Electronic Supplementary Material (ESI) for Molecular BioSystems.

Supplementary information

Evolutionarily conserved and conformationally constrained short peptides might serve as DNA-recognition elements in intrinsically disordered regions

Nitish Tayal, Preeti Choudhary, Shashi B. Pandit, Kuljeet Singh Sandhu*

Department of Biological Sciences
Indian Institute of Science Education and Research (IISER) – Mohali,
Knowledge City, Sector-81, SAS Nagar, Manauli PO 140306, India.

*To whom correspondence should be addressed
Kuljeet Singh Sandhu
Assistant Professor & Group Leader
Department of Biological Sciences
Indian Institute of Science Education and Research (IISER) – Mohali,
Knowledge City, Sector 81, SAS Nagar, Manauli PO 140306, India.
Ph: +91-9915013645
Fax: +91-172-2240266, 2240124
E-mail: sandhuks@iisermohali.ac.in

Supplementary figure legends

Figure S1. (A) Overlap of number of LDRs predicted by IUPred and FoldIndex algorithms. (B) Venn diagram for the % of CoREs (FDR ≤ 0.01) found in major life kingdoms represented in UniProt. (C) % distribution of random hexa-peptides picked from LDR containing proteins.

Figure S2. (A) Ramachandran plot showing the ϕ-Ψ ranges for extended sheet (E), right handed helical (R) and left handed (L) conformations. Letter "U" signifies "unassigned". [R=(ϕ:-120° to -10°; ψ:-120° to 20°), E=(ϕ:-180° to -60°;ψ: 60° to 180°), L= (ϕ: 0° to 90°, ψ: -10° to 90°)]. (B) Conservation analysis for DSSP assigned secondary structures of CoREs and the adjacent regions. (C) Conservation analysis for the surface accessibility of CoREs and the adjacent regions. (D) Distributions of End-to-end CA distances for CoREs and neighboring regions of the same lengths.

Figure S3. Structural analyses of control peptides (scrambled CoREs). (A) End to end distances, (B) deviation in end-to end distances and, (C) root mean squared deviation of superimpositions of scrambled CoREs found in Protein Data Bank. P-values were calculated using Mann-Whitney U test. (D)Examples of control peptides.

Figure S4. (A) Enrichment of protein domains in CoRE containing human proteins, when compared to LDR containing human proteins. The analysis was done using FatiGO. (B) Prediction of DNA Binding potential of CoREs in yeast and fruit-fly. (C) Predictions for the effect of mutation using SIFT (left panel, number of amino acid substitutions showing the SIFT score < 0.05). P-values were calculated using Mann-Whitney U test.

Figure S5. Additional worked-out examples for DNA binding function of CoREs . (A) DNA-bound examples of CoREs. DNA is represented as blue colored space-filled model. CoRE in the protein is highlighted in red color. (B) ΔΔG values (methods) for change in DNA binding upon substitution mutations. ΔΔG values for mutations in CoRE regions are highlighted in red.

Figure S6. (A-D) Bibliographic examples of CoREs, their DNA binding potential and mutations therein. Upper panel in each figure shows predicted intrinsic disorder (red: disordered, grey: ordered) and lower panel shows magnified CoRE region, DNA binding residues (red, upper case) as predicted by BindN/DP-Bind and known loss-of-function mutations.

Figure S7. (A-B) The case study of AT-Hook. (A) Conformations of RGRP motifs alignable (RMSD < 1 A°) to DNA unbound (upper panel) and bound conformations (lower panel, PDB:). DNA-binding face is indicated in orange colour. Small solid box indicates glycyl residue. Inset picture in lower panel shows DNA bound structure of AT-hook (PDB: 2EZD) (B) Ramachandran plots of glycine, proline and arginine residues occurring in RGRP motifs in DNA unbound (upper panel) and DNA-bound (lower panel) conformation in non-homologous protein structures. (C) δ_L/δ_R ratio for the glycines in the DNA bound and unbound conformations of DNA-binding sites in all the DNA-protein complexes available in PDB.

