# Subdomain interactions foster the design of two protein pairs with ~80% sequence identity but different folds

## *Supporting information*

Lauren L. Porter[1,2,*] Yanan He[1], Yihong Chen[1], John Orban[1,3], and Philip N. Bryan[1,2]
[1]University of Maryland, Institute for Bioscience and Biotechnology Research, Rockville, MD 20850, USA [2]Potomac Affinity Proteins, Rockville, MD 20850, USA [3]University of Maryland, Department of Chemistry and Biochemistry, College Park, MD 20742, USA
[*]Corresponding author: llporter@umd.edu

**Table S1A**. High-identity protein variants.

| Variant | Sequence[1,2] |
|---|---|
| $G_A$ (PSD-1) | MEAVDANSLAQAKEAAIKELKQYGIGDYYIKLINNAKTVEGVESLKNEILKALPTE |
| $^{54}GA_{MBP}$ | NGDKDANSLAEAKEAAIKELKQYGIGEYYIKLIENAKTVEGVESLKDEILKALPRF |
| $^{63}GA_{MBP}$ | NGDKDANSLAEAKEKAIKELKIYGIGEHYIKLIENAKQVEAVESLKDEILKALPRF |
| $^{64}GA_{MBP}$ | NGDKDANSLAEAKEKAIKELKIYGIGEHYIKLIENAKQVAAVESLKDEILKALPRF |
| $^{68}GA_{MBP}$ | NGDKDANSLAEAKEKAIKDLKIYGIGEHYIKLIENAKQVAAVEDLKDEILKALPRF |
| $^{70}GA_{MBP}$ | NGDKDANSLAEAKEKAIKDLKIYGIGEHYIKLIEKAKQVAAVEDLKDEILKALPRF |
| $^{75}GA_{MBP}$ | NGDKGANSLAEAKEKAIKDLKIYGIGEHYIKLIEKAKQVAAVEDLKDEILKAHDRF |
| $^{77}GA_{MBP}$ | NGDKGYNSLAEAKEKAIKDLKIYGIGEHYIKLIEKAKQVAAVEDLKDEILKAHDRF |
| $^{79}GA_{MBP}$ | NGDKGYNGLAEAKEKAIKDLKIYGIGEHYIKLIEKAKQVAAVEDLKDEILKAHDRF |
| $^{80}GA_{MBP}$ | NGDKGYNGLAEAKEKAIKDLKIYGIGEHYIKLIEKAKQVAAVEDLKDIILKAHDRF |
| MBP | NGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRF |
| $^{71}MBP_{GA}$ | NGDKGYNGLAEVGEKFEKDTGIKVIVEHYIKLEEKFKQVAAVGDGPDIIFWAHDRF |
| $^{75}MBP_{GA}$ | NGDKGYNGLAEVGEKFEKDTGIKVIVEHYIKLEEKFKQVAAVEDGPDIILWAHDRF |
| $^{79}MBP_{GA}$ | NGDKGYNGLAEVGEKAEKDLGIKVIVEHYIKLEEKFKQVAAVEDGPDIILWAHDRF |
| $^{80}MBP_{GA}$ | NGDKGYNGLAEAGEKAEKDLGIKVIVEHYIKLEEKFKQVAAVEDGPDIILWAHDRF |
| $^{59}GA_{OspA}$ | LDEKNANSLAQAKEMAIKELKEYGIDDYYIKLINNAKTVEGVESLKNNILKALEGV |
| $^{66}GA_{OspA}$ | LDEKNSVSLDQAKEMAIKELKEYGIDDYYIKLINNAKTVEGVESLKNNILKVLEGV |
| $^{70}GA_{OspA}$ | LDEKNSVSLDQAKEMAIKELKEYGIDDKYIKLITNAKTVEGVESLKNNILGVLEGV |
| $^{77}GA_{OspA}$ | LDEKNSVSLDQAKEMAIKELKEYGIDDKYILLITVAKTVLKVESLKNNILGVLEGV |
| OspA | LDEKNSVSVDLPGEMKVLVSKEKNKDGKYDLIATVDKLELKGTSDKNNGSGVLEGV |
| $^{68}GA_{OspA}$ | LDEKNSVSVDLPGEMKIKVSKEYGKDDKYILIATVDKLELKGTSDKNNGSGVLEGV |
| $^{73}GA_{OspA}$ | LDEKNSVSVDLPGEMKIKVSKEYGKDDKYILIATVDKTVLKGESDKNNGSGVLEGV |
| $^{77}GA_{OspA}$ | LDEKNSVSVDLPGEMAIKVSKEYGIDDKYILIATVDKTVLKGESDKNNGSGVLEGV |

[1]Colored backgrounds indicate new mutations. Font colors are changed thereafter.

