test set. We were curious if adding more side-chain and backbone flexibility in our simulations would decrease the number of stabilizing mutations that were not identified. In particular, relaxation may allow clashes to be relieved with new amino acids and allow for lower binding energies. In general, we observed the opposite result. Relaxing the system lowered the energy of the wild-type interaction and made more mutations appear unfavorable.

It is difficult to determine why the protocols with more flexibility did not outperform the fixed system model. There may be a combination of reasons. First of all, a number of stability and structural studies have been done that demonstrate that point mutations result in little to no movement away from the wild-type crystal structure coordinates other than in the immediate area of the mutation.[27; 38; 39] Secondly, the community-wide critical assessment of protein structure prediction test from 2004 demonstrates that while modeling has improved significantly over the past decade, it is still very difficult to refine a nearly correct protein model to a more correct model.[40] It may be that our flexible backbone and side-chain relaxation procedures are incorrectly predicting the structure of the mutant; in this case they would not be expected to provide a better estimate of binding energy. Kortemme and co-workers when computationally modeling protein–protein energetic hot spots, noted that optimizing rotameric side-chain conformations did not significantly improve their predictions except for the case of staphylococcal enterotoxin C3 bound to the T cell receptor β chain, a low resolution structure.[41] The improvement in this case seemed to be due to the changes made in the native complex.

One potential downside of our approach for increasing binding affinity is that it may lead to lower protein solubility as more hydrophobic surface area is exposed on the surface of the protein. For the 12 mutations that we characterized we did not see any reduction in solubility as evidenced by gel filtration chromatography. However, it is unlikely that one could combine several mutations of this type and still have a highly soluble protein. In general, we feel

that our protocol will be most useful for finding one or two mutations that provide a 1 or 2 kcal/mol increase in binding affinity, and then combining these mutations with other types of mutations to provide a larger increase in affinity.

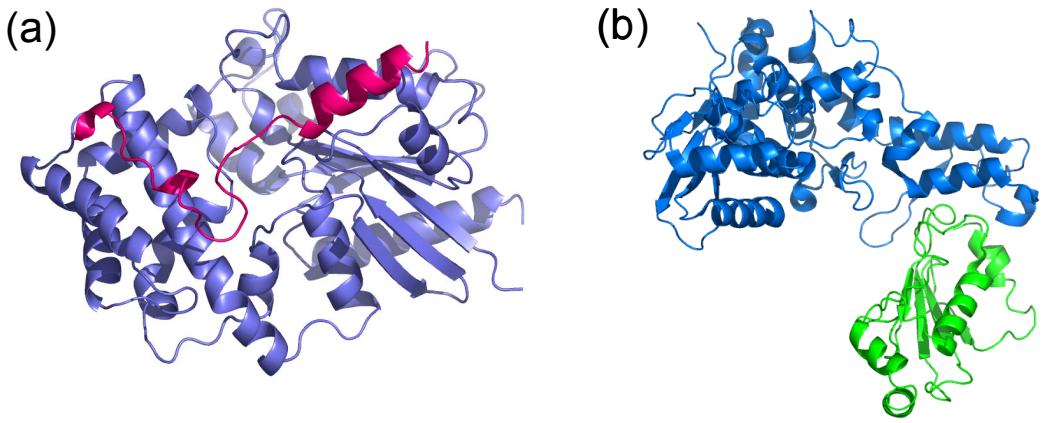


Figure 2.1. Structures of protein complexes selected for experimental validation of computational protocol. (a) $G\alpha_{i1}$ shown in purple with the GoLoco domain of the RGS14 protein shown in magenta. (b) E6AP shown in blue bound to UBCH7 shown in green.

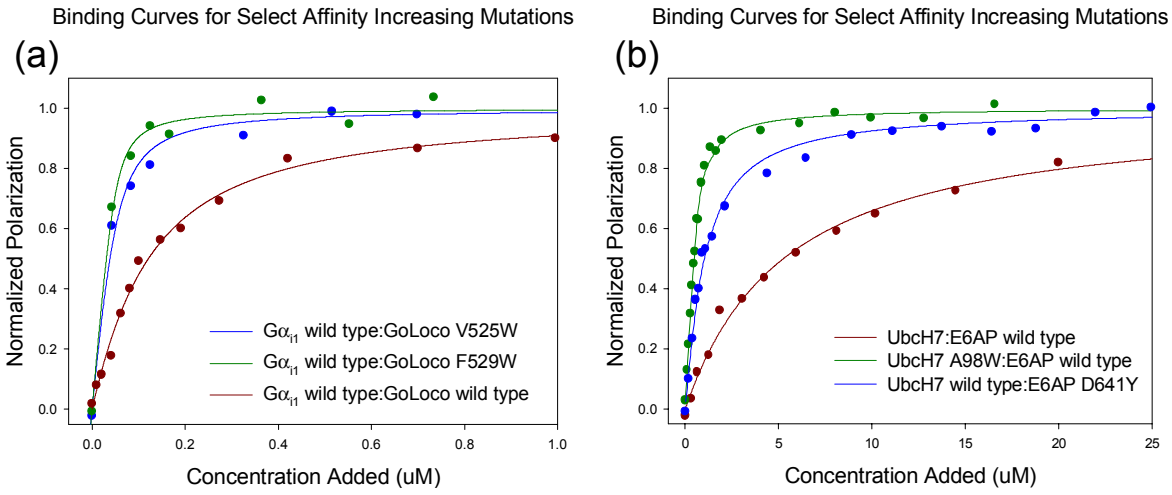


Figure 2.2. Binding curves for select affinity increasing mutations compared to wild type for (a) $G\alpha_{i1}$:GoLoco and (b) E6AP:UbcH7.

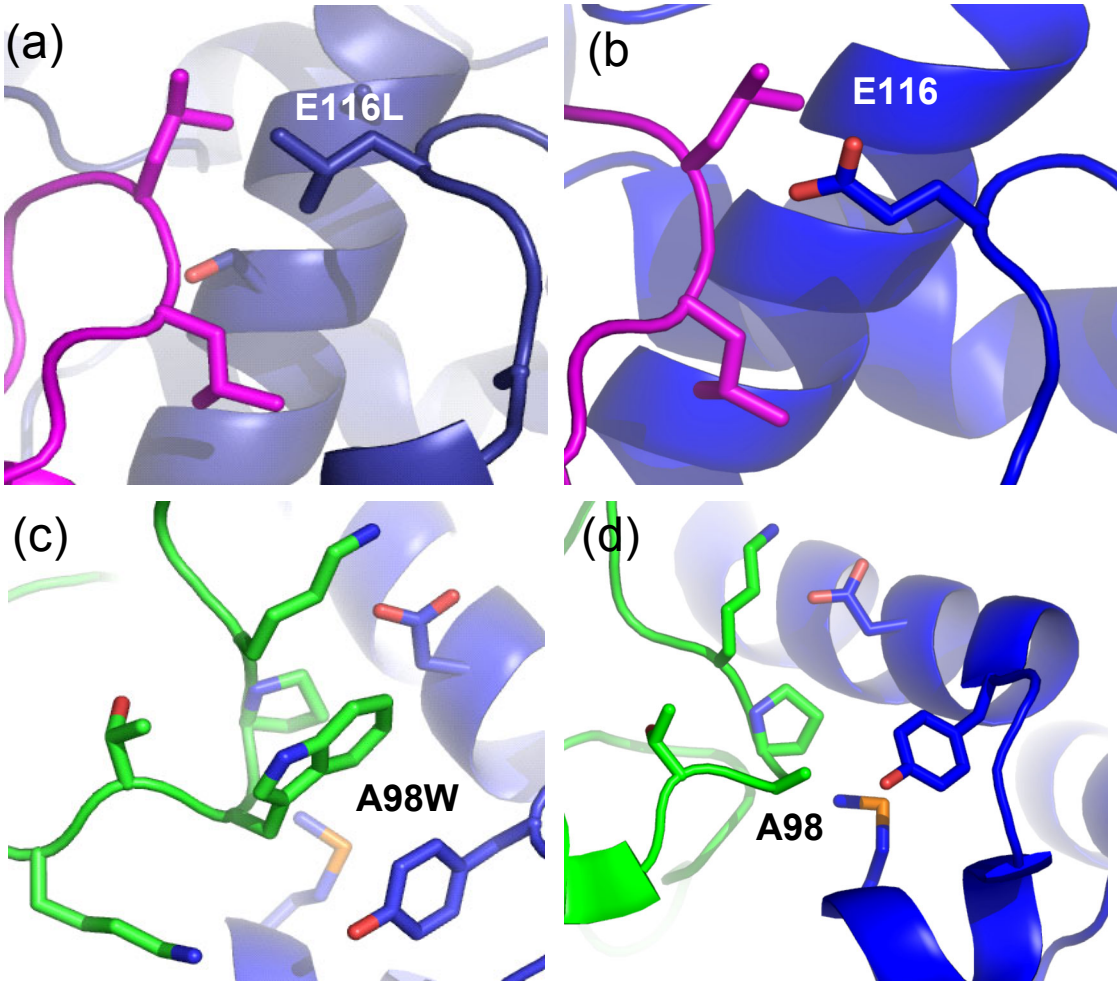


Figure 2.3. Modeled structure of the affinity enhancing design for (a) E116L $G\alpha_{i1}$ bound to wild type GoLoco compared to (b) the crystal structure of the wild type $G\alpha_{i1}$ bound to wild type GoLoco and (c) the modeled structure of A98W UbcH7 bound to wild type E6AP compared to (d) the crystal structure of the wild type UbcH7 bound to wild type E6AP..

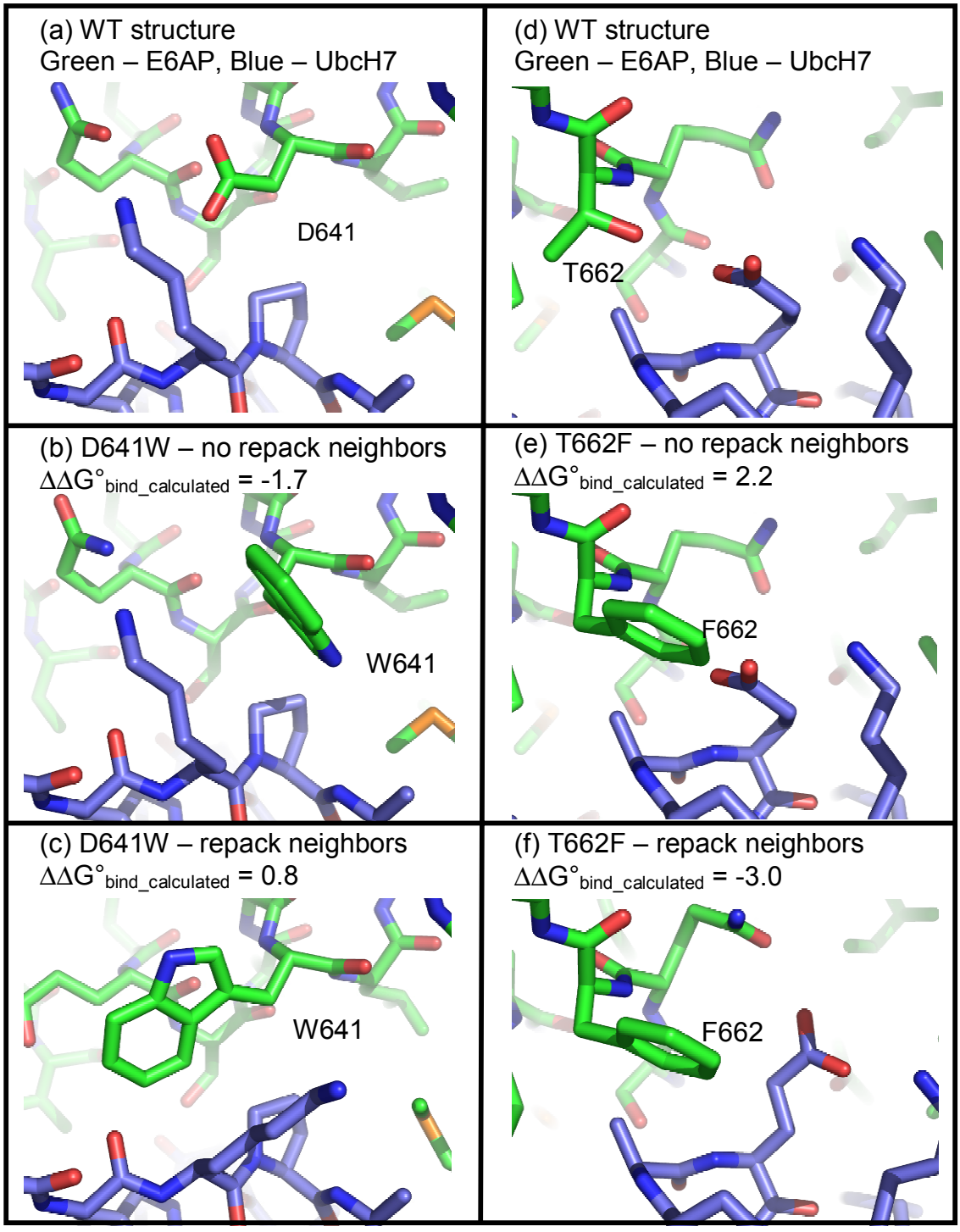


Figure 2.4. Models for the mutations D641W and T662F at the interface of E6AP and UbcH7. (a) and (d) Wild type residues; (b) and (e) the mutations modeled without repacking neighbors; (c) and (f) the mutations modeled with repacking the neighboring residues. The calculated changes in binding energy are indicated. Experimentally, D641W stabilizes binding ($\Delta\Delta G^{\circ}_{bind}$=-0.9 kcal/mol) and T662F destabilizes binding ($\Delta\Delta G^{\circ}_{bind}$=0.2 kcal/mol).

Table 2.1. Scanning for affinity enhancing mutations at the A:Gα_{i1} – B:GoLoco interface.

Mutation	*ΔΔG°$_{bind}$*[a]	*ΔΔG° $_{chain\ A}$*[b]	*ΔΔG° $_{chain\ B}$*	*# neighbors*[c]	*ΔΔG°$_{h\text{-}bond}$*[d]
A:G203W	-2.8	5.7	0	23	0
A:E116W	-2.4	1.2	0	12	0
B:I497W	-2.1	0	0.6	14	0
A:E116F	-1.8	1.7	0	12	0
A:G203Y	-1.8	3.1	0	23	-0.7
A:R86F	-1.7	-2.9	0	23	0
A:G203F	-1.7	4.4	0	23	0
A:E116L	-1.5	-0.7	0	12	0
B:L524W	-1.5	0	7.8	18	0
A:V72W	-1.4	0.1	0	25	0
A:E116Y	-1.3	0.9	0	12	-0.4
A:E116I	-1.1	0.6	0	12	0
B:S510V	-1.1	0	2.5	23	0
B:L524F	-1.1	0	9.2	18	0
B:F529W	-1.1	0	-2.4	16	0
B:L518F	-1	0	1.3	16	0
B:L518Y	-1.1	0	0.6	16	0
A:V72F	-1	-1.2	0	25	0
B:Q508L	-1	0	-8	26	0.8
A:V72Y	-1	-1.7	0	25	0
A:G112W	-1	3	0	16	0
B:Q508I	-1	0	-7.8	26	0.8
B:S510W	-0.9	0	3.7	23	0
A:S252W	-0.8	0.9	0	16	0
A:S252Y	-0.8	0.4	0	16	0
A:N256W	-0.8	1.6	0	16	0
A:R242W	-0.7	-1.1	0	27	-0.1
A:Q147M	-0.7	0.4	0	20	0
A:R86W	-0.7	-5.7	0	23	0
A:R242Y	-0.7	-1.4	0	27	0
A:Q147L	-0.7	-1.2	0	20	0
A:G203M	-0.7	4.5	0	23	0
A:S246F	-0.6	0.3	0	27	0
A:E116M	-0.6	1.2	0	12	0
A:E116V	-0.6	0	0	12	0
B:V525W	-0.6	0	-1.3	12	0
A:Q147W	-0.6	0	0	20	0
A:Q147F	-0.6	-0.7	0	20	0
A:Q147Y	-0.6	-0.7	0	20	0
A:G203L	-0.6	2.8	0	23	0
A:G202F	-0.6	7.9	0	26	0.1
B:L518W	-0.6	0	1.3	16	0
B:R516F	-0.6	0	-0.9	22	0.8
A:K46M	-0.5	0.2	0	29	0
B:E498M	-0.5	0	1.1	10	0
A:N256M	-0.5	0.4	0	16	0
B:S510M	-0.5	0	2.8	23	0
A:A101Y	-0.5	3.3	0	15	0
A:S246Y	-0.5	0.2	0	27	0
A:R242F	-0.5	-1.1	0	27	0
A:A101F	-0.5	3.8	0	15	0
A:G203I	-0.5	2.5	0	23	0
A:F215W	-0.5	-2	0	21	0
A:V72M	-0.5	-0.8	0	25	0
B:R516Y	-0.5	0	-2.3	22	0.8

Mutations selected for biophysical study are highlighted in orange. Mutations highlighted in gray were not considered for biophysical study because were predicted to destabilize the monomer.

[a] Predicted change in binding energy (kcal / mol) with the Rosetta Energy function.

[b] Predicted change in folding energy (kcal / mol) of the isolated chain.

[c] Number of residues in the complex within 10Å of the mutation.

[d] Predicted change in hydrogen bond energy across the interface.

Table 2.2. Binding Affinities for Experimentally Characterized Mutants

Mutation	$\Delta\Delta G°_{Rosetta}$	$\Delta\Delta G°_{exp.}$	K_d***exp. (μM)***
wild type $G\alpha_{i1}$:GoLoco	0	0	0.098
$G\alpha_{i1}$ E116L	-1.5	-1.05	0.016
$G\alpha_{i1}$ Q147L	-0.7	-0.84	0.023
GoLoco Q508L	-1	3.17	>20
GoLoco L518Y	-1.1	0.07	0.110
GoLoco V525W	-0.6	-1.16	0.014
GoLoco F529W	-1.1	-1.65	0.006
wild type UbcH7:E6AP	0	0	5.0
UbcH7 A98W	-1.7	-1.9	0.19
E6AP D641Y	-1.7	-1.1	0.8
E6AP D641W	-1.7	-0.86	1.2
UbcH7 K64L	-0.7	-0.54	2.0
UbcH7 F63W	-0.6	0.79	16.9
E6AP Q637W	-0.6	-0.64	1.7

Table 2.3. Scanning for affinity enhancing mutations at the E6AP – UbcH7 interface.

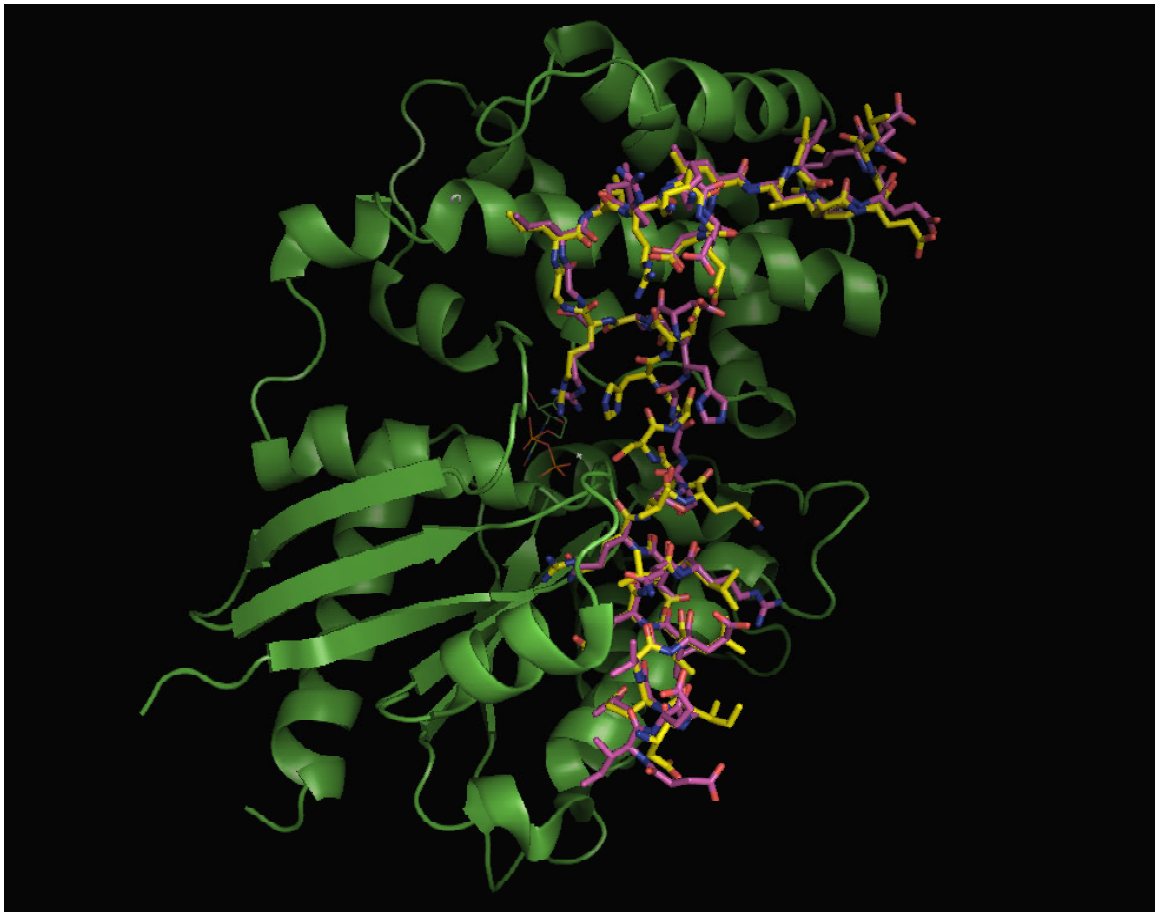
Mutation	*ΔΔG°$_{bind}$*	*ΔΔG°$_{chain\ A}$*	*ΔΔG°$_{chain\ B}$*	*# neighbors*	*ΔΔG°$_{h\text{-}bond}$*
D:A92F	-2.7	0	9	13	0
D:K9W	-1.8	0	1.6	12	0.2
A:D641F	-1.8	1.9	0	16	0.1
A:D641Y	-1.7	1.2	0	16	0.1
A:D641W	-1.7	2.1	0	16	0.1
D:A98W	-1.7	0	-0.1	15	0
D:N31W	-1.6	0	3.1	14	0
D:L33W	-1.5	0	3.2	22	0
D:A92Y	-1.5	0	7.3	13	0
D:A92W	-1.4	0	6	13	0
D:L33F	-1.1	0	0.2	22	0
D:A98F	-1.1	0	1.3	15	0
D:A98V	-1.1	0	2	15	0
D:R6W	-1.1	0	3.8	18	0.1
D:A98Y	-1	0	0.7	15	0
A:L639Y	-0.9	14.4	0	23	0
D:K64L	-0.7	0	-1	19	0
A:M653W	-0.7	-5.8	0	24	0
D:N31M	-0.6	0	1.7	14	0
A:Q637W	-0.6	-1.9	0	20	0
D:P62I	-0.6	0	7.7	26	0
A:D641V	-0.6	0.9	0	16	0.1
D:R6Y	-0.6	0	5	18	0.1
A:S660W	-0.6	13.8	0	17	0
D:F63W	-0.6	0	0.2	23	0
D:N31I	-0.5	0	0	14	0
A:T656I	-0.5	2.9	0	20	0
D:E60I	-0.5	0	-2.3	20	0.1
D:P58W	-0.5	0	4	18	0
D:A92M	-0.5	0	5	13	0
A:T662M	-0.5	-3.4	0	15	0.1

Mutations selected for biophysical study are highlighted in orange. Mutations highlighted in gray were not considered for biophysical study because were predicted to destabilize the monomer.
[a] Predicted change in binding energy (kcal / mol) with the Rosetta Energy function.
[b] Predicted change in folding energy (kcal / mol) of the isolated chain.
[c] Number of residues in the complex within 10Å of the mutation.
[d] Predicted change in hydrogen bond energy across the interface.

