ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

In Silico Identification of Potential Phosphorylation in the Cytoplasmic Domain of Epithelial Cell Adhesion Molecule

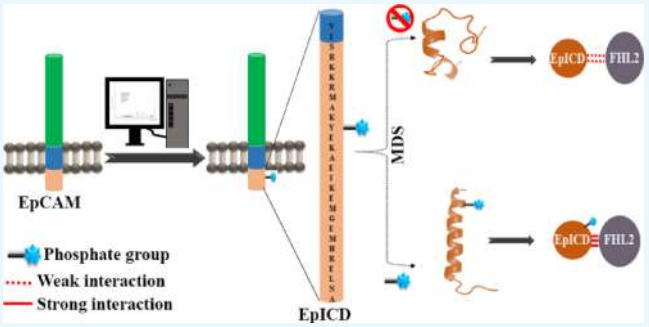
Arijit Mal, Pranay Dey, Robert Michael Hayes, Justin V. McCarthy, Arjun Ray, and Abhijit De*

Cite This: *ACS Omega* 2020, 5, 30808−30816

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The epithelial cell adhesion molecule (EpCAM) is a transmembrane cell adhesion glycoprotein, which primarily contributes to stemness, proliferation, and metastasis properties of tumor cells. Regulated intramembrane proteolysis by ADAM proteases and γ-secretase cleaves EpCAM into an ∼27 kDa soluble extracellular and an ∼4 kDa cytoplasmic domain (EpICD). After the EpICD fragment is released inside the cell, the formation of a nuclear signaling complex with the FHL2 molecule is critical for exerting its regulatory role. Trop-2, a homologous protein of EpCAM, undergoes phosphorylation in its cytoplasmic domain (Trop-IC). The phosphorylation of Trop-2 is reported to be crucial for its function. This led us to ask the fundamental question if EpCAM does undergo similar post-translational modification(PTM) like its homologous protein to carry out its diverse biological function. Here, we identify a putative phosphorylation site at Tyr297 located in the cytoplasmic domain of EpCAM. Molecular dynamic simulation (MDS) of 90 ns was carried out to understand the biological/functional relevance of the putative phosphorylation. It was observed that this phosphorylation stabilizes the α-helical structure of the EpICD. Though Tyr297 does not affect the γ-secretase mediated cleavage of EpCAM, it affects the binding of EpICD to FHL2. Docking analysis revealed that phosphorylation mediated structural stability of EpICD positively impacts its binding affinity with FHL2, which was further validated using 100 ns MDS. Phosphorylated EpICD forms higher numbers of hydrogen bonds, salt bridges, and other non-bonded interactions with FHL2, leading to enhanced interactions. This *in silico* study reveals a potential PTM in the EpICD, providing the basis for future research in understanding the mechanism behind the diverse biological function of EpCAM.



INTRODUCTION

The epithelial cell adhesion molecule (EpCAM) is a Ca^{2+} independent homotypic cell adhesion molecule.[1] Initially, it was discovered as a dominant antigen on colon carcinomas.[2] Extensive research has provided valuable insights into the role of EpCAM in promoting oncogenesis and linking it to poor prognosis in various cancer types.[3] It is considered as a target molecule for many immunotherapeutic approaches.[4] Recently, EpCAM was identified as a surface marker of cancer stem cells. EpCAM can also induce stemness in cancer cells by regulating the PTEN/AKT/mTOR signaling pathway.[5]

Human EpCAM is a type-1 transmembrane protein consisting of a large extracellular domain (EpEX), a single-spanning transmembrane domain (EpTM), and a very short cytoplasmic domain (EpICD). EpCAM signaling initiates at the plasma membrane where claudin7 recruits and transfers EpCAM to tetraspanin enriched microdomain (TEM).[6] Within TEM, initial cleavage of EpCAM by ADAM proteases leads to the release of EpEX, leaving the C-terminal fragment (CTF) anchored within the membrane.[7] The CTF is cleaved at multiple sites by the intramembrane γ-secretase protease to release the EpICD.[8] Once released, EpICD first interacts with FHL2 in the cytoplasm, which then moves to the nucleus after forming a complex with β-catenin.[7,9] In the nucleus, this EpICD-containing complex binds to LEF-1, which further acts as a transcription factor for genes like *c-MYC*, *cyclin D1*, *A*, and *E*.[10,11] Therefore, on one hand, the EpICD domain of EpCAM is known to be involved in cell adhesion, whereby it interacts with actinin. On the other hand, it is engaged in a transcription factor complex regulating cell proliferation, stemness properties, etc.[4] Thus, EpICD function is pivotal in the involvement of EpCAM in multiple cellular processes. Intriguingly, EpICD has also been reported to be a poor prognostic marker in hepatocellular carcinoma and breast cancer.[12,13]

It is well known that post-translational modification (PTM) plays a crucial regulatory role in determining protein−protein interactions, protein trafficking, protein structural conformity, and function.[14] EpCAM is known to undergo glycosylation

Received: May 7, 2020
Accepted: October 20, 2020
Published: November 23, 2020

https://dx.doi.org/10.1021/acsomega.0c02113
ACS Omega 2020, 5, 30808−30816

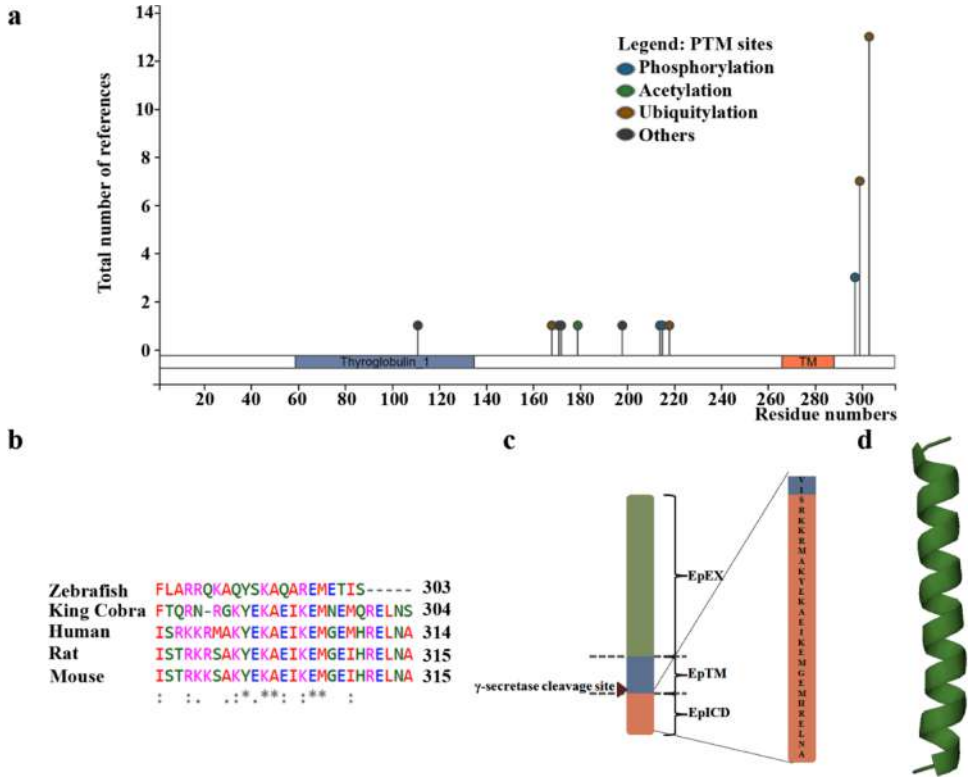


Figure 1. *In silico* phosphorylation of the EpCAM and EpICD. (a) Putative PTM of EpCAM is determined by PhosphoSitePlus. (b) Part of the multiple sequence alignment of EpCAM protein (mostly the EpICD domain) from different species. (c) Schematic representation of different domains of EpCAM. (d) Cartoon of the EpICD structure modeled with I-TASSER.

and proteolysis.[4] Considering the multifaceted role of EpCAM, identification of other PTM sites of EpCAM, specifically within the EpICD domain, is essential and may shed light on the divergent functionality of EpCAM. A recent report demonstrated that the EpCAM homologous protein Trop-2, with similar functionality, is phosphorylated in Trop-IC, enabling conformational switching.[15,16] In light of this, the study aimed to employ an *in silico* analysis for prediction of novel EpCAM PTM sites. Further, we checked the biological/functional relevance of the PTM based on the impact of the putative PTM on conformational dynamics and functionality of the human EpICD.

RESULTS AND DISCUSSION

***In silico* Phosphorylation of EpCAM at Y297.** Our analysis using PhosphoSitePlus, identified several putative PTM sites on EpCAM (Figure 1a). We observed putative ubiquitination sites at lysine 299 and 303 in the EpICD, which might be crucial for already reported proteasome-mediated degradation of EpICD. In this study, we focused on identifying novel PTMs, which might be important for EpCAM diverse biological functions. So we took interest in another potential PTM, that is, phosphorylation at tyrosine 297(Y297) in human EpCAM. This is recorded in three high-throughput studies where the site was identified by mass spectrometry. We observed that Y297 is conserved among zebrafish, king cobra, mouse, rat, and human (Figure 1b), indicating the importance of this amino acid for EpCAM protein function.

After ectodomain shedding, the remaining membrane-bound EpCAM CTF is cleaved at multiple sites by γ-secretase.[8] The shortest cytoplasmic domain of EpCAM generated by γ-secretase is only 28 amino acids in length (Figure 1c). EpICD is a key player in EpCAM signaling and the predicted phosphorylation site (Y297) is within the EpICD region.Thus, we reevaluated the tyrosine phosphorylation in EpICD using both kinase-specific (NetPhos server 3.1 and GPS 5.0) and non-kinase specific prediction tools (PhosphoSVM) (Figure S1a−c). Moreover, the presence of similar phosphorylation sites in multiple other proteins (Figure S1d) affirms the biological and functional relevance of this site.

Next, the EpICD structure was modeled using I-TASSER (Figure 1d). The model structures were generated using a variation of the 10 template structure (PDB6mzcE, PDB4pnbA, PDB6nvqB, PDB5lj3S, PDB4uotA, PDB6qbyB, PDB2qdqA, PDB5o5jB, PDB6jzcC, and PDB3iygQ) (Table S1). I-TASSER uses combinations of comparative modeling, threading, and *ab initio* modeling to model structures.[17] I-TASSER is reported to be very effective in *ab initio* modeling of small proteins (<90 residues as well as <120 residues).[18] In concordance to this, I-TASSER can predict a good quality EpICD structure, which is of only 28 amino acids long. The chances of predicting an EpICD structure with correct folds and good resolution is higher as the protein of our interest is small. Further, high *C*-score value (−0.15) and Tm score (0.69 ± 1.2) and low RMSD (1.9 ± 1.6 Å) increase our confidence in the modeled EpICD structure (Figure S1e). *C*-score determines the quality of the structure. *C*-score is generally with a range of −5 to 2. If the *C*-score is >−1.5, then the modeled structure is of correct topology, and both the false positive and false negative rate is below 0.1.[17] Tm score and RMSD are a measure of a structural similarity between the predicted structure and native structure. Here, the estimated Tm score and RMSD denote the estimated accuracy of the model.[17] A Tm score >0.5 denotes two structures with the

https://dx.doi.org/10.1021/acsomega.0c02113
ACS Omega 2020, 5, 30808−30816

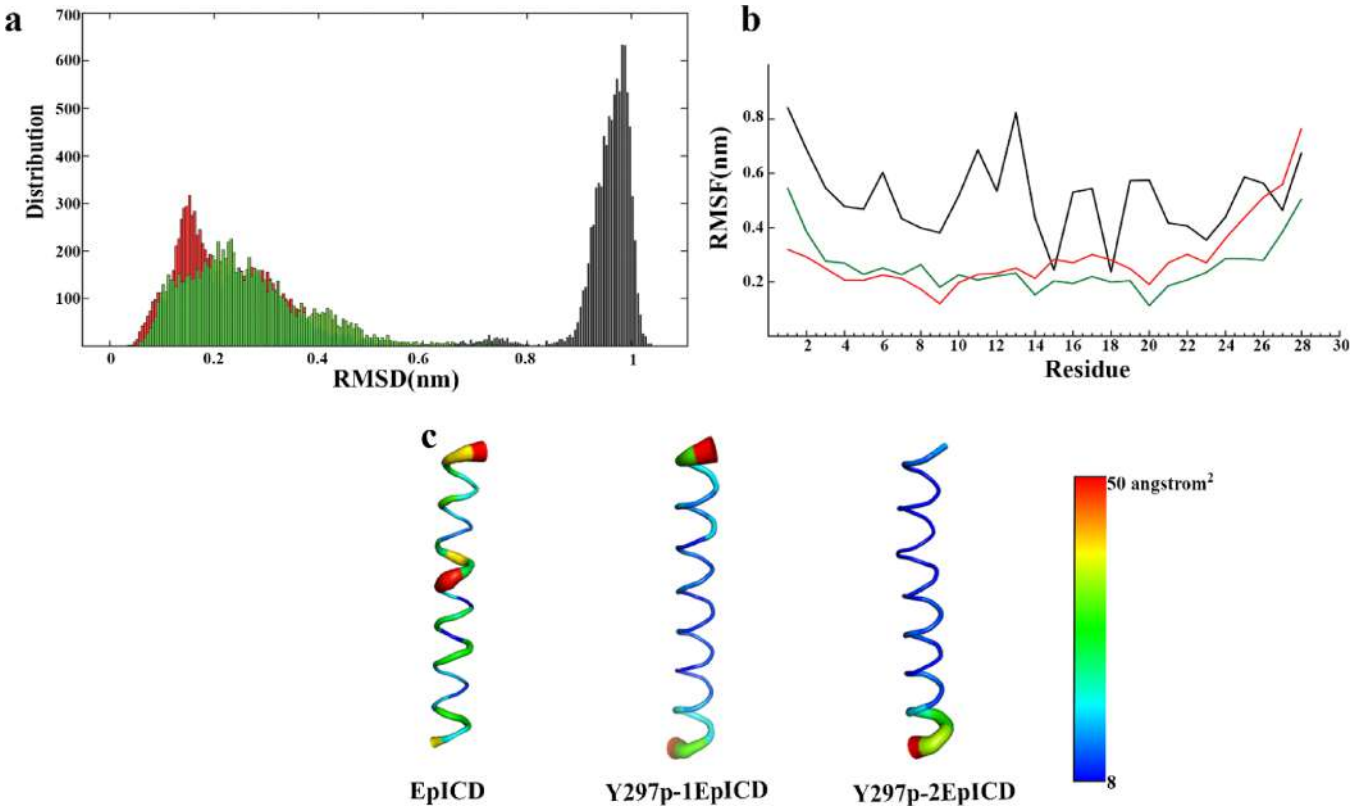


Figure 2. MDS of phosphorylated and un-phosphorylated EpICD for 90 ns each, showing the change in structural conformation with time. (a, b) RMSD and RMSF profiles of EpICD (black), Y297p-1EpICD (red), and Y297p-2EpICD (green). (c) B-factor putty representation of the structure obtained from the MD trajectory.

same folds, not a random similarity. Further, a combination of the Tm score >0.5 and *C*-score above >−1.5 denotes that the model has a false positive rate of 0.05 and a false negative rate of 0.09.[19] As the corresponding value of our modeled structure is above the cutoff, the modeled structure can be assumed to have the correct fold. This was further reflected in the I-TASSER simulation where all simulations converge to give only two clusters, an indication of a good quality model.

The Ramachandran plot of the modeled structure shows that 98% of the residues are in the favorable region, while 2% is in the allowed region and 0% is in the disallowed region (Figure S1f), providing the confidence on the predicted structure. In addition to the Ramachandran plot, we carried out quality assessment of the modeled structure using ERRAT, Modfold, and ProSA. ERRAT determines the correctness of the modeled protein structure based on the characteristic atomic interaction.[20] The ERRAT analysis of the modeled EpICD structure shows that all the residues are well below the cutoff error value for rejection, i.e., 95%. Overall quality factor (the percentage of the residues that are below 95% error value) of the EpICD model is 100 (Figure S1g). This indicates a high-resolution structure as it is reported that a structure of good high resolution generally has a value of 95 or higher.[20] Further, we assessed the quality of the modeled structure using the ModFOLD6 server. ModFOLD6 predicts the local and global quality of a modeled structure. According to CASP12 evaluation, it is one of the best in determining good models.[21] According to ModFOLD6, the EpICD model has a global model quality score of 0.8279 and a *p*-value of 6.034E-11 (Figure S1h). Global model quality score has a range of 0−1, where a higher score indicates complete and confident models, which are very similar to the native structure. Further, a *p*-value < 0.001 means that the chances of the model to be incorrect are very rare. We also found that the predicted distance of most residues from the native structure is below 2 Å (Figure S1i). These findings taken together indicate that the EpICD structure that has been modeled is of good quality with high reliability. We also find that the *Z*-score identified using ProSA[22] of our modeled structure is within the range of that of experimentally determined protein structures of the same size (Figure S1j). EpICD has two principle secondary structures, coils, and a right-handed α-helix. The coils are at the first three residues (V287, I288, and S289) and the last two residues (N313 and A314) (Figure S1k).

Effect of Phosphorylation on the Conformation of the EpICD Structure. As the addition of a phosphate group affects the structural conformation of proteins, phosphorylated and non-phosphorylated forms of the proteins can carry out different biological processes based on physiological conditions. Hence, the requirement for the protein to get phosphorylated depends on the physiological context. EpICD is involved in multiple biological processes at different physiological conditions. Thus, the likelihood of EpICD to get phosphorylated depends on the physiological condition where the addition of the phosphate group have some effect on the structural conformation, allowing it to perform a different biological role. Addition of phosphate might have a stabilizing or destabilising effect on the EpICD structure. Thus, to gain insight into the potential change in structural conformity of the EpICD domain upon phosphorylation at Y297, we conducted MDS for 90 ns. As the charge of phosphotyrosine in EpICD is not known, we considered both −1 and −2 charges of

https://dx.doi.org/10.1021/acsomega.0c02113
ACS Omega 2020, 5, 30808−30816

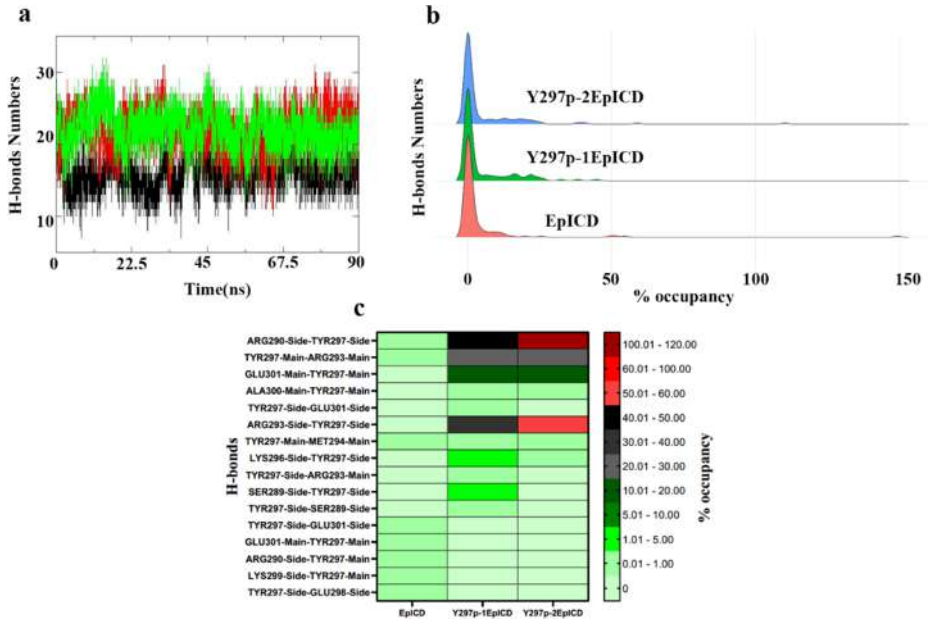


Figure 3. Intramolecular H-bond formation in the MD trajectory. (a) Number of H-bonds profile of EpICD (black), Y297p-1EpICD (red), and Y297p-2EpICD (green). (b) Ridgeline plot showing the frequency distribution of percent occupancy of H-bonds. (c) Heatmap is showing the percent occupancy of H-bonds formed by Tyr297.

phosphotyrosine for our study. MD trajectory was overall stable, which was evident from the unchanging total energy level (Figure S2a). To assess the structural stability of our protein, we calculated the RMSD plot of the Cα backbone of the proteins (Figure 2a). We observed that unphosphorylated EpICD (hereafter referred to as EpICD) undergoes a drastic conformational change.The RMSD value for EpICD increased gradually to 1.1 nM till 14 ns and then got stabilized, while both the form of phosphorylated EpICD have a low RMSD value and undergo a minimal conformational change. The phosphotyrosine having −1 charge (Y297p-1EpICD) and phosphotyrosine having −2 charge (Y297p-2EpICD) have an RMSD value in the range ∼0.05−0.55 nm and ∼0.28−0.86 nm, respectively. In line with RMSD, the RMSF plot also shows a greater conformational variation in EpICD (Figure 2b,c). High fluctuation has been observed in the N-terminus (0−4aa residue) and C-terminus (24−28aa residue) of Y297p-1EpICD and Y297p-2EpICD (Figure 2b,c). Otherwise, the structure of phosphorylated EpICD is stable. We observed that the flexibility at the Y297 position gets attenuated on phosphorylation. Low RMSD and RFSM values of phosphorylated EpICD indicate that phosphorylation reduces the conformational switching and thus stabilizing the EpICD structure. Finally, the stability of all the structures that were obtained at the end of the MDS was calculated using FoldX. The stability is predicted based on the free energy (ΔG), with lower free energy, indicating higher stability. The predicted change in the Gibbs free energy ($\Delta\Delta G$) of the Y297p-1EpICD (−10.1 kcal/mol) and Y297p-2EpICD (−3.46 kcal/mol) further validates the stabilizing effect of phosphorylation on the EpICD structure. Taken together, these findings indicate that phosphorylation might be energetically favorable for EpICD.

We have observed that there is relatively no change in the radius of gyration (R_g) of both the phosphorylated form of EpICD, whereas the R_g of EpICD changed at 14 ns (Figure S2b). Further, we observed that the solvent accessible surface area (SASA) plot in concordance with the R_g plot. SASA plot shows that EpICD has a lower surface area that is accessible to solvents for interaction in comparison to its phosphorylated form (Figure S2c).

Different non-bonded interactions like hydrogen bonds (H-bonds), hydrophobic interactions, salt bridges, etc. play a crucial role in protein structural stability and biological functions. We examined the formation of intramolecular H-bonds over the 90 ns simulation. We found that Y297p-1EpICD and Y297p-2EpICD formed a higher number of H-bonds, which also persisted longer in comparison to EpICD (Figure 3a,b). Further, it was observed that Tyr297 formed H-bonds of higher percentage occupancy with Arg290 and Arg293 in both Y297p-1EpICD and Y297p-2EpICD while these H-bonds are absent in EpICD (Figure 3c). The phosphate oxygen is known to interact with the nearby arginine and lysine side chains.[23] These interactions of the phosphoryl group with the nearby arginine might be the probable reason for stabilizing the α-helix, and in its absence, the structure gets disrupted. Furthermore, our finding is in concordance with other reports showing the role of phosphorylation in the stabilization of α-helix.[24]

Trop-2 is reported to be phosphorylated at Ser303. But, the similar site is absent in EpCAM. However, the plausible phosphorylation site of EpCAM Tyr297 is found conserved between these two proteins. Interestingly, the equivalent tyrosine site (Tyr306) in Trop-2 was not predicted to be phosphorylated, perhaps due to the steric hindrance of Ser303 and Tyr306 side chains.[15] However, in EpCAM, the absence of potential phosphorylation sites in the vicinity might make phosphorylation of Tyr297 thermodynamically feasible. Nonetheless, phosphorylation affects the structural conformity in both proteins. The NMR structure of the phosphorylated-(PDB2MVK) and non-phosphorylated (PDB2MVL) Trop-IC reveals that phosphorylation mediated salt bridges led to ordered C-terminus and reduced the tilt angle to the main α-helix of the N-terminus.[15] While in EpICD, we observed that

https://dx.doi.org/10.1021/acsomega.0c02113
ACS Omega 2020, 5, 30808−30816

phosphorylation stabilizes the α-helix by forming H-bonds (Figure S3a−e). Like Trop-2 and EpCAM, another γ-secretase substrate Notch also gets phosphorylated at its cytoplasmic domain.[25]

Phosphorylation of EpICD Leads to Increased Binding with FHL2. Further, we wanted to investigate if the phosphorylation has any effect on the EpICD function. Consequently, we generated a phospho-mutant (EpCAM Y297A) to determine whether disruption of Y297 affected γ-secretase mediated cleavage of EpCAM and subsequent generation of the EpICD. HEK293T cells were transiently transfected with EpCAM or EpCAM Y297 mutant (both tagged with mOrange) (Figure 4). Stimulation of EpCAM

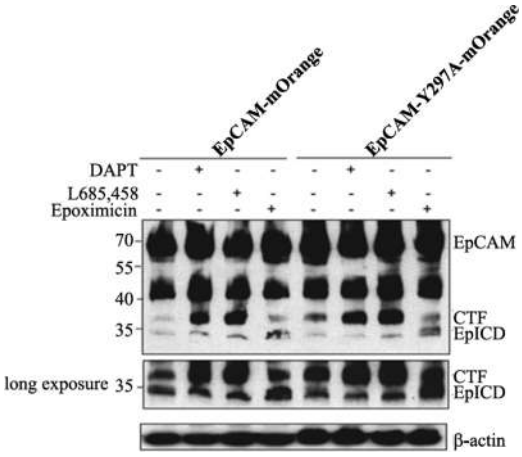


Figure 4. Effect of EpICD phosphorylation on γ-secretase cleavage. Wildtype EpCAM was tagged with a C-terminal mOrange tag (EpCAM-mOrange) and a point mutation from tyrosine to alanine at 297 was generated (EpCAM-Y297A-mOrange). HEK293T cells overexpressing wildtype and EpCAM-Y297A were left untreated or pretreated with the γ-secretase inhibitors DAPT (1 μM for 16 h) or L-685,458 (1 μM for 16 h) and/or epoxomicin (Epox) (1 μM for 2 h) and subsequently stimulated with PMA (200 nM for 2 h) as indicated, followed by Western blot analysis with an anti-EpCAM C-terminal specific antibody.

expressing cells with the phorbol 12-myristate 13-acetate (PMA), induced production of EpCAM CTF, and EpICD (lane 1). In cells transfected with EpCAM and pretreated with the γ-secretase inhibitor DAPT(1 μM/mL; 16 h) or L-685,458(1 μM/mL; 16 h), accumulation of EpCAM CTF was visibly increased, while in cells treated with the proteasome inhibitor, epoxomicin, accumulation of EpICD was evident. Importantly, in cells transfected with EpCAM Y297A, the production of EpCAM CTF and EpICD was evident, and cells had a similar response to treatment with the γ-secretase inhibitors or epoxomicin. This data implies that the Y297 site is not essential for the γ-secretase cleavage of EpCAM.

After that, we investigated the interaction of EpICD with FHL2, which is an important step of EpCAM nuclear signaling. By the yeast two-hybrid system, FHL2 was identified as the interacting partner of EpCAM. Subsequently, by co-immunoprecipitation assay it has been shown that the 4th LIM domain is crucial for FHL2 interaction EpCAM.[7] Different regions were deleted from FHL2 and were investigated for its binding to EpCAM. It was observed that if the 4th LIM domain is deleted, FHL2 cannot interact with EpICD. This indicates that the 4th LIM domain is crucial for EpICD-FHL2 interaction. Thus, we consider only this domain in our study for interaction with EpICD because the importance of this domain was experimentally confirmed.[7] Therefore, we docked the 4th LIM domain of FHL2 with the modeled structures of EpICD, Y297p-1EpICD, or Y297p-2EpICD all of which were obtained at the end of the MD trajectory. However, due to lack of experimental evidence on interaction of the 4th LIM domain and EpICD, we used CPORT for *in silico* identification of the interacting interface of the FHL2 4th LIM domain and EpICD. We used these putative interacting interfaces as restraints to carry out guided docking using HADDOCK. CPORT is a consensus method to predict interacting interfaces using a combination of interface prediction software like WHISCY, PIER, ProMate, cons-PPISP, SPPIDER, and PINUP. CPORT identified the residues, which might be actively involved in the interaction and also the residues, which might help in the interaction.[26] We observed that in comparison to the EpICD-FHL2 complex (docking score is −55.6 ± 7.2), both Y297p-1EpICD-FHL2 and Y297p-2EpICD-FHL2 complex showed lower docking score (−88.3 ± 13.2 and −78.5 ± 11.7, respectively)(Table 1). This indicates that FHL2 has a higher binding preference toward the phosphorylated EpICD than the non-phosphorylated counterpart. One of the reasons behind the higher binding affinity is the lower electrostatic energy of the phosphorylated EpICD (Table 1). Further, binding energy of the complexes (Figure S5) was calculated using FoldX. Lower binding energy of the Y297p-1EpICD-FHL2 complex (−8.00 kcal/mol) and the Y297p-2EpICD-FHL2 complex (−9.72 kcal/mol) in comparison to the EpICD-FHL2 complex (−5.29 kcal/mol) indicates stronger binding affinity of FHL2 with the phosphorylated EpICD.

Further, we observed that Y297p-1EpICD and Y297p-2EpICD interact with FHL2 by forming a higher number of H-bonds, non-bonded interactions, and salt bridges than that formed by EpICD (Figure 5a and Figure S4). We also found that the phosphoryl group at Y297 residue in EpICD is also involved in the H-bond and other non-bonded interactions with FHL2 residues (Figure S4−S5). Even higher numbers of EpICD residues are involved in complex formation with FHL2 after phosphorylation(Figure S4−S5).

To check the stability of these complexes, we carried out MDS for 100 ns. We observed that the Y297p-2EpICD-FHL2 complex is the most stable, which is evident from the RMSD

Table 1. Haddock Docking Profile of Binding of EpICD, Y297p-1EpICD, and Y297p-2EpICD with FHL2

docked complex	HADDOCK score	van der Waals energy (kcal/mol)	electrostatic energy (kcal/mol)	desolvation energy (kcal/mol)	buried surface area (Å^2)
EpICD-FHL2	−55.6 ± 7.2	−61.4 ± 5.7	−156.5 ± 39.8	−5.7 ± 3.7	1624.4 ± 45.4
Y297p-1EpICD-FHL2	−88.3 ± 13.2	−43.1 ± 9.0	−592.1 ± 66.6	19.5 ± 5.7	1712.9 ± 115.9
Y297p-2EpICD-FHL2	−78.5 ± 11.7	−61.8 ± 3.6	−291.6± 35.8	5.6 ± 1.9	1913.2 ± 131.0

https://dx.doi.org/10.1021/acsomega.0c02113
ACS Omega 2020, 5, 30808−30816

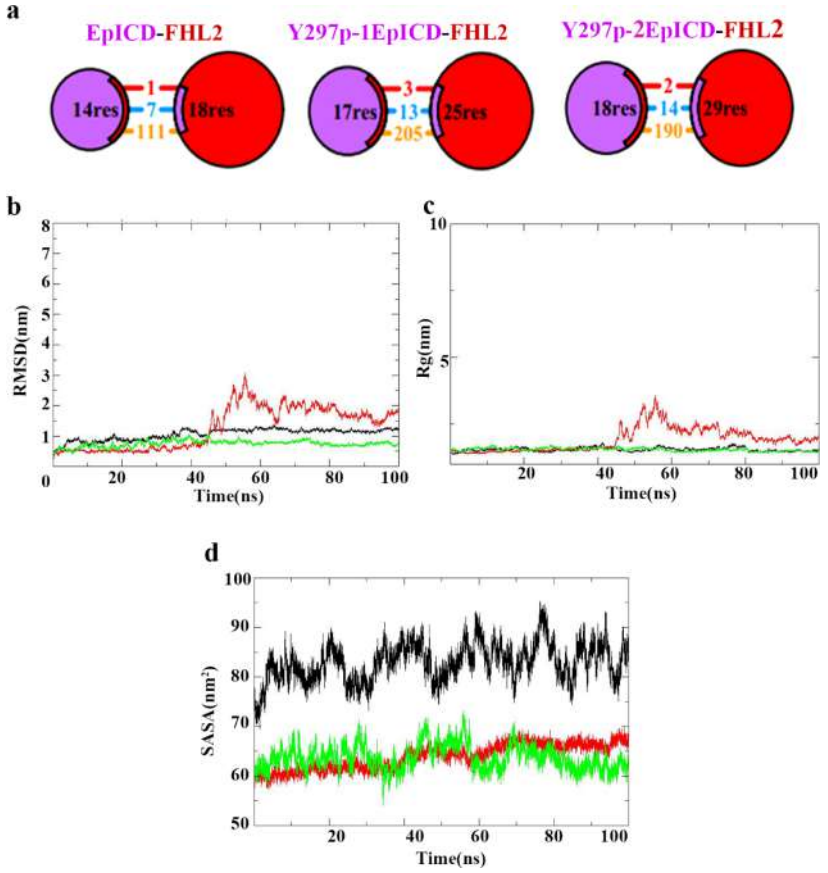


Figure 5. Binding affinity of phosphorylated and unphosphorylated EpICD toward FHL2. (a) Schematic diagrams are showing the number of residues (number inside the circle) of EpICD (purple) and FHL2 (red) interacting with each other. The mode of interaction as well as the number interaction, is also denoted. Red, indicates salt bridge; blue, indicates H-bonds; and orange, indicates non-bonded interactions. (b–d) RMSD, R_g, and SASA plots of 100 ns MD trajectory of EpICD-FHL2 (black), Y297p-1EpICD-FHL2 (red), and Y297p-2EpICD-FHL2 (green) complexes.

plot, which is stable throughout the MD trajectory (Figure 5b). The Y297p-1EpICD-FHL2 complex has a stable RMSD till ∼42 ns after which the RMSD increased to 2−3 nm and remained in that range for the rest of the trajectory. As the Y297p-1EpICD-FHL2 complex has the lowest HADDOCK score and also has low binding energy in comparison to the EpICD-FHL2 complex, we expected it to be a stable complex. So when around 90 ns, the RMSD graph showed a small dip, we extended the simulation for an additional 10 ns to check whether the RMSD falls further, and the Y297p-1EpICD-FHL2 complex comes close to its initial conformation. But, the RMSD value did not decrease any further. The EpICD-FHL2 complex at around 4 ns undergoes conformational changes, as evidenced by the increase in RMSD value to ∼1 nm (Figure 5b). The R_g values are in accordance with the RMSD values of the complexes (Figure 5c). Interestingly, Y297p-1EpICD-FHL2 and Y297p-2EpICD-FHL2 complexes have a lower SASA (Figure 5d). This is in line with the desolvation energy obtained from HADDOCK. An increase in the buried surface area probably led to a decrease in the area of the complexes accessible to solvents.

Therefore, this comprehensive *in silico* study brings in first-hand evidence of putative EpICD phosphorylation, which might be crucial for the diverse cellular function of EpCAM. We found the phosphorylation at Y297 position as stabilizing effect on the EpICD structure, which further increases the binding affinity of EpICD with FHL2. As the phosphorylation of EpICD can be dependent on the different physiological conditions, these results provide a strong incentive for future biochemical validation of the phosphorylation and exploring its role in other EpCAM mediated biological function.

MATERIALS AND METHODS

Accesion ID. Human EpCAM UniProt KB: P16422. EpCAM sequence was retrieved from NCBI protein database (Human EpCAM-Accession: NP_002345.2, Rat EpCAM-Accession: NP_612550.1, Zebrafish EpCAM-Accession: NP_001017593.1, Mouse EpCAM-Accession: NP_032558.2 and King Cobra EpCAM-Accession: ETE70163.1). The structure of FHL2 (PDB ID: 1X4L) (UniProtKB: Q14192) was obtained from RCSB protein data bank.[27]

Modeling of Structure, Phosphorylation Prediction, and Multiple Sequence Alignment. EpICD structure was modeled using I-TASSER.[28] The model quality was analyzed using ERRAT,[20] ModFOLD6, and ProSA. The secondary structure of EpICD was also obtained from I-TASSER. Phosphorylation was predicted using PhsophoSitePlus,[29] where the phosphorylation are manually curated from experimental studies and from high-throughput mass spectrometry studies. Other phosphorylation prediction resources used, NetPhos 3.1 server,[30] PhosphoSVM,[31] and PhosphonetKinexus (www.phosphonet.ca). Multiple sequence alignment was done using EMBL-EBI Clustal Omega. Validation of the EpICD model was done using the Ramachandran plot generated by PROCHECK.

https://dx.doi.org/10.1021/acsomega.0c02113
ACS Omega 2020, 5, 30808−30816

Molecular Dynamic Simulation (MDS). Phosphorylation of this model was done using the Vienna PTM tool.[32] All the three modeled structures (EpICD, Y297p-1EpICD, and Y297p-2EpICD) and their docked counterparts (EpICD-FHL2, Y297p-1EpICD-FHL2, and Y297p-2EpICD-FHL2 complex) were subjected to MDS studies using GROMACS 2018.1 with the implementation of CHARMM22 force field. Solvation of the system was done using the TIP3P water model in a cubic box with periodic boundary conditions. Required counter ions were added to neutralize the system before starting the run. The three systems were initially energy minimized using the steepest descent algorithm with a tolerance of 1000 kJ/mol/nm. The system was subsequently equilibrated by employing positional restraints on the structure using NVT and NPT ensemble for 100 ps each. The temperature of 300 K was coupled by a Berendsen thermostat with the pressure of 1 bar using the SHAKE algorithm. The three equilibrated systems were then subjected to 50 ns of the production run with time-step integration of 2 fs. Trajectories of the simulation were saved at every 2 ps and were analyzed using GROMACS 2018.1. The root mean square fluctuations (RMSF), root mean square deviation (RMSD), hydrogen bonds, radius of gyration (R_g), and solvent accessible surface area (SASA) were analyzed.

Docking Study. Molecular docking studies of EpICD, Y297p-1EpICD, and Y297p-2EpICD were performed using the HADDOCK server.[33] The structure of the FHL2 (PDB1X4L) was downloaded from the protein data bank. The residues involved in the interaction between FHL2 and EpICD and its phospho forms were identified using CPORT.[26] The chain assignment of both proteins was edited by the PDB editor. Prior to docking, all co-factors atoms were removed from FHL2. The best cluster showing the lowest HADDOCK score was selected. These docked structures were then subjected to PDBsum analysis to gain information about the residues involved in interactions.[34]

Energy Calculation. Gibbs free energy(ΔG) of the proteins and binding energy of protein complexes is calculated using FoldX.[35] The difference in the change in Gibbs free energy ($\Delta\Delta G$) between the final state (phosphorylated) and reference (wildtype) denotes the stability of the protein on the introduction of the modification. $\Delta\Delta G < 0$ indicates stability. This stability analyzed by FoldX is known to be correlated with experimentally determined stability.

Reagents and Antibodies. γ-secretase inhibitors (GSI) DAPT and L-685,458, PMA, and proteasomal inhibitor epoxomicin were purchased from Calbiochem. Antibodies used in this study, -anti-EpCAM (EpICD) antibody [E144] (ab32392) and enhanced chemiluminescence (ECL) secondary antibodies, goat anti-rabbit HRP (IgG H&L) and rabbit anti-mouse HRP (IgG H&L) were purchased from AbCAM (Dublin, Ireland). Anti-β-actin antibody was purchased from Sigma Aldrich (Dublin, Ireland). Primary antibodies were used at 1:1000 dilution in 5% non-fat milk in TBS-T. Secondary antibodies were used at a 1:10,000 dilution in 5% non-fat milk in TBS-T.

Transfections. Transient transfections of HEK293T were performed using the calcium phosphate precipitation method.

Western Blotting. Cell cultures were treated with DAPT (1 μM; 16 h) or L-685,458 (1 μM; 16 h) and/or with epoxomicin (1 μM; 2 h) before harvesting. For cells treated with both GSI and epoxomicin, cells were pretreated with GSI for 14 h, and culture media was supplemented with epoxomicin for the final 2 h of treatment. After the indicated treatments, total protein extracts were obtained from cells. Cells were washed in ice cold PBS and lysed on ice for 30 min with RIPA lysis buffer (200 μL/ 60 cm plate) freshly supplemented with 1 mM sodium orthovanadate and protease inhibitor mixture (Complete, Molecular Biochemicals) on ice. If lysates appeared too viscous, the lysates were syringed through a fine hypodermic needle 10 times to shear DNA. Lysates were centrifuged (13,200 rpm for 20 min at 4 °C), and the supernatants were collected, and protein yield was quantified using a bicinchoninic acid assay (BCA) (Pierce). Normalized samples were prepared with Laemmli sample buffer containing β-mercaptoethanol and electrophoresed on an SDS-polyacrylamide gel. Proteins were transferred to nitrocellulose membrane (Millipore). After blocking for 1 h with 5% non-fat milk in TBS-T (tris-buffered saline containing 0.1% Tween 20), the membrane were probed with the primary antibody -anti-EpCAM (EpICD) antibody[E144] (ab32392) (1 h at RT or overnight at 4 °C), washed three times in TBS-T, followed by the secondary antibody diluted in in 5% non-fat milk in TBS-T. Immunoreactivity was visualized by the Odyssey imaging system (Li-COR Biosciences) or by the ECL Western blotting detection system (GE Healthcare). Signal intensity was analyzed within a linear range using ImageJ (NIH, Bethesda, MD).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.0c02113.

> Putative phosphorylation site of EpCAM, description of threading templates and quality assessment of the modeled EpICD structure, molecular dynamic simulation profile of phosphorylated and unphosphorylated EpICD, cartoon of the phosphorylated and unphosphorylated modeled EpICD structure after simulation, and description of the docked complexes between FHL2 and phosphorylated/unphosphorylated EpICD (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Abhijit De − *Molecular Functional Imaging Laboratory, ACTREC, Navi Mumbai 410210, India; Life Science, Homi Bhaba National Institute, Mumbai 400094, India;* orcid.org/0000-0002-5818-0206; Phone: +91-22-2740 5038; Email: ade@actrec.gov.in; Fax: +91-22-2740 5085

Authors

Arijit Mal − *Molecular Functional Imaging Laboratory, ACTREC, Navi Mumbai 410210, India; Life Science, Homi Bhaba National Institute, Mumbai 400094, India*

Pranay Dey − *Molecular Functional Imaging Laboratory, ACTREC, Navi Mumbai 410210, India; Life Science, Homi Bhaba National Institute, Mumbai 400094, India*

Robert Michael Hayes − *Signal Transduction Laboratory, School of Biochemistry & Cell Biology, University College Cork, Cork T12 K8AF, Ireland*

Justin V. McCarthy − *Signal Transduction Laboratory, School of Biochemistry & Cell Biology, University College Cork, Cork T12 K8AF, Ireland*

https://dx.doi.org/10.1021/acsomega.0c02113
ACS Omega 2020, 5, 30808−30816

Arjun Ray − *Computational Biology, Indraprastha Institute of Information Technology, Delhi 110020, India;* orcid.org/0000-0002-6020-5863

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.0c02113

Funding

None.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge various research facilities at ACTREC and University College Cork for support. Salary support in form of graduate student (PhD) fellowship from ACTREC to A.M. is also acknowledged.

■ REFERENCES

(1) Litvinov, S. V.; Velders, M. P.; Bakker, H. A.; Fleuren, G. J.; Warnaar, S. O. Ep-CAM: a human epithelial antigen is a homophilic cell-cell adhesion molecule. *J. Cell Biol.* **1994**, *125*, 437−446.
(2) Herlyn, M.; Steplewski, Z.; Herlyn, D.; Koprowski, H. Colorectal carcinoma-specific antigen: detection by means of monoclonal antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76*, 1438−1442.
(3) van der Gun, B. T. F.; Melchers, L. J.; Ruiters, M. H. J.; de Leij, L. F. M. H.; McLaughlin, P. M. J.; Rots, M. G. EpCAM in carcinogenesis: the good, the bad or the ugly. *Carcinogenesis* **2010**, *31*, 1913−1921.
(4) Schnell, U.; Cirulli, V.; Giepmans, B. N. G. EpCAM: structure and function in health and disease. *Biochim. Biophys. Acta* **2013**, *1828*, 1989−2001.
(5) Wang, M. H.; Sun, R.; Zhou, X. M.; Zhang, M. Y.; Lu, J. B.; Yang, Y.; Zeng, L. S.; Yang, X. Z.; Shi, L.; Xiao, R. W.; Wang, H. Y.; Mai, S. J. Epithelial cell adhesion molecule overexpression regulates epithelial-mesenchymal transition, stemness and metastasis of nasopharyngeal carcinoma cells via the PTEN/AKT/mTOR pathway. *Cell Death Dis.* **2018**, *9*, 2.
(6) Coleman-Vaughan, C.; Mal, A.; De, A.; McCarthy, J. V. The γ-Secretase Protease Complexes in Neurodegeneration, Cancer and Immunity. In *Pathophysiological Aspects of Proteases*; Chakraborti, S.; Dhalla, N. S., Eds.; Springer Singapore: Singapore, 2017; 47−87.
(7) Maetzel, D.; Denzel, S.; Mack, B.; Canis, M.; Went, P.; Benk, M.; Kieu, C.; Papior, P.; Baeuerle, P. A.; Munz, M.; Gires, O. Nuclear signalling by tumour-associated antigen EpCAM. *Nat. Cell Biol.* **2009**, *11*, 162−171.
(8) Tsaktanis, T.; Kremling, H.; Pavšič, M.; von Stackelberg, R.; Mack, B.; Fukumori, A.; Steiner, H.; Vielmuth, F.; Spindler, V.; Huang, Z.; Jakubowski, J.; Stoecklein, N. H.; Luxenburger, E.; Lauber, K.; Lenarčič, B.; Gires, O. Cleavage and cell adhesion properties of human epithelial cell adhesion molecule (HEPCAM). *J. Biol. Chem.* **2015**, *290*, 24574−24591.
(9) Carpenter, G.; Red Brewer, M. EpCAM: another surface-to-nucleus missile. *Cancer Cell* **2009**, *15*, 165−166.
(10) Münz, M.; Kieu, C.; Mack, B.; Schmitt, B.; Zeidler, R.; Gires, O. The carcinoma-associated antigen EpCAM upregulates c-myc and induces cell proliferation. *Oncogene* **2004**, *23*, 5748−5758.
(11) Chaves-Pérez, A.; Mack, B.; Maetzel, D.; Kremling, H.; Eggert, C.; Harréus, U.; Gires, O. EpCAM regulates cell cycle progression via control of cyclin D1 expression. *Oncogene* **2013**, *32*, 641−650.
(12) Park, S. Y.; Bae, J. S.; Cha, E. J.; Chu, H. H.; Sohn, J. S.; Moon, W. S. Nuclear EpICD expression and its role in hepatocellular carcinoma. *Oncol. Rep.* **2016**, *36*, 197−204.
(13) Srivastava, G.; Assi, J.; Kashat, L.; Matta, A.; Chang, M.; Walfish, P. G.; Ralhan, R. Nuclear Ep-ICD accumulation predicts aggressive clinical course in early stage breast cancer patients. *BMC Cancer* **2014**, *14*, 726.
(14) Audagnotto, M.; