International Journal of Research Studies in Biosciences (IJRSB)
Volume 3, Issue 9, September 2015, PP 112-125
ISSN 2349-0357 (Print) & ISSN 2349-0365 (Online)
www.arcjournals.org

Sequence and Structure Analysis of Human Selenoprotein 15kDa (Sep15), an Up-Expressed Protein in the Hepatocellular Carcinoma

[1]Stefano Guariniello, [2]Giovanni Colonna, [3*]Eliana Guerriero, [3,4]Francesca Capone, [5]Maria Costantini, [6]Giovanni Di Bernardo, [7]Marina Accardo, [8]Giuseppe Castello, [8,9]Susan Costantini

[1]Dottorato in Biologia Computazionale, Dipartimento di Biochimica, Biofisica e Patologia generale, Seconda Università degli Studi di Napoli, Napoli, Italy
ste_guar@hotmail.it
[2]Servizio di Informatica Medica, Azienda Ospedaliera Universitaria, Seconda Università di Napoli, Napoli, Italy
giovanni.colonna@unina2.it
[3]CROM, Istituto Nazionale Tumori "Fondazione G. Pascale", IRCCS, Napoli, Italy
[*]*eliana8@hotmail.it*
[4]*franci.capone@gmail.com*
[5]Department of Biology and Evolution of *Marine Organisms,* Stazione Zoologica Anton Dohrn, Villa Comunale, Napoli, Italy
maria.costantini@szn.it
[6]Dipartimento di Medicina Sperimentale, Seconda Università degli Studi di Napoli, Napoli, Italy
gianni.dibernardo@unina2.it
[7] Dipartimento di Salute Mentale e Fisica- sez. Anatomia Patologica, Seconda Università degli Studi di Napoli, Napoli, Italy
marina.accardo@unina2.it
[8]CROM, Istituto Nazionale Tumori "Fondazione G. Pascale", IRCCS, Napoli, Italy
[9]*s.costantini@istitutotumori.na.it; susancostantini77@gmail.com*

Abstract: *Sep15 (selenoprotein 15kDa) may function as a regulator of the redox homeostasis. No data are already reported about its involvement in hepatocellular carcinoma (HCC) nor information related to its structural organization.*

Therefore, firstly we evaluated Sep15 expression in two human hepatoma cell lines, HepG2 and Huh7, compared to normal hepatocytes, and in HCC tissues by RT-qPCR analysis, and verified its up-expression in both HCC cell lines and tissues suggesting the necessity of further studies to verify the possibility to use it as an index of endoplasmic reticulum stress level. Then, to have an indirect proof of Sep15 biological importance we have also studied its evolutionary history showing that it is an ancient protein and its N-terminal region is less conserved in comparison to the thioredoxin-like domain. However, a detailed sequence study of the N-terminal region revealed that it is rich of cysteines and of negatively charged residues with an high propensity to intrinsic disorder. Finally, we modeled the three-dimensional structure of the human Sep15 analysing its structural stability by molecular dynamics simulations as well as evaluated the residues that play a key role in structural terms by a residue interaction network analysis (RIN), identifying three HUB residues, Glu74, Phe154 and Leu159. In general the protein is characterized by a disordered fluctuating N-terminal region and by a globularly organized core domain. Therefore, due to its high flexibility the human Sep15 can be adequately represented only through conformational ensembles characterized by fluctuating irregular secondary structures which generate numerous clusters of conformers.

Keywords: *Hepatocellular carcinoma, Sep15, Structure, Molecular dynamics, Interactomic analysis of amino acid residues, conformational ensembles.*

1. INTRODUCTION

Hepatocellular carcinoma (HCC) generally develops from chronic liver diseases [1], including hepatitis B (HBV) or C virus (HCV) infection, exposure to environmental carcinogens (particularly to

the aflatoxin), nonalcoholic fatty liver disease (NAFLD), alcohol-induced liver disease (ALD), primary biliary cirrhosis, type 2 diabetes and obesity [2]. However, it still remains one of the most fatal cancers even if some studies reported advances in its diagnosis and management [3]. In particular, the liver transplantation or the surgical resection or the local ablation represent the only curative modalities for HCC [4, 5]. Unfortunately the percentage of recurrence after the resection is very high (over 70% of patients within 5 y) [6, 7] and this explains because the researchers are always on the search of new prognostic biomarkers.

Some studies have evidenced that the selenium (Se) was able to block the cellular oxidative damage playing an important role in cancer development [8]. Moreover, the selenium has been found incorporated in the selenoproteins as selenocysteine [9]. Recently our group focused on the possible involvement of some selenoproteins in HCC by evaluating the expression of selenium binding protein-1 (SELENBP1) as well as that of Selenoprotein M (SELM) in tissue samples of HCC patients [10, 11] by immuno-histochemistry (IHC). Moreover the expression of the complete seleno-transcriptome has been recently evaluated in two HCC cell lines [12].

As regards the structural features, the only information is that SELM has a motif very similar to that shown by Sep15 (selenoprotein 15kDa). Functionally, Sep15 seems involved in the quality control of the glycoproteins folding [13] and it is also reported that it is highly expressed in organs such as prostate, liver, kidney, testes, and brain [14, 15]. Furthermore, its expression was found altered in some types of cancer, such as lung, breast and prostate [16], and the knockdown of its mRNA led to the inhibition of colony formation, tumor growth, and lung metastasis in a colon cancer cell line [17, 18].

Previous studies have evidenced that Sep15 and SELM are structurally similar to the thioredoxin superfamily, which shows a catalytic domain composed of a two-layer α/β sandwich with a four-stranded β-sheet and a pair of α-helices packed against one side of the β-sheet [19]. However, both proteins have an N-terminal signal peptide that is cleaved, as well a different structural organization of the active-site motif compared to that of thioredoxins [20-22]. In fact, the sequence of Sep15 has CGU as active-site motif, where U is a selenocysteine, whereas SELM has CGGU. The presence of selenocysteine in the catalytic motif is very important because it promotes the function of redox regulator, and thus can be involved into the control of the redox homeostasis [12, 13]. However, as reported above, SELM, as well as Sep15, shows motifs differently organized both compared to thioredoxins (CGPC) that to disulfide isomerases (CGHC) and disulfide oxidases (CPHC) [9, 20, 21]. Aim of this study is to establish the likely involvement of Sep15 in HCC, the evolutionary significance of this protein and also more information on its structure to understand the molecular mechanism at the basis of its function.

