FOR THE RECORD

Structure of a protein G helix variant suggests the importance of helix propensity and helix dipole interactions in protein design

PAVEL STROP,¹ ANDREI M. MARINESCU,² AND STEPHEN L. MAYO³

¹Biochemistry Option, California Institute of Technology, Mail Code 147-75, Pasadena, California 91125

²Division of Biology, California Institute of Technology, Mail Code 147-75, Pasadena, California 91125

³Howard Hughes Medical Institute and Division of Biology, California Institute of Technology, Mail Code 147-75, Pasadena, California 91125

(RECEIVED March 29, 2000; FINAL REVISION May 3, 2000; ACCEPTED May 11, 2000)

Abstract: Six helix surface positions of protein G (*Gβ1*) were redesigned using a computational protein design algorithm, resulting in the five fold mutant *Gβ1m2*. *Gβ1m2* is well folded with a circular dichroism spectrum nearly identical to that of *Gβ1*, and a melting temperature of 91 °C, ~6 °C higher than that of *Gβ1*. The crystal structure of *Gβ1m2* was solved to 2.0 Å resolution by molecular replacement. The absence of hydrogen bond or salt bridge interactions between the designed residues in *Gβ1m2* suggests that the increased stability of *Gβ1m2* is due to increased helix propensity and more favorable helix dipole interactions.

Keywords: helix dipole interaction; helix propensity; protein design; protein G

While many surface sites in proteins can tolerate a wide variety of side chains, surface residues can still have a significant effect on protein structure and stability (Reidhaar-Olson & Sauer, 1988; Bowie et al., 1990; O'Neil & DeGrado, 1990; Zhang et al., 1991; Minor & Kim, 1994; Smith et al., 1994; Zhao & Arnold, 1999). Previous computational protein design efforts on the surface of GCN4, a homodimeric coiled coil, suggest that helix propensity is the key factor in sequence design for surface helical positions (Dahiyat et al., 1997a). The designed GCN4 sequence was randomly scrambled, thereby disrupting the designed hydrogen bond and salt-bridge interactions, but not changing the overall helix propensity. Both the designed and the scrambled sequences were 12 °C more stable than the wild-type sequence. The increased stability was not due to the designed interactions between the surface residues, but apparently due to increased helical propensity. Results from this GCN4 study suggest that for helix surface design, hydrogen bonding and electrostatic interactions are not as impor-

tant as helix propensity and helix capping (O'Neil & DeGrado, 1990; Blaber et al., 1993; Chakrabarty et al., 1993, 1994; Huyghues-Despointes et al., 1993; Doig et al., 1994).

Protein G (*Gβ1*) consists of a four stranded β -sheet traversed by an α -helix. Protein G is a monomeric globular 56 residue protein that contains no disulfide bonds or cofactors and lacks the repetitive side chain patterns found in coiled coils and helical bundles. These properties make protein G an ideal model system for testing the importance of specific helical surface design in globular proteins.

Six residues on the helix surface of protein G were designed using a computational design algorithm called ORBIT (Optimization of Rotamers By Iterative Techniques) and yielded a variant called *Gβ1m2*. The structure of *Gβ1m2* was determined by X-ray crystallography to assess the performance of the design algorithm and to examine the origins of helix surface stabilization.

Results and discussion: *Gβ1m2* contains the following five mutations: A24E, K28R, Q32E, N35K, and D36K (position 22 retained its wild-type identity). *Gβ1m2* is well folded with a far-ultraviolet circular dichroism (CD) spectrum nearly identical to the wild-type protein G at 1 °C; thermal denaturation of this variant yields a melting temperature (T_m) of 91 °C, ~6 °C higher than the wild-type protein (data not shown).