Table 2.4. Predicting affinity enhancing mutations with varying degrees of side chain and backbone flexibility.

	% true positive	*% recovery (out of 20 possible)*
Default	83	50
R	60	15
RR	67	20
M	89	40
RM	67	20
RRM	83	25
RRM 2nd	100	15
RRMncst	100	25
Full LJ, RRM	63	50

% true positive is the percentage of mutations predicted to increase affinity that actually increase affinity. Mutations that disrupt hydrogen bonding across the interface or destabilize the monomers by more than 1 kcal / mol were filtered out. % recovery is the number of correctly predicted affinity enhancing point mutations divided by the total number of experimentally determined affinity enhancing point mutations. The test set contains 57 mutations (polar to hydrophobic, hydrophobic to bigger hydrophobic), 20 of which increase binding affinity.
Default = fixed backbone and side chains (except for the site of mutation)
M = minimize backbone
R = repack neighbors
RM = repack neighbors, minimize backbone
RR = repack neighbors, relax unbound
RRM = repack neighbors, relax unbound, minimize backbone
RRM 2nd = RRM run a second time
RRMncst = repack neighbors, relax unbound, minimize backbone with no constraints
full LJ, RRM = full Lennard-Jones repulsion, repack neighbors, relax unbound, minimize backbone



Supplementary Figure 2.1. Comparison of the 2.2 angstrom structure of R14GL with the 2.7 angstom structure. Gα subunits were aligned and is shown in green ribbon. R14GL from 1KJY is shown in yellow sticks and the R14GL peptide from 2OM2 is shown in purple stick form. GDP is shown in line form.

METHODS

Rosetta

All energy calculations and side-chain and backbone relaxation simulations were performed with the molecular modeling program Rosetta.[33] Rosetta's core full atom energy function is a linear sum of molecular mechanics and knowledge-based terms: a 6–12 Lennard-Jones potential, the Lazaridis-Karplus implicit solvation model,[42] an empirically based hydrogen bonding potential,[31] backbone-dependent rotamer probabilities,[43] a knowledge-based electrostatics energy potential, amino acid probabilities based on particular regions of Φ/ψ space, and reference energies that approximate the energies of amino acids in the unfolded state.[44] Within Rosetta, there are several variations on this core energy function. For the studies described here, we primarily use a version of the energy function that was recently parameterized to best reproduce native sequences when redesigning whole proteins in fixed backbone simulations (command line option, -soft_rep_design, Rosetta v 2.1). This variation of the energy function significantly dampens repulsion energies to allow for small atom–atom clashes that may be accommodated by small changes in side-chain and backbone conformation. See supplementary material of Dantas *et al*. for a complete description of this version of the Rosetta energy function. It is referred to as Rosetta_DampRep.[17]

Varying degrees of backbone and side-chain flexibility were used as described in RESULTS. Side-chain flexibility is modeled by allowing amino acids to adopt different rotamers. We use Dunbrack's backbone-dependent rotamer library supplemented with rotamers that vary chi 1 and chi 2 one standard deviation away from their most probable values.[43] Low energy combinations of side-chain conformations are identified using Monte Carlo optimization with simulated annealing.[44] Independent side-chain repacking simulations typically converge to very similar energies (standard deviation < 0.1 kcal/mol) but generally do not have identical structures. The structures are not identical because some amino acids have alternate conformations that are isoenergetic (in the Rosetta energy function) in the absence of non-local interactions. Hydroxyl hydrogen atoms can adopt three isoenergetic states, asparagine and glutamine flips are isoenergetic, and the tautamers of histidine are isoenergetic. Hydrogen atoms only contribute to the Lennard-Jones repulsion term in Rosetta and therefore it is common for two hydrogen placements to have equal energy.

Small motions in backbone and side-chain conformation and rigid-body displacement were modeled using gradient-based minimization (the Rosetta energy function is differentiable) with a quasi-Newton method.[33] Phi, psi, omega and chi angles for residues within 5 Å of the protein–protein interface were allowed to vary. To prevent large structural changes during the initial step of minimization, eight cyles of mimization were performed in which the weight on the repulsive energy term was ramped to 1 while distance constraints based on the starting structure were lowered from high strength to low strength.

Independent runs of gradient-based minimization converge to identical results when the starting structures are identical, however, in most cases the starting structures were not identical because Rosetta was used to repack hydrogen atoms and/or side-chains (see above) before performing minimization. Because the runs did not converge to identical structures or energies, 100 separate minimizations were performed and the lowest energy structure was used for calculating binding energies.

The affinity increase protocol is located in the analyze_interface_ddg.cc file in the Rosetta code. This protocol requires the input of a pdb and will determine which mutations to make at the interface. The command lines used were:

rosetta.gcc –s input.pdb –interface –affin_incr –Wpack_only –soft_rep_design –intout name_output –ex1 –ex2 –extrachi_cutoff 1

rosetta.gcc –s input.pdb –interface –affin_incr –Wpack_only –soft_rep_design –intout name_output –ex1 –ex2 –extrachi_cutoff 1 –repack_neighbors

rosetta.gcc –s input.pdb –interface –affin_incr –Wpack_only –soft_rep_design –intout name_output –ex1 –ex2 –extrachi_cutoff 1 –repack_neighbors –relax_unbound

rosetta.gcc –s input.pdb –interface –affin_incr –Wpack_only –soft_rep_design –intout name_output –ex1 –ex2 –extrachi_cutoff 1 –repack_neighbors –relax_unbound –min_inter –no_cst

where Wpack_only instructs Rosetta to use the default energy weight from the simulated annealing protocol, -soft_rep_design dampens the Lennard-Jones repulsion term, -repack_neighbors relaxes the entire wild type interface as well as all residues that neighbor a mutation, -relax_unbound relaxes all interface residues of the unbound structures, –min_interface performs gradient based minimization on the side chains and backbone as well as rigid body docking, and –no_cst sets constraints (designed to direct the designed structure to keep the RMSD to the wild type structure small) to the most relaxed setting.

Binding energy calculations

Binding energies were calculated by subtracting the energy of the complex from the energies of the individual chains. The energies of the individual chains is reported along with the binding energies to allow the identification of point mutations that are predicted to significantly destabilize the bound conformation of an individual chain. In our simplest protocol only the mutated residue is allowed to relax to a different rotamer, and it is assumed to adopt the same conformation in the unbound state. Our next level of complexity is to allow residues surrounding the mutated residue to relax to alternate side-chain rotamers (repack_neighbors). If the repack_neighbors option is true, the binding calculation begins by optimizing the conformation of all the side-chains at the protein–protein interface in the wild-type structure. This minimized structure is used to calculate the binding energy of the wild-type structure and serves as the starting structure for calculating the binding energy of the mutated structure.

Before calculating the energy of the mutant complex, amino acids that are close enough to have non-zero energy with the mutated residue are allowed to relax to more favorable rotamers as identified by Rosetta's side-chain repacking routine. In the simplest case, the side-chains are assumed to adopt the same rotamer in the unbound state. If the relax_unbound option is specified, side-chains near the point of mutation are relaxed separately in the bound and unbound state. The same residues are also relaxed in the bound and unbound state in the wild-type structure. Our calculations do not take into account changes in conformational entropy, and therefore, the absolute values of the binding energy calculations for a single complex do not represent true free energies. However, changes in binding energy (energy mutant–energy wild-type) do represent changes in free energy if one assumes that the conformational entropies of the various states do not change significantly with the mutation.

Our protocol with the most degrees of freedom allows side-chains to adopt alternative rotamers and performs gradient-based minimization on backbone and side-chain torsion angles and rigid-body displacement. The flexible backbone procedure begins by minimizing the wild-type complex. This serves as the starting point for the calculations with the mutant complexes. If the relax_unbound option is true, separate gradient-based minimization is performed on bound and unbound molecules. The protocol with full side-chain and backbone flexibility does not converge to the same result each time it is performed. A total of 100 separate simulations were performed and the lowest

energy wild-type and mutant complexes, as well as wild-type and mutant unbound chains were used to calculate binding energies.

Construction and cloning of protein designs

The DNA sequence for the GoLoco motif of RGS14 (residues 496–531) was cloned into pET21b as a C-terminal fusion to the small protein Tenascin. Tenascin was included to aid in the expression and purification of the peptide. The sequence for a hexahistidine tag was placed at the C terminus of the construct. Residue G498 of the wild-type GoLoco motif was mutated to a cysteine to enable the covalent labeling of the thiol-reactive fluorescent probe 6-iodoacetamidofluorescein (6-IAF) (Molecular Probes). We used an N-terminal-truncated, hexahistidine-tagged expression construct of human $G\alpha_{i1}$ with the first 25 codons of the Gα open reading frame removed, as described.[34] UbcH7 and E6AP expression plasmids have been described.[45] Point mutations were introduced using the QuickChange® site-directed mutagenesis protocol (Stratagene) and all vectors were verified by DNA sequencing.

Protein purification

GoLoco motif peptide was expressed either for 4 h at 37 °C or overnight at 25 °C with 0.5 mM IPTG in the BL21(DE3) strain of *Escherichia coli*. $G\alpha_{i1}$ was expressed overnight at 25 °C with 1 mM IPTG in the BL21(DE3) strain of *E. coli*. Cells were lysed using an Avestin emulsiflex and the resulting lysates were cleared by ultracentrifugation. The $G\alpha_{i1}$ ΔN2 and RGS14-GoLoco motif-Tenascin

fusion proteins were purified using a HiTrap (Amersham Biosciences) column by eluting the protein with an imidazole step gradient, then followed by gel filtration with a Superdex-200 column (Amersham Biosciences). Proteins were concentrated using Vivaspin 20® centrifugal concentrators. E6AP and UbcH7 were expressed and purified as described.[45] Protein concentrations were determined by measuring absorbance at 280 nm. Extinction coefficients were calculated using the method described by Gill and von Hipple.[46]

Fluorescence polarization binding analysis

The thiol-reactive fluorescent probe 6-IAF (Molecular Probes) was conjugated to the unique cysteine on the GoLoco motif using the manufacturer's recommended protocol. GoLoco motif protein was concentrated to ~100 μM then buffer exchanged into 50 mM Tris-Cl (pH 7.5) and 1 mM TCEP using a PD10 desalting column. The PD10 eluate was stirred for 1 h at room temperature. A 20 mM stock solution of 6-IAF suspended in dimethyl sulfoxide (DMSO) was diluted into the GoLoco motif protein solution to a tenfold molar excess and the conjugation reaction was allowed to proceed overnight, in the dark at 4 °C. Precipitate was pelleted and discarded, and 5 mM β-mercaptoethanol (β-ME) was added to quench the reaction. The supernatant containing fluorescein-GoLoco motif protein was run over a PD10 column to separate free probe from labeled protein. The concentration of fluorescein-GoLoco motif protein was quantified using UV/Vis, taking readings at 280 and 495 nm for the protein and fluorophore, respectively. The Tenascin-GoLoco fusion protein was used in the

binding assays. The previously published dissociation constant for $G\alpha_{i1}$ and GoLoco of 65 nM determined using surface plasmon resonance[34] is in close agreement with the 95 nM that we obtained using fluorescence anisotropy.

Fluorescence polarization assays were carried out on a Jobin Yvon Horiba Spec FluoroLog-3 instrument (Jobin Yvon Inc.) performed in L-format with the excitation wavelength set at 495 nm and the emission wavelength set at 520 nm. Titrations were performed using a 3 mm × 3 mm quartz cuvette with a starting volume of 200 µl. Fluorescein labeled wild-type or mutant GoLoco motif protein was diluted to 50 nM and the excitation and emission slit widths adjusted to give a fluorescence intensity >100,000 counts per second. Wild-type or mutant $G\alpha_{i1}$ was added in increasing volumes from a stock solution whose initial concentration depended on the strength of the interaction, generally having a concentration of 3–10 µM. Two to three polarization readings consisting of three averaged measurements were collected for increasing concentrations of $G\alpha_{i1}$. Data were averaged and analyzed using a model for single site binding according to equation (1), which was incorporated into equation (2) to account for the observed polarization:

(1)

$$[A:B] = \frac{([A_t] + [B_t] + K_d) - \sqrt{([A_t] + [B_t] + K_d)2 - 4 \cdot [A_t] \cdot [B_t]}}{2}$$

(2)

$$P_{obs} = \frac{(P_{max} - P_o) \cdot [A:B]}{[A_t]} + P_o$$

where [A:B] is the concentration of fluorescein-GoLoco motif protein and $G\alpha_{i1}$ complex formed, $[A_t]$ is the total concentration of fluorescein-GoLoco motif protein, $[B_t]$ is the concentration of $G\alpha_{i1}$, K_d is the dissociation constant for the interaction, P_o is the polarization in the absence of $G\alpha_{i1}$, P_{max} is the maximum polarization observed when all fluorescein-GoLoco motif protein is bound to $G\alpha_{i1}$, and P_{obs} is the measured polarization at a given concentration of $G\alpha_{i1}$. The data were fit according to equation (2) using non-linear regression with SigmaPlot software to obtain fitted parameters for K_d, P_{max}, and P_o.

A detailed protocol for the conjugation of the thiol-reactive fluorophore bodipy (507/545)-iodoacetamide (Molecular Probes) to UbcH7 has been described.[45] Binding assays were performed essentially as described.[45] Data analysis was performed as described in equations (1) and (2) above where [A:B] is the concentration of bodipy-UbcH7 and E6AP complex formed, $[A_t]$ is the total concentration of bodipy-UbcH7 protein, $[B_t]$ is the concentration of E6AP protein, K_d is the dissociation constant for the interaction, P_o is the polarization in the absence of UbcH7, P_{max} is the maximum polarization observed when all bodipy-UbcH7 is bound to E6AP, and P_{obs} is the measured polarization at a given concentration of UbcH7. The data were fit according to equation (2) using non-linear regression with SigmaPlot software to obtain fitted parameters for K_d, P_{max}, and P_o. Starting concentrations for bodipy-E2 depended on the extent of conjugated fluorophore and typically fell in the range of 0.5–2.0 μM. Manual

titrations were performed using wild-type and mutant E6AP(HECT) stock solutions that varied based on yield and strength of the interaction. All binding assays were performed at room temperature in 20 mM KH_2PO_4 (pH 7.0), 150 mM NaCl, 5 mM β-ME. For each binding experiment, nine polarization readings were collected and averaged at 20 concentrations of E6AP(HECT).

$G\alpha_{i1}$-RGS14-GoLoco crystal structure

The atomic coordinates used in this study are from a newly deposited crystal structure (PDB ID: 2OM2) of $G\alpha_{i1}$ complexed with a 36 amino acid residue peptide of the GoLoco motif from RGS14 (residues r496–r530, numbered according to full-length rat RGS14 protein and previously described[34]). Crystals were obtained by vapor diffusion from sitting drops containing a 1:1 (v/v) ratio of protein solution (16.5 mg ml^{-1} $G\alpha_{i1}\Delta N25$, and 1.5-fold molar excess R14GL peptide in 10 mM Tris buffer (pH 7.5), 1 mM $MgCl_2$, 10 μM GDP, 5% glycerol) to well solution (1.55 M ammonium sulfate, 100 mM sodium acetate (pH 5.0), 10% glycerol). For data collection at 100 K, the solution containing the crystals was adjusted to 25% glycerol by 2% (v/v) stepwise increases in glycerol concentration. A native data set was collected on a single crystal on SER-CAT beamline 22-ID at the Advanced Photon Source at Argonne National Labs. All data were indexed and processed using DENZO and SCALEPACK.[47] The structure of $G\alpha_{i1}\cdot GDP\cdot Mg^{2+}$ (PDB ID: 1BOF)[48] was used as a molecular replacement model using the CCP4 program AMoRe.[49] Model building was performed with the program O[50] and the program CNS was employed for

simulated annealing and torsion angle refinement.[51] The structure was refined to 2.2 angstroms.

This structure (PDB ID: 2OM2) provides a higher maximum resolution than that reported previously (PDB ID: 1KJY).[34] There are some important similarities and differences that can be noted (see Supplementary Data Figure 2.1). Briefly, the GoLoco motif "arginine finger" (Arg r516) maintains its previously described contacts with GDP. In addition, the carboxy-terminal portion of the GoLoco motif peptide maintains all its contacts with the all-helical domain of $G\alpha_{i1}$, supporting our previous findings that the all-helical domain is important for Gα specificity. In the higher resolution structure, both Leu r530 and Phe r529 are ordered, with the side-chain of Leu r530 stacking with the ring of Phe r529. The switch regions of $G\alpha_{i1}$ do not change their conformation. The most significant differences are seen in the N-terminal alpha-helix of the GoLoco motif that nestles between switch region II (α2 helix) and the α3 helix of $G\alpha_{i1}$. His r513 has rotated out of the binding interface with its γ-carbon moving by 6.6 Å. This rotation results in a two amino acid frame shift in the preceding alpha helix. For example, the hydrogen bonds formed by Arg r506 in the initial crystal structure are now being formed by Gln r508. Arg r506 maintains hydrogen bonding distance to $G\alpha_{i1}$.

REFERENCES

1. Goh, Y. Y., Frecer, V., Ho, B. & Ding, J. L. (2002). Rational design of green fluorescent protein mutants as biosensor for bacterial endotoxin. *Protein Eng* **15**, 493-502.

2. Marshall, S. A., Lazar, G. A., Chirino, A. J. & Desjarlais, J. R. (2003). Rational design and engineering of therapeutic proteins. *Drug Discov Today* **8**, 212-21.

3. Rao, B. M., Lauffenburger, D. A. & Wittrup, K. D. (2005). Integrating cell-level kinetic modeling into the design of engineered protein therapeutics. *Nat Biotechnol* **23**, 191-4.

4. Crameri, A., Cwirla, S. & Stemmer, W. P. (1996). Construction and evolution of antibody-phage libraries by DNA shuffling. *Nat Med* **2**, 100-2.

5. Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H. R. & Pluckthun, A. (1998). Ribosome display efficiently selects and evolves high-affinity antibodies in vitro from immune libraries. *Proc Natl Acad Sci U S A* **95**, 14130-5.

6. Martin, L., Stricher, F., Misse, D., Sironi, F., Pugniere, M., Barthe, P., Prado-Gotor, R., Freulon, I., Magne, X., Roumestand, C., Menez, A., Lusso, P., Veas, F. & Vita, C. (2003). Rational design of a CD4 mimic that inhibits HIV-1 entry and exposes cryptic neutralization epitopes. *Nat Biotechnol* **21**, 71-6.

7. Marvin, J. S. & Lowman, H. B. (2003). Redesigning an antibody fragment for faster association with its antigen. *Biochemistry* **42**, 7077-83.

8. Selzer, T., Albeck, S. & Schreiber, G. (2000). Rational design of faster associating and tighter binding protein complexes. *Nat Struct Biol* **7**, 537-41.

9. Kiel, C., Selzer, T., Shaul, Y., Schreiber, G. & Herrmann, C. (2004). Electrostatically optimized Ras-binding Ral guanine dissociation stimulator mutants increase the rate of association by stabilizing the encounter complex. *Proc Natl Acad Sci U S A* **101**, 9223-8.

10. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L. & Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364-8.

11. Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Rothlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Barbas, C. F.,

3rd, Hilvert, D., Houk, K. N., Stoddard, B. L. & Baker, D. (2008). De novo computational design of retro-aldol enzymes. *Science* **319**, 1387-91.

12. Rothlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J. L., Althoff, E. A., Zanghellini, A., Dym, O., Albeck, S., Houk, K. N., Tawfik, D. S. & Baker, D. (2008). Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190-5.

13. Bolon, D. N. & Mayo, S. L. (2001). Enzyme-like proteins by computational design. *Proc Natl Acad Sci U S A* **98**, 14274-9.

14. Wunderlich, M., Martin, A., Staab, C. A. & Schmid, F. X. (2005). Evolutionary protein stabilization in comparison with computational design. *J Mol Biol* **351**, 1160-8.

15. Korkegian, A., Black, M. E., Baker, D. & Stoddard, B. L. (2005). Computational thermostabilization of an enzyme. *Science* **308**, 857-60.

16. Malakauskas, S. M. & Mayo, S. L. (1998). Design, structure and stability of a hyperthermophilic protein variant. *Nat Struct Biol* **5**, 470-5.

17. Dantas, G., Corrent, C., Reichow, S. L., Havranek, J. J., Eletr, Z. M., Isern, N. G., Kuhlman, B., Varani, G., Merritt, E. A. & Baker, D. (2007). High-resolution structural and thermodynamic analysis of extreme stabilization of human procarboxypeptidase by computational protein design. *J Mol Biol* **366**, 1209-21.

18. Dantas, G., Kuhlman, B., Callender, D., Wong, M. & Baker, D. (2003). A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J Mol Biol* **332**, 449-60.

