In silico prediction of the intrinsic disorder and ligand binding sites of human 3β-hydroxysteroid dehydrogenase type 2 enzyme

Goswami Achintya Mohan
Department of Physiology, Krishnagar Govt. College, Krishnagar, Nadia, West Bengal, INDIA
achintya02@gmail.com

Abstract

In human, the 3β-hydroxysteroid dehydrogenase (HSD3B)/$\Delta^{4,5}$-isomerase is an important enzyme in steroid hormone biosynthesis. It catalyzed the conversion of Δ^5-steroid precursors to respective Δ^4-ketosteroids. There were two HSD3B isoenzymes designated as type 1(HSD3B1) and type 2 (HSD3B2), encoded by two different genes located in chromosome 1p13.1. The mutations in HSD3B2 gene could lead to congenital adrenal hyperplasia and impaired steroidogenesis in both the adrenals and gonads leading to varieties of phenotypes. It was known that many proteins had intrinsically disordered regions (IDRs) and sometimes the entire protein lacked a stable tertiary structure, known as intrinsically disordered proteins (IDPs). These IDRs were attractive target for potential drug discovery.

The HSD3B2 was found to possess several IDRs and the regions of IDRs varied according to the algorithm or server used to predict them. The HSD3B2 was predicted to contain two ligand binding sites for NAD and ATP and a single cavity. Some amino acid residues involved in formation of cavity and ligand binding sites were within the intrinsically disordered regions.

Keywords: 3β-hydroxysteroid dehydrogenase, *in silico*, Intrinsically disordered regions, Ligand binding, Cavity.

Introduction

In human, the 3β-hydroxysteroid dehydrogenases (HSD3B)/$\Delta^{4,5}$-isomerase is an important enzyme in steroid metabolism and is involved in both dehydrogenation and $\Delta^{4,5}$-isomerization of the Δ^5-steroid precursors leading to formation of respective Δ^4-ketosteroids and thus catalyzes the conversion of pregnenolone to progesterone, 17α-hydroxypregnenolone to 17α-hydroxyprogesterone, dehydroepiandrosterone to androstenedione, androstenediol to testosterone and androstadienol to androstadienone.[9,15] The 3β-hydroxysteroid dehydrogenases enzyme requires NAD^+ which is then reduced to NADH during dehydrogenation and this NADH is then recruited for the next step of isomerization reaction by the same enzyme.[28] It is well known that these steroid hormones play important roles in physiological processes such as differentiation, development and growth in the most human tissues.[12] There are two HSD3B isoenzymes designated as Type 1 and type 2. These isoenzymes are 93.5% homologous and are encoded by two different genes located on chromosome 1p13.1.[20] The type 1 gene (called *HSD3B1*) is generally expressed in the placenta and peripheral tissues like prostate, mammary gland and skin. On the other hand, type 2 gene (called *HSD3B2*) is predominantly expressed in adrenal cortex, ovary and testicular tissues where HSD3B2 exists as a membrane-bound form.[24] HSD3B2 is the key enzyme required for the synthesis of cortisol and aldosterone in adrenal cortical tissues.[19]

It was observed that mutations in *HSD3B2* gene could lead to congenital adrenal hyperplasia (CAH; OMIM # +201810) which was an autosomal recessive disease that impaired steroidogenesis in both the adrenals and gonads.[17,21,25,26] The clinical manifestation of adrenal hyperplasia ranges from salt-losing to non salt-losing forms in both sexes. HSD3B2 deficiency in newborn males results in ambiguity of the external genitalia whereas in female newborns, partial virilization is observed. HSD3B2 deficiency during adolescence results in hypogonadism (variable degrees) in boys and hyperandrogenism (premature pubarche and hirsutism) in girls.[5, 23]

Previous *in silico* analysis showed that out of 16 nsSNPs, eight point mutations in the coding region may have significant effect in the HSD3B2 structure as well as in its function.[7]

