Supporting material for:

Dissecting the Energetics of Intrinsically Disordered Proteins via a Hybrid Experimental and Computational Approach

Junjie Zou[1,2], Carlos Simmerling[1,2,*], Daniel P. Raleigh[1,2,*]

[1]Department of Chemistry, Stony Brook University, Stony Brook, New York 11794-3400, United States

[2]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York 11794-3400, United States

* Authors to whom correspondence should be addressed:

DPR email: daniel.raleigh@stonybrook.edu phone: (631)-632-9547;

CS email: carlos.simmerling@stonybrook.edu, phone: (631)-632-5424.

Methods

Free Energy Calculations for the Binding Between the Ovomucoid Inhibitor Third Domain (OMTKY3) and Its Target Proteases

The PDB codes for the structures of monomeric OMTKY3, *Streptomyces griseus* proteinase B (SGPB)/OMTKY3 complexes and subtilisin Carlsberg (CARL)/OMTKY3 complex used in the free energy calculations are listed in **Table S1**. Hydrogen atoms were added using the MolProbity program[1]. Side chain rotamer states for ASN/GLN were corrected based on suggestions provided by MolProbity. LYS, ARG side chains and N-termini were set to be protonated and ASP, GLU side chains and C-termini were set to be unprotonated. All HIS side chains were set to be neutral with protons on Nε except HIS57 in SGPB and HIS64/HIS226 in CARL which have protons on Nδ. Waters present in X-ray structures were kept while all salt ions were deleted. Truncated octahedron boxes were used to solvate the proteins. Free energy calculations were performed using non-softcore thermodynamic integration (TI) implemented in Amber[2-3]. The Amber force field ff14SB and the TIP3P water model were used for the TI calculations[4-5] with the implementation of GPU-accelerated thermodynamic integration, using pmemdGTI[6].

Minimization and equilibration under constant pressure[7] were conducted to heat up and relax the X-ray structures. Production runs were conducted using the implementation of GPU-accelerated thermodynamic integration, pmemdGTI[6], under constant volume. The temperature was set to 294K and no salt ions were included. Langevin dynamics was used to control temperature and the collision frequency was set to be 1.0 ps^{-1}. Particle mesh Ewald methods were used to calculate electrostatic energies[8]. Hydrogen atoms were constrained using the SHAKE algorithm[9]. The cutoff

of non-bonded interactions was set to 8 Å. A timestep of 2fs was used. The simulation time length for each λ window is 4ns. The trapezoidal rule was used for the integration of all λ windows.

Free Energy Calculations for D-to-A Mutations in SGPB/OMTKY3.

Analysis of the D-to-A mutations requires knowledge of the protonation state of the Asp in the free and complexed state. The free state pKa is available from NMR measurements and the bound state pKa has been estimated from experimental pKa dependent binding free energies. The NMR-based pKa of Asp18 in the free state of OMTKY3-Asp18 is 3.87[10]. The pKa values of Asp18 in the SGPB/OMTKY3-Asp18 complex were estimated to be 9.26 using the experimentally determined pH dependence of association equilibrium constants[11]. The experimental pH for the binding affinity measurement was 8.30[12], which indicates that Asp18 is fully deprotonated in the unbound state, but about 90% protonated in the complex state. The thermodynamic cycle of OMTKY3-Asp18 to SGPB is described in **Fig. S1**. To calculate the binding free energy difference between SGPB/OMTKY3-Ala18 and SGPB/OMTKY3-Asp18, which is equivalent to ③ - ⑥, one needs to calculate ④ and ⑤ in addition to ② - ① due to the change of protonation state of Asp18 upon binding. The protonation free energy of Asp18 in the free and complex state of OMTKY3 at pH=8.30 can be calculated as:

$$\Delta G = 2.303RT(pH - pKa_{free/com}^{Asp18})$$

$pKa_{free/com}^{Asp18}$ is the pKa of Asp18 of OMTKY3 in the free or complex state. The binding free energy difference between SGPB/OMTKY3-Ala18 and SGPB/OMTKY3-Asp18 can be obtained by ②+⑤-①-④.

Molecular Dynamics (MD) Simulations of the NTAIL/XD Mutants.

