Trakia Journal of Sciences, Vol. 12, Suppl. 1, pp 110-114, 2014

Available online at:
http://www.uni-sz.bg

ISSN 1313-7050 (print)
ISSN 1313-3551 (online)

INTRINSIC PROTEIN DISORDER AND DISORDERED PROTEIN BINDING REGIONS IN EWING'S SARCOMA

R. Todorova*

ABSTRACT

Intrinsic Protein Disorder (IPD) of native Ewing's sarcoma protein (EWS) and oncogenic EWS fusion proteins (EFPs) was estimated by Predictors, and relationship was found between Amino acid (AA) composition, IPD and functional regions. EFPs show similar IPD in N-terminal domain (NTD) (AA 1-264, EAD) and different IPD in C-terminal domain (CTD). The IPD of different isoforms of EFPs are shifted with some AAs depending from the translocation point. The EFPs oncogenic function may be related to decreased probability for IPD and Protein Binding Regions (PBRs) in CTD, due to the fused Transcription Factor (TF). IPD and PBRs probabilities are high around the breakpoint, and relatively low in EFP CTDs. IPD is high around Y in the EAD repetitive segments. The EAD Y-free box, IQ domain (Y-free region flanked by Y-boxes) as well as RGG1, RGG2, RGG3 are disordered. EWS shows high IPD and high probability for protein–protein interactions (64% predicted DBRs) in both EAD and CTD. In EFPs the EAD has high DBRs, a peak around and after the breakpoint, and relatively low DBRs in CTD, corresponding to functional interactions. The reported results could be used in the design of antitumor agents against the corresponding malignances.

Key words: Intrinsically disordered protein; Ewing's sarcoma protein; EWS fusion proteins (EFPs); Function-structure-disorder; Predictors; Targeted inhibition; Protein binding regions

INTRODUCTION

The discovery of Intrinsically disordered proteins (IDPs) challenges the conventional protein "sequence-structure-function" paradigm (1). IDPs lack stable structure and have important biological functions in signaling, regulation and control. Proteins, associated with cancer, are enriched in intrinsic disorder (ID): they enter in high-specificity-low-affinity interactions and bind to multiple structurally diverse partners, thus participating in multiprotein complexes like transcription (2). The characterization and role of disordered proteins in human diseases represents a novel intersection of biochemistry and pathology.

Ewing's Sarcoma Oncogene (*EWS*) on chromosome 22q12 is encoding a RNA binding protein that is the target of tumor-specific _ chromosomal translocations in childhood sarcomas with poor prognosis: Ewing sarcoma family of tumours (ESFTs), Soft tissue clear sarcoma, Myxoid liposarcoma, Malignant melanoma of soft parts, Desmoplastic small round cell tumor, Angiomatoid fibrous histocytoma and others. The incidence of *EWS-Fli1* translocation t(11;22)(q24;q12) in ESFTs is of 85%. Breakpoints in *EWSR1* and *FLI1* genes occur in many introns, and following splicing, the exons join together to generate various subtypes of *EWS/FLI*. The malignancy of these tumours, affecting predominantly children and adolescents imposes to search for new targets and therapies to treat these sarcomas. Such promising targets are the native Ewing's sarcoma protein (EWS), involved in transcription, splicing and regulation in normal tissues, and the oncogenic EWS fusion proteins (EFPs), involved in initiation and maintaining of these sarcomas with high potency to metastasis.
The available experimental data about the structure and function of native EWS and its fusions with different transcription factors (TFs) or EFPs are compared to the intrinsic disorder predictions of these proteins with known primary structure. There is found a link between the amonoacid (AA) sequence or primary protein structure, IPD, Disordered binding regions (DBRs) and function of these proteins.

***Correspondence to**: *Roumiana Todorova, IBPhBME-BAS, 1113 Sofia, Bulgaria, E-mail: todorova@bio21.bas.bg*

MATERIAL AND METHODS

Predictors IUPred (3), GlobPlot2 (4), DisEMBL (5), FoldIndex (6), RONN (7), PONDR (8) were used to estimate the IPD of native EWS and different isoforms of its oncogenic fusions EWS-FLI1, EWS-ATF1 and EWS-ZSG (NCBI). The Protein Binding Regions (PBRs) were predicted by ANCHOR (9).