Figure S1

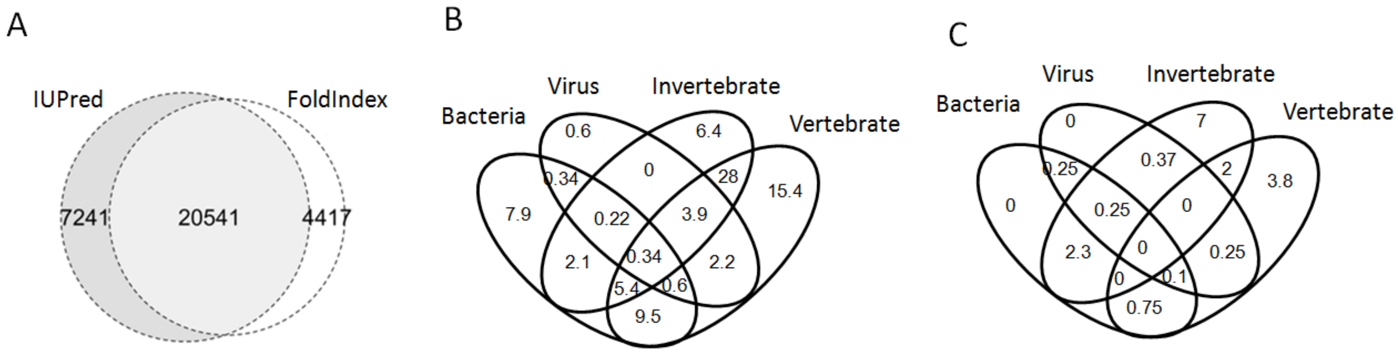

A
IUPred
FoldIndex
7241
20541
4417
B
Bacteria
Virus
Invertebrate
Vertebrate
0.6
6.4
0.34
0
28
15.4
7.9
0.22
3.9
2.1
0.34
2.2
5.4
0.6
9.5
C
Bacteria
Virus
Invertebrate
Vertebrate
0
7
0.25
0.37
2
3.8
0
0.25
0
2.3
0
0.25
0
0.1
0.75

Figure S2

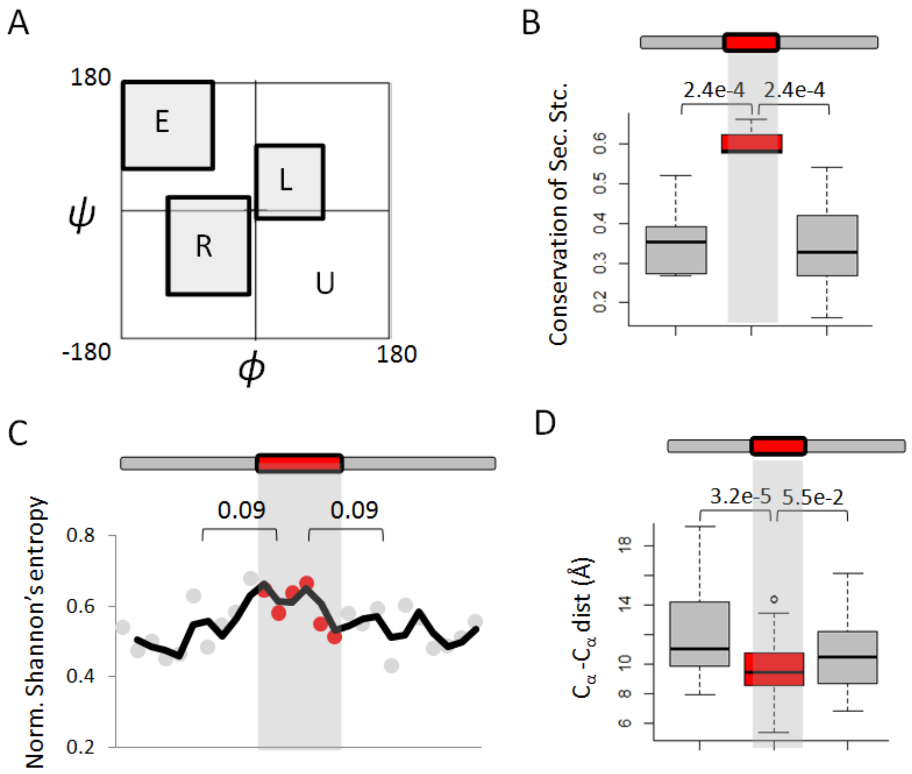

A
180
E
L
ψ
R
U
-180
φ
180
B
2.4e-4
2.4e-4
Conservation of Sec. Stc.
0.2
0.3
0.4
0.5
0.6
C
0.09
0.09
Norm. Shannon's entropy
0.8
0.6
0.4
0.2
D
3.2e-5
5.5e-2
Cα -Cα dist (Å)
6
8
10
14
18

