Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/3/7/e1700479/DC1

Supplementary Materials for

Solving protein structures using short-distance cross-linking constraints as a guide for discrete molecular dynamics simulations

Nicholas I. Brodie, Konstantin I. Popov, Evgeniy V. Petrotchenko,
Nikolay V. Dokholyan, Christoph H. Borchers

Published 7 July 2017, *Sci. Adv.* **3**, e1700479 (2017)
DOI: 10.1126/sciadv.1700479

This PDF file includes:

table S1. Mb cross-links used as constraints in DMD simulations. Samples cross-linked with ABAS or TATA were digested with proteinase K, while samples cross-linked with SDA, DSA, or DSG were digested with trypsin.

Cross-linker	M+H	m/z	z	ppm error	Residues 1	Sequence	Residues 2	Sequence	Cross-link
ABAS	1094.5079	547.7579	2	0.8	55-56	E.MK.A	118-125	S.KHPGDF.G	K56-K118
ABAS	1225.5476	613.2777	2	0.7	133-134	T.KA.L	1-8	M.GLSDGEWQ.Q	N-term-K133
TATA	787.3334	394.1706	2	1.1	1-2	-.GL.S	80-83	K.GHHE.A	N-term-G80
TATA	821.4743	411.2408	2	-0.3	88-90	K.PLA.Q	142-145	D.IAAK.Y	A90-K145
TATA	922.5115	461.7597	2	-1.1	37-39	H.PET.L	98-101	H.KIPI.K	T39-I99
TATA	1031.4278	516.2178	2	1.1	1-7	-.GLSDGEW.Q	129-130	Q.GA.M	N-term-A130
TATA	1035.5952	518.3015	2	0.7	45-48	D.KFKH.L	98-100	H.KIP.P	K45-K98
TATA	1090.4008	1090.4008	1	0	3-8	L.SDGEWQ.Q	129-131	Q.GAM.T	S3-A130
DSA	1785.9883	893.4978	2	0.5	97-102	K.HKIPIK.Y	146-153	K.YKELGFQG.-	K98-K147
DSA	2853.4494	951.8213	3	0.5	1-16	-.GLSDGEWQQVLNVWGK.V	140-147	R.NDIAAKYK.E	N-term-K145
DSA	2904.4621	968.8255	3	0.4	80-96	K.GHHEAELKPLAQSHATK.H	146-153	K.YKELGFQG.-	K87-K147
DSA	3013.5848	1005.1998	3	0	79-96	K.KGHHEAELKPLAQSHATK.H	140-147	R.NDIAAKYK.E	K87-K145
DSG	2492.3603	623.8455	4	-2.5	97-102	K.HKIPIK.Y	32-45	R.LFTGHPETLEKFDK.F	K98-K42
SDA	3194.8229	799.4612	4	0.8	64-78	K.HGTVVLTALGGILKK.K	17-31	K.VEADIAGHGQEVLIR.L	K77-E18
SDA	3194.8225	799.4611	4	0.9	64-78	K.HGTVVLTALGGILKK.K	17-31	K.VEADIAGHGQEVLIR.L	K77-V17
SDA	3167.6229	792.6612	4	0.6	51-63	K.TEAEMKASEDLKK.H	17-31	K.VEADIAGHGQEVLIR.L	K56-E27
SDA	3167.6252	792.6617	4	-0.2	51-63	K.TEAEMKASEDLKK.H	17-31	K.VEADIAGHGQEVLIR.L	K56-Q26
SDA	3167.6245	792.6616	4	0	51-63	K.TEAEMKASEDLKK.H	17-31	K.VEADIAGHGQEVLIR.L	K56-H24
SDA	3349.7411	838.1907	4	0.3	32-45	R.LFTGHPETLEKFDK.F	17-31	K.VEADIAGHGQEVLIR.L	K42-L29
SDA	3750.9015	938.4808	4	-0.9	80-96	K.GHHEAELKPLAQSHATK.H	1-16	-.GLSDGEWQQVLNVWGK.V	N-term-H81

table S2. FKBP-25 cross-links used as distance constraints in DMD simulations. Peptide information is shown below, with an "i" indicating an intra-peptide cross-link. Samples cross-linked with TATA were digested using proteinase K, while samples cross-linked with DSA samples were digested with trypsin.

Cross-linker	M+H	m/z	z	ppm	Residues 1	Sequence	Residues 2	Sequence	Cross-link
DSA	1144.5990	572.8034	2	0.6	30-38	K.KGDKTNFPK.K	-	-.i.-	K30-K33
DSA	1193.5875	597.2977	2	-0.3	23-25	K.YTK.S	16-22	R.GSMGPPK.Y	K22-K25
DSA	1473.8036	491.9398	3	1.6	113-125	K.KGQPDAKIPPNAK.L	-	-.i.-	K113-K119
DSA	1679.9786	420.7505	4	1	113-119	K.KGQPDAK.I	78-85	K.VGVGKVIR.G	K82-K113
DSA	1701.9623	851.4851	2	1.4	97-101	K.GEKAR.L	69-77	K.KNAKPLSFK.V	K72-K99
DSA	1709.0443	855.0261	2	1.7	71-85	N.AKPLSFKVGVGKVIR.G	-	-.i.-	K72-K77
DSA	1748.0042	437.7569	4	1.3	67-69	K.KKK.N	114-125	K.GQPDAKIPPNAK.L	K119-K68
DSA	1901.9437	951.4758	2	-0.6	86-101	R.GWDEALLTMSKGEKAR.L	-	-.i.-	K99-K96
DSA	1917.9720	480.2489	4	1	113-119	K.KGQPDAK.I	16-25	R.GSMGPPKYTK.S	K113-K22
DSA	1982.0948	496.2796	4	0.5	23-29	K.YTKSVLK.K	30-38	K.KGDKTNFPK.K	K25-K30
DSA	2077.1019	693.0392	3	-1	67-69	K.KKK.N	86-99	R.GWDEALLTMSKGEK.A	K96-K68
DSA	2271.2424	757.7527	3	-1.8	78-85	K.VGVGKVIR.G	102-112	R.LEIEPEWAYGK.K	K82-Y110
DSA	2351.2040	588.5569	4	-1	34-39	K.TNFPKK.G	3-15	R.GSHHHHHHGLVPR.G	S3-K38
DSA	2399.3342	600.5894	4	-0.4	78-85	K.VGVGKVIR.G	102-113	R.LEIEPEWAYGKK.G	K112-K82
DSA	2472.2611	824.7589	3	-0.8	16-22	R.GSMGPPK.Y	100-113	K.ARLEIEPEWAYGKK.G	S17-K112
DSA	2523.2475	421.3811	6	1	31-38	K.GDKTNFPK.K	3-15	R.GSHHHHHHGLVPR.G	S3-K33
DSA	2637.3286	1319.1682	2	-0.6	16-25	R.GSMGPPKYTK.S	102-113	R.LEIEPEWAYGKK.G	K22-K112
DSA	2637.3296	879.7817	3	-1	16-25	R.GSMGPPKYTK.S	102-113	R.LEIEPEWAYGKK.G	K22-K113
DSA	2807.4643	702.6220	4	-1	102-113	R.LEIEPEWAYGKK.G	114-125	K.GQPDAKIPPNAK.L	K113-K119
DSA	2864.4653	716.8722	4	0	16-25	R.GSMGPPKYTK.S	100-113	K.ARLEIEPEWAYGKK.G	K22-K112
TATA	1591.7847	531.2668	3	-1.5	102-107	R.LEIEPE.W	113-119	K.KGQPDAK.I	E107-G114
TATA	1743.8944	581.9700	3	0.8	37-39	F.PKK.G	86-96	R.GWDEALLTMSK.G	K39-E89
TATA	1866.9105	622.9754	3	1.5	23-25	K.YTK.S	102-112	R.LEIEPEWAYGK.K	K25-W108
TATA	2164.0559	722.0239	3	0.5	102-109	R.LEIEPEWA.Y	83-91	K.VIRGWDEAL.L	A109-I84

table S3. Residues modified by PCAS- $^{12}C_6{}^{13}/C_6$ in the urea-PCAS SM experiments.
The Heavy/Light (H/L) ratio indicates the ratio of the exposure a specific residue to solvent when treated with urea compared to its exposure in the folded state. An H/L ratio of > 1.5 indicates that the residue was protected from solvent in the folded, native protein.

