Biophysical Journal, Volume 99

Supporting Material

Refolding the Engrailed Homeodomain: Structural basis for the accumulation of a folding intermediate

Michelle McCully, David Beck, Alan Fersht, and Valerie Daggett

Supporting Material

Supplemental Discussion: Property Space Descripton of the Native State and Unfolding Pathway of EnHD

Our 35-dimensional property space analysis allows us to compare the native state of EnHD at different temperatures better than any one of the individual properties. The 310, 314, and 319 K native state overlapped in property space with the 298 K native state, but the higher-temperature native states spanned regions that the 298 K native state did not. In other words, the 298 K ensemble was a subset of the 310, 314, and 319 K native states. This is due to the higher prevalence of the N′ state at slightly elevated temperature. In N′, HIII translates ~10 Å towards the N-terminus (1), which has little effect on most of the reported properties. For example, the COM Distances are slightly, though not significantly, higher for residue pairs that have a member in the C-terminal portion of HIII as is the core Cα RMSD. The number of native contacts is slightly, though again not significantly, lower. The percentage of NOEs satisfied for the native state at 310, 314, and 319 K (86 ± 2%) was the same as for the native state at 298 K (85 ± 3%). So, despite the fact that the native state in the elevated temperature simulations is broader than at 298 K, its agreement with the native state as measured by experiment is comparable. Because the temperature is slightly elevated, EnHD is more likely to overcome the energy barriers that confine it to a smaller portion of the native energy well at lower temperatures.

From our 35-dimensional property space, we calculated a multidimensional embedded 1D reaction coordinate. The reaction coordinate for the high-temperature unfolding shows that the TS selected previously (2-5) falls just outside the reference distribution (Fig. 1), as would be expected for this method. The intermediate and denatured populations formed separate but connected peaks in the 498 K distribution. Instead of the generic 15 properties used for high-throughput analysis of data in the lab's Dynameomics database (6), many of the properties employed here were specific to EnHD. Better discrimination between the native, transition, intermediate, and denatured states was possible using the 35 properties listed in Table S3 than our standard 15 property subset. This underscores the importance of using multiple, protein-specific properties to determine the state of a protein.

Supplemental References

1. McCully, M. E., D. A. C. Beck, and V. Daggett. 2008. Microscopic reversibility of protein folding in molecular dynamics simulations of the engrailed homeodomain. *Biochemistry*. 47:7079-7089.
2. Gianni, S., N. R. Guydosh, F. Khan, T. D. Caldas, U. Mayor, G. W. White, M. L. DeMarco, V. Daggett, and A. R. Fersht. 2003. Unifying features in protein-folding mechanisms. *Proc Natl Acad Sci U S A*. 100:13286-13291.
3. Mayor, U., C. M. Johnson, V. Daggett, and A. R. Fersht. 2000. Protein folding and

unfolding in microseconds to nanoseconds by experiment and simulation. *Proc Natl Acad Sci U S A*. 97:13518-13522.

4. Mayor, U., N. R. Guydosh, C. M. Johnson, J. G. Grossmann, S. Sato, G. S. Jas, S. M. Freund, D. O. V. Alonso, V. Daggett, and A. R. Fersht. 2003. The complete folding pathway of a protein from nanoseconds to microseconds. *Nature*. 421:863-867.
5. DeMarco, M. L., D. O. V. Alonso, and V. Daggett. 2004. Diffusing and colliding: the atomic level folding/unfolding pathway of a small helical protein. *J Mol Biol*. 341:1109-1124.
6. Toofanny, R. D., A. L. Jonsson, and V. Daggett. 2010. A comprehensive multidimensional-embedded one-dimensional reaction coordinate for protein unfolding/folding. *Biophys J*. 98:2671-2681.
7. Religa, T. L. 2008. Comparison of multiple crystal structures with NMR data for engrailed homeodomain. *J Biol NMR*. 40:189-202.
8. Kabsch, W., and C. Sander. 1983. Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers*. 22:2577-2637.
9. Lee, B., and F. M. Richards. 1971. Interpretation of Protein Structures - Estimation of Static Accessibility. *J Mol Biol*. 55:379-380.