Figure S3

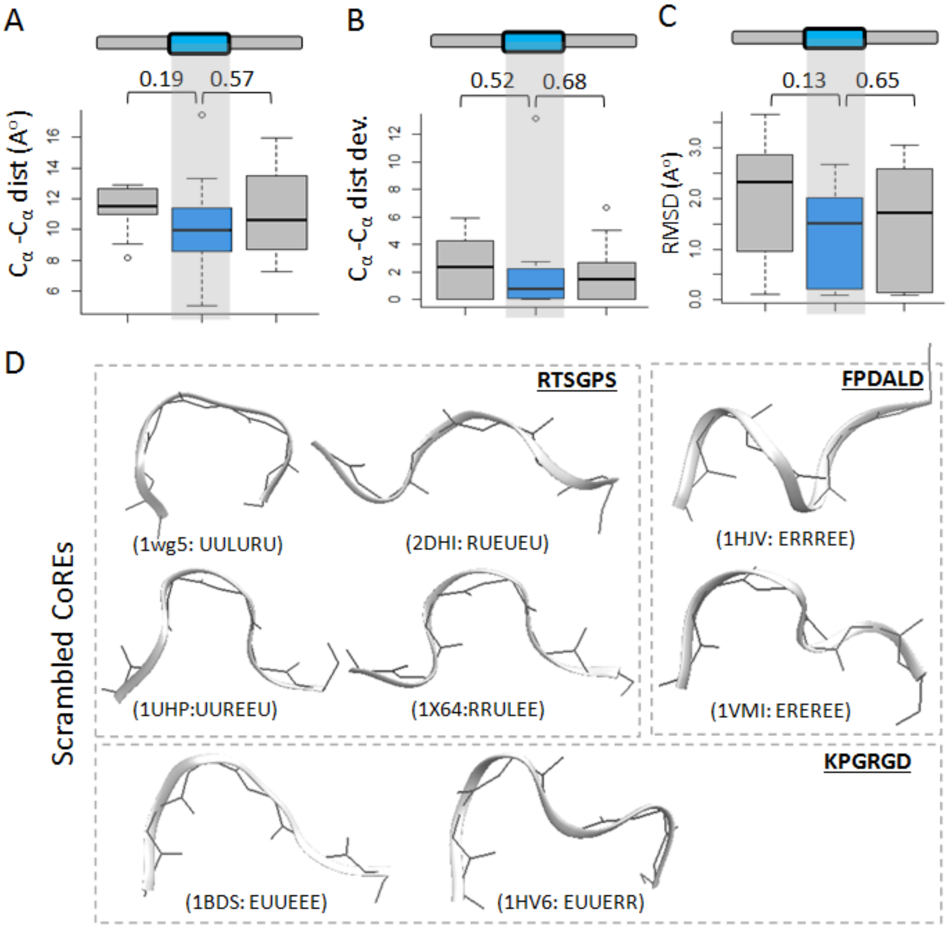

A
0.19
0.57
C_α -C_α dist (A°)
6
8
10
12
14
16
B
0.52
0.68
C_α -C_α dist dev.
0
2
4
6
8
10
12
C
0.13
0.65
RMSD (A°)
0.0
1.0
2.0
3.0
D
Scrambled CoREs
RTSGPS
(1wg5: UULURU)
(2DHI: RUEUEU)
(1UHP:UUREEU)
(1X64:RRULEE)
FPDALD
(1HJV: ERRREE)
(1VMI: EREREE)
KPGRGD
(1BDS: EUUEEE)
(1HV6: EUUERR)

