Supporting material for:

Experimental and Computational Analysis of Protein Stabilization by Gly-to-D-Ala substitution: A Convolution of Native State and Unfolded State Effects

Junjie Zou, Benben Song, Carlos Simmerling, Daniel Raleigh

Methods

Protein Solid Phase Synthesis

The proteins and their Gly-to-D-Ala variants were chemically synthesized using Fmoc chemistry [1]. Sequences of these proteins are provided below. EH, GA and PSBD have a free N-terminus and amidated C-terminus, while HP35 has a free N-terminus and free C-terminus. Peptide identity was confirmed using MALDI or ESI and purity was greater than 95%. EH, observed mass 7453.97, expected mass 7453.52; EH D-Ala, observed mass 7467.75, expected mass 7467.55; GA D-Ala, observed mass 5143.96, expected mass 5143.91; HP35, observed mass 4065.16, expected mass 4064.13; HP35 D-Ala, observed mass 4079.32, expected mass 4078.15. PSBD, observed 4400.72, expected 4402.10.

Sequences of the Proteins Synthesized for This Study

dA refers to D-Ala and L_N refers to nor-leucine.

EH: MDEKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFQNKRAKIKKS

EH-G39D-Ala: MDEKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELdALNEAQIKIWFQN KRAKIKKS

GA: LKNAIEDAIAELKKAGITSDFYFNAINKAKTVEEVNALVNEILKAHA

GA-G16D-Ala: LKNAKEDAIAELKKAdAITSDFYFNAINKAKTVEEVNALVNEILKAHA

HP35: LSDEDFKAVFGMTRSAFANLPLWL_NQQHLKKEKGLF

HP35-G11D-Ala: LSDEDFKAVFdAMTRSAFANLPLWL_NQQHLKKEKGLF

PSBD: AMPSVRKYAREKGVDIRLVQGTGKNGRVLKEDIDAFLAGGA

PSBD-G15D-Ala: AMPSVRKYAREKdAVDIRLVQGTGKNGRVLKEDIDAFLAGGA

Backbone phi/psi Angles and Calculation of the Solvent Accessibility of the Gly Backbone

The φ/ψ angles of C-capping glycines were calculated by using VMD [2]. The same PDB structures used for molecular dynamics simulations were used and missing hydrogen atoms were added using tLeap in Amber [3]. The solvent accessible surface area (SASA) of C-capping glycines was calculated by using VMD with a water probe radii of 1.4 Å. The extended tetrapeptides were constructed using tLeap with the same local sequence as the respective full length proteins. The C-termini of the tetrapeptides were amidated and the N-termini were acetylated. Residues in the extended peptides all have φ and ψ angles equal to 180°. Fractional SASA is defined as the ratio between the SASA found for the PDB structure and the SASA found for the extended tetrapeptide.

Table S1. Backbone phi/psi and solvent accessibility of Gly

Protein	φ (°)	ψ (°)	SASA (Å^2)	SASA in extended tetrapeptide (Å^2)	Fractional SASA (%)
EH	51.8	35.8	64.0	88.5	72.4
GA	107.8	-21.7	55.5	120.8	45.9
HP35	75.7	19.8	66.9	73.1	91.5
NTL9	70.4	26.9	36.7	98.9	37.1
PSBD	84.0	48.1	63.5	94.0	67.6
Trp-cage	119.9	10.0	31.6	113.0	28.0
UBA	127.0	1.3	64.2	95.9	67.0
Ubiquitin	81.2	5.2	53.7	100.5	53.4

Thermal and Urea/Guanidine Denaturation

The unfolding free energy of each protein was measured by CD-monitored urea/guanidine hydrochloride denaturation at 222nm under the conditions listed in **Table S2**. Thermal denaturation experiments were also conducted at 222nm using the same buffer and pH employed for the urea/guanidine hydrochloride denaturation experiments. The concentration of urea/guanidine was determined by measuring the refractive index on a refractometer. Urea/guanidine denaturation experiments were carried out with a titrator unit interfaced to the CD spectrometer. Unfolding curves for EH, GA, PSBD were recorded using Aviv model 62A DS and 202SF circular dichroism spectrophotometers. Unfolding curves for HP35 were recorded using an Applied Photophysics Chirascan instrument. ΔG^o of unfolding was determined by fitting the urea/guanidine denaturation curves to the following equation:

$$\theta[denaturant] = \frac{(a_n + b_n[denaturant]) + (a_d + b_d[denaturant])e^{-\left(\frac{\Delta G^o([denaturant])}{RT}\right)}}{1 + e^{-\left(\frac{\Delta G^o([denaturant])}{RT}\right)}} \quad (1)$$

$$\Delta G^o([denaturant]) = \Delta G^o(H_2O) - m[denaturant] \quad (2)$$

where θ is the measured ellipticity, a_n ,b_n ,a_d ,b_d are the parameters that define the signals of the native state and denatured state. $\Delta G^o([denaturant])$ is the free energy change upon unfolding as a function of denaturant and $\Delta G^o(H_2O)$ is the free energy change in the absence of denaturant. Thermal unfolding data was fit using standard methods and the Gibbs-Helmholtz equation to obtain the melting temperature T_m and ΔH^0 at T_m.

$$\theta[T] = \frac{(a_n + b_n T) + (a_d + b_d T)e^{-\left(\frac{\Delta G^o(T)}{RT}\right)}}{1 + e^{-\left(\frac{\Delta G^o(T)}{RT}\right)}} \quad (3)$$

$$\Delta G^o(T) = \Delta H^o(T_m)\left(1 - \frac{T}{T_m}\right) - \Delta C^o{}_p[T_m - T + Tln(\frac{T}{T_m})]$$

(4)

Where T_m is the melting temperature. $\Delta H^o(T_m)$ is the change of enthalpy upon unfolding at the melting temperature. $\Delta C^o{}_p$ is the change of heat capacity upon unfolding.

Thermal and Urea/Guanidine Denaturation Conditions

Table S2. Conditions for thermal and urea/guanidine denaturation experiments

Protein	Urea/guanidine hydrochloride	Buffer	pH	Temperature (˚C)
EH	Urea	50mM sodium acetate	5.7	5
GA	Guanidine hydrochloride	50mM phosphate	7.0	25
HP35	Urea	100mM sodium chloride and 20mM sodium acetate	4.8	25
PSBD	Guanidine hydrochloride	2mM sodium phosphate, 2mM sodium borate and 50mM sodium chloride	8.0	25