It was known that many protein regions and even the entire protein lack a stable tertiary structure and show corresponding high degree of flexibility under physiological conditions.[29] These naturally flexible proteins and regions are known as intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) respectively. This conformational flexibility of IDPs and proteins with IDRs allowed them to involve in essential cellular processes. The regions of intrinsic disorder frequently possess sites of proteolytic attack and sites of various post-translational modifications.[32] The capability of IDPs and IDRs to interact with collections of partners is utilized in organizing protein-protein interaction networks. Furthermore, often IDPs and IDRs are involved in protein aggregation and misfolding related pathogenesis; making them attractive targets for potential drug discovery.[16]

There was no crystal structure available right now for human HSD3B2 and there was no information available about the intrinsically disordered regions of the protein and the ligand binding sites of the enzyme. Here in this study, extensive

computational analysis has been performed to predict its disordered regions, ligand binding sites and possible cavities.

Material and Methods

Sequence: HSD3B2 protein sequences were retrieved from the Uniprot database (https://www.uniprot.org/uniprotkb/P26439/entry) for computational analysis in this study. Both amino acid sequence and homology based model of HSD3B2, developed by Raptor X in our previous study[7] were utilized for the further analysis of the protein intrinsic disorder and ligand binding sites of the enzyme.

Prediction of disordered region: The IDRs of human HSD3B2 were predicted by GlobPlot 2.3, DisEMBL 1.5, FoldIndex and IUPRED server.[8] GlobPlot 2.3 was used to find out the order/globularity or disorder tendency of HSD3B2 based on a running sum of the propensity for an amino acid to be in ordered or disordered state by searching the domain databases and known disordered regions in proteins.[14] DisEMBL 1.5 (http://dis.embl.de) was utilized to predict three kinds of disordered structure including loops/coils, hot loops and those that are missing from the PDB X-ray structures as defined by REMARK-465 entries.[13]

FoldIndex, a dynamic and interactive process that estimates the local and general probability for the provided sequence, under specified conditions, to fold (http://bip.weizmann.ac.il/fldbin/findex)[18] was used to predict the intrinsic disordered regions of HSD3B2. The programme uses Kyte and Doolittle hydrophobicity algorithm during prediction. IUPRED server (http://iupred.enzim.hu/) was used to predict intrinsically unstructured/disordered regions of HSD3B2 from the amino acid sequence based on the estimated pairwise energy content.[3]

Prediction of ligand binding site or cavities of HSD3B2: Prediction of ligand binding sites of HSD3B2 was carried out in COACH server (http://zhanglab.ccmb.med.umich.edu/COACH/) for the present study. COACH is a meta-server to protein-ligand binding site prediction.[33] The homology based structure of HSD3B2 obtained from RaptorX[10] was energy minimized in NOMAD Ref (http://lorentz.immstr.pasteur.fr/gromacs/minimization_submission.php) and the structure was then provided for the predictions by COACH server. These predictions were also combined with results from other methods including COFACTOR[22], FINDSITE[1] and ConCavity[2] to generate final ligand binding site predictions.

Results and Discussion

Disordered regions in HSD3B2: High degree of flexibility in polypeptide chain and insufficiency of regular secondary structure are considered as disordered regions in a protein. Disordered regions might contain functional sites or linear motifs and many proteins are found as intrinsically disordered *in vivo*. In this study, per-residue intrinsic disorder predictors were utilized. Per-residue predictors provide the distribution of the propensity for intrinsic disorder over the amino acid sequence whereas binary classifiers identify entire protein as wholly ordered or wholly disordered.[8] Five intrinsically disordered regions from HSD3B2 have been identified by GlobPlot (Figure 1). The regions were from amino acid number 145 to 155, 189 to 200, 256 to 264, 310 to 315 and 365 to 370.

In fig. 1, the blue colored sections on the X-axis were disordered regions and green colored regions were globular or ordered domains. Disordered regions as predicted by DisEMBL 1.5 for HSD3B2 were shown in figure 2A and 2B. The IDRs of HSD3B2, when defined by loops/coils definition, showed eleven disordered regions viz. residue numbers 1 to 13, 51 to 70, 124 to 134, 139 to 158, 170 to 197, 206 to 225, 239 to 267, 276 to 287, 305 to 326, 333 to 345, 365 to 372. When disordered regions were predicted by Hot-loops definition, six disordered regions were observed viz. residue numbers 1 to 16, 30 to 66, 140 to 149, 238 to 246, 254 to 263, 361 to 372. DisEMBL 1.5, which utilized Remark-465 definition for prediction of IDRs, showed only one disordered region from residues 364 to 372 in HSD3B2.

