Structural basis for high affinity binding of LEDGF PWWP to mononucleosomes

Jocelyn O. Eidahl, Brandon L. Crowe, Justin A. North, Christopher J. McKee, Nikoloz Shkriabai, Lei Feng, Matthew Plumb, Robert L. Graham, Robert J. Gorelick, Sonja Hess, Micheal G. Poirier, Mark P. Foster, Mamuka Kvaratskhelia

Supplemental Information

Table S1. LEDGF PWWP binds nucleosomes directly. The "Unweighted Spectrum Count" values for identified peptides are listed. Proteins were identified by MASCOT search engine and further organized using Scaffold3 software. Scaffold3 viewer settings were as follows: minimum identification probabilities for both protein and peptide were set at 95% resulting in a false discovery rate of less than 1%.

Identified Proteins	Load control	Pull-down
LEDGF/p75	96	130
Histone H4	103	116
Histone H2A	77	109
Histone H2B	45	33
Histone H3	21	18
Histone H1	10	17
X-ray repair cross-complementing protein 6	6	0
Ubiquinone biosynthesis monooxygenase COQ6	2	2
X-ray repair cross-complementing protein 5	4	0
DNA topoisomerase 2	11	0
FACT complex subunit SSRP1	3	0
FACT complex subunit SPT16	7	0
Topoisomerase (DNA) I	3	1
Serine/Arginine-rich splicing factor	4	0
Heterochromatin protein 1-β	9	0
DNA-dependent protein kinase catalytic subunit	6	0
Dermcidin isoform 2	1	0
NOP56 protein	4	0
H/ACA ribonucleoprotein complex subunit 1	1	0
RNA binding motif protein, X-linked, isoform CRA	2	0
SNRNP200 protein	3	0

Table S2. NMR data

NMR Constraints	
NOE-Based Distance Constraints:	
Total	**1429**
Intra-residue [i=j]	**302**
Sequential [|i-j|=1]	**421**
Medium range [1<|i-j|<5]	**245**
Long range [|i-j|>5]	**471**
Dihedral angle constraints:	**127**
Total structures calculated	**150**
Number of structures used	**20**
Residual constraint violations	
Distance violations / structure	
0.1-0.2 Å	**2.15**
0.2-0.5 Å	**0.15**
> 0.5 Å	**0**
Dihedral angle violations / structure	
1-10 °	**9.90**
> 10 °	**0.35**
Ramachadran plot statistics (%)[a]	
Most favoured regions	**79.3**
Additionally allowed regions	**20.7**
Generously allowed regions	**0.01**
Disallowed regions	**0.0**
Precision (RMSD values) (Å)[b]	
All backbone atoms	**0.4**
All heavy atoms	**0.9**

[a] Calculated via Procheck.
[b] Ordered residues used ranges from 2-36 and 39-91, selected based on: Dihedral angle order parameter, with S(phi)+S(psi)≥1.8.

Supplemental Figures:

Figure S1: SDS-Page analysis of GST-tagged p75, IBD and PWWP used for pulldown experiments (1/4 input). Lane contents are as follows: lane 1 contains molecular weight marker, lanes 2-4 contain GST-tagged LEDGF/p75, PWWP and IBD, respectively.

Figure S2: The solution structure differs from coordinates deposited by Structural Genomics Consortium. A) Overlay of the NMR solution structure (Green, PDB ID: 2M16) and the crystal structure (Blue, PDB ID: 4FU6). B) Strips from the ^{15}N-edited NOESY spectrum corresponding to the H^N from residues 76-84. Grey lines indicate the position of the diagonal signal; blue lines indicate the chemical shift the H^{α} of the (i-3) residue. C) Close-up view of the crystal structure at the beginning of α3 with the sidechains of Pro71 and Phe77 shown. D) The NMR structure in the same view. E) Strips from the ^{13}C-edited aromatic region NOESY spectrum corresponding to the aromatic protons of Phe77. Position of signals corresponding to same residue are indicated by grey lines. The absence of NOEs to other protons argue strongly that the sidechain of Phe77 is not buried in the interior of the protein, as observed in the crystal, but rather points towards solvent. F) The crystal structure zoomed in to α2 with the sidechains of His20, Phe59, Pro60, Asn64, Lys67, and Tyr68 shown. G) The NMR structure in the same view as G. H) Strips from the ^{13}C-edited aromatic region NOESY spectrum corresponding to the aromatic protons of Tyr68. Grey lines indicate chemical shifts for protons of the same residue. Positions of expected NOE crosspeaks to Tyr68 in both the NMR and crystal structure are indicated by cyan lines, while expected NOEs to His20 protons from the crystal structure are indicated by red lines. The absence of these NOEs indicates the stacking between His and Tyr rings observed in the crystal is not observed in solution.

Figure S3: Comparison and analysis of the present work with deposited coordinates. A) Solution and crystal structures of LEDGF PWWP domain differ. (Top) Blue, crystal structure 4FU6; green, NMR ensemble. Despite a high degree of structural similarity, significant differences are observed in the loop between helices α2 and α3 that repackage the aromatic residues Tyr68 and Phe77. B) Superposition of the NMR solution (blue) and crystal (4FU6, pink and cyan) structures. Overall, the structures are highly similar (1.0 Å RMSD for C^{α} excluding the region spanning residues 68-86). In this region, there are significant differences in the backbone structure that accompany differences in packing of the hydrophobic core. C) Fo – Fc map contoured at 3 σ. Mesh illustrates unexplained density in the interface between two molecules in the asymmetric unit, and in the vicinity of structural differences between the solution and crystal structures.

Figure S4: The HSQC spectra for the titrations of H3K36me3 (A), the unmodified H3 peptide (B) and DNA (C). Free LEDGF PWWP is in black. LEDGF PWWP with 80:1 equivalence of H3K36me3 is in red (A); LEDGF PWWP with 80:1 equivalence of the unmodified H3 peptide is in blue (B); LEDGF PWWP with 3.2:1 equivalence of DNA is in green (C).

