Supporting Material

Early Folding Events, Local Interactions, and Conservation of Protein Backbone Rigidity

Rita Pancsa,[1] Daniele Raimondi,[1] Elisa Cilia,[1] and Wim F. Vranken[1,]*

[1]Structural Biology Brussels, Vrije Universiteit Brussel, Brussels, Belgium

*Correspondence: wvvranken@vub.ac.be

Text S1: Composition of the early folding residues data set

The sequence lengths of the proteins in the **earlyFold** dataset range from 56 to 164 (Figure S1), with Ile, Leu, Pro and Arg under-represented, and Cys, Lys and Asn over-represented compared to SwissProt (Figure S2). The reported temperature range is from 0° to 30° C (except for the BPTI outlier at 70° C), with pH values from 3.0 to 8.0 due to a variety of experimental reasons (Table S1). Within each protein, the percentage of reported early folding residues is between 0.5% up to 45% (Figure S3) while ensuring that the experiments did not report on back-unfolding or aggregation (1). At the amino acid level, between 2-30% of residues are reported to be early folding, with a strong bias depending on amino acid type: the most likely residues are, in order of occurrence, Tyr, Phe, Trp, Val, Ile, Leu, Met and Cys, whereas the least likely are Gly, Asn, Asp, Ser and His (Figure S4). No data is available for Pro as it lacks an amide proton and cannot be detected in the experiments. Finally, the reported continuous fragments of early folding residues are very short (often a single residue) in comparison to the length of the fragments that connect them (Figure S5).

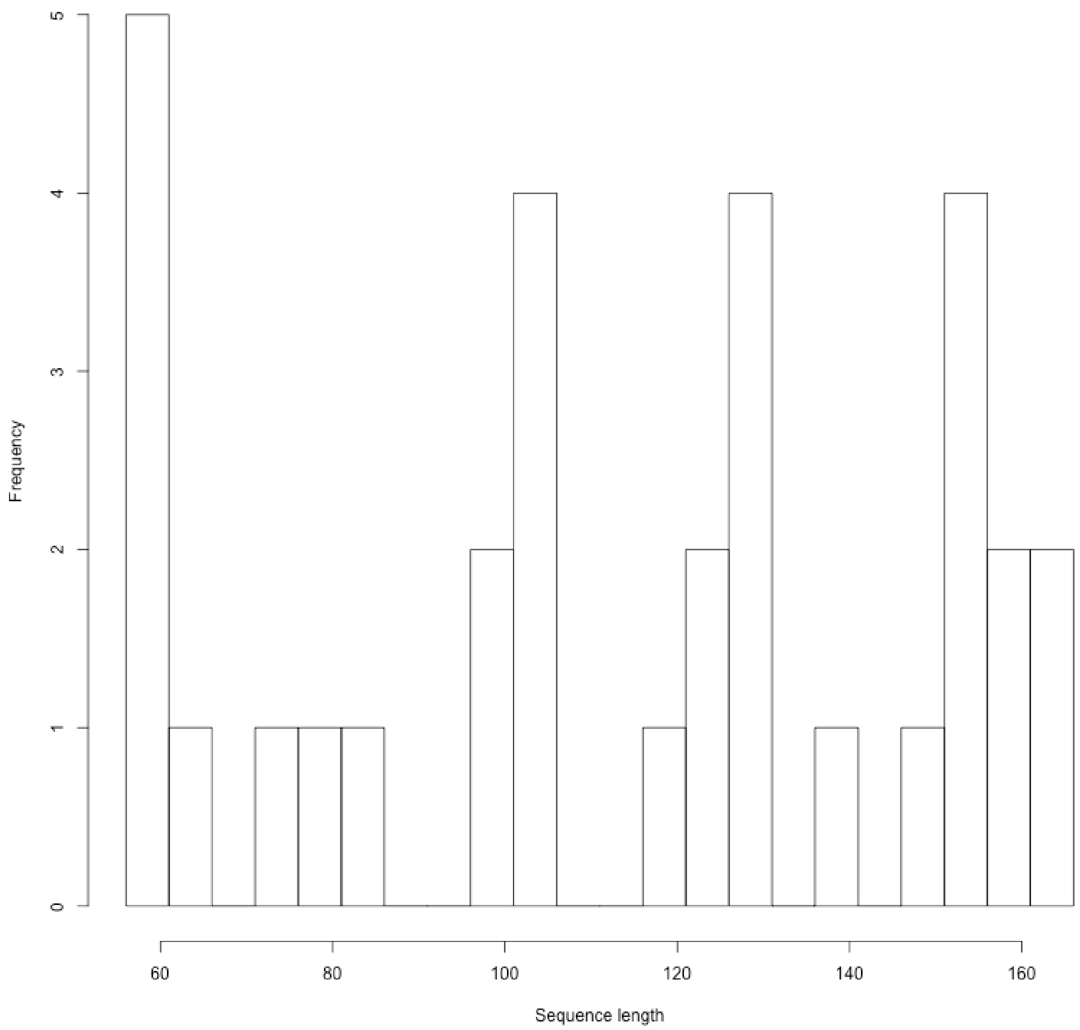

Figure S1: Sequence length distribution of proteins in the **earlyFold** dataset.

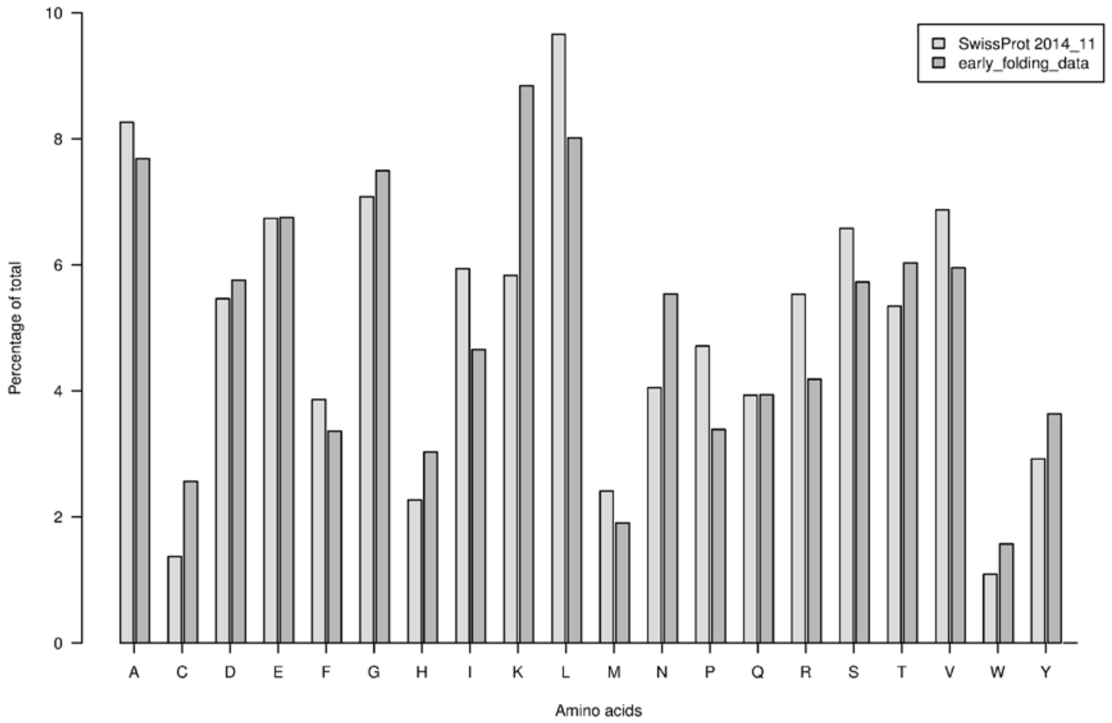

Figure S2: Amino acid distribution in the SwissProt database (November 2014) and the earlyFold dataset.

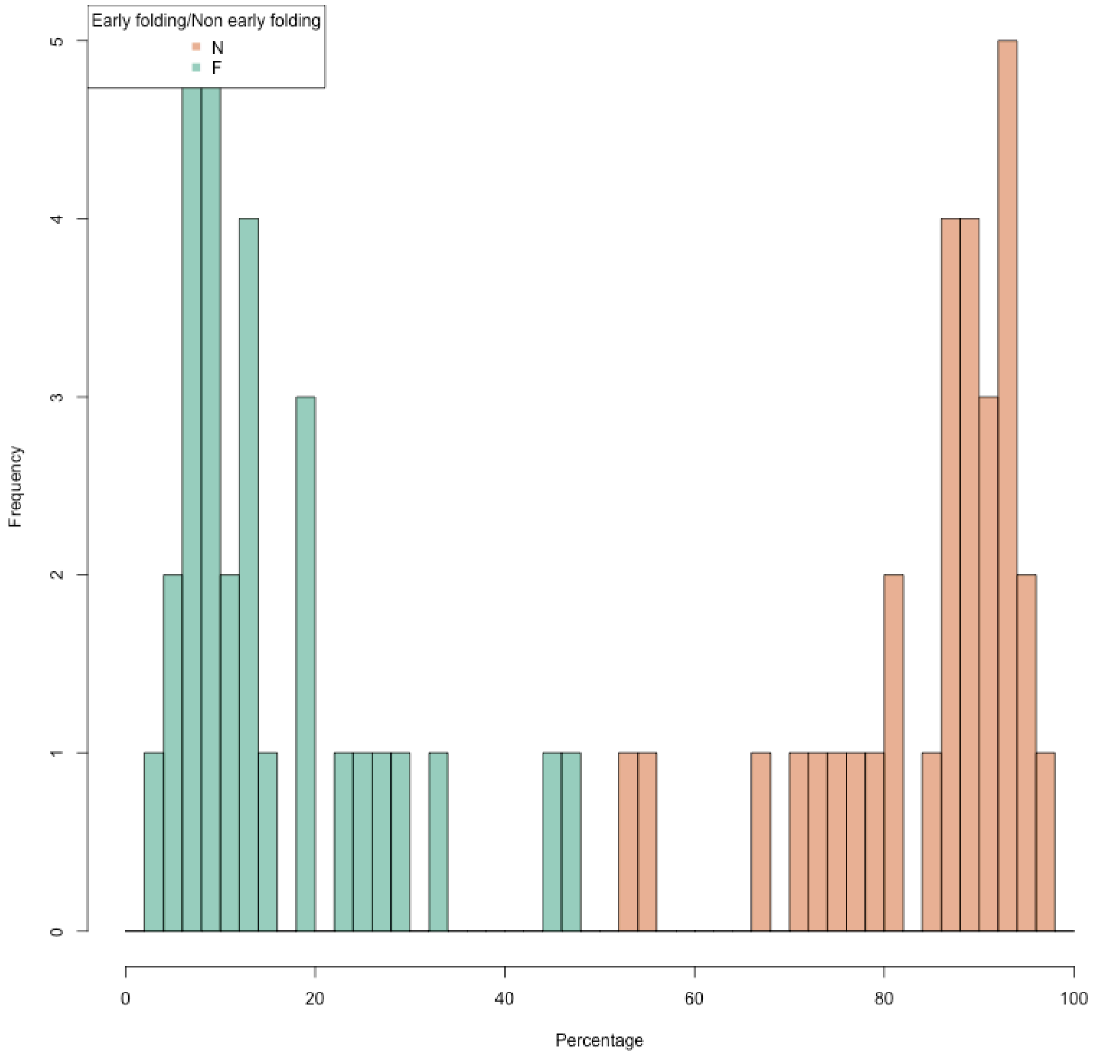

Figure S3: Per-protein percentages of reported early folding residues (F) and not early folding residues (N) in the **earlyFold** dataset.

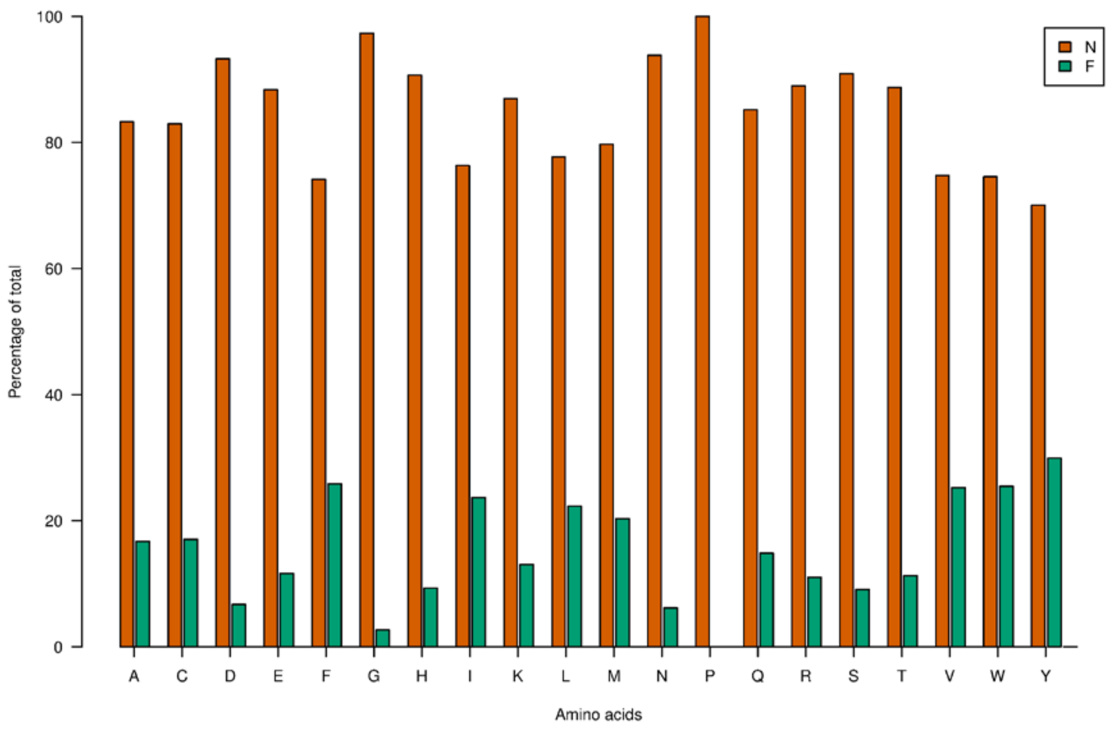

Figure S4: Per-amino acid frequencies of reported early folding residues (F) and not early folding residues (N) in the **earlyFold** dataset.

