

Research Article

Computational Biophysical, Biochemical, and Evolutionary Signature of Human R-Spondin Family Proteins, the Member of Canonical Wnt/ β -Catenin Signaling Pathway

Ashish Ranjan Sharma,^{1,2} Chiranjib Chakraborty,^{1,3} Sang-Soo Lee,¹ Garima Sharma,¹ Jeong Kyo Yoon,⁴ C. George Priya Doss,⁵ Dong-Keun Song,¹ and Ju-Suk Nam¹

¹ Institute for Skeletal Aging & Orthopedic Surgery, Hallym University-Chuncheon Sacred Heart Hospital, Chuncheon 200704, Republic of Korea

² Institute for Skeletal Aging & Orthopedic Surgery, Hallym University Hospital, College of Medicine, Chuncheon-si, Gangwon-do 200-704, Republic of Korea

³ Department of Bioinformatics, School of Computer Sciences, Galgotias University, Greater Noida 203201, India

⁴ Center for Molecular Medicine, Maine Medical Center Research Institute, 81 Research Drive, Scarborough, ME 04074, USA

⁵ Medical Biotechnology Division, School of Biosciences and Technology, VIT University, Vellore, Tamil Nadu 632014, India

Correspondence should be addressed to Chiranjib Chakraborty; drchiranjib@yahoo.com and Ju-Suk Nam; jsnam88@hallym.ac.kr

Received 4 April 2014; Revised 12 July 2014; Accepted 12 July 2014; Published 8 September 2014

Academic Editor: Tun-Wen Pai

Copyright © 2014 Ashish Ranjan Sharma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In human, Wnt/ β -catenin signaling pathway plays a significant role in cell growth, cell development, and disease pathogenesis. Four human (Rspo)s are known to activate canonical Wnt/ β -catenin signaling pathway. Presently, (Rspo)s serve as therapeutic target for several human diseases. Henceforth, basic understanding about the molecular properties of (Rspo)s is essential. We approached this issue by interpreting the biochemical and biophysical properties along with molecular evolution of (Rspo)s thorough computational algorithm methods. Our analysis shows that signal peptide length is roughly similar in (Rspo)s family along with similarity in aa distribution pattern. In Rspo3, four N-glycosylation sites were noted. All members are hydrophilic in nature and showed alike GRAVY values, approximately. Conversely, Rspo3 contains the maximum positively charged residues while Rspo4 includes the lowest. Four highly aligned blocks were recorded through Gblocks. Phylogenetic analysis shows Rspo4 is being rooted with Rspo2 and similarly Rspo3 and Rspo1 have the common point of origin. Through phylogenomics study, we developed a phylogenetic tree of sixty proteins ($n = 60$) with the orthologs and paralogs seed sequences. Protein-protein network was also illustrated. Results demonstrated in our study may help the future researchers to unfold significant physiological and therapeutic properties of (Rspo)s in various disease models.

1. Introduction

R-spondins (Rspo)s are a recently discovered family of genes that encodes cysteine-rich secretory proteins containing a thrombospondin type 1 domain/repeat-1 [1]. The (Rspo)s family includes four conserved proteins (Rspo1, Rspo2, Rspo3, and Rspo4), showing overall similarity of 40–60% sequence homology and domain organization [2]. Besides the existence of TSR-1 domain, all four (Rspo)s can be recognized by the existence of a carboxy-terminal region with

positively charged amino acids and two furin-like cysteine-rich repeats adjacent to the amino terminus of the mature protein. Numerous studies have implicated (Rspo)s for acting synergistically with extracellular components of the Wnt signaling pathway (Figure 1) [3–5]. Studies showed close or overlapped gene expression of Wnt and (Rspo)s during developmental events, implying a possible coupling of the (Rspo)s with Wnt signaling [6–8]. Consistent with this, a significant reduction in mRNA expression of Rspo1 was observed in a Wnt1/3a double knockout mouse [1]. Rspo1

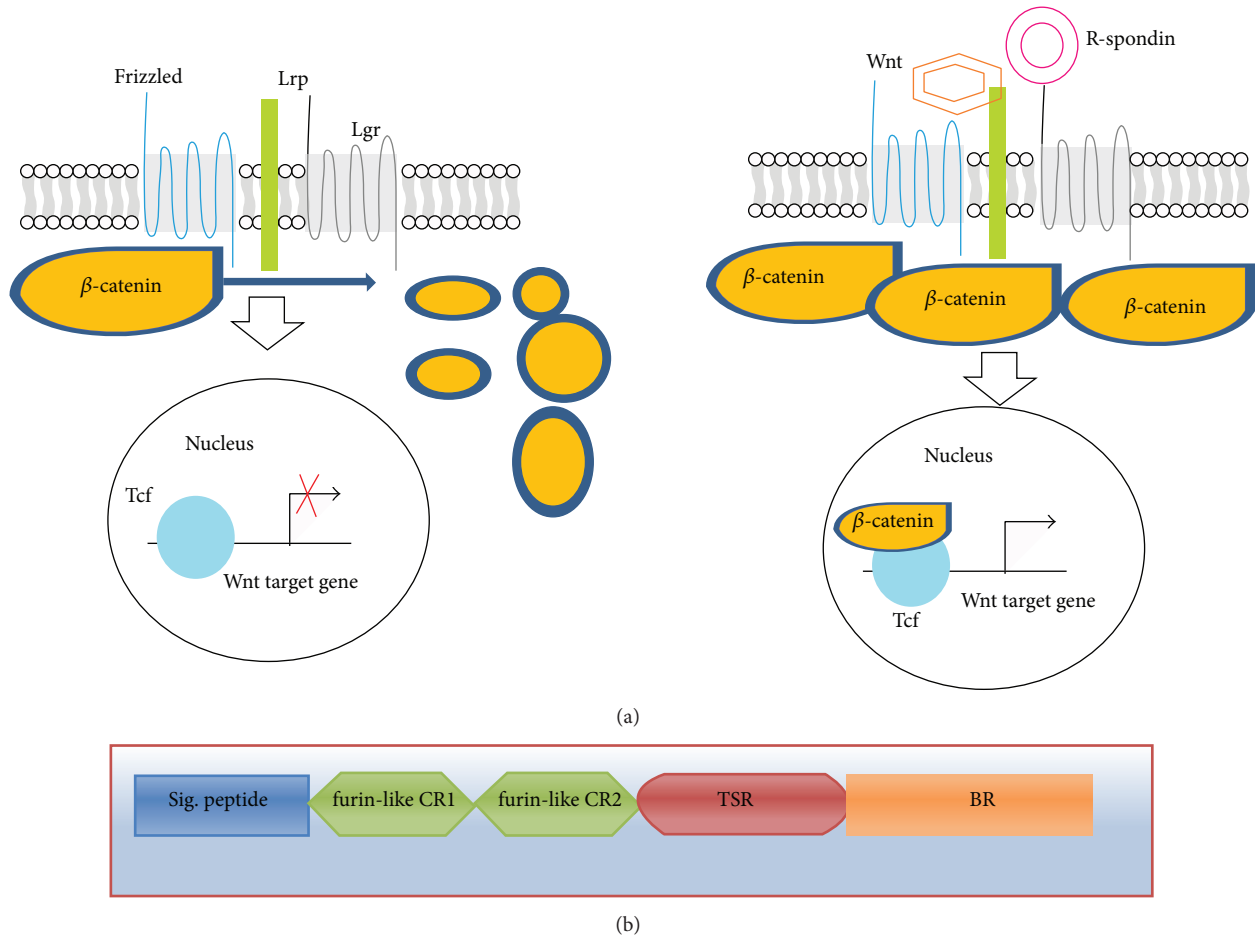


FIGURE 1: The role of (Rspo)s in canonical Wnt signaling pathway and the general architecture of (Rspo)s. (a) Schematic diagram of Wnt and R-spondin signaling models. In absence of Wnt ligand, constitutively synthesized cytoplasmic β -catenin is destroyed by the β -catenin destruction complex causing no β -catenin complex formation with T-cell transcription factor (Tcf)/Lef transcription factors for an active transcriptional response. Canonical Wnt signaling is instigated by the binding of Wnt ligands to the frizzled/LRP receptor complex which in turn deactivates the β -catenin destruction complex increasing its concentration in cytoplasm. The Wnt-frizzled/LRP complex-induced cytoplasmic buildup of β -catenin leads to its import into the nucleus and binding to (Tcf)/Lef transcription factors initiating transcription of Wnt targeted genes. (Rspo)s also act in similar manner but induce this unique property of enhancing Wnt activity by binding to recently discovered seven transmembrane G protein coupled receptors, Lgr (4, 5, and 6). (b) Schematic diagram shows the general domain architecture of human (Rspo)s. The architecture shows (i) signal peptide at N terminal end, (ii) two cysteine-rich furin-like repeats/domains, (iii) a single thrombospondin domain, and (iv) a basic amino-acid-rich domain at C-terminal basic region.

has been shown to augment Wnt signaling by interacting with the low-density lipoprotein receptor related protein 5 or 6 (LRP5/6) coreceptor and inhibiting Dickkopf-1 (Dkk-1) mediated receptor internalization [9]. Rspo2 deficient mice show death at early stages and have limb patterning defects associated with altered Wnt signaling [10, 11]. Rspo3 interacts with Frizzled 8 and LRP-6 and enhances Wnt ligand signaling [3, 4]. In addition to interaction with Wnt/ β -catenin signaling, (Rspo)s can also regulate noncanonical Wnt signaling [12]. It was found that furin domain repeats are essential and sufficient for (Rspo)s to mediate Wnt-potentiating effects [13, 14]. Most recently, several studies conclusively determined that the (Rspo)s are the ligands for the leucine-rich repeat containing G protein-coupled receptor 4/5/6 (LGR4/5/6 receptors) [15–18].

Wnt signaling plays a fundamental role during fate determination steps of embryonic development and has been shown to govern process like cell differentiation, cell proliferation, and stem cell maintenance [19, 20]. Due to (Rspo)s ability to function as regulators of Wnt signaling pathways, various potential roles of (Rspo)s have been proposed and have been suggested as novel therapeutic targets [17, 21]. Rspo1 has been shown to control sex phenotypes between individuals. A study by Parma et al. [22] observed sex reversal due to the homozygous Rspo1 gene mutations in affected individuals. In addition, palmoplantar hyperkeratosis and predisposition to squamous cell carcinoma of the skin were also observed in these individuals. Rspo1 has also been recognized as a potent and specific mitogen for the gastrointestinal epithelium [13, 23]. Various studies have also

implicated the importance of Rspo1 in skeletal biology. Rspo1 has been shown to synergize with Wnt3a to promote the process of osteoblast differentiation and inhibit the process of osteoclastogenesis by inducing expression of osteoprotegerin (OPG) [24–26]. Expression of Rspo2 has been shown to promote myogenesis via the Wnt/ β -catenin signaling pathway in *Xenopus* [6]. A study with Rspo2 gene-targeted mutant mice observed that Rspo2 is requisite for normal development of several tissues, including craniofacial structures, lung, kidney, and limbs [27]. Moreover, study reported that Rspo2 is required for the maintenance of apical ectoderm ridge in the hind limbs of the mice. In other studies on Rspo2 mutant mice, hypoplasia and branching defects within the lungs were also being reported [11, 28]. It was observed that Lrp6-mediated Rspo2 signaling via the canonical Wnt pathway is essential for normal morphogenesis of the respiratory tract and for limbs as well [11]. Investigation into the genes responsible for coat features in domestic dogs revealed that Rspo2 is also supported in the Wnt-mediated hair follicle growth [29]. More recently, the role of recurrent Rspo2 gene fusion exclusively with APC mutations has been linked to the activation of Wnt signaling and colon tumorigenesis [30]. Like Rspo2 gene, recurrent Rspo3 gene fusions were also found to be associated with human colon tumors [30]. In recent time, it was proposed that Rspo3 gene may function along with Rspo2 gene in hind limb development, since the knockout of both Rspo2 and Rspo3 in limb mesenchymal cells caused more severe hind limb defects than those of Rspo2 mutant mice [31]. Rspo2 and Rspo3 genes were also identified for their oncogenic potential in mouse mammary tumor virus associated with mammary tumorigenesis in mice [32, 33]. Expression of Rspo4 has been shown to play a key role during nail development and mutations in Rspo4 gene results into absence of the nails in humans termed as anonychia/hyponychia congenita [34].

Given the diverse role of (Rspo)s in dynamic processes of life, like embryogenesis, tumor progression, angiogenesis, myogenesis, development of skeletal system, and so forth, we can expect (Rspo)s as vital therapeutic targets for a number of disabilities. Therefore, we tried to decipher biochemical, biophysical, molecular evolution, and protein-protein interaction characteristics of (Rspo)s by a series of computer based analysis. It may help us to understand the basic molecular properties of these molecules and thus their participation in critical events regulating essential life processes.

2. Materials and Methods

2.1. Data Mining for Human R-Spondin Protein Family Sequences and Their Feature of the Different Regions. We gathered the information on the sequences of human (Rspo)s family members based on searches in the National Centre for Biotechnology Information database (<http://www.ncbi.nlm.nih.gov/protein>) [35] and UniProt (<http://www.uniprot.org/>) [36, 37]. The FASTA formats of the sequences were further retrieved for analysis. To investigate the features of the primary structure such as the signal peptide in the protein chain and the chain other than the signal peptide portion,

we used UniProt server (<http://www.uniprot.org/>), a database for information on proteins [36–38]. To understand signal peptide with “C-score” (predicted cleavage site value), “S-score” (the predicted signal peptide value), and “Y-score” (a combination of C- and S-scores), SignalP 4.0 server was used [39]. In addition, different repeats and domain in the R-spondin family members have been analysed using UniProt server.

2.2. Investigation of Amino Acid Distribution, Amino Acid Composition, and Some Parameters Related to the Primary Structure Such as Charge Distribution Analysis, Repetitive Structures, Cysteine Positions, and Disulphide Bonds of Human R-Spondin Family Proteins. To understand the amino acid distribution in the investigated proteins, we used protein calculator (<http://spin.niddk.nih.gov/clare/Software/A205.html>) [40]. In order to examine the amino acid prototype and protein sequence properties, such as amino acid composition percentage, high scoring hydrophobic segments, and tandem and periodic repeats of structure data of the human (Rspo)s, we used the statistical analysis of protein sequences (SAPS) [41], which is one of the most significant tools to bring out the details about protein sequence properties.

For the study of the secondary structural aspect of R-spondin family members such as cysteine positions and disulphide bond topology prediction, we used “SCRATCH protein predictor” for cysteine positions [42] as well as UniProt (<http://www.uniprot.org/>) [36, 37] server.