2. MATERIALS AND METHODS

2.1. Reverse Transcription-qPCR (RT-qPCR) Analysis on Tissue Samples and Cell Cultures

Paraffin-embedded HCC tissues obtained by biopsy from thirty patients (13 women and 17 men) with ages ranged from 57 to 84 years were subjected to reverse transcription-qPCR. All patients had HCV-related cirrhosis, higher serum transaminase alanine aminotransferase (ALT) and aspartate aminotransferase (AST) levels and AFP (alpha feto protein) higher than 30 ng/mL compared to the healthy individuals. No information related to follow-up data of these patients is known. All patients in this study provided the written informed consent, and the study was approved by the Second University of Naples Ethics Committee.

Human hepatocellular carcinoma cell lines Huh7 and HepG2, and Human Hepatocytes (Lonza, Basel, Switzerland) were kept in culture as already reported in our recent papers [12, 23]. Total RNA from cells and tissues were obtained using the TRizol Reagent (Invitrogen, Milan, Italy) and RecoverAll Total Nucleic Acid Isolation Kit (Life Technologies-Ambion, Carlsbad, CA, USA), respectively. The procedure and the primers used for RT-qPCR analysis were reported in [12]. The 2 x-fold expression level was chosen as the threshold for significance of target genes. Statistical analyses (paired Student's t) were performed using the Prism software (Graphpad Software, La Jolla, CA, USA). Differences in Sep15 gene expression between HCC and healthy samples are considered significant when p-value is lower than 0.05.

2.2. Sequence Analysis

The sequences of Sep15 in human or in other organisms, and of human SELM were obtained from protein databases, Uniprot (www.uniprot.org/) and SCOP (http://scop.mrc-lmb.cam.ac.uk/scop/), and

aligned by using the CLUSTALW program [24]. Secondary structure and disorder propensity predictions were performed by Jpred [25] and Gensilico Meta Disorder Service (http://iimcb.genesilico.pl/metadisorder/) running a wide range of disorder prediction software [26].

The putative phosphorylation, sulfation and glycosylation sites of human Sep 15 were predicted by the neural network based NetPhos Server [27], Sulfinator [28], and NetNGlyc and NetOGlyc servers, respectively [29]. Moreover, we used SLIMPRED [30] servers for the prediction of the short linear motifs, and ANCHOR [31], and α-MoRF-PredII [32] for the prediction of the binding regions in disordered proteins. Finally, we have evaluated the charge distribution of the N-terminal region of human Sep15 according to Das and Pappu [33].

2.3. Modeling of Human Sep15

The human Sep15 sequence was analyzed with BLAST to search similar proteins of known 3D structures [34]. We have evidenced a sequence identity percentage equal to 50.5% between the region (84-162) of the human sequence and the sequence from *Drosophila Melanogaster* (PDB: 2A4H) [19]. Thus, we have aligned the related sequences by CLUSTALW [24], and modeled the human segment using the MODELLER9v11 program by setting at 4.0 Å the root mean square deviation (RMSD) among the structures of the template and fully optimized models [35].

On the other hand, since for the human region (29-83) were missing proteins with an acceptable percentage of sequence identity to be used by comparative modeling as templates, we modeled this segment by Phyre2 program using the "Intensive Method", which attempts to generate a complete full-length model of a sequence through a combination of multiple template modeling and simplified ab initio folding simulation [36].

By means of these procedures, we have created the whole three-dimensional model for human Sep15 by MODELLER9v11 [35] by using as templates for the region (29-83) the model obtained by Phyre server [36] and for the region (84-162) the model obtained from the Sep15 structure from *Drosophila Melanogaster* (PDB: 2A4H) [19] by a procedure of comparative modeling already used by our group [37].

All the models were analyzed in terms of energetic quality by ProSA program [38] and in terms of stereo-chemical quality by the Ramachandran plot [39]. The solvent accessibility of the human Sep15 structure was evaluated by ASAview program [40] while H-bonds, π-cations and π-stacking interactions by the Protein Interactions Calculator (PIC) [41] and COCOMAPS tools [42]. Moreover, to check the similarities between our model of the human Sep15 and the structures already reported in PDB, we used VAST [43-44].

2.4. Molecular Dynamics (MD) Simulation and MD Analysis

The best model of the human Sep15 was subjected to Molecular Dynamics (MD) simulations for 20 ns by GROMACS software package (v3.3.1) [45] using GROMOS43a1 all-atom force field and setting a neutral pH, as previously described [46-47]. In details, the protein was inserted into a cubic box with sides of 88.6 Å and filled with 22519 SPC216 water molecules, and five sodium ions were added to maintain the net charge of the system equal to zero.

We used different GROMACS routines to analyze the trajectories (RMSD, gyration radius, RMSF, secondary structure evolution etc), and to calculate correlated motions by Principal Components Analysis (PCA). Conformations of Sep15 at zero time and after 5, 10, 15 and 20 ns of MD simulation were also studied by RING, a tool [48] that considers them as networks of interacting amino acid pairs (nodes), and focusing on the type of interaction between the different nodes represented as edges in terms of H-bonds, π-cation, π-stacking, salt bridges and interactions with their closest atoms (IAC). In particular, we have also used two Cytoscape plugins, NetworkAnalyzer and RINalyzer, for the standard and advanced analyses of network topologies, and a third plugin, CytoHubba, to explore the interaction network, and to detect the essential nodes (HUB residues) by analyzing the network in terms of seven different statistics, i.e. degree of nodes, betweenness and closeness centrality, bottleneckness, Density of Maximum Neighborhood Component (DMNC), Maximum Neighborhood Component (MNC) and Maximal Clique Centrality (MCC) [49-50].

3. RESULTS AND DISCUSSION

3.1. Analysis of Sep15 Expression by RT-qPCR Analysis

Recently human SELM resulted up-expressed in two HCC cell lines and tissues [11] but until now no data are reported in the literature on a possible involvement of Sep15 in HCC, therefore, we have evaluated its gene expression levels by RT-qPCR in HepG2 and Huh7 cell lines compared to normal hepatocytes. In details, we obtained that i) Sep15 has a statistically significant up-expression in both HCC cell lines compared to control cells, and ii) its expression in HepG2 and Huh7 cells is very similar, as shown by the mRNA fold change equal to about 4 times in both cases. To verify these results also in human tissues, we have evaluated the gene expression of Sep15 also in thirty HCC tissues compared to the normal liver tissue. Sep15 resulted up-expressed in HCC tissues compared to normal tissues (with mRNA fold change equal to 2.9) as well in HepG2 and Huh7 cells. In the liver we find that Sep15 is involved in a complex with another endoplasmic reticulum (ER)-resident protein, the UDP-glucose:glycoprotein glucosyltransferase (UGTR), the function of which is the glucosylation of the misfolded proteins, which are retained in the ER until correctly folded or transferred to degradation pathways. The association between Sep15 and UGTR is reported to play an important role for maintaining this selenoprotein in the ER [13]. In general, the ER stress represents the biochemical basis of many patho-physiological conditions, including cancer, with the formation of toxic protein aggregates, tumor cell anti-apoptosis and drug resistance [51]. Thus, these effects might justify the involvement of this protein in the pathophysiology of HCC, contributing to dysfunction and death of hepatocytes [52]. Moreover, since the ER stress is known to up-regulate the expression of Sep15, this could be a necessary requirement for the protection of cells from stress [14, 53]. On this basis, we suggest that the up-expression of Sep15 could be usefully used as an index of ER stress level in the presence of HCC.