Myoglobin				
Peptide	**Mass**	**ppm error**	**H/L ratio**	**Residue**
L.NVWGK(+105.02)VE.A	935.4501	0.3	1.30	17
K.FK(+105.02)HLKT.E	877.4810	0	2.63	48
L.K(+105.02)TEAEM.K	812.3375	-0.8	1.43	51
A.SEDLK(+105.02)KHGTVVL.T	1429.7570	-1.2	2.24	63
M.KASEDLKK(+105.02)HGTVV.L	1515.8050	-0.1	1.43	64
K.K(+105.02)GHHEAEL.K	1024.4730	-1.6	1.46	80
L.KPLAQSHATK(+105.02).H	1184.6300	-0.2	2.20	97
I.IHVLHSK(+105.02)HPGDF.G	1490.7420	-0.3	2.54	119
D.AQGAMTK(+105.02)A.L	881.4066	0	5.14	134
A.K(+105.02)YKELG.F	841.4334	-0.1	2.88	146
K.YK(+105.02)ELGF.Q	860.4069	0.1	1.31	148

FKBP25				
Peptide	**Mass**	**ppm error**	**H/L ratio**	**Residue**
L.KK(+105.02)GDKTNF.P	1041.5240	-1	3.05	30
F.K(+109.05)VGVGKVIRG.W	1120.7060	-4.1	1.89	77
G.K(+105.02)VIRGWD.E	977.5083	0.3	1.89	82
M.SK(+105.02)GEKARL.E	992.5403	-0.5	1.72	96
L.TMSKGEK(+109.05)ARL.E	1228.6580	-4.5	2.87	99
L.EIEPEWAY(+105.02)G.K	1197.4980	-0.6	1.33	110
A.YGK(+105.02)KGQPDAKIPPNAKLTF.E	2177.1630	0.3	1.48	112

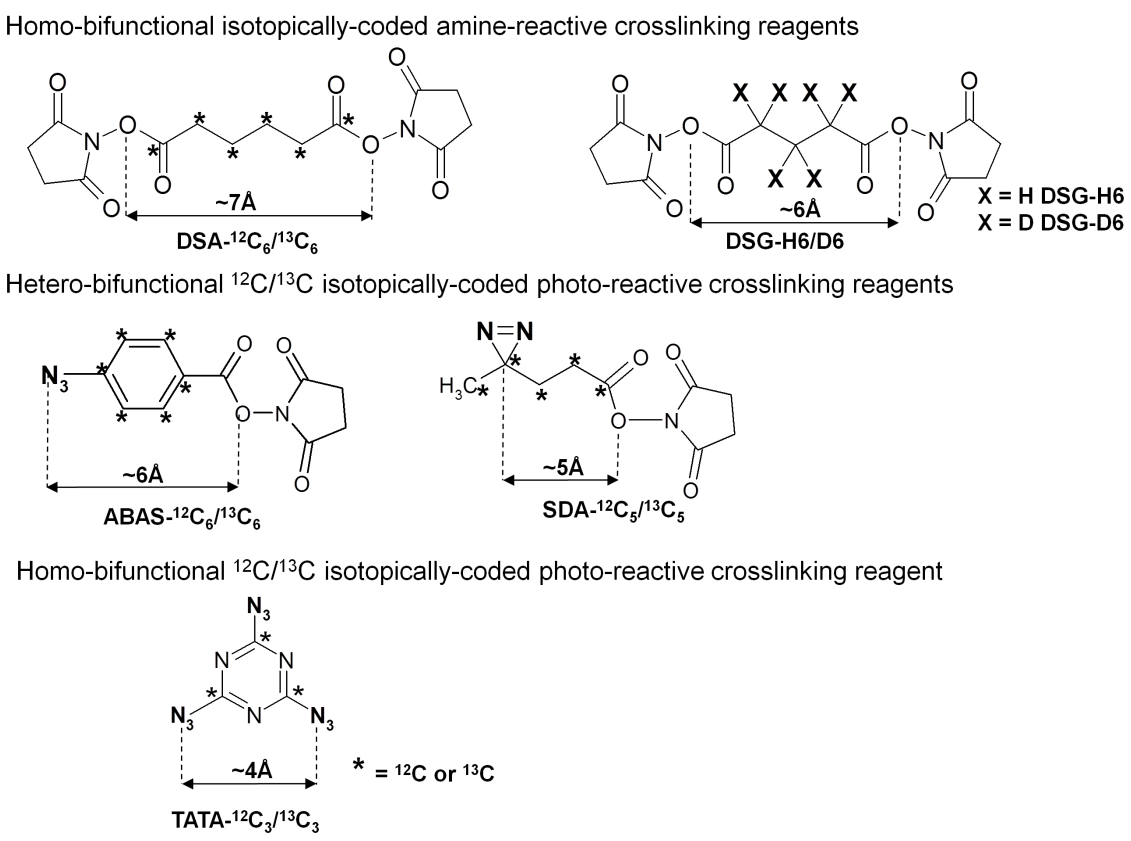


fig. S1. Panel of isotopically coded cross-linking reagents used for the structural characterization of Mb and FKBP. All cross-links used in the DMD simulations were less than 8Å in length.