RESULTS AND DISCUSSION

The functional regions of EWS native protein and its oncogenic EFPs (EWS-FLI, EWS-ATF1 and EWS-ZSG) are shown on **Figure 1.**

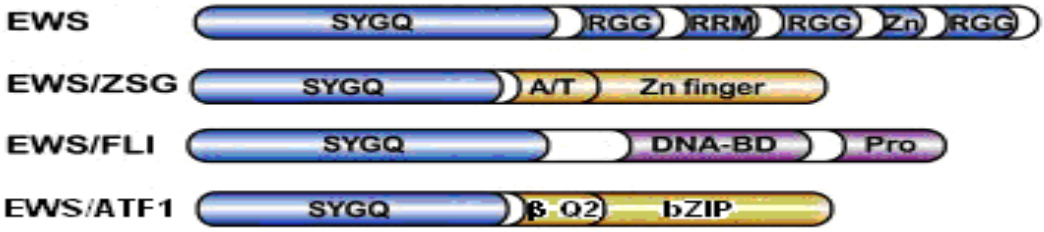

Figure 1. Schematic representation of domain structures of protein EWS and EFPs. Native EWS: EAD (SYGQ rich transactivation domain (TAD)); RGG rich domain; RRM (RNA-recognition motif); Zn (zinc finger). EWS-ZSG: A/T (A-T hook DNA binding motif); Zn finger (Cys2-His2 zinc finger). EWS-FLI: DNA-BD (DNA binding domain); Pro (Pro rich activation domain). EWS-ATF1: β (critical motif DLSSD); Q2 (Glu rich constitutive activation domain); bZIP (dimerization and DNA-BD).

IPD predictions of native EWS and its oncogenic fusion proteins EFPs

Predictors were used to estimate the protein disorder of EWS and EFPs. A difference in the IPD of C-terminal domain (CTD) was detected between EWS and EFPs, due to different C-terminal content. A relationship between protein structure, disorder and function was found in some regions of EFPs.

EWS native protein. By IUPred the EAD (AAs 1-264) (N-terminal domain (NTD, TAD)) consists of large disordered regions, while the CTD is almost completely disordered. A common disordered region (free of Y motifs) in EAD (AAs 132-156) was found by all Predictors, involved in interactions with different factors. The IQ domain, binding calmodulin (AA 258-280), and RGG1, RGG2 and RGG3 are disordered by all Predictors. The EAD is disordered with highest ID probability for regions of AAs 80-95, 130-160, 195-205, and 240-250. The AAs 95-120 are folded. The Zing finger domain (AAs 518-549) is partially ordered by RONN and flanked by disordered AAs regions.

EWS-FLI1 protein. By Predictors, EWS-FLI1 approaches a largely unfolded conformation. The EAD shows similar intrinsic disorder as native EWS, while the CTD, originated from Fli1 is almost ordered with low propensity for disorder. The globular domains in all isoforms are close to C-terminus consisting of about 60 AAs. The DNA-BD, composed of helix-turn-helix is relatively ordered, whereas the far-carboxy region is intrinsically disordered.

EWS-ATF1 protein. By IUPred, the fusions of EWS and ATF1 are intrinsically disordered with similar disorder in CTD, increased in regions flanking bZIP domain, while the rest of CTD is almost ordered. The bZIP domain is folded and linked by highly conserved sequences, mobile and unstructured. By all Predictors the critical elements and breakpoints are connected by long segments of structural disorder.

EWS-ZSG protein. The EWS/ZSG chimeric fusion is a non-TET/ETS fusion, containing a Zn finger at his C-terminus, originated from the *ZSG* gene. By IUPred, the IPD of EWS-ZSG shown a long disordered region originated from EAD, followed by a globular domain in the CTD comprising the A-T hook DNA binding motif and the Zn finger, and a short disordered region at C-terminal end.

Protein Binding Regions (PBRs) by ANCHOR for EWS and EFPs

The predicted DBRs of native EWS are about 64% from the complete protein sequence, comprising functional domains, such as Tyr

reach regions, IQ domain, RRM and Zing finger domain. These regions are sites of post-translational modifications and functional interactions with partners **(Table 1).**

Table 1. *Predicted DBRs by ANCHOR for EWS and EFPs, corresponding to functional domains.*