Figure S4

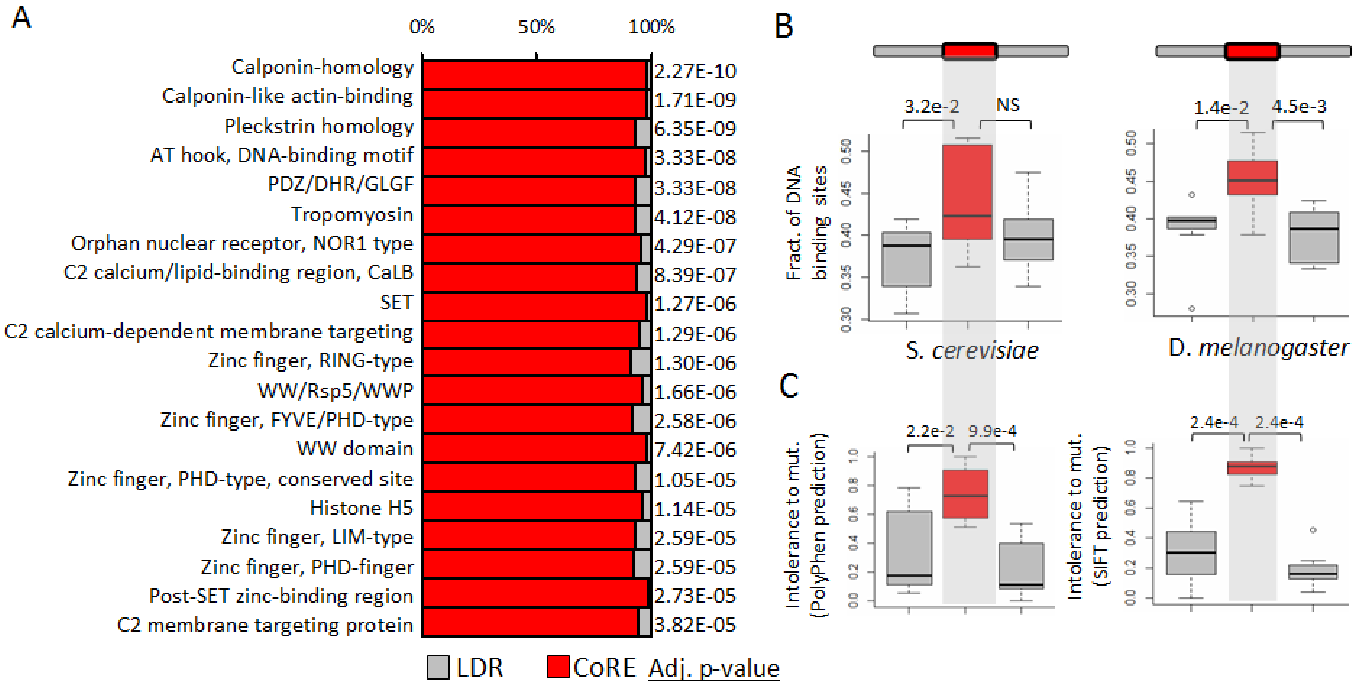
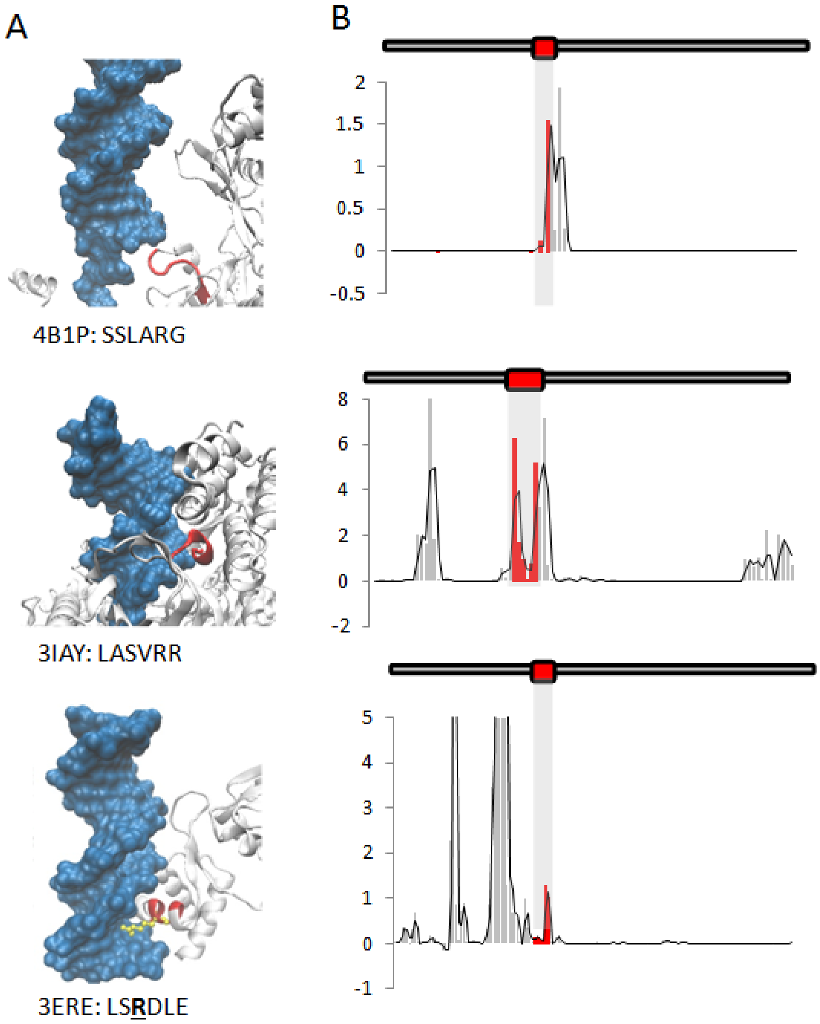

