JPP JOURNAL OF PROTEINS AND PROTEOMICS
3(2), July-December 2012, pp. 119-127

RESEARCH ARTICLE

SEQUENCE COMPOSITION OF BINDING SITES IN NATIVELY UNFOLDED HUMAN PROTEINS

Mritunjoy Maity and Nakul C. Maiti*

Structural Biology and Bioinformatics Division, CSIR-Indian Institute of Chemical Biology, Kolkata 700032, India

Abstract: Natively unfolded proteins that lack any globular fold and remain very flexible in solution state exhibit diverse cellular activity and play a significant role in cell signalling and reorganization processes. Further, binding of unstructured protein to a target molecule often causes folding and conformational alteration and functional cooperativity. Engaging several algorithm and computational methods we measured the compositional aspects of the binding region of fully unfolded native human proteins. Number of binding regions was found to increase with the protein sequence length. However, most of the proteins contained multiple binding regions with variable length. The regions were distributed throughout the protein sequence and not localized. Content of different amino acids in binding regions followed a similar trend for total protein sequences; however, the regions were enriched with Ala, Gly, Leu, Phe and Tyr. Grand average of hydropathy of all the proteins was negative, although many of their binding regions showed positive values. Most of the residues in both the proteins and in the binding region favoured random-coil conformation.

Keywords: Natively unfolded human protein; Binding region; ANCHOR

Introduction

The available genome sequences and bioinformatics prediction method revealed unique presence of natively unfolded proteins (NUPs) in eukaryotes (Dosztányi *et al.*, 2010). These proteins are also known by other names like intrinsically disordered (Orosz & Ovadi, 2011) natively denatured (Schweers *et al.*, 1994) or intrinsically unstructured proteins (Orosz & Ovadi, 2011). It is established that more than 30% of eukaryotic proteins are partially or completely disordered and disordered proteins have no consistency in their sizes.

These proteins lack any well-defined three dimensional folded structure in solution (Wright & Dyson, 1999; Uversky *et al.*, 2000; Dunker *et al.*, 2001; Uversky, 2002; Iwasa *et al.*, 2011) and remain as rapidly inter converting conformations even under physiological conditions (Weinreb *et al.*, 1996; Uversky, 2002; Ahmad *et al.*, 2005). Structurally they resemble the denatured states of ordered protein and remain very dynamic under physiological conditions (Weinreb *et al.*, 1996; Uversky, 2002; Ahmad *et al.*, 2005) and many of these proteins in solution posses no stable globular folds (Ward *et al.*, 2004) under physiological pH. The lack of a rigid stable folded structure under physiological conditions might provide large plasticity to interact efficiently with several different targets, as compared with a globular protein with limited conformational ûexibility (Wright & Dyson, 1999; Dunker *et al.*, 2005). It allows them to bind several target molecules and involve in different pathological and biochemical processes (Wright & Dyson, 1999; Uversky *et al.*, 2000; Dunker *et al.*, 2001; Dunker *et al.*, 2002; Uversky, 2003). As such unfolded proteins and disordered regions are involved in DNA binding and other types of molecular recognition, cell cycle regulation,

Corresponding Author: **Nakul C. Maiti**
E-mail: ncmaiti@iicb.res.in

Received: November 25, 2012
Accepted: November 28, 2012
Published: November 30, 2012

membrane transport and other important cellular functions (Dunker *et al.*, 2002; Dyson & Wright, 2002; Tompa, 2002; Uversky & Fink, 2004; Dyson & Wright, 2005).

Some regions of the protein are much prone to interact with target molecules and act as binding region or functional region of the molecule. The binding of unfolded proteins to a target molecule often cause folding and structural transformation, particularly when it binds to structured partner/proteins. Structural alteration of the binding region may render the protein function more effectively and with certain specificity. Further conformational flexibility and dynamic nature of the molecule allow NUPs to target several binding partner and become important in cell signalling and regulatory process.