[2]Colors correspond to classifications in Table S1B, except yellow backgrounds, which indicate residues identical to the highest-identity folded $G_A$ variant

**Table S1B**. Anticipated destabilizing effects of mutations

| Minor | Moderate | Significant |
|---|---|---|
| Irregularly structured N- and C-terminal residues with that do not form hydrophobic contacts or hydrogen bonds | Hydrophobic/hydrophilic surface residues ↔ other hydrophobic/hydrophilic residues (e.g. I↔V↔T, L↔N↔D, L↔M↔K, L→A, V→A, etc.) | Glycine populating disallowed regions of the alanyl Ramachandran plot→ anything else |
| Charged and polar surface residues whose side chains do not hydrogen bond to the backbone → residues with similar steric effects and electrostatic properties (e.g. N↔D, Q↔E, K↔R, D↔E, N↔Q, S↔T) | Helical residue →high moderate helix propensity (e.g. A → L) | Residues with ϕ–ψ angles outside of proline's sterically-allowed boundaries →proline |
| Attractive surface mutation: changing a residue that had repulsive electrostatic interactions with neighboring residues to one that no longer causes repulsive interactions | Strand residue→high moderate strand propensity (e.g. V→T) | Helical residue →residue much lower helix propensity (e.g. A → G) |
| | Surface hydrophobic to surface hydrophilic (e.g. L->K) | Strand residue→residue much lower strand propensity (e.g. V→P) |
| | Change in loop residue | Repulsive surface mutation from accumulating multiple residues with the same charge in close proximity on the protein surface |
| | | Any changes to hydrophobic core residues |
| | | Hydrophobic surface residue likely to cause aggregation |

**Table S2**: Summary of structure statistics for $^{79}GA_{MBP}$

| | |
|---|---|
| Experimental restraints | |
| CS-Rosetta input | |
| $^{13}C^{\alpha}$ shifts | 56 |
| $^{13}C^{\beta}$ shifts | 51 |
| $^{13}C'$ shifts | 55 |
| $^{15}N$ shifts | 55 |
| $^{1}H^{N}$ shifts | 55 |
| $^{1}H^{\alpha}$ shifts | 52 |
| RMSDs to the mean structure (Å) | |
| Over residues 8-54 | |
| Backbone atoms | 0.54 ± 0.16 |
| Heavy atoms | 1.07 ± 0.24 |
| Secondary structures[a] | |
| Backbone atoms | 0.46 ± 0.15 |
| Heavy atoms | 0.88 ± 0.22 |
| Measures of structure quality | |
| Ramachandran distribution | |
| Most favored regions (%) | 98.6 ± 1.6 |
| Additionally allowed regions (%) | 1.4 ± 1.6 |
| Generously allowed regions (%) | 0.0 ± 0.0 |
| Disallowed regions (%) | 0.0 ± 0.0 |
| Bad contacts/100 residues | 0.4 ± 0.5 |
| Overall dihedral *G* factor | 0.55 ± 0.03 |

[a] The secondary structure elements used were residues 8-23, 27-35 and 39-52.