Molecular Dynamics Simulations Using an Explicit-water Model

The starting structures used for the simulations of EH, GA, HP35, NTL9, PSBD, Trp-cage, UBA and ubiquitin were obtained from the pdb files 1ENH [4] , 1PRB [5], 1WY4 [6], 2HBB [7], 2PDD [8], 1L2Y [9], 1DV0 [10] and 1UBQ [11] respectively. Residues not included in the sequences listed above were deleted from the pdb file and the actual missing residues were added by Swiss PDB [12] and equilibrated by MD simulations with restraints on all other residues. C-Terminal amidation and N-terminal acetylation was added if the studied proteins had these modifications. X-ray structures are available for EH, HP35, NTL9 and ubiquitin, while only NMR structures are available for GA, PSBD, Trp-cage and UBA. For proteins with multiple models from NMR studies, the RMSD of each model was calculated using the average conformation as the reference. The model with the lowest RMSD was chosen as the starting structure for MD simulations. Starting structures for D-Ala mutants were created using tLeap in Amber [3]. Four independent MD simulations were run for each protein and for the D-Ala variant with different initial velocities, which results in eight simulations in total. The length of the simulations were 200 ns with the stepsize set to 2 fs. All simulations were performed using the Amber software package with the Amber ff14SB force field [13] and TIP3P water [14]. Parameters for nor-leucine were obtained from Forcefield_NCAA [15]. No ions were included in the system. All simulations were conducted under constant pressure conditions at 298K using Berendsen barostat to control pressure [16]. Temperature was controlled using a weak-coupling algorithm with the coupling constant set to 1 ps [16]. Truncated octahedron boxes with periodic boundary condition were used. Particle mesh Ewald methods were used to calculate electrostatic energies [17]. Hydrogen atoms were constrained using the SHAKE algorithm [18]. The cutoff of non-bonded interactions was set to 8 Å. The N-terminus was acetylated and C-terminus was amidated for proteins which had free termini and in which the termini were calculated to be neutral since deprotonated N-terminus and

protonated C-terminus are not currently available in the Amber force field [13]. Regular terminal residues defined in the Amber force field [13] were used for cases where the N and C termini were charged.

Local effects in the unfolded state were modeled as blocked tetrapeptides with sequence ACE-Xaa_1-Gly/dAla-Xaa_2-NH2. Xaa_1 and Xaa_2 are the two residues adjacent to the C-capping Gly/dAla in the full length protein sequences. This approach provides a model of purely local interactions and is not meant to mimic the actual unfolded chain. In order to enhance sampling, the tetrapeptides were simulated at 500K for 0.4ns, followed by cooling from 500K to 298K in 0.4ns and 0.4ns at 298K. This annealing cycle was repeated 120 times. Only data from 298K was collected for all cycles. These procedures were repeated thrice with different initial velocities which resulted in 3 sets of 4 independent folded state simulations and 3 sets of 120 annealing cycles of unfolded state simulations. A total of 96,000 frames from the folded state simulations and 144,000 frames from the unfolded state simulations at 298K were saved for analysis.

Starting Structures of PSBD, Trp-cage and UBA used for MD Simulations

PSBD, Trp-cage and UBA have multiple models obtained through NMR experiments. For each model, the backbone RMSD was calculated using VMD [2]. The reference coordinates are the averaged coordinates of all the models. The models used as starting structures are as follows:

Protein	PDB code	Model number
PSBD	2PDD	Model 32
Trp-cage	1L2Y	Model 32
UBA	1DV0	Model 15

Assignment of Protonation States of Titratable Residues during MD Simulations

Protonation states of titratable residues were set to reflect the pH at which thermodynamic properties of proteins were measured. The H++ server was used to determine the protonation state [19]. Experimental $\Delta\Delta G^{\circ}$ have been reported for the ubiquitin variants over the pH range of 2.5 to 3.5 [20]. The value of $\Delta\Delta G^{\circ}$ at pH 2.5 was compared to the calculated value since the TI approach only allows fixed protonation states. By fixing all the acidic residues and the C-terminus to be protonated, the system resembles that expected at pH=2.5.

Protonation states for titratable residues and terminus are listed in the table below. Asp, Glu, and C-termini which are not listed were fixed in the deprotonated state. Lys, Arg and N-termini which are not listed were fixed in the protonated state.

Protein	pH	Asp and Glu	His	C-terminus and N-terminus
EH	5.7			
GA	7.0		52, doubly protonated	
HP35	4.8		68, doubly protonated	
NTL9	5.5			
PSBD	8.0			Deprotonated N-terminus
Trp-cage	7.0			
UBA	6.5			
Ubiquitin	2.5	All Asp and Glu are protonated	68, doubly protonated	Protonated C-terminus

Free Energy Calculations

Free energy calculations were performed using non-softcore thermodynamic integration implemented in Amber [3, 21]. Gly was turned into D-Ala in three stages. In the first stage, partial charges on the CA/HA2/HA3 of Gly were turned off. In the second stage, three dummy atoms were added to the disappearing glycine and van der Waals interaction of these dummy atoms were turned on so a D-Ala with no partial charges on the CA/HA/CB/HB1/HB2/HB3 atoms appeared. In the third stage, partial charges on the CA/HA/CB/HB1/HB2/HB3 atoms of D-Ala were turned on. The first and third stages have λ evenly distributed from 0.0 to 1.0 with an interval of 0.1 including 0.0 and 1.0. In order to avoid singularity at λ = 0.0 and λ = 1.0 and have more sampling at where $dV/d\lambda$ has a steep change, the second stage has λ equal to 0.00922, 0.04794, 0.115, 0.20634, 0.316, 0.43738, 0.56262, 0.68392, 0.79366, 0.88495, 0.95206, 0.99078. For the folded state, one set of the TI calculations began with the C-capping glycine in place and used the crystal structures. Dummy atoms were added to the experimental structures to give the starting structures for the second stage of the calculations. Starting structures for the third stage were obtained by changing the Gly in the experimental structures to D-Ala. The alternate set of TI calculations was derived from the last frames of a 50 ns standard MD simulations of the D-Ala mutants. The structures resulting from these simulations were converted back to the Gly containing variants to provide starting structures for the first stage of the calculations.

For the folded state, MD simulations used the same set up as the standard MD simulations described above except that the length of the simulation was set to 12ns for each window. The blocked peptides, which model local interactions in the unfolded state, were converted from Gly to D-Ala in three stages using the same λ values that were used for the folded states. The same sampling enhancement strategy described above was used for all stages and λ windows. Only data from 298K was collected. Numerical integration was performed using trapezoidal

integration. Three ΔΔG° values were obtained by dividing simulations of each λ window for the folded states and unfolded states into three blocks. Error bars for the calculated ΔΔG° were the standard deviation of the three ΔΔG° values.

Energy Decomposition and Analysis of First Shell Water Molecules

The van der Waals potential energy between Gly or D-Ala and the rest of protein was calculated by post processing MD simulation trajectories. 1-4 van der Waals interactions were considered as van der Waals interactions with a scaling factor of 0.5. $\Delta\Delta E_{vdw}$ is defined as:

$$\Delta\Delta E(VDW) = [E^{u}_{D-ala}(VDW) - E^{u}_{Gly}(VDW)] - [E^{f}_{D-ala}(VDW) - E^{f}_{Gly}(VDW)] \quad (5)$$

where "u" and "f" indicate unfolded and folded states respectively. For example, $E^{u}_{D-ala}(VDW)$ is the van der Waals interaction between D-Ala residue and the rest of the protein in the unfolded state.