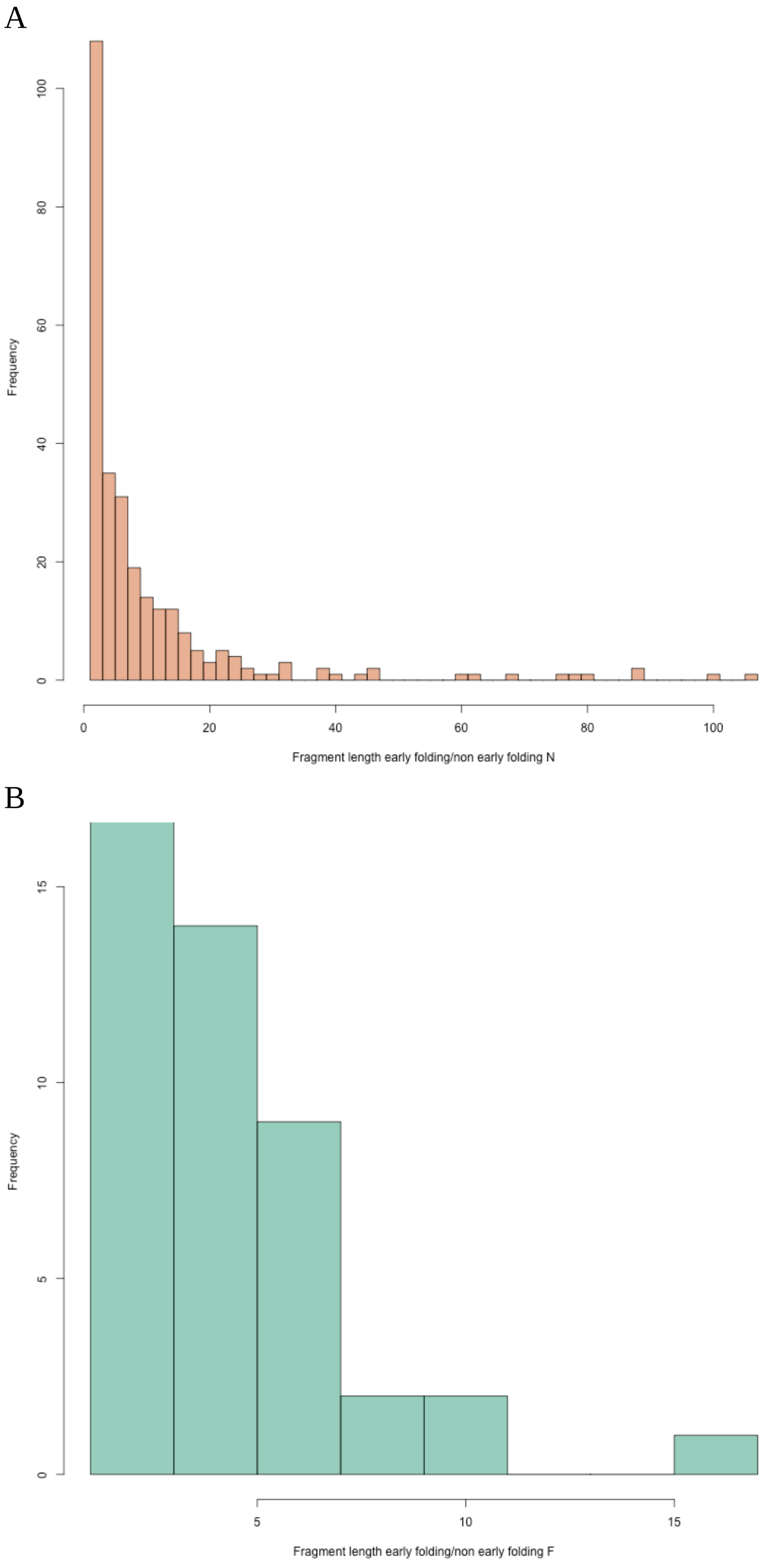

Figure S5: Length of continuous sequence fragments without early folding (A) and containing only early folding (B) residues.

Comparison between early folding and native exchange residues.

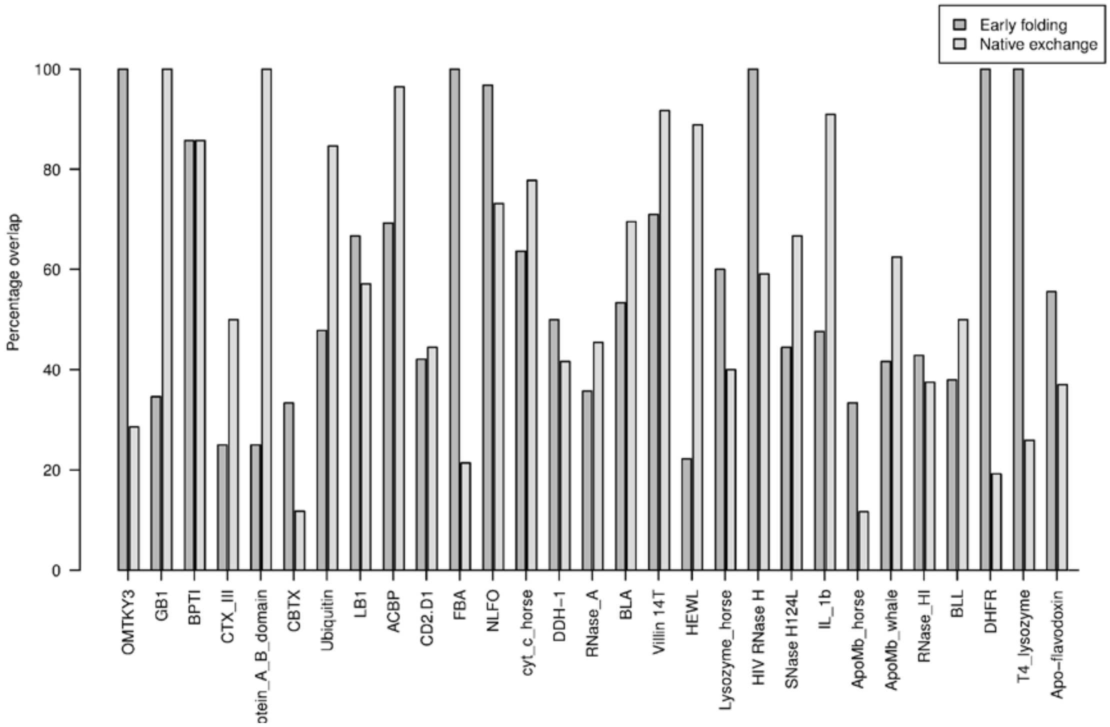

Figure S6: Overlap between early folding residues (from pulsed labelling HDX experiments) and residues that have strong resistance against unfolding (from native exchange HDX experiments) for 29 proteins from the Start2Fold database. For only 7 proteins the early folding residues fully or extensively (more than 80%) overlap with the residues that are very resistant to unfolding (OMTKY3, BTPI, apo-Pc (FBA), onconase (NLFO), HIV RNase H, DHFR and T4_lysozyme), but only BPTI, NLFO and HIV RNAse H have more than 30% reverse overlap. This shows that there is, as expected, some overlap between the pulsed labeling HDX dataset and the native exchange HDX dataset for each protein, but importantly that overall the datasets cover distinct residue sets. These data therefore cannot be simply equated to each other.

Additional per-amino acid distributions of backbone rigidity predictions

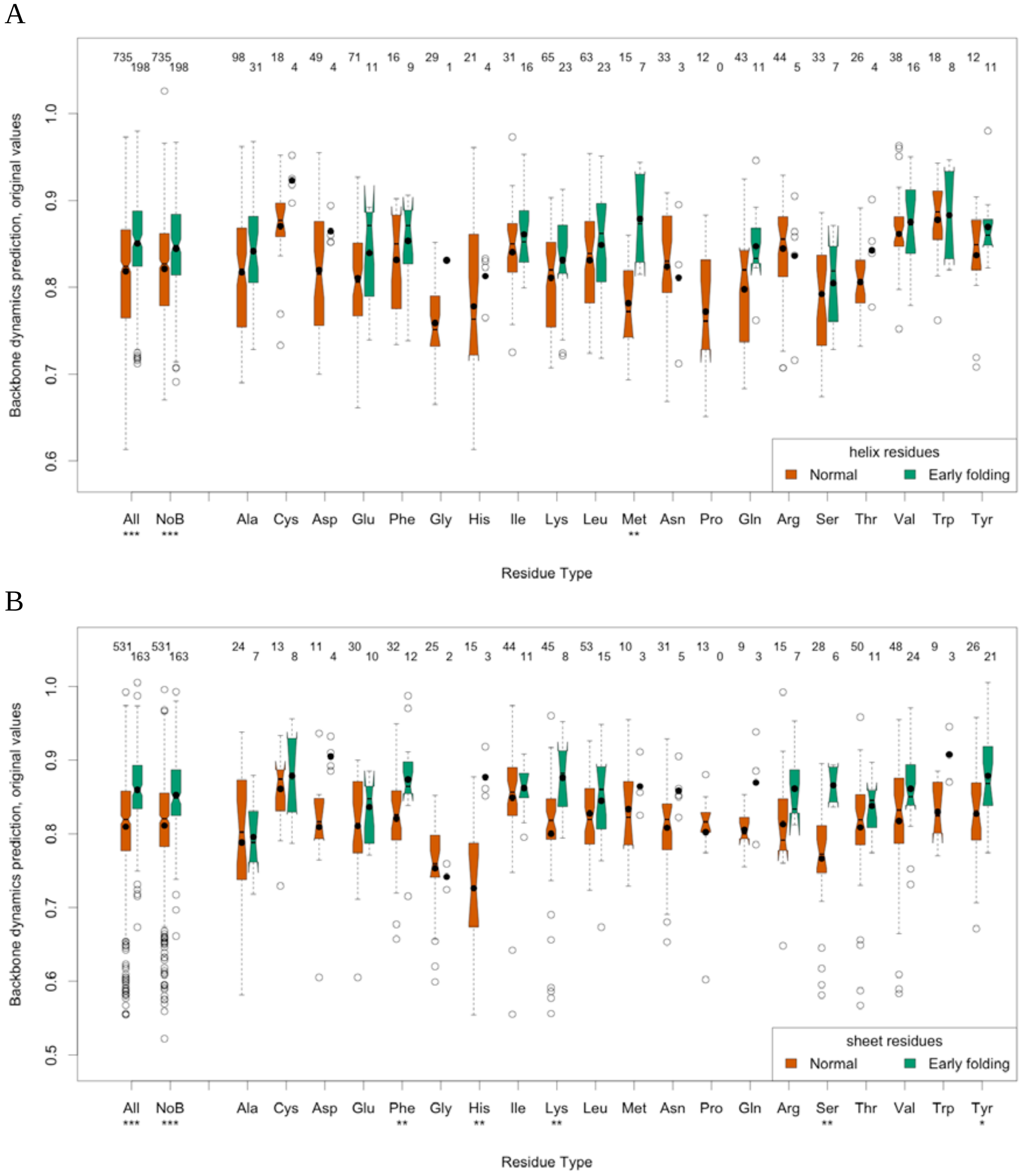

Figure S7: Per amino acid, all amino acid (All) and bias-corrected (NoB) distributions of predicted backbone rigidity for early folding and non-early folding residues. They are here subdivided by the secondary structure element as found in the related PDB structure with the native fold for helices (A) and beta sheets (B).