19. Bolon, D. N., Wah, D. A., Hersch, G. L., Baker, T. A. & Sauer, R. T. (2004). Bivalent tethering of SspB to ClpXP is required for efficient substrate delivery: a protein-design study. *Mol Cell* **13**, 443-9.

20. Green, D. F., Dennis, A. T., Fam, P. S., Tidor, B. & Jasanoff, A. (2006). Rational design of new binding specificity by simultaneous mutagenesis of calmodulin and a target peptide. *Biochemistry* **45**, 12547-59.

21. Havranek, J. J. & Harbury, P. B. (2003). Automated design of specificity in molecular recognition. *Nat Struct Biol* **10**, 45-52.

22. Joachimiak, L. A., Kortemme, T., Stoddard, B. L. & Baker, D. (2006). Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. *J Mol Biol* **361**, 195-208.

23. Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L. & Baker, D. (2004). Computational redesign of protein-protein interaction specificity. *Nat Struct Mol Biol* **11**, 371-9.

24. Song, G., Lazar, G. A., Kortemme, T., Shimaoka, M., Desjarlais, J. R., Baker, D. & Springer, T. A. (2006). Rational design of intercellular adhesion molecule-1 (ICAM-1) variants for antagonizing integrin lymphocyte function-associated antigen-1-dependent adhesion. *J Biol Chem* **281**, 5042-9.

25. Clark, L. A., Boriack-Sjodin, P. A., Eldredge, J., Fitch, C., Friedman, B., Hanf, K. J., Jarpe, M., Liparoto, S. F., Li, Y., Lugovskoy, A., Miller, S., Rushe, M., Sherman, W., Simon, K. & Van Vlijmen, H. (2006). Affinity enhancement of an in vivo matured therapeutic antibody using structure-based computational design. *Protein Sci* **15**, 949-60.

26. Pace, C. N. (1992). Contribution of the hydrophobic effect to globular protein stability. *J Mol Biol* **226**, 29-35.

27. Takano, K., Yamagata, Y., Fujii, S. & Yutani, K. (1997). Contribution of the hydrophobic effect to the stability of human lysozyme: calorimetric studies and X-ray structural analyses of the nine valine to alanine mutants. *Biochemistry* **36**, 688-98.

28. Hendsch, Z. S., Jonsson, T., Sauer, R. T. & Tidor, B. (1996). Protein stabilization by removal of unsatisfied polar groups: computational approaches and experimental tests. *Biochemistry* **35**, 7621-5.

29. Hendsch, Z. S. & Tidor, B. (1994). Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci* **3**, 211-26.

30. Waldburger, C. D., Schildbach, J. F. & Sauer, R. T. (1995). Are buried salt bridges important for protein stability and conformational specificity? *Nat Struct Biol* **2**, 122-8.

31. Kortemme, T., Morozov, A. V. & Baker, D. (2003). An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J Mol Biol* **326**, 1239-59.

32. Kobko, N., Paraskevas, L., del Rio, E. & Dannenberg, J. J. (2001). Cooperativity in amide hydrogen bonding chains: implications for protein-folding models. *J Am Chem Soc* **123**, 4348-9.

33. Rohl, C. A., Strauss, C. E., Misura, K. M. & Baker, D. (2004). Protein structure prediction using Rosetta. *Methods Enzymol* **383**, 66-93.

34. Kimple, R. J., Kimple, M. E., Betts, L., Sondek, J. & Siderovski, D. P. (2002). Structural determinants for GoLoco-induced inhibition of nucleotide release by Galpha subunits. *Nature* **416**, 878-81.

35. Huang, L., Kinnucan, E., Wang, G., Beaudenon, S., Howley, P. M., Huibregtse, J. M. & Pavletich, N. P. (1999). Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science* **286**, 1321-6.

36. Pickart, C. M. (2001). Mechanisms underlying ubiquitination. *Annu Rev Biochem* **70**, 503-33.

37. Johnston, C. A., Lobanova, E. S., Shavkunov, A. S., Low, J., Ramer, J. K., Blaesius, R., Fredericks, Z., Willard, F. S., Kuhlman, B., Arshavsky, V. Y. & Siderovski, D. P. (2006). Minimal determinants for binding activated G alpha from the structure of a G alpha(i1)-peptide dimer. *Biochemistry* **45**, 11390-400.

38. Takano, K., Ogasahara, K., Kaneda, H., Yamagata, Y., Fujii, S., Kanaya, E., Kikuchi, M., Oobatake, M. & Yutani, K. (1995). Contribution of hydrophobic residues to the stability of human lysozyme: calorimetric studies and X-ray structural analysis of the five isoleucine to valine mutants. *J Mol Biol* **254**, 62-76.

39. Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P. & Matthews, B. W. (1992). Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **255**, 178-83.

40. Moult, J. (2005). A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr Opin Struct Biol* **15**, 285-9.

41. Kortemme, T. & Baker, D. (2002). A simple physical model for binding energy hot spots in protein-protein complexes. *Proc Natl Acad Sci U S A* **99**, 14116-21.

42. Lazaridis, T. & Karplus, M. (1999). Effective energy function for proteins in solution. *Proteins* **35**, 133-52.

43. Dunbrack, R. L., Jr. & Cohen, F. E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci* **6**, 1661-81.

44. Kuhlman, B. & Baker, D. (2000). Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A* **97**, 10383-8.

45. Eletr, Z. M., Huang, D. T., Duda, D. M., Schulman, B. A. & Kuhlman, B. (2005). E2 conjugating enzymes must disengage from their E1 enzymes before E3-dependent ubiquitin and ubiquitin-like transfer. *Nat Struct Mol Biol* **12**, 933-4.

46. Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* **182**, 319-26.

47. Otwinowski, Z. M., W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307-326.

48. Coleman, D. E. & Sprang, S. R. (1998). Crystal structures of the G protein Gi alpha 1 complexed with GDP and Mg2+: a crystallographic titration experiment. *Biochemistry* **37**, 14376-85.

49. Navaza, J. (2001). Implementation of molecular replacement in AMoRe. *Acta Crystallogr D Biol Crystallogr* **57**, 1367-72.

50. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47 (Pt 2)**, 110-9.

51. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-21.

CHAPTER III

REDESIGNING PROTEIN-PEPTIDE BINDING SPECIFICITY

ABSTRACT

We automated a computational protocol designed to increase the binding specificity of a protein-peptide interaction. A conserved 19-amino acid motif, termed the GoLoco motif, is able to bind with the G-alpha subunits from the adenylyl-cyclase-inhibitory subclass. This motif acts as a guanine nucleotide dissociation inhibitor for the GDP-bound G-alpha proteins. We selected a complex from these two protein families that includes the RGS14 GoLoco motif and the $G\alpha_{i1}$ protein. We sought to redesign this interface so that these proteins would interact in a comparable manner to the wild type complex but will no longer interact with the other protein family members or their wild type counterparts. The ability to manipulate complexes in signaling networks can be used to develop biosensors for live cell imaging or to rewire or inhibit cell signaling pathways. Increasing the binding specificity of a protein has additional applications including developing novel affinity chromatography columns or improving the specificity of protein therapeutics. Our protocol searches for mutations that disrupt the wild-type interaction, followed by optimizing the neighboring sequence positions on the partner protein in an effort to compensate for the destabilizing effects. We experimentally characterized 8 selected designs. 3 of the 8 redesigned complexes have a tighter binding affinity than their mixed mutant-wild type counterparts, demonstrating successful redesign of binding specificity. We recovered a binding affinity comparable to that of the wild type complex for 2 of the 8 designs, showing a redesigned protein-peptide complex can behave similarly to the wild-type complex. We then selected a

design with a specificity switch, but not recovery of wild-type binding affinity. We are able to add affinity enhancing mutations to the design in order to recover a wild-type binding affinity while achieving a specificity switch of approximately 100-fold.

INTRODUCTION

Computational design of protein-protein interactions is a rigorous test of our understanding of the energetics regulating protein-protein recognition. The accurate manipulation of the free energy of binding for a given protein complex tests the energy function and computational search algorithm of protein design software. The energy function must be detailed enough to have predictive power. In addition the protocol needs to sufficiently sample conformational space. In many cases, where designs contain only a few amino acid substitutions, a fixed-backbone approximation can be used, where the protein backbones coordinates come from the experimentally determined structure and are not allowed to vary. Thus the only parts of the protein structure that will be modeled are the amino acid side chains, greatly reducing the degrees of freedom that need to be considered. Even with this approximation it is important to adequately model the side chain torsion angles. Amino acid side chains are represented in discrete conformations, or rotamers, based on the torsion angles that are most often seen in a sample of non-homologous structure coordinate files taken from the Protein Data Bank.[1; 2] Destabilizing effects of a mutation can be incorrectly predicted or overestimated due to an insufficient sampling of side-chain torsion angles. We evaluated the energy function, search procedure and coverage of conformational space of our protocol by both designing amino acid sequences to stabilize desired interactions while at the same time disrupting the undesired interactions.

Several approaches have been used to successfully redesign protein binding specificity. Shifman et al. redesigned the interface between calmodulin (CaM) and one of its native binding partners, smooth muscle myosin light chain kinase (smMLCK).[3] The redesigned CaM maintains wild type-like binding affinity to smMLCK but binds less tightly to a group of other native binding partners. They were able to accomplish this using only positive protein design, by redesigning the CaM sequence to stabilize the desired interaction, and without using negative protein design, or designing a sequence to disrupt the undesired interactions. The decrease in binding affinity for most of the undesired interactions ranges from approximately equal for a homologue of smMLCK to over 100 fold for a peptide containing a non-natural amino acid. These are impressive results, but our goal was to redesign protein-protein interfaces so that the proteins will not interact with the homologous family members of their binding partner. Additionally, the successfully redesigned CaM protein contain 8 mutations at the interface. Maintaining the biochemical function of a protein might be a challenge with this many mutations at a binding interface. Havranek et al. took a different approach by combining positive and negative protein design to design specificity in coiled-coil interactions.[4] This protocol selects amino acids by simultaneously considering how much the substitution stabilizes the desired interaction and destabilizes the undesired interactions. While the results are quite impressive, this protocol could be difficult to implement with other protein interactions because it requires structures for all protein-protein complexes that either binding partner is a part of. Kortemme et al. redesigned the interface of

the colicin E6 DNase-Im7 immunity protein complex[5] using a protocol that also uses both positive and negative protein design. Their goal was to create an orthogonal interface where the components of the redesigned complex do not interact with the components of the wild type complex. The idea is that if a redesigned mutant protein will not interact with its wild type counterpart then this specificity might extend to homologous family members. A design of this nature could allow the isolation of a specific protein-protein interaction from a large network of interactions. Their protocol uses negative protein design to search for point mutations that will destabilize the wild type interactions followed by positive design to search for mutations that can compensate for the destabilizing effects. The destabilizing point mutation explicitly designs against the mutant protein-wild type protein interaction. The other mixed complex, the wild type protein-mutant protein, is implicitly designed against in the same way that Shifman et al. redesigned the calmodulin binding specificity.[3] It is not certain if this type of implicit negative design will accomplish the desired destabilization of the mixed complex. In fact, Kortemme et al. combined two sets of orthogonal mutations to obtain the desired destabilization for both mixed protein-protein complexes. In this case, the redesigned complex is significantly destabilized compared to the wild-type interaction, however. Recently Joachimiak et al. compared this same protocol to a protocol that uses rigid body docking about an axis of rotation to generate an ensemble of protein-protein complex structures. They were able to design a hydrogen bond network at the interface of the colicin E6 DNase-Im7 immunity protein complex using the rigid body docking protocol.[6] This is the first

known example of a rationally designed hydrogen bond network confirmed by an x-ray crystal structure. The redesigned complex does not recover wild type binding affinity. A crystal structure of the redesign was used for a second round of design. The resulting complex did regain wild type binding affinity but lost specificity against the mixed wild type-mutant complexes. A third round of design added a point mutation that successfully destabilized one of the mutant protein-wild type protein complexes by 300 fold.

We automated the orthogonal design protocol proposed by Kortemme et al.[5] in order to evaluate it with a protein peptide complex. Our aim was to obtain a destabilization of the wild type-redesigned complexes while maintaining a binding affinity for the redesigned complex comparable to that of the wild type interaction. We use the Rosetta protein design software[7]. Kortemme et al. used essentially the same software, although it was maintained separately from Rosetta. Rosetta was created for protein structure prediction[8]. Rosetta has had tremendous success in the critical assessment of structure prediction experiments (CASP)[9; 10] and was subsequently applied to protein docking prediction[11; 12] and protein design.[13] Rosetta has been applied to protein-protein interface design to predict the importance of amino acid residues at protein interfaces,[14] to predict affinity enhancing point mutations,[15] and to design peptide extensions that will increase buried surface area and enhance protein peptide binding affinity.[16]

We select the $G\alpha_{i1}$ protein and RGS14 GoLoco motif for our model system.[17] The $G\alpha_{i1}$-GoLoco interface, with over 1900 $Å^2$ of buried surface area, provides

both hydrophobic as well as limited electrostatic regions for redesign. The GoLoco binds to both the Ras-like domain and the all-helical domain of the $G\alpha_{i1}$ protein, offering two distinctly separate and structurally diverse regions for design.

RESULTS

Each residue at the protein-protein interface of the $G\alpha_{i1}$-GoLoco model system was sequentially mutated to the twenty amino acids in silico and the free energy of the structure was predicted with the Rosetta energy function. In cases where the mutation was predicted to destabilize binding of the protein-peptide complex, neighboring residues were redesigned in search of compensating mutations. Neighboring residues were defined as any sequence position where a side chain heavy atom is within 5.5 Å of a side chain heavy atom on the mutated sequence position. The residues on the partner protein were redesigned in an effort to maintain a binding affinity comparable to the wild type complex (Figure 3.1 (b)). Neighboring residues on the same chain as the destabilizing point mutation were also redesigned to prevent destabilization of the protein from the point mutation. In this way each destabilizing point mutation became an orthogonal design. In order to evaluate the quality of the design, the free energy of binding was predicted for each redesigned complex as well as the mixed complexes. The mixed complexes need to be destabilized with respect to the redesigned complex in order to achieve orthogonality (Figure 3.1 (a)). Binding energies were calculated by subtracting the calculated energy of each unbound protein from the calculated energy of the complex. In the first round of calculations, the backbone and side chain conformations of the unbound proteins were assumed to be identical to those in the bound state. The side chains of the redesigned residues were built by choosing the rotamer with the lowest energy when modeled in the context of the complex. The backbone coordinates were

held fixed to those found in the crystal structure (PDB ID: 2OM2). All structures being evaluated, including the wild type $G\alpha_{i1}$-GoLoco complex, redesigned $G\alpha_{i1}$-GoLoco and the mixed complexes, were modeled independently to allow the redesigned residues to find the lowest energy rotamer for each complex.

Selection of orthogonal designs for experimental characterization

Each interface position was sequentially mutated to the 20 amino acids in search of destabilizing point mutations. Each predicted destabilizing point mutation became a design. We considered more than 1500 designs. Searching for possible orthogonal designs, we narrow the results to designs with binding energy for the mixed complexes that are less favorable than the wild type complex. The redesigned complexes should perform like the wild type complexes, so they need to have a binding affinity in the range of the wild type complex. We searched the remaining designs for predicted change in binding energy approximately equal to zero. Finally, we removed designs that alter residues on the $G\alpha_{i1}$ protein or GoLoco motif that have been shown to play an important role in biochemical function. For example, the RGS14 GoLoco motif interacts with both the Ras-like, guanine nucleotide binding domain and the all-helical domain of $G\alpha_{i1}$.[17] The Ras-like domain of $G\alpha_{i1}$ contains a flexible region termed the switch II which is able to open and expose a large hydrophobic surface when guanosine diphosphate (GDP) is bound.[18] Mutations in the switch II can alter the intrinsic GTPase activity of the $G\alpha_{i1}$[19] so we try to avoid designs

that require mutations in this region. Additionally, we eliminate designs that alter $G\alpha_{i1}$ or GoLoco residues that interact with the bound GDP.

We found four designs that meet our criteria. These four designs are comprised of mutations with different chemical properties. Three of the designs contain hydrogen bond interactions. The fourth design increases the buried hydrophobic surface area across the interface by replacing a polar residue with a hydrophobic residue. In order to thoroughly evaluate our protocol we wished to characterize two additional categories of mutations; a charge reversal and a hydrophobic core redesign. The $G\alpha_{i1}$-GoLoco interface is largely hydrophobic, with most of the hydrophilic residues at the interface participating in stabilization of the nucleotide or backbone hydrogen bonds. We find one possibility for a charge reversal design, which alters a residue in the switch II region of $G\alpha_{i1}$. This charge reversal design ($G\alpha_{i1}$ R208E and GoLoco E498R) was selected by eye and all complexes were predicted by Rosetta to be stabilizing. Charge reversal designs can be a highly effective way to redesign binding specificity since other native binding partners for each protein have likely evolved to have complementary charges to the wild type, as long as they bind at the same location. Charge reversal design has been successfully used to improve a biosensor specificity.[20] We have a thorough understanding of the hydrophobic effect and have been able to model mutations that enhance the stability of monomeric proteins and protein complexes.[15; 21] We therefore add four hydrophobic core designs (Figure 3.2 panels c,d,g and h).

Experimental characterization of orthogonal designs

Of the four designs that passed our initial filters, experimental validation showed that two designs retain a binding affinity comparable to the wild type complex. Those are design 1, $G\alpha_{i1}$ E245L,L249A and GoLoco V507M, and design 2, $G\alpha_{i1}$ K248E,S252L and GoLoco L503K. (Figure 3.2 panels a and b). Additionally, both designs have higher binding affinity for the redesigned complexes than for the mixed complexes, although the specificity switch is modest. One of the designs, design 1 ($G\alpha_{i1}$ E245L,L249A and GoLoco V507M), is stabilized relative to the wild type complex. (Table 3.1 and Figure 3.3) This redesigned $G\alpha_{i1}$ ($G\alpha_{i1}$ E245L,L249A) with the wild type GoLoco has a binding affinity comparable to that of the wild type complex. The redesigned GoLoco (GoLoco V507M) with the wild type $G\alpha_{i1}$ is destabilized by approximately one order of magnitude. Thus the desired specificity was achieved, but only in one direction. ($\Delta\Delta G^{\circ}_{bind}$ Table 3.1 design 1 and Figure 3.3 - $G\alpha_{i1}$ E245L,L249A and GoLoco V507M) The second design does have the desired specificity switch ($\Delta\Delta G^{\circ}_{bind}$ in Table 3.1 design 2 and Figure 3.3 - $G\alpha_{i1}$ K248E,S252L and GoLoco L503K). Design 2 ($G\alpha_{i1}$ K248E,S252L and GoLoco L503K) is located on the periphery of the interface and introduces an electrostatic interaction across the interface accommodated by a charge reversal on the $G\alpha_{i1}$. The specificity switch for this design is modest. The remaining two designs were predicted to bind similarly to the wild type complex but were in fact destabilized by close to two orders of magnitude. Both contain designed hydrogen bonds. The hydrogen bonds in these designs are not ideal in their distance or angle.[22] (Figure 3.2)

Only one of the ten designs was not characterized due to a protein, GoLoco F529P,L530Y, that was insoluble. One design, $G\alpha_{i1}$ I78S,A111Q and GoLoco L519T, was not fully characterized when the redesigned structure measured a binding affinity estimated to be greater than 45 μM. Of the remaining 8 designs, three have the specificity intended by the protocol, where the redesigned complex binds more tightly than the mixed complexes. 5 of the 8 exhibit specificity in at least one direction, in that the redesigned complex binds more tightly than at least one of the mixed complexes. (Figure 3.3). These results are encouraging given that the design protocol explicitly designs specificity in only one direction due to the destabilizing point mutation. Ideally the compensating mutations made on the opposite chain would destabilize the interaction with its wild type counterpart. This is based on the work by Shifman et al. who were able to achieve increased specificity with calmodulin simply by redesigning the wild type interaction with smooth muscle myosin light chain kinase[3].

Combining affinity enhancing mutations with orthogonal designs

The results for design 2, are encouraging given that specificity was achieved and the redesigned complex maintained the desired binding affinity. Applying this protocol to design biosensors or rewire a cellular network may require a larger specificity switch than was achieved with designs 1 and 2. It is difficult to know what binding specificities are needed to maintain desired interactions while inhibiting undesired interactions without knowing the kinetics of

all relevant protein interactions as well as the cellular protein concentrations. We therefore wanted to see if we could design a complex that could maintain the biochemically relevant binding affinity while achieving more destabilization of the wild type-mutant mixed complexes. Kortemme et al., using the protocol that this work is based on, combined two designs in order to weaken binding for both mixed complexes.[5] The specificity switch was modest and the redesigned complex was destabilized compared to the wild type complex. We have two designs that bind similarly to the wild type. These designs are immediately adjacent to one another. We therefore looked for an alternative approach.

We have developed a protocol to efficiently predict point mutations to enhance protein-protein binding affinity. Looking for an approach to develop an orthogonal design with the desired specificity we try combining affinity enhancing mutations with orthogonal designs to bring the binding affinity of the redesigned complex into the range of the wild type complex. The designs that did not pass our initial filters were not predicted to bind with an affinity comparable to the wild type complex and they do not. Three of these designs did achieve at least partial specificity. We selected design 6 ($G\alpha_{i1}$ I78A,A111F and GoLoco L519G,L524T) and design 7 ($G\alpha_{i1}$ F223V,L249F and GoLoco L504A) to combine with affinity enhancing point mutations $G\alpha_{i1}$ E245L, $G\alpha_{i1}$ Q147L and GoLoco F529W. Design 6 has the desired specificity but the redesigned complex is destabilized by an order of magnitude. The binding affinity of design 6 is not enhanced by any of the mutations tried, however. (Table 3.2) Perhaps the GoLoco L519G mutation has an entropic effect that is not compensated for by the affinity enhancing

mutations. Design 7 is selected because it is a simple hydrophobic core redesign and may be simpler to manipulate with affinity enhancing mutations. The addition of the $G\alpha_{i1}$ Q147L mutation brings the dissociation constant down from greater than 2 μM to 1.1 μM. Combining this with the $G\alpha_{i1}$ E245L mutation brings the dissociation constant down to 0.14 μM, within the range of the wild type (see Figure 3.4 and Table 3.2). Thus a specificity switch of approximately 17 fold has been achieved while maintaining a binding affinity similar to the wild type. Adding one final affinity enhancing point mutation, F529W on the GoLoco domain, brings the dissociation constant down to 0.04 μM and the dissociation constant for the GoLoco V507M,F529W with the wild type $G\alpha_{i1}$ is measure at greater than 5 μM, giving a specificity switch of greater than 125 fold against the redesigned GoLoco domain. (Figure 3.5) Joachimiak et al. obtained a specificity switch of approximately 300 fold.[6] They reported designing a hydrogen bond across a protein-protein interface, but destabilized the complex with respect to wild type. Solving the crystal structure of the design they performed a second round of design to recover binding affinity, followed by a third design where they added a single mutation to obtain the 300 fold specificity switch.