The first shell water molecules were counted by using Cpptraj [22] in Amber, with a cutoff of 3.4 Å. For the folded states, the first shell water molecules around the amide nitrogen, amide proton, carbonyl carbon and carbonyl oxygen of residues i-4 to i+1 (i=Gly/D-Ala) were counted because these atoms are structurally close to the C-capping residues. For the unfolded states, the water molecules around amide nitrogen, amide proton, carbonyl carbon and carbonyl oxygen of residues i-1 to i+1 (i=Gly/D-Ala) were counted.

$$\text{Number of water molecules (unfolded} - \text{folded)} = (n^{u}_{D-ala} - n^{u}_{Gly}) - (n^{f}_{D-ala} - n^{f}_{Gly}) \quad (6)$$

Where n is the number of first shell water molecules. The error bars of $\Delta\Delta E_{vdw}$ and number of water molecules (unfolded-folded) are the standard deviation of the 3 sets of simulations.

The desolvation effect on the backbone was also quantified by using Poisson Boltzmann (PB) equation solved by DelPhi[23]. The Amber ff14SB partial charges[13] and Yamagishi, J's radii set[24] were used.

$$\Delta\Delta G(bb\_solvation) =$$

$$[G^{u}_{D-ala}(bb\_solvation) - G^{u}_{Gly}(bb_{solvation})] -$$

$$[G^{f}_{D-ala}(bb\_solvation) - G^{f}_{Gly}(bb\_solvation)]$$

(7)

Since PB equation is non-linear, the solvation energy of each term on the right side of equation 7 was calculated in two steps. In the first step, we calculated the solvation energy of the whole protein with partial charges on the amide nitrogen, amide proton, carbonyl carbon and carbonyl oxygen of residues i-4 to i+1 (i-1 to i+1 for the unfolded state; i=Gly/D-Ala). In the second step, the partial charges on the amide nitrogen, amide proton, carbonyl carbon and carbonyl oxygen of residues i-4 to i+1 (i-1 to i+1 for the unfolded state; i=Gly/D-Ala) were set to 0 and the solvation energy of the whole protein was calculated again. The difference in the solvation energy obtained from these two step was considered as the solvation energy of the backbone around the Gly/D-Ala.

Calculation of $\Delta\Delta E_{vdw\text{-}gb}$ Using an Implicit-solvent Model

The length of the simulations were 5 ns with stepsize set to 1fs. Amber ff14SBonlysc [25] was used and igb was set to 8 which corresponds to GBneck2 implicit solvent model [26]. Mbondi3 radii set was used [26]. Simulations were conducted under 200K due to low thermostability of proteins in the implicit-solvent model used here[25]. Langevin dynamics was employed with the collision

frequency set to 1 ps^{-1}. No cutoff of non-bond interactions was used. The salt concentration was set to 0.0 M.

For the experimentally tested proteins (EH, GA, HP35, NTL9, PSBD, Trp-cage, UBA and ubiquitin), the starting structures were prepared in the same way as for the simulations in explicit solvent except no solvent was added. For the 120 proteins and their D-Ala variants listed in **Table S3**, any selenomethionines were converted to methionines and all acidic residues were deprotonated and all basic residues except histidines were protonated. The protonation states of histidines depends on whether the hydrogen on δ or ε nitrogen is resolved by X-ray. If neither of the hydrogens is resolved, the ε nitrogen was protonated. Disulphide bonds were added as indicated by the authors of the structures. All non-protein molecules and ions were deleted. Local effects in the unfolded states of proteins were modeled as blocked tetrapeptides. The tetrapeptides were simulated at 400K for 0.4ns, followed by cooling from 400K to 200K in 0.4ns and 0.4ns at 200K. This annealing cycle was repeated 160 times. The van der Waals potential energy between Gly or D-Ala and the rest of protein was calculated by post processing MD simulation trajectories. 1-4 van der Waals interactions were considered as van der Waals interactions instead of bonded interactions. $\Delta\Delta E_{vdw\_gb}$ is defined as:

$$\Delta\Delta E(VDW\_gb) = [E^{u}_{D-ala}(VDW\_gb) - E^{u}_{Gly}(VDW\_gb)] - [E^{f}_{D-ala}(VDW\_gb) - E^{f}_{Gly}(VDW\_gb)] \quad (8)$$

where "u" and "f" indicate unfolded and folded states respectively. For example, $E^{u}_{D-ala}(VDW\_gb)$ is the van der Waals interaction between the D-Ala residue and the rest of the protein in the unfolded state calculated using the implicit-solvent model.

For the 8 experimentally tested proteins, each $E^{f}_{D-ala}(VDW\_gb)$ value and each $E^{f}_{Gly}(VDW\_gb)$ value is the average over 100,000 frames from 10 independent simulations with different random number seeds for Langevin dynamics. For the 120 target proteins and their variants, $E^{f}_{D-ala}(VDW\_gb)$ values and $E^{f}_{Gly}(VDW\_gb)$ values were averaged over 30,000 frames from 3 independent simulations. For all of the proteins, $E^{u}_{D-ala}(VDW\_gb)$ values and $E^{u}_{Gly}(VDW\_gb)$ values were averaged over 40,000 frames collected from the simulations at 200K.

Protein Chains Dataset and $\Delta\Delta E_{vdw\_gb}$

All protein chains listed here are non-redundant protein chains with BLAST [27] pvalue less than 10e-7. According to the authors of the structures, all of the protein chains are monomeric. All proteins have at least one α-helical C-capping Gly. The criteria for defining a helix was at least 5 sequential residues with $-140^{\circ} \leqslant \varphi \leqslant -30^{\circ}$ and $-90^{\circ} \leqslant \psi \leqslant 45^{\circ}$. A C-capping Gly is the first non-helical residue at the C-terminus of a helix with $20^{\circ} \leq \varphi \leq 125^{\circ}$ and $-45^{\circ} \leq \psi \leq 90^{\circ}$ [28]. $\Delta\Delta E_{vdw\_gb}$ values were only calculated for proteins with high sequence diversity. In order to do so, a table of sequence redundancy in protein data bank was obtained from Molecular Modelling Database [29]. A representative of each non-redundant sequence was chosen according to the ranking provided by this table.

Table S3. Calculated values of $\Delta\Delta E_{vdw\_gb}$ for 160 C-capping sites from 120 non-redundant proteins taken from the pdb bank. Positive $\Delta\Delta E_{vdw\_gb}$ values indicate a stabilizing effect.

