Supporting Information
for

Systematic enhancement of protein crystallization efficiency by bulk lysine-to-arginine (KR) substitution

Nooriel E. Banayan [1†] | Blaine J. Loughlin [1†] | Shikha Singh [1] | Farhad Forouhar [1] | Guanqi Lu [1] | Kam-Ho Wong [1§] | Matthew Neky [1] | Henry S. Hunt [2] | Larry B. Bateman Jr. [3] | Angel Tamez [3] | Samuel K. Handelman [1§] | W. Nicholson Price [1§] | John F. Hunt [1*]

[1] Department of Biological Sciences, 702A Sherman Fairchild Center, MC2434, Columbia University, New York, NY, 10027, USA;

[2] Physics Department, Stanford University, Stanford, CA 94305, USA; and

[3] Accendero Software, P.O. Box 2826, Idaho Falls, ID 83404, USA.

[†] These authors contributed equally to the work reported in this paper.

[§] Current addresses: SKH, 307 E. Merrill St, Indianapolis, IN 46225; MN, Mailman School of Public Health, Columbia University, 722 W. 168th Street, New York, NY 10032; WNP, University of Michigan Law School, 625 S. State St, Ann Arbor, MI 48109; KHW, 401 N Middletown Rd, Pearl River, NY 10965.

Running Title: **Bulk K-to-R substitution enhances protein crystallization**

Supporting Information Table S1.

Table S1. ***Data collection & refinement statistics for Bulk-KR crystal structures.***

Protein	**hPDIa-9KR**	**MA_2137-D65R-11KR**
Crystal parameters		
Space group	$P2_12_12_1$	$P4_22_12$
Cell dimensions *a*, *b*, *c* (Å)	58.7, 61.3, 68.6	109.9, 109.9, 39.1
Cell angles α, β, γ (°)	90, 90, 90	90, 90, 90
Data Collection		
Resolution (Å)	68.6-1.89 (1.93-1.89)*	110-1.91 (1.95-1.92)*
R_{merge} (%)	0.24 (2.15)	0.135 (2.828)
I/σI	7 (0.9)	20.4 (0.9)
Completeness (%)	98.8 (82.2)	99.8 (97.7)
Redundancy	12.8 (10.4)	25.4 (17.3)
CC1/2	0.99 (0.46)	1.00 (0.49)
Model contents		
Protein residues	A20-137, B18-137	A1-140,153-194
# protein atoms	1,897	1,464
Ligands	2 SCN, 1 EDO	1 SAH
# ligand atoms	10	26
# waters	144	82
Refinement		
Resolution (Å)	45.70-2.05 (2.12-2.05)*	77.75-2.01 (2.07-2.01)*
No. reflections	16,013 (1,574)	16,541 (1,642)
R_{work} / R_{free} (%)	19.5 (28.8) / 23.2 (34.1)	18.2 (20.9) / 22.1 (24.4)
Ramachandran Plot (%)		
Outliers	0.00	0.00
Allowed	2.99	2.25
Favored	97.01	97.75
B-factors		
Protein	34.6	43.2
Ligand/ion	37.9	31.8
Water	39.5	53.3
r.m.s deviations		
Bond lengths (Å)	0.007	0.007
Bond angles (º)	0.841	0.935
PDB ID	**8GDY**	**8GDU**

* Values in parentheses represent statistics for the highest resolution shell.

Supporting Information Figure S1.