Figure S5: The CSP values for the titrations of H3K36me3, unmodified H3 peptide, and DNA. The H3K36me3 titration CSP are red (top). The unmodified H3 peptide titration CSP are blue (Middle). The DNA titration CSP are green (bottom). All CSP values are determined by the equation $\Delta\delta(ppm) = \sqrt{0.5\left(\Delta\delta_H{}^2 + \left(\Delta\delta_N{}^2/25\right)\right)}$. The secondary structure is show above the graph. The gradients on the right side of the graph show the coloring used in Figure 4 C and D to map the CSP data.

Figure S6: The quantitative analysis of the NMR titration experiments. A and B show CSP values for W21 and A51 with addition of increasing concentrations of the H3K36me3 peptide (red squares) and its unmodified counterpart (blue circles) to LEDGF PWWP. C and D show CSP values for K75 and F77 upon titration of DNA (green squares) into LEDGF PWWP.

Figure S7: EMSA analysis of the LEDGF PWWP interactions with DNA. Increasing concentrations of GST-LEDGF PWWP were incubated with 300 nM of native DNA (A), SMYD1 DNA (B) or non-specific DNA (C), and the complexes were analyzed by EMSA. Lane 1 in A and C contain DNA marker ranging from 250bp to 12,000bp (Bioline Quanti-Marker 1kb). To quantitatively analyze the data, the disappearance of free DNA was monitored. Each experiment was repeated at least twice and yielded highly reproducible results. Plots in (D), (E), and (F) quantitatively represent DNA binding in (A), (B), and (C), respectively.

Figure S8: A molecular model of the LEDGF PWWP:MN complex. Front (A) and side (B) views are shown. The histone core is in black, the wrapped DNA is in magenta, the electrostatic potential of LEDGF PWWP surface is shown. A zoomed view for the LEDGF PWWP binding to MN is depicted in Figure 6.

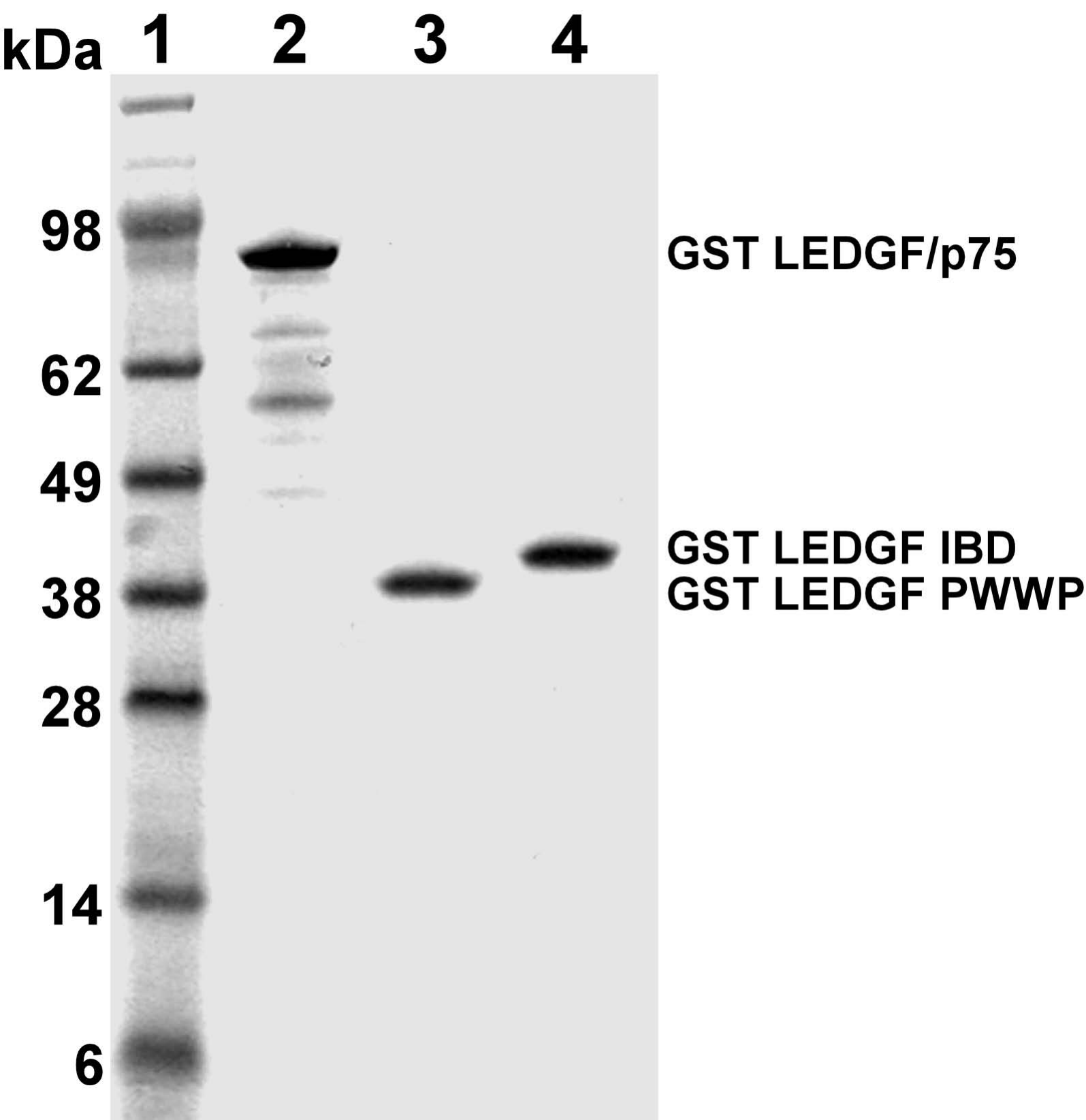
Figure S9: A) Sequence alignment of various PWWP domains. Block letters indicate invariant residues and boxes indicate residues with a global similarity scores >0.7. Alignment was performed using ClustalW (50) and colored according to similarity with a Risler scoring matrix (51) using the program ESPript (52). B) Side by side comparison of the PWWP domain from various proteins: (13,14,18,19). All PWWPs are shown in the same orientation to the β-barrel core. All PWWP domains are labeled by protein containing the PWWP domain and with the PDB ID. The structural elements that contribute to the hydrophobic cavity are shown in green. The aromatic conserved side chain located in the hydrophobic pocket of each domain is shown is blue. The C-terminal segments contributing to the top wall of the hydrophobic cavity are shown in red.