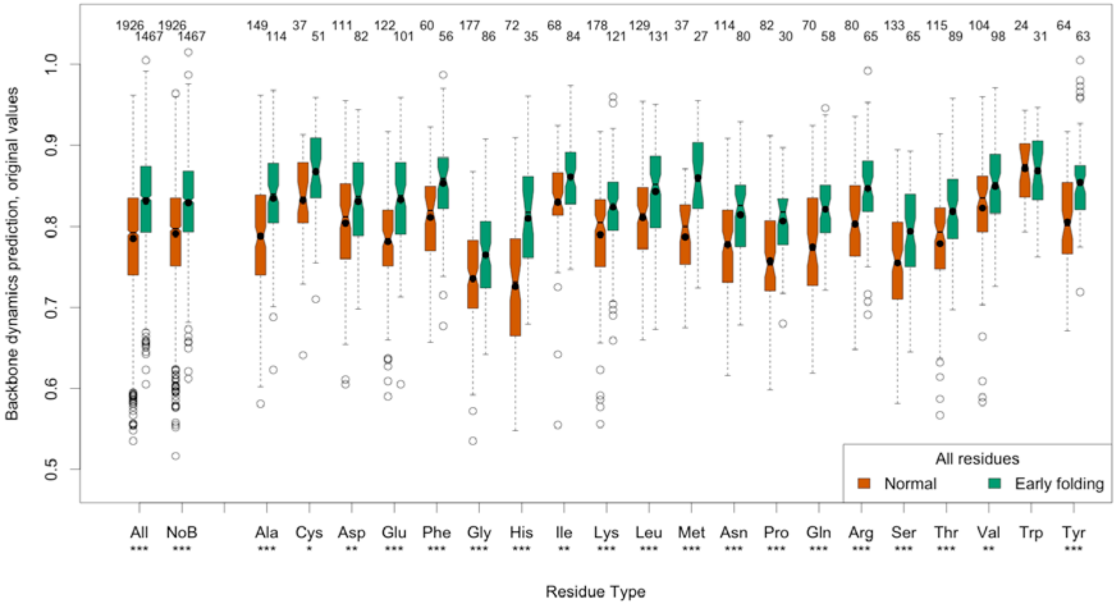

Figure S8: Per amino acid, all amino acid (All) and bias-corrected (NoB) distributions of the original predicted backbone rigidity for early folding and non-early folding residues in early folding fragments, where 3 residues preceding and following early folding residues were also included.

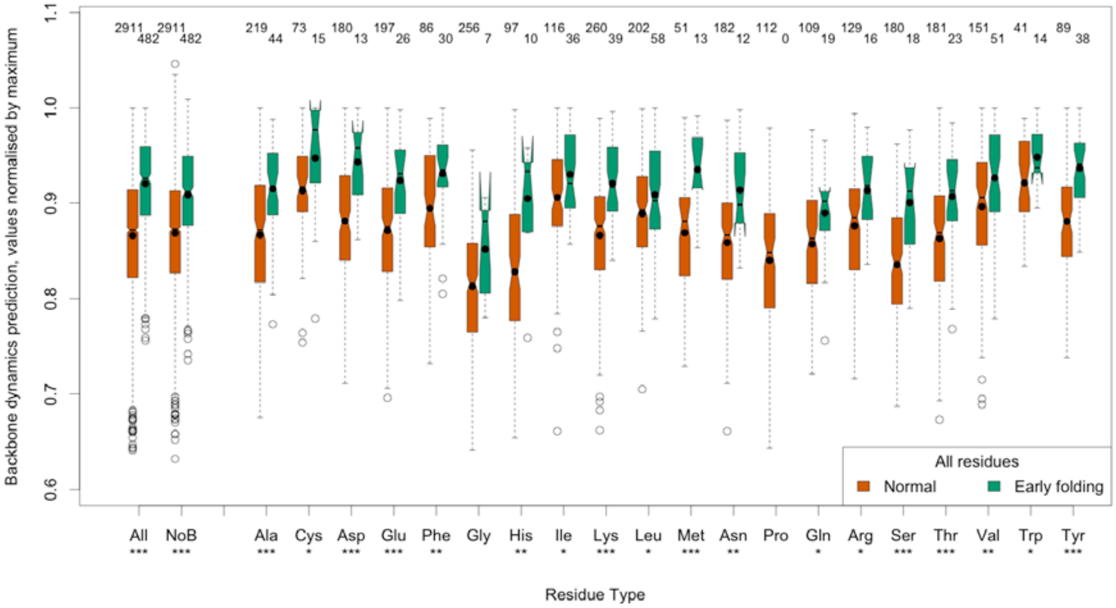

Figure S9. Per amino acid, all amino acid (All) and bias-corrected (NoB) distributions of the normalised predicted backbone rigidity for early folding and non-early folding residues.

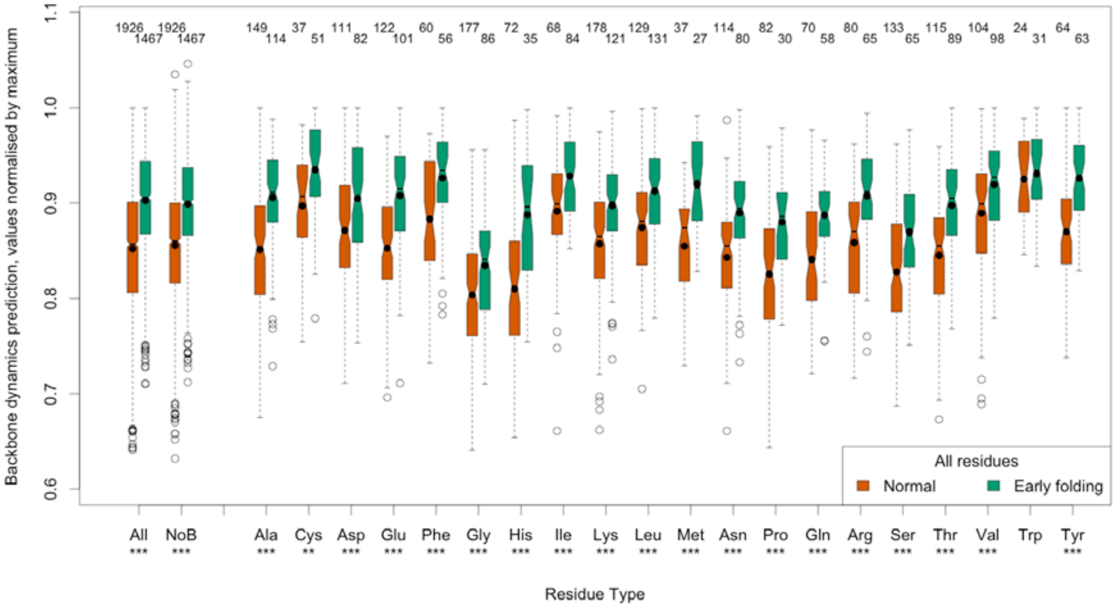

Figure S10: Per amino acid, all amino acid (All) and bias-corrected (NoB) distributions of the normalised predicted backbone rigidity for early folding and non-early folding residues in early folding fragments, where 3 residues preceding and following early folding residues were also included.

Text S2: Comparison with hydrophobicity, surface accessibility, order/disorder and secondary structure population.

The relative (rsa) and absolute (asa) solvent accessibilities were determined using the online NetSurfP server (2), which includes evolutionary information from PSI-BLAST in the form of Position Specific Scoring Matrix (PSSM). The order/disorder scores were predicted using the online ESpritz server (3) with both the versions trained on NMR and on X-ray data, and using default parameters. Residue hydrophobicity was calculated based on 22 different hydrophobicity scales using a linear 15 residue sliding window.

To investigate the link with hydrophobicity we employed 22 different hydrophobicity scales available from ProtScale (4) on the **earlyFold** sequences using a linear 15 residue window to calculate the central value. In only 1 case (Bull (5)) is there a significantly different distribution for the bias-corrected overall case (Figure S11). There is no significance at the amino acid level for any scale after employing the Benjamin-Hochberg correction. The local interactions required to form foldons therefore go beyond simple hydrophobic collapse (6, 7); this is not surprising as it is already surpassed as a folding model (8). Hydrophobicity values are also an amino acid property: specific interactions between amino acids are not taken into account, resulting in an oversimplified picture of hydrophobicity in a sequence context.

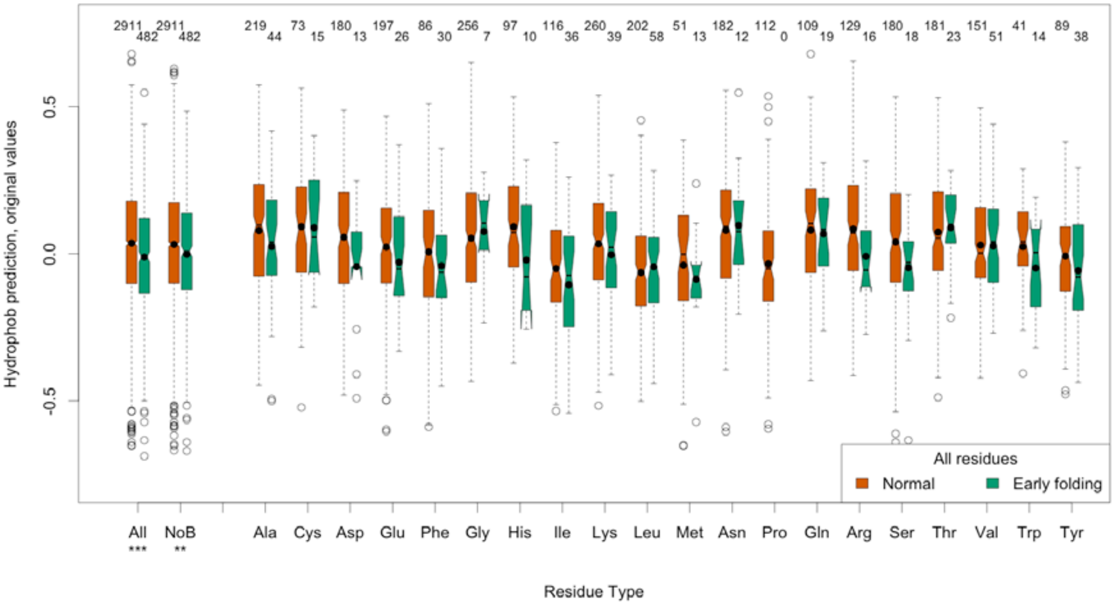

Figure S11: Per amino acid, all amino acid (All) and bias-corrected (NoB) distributions of hydrophobicity calculated for early folding and non-early folding residues using the Bull hydrophobicity scales.

Residues in the core of the protein, which are less exposed to solvent, tend to be involved in early folding events (9). The NetSurfP solvent accessibilities predictions (2), which include evolutionary information, show 7 highly significant, 2 very significant and 3 significant differences on a per-amino acid basis, totalling 12 residue types (Figure S12). The overall bias-corrected distributions also show a highly significant difference. Normalisation by minimum value, similar to the one performed on the DynaMine predictions, did not change the results.

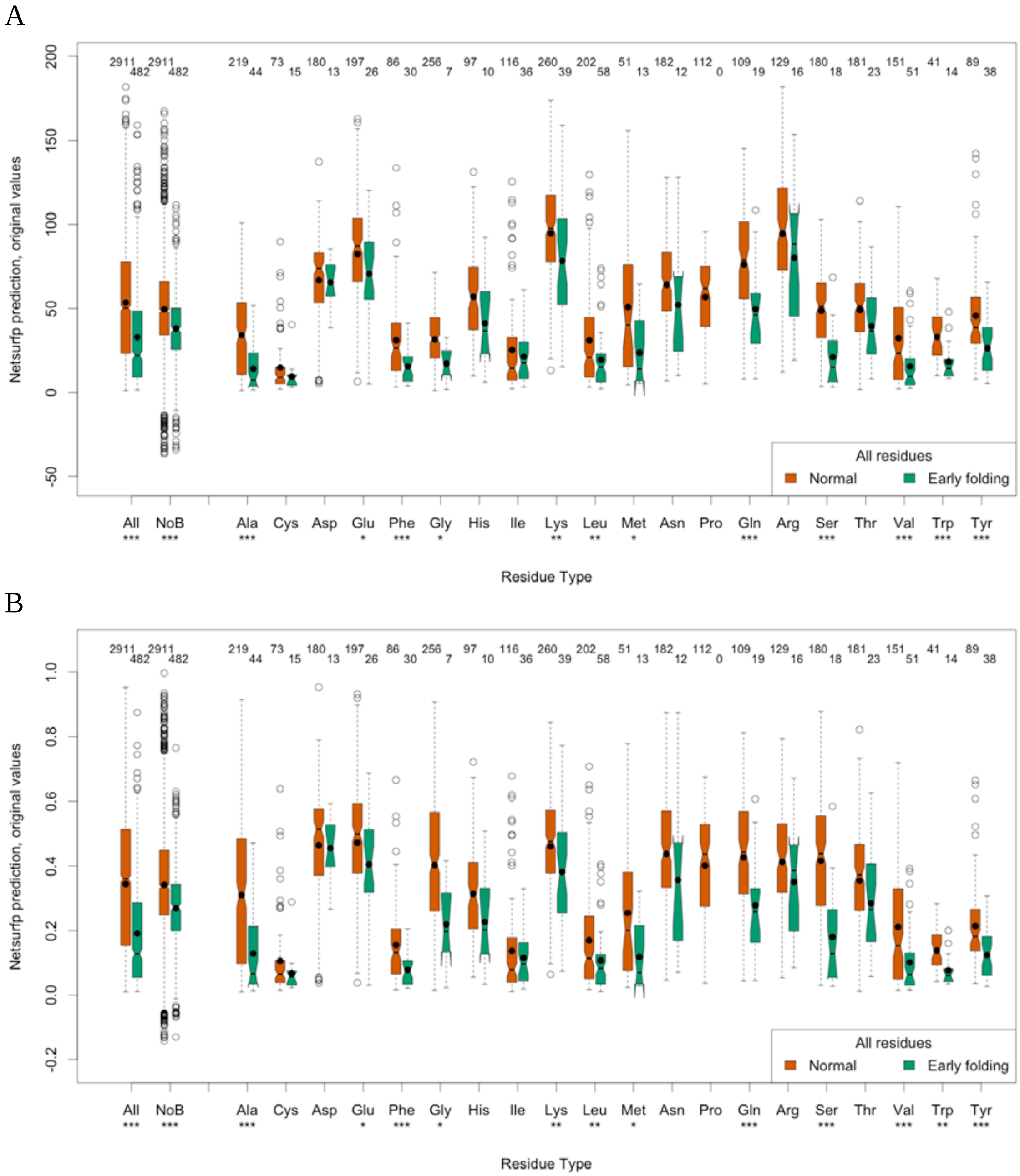

Figure S12: Per amino acid, all amino acid (All) and bias-corrected (NoB) distributions of NetSurfP predicted absolute (A) and relative (B) solvent accessibilities for early folding and non-early folding residues.