## Table S3

| Variant Name[1] | Experimentally-measured $T_M$ (°C) | Experimentally-derived $\Delta\Delta G$[2,4] (kcal/mol) | Predicted $\Delta\Delta G$[3,4] (kcal/mol) |
|---|---|---|---|
| $^{63}GA_{MBP}$ | 80 | 0.29 | 0.55 |
| $^{64}GA_{MBP}$ | 76 | 0.23 | 0.18 |
| $^{68}GA_{MBP}$ | 62 | 0.86 | 1.43 |
| $^{70}GA_{MBP}$ | 63 | -0.06 | 0.00 |
| $^{79}GA_{MBP}$ | 55 | 0.50 | 0.68 |
| $^{80}GA_{MBP}$ | 53 | -0.06 | 0.09 |
| | | | |
| $^{71}MBP_{GA}$ | 58 | 3.6 | 3.17 |
| $^{75}MBP_{GA}$ | 56 | 1.42 | 3.41 |
| $^{79}MBP_{GA}$ | 56 | 0 | 3.65 |
| $^{80}MBP_{GA}$ | 55 | 0.72 | 3.43 |

[1]Variant names correspond to constructs in Figure 4. Mutations with disordered ends were excluded

[2]Determined by multiplying $\Delta T_M$ by $\Delta S$. $\Delta S$ values were determined from calorimentric measurements in [1] for $G_A$ and reference [2] for MBP.

[3]Values determined from PoPMuSiC 2.1 (reference 36 in the main text).

[4]Values in red denote a significant deviation from predicted $\Delta\Delta G$. We defined significant > 1 kcal/mol difference between experimental and predicted values.

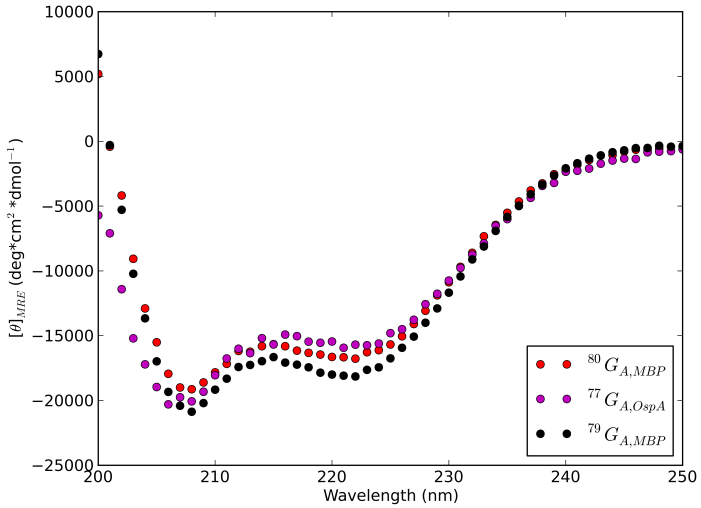


**Figure S1A.** CD Spectra of all 3 $G_A$ variants ($^{79}GA_{MBP}$, black; $^{80}GA_{MBP}$, red; $^{77}GA_{OspA}$, purple) are consistent with one another. $^{79}GA_{MBP}$ is known to adopt a 3-α–helix bundle structure (see main text). All spectra were measured in 100 mM potassium phosphate buffer, pH 6.8.

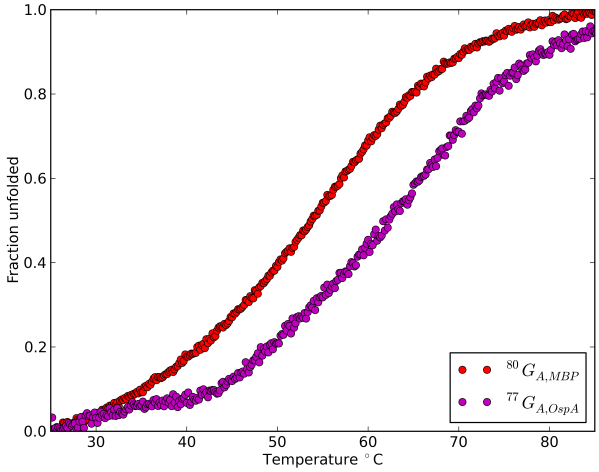


**Figure S1B.** Both $G_A$ variants ($^{80}GA_{MBP}$, red; $^{77}GA_{OspA}$, purple) unfold cooperatively in response to heat. All measurements were taken in 100 mM potassium phosphate buffer, pH 6.8.