Varying levels of side chain flexibility in predicting $\Delta\Delta G°_{bind}$

The first round of calculations maintained the same backbone and side chain conformations for the unbound proteins and the bound state. The side chains that were allowed to vary their identity were the only side chains allowed to search for lower conformational states. The results are shown in Table 3.1. It is possible that residues neighboring computational mutation sites could also find

lower energy conformations, relieving steric clashes or improving hydrogen bonds. The command for this protocol is –repack_neighbors. We also consider the interface residues could find lower energy conformation in the unbound structures, termed –relax_unbound. We combine –repack_neighbors and –relax_unbound to get maximum side-chain flexibility. This protocol with increased flexibility did not improve the predictions for electrostatic interactions, as seen with designs 2 and 3. (Supplementary Table 3.1) The redesigned complexes for designs 1 and 2 were correctly predicted to be affinity enhancing, although with only two data points it is not clear what led to the improvement. Overall the results from the initial protocol which only models amino acids allowed to change identity are comparable to the more computationally expensive protocols that allow for more extensive redesign.

DISCUSSION

One of the four designs that passed our set of filters was in fact a modestly orthogonal interface. The redesigned proteins bound with less affinity to the wild type binding partners than the redesigned or wild type complexes. A second design gave a specificity switch of one order of magnitude against the redesigned GoLoco motif only. Additionally, both of these redesigned complexes bound with a binding affinity similar to that of the wild type complex. The results of this work demonstrate that the specificity of a protein peptide complex can be redesigned using this simple approach, introducing a destabilizing point mutation at the interface and designing compensating mutations around it. These results also suggest that there is reasonable predictive power with the Rosetta energy function. The specificity switch for the initial round of designs is modest, however. Redesigning protein-protein binding specificity is a hard problem. Applying this protocol to the $G\alpha_{i1}$-GoLoco interface, which buries approximately 1900 square angstroms of surface area, resulted in more than 1500 designs. Yet only 4 designs passed our filters. Of the designs selected for experimental characterization that did not pass our initial set of filters, none recovered wild type binding affinity. One possibility is that it is very difficult to find amino acids that can compensate for the introduction of a destabilizing point mutation. This can be seen in the quality of hydrophobic packing and hydrogen bonds in the designs (Figure 3.6). The incorporation of small backbone movements in the design protocol could increase the sequence space that is sampled, thereby increasing the number and quality of the designs. Our work automating a

protocol to predict affinity enhancing mutations evaluated a gradient based backbone and side chain minimization method.[15] We did not, however, see significant improvement in the predictive power of the protocol. Davis et al. found that the reorientation of two consecutive peptide backbone torsion angles, termed backrub motion, could significantly could alter accessible side chain conformations and greatly improve interactions in crystal structures.[23] Adding movement of the backbone torsion angles such as the backrub motion might increase the sequence space that is sampled, thereby improving the quality of the designs. At least five research groups including ours have addressed this problem in a variety of ways.[3; 4; 5; 6; 24] The most impressive results to date are those by Joachimiak et al. who implemented backbone movement in the form of rigid body docking of the protein-protein complex, ultimately achieving a specificity switch of approximately 300 fold.[6] These results were achieved after solving the crystal structure of the first design, followed by two additional rounds of design.

While the results of our orthogonal design are encouraging, if this protocol were applied to rewire a cellular network or develop a biosensor, any undesired protein-protein interactions decrease the accuracy of these tools. It is difficult to know exactly how much destabilization of the mixed complexes is required without knowing the kinetics of all involved protein-protein interactions as well as cellular protein concentrations. We therefore wanted to achieve greater destabilization of the redesigned protein – wild type protein complexes while still maintaining a wild type binding affinity to evaluate how much specificity we could

achieve with this protocol. Many of the designs that did not recover wild type binding affinity did achieve significantly more destabilization of the mixed complexes. By combining the protocol to predict affinity enhancing point mutations with the specificity redesign protocol we were able to create a redesigned complex that maintained wild type binding affinity with a specificity switch of 17 fold against one of the mutant protein-wild type protein complexes. An additional affinity enhancing mutation gave the same complex a specificity switch of at least 125 fold against the redesigned GoLoco domain. This redesigned complex has four mutations on the $G\alpha_{i1}$ protein (Q147L,F223V,E245L,L249F) and two mutation on the GoLoco peptide (V507M,F529W).

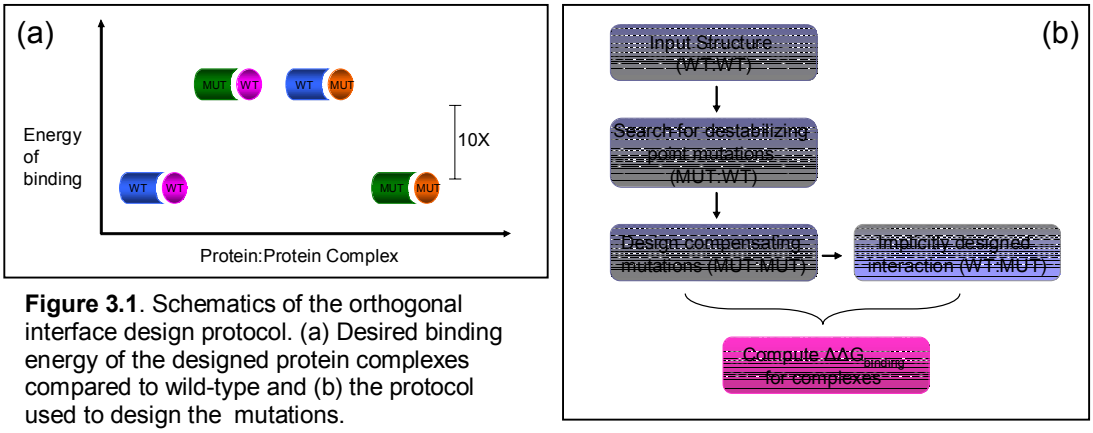


Figure 3.1. Schematics of the orthogonal interface design protocol. (a) Desired binding energy of the designed protein complexes compared to wild-type and (b) the protocol used to design the mutations.

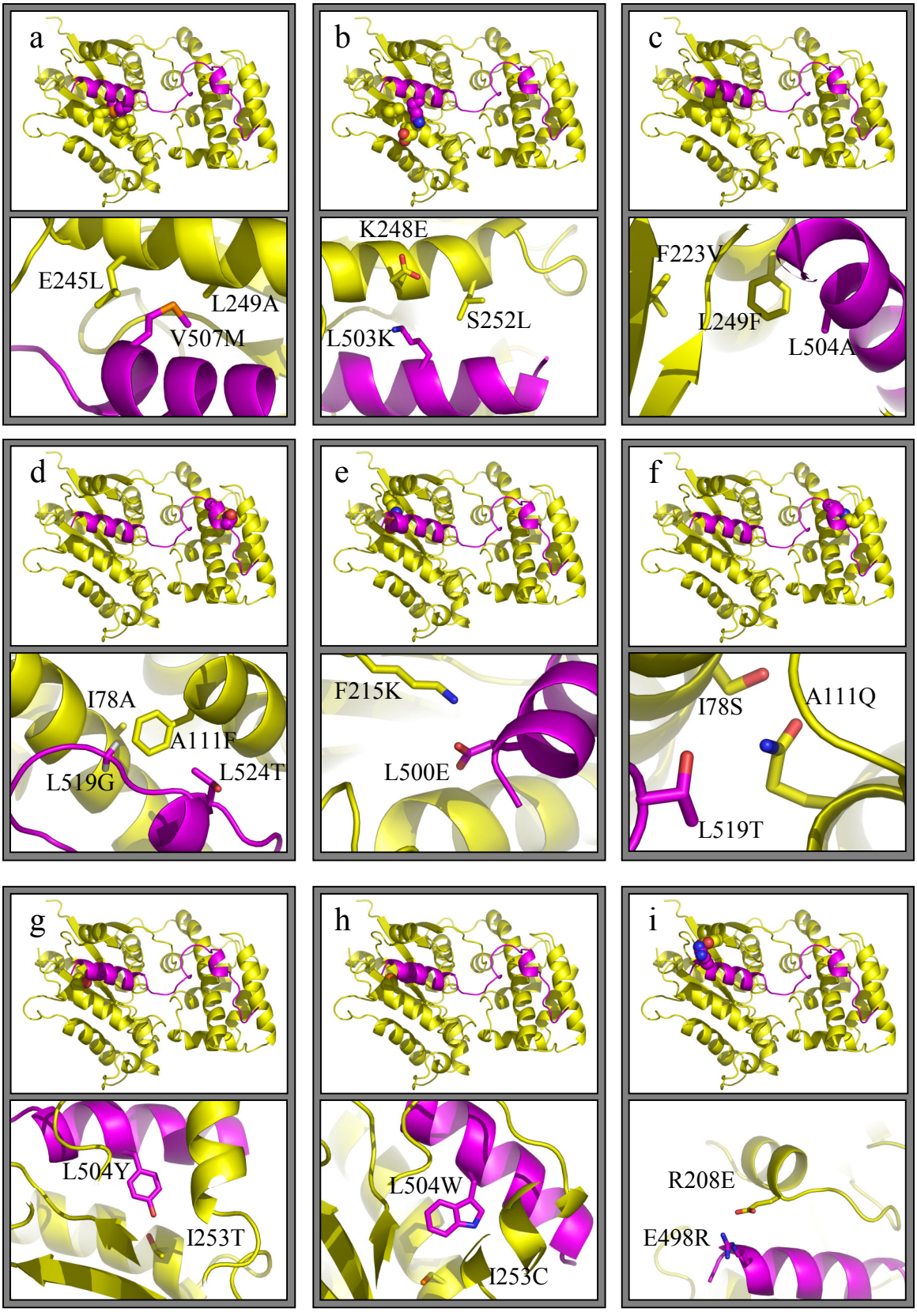


Figure 3.2. Modeled structures of orthogonal designs selected for experimental characterization. (a) Design 1: $G\alpha_{i1}$ E245L, L249A and GoLoco V507M, (b)

Design 2: $G\alpha_{i1}$ K248E, S252L and GoLoco L503K, (c) Design 3: $G\alpha_{i1}$ F223V, L249F and GoLoco L504A, (d) $G\alpha_{i1}$ I78A, A111F and GoLoco L519G, L524T, (e) $G\alpha_{i1}$ F215K and GoLoco L500E, (f) $G\alpha_{i1}$ I78S, A111Q and GoLoco L519T, (g) $G\alpha_{i1}$ I253T and GoLoco L504Y, (h) $G\alpha_{i1}$ I253C and GoLoco L504W, (i) $G\alpha_{i1}$ R208E and GoLoco E498R.

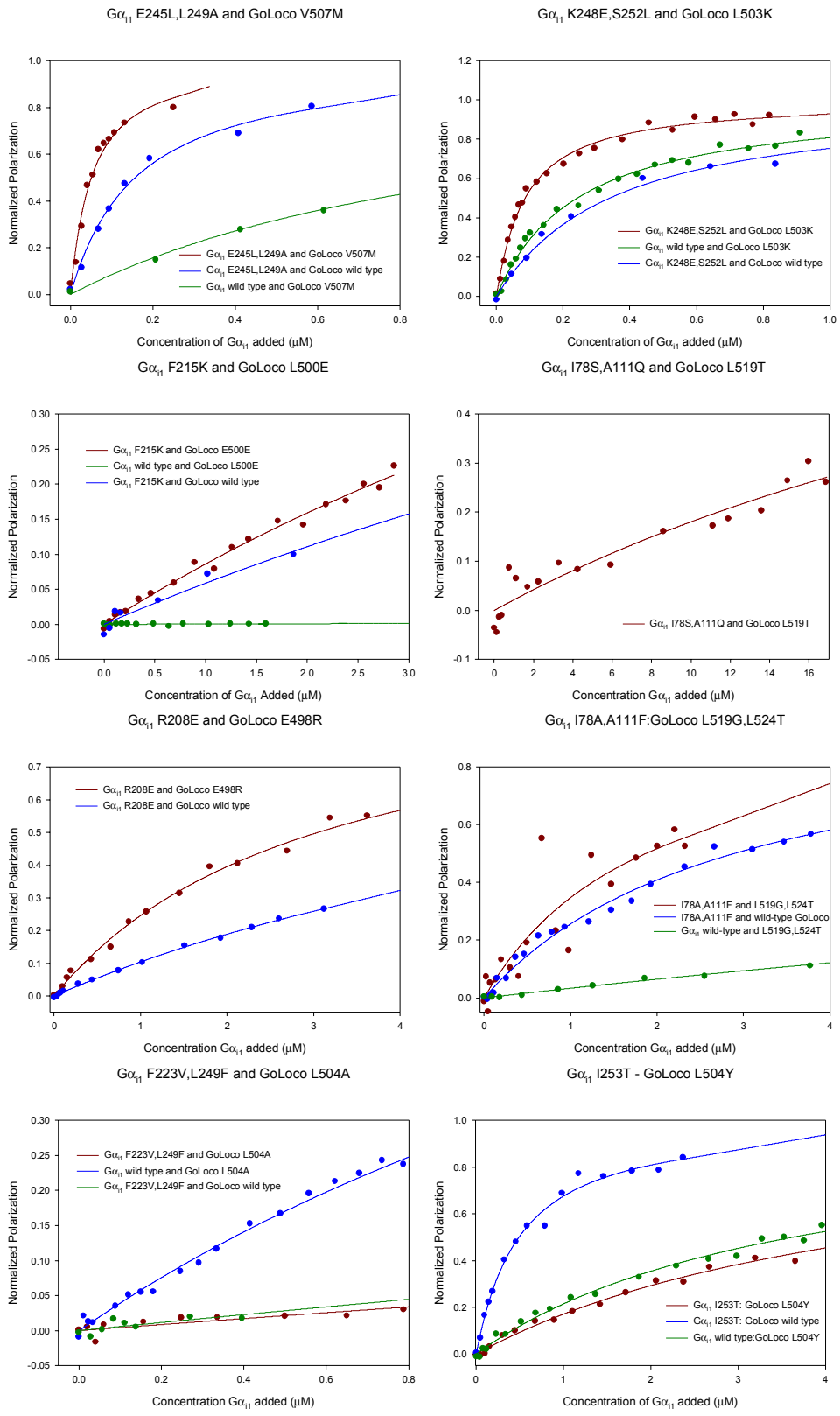


Figure 3.3. Binding curves for experimentally characterized orthogonal designs.

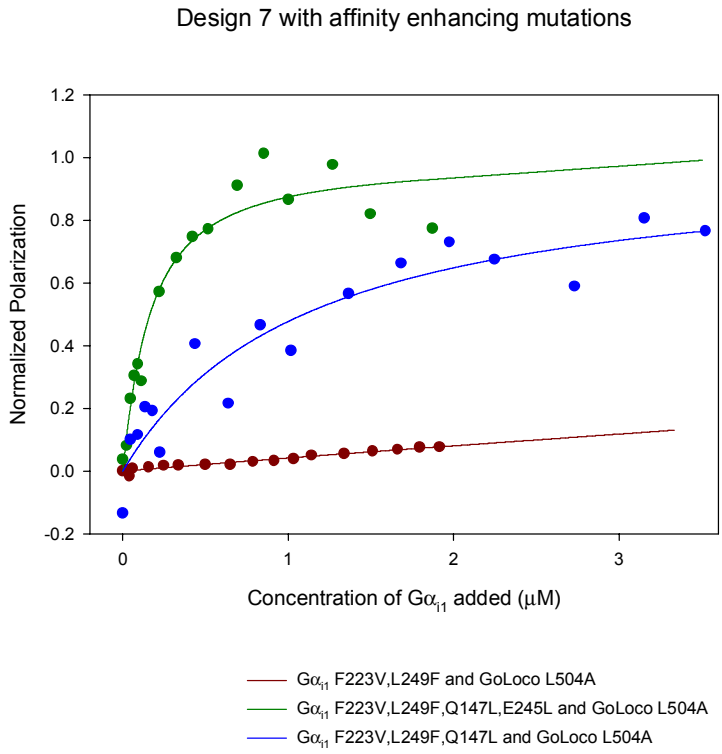


Figure 3.4. Binding curves for Design 7 and Design 7 with affinity enhancing point mutations. Binding for Gαi1 F223V,L249F-GoLoco L504A complex compared to Gαi1 F223V,L249F,**Q147A**-GoLoco L504A complex and Gαi1 F223V,L249F,**Q147A,E245L**-GoLoco L504A complex.

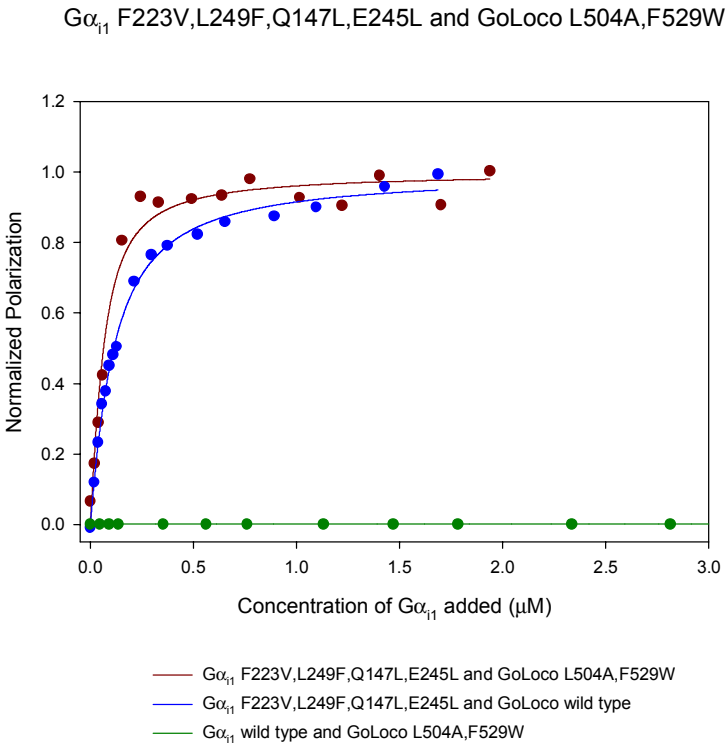


Figure 3.5. Specificity of design 7 with addition of affinity enhancing mutations, Gαi1 F223V,L249F,Q147L,E245L and GoLoco L504A,F529W

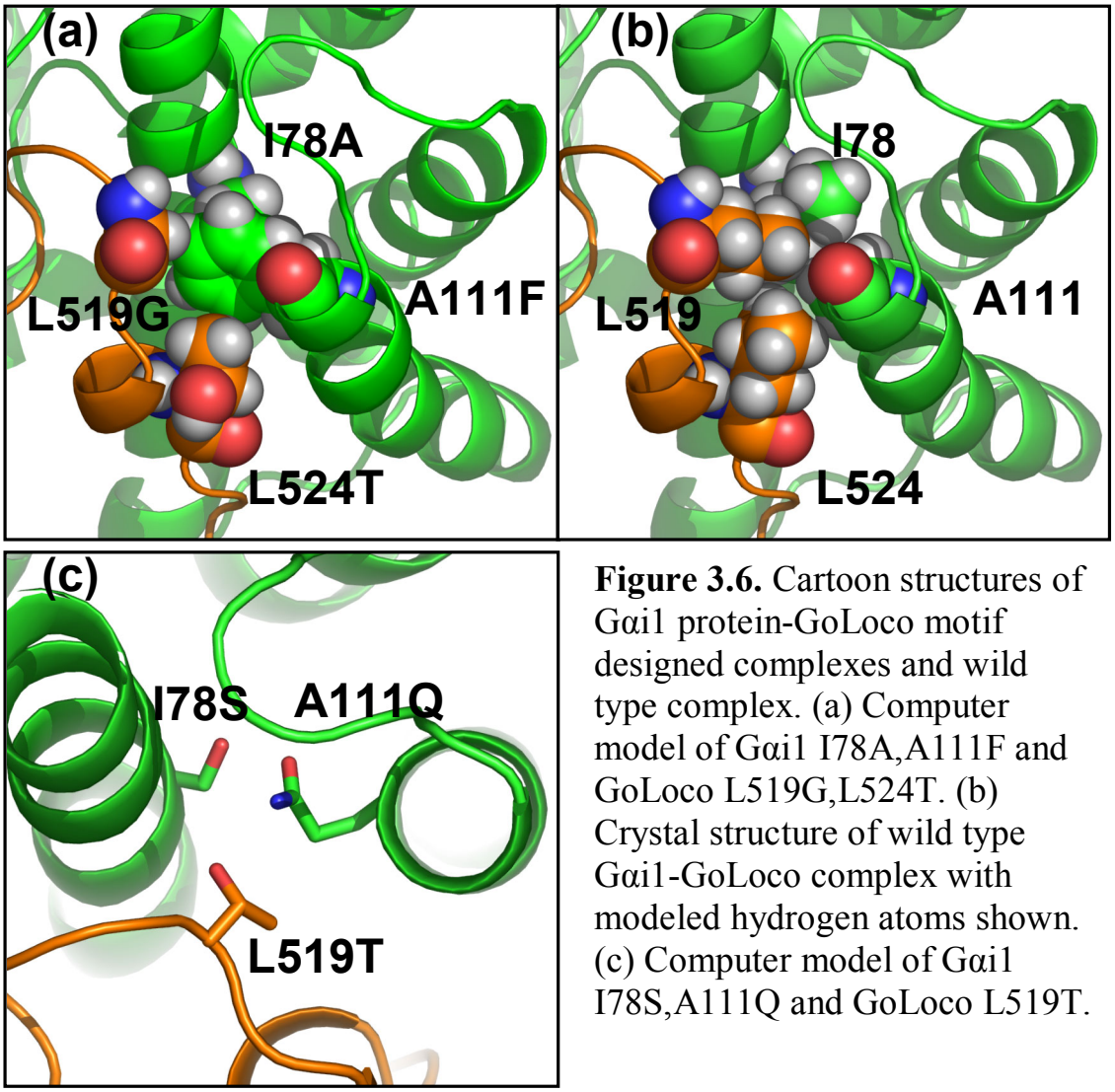


Figure 3.6. Cartoon structures of Gαi1 protein-GoLoco motif designed complexes and wild type complex. (a) Computer model of Gαi1 I78A,A111F and GoLoco L519G,L524T. (b) Crystal structure of wild type Gαi1-GoLoco complex with modeled hydrogen atoms shown. (c) Computer model of Gαi1 I78S,A111Q and GoLoco L519T.