The DynaMine predictions already relate well to the order/disorder distinction (10), but we further analysed results from ESpritz (3), a sophisticated order/disorder predictor that was recently shown to be one of the top five predictors in the field (11), to investigate the connection of early folding with the tendency of the protein to be ordered (folded into a specific conformation) or disordered (dynamic and adopting multiple conformations). Both the ESpritz predictors based on X-ray and on NMR data perform well, with results slightly improving when shifted by minimum value to zero, similar to the maximum value correction we do here for DynaMine (Figure S13). The X-ray predictor performs worse than the NMR-based approach, providing significant differences for 8 amino acid types compared to 14. These are excellent results, and it is in this context important to note that the ESpritz-NMR predictor is based on the variability observed in NMR structure ensembles, while DynaMine is based on estimations of backbone dynamics directly from NMR chemical shift data for proteins in solution. Variability in NMR structures is related to lack of meaningful restraints in the structure calculation, which can be due to dynamics but also, for example, extensive signal overlap. The better performance of DynaMine in detecting early folding residues shows that NMR chemical shift data more accurately probe the residue-level behaviour of proteins in solution by also covering fast transient unfolding and proteins without a well-defined fold.

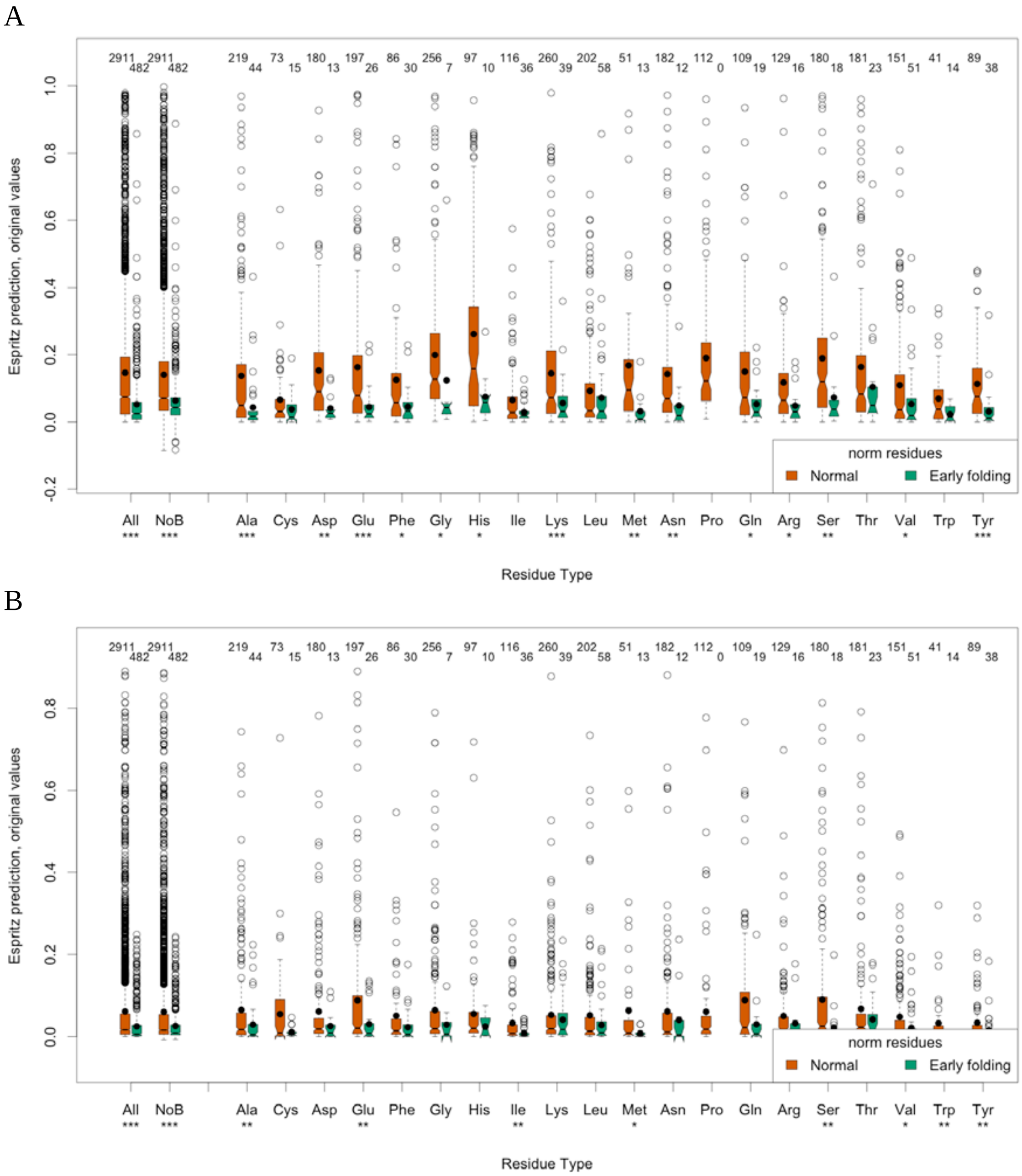

Figure S13: Per amino acid, all amino acid (All) and bias-corrected (NoB) distributions of predicted disorder tendency for early folding and non-early folding residues with the normalised NMR (A) and X-ray (B) ESpritz predictions.

The s2D sequence-based predictions of secondary structure population (12) show only 2 very significant differences on a per-amino acid basis for helix, with the overall bias-corrected distributions showing a highly significant difference (Figure S14A). No significance is present for sheet (data not shown), while the coil predictions show highly significantly lower populations for Glu, Lys and Met, and very significantly lower ones for Ala, His and Ser (Figure S14B). This shows there is a relation between conformational preference and early folding, but that it is not in itself a key characteristic.

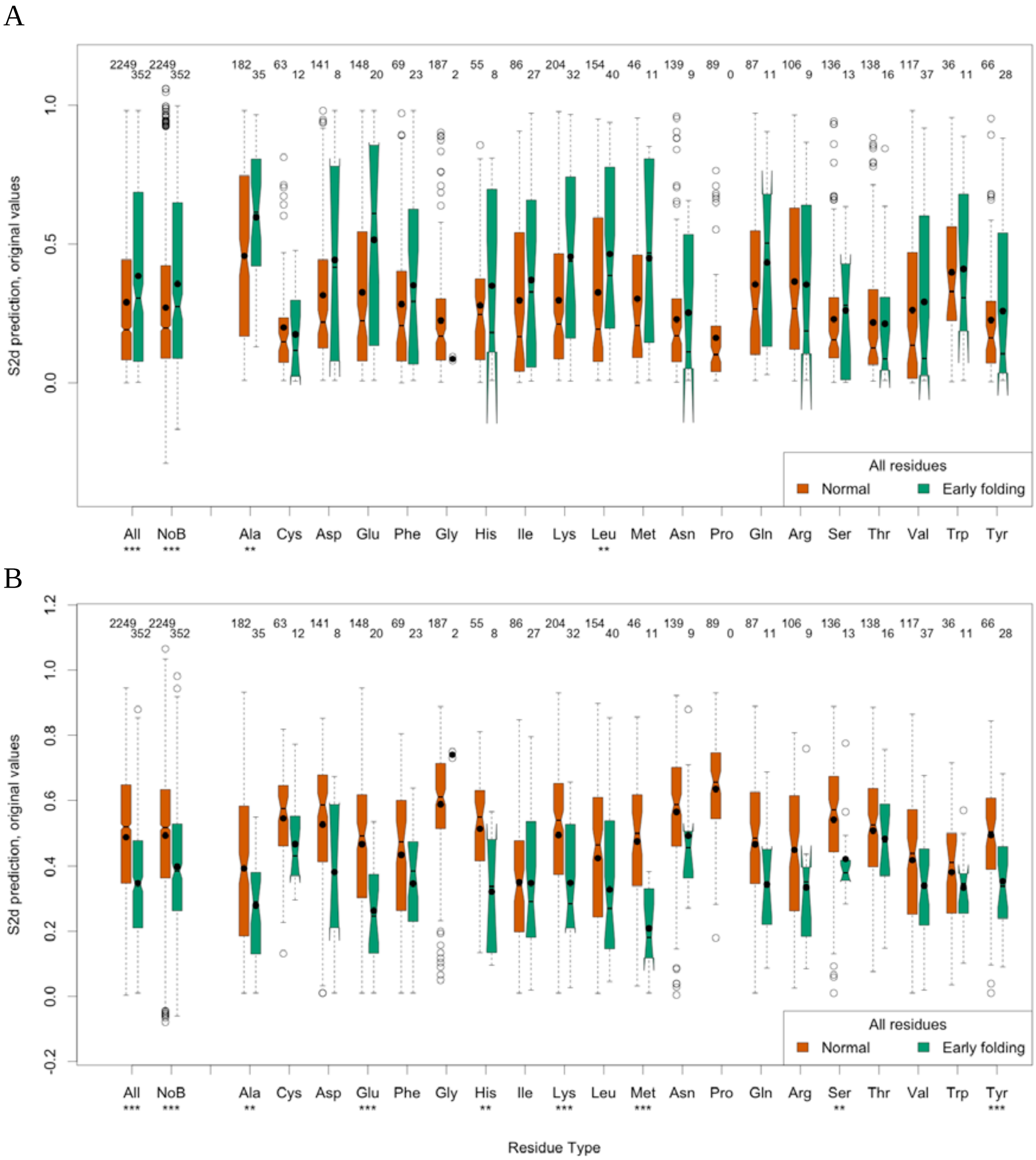

Figure S14: Per amino acid, all amino acid (All) and bias-corrected (NoB) distributions of predicted disorder tendency for early folding and non-early folding residues with the s2D helix (A) and coil (B) predictions

Text S3: Comparison with structure-based parameters.

For the 28 proteins where PDB entries were available, we calculated the relative (rsa) solvent accessibilities using DSSP (13), and the contact S^2 parameter, which is predictive of the backbone dynamics of the protein (14).The DSSP rsa is overall significantly lower for early folding residues, also after correcting for bias, and on an amino acid level is highly significant for Ala, Leu and Ser, very significant for Lys, Gln and Val, and significant for Glu, Phe and Tyr, totalling 9 residues (Figure S15A). The contact S^2 is a much better indicator, with the distributions highly significantly different for 7 residues, very significant for 6 residues and significant for another 2 residues, totalling 15 (Figure S15B). The structure-based contact S^2 therefore has a performance similar to the sequence-based normalised DynaMine values (Table 1). Since the contact S^2 reflects the amount of heavy atoms close to the backbone amide H and the carbonyl oxygen of the preceding residue, it takes all atoms into account in the folded protein, not just local interactions, and indicates that the early folding residues tend to become the residues with the most and closest backbone interactions in the folded protein.

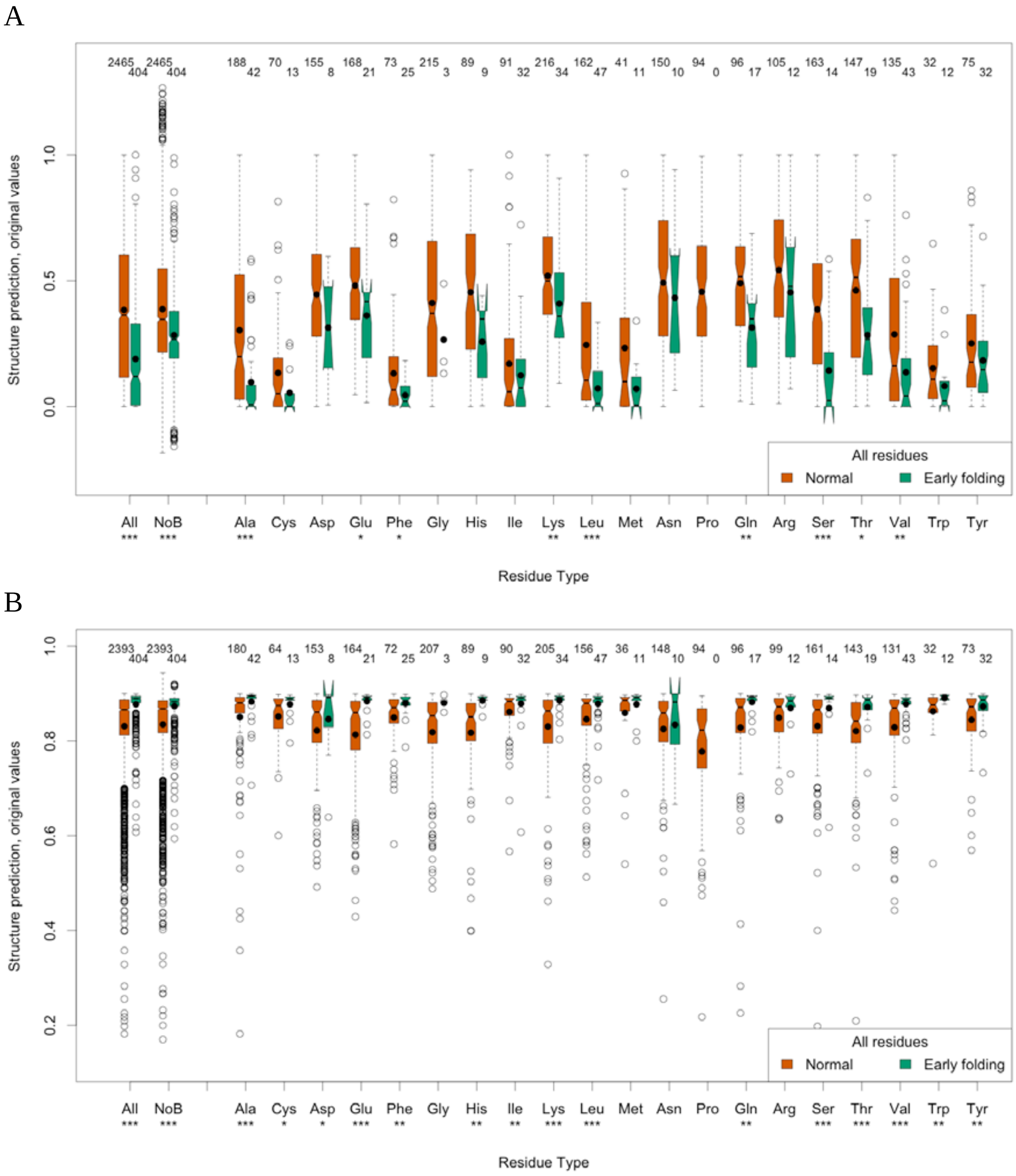

Figure S15: Per amino acid, all amino acid (All) and bias-corrected (NoB) distributions of the PDB structure derived DSSP relative solvent accessibility (A) and the contact S^2 parameter (B).