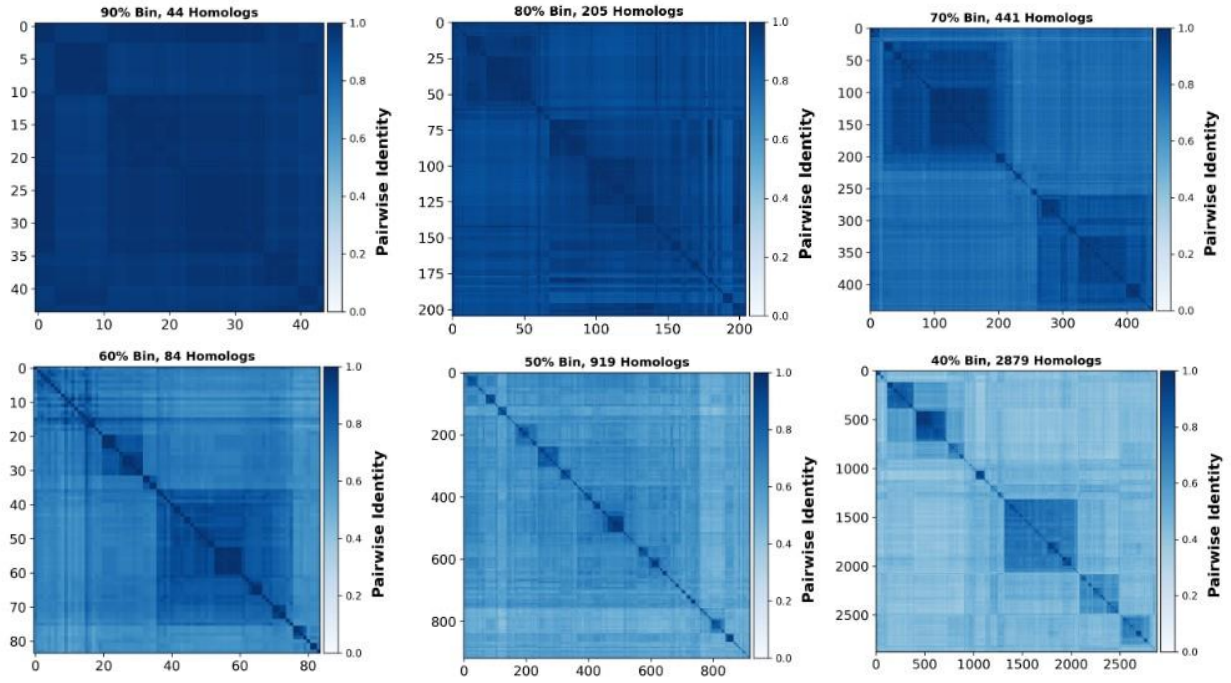


Figure S1. ***Pairwise percent sequence identity distributions between homologs in mutually exclusive bins spanning 10% ranges of sequence identity to target protein hPDIa.***

Supporting Information Figure S2.

