

OPEN

Structural analysis of human SEPHS2 protein, a selenocysteine machinery component, over-expressed in triple negative breast cancer

Carmine Nunziata^{1,5}, Andrea Polo^{1,5}, Angela Sorice^{1,5}, Francesca Capone¹, Marina Accardo², Eliana Guerriero³, Federica Zito Marino², Michele Orditura⁴, Alfredo Budillon^{1,6*} & Susan Costantini^{1,6*}

Selenophosphate synthetase 2 (SEPHS2) synthesizes selenide and ATP into selenophosphate, the selenium donor for selenocysteine (Sec), which is cotranslationally incorporated into selenoproteins. The action and regulatory mechanisms of SEPHS2 as well as its role in carcinogenesis (especially breast cancer) remain ambiguous and need further clarification. Therefore, lacking an experimentally determined structure for SEPHS2, we first analyzed the physicochemical properties of its sequence, modeled its three-dimensional structure and studied its conformational behavior to identify the key residues (named HUB nodes) responsible for protein stability and to clarify the molecular mechanisms by which it induced its function. Bioinformatics analysis evidenced higher amplification frequencies of SEPHS2 in breast cancer than in other cancer types. Therefore, because triple negative breast cancer (TNBC) is biologically the most aggressive breast cancer subtype and its treatment represents a challenge due to the absence of well-defined molecular targets, we evaluated SEPHS2 expression in two TNBC cell lines and patient samples. We demonstrated mRNA and protein overexpression to be correlated with aggressiveness and malignant tumor grade, suggesting that this protein could potentially be considered a prognostic marker and/or therapeutic target for TNBC.

The selenoproteins include selenocysteine (Sec), which is a nonstandard amino acid in the UGA codon. These proteins are found in all organisms throughout the tree of life. The incorporation of this 21st amino acid into proteins in mammalian cells is guided by the Sec biosynthesis machinery¹, of which selenophosphate synthetase 2 (SEPHS2) is an important component². SEPHS2 catalyzes the synthesis of selenophosphate from selenide, adenosine triphosphate (ATP), and water and produces adenosine monophosphate (AMP) and inorganic phosphate. Selenophosphate is the selenium donor for Sec synthesis, which, in contrast to other amino acids, takes place on its own tRNA, tRNA^{Sec}^{3,4}. The selenocysteine incorporation machinery requires certain protein-protein and protein-RNA interactions to function and is guided by stem-loop structures localized in the 3' untranslated regions of selenoprotein-encoding genes. Recently, SEPHS2 has been demonstrated to interact with selenocysteine synthase (SepSecS) and SEPHS1, which plays a nonessential role in selenoprotein metabolism⁵.

Although some papers have reported that a decrease or increase in selenoprotein expression can induce a cancer phenotype⁶, the role of selenoproteins in carcinogenesis and their mechanisms of action and regulation remain ambiguous and need further clarification. In fact, few data are available regarding changes in SEPHS2

¹Experimental Pharmacology Unit, Laboratori di Mercogliano, Istituto Nazionale Tumori – IRCCS – Fondazione G. Pascale, Napoli, Italia. ²Department of Mental and Physical Health and Preventive Medicine, Università degli Studi della Campania “Luigi Vanvitelli”, Pathology Unit, Napoli, Italia. ³Istituto Tecnico Industriale (ITIS) “Guido Dorso”, Avellino, Italia. ⁴Division of Medical Oncology, Department of Precision Medicine, School of Medicine, Università degli Studi della Campania “Luigi Vanvitelli”, Napoli, Italia. ⁵These authors contributed equally: Carmine Nunziata, Andrea Polo and Angela Sorice. ⁶These authors jointly supervised this work: Alfredo Budillon and Susan Costantini. *email: a.budillon@istitutotumori.na.it; s.costantini@istitutotumori.na.it

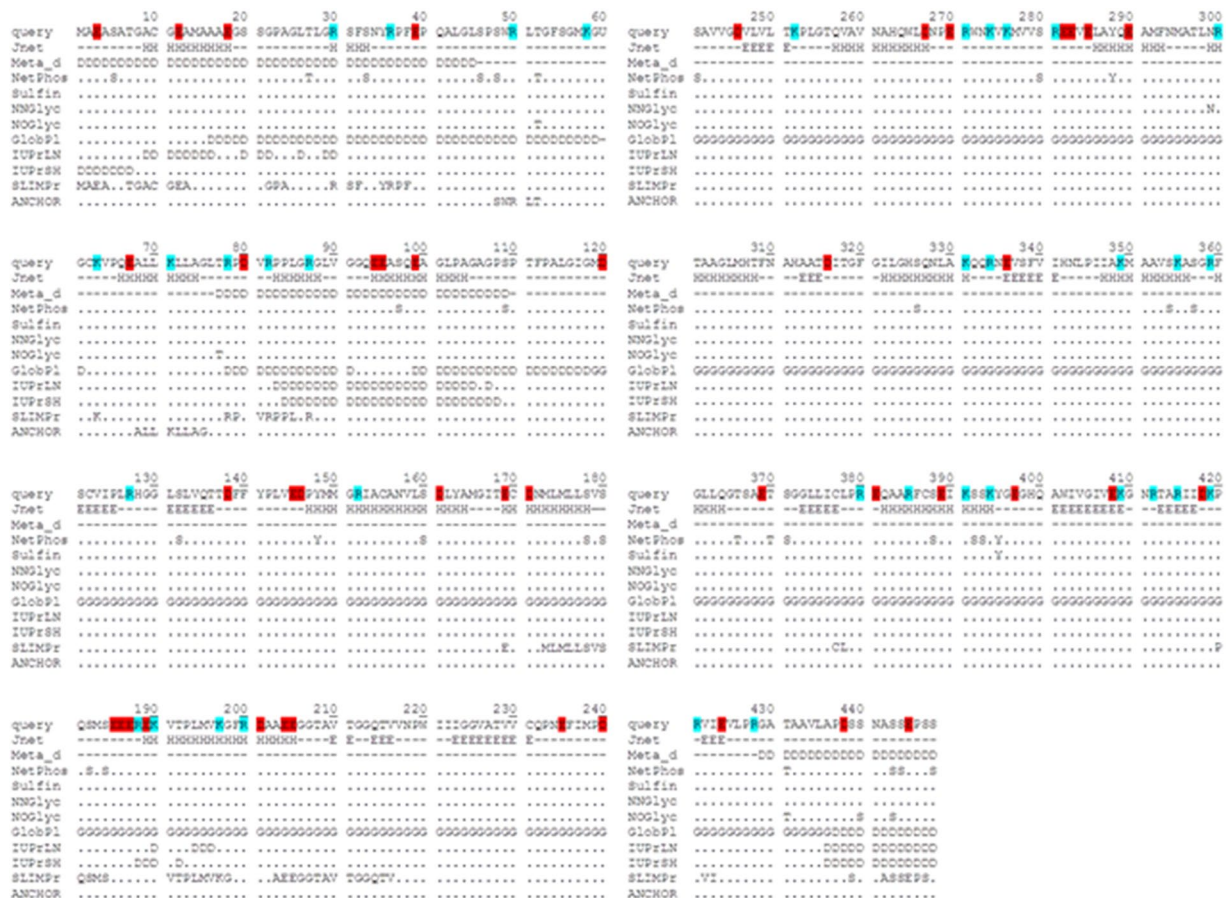


Figure 1. Sequence analysis of human SEPHS2. The negatively and positively charged residues are shown in red and cyan, respectively. We report predictions for secondary structure (Jnet), disorder propensity (Meta_d and IUPred), globularity propensity (GlobPL), predicted and experimentally determined phosphorylation sites (NetPhos and Phosphosite), sulfination sites (Sulfinator), glycosylation sites (NetNGlyc and NetOGlyc), and molecular recognition features (MorfPred and ANCHOR).

expression in cancer. Recently, researchers demonstrated that SEPHS2 expression was significantly decreased in gastric cancer and para-carcinoma tissues⁷. Moreover, Maciel-Dominguez *et al.* (2013) evaluated whether the effects of selenium on gene expression were exerted through miRNAs in a human colon adenocarcinoma cell line (Caco-2)⁸ and demonstrated that miR-185 played a role in the upregulation of SEPHS2 expression the maintenance of selenoprotein synthesis⁸. SEPHS2 levels were reported to be decreased in BRCA1-linked breast cancer versus BRCA1-negative cancer⁹, although its expression in specific breast cancer subtypes is unknown. In fact, breast cancer is a heterogeneous cancer with three different molecular subtypes based on overexpression of the estrogen and progesterone receptors, overexpression of the epidermal growth factor receptor 2 (HER2) receptor or the absence of these three receptors¹⁰. This latter subtype is defined as TNBC. TNBC constitutes 10%–20% of all breast cancers, affects younger patients, has lymph node involvement at diagnosis, and is the most aggressive type¹¹. Treatment of TNBC patients still represents a challenge due to the absence of well-defined molecular targets^{10,11}.