pdb code	chain ID	Short description of protein	Organism	Site No.	Calculated $\Delta\Delta E_{vdw\_gb}$ (kcal/mol)
1ABA	A	T4 glutaredoxin	Enterobacteria phage T4 sensu lato	56	0.61
1C44	A	Sterol carrier protein 2	Oryctolagus cuniculus	32	0.15
				86	1.20
				97	0.95
1KAF	A	The DNA Binding Domain Of Phage T4 Transcription Factor MotA	Enterobacteria phage T4 sensu lato	125	0.59
				179	0.36
1KP6	A	Killer toxin kp6 alpha-subunit	Ustilago maydis	9	-0.35
1L8R	A	Dachshund protein	Homo sapiens	255	0.76
1L9L	A	Granulysin from cytolytic T lymphocytes	Homo sapiens	63	0.64
1LWB	A	Phospholipase A2 protein	Streptomyces violaceoruber	75	0.35
1MC2	A	Phospholipase A2 protein	Deinagkistrodon acutus	14	0.34
1MK0	A	The catalytic domain of intron endonuclease I-TevI	Enterobacteria phage T4 sensu lato	38	0.27
1MOL	A	Monellin	Dioscoreophyllum cumminsii	27	0.87
1NWZ	A	Light receptor photoactive yellow protein	Halorhodospira halophila	51	0.31
				86	0.54
1OOH	A	An odorant binding protein LUSH	Drosophila melanogaster	34	1.40
				56	1.08
1ORG	A	A pheromone-binding protein	Rhyparobia maderae	53	1.03
1OSD	A	A mercury-binding protein	Cupriavidus metallidurans	65	0.24
1PBJ	A	A hypothetical protein	Methanothermobacter thermautotrophicus	59	0.40
1Q6V	A	Phospholipase A2 protein	Daboia russelii	14	0.24
1R6J	A	The PDZ2 domain of syntenin	Homo sapiens	231	0.88
1SBX	A	The dachshund-homology domain of Nuclear protooncoprotein SKI	Homo sapiens	165	0.66
1T1J	B	A hypothetical protein	Pseudomonas aeruginosa	43	0.52
				111	0.51
1T8K	A	An apo acyl carrier protein	Escherichia coli	16	0.53
				33	0.37
1TP6	A	A hypothetical protein	Pseudomonas aeruginosa	22	1.21
1TQG	A	CheA phosphotransferase domain	Thermotoga maritima	55	0.20

1U8T	B	CheY protein	Escherichia coli	29	0.37
				102	1.00
1VCD	A	Nudix protein Ndx1	Thermus thermophilus	52	0.33
1VYI	A	The C-terminal domain of a polymerase cofactor	Rabies virus	254	0.70
1WHZ	A	A hypothetical protein	Thermus thermophilus	18	0.60
1WOL	A	An HEPN homologue	Sulfolobus tokodaii	25	0.37
				50	0.67
1WY4	A	A villin headpiece	Gallus gallus	51	-0.31
1XLQ	A	Putidaredoxin	Pseudomonas putida	31	0.44
1XMK	A	The Zβ domain from the RNA editing enzyme ADAR1	Homo sapiens	341	0.78
1YN3	A	An extracellular adherence protein	Staphylococcus aureus	203	0.12
1Z96	A	Mud1 UBA domain	Schizosaccharomyces pombe	307	0.07
1ZMA	A	A bacterocin transport accessory protein	Streptococcus pneumoniae	81	0.38
2ACY	A	An acyl-phosphatase	Bos taurus	34	0.72
2B1L	B	A thiol:disulfide oxidoreductase	Escherichia coli	97	0.37
2B8I	A	A putative bacterial secretion factor	Vibrio vulnificus	56	0.99
2BO1	A	Ribosomal protein L30E	Thermococcus celer	30	0.87
				57	0.65
				75	0.47
2BWF	A	The UBL domain of Dsk2	Saccharomyces cerevisiae	36	-0.05
2CWY	A	A hypothetical protein	Thermus thermophilus	15	0.59
				55	0.48
2CX7	B	Sterol carrier protein 2	Thermus thermophilus	89	0.94
				100	0.65
2D48	A	Interleukin 4	Homo sapiens	95	0.43
2D58	A	An ionized calcium-binding adaptor	Homo sapiens	78	0.55
2FB6	A	A hypothetical protein	Bacteroides thetaiotaomicron	34	0.42
				68	0.59
				82	0.38
				91	0.43
2FC3	A	Ribosomal protein L7Ae	Aeropyrum pernix	46	0.66
				91	0.79
2FE5	A	The second PDZ domain of DLG3	Homo sapiens	270	0.56
2FYG	A	Nsp10	Severe acute respiratory syndrome-related coronavirus	34	0.13
2HC8	A	The actuator domain from Cu^{+}-ATPase	Archaeoglobus fulgidus	277	-0.01
2HL7	A	The periplasmic domain of cytochromes C maturation protein H	Pseudomonas aeruginosa	55	0.72
2HU9	A	A Zn2+ and [2Fe-2S]-containing copper chaperone	Archaeoglobus fulgidus	102	0.78

2I6V	A	Epsc, a crucial component of the type 2 secretion system	Vibrio cholerae	254	0.55
2IAY	A	LP2179, a member of the PF08866 family	Lactobacillus plantarum	31	0.87
2ICT	A	Antitoxin HigA	Escherichia coli	44	0.90
2J5Y	A	An albumin-binding domain	Finegoldia magna	22	0.73
2NT4	A	A response regulator homolog	Myxococcus xanthus	26	0.24
2O0Q	A	A hypothetical protein	Caulobacter vibrioides	20	0.40
				32	0.44
2OGB	A	The C-terminal domain of neuregulin receptor degrading protein 1	Mus musculus	237	0.51
2OY3	A	A macrophage receptor	Mus musculus	463	0.15
2P1H	A	The caspase recruitment domains of apoptotic protease activating factor 1	Homo sapiens	35	0.58
				81	0.22
2P3H	A	The CorC_HlyC domain of a putative hemolysin	Corynebacterium glutamicum	31	0.40
2POS	A	Sylvaticin	Pythium sylvaticum	32	0.30
2PSP	A	A pancreatic spasmolytic polypeptide	Sus scrofa	33	0.39
2PVB	A	Parvalbumin	Esox lucius	34	0.96
2PYQ	C	An uncharacterized protein	uncharacterized protein	20	0.40
				68	0.60
2QJL	A	A ubiquitin-related modifier	Saccharomyces cerevisiae	17	0.30
2RH3	A	The C-terminal domain of VirC2	Agrobacterium tumefaciens	130	0.37
2VB1	A	Triclinic hen egg-white lysozyme	Gallus gallus	16	0.60
				102	0.17
2VSV	A	The PDZ domain of human rhophilin-2	Homo sapiens	55	0.55
2VWR	A	The second PDZ domain of the human numb-binding protein 2	Homo sapiens	379	0.60
2W50	A	The N-terminal domain of human conserved dopamine neurotrophic factor	Homo sapiens	29	0.60
				60	0.71
2WFB	A	The apo Form of the Orange Protein	Desulfovibrio gigas	67	0.54
				88	0.64
2WT8	A	The N-terminal Brct domain of human microcephalin	Homo sapiens	36	0.62
				67	0.59
				83	0.47
2XEV	B	The TPR domain of YbgF	Xanthomonas campestris	15	0.50
				89	0.46
2ZQE	A	The endonuclease domain of an anti-recombination enzyme	Thermus thermophiles	31	0.63
3A0S	A	The PAS domain of histidine kinase ThkA	Thermotoga maritima	448	-0.35
3A0U	A	Response regulator protein TrrA	Thermotoga maritima	25	0.23