Table S1: Simulation Properties for Native Simulations

Temperature (K)	Run	Total Time (ns)	Reference Set*	% NOEs Satisfied †
298	1	80	all	83.0
	2	70	all	82.6
	3	50	all	88.8
	4	50	all	87.5
	5	20	all	85.6
	6	20	all	90.7
	7	20	all	88.1
310	1	100	all	82.0
	2	50	all	87.8
	3	100	0-94 ns	87.2
	4	100	0-4.6, 5.3-5.9, 6.2-100 ns	82.0
	5	20	all	88.5
	6	20	all	88.1
314	1	78	all	87.9
	2	79	all	85.9
	3	79	all	87.8
319	1	335	0-37 ns	83.5
	2	334	0-61 ns	87.5
	3	166	0-21 ns	85.8
	4	166	0-64 ns	88.1
	5	166	0-137 ns	87.2
Summary	21	2103	1249 ns	86.4

* Time during which EnHD was in the native state. Only the time spans in this column were used when calculating the reference set for the reaction coordinate.

† A total of 654 reported NOEs were used for comparison (7). NOEs were calculated over the reference set in column 4. An NOE was considered satisfied if the r^{-6} weighted distance between protons was ≤ 5.5 Å.

Table S2: Simulation Properties for Quench Simulations

Temperature (K)	Number of Simulations	Total Time (μs)
310	7	4.682
314	9	4.743
319	30	5.495
Total	46	14.920