Therefore, since no three-dimensional structure of this protein has been experimentally determined, we focused our attention first on the utility of determining the SEPHS2 structure and studying the related structure-fluctuation relationships to examine the putative molecular mechanisms through which it induced its function. Additionally, because a bioinformatics analysis of databases found higher amplification frequencies in breast cancer, we evaluated SEPHS2 expression in two TNBC cell lines (MDA-MB231 and MDA-MB468) and TNBC patients to elucidate whether this protein could be correlated with the initiation and progression of this breast cancer subtype.

Results and Discussion

Sequence analysis of human SEPHS2. First, the human SEPHS2 sequence (UniProt code: Q99611) was analyzed to identify the relationships between its amino acid composition and conformation and particularly to assess how the amino acid composition of SEPHS2 could affect the conformational propensity to form regular secondary structures or be disordered using the ProtParam tool¹². The sequence is characterized by 27 prolines, 48 glycines and 43 negatively ($f^- = 0.096$) and 36 positively ($f^+ = 0.080$) charged residues (Fig. 1). SEPHS2 is located in Region 1 of the state diagram, which contains weak polyampholytes and polyelectrolytes. The fraction

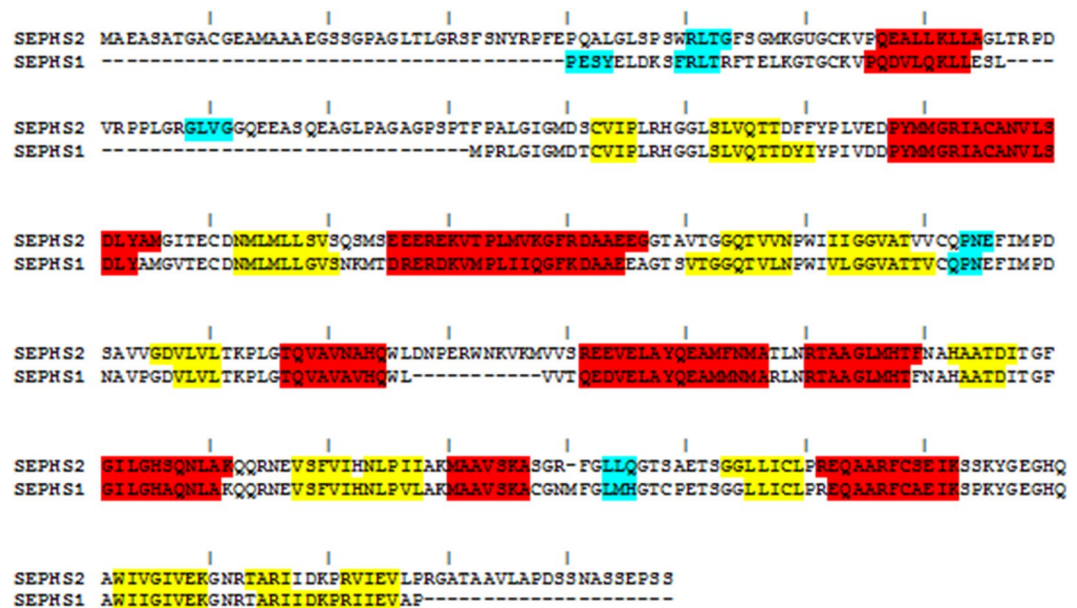


Figure 2. BLAST alignment between human SEPHS2 and human SEPHS1. The 3_{10} -helices, α -helices and β -strands are shown in cyan, red and yellow, respectively.

of charged residues (FCR) in the protein is <0.25 , the net charge per residue (NCPR) is <0.25 , and the protein has a propensity to form globule and tadpole ensembles in agreement with Das and Pappu¹³ (Supplementary Fig. 1).

The secondary structure prediction by Jnet¹⁴ suggested the putative presence of 15 α -helices, 12 β -strands and some disordered regions. This finding was confirmed by the prediction of putative disordered regions by Genesilico MetaDisorder, GlobPlot and IUPred¹⁵. Some regions with high disorder propensities are located at the N-terminal (positions 1–60 and 77–110) and C-terminal regions (positions 428–448) (Fig. 1).

Since posttranslational modifications are important for the functional features of intrinsically disordered proteins (IDPs), including modulating the recognition of molecular partners¹⁶, we evaluated the putative presence of phosphorylation, sulfation and glycosylation sites as well as linear motifs associated with protein interactions. As shown in Fig. 1, analysis of the SEPHS2 sequence revealed 31 phosphorylation sites predicted by NetPhos¹⁷. Among these sites, four located at positions 33, 46, 97 and 109 have also been found experimentally, as shown by the PhosphoSitePlus analysis¹⁸. Sulfation prediction by the Sulfinator tool¹⁹ showed only one sulfation site at position 395, whereas the prediction of glycosylation sites by the NetNGlyc and NetOGlyc tools²⁰ indicated a potential N-linked site at position 299 and five O-linked sites at positions 52, 77, 431, 440 and 443. The prediction of binding regions in disordered proteins by the ANCHOR²¹ and α -MoRF-PredII²² tools indicated potential binding positions in regions 1–80, 170–218, 379–380 and 420–448.

All these analyses demonstrated that three regions of SEPHS2 with high disordered propensities (positions 1–60, 77–110 and 428–448) contained posttranslational modifications that could induce a conformational change on the SEPHS2 structure with their charges.

Molecular modeling of human SEPHS2. The SEPHS2 structure was modeled by an integrated procedure based on comparative modeling and fold recognition that was described in our recent papers^{23,24}. A BLAST search²⁵ showed that region 41–427 of SEPHS2 had 75% sequence identity with human SEPHS1 (PDB code: 3FD5)²⁶. No possible templates were available for the first 40 residues in the N-terminal region and the last 19 residues in the C-terminal region. Therefore, we created ten models for region 41–427 of SEPHS2 using the SEPHS1 structure as the template in the MODELLER program²⁷ (Fig. 2) and analyzed their energetic and stereochemical quality. The best selected model showed 98.1% of residues in the allowed regions of the Ramachandran plot²⁸ and an energetic Z-score²⁹ of -9.01 , demonstrating the good quality of the model. Then, this model was subjected to loop refinement using the MODELLER tool to energetically improve the conformation of the disordered 77–110 region loop. The final SEPHS2 model for the 41–427 region is composed of 3 3_{10} -helices, 9 α -helices and 13 β -sheets. Comparison of the secondary structures between human SEPHS2 and SEPHS1 showed that the helices and β -strands were sufficiently conserved along the sequences, with only a few changes in their lengths (Fig. 2).

To gain further insights into the complete structure of SEPHS2, we modeled the N-terminal (1–40) and C-terminal (428–448) regions using the “*ab initio*” folding method with the MUSTER program³⁰. The best 3D model of the N-terminal region had a Z-score of -0.26 and a total percentage of residues in the allowed regions of the Ramachandran plot of 97.4%, whereas that of the C-terminal region had a Z-score of -1.99 and 100% of residues in the favored region.

Finally, we modeled the complete SEPHS2 structure using the three models reported above as templates for regions 1–40, 41–427 and 428–448. The 3D model of complete SEPHS2 had an energetic Z-score of -8.5 and 98.7% of the residues in the allowed regions. As shown in Fig. 3, the entire SEPHS2 model showed an N-terminal

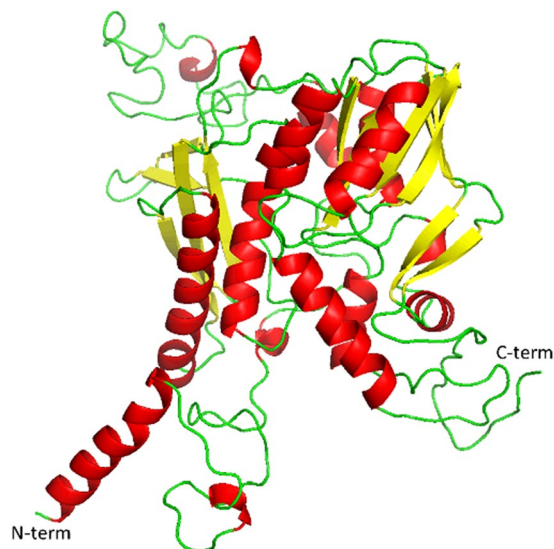


Figure 3. Complete SEPHS2 model obtained by the molecular modeling approach. In detail, 3_{10} -helices and α -helices are reported in red, β -strands in yellow and loops in green.

domain with an α -helix and a long disordered loop, a central core with an α - β 2-layer sandwich architecture and a disordered C-terminal domain.