3.2. Multiple Alignment of Sep15 in Different Organisms

Considering the functional importance of Sep15, we wanted to ascertain its real biological importance, going to investigate how much this protein was evolutionarily ancient. Therefore, all the known Sep15 sequences from different organisms were aligned by CLUSTALW. We found 131 sequences, including: 57 mammals, 11 birds, 7 reptiles, 1 sauropside, 2 amphibians, 17 Osteichthyes, 1 Chondrichthyes, 6 simple eukaryotes, 8 plants and 21 arthropods. As one can see, the protein is widely distributed in the living world and thus it can be considered very ancient.

We performed also a detailed analysis of the sequence identity percentages between human Sep15 and other organisms, and the data showed that human Sep15 had the highest sequence identity percentages within mammals, birds, and reptiles whereas the lowest percentage is in the plants, which represent the more ancient species (Table 1).

Table1. *We report the mean identity sequence percentages (%) with the related standard deviation (ST.DEV.) values between Sep15 in human and in different phyla.*

Phyla	**% ± ST.DEV.**
Plants	37.31 ± 0.97
Insecta	48.31 ± 3.26
Simple Eukaryotes	49.62 ± 8.42
Amphibious	80.97 ± 0.52
Fishes	85.71 ± 2.98
Birds	90.65 ± 0.56
Reptiles	91.13 ± 1.68
Mammals	97.40 ± 1.79

About the conservation of human Sep15 sequence compared to other organisms, we found that the percentage of residue conservation in the regions 29-83 (N-terminal segment) and 84-162 (thioredoxin-like segment) was of 74% and 85%, respectively. Moreover, the great part of the Sep15 sequences shows a well conserved CGU motif (where U is the selenocysteine), which is defined as the active-site motif. The only exceptions were found in the *Arthropod phylum*, in *planta* kingdom, and in *nematodes*, where we observe CTC, CMR and CEC, respectively, as motifs, whereas the CGU motif is completely lost in the reptile *Ophiophagus Hannah*. However, we have found that even the region around the CGU motif is highly conserved in all the examined organisms except for two nematodes (*Brugia malayi* and *Ascaris suum*) where the percentage of conservation of residues in the motif and in the surrounding region was equal to 83% and 88%, respectively.

In overall these data showed that: i) Sep15 is a very ancient protein because it is present from plants to humans, ii) the thioredoxin-like segment in human Sep15 is more conserved compared to the N-terminal, and iii) the CGU motif is evolutionary well conserved. From these observations we conclude that Sep15 is widely present in the living organisms because has to share an important functional role, even if in humans this role still is not clearly connected to the HCC.

3.3. Multiple Alignment of Human Sep15 and SELM

Since human Sep15 is homologous to human SELM, we have aligned their sequences to evidence similarities and differences. Both sequences contain a thioredoxin-like domain with a total sequence identity percentage of 24.8%. This value increases to 28.6% if we concern only their thioredoxin-like domains. Moreover, they show two different active site motifs, i.e., CGU in Sep15 and CGGU in SELM. On the other hand, SELM presents a short N-terminal segment and a longer C-terminal region, conversely Sep15 a longer N-terminal region but a shorter C-terminal segment. However, the shortness of the N-terminal region in SELM is highly conserved also in many different organisms as reported from Ferguson et al. (2009) [19].

3.4. Sequence Analysis on N-terminal Region of Human Sep15

Since the N-terminal region of human Sep15 (from residue 29 to 83) has been found scarcely conserved in the multiple alignment compared to thioredoxin-like domain, we focused our attention on its physico-chemical properties. The most remarkable features are six cysteines, nine negatively and four positively charged residues, with a net total charge of the protein at neutral pH of -5.3, and some residues classified as intrinsically disordered (**Fig. 1A**). Moreover, we have also evaluated the putative presence of phosphorylation, sulfation, glycosylation sites and of linear motifs for protein interaction because the presence of even small functional regions may play important functional roles as a consequence of post-translational modifications [27-32]. We found only a single putative phosphorylation site, whereas no sites of sulfation and glycosylation are predictable as well as no binding region has been detected by Anchor [31] and α-MoRF-PredII [32]. Moreover, the SliMPred algorithm, using a neural network approach, has found three residues involved in short linear motifs [30].

A

```
                    29                                                  84
                     |                                                   |
                     FGAEFSSEACRELGFSSNLLCSSCDLLGQFNLLQLDPDCRGCCQEEAQFETKKLY
metadisorder         DDDDDDDD------------------------------------DDDDDDDD
Netphos              ***************S************************************
SLiMPred             ***********MM*****M************************************
```

B

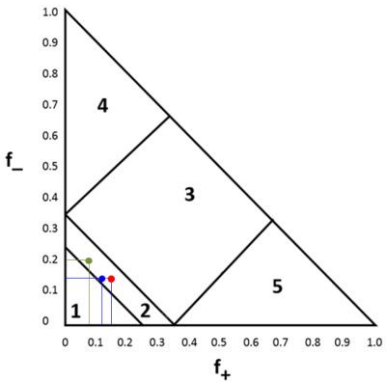

Figure1. *A) We report for the N-terminal region (1-55 amino acids) of Sep15 the positively and negatively charged residues in blue and red, respectively. Moreover, we show i) the propensity to disorder by Gensilico Meta Disorder Service [26] running a wide range of disorder predictions and indicating with "D" the disorder propensity, and with "-" the coil propensity, ii) the putative phosphorylation sites (S) by NetPhos Server [27], and iii) the prediction of the short linear motifs (M) by SLIMPRED servers [30] B) The Diagram of State for IDPs evaluated according to Das and Pappu [33] for the complete sequence of Sep15 (in blu), and the regions 29-83 (in green) and 84-162 (in red).*