The starting structure for the MD simulations of wildtype NTAIL/XD complexes was obtained from PDB code 1T6O[13]. *In silico* mutations of A494G, L495A and L498A were made using Swiss PDB[14]. Hydrogen atoms were added using the MolProbity program[1]. Residues G484 and S485 which belong to the artificial linker between NTAIL and XD in the X-ray structure were removed and the C-terminus and the N-terminus of NTAIL were amidated and acetylated respectively. LYS, ARG and free N-termini were set to be protonated and ASP, GLU and free C-termini were set to be unprotonated. His498 of XD was set to be neutral with the proton on Nε. The protocol for the simulations was the same as the one used for the TI calculations of SGPB/OMTKY3 except that a temperature of 298K was used here. The simulations were 100ns long for each mutant and the last frames of simulations were saved for the backward TI calculations of NTAIL/XD.

Free Energy Calculations for the NTAIL/XD Complexes, Capped tripeptides and Fully Helical NTAIL (486-504)

The procedure used for the TI calculations of NTAIL/XD complexes, capped tripeptides and fully helical α-MoRE was the same as the one used for the TI calculations of SGPB/OMTKY3, except that a length of 20ns was used for all windows of the TI calculations on capped tripeptides, and a window of 4ns was used for the other two calculations.

The fully helical NTAIL (486-504) segment was obtained by stripping off the XD and the artificial linker in the X-ray structure of NTAIL/XD complex (PDB code 1T6O). The C-terminus and the N-terminus were amidated and acetylated respectively. Three independent TI calculations with different initial velocities were conducted for the capped tripeptides and the fully helical NTAIL (486-504).

Analysis of the Origin of Outliers in the TI Calculations

One of the outliers in **Fig. 3**, the S-to-C mutation, may be caused by problematic Lennard-Jones (LJ) parameters for sulfur in ff14SB. ε, which defines the minimum of the 6-12 LJ interaction profile, for sulfur in Cys is only 1.2 times that of the hydroxyl oxygen in Ser, but the difference is larger in other force fields: 2.5 times in the OPLS-AA/m[15] and 3.0 times in the CHARMM36m force field[16]. Calculations using LJ parameters from OPLS-AA/m reduced the error to 0.8 kcal/mol. The other outlier, the Y-to-F mutation, may due to a buried and structured water forming a hydrogen bond with the hydroxyl group of the Tyr as observed in the X-ray structures (**Fig. S2**). This water molecule was included in the TI calculations. The X-ray structures were solved at 287K. However, it is unclear whether this water should be included, since the water may be displaced at the condition of the experiments. Moreover, the direction dependency of hydrogen bonds is not well described in fixed-charge force fields which may also lead to inaccurate interaction strengths of hydrogen bonds[17-18]. In addition, there may be an issue with the TIP3P water model that leads to inaccuracies in modeling the geometries of ordered waters. These observations suggest caution needs to be employed when conducting TI calculations on systems with bound water that participate in hydrogen bond interactions. It is surprising that the calculated $\Delta\Delta G_{calc}$ for the V-to-A mutation has the largest error because the difference between V and A in terms of size and hydrophobicity is relatively small among the mutations studied here. Moreover, the calculated $\Delta\Delta G$ values of I-to-V and V-to-T are in good agreement with the experimental results as are most non-beta-branched to non-beta-branched mutations. The likely explanation is that the force field has a poor transferability between beta-branched and non-beta-branched amino acids.

Table S1. PDB codes for the structures of SGPB/OMTKY3 and CARL/OMTKY3 studied using TI calculations.

	PDB code
OMTKY3-LEU18	2GKR[19]
SGPB/OMTKY3-LEU18	1SGR[20]
SGPB/OMTKY3-ALA18	1SGP[20]
SGPB/OMTKY3-GLY18	1SGQ[20]
SGPB/OMTKY3-ASN18	1SGN
SGPB/OMTKY3-SER18	1CT0[21]
SGPB/OMTKY3-VAL18	1CT4[21]
SGPB/OMTKY3-ILE18	1CSO[21]
SGPB/OMTKY3-THR18	1CT2[21]
SGPB/OMTKY3-TYR18	1SGY
SGPB/OMTKY3-PHE18	2SGF
SGPB/OMTKY3-ASP18	1SGD
CARL/OMTKY3-LEU18	1R0R[22]

Table S2. $\Delta\Delta G_{calc}$ and $\Delta\Delta G_{exp}$ values (kcal/mol) for the mutations studied in the SGPB/OMTKY3 and CARL/OMTKY3 complexes.