The efficiency and energetic of interaction of binding region to a partner is largely influenced by amino acid composition and structural preferences present in binding regions of NUPs. The amino acid compositions in NUPs differ from globular proteins and often consist of charged and polar residues with low sequence complexity and high content of highly fluctuating amino acids. In this investigation several computational algorithms were applied to 100% disordered proteins to detect and define structure and many other sequence aspects of the binding regions present in fully unfolded native human protein (FUNHPs).

Materials and Methods

Natively unfolded proteins (NUP) - Fully unfolded (100% disorder) native human proteins were selected from DisProt database (Sickmeier *et al.*, 2007) (release 5.4, 2010-10-14). We found 25 fully unfolded (100% disorder) native human proteins (FUNHP). DisProt provides the proteins with a disport ID number and disorder regions. The database provides the name of the protein, accession codes, amino acid sequence, location of the disordered region(s), and methods used for structural (disorder) characterization. It also provides the information related to biological function (s) of each disordered region and its structural implication. Sequences of each protein were retrieved in FASTA format. It is available at *http://www.disprot.org/.*

Calculation of Binding region - ANCHOR method was engaged to detect the binding region of the proteins. The web based methods analyze the imputed sequences of unfolded protein and predict the binding region based on certain scoring values (Dosztanyi *et al.*, 2009). The server is available at *http://anchor.enzim.hu.* The sequences of the protein as obtained in the FASTA format in DisProt data base analysis was further analyzed by ANCHOR and binding regions were obtained based on some probability score.

Amino acid distribution - The amino acid distribution in binding region is predicted by using the server ProtParam which is available at *http://web.expasy.org/protparam*. Length, sequence (amino acid composition), residue charge isoelectric point (pI), GRAVY and molecular weight were obtaind using the ProtParam. The total charge of the proteins was calculated using 'protein calculator' server *www.-scripps.-edu/-~cdputnam/-protcalc.-html*) at pH 7.0.

Determination of Hydrophobilcity - Overall hydrophobicity was calculated using ProtParam tool of ExPASy proteomic server in the form of GRAVY (grand average of hydropathy) (Kyte & Doolittle, 1982). GRAVY of individual binding regions was also calculated separately using the same protocol.

Prediction of secondary structure - Computational algorithm, PSIPRED (Jones, 1999; McGuffin *et al.*, 2000) was used to define the structural propensities of all the protein residues. The algorithm uses a sequence of amino acids as a query sequence input and predicts the secondary structure (helix, β-strand/sheet and coil) propensity with certain confidence for each residues (Jones, 1999). Percentage of residues in a protein with preference for a particular conformation was measured by taking a ratio of the total number of residues with a particular conformation to the protein sequence length. Structural preferences of binding regions were calculated as a part of a whole protein analysis as mentioned above and the sequence length of individual binding region was considered as the total to determine the percentage of residues with defined conformation in that region.

Results and Discussion

Short segments of unfolded proteins that showed propensity to interact with some target molecules with a possibility of structural reorganization are key in detection of binding region (BR) by ANCHOR method. This report analyzed binding regions (BR) of a special class of human proteins which are fully unfolded (100%) in native condition. These fully unfolded native human proteins (FUNHPs) were selected from DisProt database and analyzed by ANCHOR. The DisProt database analysis revealed 25 such proteins and Table 1 lists them with their unique disport ID. Many of these proteins contained large number of charged and polar residues and showed common characteristic of natively unfolded proteins (NUPs). Intrinsic features such as theoretical isoelectric point (pI), number of charged residues and length of individual proteins are also shown in the table. The shortest protein, Antibacterial protein LL-37 was 37 residues long and the largest protein was microtubule associate tau with 441 amino acid residues.

ANCHOR analysis detected 90 BRs in the set of 25 FUNHPs as shown in Table 2. Most of the proteins contained multiple binding regions and two proteins found without any BR. 45% of the total residues in the set of FUNHPs found to be in BR. BR occupied 40% sequences in DP00016, 82% in DP00039 and 11% in the DP00069. However, the number of sequences in BR increased with the sequence length and showed a linear correlation (Figure 1).