The backbone of *Gβ1m2*, as determined from X-ray crystallographic analysis (Table 1), superimposes onto the *Gβ1* backbone with a root-mean-square deviation (RMSD) of 0.92 Å (Fig. 1A). The backbone of the helix alone (residues 22–37) superimposes onto the helix of *Gβ1* with an RMSD of 0.17 Å. Residue D22 is a helix capping position. In the *Gβ1m2* structure, D22 makes hydrogen bonds to both the backbone and side chain of T25, as in the wild-type structure (Fig. 1B). In the predicted structure of the designed sequence, R28 makes a putative salt bridge to E24. Poor electron density for R28 suggests either low occupancy or the absence of this salt bridge in the *Gβ1m2* structure. Side chains of residues E32 and K35 are mostly disordered and do not appear to form the predicted E32–K35 interaction. Residues E24 and K36

Reprint requests to: Stephen L. Mayo, Howard Hughes Medical Institute, Division of Biology, California Institute of Technology, Mail Code 147-75, Pasadena, California 91125; e-mail: steve@mayo.caltech.edu.

Fig. 1. Stereoview of the superposition of the Gβ1 and Gβ1m2 X-ray structures. **A:** Structural comparison of the Gβ1 (blue) and Gβ1m2 (red) backbones. **B:** Superimposition of the redesigned helix surface residues. Wild-type residues shown in yellow and Gβ1m2 residues shown in green. The side chains of residues R28, E32, and K35 in Gβ1m2 are mostly disordered. Figures were generated using BOBSCRIPT and RASTER3D (Esnouf, 1997; Merritt & Bacon, 1997).

are well ordered and make hydrogen bonds through water molecules to symmetry related molecules and not to their designed partners (R28 and E32, respectively).

These results suggest that the stabilization of the Gβ1m2 mutant is not due to the increased number of hydrogen bonds or salt bridges on the helix surface, as predicted by the design algorithm. However, the amino acids at four of the five helix surface positions in the designed sequence have higher helical propensity than the wild-type sequence. The mutation to a more negatively charged residue at the N-terminus of the helix (A24E), as well as mutations to more positively charged residues at the C-terminus of the helix (N35K and D36K), could also result in more favorable interactions with the helix dipole. Electrostatic interactions with helix dipoles have been shown to be favorable and stabilizing in proteins (Nicholson et al., 1988, 1991; Lockhart & Kim, 1992, 1993;

Huyghues-Despointes et al., 1993; Creamer & Rose, 1994; Doig et al., 1994; Walter et al., 1995). The increased stability of Gβ1m2 relative to Gβ1, together with the absence of hydrogen bonds or salt bridges between the designed residues, suggests that the stabilization is due to increased helix propensity and more favorable helix dipole interactions.

Materials and methods: *Computational design:* In the calculation, protein residues were classified into core, boundary, and surface as described (Dahiyat & Mayo, 1997a). Six residues were classified as helix surface residues (22, 24, 28, 32, 35, and 36) and were restricted to the following ten amino acids: Ala, Arg, Asn, Asp, Gln, Glu, His, Lys, Ser, Thr. The resulting combinatorial complexity is 10^6 possible amino acid sequences. Protonated his-

Table 1. Refinement and model statistics

Space group	P4 ₂ 1 ₂
Cell dimensions <i>a</i> , <i>c</i> (Å)	52.61, 44.59
Resolution (Å)	18.0–2.0
Measured reflections	81,792
Unique reflections	4,020
R_{merge}^a	4.1 (9.1)
Completeness (%)	95.0 (98.0)
Number of protein atoms	439
Number of water molecules	29
<i>B</i> -factor values (Å ²)	
Overall	29.8
Protein atoms	29.6
Solvent	38.6
R_{cryst}^b	0.235
R_{free}^c	0.267
RMSD	
Bonds	0.016
Angles	1.74
Ramachandran	
Most favored regions	96.0
Additional allowed regions	4.0
Generously allowed regions	0.0
Disallowed regions	0.0

^a $R_{\text{merge}} = \sum |I_{\text{obs}} - \langle I \rangle| / \sum \langle I \rangle$ summed over all observations and reflections.

^b $R_{\text{cryst}} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum |F_{\text{obs}}|$.