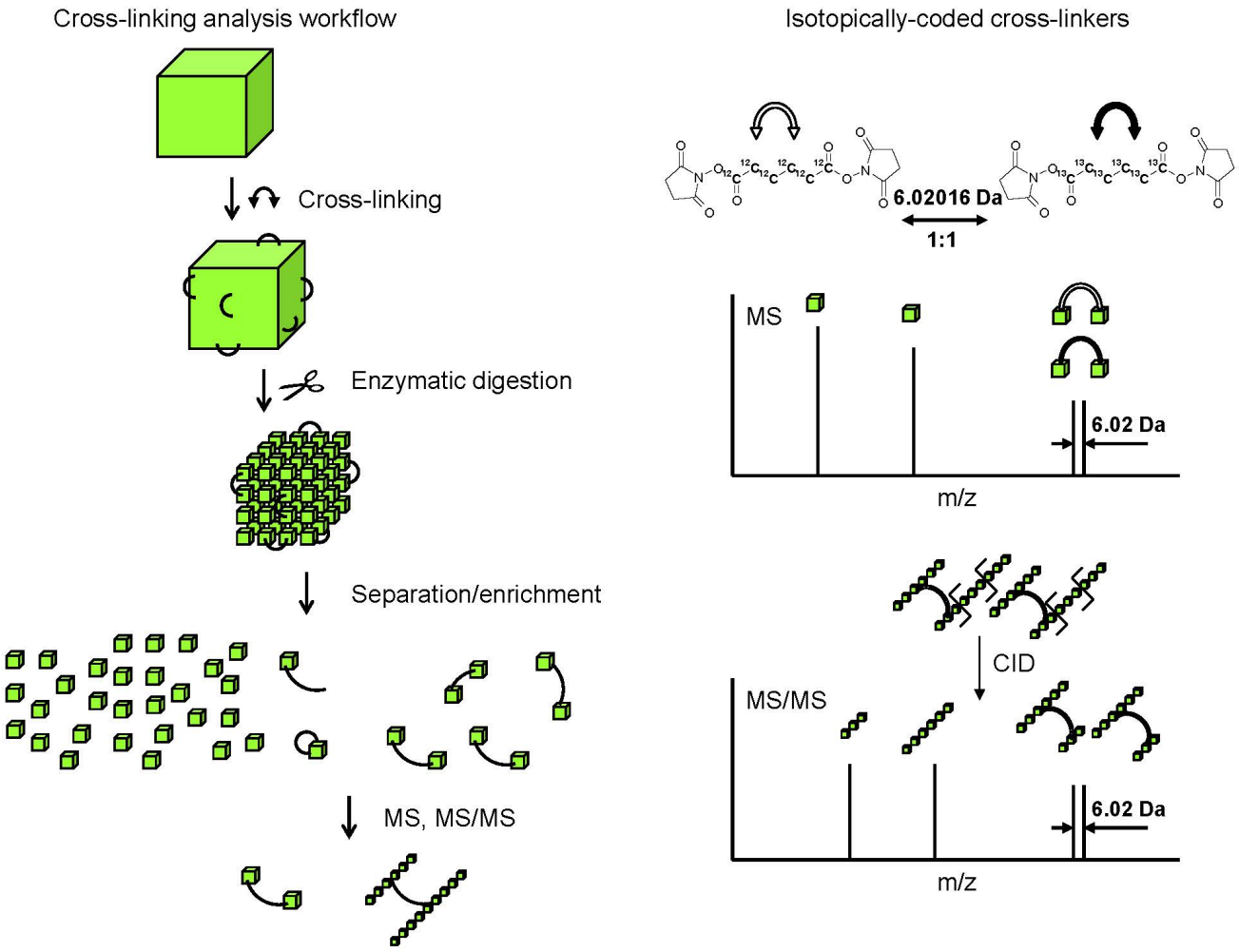
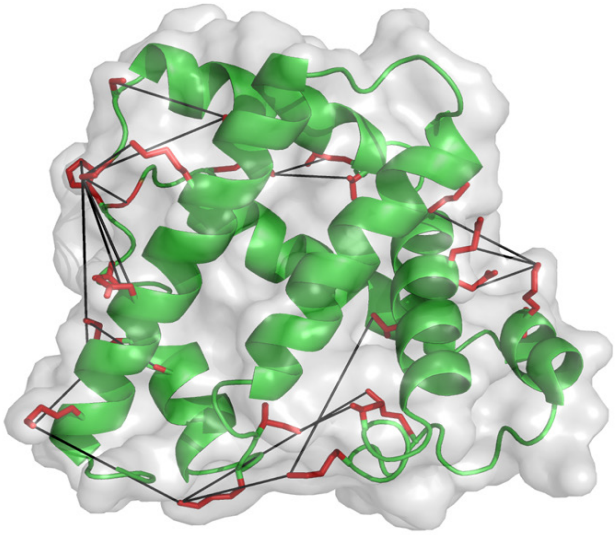


fig. S2. Cross-linking analysis workflow. A) Proteins are cross-linked, enzymatically digested, and analyzed by LC-MS/MS. B) Isotopic labeling of cross-linking reagents enables the specific detection and acquisition of cross-linked peptides in MS1. The doublets present in the MS/MS spectra of the combined light and heavy forms of the cross-link also provide additional confirmation of the correct assignment of the cross-link.

A



B

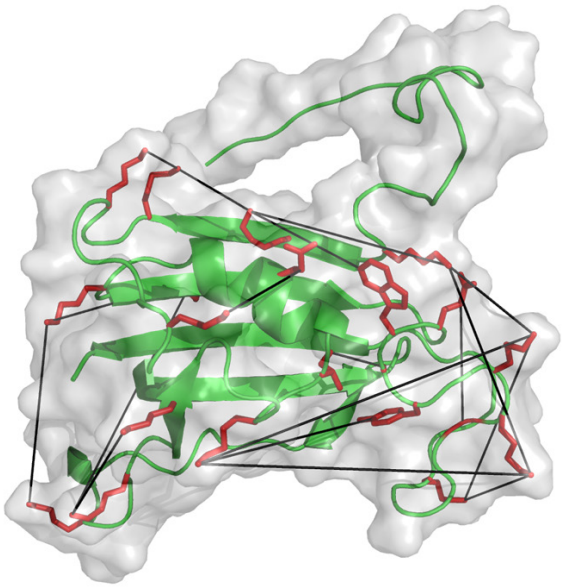
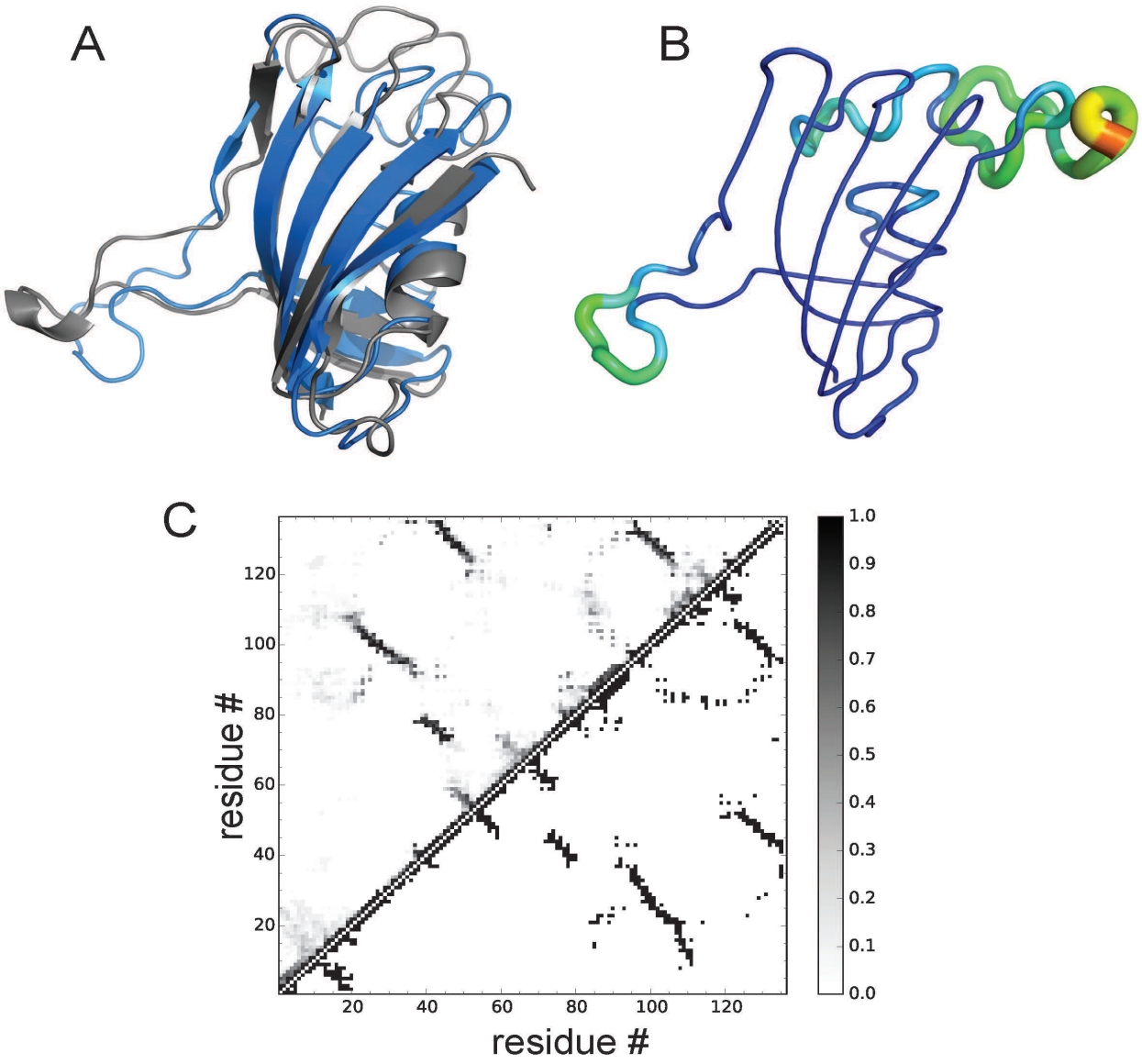