Overall, these data highlighted that the SEPHS2 model conserved the structure of the SEPHS family. This finding was also confirmed by the circular dichroism spectrum analysis obtained from the protein atom coordinates by the PDB2CD tool (<http://pdb2cd.cryst.bbk.ac.uk/>). This analysis demonstrated overlap of the spectra and similarity of secondary structures related to our model and crystallographic structures of SEPHS1 from four different species (human, *Escherichia coli*, *Aquifexaeolicus* and *Leishmania major*) (Supplementary Fig. 2).

Subsequently, the whole model was subjected to molecular dynamics (MD) simulations at neutral and acidic pH in water by the GROMACS software³¹ to obtain more information on its structural dynamics features.

Comparison between MD simulations at neutral and acidic pH. MD simulations were conducted at neutral and acidic pH to study the possible roles played by charged amino acids in SEPHS2 and the related electrostatic interactions. Importantly, an acidic extracellular pH is reported to be a major feature of cancer tissues³².

The analysis of the trajectories demonstrated that the MD simulations at both neutral and acidic pH reached convergence after 2 ns (Fig. 4A). However, compared to the MD simulation at neutral pH, the MD simulation at acidic pH reached (i) a greater compactness with a gyration radius (Rg) equal to 2.15 nm (Fig. 4B) and (ii) more stability due to an increase in the H-bond number (Fig. 4C) even if the protein floated more, as was visible in the root mean square fluctuation (RMSF) plot (Fig. 4D). Moreover, evaluation of the solvent accessible area (ASA) (Fig. 4E), using protein a-sphericity as an indirect measure, showed that SEPHS2 tended to be more compact at acidic pH and tended to move toward the center of the box (Fig. 4F).

Analysis of the pattern of secondary structure evolution (Supplementary Fig. 3) suggested that most of the helices and β -strands remained relatively stable during MD simulation at neutral pH, with the exception of some terminal residues in short helices that tended to lose their regular structure. On the other hand, during the simulation at acidic pH, the secondary structure evolution of SEPHS2 revealed an increase in structural fluctuations compared to those found at neutral pH; in fact, some short α -helices and β -strands were completely lost during the simulation (Supplementary Fig. 3).

Subsequently, to better understand the collective fluctuations of SEPHS2 during the MD simulations, we performed principal component analysis (PCA), cluster analysis, and covariance matrix analysis. Specifically, the projections of the principal eigenvectors were plotted onto the plane over the $C\alpha$ coordinates, and cluster analysis was performed based on the root mean square deviation (RMSD) values between the different conformations by selecting structures sharing similar conformational features during the simulation. The PCA plot and covariance matrix showed that during the MD simulation at acidic pH: (i) the atoms of SEPHS2 moved much faster and favored conformational interchanges to reach increased structural stability (Fig. 5); and (ii) the fluctuating regions, indicated by a more intensive red color, were located at both the N-terminal and C-terminal regions, which included negatively charged residues (Supplementary Fig. 4). The cluster analysis showed the presence of eight and seven clusters, among which three and two were the most populated at neutral and acidic pH, respectively (Fig. 5). However, although the total number of clusters was similar in the simulations at neutral and acidic pH, the transition rate of the conformational interchange among clusters was very high at the acidic pH, indicating that more fluctuations were necessary to reach a compact structure.

Finally, we focused on residue Sec. 60. During the MD simulation at acidic pH, this residue was more exposed on the surface and interacted with Lys 58, Cys 62 and Arg 359 during all simulations.

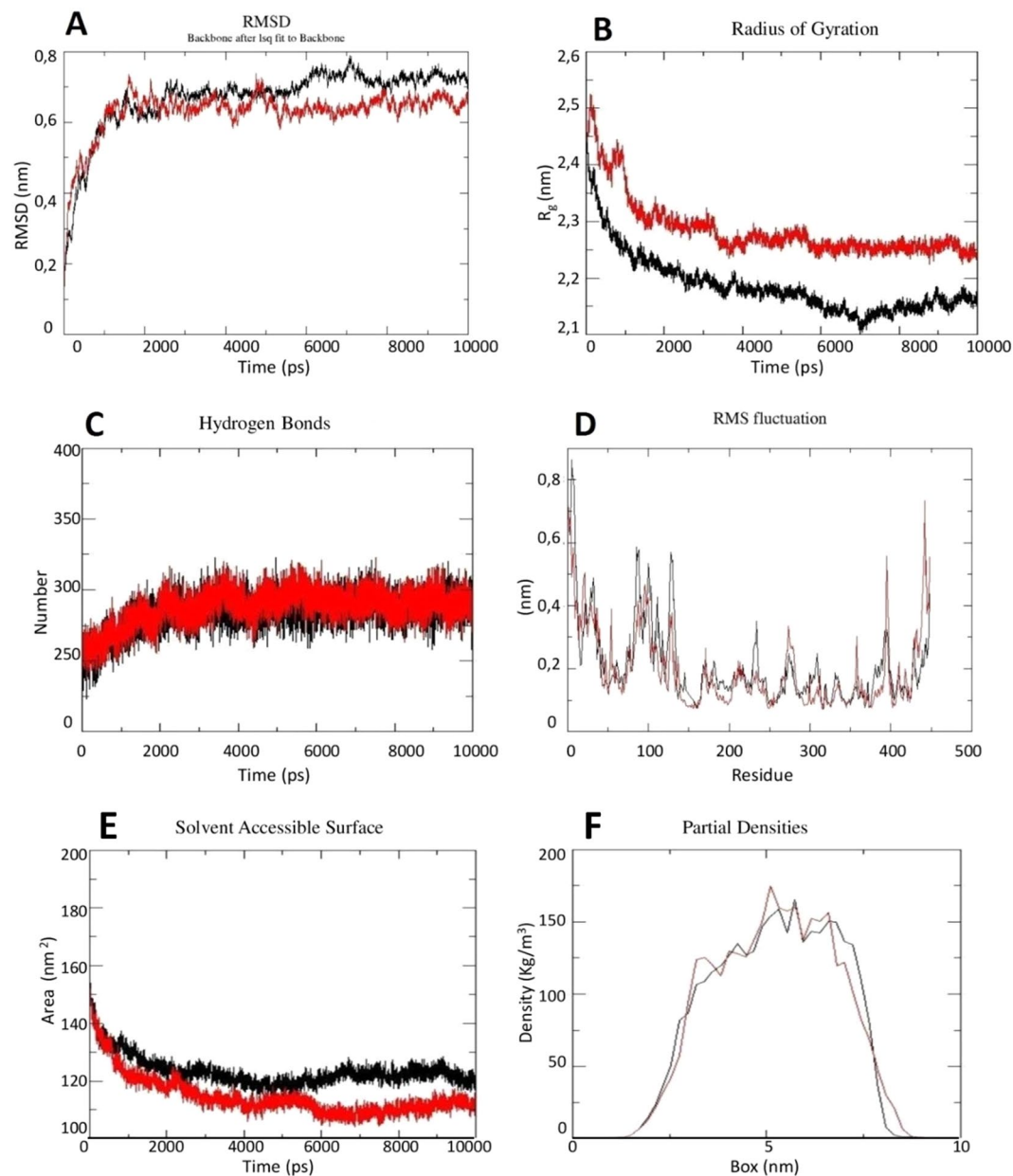


Figure 4. (A) RMSD, (B) gyration radius, (C) H-bond, (D) RMSF, and (E) solvent accessible surface plots and (F) partial densities for SEPHS2 during MD simulations at neutral (in red) and acidic (in black) pH.

Identification of key residues (HUB) by residue interaction network (RIN) analysis during MD simulations at neutral and acidic pH.

For studies of the specific role of individual residues in the general organization of a protein, RIN analysis is a very useful method to identify residues with the strongest coordinating role (i.e., HUB residues) and to obtain information about protein stability^{23,24}. Our model was dynamically stabilized by a certain number of H-bonds and interactions with the closest atoms (IACs) as well as by salt bridges and π -cation and π -stacking interactions during the MD simulations at neutral and acidic pH based on evaluation with the Protein Interactions Calculator (PIC)³³, HBPLUS³⁴ and COCOMAPS³⁵ tools (Table 1). In detail, the RIN analysis was performed on the SEPHS2 conformation at 0, 2, 4, 6, 8 and 10 ns of simulation time by evaluating the following seven topological parameters to identify HUB nodes: (i) density of the maximum neighborhood component (DMNC), (ii) degree, (iii) maximal clique centrality (MCC), (iv) maximum neighborhood component (MNC), (v) betweenness, (vi) bottleneck, and (vii) closeness^{36,37}.

