Electronic Supplementary Material (ESI) for Molecular Omics.

Supporting Information

Structural motif, *Topi* and its role in protein function and fibrillation

Jesmita Dhar,[a] Pinak Chakrabarti[a,b,*]

[a] Bioinformatics Centre, [b]Department of Biochemistry, Bose Institute, P1/12 CIT Scheme VIIM, Kolkata 700054, India

***Author for correspondence:**

E-mail: pinak@jcbose.ac.in, Tel: +91 33 2569 3253

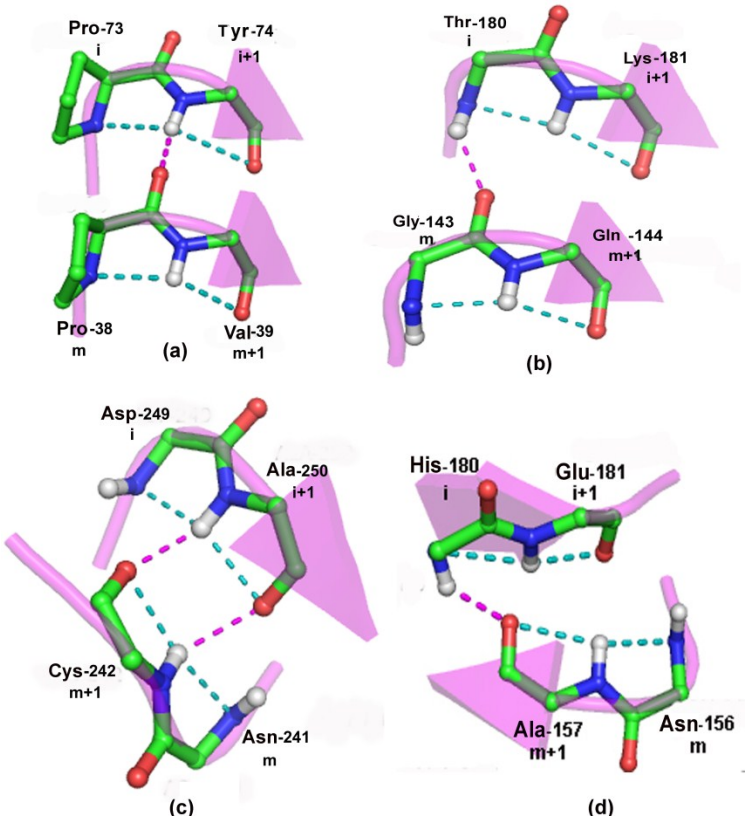


Figure S1. *Topi* motif, where two fused-rings motifs (involving residue pairs, (i, i+1) and (m,m+1)) are connected by a long range intermolecular hydrogen bond, (a) (i+1) → m , type I; (b) i→m, type II; (c) (i+1)→(m+1), type III and (d) i→(m+1), type IV[9].