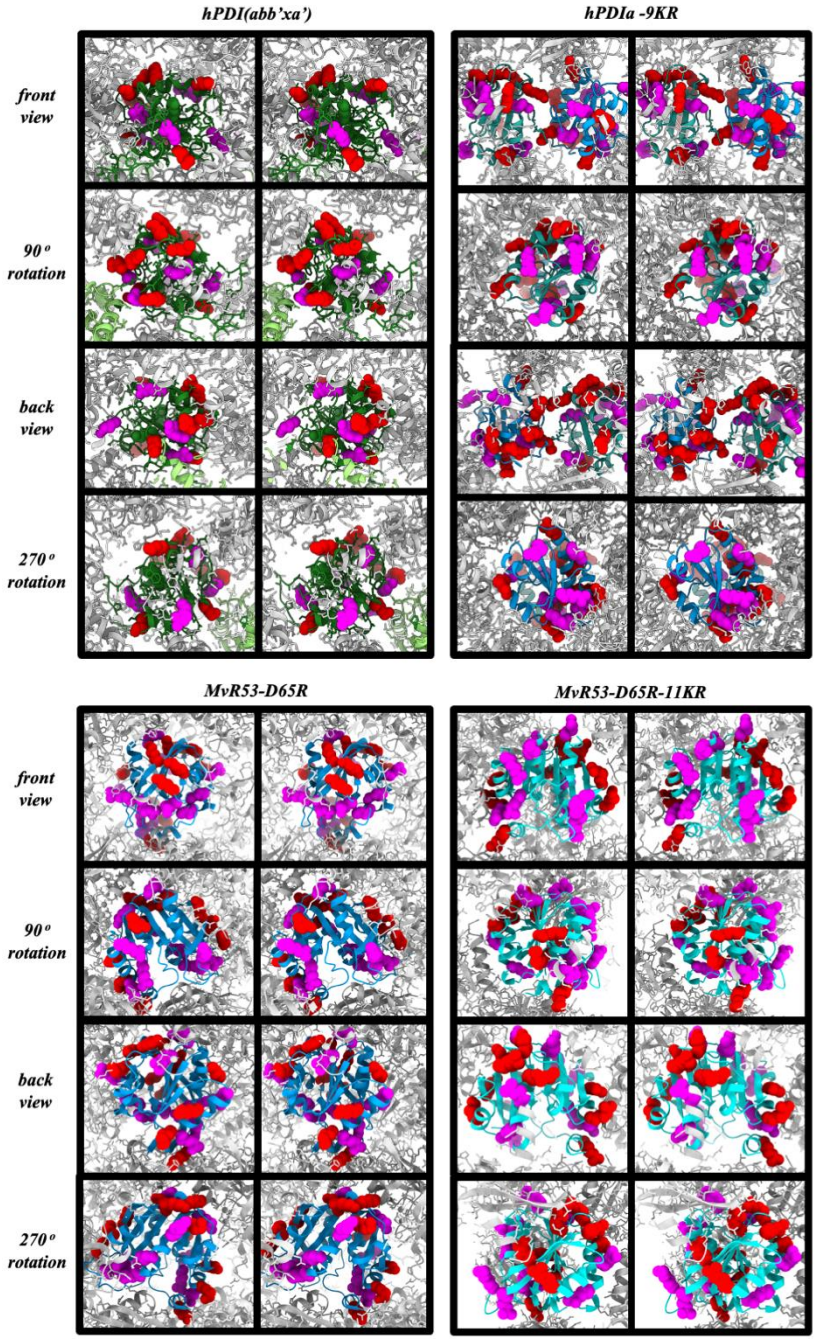
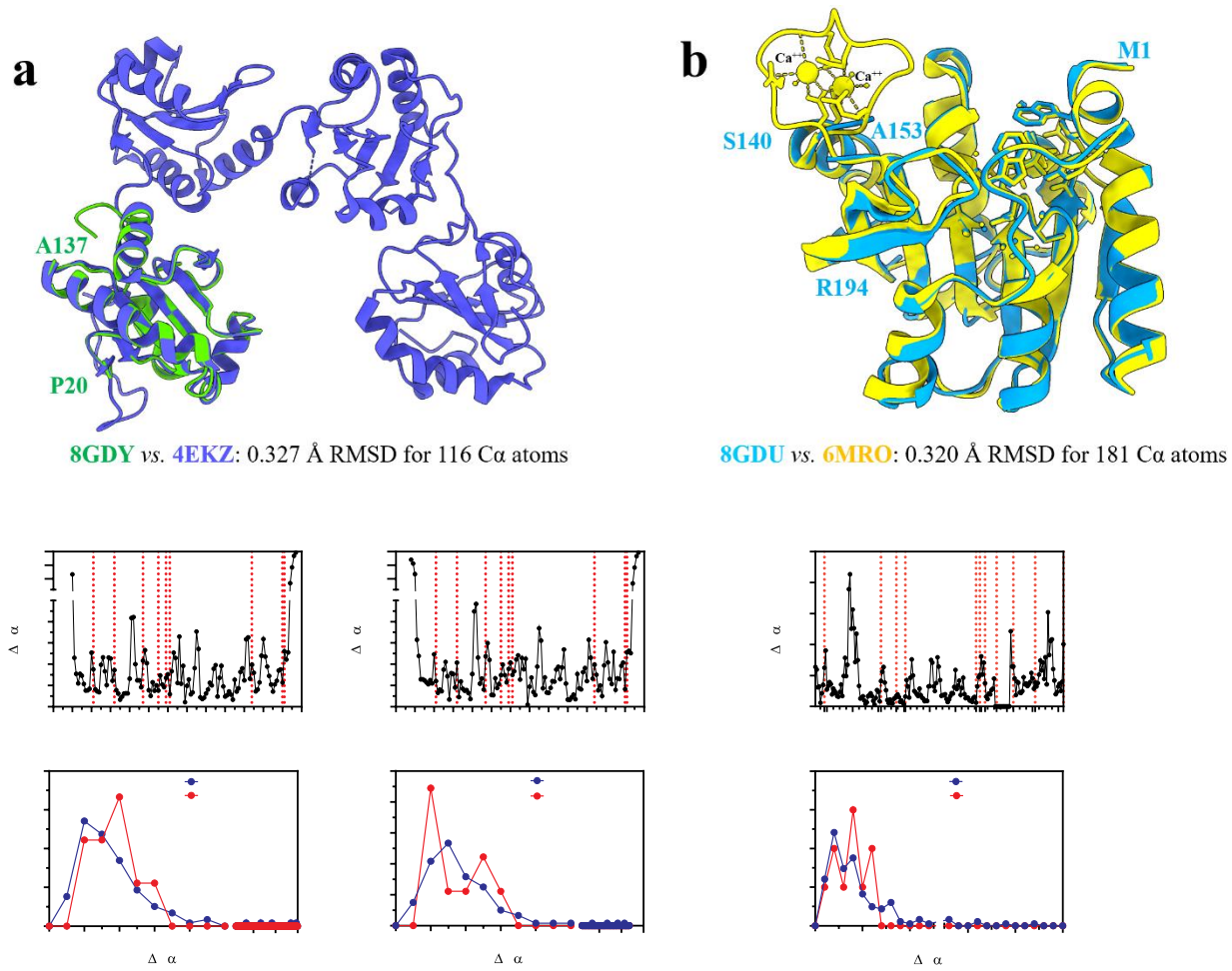