This analysis showed that in the MD simulation at neutral pH (Fig. 6), Phe 199 was the most conserved HUB residue during the simulation based on five out of seven centrality measures, suggesting that this residue might play an important structural role during MD evolution. Moreover, the analysis identified the following HUB residues: Lys 252 after 2, 4 and 8 ns of simulation, Tyr 163 after 4, 6 and 10 ns, and Phe 140 after 6, 8 and 10 ns. These data suggested that Lys 252, Tyr 163 and Phe 140 followed the conformation changes of the protein during the simulation at neutral pH and played a role in stabilizing the protein structure (Table 2). In fact, during the MD

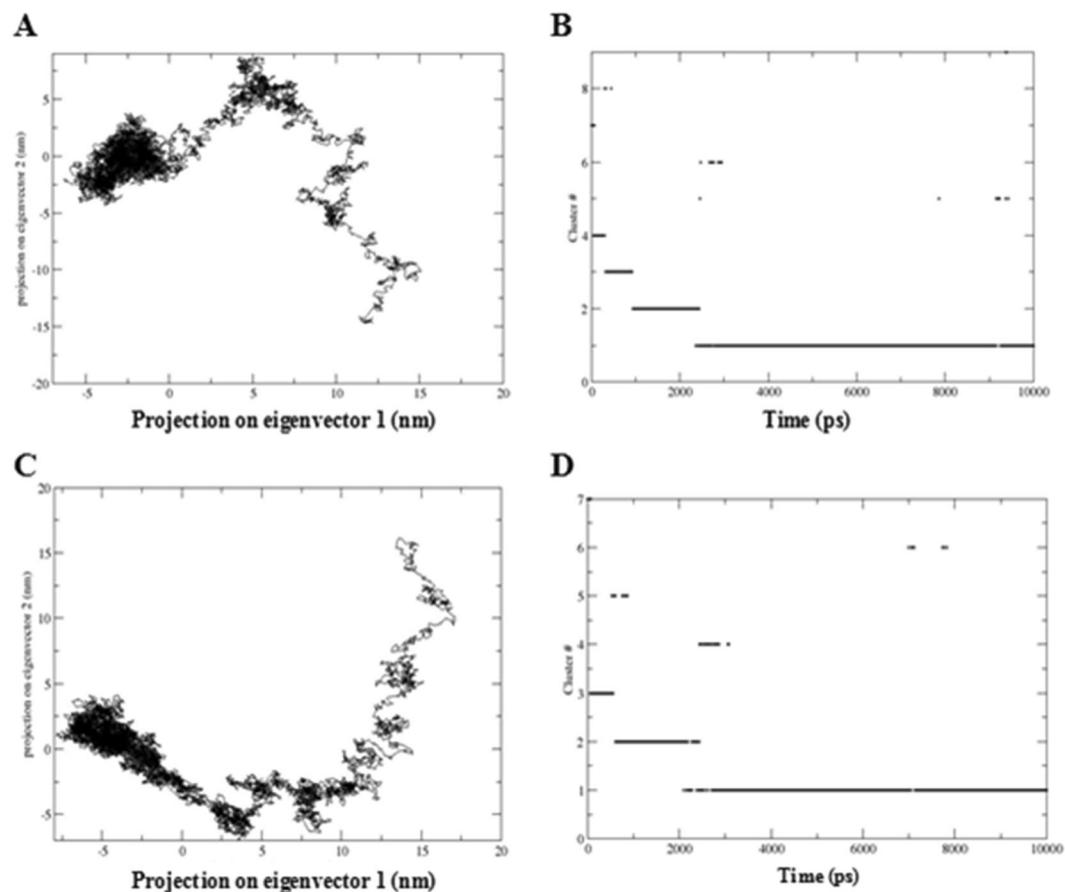


Figure 5. Projection plot and cluster analysis results for SEPHS2 during MD simulations at neutral (A,B) and acidic pH (C,D), respectively.

Neutral pH	H-bonds				π -cation	π -stacking	Salt bridges	IAC
	MM	MS	SS	SM				
0 ns	199	20	24	35	7	4	7	1990
2 ns	218	33	34	30	3	5	7	2202
4 ns	230	28	28	36	2	6	6	2235
6 ns	217	42	21	39	1	4	5	2238
8 ns	244	31	27	29	1	3	6	2264
10 ns	227	38	25	41	2	4	5	2241
Acidic pH								
0 ns	205	23	19	32	7	5	11	1990
2 ns	222	21	14	39	8	2	5	2330
4 ns	222	29	16	36	11	4	5	2354
6 ns	233	27	12	37	8	4	3	2360
8 ns	229	27	15	37	8	3	6	2389
10 ns	228	25	11	38	7	4	6	2351

Table 1. Number of H-bonds (subdivided into main chain - main chain (MM), main chain - side chain (SM), side chain - side chain (SS) and side chain - main chain (SM)), π -cation and π -stacking interactions, salt bridges and interactions of closest atoms (IACs) during MD simulations in water at neutral and acidic pH.

simulation, these four HUB residues tended to form a higher number of H-bonds (Supplementary Table 1) and to become more buried, as shown by the ASA analysis (Supplementary Table 2), confirming their roles in promoting the compactness of this protein.

By analyzing the network obtained for MD simulation at acidic pH (Fig. 7), we showed that no HUB node was identified at 0 ns, whereas nine residues (Phe 320, Glu 283, Lys 190, Phe 199, His 325, Met 295, Phe 140, Val

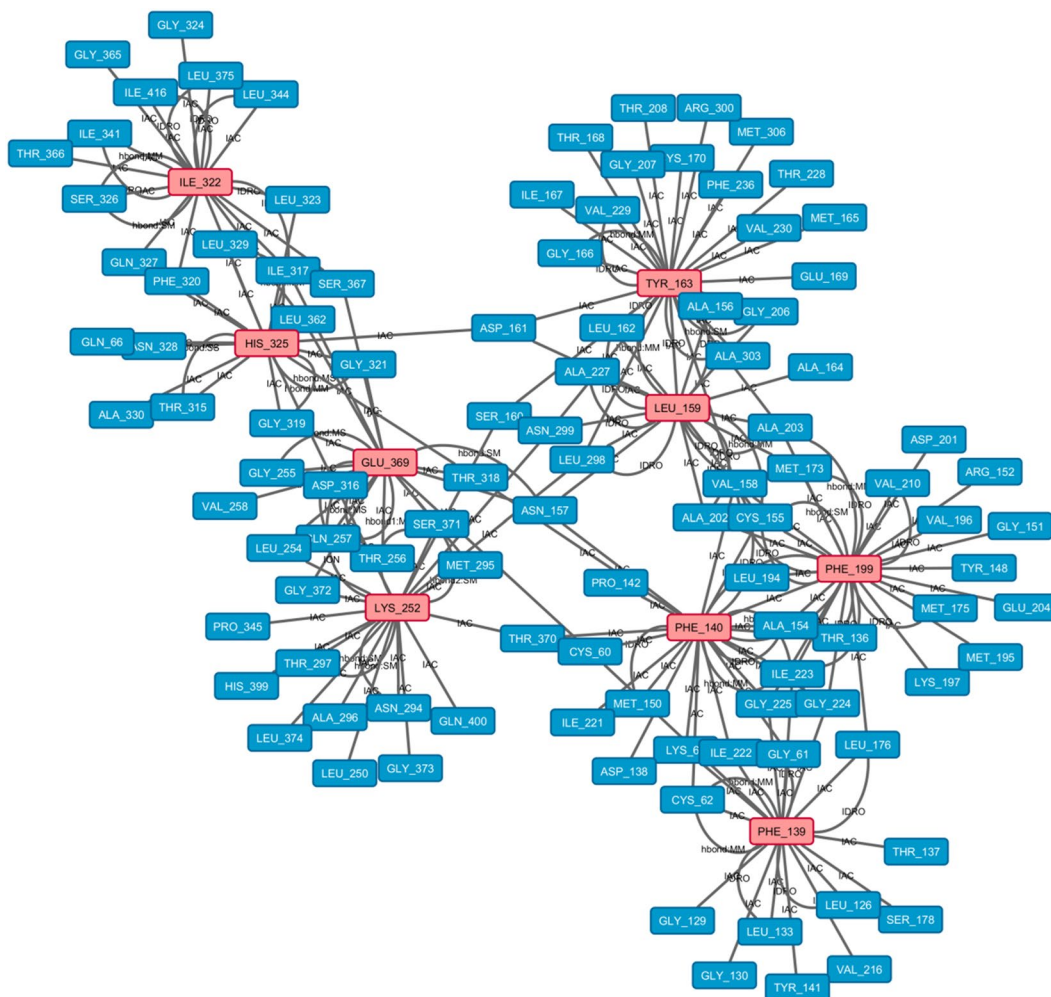


Figure 6. Residue interaction network for SEPHS2 at the end of the neutral pH simulation. HUB nodes are in pink, and the other residues are in light blue.