Table 3.1. Predicted and experimentally characterized binding energies for orthogonal designs selected for experimental characterization.

Design	$G\alpha_{i1}$ mutations	GoLoco mutations	$\Delta\Delta G°_{Rosetta}$	$\Delta\Delta G°_{bind}$	K_d exp	Std Dev exp
	Wild type	wild type	0	0	0.10	0.02
1	E245L,L249A	V507M	0.1	-0.41	0.05	0.01
	E245L,L249A	wild type	2	0.24	0.15	0.03
	Wild type	V507M	0.6	1.39	1.04	0.06
2	K248E,S252L	L503K	0.3	-0.13	0.08	0.001
	K248E,S252L	wild type	0.4	0.69	0.32	0.06
	Wild type	L503K	2.5	0.52	0.24	0.02
3	F215K	L500E	0.1	>2	>3	4.9
	F215K	wild type	0.3	>2.3	>5	12.9
	Wild type	L500E	0.4	-	no binding detected	
4	I78S,A111Q	L519T	-0.5	>3	>17	56
	I78S,A111Q	wild type	0.9	-	not characterized	
	Wild type	L519T	0.9	-	not characterized	
5	R208E	E498R	-4.9	2.02	3.03	0.47
	R208E	wild type	-3.7	2.64	8.55	1.63
	Wild type	E498R	-1	-1.37	0.01	
6	I78A,A111F	L519G,L524T	1.2	1.73	1.85	1.65
	I78A,A111F	wild type	0.5	1.99	2.87	0.57
	Wild type	L519G,L524T	1.9	>2.6	>8	15.22
7	F223V,L249F	L504A	1.5	>1.8	>2	64.3
	F223V,L249F	wild type	108.7	>2	>3	23.7
	Wild type	L504A	1.5	1.89	2.41	0.95
8	I253T	L504Y	6.1	2.14	3.67	0.60
	I253T	wild type	0.8	0.92	0.47	0.07
	Wild type	L504Y	437.9	2.13	3.60	0.63
9	I253C	L504W	84.8	>2.3	>5	21.6
	I253C	wild type	0.8	0.95	0.50	
	Wild type	L504W	220.5	1.99	2.85	0.23
10	F95H,A101W	F529PL530Y	3.1	-	not characterized	
	F95H,A101W	wild type	0.1	-	not characterized	
	wild type	F529PL530Y	14.7	-	insoluble	
8.5	L249A,I253C	L504W		1.82	2.16	0.21

Table 3.2. Experimentally determined binding energies for designs 6 and 7 with affinity enhancing mutations.

	$G\alpha_{i1}$ mutations	GoLoco mutations	K_d exp.	Standard Deviation exp.
	wild type	wild type	0.10	0.02
Design 7	F223V,L249F	L504A	>2	64
	F223V,L249F	wild type	>3	24
	wild type	L504A	2.41	0.95
+ Q147L	F223V,L249F,Q147L	L504A	1.10	0.5
	F223V,L249F,Q147L	wild type	>3.5	3.4
	wild type	L504A	2.41	0.95
+Q147L, +E245L	F223V,L249F,Q147L,E245L	L504A	0.14	0.05
	F223V,L249F,Q147L,E245L	wild type	0.09	0.01
	wild type	L504A	2.41	0.95
+Q147L, +E245L, +F529W	F223V,L249F,Q147L,E245L	L504A,F529W	0.04	0.01
	F223V,L249F,Q147L,E245L	wild type	0.09	0.01
	wild type	L504A,F529W	>5	-
Design 6	I78A,A111F	L519G,L524T	1.85	1.65
	I78A,A111F	wild type	2.87	0.57
	wild type	L519G,L524T	>8	15
+E245L	I78A,A111F,E245L	L519G,L524T	>7.5	2956
+F529W	I78A,A111F	L519G,L524T,F529W	>2.7	3.8

Supplementary Table 3.1. Computationally predicting $\Delta\Delta G_{bind}$ with varying degrees of side chain flexibility.

	$G\alpha_{i1}$ mutations	GoLoco mutations	$\Delta\Delta G°_{Rosetta}$ (default)	$\Delta\Delta G°_{Rosetta}$ (RN,RU)	$\Delta\Delta G°_{bind}$
	wild type	wild type	0	0	0
Design 1	E245L,L249A	V507M	0.1	-1.6	-2.05
	E245L,L249A	wild type	2	-2.2	-0.72
	wild type	V507M	0.6	12.6	0.28
Design 2	K248E,S252L	L503K	0.3	-0.2	-0.10
	K248E,S252L	wild type	0.4	16.7	0.72
	wild type	L503K	2.5	2.8	0.55
Design 3	F215K	L500E	0.1	-1.3	>2
	F215K	wild type	0.3	0.7	>2.3
	wild type	L500E	0.4	-0.7	-
Design 4	I78S,A111Q	L519T	-0.5	-0.1	>3
	I78S,A111Q	wild type	0.9	21.6	-
	wild type	L519T	0.9	0.3	-
Design 5	R208E	E498R	-4.9	0.3	2.05
	R208E	wild type	-3.7	1.5	2.67
	wild type	E498R	-1	0.3	-1.40
Design 6	I78A,A111F	L519G,L524T	1.2	2.5	1.76
	I78A,A111F	wild type	0.5	141.2	2.02
	wild type	L519G,L524T	1.9	3.3	>2.6
Design 7	F223V,L249F	L504A	1.5	0.4	>1.8
	F223V,L249F	wild type	108.7	215.4	>2
	wild type	L504A	1.5	1	1.92
Design 8	I253T	L504Y	6.1	3	2.17
	I253T	wild type	0.8	-0.9	0.95
	wild type	L504Y	437.9	35.9	2.16
Design 9	I253C	L504W	84.8	29.8	>2.4
	I253C	wild type	0.8	-1	0.99
	wild type	L504W	220.5	335.2	2.02
Design 10	F95H,A101W	F529PL530Y	3.1	17.3	-
	F95H,A101W	wild type	0.1	0.8	-
	wild type	F529PL530Y	14.7	25.2	-
Design 8.5	L249A,I253C	L504W			1.85

$\Delta\Delta G°_{Rosetta}$ (default) = fixed backbone and side chains (except for the site of mutation)

$\Delta\Delta G°_{Rosetta}$ (RN,RU) = repack neighbors, relax unbound

METHODS

Rosetta

All energy calculations and side chain and backbone relaxation simulations were performed with the molecular modeling program Rosetta[7].
We used the following command lines to run the alter_spec protocol and generate sets of mutations designed to alter binding specificity.

The command line for design 1 ($G\alpha_{i1}$ E245L,L249A and GoLoco V507M):

rosetta.gcc –s gpep1.pdb –design –alter_spec –alter_spec_mutlist mutlist.

The input pdb, gpep1.pdb, is from the crystal structure 2OM2. The mode of rosetta used is the protein design mode, indicated by –design. The specific protocol used in design mode is alter_spec. The output file containing the sets of mutations selected is renamed using the –alter_spec_mutlist command and is renamed to mutlist.

The GoLoco mutation V507M was predicted to significantly destabilize the interaction due to van der Waals clashes. The experimental characterization demonstrated that this mutation did not significantly destabilize the $G\alpha_{i1}$:GoLoco interaction. Methionine is a very flexible residue and was probably not represented by enough rotamers, or discrete representations of side-chain

conformations[1], resulting in an incorrect prediction. Negative protein design requires a large rotamer library, and so we added the command "select_rotamer_set = large" to the code.

Design 2 ($G\alpha_{i1}$ K248E,S252L and GoLoco L503K), design 3 ($G\alpha_{i1}$ F215K and GoLoco L500E), design 6 ($G\alpha_{i1}$ I78A,A111F and GoLoco L519G,L524T) and design 7 ($G\alpha_{i1}$ F223V,L249F and GoLoco L504A) were generated using the following command line:

rosetta.gcc -s gpep1.pdb –design –alter_spec –alter_spec_mutlist mutlist_10_2_fix –fix fix_9_17

The –fix command allows for a file to be read into the program that specifies residues which are not to be mutated. In this case the file was named fix_9_17 and included residues 39,40,41,42,43 and 178 from the $G\alpha_{i1}$ and residues 508,514,515 and 516 from the GoLoco. These residues have been shown to contact or be important in stabilizing the bound nucleotide or appear to form hydrogen bonds with a backbone residue. A design run was also performed using the –fix input file to hold fixed the $G\alpha_{i1}$ residues that are in the region termed switch II, as mutations in this region can alter the intrinsic GTPase activity[19]. Limiting the number of residues that may be used in a redesign decreased the total number of output designs but did not alter the designs that

were generated and so no designs were selected for experimental characterization from this run.

Design 4 ($G\alpha_{i1}$ I78S,A111Q and GoLoco L519T) and design 10 ($G\alpha_{i1}$ F95H,A101W and GoLoco L529P,L530Y) were generated using the following command line:

rosetta.gcc –s gpep1.pdb –design –alter_spec –alter_spec_mutlist mutlist –ex1 –ex2 –ex3

The –ex1 –ex2 and –ex3 flags are used to increase the rotamer library by adding rotamers that are expanded about the chi 1, chi 2 and chi 3 angles respectively. Design 10 ($G\alpha_{i1}$ R208E and GoLoco E498R) was selected by eye. It was selected because it is a charge-swap design and the alter_spec runs did not produce any clear charge-exchanging designs.

Binding Energy Calculations

The second step in our protocol computationally evaluates the $\Delta\Delta G_{binding}$ of all the designs produced in the alter_spec run. This protocol is called analyze_interface and can be found in the analyze_interface_ddg.cc file in the code. This command lines used are as follows:

(1) rosetta.gcc –s gpep1.pdb –interface –Wpack_only –repack_neighbors –alter_spec_format –mutlist mutlist –intout intout –output_structure –ex1 –ex2 –ex3

(2) rosetta.gcc –s gpep1.pdb –interface –Wpack_only –repack_neighbors –soft_rep_design –alter_spec_format –mutlist mutlist –intout intout –output_structure –ex1 –ex2 –ex3

(3) rosetta.gcc –s gpep1.pdb –interface –Wpack_only –repack_neighbors –relax_unbound –soft_rep_design –alter_spec_format –mutlist mutlist –intout intout –output_structure –ex1 –ex2 –ex3

We evaluated the designs by comparing the predicted $\Delta\Delta G_{binding}$ results from runs using different combinations of commands for the analyze_interface protocol, ranging from the most conservative (shown by command line 1 above), which designs only the residues that will alter their identity, to more flexible where the wild type interface is relaxed as well as all residues that neighbor mutation sites and all interface residues on the unbound structures (command line 3 above). The –interface flag indicate the rosetta mode designed to predict the free energy of binding for two proteins given structural coordinates and a list of mutations. The file containing a list of mutations is the file title mutlist. The output file can be renamed using the –intout command, with the name immediately following the command. The –output_structure command specifies that the structures generated, both wild-type and mutant, be output as pdb files. The –Wpack_only command specifies that the energy function used for both modeling and energy prediction is the rosetta energy function that has been parameterized to reproduce native amino acid sequences. The default energy function for the –interface mode was parameterized to predict experimentally

determined $\Delta\Delta G_{binding}$ data from alanine scan studies culled from literature[25]. The -repack_neighbors flag indicates that any residues neighboring a mutated residue should be allowed to relax. The –relax_unbound flag allows all residues at the protein-protein interface to relax in the unbound structure. The –relax_unbound flag will only be utilized if –repack_neighbors and/or min_interface is also indicated in the command line. The reason can be understood when considering how analyze_interface determines the computational $\Delta\Delta G_{bind}$;

$$\Delta\Delta G_{bind} = \Delta G_{mutant} - \Delta G_{wild\ type}$$

$$\text{where } \Delta G_{mutant} = G_{mut\ complex} - G_{mut\ chA} - G_{mut\ chB}$$

$$\text{and } \Delta G_{wild\ type} = G_{wt\ complex} - G_{wt\ chA} - G_{wt\ chB}.$$

The analyze_interface protocol by default only moves residues that are changing identity; ideally this would be only three to six residues. The repack_neighbors flag, on the other hand, moves all residues at the protein-protein interface. If a user were to use the repack_neighbors flag in the default mode then the predicted energy of the unbound structures could be significantly lowered and might no longer compare to the predicted energy of the crystal structure coordinates, resulting in a significant and artificial increase in the number of positive ΔG_{mutant} and $\Delta G_{wild\ type}$ predictions. The repack_neighbors flag not only allows all residues determined to be neighbors of a mutated residue to adopt a lower-energy conformation but also relaxes the entire wild-type interface. Thus

the bound and unbound structures should start from the same or very similar conformations. The –soft_rep_design flag dampens repulsion energies to allow for small atom-atom clashes that may be accommodated by small changes in side chain and backbone conformation. The –alter_spec_format flag outputs the results in a format that visually clusters the results by design, with the mutant:mutant, mutant:wild-type and wild-type:mutant $\Delta\Delta G_{bind}$ are grouped.

Binding energies were calculated by subtracting the energy of the complex from the energies of the individual chains. In our simplest protocol only the mutated residue is allowed to relax to a different rotamer, and it is assumed to adopt the same conformation in the unbound state. Our next level of complexity is to allow residues surrounding the mutated residue to relax to alternate side chain rotamers (repack_neighbors). If the repack_neighbors option is true, the binding calculation begins by optimizing the conformation of all the side chains at the protein-protein interface in the wild type structure. This minimized structure is used to calculate the binding energy of the wild type structure and serves as the starting structure for calculating the binding energy of the mutated structure. Before calculating the energy of the mutant complex, amino acids which are close enough to have non-zero energy with the mutated residue are allowed to relax to more favorable rotamers as identified by Rosetta's side chain repacking routine. In the simplest case, the side chains are assumed to adopt the same rotamer in the unbound state. If the relax_unbound option is specified, side chains near the point of mutation are relaxed separately in the bound and

unbound state. The same residues are also relaxed in the bound and unbound state in the wild type structure.

Our protocol with the most degrees of freedom allows side chains to adopt alternative rotamers and performs gradient-based minimization on backbone and side chain torsion angles and rigid body displacement. The flexible backbone procedure begins by minimizing the wild type complex. This serves as the starting point for the calculations with the mutant complexes. If the relax_unbound option is true, separate gradient based minimization is performed on bound and unbound molecules. The protocol with full side chain and backbone flexibility does not converge to the same result each time it is performed. 100 separate simulations were performed and the lowest energy wild type and mutant complexes, as well as wild type and mutant unbound chains were used to calculate binding energies.

Construction and cloning of protein designs

The DNA sequence for the GoLoco motif of RGS14 (residues 496-531) was cloned into pET21b as a C-terminal fusion to the small protein Tenascin. Tenascin was included to aid in the expression and purification of the peptide. The sequence for a hexahistidine tag was placed at the C-terminus of the construct. Residue G498 of the wild type GoLoco motif was mutated to a cysteine to enable the covalent labeling of the thiol-reactive fluorescent probe 6-iodoacetamidofluorescein (6-IAF) (Molecular Probes). We used an N-terminal-truncated, hexahistidine-tagged expression construct of human $G\alpha_{i1}$ with the first 25 codons of the Gα open reading frame removed, as previously described.[50]

Point mutations were introduced using the QuickChange® site-directed mutagenesis protocol (Stratagene) and all vectors were verified by DNA sequencing.

Protein purification

As described in chapter II, GoLoco motif peptide was expressed either for 4 hours at 37 ˚C or overnight at 25 ˚C with 0.5 mM IPTG in the BL21(DE3) strain of E. coli. $G\alpha_{i1}$ was expressed overnight at 25 ˚C with 1 mM IPTG in the BL21(DE3) strain of E. coli. Cells were lysed using an Avestin emulsiflex and the resulting lysates were cleared by ultracentrifugation. The $G\alpha_{i1}$ ΔN2 and RGS14-GoLoco motif-Tenascin fusion proteins were purified using a HiTrap (Amersham Biosciences) column by eluting the protein with an imidazole step gradient, then followed by gel filtration with a Superdex-200 column (Amersham Biosciences). Proteins were concentrated using Vivaspin 20® centrifugal concentrators. Protein concentrations were determined by measuring absorbance at 280nm. Extinction coefficients were calculated using the method of Gill and von Hipple[25].

Fluorescence polarization binding analysis

A thiol-reactive fluorescent probe 6-iodoacetamidofluorescein (6-IAF) (Molecular Probes) was conjugated to the unique cysteine on the GoLoco motif using the manufacture's recommended protocol. We do not consider any designs that include the addition of a cysteine to the GoLoco motif to prevent multiple fluorescent labeling. GoLoco motif protein was buffer exchanged into 50 mM Tris-Cl pH 7.5 using a PD10 desalting column, concentrated to ~50 to 100

uM. Next 1 mM TCEP was added to the PD10 eluate and stirred for one hour at room temperature. A 20 mM stock solution of 6-IAF suspended in dimethyl sulfoxide (DMSO) was diluted into the GoLoco motif protein solution to a 10-fold molar excess and the conjugation reaction was allowed to proceed overnight, in the dark at 4 °C. Precipitate was pelleted and discarded, and 5 mM β-mercaptoethanol (β-ME) was added to quench the reaction. The supernatant containing fluorescein-GoLoco motif protein was run over a PD10 column to separate free probe from labeled protein. The concentration of fluorescein-GoLoco motif protein was quantified using UV/Vis, taking readings at 280 and 495 nm for the protein and fluorophore respectively. The Tenascin-GoLoco fusion protein was used in the binding assays.

Fluorescence polarization assays were carried out on a Jobin Yvon Horiba Spec FluoroLog-3 instrument (Jobin Yvon Inc.) performed in L-format with the excitation wavelength set at 495 nm and the emission wavelength set at 520 nm. Titrations were performed using a 3 x 3-mm quartz cuvette with a starting volume of 200 µL. Fluorescein labeled wild type or mutant GoLoco motif protein was diluted to 50 to 150 nM and the excitation and emission slit widths adjusted to give a fluorescence intensity >100,000 counts per second. Wild type or mutant $G\alpha_{i1}$ was added in increasing volumes from a stock solution whose initial concentration depended on the strength of the interaction, generally having a concentration of 3-10 µM. Two to three polarization readings consisting of 3 averaged measurements were collected for increasing concentrations of $G\alpha_{i1}$.

Data was averaged and analyzed using a model for single site binding according, as described previously[15].

REFERENCES

1. Dunbrack, R. L., Jr. & Cohen, F. E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci* **6**, 1661-81.

2. Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D. & Zardecki, C. (2002). The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* **58**, 899-907.

3. Shifman, J. M. & Mayo, S. L. (2002). Modulating calmodulin binding specificity through computational protein design. *J Mol Biol* **323**, 417-23.

4. Havranek, J. J. & Harbury, P. B. (2003). Automated design of specificity in molecular recognition. *Nat Struct Biol* **10**, 45-52.

5. Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L. & Baker, D. (2004). Computational redesign of protein-protein interaction specificity. *Nat Struct Mol Biol* **11**, 371-9.

6. Joachimiak, L. A., Kortemme, T., Stoddard, B. L. & Baker, D. (2006). Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. *J Mol Biol* **361**, 195-208.

7. Rohl, C. A., Strauss, C. E., Misura, K. M. & Baker, D. (2004). Protein structure prediction using Rosetta. *Methods Enzymol* **383**, 66-93.

8. Bradley, P., Misura, K. M. & Baker, D. (2005). Toward high-resolution de novo structure prediction for small proteins. *Science* **309**, 1868-71.

9. Bradley, P., Chivian, D., Meiler, J., Misura, K. M., Rohl, C. A., Schief, W. R., Wedemeyer, W. J., Schueler-Furman, O., Murphy, P., Schonbrun, J., Strauss, C. E. & Baker, D. (2003). Rosetta predictions in CASP5: successes, failures, and prospects for complete automation. *Proteins* **53 Suppl 6**, 457-68.

10. Lesk, A. M., Lo Conte, L. & Hubbard, T. J. (2001). Assessment of novel fold targets in CASP4: predictions of three-dimensional structures, secondary structures, and interresidue contacts. *Proteins* **Suppl 5**, 98-118.

11. Schueler-Furman, O., Wang, C., Bradley, P., Misura, K. & Baker, D. (2005). Progress in modeling of protein structures and interactions. *Science* **310**, 638-42.

12. Gray, J. J., Moughon, S. E., Kortemme, T., Schueler-Furman, O., Misura, K. M., Morozov, A. V. & Baker, D. (2003). Protein-protein docking predictions for the CAPRI experiment. *Proteins* **52**, 118-22.

13. Kuhlman, B. & Baker, D. (2000). Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A* **97**, 10383-8.

14. Kortemme, T., Kim, D. E. & Baker, D. (2004). Computational alanine scanning of protein-protein interfaces. *Sci STKE* **2004**, pl2.

15. Sammond, D. W., Eletr, Z. M., Purbeck, C., Kimple, R. J., Siderovski, D. P. & Kuhlman, B. (2007). Structure-based protocol for identifying mutations that enhance protein-protein binding affinities. *J Mol Biol* **371**, 1392-404.

16. Sood, V. D. & Baker, D. (2006). Recapitulation and design of protein binding peptide structures and sequences. *J Mol Biol* **357**, 917-27.

17. Kimple, R. J., Kimple, M. E., Betts, L., Sondek, J. & Siderovski, D. P. (2002). Structural determinants for GoLoco-induced inhibition of nucleotide release by Galpha subunits. *Nature* **416**, 878-81.

18. Sprang, S. R. (1997). G protein mechanisms: insights from structural analysis. *Annu Rev Biochem* **66**, 639-78.

19. Thomas, C. J., Du, X., Li, P., Wang, Y., Ross, E. M. & Sprang, S. R. (2004). Uncoupling conformational change from GTP hydrolysis in a heterotrimeric G protein alpha-subunit. *Proc Natl Acad Sci U S A* **101**, 7560-5.