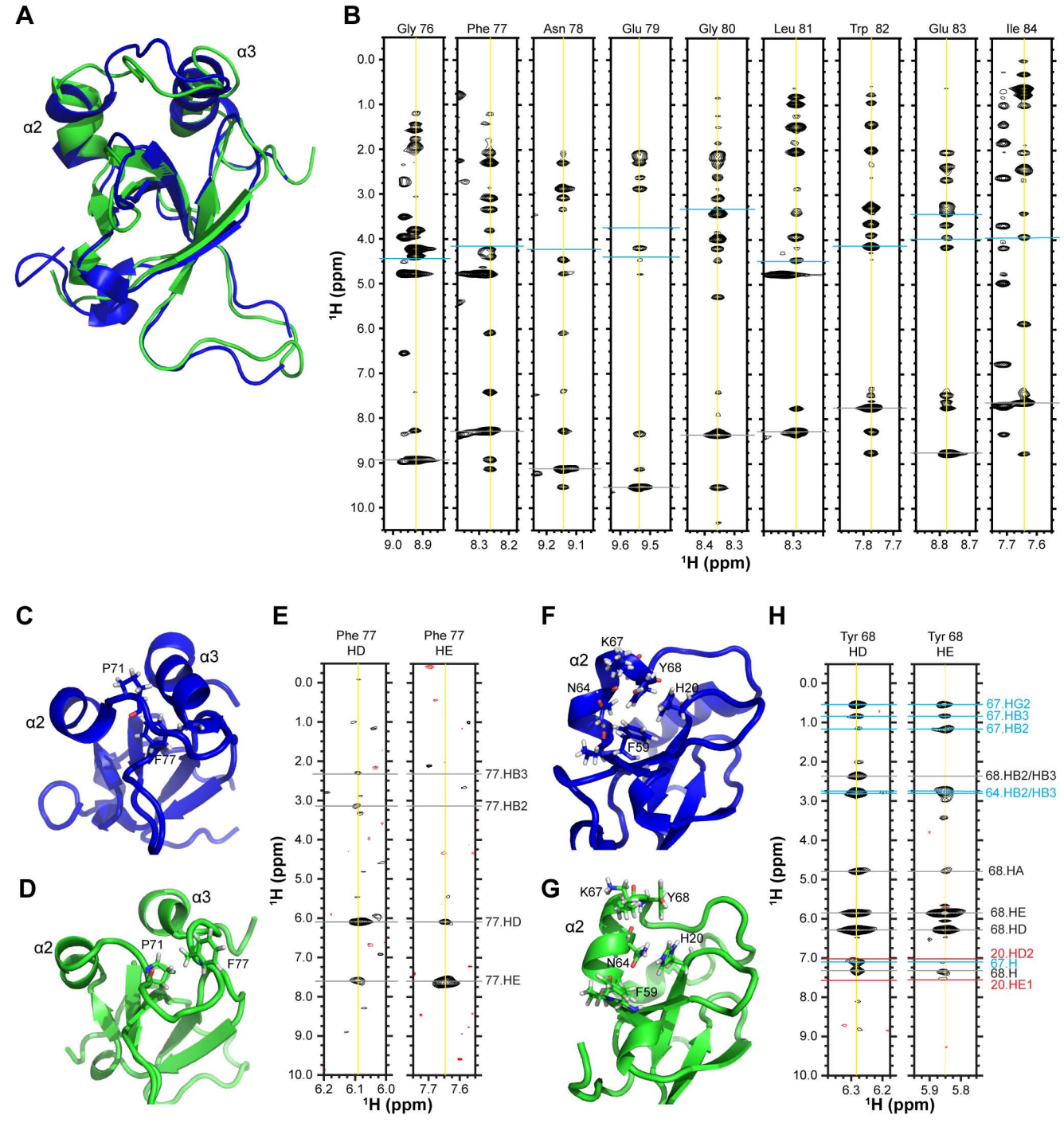
Figure S10: Side by side comparison of LEDGF PWWP domain (PDB ID: 2M16) and the Brpf1 PWWP domain with the H3K36Me3 peptide (PDB ID: 2X4W) (14). The sidechain of the conserved aromatic residue (Trp21 in LEDGF, Tyr1099 in Brpf1) located in the hydrophobic pocket in both proteins are shown in blue. The Brpf1 segments interacting with the cognate peptide, which include α-helix (1125-1139), two short β barrels (1113-1115 and 1118-1120), and their connecting loop (1121-1124) are colored red. In contrast, the corresponding region of LEDGF PWWP that interacts with the H3K36Me3 peptide contains a loop connecting β2 and β3 (colored red).