fig. S3. Cross-linking results for A) Mb and B) FKBP, overlaid on the crystal structures. Cross-linked residues are represented in red. Distances were well within the maximum distance expected for each of the cross-linking reagents.


A
B
C
residue #
residue #
20
40
60
80
100
120
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

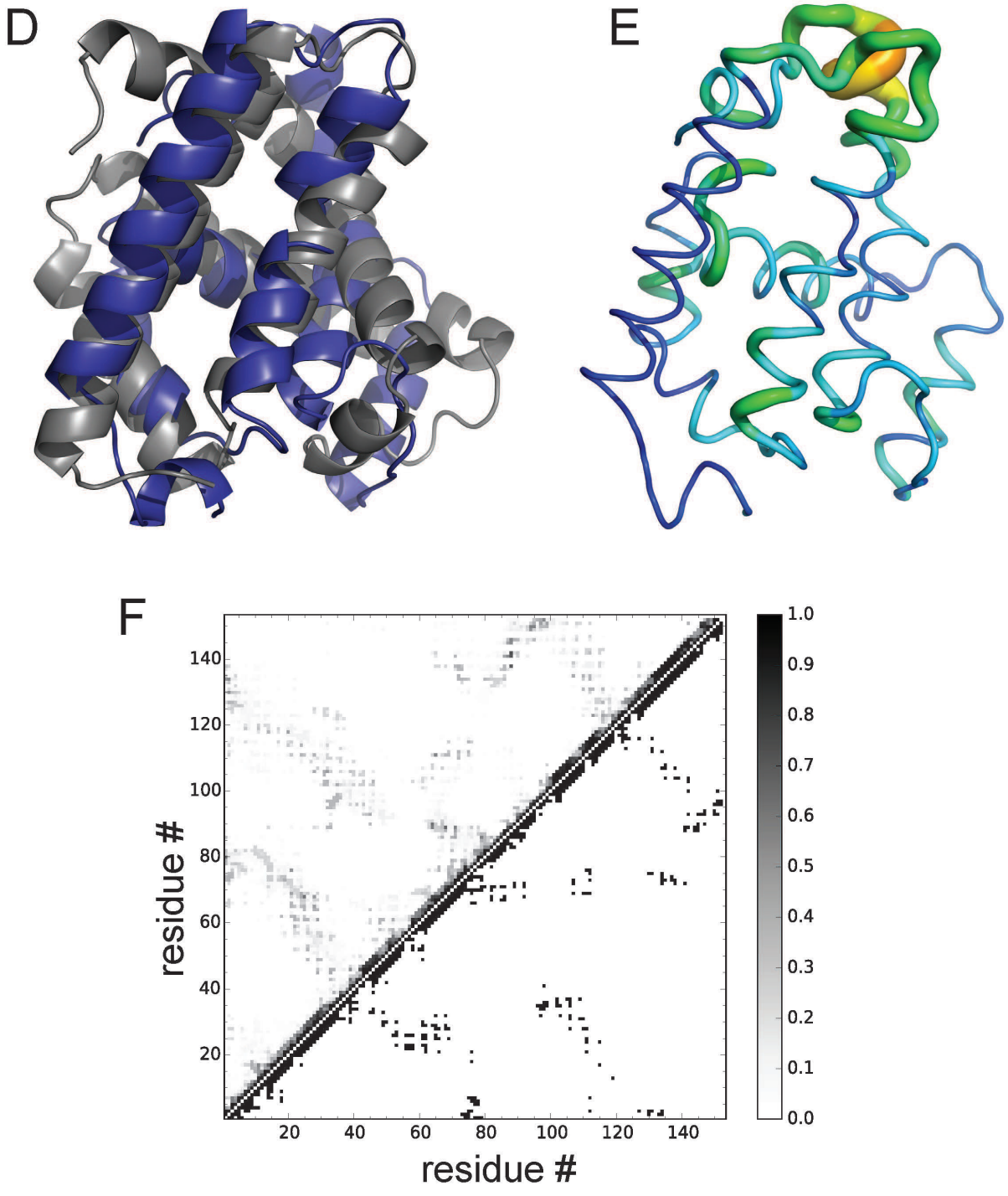

D
E
F
residue #
residue #
20
40
60
80
100
120
140
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

fig. S4. Conformational dynamics of predicted structures. Shown here are the predicted models (blue) for (A) FKBP and (D) myoglobin aligned with their respective X-ray structures (grey), along with a tube representation of the fluctuations of the (B) FKBP and (E) myoglobin models. The thickness and color of the tubes indicate the dynamics of the corresponding regions during the simulations. The tubes are colored from blue (low flexibility) to red (high flexibility).
The datapoints below the diagonals in plots (C) and (F) represent binary static-contact maps between the residues of the predicted structures shown in blue in (A) and (D), respectively. Two residues form a contact if their Cα atoms are within 8Å of each other; every black dot indicates a contact between corresponding residues. Data points above the diagonal in panels (C) and (F) show how often each particular contact between two residues can be found within the clusters for which our models (blue in (A) and (D) are centroids. The grayscale color code indicates the frequencies of the corresponding contact, where black = always in contact and white = never in contact.

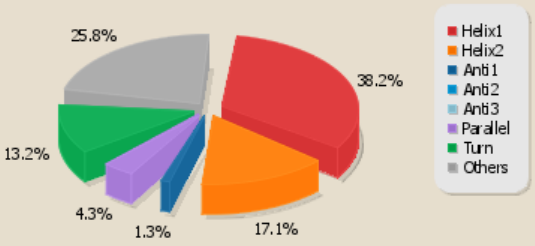
A

Title	2016091500414	Wavelength range	190-250 nm
Date/time	15-09-2016 0:04 (GMT+1:00)	Scale factor	1