	ΔG_{com} (forward)	ΔG_{com} (backward)	ΔG_{free}	$\Delta\Delta G_{calc}$ (forward)	$\Delta\Delta G_{calc}$ (backward)	$\Delta\Delta G_{exp}$
A to G	-6.64	-6.62	-8.92	2.27	2.31	1.99
S to A	7.56	7.64	8.54	-0.98	-0.90	-1.15
I to V	-21.78	-22.03	-20.21	-1.57	-1.82	-1.42
L to A	17.18	17.19	14.41	2.76	2.78	2.95
L to A (CARL)	14.54		14.41	0.14		0.31
L to N	-51.20	-51.82	-54.79	3.59	2.97	3.35
V to T	-9.60	-9.58	-9.62	0.03	0.05	0.16
L to F	22.72	23.22	21.42	1.31	1.81	1.36
V to A	16.83	16.65	14.82	2.01	1.83	-0.10
C to S		-8.37	-10.72		2.33	4.11
Y to F	22.45	22.25	21.25	1.20	1.00	-0.30
D to A *	54.71	54.02	50.86	-2.24	-2.93	-2.64

*: The ΔG_{com} and ΔG_{free}, which correspond to steps ② and ① in Fig S1 respectively, are calculated by changing neutral Asp to Ala. $\Delta\Delta G_{calc}$ is calculated as explained in the previous paragraphs.

Table S3. Comparison of ΔG_{free} obtained through $\Delta G_{free} = \Delta G_{bind} - \Delta G'_{bind} + \Delta G_{com}$ and the directly calculated ΔG_{free} (kcal/mol) for the mutations studied in the SGPB/OMTKY3 and CARL/OMTKY3 complexes.

	$\Delta G_{bind} - \Delta G'_{bind} + \Delta G_{com}$ (forward)	$\Delta G_{bind} - \Delta G'_{bind} + \Delta G_{com}$ (backward)	Directly calculated ΔG_{free}	Error (forward)	Error (backward)
A to G	-8.63	-8.61	-8.92	0.29	0.31
S to A	8.71	8.79	8.54	0.17	0.25
I to V	-20.36	-20.61	-20.21	-0.15	-0.4
L to A	14.23	14.24	14.41	-0.18	-0.17
L to A (CARL)	14.23		14.41	-0.18	
L to N	-54.55	-55.17	-54.79	-0.24	-0.38
V to T	-9.76	-9.74	-9.62	-0.14	-0.12
L to F	21.36	21.86	21.42	-0.06	-0.44
V to A	16.93	16.75	14.82	2.11	1.93
C to S		-12.48	-10.72		-1.76
Y to F	22.75	22.55	21.25	1.50	1.30
D to A *	51.26	50.57	50.86	0.40	-0.29

*: The ΔG_{com} and ΔG_{free}, which are ② and ① in Fig S1 respectively, are calculated by changing neutral Asp to Ala. $\Delta\Delta G_{calc}$ is calculated as explained in the previous paragraphs.