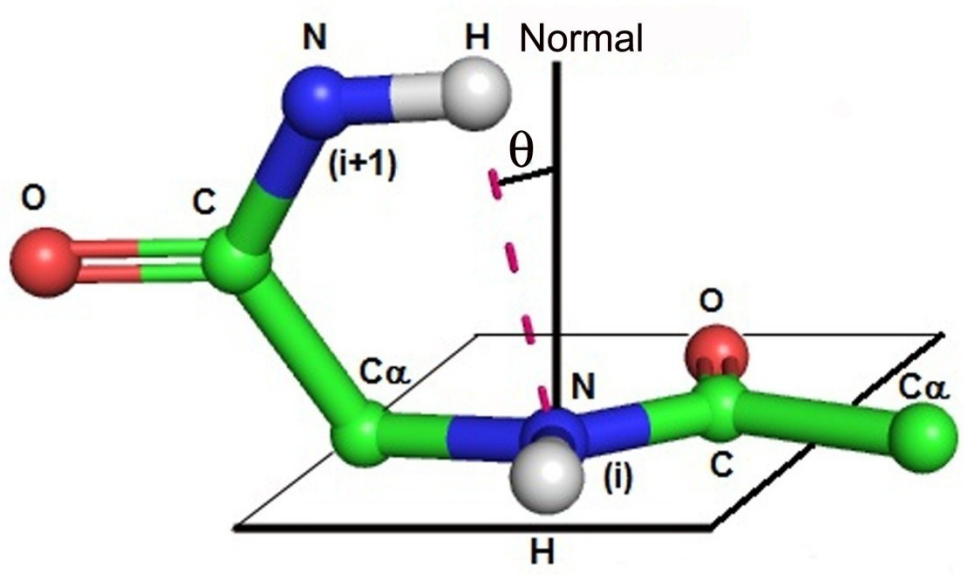


Figure S2. Definition of angle θ – the angle made by the N(i) ⋯ H(i+1) direction with the normal to the peptide plane passing through N(i).

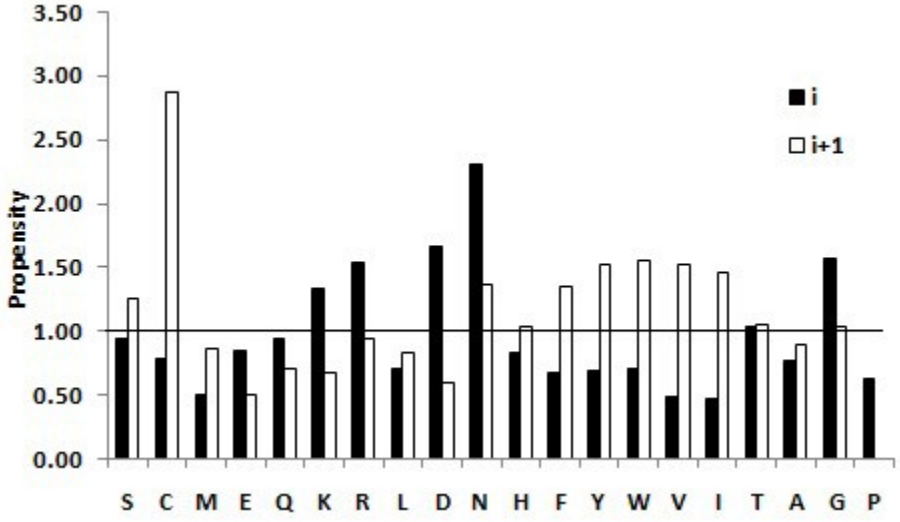


Figure S3. Propensities of residues to occur at i and i+1 positions of type I *topi*.

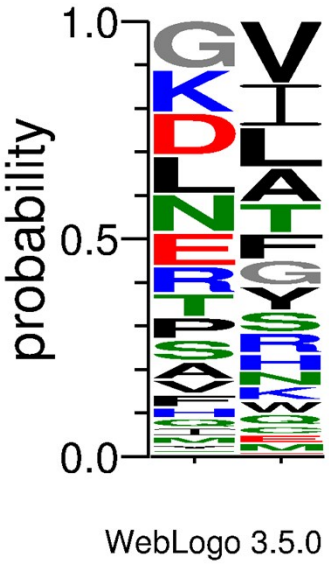


Figure S4. The WebLogo for two residues (i and i+1) (and using both the segments) in type I *topi*.