2.3. Structural Prediction of Thrombospondin-1 Domain Type 1 (TSP1) Repeats and Its Molecular Dynamics and Geometry. To understand the thrombospondin-1 domain type 1 repeats, we used the PDB file (1LSL.pdb) extracted from the protein data bank (<http://www.rcsb.org>); for further analysis see [43, 44]. The structure was visualized using Jmol Applet. We used InterPro, a database for protein families, domains, and functional sites, to understand domain structure [45]. The geometry of thrombospondin-1 domain type 1 repeats such as B factor plot, Omega plot, and FDS (fold deviation score) plot was developed using PDB server. Furthermore, we also developed Ramachandran plot for thrombospondin-1 domain type 1 repeats using PROCHECK server (<http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>).

2.4. Prediction of Glycosylation Sites. Analyses of the sequence location of the posttranslational modifications assist to determine the functional characteristics of the proteins. Glycosylation is a type of posttranslational modification (PTM) that assists in protein structural folding, transport, and different types of functions. We predicted the two kinds of glycosylation such as O-glycosylation and N-glycosylation sites by using the NetNGlyc and NetOGlyc servers of the four human (Rspo)s [46–48].

2.5. Prediction of R-Spondin Family Proteins Instability Index, Grand Average of Hydrophobicity (GRAVY), Aliphatic Index,

and Total Number of Positively/Negatively Charged Residues. A comparison of the various biophysical and biochemical parameters of the proteins coded by the human (Rspo)s was carried out using the ProtParam tool from the ExPASy portal (<http://web.expasy.org/protparam/>) [49]. The different computed parameters for the (Rspo)s includes instability index, aliphatic index, grand average of hydrophobicity (GRAVY), total number of negative charged residues (Asp, Glu), and the total number of positive charged residues (Arg, Lys).

2.6. Prediction of Globularity in the R-Spondin Protein Family. Globular (globe-like) domain of the protein is having spherical domain. The ability to discover the functional sites of domains in proteins is becoming increasingly important. GlobPlot was used to predict the globularity in the domains [17]. The algorithm was as follows:

$$\Omega(a_i) = \sum_{j=1}^{i-1} \Omega(a_j) + \ln(i+1) \cdot P(a_i) \quad \text{for } i = 1, \dots, L. \quad (1)$$

For the protein sequence which is used for analysis, the length of the sequence is L ; Linding et al. [50] defined the sum function Ω as $P(a_i) \in R$. $P(a_i)$ is the propensity of the i th amino acid and \ln is the natural logarithm. The globularity in the domains of the regulatory subunit p85 α was determined using the GlobPlot Web server.

2.7. Multiple Sequences Alignment (MSA) Analysis among R-Spondin Family Proteins. Four sequences of R-spondin family proteins were used to understand the sequences similarity and alignment positions using MSA analysis. For that, we used clustal-omega to understand the sequence similarities and to elucidate the respective pairwise alignment scores. Clustal-omega has a graphical interface that is easy to use [51]. The clustal-omega server was organized on the basis of “progressive algorithm” [52] and the scoring system of the pairwise alignment algorithm is possibly the powerful component of the progressive algorithm. During the best alignment between N sequences, a computational complexity is found (L^N) for N sequences of length L . The basic algorithm to elucidate respective pairwise alignment scores is based on Needleman and Wunsch’s algorithm [53].

Additionally, other MSA tools were used known as “multiple sequence comparison by log-expectation” (MUSCLE) to locate the conserved pattern across R-spondin protein family [54]. MUSCLE uses a function that can be described as the following log-expectation (LE) score function:

$$LE^{xy} = (1 - f_G^x)(1 - f_G^y) \log \sum_i \sum_j f_i^x \frac{f_j^y p_{ij}}{p_i p_j}. \quad (2)$$

This function is a modified version of the log-average function expressed as follows:

$$LA^{xy} = \log \sum_i \sum_j f_i^x \frac{f_j^y p_{ij}}{p_i p_j}, \quad (3)$$

where i and j are amino acid types; p_i is the background probability of i ; p_{ij} is the joint probability of i and j being aligned to each other; f_i^x is the observed frequency of i in column x of the first profile; and f_G^x is the observed frequency of gaps in that column at position x in the family and likewise for position y in the second profile. The approximate probability α_i^x of experimental amino acid i in location x can be derived from f_x . The graphical yield of MUSCLE was visualised through JalView. Finally, Gblocks server was used to observe the aligned blocks of the sequences, which describes a set of conserved blocks from an MSA according to a set of simple requirements [55].

2.8. Multiple Sequences Alignment (MSA) Analysis of R-Spondin Family Proteins with Other Species. To understand the sequence similarity of human four (Rspo)s with other species, we used PhylomeDB server [56, 57]. This server performed homology searches by means of the Smith-Waterman algorithm [58] and ultimately filtered the sequences according to specific e -value and overlap cut-offs.

2.9. Analysis of Molecular Phylogenetics of Human R-Spondin Family Proteins. For the molecular phylogenetics, we used three servers to develop two phylogenetic trees. First we used accessible computer software and constructed the phylogenetic tree using Phylogeny.fr and performed computational biology [59]. This software uses several kinds of software for the workflow such as MUSCLE multiple alignment, Gblocks for the alignment curation, PhyML for the construction of the phylogenetic tree, and TreeDyn for the visualisation of phylogenetic tree. We have developed two types of the phylogenetic tree, namely, phylogram and cladogram (without branch distance). The phylogram depicted distances among protein sequences within the (Rspo)s. Then, another tree known as the “circular alpha phylogenetic tree” has been developed using MAFFT (version 7) [60]. Again using the four family sequences, we used clustal-omega to develop another phylogenetic tree [51]. The servers implemented either a neighbour-joining method or the bottom-up clustering method developed by Saitou and Nei [61] and the algorithm used a distance matrix to specify the distance between each pair of taxa. In this case, the matrix had a magnitude which is $N \times N$. In this case, N is the number of points or nodes.

2.10. Prediction of Phylogenomics of Human R-Spondin Family Proteins Using Molecular Phylogenetics to Understand the Framework Topology of Other Related Species. To understand the phylogenomics of four human (Rspo)s and framework topology of other related species, we developed another phylogenetic tree using the sequence similarity of four human (Rspo)s with other species. For this analysis, we use PhylomeDB server, one of the largest phylogenetic repository [56, 57]. This server performed homology searches by means of the Smith-Waterman algorithm [58] and ultimately filtered the sequences according to specific e -value and overlap cut-offs. The server is a resulting collection of trees which characterize the full complement of evolutionary histories

of all genes determined in a given genome. This has been entitled with the term phylome [59]. For phylogenomics analysis, the method used in this study is more closely a gene-centered method. And it is computationally more extensive compared to developing a family-based approach.

2.11. Understanding the Protein-Protein Interaction Network of R-Spondin Family Proteins. We have developed protein-protein interaction network using STRING server to understand the possible protein interactions with (Rspo)s [62, 63]. We developed four interaction networks, one for each (Rspo). Finally, we also developed scores to understand the interaction among possible interacting proteins with (Rspo)s.

3. Results

3.1. Searched Data for Corresponding Proteins and Their Features Such as Signal Peptide, Repeats, and Domains. Supplementary Table S1 (available online at <http://dx.doi.org/10.1155/2014/974316>) shows the protein sequence information related to the human (Rspo)s analysed in this study, while the genes and proteins information related to human (Rspo)s have been displayed in Supplementary Table S2. The sequence lengths of Rspo1, Rspo2, Rspo3, and Rspo4 have been plotted in Figure 2(a). The figure shows that Rspo3 contains the highest sequence length of amino acids (aa) (272 aa), while Rspo4 contains the lowest sequence length (234 aa). Next, we plotted the sequence of amino acid number in the scattered distribution ($R^2 = 0.1824$) (Figure 2(b)). The features of the primary structure such as the signal peptide in the protein chain and the chain other than the signal peptide and information of different regions such as repeat and domain of human R-spondin family members were analysed. We depicted the position of regions, length, and graphical view of such regions in Figures 2(c), 2(d), 2(e), and 2(f). Thereafter, we compared the amino acid length of the signal peptide in the protein chains and the chain other than the signal peptide of these four proteins (Figure 2(g)). We observed that the length of the signal peptide portions is more or less similar (19 to 21 aa length) among the four proteins. Conversely, differences in the amino acid length have been noted in the chain other than the signal peptide portion where Rspo3 comprises the highest sequence length (251 aa) while Rspo4 contains the lowest sequence length (215 aa). Furthermore, we have analyzed the signal peptides of four human (Rspo)s and depicted their “C-score” (predicted cleavage site value), “S-score” (the predicted signal peptide value), and “Y-score” (a combination of C- and S-scores) (Figure 3).

3.2. Investigation of Amino Acid Distribution, Amino Acid Composition, and Some Parameters Related to the Primary Structure Such as Charge Distribution Analysis, Repetitive Structures, Cysteine Positions, and Disulphide Bonds of Human R-Spondin Protein Family. Amino acid distributions of human R-spondin protein family have been reprinted in Figures 4(a), 4(b), 4(c), and 4(d). Furthermore, we exposed the four amino acid distributions at a time to understand

the distribution pattern of these proteins (Figure 4(e)). The composition analysis of the amino acids of human (Rspo)s has been represented in Supplementary Table S3 and Supplementary Figure S1. From the calculated distribution of amino acids as well as the composition of the amino acids of human (Rspo)s, we found the following data: Rspo1 with highest Arg 27 (10.3%) and lowest Trp 3 (1.1%) and Tyr 3 (1.1%) both, Rspo2 with highest Arg 38 (11.1%) and lowest Trp 4 (1.6%), Rspo3 with highest Cys 22 (8.1%) and Ser 22 (8.1%) both and lowest Trp 3 (1.1%), and Rspo4 with highest Gly 27 (11.5%) and Arg 27 (11.5%) both and lowest Trp 3 (1.3%), respectively. From the distribution and composition of amino acid, it was noted that the highest of number of Arg residue was noted in the three proteins (Rspo1, Rspo2, and Rspo4), and the lowest number of residue was Trp in all of the R-spondin family proteins. The charge distribution analysis, repetitive structures, and cysteine positions of human (Rspo)s has been noted in Supplementary Table S3. Total numbers of cysteine and disulphide bonds present among (Rspo)s have been illustrated in Figure 4(f) showing maximum number of cysteine residues in Rspo2 protein (twenty-four). However, disulphide bonds are same in all of the human (Rspo)s (eleven).

3.3. Structural Prediction of Thrombospondin-1 Type 1 (TSP1) Repeats and Its Geometry. Human R-spondin family proteins contain a thrombospondin type 1 domain type 1 repeats [1] (Figure 5(a)). The structure of monomeric assembly of the thrombospondin type 1 domain type 1 repeats has been depicted in Figure 5(b). This domain structure has been illustrated through the CATH and Pfam database and described in Figures 5(c) and 5(d). The surface structure of this domain has been developed with atomic properties described through different colours (Figure 5(e)). TSP1 domain(s) has been identified in a number of proteins, but generally in multiple copies. From this aspect, R-spondin is very unique since it has only one copy and predicted structure of this domain is hinge-like structure. This specific hinge-like structure of TSP1 domain may play a vital role in binding activity with the receptors. It has been found that TSR motifs especially the WSGWSSCSVSCG sequence are most significant for different neuronal responses such as neurite extension, neuronal survival, neuronal aggregation, and so forth [64].

B factors plot signifies the convolution of static and dynamic disorder in the crystal structure. While, dynamic disorder present in a crystal can be recognized through the local motions of individual atoms. Conversely, static disorder signifies the different atomic positions in a particular protein molecule [65]. Omega plot is helpful to understand the proper residue. Fold Deviation Score (FDS) plot is important to understand the structural geometry of the protein [32]. Ramachandran plot is also significant to comprehend residues in a generously allowed region [66]. Therefore, we developed the geometry of the thrombospondin type 1 domain type 1 repeats and the associated different geometry of these domain, such as B factor Plot, Omega plot, Fold Deviation Score (FDS) plot and Ramachandran plot and

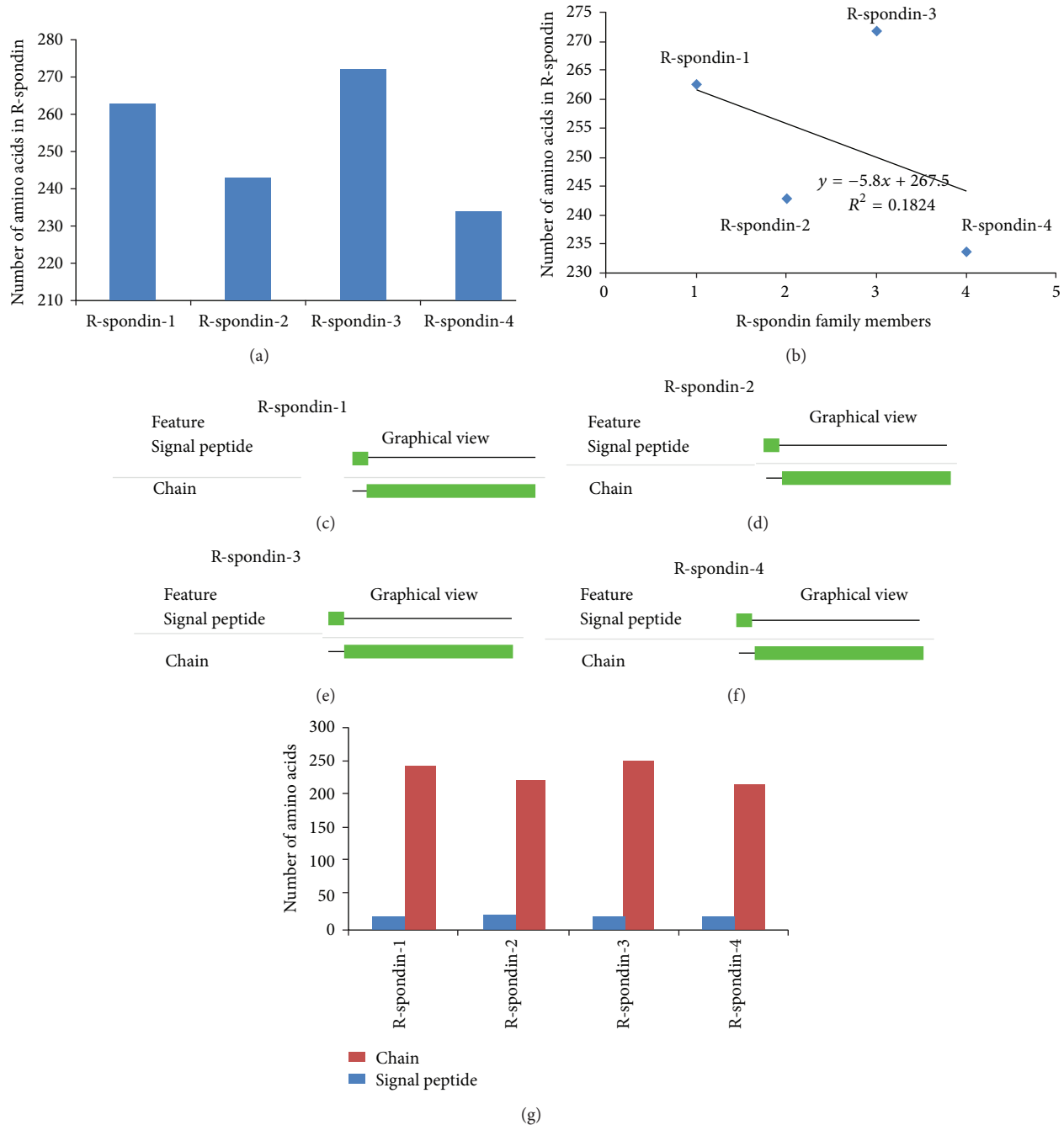


FIGURE 2: General architecture of human (Rspo)s in respect to amino acid sequence. (a) Comparison between the numbers of amino acids in all of the R-spondin family proteins. (b) Plot showing the scattered distribution of amino acid numbers along with the (Rspo)s and their correlations. ((c), (d), (e), (f)) Graphical overview of signal peptide and other parts of the amino acid chain-Rspo1, Rspo2, Rspo3, and Rspo4. (g) Comparison between number of amino acids in signal peptide and other parts of the protein in four human (Rspo)s.

recorded in Supplementary Figures S2(a), S2(B), S2(C), and S2(D), respectively.