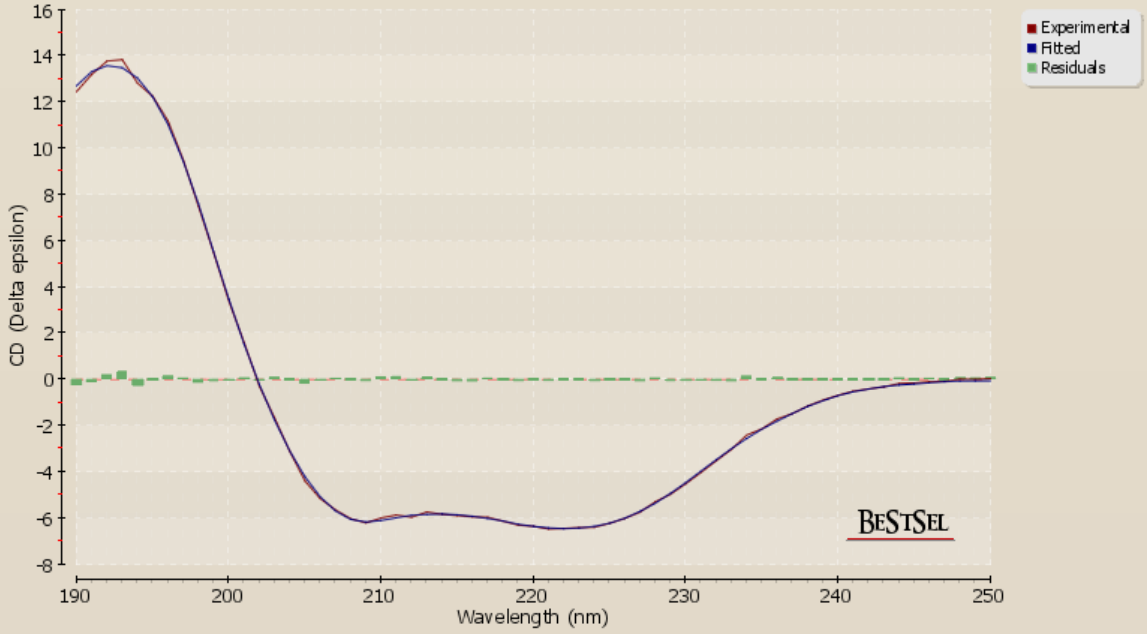
Secondary structure

Estimated secondary structure content (%)

Helix	55.3	Helix1 (regular)	38.2
		Helix2 (distorted)	17.1
Antiparallel	1.3	Anti1 (left-twisted)	1.3
		Anti2 (relaxed)	0.0
		Anti3 (right-twisted)	0.0
Parallel	4.3		
Turn	13.2		
Others	25.8		



Fitting results



RMSD	0.093	NRMSD	0.00458

B

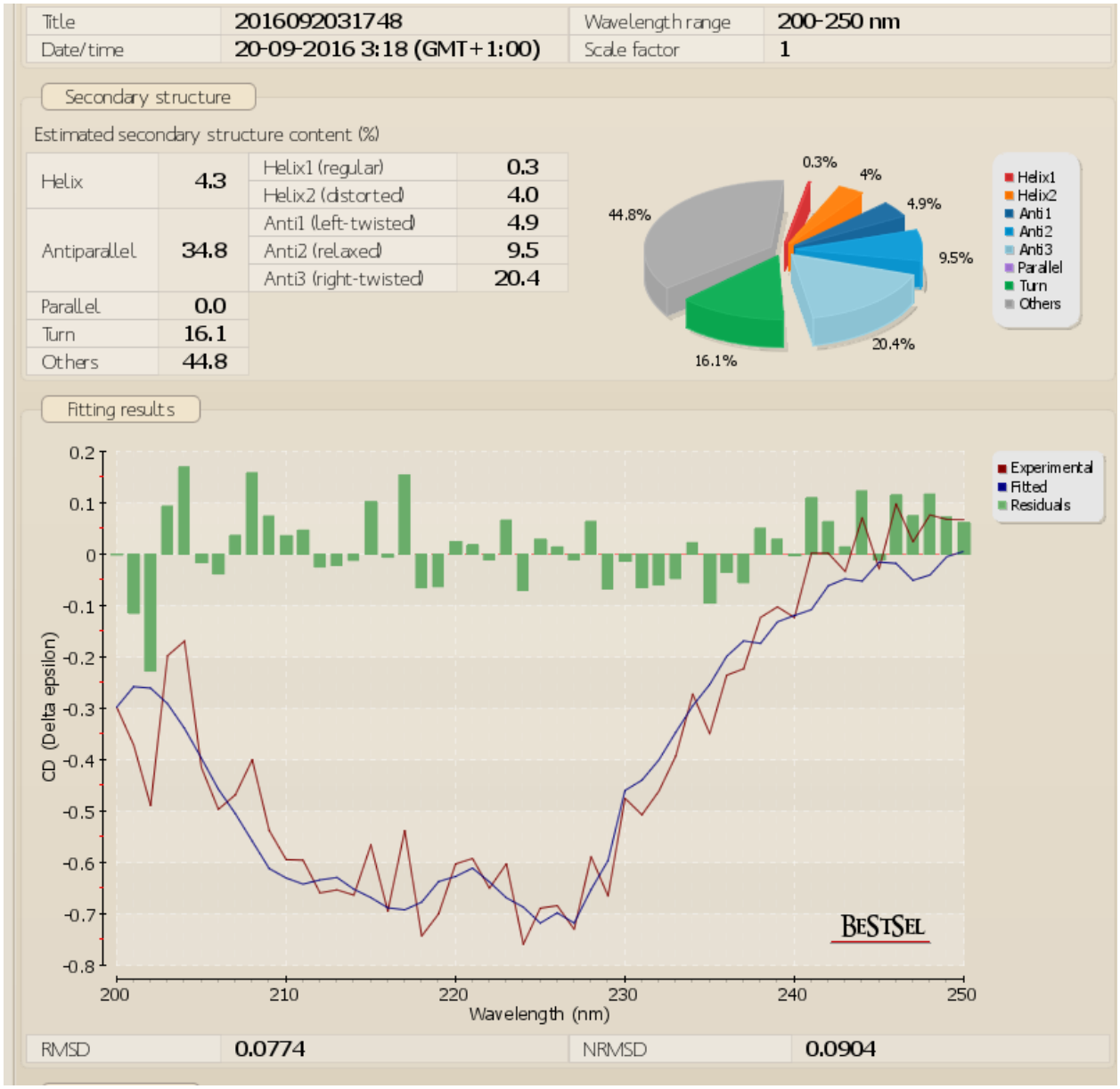


fig. S5. **CD results for A) Mb and B) FKBP.** CD results were calculated using the BestSel server.