FIGURE S2. ***Stereo views of crystal-packing interactions in structures from the parental proteins and protein constructs containing bulk-KR mutations.*** The protein backbone is shown in ribbon representation, colored gray for symmetry mates and shades of green, blue, cyan, and teal for the subunits modeled in the asymmetric unit of each crystal structure. The sidechains of the native arginines (magenta) and residues mutated from lysine-to-arginine (red) are shown in space-filling representation. Note that the red residues in the parental hPDI(abb'xa') and MA_2137-D65R structures on the left are the native lysine residues, while the red residues in the hPDIa-9KR and MA_2137-D65R-11KR structures on the right are the mutated arginine residues.

Supporting Information Figure S3.



e **Fraction of Ca atoms within increasing cutoff distances.**

		Cutoff distance			
Protein construct	**Chain**	**0.5 Å**	**1.0 Å**	**2.0 Å**	**4.0 Å**
hPDIa-9KR	A	0.88	0.96	0.97	0.97
hPDIa-9KR	B	0.84	0.94	0.95	0.95
MA_2137-D65R-11KR	A	0.82	0.96	1.00	1.00

Figure S3. ***Crystal structures of Bulk-KR-substituted protein constructs show no significant conformational or stereochemical changes compared to reference structures.*** **(a)** Comparison of our crystal structure of hPDIa-9KR (**Fig. 5b** and **Table S1**) to an earlier structure of the same domain as part of a much larger multi-domain hPDI(abb'xa') construct (PDB ID 4EKZ) [87]. **(b)** Comparison of our crystal structures of MA_2137-D65R-11KR (**Fig. 5d** and **Table S1**) and MA_2137-D65R (PDB ID 6MRO). The loop at residues 141-152 is ordered by a high concentration of Ca^{++} present in the mother liquor for the latter structure, while the was no Ca^{++}

present in the mother liquor for former structure. Least-squares alignments were performed in PyMOL (www.pymol.org) and ChimeraX [136] using the default algorithms. **(c-e)** More detailed analyses of the structural differences in the Bulk KR constructs compared to the parental domains were performed using the Global Distance Test (GDT) [137]. The top graphs in panels **c-d** show the Cα displacements between all shared residues in the alignments used for the test, with the vertical dotted red lines indicating the sites of the KR mutations. The bottom graphs in these panels show the corresponding histograms of the Cα displacements for all residues (blue) and the engineered arginine residues (red). Panel **e** shows the fraction of Cα atoms within the indicated cutoff distances as calculated by the High Accuracy GDT algorithm.