A
0%
50%
100%
Calponin-homology
Calponin-like actin-binding
Pleckstrin homology
AT hook, DNA-binding motif
PDZ/DHR/GLGF
Tropomyosin
Orphan nuclear receptor, NOR1 type
C2 calcium/lipid-binding region, CaLB
SET
C2 calcium-dependent membrane targeting
Zinc finger, RING-type
WW/Rsp5/WWP
Zinc finger, FYVE/PHD-type
WW domain
Zinc finger, PHD-type, conserved site
Histone H5
Zinc finger, LIM-type
Zinc finger, PHD-finger
Post-SET zinc-binding region
C2 membrane targeting protein
2.27E-10
1.71E-09
6.35E-09
3.33E-08
3.33E-08
4.12E-08
4.29E-07
8.39E-07
1.27E-06
1.29E-06
1.30E-06
1.66E-06
2.58E-06
7.42E-06
1.05E-05
1.14E-05
2.59E-05
2.59E-05
2.73E-05
3.82E-05
LDR
CoRE
Adj. p-value
B
3.2e-2
NS
1.4e-2
4.5e-3
Fract. of DNA binding sites
0.30
0.35
0.40
0.45
0.50
S. cerevisiae
D. melanogaster
C
2.2e-2
9.9e-4
2.4e-4
2.4e-4
Intolerance to mut. (PolyPhen prediction)
Intolerance to mut. (SIFT prediction)
0.0
0.2
0.4
0.6
0.8
1.0