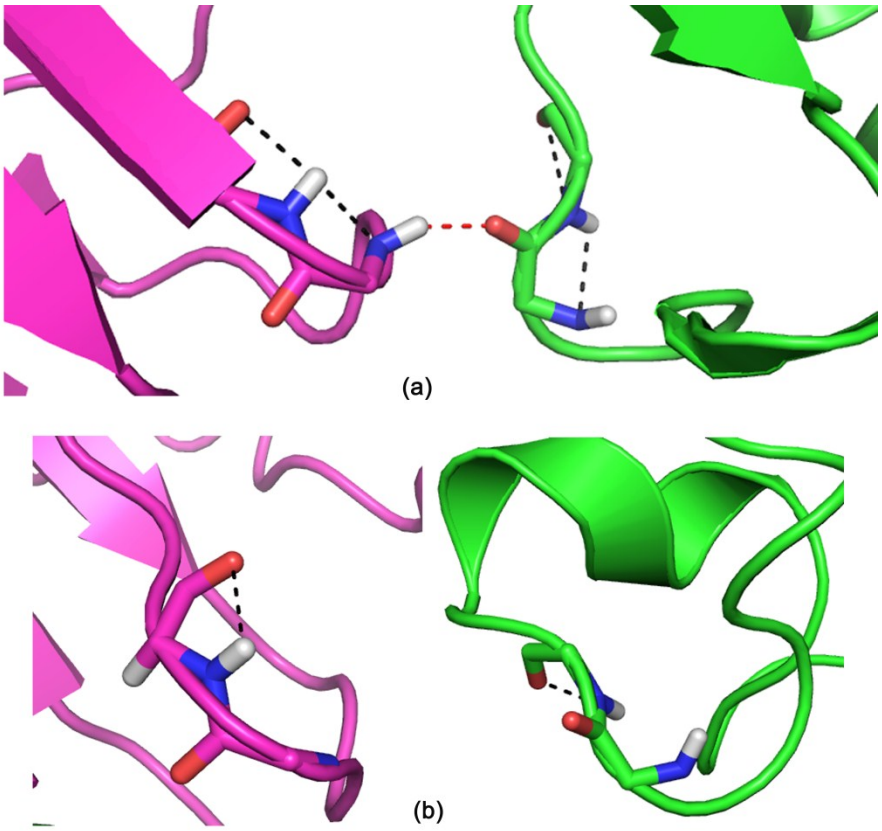


Figure S5. (a) *Topi* structure at the interface of a heterodimer (PDB: 1XU1). The fused-rings motif in the component chains is missing in (b) the unbound forms of the subunits (PDB: 1U5Y and 1XUT).

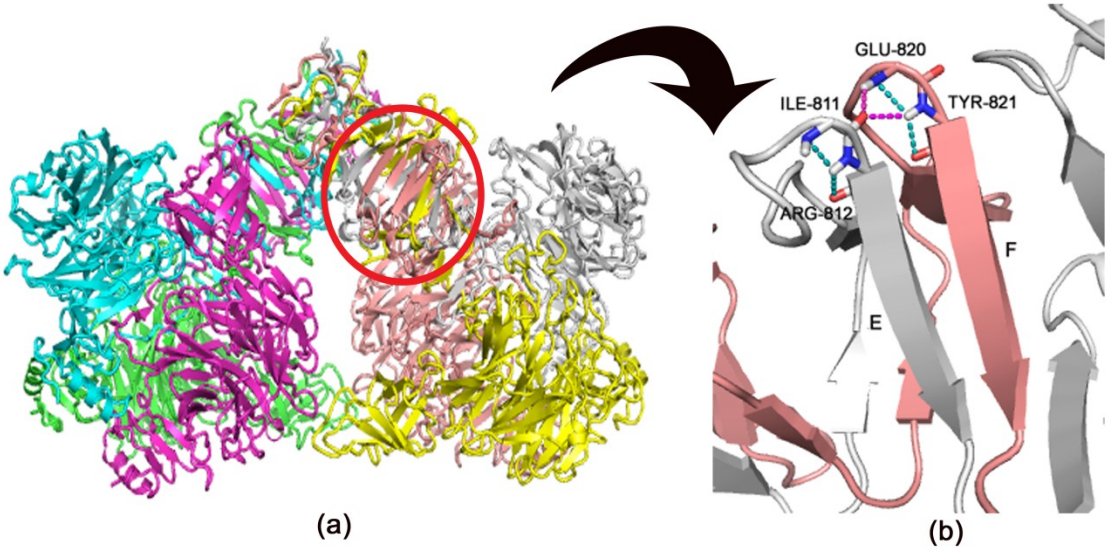
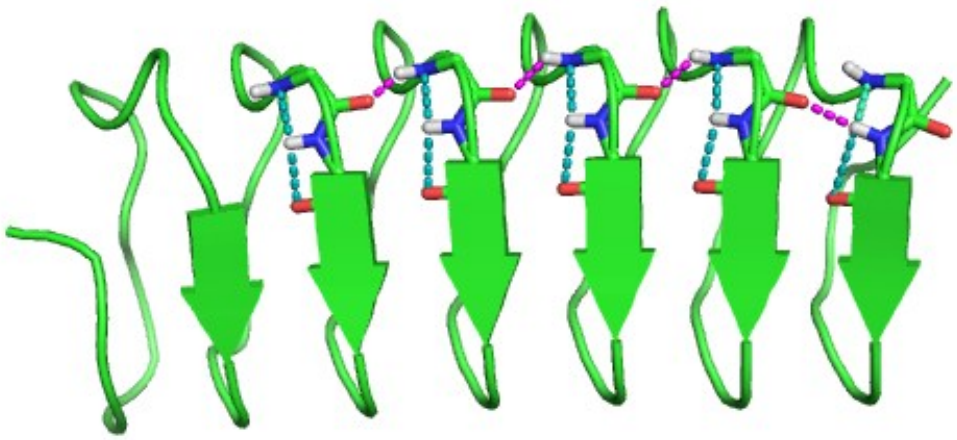
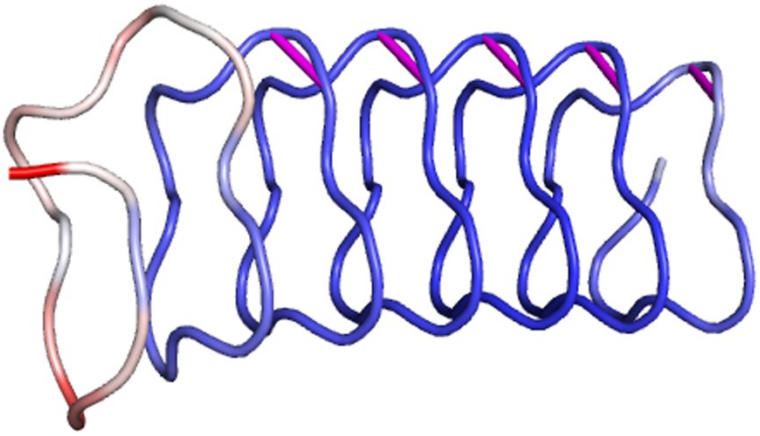