Additional per-amino acid secondary structure based distributions.

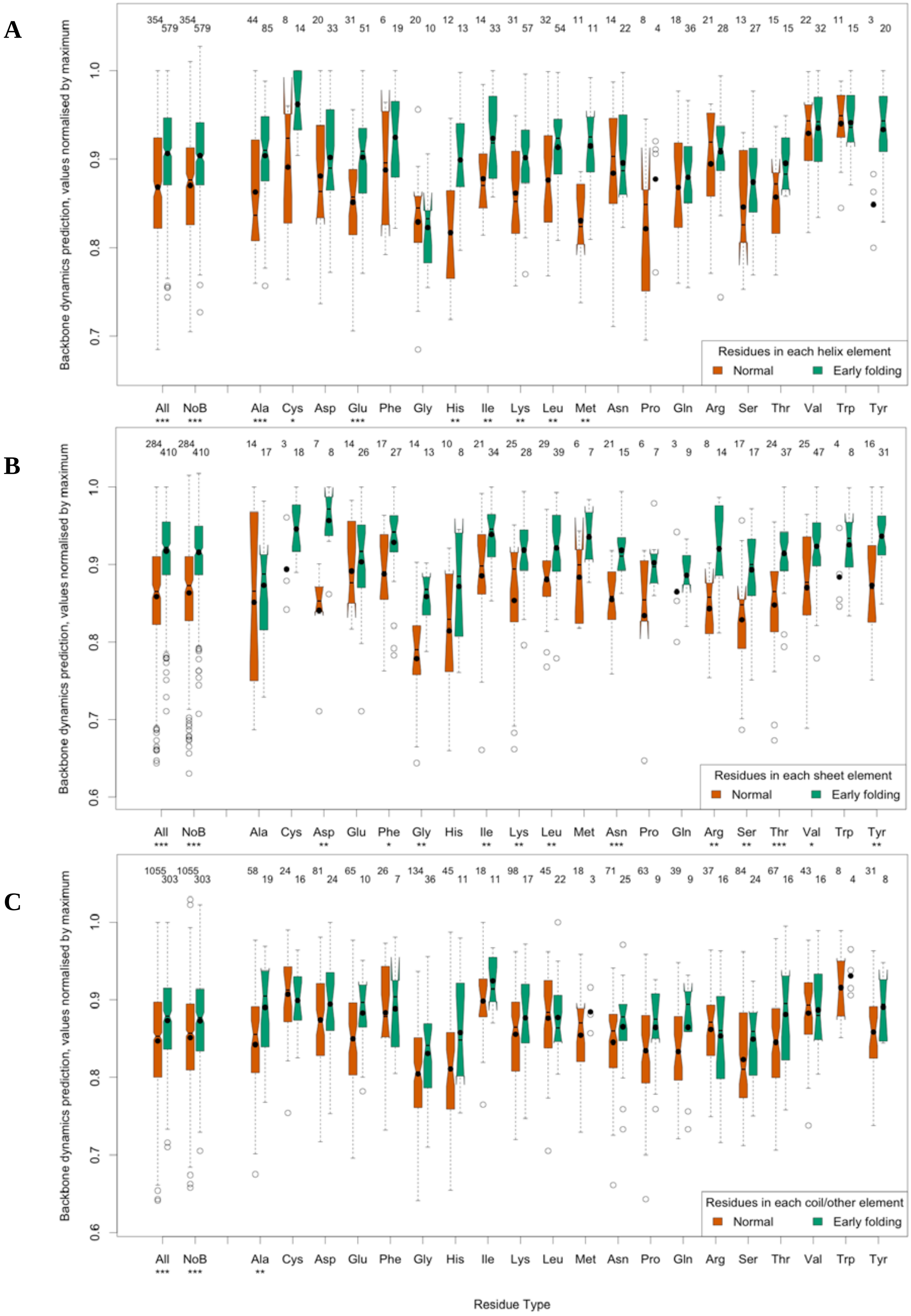

Figure S16: Boxplots showing the distribution per amino acid residue of the normalised predicted backbone rigidity divided by their absence or presence in α-helices (A) and β-sheets (B) and coil/other (C) that contain early folding residues. The number of amino acids in each distribution is indicated at the top of each graph, while the significance of the difference between the distributions is reported under the amino acid three-letter code.

Correlation between SSE length and average backbone rigidity.

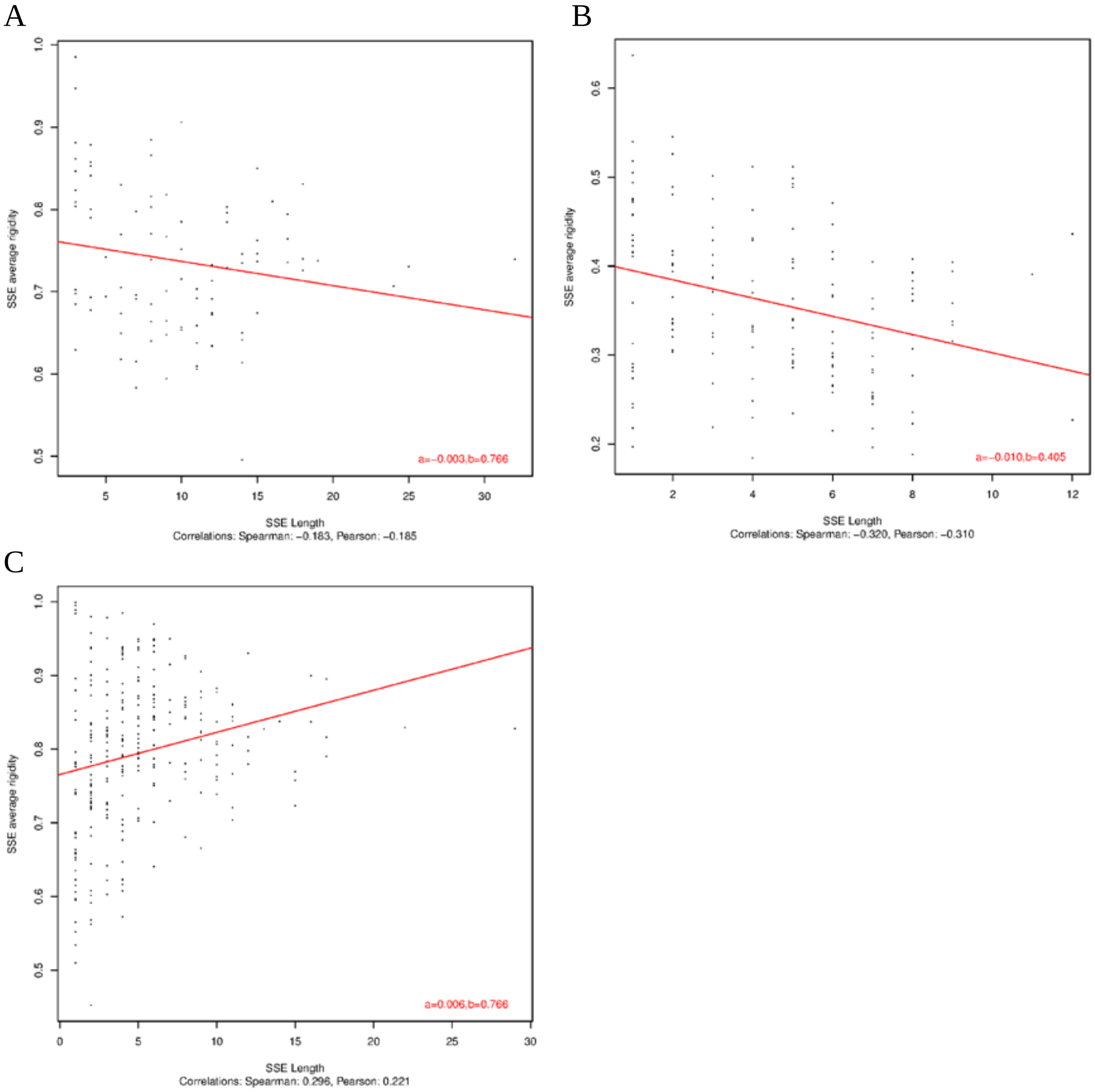

Figure S17: Scatter plots showing the relation between SSE length and the average of the normalised backbone rigidity prediction for all residues in that SSE for α-helices (A), β-sheets (B) and coil/other (C). The slope (a) and intercept (b) of the line of best fit are indicated in red, the Spearman and Pearson correlations are indicated underneath each plot.

Relation of backbone rigidity predictions to (reduced) sequence entropy.

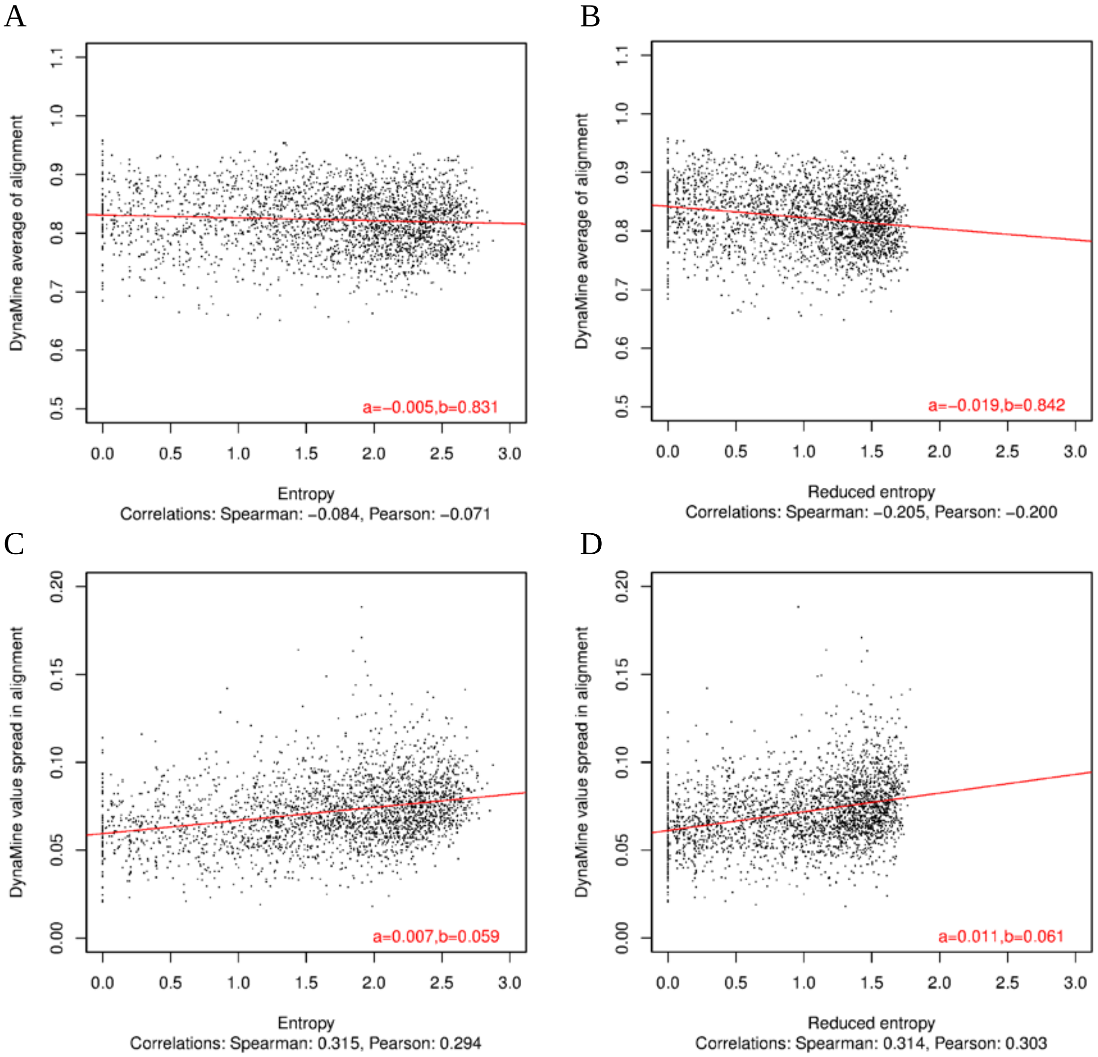

Figure S18: Scatter plots showing the relation between median value (A,B) and spread of values (C,D) of the MSAs with entropy (A,C) and reduced entropy (B,D). The slope (a) and intercept (b) of the line of best fit are indicated in red, the Spearman and Pearson correlations are indicated underneath each plot.