Supporting Information Figure S4.

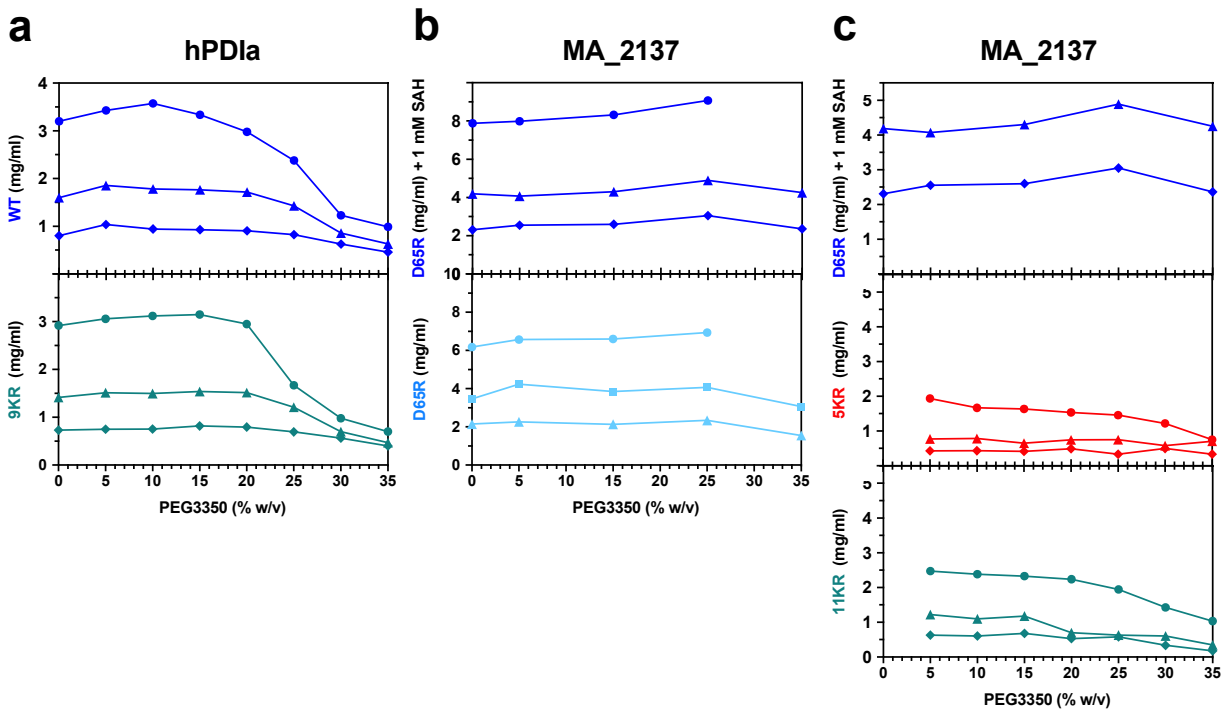


Figure S4. ***Solubility assays on selected hPDIa and MA_2137 constructs.*** Assays were conducted at room temperature. The lower graph in the center column shows data collected without S-adenosylhomocysteine (SAH) in the buffer, which is a product of the methyltransferase reaction catalyzed by MA_2137, while all of the other solubility assays on this protein were conducted in the presence of 1 mM SAH. The top graphs in both MA_2137 columns show the same data, which are duplicated but matched in scale to the graphs beneath them in order to provide a visual reference.

Supporting Information Figure S5.