20. Palmer, A. E., Jin, C., Reed, J. C. & Tsien, R. Y. (2004). Bcl-2-mediated alterations in endoplasmic reticulum Ca2+ analyzed with an improved genetically encoded fluorescent sensor. *Proc Natl Acad Sci U S A* **101**, 17404-9.

21. Dantas, G., Kuhlman, B., Callender, D., Wong, M. & Baker, D. (2003). A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J Mol Biol* **332**, 449-60.

22. Kortemme, T., Morozov, A. V. & Baker, D. (2003). An orientation-dependent hydrogen bonding potential improves prediction of specificity

and structure for proteins and protein-protein complexes. *J Mol Biol* **326**, 1239-59.

23. Davis, I. W., Arendall, W. B., 3rd, Richardson, D. C. & Richardson, J. S. (2006). The backrub motion: how protein backbone shrugs when a sidechain dances. *Structure* **14**, 265-74.

24. Green, D. F., Dennis, A. T., Fam, P. S., Tidor, B. & Jasanoff, A. (2006). Rational design of new binding specificity by simultaneous mutagenesis of calmodulin and a target peptide. *Biochemistry* **45**, 12547-59.

25. Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* **182**, 319-26.

CHAPTER IV

PARTIAL *de novo* PROTEIN-PEPTIDE INTERFACE DESIGN

ABSTRACT

The ability to design a peptide *de novo* to bind with any desired target has biomedical and industrial applications ranging from vaccine development to biosensor design. Taking a step towards this ultimate goal, we describe a method to computationally redesign the secondary structure of a component of a protein-peptide interface while preserving a biochemically relevant binding affinity. Using the $G\alpha_{i1}$ protein in complex with the RGS14 GoLoco motif as our model system, we redesigned the C-terminal portion of the GoLoco, converting it from a random coil to an alpha-helix, in the context of binding the wild type $G\alpha_{i1}$ protein. We used the Rosetta software for the backbone and side chain redesign. We used fluorescence anisotropy to experimentally evaluate the dissociation constants for the wild type interaction, 95 nM, and the newly designed interaction, 810 nM. We then performed mutational studies on both the Gαi1 protein and the redesigned RGS14 GoLoco motif to confirm the location of binding for the redesigned GoLoco motif.

INTRODUCTION

Protein-protein interactions, which govern much of the functioning of cells, are manipulated and utilized for both research and industrial uses. Biosensors have been developed and optimized by creating or changing protein interactions.[1; 2] Therapeutics, both small molecule and protein therapeutics, seek to alter protein interaction pathways. The energetic forces that lead to protein-protein recognition, including hydrogen bonding and hydrophobic interactions, have stringent geometric requirements.[3; 4; 5; 6; 7; 8] Computational protein design seeks to model these interactions by balancing physical accuracy with speed and memory limitations.[9] Additional geometric constraints are imposed when trying to redesigning a protein backbone and sequence while trying to maintain the proteins' naturally occurring interactions. The promise of computational protein design is to increase the scope of what we can attempt while decreasing the time required for development.

Realizing the potential for protein interface design will require improvements in how we model protein backbone flexibility.[10] A number of approaches have been described to date. Small rotations in backbone angles[11] or slight perturbations in backbone coordinates[12] have the potential to increase sequence search space and as well as accuracy. A docking algorithm has been applied to convert a monomeric protein into a dimer.[13] New backbone structures have been designed for helix bundles using a coiled-coil topology,[14] and, in the case of a monomeric protein, an entirely new fold has been designed.[15]

Here we describe a computational method that will design a distinctly different backbone secondary structure and compatible sequence for a portion of a peptide while maintaining a biochemically relevant binding affinity for one of its naturally occurring binding partners. (Figure 4.1) We use the protein design software, Rosetta, to build the new backbone coordinates, design new compatible sequences and evaluate the free energies of binding for the designs. Rosetta began as protein structure prediction software.[16] Rosetta has made significant contributions toward the field of structure prediction,[17] and was further developed to include computational protein design.[18] Rosetta has been used for alanine scanning of protein-protein interfaces,[19] altering the binding specificity of protein interactions,[20; 21] and we have used it to predict point mutations that enhance protein-protein binding affinity.[22] Especially relevant to the work presented here, Rosetta has been used for protein backbone design, including the redesign of a loop region of a protein,[23] design of a novel monomeric protein[15] and increasing binding affinity by lengthening peptides thereby increasing the buried surface area of the protein-peptide interface.[24] The protocol we use to redesign the backbone of the C-terminal portion of a 36-amino acid peptide is based on methods used by the Rosetta structure prediction protocol. Fragments of protein backbones taken from a non-homologous set of structures from the Protein Data Bank[25] were combined, generating ensembles of redesigned protein structures. The amino acid sequences for these new backbone structures were generated using the Rosetta simulated annealing protocol.[18]

The model system we used is the $G\alpha_{i1}$ protein from the G-protein signaling system and one of its binding partners, the GoLoco motif from the RGS14 multidomain protein.[26] The RGS14 GoLoco motif interacts with both the Ras-like, guanine nucleotide binding domain and the all-helical domain of $G\alpha_{i1}$.[26] The C-terminal portion of the GoLoco binds between αA and αB of the all-helical $G\alpha_{i1}$ domain. We redesigned and lengthened the C-terminal portion of the GoLoco peptide in order to create an alpha helical backbone while maintaining the contacts with the $G\alpha_{i1}$, holding the sequence of the $G\alpha_{i1}$ protein fixed.

RESULTS

The C-terminal portion of the RGS14 GoLoco motif in complex with the $G\alpha_{i1}$ protein, starting from residue 519 on the $G\alpha_{i1}$-GoLoco domain (PDB ID: 2OM2),[22] was computationally redesigned from a random coil to an alpha helix. (Figure 4.2) A compatible sequence was designed for the new, redesigned GoLoco backbone, allowing the contacting $G\alpha_{i1}$ side-chains and backbone to relax during sequence design. Designs were selected for experimental characterization based on the proper location of the C-terminal GoLoco helix, the predicted $\Delta G_{binding}$, and the absence of the introduction of unsatisfied hydrogen bonding partners and high quality hydrogen bonds in the helix.

About the Gαi1:GoLoco structure: Selection of the design site

The RGS14 GoLoco motif spans the GTP-binding domain and the all-helical domain of the $G\alpha_{i1}$ protein. The portion of the interface comprising the all-helical domain of the $G\alpha_{i1}$ protein and the C-terminal region of the GoLoco motif is thought to dictate specificity for the GoLoco-Gα protein interfaces.[27] This segment of the interface buries approximately 660 $Å^2$ of surface area of the roughly 1900 $Å^2$ of buried surface area across the whole interface. Members of the GoLoco family bind the $G\alpha_{i/o}$ proteins with affinities ranging from 4 uM to 19 nM. [28],[29] The residues in the C-terminal portions of the GoLoco motif aid in affinity, as shown by Adhikari et al.[29] We verify the importance of this region of the RGS14 GoLoco motif by comparing the binding affinity of the wild-type interaction to that of a truncated version, with all residues following the conserved

DQRG motif removed, obtaining a dissociation constant of 95nM and 20 μM respectively.

Residues just preceding the C-terminal random coil of the RGS14 GoLoco motif are crucial to the biochemical function of this G-protein regulatory domain. The residues in the conserved set of residues act to stabilize the guanosine diphosphate (GDP) bound to the $G\alpha_{i1}$ protein. Mutational analysis done on these residues show them to be crucial for binding.[26; 30; 31] Redesign in this area would likely result in the loss of binding and/or of guanine nucleotide dissociation inhibitor (GDI) activity for the GoLoco motif. We therefore held this region, the conserved DQR triad as well as the residues immediately adjacent to it, fixed. The αA and αB helices in the all-helical domain of the $G\alpha_{i1}$ protein make a groove where the C-terminal random coil of the GoLoco binds. A phenylalanine at position 108 in the $G\alpha_{i1}$ structure appears to occlude the groove where we wish to design our helix. We sought to design the helix to bind in the groove, keeping in mind that the design had to include realistic chi angles for Phenylalanine 108.

Computational Redesign of backbone: Design of the GoLoco C-terminal helix

The C-terminal portion of the GoLoco motif adopts a random coil when bound to the all-helical domain of the $G\alpha_{i1}$.[26] Redesigning this region to be an alpha helix accomplishes two things; first, the secondary structure will be starkly different from that of the wild type, and second the coil can interact with two of the helices in the $G\alpha_{i1}$ all-helical domain making a three coiled bundle. We first

decided on the location to begin our design. We selected sequence position 519 of the GoLoco domain to begin our redesign based on the known biochemically crucial GoLoco region of the interface, namely the conserved three residues termed the conserved triad, which play a role in stabilizing the bound GDP. We held these highly conserved residues fixed as well as two residues C-terminal to the conserved triad to protect this region of the interface. The additional two residues C-terminal to the conserved triad were selected to give a buffer between the designed helix and the conserved triad in case altering the GoLoco backbone caused torsion angle movement that could propagate towards the conserved triad. The challenge here was to stabilize the αB helix on the $G\alpha_{i1}$ protein in the conformation seen in the Gαi1-GoLoco complex (PDB ID: 2OM2).[26] Residues 83 through 88 on the $G\alpha_{i1}$ protein adopt a loop conformation, opening up the region of the binding interface with the GoLoco motif. The same residues (numbered 112 to 117 in PDB ID: 1ASO) adopt an alpha helical conformation in the GTPgammaS, unbound $G\alpha_{i1}$ structure,[32] displacing the residues by up to 5 Å and appearing to occlude this portion of the GoLoco binding site. We addressed this structural issue by designing the new helix to start at the same height as the wild type random coil.

Next we generated libraries of protein backbone fragments that would be used to build the C-terminal GoLoco helical backbone. These fragments consist of helical backbone segments from a set of non-homologous pdb structures taken from the Protein Data Bank[25]. The random coil region in the GoLoco motif was replaced by a newly designed alpha helix using Rosetta to piece together

the alpha helical pdb fragments to generate the ensembles of structures. Each library contained 1000 structures. The redesigned alpha helical regions in the GoLoco motif were directed to bind in the groove between αA and αB helices of the $G\alpha_{i1}$ all-helical region.

We additionally set a starting sequence for the new helical region. The side-chain coordinates for all residues in the $G\alpha_{i1}$ protein and the wild-type and redesigned portions of the GoLoco domain were replaced by reduced van der Walls spheres to aid in computational speed during the backbone design protocol.[16] Some side chain information including the chemical character of the amino acids was retained in this portion of the protocol, even though the side-chain coordinates are removed. For example, the size of the van der Waals sphere changed based on the size of the all-atom side chain. Using this side chain information in the centroid design mode, we set a starting sequence for the redesigned helix. We set the buried residues to large hydrophobic residues to encourage sufficient space between the C-terminal GoLoco helix and the all-helical region of the $G\alpha_{i1}$. We ran a risk of designing a protein that would either self-associate or form higher-order oligomers because we were designing an amphipathic helix in a peptide already containing an amphipathic helix in the N-terminal region. We believe our previous designs may have failed at least in part due to self-association. (Figure 4.3 panels a through i). The truncated GoLoco bound with a dissociation constant of 20 μM while the previous designs ranged from a K_d of greater than 57 μM to no binding detected. We addressed the solubility issue by setting the starting residues for the solvent exposed region of

the helix to have similar charge to those of the N-terminal wild-type GoLoco helix so that the two helices will not be electrostatically complementary. The final step in our protocol iterates for four cycles between designing the protein sequence for the newly designed backbone using the simulated annealing protocol,[18] and performing gradient based minimization on the backbone and side chains. The Lennard-Jones repulsion term is down-weighted for the initial cycle, and gradually increased through the final cycle so that Lennard-Jones repulsion energy does not push the two proteins apart or promote the exclusive use of small amino acids at the interface.

We selected designs for experimental validation based upon several factors. First, we removed from consideration any computational designs that did not bind in the groove between the αA and αB helices on the all-helical domain of $G\alpha_{i1}$. Second, we also did not consider designs where buried polar residues not participating in hydrogen bonds had been introduced, as well as poor hydrogen bonding in the newly designed C-terminal alpha helix. We then ranked the designs based on the predicted $\Delta G_{binding}$ and the quality of the hydrophobic packing at the binding interface. Both the $\Delta G_{binding}$ and the quality of the hydrophobic packing were evaluated using the interface mode of Rosetta. Binding energies were calculated by subtracting the calculated energy of each unbound protein from the calculated energy of the complex. The predicted $\Delta G_{bindings}$ vary significantly for designs selected from starting structures created using different amino acid initiation sequences or helix lengths. We therefore

select the best designs from each run for experimental validation. (Figure 4.3 and Supplementary Figure 4.1).

Experimental characterization of the redesigned GoLoco-$G\alpha_{i1}$ complex

The binding affinities were experimentally evaluated using fluorescence anisotropy, with the fluorophor covalently attached to the GoLoco peptide on a unique cysteine residue pointing away from the $G\alpha_{i1}$-GoLoco interface. We obtained a dissociation constant of 95 nM for the wild-type GoLoco-$G\alpha_{i1}$ complex and a dissociation constant of 810 nM for the redesigned GoLoco_Design11-$G\alpha_{i1}$ complex. Dissociation constants have been reported for members of the GoLoco domain family bound to $G\alpha_{i/o}$ proteins ranging from low nano-molar, including 19 nM for one of the four GoLoco motifs from the AGS3-C protein with GDP-bound $G\alpha_{i1}$ protein[29] to low micro-molar, including 2.7 to 5.6 μM for two separate GoLoco motifs from the GPSM2/LGN protein[28]. A dissociation constant of 810 nM falls well within a biochemically relevant range.

We then investigated whether the newly designed C-terminal GoLoco helix was binding to the αA and αB helices on the $G\alpha_{i1}$ by performing a mutational analysis on both the $G\alpha_{i1}$ protein as well as the redesigned GoLoco motif. We selected single and multiple mutations on the $G\alpha_{i1}$ based on Rosetta $\Delta\Delta G_{binding}$ predictions indicating a loss of stability. The mutations cover the full span of the αA and αB groove. (Figure 4.4) Significant loss of binding affinity is seen with all of the mutations tested. (Table 4.1 and Figure 4.5 panel a) We then performed circular dichroism on the wild-type $G\alpha_{i1}$ and the mutants to

ensure the mutations have not significantly destabilized the proteins. The spectra for the mutants match that of the wild-type $G\alpha_{i1}$ protein. We selected the buried hydrophobic residues on the redesigned GoLoco helix and mutated each residue to an alanine. Again, the mutations selected cover the entire groove between the αA and αB helices of the $G\alpha_{i1}$ protein where the GoLoco helix is designed to bind. The fully buried phenylalanine at position 531 results in a significant loss of binding affinity when mutated to an alanine. The Tryptophan at sequence position 534 at the very bottom of the interface results in a modest loss of affinity. The L528A, at the edge of the interface, and V524A, at the top of the interface, do not appear to have an effect on the binding affinity (Table 4.1 and Figure 4.5 panel b). The Leucine is perhaps too solvent exposed to result in a significant loss. The Valine is near a water pocket, which could decrease the effect from the loss of buried hydrophobic surface area. (Figure 4.4 panel b) We have not solved an x-ray crystal structure for the designed GoLoco_Design11-$G\alpha_{i1}$ complex to confirm that the C-terminal region of the GoLoco is forming an alpha helix.

Past Designs

The phenylalanine at position 108 of the $G\alpha_{i1}$ protein appears to occlude the hydrophobic groove between αA and αB helices. Our initial designs, shown in Figure 4.3, designs 1 through 9, afforded one mutation on the $G\alpha_{i1}$, changing the phenylalanine to an alanine. (Figure 4.3 panels a through i) Mutational analysis of $G\alpha_{i1}$ F108A showed a loss of binding, with the dissociation constant for the wild type complex at 95 nM and the F108A $G\alpha_{i1}$ complex at 17 μM.

Subsequent designs kept the wild type phenylalanine at position 108. (Figure 4.3 panels j through l). The designs (a) through (h) were initialized with a starting sequence composed of all alanines for the redesigned C-terminal GoLoco alpha helix. For some of the designs the redesigned GoLoco helix was so close to the $G\alpha_{i1}$ αA and αB helices that it appears larger hydrophobic amino acids were not accepted during the simulated annealing protocol due to an increase in the Lennard-Jones repulsive energy. We altered the initial design sequence to have large hydrophobic amino acids at the buried positions and amino acids that are found both buried and solvent exposed at the peripheral interface positions. Solubility of the redesigned GoLoco peptide may have been another issue that could be addressed by the C-terminal helix initiation sequence. In the event that the designed GoLoco amphipathic helices were self-associating, we introduced polar residues at the solvent exposed positions.

We tried designing the C-terminal region in the absence of the N-terminal region of GoLoco (Figure 4.3, (d)-(g)). Despite detecting binding using fluorescence polarization for two of the designs, we were unable to disrupt the binding with the introduction of point mutations on the $G\alpha_{i1}$. It is possible that the designed C-terminal helices were binding in the large hydrophobic groove in the Ras-like domain of the GDP-bound $G\alpha_{i1}$.

DISCUSSION

We designed a new backbone and sequence for the C-terminal portion of the RGS14-GoLoco motif, with the goal of changing the secondary structure from a random coil to an alpha helix. The redesign was done in the context of the $G\alpha_{i1}$ binding partner. Maintaining a nanomolar dissociation constant between a redesigned peptide with an altered secondary structure and an unaltered protein binding partner is an important step towards de novo protein-protein interface design. We confirmed the location of binding for the redesigned C-terminal GoLoco helix using site-directed mutagenesis. We have not confirmed the secondary structure of the redesigned GoLoco with x-ray crystallography.

The goal of designing protein-protein interfaces has led to several successful approaches. Huang and co-workers utilized a protocol with protein docking and sequence design to convert a monomeric protein to a dimer.[13] Sood and co-workers used Rosetta to extend peptides, thereby enhancing the binding affinity for wild-type binding partners.[24] The wild-type $G\alpha_{i1}$ protein has numerous binding partners, and several crystal structures of the $G\alpha_{i1}$ in complex with some of its binding partners have been solved. The hydrophobic groove in the all-helical domain of the $G\alpha_{i1}$ protein that the redesigned GoLoco motif binds to is only known to bind members of the GoLoco motif family. It is therefore a more rigorous target and highlights the success of this approach. The computational protocol used in this work can be applied to any protein-protein interface and is a rigorous test of the computational modeling of not just side-chain movement and energetics but also the protein backbone.

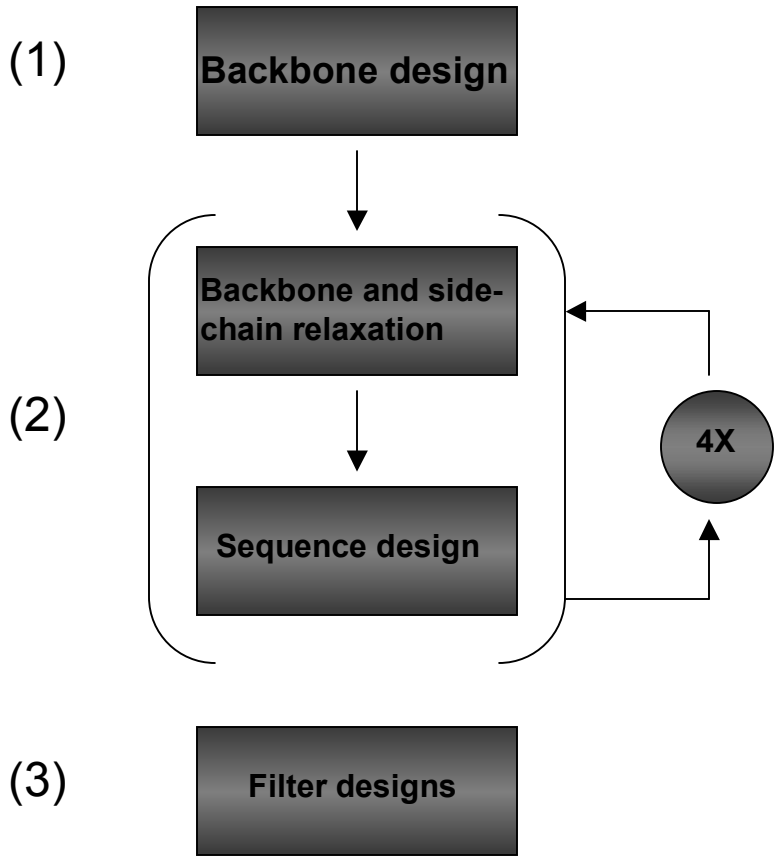
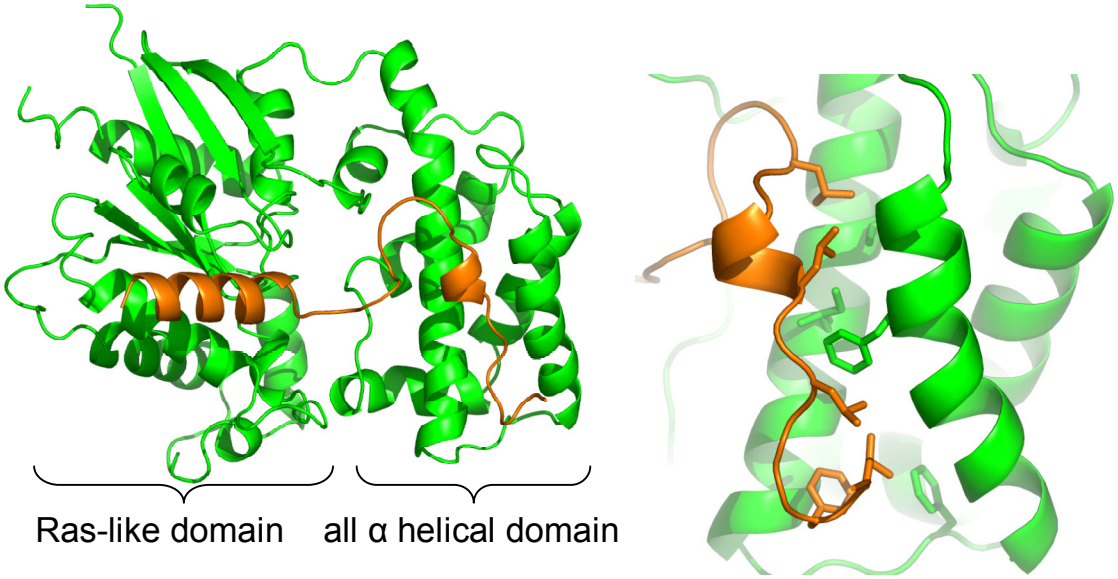


Figure 4.1. Schematic of the protocol used to redesign the C-terminal helix of RGS14 GoLoco motif in the context of the $G\alpha_{i1}$-GoLoco bound complex. Step (1) is the design of the initial backbone coordinates for the redesigned portion of the GoLoco motif. Step (2) relaxes the backbone and side chain angles followed by a simulated annealing search for the lowest energy sequence for the given structure, iterating between these steps 4 times. Step (3) involves filtering the output designs and ultimately selecting designs for experimental characterization.