Wanting to translate in an overall structural view all the peculiarities found in the sequence of the N-terminal region, we have estimated the net charge per residue according to Das and Pappu (2013) **[33]** computing the fractions of positively and negatively charged residues in the sequence (f+ = 0.07, f− = 0.19, FCR=0.26 and NCPR=0.12). These values characterize the Sep15 N-terminus as polyelectrolyte with a charge distribution value which places it in the border region 2 of their Diagram of State for IDPs **[**33**]** (**Fig. 1B**), that is, between the weak polyampholytes and polyelectrolytes, with propensity for ensembles of globule and tadpole and the strong Polyampholytes (or Polyelectrolytes) with propensity for ensemble of Coils, Hairpins, and Chimeras. This also means that the addition of any other negative charge due to post-translational modifications, will alter the conformational state of Sep15 leading to migration to different structural regions, more appropriate for the biological function to be performed. We have to note that the term globule does not mean "globular" but a compacted disordered state of the structure as consequence of the poor interaction of residues with the solvent but with a predominance of inter-residue interactions, mainly of MM-MM type. However all these data are interesting because the N-terminal region of Sep15 is reported in literature to be important in the binding with UGTR and its deletion or the mutation of its cysteines prevents the formation of the complex [53]. In fact, this region was suggested as crucial for the protein-protein interaction with UGTR [53]. Of course the interaction lies on the three-dimensional structure and this was a good reason for the protein modeling.

3.5. Sep 15 Molecular Modelling

The 3D model of the human Sep15 (Uniprot code O60613, residues 29–162) was built by means of a combined approach of *ab initio* and comparative modeling as already used in our recent papers [37, 54, 55]. In details, we have chosen as template for the human region 84-162, the structure of Sep15 from *Drosophila Melanogaster* (PDB: 2A4H) [19] because the sequence identity percentage between the human sequence and that of the template resulted equal to 50.5%. Before we aligned these regions of human and D. Melanogaster Sep15 sequences, as shown in **Fig. 2,** and, then, we modeled the structure by comparative modeling using MODELLER9v11 [35]. In details, starting from this alignment, a set of 10 all-atoms models was generated and the best model was selected in terms of its energetic and stereo-chemical quality. It showed 89.53% of residues in the most favored regions of the Ramachandran Plot [39] and a ProSA Z-score of −3.61 [38].

Figure2. *Alignment of Sep15 sequences in Drosophila Melanogaster (PDB: 2A4H) and in human wild type (indicated with Sep15 wt) and in the human mutants. Amino acids in the β-strands and helices are evidenced in yellow and red, respectively. Amino acids predicted in α-helices, β-strands and coil by Jpred program [25] are indicated with H (red), E (light blue) and "-", respectively.*

These values, compared with those of the template structure that had a ProSA Z-score = −3.09 and 88.50% of residues in the most favored regions, indicated the good quality of our model. The comparison of the secondary structures between human and Drosophila models has evidenced that helices and β-strands were well conserved along the sequences, and only few changes were observed in their length (**Fig. 2**). Further, the RMSD value between the two structures was equal to 0.8 Å. Also the secondary structure predictions, made by Jpred tool [25], agreed with the modeled structure except few differences (**Fig. 2**). However, to check similarities and differences of human Sep15 structure compared to PDB structures, we performed a structure-based multiple sequence alignment by VAST tool [43], and found some structures among which that of SELM from *Mus Musculus*, of human ubiquinone oxidoreductase, and of peroxiredoxin, thioredoxin and glutathione S-transferase from

different organisms with RMSD values lesser than 2.5 Å compared to our human Sep15 structure. These data are interesting because they demonstrated that, even if the sequence identity percentages between these proteins and human Sep15 were lower, nonetheless they conserved an acceptable structural similarity.

Moreover, since the Blast search did not show proteins with high identity percentage for the human region 29-83 of Sep15 to be used as a template, we modeled this segment by the "Intensive Method" of the Phyre server, which performs more steps in the case of disordered proteins [36]. The structure obtained for this region had a ProSA Z-score = −2.41 and 86.37% of residues in the most favored regions of the Ramachandran Plot [38-39].

The whole model of the human Sep15 was obtained by comparative modeling using as templates the models obtained for the region 29-83 by ab-initio methods and for the region 84-162 by comparative modeling. The best model (**Fig. 3A**) showed 87.12% of residues in the most favored regions of the Ramachandran Plot [39] and a ProSA Z-score of −2.67 [38].

In details, the human model shows: i) the N-terminal domain with only two helices of different length, and ii) a single α/β domain in the region 84-162 that is composed by 3 α-helices, 4 antiparallel β-strands and long irregular regions, typical of thioredoxin-like proteins. Further, this structure is characterized by only 42% of residues in regular secondary structure elements (i.e. helices and β-strands) because we have to include also many residues of the N-terminal region (29-83), dispersed in not-allowed regions as shown in the Ramachandran Plot [39].

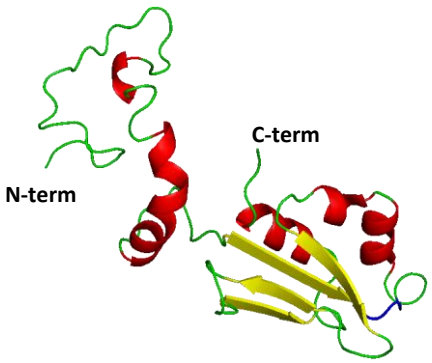

Figure3. *A) The best model of human Sep15. The secondary structures, α-helices and β-strands, are shown with red helices and yellow arrows, respectively, while the loops and bends are colored in green, and the active site in blue.*

These structural features are in excellent agreement with the analysis performed on the sequence (**Fig. 1B**), however, comparing the fractions of charged residues calculated for the entire sequence (f+ = 0.12, f− = 0.15, FCR=0.27 and NCPR=0.03) with those of the region 84-162 (f+ = 0.15, f− = 0.15, FCR = 0.30 and NCPR = 0), we can conclude that both are located in the boundary region 2 of the Diagram of State for IDPs, thus confirming the large propensity of the human Sep15 to have residues in non regular secondary structure elements.

Concerning that the IDPs are known for their high flexibility, we have studied this aspect by MD simulation both to know the intrinsic dynamical features of the N-terminal region and to evaluate the contribute to the protein fluctuation of the long disordered flexible segments of the thioredoxin-like domain.

3.6. Molecular Dynamics on Human Sep15

MD simulation on human Sep15 reached a stable equilibrated state after 7 ns of simulation; in fact, the related RMSD values **(Fig. 4A)**, computed by superposing the various structures, were quite constant during the remaining simulation time (from 7 to 20 ns). The gyration radius trend decreased reaching the value of about 1.6 nm after 7 ns **(Fig. 4B),** suggesting that the protein becomes more compact during the simulation. This was also confirmed by the H-bond plot that evidenced an increase of the these bonds to stabilize the structure **(Fig. 4C).** However, the overall architecture of the protein was quite stable during the simulation time except the N-terminal segment (region 29-83),

as shown by the RMSF plots **(Fig. 4D)** and secondary structure evolution **(Fig. 4E)**. In fact, this region is more fluctuating if compared to the thioredoxin-like domain (region 84-162) where the regular secondary structure elements (three helices and four β-strands, reported in blue and red, respectively) remained almost unchanged as reported by the secondary structure evolution plot (Fig. 4E), while the first short helix of the N-terminal region is very fluctuating.