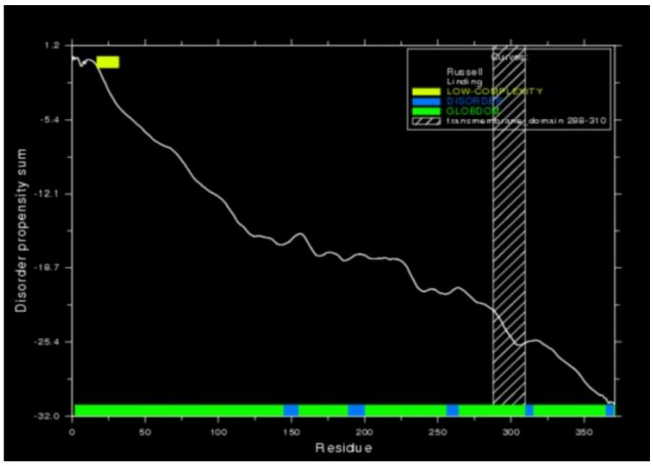

Figure 1: Prediction of intrinsic disorder by FlobPlot. The disordered regions were shown in blue colour. The Russel-Linding algorithm was used for prediction

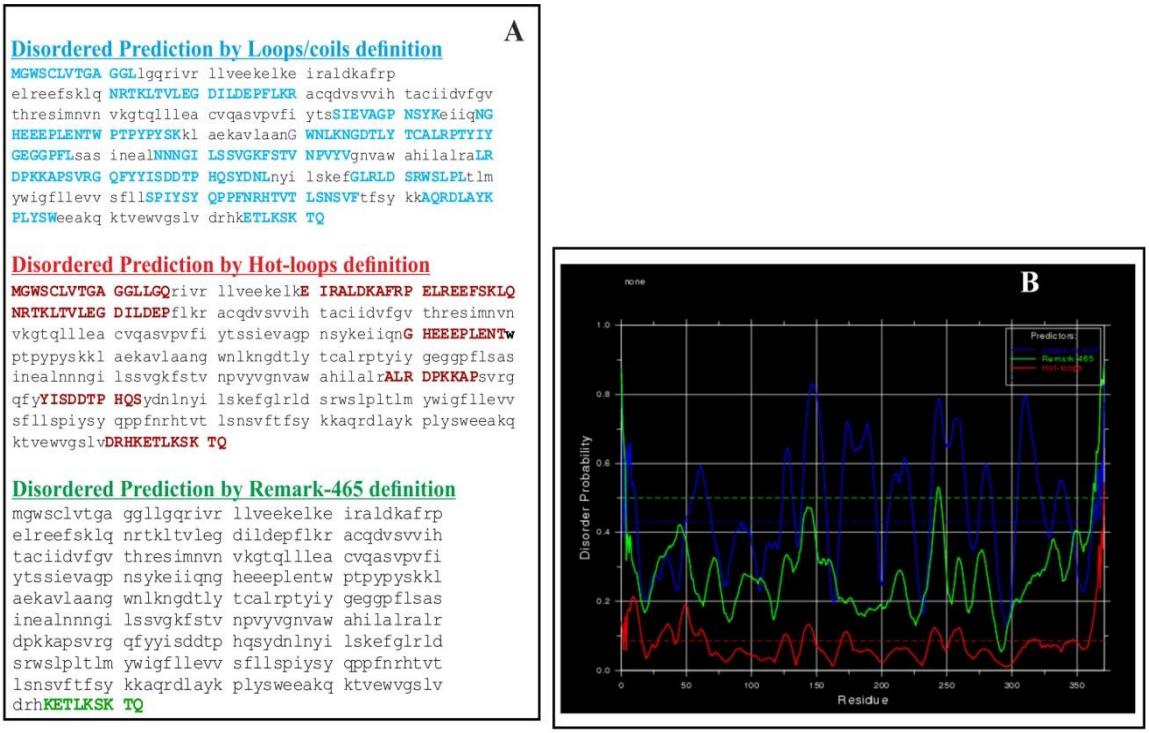

Figure 2: Disordered regions as predicted by DisEMBL1.5 for HSD3B2. Figure 2A showed the disorder prediction by Loop/coils definition (Blue colour; capitalized), Hot-loops definition (Red colour; capitalized) and by Remark-465 definition (Green colour; capitalized). Figure 2B showed the graphical representation of residue wise disorder probability

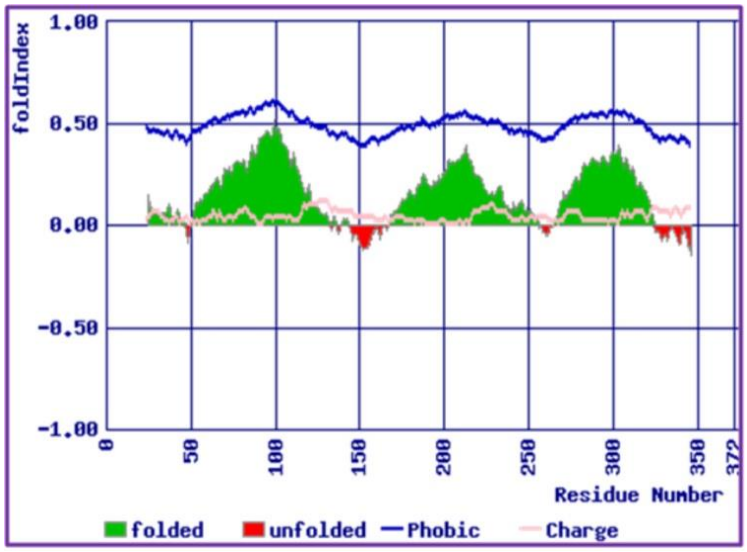

Figure 3: Disorders regions of HSD3B2 as predicted by FoldIndex server

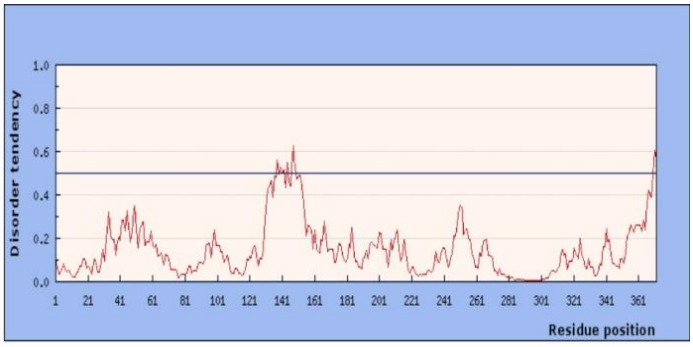

Figure 4: Intrinsic disorder prediction by IUPRED server

Disordered regions as predicted by FoldIndex for HSD3B2 also showed the four disordered regions viz. 143 to 159, 161 to 166, 256 to 262 and 325 to 372 (Figure 3). Results from FoldIndex server also predicted the unfoldability of 0.176 and Phobic of 0.479. Figure 4 showed the results obtained from IUPRED server which predicted that amino acids in the region of 138 to 149 had intrinsic disorder tendency. The conformational flexibility of IDPs and proteins with IDRs allowed them to involve in essential cellular processes.[4] Several studies have shown that intrinsically disordered regions are enriched with uncharged and polar amino acids and lacked bulky hydrophobic residues.[30]

Disordered regions are important because many intrinsically disordered proteins exist as unstructured and become structured when bound to another molecule or ligand.[27] Moreover, IDRs can bind to multiple interacting partners to attain distinct conformational ensemble with each of them, resulting is physiological functions of the protein.[31] Post-translational modifications also enable IDPs/IDRs to attain diverse conformations, thus increasing the total ensemble of conformations to perform multiple apparently unrelated biological functions.[6]

Analysis of ligand binding sites and cavities of HSD3B2: The energy minimized homology based structure of HSD3B2 obtained from RaptorX was utilized for prediction of its ligand binding sites and cavities. The COACH server was used to predict the ligand binding sites of HSD3B2. Figures 5A and 5B showed the NAD binding and uridine diphosphate N-acetylgalactosamine (UD2) binding residues of HSD3B2.