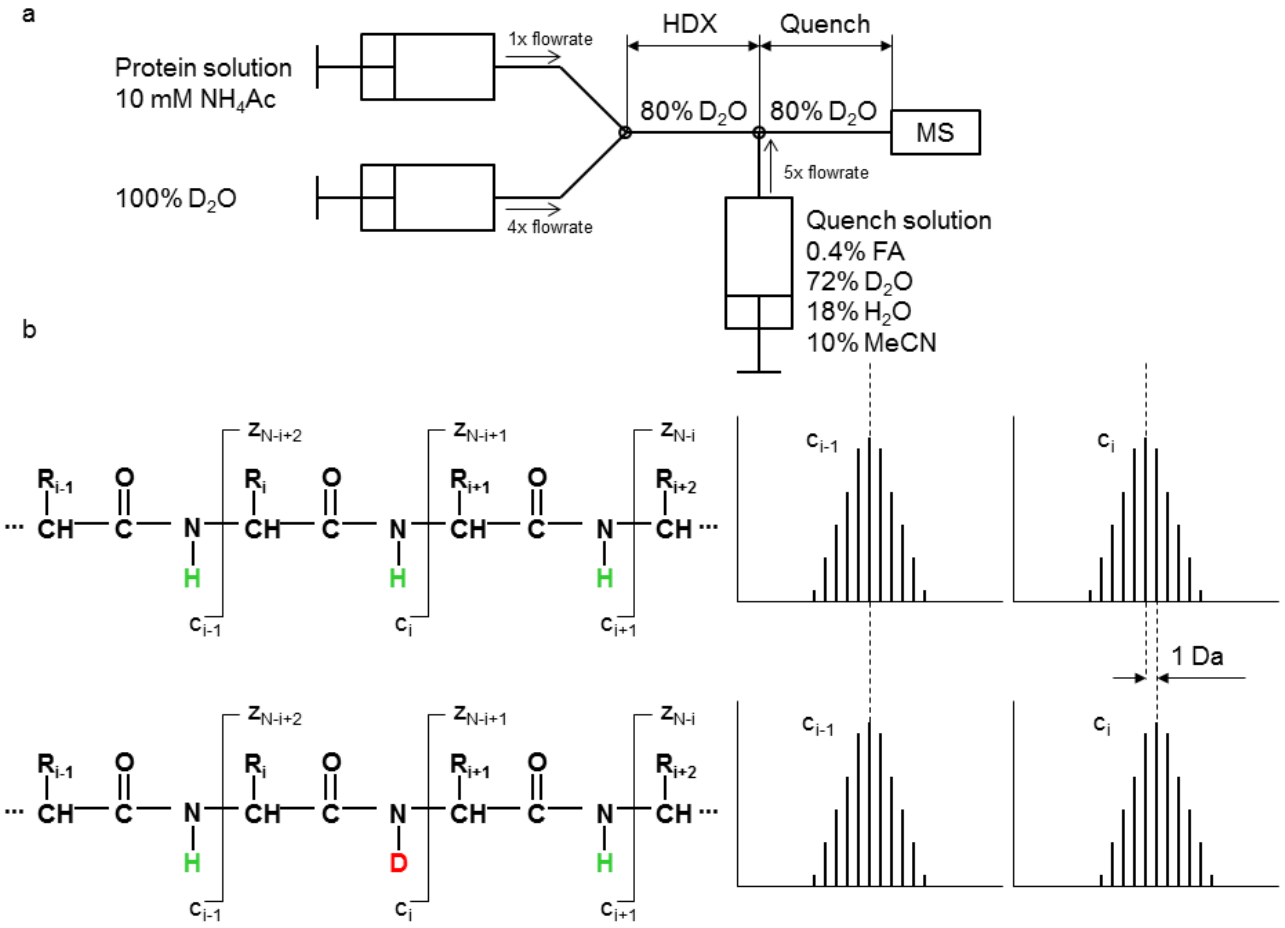
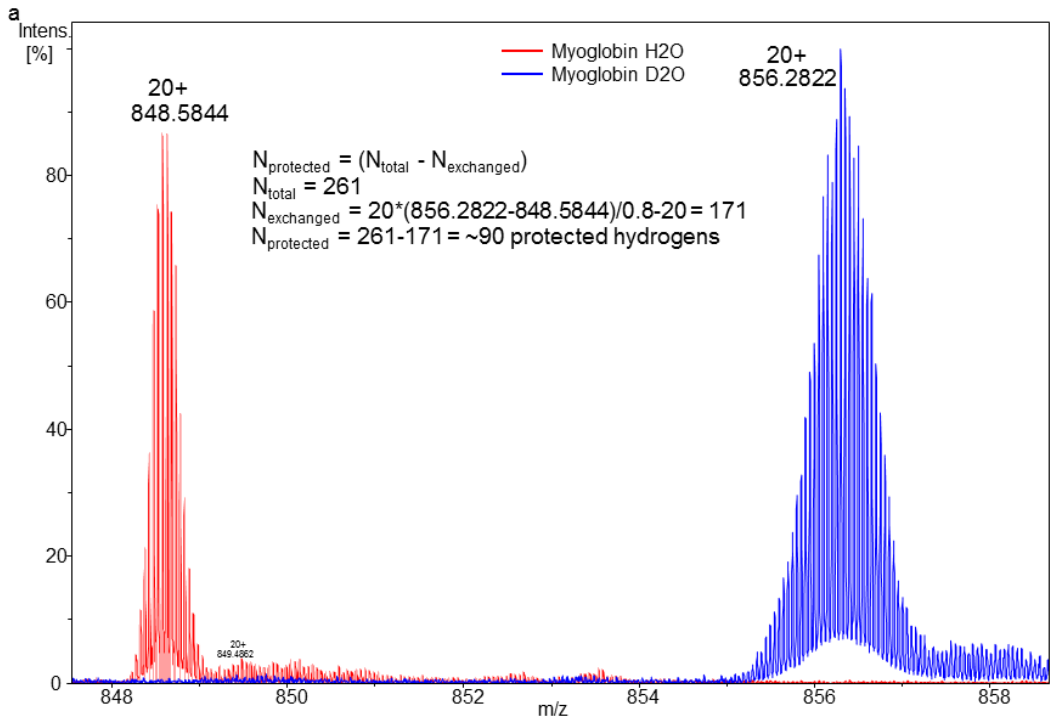


fig. S6. HDX analysis workflow. A) Schematic of a top-down HDX experiment. Protein and D $_2$O are combined in-line at a 1:4 ratio and exchanged in a short capillary (approximately 2 s at 10 µL/minute). A quenching solution -- which contains D_2O in the same proportion as in the sample mixture -- is added at a rate of 10 µL/minute, and the whole mixture flows directly into the source of the mass spectrometer for analysis. B) ECD fragmentation of the intact protein yields c- and z-ions which can provide residue-specific information on the incorporation of deuterium atoms into the backbone amides.

A



B

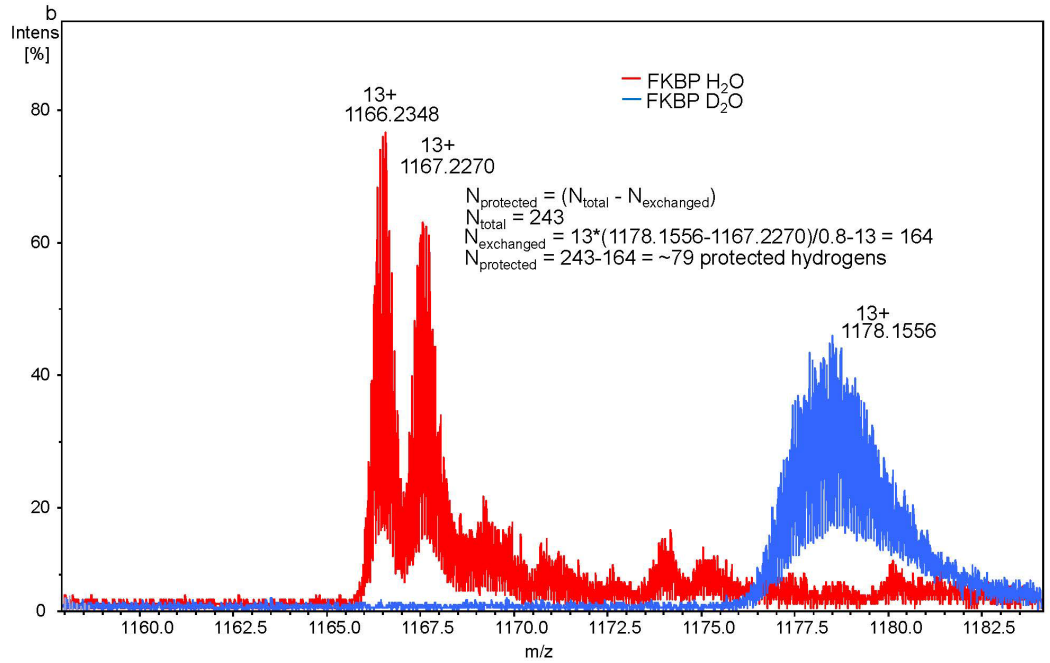


fig. S7. Hdx of intact proteins. A) Hydrogen/deuterium exchange of myoglobin. B) Hydrogen/ deuterium exchange of FKBP25.