Sequence length of the BRs varied and most of the region consisted of less than 40 aa residues. Figure 2 shows the no of binding regions in each protein with different sequence length. Nearly 18% binding regions were less than 10 residues long, 42% BRs were 11-20 residues long. Very few BRs were detected with sequence length greater than 40. The shortest BR was six residues long and found in DP00016, DP00174 and DP00219 and the longest BR with 118 residues were detected in DP00126. Only one binding region was detected with sequences greater than 100. However, longer binding regions were also found with lower frequency. The analysis found the smallest binding region with six residues and the largest binding region 118 residues.

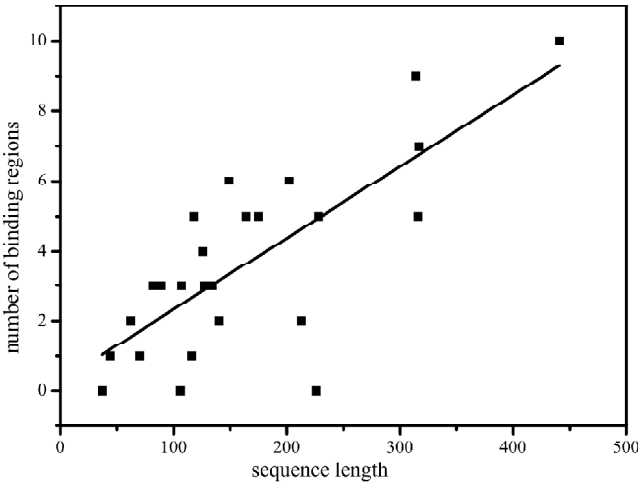

Figure 1: **Correlation of binding regions and their sequence length. Plot shows the number of binding regions against sequence length of the protein. Black squares (■) represent the total number of binding regions in each individual protein. Straight line in the plot shows a tentative linear correlation with no of binding region and sequence length of the protein.**

BRs were distributed along the protein sequence. To examine the positions distribution of BRs located in the amino- (N-), carboxyl- (C-) the total sequence length of each protein was divided into three parts: (i) N-terminal region consisting of first fifteen residues (ii) C-terminal region consisting of the last 15 residues and (iii) the middle region covering the remaining residues in between N terminal and C terminal. Figure 3 shows in a pi diagram that BRs distributed along all the sequence length. ~13% and ~10% BR were found in N- and C-terminal region, respectively. However 77% BRs were in the middle (between C- and N-terminal) part of the protein sequence.

The amino acid composition of binding regions is shown in Figure 4 and compared with the total protein sequence of individual protein. The results indicated that content (%) of different aa in the binding region follow a similar trend, except a few exception, with the aa distribution of individual protein residues. Though there was an almost similar trend in amino acid distribution between BR sequences and total protein sequence, the relative individual amino acid distribution at the binding regions showed some deviations (Panel B, Figure 4). In binding regions, the most abundant residues were Ala, Ser, Gly and Lys. Ala, Gly, Leu, Asp and Tyr were among the

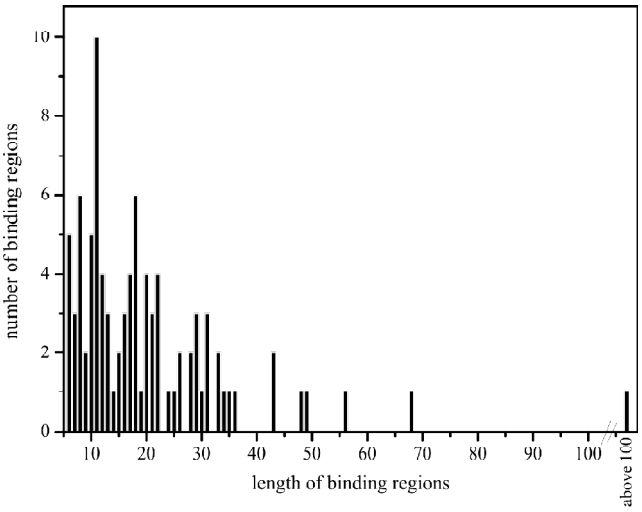

Figure 2: **Number distribution of binding region according to their sequence length. X-axis represents sequence length and Y-axis indicates their population in the total group of proteins (FUNHP).**