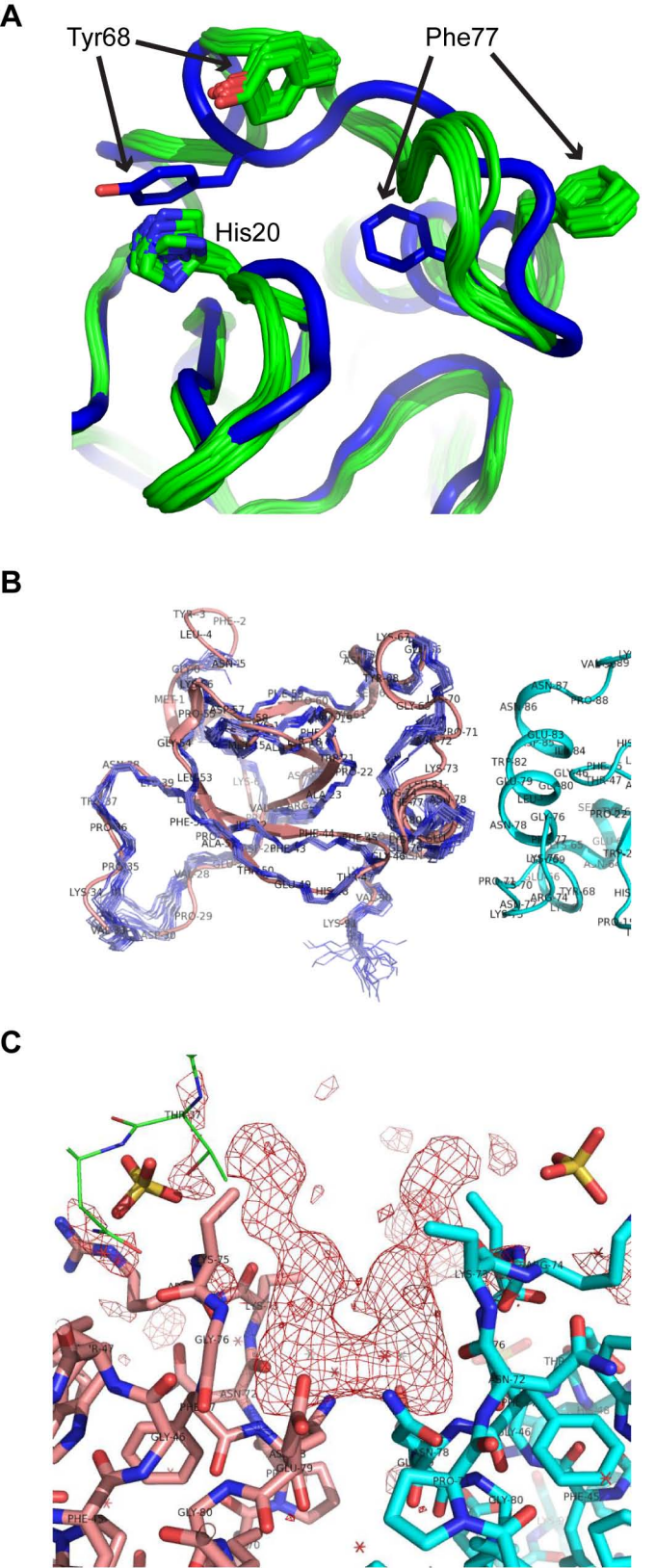
References:

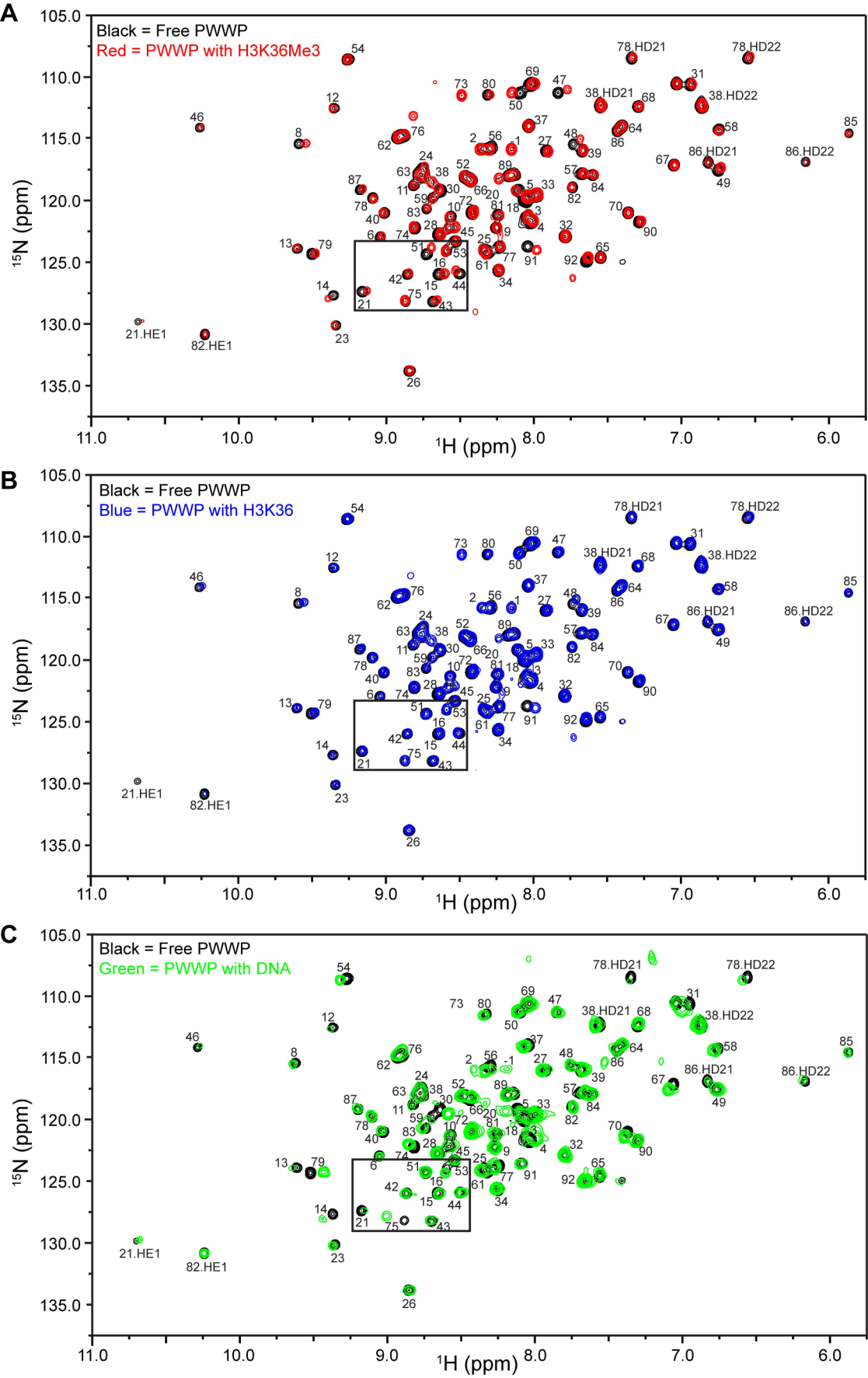
13. Qiu, Y., Zhang, W., Zhao, C., Wang, Y., Wang, W., Zhang, J., Zhang, Z., Li, G., Shi, Y., Tu, X. *et al.* (2012) Solution structure of the Pdp1 PWWP domain reveals its unique binding sites for methylated H4K20 and DNA. *Biochem J*, **442**, 527-538.
14. Vezzoli, A., Bonadies, N., Allen, M.D., Freund, S.M., Santiveri, C.M., Kvinlaug, B.T., Huntly, B.J., Gottgens, B. and Bycroft, M. (2010) Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. *Nat Struct Mol Biol*, **17**, 617-619.
18. Wu, H., Zeng, H., Lam, R., Tempel, W., Amaya, M.F., Xu, C., Dombrovski, L., Qiu, W., Wang, Y. and Min, J. (2011) Structural and histone binding ability characterizations of human PWWP domains. *PloS one*, **6**, e18919.
19. Lukasik, S.M., Cierpicki, T., Borloz, M., Grembecka, J., Everett, A. and Bushweller, J.H. (2006) High resolution structure of the HDGF PWWP domain: a potential DNA binding domain. *Protein Sci*, **15**, 314-323.