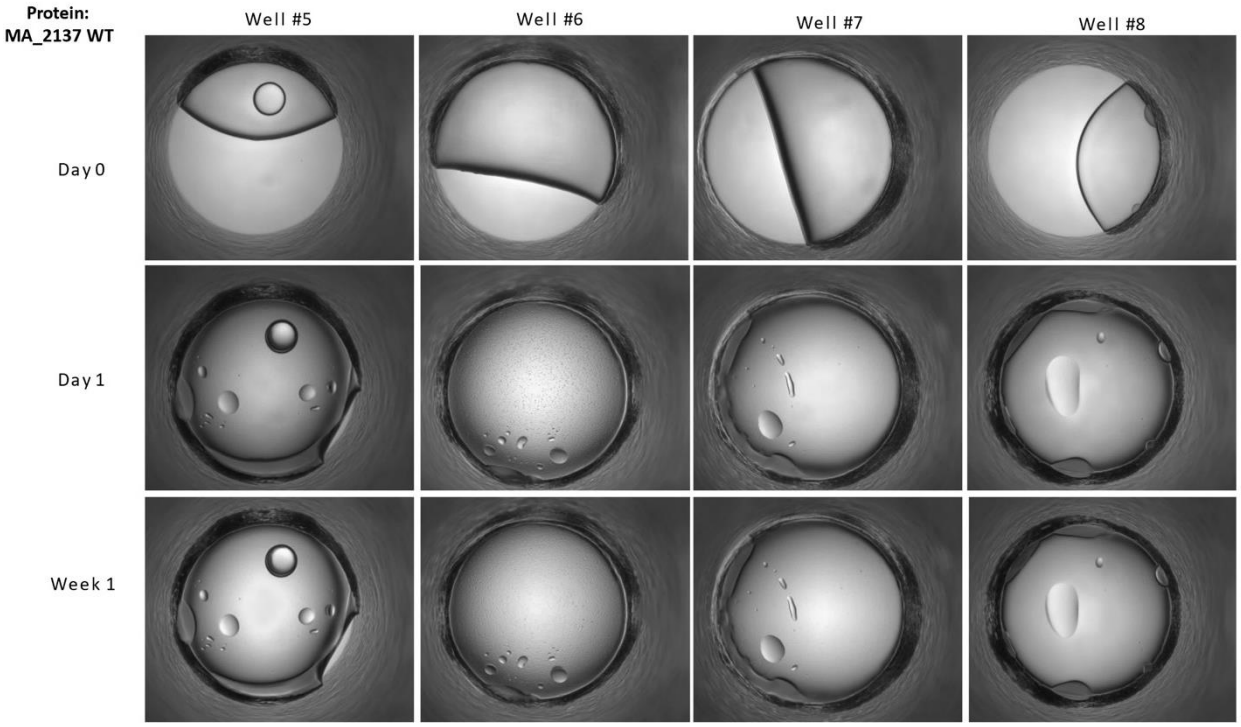


Figure S5. ***Optical micrographs showing time evolution of representative MA_2137-D65R crystallization screening reactions showing evidence of liquid-liquid phase separation.*** Representative micrographs from the HWI generation #19 1,536-well crystal screen (https://hwi.buffalo.edu/crystallization-cocktails/) show evidence of liquid-liquid phase separation for MA_2137-D65R. Other than a short engineered C-terminal affinity tag and a 12-residue internal loop, likely to be a calcium binding site (**Fig. S3b**), the backone atoms of all residues in this protein are well-ordered based on our crystal structures. The cocktail recipes are as follows for the individual wells:

#5: 0.1 M $Na_2S_2O_3$, 0.1 M Tris, pH 8, 20% (w/v) PEG 4000;

#6: (HR SaltRx Screen HT A12) 3.2 M NaCl, 0.1 M NaOAc, pH 4.6;

#7: (HR SaltRx Screen HT D4) 6.0 M NH_4NO_3, 0.1M NaOAc, pH 4.6;

#8: 0.1 M LiBr, 0.1 M Tris, pH 8, 20% (w/v) PEG 4000.

Supporting Information Figure S6.