^c $R_{\text{free}} = R_{\text{cryst}}$ calculated for 5% of reflections omitted from the refinement.

tidine was used in the calculation. Both amino acid identities and side-chain conformations were determined by the design algorithm for the six optimized residues. The remainder of the structure was used as a template for the sequence selection calculation. Computational details, potential functions, and parameters for van der Waals interactions, hydrogen bonding, and electrostatic interactions are described in our previous work (Dahiyat & Mayo, 1996, 1997a, 1997b; Dahiyat et al., 1997a, 1997b).

Expression and CD analysis: Variant Gβ1m2 was constructed by inverse polymerase chain reaction (Hemsley et al., 1989) in plasmid pet11c and verified by sequencing. Recombinant protein was expressed by IPTG induction in BL21(DE3) cells (Invitrogen, Carlsbad, California) as described (Alexander et al., 1992) and isolated using a freeze/thaw method (Johnson & Hecht, 1994). Purification was accomplished by reversed-phase high-performance liquid chromatography using an acetonitrile/water gradient containing 0.1% trifluoroacetic acid. Molecular weight was verified by mass spectrometry. CD data were collected on an Aviv 62DS spectrometer equipped with a thermoelectric unit. Protein samples were studied in 50 mM sodium phosphate buffer at pH 5.0. Wavelength scans were performed at 1 °C. Thermal denaturations were monitored by CD at 222 nm and fitted to a two-state model as described (Minor & Kim, 1994).

Crystallization, data collection, and refinement: Gβ1m2 crystals were grown by the sitting drop method at 4 °C using a precipitant solution containing 100 mM sodium acetate, pH 4.6, 30% poly-

ethylene glycol monomethyl ether 2000, and 200 mM ammonium sulfate. The crystals belong to the space group P4₂1₂ with unit cell dimensions *a* = *b* = 52.61 Å, *c* = 44.59 Å. Crystals were transferred stepwise to mother solution containing 20% glycerol and flash frozen. Diffraction data were collected on a R-axis IV image plate detector with CuK_α radiation. All data were processed with DENZO and SCALEPACK (Otwinowski & Minor, 1997). Wild-type protein G, Protein Data Bank (PDB) code 1PGA (Gallagher et al., 1994), was used for molecular replacement with the CCP4 program suit (Bailey, 1994) yielding an initial *R*-factor of 38.3% at 3 Å. Alternate cycles of manual model building using the program O (Jones et al., 1991), positional and individual *B*-factor refinement with the program CNS (Brünger, 1992), and addition of water molecules to the model reduced the *R* and *R*-free to 23.5 and 26.7, respectively, for all of the reflections (2.0 Å resolution). A total of 96% of the residues in Gβ1m2 are in the most favored regions of the Ramachandran plot, 4% in the additional allowed regions as calculated with PROCHECK (Laskowski et al., 1993). The coordinates for the structure have been deposited in the Protein Data Bank (PDB code 1EM7).

Acknowledgments: This work was supported by the Howard Hughes Medical Institute (S.L.M.) and a NSF fellowship (P.S.).