Predicted Disordered Binding Regions	**Functional domains**	**Post-translational modifications or interactions**
EWS native isoform 2 (full protein - EAD and CTD)		
AA 89-98	Y reach region	Thr95-*O*-GlcNAcylation
AA 110-131	Y reach region	Ser111, Thr120-*O*-GlcNAcylation
AA 156-185	Y reach region	Ser162, Ser168-*O*-GlcNAcylation
AA 200-234	Pfam-B	AAs (228–264)–ZFM1
AA 242-293	IQ domain	AAs (258–280) – ZFM1, calmodulin Ser266 – PKC phosphorylation
AA 360-388	RRM region	RNA-recognition motif
AA 396-403	RRM region	RNA-recognition motif
AA 423-460	RRM region	RNA-recognition motif
AA 512-559	zf-RanBP	Zing finger domain
EWS/FLI1 type1 (far EAD, breakpoint and CTD)		
AA 206-240	Pfam B	
AA 331-351	ETS	DNA-binding
AA 410-415	ETS	
EWS/ATF1 type1 (far EAD, breakpoint and CTD)		
AA 206-240	Pfam B	
AA 325-340	pKID	
AA 502-512	bZIP	Dimerization and DNA-binding
EWS-ZSG long B isoform (far EAD, breakpoint and CTD)		
AA 200-234	Pfam B	
AA 344-375	Zf-C2H2	Zn finger

The IPD and PBRs for native EWS were predicted by ANCHOR (9) and IUPred **(Figure 2).** The EWS PBRs are distributed almost regularly in all protein sequence. There is not high PBRs in the first 82 AAs of EAD, but they are important for trans-activation and binding of RNA Polymerase II. Small IPD is around AAs 82 and 100, and a small PBR around AAs 50-60, consistent with experimentally found physical interaction between AAs 1-57 from EAD and hsRPB7 (10). The rest of EAD and the full-length CTD have high probability to form PBRs, except of short segments of the sequence.

ANCHOR shows a high probability of PBRs formation, or protein-protein interactions with other partners, in the EAD of EFPs **(Figure 2).** A peak of PBRs formation is detected around the breakpoint and especialy after it in all fusions (11). Thus the oncogenic function of EFPs may be related to diminished PBRs probability, due to fused TFs, leading to breaking of EWS functional interactions.

EAD function and targeted inhibition against IPD regions

EWS as disordered protein binds to structured partners, thus undergoing disorder-to-order transition. Predictors are used to study native EWS and its oncogenic fusions, and relationship structure-function-protein disorder was found in some regions of these proteins (12). EAD as a flexible string-like structure with several Tyr residues is making contact with many different proteins (13). Tyr residues are important as flanking elements in EAD regions with high protein disorder, involved in functional protein-protein interactions, including trans-activation by EAD.

Targeted inhibition of EWS-FLI1 is challenged because of lack of intrinsic enzymatic function. No direct, targeted therapeutics against EWS-FLI1 are used in clinics. The small molecule YK-4-279 ($C_{17}H_{13}C_{l2}NO_4$) dissociates the recombinant RNA helicase A (RHA) from EWS-FLI1. The location of the YK-4-279 binding site and whether it is fully disordered is not known.

YK-4-279 induces apoptosis of ESFTs cell lines and reduces the growth of ESFTs orthotopic xenografts in mice (14), thus targeting directly EWS-FLI1 against ESFTs.