Neutral pH	0 ns	Residue interaction network				
		2 ns	4 ns	6 ns	8 ns	10 ns
	Ile 322	Lys 252	His 325	Phe 140	Phe 140	Phe 139
	Phe 199	Phe 199	Leu 159	Tyr 163	Glu 369	Phe 140
			Lys 252	Phe 199	Ile 322	Phe 199
			Phe 199		Lys 252	Tyr 163
			Tyr 163		Phe 199	
Acidic pH		Phe 320	Phe 199	Phe 199	Phe 199	Phe 199
		Glu 283	His 325	Phe 320	Phe 140	Val 158
		Lys 190	Met 295	Phe 140	Phe 320	Asn 157
			Phe 140			
			Phe 360			

Table 2. Differences in residue interaction network (RIN) analysis of SEPHS2 during MD simulations in water at neutral and acidic pH.

158 and Asn 157) were considered HUB nodes during the rest of the simulation (Table 2). Among these residues, an increase was observed in the formation of H-bonds during the simulation for Phe 320, Glu 283 and Met 295 (Supplementary Table 1), whereas a decrease in ASA values was observed for Phe 320, Glu 283, Lys 190, Phe 199, His 325 and Phe 140. These findings confirmed that these latter residues were buried during the simulation and tended to play a structural role (Supplementary Table 2).

Then, to examine whether these HUB nodes created subnetworks able to stabilize the structure, we focused on the interactions between them. As shown in Figs 6 and 7, the pivotal HUB residues formed IAC, π -cation and

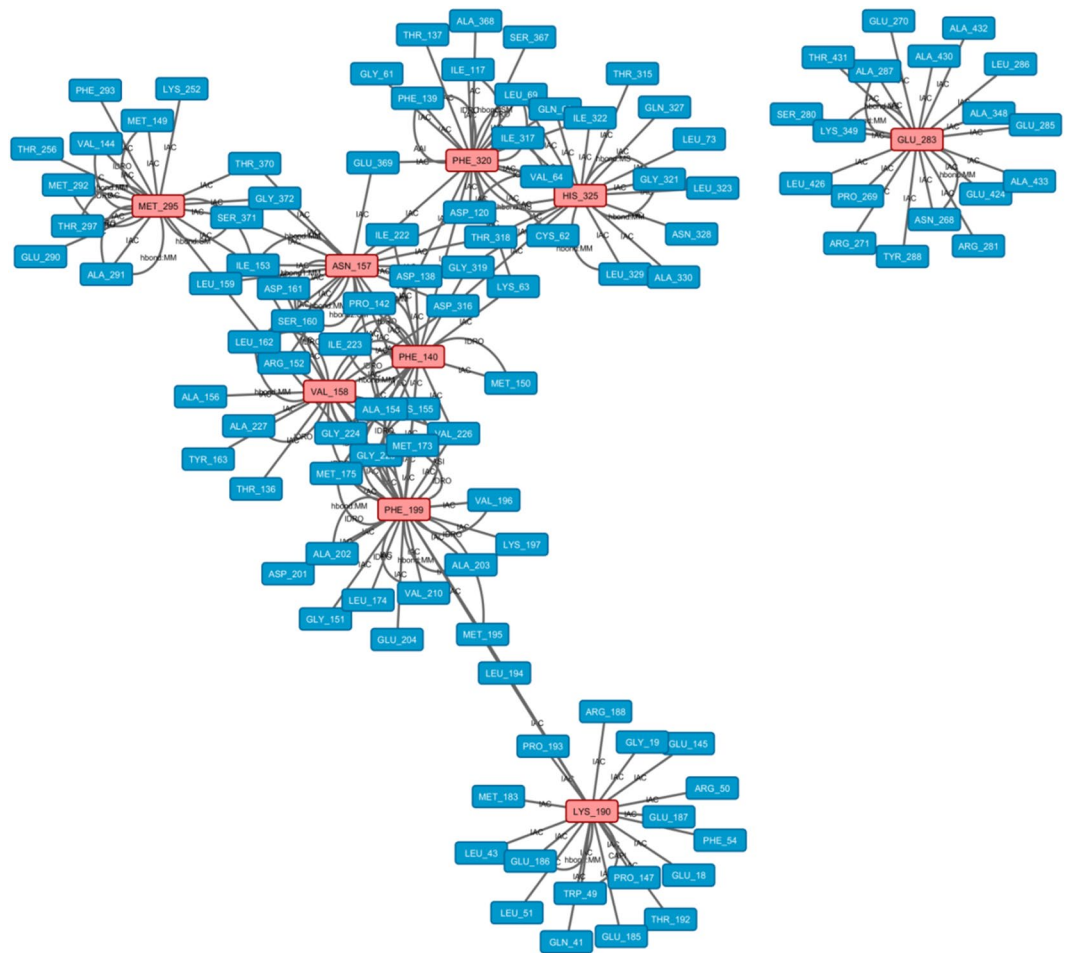


Figure 7. Residue interaction network for SEPHS2 at the end of the acidic pH simulation. HUB nodes are in pink, and the other residues are in light blue.

hydrophobic interactions and H-bonds during the MD simulations at neutral and acidic pH, confirming a structural role for these HUB nodes in driving the conformational organization of SEPHS2 for all MD simulations and in maintaining its compactness.

Moreover, to understand the possible biological relevance of these HUB residues, we evaluated whether they were part of the active site composed by Sec. 60, Lys 63 and 319–325 region. In fact, it is notable that (i) Phe 320 (a HUB residue at acidic pH), an Ile 322 (HUB residue at neutral pH) and a His 325 (HUB residue at neutral and acidic pH) were present in the active site; (ii) His 325 formed a H-bond with Gly 319 of the active site; and (iii) Ile 322 formed a hydrophobic interaction with Leu 323. However, the other HUB residues interacted with the active site by proximity interactions (IAC), suggesting a potential molecular advantage of the HUB residues in stabilizing the active site and permitting SEPHS2 to develop its activity.

To verify if there were mutations of HUB residues reported in the SNP databases and correlated with breast cancer and/or the TNBC subtype, we interrogated different online databases containing SNPs such as dbSNP (<https://www.ncbi.nlm.nih.gov/snp/>), CbioPortal (<http://cbioportal.org>)^{38,39} and COSMIC (<https://cancer.sanger.ac.uk/cosmic>). We found 1252 SNPs associated with SEPHS2 by dbSNP database, and in particular, we found one missense mutation of Tyr163 (rs1284481836) and Phe140 (rs770633929), two missense mutations (rs1366381318, rs746372646) of Met295 and four missense mutations (rs376948016, rs1446115907) of Asn157. However, all of them are not still associated with a pathologic state and no clinical significance was still evidenced. Moreover, we searched the CbioPortal and cosmic databases for mutations of the residues correlated with breast cancer, particularly, the TNBC subtype. In detail, we found three missense mutations (Val404Ile, Met183Ile and Phe360Val) correlated with the breast invasive carcinoma by CbioPortal, and four missense mutations (and Gly358Glu) were correlated to the breast ductal carcinoma by COSMIC. No specific mutations in HUB residues were found to correlate to the TNBC subtype.

In conclusion, all these data related to RINs at neutral and acidic pH revealed the following: (i) three HUB residues (two at neutral pH and two at acidic pH) were present in the active site; (ii) at acidic pH, a greater number of the HUB residues were buried, increasing the compactness of the structure; and (iii) these buried residues included three residues, Met 295, Phe 140 and Asn 157, reported to be mutated by the dbSNP database, which suggests that at acidic pH, their mutation can decrease the compactness of the structure and induce conformational fluctuations and/or changes to the biological activity of SEPHS2.

	Number of patients
Female	30
Age (mean \pm st.dev.)	52.6 \pm 13.5
ER-	30
PR-	30
HER2-	30
Primitive TNBC	30
Histotype	
No special type (ductal)	28
Lobular carcinoma	1
Metaplastic carcinoma	1
Grading	
G2	10
G3	20
Ki67	
<50%	15
\geq 50	15
Status	
Live	25
Dead	5

Table 3. Clinicopathological assessment of the patients.

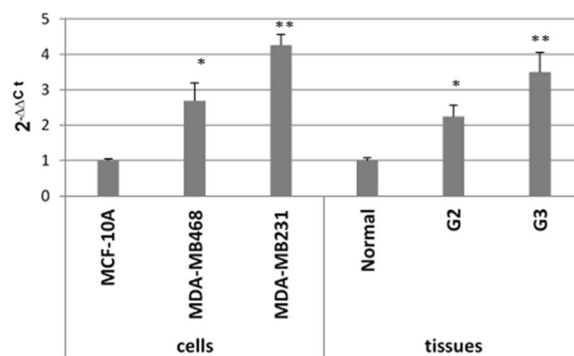


Figure 8. SEPHS2 levels in MDA-MB468 and MDA-MB231 cells compared to MCF10A cells and in TNBC tissues compared to their normal counterparts based on RT-qPCR analysis. The p-values <0.05 and <0.01 are indicated by * and **, respectively.

Therefore, it could be useful to focus research on understanding the role of this protein in cancer development and progression.