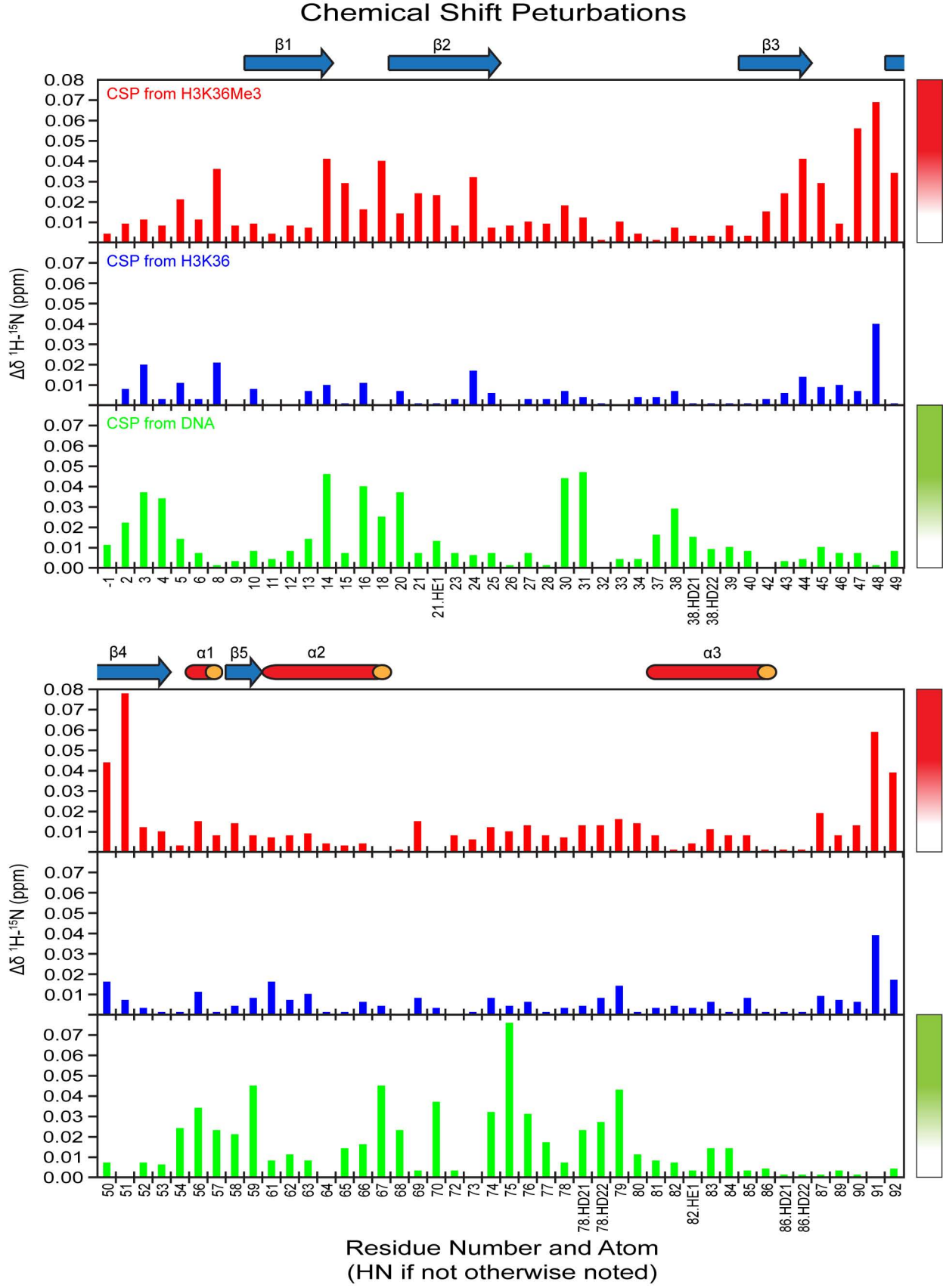
50. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22**, 4673-4680.
51. Risler, J.L., Delorme, M.O., Delacroix, H. and Henaut, A. (1988) Amino acid substitutions in structurally related proteins. A pattern recognition approach. Determination of a new and efficient scoring matrix. *J Mol Biol*, **204**, 1019-1029.
52. Gouet, P., Courcelle, E., Stuart, D.I. and Metoz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305-308.