Figure S6. Occurrence of *topi* motif in the interface (marked by red circle) of two protein chains, labelled E and F, of a multimeric protein (PDB, 1V0F).



(a)



(b)

Figure S7. (a) The occurrence of a series of linked ((i+1) → m and i→m, pink dashed line) fused-rings motifs in the pentapeptide repeat protein (PDB, 1EZG, antifreeze protein from the beetle)[48]. (b) Fluctuation of the backbone (taking C, Cα, and N only) is shown in color (blue to red signifies rigid to flexible). Magenta stick indicates the position of *topi* motifs is parallel peptides.

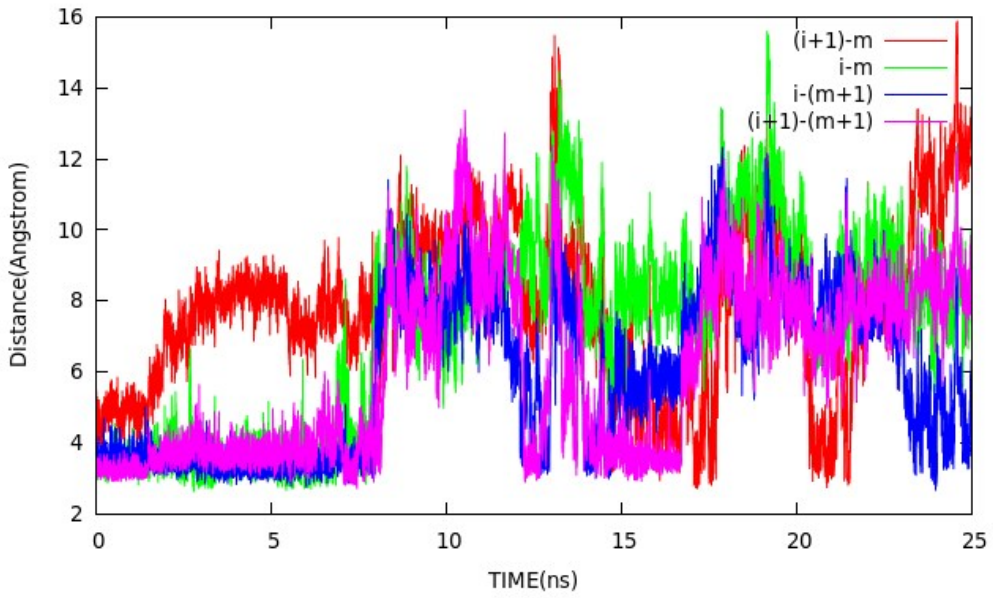


Figure S8. Different distances (i+1→m, i→m, i→m+1 or i+1→m+1) between *topi* segments, plotted against the simulation time (in ns) for the pair of 6-residue peptides (with *topi* at center). The sequence of the segment is (200)Val-Asp-His-Cys-Phe-Phe (PDB file: 1OFL).[32] The peptide segments fall apart beyond 25 ns.