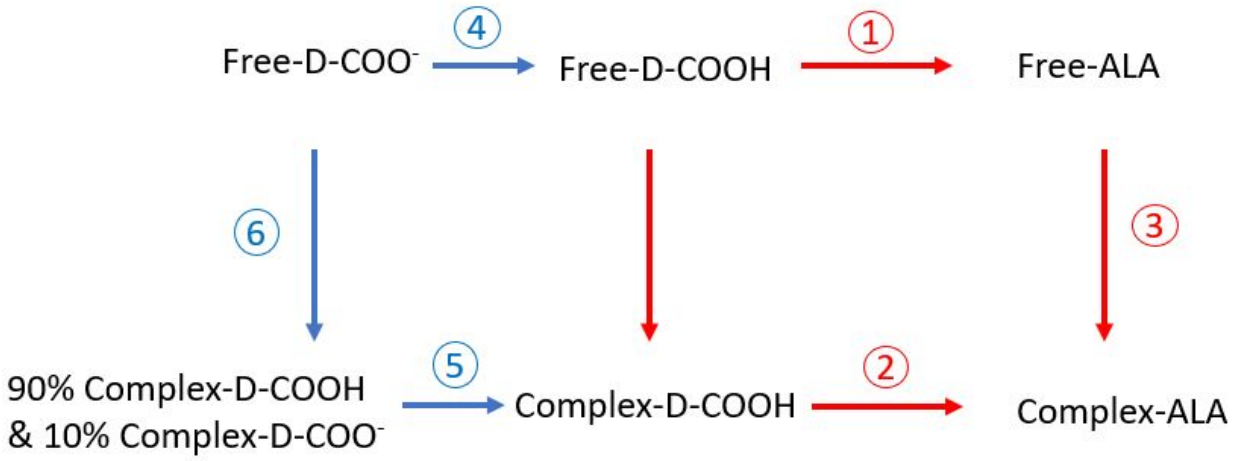

Figure S1. A thermodynamic cycle of OMTKY3-ASP18 forming a complex with SGPB with an additional process of protonation. D-COO^- indicates the deprotonated state of Asp18 and D-COOH indicates the protonated state of Asp18. ③ and ⑥ are the binding free energy of OMTKY3-Ala18 to SGPB and OMTKY3-Asp18 to SGPB respectively. ① and ② are the free energy changes of mutation from protonated aspartate to alanine calculated using TI in the unbound state of OMTKY3 and the SGPB/OMTKY3 respectively. ④ and ⑤ are the protonation free energy of Asp18 in the free state and complex state of OMTKY3 at pH 8.30 respectively. The red cycle is the thermodynamic cycle used for non-titratable residues in this study.

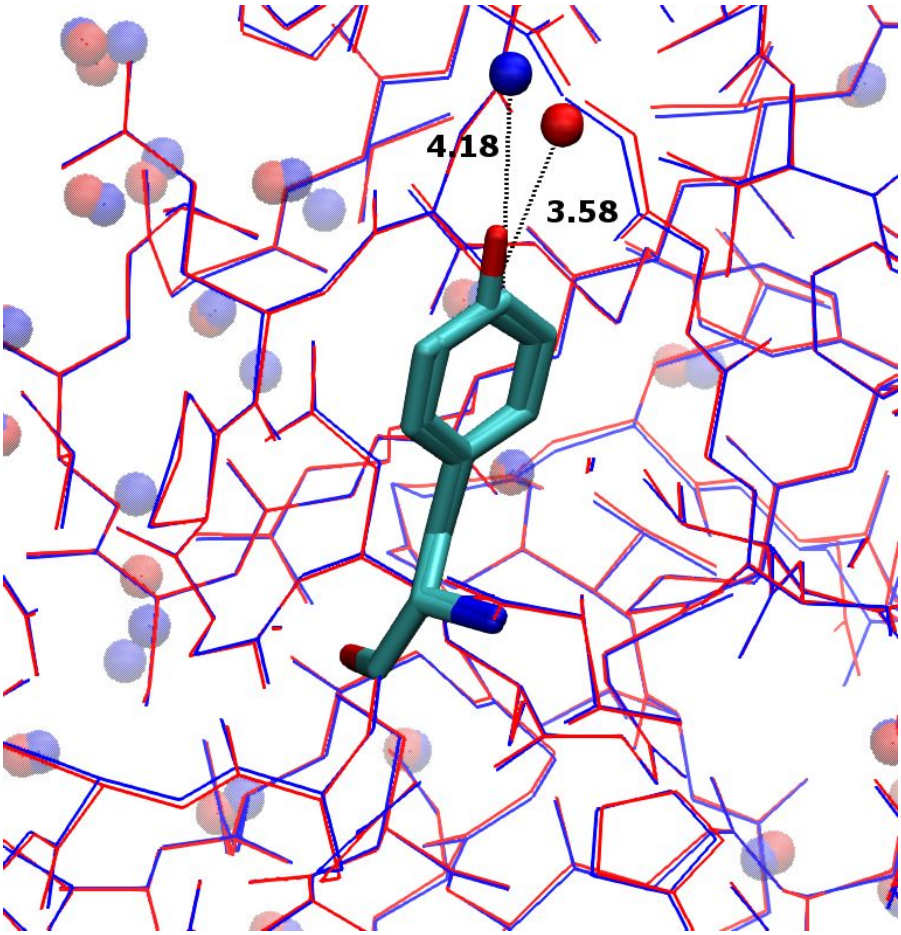

Figure S2. The X-ray structures of SGPB/OMYKY3-Phe18 (blue, PDB code 2sgf) and SGPB/OMTKY3-Tyr18 (red, PDB code 1sgy). Phe18 and Tyr18 are shown licorice. The other residues and water molecules are shown as lines and spheres respectively. Two water molecules found adjacent to the rings of Tyr18 and Phe18 are shown as opaque spheres. The distances between these two waters and the ζ-carbons of Phe18 and Tyr18 were measured to be 4.18 and 3.58 Å respectively. The shorter distance between the water and the ζ-carbon of Ty18 may be caused by a hydrogen bond involving the hydroxyl group and the water.