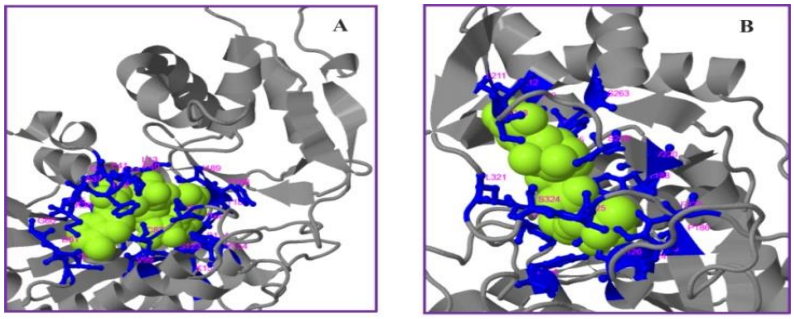

Figure 5: Prediction of ligand binding sites from COACH server.
Figure 5A and 5B showed the NAD binding and uridine diphosphate N-acetylgalactosamine binding residue of HSD3B2 respectively. The interacting residues were shown in blue colour and ligand was shown in yellow colour

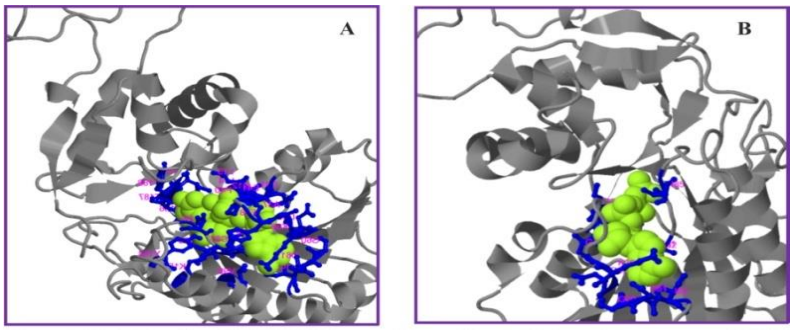

Figure 6: Prediction from COFACTOR server.
Figure 6A and 6B showed the presence of NAD and ATP binding cavities in HSD3B2 respectively. The interacting residues were shown in blue colour and ligand was shown in yellow colour

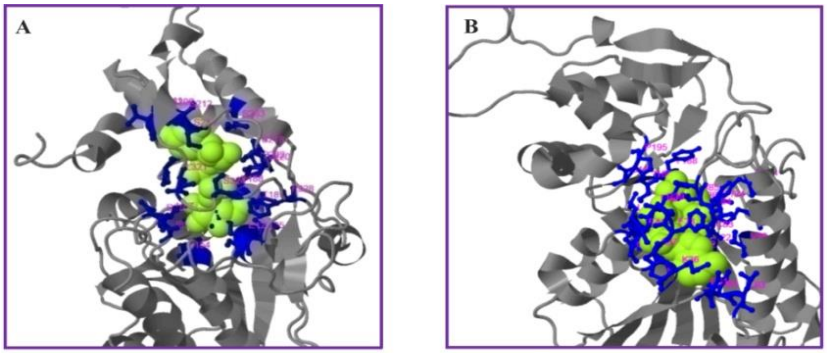

Figure 7: Prediction from FINDSITE server.
Figure 7A showed the presence of ATP binding regions whereas 7B also showed the presence of NAD binding regions in HSD3B2 structure. The interacting residues were shown in blue colour and ligand was shown in yellow colour