Table S1. List of fused-rings motifs found in misfolded proteins

PDB ID	Chain ID	Residue (i) number	Residue (i)	Residue (i+1)
2kj3	C	221	ALA	ILE
	A	224	GLY	ARG
	B	224	GLY	ARG
	C	224	GLY	ARG
	A	229	LYS	ASP
	B	229	LYS	ASP
	C	229	LYS	ASP
	B	236	ARG	ALA
	B	251	HIS	GLY
	A	252	GLY	GLY
	A	254	ILE	ARG
	B	254	ILE	ARG
	B	257	SER	ASP
	A	265	GLU	THR
	B	265	GLU	THR
	A	272	GLU	SER
	A	289	ASN	HIS
	B	292	HIS	HIS
2kkw	A	39	TYR	VAL
	A	94	PHE	VAL
	A	98	ASP	GLN
	A	108	PRO	GLN
	A	110	GLU	GLY
2lmp	L	23	ASP	VAL
	N	23	ASP	VAL
	Q	23	ASP	VAL
	C	25	GLY	SER
	F	25	GLY	SER
	G	25	GLY	SER
	H	25	GLY	SER
	K	25	GLY	SER
	L	25	GLY	SER
	O	25	GLY	SER
	Q	25	GLY	SER
	R	25	GLY	SER
	A	29	GLY	ALA
	C	29	GLY	ALA
	R	29	GLY	ALA

	E	38	GLY	VAL
	F	38	GLY	VAL
	I	38	GLY	VAL
	L	38	GLY	VAL
	O	38	GLY	VAL
	Q	38	GLY	VAL
2n0a	H	3	VAL	PHE
	B	4	PHE	MET
	C	6	LYS	GLY
	D	6	LYS	GLY
	E	11	ALA	LYS
	I	11	ALA	LYS
	G	12	LYS	GLU
	J	12	LYS	GLU
	B	14	GLY	VAL
	F	14	GLY	VAL
	C	25	GLY	VAL
	J	26	VAL	ALA
	E	27	ALA	GLU
	C	28	GLU	ALA
	H	28	GLU	ALA
	D	29	ALA	ALA
	E	99	GLN	LEU
	J	99	GLN	LEU
	H	103	ASN	GLU
	J	104	GLU	GLU
	F	109	GLN	GLU
	I	109	GLN	GLU
	G	110	GLU	GLY
	H	111	GLY	ILE
	B	112	ILE	LEU
	C	117	PRO	VAL
	I	117	PRO	VAL
	F	120	PRO	ASP
	G	121	ASP	ASN
	D	123	GLU	ALA
	B	125	TYR	GLU
	I	125	TYR	GLU
	B	128	PRO	SER
	C	128	PRO	SER

	I	129	SER	GLU
	G	134	GLN	ASP
	J	134	GLN	ASP
	B	139	GLU	ALA
	I	139	GLU	ALA
3sgr	A	13	GLY	LYS
	B	13	GLY	LYS
	C	13	GLY	LYS
	D	13	GLY	LYS
	E	13	GLY	LYS
	F	13	GLY	LYS
1v0f	E	264	ASP	THR
	F	264	ASP	THR
	E	281	THR	SER
	F	281	THR	SER
	E	288	ARG	PHE
	F	288	ARG	PHE
	E	291	ASN	THR
	F	291	ASN	THR
	E	298	ARG	ILE
	F	298	ARG	ILE
	E	311	PHE	VAL
	F	311	PHE	VAL
	E	319	ILE	THR
	F	319	ILE	THR
	E	338	ASN	VAL
	F	338	ASN	VAL
	E	348	ASP	ARG
	F	348	ASP	ARG
	E	367	GLN	THR
	F	367	GLN	THR
	F	375	LEU	THR
	E	384	THR	VAL
	F	384	THR	VAL
	E	390	MET	SER
	F	390	MET	SER
	E	397	ASN	ARG
	F	397	ASN	ARG
	E	432	GLY	ILE
	F	432	GLY	ILE