3.4. Prediction of Glycosylation Sites. Similar to phosphorylation, in some eukaryotic proteins, glycosylation plays a significant role in protein function and interaction during the signalling process [67]. In biophysical and biochemical point

of view, N-glycosylation sites and O-glycosylation sites are important for functionality of the protein. In reviewing the presence of N-glycosylation sites (Supplementary Table S5) among (Rspo)s, we found the following: Rspo1 with 1 site (at the residue position of 137), Rspo2 with 1 site (at the residue position of 160), Rspo3 with 4 sites (at the residue position of 23, 36, 137 and 194) and Rspo4 with 1 site. The results of Rspo3

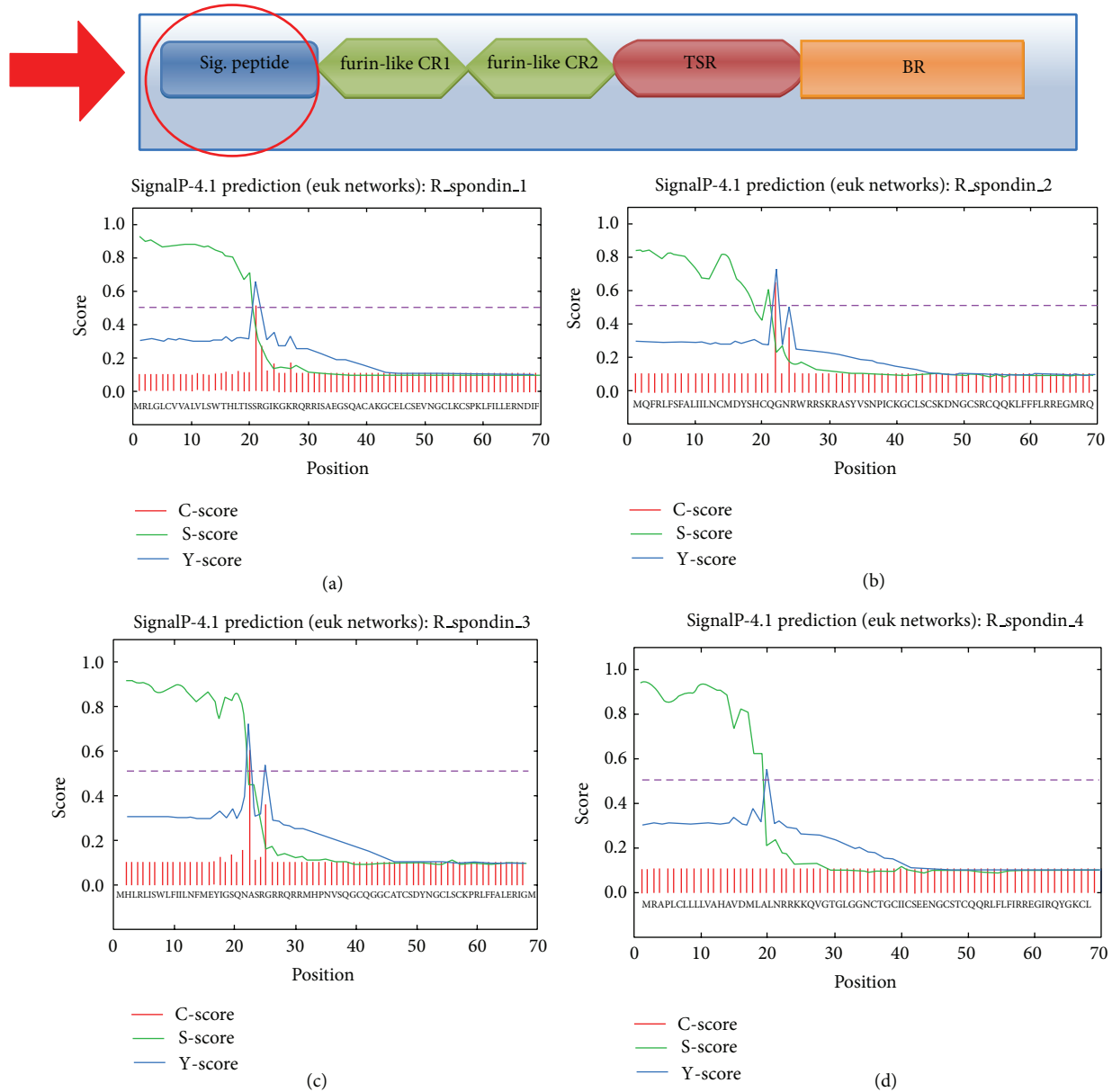


FIGURE 3: Predicted architecture of signal peptide of different human (Rspo)s with “C-score,” “S-score,” and “Y-score” (C-score represents predicted cleavage site value; S-score represents the predicted signal peptide value; Y-score represents a combination of C- and S-scores). (a) Rspo1, (b) Rspo2, (c) Rspo3, and (d) Rspo4 (the schematic diagram shows the location of signal peptide in the general domain architecture of human (Rspo)s and our region of analysis).

showed highest N-glycosylation sites. While reviewing the O-glycosylation potentiality and location (Supplementary Table S6), we found that only Rspo1 has one site. No other (Rspo)s have O-glycosylation sites. However, several O-glycosylation sites potentialities were recorded among (Rspo)s; although, the values of these sites were below the threshold limit (Figure 6).

3.5. Prediction of R-Spondin Family Proteins Instability Index, Grand Average of Hydrophobicity (GRAVY), Aliphatic Index, and Total Number of Positively/Negatively Charged Residues.

The protein stability is associated with different structural properties and functionality of the proteins such as metabolic stability [68], protein-protein interactions [69], and so forth. An instability index provides the knowledge about a protein’s stability, in particular in an *in vitro* environment. The instability index value greater than 40 designates an unstable protein, and one less than 40 designates a stable protein. Several factors such as the arrangement of amino acids in a sequence and some peptide bonds make *in vivo* proteins stable [70]. The results of our instability index analysis of the R-spondin family proteins are shown in Figure 7(a). The Rspo1 was found to have the highest instability index, whereas

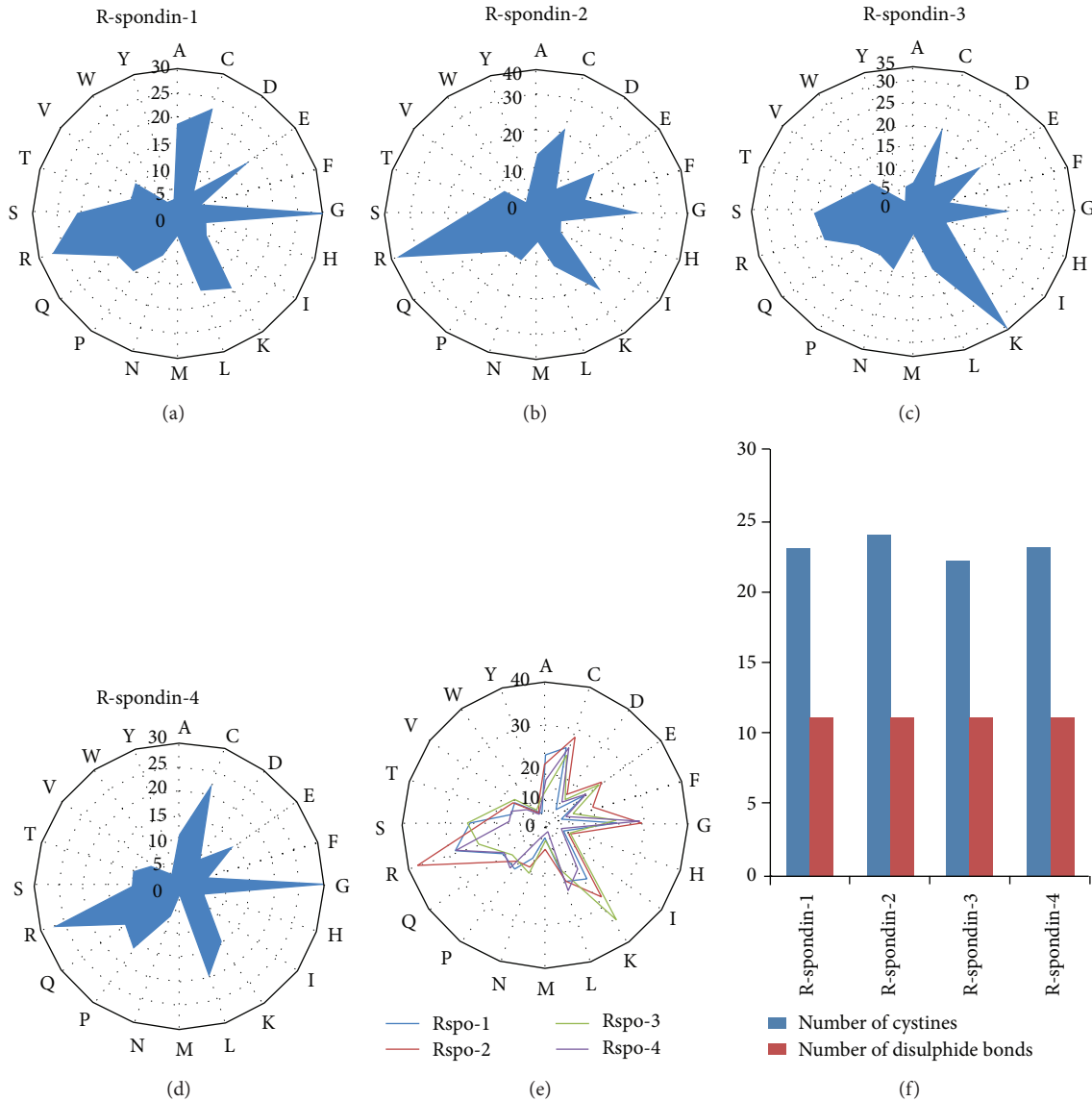


FIGURE 4: General amino acid distribution of amino acids in human (Rsp)s. (a) Rspo1, (b) Rspo2, (c) Rspo3, (d) Rspo4, (e) a general trend of amino acid distribution of amino acids for all human (Rsp)s where we have exposed the four protein's amino acid distribution at a time, and (f) comparison between number of cystine residue and disulphide bond in four human (Rsp)s.

Rspo3 was found to have the lowest. Every R-spondin protein was found to be unstable as per their instability index, since the values are greater than 40. The changes in amino acid composition and hydrophobicity may have caused the observed distinct stability of the protein.

Kyte and Doolittle have formulated the scale of hydropathy in which the hydrophilic and hydrophobic possessions of amino acid chain are assessed in a protein [71]. Grand average of hydrophobicity (GRAVY) score can be computed as the sum of the hydropathy values for all the amino acids in a protein that can be divided by the total number of residues in the protein. Grand average of hydrophobicity (GRAVY) is associated with protein solubility. It has been noted that the positive GRAVY value is positively associated with hydrophobicity and negatively associated with the

hydrophilicity. Because a more hydrophilic protein forms a larger amount of hydrogen bonds with water, therefore, the solubility is more. A ProtParam GRAVY study predicted grand average of hydrophobicity in the (Rspo)s (Figure 7(b)). Our analysis revealed that all (Rspo)s were hydrophilic in nature, Rspo3 being the most hydrophilic. The GRAVY value shows approximate similar values for the Rspo1, Rspo2, and Rspo4 (-0.717, -0.769, and -0.701, resp.).