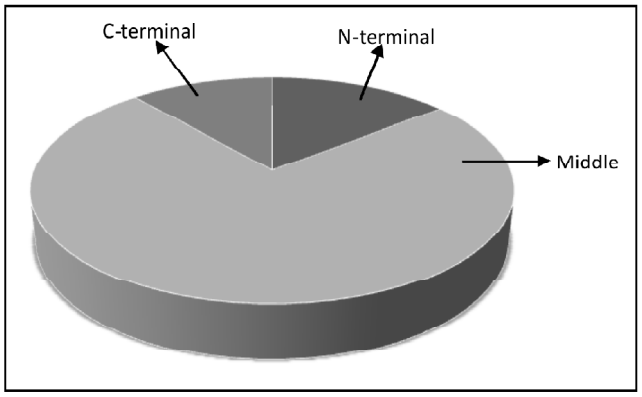

Figure 3: **Position of binding regions along the protein sequence length. This pi-diagram indicates the percentage of BRs in N- terminal, C- terminal and in the middle range of the protein sequence. Regions are marked according to their positions.**

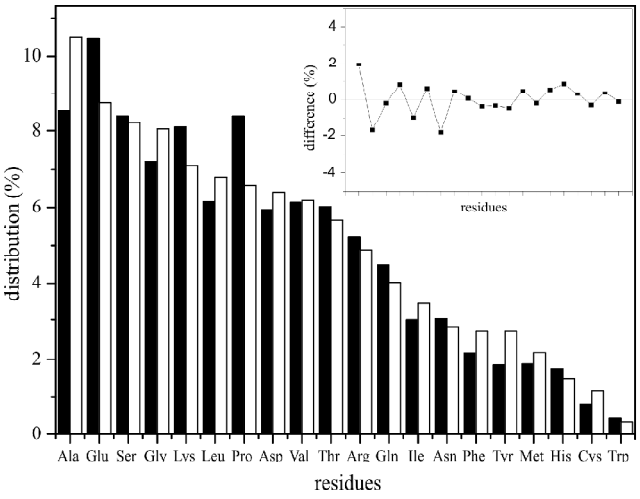

Figure 4: **Average content of different amino acids in the binding region (□) and in total protein (■). Y-axis indicates the % of residues and x-axis indicates different amino acids. Selection of amino acids along x-axis was based on the content of amino acids in the binding regions and arranged from most abundant to least. Top panel in the figures shows the difference in content (%) of amino acids in binding region when compared to the content of total protein sequence (percentage of a particular aa in binding region – percentage of the same aa in total protein sequence). The selection of amino acids along x-axis was same as in the main figure and names of amino acids were omitted for clarity.**

residues that showed some increase compared to the average number of the same residues found in the binding regions. A decrease in proportion at the BR compared to its overall average occurred for Pro, Lys. Other residues that showed less in numbers were Val, Asn and Trp. A slight dominance of Ala, Gly, Leu, Tyr and Asp in BRs may have an impact on binding to their target molecule. Presence of Gly has a tendency to be in turn region and provide some structural specificity. Tyr is often involved in hydrogen bond formation with target molecules.

Using ProtParam algorithm we derived grand average hydropathy (GRAVY) indexes of individual BRs and total proteins (Table1). The calculated GRAVY indexes of all the proteins was negative and varied between –0.21 (DP00017) and –1.87 (DP00040) it was expected as the proteins were rich in polar and charged residues. However GRAVY values for many BRs were positive (Table 2). Most of the BRs detected in DP00016, DP00017 and DP00069 were positive. However, other proteins contained BRs with both –ve and +ve charge. The differences could be due to some percentage of change in sequence composition of the BR regions.