Per-amino acid distributions of entropy and reduced entropy.

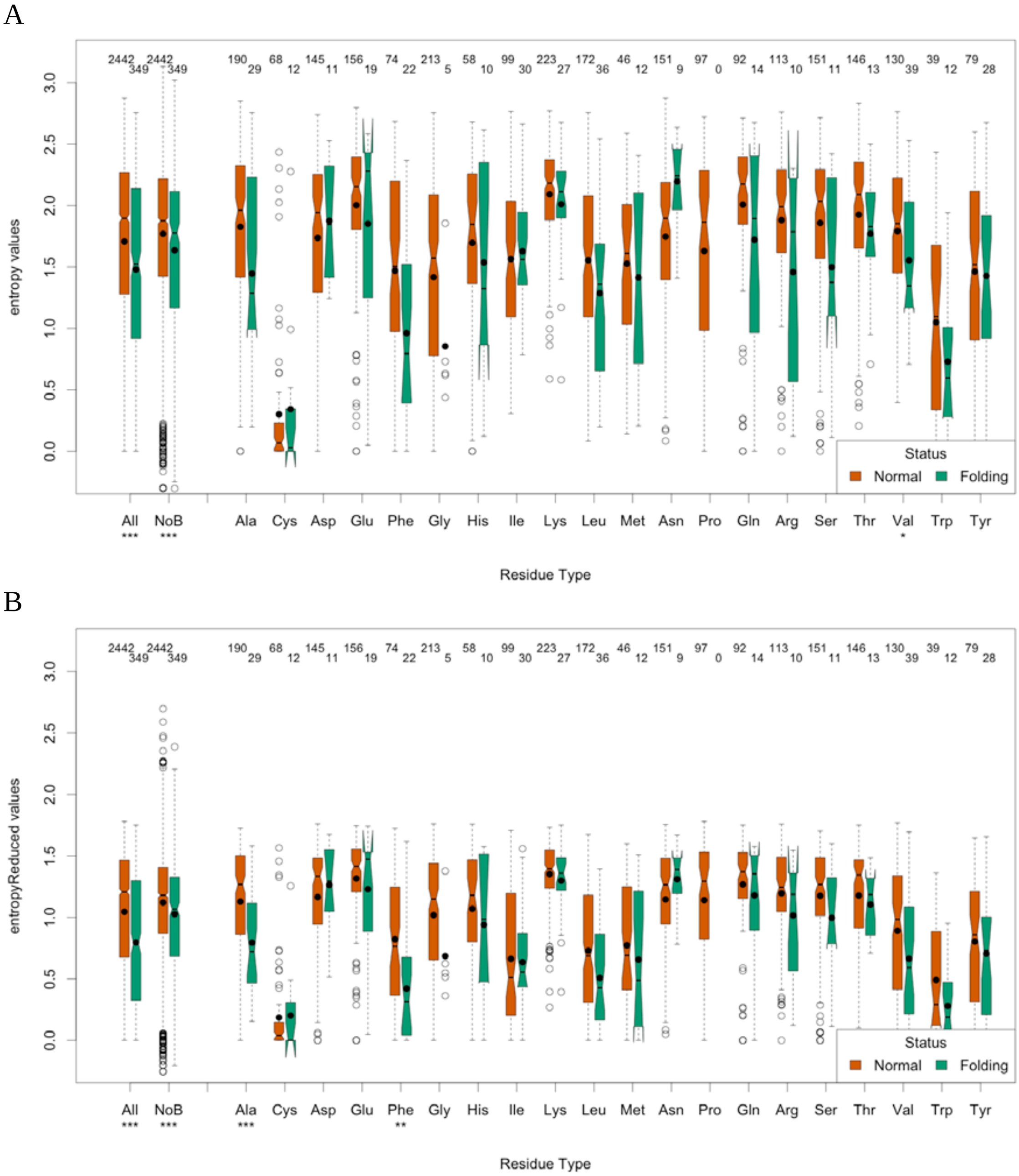

Figure S19: Entropy (A) and reduced entropy (B) distributions per amino acid, over all amino acids and for the bias-corrected set (NoB) in the **HHBLITS_lowSeqId** dataset.

Predicted median backbone rigidity per MSA column divided by SSE.

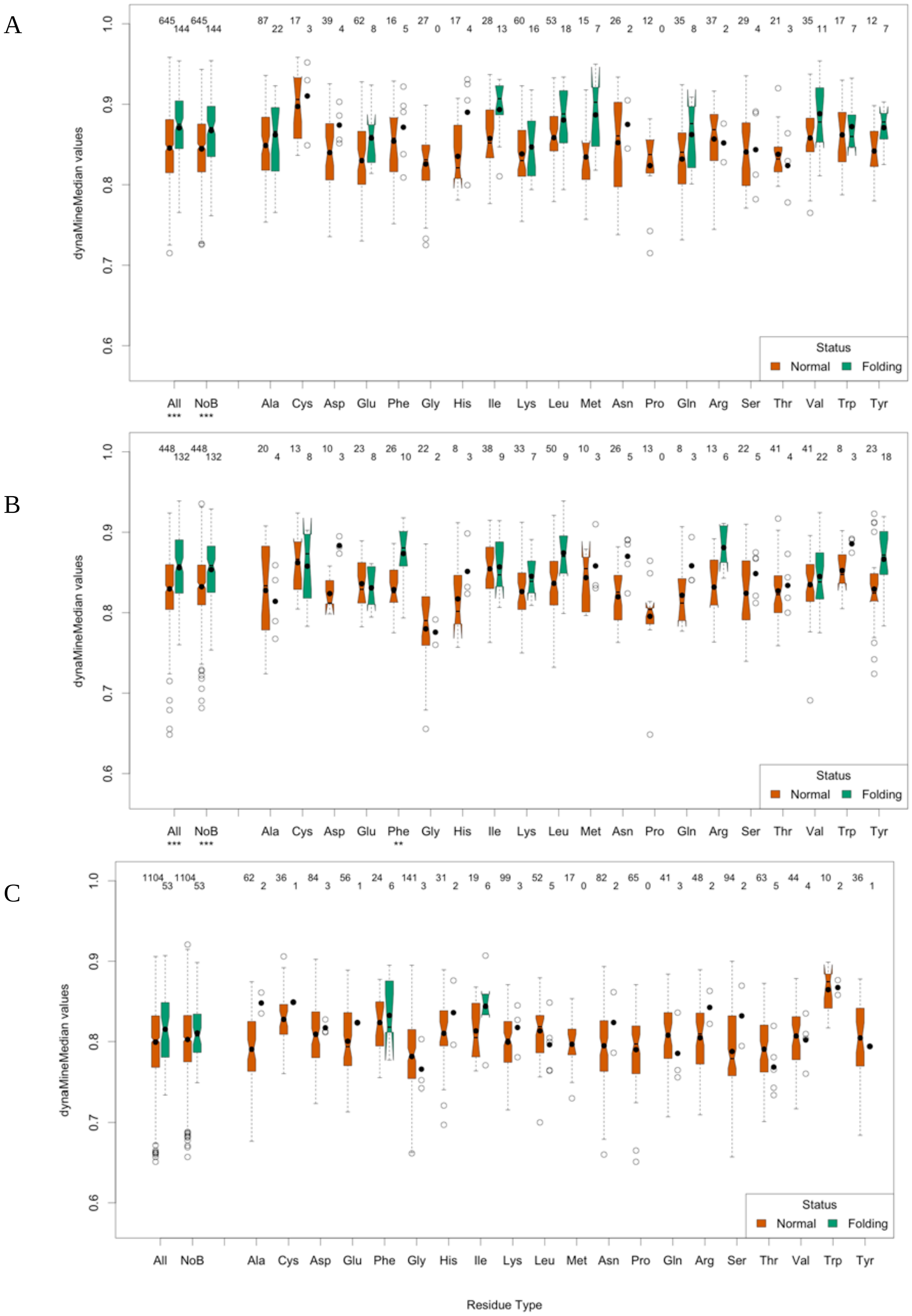

Figure S20: Distribution of median values of the predicted backbone rigidity in the **HHBLITS_lowSeqId** dataset for early folding versus other residues subdivided by helix (A), sheet (B) and coil/other (C) secondary structure elements as observed in the native fold. The distribution is shown per amino acid, over all amino acids and for the bias-corrected values (NoB).

Predicted median backbone rigidity per MSA column by early folding SSE.

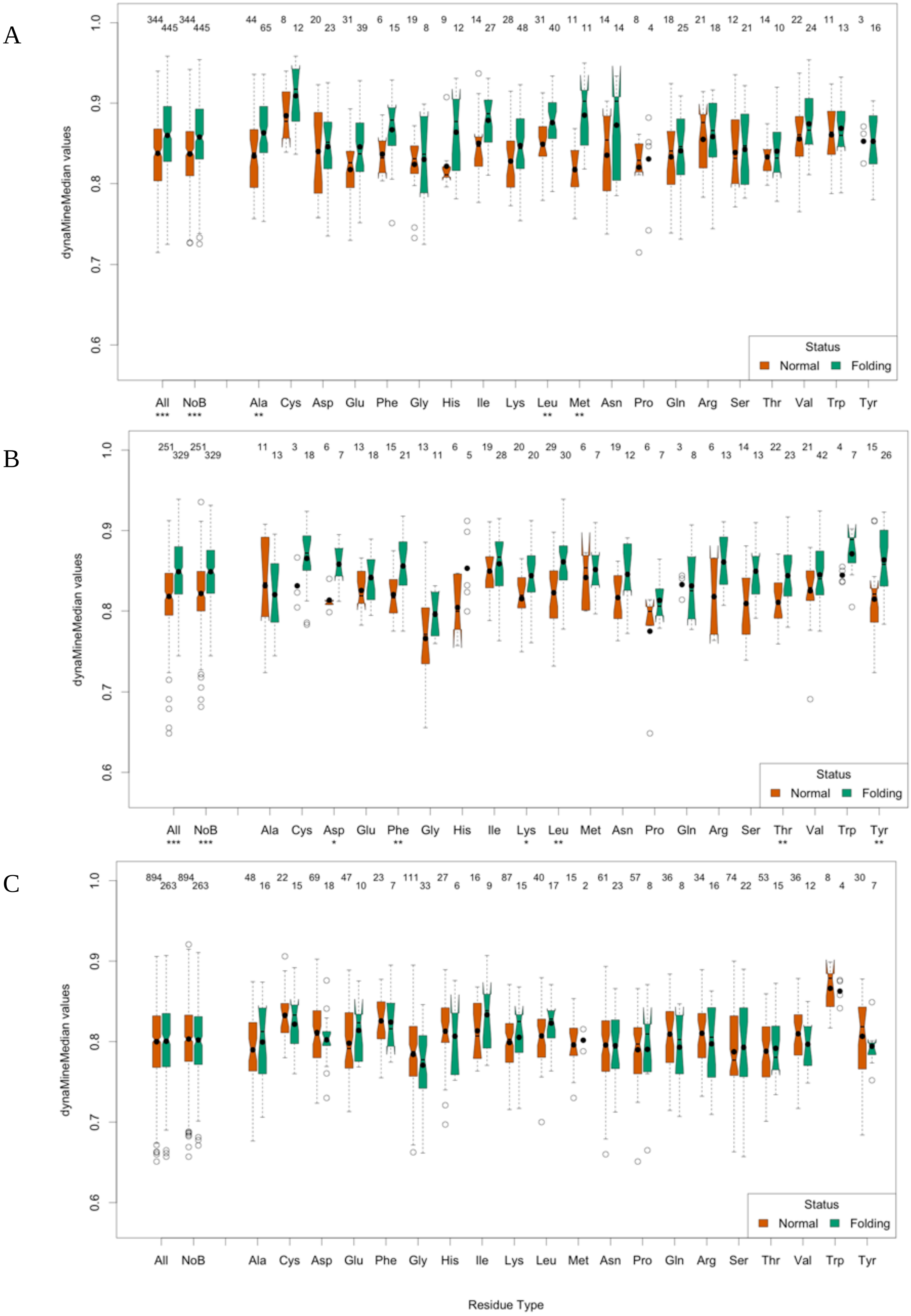

Figure S21: Distribution of median values of the predicted backbone rigidity in the **HHBLITS_lowSeqId** dataset for residues that are part of early folding helix (A), sheet (B) and coil/other (C) secondary structure elements, and ones that are not. The distribution is shown per amino acid, over all amino acids and for the bias-corrected values (NoB).