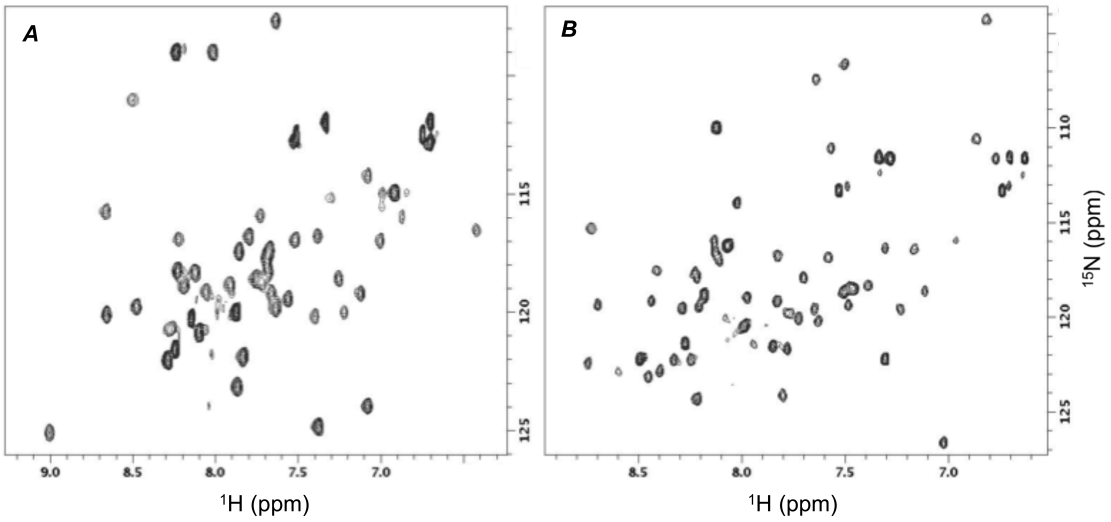


**Fig. S2**: The 2D $^{15}$N HSQC spectra for A) $^{80}GA_{MBP}$ and B) $^{77}GA_{OspA}$ indicate well-folded proteins. Both were acquired at 10°C in 100 mM potassium phosphate buffer, pH 7.0.

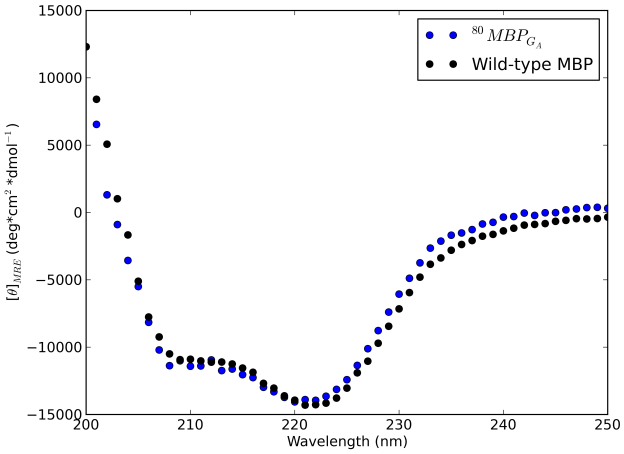


**Figure S3A.** The CD Spectrum of $^{80}MBP_{GA}$ (blue) is consistent with that of wild-type MBP (black). Both spectra were measured in 100 mM potassium phosphate buffer, pH 6.8.

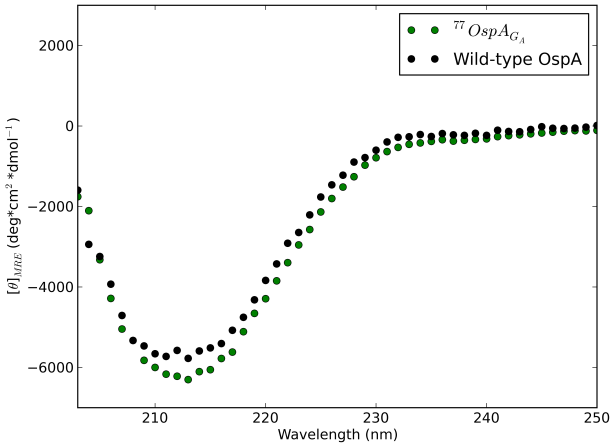


**Figure S3B.** The CD Spectrum of $^{77}OspA_{GA}$ (green) is consistent with that of wild-type OspA (black). Both spectra were measured in 100 mM potassium phosphate buffer, pH 6.8.