The conformations (helix, β-strand/sheet and coils) that the residues in total protein and total BRs preferred to adopt are shown in Figure 5. Overall structural content of BR sequences was: helix ~18 %, β-sheet/strand ~3% and random/ extended conformation ~79% indicating that binding region was dominated with sequences that preferred to be flexible. In total protein however structural preference of the sequences was helix ~23% β-sheet/strand ~4% and random/extended conformation ~73%. BR

sequence in proteins such as DP00028, DP00219 and DP00287 preferred more than 10% β-sheet/ strand conformation. BRs in DP00069, DP00357 and DP00630 contained more than 60% residues with preference for helical conformation. However, BR residues in DP00017, DP00039, DP00040 and DP00126 showed propensity for random/extended conformation. There are only two proteins DP00069 and DP00630 that contained BRs of which less than 30% coil conformation. The analysis proved that very few residues preferred β-sheet/strand conformation and both the BRs and the parent protein molecules are rich in sequences most of which preferred random/extended conformation and indicated that a sequences in BRs did not favor β-sheet/strand backbone conformation.

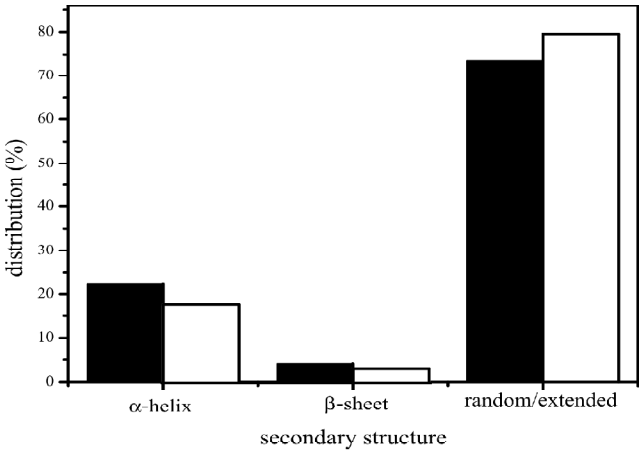

***Figure 5*: Structural preferences of residues in BRs and in total protein. Percentage of total residues with preference to helix, strand and random/disorder conformation are shown for FUNHPs (■) and binding regions (□). X-axis represent different structural component and Y-axis represents average percentage of residues that prefer to adopt particular conformation.**

Table 1
Fully unfolded native human proteins (FUNHPs)

serial no	*DisPort ID*	*name*	*sequence length*	*pI*	*charge*	*GRAVY*
1	DP00004_C002	Antibacterial protein LL-37	37	10.61	5.9	-0.724
2	DP00016	Cyclin-dependent kinase (I)	164	8.69	3.3	-0.845
3	DP00017	Cyclin-dependent kinase (IC)	316	5.39	-3.6	-0.21
4	DP00028	factor 4E-binding protein 1	118	5.32	-1.9	-0.701
5	DP00039	protein HMG-17	89	10.00	11.9	-1.761
6	DP00040	HMG-I/HMG-Y	107	10.31	11.9	-1.807
7	DP00069	membrane protein 2	116	7.84	11.9	-0.017
8	DP00070	Alpha-synuclein	140	4.67	-8.8	-0.403
9	DP00126	protein tau [Isoform Tau-F]	441	8.24	4.8	-0.868
10	DP00174	Stathmin	149	5.75	-3.1	-1.311
11	DP00199	Beta-casein	226	5.52	-3.9	-0.094
12	DP00214	Osteopontin	314	4.37	-42.2	-1.1
13	DP00219	Protein phosphatase 1	126	6.52	0	-1.205
14	DP00287	tumor suppressor [Isoform 1]	213	4.68	-16.9	-0.777
15	DP00332	Bone sialoprotein 2	317	4.12	-51.8	-1.35
16	DP00357	Thymosin beta-4	44	5.02	-2.1	-1.62
17	DP00372	Uncharacterized protein	106	10.14	11.9	-0.717
18	DP00510	Nuclear protein 1	82	9.98	5.4	-1.33
19	DP00521	Securin	202	6.18	-0.6	-0.409
20	DP00546	protein K [Isoform 1]	175	5.35	-4.6	-0.811
21	DP00555	Beta-synuclein	134	4.41	-14.8	-0.537
22	DP00592	Purkinje cell protein 4	62	6.21	-0.1	-1.332
23	DP00617	subunit DSS1	70	3.81	-21.5	-1.36
24	DP00630	Gamma-synuclein	127	4.86	-6.0	-0.53
25	DP00694	RING1 and YY1-binding protein	228	9.66	14.1	-1.16