Table S3: Property Space Weights and Values of the Reference Sets

	PC 1 Weight*	298 K	310 K	314 K	319 K
Core Cα RMSD[†] (Å)	0.99	2.19 ± 0.58	2.37 ± 0.54	2.31 ± 0.69	2.20 ± 0.60
Fraction α-Helix[‡]	0.90	0.72 ± 0.05	0.66 ± 0.07	0.67 ± 0.03	0.68 ± 0.04
COM Distance[§] Arg 30 - Glu 19 (Å)	0.95	6.99 ± 0.79	6.70 ± 0.61	6.50 ± 0.60	6.80 ± 0.86
COM Distance Leu 34 - Arg 15 (Å)	0.95	7.49 ± 0.91	7.80 ± 1.47	7.03 ± 0.88	7.27 ± 1.09
COM Distance Leu 34 - Leu 16 (Å)	0.96	8.37 ± 0.75	8.25 ± 0.94	7.52 ± 0.75	8.01 ± 0.87
COM Distance Glu 37 - Arg 15 (Å)	0.94	8.16 ± 0.68	8.49 ± 1.24	8.02 ± 0.64	8.47 ± 1.49
COM Distance Leu 38 - Gln 12 (Å)	0.95	7.17 ± 1.19	7.43 ± 1.27	6.87 ± 1.02	7.14 ± 1.25
COM Distance Ile 45 - Leu 38 (Å)	0.72	9.03 ± 1.10	9.54 ± 1.13	9.65 ± 1.29	9.54 ± 1.17
COM Distance Ile 45 - Leu 40 (Å)	0.67	6.74 ± 0.82	6.56 ± 0.62	6.61 ± 0.66	6.64 ± 0.70
COM Distance Trp 48 - Leu 16 (Å)	0.94	9.84 ± 1.72	10.08 ± 2.07	9.37 ± 2.03	9.15 ± 1.56
COM Distance Phe 49 - Leu 16 (Å)	0.93	9.41 ± 1.41	9.08 ± 1.62	9.75 ± 2.07	8.96 ± 1.60
COM Distance Phe 49 - Phe 20 (Å)	0.91	8.32 ± 1.59	7.50 ± 1.23	7.63 ± 1.02	7.38 ± 0.98
COM Distance Phe 49 - Arg 24 (Å)	0.90	9.70 ± 2.00	10.98 ± 3.01	11.39 ± 2.99	11.25 ± 2.64
COM Distance Phe 49 - Leu 26 (Å)	0.92	7.96 ± 1.69	8.92 ± 2.16	9.18 ± 2.35	8.87 ± 2.43
COM Distance Lys 52 - Phe 20 (Å)	0.90	9.55 ± 2.03	10.37 ± 2.65	10.65 ± 2.44	10.22 ± 2.69
COM Distance Arg 53 - Phe 20 (Å)	0.92	9.14 ± 1.94	8.78 ± 2.08	9.34 ± 2.14	9.19 ± 2.05
COM Distance Arg 53 - Arg 24 (Å)	0.92	8.33 ± 2.83	10.02 ± 4.11	10.95 ± 3.72	10.31 ± 3.85
COM Distance Arg 53 - Leu 26 (Å)	0.91	9.78 ± 2.59	12.09 ± 2.64	12.51 ± 2.42	11.92 ± 3.14
Core Main Chain SASA[¶] ($Å^2$)	0.93	459 ± 33	462 ± 42	455 ± 44	465 ± 39
Core Main Chain Polar SASA ($Å^2$)	0.93	276 ± 27	288 ± 34	282 ± 37	288 ± 35
Core Side Chain SASA ($Å^2$)	0.91	3282 ± 184	3255 ± 171	3217 ± 166	3233 ± 183
Core Side Chain Non-Polar SASA ($Å^2$)	0.92	1729 ± 152	1710 ± 145	1647 ± 144	1665 ± 131
Core Non-Polar SASA ($Å^2$)	0.91	1912 ± 158	1884 ± 151	1821 ± 147	1842 ± 135
Core Polar SASA ($Å^2$)	0.88	1829 ± 84	1833 ± 89	1852 ± 102	1856 ± 125
Core Total SASA ($Å^2$)	0.95	3741 ± 191	3718 ± 187	3672 ± 176	3697 ± 200
Trp 48 SASA ($Å^2$)	0.60	95 ± 32	98 ± 36	77 ± 38	76 ± 38
Native MC-MC Contacts[‖]	0.96	122.4 ± 2.1	120.9 ± 2.9	121.0 ± 2.5	121.4 ± 2.7
Native MC-SC Chain Contacts	0.96	117.1 ± 5.8	111.7 ± 7.9	112.6 ± 7.5	113.2 ± 8.4
Native SC-SC Contacts	0.93	63.1 ± 6.4	59.4 ± 6.6	60.3 ± 7.6	60.4 ± 7.0
Total Native Contacts	0.97	152.7 ± 7.0	149.7 ± 7.4	151.0 ± 8.4	151.5 ± 7.7
Nonnative MC-MC Contacts	0.85	4.1 ± 2.4	7.4 ± 3.9	6.7 ± 3.4	6.3 ± 3.0
Nonnative MC-SC Chain Contacts	0.87	4.9 ± 3.0	10.8 ± 5.6	11.1 ± 6.4	9.6 ± 5.0
Nonnative SC-SC Contacts	0.77	6.8 ± 3.3	12.6 ± 6.2	12.5 ± 5.7	12.2 ± 5.8
Total Nonnative Contacts	0.83	12.0 ± 4.6	20.2 ± 8.4	19.8 ± 8.3	18.6 ± 7.1
Total Contacts	0.90	164.8 ± 6.6	169.9 ± 7.3	170.7 ± 6.2	170.1 ± 6.3

* Weights are reported for the first principal component of the property space of a previously validated 498 K unfolding trajectory and the 298 K reference set.

[†] RMSD was calculated over the Cα atoms of residues 8-53. All properties are given as the average ± 1 s.d.