The aliphatic index (AI) is very significant for understanding a protein, as it describes the relative volume occupied by aliphatic side chains such as alanine, valine, isoleucine and leucine. Aliphatic hydrophobicity is amplified with a rise in temperature and is, therefore, a positive factor enhancing the thermal stability of globular proteins [72]. Our analyses (Figure 7(c)) showed that, Rspo4 have the highest aliphatic

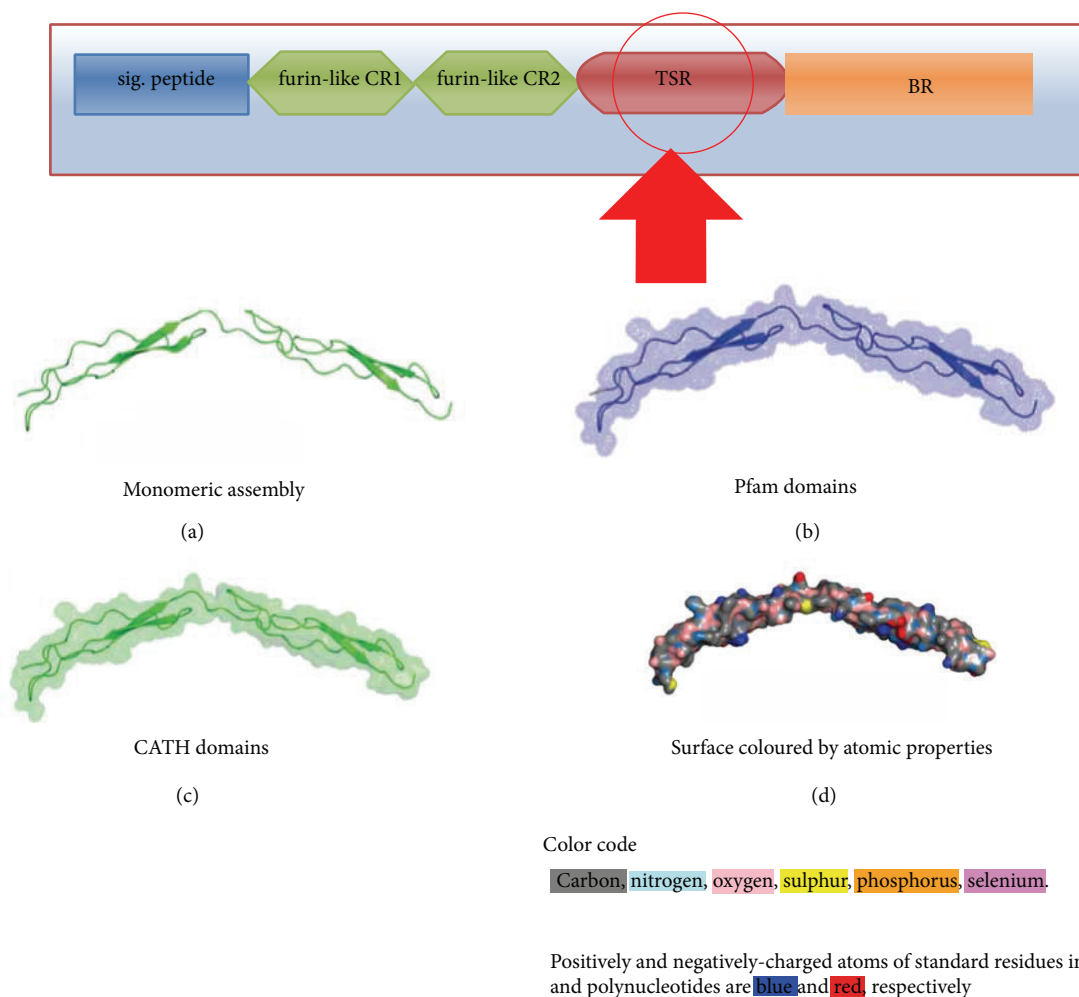


FIGURE 5: Unique backbone structure of thrombospondin-1 type 1 repeats/domain (TSR). (a) Monomeric assembly structure of TSR, (b) structure of TSR domain generated through Pfam domain database, (c) structure of TSR domain generated through CATH domain database, and (d) surface structure of TSR domain shows the atomic properties (the schematic diagram shows the location of thrombospondin-1 type 1 repeats/domain peptide in the general domain architecture of human (Rspo)s).

index among (Rspo)s and the Rspo2 have the lowest. The AI value of Rspo1 (54.94) was approximately closer to the value of Rspo3 (51.58).

It has been reported that AI value is directly proportional to the structural stability of the protein. The procedure is generally used to calculate the AI of a protein [72, 73], which is as follows:

$$AI = X_A + aX_V + b(X_I + X_L), \quad (4)$$

where, X_A , X_V , X_I , and X_L represent the mole percentage of the four residues in a protein which are Ala, Val, Ile, and Leu, respectively. The notation “a” and “b” are coefficients representing the relative volumes of aliphatic side chains and the values are ($a = 2.9 \pm 0.1$ and $b = 3.9 \pm 0.1$), calculated from the volume occupied by the aliphatic amino acids in a protein.

Positively charged residues (PCR) and negatively charged residues (NCR) control several cell properties such as PCR

controlled ribosomal velocity [74], NCR controlled K⁺ channels [75]. These two parameters are helpful to determine the topology of protein [76, 77]. A sum of Arg and Lys are calculated for the presence of the total number of positively charged residues in a protein. Conversely, totality of Asp and Glu are used to calculate the total number of negatively charged residues. Our analysis revealed that, Rspo3 contains the maximum number of positively charged residues while Rspo4 consisted of lowest number. Similarly, Rspo3 consisted of the highest number of negatively charged residues while Rspo4 had the lowest number (Figure 7(d)). The results signify that total numbers of positively charged residues are more than the total number of negatively charged residues for all (Rspo)s.

3.6. Prediction of Globularity in the R-Spondin Family Proteins. From globular domains, several conventional concepts of protein science were initially developed and it challenge by essentially disordered domains [78]. It is frequently analysed

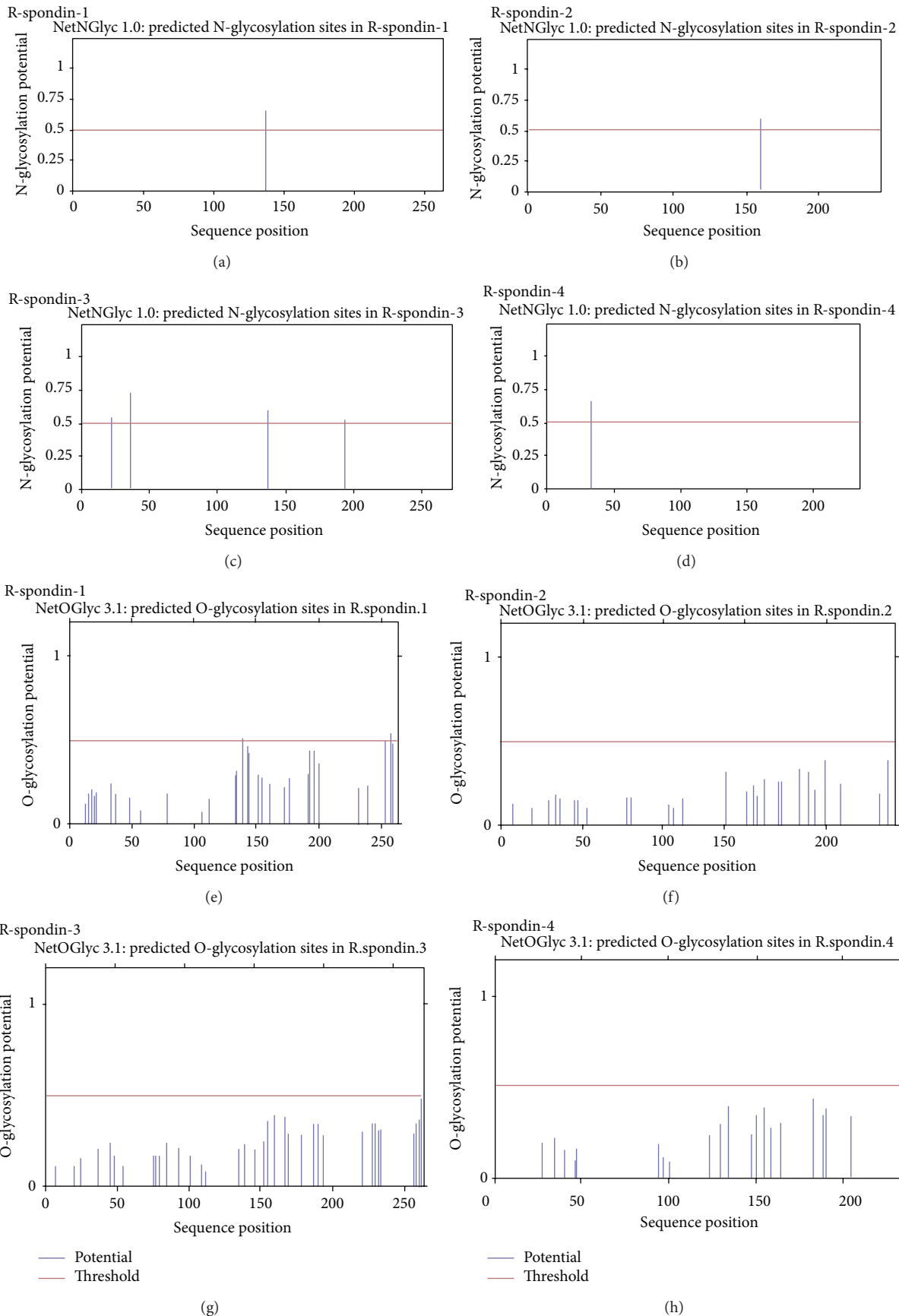
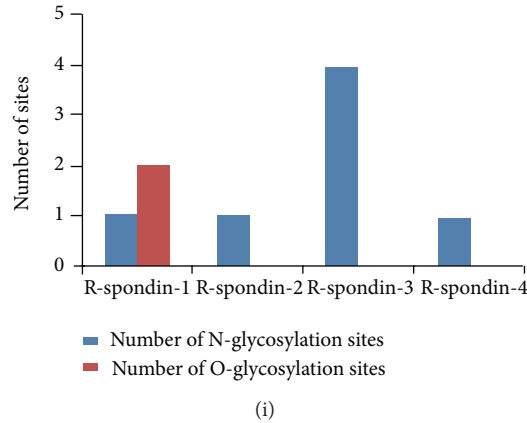
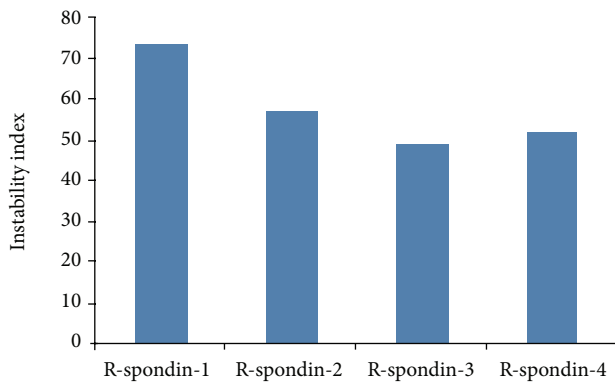


FIGURE 6: Continued.

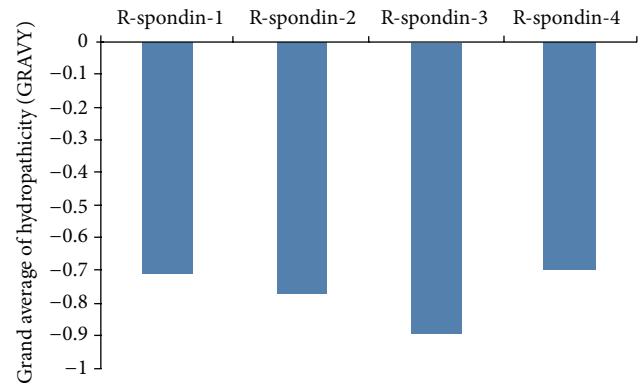


(i)

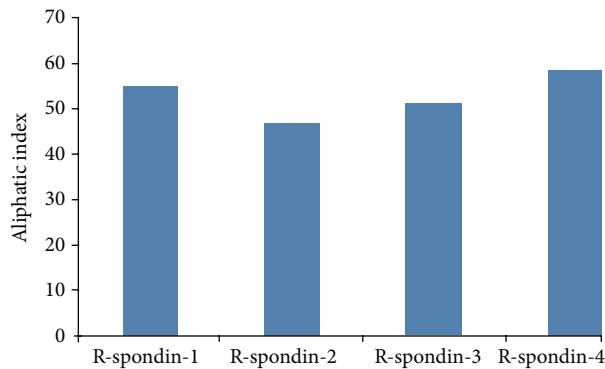
FIGURE 6: Predicted N-glycosylation and O-glycosylation potentialities and their positions in the different human (Rspo)s. (a) N-glycosylation potentialities of Rspo1, (b) N-glycosylation potentialities of Rspo2, (c) N-glycosylation potentialities of Rspo3, (d) N-glycosylation potentialities of Rspo4, (e) O-glycosylation potentialities of Rspo1, (f) O-glycosylation potentialities of Rspo2, (g) O-glycosylation potentialities of Rspo3, (h) O-glycosylation potentialities of Rspo4, and (i) comparison of predicted N-glycosylation and O-glycosylation sites for four human (Rspo)s.



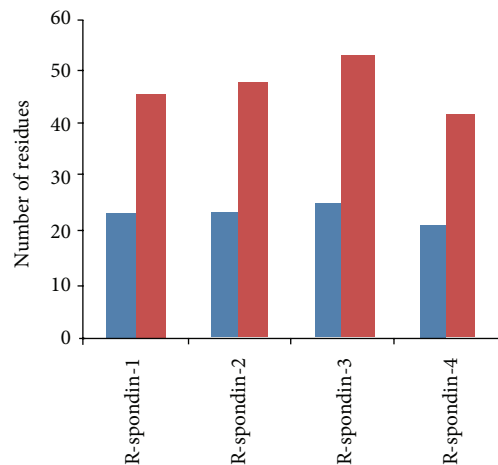
(a)



(b)



(c)



(d)

■ Total number of negatively charged residues (Asp + Glu)
 ■ Total number of positively charged residues (Arg + Lys)

FIGURE 7: Comparison of biophysical and biochemical properties of four human (Rspo)s. (a) Comparison of instability index, (b) comparison of grand average of hydrophobicity (GRAVY), (c) comparison of aliphatic index, and (d) comparison of total number of positively/negatively charged residues.

to understand the structure-function relationships, because the structure is having one or numerous catalytic or binding sites on its surface [79]. The globular domains which we analysed are shown in Figure 8. The amino acid sequence alignment in the upper portion of the figure illustrates the differences between the domains. All the proteins were found to contain disordered regions on its surfaces which are as following: Rspo1 (5), Rspo2 (6), Rspo3 (7) and Rspo4 (6). Although globular domain analysis found that Rspo1, Rspo2 and Rspo3 contains globular domain, but no globular domain was observed in Rspo4.

3.7. Multiple Sequences Alignment (MSA) Analysis among R-Spondin Protein Family. The alignment of the (Rspo) sequences using Clustal Omega is illustrated in supplementary Figure S3. The MUSCLE output was visualised through JalView and is shown in Figure 9(a). As mentioned, 37 small and large aligned divisions were found. We observed best aligned parts between the Rspo4 and Rspo2 sequences, as well as between Rspo4 and Rspo2. We also analysed the highly aligned blocks through Gblocks. The alignment results of Gblocks are shown in Figure 9(b) which shows four highly aligned blocks. From this result, we found highly conserved amino acids such as Leu, Arg, Ser, Gly, Cys, Asn, and Phe.

3.8. Multiple Sequences Alignment (MSA) Analysis of R-Spondin Family Proteins with Other Species. Thereafter, we performed MSA analysis of R-spondin family proteins with other species ($n = 53$). The MSA result is shown in Figure 10. The maximum conservation found was up to 270 sequence and some amino acids such as glycine, cysteine, valine, serine, proline, histidine, leucine and tyrosine were found highly conserved between the sequences.