Figure S5

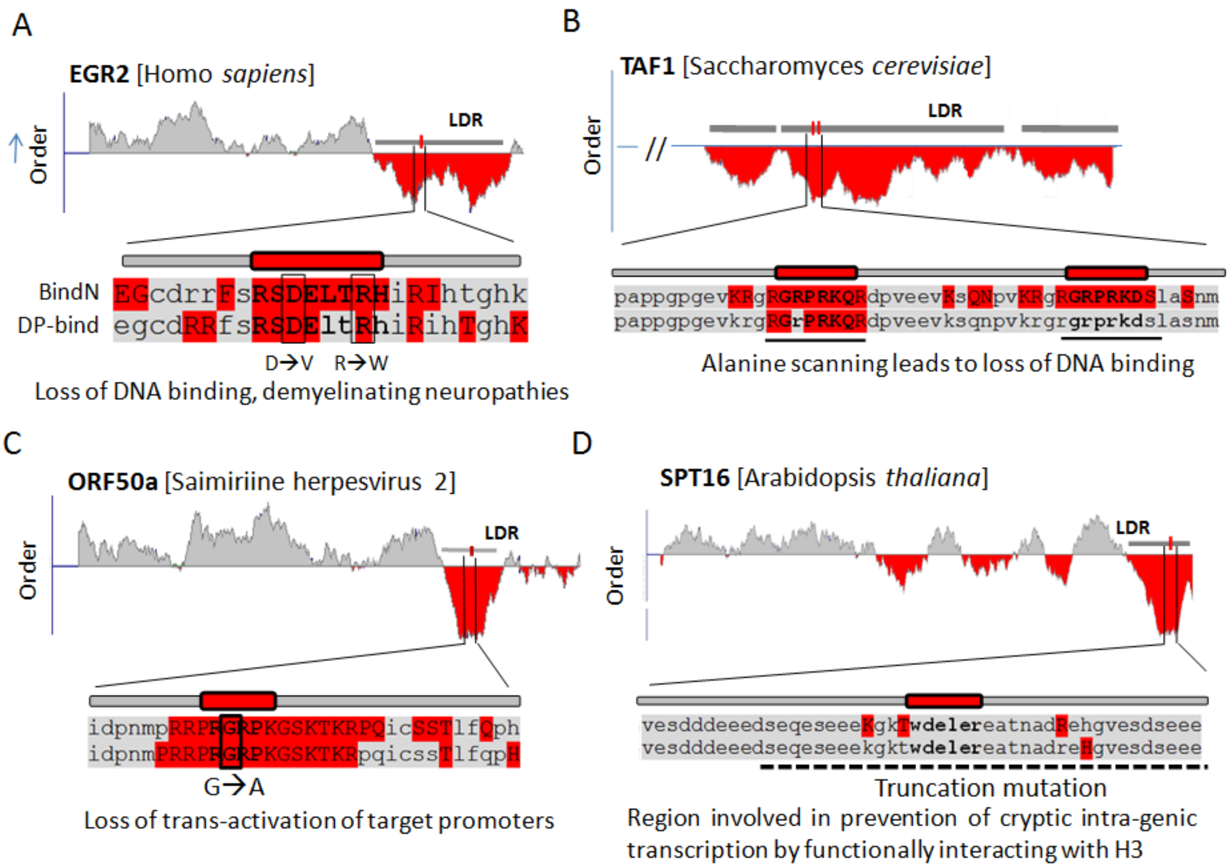


4B1P: SSLARG

3IAY: LASVRR

3ERE: LS**R**DLE

Figure S6



Loss of DNA binding, demyelinating neuropathies

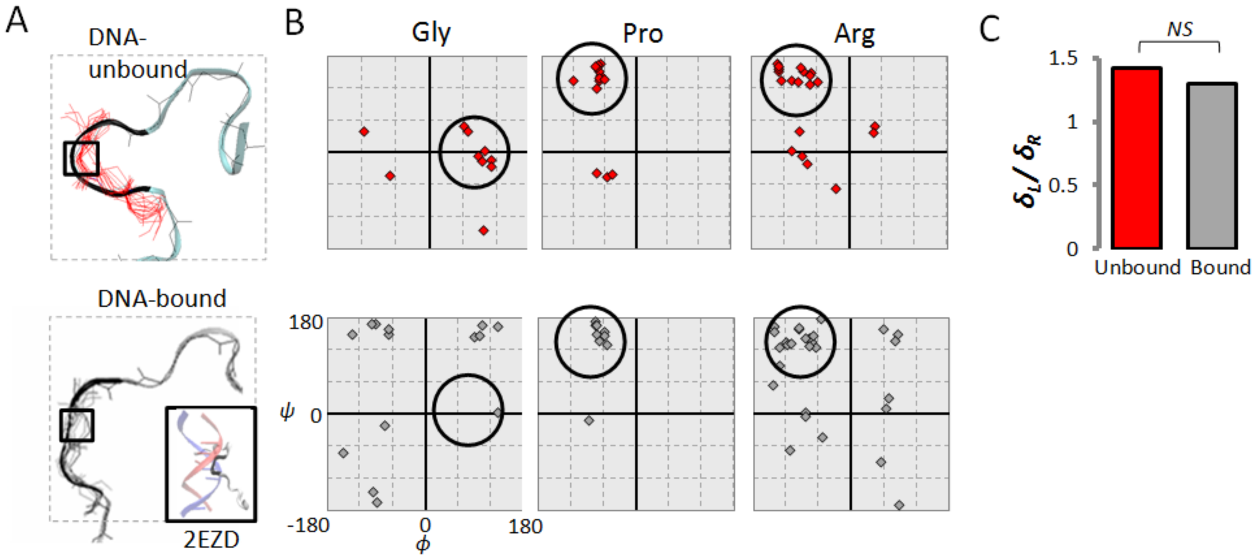
Alanine scanning leads to loss of DNA binding

Loss of trans-activation of target promoters

Truncation mutation

Region involved in prevention of cryptic intra-genic transcription by functionally interacting with H3

Figure S7


A
DNA-unbound
DNA-bound
2EZD
B
Gly
Pro
Arg
180
ψ 0
-180
0
180
ϕ
C
NS
1.5
1
0.5
0
δ_L / δ_R
Unbound
Bound