SEPHS2 involvement in cancer based on TCGA datasets. To search for information related to SEPHS2 involvement, we first analyzed data in TCGA databases related to SEPHS2 amplification and mutation frequencies in cancer with the CbioPortal tool (<http://cbioportal.org>)^{38,39}. As shown in Supplementary Table 3, higher amplification frequencies were found in breast cancer.

Considering that TNBC represents the most aggressive breast cancer subtype and that its specific molecular targets have not been well defined, we evaluated SEPHS2 expression in TNBC cell lines and patient tissues.

SEPHS2 expression in TNBC cells and tissues by RT-qPCR. SEPHS2 expression was evaluated using RT-qPCR in two TNBC cell lines (MDA-MB231 and MDA-MB468), normal breast cells (MCF-10A), and thirty TNBC tissues and their adjacent normal breast tissue counterparts (Table 3). As shown in Fig. 8, the SEPHS2 level (expressed as $2^{-\Delta\Delta C_t}$) was higher in the cell lines MDA-MB231 (p-value <0.01) and MDA-MB468 (p-value <0.05) cell lines than in the MCF-10A cells and in the MDA-MB231 cells than in the MDA-MB468 cells (Fig. 8). The difference between these two cell lines is certainly due to their different phenotypic characteristics⁴⁰. In fact, both MDA-MB231 and MDA-MB468 cells are poorly differentiated TNBC cells with mutations in p53. MDA-MB231 cells also present mutations in BRAF and KRAS, low levels of claudin-3, claudin-4, claudin-7, E-cadherin and proliferation marker Ki67, and high levels of vimentin and CD44⁺CD24⁻, which are markers associated with the epithelial-mesenchymal transition and mammary cancer stem cells (CSCs), respectively. Moreover, MDA-MB468 cells have mutations in PTEN and high levels of EGFR⁺, cytokeratin 5/6⁺ and Ki67 gene

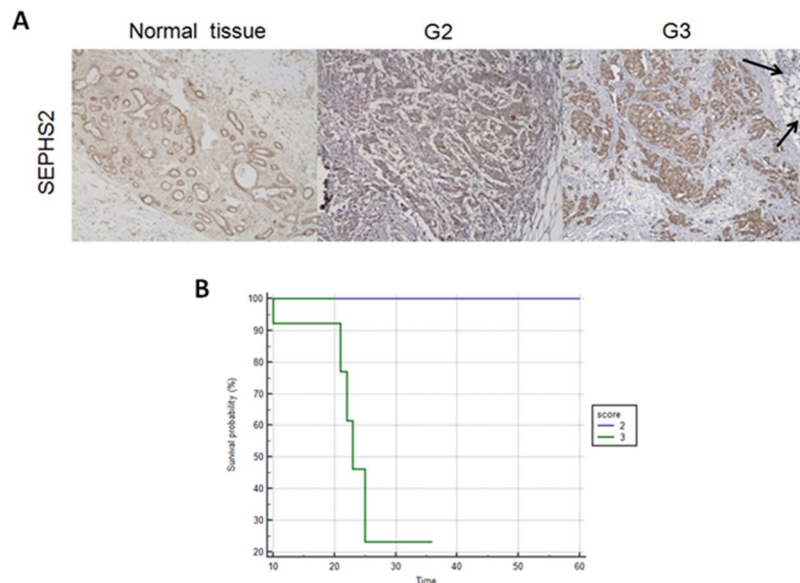


Figure 9. (A) Immunohistochemical observation of human SEPHS2 expression at 200x magnification in normal mammary tissues and grade 2 (G2) and 3 (G3) TNBC patients. In the panel related to a tissue section from a G2 patient, only the neoplastic lesion is visible. In the panel related to a tissue section from a G3 patient, two black arrows indicate a lobular ductule of a normal mammary gland immersed in the adipose stroma; moreover, clusters of neoplastic cells are found in the same stroma. (B) Kaplan–Meier curves of the overall survival of TNBC patients. Blue line: score = 2 and green line: score = 3.

expression⁴⁰. In this context, MDAMB231 cells are more invasive⁴⁰, which may explain why the SEPHS2 levels are higher in MDA-MB231 cells than in MDA-MB468 cells.

Moreover, in accordance with the cellular results, the SEPHS2 level was increased in the TNBC tissues of with grade 2 (p-value < 0.05) and grade 3 (p-value < 0.01) compared to that in their normal counterparts and, in particular, increased with the malignant grade (Fig. 8).

To confirm these data at the protein level, we subjected the collected tissues to SEPHS2 staining by immunohistochemistry (IHC).

IHC of TNBC tissues. Microscopically, a solid architecture of the TNBC tissues was observed, with little or no formation of tubular structures and a poor intercellular stroma. Some carcinomas showed medullary-like features with lymphocyte infiltration on the periphery of the neoplastic population. In other cases, cellular cords associated with necrotic areas were observed. The neoplastic cells showed a high nucleus/cytoplasm ratio, evident nucleoli and a high mitotic index. Infiltrative lobular carcinoma showed a classical and diffuse mixed growth pattern with pleomorphic cells. Metaplastic carcinoma showed giant cells and areas with squamous metaplasia.

All the TNBC samples were immunoreactive for SEPHS2, with 26 cases higher than 30% and the other 4 cases between 5 and 30%. The positivity was expressed both at the cytoplasmic and nuclear levels, with a staining intensity ranging from weak to moderate. In context, a more diffuse and moderate intensity (score = 3) was observed in poorly differentiated TNBC patients with grade 3 than in patients with grade 2, for whom weak staining (score = 2) was evident (Fig. 9A). In detail, (i) for 80% of patients (8 of 10) with grade 2, a score = 2 has been assigned (chi-square test, df = 6.316 and p-value = 0.0120); (ii) for 85% of patients (17 of 20) with grade 3, a score = 3 has been assigned (chi-square test, df = 4.787 and p-value = 0.0287). The “*in situ*” carcinomatous component was less than 25% and showed no significant differences in immunoreactivity from the infiltrating components. As shown in Fig. 9A, the healthy glandular component also expressed extremely weak cytoplasmic positivity (score = 1) at the level of myoepithelial cells.

Overall, these evaluations suggest that the gradual increase in SEPHS2 expression is associated with an increased malignant grade in TNBC patients. In fact, higher SEPHS2 levels correlate with poor overall survival (with a p-value = 0.0067) (Fig. 9B).

Conclusions

In this paper, we obtained for the first time a structural model of the SEPHS2 protein, which is an important component of the Sec biosynthesis machinery for which no three-dimensional structure has been experimentally determined. The obtained model consists of an N-terminal domain with an α -helix and a long disordered loop, a central core with an α - β 2-layer sandwich architecture and a disordered C-terminal domain. MD simulation of the SEPHS2 structure at neutral pH highlighted that the structure was stabilized through formation of π -cation interactions and H-bonds and showed that Sec. 60 remained exposed on the surface throughout the MD simulation, whereas other residues, termed HUB residues, tended to stabilize the structure. During the simulation at acidic pH, the stability and compactness of SEPHS2 were greater than those during the simulation at neutral pH,

and Sec. 60 was more exposed on the surface. Hence, considering that acidic extracellular pH is a major feature of cancer tissues³², these structural features of SEPHS2 at acidic pH can help explain how this protein functions in cancer. Therefore, we suggest the utility of further experimental studies on these identified HUB residues to understand their specific roles in dysregulation of the Sec biosynthesis machinery in cancer.

We also highlighted for the first time that the SEPHS2 levels were higher in TNBC than in normal epithelial breast tissues and that the levels increased with the malignant grade, suggesting that this protein could be a potential novel prognostic marker and/or therapeutic target. Further studies are warranted to validate SEPHS2 as a prognostic marker in larger cohorts of TNBC patient samples.

Notably, selenoproteins can be secreted and detected in the blood⁴¹. Blood plasma and its proteins are an ideal source of biomarkers, since they represent a snapshot of a patient's pathophysiological state at a given time. Indeed, so-called "liquid biopsies" allow dynamic monitoring with an insight into the process of spatial and temporal clonal evolution of the tumor, including secondary resistance to treatment, which is denied by the invasiveness of tissue biopsies. In the literature, it is reported that thioredoxin reductase 1 (TrxR1), a cytoplasmic and nuclear selenoprotein, is secreted and that its serum levels are associated with poor prognosis in non-small cell lung cancer⁴². Hence, to verify whether SEPHS2, a cytoplasmic protein, is also secreted, our laboratory will perform further studies on biological fluids.

Methods

Sequence analysis. The ProtParam tool was used to analyze the amino acid composition of the human SEPHS2 sequence (UniProt code: Q99611)¹², and Jpred¹⁴ and MetaDisorder server¹⁵ (<http://iimcb.genesisilico.pl/metadisorder/>) were used to predict the propensity to form regular secondary structure elements or to be disordered.