3A4R	A	The small ubiquitin-like modifier domain in Nip45	Mus musculus	376	0.66
3B79	A	The N-terminal peptidase C39 like domain of the toxin secretion ATP-binding protein	Vibrio parahaemolyticus	17	0.25
				49	0.73
3BS7	A	The sterile alpha motif domain of hyphen/aveugle	Drosophila melanogaster	71	0.50
3C9P	A	An uncharacterized protein	Streptococcus pneumoniae	25	0.40
				40	0.47
				106	0.41
3CJK	A	Copper transport protein ATOX1	Homo sapiens	59	0.20
3D2Q	B	The tandem zinc finger 3 and 4 domain of muscleblind-like protein 1	Homo sapiens	19	0.57
3E0Z	B	A putative imidazole glycerol phosphate synthase homolog	Agathobacter rectalis	39	0.10
3E11	B	A predicted zincin-like metalloprotease	Acidothermus cellulolyticus	102	0.51
3EZI	B	Histidine kinase NarX sensor domain	Escherichia coli	94	0.75
3FBL	A	An uncharacterized protein	Acidianus filamentous virus 1	66	0.60
3FZ4	A	A possible arsenate reductase	Streptococcus mutans	50	0.26
				68	1.07
3ID4	A	RseP PDZ2 domain	Escherichia coli	239	0.58
3IPJ	A	A domain of the PTS system	Peptoclostridium difficile	83	0.34
3L2A	A	VP35 interferon inhibitory domain	Reston ebolavirus	259	0.31
3LJW	B	The second bromodomain of human polybromo	Homo sapiens	240	1.67
3LLB	A	An uncharacterized protein	Pseudomonas aeruginosa	27	-0.13
				47	-0.06
3M3G	A	An elicitor of plant defense responses	Trichoderma virens	115	0.70
3NIR	A	Crambin	Crambe hispanica	20	0.17
				31	0.27
3NUF	A	A PRD-containing transcription regulator	Lactobacillus paracasei	67	0.59
3O79	B	A Prion protein	Oryctolagus cuniculus	195	0.23
3ODV	A	Kaliotoxin	Androctonus mauritanicus	22	1.28
3PO0	A	A Ubiquitin-like small archaeal modifier proteins	Haloferax volcanii	14	0.27
3QMX	A	Glutaredoxin A	Synechocystis sp. PCC 6803	29	-0.32
3S0A	A	An odorant-binding protein	Apis mellifera	22	0.47
				34	0.90
3SNS	A	The C-terminal domain of lipoprotein BamC	Escherichia coli	263	1.12
				292	0.82
3SVI	A	The Pto-binding domain of	Pseudomonas syringae	157	0.48

		HopPmaL	group genomosp. 3	173	0.44
3SZS	B	Hellethionin D	Helleborus purpurascens	20	0.33
3T7Z	A	Nop N-terminal domain	Methanocaldococcus jannaschii	60	-0.13
				91	0.56
3UI6	A	Parvulin 14	Homo sapiens	61	0.29
3V1A	A	A Metal interface design	synthetic construct	22	0.61
3W1O	A	A hypothetical protein	Neisseria meningitidis	51	0.86
3WCQ	A	Ferredoxin	Cyanidioschyzon merolae	33	0.31
				73	0.55
3ZR8	X	Rxlr effector AVR3a11	Phytophthora capsici	100	0.65
4CVD	A	A cell wall binding module	Streptococcus phage Cp-1	263	1.08
				279	0.33
4D40	A	Type IV pilin	Shewanella oneidensis	28	0.17
4F55	A	The catalytic Domain of SleB rotein	Bacillus cereus	202	0.56
				222	0.23
4FQN	A	CCM2 C-terminal harmonin homology domain	Homo sapiens	328	0.50
4G9S	A	A goose-type lysozyme	Escherichia coli	60	1.04
4GOQ	A	A hypothetical protein	Caulobacter vibrioides	20	0.41
4HRO	A	Small archaeal modifier protein 1	Haloferax volcanii	14	0.22
4HS5	A	Protein CyaY	Psychromonas ingrahamii	25	0.60
4JIU	A	An uncharacterized protein	Pyrococcus abyssi	14	0.50
4N6X	A	Na(+)/H(+) exchange regulatory cofactor NHE-RF1/Chemokine receptor CXCR2 fusion protein	Homo sapiens	52	0.89
4PXV	A	The LysM domain of chitinase A	Pteris ryukyuensis	32	0.62
4XPX	A	Hemerythrin	Methylococcus capsulatus	69	0.37
				97	0.54