(a) RGS14 GoLoco motif bound to the $G\alpha_{i1}$ protein.



(b) Redesigned RGS14 GoLoco_Helix9 bound to the $G\alpha_{i1}$ protein.

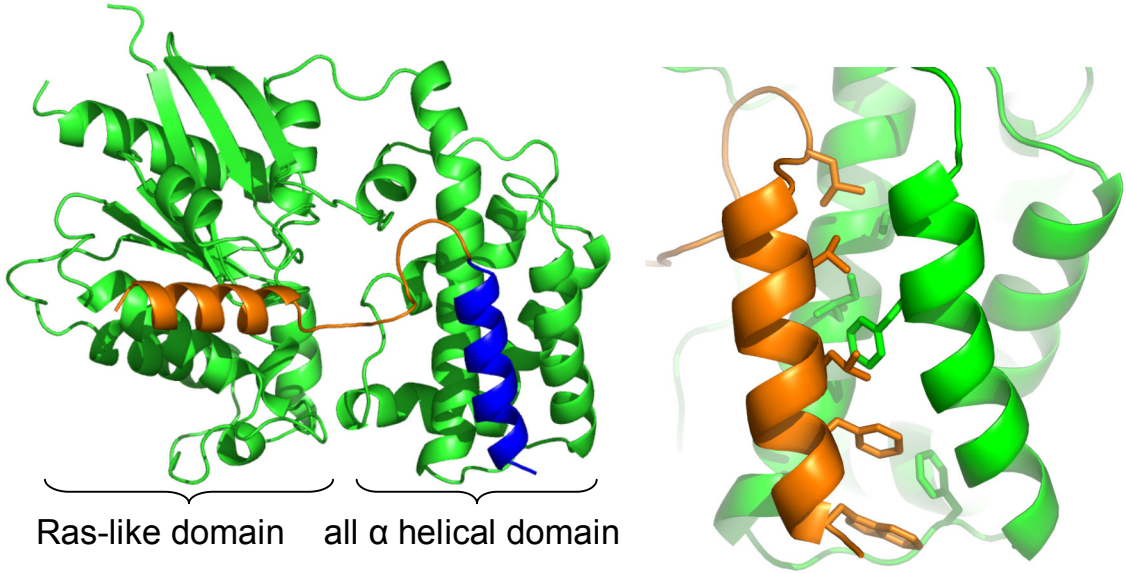
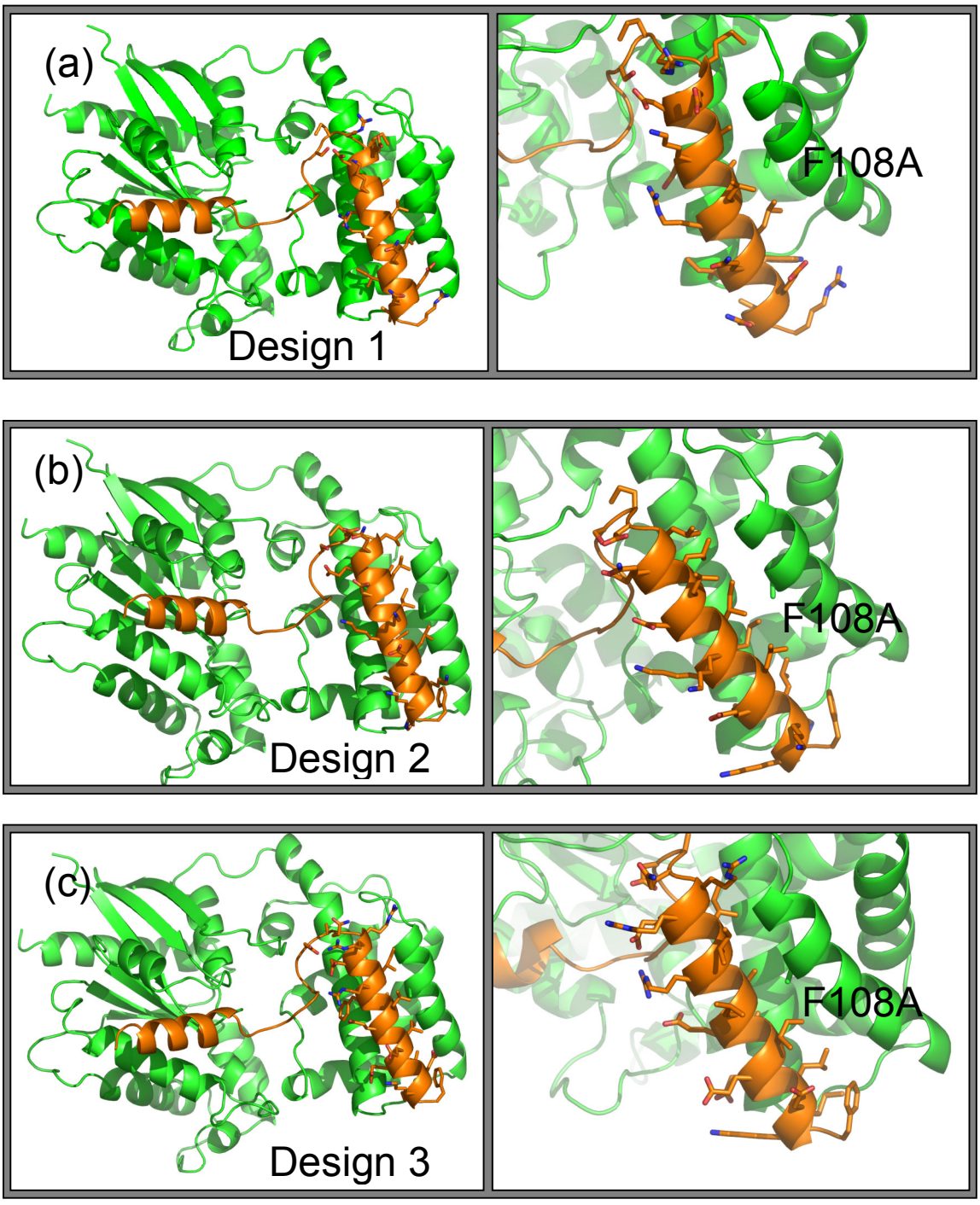
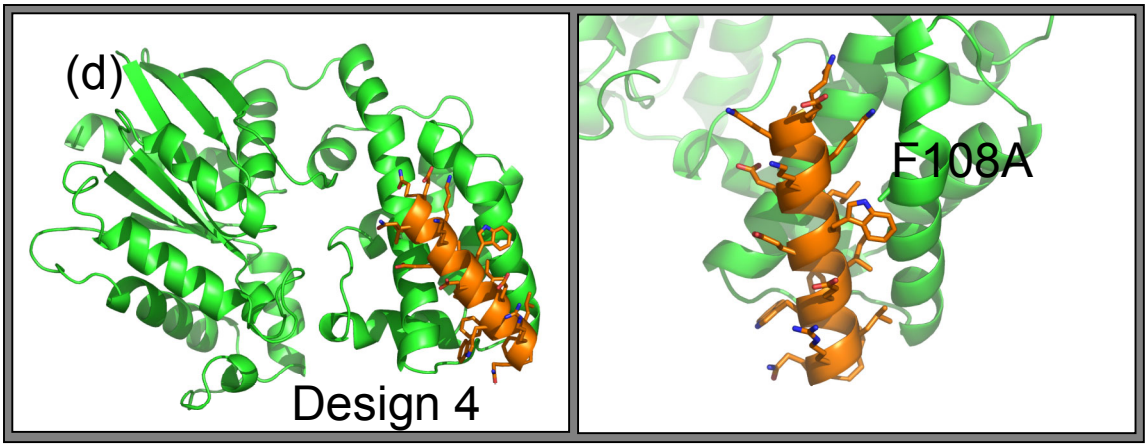
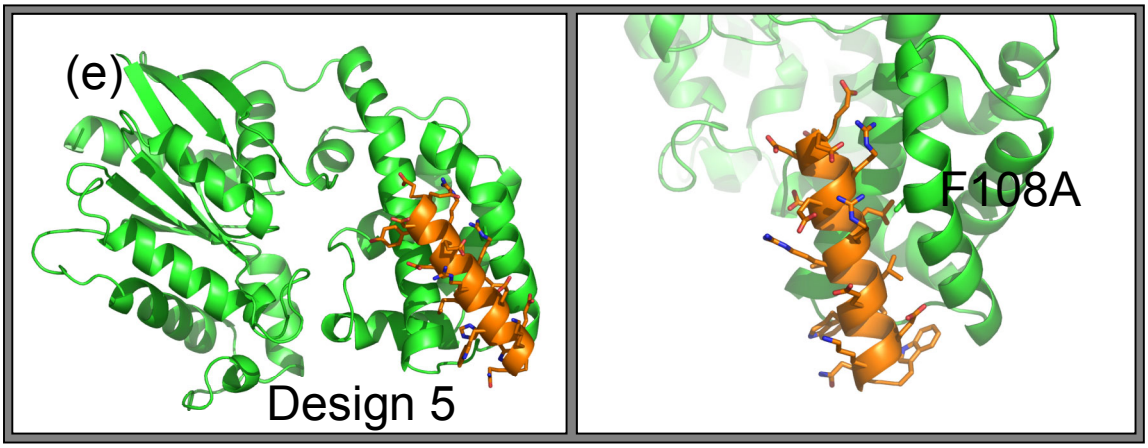
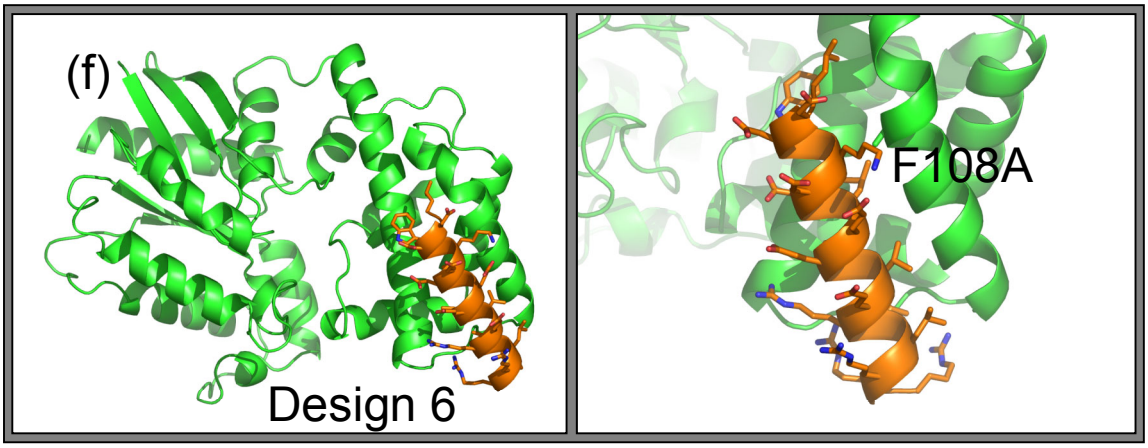


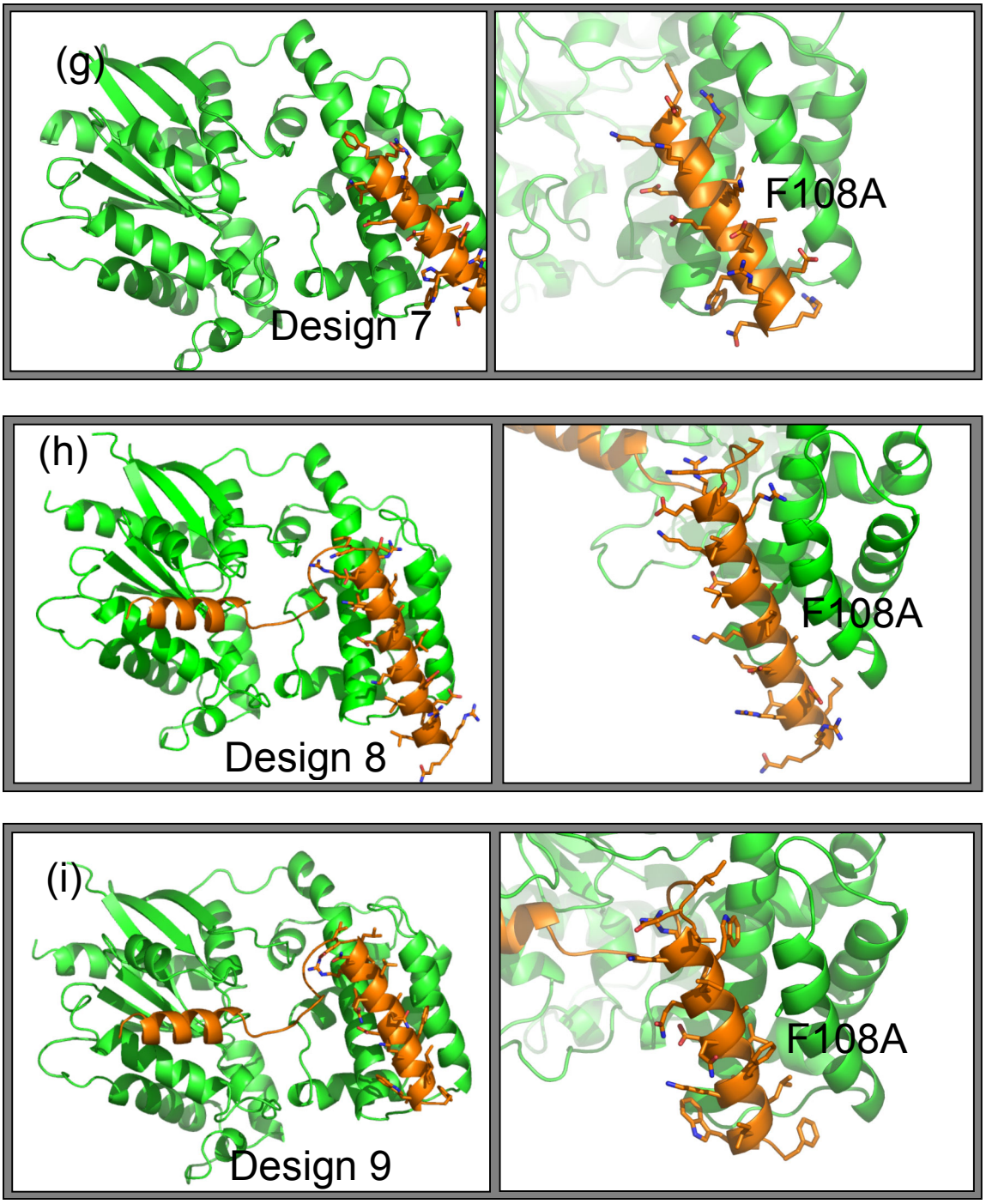
Figure 4.2. (a) The x-ray crystal structure of the RGS14 GoLoco motif bound to the $G\alpha_{i1}$ protein (PDB ID 2OM2), and (b) the model of the redesigned GoLoco motif with the C-terminal region redesigned (shown in blue) to form a helix, bound to the $G\alpha_{i1}$ protein


(a)
Design 1
F108A
(b)
Design 2
F108A
(c)
Design 3
F108A


(d)
Design 4
F108A


(e)
Design 5
F108A


(f)
Design 6
F108A


(g)
Design 7
F108A
(h)
Design 8
F108A
(i)
Design 9
F108A

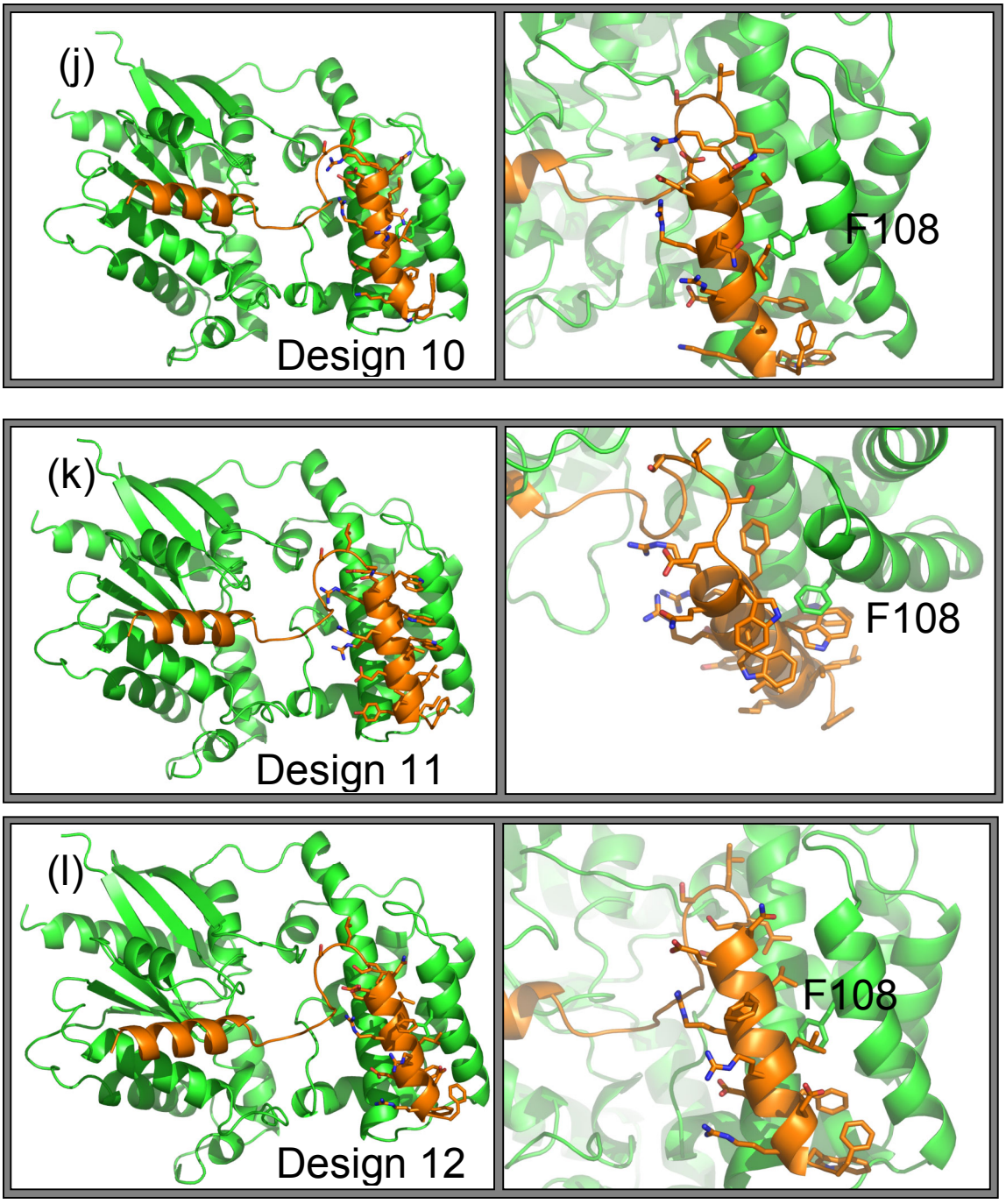


Figure 4.3. Computationally modeled C-terminal GoLoco designs (a) Design 1 through (l) Design 12. The GoLoco designs 1 through 9, panels (a) through (i), were designed to bind $G\alpha_{i1}$ with a single mutation, F108A. The GoLoco designs 10 through 12, panels (j) through (l), were designed to bind the wild type $G\alpha_{i1}$ protein.

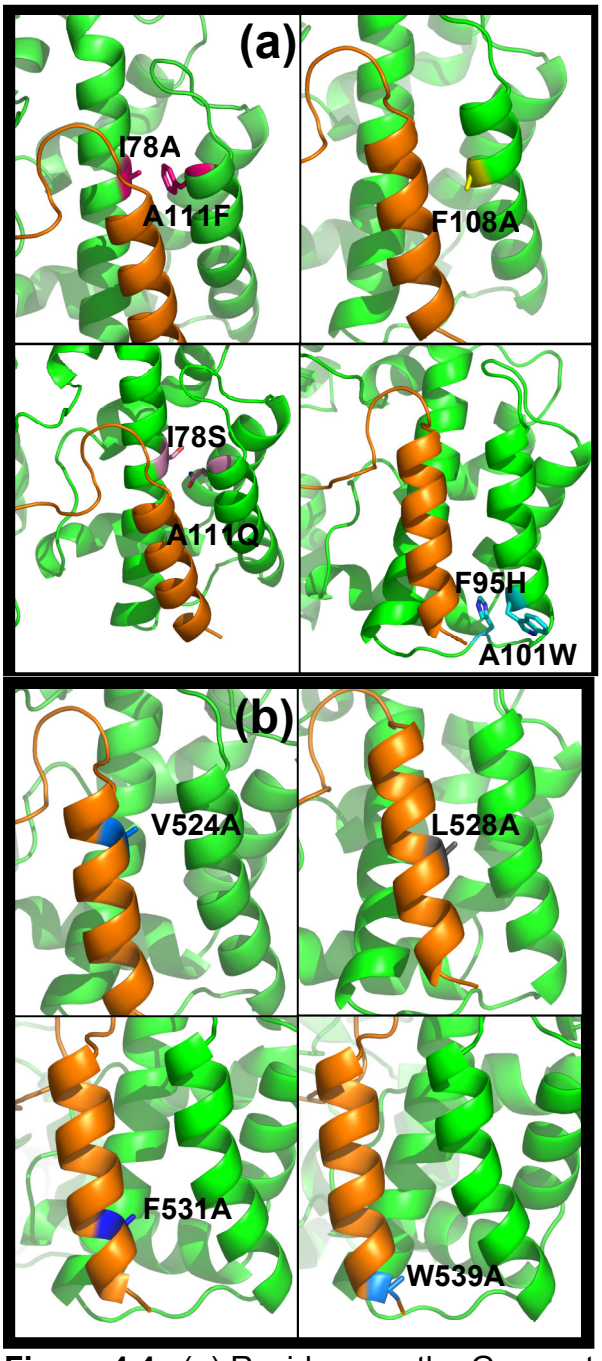
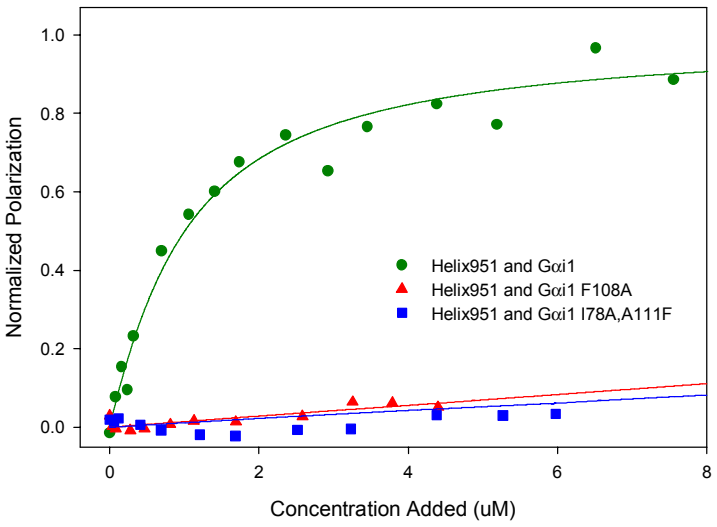


Figure 4.4. (a) Residues on the $G\alpha_{i1}$ protein at the $G\alpha_{i1}$-GoLoco interface that were mutated to test for binding of the GoLoco C-terminal helix, and (b) residues on the GoLoco at the interface that were mutated to test for binding.