The charge distribution on the human SEPHS2 sequence was evaluated in agreement with Das and Pappu¹³. In particular, the FCR and NCPR were calculated using $FCR = |f_+ + f_-|$ and $NCPR = |f_+ - f_-|$, where f_+ and f_- represent the fractions of positive and negative charges, respectively. This calculation allows classification of the protein sequences in the following four regions of the state diagram: (i) Region 1 ($FCR < 0.25$ and $NCPR < 0.25$), which contains weak polyelectrolytes and polyampholytes and shows a tendency to form tadpole and globule ensembles; (ii) Region 2 ($0.25 \leq FCR \leq 0.35$ and $NCPR \leq 0.35$); (iii) Region 3 ($FCR > 0.35$ and $NCPR \leq 0.35$) which contains strong polyampholytes and has a tendency to form ensembles of hairpins, coils and chimeras; and (iv) Region 4 ($FCR > 0.35$ and $NCPR > 0.35$), which contains strong polyelectrolytes and tends to form ensembles of swollen coils¹³.

Posttranslational modifications, such as sulfation, phosphorylation and glycosylation, were predicted by the Sulfinator¹⁹, NetPhos¹⁷, and NetNGlyc and NetOGlyc²⁰ tools, respectively. We also searched for experimental phosphorylation sites using the PhosphoSitePlus server¹⁸. Finally, the binding regions in disordered proteins were predicted by the ANCHOR²¹ and α -MoRF-PredII²² tools. All these procedures were performed in accordance with the relevant guidelines and regulations.

Molecular modeling. The SEPSH2 structure was modeled using an integrated procedure based on comparative modeling and fold recognition that we described previously^{23,24}. BLAST analysis²⁵ showed that the 41–427 region of SEPSH2 had 73% sequence identity with human SEPHS1. Hence, human SEPHS1 was used as a starting template. We created ten structures using the MODELER program²⁷ and selected the best model based on the energetic and stereochemical quality. In detail, the structures were analyzed with the ProSA²⁹ and Ramachandran Plot 2.0²⁸ tools to calculate the energetic stability (Z-score) and the numbers of residues in allowed and disallowed positions in the Ramachandran plot, respectively. The best selected model was subjected to a loop refinement tool to obtain a better structure of the unstructured disordered loop regions.

The N-terminal (1–40) and C-terminal (428–448) regions were modeled by MUSTER, which is a fold recognition tool based on a sequence profile-profile alignment algorithm (PPA)³⁰.

Then, the complete 3D structure of SEPHS2 was modeled using as reference the models obtained, as reported above, for the N-terminal, C-terminal and 41–427 regions. The complete best model was chosen always by evaluating the energetic quality using the ProSA program²⁹ and the stereochemical quality using the Ramachandran plot²⁸.

The final SEPHS2 model was deposited in the ModelArchive database (<https://doi.org/10.5452/ma-y6ovo>).

MD simulations and analysis. The complete SEPHS2 model was subjected to MD simulations by the GROMACS software (v4.5.6)³¹, at neutral and acidic pH, for 10 ns at 300 K, using GROMOS43a1 as the force field. A cubic box ($88.6 \times 88.6 \times 88.6 \text{ \AA}^3$) was designed to contain the model and covered with 6775 SPC216 water molecules.

Initially, 2000 energy minimization and 25000 position restraint steps were performed to minimize the protein system and the location of the water molecules.

The final trajectories were analyzed through different routines present in GROMACS, such as RMSD, gyration radius (R_g), secondary structure evolution, RMSE, total solvent accessible area, PCA and partial densities.

For each conformer, the numbers of π -cations, H-bonds, π -stacking and IACs were calculated before MD and after 2, 4, 6, 8 and 10 ns of MD simulation using the PIC³³, HBPLUS³⁴ and COCOMAPS³⁵ tools.

Network analysis. Taking advantage of the NetworkAnalyzer and RINalyzer Cytoscape plugins, a network of interacting amino acids was created using as nodes the amino acids and as edges the π -cations, H-bonds, π -stacking and IACs to represent the interactions between them. The network topologies were analyzed by the CytoHubba plugin, and the HUB residues (the most correlated nodes) were identified. In detail, the following

statistical network parameters were evaluated: (i) betweenness centrality; (ii) degree of nodes; (iii) bottleneck; (iv) closeness centrality; (v) DMNC; (vi) MNC and (vii) MCC^{36,37}.

Cancer evaluation using TCGA datasets. To determine the mutation and amplification frequencies of SEPHS2 in different cancers, we analyzed TCGA datasets using the CbioPortal tool (<http://cbioportal.org>)^{38,39}.

Cell culture. Normal human breast epithelial cells (MCF-10A) and two TNBC cell lines (MDA-MB231 and MDA-MB468) were used in this work. MDA-MB231 and MDA-MB468 cells were grown in RPMI 1640 (Lonza) supplemented with penicillin/streptomycin (100×) (Euroclone), fetal bovine serum (10%) (Invitrogen), and non-essential amino acids (100×) (Invitrogen), and Glutamax (100×) (Invitrogen) whereas MCF-10A cells in DMEM were supplemented with 20 ng/mL human epidermal growth factor (Life Technologies), 10 µg/mL human insulin (Life Technologies Corporation), and 0.5 µg/mL of hydrocortisone (Sigma-Aldrich). In the case of MDA-MB468 cells, Ham's F-12 medium was also added to the medium (1:1 ratio). Finally, the cells were kept at 37 °C in an incubator in a humidified atmosphere (5% CO₂ and 95% air).

Tissue samples. We performed RT-qPCR and IHC on formalin-fixed paraffin-embedded (FFPE) blocks of TNBC tissues obtained by surgical resection from thirty patients. In this study, the adjacent normal breast tissue counterparts were used as tissue controls. This study was approved by the Ethics Committee of University of Campania. All the patients provided informed consent for study participation. The clinicopathological assessment of the patients is listed in Table 3. All thirty patients had primitive TNBC and did not present the progesterone and estrogen receptors, and HER2/neu genes. The tissues were obtained from 28 patients with ductal carcinoma, 1 with lobular carcinoma and 1 with metaplastic carcinoma. Moreover, 10 TNBC patients were grade 2 and 20 patients were grade 3. Five of these patients died.

RNA preparation and RT-qPCR analysis of the cells and tissues. RNeasy Mini Kit (Qiagen Inc) was used to extract total RNA from three cell lines. For the tissue samples, the paraffin was removed by xylene extraction for RNA isolation, and the Recoverall (TM) Total RNA Isolation Kit (Life Technologies-Ambion) was used to extract total RNA from tissue sections equivalent to 60 µm (three 20 µm sections). All the obtained RNA extracts were dissolved in diethyl pyrocarbonate-treated water, and their concentrations and purities were assessed by NanoDrop 2000 spectrophotometer (Thermo Scientific) at 260/280 nm of optical density. SuperScript VILO cDNA Synthesis Kit (Life Technologies Corporation) and nuclease-free water (Life Technologies-Ambion) were used to reverse-transcribe 2 µg of total RNA obtained for each sample and to dilute them, respectively.

Primer pairs for RT-qPCR were designed using the mRNA sequences deposited in the nucleotide data bank (NCBI) with an aim of obtaining a <100 bp amplicon. Oligonucleotides were obtained from Eurofins. The primer sequences were CATCCCCCTGAGGCACGG and TCACTCAGCACGTTGGCACA for SEPHS2 and TCTGGCACCACACCTTCTACAATG and AGCACAGCCTGGATAGCAAG for β-actin.

A Step-One Real Time PCR System (Applied Biosystems) was used to perform RT-qPCR experiments. An aliquot of cDNA (2 µL) was amplified in a mixture (25 µL) consisting of the reverse and forward primers (300 nM) and 2X SYBR Green PCR Master Mix (Applied Biosystems) using the following conditions: denaturation for 5 min at 95 °C, and, then, 44 cycles of denaturation for 30 s at 95 °C plus annealing for 1 min at 60 °C. Each reaction was performed in triplicate. β-Actin mRNA was used to normalize the data⁴³ and Prism software (GraphPad Software) to perform statistical analysis (a paired t test). All these procedures were performed in accordance with the relevant guidelines and regulations.

Immunohistochemistry of TNBC tissues. The sections were inserted in jars containing trisodium citrate solution (0.01 M) after alcohol rehydration and xylene dewaxing, microwaved, and, then, rinsed for 5 min in cool H₂O and for 30 min at room temperature in H₂O₂ (3%). Successively, the samples were washed in Tris-buffered saline and incubated overnight at 4 °C with an anti-SEPHS2 antibody produced in rabbits (Sigma-Aldrich) using a 1:100 dilution. After incubation, biotinylated secondary antibodies plus streptavidin (Dako) were used to stain the samples, DAB chromogen (Dako) as substrate, and hematoxylin solution for nuclear counterstaining. All these procedures were performed in accordance with the relevant guidelines and regulations. However, we used some FFPE tissue samples of human ovarian cancer as positive control to verify the reliability of anti-SEPHS2 antibody, as suggested by the Sigma-Aldrich data sheet.