Experimental Thermal Denaturation of EH, HP35, PSBD and Their D-Ala Variants

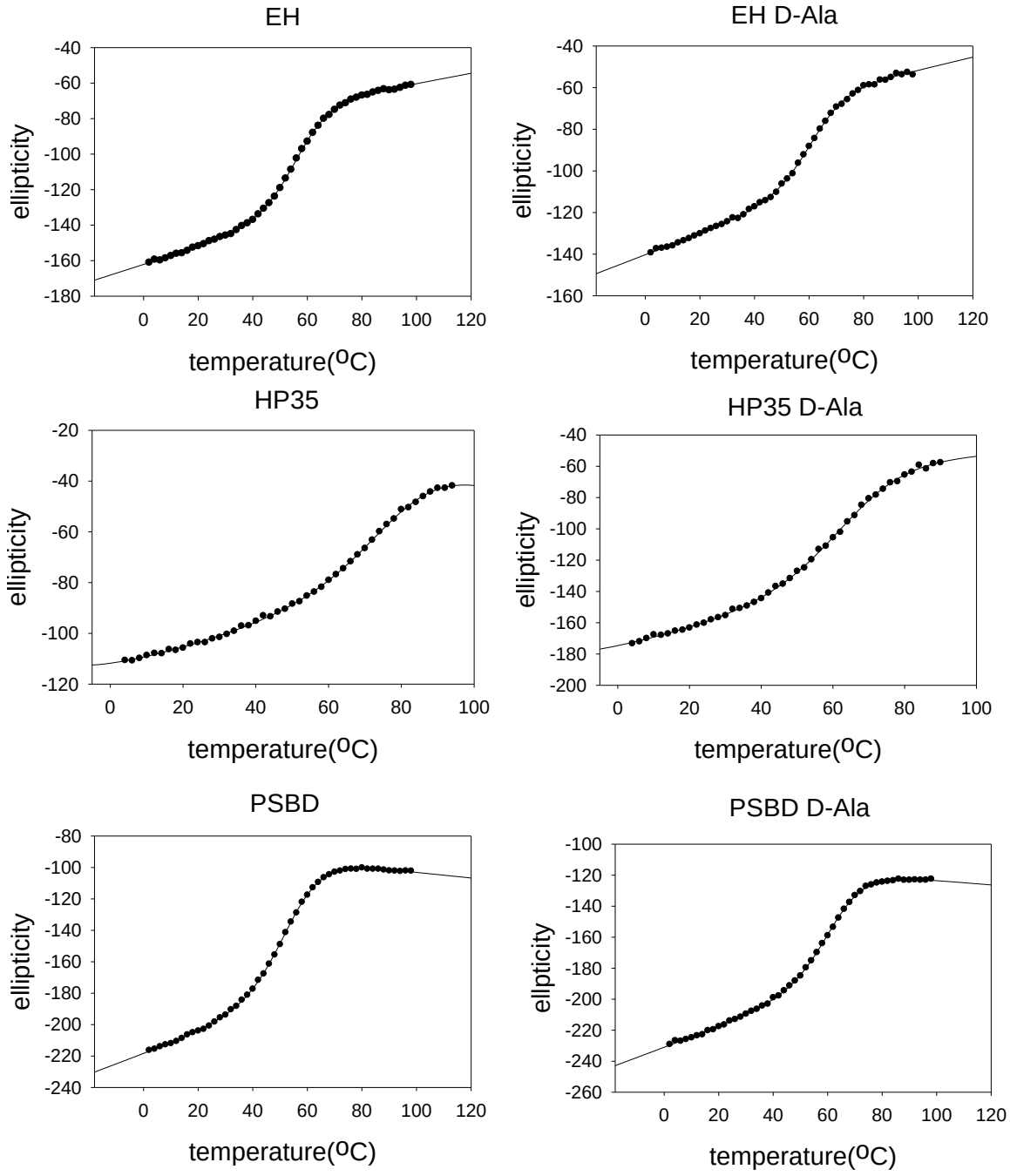
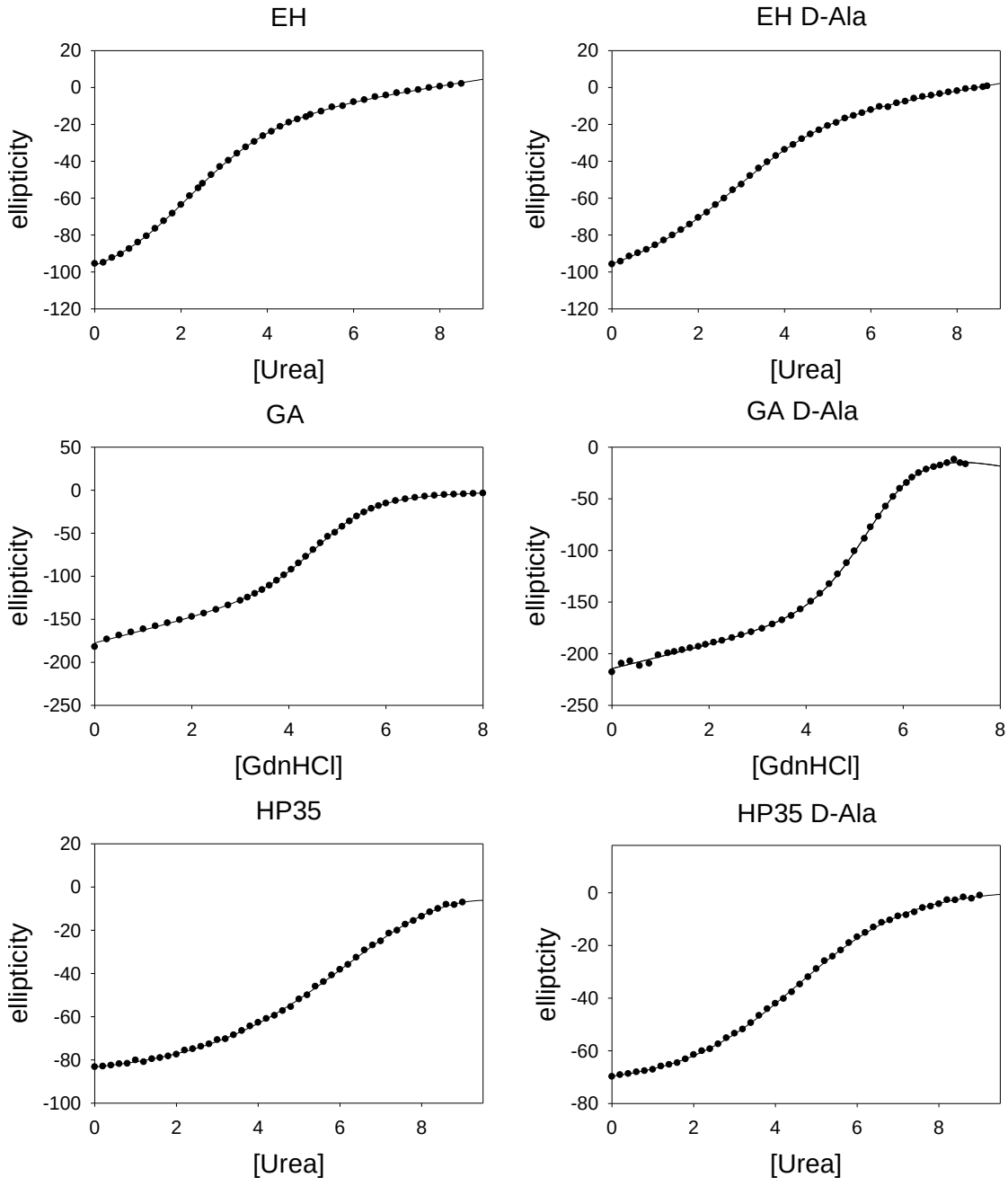


Figure S1. Thermal denaturation of EH, HP35, PSBD and their D-Ala variants. The solid line are the fitted curves.

Urea/Guanidine Hydrochloride Denaturation of EH, GA, HP35, PSBD and Their D-Ala Variants



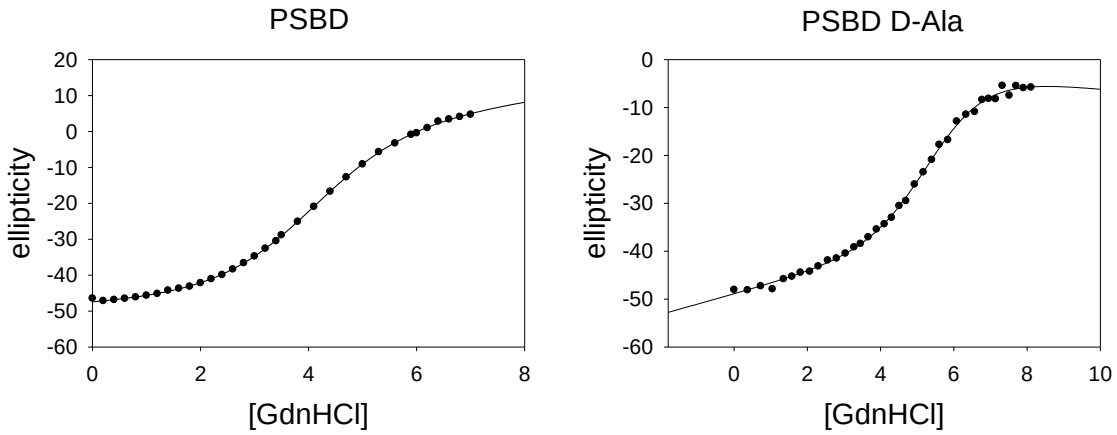


Figure S2. Urea/Guanidine hydrochloride denaturation of EH, GA, HP35, PSBD and their D-Ala variants. The solid lines are the fitted curves.