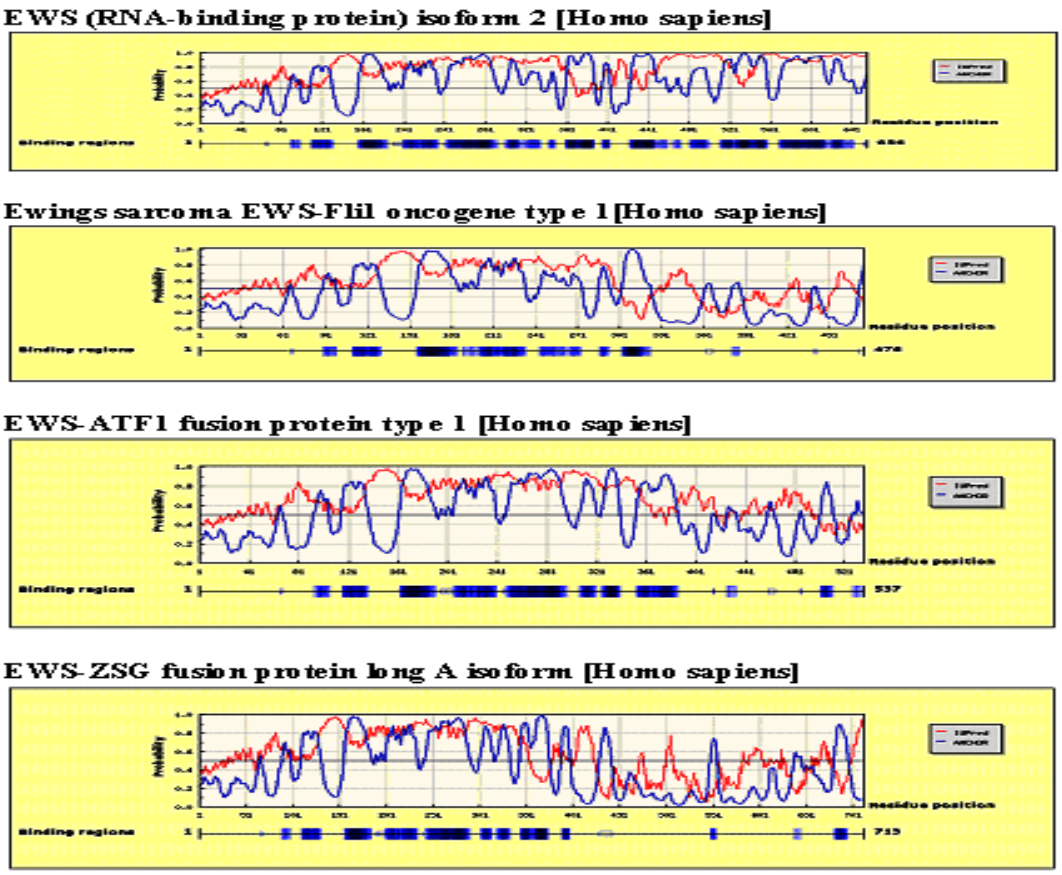

Figure 2. Predictions of IPD and PBRs by IUPred and ANCHOR for EWS and oncogenic EFPs.

CONCLUSIONS

The IPD and PBRs of EWS and EFPs are consistent with the experimental findings for their functional sequences. The results for IPD

obtained by different Predictors are similar. It seems that the oncogenic function is related to IPD decrease in CTD due to the fused TF. The vicinity of breakpoint in EFPs is significantly more disordered. The relationship function-structure-disorder could help the design of potential antitumor agents against ESFTs. Targeted inhibition against IPD in Ewing's sarcoma should be developed because of high mortality of patients with metastatic disease. The disruption of EFPs-involved functional protein-protein interactions is a promising target for small molecule therapeutics.

REFERENCES

1. Huang, Y.-Q. and Liu, Z.-R., Intrinsically Disordered Proteins: the New[J]. *Acta Phys Chim Sin*, 26 (08): 2061-2072, 2010.
2. Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C.J. and Dunker, A.K., Predicting intrinsic disorder from amino acid sequence. *Proteins*, 53, Suppl 6: 566-572, 2003.
3. Dosztanyi, Z., Csizmok, V., Tompa, P. and Simon, I., IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, 21(16): 3433-3434, 2005.
4. Linding R., Russell, R.B., Neduva, V. and Gibson, T.J., GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res*, 31: 3701-3708, 2003.
5. Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J. and Russell, R.B., Protein disorder prediction: implications for structural proteomics. *Structure*, 11(11): 1453-1459, 2003.
6. Prilusky, J., Felder, C.E., Zeev-Ben-Mordehai T., Rydberg, E.H., Man, O., Beckmann, J.S., Silman, I. and Sussman,

J.L., FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*, 21(16): 3435-3438, 2005.

7. Yang, Z.R., Thomson, R., McNeil, P. and Esnouf, R.M., RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, 21: 3369-3376, 2005.
8. Xue, B., Dunbrack, R.L., Williams, R.W., Dunker, A.K. and Uversky, V.N., PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta*, 1804(4): 996-1010, 2010.
9. Dosztányi Z., Mészáros, B. and Simon, I., ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics*, 25(20): 2745-2746, 2009.
10. Todorova, R., In vitro interaction between the N-terminus of the Ewing's sarcoma protein and the subunit of RNA polymerase II hsRPB7. *Mol Biol Rep*, 36(6): 1269-74, 2009.
11. Todorova, R., Disordered Binding Regions of Ewing's Sarcoma Fusion Proteins. *Russ J of Bioorg Chem*, 40 (1): 16–25, 2014.
12. Todorova, R., Disorder structural predictions of the native EWS and its oncogenic fusion proteins in rapport with the function. *Adv Biosc Biotech*, 3: 25-34, 2012.
13. Lee K.A., Molecular Recognition by the EWS Transcriptional Activation Domain. *Adv Exp Med Biol*, 725: 106-25, 2012.
14. Erkizan, H.V., Kong, Y., Merchant, M., Schlottmann, S., Barber-Rotenberg, J.S., Yuan, L., Abaan, O.D., Chou, T.-h, Dakshanamurthy, S., Brown, M.L., Uren, A. and Toretsky, J.A., A small molecule blocking oncogenic protein EWS-FLI1 interaction with RNA helicase A inhibits growth of Ewing's sarcoma. *Nature Med.* 15 (7): 750-757, 2009.