Table 2
Binding regions (BR) of FUNHPs

Proteins [a] *DP00004_C002*	*startposition* *no*	*end position* *no*	*length* *no*	*m.w.* *no*	*pI* *no*	*net charge* *no*	*GRAVY* *no*
DP00016							
1	36	41	6	564.7	5.56	-0.1	1.9
2	68	79	12	1325.6	9.99	1.9	0.475
3	100	105	6	601.7	5.57	-0.1	0.8
4	109	123	15	1641.8	5.21	-0.9	0.193
5	145	164	20	2498.9	11.07	5.1	-1.14
DP00017							
1	22	33	12	1301.5	10.39	1.9	0.617
2	96	112	17	1819.2	11.7	2.9	0.665
3	167	182	16	1413.7	5.57	-0.1	1.675
4	215	282	68	6827.5	9.52	2.2	-0.618
5	291	316	26	2658	10.76	2.9	-0.523
DP00028							
1	13	31	19	2015.3	11.7	1.9	0.079
2	38	66	29	3215.6	9.37	1.9	-0.462
3	75	82	8	796.9	5.52	-0.1	0.938
4	92	99	8	926.9	6.85	0.2	-1.337
5	108	118	11	1241.2	3.5	-4.1	-0.9
DP00039							
1	1	33	33	3587.1	10.37	5.9	-1.706
2	42	63	22	2221.6	10.3	6.9	-1.614
3	72	89	18	1760.8	4.68	-1.1	-1.422
DP00040							
1	1	18	18	1939.1	8.25	0.9	-1.411
2	28	48	21	2229.5	11	2.9	-1.162
3	62	75	14	1474.7	11.39	5.9	1.921
DP00069							
1	1	13	13	1156.3	5.28	-0.1	0.438
DP00070							
1	87	96	10	964.1	8.47	0.9	1.11
2	111	140	30	3433.5	3.12	-11.1	-1.25
DP00126							
1	1	56	56	6157.5	4.14	-9.6	-1.355
2	59	176	118	11740.7	5.13	-3.6	-0.904
3	191	202	12	1156.1	5.55	-0.1	-1.342
4	208	215	8	903	12	1.9	-1.238
5	221	263	43	4612.4	11.2	7.9	-0.765
6	272	291	20	2172.5	9.63	2.9	0.4
7	305	326	22	2308.7	9.1	1.9	0.232
8	346	358	13	1492.7	8.59	0.9	-0.731
9	390	398	9	1005.1	6.05	-0.1	0.4
10	424	441	18	1815	4.37	-1.1	0.333
DP00174							
1	1	6	6	622.6	3.8	-1.1	0.517
2	16	26	11	1161.3	4	-1.1	0.555
3	47	52	6	758.8	4.53	-1.1	-1.017
4	67	73	7	799.9	6.1	-0.1	0.386
5	82	105	24	2833.2	6.74	0.2	-0.908
6	113	130	18	2166.5	8.6	1.2	-1.222
DP00199	no	no	no	no	no	no	no