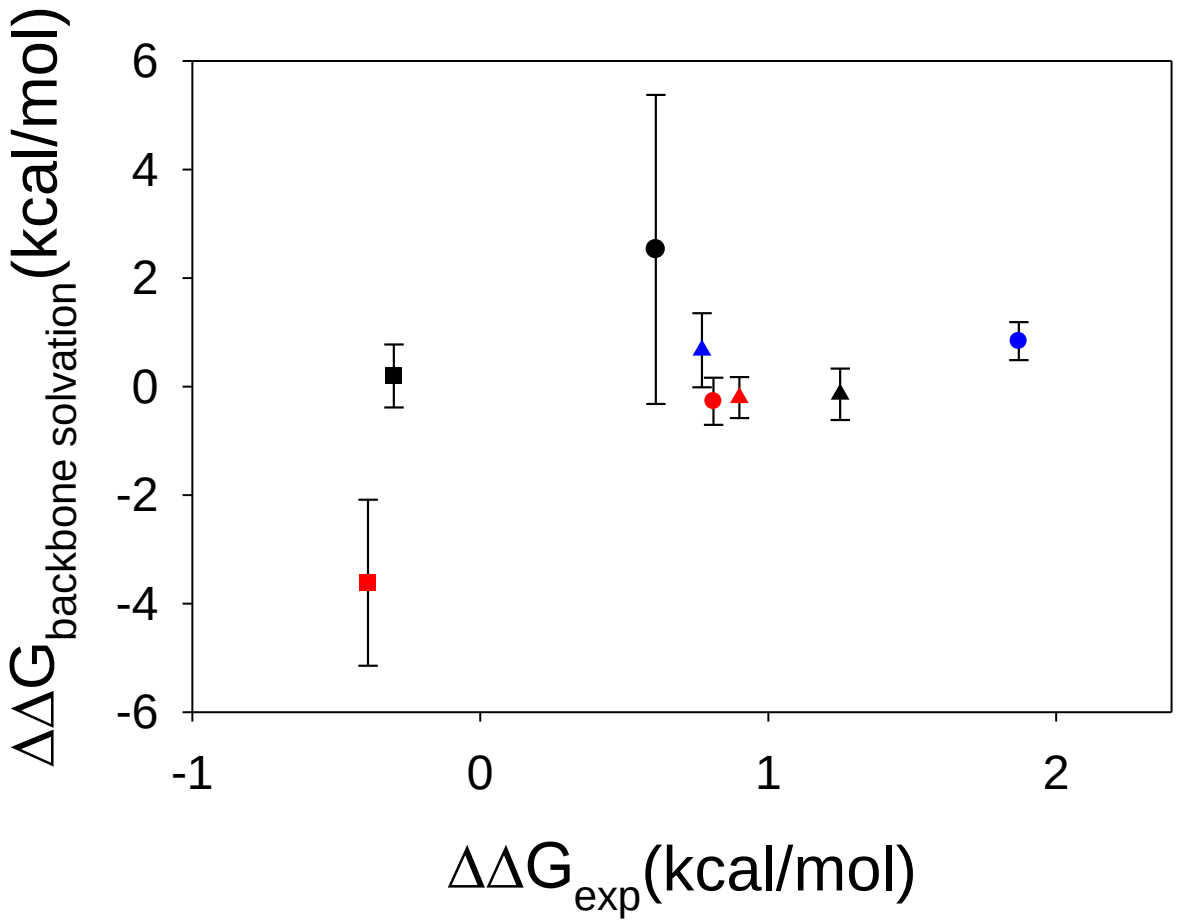


Figure S3. Correlation between $\Delta\Delta G_{backbone\ solvation}$ and $\Delta\Delta G_{exp}$. r=0.52, p=0.19. If only proteins with good convergence are included (GA, NTL9, PSBD, Trp-cage, UBA and ubiquitin), r=0.28, p-value=0.58, slope=0.20. EH ●; GA ●; HP35 ■; NTL9 ●; PSBD ▲; Trp-cage ▲; UBA ▲; Ubiquitin ■;

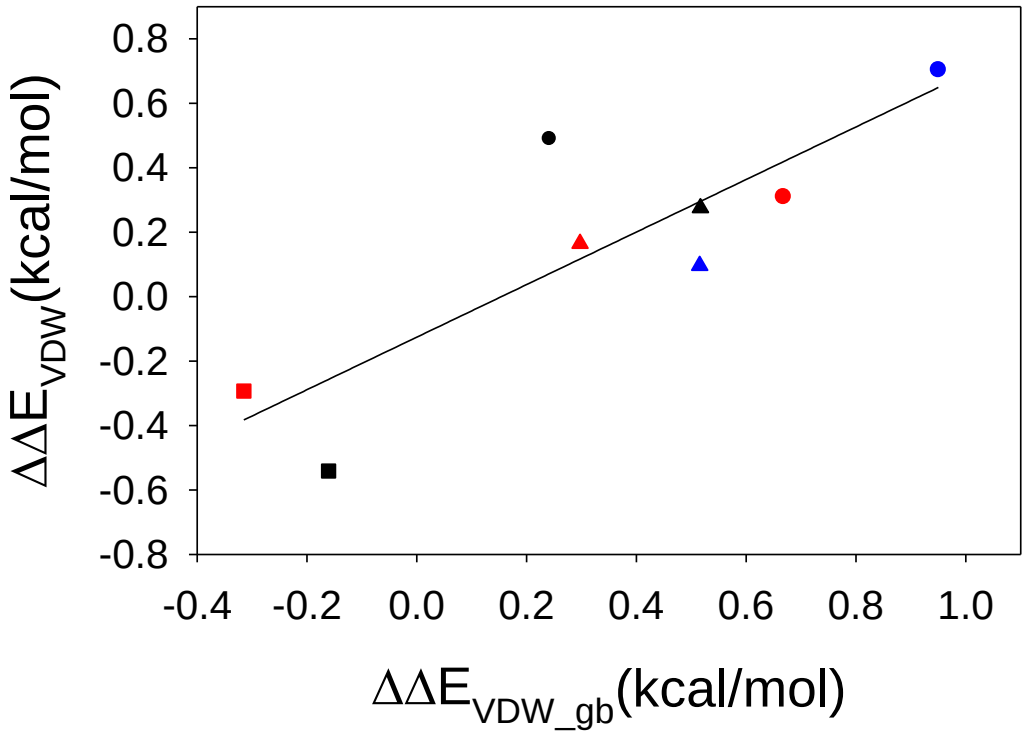


Figure S4. Correlation between $\Delta\Delta E_{vdw}$ and $\Delta\Delta E_{vdw\_gb}$. r=0.84, p=0.0079. EH ●; GA ●; HP35 ■; NTL9 ●; PSBD ▲; Trp-cage ▲; UBA ▲; Ubiquitin ■;