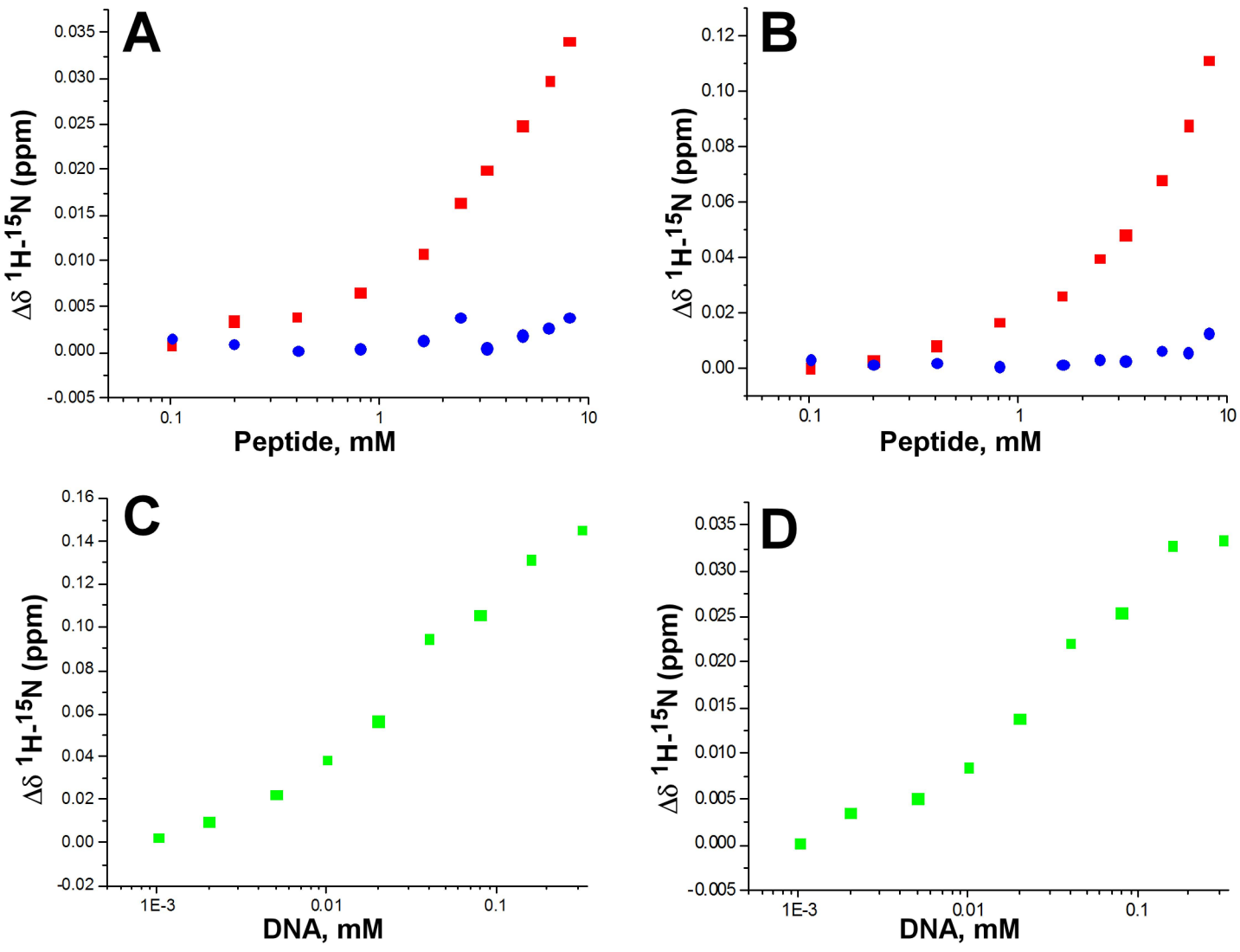

kDa
1
2
3
4
98
62
49
38
28
14
6
GST LEDGF/p75
GST LEDGF IBD
GST LEDGF PWWP

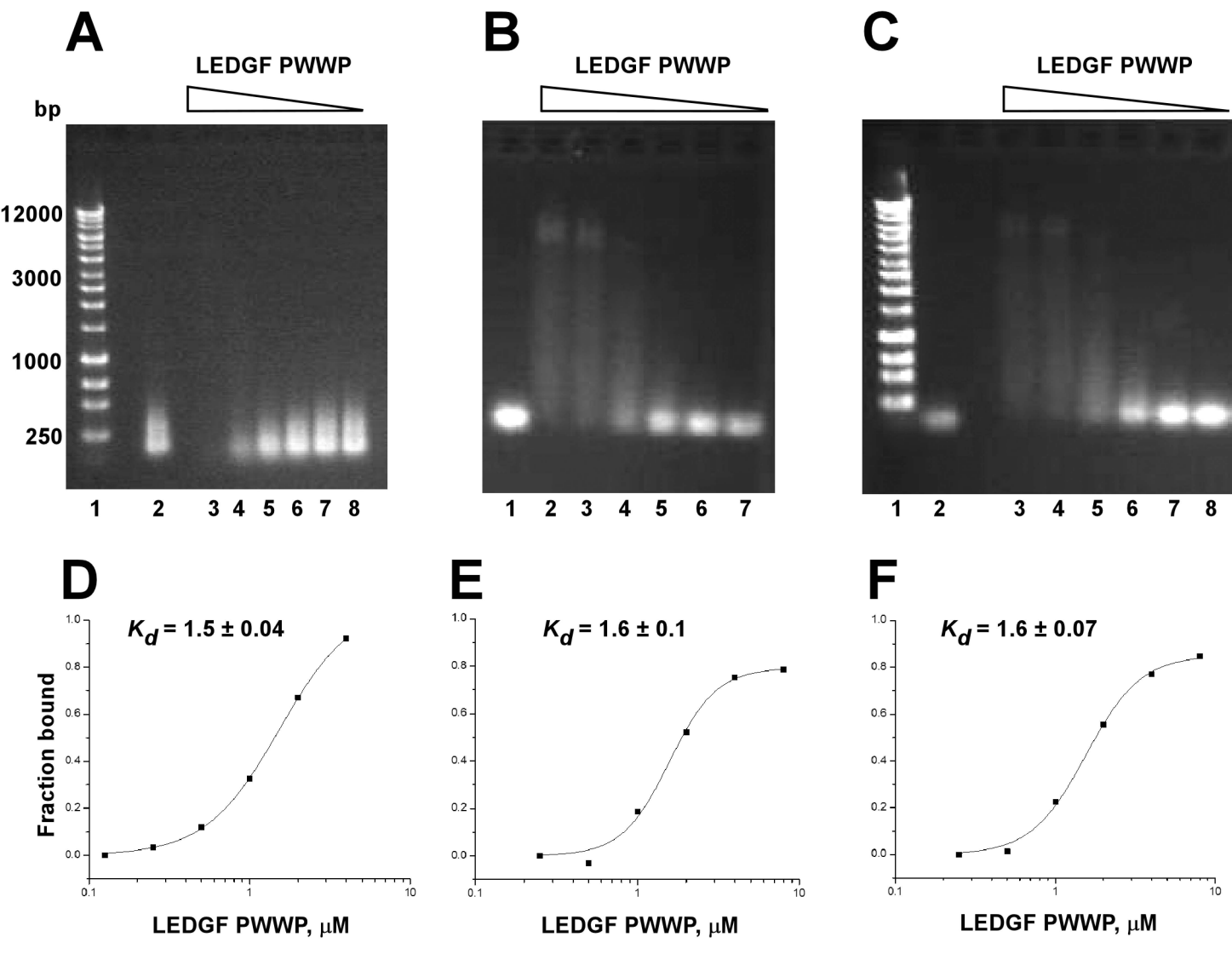

A
α3
α2
B
Gly 76
Phe 77
Asn 78
Glu 79
Gly 80
Leu 81
Trp 82
Glu 83
Ile 84
¹H (ppm)
¹H (ppm)
C
P71
α3
α2
F77
D
α3
α2
P71
F77
E
Phe 77 HD
Phe 77 HE
77.HB3
77.HB2
77.HD
77.HE
¹H (ppm)
¹H (ppm)
F
α2
K67
Y68
N64
H20
F59
G
K67
Y68
α2
H20
N64
F59
H
Tyr 68 HD
Tyr 68 HE
67.HG2
67.HB3
67.HB2
68.HB2/HB3
64.HB2/HB3
68.HA
68.HE
68.HD
20.HD2
67.H
68.H
20.HE1
¹H (ppm)
¹H (ppm)