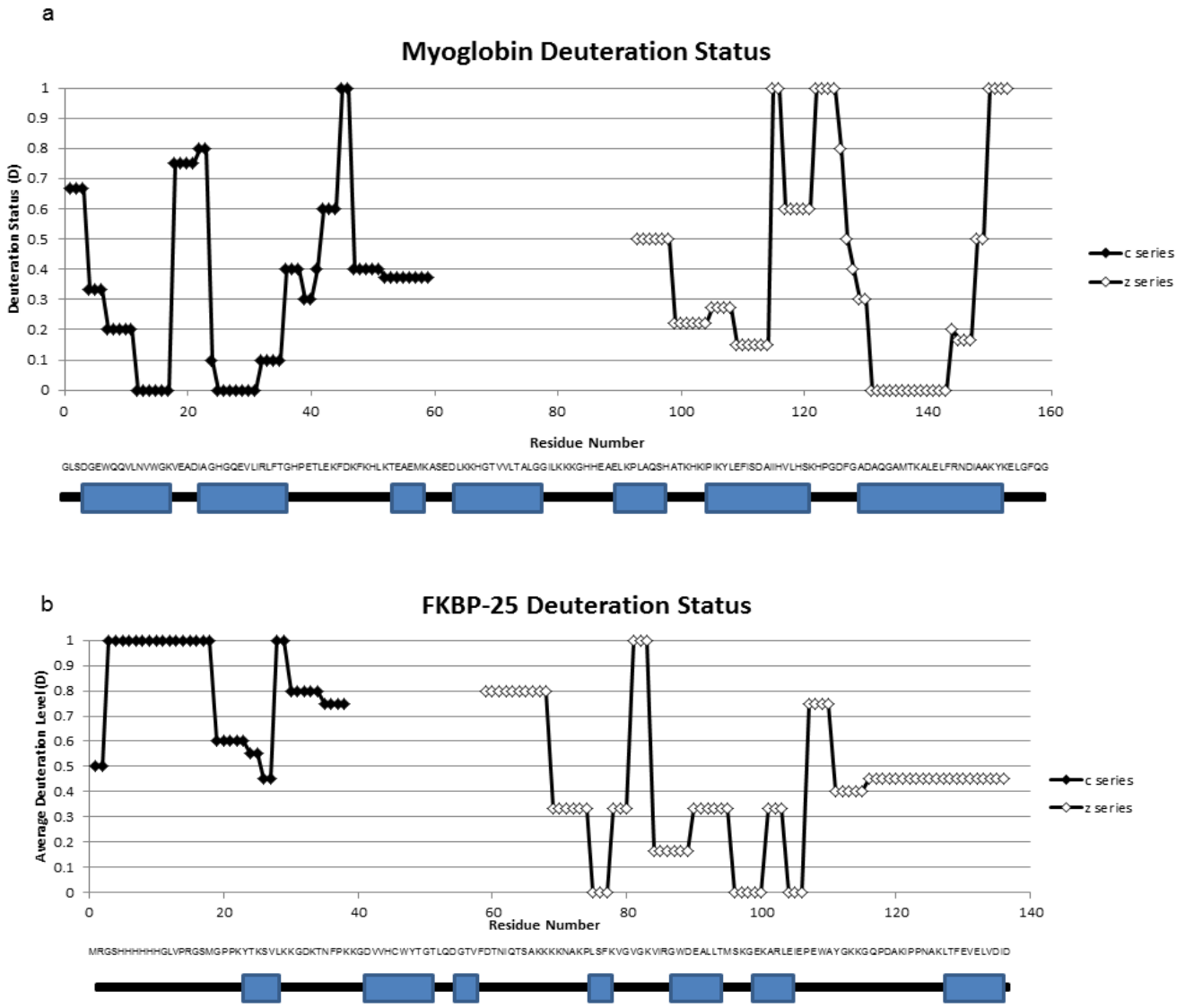


fig. S8. Deuteration status of backbone amides for Mb and FKBP.

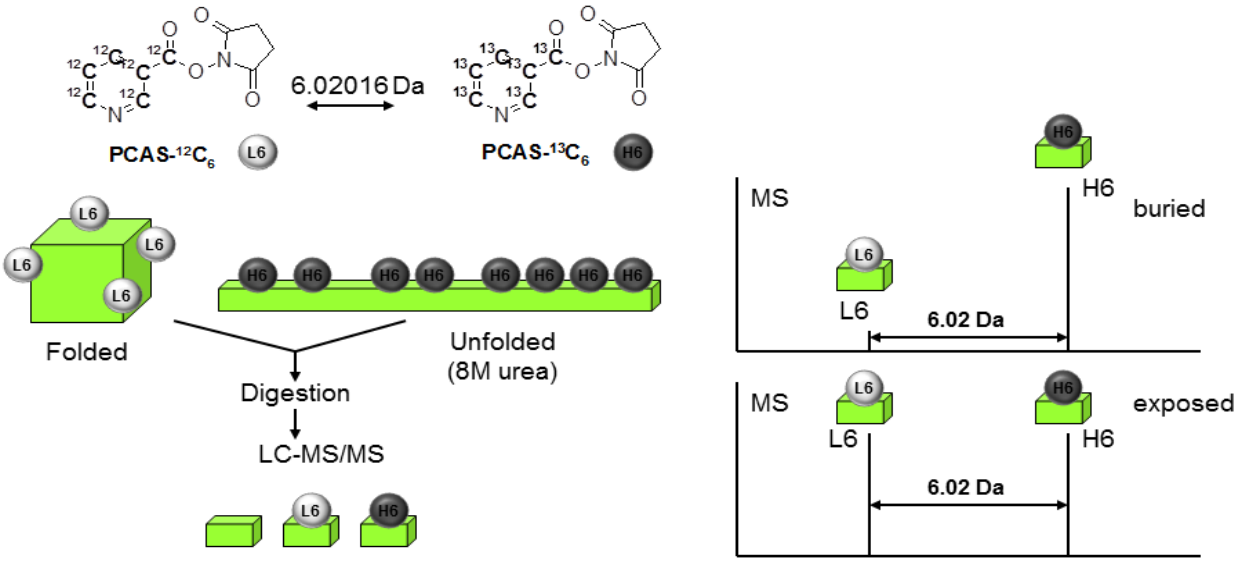
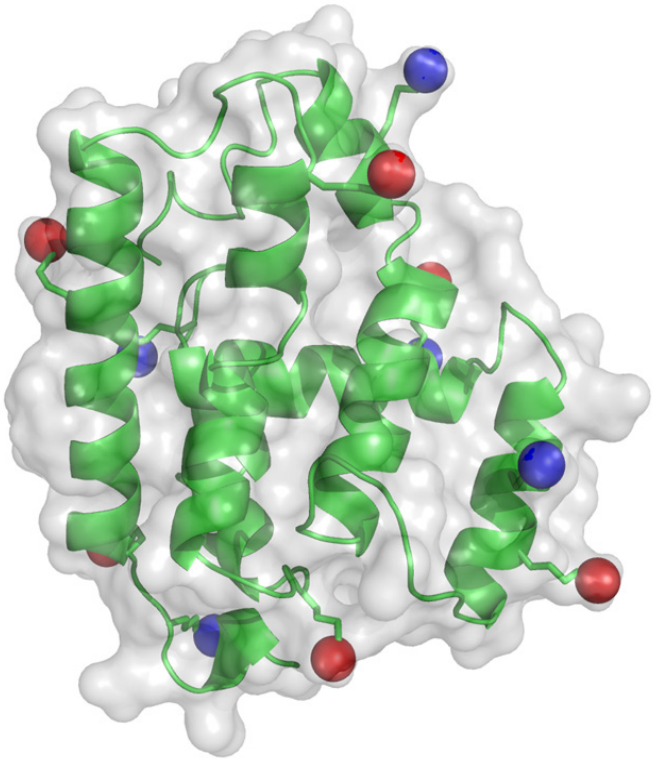