References

- Alexander P, Fahnestock S, Lee T, Orban J, Bryan P. 1992. Thermodynamic analysis of the folding of the streptococcal protein G IgG-binding domains β1 and β2—Why small proteins tend to have high denaturation temperatures. *Biochemistry* 31:3597–3603.
- Bailey S. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D50*:760–763.
- Blaber M, Zhang XJ, Matthews BW. 1993. Structural basis of amino acid α helix propensity. *Science* 260:1637–1640.
- Bowie JU, Reidhaar-Olson JF, Lim WA, Sauer RT. 1990. Deciphering the message in protein sequences: Tolerance to amino acid substitutions. *Science* 247:1306–1310.
- Brünger AT. 1992. *X-PLOR, version 3.1: A system for X-ray crystallography and NMR*. New Haven, Connecticut: Yale University Press.
- Chakrabarty A, Doig AJ, Baldwin RL. 1993. Helix capping propensities in peptides parallel those in proteins. *Proc Natl Acad Sci USA* 90:11332–11336.
- Chakrabarty A, Kortemme T, Baldwin RL. 1994. Helix propensities of the amino-acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci* 3:843–852.
- Creamer TP, Rose GD. 1994. Alpha-helix-forming propensities in peptides and proteins. *Proteins* 19:85–97.
- Dahiyat BI, Gordon DB, Mayo SL. 1997a. Automated design of the surface positions of protein helices. *Protein Sci* 6:1333–1337.
- Dahiyat BI, Mayo SL. 1996. Protein design automation. *Protein Sci* 5:895–903.
- Dahiyat BI, Mayo SL. 1997a. De novo protein design: Fully automated sequence selection. *Science* 278:82–87.
- Dahiyat BI, Mayo SL. 1997b. Probing the role of packing specificity in protein design. *Proc Natl Acad Sci USA* 94:10172–10177.
- Dahiyat BI, Sarisky CA, Mayo SL. 1997b. De novo protein design: Towards fully automated sequence selection. *J Mol Biol* 273:789–796.
- Doig AJ, Chakrabarty A, Klingler TM, Baldwin RL. 1994. Determination of free energies of N-capping in alpha-helices by modification of the Lifson-Roig helix-coil theory to include N- and C-capping. *Biochemistry* 33:3396–3403.
- Esnouf RM. 1997. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J Mol Graph* 15:133–138.
- Gallagher T, Alexander P, Bryan P, Gilliland GL. 1994. 2 crystal-structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry* 33:4721–4729.
- Hemsley A, Arnheim N, Toney MD, Cortopassi G, Galas DJ. 1989. A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acid Res* 17:6545–6551.
- Huyghues-Despointes BMP, Scholtz JM, Baldwin R. 1993. Effect of a single aspartate on helix stability at different positions in a neutral alanine-based peptide. *Protein Sci* 2:16604–16611.

- Johnson BH, Hecht MH. 1994. Recombinant proteins can be isolated from *E. coli* cells by repeated cycles of freezing and thawing. *Biotechnology* 12:1357–1360.
- Jones TA, Zou JY, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron: Density maps and the location of errors in these models. *Acta Crystallogr A* 47:110–119.
- Laskowski RA, Macarthur MW, Moss DS, Thornton JM. 1993. Procheck: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283.
- Lockhart DJ, Kim PS. 1992. Internal Stark effect measurement of the electric field at the amino terminus of an alpha helix. *Science* 257:947–951.
- Lockhart DJ, Kim PS. 1993. Electrostatic screening of charge and dipole interactions with the helix backbone. *Science* 260:198–202.
- Merritt EA, Bacon DJ. 1997. Raster3D photorealistic molecular graphics. *Methods Enzymol* 277:505–524.
- Minor DL, Kim PS. 1994. Measurement of the β -sheet-forming propensities of amino acids. *Nature* 367:660–663.
- Nicholson H, Anderson DE, Dao-pin S, Matthews BW. 1991. Analysis of the interaction between charged side chains and the alpha-helix dipole using designed thermostable mutants of phage T4 lysozyme. *Biochemistry* 30:9816–9828.
- Nicholson H, Becktel WJ, Matthews BW. 1988. Enhanced protein thermostability from designed mutations that interact with alpha-helix dipoles. *Nature* 336:651–656.
- O'Neil KT, DeGrado WF. 1990. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* 250:646–651.
- Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
- Reidhaar-Olson JF, Sauer RT. 1988. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science* 241:53–57.
- Smith CK, Withka JM, Regan L. 1994. A thermodynamic scale for the β -sheet forming tendencies of the amino acids. *Biochemistry* 33:5510–5517.
- Walter S, Hubner B, Hahn U, Schmid FX. 1995. Destabilization of a protein helix by electrostatic interactions. *J Mol Biol* 252:133–143.
- Zhang XJ, Baase WA, Matthews BW. 1991. Toward a simplification of the protein folding problem: A stabilizing polyalanine α helix engineered in T4 lysozyme. *Biochemistry* 30:2012–2017.
- Zhao H, Arnold FH. 1999. Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Eng* 12:47–53.