contd. table 2

DP00214							
1	33	45	13	1554.7	5.83	-0.1	-0.662
2	52	63	12	1241.3	5.52	-0.1	-0.15
3	66	76	11	1279.3	4.37	-1.1	-1.373
4	84	111	28	3197	3.39	-14.6	-1.871
5	128	170	43	4689.1	3.93	-6.1	-0.302
6	197	213	17	1787	4.37	-1.1	-0.071
7	242	248	7	950.1	11	2.9	-2.1
8	260	287	28	3322.6	5.27	-3.1	-0.475
9	298	308	11	1344.5	6.75	0.2	-0.482
DP00219							
1	1	10	10	935	3.79	-2.1	0.25
2	27	32	6	673.8	8.47	0.9	1.117
3	56	71	16	1848.1	9.11	2.8	-0.537
4	88	104	17	1961.2	12.18	6.4	-1.359
DP00287							
1	5	26	22	2380.3	3.28	-9.1	-1.109
2	72	91	20	2374.8	10.35	1.9	0.59
DP00332							
1	6	16	11	1104.4	5.52	-0.1	2.473
2	39	59	21	2879.3	9.99	5.4	-1.205
3	101	143	43	4311.7	4.51	-2.1	-0.407
4	164	184	21	2211.1	3.4	-6.1	-1.419
5	186	214	29	2767.6	3.31	-8.1	-1.2
6	233	281	49	5551.8	3.84	-8.1	-1.231
7	289	317	29	3631.7	4.8	-2.6	-2
DP00357							
1	1	18	18	2112.4	4.86	-1.1	-1.089
DP00372	no	no	no	no	no	no	no
DP00510							
1	1	10	10	993.1	5.28	-0.1	0.47
2	23	53	31	3280.5	8.18	1.2	-0.906
3	61	76	16	1923.2	10.28	3.2	-1.75
DP00521							
1	1	8	8	925.1	3.8	-1.1	1.337
2	19	35	17	1653.9	9.7	1.9	0.241
3	45	52	8	941	8.75	0.9	-0.7
4	115	125	11	1400.6	4.37	-1.1	-0.236
5	142	148	7	715.9	5.24	-0.1	1.657
6	192	202	11	1216.4	3.49	-3.1	0.755
DP00546							
1	1	22	22	2334.7	11.52	2.9	-0.395
2	32	67	36	3933.3	4.48	-4.1	-0.414
3	84	114	31	3314.5	4.00	-5.1	-0.403
4	133	150	18	2163.6	4.59	-2.1	-0.083
5	165	175	11	1156.3	4.00	-1.1	1.173
DP00555							
1	1	9	9	1027.2	5.59	-0.1	0.667
2	65	89	25	2472.7	6.76	0.2	0.38
3	100	134	35	4096.2	3.14	-14.1	-1.566
DP00592							
1	1	8	8	834.9	5.75	-0.1	-1.175
2	26	58	33	3855.3	5.1	-1.1	-0.894
DP00617							
1	7	40	34	3962	3.3	-15.8	-0.894

contd. table 2

DP00630							
1	1	10	10	1171.4	8.35	0.9	0.37
2	68	77	10	1004.1	4	-1.1	0.93
3	87	97	11	1143.3	11	1.9	0.464
DP00694							
1	22	47	26	2974.3	4.78	-1.2	0.004
2	57	71	15	1660.8	5.52	-0.1	-0.067
3	114	133	20	2085.2	4.56	-1.1	-0.85
4	146	176	31	3441.9	9.72	1.9	-0.219
5	218	228	11	1171.2	3.49	-3.1	-0.482

a: different numbers (1, 2, 3.....) under the proteins' ID are the serial no of binding regions of the corresponding protein

Conclusion

Our investigation computationally analyzed the composition of binding region in a particular class of human proteins which remains fully unfolded under native condition. Many of these proteins possessed multiple binding regions in their sequence. Sequence length of the BRs varied from region to region and BRs detected with sequence length greater than 40 was very uncommon. The sequence content in binding region increased with the protein length. The regions contained residues with preference for random-coil conformation. It would be interesting to test the binding and functional efficacy of the regions with some of the target molecules. We are in the process of designing some of the binding regions and to test them with suitable target proteins.