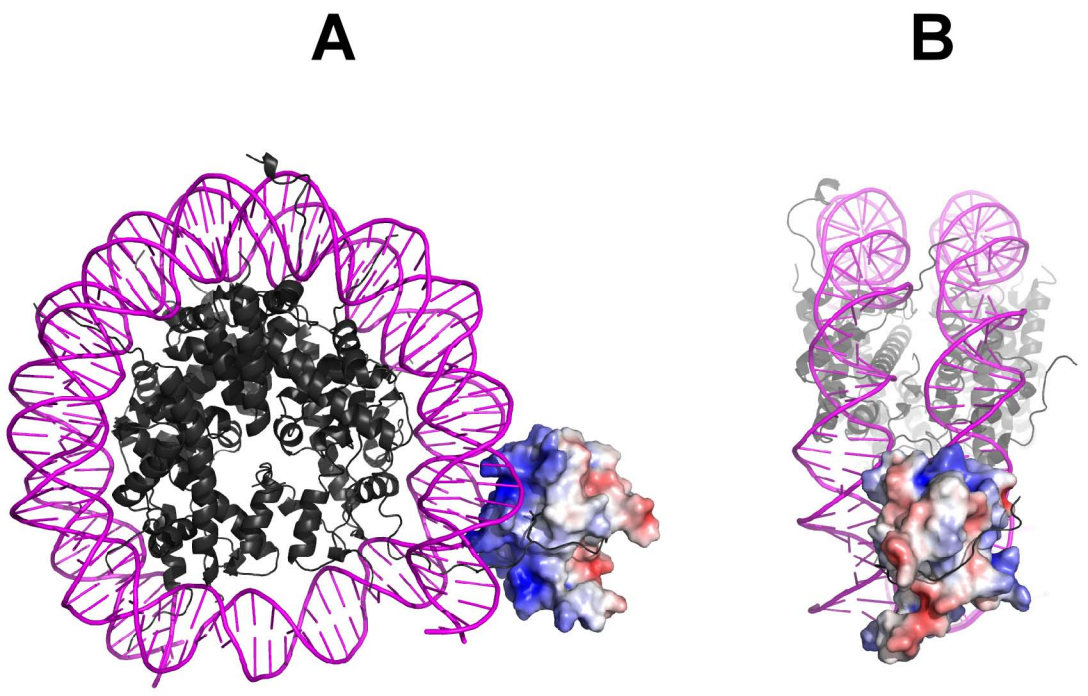

A
Tyr68
Phe77
His20
B
C

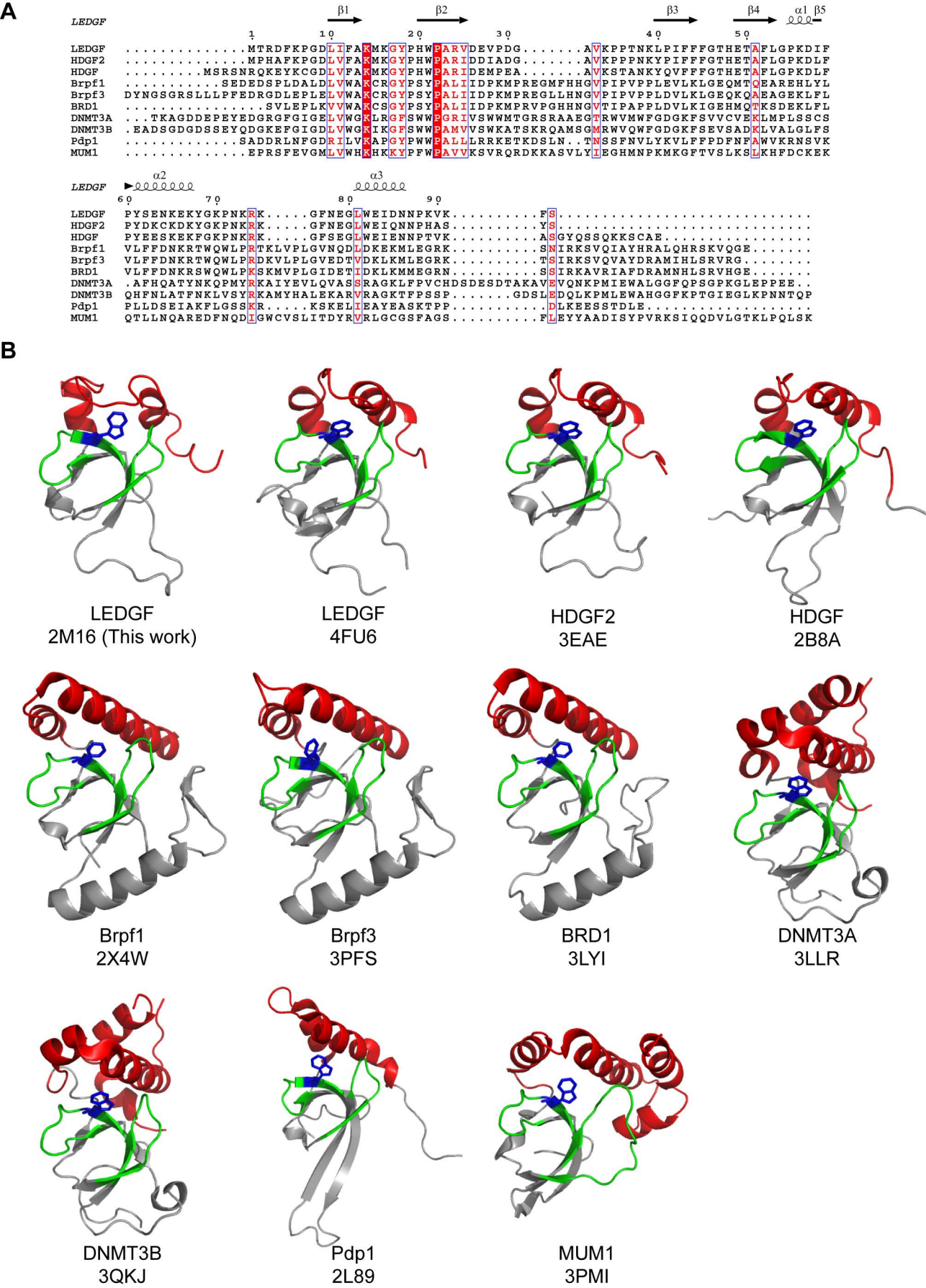

A
Black = Free PWWP
Red = PWWP with H3K36Me3
B
Black = Free PWWP
Blue = PWWP with H3K36
C
Black = Free PWWP
Green = PWWP with DNA
^{15}N (ppm)
^{1}H (ppm)
105.0
110.0
115.0
120.0
125.0
130.0
135.0
11.0
10.0
9.0
8.0
7.0
6.0
78.HD21
78.HD22
38.HD21
38.HD22
86.HD21
86.HD22
21.HE1
82.HE1

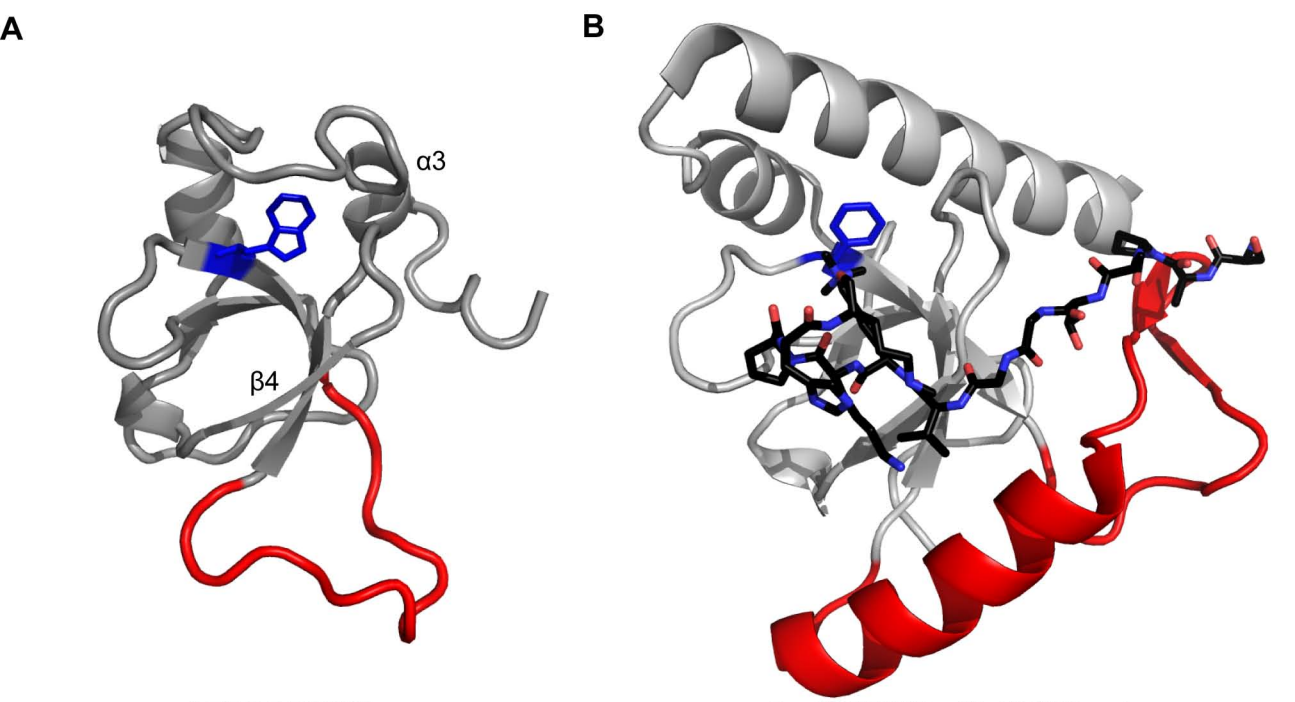

Chemical Shift Peturbations
β1
β2
β3
CSP from H3K36Me3
CSP from H3K36
CSP from DNA
Δδ ¹H-¹⁵N (ppm)
β4
α1
β5
α2
α3
Residue Number and Atom
(HN if not otherwise noted)


A
B
C
D
Δδ 1H-15N (ppm)
Peptide, mM
DNA, mM
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
-0.005
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0.16
0.14
-0.02
0.1
1
10
1E-3
0.01


A
LEDGF PWWP
bp
12000
3000
1000
250
1 2 3 4 5 6 7 8
B
LEDGF PWWP
1 2 3 4 5 6 7
C
LEDGF PWWP
1 2 3 4 5 6 7 8
D
K_d = 1.5 ± 0.04
Fraction bound
LEDGF PWWP, μM
E
K_d = 1.6 ± 0.1
LEDGF PWWP, μM
F
K_d = 1.6 ± 0.07
LEDGF PWWP, μM


A
B


A
LEDGF
β1
β2
β3
β4
α1
β5
α2
α3
LEDGF
HDGF2
HDGF
Brpf1
Brpf3
BRD1
DNMT3A
DNMT3B
Pdp1
MUM1
B
LEDGF
2M16 (This work)
LEDGF
4FU6
HDGF2
3EAE
HDGF
2B8A
Brpf1
2X4W
Brpf3
3PFS
BRD1
3LYI
DNMT3A
3LLR
DNMT3B
3QKJ
Pdp1
2L89
MUM1
3PMI



LEDGF PWWP

Brpf1 PWWP with H3K36me3