	E	438	ASN	GLN
	F	438	ASN	GLN
	E	440	ARG	TYR
	F	440	ARG	TYR
	E	450	GLY	LEU
	F	450	GLY	LEU
	E	463	ALA	VAL
	F	463	ALA	VAL
	E	477	ILE	ASP
	F	477	ILE	ASP
	E	480	ASP	ASN
	F	480	ASP	ASN
	E	487	PRO	ASN
	F	487	PRO	ASN
	E	491	THR	SER
	F	491	THR	SER
	F	498	GLY	LYS
	E	509	LYS	SER
	F	509	LYS	SER
	E	522	SER	VAL
	F	522	SER	VAL
	E	545	ASP	VAL
	F	545	ASP	VAL
	E	557	PRO	ASP
	F	557	PRO	ASP
	E	578	ASP	ALA
	F	578	ASP	ALA
	E	589	GLY	VAL
	F	589	GLY	VAL
	E	598	THR	ARG
	F	598	THR	ARG
	E	601	ASP	ARG
	F	601	ASP	ARG
	E	615	GLN	THR
	F	615	GLN	THR
	E	628	HIS	THR
	F	628	HIS	THR
	E	638	ASP	ASP
	F	638	ASP	ASP
	E	647	ARG	ALA

	F	647	ARG	ALA
	F	659	ASP	ARG
	E	661	TYR	LYS
	F	661	TYR	LYS
	E	678	TRP	ASN
	F	678	TRP	ASN
	E	696	GLY	ILE
	F	696	GLY	ILE
	E	712	ASN	TYR
	F	712	ASN	TYR
	E	738	PHE	LYS
	F	738	PHE	LYS
	E	758	ASN	ARG
	F	758	ASN	ARG
	E	767	GLY	ALA
	E	773	ALA	VAL
	E	782	GLY	VAL
	F	782	GLY	VAL
	E	811	ILE	ARG
	F	811	ILE	ARG
	E	820	GLU	TYR
	F	820	GLU	TYR
	E	842	GLY	GLY
	F	842	GLY	GLY
	E	844	GLU	GLY
	F	844	GLU	GLY
	E	850	THR	GLY
	F	850	THR	GLY
	E	872	ASP	GLU
	F	872	ASP	GLU
	E	884	TYR	ASN
	F	884	TYR	ASN
	F	896	ASN	ARG
	E	899	THR	THR
	F	899	THR	THR
	E	904	GLY	SER
	F	904	GLY	SER
2beg	B	27	ASN	LYS
	C	27	ASN	LYS
	E	17	LEU	VAL

1k28	D	18	ASP	SER
	A	34	LEU	HIS
	A	62	THR	SER
	A	65	ALA	MET
	A	68	GLY	ILE
	A	73	VAL	THR
	D	81	ASN	ASN
	A	87	LEU	ASP
	D	91	LYS	HIS
	A	97	LEU	GLY
	A	105	ARG	GLU
	A	111	LEU	GLU
	A	114	PHE	SER
	D	115	ILE	GLU
	A	120	GLN	TYR
	A	123	ARG	ARG
	D	126	PRO	ASP
	D	183	ASP	LYS
	A	188	LEU	LYS
	D	237	LEU	THR
	D	239	LYS	SER
	D	259	LEU	ASP
	D	261	SER	SER
	D	272	GLU	ASN
	D	279	ARG	SER
	D	289	ARG	ASN
	D	321	GLY	VAL
	D	332	GLN	PHE
	A	334	GLY	ASN
	D	334	LYS	THR
	D	340	ASP	GLU
	D	361	ALA	THR
	A	371	ASP	SER
	A	375	LEU	PHE
	A	398	GLY	HIS
	A	467	MET	ASP
	A	555	ALA	SER
	A	570	SER	ARG
1z0q	A	23	ASP	VAL