References

1. Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 1), 12-21.
2. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cruzeiro, I., V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K., et al., *AMBER 2018*. University of California, San Francisco, 2018.
3. Kirkwood, J. G., Statistical mechanics of fluid mixtures. *J Chem Phys* **1935,** *3* (5), 300-313.
4. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput.* **2015,** *11* (8), 3696-3713.
5. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **1983,** *79* (2), 926-935.
6. Lee, T. S.; Hu, Y.; Sherborne, B.; Guo, Z.; York, D. M., Toward fast and accurate binding affinity prediction with pmemdGTI: An efficient implementation of GPU-accelerated thermodynamic integration. *J Chem Theory Comput* **2017,** *13* (7), 3077-3084.
7. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., Molecular-dynamics with coupling to an external bath. *J Chem Phys* **1984,** *81* (8), 3684-3690.
8. Darden, T.; York, D.; Pedersen, L., Particle mesh ewald - an N.Log(N) method for ewald sums in large systems. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.
9. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of N-alkanes. *J Comput Phys* **1977,** *23* (3), 327-341.
10. Song, J.; Laskowski, M., Jr.; Qasim, M. A.; Markley, J. L., NMR determination of pKa values for Asp, Glu, His, and Lys mutants at each variable contiguous enzyme-inhibitor contact position of the turkey ovomucoid third domain. *Biochemistry* **2003,** *42* (10), 2847-2856.
11. Abul Qasim, M.; Ranjbar, M. R.; Wynn, R.; Anderson, S.; Laskowski, M., Jr., Ionizable P1 residues in serine proteinase inhibitors undergo large pK shifts on complex formation. *J Biol Chem* **1995,** *270* (46), 27419-27422.
12. Lu, W.; Apostol, I.; Qasim, M. A.; Warne, N.; Wynn, R.; Zhang, W. L.; Anderson, S.; Chiang, Y. W.; Ogin, E.; Rothberg, I., et al., Binding of amino acid side-chains to S1 cavities of serine proteinases. *J Mol Biol* **1997,** *266* (2), 441-461.
13. Kingston, R. L.; Hamel, D. J.; Gay, L. S.; Dahlquist, F. W.; Matthews, B. W., Structural basis for the attachment of a paramyxoviral polymerase to its template. *Proc Natl Acad Sci U S A* **2004,** *101* (22), 8301-8306.
14. Guex, N.; Peitsch, M. C., SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **1997,** *18* (15), 2714-2723.
15. Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L., Improved peptide and protein torsional energetics with the OPLSAA force field. *J Chem Theory Comput* **2015,** *11* (7), 3499-3509.
16. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., Jr., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **2017,** *14* (1), 71-73.

17. Oroguchi, T.; Nakasako, M., Influences of lone-pair electrons on directionality of hydrogen bonds formed by hydrophilic amino acid side chains in molecular dynamics simulation. *Sci Rep* **2017,** *7* (1), 15859.
18. Ren, P.; Wu, C.; Ponder, J. W., Polarizable atomic multipole-based molecular mechanics for organic molecules. *J Chem Theory Comput* **2011,** *7* (10), 3143-3161.
19. Lee, T. W.; Qasim, M. A.; Laskowski, M., Jr.; James, M. N., Structural insights into the non-additivity effects in the sequence-to-reactivity algorithm for serine peptidases and their inhibitors. *J Mol Biol* **2007,** *367* (2), 527-546.
20. Huang, K.; Lu, W.; Anderson, S.; Laskowski, M., Jr.; James, M. N., Water molecules participate in proteinase-inhibitor interactions: crystal structures of Leu18, Ala18, and Gly18 variants of turkey ovomucoid inhibitor third domain complexed with Streptomyces griseus proteinase B. *Protein Sci* **1995,** *4* (10), 1985-1997.
21. Bateman, K. S.; Anderson, S.; Lu, W.; Qasim, M. A.; Laskowski, M., Jr.; James, M. N., Deleterious effects of beta-branched residues in the S1 specificity pocket of Streptomyces griseus proteinase B (SGPB): crystal structures of the turkey ovomucoid third domain variants Ile18I, Val18I, Thr18I, and Ser18I in complex with SGPB. *Protein Sci* **2000,** *9* (1), 83-94.
22. Horn, J. R.; Ramaswamy, S.; Murphy, K. P., Structure and energetics of protein-protein interactions: the role of conformational heterogeneity in OMTKY3 binding to serine proteases. *J Mol Biol* **2003,** *331* (2), 497-508.