fig. S9. Surface modification experimental scheme. Proteins are unfolded with 8M urea, and then both the folded and unfolded forms of the protein are modified with either the heavy or light version of PCAS-$^{12}C_6$/$^{13}C_6$. Samples are then combined pairwise, digested, and analyzed by LC-MS. Residues which exhibit a higher level of modification in the unfolded protein than in the folded protein are more likely to be buried in the folded protein than residues which exhibit similar levels of modification in both the folded and unfolded states of the protein.

A



B

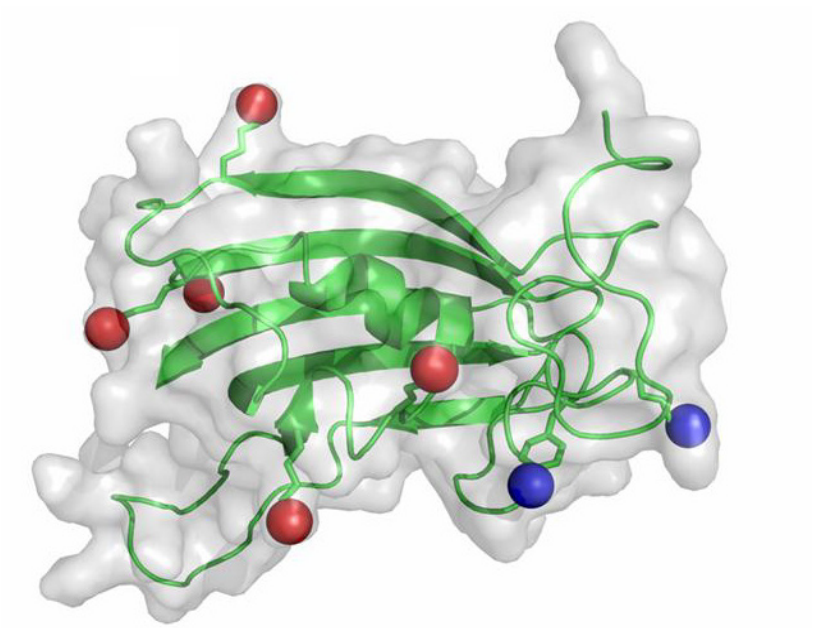
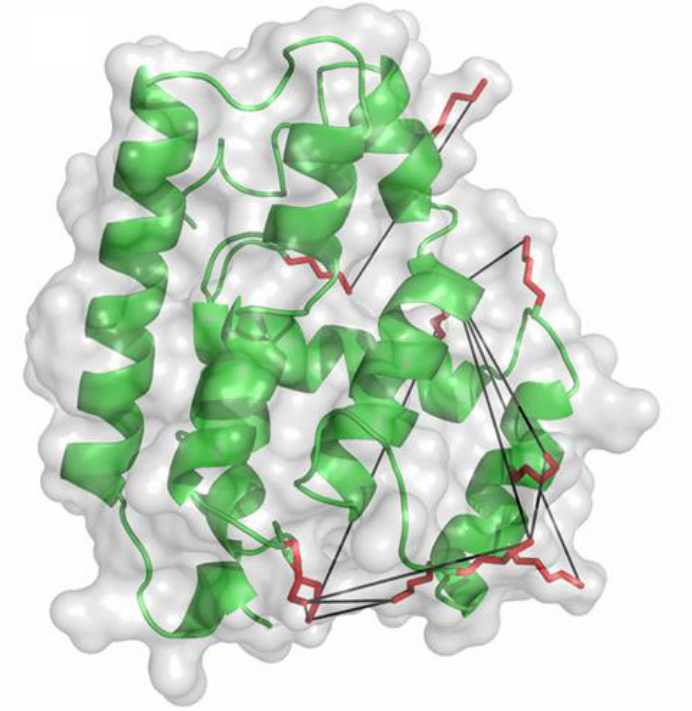


fig. S10. Surface modification results for A) Mb and B) FKBP. Residues with heavy/light peak intensity ratios of ≥1.5 are shown as red spheres, while those with heavy/light peak intensity ratios below 1.5, shown here as blue spheres.

A



B

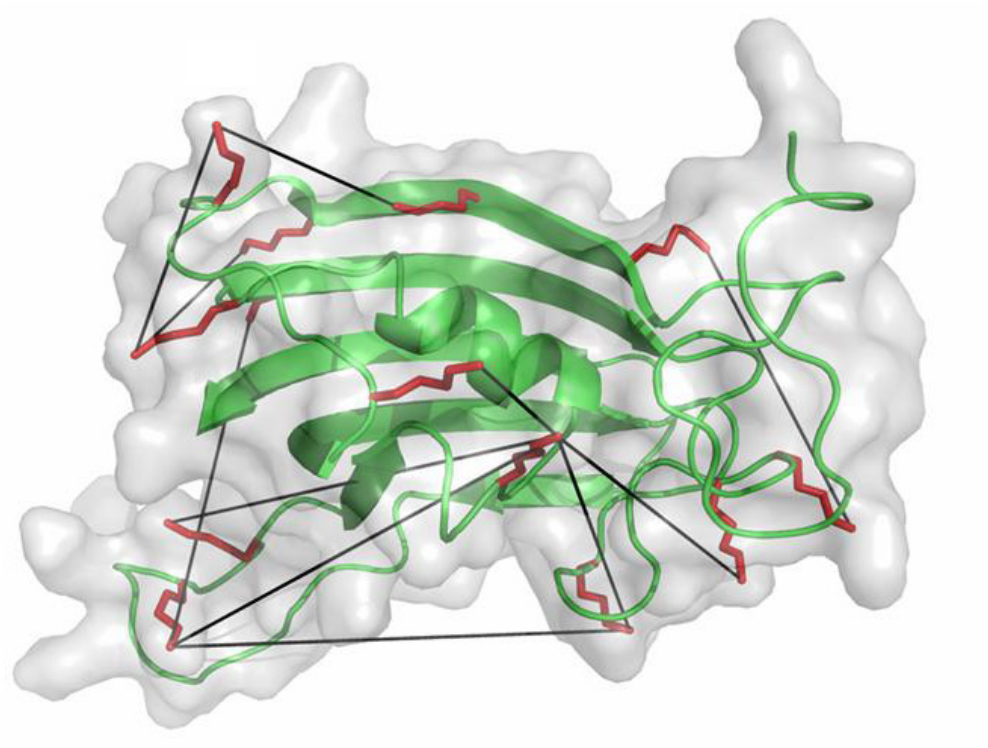


fig. S11. LD-CL analysis using CBDPS for A) Mb and B) FKBP.
All results were well within the 25Å maximum range for CBDPS cross-links.