Distributions of average over secondary structure element of the predicted median backbone rigidity per MSA column.

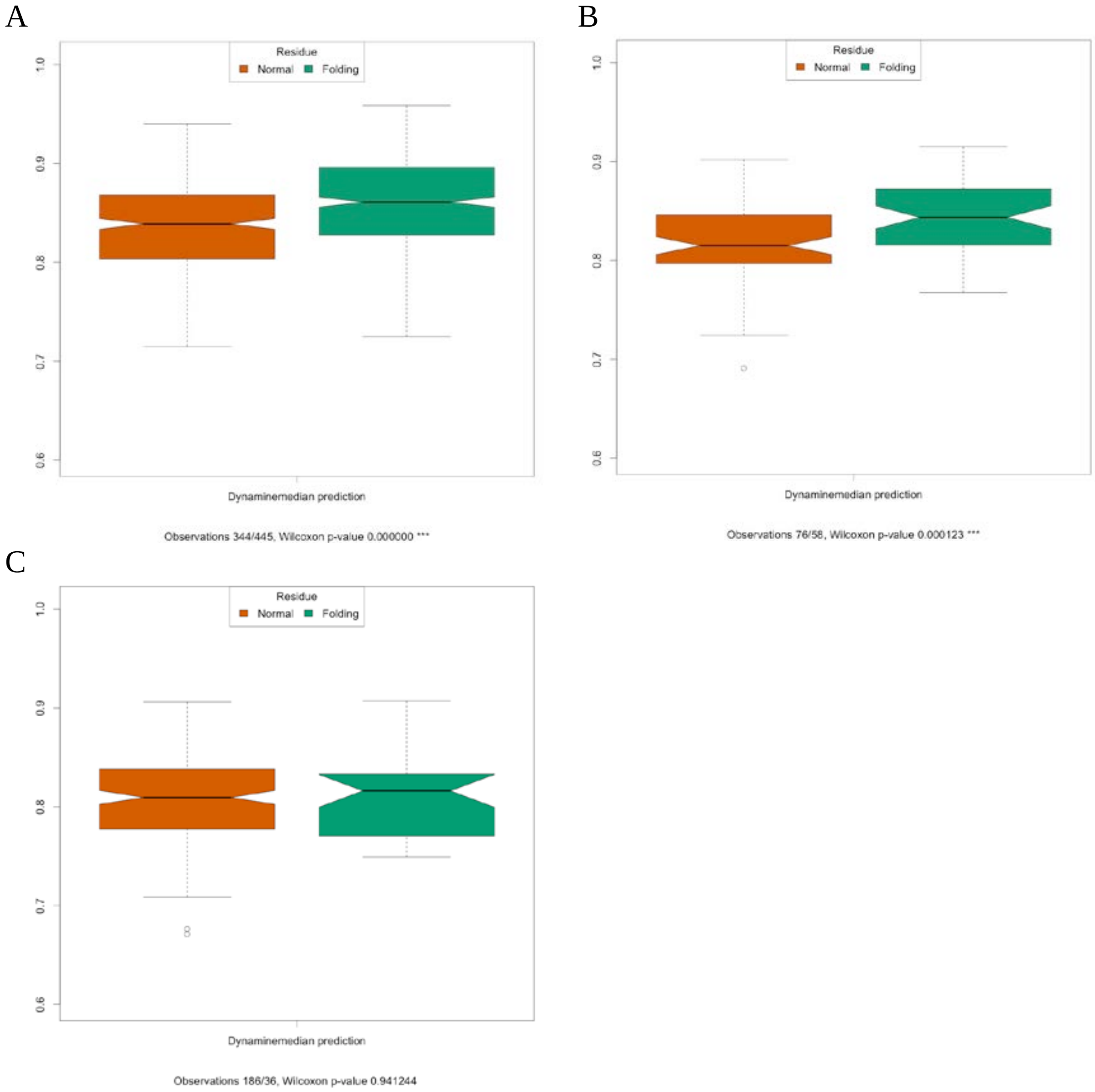

Figure S22: Distribution of the average value of the median predicted backbone rigidity in the **HHBLITS_lowSeqId** dataset over secondary structure elements as observed in the native fold. The early folding versus normal secondary structure are shown for helix (A), sheet (B) and coil/other (C).

Additional case study.

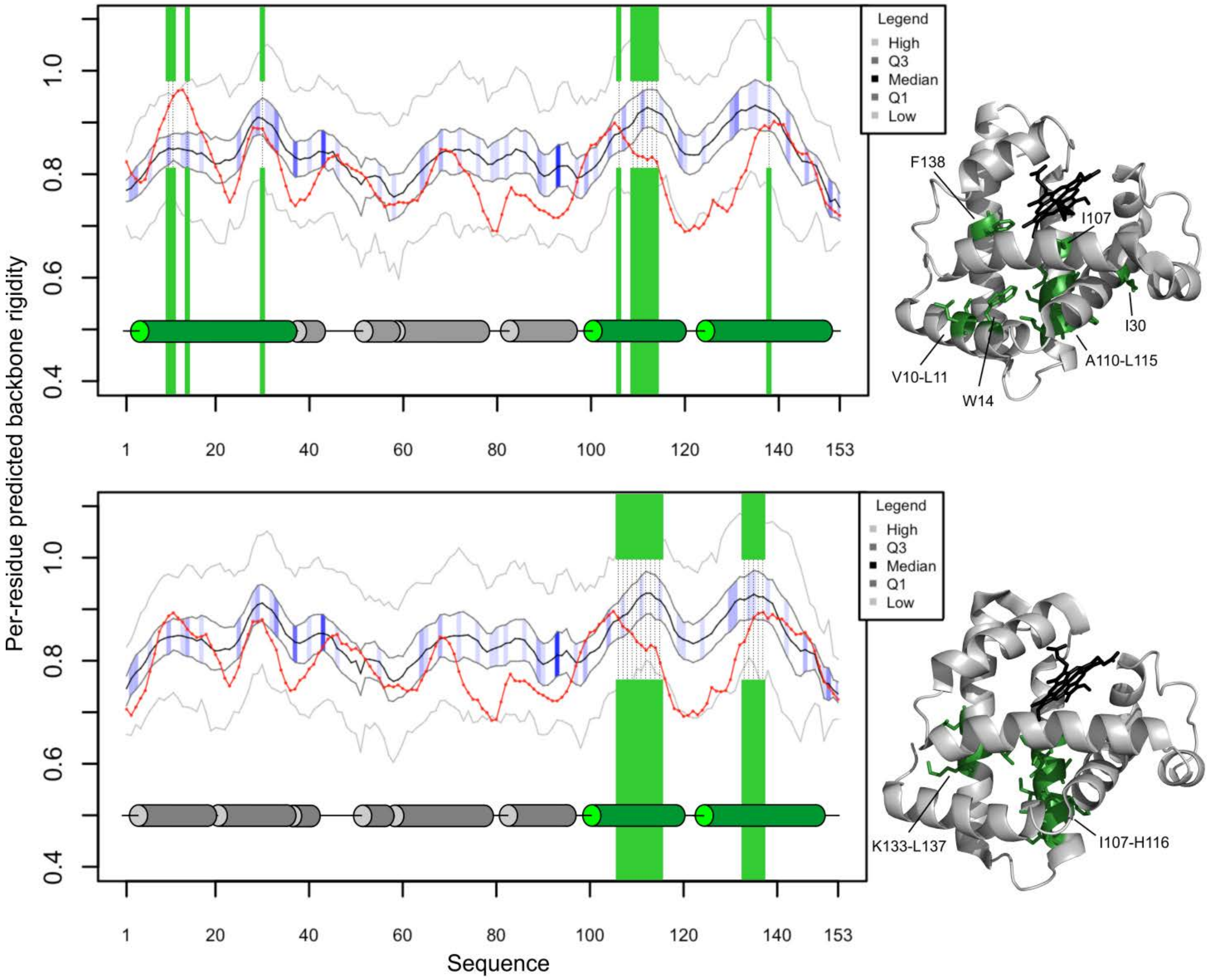

Figure S23. Dynamics and evolutionary properties of early folding residues compared between myoglobins from sperm whale and horse.

The structural, dynamics and evolutionary properties of sperm whale apo-Mb (top) and horse apo-Mb (bottom) are shown as a function of their residue positions on the left, while the corresponding 3D structures are on the right (there are no structures available for the apo forms, thus the ones with the heme cofactor (in black) are shown; PDB IDs are 1mbc and 1ymb, respectively). Early folding residues are marked with green shading on the graphs and with green stick representations within the 3D structures, with their residue positions and types indicated. The per-residue DynaMine-predicted backbone rigidity is depicted by a red line. The medians of predicted values in the corresponding **HHBLITS_lowSeqId** alignment columns are shown as a black line, while their first and third quartiles are marked in dark grey and their minima and maxima with lighter grey. The blue shading between the quartile lines represents the sequence entropy for each alignment position, with darker blue indicating lower entropy (high evolutionary conservation). The secondary structure elements assigned by the Polyview server are also provided, with early folding helixes shown as green cylinders and others as grey cylinders.

Table S1: Information on proteins in the earlyFold dataset.[§]

Protein[$]	Method[¥]	PDB	FT	SST	NR	NP	pH[£]	T	Protection threshold	MSA[*]	Ref.	PL
ACBP (bovine)	QF HDX NMR	2abd	2	all-a	39	82	5.3	5	protection rate (s-1) ~20±5	226	(15)	8.5
apo-Mb (horse)	PL HDX MS	1ymb	3	all-a	15	all	2.0→10.0	0	deuteration level <0.5 at 10ms	227	(16)	20
apo-Mb (whale)	PL HDX NMR	1mbc	3	all-a	12	51	5.8	25	k_{cl}/k_{op} >80 (strong protection)	223	(17)	3.6
apo-Pc (French bean)	CO HDX NMR	9pcy	3	a+b	6	29	5.0, 5.9	25	protection factor (P)>20	165	(18)	-
Lysozyme (phage λ)	PL HDX NMR, MS	1am7	2	a+b	29	54	5.6	20	RTC <180 ms	20	(19)	8.4
BPTI	CO HDX NMR	5pti	3	all-b	7	8	4.0,…,7.5	70	-	711	(20)	-
CD2.D1 (rat)	CO HDX NMR	(1a64)	2	all-b	19	42	6.0,…,10.0	25	lnP >1.0	62	(21)	-
cobrotoxin (CBTX)	QF HDX NMR	1coe	3	all-b	6	24	3.0	5	RTC <20 ms	62	(22)	10
CTX III	QF HDX NMR	2crt	3	all-b	12	32	3.0	5	RTC < 30 ms	58	(23)	10
DHFR (E. coli)	PL HDX NMR	5dfr	3	a+b	5	26	6.3	15	strong protection in 13 ms	576	(24)	20
Fadd-DD	QF HDX NMR	(1e3y)	2	all-a	24	24	6.2	20	folding rate constant: 20.9±1.7 s^{-1}	(6)	(25)	5.4
ferricytochrome c (horse) H33N	CO HDX NMR	(1hrc)	3	all-a	13	all	2.0→9.8	22	folding eq. constant K_{UI}^{loc} >3 at 140 µs	482	(26)	-
GB1	PL HDX NMR	1pga	2	b+a	26	26	4.1	5	rate constant (s^{-1}) ~ 133	(4)	(27)	25[a]
hen egg white lysozyme (HEWL)	PL HDX NMR	1hel	3	a+b	7	48	5.2	20	RTC < 3 ms	81	(28), (29)	8.4
human acidic FGF	QF HDX NMR	(1rg8)	3	all-b	39	75	5.0	20	fast protection rate (s^{-1}) 1–0.3	(7)	(30, 31)	10
hisactophilin-1	QF HDX NMR	1hce	3	all-b	10	31	7.8	20	fast protection rate (s^{-1}) >20	n.a.	(30)	29[b]
RNase H (HIV)	PL HDX NMR	1hrh	3	a+b	13	23	5.5	25	P >10 at 74 ms	567	(32)	20
IL_1β	QF HDX NMR	1i1b	3	all-b	21	47	5.0	4	amide protection half-lives between 0.7 and 1.5 s	42	(33)	16
LB1	DT HDX NMR	2ptl	2	b+a	12	24	6.8→9.0	n.a.	P > 1.4		(34)	3.5
lysozyme (horse)	CO HDX NMR	2eql	3	a+b	10	46	7.5	25	high protection within 3.5 ms	80	(35)	-
lysozyme (human)	PL HDX NMR	1lz1	3	a+b	13	47	5.3	20	high protection within 3.5 ms	83	(36)	8
ovomucoid third domain (OMTKY3)	pH-dependent HDX NMR	1omu	3	a+b	4	13	6.0,…,10.0	30	high protection within 170 µs	51	(37)	-
protein A, B-	PL HDX NMR	1bdd	2	all-a	20	20	5.0	5	proton occupancy >0.7 at 6 ms	n.a.	(38)	10

domain									refolding time			
RNase A	PL HDX NMR	1rbx	3	a+b	14	27	4.25	10	strong protection in I1	70	(39)	37[c]
RNase H* (E. coli)	PL HDX MS	1f21	3	a+b	16	all	5.0	10	high protection in 9 ms	118	(40)	10
RNase T1	PL HDX NMR	(1ygw)	3	a+b	13	24	5.0	10	rate constant (s-1) > 25		(41)	50[d]
SNase H124L	PL HDX NMR	1joo	3	a+b	9	60	5.3	15	P >= 5	30	(42)	57[e]
lysozyme (phage T4) C54T/C97A	DT HDX NMR	(1am7)	3	a+b	7	60	6.8→10.2	25	proton occupancy <~ 0.6	19	(43)	13
Villin 14T	QF HDX NMR	2vil	2	a+b	31	31	4.1	21	RTC ~ 60 ms	42	(44)	20
Onconase (NLFO)	QF HDX NMR	1onc	3	b+a	31	42	5.5	20	highly protected within 250 ms	n.a.	(45)	8.3

[§] FT: Folding Type, SST: Secondary Structure Type: PDB; PDB code; NR: Number of early folding residues; NP: Number of probes; T: Temperature in °C; MSA: Number of sequences in Multiple Sequence Alignment; Ref.: Reference to paper; PL: Length of labelling pulse in ms, with justification if longer than 20 ms as footnote to this table. In the Protection threshold column, P stands for protection factor, while RTC stands for refolding time constant. In the PDB code column the ids are in brackets if the PDB sequence does not completely match the sequence measured in the folding experiment.