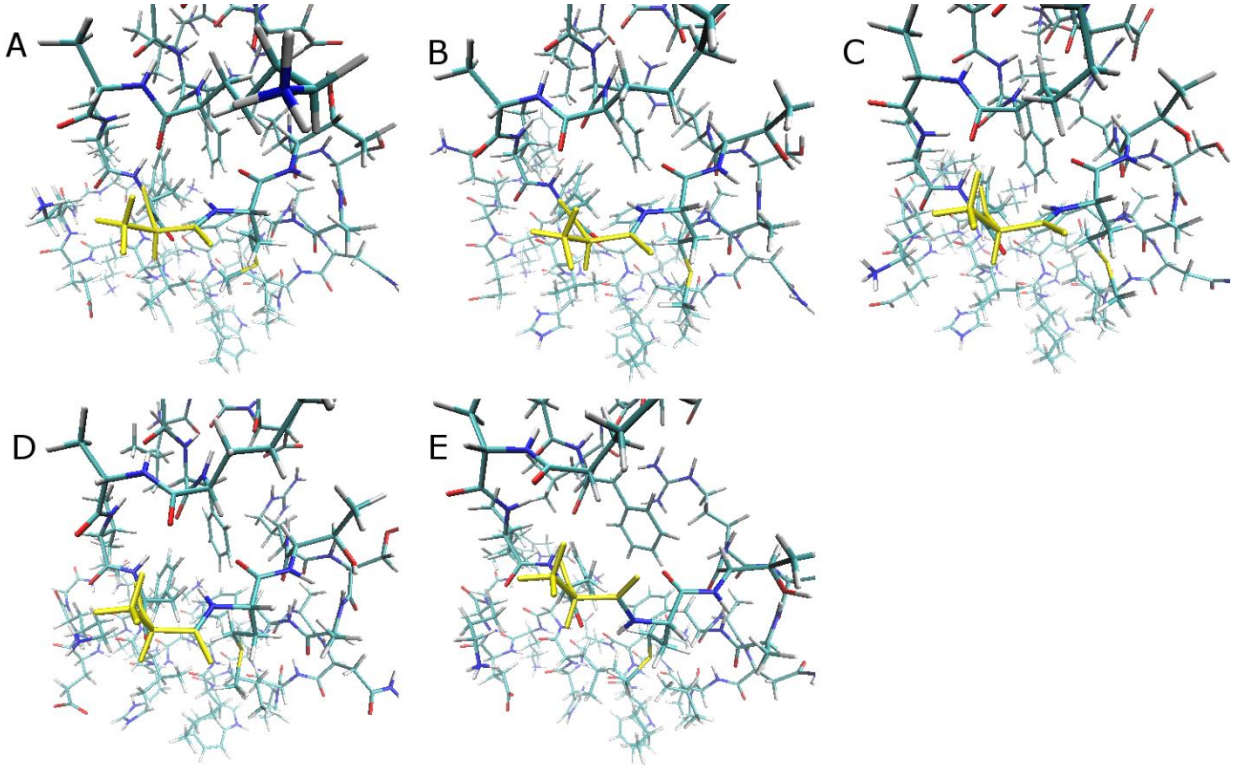


Figure S5. Structure of the HP35 G11D-Ala mutant taken from an MD simulation. 5 snapshots at 40 ns (A), 80 ns (B), 120 ns (C), 160 ns (D) and 200 ns (E) are shown with hydrogen included. The D-Ala residues are colored yellow.

1. Carpino, L. A.; Han, G. Y., *J. Am. Chem. Soc.* **1970,** *92*, 5748-&.
2. Humphrey, W.; Dalke, A.; Schulten, K., *J. Mol. Graph.* **1996,** *14*, 33-38, 27-38.
3. D.A. Case, J. T. B., R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D.M. York and P.A. Kollman *AMBER 2015*. University of California, San Francisco, 2015.
4. Clarke, N. D.; Kissinger, C. R.; Desjarlais, J.; Gilliland, G. L.; Pabo, C. O., *Protein Sci.* **1994,** *3*, 1779-1787.
5. Johansson, M. U.; de Chateau, M.; Wikstrom, M.; Forsen, S.; Drakenberg, T.; Bjorck, L., *J. Mol. Biol.* **1997,** *266*, 859-865.
6. Chiu, T. K.; Kubelka, J.; Herbst-Irmer, R.; Eaton, W. A.; Hofrichter, J.; Davies, D. R., *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102*, 7517-7522.
7. Cho, J. H.; Meng, W.; Sato, S.; Kim, E. Y.; Schindelin, H.; Raleigh, D. P., *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111*, 12079-12084.
8. Kalia, Y. N.; Brocklehurst, S. M.; Hipps, D. S.; Appella, E.; Sakaguchi, K.; Perham, R. N., *J. Mol. Biol.* **1993,** *230*, 323-341.
9. Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H., *Nat. Struct. Biol.* **2002,** *9*, 425-430.
10. Withers-Ward, E. S.; Mueller, T. D.; Chen, I. S. Y.; Feigon, J., *Biochemistry* **2000,** *39*, 14103-14112.
11. Vijaykumar, S.; Bugg, C. E.; Cook, W. J., *J. Mol. Biol.* **1987,** *194*, 531-544.
12. Guex, N.; Peitsch, M. C., *Electrophoresis* **1997,** *18*, 2714-2723.
13. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., *J. Chem. Theory Comput.* **2015,** *11*, 3696-3713.
14. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., *J. Chem. Phys.* **1983,** *79*, 926-935.
15. Khoury, G. A.; Smadbeck, J.; Tamamis, P.; Vandris, A. C.; Kieslich, C. A.; Floudas, C. A., *ACS Synth. Biol.* **2014,** *3*, 855-869.
16. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., *J. Chem. Phys.* **1984,** *81*, 3684-3690.
17. Darden, T.; York, D.; Pedersen, L., *J. Chem. Phys.* **1993,** *98*, 10089-10092.
18. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., *J. Comput. Phys.* **1977,** *23*, 327-341.
19. Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A., *Nucleic Acids Res.* **2005,** *33*, W368-371.
20. Bang, D.; Gribenko, A. V.; Tereshko, V.; Kossiakoff, A. A.; Kent, S. B.; Makhatadze, G. I., *Nat. Chem. Biol.* **2006,** *2*, 139-143.
21. Kirkwood, J. G., *J. Chem. Phys.* **1935,** *3*, 300-313.
22. Roe, D. R.; Cheatham, T. E., 3rd, *J. Chem. Theory Comput.* **2013,** *9*, 3084-3095.
23. Li, L.; Li, C.; Sarkar, S.; Zhang, J.; Witham, S.; Zhang, Z.; Wang, L.; Smith, N.; Petukh, M.; Alexov, E., *BMC Biophys.* **2012,** *5*, 9.
24. Yamagishi, J.; Okimoto, N.; Morimoto, G.; Taiji, M., *J. Comput. Chem.* **2014,** *35*, 2132-2139.
25. Nguyen, H.; Maier, J.; Huang, H.; Perrone, V.; Simmerling, C., *J. Am. Chem. Soc.* **2014,** *136*, 13959-13962.
26. Nguyen, H.; Roe, D. R.; Simmerling, C., *J. Chem. Theory Comput.* **2013,** *9*, 2020-2034.

27. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., *J. Mol. Biol.* **1990,** *215*, 403-410.
28. Gunasekaran, K.; Nagarajaram, H. A.; Ramakrishnan, C.; Balaram, P., *J. Mol. Biol.* **1998,** *275*, 917-932.
29. Madej, T.; Lanczycki, C. J.; Zhang, D.; Thiessen, P. A.; Geer, R. C.; Marchler-Bauer, A.; Bryant, S. H., *Nucleic Acids Res.* **2014,** *42*, D297-303.