3.9. Analysis of Molecular Phylogenetics of Human R-Spondin Family Proteins. Phylogram, cladogram and binary tree (equivalent to cladogram) have been depicted (Figures 11(a), 11(b), and 11(c)) and it demonstrates a significant relationship among the proteins of R-spondin family. A molecular phylogenetic analysis of R-spondin members would represent a significant feature of (Rspo)s evolution. In the constructed phylogenetic tree, the distance of branches was illustrated through the likelihood ratio mapping for evolutionary relationships among distinct members of R-spondin family. During the analysis of the tree algorithm, another figure have been described (Figure 11(c)) from the cladogram (Figure 11(b)), that clearly shows the phylogenetic tree rooted with ideal binary numbers (Figure 11(c)). The rooted tree contains two internal nodes and each internal node is further divided into two children nodes, highlighting proteins at their tips. We observed that the altitude of the binary tree was 2 stage. To cross check, other phylogenetic tree called "circular alpha phylogenetic tree", was developed (Figure 11(d)) using MAFFT server. Developed tree resembled the first phylogenetic tree when compared (Figure 11(a)). Both the tree shows, Rspo4 being rooted with Rspo2 and likewise, Rspo3 and Rspo1 have the common point of origin. Again using the four sequences, we developed another phylogenetic tree

using Clustal Omega (Figure 11(e)). This tree also showed that Rspo4 and Rspo2 have the common point of origin. We plotted the branch length from the tree in (Figure 11(f)), where Rspo4 showed longest branch length while Rspo3 had the shortest branch length.

3.10. Prediction of Phylogenomics of Human R-Spondin Proteins Using Molecular Phylogenetics to Understand the Framework Topology of Other Related Species. Presently, the phylogenomics, the study of genomes from an evolutionary perspective, is one of the most significant branches to understand the molecular phylogenetics [80, 81]. Phylogenomics provides an understanding about the framework topology of other related species containing orthologous and paralogous genes. The phylogenomics and the framework topology may provide an understanding about the speciation event or duplication event [82, 83]. The phylogenomics (molecular phylogenetics) of human four (Rspo)s with other species have been depicted in Figure 12. Here, phylogenetic tree has been developed using sixty proteins ($n = 60$) and it is an interactive tree with the orthologs and paralogs of the seed sequences. From Phylogenetic tree, it is very clear that R-spondin family is only distributed among vertebrate species. Our interactive tree shows the origin and evolution of R-spondin among vertebrate family members and it illustrated that non-vertebrate members (*Drosophila* and *C. Elegans*) are not having domains similar to that of R-spondins. We also specified the tree legend containing different color codes of the different tree nodes. In front of the figure, the domain and sequence panel have been illustrated showing PFAM motifs. The motifs are represented by different shapes. Inter-domain coding regions have been demonstrated as the standard amino acid colour codes and the gap regions are pointed up as a flat line.

Our tree not only describes the phylogenomics of the R-spondin family but also offers an ideal framework topology based on the biological knowledge of R-spondin family and other related sequences. Our result shows the state-of-the-art evolutionary patterns of R-spondin family and the related gene families.

3.11. Understanding the Protein-Protein Interaction Network of R-Spondin Protein Family. Complete knowledge about the protein-protein interaction networks offers direct and indirect interactions between proteins in a cell, helping us to depict a comprehensive description of cellular mechanisms and functions [84, 85].

The protein-protein network of R-spondin protein family is illustrated in Figure 13. The input file for the development of protein-protein network of (Rspo)s has been noted in the Supplementary Figure S4. Four different developed protein-protein interaction networks exemplified the different interactive proteins with the four members of R-spondin family. Rspo1 shows interaction network between FURIN, DKK1, ZNRF3, LRP6, FZR8, SRY, FOXL2, SOX9, MYF5, and STRA8 (Figure 13(a)). The interaction network is more condensed among DKK1, LRP6 and FZR8. Rspo2 shows interaction network between SP8, KRT71, FGF5,

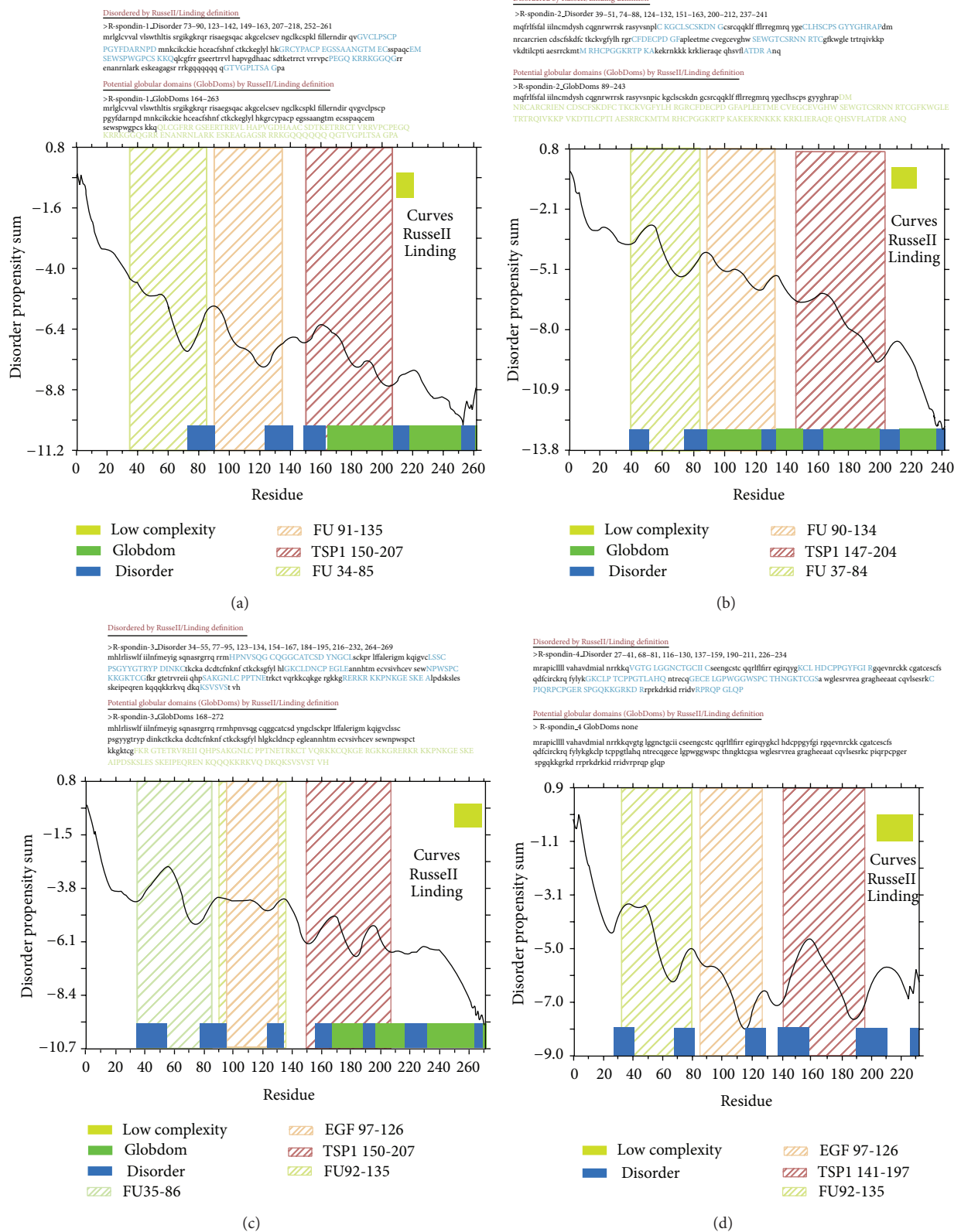
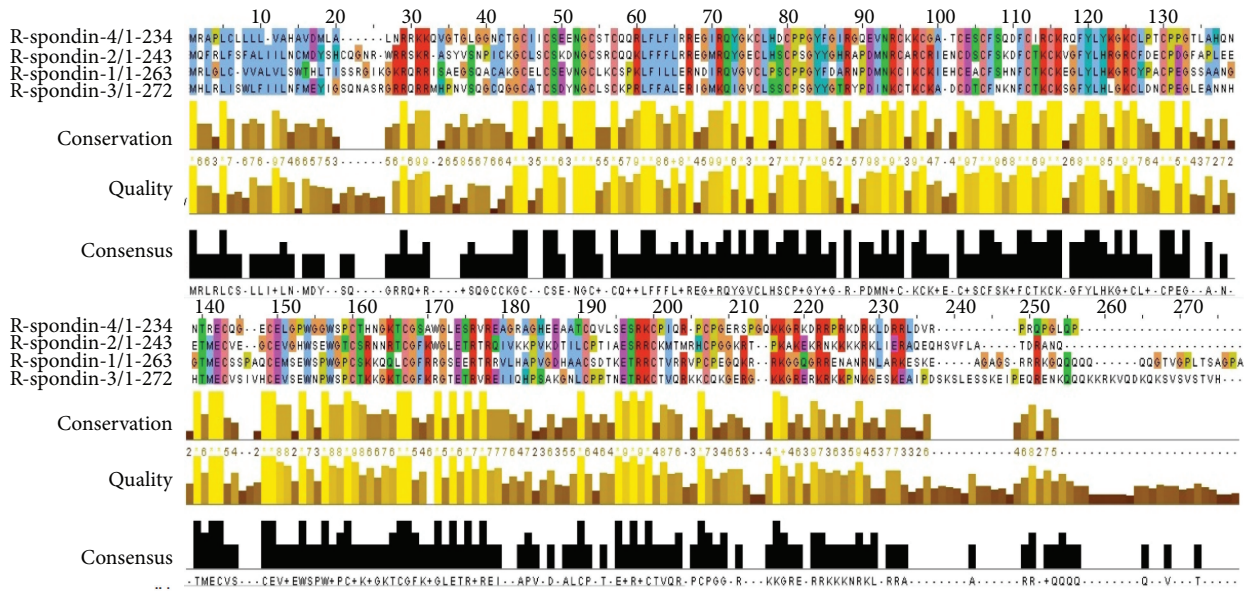
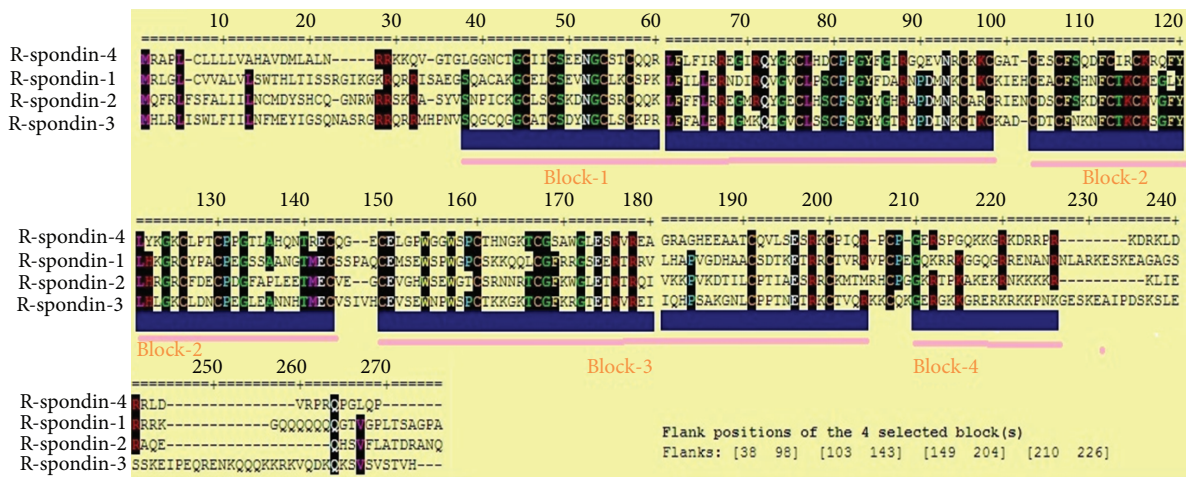


FIGURE 8: Globular domain gain/loss as a function of the variation between the four human (Rspo)s. The disorder propensity of the protein stretch was calculated using GlobPlot analyses to identify the disorder region (blue). The upper portion in the figure illustrates the differences between the amino acid sequence alignments among the domains. The tool uses a simple peak-finder algorithm to select the putative globular and disorder segments. (a) Rspo1, (b) Rspo2, (c) Rspo3, and (d) Rspo4.



(a)



(b)

FIGURE 9: Multiple sequence alignment (MSA) of the different human (Rspo)s. (a) MSA output visualised through JalView and (b) the Gblocks results of human (Rspo)s show blocks from the alignments. The results show highly aligned four blocks.

GORAB, PTPRK, KIAA1804, PDIK1L, GUCY2F, MYLK2, and WNT3A (Figure 13(b)). In this network, no condensed part was found. Rspo3 shows interaction network between FZD8, SDC4, MYF5, FURIN, FAM70A, WNT1, LRP6, KREMEN2, DVLI, and CTNBN1 (Figure 13(c)). The interaction network is more condensed among the proteins which are located in the upper portion of the network such as FZD8, SDC4, WNT1, LRP6, KREMEN2, DVLI, and CTNBN1. Rspo4 shows network between only one protein that is, FURIN (Figure 13(d)) and it is the shortest network among R-spondin protein family.