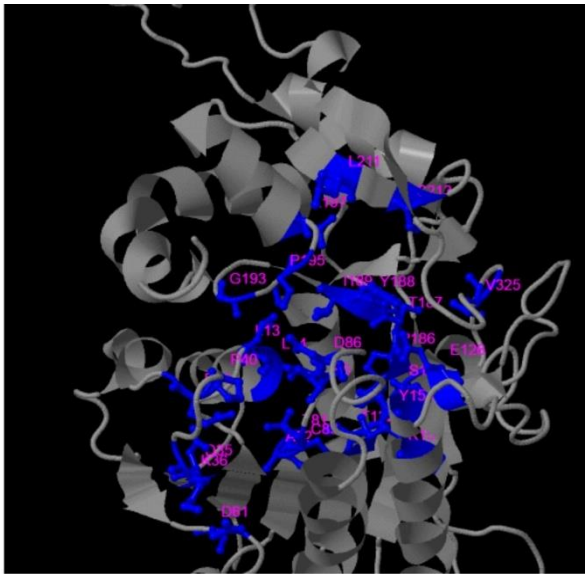

Figure 8: The cavity prediction by ConCavity server. The residues involved in cavity formation were marked in blue.

Table 1
Distribution of amino acids in the intrinsic disordered regions which were present within the predicted ligand binding sites of HSD3B2.

Ligand Name (Server used for prediction)	**Binding residue number**	**No. of amino acids in disordered region**			
		GlobPlot 2.3	**DisEMBL 1.5**	**FoldIndex**	**IUPRED**
NAD (COACH server)	9, 11, 12, 13, 14, 35, 36, 37, 38, 40, 41, 60, 61, 62, 81, 82, 83, 85, 99, 122, 123, 124, 154, 158, 186, 187, 188 and 189	02	14	02	00
Uridine diphosphate N-acetylgalactosamine (COACH server)	85, 86, 124, 125, 126, 154, 186, 188, 195, 196, 199, 211, 212, 213, 218, 220, 263, 321, 324, 325 and 328	05	17	03	00
NAD (COFACTOR server)	11, 12, 13, 14, 35, 36, 37, 38, 40, 41, 60, 61, 62, 81, 82, 83, 85, 99, 122, 123, 154, 158, 186, 187, 188 and 189	02	12	02	00
ATP (COFACTOR server)	11, 12, 13, 14, 35, 36, 60, 61, 62, 81, 82 and 85	00	06	00	00
NAD (FINDSITE server)	9, 11, 12, 13, 14, 35, 36, 37, 38, 39, 40, 41, 61, 62, 63, 81, 82, 83, 84, 85, 99, 122, 124, 154, 158, 186, 188, 189, 194 and 195	04	15	02	00
ATP (FINDSITE server)	85,86,87,124,125,126,154,187,188,195,196,199,200,211,212, 213,218,220,222, 263, 321,324 and 328	06	17	02	00
Cavity (ConCavity server)	13, 14, 35, 36, 40, 41, 61, 81, 82, 83, 85, 86, 122, 123, 124, 126, 154, 158, 186, 187, 188, 189, 193, 195, 197, 211, 213 and 325	05	16	03	00

Residue numbers 9, 11, 12, 13, 14, 35, 36, 37, 38, 40, 41, 60, 61, 62, 81, 82, 83, 85, 99, 122, 123, 124, 154, 158, 186, 187, 188 and 189 were involved in NAD binding with a confidence score of 0.89 whereas residue numbers 85, 86, 124, 125, 126, 154, 186, 188, 195, 196, 199, 211, 212, 213, 218, 220, 263, 321, 324, 325 and 328 were involved in UD2 binding with a confidence score of 0.15.

Figure 6A and 6B represented the results obtained from COFACTOR server showing NAD and ATP binding cavities in HSD3B2. Residue numbers 11, 12, 13, 14, 35, 36, 37, 38, 40, 41, 60, 61, 62, 81, 82, 83, 85, 99, 122, 123, 154, 158, 186, 187, 188 and 189 were shown to bind NAD with a confidence score of 0.60 and BS score of 1.34 and residue numbers 11, 12, 13, 14, 35, 36, 60, 61, 62, 81, 82 and 85 were shown to bind ATP with a lower confidence score of 0.26 and BS score of 1.04. Results from FINDSITE server (Figure 7A and 7B) also showed the presence of both ATP and NAD binding regions in HSD3B2 with a better confidence score for prediction of ATP binding site than NAD binding site.

The residues 85, 86, 87, 124, 125, 126, 154, 187, 188, 195, 196, 199, 200, 211, 212, 213, 218, 220, 222, 263, 321, 324 and 328 formed the ATP binding site with a confidence score of 0.41 whereas NAD binding site was formed by residues 9, 11, 12, 13, 14, 35, 36, 37, 38, 39, 40, 41, 61, 62, 63, 81, 82, 83, 84, 85, 99, 122, 124, 154, 158, 186, 188, 189, 194 and 195 with a confidence score of 0.38. The results from the ConCavity server showed the presence of a single cavity in HSD3B2 with good confidence score (Figure 8). Cavity was shown to be formed by residue numbers 13, 14, 35, 36, 40, 41, 61, 81, 82, 83, 85, 86, 122, 123, 124, 126, 154, 158, 186, 187, 188, 189, 193, 195, 197, 211, 213 and 325 having a confidence score of 0.51. It was evident that many residues involved in the ligand binding and cavity formation were within the disordered regions (Table 1). This implies that these amino acid resides in the unstructured or disordered regions favour ligand binding and it might be possible that upon ligand binding, these regions would be structured.[16] It is known that IDPs/IDPRs also exhibit some structural and dynamical ordering and their larger structural plasticity emphasizes the importance of entropically driven features.[11]

Conclusion

In conclusion, the computational analysis of human HSD3B2 showed the presence of several intrinsic disordered regions within the protein. According to the prediction algorithm of each disorder predictor, the varying patterns of disordered regions were found and some of these IDRs collide with other prediction method used. The HSD3B2 contained ligand binding pockets and the NAD binding residues were identified. Some residues involved in formation of cavity and ligand binding sites were within the disordered regions.

Acknowledgement

I want to acknowledge UGC-Minor research grant [F. No. PSW-092/14-15 (ERO)] for instrumentation facility. I also want to acknowledge Department of Higher Education, Government of West Bengal.

References

1. Brylinski M. and Skolnick J., A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation, Proc Natl Acad Sci U.S.A., **105**, 129-134 **(2008)**

2. Capra J.A. et al, Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure, *PLoS Comput Biol*, **5**, 12 **(2009)**

3. Dosztányi Z. et al, IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, *Bioinformatics*, **21**, 3433-3434 **(2005)**

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

5. Eldar-Geva T. et al, Secondary biosynthetic defects in women with women with late-onset congenital adrenal hyperplasia, *N Engl J Med*, **323**, 855-863 **(1990)**

6. Follis A.V. et al, Regulation of apoptosis by an intrinsically disordered region of Bcl-xL, *Nat Chem Biol*, **14**, 458–465 **(2018)**

7. Goswami A.M., Structural modeling and in silico analysis of non-synonymous single nucleotide polymorphisms of human 3β-hydroxysteroid dehydrogenase type 2, *Meta Gene*, **5**, 162-172 **(2015)**

8. He B. et al, Predicting intrinsic disorder in proteins: an overview, *Cell Research*, **19**, 929-949 **(2009)**

9. Jadot O., Thiry G. and Bury F., Congenital adrenal hyperplasia and ambiguous genitalia due to 3 beta hydroxysteroid dehydrogenase deficiency, *Rev Med Liege*, **59**, 485-488 **(2004)**

10. Källberg M. et al, Template-based protein structure modeling using the RaptorX web server, *Nat Protoc*, **7**, 1511–1522 **(2012)**

11. Kulkarni P. and Uversky V.N., Intrinsically Disordered Proteins in Chronic Diseases, *Biomolecules*, **9**, 147 **(2019)**

12. Labrie F. et al, Structure, function and tissue-specific gene expression of 3b-hydroxysteroid dehydrogenase/D5-D4 isomerase enzymes in classical and peripheral intracrine steroidogenic tissues, *J Steroid Biochem Molec Biol*, **43**, 805-826 **(1992)**

13. Linding R. et al, Protein disorder prediction: implications for structural proteomics, *Structure*, **11**, 1453-1459 **(2003)**

14. Linding R. et al, GlobPlot, exploring protein sequences for globularity and disorder, *Nucl Acid Res*, **31**, 3701–3708 **(2003)**

15. Mason J.I. et al, The regulation of 3b-hydroxysteroid dehydrogenase expression, *Steroids*, **62**, 164–168 **(1997)**

16. Metallo S.J., Intrinsically disordered proteins are potential drug targets, *Curr Opin Chem Biol*, **14**, 481–488 **(2010)**

17. Pang S., Congenital adrenal hyperplasia owing to 3b-hydroxysteroid dehydrogenase deficiency, *Endocrinol Metab Clin North Am*, **1**, 81-99 **(2001)**

18. Prilusky J., Felder C.E. and Zeev-Ben-Mordehai T., FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded, *Bioinformatics*, **21**, 3435-3438 **(2005)**

19. Rainey W. et al, The adrenal genetic puzzle: how do the fetal and adult pieces differ?, *Endocr Res*, **28**, 611-622 **(2002)**

20. Rhe´aume E. et al, Structure and expression of a new complementary DNA encoding the almost exclusive 3b-hydroxysteroid dehydrogenase/delta 5-delta 4-isomerase in human adrenals and gonads, *Mol Endocrinol*, **5**, 1147–1157 **(1991)**

21. Rheaume E. et al, Congenital adrenal hyperplasia due to point mutations in the type II 3b hydroxysteroid dehydrogenase gene, *Nature Gen*, **1**, 239-245 **(1992)**

22. Roy A., Yang J. and Zhang Y., COFACTOR: An accurate comparative algorithm for structure-based protein function annotation, *Nucl Acid Res*, **40**, W471-W477 **(2012)**

23. Sakkal-Alkaddour H. et al, Studies of 3 beta-hyddroxysteroid dehydrogenase genes in infants and children manifesting

premature pubarche and increased ACTH stimulation delta 5 steroid levels, *J Clin Endocrinol Metab*, **81**, 3961-3965 **(1996)**

24. Simard J. et al, Molecular biology and genetics of the 3b-hydroxysteroid dehydrogenase/ delta5-delta4 isomerase gene family, *Journal of Endocrinology*, **150**, S189–S207 **(1996)**

25. Simard J., Moisan A. and Morel Y., Congenital adrenal hyperplasia due to 3b hydroxysteroid dehydrogenase/D5-D4 isomerase deficiency, *Semin Reprod Med*, **3**, 255-276 **(2002)**

26. Simard J. et al, Molecular basis of congenital adrenal hyperplasia due to 3b-hydroxysteroid dehydrogenase deficiency, *Mol Endocrinol*, **7**, 716-728 **(1993)**

27. Strodel B., Energy landscapes of protein aggregation and conformation switching in intrinsically disordered proteins, *J Mol Biol*, **433**, 167182 **(2021)**

28. Thomas J. et al, Structure/function relationships responsible for coenzyme specificity and the isomerase activity of human type 1 3 beta-hydroxysteroid dehydrogenase/isomerase, *J Biol Chem*, **278**, 35483-35490 **(2003)**

29. Tompa P., Unstructural biology coming of age, *Curr Opin Struct Biol*, **21**, 419–425 **(2011)**

30. Trivedi R. and Nagarajaram H.A., Intrinsically Disordered Proteins: An Overview, *Int J Mol Sci*, **23**, 14050 **(2022)**

31. Uversky V.N., Analyzing IDPs in Interactomes, *Methods Mol Biol*, **2141**, 895–945 **(2020)**

32. Xie H. et al, Functional Anthology of Intrinsic Disorder, 3, Ligands, Post-Translational Modifications and Diseases Associated with Intrinsically Disordered Proteins, *J. Proteome. Res.*, **6**, 1917-1932 **(2007)**

33. Yang J., Roy A. and Zhang Y., Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment, *Bioinformatics*, **29**, 2588-2595 **(2013)**.

(Received 11th January 2023, revised 08th September 2024, accepted 15th October 2024)