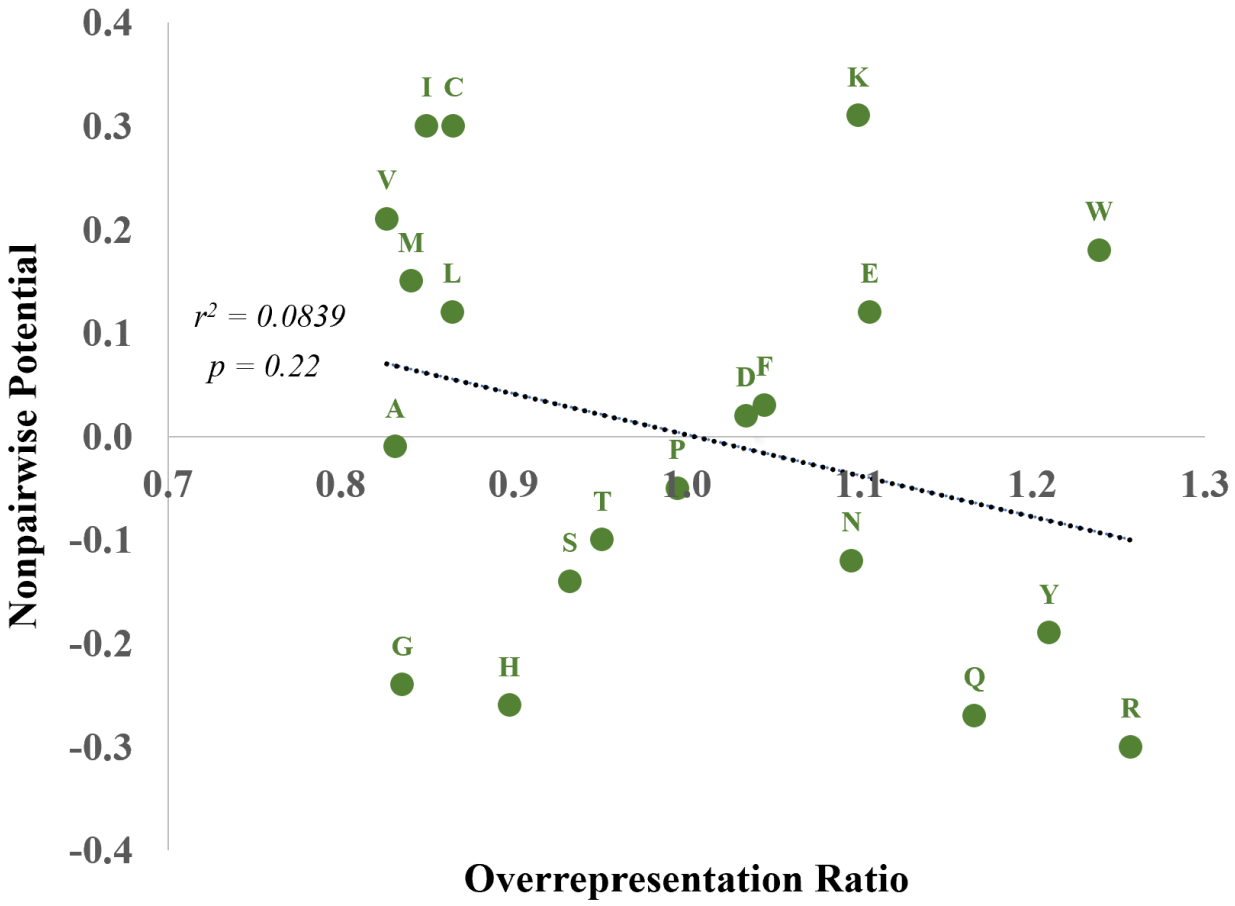


Figure S6. ***Comparison of overrepresentation ratios of amino acids in crystal-packing interfaces in our analysis of 87,684 crystal structures [37] (Fig. 1) to an earlier analysis of contact potentials in crystal-packing interfaces in 233 crystal structures [113].***