(a) GoLoco_helix bound to Gαi1 wild-type and mutants



(b) Gαi1 bound to GoLoco_helix wild-type and mutants

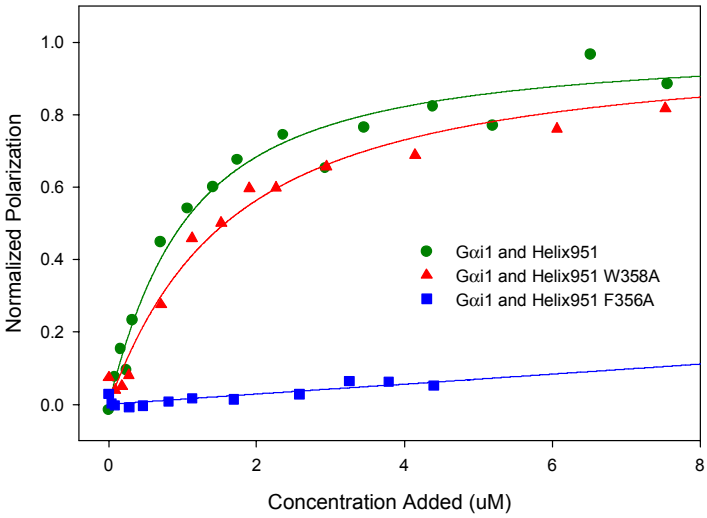
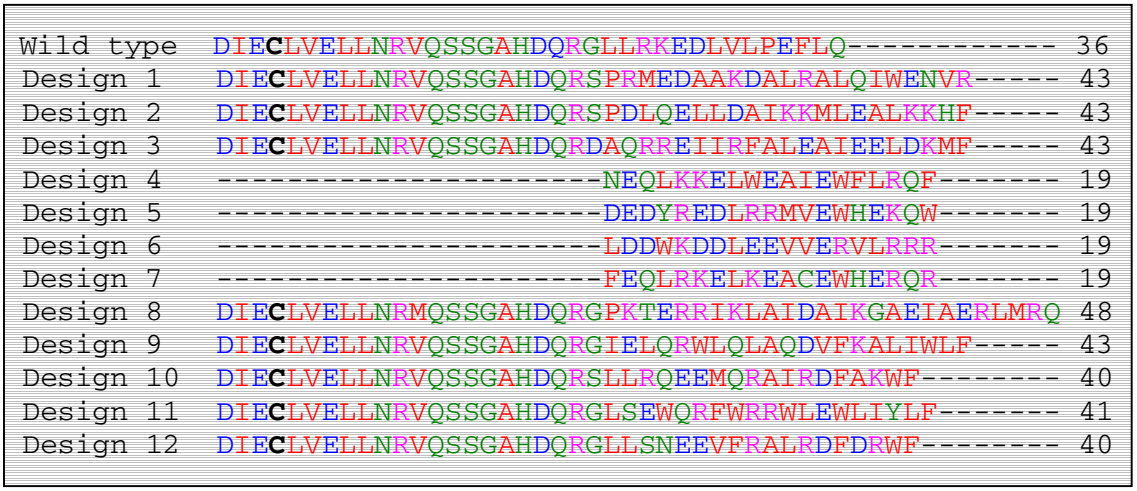


Figure 4.5. Comparing binding curves for (a) the redesigned GoLoco motif bound to wild type $G\alpha_{i1}$ and $G\alpha_{i1}$ mutations designed to destabilize the complex and (b) $G\alpha_{i1}$ wild type protein bound to the redesigned GoLoco motif and the redesigned GoLoco motif with mutations designed to destabilize the complex. .

Table 4.1. Binding affinities for the $G\alpha_{i1}$ protein bound to the redesigned GoLoco motif, GoLoco_Helix951 and point mutations on the $G\alpha_{i1}$ protein or the designed GoLoco helix at the redesigned interface.

Mutation	$\Delta\Delta G_{Rosetta}$	$\Delta\Delta G_{exp.}$	$K_{d\ exp.}$(uM)
GoLoco_Helix951:$G\alpha_{i1}$	—	—	0.8
$G\alpha_{i1}$:I78A,A111F	0.2	>1.9	>20
$G\alpha_{i1}$:F95H,A101W	3.9	NBD	NBD
$G\alpha_{i1}$:F108A	3.5	NBD	NBD
$G\alpha_{i1}$:I78S,A111Q	0.9	NBD	NBD
Helix951:V524A	1.2	0.1	1.0
Helix951:L528A	0.7	-0.1	0.7
Helix951:F531A	3.2	>1.9	>20
Helix951:W534A	4.1	0.2	1.1

NBD - no binding detected



Supplementary Figure 4.1. Amino acid sequences of the wild type RGS14 GoLoco aligned with Designs 1 through 12. Designs 4 through 7 were designed with just the C-terminal portion of the GoLoco motif. The unique Cystine residue located at sequence position 4 is a Glycine in the wild type GoLoco motif. The fluorophor is attached to this residue.

METHODS

Rosetta

All energy calculations and side chain and backbone relaxation simulations were performed with the molecular modeling program Rosetta[32].
We used the following command lines to run the backbone modeling protocol, called loop modeling in the Rosetta code:

For the centroid-mode backbone design we first created a fasta file for the GoLoco-$G\alpha_{i1}$ complex. We altered the sequence for the C-terminal region of the GoLoco that was to be redesigned, extending the sequence to allow for the conversion from a random coil to an alpha helix. We then altered the pdb file by, removing the atomic coordinates for the C-terminal region of the GoLoco that was to be redesigned, replacing the coordinates with zeros. We also added residues, extending the C-terminal region. The sequence in the fasta file had to match the sequence in the altered pdb file. We generated a set of alpha helical backbone fragments from a non-homologous set of pdb files from the Protein Data Bank.

rosetta.gcc aa g000 _ -s g000.pdb –loops –cst cst

where the fasta file is named g000_.fasta, the pdb file is named g000.pdb, the fragment files are named aag000_03_04.200_v1_3 and aag000_09_04.200_v1_3. The –cst cst command can be used for a constraint

file, containing constraints designed to direct the location of binding for the redesigned C-terminal helix. We did not use constraint files for the majority of the designs here, including the redesigned GoLoco discussed in the results section.

The resulting ensemble of designed structure are missing the coordinates for the side chain atoms. The original pdb file was then merged with the designed structures so that the side chain coordinates for all wild type residues could be recovered and not modeled. The redesigned sequence positions and neighboring sequence positions were relaxed using the Rosetta simulated annealing protocol, with the following command line:

rosetta.gcc –design –l list_of_pdb_structures –tail -begin 343 –end 368 –chain _ -series bb –protein g000 –resfile tail.resfile –ex1 –ex2 –extrachi_cutoff 1

where list_of_pdb_structure is a file containing the name of the pdb files generated using the centroid design mode and merged with the original pdb file. –begin 343 and –end 368 indicate the start and end of the region to be redesigned. The sequence positions here do not match those from the original pdb file because Rosetta renumbers all sequence positions, starting with 1 and continuing sequentially through both protein chains.

Construction and cloning of protein designs

The DNA sequence for the GoLoco motif of RGS14 (residues 496-531) was cloned into pET21b as a C-terminal fusion to the small protein Tenascin.

Tenascin was included to aid in the expression and purification of the peptide. The sequence for a hexahistidine tag was placed at the C-terminus of the construct. Residue G498 of the wild type GoLoco motif was mutated to a cysteine to enable the covalent labeling of the thiol-reactive fluorescent probe 6-iodoacetamidofluorescein (6-IAF) (Molecular Probes). We used an N-terminal-truncated, hexahistidine-tagged expression construct of human $G\alpha_{i1}$ with the first 25 codons of the Gα open reading frame removed, as previously described.[50] Point mutations were introduced using the QuickChange® site-directed mutagenesis protocol (Stratagene) and all vectors were verified by DNA sequencing.

Site-directed mutagenesis was used to remove a Sac1 cut site from the pET21b vector, leaving a unique Sac1 cut site in the GoLoco gene. The C-terminal portion of the GoLoco was removed by incubating the DNA at 37° C for 1 hour with Sac1 and Xho1 with the addition of BSA. The linearized plasmid was purified using gel electrophoresis and then the Qiagen gel extraction kit. Oligos for the designs 1 through 7 were ordered from Lineberger Tissue Culture Facility. The oligoes were annealed and ligated into the plasmid over night at 4° C.

Protein purification

As described in chapter 1, GoLoco motif peptide was expressed either for 4 hours at 37 ˚C or overnight at 25 ˚C with 0.5 mM IPTG in the BL21(DE3) strain of E. coli. $G\alpha_{i1}$ was expressed overnight at 25 ˚C with 1 mM IPTG in the BL21(DE3) strain of E. coli. Cells were lysed using an Avestin emulsiflex and the resulting lysates were cleared by ultracentrifugation. The $G\alpha_{i1}$ ΔN2 and RGS14-

GoLoco motif-Tenascin fusion proteins were purified using a HiTrap (Amersham Biosciences) column by eluting the protein with an imidazole step gradient, then followed by gel filtration with a Superdex-200 column (Amersham Biosciences). Proteins were concentrated using Vivaspin 20® centrifugal concentrators. Protein concentrations were determined by measuring absorbance at 280nm. Extinction coefficients were calculated using the method of Gill and von Hipple[88].

Fluorescence polarization binding analysis

A thiol-reactive fluorescent probe 6-iodoacetamidofluorescein (6-IAF) (Molecular Probes) was conjugated to the unique cysteine on the GoLoco motif using the manufacture's recommended protocol. We do not consider any designs that include the addition of a cysteine to the GoLoco motif to prevent multiple fluorescent labeling. GoLoco motif protein was buffer exchanged into 50 mM Tris-Cl pH 7.5 using a PD10 desalting column, concentrated to ~50 to 100 uM. Next 1 mM TCEP was added to the PD10 eluate and stirred for one hour at room temperature. A 20 mM stock solution of 6-IAF suspended in dimethyl sulfoxide (DMSO) was diluted into the GoLoco motif protein solution to a 10-fold molar excess and the conjugation reaction was allowed to proceed overnight, in the dark at 4 °C. Precipitate was pelleted and discarded, and 5 mM β-mercaptoethanol (β-ME) was added to quench the reaction. The supernatant containing fluorescein-GoLoco motif protein was run over a PD10 column to separate free probe from labeled protein. The concentration of fluorescein-GoLoco motif protein was quantified using UV/Vis, taking readings at 280 and

495 nm for the protein and fluorophore respectively. The Tenascin-GoLoco fusion protein was used in the binding assays.

Fluorescence polarization assays were carried out on a Jobin Yvon Horiba Spec FluoroLog-3 instrument (Jobin Yvon Inc.) performed in L-format with the excitation wavelength set at 495 nm and the emission wavelength set at 520 nm. Titrations were performed using a 3 x 3-mm quartz cuvette with a starting volume of 200 µL. Fluorescein labeled wild type or mutant GoLoco motif protein was diluted to 50 to 150 nM and the excitation and emission slit widths adjusted to give a fluorescence intensity >100,000 counts per second. Wild type or mutant $G\alpha_{i1}$ was added in increasing volumes from a stock solution whose initial concentration depended on the strength of the interaction, generally having a concentration of 3-10 µM. Two to three polarization readings consisting of 3 averaged measurements were collected for increasing concentrations of $G\alpha_{i1}$. Data was averaged and analyzed using a model for single site binding according, as described previously[26].

REFERENCES

1. Palmer, A. E., Giacomello, M., Kortemme, T., Hires, S. A., Lev-Ram, V., Baker, D. & Tsien, R. Y. (2006). Ca2+ indicators based on computationally redesigned calmodulin-peptide pairs. *Chem Biol* **13**, 521-30.

2. Palmer, A. E., Jin, C., Reed, J. C. & Tsien, R. Y. (2004). Bcl-2-mediated alterations in endoplasmic reticulum Ca2+ analyzed with an improved genetically encoded fluorescent sensor. *Proc Natl Acad Sci U S A* **101**, 17404-9.

3. Kortemme, T., Morozov, A. V. & Baker, D. (2003). An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J Mol Biol* **326**, 1239-59.

4. Ippolito, J. A., Alexander, R. S. & Christianson, D. W. (1990). Hydrogen bond stereochemistry in protein structure and function. *J Mol Biol* **215**, 457-71.

5. Stickle, D. F., Presta, L. G., Dill, K. A. & Rose, G. D. (1992). Hydrogen bonding in globular proteins. *J Mol Biol* **226**, 1143-59.

6. Fabiola, F., Bertram, R., Korostelev, A. & Chapman, M. S. (2002). An improved hydrogen bond potential: impact on medium resolution protein structures. *Protein Sci* **11**, 1415-23.

7. Pace, C. N. (1992). Contribution of the hydrophobic effect to globular protein stability. *J Mol Biol* **226**, 29-35.

8. Takano, K., Yamagata, Y., Fujii, S. & Yutani, K. (1997). Contribution of the hydrophobic effect to the stability of human lysozyme: calorimetric studies and X-ray structural analyses of the nine valine to alanine mutants. *Biochemistry* **36**, 688-98.

9. Schueler-Furman, O., Wang, C., Bradley, P., Misura, K. & Baker, D. (2005). Progress in modeling of protein structures and interactions. *Science* **310**, 638-42.

10. Kortemme, T. & Baker, D. (2004). Computational design of protein-protein interactions. *Curr Opin Chem Biol* **8**, 91-7.

11. Davis, I. W., Arendall, W. B., 3rd, Richardson, D. C. & Richardson, J. S. (2006). The backrub motion: how protein backbone shrugs when a sidechain dances. *Structure* **14**, 265-74.

12. Desjarlais, J. R. & Handel, T. M. (1999). Side-chain and backbone flexibility in protein core design. *J Mol Biol* **290**, 305-18.

13. Huang, P. S., Love, J. J. & Mayo, S. L. (2007). A de novo designed protein protein interface. *Protein Sci* **16**, 2770-4.

14. Havranek, J. J. & Harbury, P. B. (2003). Automated design of specificity in molecular recognition. *Nat Struct Biol* **10**, 45-52.

15. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L. & Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364-8.

16. Rohl, C. A., Strauss, C. E., Misura, K. M. & Baker, D. (2004). Protein structure prediction using Rosetta. *Methods Enzymol* **383**, 66-93.

17. Bradley, P., Chivian, D., Meiler, J., Misura, K. M., Rohl, C. A., Schief, W. R., Wedemeyer, W. J., Schueler-Furman, O., Murphy, P., Schonbrun, J., Strauss, C. E. & Baker, D. (2003). Rosetta predictions in CASP5: successes, failures, and prospects for complete automation. *Proteins* **53 Suppl 6**, 457-68.

18. Kuhlman, B. & Baker, D. (2000). Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A* **97**, 10383-8.

19. Kortemme, T., Kim, D. E. & Baker, D. (2004). Computational alanine scanning of protein-protein interfaces. *Sci STKE* **2004**, pl2.

20. Joachimiak, L. A., Kortemme, T., Stoddard, B. L. & Baker, D. (2006). Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. *J Mol Biol* **361**, 195-208.

21. Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L. & Baker, D. (2004). Computational redesign of protein-protein interaction specificity. *Nat Struct Mol Biol* **11**, 371-9.

22. Sammond, D. W., Eletr, Z. M., Purbeck, C., Kimple, R. J., Siderovski, D. P. & Kuhlman, B. (2007). Structure-based protocol for identifying mutations that enhance protein-protein binding affinities. *J Mol Biol* **371**, 1392-404.

23. Kuhlman, B., O'Neill, J. W., Kim, D. E., Zhang, K. Y. & Baker, D. (2002). Accurate computer-based design of a new backbone conformation in the second turn of protein L. *J Mol Biol* **315**, 471-7.

24. Sood, V. D. & Baker, D. (2006). Recapitulation and design of protein binding peptide structures and sequences. *J Mol Biol* **357**, 917-27.

25. Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D. & Zardecki, C. (2002). The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* **58**, 899-907.

26. Kimple, R. J., Kimple, M. E., Betts, L., Sondek, J. & Siderovski, D. P. (2002). Structural determinants for GoLoco-induced inhibition of nucleotide release by Galpha subunits. *Nature* **416**, 878-81.

27. Willard, F. S., Kimple, R. J. & Siderovski, D. P. (2004). Return of the GDI: the GoLoco motif in cell division. *Annu Rev Biochem* **73**, 925-51.

28. McCudden, C. R., Willard, F. S., Kimple, R. J., Johnston, C. A., Hains, M. D., Jones, M. B. & Siderovski, D. P. (2005). G alpha selectivity and inhibitor function of the multiple GoLoco motif protein GPSM2/LGN. *Biochim Biophys Acta* **1745**, 254-64.

29. Adhikari, A. & Sprang, S. R. (2003). Thermodynamic characterization of the binding of activator of G protein signaling 3 (AGS3) and peptides derived from AGS3 with G alpha i1. *J Biol Chem* **278**, 51825-32.

30. Peterson, Y. K., Bernard, M. L., Ma, H., Hazard, S., 3rd, Graber, S. G. & Lanier, S. M. (2000). Stabilization of the GDP-bound conformation of Gialpha by a peptide derived from the G-protein regulatory motif of AGS3. *J Biol Chem* **275**, 33193-6.

31. Takesono, A., Cismowski, M. J., Ribas, C., Bernard, M., Chung, P., Hazard, S., 3rd, Duzic, E. & Lanier, S. M. (1999). Receptor-independent activators of heterotrimeric G-protein signaling pathways. *J Biol Chem* **274**, 33202-5.

32. Raw, A. S., Coleman, D. E., Gilman, A. G. & Sprang, S. R. (1997). Structural and biochemical characterization of the GTPgammaS-, GDP.Pi-, and GDP-bound forms of a GTPase-deficient Gly42 --> Val mutant of Gialpha1. *Biochemistry* **36**, 15660-9.

33. Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* **182**, 319-26.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

We developed a protocol to predict point mutations to enhance protein-protein binding affinity, evaluated a protocol to redesign protein-peptide binding specificity and developed a protocol to redesign the backbone and sequence of a peptide in the context of a wild type binding partner. We used the protein design software, Rosetta, for these three applications. The model system that we used to validate these protocol is the $G\alpha_{i1}$ protein from the heterotrimeric G-protein complex bound to the GoLoco regulatory domain from RGS14. We also use the E2-E3 complex, E6AP bound to UbcH7 for the protocol to predict affinity enhancing point mutations. We found that our protocol to model single point mutations that will increase buried hydrophobic surface area is an efficient and reliable way to enhance protein-protein binding affinity. This protocol required modeling of the fewest degrees of freedom and relied on our ability to model the hydrophobic effect, which we understand and can model well. Rosetta's energy function exhibited good predictive power in this context.

We then evaluated an automated protocol to redesign protein-peptide binding specificity. This application requires the protocol to select sequences to stabilize desired protein-peptide interactions while destabilizing undesired interactions and thus is a more stringent evaluation of our protein design protocol. The goal was to design an orthogonal binding interface, with a redesigned protein-peptide complex that has a binding affinity similar to that of the wild type complex but that will not interact with the wild type protein components. We did have some success using this protocol. Four designs passed our set of filters, and two of the four designs have the intended binding

specificity. The redesigned specificity for these two designs is more modest than desired, however. Perhaps more importantly is the fact that our protocol generated more than 1500 designs, yet only four passed our set of filters. We relaxed our filters and selected additional designs, but none were able to recover wild type binding affinity. It is possible that fixing the backbone torsion angles to those found in the x-ray crystal structure limited the sequence space that this protocol was able to search. Expanding this protocol to include flexibility in the backbone torsion angles could improve the quality of the designed hydrogen bonds or redesigned hydrophobic cores. This could increase the number of designs generated that pass our initial set of filters.

Lastly, we examined our ability to design a new peptide backbone in the context of a wild type binding partner. Our approach redesigned the C-terminal region of the 36-amino acid GoLoco domain, allowing the N-terminal region to direct binding to the wild type partner, $G\alpha_{i1}$. We sought to change the secondary structure of the GoLoco C-terminal region from a random coil to an alpha helix. We confirmed that the new C-terminal GoLoco helix does bind the $G\alpha_{i1}$ as designed using site-directed mutagenesis. Solving the crystal structure of the redesigned interface is desired to determine how close the redesigned GoLoco-$G\alpha_{i1}$ complex is to the computationally generated structure. Ultimately the goal is to design a peptide or protein in the context of a binding partner without relying on an initial wild type region to direct and enhance binding. Further evaluation of the redesigned GoLoco by comparison with an experimentally determined

structure is needed to determine how successful this protocol is with this more conservative design approach.