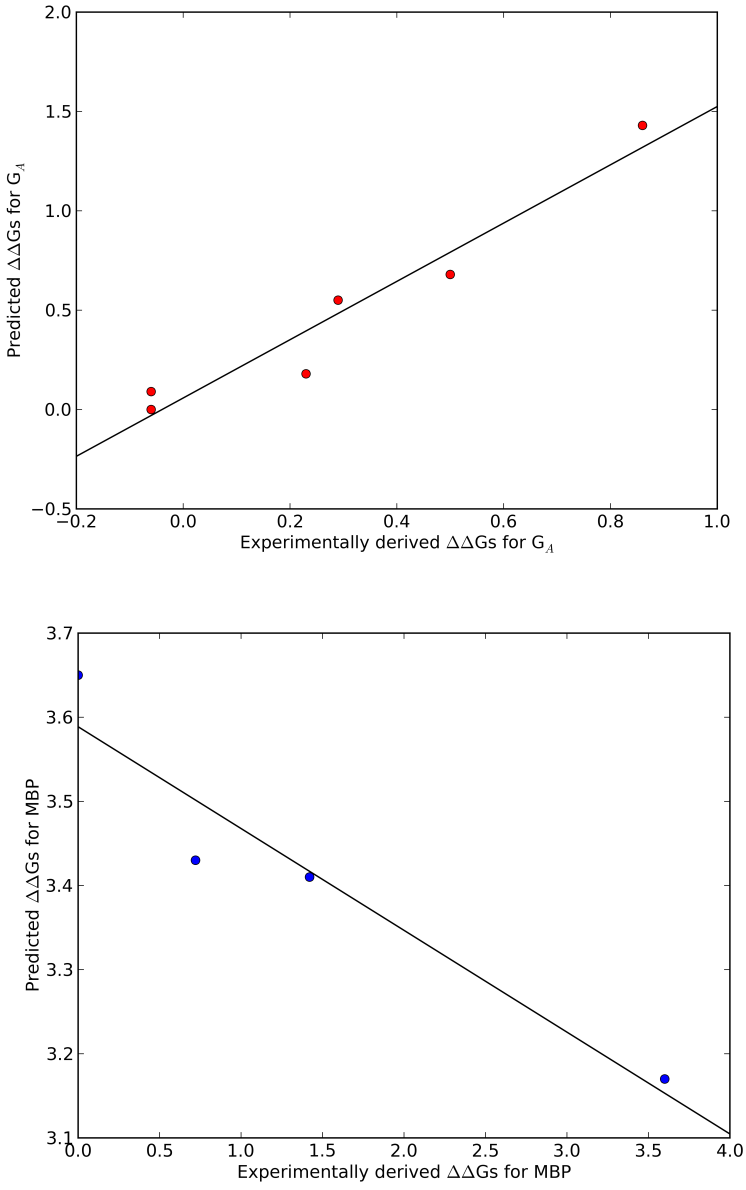


**Fig. S4**: Experimentally-measured ΔΔGs correspond well to predicted values (positive correlation, $R^2$=0.94) for $G_A$ (red) but not for MBP (blue, negative correlation, $R^2$=0.98). This suggests that interactions with neighboring subdomains compensated for mutations that would have likely destabilized the MBP destination fold in isolation. Lines of best fit are shown in black for both plots.

**SUPPORTING REFERENCES**

[1] He, Y., Rozak, D.A., Sari, N., Chen, Y., Bryan, P.N., and J. Orban, 2006. Structure, Dynamics, and Stability Variation in Bacterial Albumin Binding Modules: Implications for Species Specificity. *Biochemistry* 45:10102-10109

[2] Ganesh, C., A.N. Shah, C.P. Swiminathan, A. Surolia, and R. Varadarajan, 1997. Thermodynamic characterization of the reversible, two-state unfolding of maltose binding protein, a large two-domain protein. *Biochemistry* 36:5020-5028.