4. Discussion

R-spondin protein family is an immensely important protein family, which acts as a key regulator factor during vertebrate

development and several signalling pathways, especially as agonists for the canonical Wnt/ β -catenin signalling pathway [17]. Association with different diseases has been found with R-spondin family proteins. (Rspo)s are associated with various developmental stages as an essential regulator. For example, Rspo1 has been found to be associated with sex determination and skin differentiation [22]; Rspo2 is a crucial protein for development of limbs; lungs and hair follicles [11, 27, 86]; Rspo3 is essential for placental development [10] and Rspo4 is a significant protein for nail deployment [17]. (Rspo)s have therapeutic potential for various diseases such as skeletal diseases [87], inflammatory bowel disease and chemotherapy-induced mucositis [23], cancer [21], and diabetes [88]. Therefore, basic understanding about the biophysical, biochemical properties of (Rspo)s may provide more understanding about their functional mechanism associated

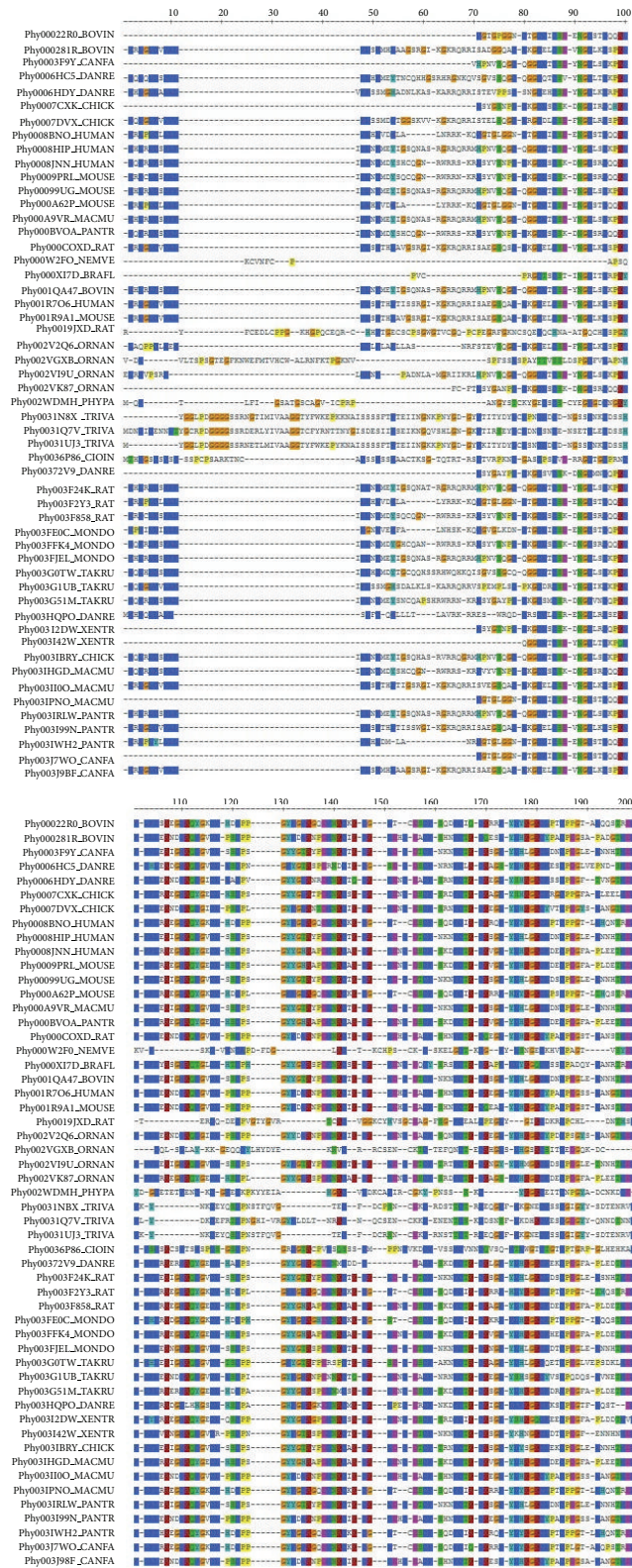


FIGURE 10: Multiple sequence alignment (MSA) of the different human (RspO)s with other species ($n = 53$) which are having sequence similarity.

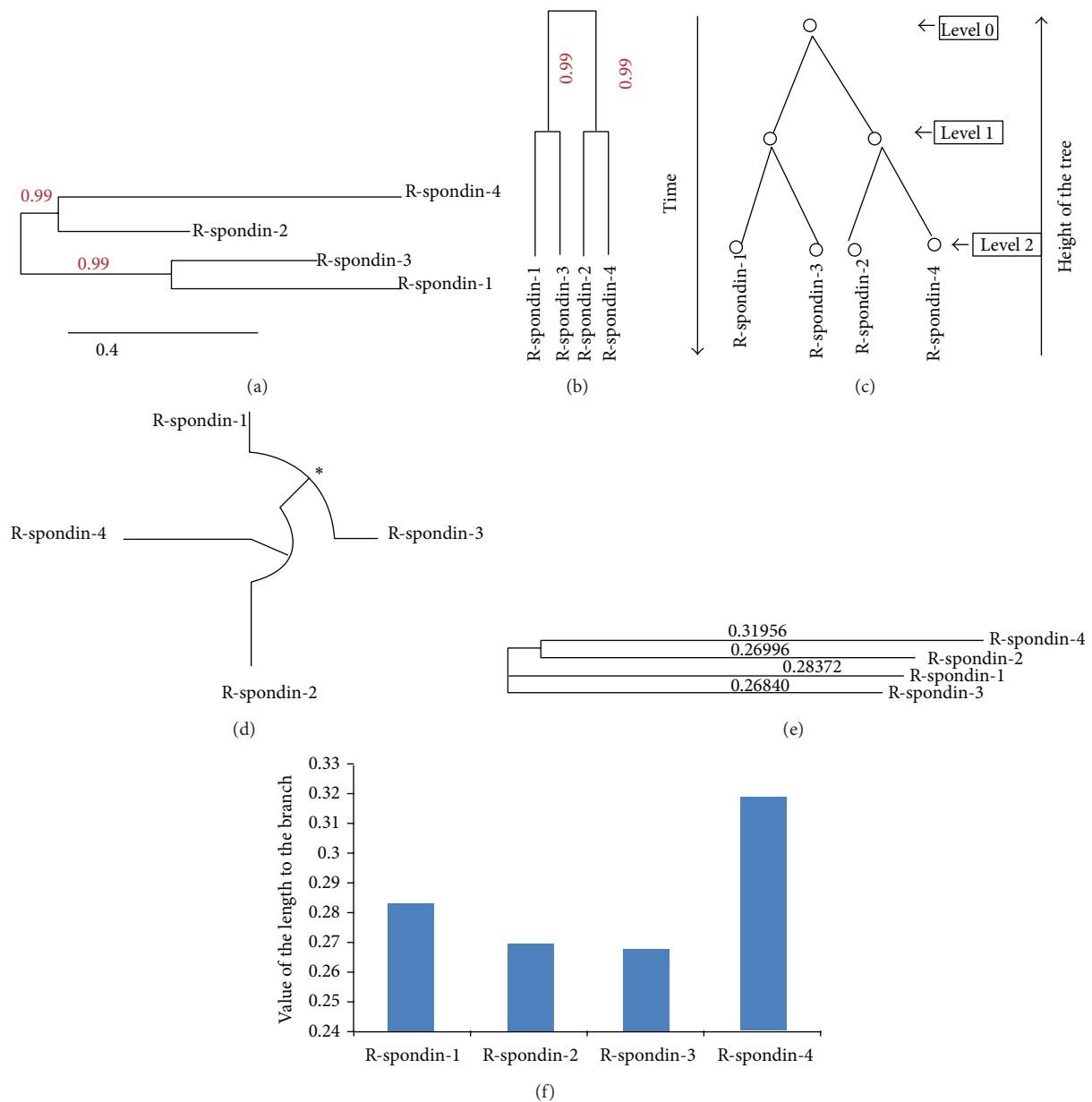


FIGURE 11: Phylogenetic analysis using different human (Rspo)s which shows the relationship between the family members. (a) A phylogenetic tree showing the evolution in the relationships between human (Rspo)s. Bootstrap values are pointed out at nodes. R-spondins protein family members names at the clade, (b) cladogram of protein sequences of the (Rspo)s members, (c) binary tree representation equivalent to cladogram, (d) phylogenetic tree (alpha circular type) reconstructed using the maximum likelihood method. The * symbol adjacent to the node indicates the origin point of the proteins, (e) phylogenetic tree developed through clustal-omega server where the value of the branch length is mentioned immediately leading to the node, and (f) comparison between the value of the branch lengths.

with the diseases and the developmental processes. In this work, to decipher more about the biophysical, biochemical and evolutionary relationship of the R-spondin family, we carried out biophysical, biochemical, and evolutionary based computational mapping of human (Rspo)s.

In general, proteins have a small signal peptide sequence which helps them to enter into the secretory pathway. The N-terminal signal peptide sequence direct proteins to the membrane of the endoplasmic reticulum (ER) and initiate translocation into the ER lumen [89]. From our database

analysis, we identified sequences similarity for signal peptides within (Rspo)s (20 to 21 sequence) (Figure 2) which corroborated to the finding of Kim et al. [2]. In addition to known findings, herein we analyzed signal peptides of human (Rspo)s in the more detail way along with their C-score, S-score and Y-score (Figure 3). Computational methods for estimating N-terminal signal peptides have been detected previously. But, our used server is an advanced tool which uses HMM-based better neural network scheme [39]. Using this tool, we have illustrated the predicted cleavage site value

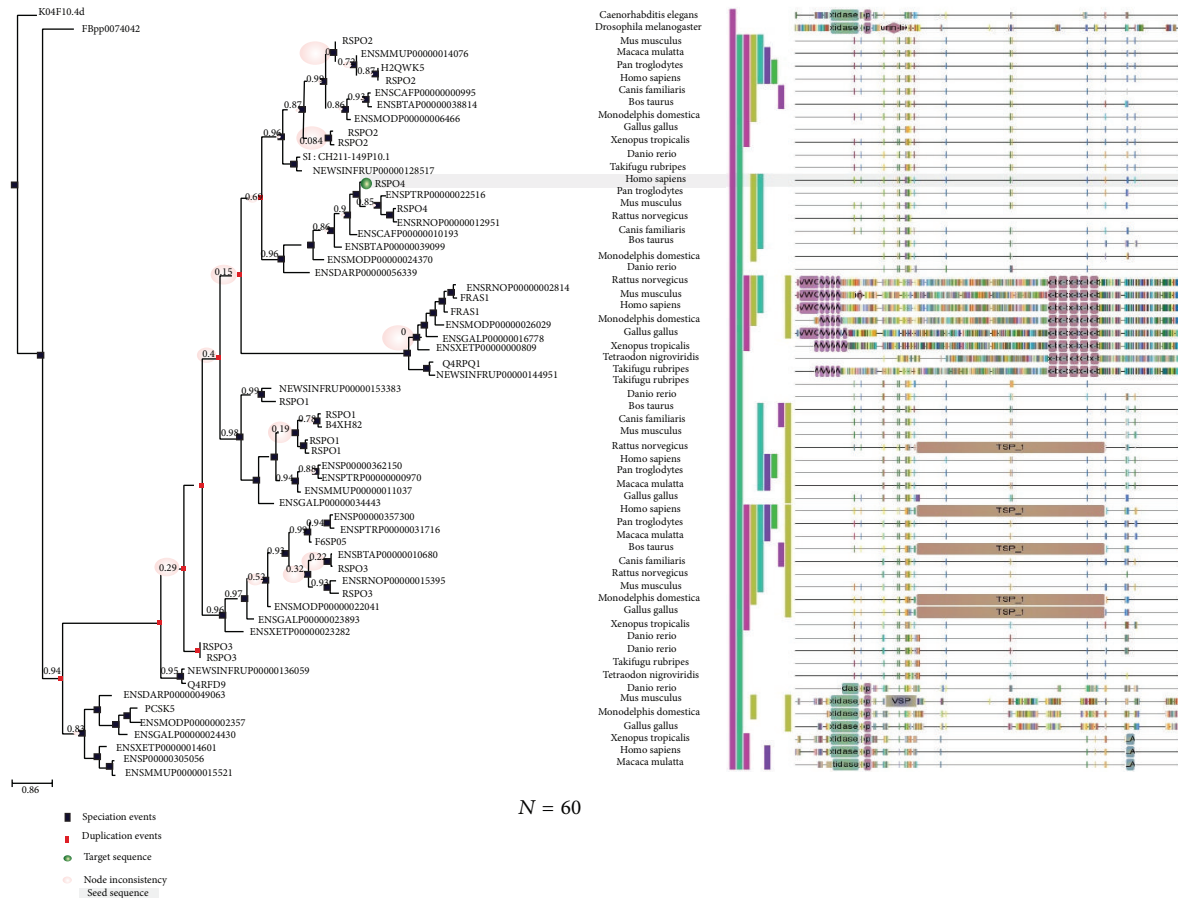


FIGURE 12: Phylogenomics of human four (Rspos) and similar proteins from other species ($n = 60$) which are having sequence similarity. In front of the figure, the domain and sequence panel have been depicted which uses PFAM motifs, and the motifs are represented by different shapes.

(C-score) in the signal peptide of human (Rspo)s where possible two signals are noted in a single signal peptidase cleavage site (Rspo2 and Rspo3) (Figure 3). Hiss and Schneider [89] revealed that long signal peptides mingle two or more signals of signal peptidase cleavage site.

From the amino acids distributed pattern of human (Rspo)s especially from the exposed distribution analysis at a time (Figure 4(e)), we observed the similarity of the amino acids distributed pattern is more or less same. However, Rspo1, Rspo2 and Rspo4 showed more similarity in the distribution pattern. At the same time, our analysis revealed identical amino acid composition pattern in the Rspo1, Rspo2 and Rspo4 (Figures 4(a), 4(b), 4(c) and 4(d)). Recently, it was reported that there is an association between amino acid composition and distribution with mutation. Researchers have shown the correlation between the amino acids distribution pattern; missense mutations and genetic disorders [90]. Conversely, amino acid composition was linked with the deleterious impact of mutations [91]. Therefore, amino acid composition and distributed pattern of human (Rspo)s may help to the future researcher to understand the impact and association with genetic disorders. Further analysis with Cys residues revealed that all these four (Rspo)s are Cys rich

protein. Also, the Cys architecture and the disulphide bond pattern show a common architecture and may be necessary for the stability of these proteins (Figure 4(f)). Recent *in vitro* study with mass spectrometry documented the pattern of disulfide bonds between the 15 available Cys residues present in furin domains in (Rspo)s [14]. However, they found five free cysteine residues in Rspo2.

Our analysis found some glycosylation sites for (Rspo) which may be necessary for their functionality and signalling process (Figure 6). Previously, Kamata et al. [1] has indicated the N-linked glycosylation sites for (Rspo)s. Our previous similar kind of computational analysis shows that the N-glycosylation sites and O-glycosylation sites are vital for the functionality of the proteins in the insulin signalling pathway proteins such as IRS and GLUT4 [67, 92]. However, identified O- and N-glycosylation sites by our analysis with (Rspo)s needs to be confirmed with molecular and biochemical experiments.

Previously, Kim et al. [2] and Nam et al. [4] performed multiple sequence analysis with four (Rspo)s. We also performed MSA among four (Rspo)s as well as with several other species proteins using different computational server (Figure 10). Compared to the previous analysis, our MSA

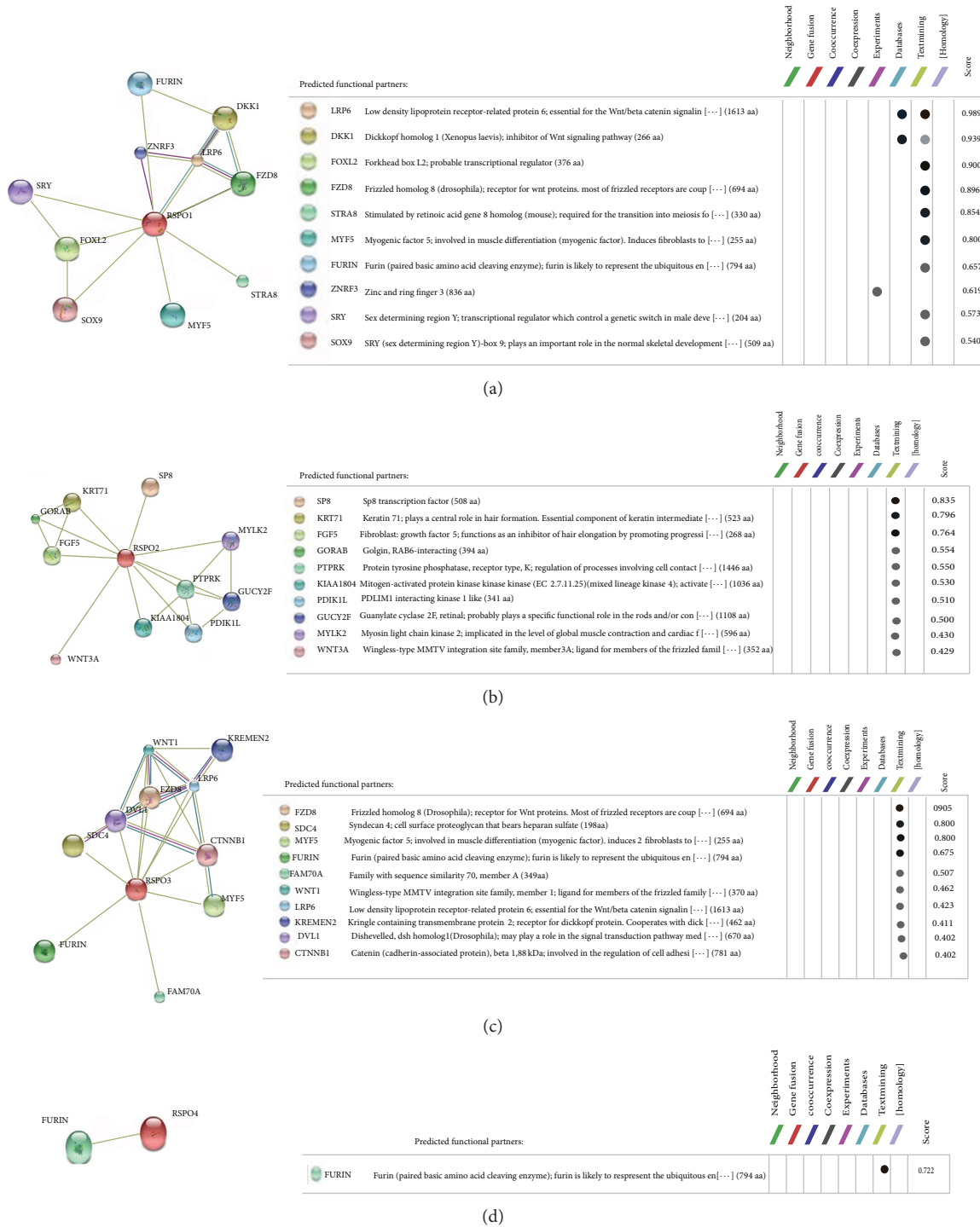


FIGURE 13: Protein-protein interaction network of R-spondin family proteins using STRING server. (a) Rspo1, (b) Rspo2, (c) Rspo3, and (d) Rspo4.

investigation provides a very clear picture about the aligned and conserved residues with different colour codes visualised through JalView. We then analysed through Gblocks server to understand the conserve blocks within the R-spondin family. Our data showed four highly conserved blocks within depicted Gblocks (Figure 9). Furthermore, another MSA analysis of (Rspo)s with other species ($n = 53$) was performed to understand more conserved residues among different

species where we found several small conserved blocks and residues such as glycine, cysteine, valine, serine, proline, histidine, leucine and tyrosine (Figure 10).

The evolutionary history of R-spondin family and the phylogenetic relationships prototype can be investigated through the molecular approach involving amino acid sequencing. Utilizing similar approach, we developed phylogenetic relationships among the members of the R-spondin

family, and we found that Rspo4 and Rspo2 were siblings in 99% bootstrap replications and likewise, Rspo3 and Rspo1 were siblings in 99% bootstrap replications (Figure 11). Previously, de lau et al. [17] and our group also [4] analysed phylogenetic relationships. Here, we performed more advanced two types of phylogenetic analyses: (i) phylogenetic relationships pattern of R-spondin family (Figure 11) and (ii) phylogenetic relationships using R-spondin family using sixty species ($n = 60$) (Figure 12). Second one is the interactive tree with the orthologs and paralogs of the seed sequences which describe the phylogenomics of the R-spondin family and also determines evolutionary relationship of different species (Figure 12). This analysis directs the study towards next generation phylogenomics [93] which may be robust and alignment-free.

From our protein-protein interaction network analysis, we noted an interaction among the Rspo1 with the LRP6 and FZR8 receptor confirming them as candidate protein for Wnt signaling pathway (Figure 13(a)). Hao et al. [94] reported that LRP6 and FZD receptors are present on the membrane and these receptors permit the Wnt ligands to generate much stronger signals. The network of ZNRF3 with Rspo1 confirms that ZNRF3 is associated with Wnt receptor yield in an R-spondin sensitive manner [94]. The network also shows Rspo1 interaction with DKK1 (an antagonist of Wnt signaling). Binnerts et al. [9] reported that Rspo1 binds to the Kremen family of transmembrane proteins and it negatively regulates the LRP6 receptor through the DKK1-associated endocytosis. Due to the controlling property of individual's sex phenotype, Rspo1 networks with SRY and SOX9 protein [95]. The network of Rspo2 with FGF shows that damage Wnt signal directs to defective expression of the important apical ectodermal ridge maintenance factors, FGF4 and FGF8, which is related with the lung and limb development (Figure 13(b)). Similar to Rspo1, we observed a strong association between Rspo3 and the LRP6/FZR8 receptor as well as DVL for Wnt signaling pathway (Figure 13(c)). Rspo4 shows an interaction between FURIN proteins. It has been known that FURIN like domain is necessary for the activity of Rspo4. Blaydon et al. [96] demonstrated that mutations interrupting furin-like domains in Rspo4 may affect its signaling activity. Recent studies showed that (Rspo)s are the ligands for the leucine-rich repeat containing G protein-coupled receptor 4/5/6 (LGR4/5/6) receptors [15–18]. However, in our analysis we have not found any network between the (Rspo)s with LGR4/5/6. This might be due to the lack of updated data in server database (STRING database) containing information about the LGR4/5/6.

In summary, through computational analysis, we performed biophysical, biochemical, and evolutionary topology of human R-spondin family proteins. In this work, we have applied innovative and rapid approach to study the structural based biophysical, biochemical, and evolutionary relationship among (Rspo)s. The difficult and time-consuming nature of the experimental analysis led us to attempt to develop a cost-effective computational research of biophysical, biochemical and evolutionary topology of human R-spondin family. In this study, we have tried to highlight the possible potent sites for O- and N-glycosylation, distribution and

conservation of amino acids and to predict phylogenetic and protein-protein interaction among (Rspo)s with the available data base. However, experimental biochemical and functional studies are required to further establish these finding. Our attempt to decipher the biophysical and biochemical properties of (Rspo)s may provide useful platform and a starting point for scientists to unfold significant physiological and therapeutic properties of R-spondin protein family in various disease models.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Authors' Contribution

Ashish Ranjan Sharma, Chiranjib Chakraborty, and Sang-Soo Lee contributed equally to this work.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2014R1A1A4A03009388 and 2011-001-4792) and by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (HI12C1265). The authors also take this opportunity to thank the management of VIT and Galgotias University for providing the facilities and encouragement to carry out this work.

References

- [1] T. Kamata, K.-I. Katsube, M. Michikawa, M. Yamada, S. Takada, and H. Mizusawa, "R-spondin, a novel gene with thrombospondin type 1 domain, was expressed in the dorsal neural tube and affected in Wnts mutants," *Biochimica et Biophysica Acta—Gene Structure and Expression*, vol. 1676, no. 1, pp. 51–62, 2004.
- [2] K.-A. Kim, J. Zhao, S. Andarmani et al., "R-spondin proteins: a novel link to β -catenin activation," *Cell Cycle*, vol. 5, no. 1, pp. 23–26, 2006.
- [3] K.-A. Kim, M. Wagle, K. Tran et al., "R-Spondin family members regulate the Wnt pathway by a common mechanism," *Molecular Biology of the Cell*, vol. 19, no. 6, pp. 2588–2596, 2008.
- [4] J.-S. Nam, T. J. Turcotte, P. F. Smith, S. Choi, and K. Y. Jeong, "Mouse cristin/R-spondin family proteins are novel ligands for the frizzled 8 and LRP6 receptors and activate β -catenin-dependent gene expression," *The Journal of Biological Chemistry*, vol. 281, no. 19, pp. 13247–13257, 2006.
- [5] Q. Wei, C. Yokota, M. V. Semenov, B. Doble, J. Woodgett, and X. He, "R-spondin1 is a high affinity ligand for LRP6 and induces LRP6 phosphorylation and β -catenin signaling," *The Journal of Biological Chemistry*, vol. 282, no. 21, pp. 15903–15911, 2007.
- [6] O. Kazanskaya, A. Glinka, I. del Barco Barrantes, P. Stannek, C. Niehrs, and W. Wu, "R-Spondin2 is a secreted activator of

- Wnt/ β -catenin signaling and is required for *Xenopus* myogenesis," *Developmental Cell*, vol. 7, no. 4, pp. 525–534, 2004.
- [7] J.-S. Nam, T. J. Turcotte, and J. K. Yoon, "Dynamic expression of R-spondin family genes in mouse development," *Gene Expression Patterns*, vol. 7, no. 3, pp. 306–312, 2007.
- [8] B. A. Parr, M. J. Shea, G. Vassileva, and A. P. McMahon, "Mouse Wnt genes exhibit discrete domains of expression in the early embryonic CNS and limb buds," *Development*, vol. 119, no. 1, pp. 247–261, 1993.
- [9] M. E. Binnerts, K.-A. Kim, J. M. Bright et al., "R-Spondin1 regulates Wnt signaling by inhibiting internalization of LRP6," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 37, pp. 14700–14705, 2007.
- [10] M. Aoki, M. Mieda, T. Ikeda, Y. Hamada, H. Nakamura, and H. Okamoto, "R-spondin3 is required for mouse placental development," *Developmental Biology*, vol. 301, no. 1, pp. 218–226, 2007.
- [11] S. M. Bell, C. M. Schreiner, S. E. Wert, M. L. Mucenski, W. J. Scott, and J. A. Whitsett, "R-spondin 2 is required for normal laryngeal-tracheal, lung and limb morphogenesis," *Development*, vol. 135, no. 6, pp. 1049–1058, 2008.
- [12] B. Ohkawara, A. Glinka, and C. Niehrs, "Rspo3 binds syndecan 4 and induces Wnt/PCP signaling via clathrin-mediated endocytosis to promote morphogenesis," *Developmental Cell*, vol. 20, no. 3, pp. 303–314, 2011.
- [13] K.-A. Kim, M. Kakitani, J. Zhao et al., "Mitogenic influence of human R-spondin1 on the intestinal epithelium," *Science*, vol. 309, no. 5738, pp. 1256–1259, 2005.
- [14] S.-J. Li, T.-Y. Yen, Y. Endo et al., "Loss-of-function point mutations and two-furin domain derivatives provide insights about R-spondin2 structure and function," *Cellular Signalling*, vol. 21, no. 6, pp. 916–925, 2009.
- [15] K. S. Carmon, X. Gong, Q. Lin, A. Thomas, and Q. Liu, "R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/ β -catenin signaling," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 28, pp. 11452–11457, 2011.
- [16] K. S. Carmon, Q. Lin, X. Gong, A. Thomas, and Q. Liu, "LGR5 interacts and cointernalizes with Wnt receptors to modulate Wnt/ β -catenin signaling," *Molecular and Cellular Biology*, vol. 32, no. 11, pp. 2054–2064, 2012.
- [17] W. B. de Lau, B. Snel, and H. C. Clevers, "The R-spondin protein family," *Genome Biology*, vol. 13, no. 3, article 242, 2012.
- [18] A. Glinka, C. Dolde, N. Kirsch et al., "LGR4 and LGR5 are R-spondin receptors mediating Wnt/ β -catenin and Wnt/PCP signalling," *EMBO Reports*, vol. 12, no. 10, pp. 1055–1061, 2011.
- [19] T. Grigoryan, P. Wend, A. Klaus, and W. Birchmeier, "Deciphering the function of canonical Wnt signals in development and disease: conditional loss- and gain-of-function mutations of β -catenin in mice," *Genes and Development*, vol. 22, no. 17, pp. 2308–2341, 2008.
- [20] R. Nusse, "Wnt signaling in disease and in development," *Cell Research*, vol. 15, no. 1, pp. 28–32, 2005.
- [21] Y. R. Jin and J. K. Yoon, "The R-spondin family of proteins: emerging regulators of WNT signaling," *The International Journal of Biochemistry and Cell Biology*, vol. 44, no. 12, pp. 2278–2287, 2012.
- [22] P. Parma, O. Radi, V. Vidal et al., "R-spondin1 is essential in sex determination, skin differentiation and malignancy," *Nature Genetics*, vol. 38, no. 11, pp. 1304–1309, 2006.
- [23] J. Zhao, J. de Vera, S. Narushima et al., "R-spondin1, a novel intestinotrophic mitogen, ameliorates experimental colitis in mice," *Gastroenterology*, vol. 132, no. 4, pp. 1331–1343, 2007.
- [24] G. Krönke, S. Uderhardt, K.-A. Kim et al., "R-spondin 1 protects against inflammatory bone damage during murine arthritis by modulating the Wnt pathway," *Arthritis & Rheumatism*, vol. 62, no. 8, pp. 2303–2312, 2010.
- [25] W. Lu, K.-A. Kim, J. Liu et al., "R-spondin1 synergizes with Wnt3A in inducing osteoblast differentiation and osteoprotegerin expression," *FEBS Letters*, vol. 582, no. 5, pp. 643–650, 2008.
- [26] A. R. Sharma, B. S. Choi, J. M. Park et al., "Rspo 1 promotes osteoblast differentiation via Wnt signaling pathway," *Indian Journal of Biochemistry and Biophysics*, vol. 50, no. 1, pp. 19–25, 2013.
- [27] J.-S. Nam, E. Park, T. J. Turcotte et al., "Mouse R-spondin2 is required for apical ectodermal ridge maintenance in the hindlimb," *Developmental Biology*, vol. 311, no. 1, pp. 124–135, 2007.
- [28] W. Yamada, K. Nagao, K. Horikoshi et al., "Craniofacial malformation in R-spondin2 knockout mice," *Biochemical and Biophysical Research Communications*, vol. 381, no. 3, pp. 453–458, 2009.
- [29] E. Cadieu, M. W. Neff, P. Quignon et al., "Coat variation in the domestic dog is governed by variants in three genes," *Science*, vol. 326, no. 5949, pp. 150–153, 2009.
- [30] S. Seshagiri, E. W. Stawiski, S. Durinck et al., "Recurrent R-spondin fusions in colon cancer," *Nature*, vol. 488, no. 7413, pp. 660–664, 2012.
- [31] S. Neufeld, J. M. Rosin, A. Ambasta et al., "A conditional allele of Rspo3 reveals redundant function of R-spondins during mouse limb development," *Genesis*, vol. 50, no. 10, pp. 741–749, 2012.
- [32] M. Klauzinska, B. Baljinnyam, A. Raafat et al., "Rspo2/Int7 regulates invasiveness and tumorigenic properties of mammary epithelial cells," *Journal of Cellular Physiology*, vol. 227, no. 5, pp. 1960–1971, 2012.
- [33] W. Lowther, K. Wiley, G. H. Smith, and R. Callahan, "A new common integration site, Int7, for the mouse mammary tumor virus in mouse mammary tumors identifies a gene whose product has furin-like and thrombospondin-like sequences," *Journal of Virology*, vol. 79, no. 15, pp. 10093–10096, 2005.
- [34] Y. Ishii, M. Wajid, H. Bazzi et al., "Mutations in R-spondin 4 (RSPO4) underlie inherited onychia," *Journal of Investigative Dermatology*, vol. 128, no. 4, pp. 867–870, 2008.
- [35] E. W. Sayers, T. Barrett, D. A. Benson et al., "Database resources of the national center for biotechnology information," *Nucleic Acids Research*, vol. 39, no. 1, pp. D38–D51, 2011.
- [36] UniProt Consortium, "Update on activities at the Universal Protein Resource (UniProt) in 2013," *Nucleic Acids Research*, vol. 41, pp. D43–D47, 2013.
- [37] UniProt Consortium, "The universal protein resource (UniProt) in 2010," *Nucleic Acids Research*, vol. 38, pp. D142–D148, 2010.
- [38] UniProt Consortium, "Ongoing and future developments at the Universal Protein Resource," *Nucleic Acids Research*, vol. 39, pp. D214–D219, 2011.
- [39] T. N. Petersen, S. Brunak, G. von Heijne, and H. Nielsen, "SignalP 4.0: discriminating signal peptides from transmembrane regions," *Nature Methods*, vol. 8, no. 10, pp. 785–786, 2011.
- [40] N. J. Anthis and G. M. Clore, "Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm," *Protein Science*, vol. 22, no. 6, pp. 851–858, 2013.

- [41] V. Brendel, P. Bucher, I. R. Nourbakhsh, B. E. Blaisdell, and S. Karlin, "Methods and algorithms for statistical analysis of protein sequences," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 6, pp. 2002–2006, 1992.
- [42] J. Cheng, A. Z. Randall, M. J. Sweredoski, and P. Baldi, "SCRATCH: a protein structure and structural feature prediction server," *Nucleic Acids Research*, vol. 33, no. 2, pp. W72–W76, 2005.
- [43] H. M. Berman, J. Westbrook, Z. Feng et al., "The protein data bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.
- [44] R. J. Read, P. D. Adams, W. B. Arendall III et al., "A new generation of crystallographic validation tools for the protein data bank," *Structure*, vol. 19, no. 10, pp. 1395–1412, 2011.
- [45] S. Hunter, P. Jones, A. Mitchell et al., "InterPro in 2011: new developments in the family and domain prediction database," *Nucleic Acids Research*, vol. 40, no. 1, pp. D306–D312, 2012.
- [46] S. E. Hamby and J. D. Hirst, "Prediction of glycosylation sites using random forests," *BMC Bioinformatics*, vol. 9, article 500, 2008.
- [47] K. Julenius, A. Mølgaard, R. Gupta, and S. Brunak, "Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites," *Glycobiology*, vol. 15, no. 2, pp. 153–164, 2005.
- [48] J. E. Hansen, O. Lund, N. Tolstrup, A. A. Gooley, K. L. Williams, and S. Brunak, "NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility," *Glycoconjugate Journal*, vol. 15, no. 2, pp. 115–130, 1998.
- [49] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, and A. Bairoch, "ExPASy: the proteomics server for in-depth protein knowledge and analysis," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3784–3788, 2003.
- [50] R. Linding, R. B. Russell, V. Neduva, and T. J. Gibson, "GlobPlot: exploring protein sequences for globularity and disorder," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3701–3708, 2003.
- [51] F. Sievers, A. Wilm, D. Dineen et al., "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega," *Molecular Systems Biology*, vol. 7, article 539, 2011.
- [52] P. Hogeweg and B. Hesper, "The alignment of sets of sequences and the construction of phyletic trees: an integrated method," *Journal of Molecular Evolution*, vol. 20, no. 2, pp. 175–186, 1984.
- [53] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *Journal of Molecular Biology*, vol. 48, no. 3, pp. 443–453, 1970.
- [54] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Research*, vol. 32, no. 5, pp. 1792–1797, 2004.
- [55] C. Chakraborty, S. S. Roy, M. J. Hsu, and G. Agoramoorthy, "Can computational biology improve the phylogenetic analysis of insulin?" *Computer Methods and Programs in Biomedicine*, vol. 108, no. 2, pp. 860–872, 2012.
- [56] J. Huerta-Cepas, S. Capella-Gutierrez, L. P. Pryszcz et al., "PhylomeDB v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions," *Nucleic Acids Research*, vol. 39, no. 1, pp. D556–D560, 2011.
- [57] J. Huerta-Cepas, A. Bueno, J. Dopazo, and T. Gabaldón, "PhylomeDB: a database for genome-wide collections of gene phylogenies," *Nucleic Acids Research*, vol. 36, no. 1, pp. D491–D496, 2008.
- [58] T. F. Smith and M. S. Waterman, "Identification of common molecular subsequences," *Journal of Molecular Biology*, vol. 147, no. 1, pp. 195–197, 1981.
- [59] A. Dereeper, V. Guignon, G. Blanc et al., "Phylogeny.fr: robust phylogenetic analysis for the non-specialist," *Nucleic Acids Research*, vol. 36, pp. W465–W469, 2008.
- [60] K. Katoh, K. Misawa, K. Kuma, and T. Miyata, "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform," *Nucleic Acids Research*, vol. 30, no. 14, pp. 3059–3066, 2002.
- [61] N. Saitou and M. Nei, "The neighbor-joining method: a new method for reconstructing phylogenetic trees," *Molecular Biology and Evolution*, vol. 4, no. 4, pp. 406–425, 1987.
- [62] L. J. Jensen, M. Kuhn, M. Stark et al., "STRING—a global view on proteins and their functional interactions in 630 organisms," *Nucleic Acids Research*, vol. 37, no. 1, pp. D412–D416, 2009.
- [63] D. Szklarczyk, A. Franceschini, M. Kuhn et al., "The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored," *Nucleic Acids Research*, vol. 39, no. 1, pp. D561–D568, 2011.
- [64] A. Meiniel, R. Meiniel, N. Gonçalves-Mendes, I. Creveaux, R. Didier, and B. Dastugue, "The thrombospondin type 1 repeat (TSR) and neuronal differentiation: roles of SCO- spondin oligopeptides on neuronal cell types and cell lines," *International Review of Cytology*, vol. 230, pp. 1–39, 2003.
- [65] T. L. Rodgers, P. D. Townsend, D. Burnell et al., "Modulation of global low-frequency motions underlies allosteric regulation: demonstration in CRP/FNR family transcription factors," *PLoS Biology*, vol. 11, no. 9, Article ID e1001651, 2013.
- [66] C. G. P. Doss, C. Chakraborty, B. Rajith, and N. Nagasundaram, "In silico discrimination of nsSNPs in hTERT gene by means of local DNA sequence context and regularity," *Journal of Molecular Modeling*, vol. 19, no. 9, pp. 3517–3527, 2013.
- [67] C. Chakraborty, G. Agoramoorthy, and M. J. Hsu, "Exploring the evolutionary relationship of insulin receptor substrate family using computational biology," *PLoS ONE*, vol. 6, no. 2, Article ID e16580, 2011.
- [68] T. Huang, X.-H. Shi, P. Wang et al., "Analysis and prediction of the metabolic stability of proteins based on their sequential features, subcellular locations and interaction networks," *PLoS ONE*, vol. 5, no. 6, Article ID e10972, 2010.
- [69] H. Jubb, A. P. Higuero, A. Winter, and T. L. Blundell, "Structural biology and drug discovery for protein-protein interactions," *Trends in Pharmacological Sciences*, vol. 33, no. 5, pp. 241–248, 2012.
- [70] K. Guruprasad, B. V. B. Reddy, and M. W. Pandit, "Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence," *Protein Engineering*, vol. 4, no. 2, pp. 155–161, 1990.
- [71] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," *Journal of Molecular Biology*, vol. 157, no. 1, pp. 105–132, 1982.
- [72] A. Ikai, "Thermostability and aliphatic index of globular proteins," *The Journal of Biochemistry*, vol. 88, no. 6, pp. 1895–1898, 1980.
- [73] P. Argos, M. G. Rossmann, U. M. Grau, H. Zuber, G. Frank, and J. D. Tratschin, "Thermal stability and protein structure," *Biochemistry*, vol. 18, no. 25, pp. 5698–5703, 1979.

- [74] C. A. Charneski and L. D. Hurst, "Positively charged residues are the major determinants of ribosomal velocity," *PLoS Biology*, vol. 11, no. 3, Article ID e1001508, 2013.
- [75] B. A. Wible, M. Tagliatalata, E. Ficker, and A. M. Brown, "Gating of inwardly rectifying K⁺ channels localized to a single negatively charged residue," *Nature*, vol. 371, no. 6494, pp. 246–249, 1994.
- [76] H. Nakashima and K. Nishikawa, "Discrimination of intracellular and extracellular proteins using amino acid composition and residue-pair frequencies," *Journal of Molecular Biology*, vol. 238, no. 1, pp. 54–61, 1994.
- [77] A. Krogh, B. Larsson, G. von Heijne, and E. L. L. Sonnhammer, "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes," *Journal of Molecular Biology*, vol. 305, no. 3, pp. 567–580, 2001.
- [78] V. N. Uversky, "Intrinsically disordered proteins from A to Z," *The International Journal of Biochemistry and Cell Biology*, vol. 43, no. 8, pp. 1090–1103, 2011.
- [79] L. B. Chemes, J. Glavina, L. G. Alonso, C. Marino-Buslje, G. de Prat-Gay, and I. E. Sánchez, "Sequence evolution of the intrinsically disordered and globular domains of a model viral oncoprotein," *PLoS ONE*, vol. 7, no. 10, Article ID e47661, 2012.
- [80] J. A. Eisen and C. M. Fraser, "Phylogenomics: intersection of evolution and genomics," *Science*, vol. 300, no. 5626, pp. 1706–1707, 2003.
- [81] J. Huerta-Cepas, S. Capella-Gutierrez, L. P. Pryszcz, M. Marcet-Houben, and T. Gabaldon, "PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome," *Nucleic Acids Research*, vol. 42, pp. D897–D902, 2014.
- [82] B. Boeckmann, M. Robinson-rechavi, I. Xenarios, and C. Dessimoz, "Conceptual framework and pilot study to benchmark phylogenomic databases based on reference gene trees," *Briefings in Bioinformatics*, vol. 12, no. 5, pp. 423–435, 2011.
- [83] K. Sjölander, "Phylogenomic inference of protein molecular function: advances and challenges," *Bioinformatics*, vol. 20, no. 2, pp. 170–179, 2004.
- [84] A. Franceschini, D. Szklarczyk, S. Frankild et al., "STRING v9.1: protein-protein interaction networks, with increased coverage and integration," *Nucleic Acids Research*, vol. 41, no. 1, pp. D808–D815, 2013.
- [85] A. Pottekat, S. Becker, K. Spencer et al., "Insulin biosynthetic interaction network component, TMEM24, facilitates insulin reserve pool release," *Cell Reports*, vol. 4, no. 5, pp. 921–930, 2013.
- [86] R. DasGupta and E. Fuchs, "Multiple roles for activated LEF/TCF transcription complexes during hair follicle development and differentiation," *Development*, vol. 126, no. 20, pp. 4557–4568, 1999.
- [87] J. M. Kerkhof, A. G. Uitterlinden, A. M. Valdes et al., "Radiographic osteoarthritis at three joint sites and FRZB, LRP5, and LRP6 polymorphisms in two population-based cohorts," *Osteoarthritis and Cartilage*, vol. 16, no. 10, pp. 1141–1149, 2008.
- [88] V. S. C. Wong, A. Yeung, W. Schultz, and P. L. Brubaker, "R-spondin-1 is a novel β -cell growth factor and insulin secretagogue," *The Journal of Biological Chemistry*, vol. 285, no. 28, pp. 21292–21302, 2010.
- [89] J. A. Hiss and G. Schneider, "Architecture, function and prediction of long signal peptides," *Briefings in Bioinformatics*, vol. 10, no. 5, pp. 569–578, 2009.
- [90] S. Khan and M. Vihinen, "Spectrum of disease-causing mutations in protein secondary structures," *BMC Structural Biology*, vol. 7, article no. 56, 2007.
- [91] S. Hormoz, "Amino acid composition of proteins reduces deleterious impact of mutations," *Scientific Reports*, vol. 3, article 2919, 2013.
- [92] C. Chakraborty, S. Bandyopadhyay, U. Maulik, and G. Agoramoorthy, "Topology mapping of insulin-regulated glucose transporter GLUT4 using computational biology," *Cell Biochemistry and Biophysics*, vol. 67, no. 3, pp. 1261–1274, 2013.
- [93] C. X. Chan and M. A. Ragan, "Next-generation phylogenomics," *Biology Direct*, vol. 8, no. 1, article 3, 2013.
- [94] H.-X. Hao, Y. Xie, Y. Zhang et al., "ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner," *Nature*, vol. 484, no. 7397, pp. 195–200, 2012.
- [95] K. Kashimada and P. Koopman, "Sry: the master switch in mammalian sex determination," *Development*, vol. 137, no. 23, pp. 3921–3930, 2010.
- [96] D. C. Blaydon, Y. Ishii, E. A. O'Toole, H. C. Unsworth, and M. T. Teh, "The gene encoding R-spondin 4 (*RSPO4*), a secreted protein implicated in Wnt signaling, is mutated in inherited anonychia," *Nature Genetics*, vol. 38, no. 11, pp. 1245–1247, 2006.