Immunoreactivity was evaluated in terms of positive stained cancer cellular percentage and staining intensity as described by Sinicrope *et al.*⁴⁴. The staining intensity was scored in the following way: 1, extremely weak; 2, weak; and 3, moderate. The statistical significance was calculated by the Fisher and chi-square tests. Survival curves were evaluated and compared by log-rank test and the Kaplan–Meier method.

Received: 26 February 2019; Accepted: 22 October 2019;

Published online: 06 November 2019

References

1. Turanov, A. A. *et al.* Biosynthesis of selenocysteine, the 21st amino acid in the genetic code, and a novel pathway for cysteine biosynthesis. *Adv. Nutr.* **2**, 122–8 (2011).
2. Xu, X. M. *et al.* Selenophosphatesynthetase 2 is essential for selenoprotein biosynthesis. *Biochem. J.* **404**, 115–20 (2007).
3. Xu, X. M. *et al.* Biosynthesis of selenocysteine on its tRNA in eukaryotes. *PLoS Biol.* **5**, e4 (2007).
4. Palioura, S., Sherrer, R. L., Steitz, T. A., Söll, D. & Simonovic, M. The human SepSecs-tRNA^{Sec} complex reveals the mechanism of selenocysteine formation. *Science*. **325**, 321–325 (2009).
5. Oudouhou, F., Casu, B., Dopgwa Puemi, A. S., Sygusch, J. & Baron, C. Analysis of Novel Interactions between Components of the Selenocysteine Biosynthesis Pathway, SEPHS1, SEPHS2, SEPSECS, and SECp43. *Biochemistry*. **56**, 2261–2270 (2017).
6. Davis, C. D., Tsuji, P. A. & Milner, J. A. Selenoproteins and cancer prevention. *Annu. Rev. Nutr.* **32**, 73–95 (2012).

7. Lan, X. *et al.* Decreased Expression of Selenoproteins as a Poor Prognosticator of Gastric Cancer in Humans. *Biol. Trace Elem. Res.* **178**, 22–28 (2017).
8. Maciel-Dominguez, A., Swan, D., Ford, D. & Hesketh, J. Selenium alters miRNA profile in an intestinal cell line: evidence that miR-185 regulates expression of GPX2 and SEPSH2. *Mol. Nutr. Food Res.* **57**, 2195–205 (2013).
9. Dudaladava, V. *et al.* Gene Expression Profiling in Hereditary, BRCA1-linked Breast Cancer: Preliminary Report. *Hered. Cancer Clin. Pract.* **4**, 28–38 (2006).
10. Braicu, C. *et al.* Novel insight into triple-negative breast cancers, the emerging role of angiogenesis, and antiangiogenic therapy. *Expert Rev. Mol. Med.* **18**, e18 (2016).
11. Xu, F. *et al.* Repositioning antipsychotic fluphenazine hydrochloride for treating triple negative breast cancer with brain metastases and lung metastases. *Am. J. Cancer Res.* **9**(3), 459–478 (2019).
12. Gasteiger, E. *et al.* Protein identification and Analysis tools on the Expasy Server; in John M. Walker (ed.): *The Proteomics Protocols Handbook*. Humana Press. 571–607 (2005).
13. Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. USA* **110**, 13392–7 (2013).
14. Cuff, J. A. & Barton, G. J. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins*. **40**, 502–511 (2000).
15. Kozlowski, L. P. & Bujnicki, J. M. MetaDisorder: a meta-server for the prediction of intrinsic disorder proteins. *BMC Bioinformatics*. **13**, 111 (2012).
16. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell. Biol.* **16**, 18–29 (2015).
17. Blom, N., Gammeltoft, S. & Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **294**, 1351–62 (1999).
18. Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–20 (2014).
19. Monigatti, F., Gasteiger, E., Bairoch, A. & Jung, E. The Sulfinator: Predicting tyrosine sulfination sites in protein sequences. *Bioinformatics*. **18**, 769–70 (2002).
20. Steentoft, C. *et al.* Precision mapping of the human O-GalNAcglycoproteome through SimpleCell technology. *EMBO J.* **32**, 1478–88 (2013).
21. Meireles, L. M., Dömling, A. S. & Camacho, C. J. ANCHOR: a web server and database for analysis of protein-protein interaction binding pockets for drug discovery. *Nucleic Acids Res.* **38**, W407–11 (2010).
22. Disfani, F. M. *et al.* 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).
23. Polo, A., Colonna, G., Guariniello, S., Ciliberto, G. & Costantini, S. Deducing functional characteristics of the human selenoprotein SELK from the structural properties of its intrinsically disordered C-terminal domain. *Mol. Biosyst.* **12**, 758–72 (2016).
24. Polo, A., Guariniello, S., Colonna, G., Ciliberto, G. & Costantini, S. A study on the structural features of SELK, an over-expressed protein in hepatocellular carcinoma, by molecular dynamics simulations in a lipid-water system. *Mol. Biosyst.* **10**, 3209–22 (2016).
25. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Best local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
26. Wang, K. T., Wang, J., Li, L. F. & Su, X. D. Crystal structures of catalytic intermediates of human selenophosphatesynthetase 1. *J. Mol. Biol.* **390**, 747–59 (2009).
27. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
28. Gopalakrishnan, K., Sowmiya, G., Sheik, S. S. & Sekar, K. Ramachandran plot on the web (2.0). *Protein Pept. Lett.* **14**, 669–71 (2007).
29. Sippl, M. J. Recognition of errors in three-dimensional structure of proteins. *Proteins*. **17**, 355–62 (1993).
30. Wu, S. & Zhang, Y. MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins*. **72**, 547–556 (2008).
31. Van Der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005).
32. Kato, Y. *et al.* Acidic extracellular microenvironment and cancer. *Cancer Cell Int.* **13**, 89 (2013).
33. Tina, K. G., Bhadra, R. & Srinivasan, N. PIC: Protein Interactions Calculator. *Nucleic Acids Res.* **35**, W473–W476 (2007).
34. McDonald, I. K. & Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **283**, 777–93 (1994).
35. Vangone, A., Spinelli, R., Scarano, V., Cavallo, L. & Oliva, R. COCOMAPS: a web application to analyze and visualize contacts at the interface of biomolecular complexes. *Bioinformatics*. **27**, 2915–2916 (2011).
36. Assenov, Y., Ramirez, F., Schelhorn, S. E., Lengauer, T. & Albrecht, M. Computing topological parameters of biological networks. *Bioinformatics*. **24**, 82–284 (2008).
37. Doncheva, N. T., Klein, K., Dominiques, F. S. & Albrecht, M. Analyzing and visualizing residue networks of protein structures. *Trends Biochem. Sci.* **36**, 179–82 (2011).
38. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* **6**, pl1 (2013).
39. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov.* **2**, 401–404 (2012).
40. Holliday, D. L. & Speirs, V. Choosing the right cell line for breast cancer research. *Breast Cancer Res.* **13**, 215 (2011).
41. Raymond, F. *et al.* Selenoprotein P- Expression, Functions, and Roles in Mammals. *Biochim Biophys Acta.* **1790**, 1441–1447 (2009).
42. Chen, G. *et al.* The serum activity of thioredoxin reductases 1 (TrxR1) is correlated with the poor prognosis in EGFR wild-type and ALK negative non-small cell lung cancer. *Oncotarget.* **8**, 115270–115279 (2017).
43. Arniogiannaki, N. *et al.* BCL-2, BAX and P53 expression profiles in endometrial carcinoma as studied by real-time PCR and immunohistochemistry. *Anticancer Res.* **29**, 3977–3982 (2009).
44. Sinicrope, F. A. *et al.* Bcl-2 and p53 oncoprotein expression during colorectal tumorigenesis. *Cancer Res.* **55**, 237–241 (1995).

Acknowledgements

Susan Costantini and Francesca Capone were supported by the Italian Ministry of Health to IRCCS Istituto Nazionale Tumori “Fondazione G. Pascale”, Napoli (Italia) (progetto 5 × 1000 2016–2019). Andrea Polo and Angela Sorice were supported by the Italian Ministry of Health to IRCCS Istituto Nazionale Tumori “Fondazione G. Pascale”, Napoli (Italia) (Ricerca Corrente 2016–2019). We are grateful to Dr. Alessandra Trocino, Librarian at IRCCS “G. Pascale” of Naples, Italy, for bibliographic assistance.

Author contributions

Alfredo Budillon and Susan Costantini conceived this study and drafted the manuscript; Carmine Nunziata, Andrea Polo, Angela Sorice, Francesca Capone, Eliana Guerriero, Marina Accardo and Federica Zito Marino performed the experiments. Michele Orditura collected the TNBC samples and clinical data. All authors approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-52718-0>.

Correspondence and requests for materials should be addressed to A.B. or S.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019