Concerning the active site, it is located in the disordered region between the β-strand (85-89) and the α-helix (99-106), and it has been found to be fluctuating during the simulation in terms of RMSF values (as shown in **Fig. 4D**). We have also noted that unexpectedly the CGU motif interacts with the residues of another disordered region (120-126). In fact, Gly92 formed one H-bond with Arg122, while the Cys91 showed a tendency to form H-bonds with Gly123, Ser124 and Asp125 during the simulation. This confirms that the disordered regions have an impact on the active site. Moreover, we focused also on the interactions between the N-terminal and thioredoxin-like domains during the MD simulation assessing that: i) a π-cation interaction is formed between Phe77 and Lys158 after 10ns of the simulation time and remaining stable until the end of the simulation, ii) no π-stacking interaction were formed between N-terminal and thioredoxin-like domains during the entire simulation time, and iii) an increasing number of H-bonds was detected between residues present in the two domains during the simulation.

Then, a principal component analysis (PCA) was also performed to have a deeper insight in the collective fluctuations of Sep15 during the MD. In parallel to the PCA, the cluster analysis allowed us to determine groupings of structures that shared similar conformational features on the basis of their RMSD values. Even if the total number of the clusters was equal to 23, the number of those most populated was of only 5. In details, as it can be seen in **Fig. 5**, the N-terminal domain (region 29-83) and the segments in the region 84-162, characterized by a poor content of regular secondary structure elements, have been found highly fluctuating. For this reason, the Sep15 structure has to be considered as conformational ensembles, characterized by fluctuating irregular secondary structures. In fact, Sep15 does not assume stable and fixed in time conformations but is made of fluctuating conformers that change continuously, going from one cluster to another. Because of the high segmental flexibility of Sep15, its representation can be done only by superimposing the most abundant conformers (**Fig. 5**).

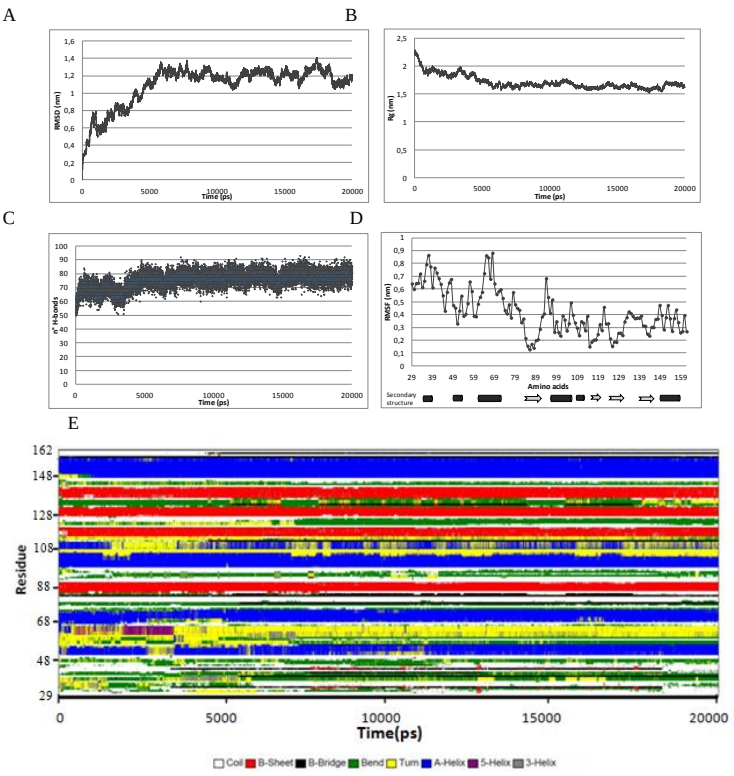

Figure4. *Analysis of molecular dynamics simulations conducted on Sep15 at physiological pH in terms of: (A) root mean square deviation (RMSD) plot, (B) gyration radius plot, (C) H-bonds plot, (D) root mean square fluctuation (RMSF), and (E) secondary structure evolution.*

Figure5. *Superposition of the most populated clusters of Sep15 from the molecular dynamics simulations, where α-helices and β-strands are shown with red helices and yellow arrows, respectively, while the loops and bends are colored in green.*

3.7. RIN (Residue Interaction Network) Analysis

To study the specific role of individual residues in the general organization of a protein, the RIN analysis is resulted very useful by creating an interaction network starting from the protein 3D structure. It permits to identify those residues which have the strongest coordination role, namely HUB residues, and, therefore, it allows to assess the structural stability of the protein. We have performed this analysis on Sep15 structure before MD and after 5, 10, 15 and 20 ns of simulation time (**Fig. 6**). As shown in **Table 2**, the analysis has evidenced that Sep15 was dynamically stabilized by a high number of H-bonds and IAC interactions, to which also π-cation and π-stacking interactions and salt bridges have to be added. Further, we have also calculated seven topological centrality parameters (betweenness, bottleneck, closeness, DMNC, degree, MCC and MNC) to identify the HUB nodes, i.e., the highly connected residues.

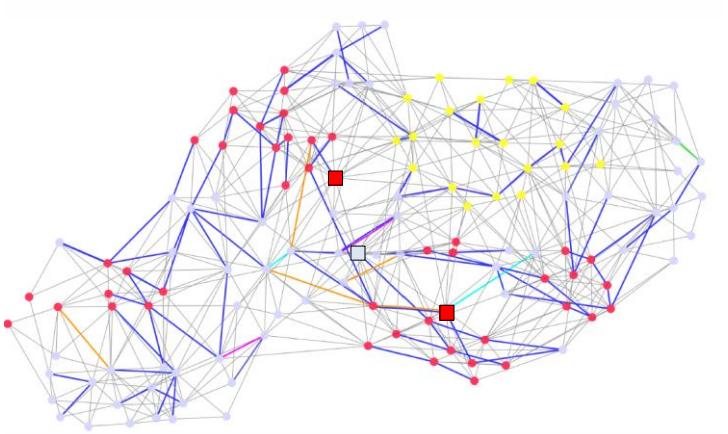

Figure6. *Residue interaction networks and after 20 ns of molecular dynamics simulation. The nodes correspond to Sep15 residues and the edges correspond to binding interactions between residues. The red and yellow nodes are the residues in α-helices and in β-strands, respectively. The colored edges indicate different bonds: grey edges indicate peptide bonds, cyan edges indicate π-stacking interactions, orange edges indicate π-cation interactions, blue edges indicate H-bonds, magenta edges indicate salt-bridges and green edges indicate disulfide-bonds. HUB nodes are evidenced by squares.*

Table2. *Evaluation of H-bonds, π-cation and π-stacking interactions, salt bridges and interactions with their closest atoms (IAC) in Sep15 model before and after 5, 10, 15 and 20 ns of molecular dynamics. In details, we report main chain—main chain (MM), main chain—side chain (MS) and side chain—side chain (SS) H-bonds.*

		H-bonds		**π-cation**	**π-stacking**	**salt bridges**	**IAC**
	MM	**MS**	**SS**				
Sep15							
Frame 0	40	7	4	4	2	0	422
Frame 5	56	12	9	1	3	1	493
Frame 10	56	16	6	3	3	0	504
Frame 15	54	17	10	6	4	2	486
Frame 20	56	23	4	5	1	0	498

Table3. *HUB nodes evaluated in the residue interaction network (RIN) performed on Sep15 structure before and after 5, 10, 15 and 20 ns of MD simulation concerning a consensus of five out to seven topological centrality parameters.*

0 ns	**5 ns**	**10 ns**	**15 ns**	**20 ns**
Pro 126	Tyr 83	Glu 74	Glu 74	Glu 74
Phe 154		Phe 154	Phe 154	Phe 154
Leu 159			Leu 155	Leu 159
			Leu 159	

Considering a consensus of five out of seven measures of centrality, we found that three residues Glu74, Phe154 and Leu159 resulted conserved during the simulation (**Table 3**), and for this reason we have considered them as HUB nodes. A detailed analysis of the connectivity has evidenced that during the MD simulation: i) Leu159 shows IAC interaction with both Glu74 and Phe154, and ii) Glu74 and Leu159 interacted with the rest of Sep15 only by H-bonds whereas Phe 154 formed some MM-type H-bonds, one π-cation, and one π-stacking interaction). However, all three HUB residues were found shielded from the solvent, at beginning of the MD, becoming more buried during the simulation**.** The large number of interactions, and the inside position suggest that these residues have a key role in the whole structure of the protein. From the multiple alignment, we have also verified that these structurally critical residues are almost conserved in the different organisms. In fact, we found: i) Glu or Asp always in the position 74 except for the *Anphimenodon* (porifera) and the *diaphorina* (insect), supporting the importance of a negative charge in this position ii) Phe is always in the position 154 except for the *Medicago* and the *Cicer* (plants), and iii) Leu always in the position 159 except for the *plantae* and the *Ascaris* (nematode).

In conclusion these analysis confirm the critical significance of these residues for the stability and structural organization of Sep15 that deserves further studies in solution.

3.8. Comparison between Human Sep15 and SELM Structures

Since human Sep15 and SELM have a common core, belonging to the thioredoxin-like family, we have compared their structures and the location of their HUB residues. The two proteins have very similar thioredoxin cores with a RMSD value of 1.8 Å while the N-terminal regions are very different because human Sep15 shows a longer tail (**Fig. 7**).

Concerning the HUB residues, we found that those of SELM, i.e., Phe59, Leu82 and Leu84 correspond to Phe104, Tyr120 and Leu130, respectively, in Sep 15, where we also find two HUB residues (Phe154 and Leu159) corresponding to residues Leu108 and Gly113 of SELM. However, it is important to underline that the third HUB residue in Sep15, i.e., Glu74, did not show any structural correspondence in SELM because it was not in the common thioredoxin-like core but in the N-terminal region of Sep15.

However, to have more information on the lack of correspondence between the HUB positions of SELM and Sep 15, we focused on the values of the seven topological centrality parameters evaluated from the RINs for the three residues, Phe104, Tyr120 and Leu130 of Sep 15. In this way we have verified that both Phe104 and Leu130 were deeply buried in the structures even if they showed suitable high values only for three or four out of seven values related to topological centrality parameters. On the other hand, Tyr120, showing low values for these parameters**,** has to be considered less shielded in terms of solvent accessibility.

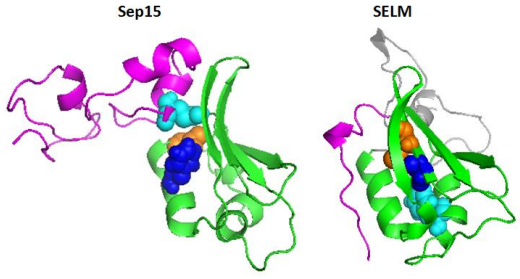

Figure7. *Models of human Sep15 and SELM after 20 ns of molecular dynamics simulation. For Sep15 we report the N-terminal and thioredoxin-like regions by magenta and green topologies, respectively, and three HUB nodes, Glu74, Phe154 and Leu159, by cyan, blue and orange spheres. Moreover, for SELM we report the N-terminal, thioredoxin-like and C-terminal regions by magenta, green and grey topologies, respectively, and three HUB nodes, Phe59, Leu82 and Leu84, by cyan, blue and orange spheres*

4. CONCLUSIONS

In the wake of our previous articles in which we have suggested an up-expression of SELM in HCC cell lines and tissues [11, 36] we have also confirmed the up-expression of Sep15, evaluated by RT-qPCR, in the two HCC cell lines, HepG2 and Huh7, compared to normal hepatocytes, as well as in thirty HCC tissues of patients, compared to normal liver samples. Furthermore, since in the literature it is reported that Sep15 is up-regulated by ER stress [53], our results also suggest that an increased expression of Sep15 in HCC opens to the possibility of using this finding as reasonable index of ER stress levels in the presence of HCC and probably also of cancer progression, but this last issue deserves further studies to confirm this hypothesis.

On the basis of these considerations and on the fact that Sep15 shares with SELM some structural features (a thioredoxin-like domain and a structurally similar surface accessible redox active site) [12], we have performed experiments to have deeper structural information as well as details on its evolutionary history.

Moreover, our studies have also evidenced that: i) Sep15 is a very ancient protein being present from the plants to mammals accumulating mutations specific for each phylum, whereas it is very well conserved in all the mammalians. This finding supports a remarkable functional role evolutionary well retained. ii) the cysteine-rich N-terminal region is the least conserved in all the organisms compared to thioredoxin-like domain, iii) the N-terminal region of Sep15 shows an high negative net charge at neutral pH which allows us to rank this segment as a polyelectrolyte with a propensity towards structural organizations other than those classically globular, and functionally driven by post-translational modifications, which can induce changes in the overall structural organization as a consequence of its negative charge increase, iv) the 3D-model of Sep15 shows few regular elements of secondary structure at level of the N-terminal region, but a core of α/β structure mixed with some disordered loops, similarly to the SELM protein, v) the structural organization of our model has been found stable during the molecular dynamics simulation with a highly fluctuating N-terminal region as confirmed by the PCA and covariance matrix and vi) the 3D-model of Sep15 is stabilized during and after MD by H-bonds, π-cation interactions, salt bridges and π-stacking interactions and shows three well conserved residues, Glu74, Phe154 and Leu159, which possessed pivotal characteristics to be considered as structural HUB nodes.

In conclusion, the picture that comes out of our study, is that of a protein with a very peculiar structural mobility, a property to be considered carefully before any future study on its biological involvements.

ACKNOWLEDGMENT

We thank Dr. Lawrence Kelley for very useful suggestions about the modeling obtained by Phyre server.

REFERENCES

[1] El-Serag H. B., Rudolph K. L., Hepatocellular carcinoma: epidemiology and molecular carcinogenesis, Gastroenterology 132, 2557–2576 (2007).

[2] Costantini S., Capone F., Guerriero E., Marfella R., Sorice A., Maio P., Di Stasio M., Paolisso G., Castello G., Colonna G., Cytokinome profile of patients with type 2 diabetes and/or chronic hepatitis C infection, PLoS ONE 7, e39486 (2012).

[3] El-Serag H. B., Hepatocellular carcinoma: recent trends in the United States, Gastroenterology 127, S27–S34 (2004).

[4] Bruix J., Llovet J. M., Prognostic prediction and treatment strategy in hepatocellular carcinoma, Hepatology 35, 519–524 (2002).

[5] Bruix J., Sherman M., Management of hepatocellular carcinoma, Hepatology 42, 1208–1236 (2005).

[6] Imamura H., Matsuyama Y., Tanaka E., Ohkubo T., Hasegawa K., Miyagawa S., Sugawara Y. Minagawa M., Takayama T., Kawasaki S., Makuuchi M., Risk factors contributing to early and late phase intrahepatic recurrence of hepatocellular carcinoma after hepatectomy, Journal of Hepatology 38, 200–207 (2003).

[7] Llovet J. M., Fuster J., Bruix J., Intention-to-treat analysis of surgical treatment for early hepatocellular carcinoma: resection versus transplantation, Hepatology 30, 1434–1440 (1999).

[8] Valko M., Rhodes C. J., Moncol J., Izakovic M., Mazur M., Free radicals, metals and antioxidants in oxidative stress-induced cancer, Chemico-Biological Interactions 160, 1–40 (2006).

[9] Bellinger F. P., Raman A. V., Reeves M. A., Berry M.J ., Regulation and function of selenoproteins in human disease, Biochemical Journal 422, 11–22 (2009).

[10] Raucci R., Colonna G., Guerriero E., Capone F., Accardo M., Castello G., Costantini S., Structural and functional studies of the human selenium binding protein-1 and its involvement in hepatocellular carcinoma, Biochimica et Biophysica Acta 1814, 513-22 (2011).

[11] Guerriero E., Accardo M., Capone F., Colonna G., Castello G., Costantini S., Assessment of the Selenoprotein M (SELM) over-expression on human hepatocellular carcinoma tissues by immunohistochemistry, European Journal of Histochemistry 58, 2433 (2014).

[12] Guariniello S., Di Bernardo G., Colonna G., Cammarota M., Castello G., Costantini S., Evaluation of the Selenotranscriptome Expression in Two Hepatocellular Carcinoma Cell Lines, Analytical Cellular Pathology 2015, 1-6 (2015).

[13] Korotkov K. V., Kumaraswamy E., Zhou Y., Hatfield D. L., Gladyshev V. N., Association between the 15-kDa Selenoprotein and UDP-glucose: Glycoprotein Glucosyltransferase in the Endoplasmic Reticulum of Mammalian Cells, The Journal of Biological Chemistry 276, 15330–15336 (2001).

[14] Labunskyy V. M., Yoo M. H., Hatfield D. L., Gladyshev V. N., Sep15, a Thioredoxin-like Selenoprotein, Is Involved in the Unfolded Protein Response and Differentially Regulated by Adaptive and Acute ER Stresses, Biochemistry 48, 8458-8465 (2009).

[15] Kasaikina M. V., Fomenko D. E., Labunskyy V. M., Lachke1 S. A., Qiu W., Moncaster J. A., Zhang J., Wojnarowicz Jr M. W., Natarajan S. K., Malinouski M., Schweizer U., Tsuji P. A., Carlson B. A., Maas R. L., Lou M. F., Goldstein L. E., Hatfield D. L., Gladyshev V. N., 15 kDa selenoprotein (Sep15) knockout mice: roles of Sep15 in redox homeostasis and cataract development, Journal of Biological Chemistry M111, 259218(2011).

[16] Kumaraswamy E., Korotkov K. V., Diamond A. M., Gladyshev V. N., Hatfield D. L., Genetic and functional analysis of mammalian Sep15 selenoprotein, Methods in Enzymology 347, 187-197 (2002).

[17] Irons R., Tsuji P. A., Carlson B. A., Ouyang P., Yoo M. H., Xu X. M., Hatfield D. L., Gladyshev V. N., Davis C. D., Deficiency in the 15-kDa selenoprotein inhibits tumorigenicity and metastasis of colon cancer cells, Cancer Prevention Research (Phila) 3, 630–639 (2010).

[18] Tsuji P. A., Naranjo-Suarez S., Carlson B. A., Tobe R., Yoo M. H., Davis C. D., Deficiency in the 15 kDa selenoprotein inhibits human colon cancer cell growth, Nutrients 3, 805–817 (2011).

[19] Ferguson A. D., Labunskyy V. M., Fomenko D. E., Arac D., Chelliah Y., Amezcua C. A., Rizo J., Gladyshev V. N., Deisenhofer J., NMR structures of the selenoproteins Sep15 and SelM reveal redox activity of a new thioredoxin-like family, The Journal of Biological Chemistry 281, 3536–3543 (2006).

[20] Qi Y., Grishin N. V., Structural classification of thioredoxin-like fold proteins, Proteins: Structure, Function, and Bioinformatics 58, 376 – 388 (2005).

[21] Labunskyy V. M., Hatfield D. L., Gladyshev V. N., The Sep15 Protein Family: Roles in Disulfide Bond Formation and Quality Control in the Endoplasmic Reticulum, IUBMB Life 59, 1-5 (2007).

[22] Korotkov K. V., Novoselov S. V., Hatfield D. L., Gladyshev V. N., Mammalian selenoprotein in which selenocysteine (Sec) incorporation is supported by a new form of Sec insertion sequence element, Molecular and Cellular Biology 22, 1402 – 1411 (2002).

[23] Costantini S., Di Bernardo G., Cammarota M., Castello G., Colonna G., Gene expression signature of human HepG2 cell line, Gene 518, 335–345 (2013).

[24] Thompson J. D., Higgins D. G., Gibson T., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positionspecific gap penalties and weight matrix choice, Nucleic Acids Research 22, 4673–4680 (1994).

[25] Cuff J. A., Barton G. J., Application of enhanced multiple sequence alignment profiles to improve protein secondary structure prediction, Proteins 40, 502–511 (2000).

[26] Kozlowski L. P., Bujnicki J. M., MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins, BMC Bioinformatics 13, 111 (2012)

[27] Blom N., Gammeltoft S., Brunak S., Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites, Journal of Molecular Biology 294 1351–62 (1999).

[28] Monigatti F., Gasteiger E., Bairoch A., Jung E., The Sulfinator: predicting tyrosine sulfation sites in protein sequences, Bioinformatics 18, 769–70 (2002).

[29] Steentoft C., Vakhrushev S. Y., Joshi H. J., Kong Y., Vester-Christensen M. B., Schjoldager K. T., Lavrsen K., Dabelsteen S., Pedersen N. B., Marcos-Silva L., Gupta R., Bennett E. P., Mandel U., Brunak S., Wandall H. H., Levery S. B., Clausenc H., Precision mapping of the human OGalNAcglycoproteome through SimpleCell technology, EMBO Journal 32, 1478–88 (2013).

[30] Mooney C., Pollastri G., Shields D. C., Haslam N. J., Prediction of short linear protein binding regions, Journal of Molecular Biology 415, 193-204 (2012).

[31] Dosztányi Z., Mészáros B., Simon I., ANCHOR: web server for predicting protein binding regions in disordered proteins, Bioinformatics 25, 2745-2746 (2009) .

[32] Disfani F. M., Hsu W. L., Mizianty M. J., Oldfield C. J., Xue B., Dunker A. K., Uversky V. N., Kurgan L., MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins, Bioinformatics 28, i75-83 (2012).

[33] Das R. K., Pappu R. V., Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues, PNAS 110, 13392–13397 (2013).

[34] Altschul S. F., Gish W., Miller W., Myers E., Lipman D. J., Best local alignment search tool, Journal of Molecular Biology 215, 403–410 (1990).

[35] Sali A., Blundell T. L., Comparative protein modelling by satisfaction of spatial restraints, Journal of Molecular Biology 234, 779–815 (1993).

[36] Kelley L. A., Mezulis S., Yates C. M., Wass M. N., Sternberg M. J., The Phyre2 web portal for protein modeling, prediction and analysis, Nature Protocol 10, 845-58 (2015).

[37] Autiero I., Costantini S., Colonna G., Human sirt-1: molecular modeling and structure–function relationships of an unordered protein, PLoS One 4, e7350 (2009).

[38] Sippl M. J., Recognition of errors in three-dimensional structures of proteins, Proteins 17, 355–362(1993).

[39] Gopalakrishnan K., Sowmiya G., Sheik S. S., Sekar K., Ramachandran plot on the web (2.0), Protein and Peptide Letters, 14, 669–671 (2007).

[40] Ahmad S., Gromiha M. M., Fawareh H., Sarai A., ASAView: Solvent Accessibility Graphics for proteins, BMC Bioinformatics 5, 51 (2004).

[41] Tina K. G., Bhadra R., Srinivasan N., PIC: Protein Interactions Calculator, Nucleic Acids Research 35, W473–W476 (2007).

[42] Vangone A., Spinelli R., Scarano V., Cavallo L., Oliva R., COCOMAPS: a web application to analyse and visualize contacts at the interface of biomolecular complexes, Bioinformatics 27, 2915-6 (2011).

[43] Gibrat J. F., Madej T., Bryant S. H., Surprising similarities in structure comparison, Current Opinion on Structural Biology 6, 377-85 (1996).

[44] McWilliam H., Li W., Uludag M., Squizzato S., Park Y. M., Buso N., Cowley A. P., Lopez R., Analysis Tool Web Services from the EMBL-EBI, Nucleic Acids Research 41, W597-600 (2013).

[45] Van Der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A. E., Berendsen H. J., GROMACS: fast, flexible, and free, Journal of Computational Chemistry 26, 1701–1718 (2005).

[46] Guariniello S., Colonna G., Raucci R., Costantini M., Di Bernardo G., Bergantino F., Castello G., Costantini S., Structure–function relationship and evolutionary history of the human selenoprotein M (SelM) found over-expressed in hepatocellular carcinoma, Biochimica et Biophysica Acta 1844, 447–456 (2014).

[47] Raucci R., Colonna G., Giovane A., Castello G., Costantini S., N-terminal region of human chemokine receptor CXCR3: Structural analysis of CXCR3(1–48) by experimental and computational studies, Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1884, 1868-1880 (2014).

[48] Martin A. J., Vidotto M., Boscariol F., Di Domenico T., Walsh I., Tosatto S. C., RING: networking interacting residues, evolutionary information and energetics in protein structures, Bioinformatics 27, 2003-5 (2011).

[49] Assenov Y., Ramírez F., Schelhorn S. E., Lengauer T., Albrecht M., Computing topological parameters of biological networks, Bioinformatics 24, 282–284 (2008).

[50] Doncheva N. T., Klein K., Domingues F. S., Albrecht M., Analyzing and visualizing residue networks of protein structures, Trends in Biochemical Sciences 36,179–182 (2011).

[51] Fan L., Song B. , Sun G. , Ma T. , Zhong F., Wei W., Endoplasmic reticulum stress-induced resistance to doxorubicin is reversed by paeonol treatment in human hepatocellular carcinoma cells, PLoS One 8, e62627 (2013).

[52] Honma Y., Harada M., New therapeutic strategy for hepatocellular carcinoma by molecular targeting agents via inhibition of cellular stress defense mechanisms, J UOEH 36, 229-35 (2014).

[53] Labunskyy V. M., Ferguson A. D., Fomenko D. E., Chelliah Y., Hatfield D. L., Gladyshev V. N., A novel cysteine-rich domain of Sep15 mediates the interaction with UDP-glucose:glycoprotein glucosyltransferase, Journal of Biological Chemistry 280, 37839-45 (2005).

[54] Bergantino F., Guariniello S., Raucci R., Colonna G., De Luca A., Normanno N., Costantini S., Structure-fluctuation-function relationships of seven pro-angiogenic isoforms of VEGFA, important mediators of tumorigenesis, BBA - Proteins and Proteomics 1854, 410-25 (2015).