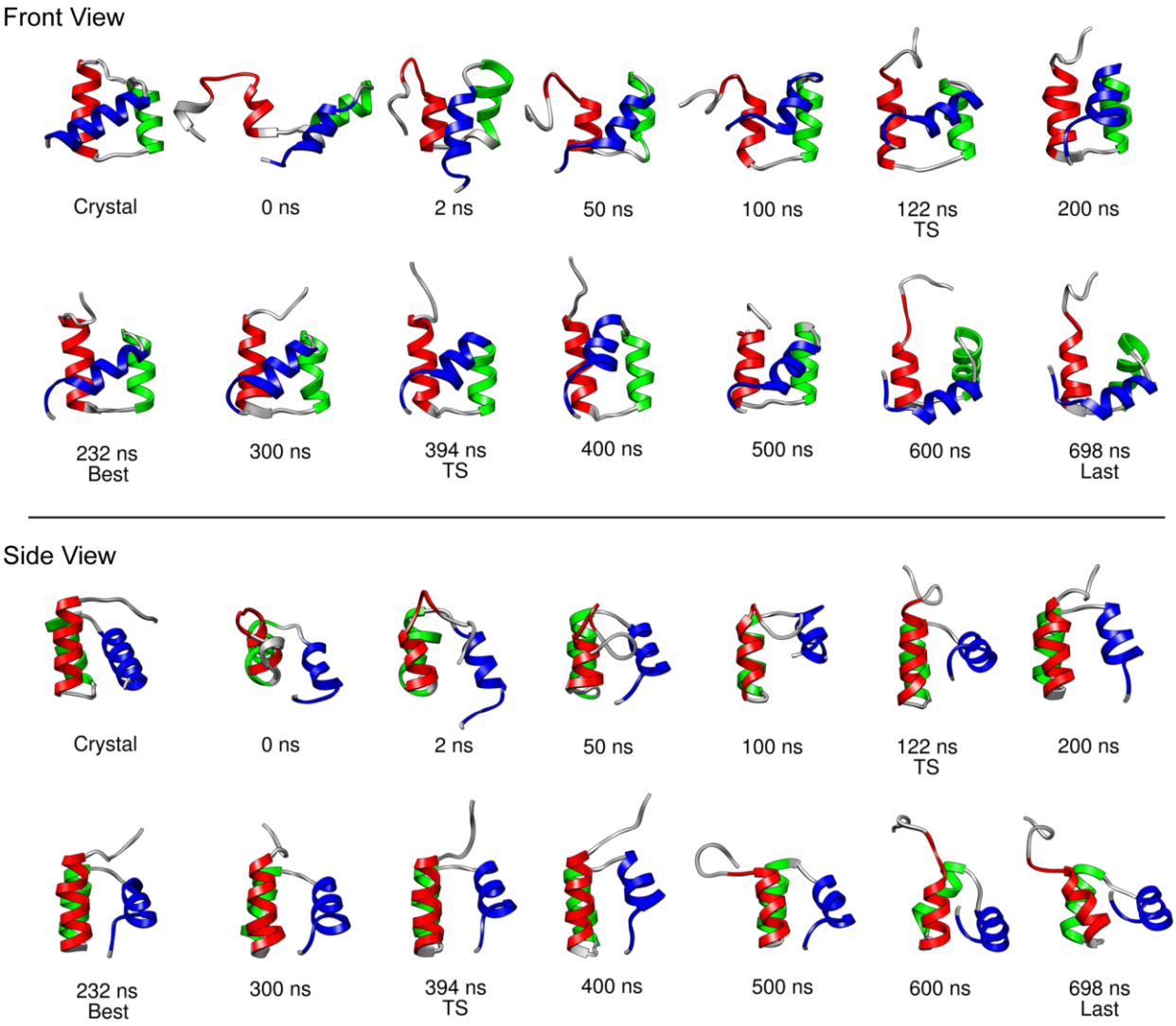
[‡] DSSP (8) was used to determine what fraction of the 54 residues was in α-helix.

[§] Center of mass was calculated over all heavy atoms for both residues listed, and the distance between the two centers of mass is reported.

[¶] SASA was calculated as specified over residues 8-53 or just Trp 48 using the Lee and Richards algorithm (9).

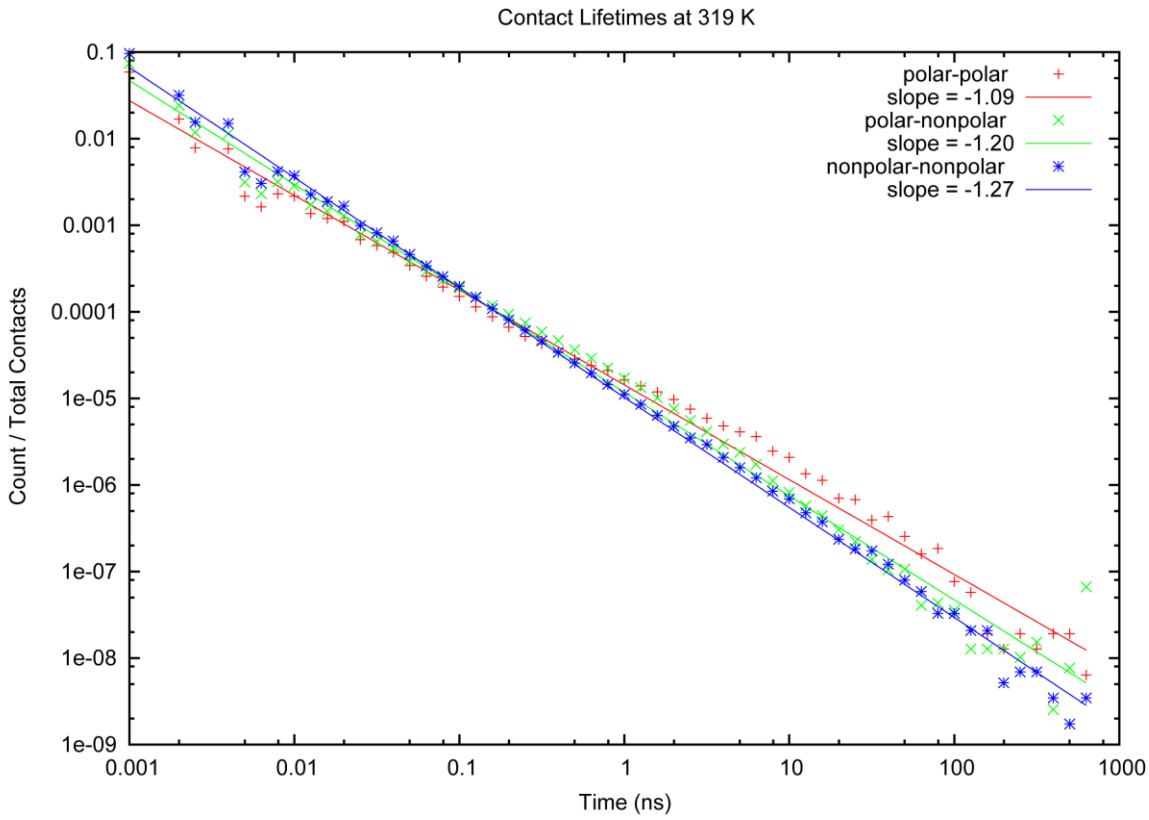
[‖] Two residues were considered to be in contact if the distance between at least one heavy atom from each was less than 5.4 Å for carbon/carbon pairs and 4.6 Å for all other pairs.

Fig. S4: Structures from the Successful Quench Simulation



Front and side views of structures from the quench simulation are shown every 100 ns as are structures from significant time points in the simulation. 2 ns: 4 HI-HII contacts formed; 100 ns: α-helix formed in the N-terminus and the final HI-HII contact formed; 122.210 ns: refolding TS; 232.260 ns: lowest mean distance to 319 K reference set (0.15); 394.596 ns: unfolding TS; 570 ns: HI-HII break apart; 698.600 ns: final structure from the simulation.

Fig. S5: Contact Lifetimes by Contact Type



Contact lifetimes for polar-polar (N=16,664,486), polar-nonpolar (N=55,431,889), and nonpolar-nonpolar (N=106,420,355) contacts for one 700-ns simulation at 319 K. A line was fitted to the data on a log-log plot, and the resulting slopes are indicated in the legend. More negative slopes indicate that shorter lifetimes dominate over longer, thus polar-polar contacts are longer-lived than nonpolar-nonpolar contacts at this temperature.