Acknowledgement

We thank CSIR-IICB for providing all kinds of support to implement this project. We also thank those who provided online computational methods, which were used in this work, made available freely.

Abbreviations

Aa, amino acid; NUP, natively unfolded protein; FUNHP, fully unfolded native human protein; GRAVY, grand average of hydropathy; PSIPRED, protein structure prediction server; pI, isoelectric point; PASTA, prediction of amyloid structure aggregation; DisProt, database of protein disorder; APP, amyloid precursor protein; BR, binding region.

References

Ahmad, A., Uversky, V. N., Hong, D.P., and Fink, A.L. (2005). Early events in the fibrillation of monomeric insulin. J. Biol. Chem., 280, 42669-42675.

Dosztanyi, Z., Meszaros, B., and Simon, I. (2009). ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics*, 25, 2745-2746.

Dosztányi, Z., Mészáros, B., and Simon, I. (2010). Bioinformatical approaches to characterize intrinsically disordered/unstructured proteins. *Brief Bioin*, 11, 225-243.

Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M., and Obradovic, Z. (2002). Intrinsic Disorder and Protein Function. *Biochemistry*, 41, 6573-6582.

Dunker, A.K., Cortese, M.S., Romero, P., Iakoucheva, L.M., and Uversky, V.N. (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J*, 272, 5129-5148.

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C., and Obradovic, Z. (2001). Intrinsically disordered protein. *J Mol Graph Mod*, 19, 26-59.

Dyson, H., and Wright, P. (2005). Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*, 6, 197-208.

Dyson, H.J., and Wright, P.E. (2002). Coupling of folding and binding for unstructured proteins. *Curr Opin Struc Biol*, 12, 54-60.

Iwasa, H., Meshitsuka, S., Hongo, K., Mizobata, T., and Kawata, Y. (2011). Covalent Structural Changes in Unfolded GroES That Lead to Amyloid Fibril Formation Detected by NMR. Insight Into Intrinsically Disordered Proteins. *J. Biol. Chem*., 286, 21796-21805.

Jones, D.T. (1999). Protein Secondary Structure Prediction Based on Position-specific Scoring Matrices. *J Mol Biol*, 292, 195-202.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J Mol Biol, 157, 105-132.

McGuffin, L., Bryson, K., and Jones, D. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, 16, 404-405.

Orosz, F., and Ovadi, J. (2011). Proteins without 3D structure: definition, detection and beyond. *Bioinformatics*, 27, 1449-1454.

Schweers, O., SchÃ¶nbrunn-Hanebeck, E., Marx, A., and Mandelkow, E. (1994). Structural studies of tau protein and Alzheimer paired helical filaments show no

evidence for beta-structure. *J. Biol. Chem*., 269, 24290-24297.

Sickmeier, M., Hamilton, J.A., LeGall, T., Vacic, V., Cortese, M.S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V.N., Obradovic, Z., and Dunker, A.K. (2007). DisProt: the Database of Disordered Proteins. *Nuc Acids Res*, 35, D786-D793.

Tompa, P. (2002). Intrinsically unstructured proteins. *Tr Biochem Sci*, 27, 527-533.

Uversky, V.N. (2002). What does it mean to be natively unfolded? *Eur J Biochem*, 269, 2-12.

Uversky, V.N. (2003). Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: which way to go? *Cell Mol Life Sci : CMLS*, 60, 1852-1871.

Uversky, V.N., and Fink, A.L. (2004). Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochim Biophys Acta*, 1698, 131-153.

Uversky, V.N., Gillespie, J.R., and Fink, A.L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins Struct Func Gen*, 41, 415-427.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol*, 337, 635-645.

Weinreb, P.H., Zhen, W., Poon, A.W., Conway, K.A., and Lansbury, P.T. (1996). NACP, A Protein Implicated in Alzheimer's Disease and Learning, Is Natively Unfolded. *Biochemistry*, 35, 13709-13715.

Wright, P.E., and Dyson, H.J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol*, 293, 321-331.