[$] Abbreviations used in this table column: ACBP, acyl-coenzyme A binding protein; apo-Mb, apo-myoglobin; apo-Pc, apo-plastocyanin; BPTI, bovine pancreatic trypsin inhibitor; CTX III, cardiotoxin analogue III; DHFR, dihydrofolate reductase; Fadd-DD, C-terminal domain of the Fas-associated death domain; CD2.D1, C-terminal domain of rat CD2; GB1, B1 immunoglobulin-binding domain of streptococcal protein G; IL-1β, interleukin-1bsubunit; LB1, B1 immunoglobulin-binding domain of peptostreptococcal protein L; RNase, ribonuclease; SNase, staphylococcal nuclease.

[¥] Quenched flow (QF), pulsed labelling (PL), exchange/folding competition (CO) and dead time labelling (DT) HDX experiments are distinguished indicating also the detection approach (NMR or MS). In case of OMTKY3, the pH dependence of exchange was used to obtain the unfolding and folding rates.

[£] An arrow connecting two pH values represents a pH jump in the experiment, values separated by a single comma mean measurements at different pH values, while those separated by ",…," mean multiple measurements within the given pH range.

[*] The number of sequences in the MSAs from the **HHBLITS_lowSeqId** approach is indicated. If in brackets, this MSA was not included in the analysis. If n.a., no alignments could be generated for this protein.

[a] The protein remains fully folded over a range extending from pH 2 to pH 11.3 at 25 C. Unfolding begins to occur above pH 11.3, conditions for the quenched flow D-H experiments were therefore chosen so that no significant contribution from the reverse reaction (i.e., unfolding) could occur.

[b] Using a labeling buffer of pH 8.93 gave the same results as using pH 9.52, confirming sufficient intensity of labeling pulse. At pH 9.52 the average time constant for exchange for the amides monitored is ~0.5 msec; therefore, protons are excluded from sites where exchange is retarded >60-fold.

[c] The duration of the pulse starts to limit complete labeling"

[d] There is no back exchange by varying pulse pHs and length; proton occupancies do not change.

[e] The stability of H124L SNase to the conditions of the labeling pulse were verified by equilibrium CD measurements and kinetically by using stopped-flow fluorescence to monitor the protein following pH jumps from pH 5 to pH 9 and from pH 5 to pH 10.

Text S4: Supporting Material on materials and methods

Modifications to original DynaMine version

The predictions from the original DynaMine linear model are affected at the termini by the lack of sequence context information, which we compensated for by assigning a weight in the linear model that captures the median behaviour over all residue types at the missing sequence positions (pre-N and post-C terminus). This adaptation did not require training a new model and practically only affects the first and last 25 residues in each sequence. With the adapted model the Root Mean Square Error (RMSE) decreases from 0.221 to 0.199, almost 10%, when run on the original RCI-S2_UNION_DP dataset of 1952 proteins in (10).

Supporting references

1. Silow, M., and M. Oliveberg. 1997. Transient aggregates in protein folding are easily mistaken for folding intermediates. Proceedings of the National Academy of Sciences of the United States of America 94:6084-6086.
2. Petersen, B., T. N. Petersen, P. Andersen, M. Nielsen, and C. Lundegaard. 2009. A generic method for assignment of reliability scores applied to solvent accessibility predictions. BMC structural biology 9:51.
3. Walsh, I., A. J. M. Martin, T. Di Domenico, and S. C. E. Tosatto. 2012. ESpritz: accurate and fast prediction of protein disorder. Bioinformatics (Oxford, England) 28:503-509.
4. Wilkins, M. R., E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel, and D. F. Hochstrasser. 1999. Protein identification and analysis tools in the ExPASy server. Methods in molecular biology (Clifton, N.J.) 112:531-552.
5. Bull, H. B., and K. Breese. 1974. Surface tension of amino acid solutions: a hydrophobicity scale of the amino acid residues. Arch Biochem Biophys 161:665-670.
6. Agashe, V. R., M. C. Shastry, and J. B. Udgaonkar. 1995. Initial hydrophobic collapse in the folding of barstar. Nature 377:754-757.
7. Udgaonkar, J. B. 2013. Polypeptide chain collapse and protein folding. Archives of biochemistry and biophysics 531:24-33.
8. Nickson, A. A., B. G. Wensley, and J. Clarke. 2013. Take home lessons from studies of related proteins. Current opinion in structural biology 23:66-74.
9. Li, R., and C. Woodward. 1999. The hydrogen exchange core and protein folding. Protein science : a publication of the Protein Society 8:1571-1590.

10. Cilia, E., R. Pancsa, P. Tompa, T. Lenaerts, and W. F. Vranken. 2013. From protein sequence to dynamics and disorder with DynaMine. Nature communications 4:2741.

11. Walsh, I., M. Giollo, T. Di Domenico, C. Ferrari, O. Zimmermann, and S. C. E. Tosatto. 2015. Comprehensive large-scale assessment of intrinsic protein disorder. Bioinformatics (Oxford, England) 31:201-208.

12. Sormanni, P., C. Camilloni, P. Fariselli, and M. Vendruscolo. 2015. The s2D method: simultaneous sequence-based prediction of the statistical populations of ordered and disordered regions in proteins. J Mol Biol 427:982-996.

13. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577-2637.

14. Zhang, F., and R. Brüschweiler. 2002. Contact model for the prediction of NMR N-H order parameters in globular proteins. Journal of the American Chemical Society 124:12654-12655.

15. Teilum, K., B. B. Kragelund, J. Knudsen, and F. M. Poulsen. 2000. Formation of hydrogen bonds precedes the rate-limiting formation of persistent structure in the folding of ACBP. J Mol Biol 301:1307-1314.

16. Pan, J., J. Han, C. H. Borchers, and L. Konermann. 2010. Characterizing short-lived protein folding intermediates by top-down hydrogen exchange mass spectrometry. Analytical chemistry 82:8591-8597.

17. Uzawa, T., C. Nishimura, S. Akiyama, K. Ishimori, S. Takahashi, H. J. Dyson, and P. E. Wright. 2008. Hierarchical folding mechanism of apomyoglobin revealed by ultra-fast H/D exchange coupled with 2D NMR. Proceedings of the National Academy of Sciences of the United States of America 105:13859-13864.

18. Koide, S., H. J. Dyson, and P. E. Wright. 1993. Characterization of a folding intermediate of apoplastocyanin trapped by proline isomerization. Biochemistry 32:12299-12310.

19. Di Paolo, A., D. Balbeur, E. De Pauw, C. Redfield, and A. Matagne. 2010. Rapid collapse into a molten globule is followed by simple two-state kinetics in the folding of lysozyme from bacteriophage lambda. Biochemistry 49:8646-8657.

20. Roder, H., and K. Wuthrich. 1986. Protein folding kinetics by combined use of rapid mixing techniques and NMR observation of individual amide protons. Proteins 1:34-42.

21. Parker, M. J., C. E. Dempsey, M. Lorch, and A. R. Clarke. 1997. Acquisition of native beta-strand topology during the rapid collapse phase of protein folding. Biochemistry 36:13396-13405.

22. Hsieh, H. C., T. K. Kumar, T. Sivaraman, and C. Yu. 2006. Refolding of a small all beta-sheet protein proceeds with accumulation of kinetic intermediates. Arch Biochem Biophys 447:147-154.

23. Sivaraman, T., T. K. Kumar, D. K. Chang, W. Y. Lin, and C. Yu. 1998. Events in the kinetic folding pathway of a small, all beta-sheet protein. J Biol Chem 273:10181-10189.

24. Jones, B. E., and C. R. Matthews. 1995. Early intermediates in the folding of dihydrofolate reductase from Escherichia coli detected by hydrogen exchange and NMR. Protein science : a publication of the Protein Society 4:167-177.

25. Greene, L. H., H. Li, J. Zhong, G. Zhao, and K. Wilson. 2012. Folding of an all-helical Greek-key protein monitored by quenched-flow hydrogen-deuterium exchange and NMR spectroscopy. Eur Biophys J 41:41-51.
26. Fazelinia, H., M. Xu, H. Cheng, and H. Roder. 2014. Ultrafast hydrogen exchange reveals specific structural events during the initial stages of folding of cytochrome c. J Am Chem Soc 136:733-740.
27. Kuszewski, J., G. M. Clore, and A. M. Gronenborn. 1994. Fast folding of a prototypic polypeptide: the immunoglobulin binding domain of streptococcal protein G. Protein science : a publication of the Protein Society 3:1945-1952.
28. Miranker, A., S. E. Radford, M. Karplus, and C. M. Dobson. 1991. Demonstration by NMR of folding domains in lysozyme. Nature 349:633-636.
29. Radford, S. E., C. M. Dobson, and P. A. Evans. 1992. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. Nature 358:302-307.
30. Liu, C., J. A. Gaspar, H. J. Wong, and E. M. Meiering. 2002. Conserved and nonconserved features of the folding pathway of hisactophilin, a beta-trefoil protein. Protein science : a publication of the Protein Society 11:669-679.
31. Samuel, D., T. K. Kumar, K. Balamurugan, W. Y. Lin, D. H. Chin, and C. Yu. 2001. Structural events during the refolding of an all beta-sheet protein. J Biol Chem 276:4134-4141.
32. Kern, G., T. Handel, and S. Marqusee. 1998. Characterization of a folding intermediate from HIV-1 ribonuclease H. Protein science : a publication of the Protein Society 7:2164-2174.
33. Varley, P., A. M. Gronenborn, H. Christensen, P. T. Wingfield, R. H. Pain, and G. M. Clore. 1993. Kinetics of folding of the all-beta sheet protein interleukin-1 beta. Science 260:1110-1113.
34. Yi, Q., M. L. Scalley, K. T. Simons, S. T. Gladwin, and D. Baker. 1997. Characterization of the free energy spectrum of peptostreptococcal protein L. Folding & design 2:271-280.
35. Morozova-Roche, L. A., J. A. Jones, W. Noppe, and C. M. Dobson. 1999. Independent nucleation and heterogeneous assembly of structure during folding of equine lysozyme. J Mol Biol 289:1055-1073.
36. Hooke, S. D., S. E. Radford, and C. M. Dobson. 1994. The refolding of human lysozyme: a comparison with the structurally homologous hen lysozyme. Biochemistry 33:5867-5876.
37. Arrington, C. B., and A. D. Robertson. 1997. Microsecond protein folding kinetics from native-state hydrogen exchange. Biochemistry 36:8686-8691.
38. Bai, Y., A. Karimi, H. J. Dyson, and P. E. Wright. 1997. Absence of a stable intermediate on the folding pathway of protein A. Protein science : a publication of the Protein Society 6:1449-1457.
39. Udgaonkar, J. B., and R. L. Baldwin. 1990. Early folding intermediate of ribonuclease A. Proceedings of the National Academy of Sciences of the United States of America 87:8197-8201.
40. Hu, W., B. T. Walters, Z. Y. Kan, L. Mayne, L. E. Rosen, S. Marqusee, and S. W. Englander. 2013. Stepwise protein folding at near amino acid resolution by

hydrogen exchange and mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America 110:7684-7689.

41. Mullins, L. S., C. N. Pace, and F. M. Raushel. 1993. Investigation of ribonuclease T1 folding intermediates by hydrogen-deuterium amide exchange-two-dimensional NMR spectroscopy. Biochemistry 32:6152-6156.

42. Walkenhorst, W. F., J. A. Edwards, J. L. Markley, and H. Roder. 2002. Early formation of a beta hairpin during folding of staphylococcal nuclease H124L as detected by pulsed hydrogen exchange. Protein science : a publication of the Protein Society 11:82-91.

43. Kato, H., N. D. Vu, H. Feng, Z. Zhou, and Y. Bai. 2007. The folding pathway of T4 lysozyme: an on-pathway hidden folding intermediate. J Mol Biol 365:881-891.

44. Choe, S. E., P. T. Matsudaira, J. Osterhout, G. Wagner, and E. I. Shakhnovich. 1998. Folding kinetics of villin 14T, a protein domain with a central beta-sheet and two hydrophobic cores. Biochemistry 37:14508-14518.

45. Schulenburg, C., C. Löw, U. Weininger, C. Mrestani-Klaus, H. Hofmann, J. Balbach, R. Ulbrich-Hofmann, and U. Arnold. 2009. The folding pathway of onconase is directed by a conserved intermediate. Biochemistry 48:8449-8457.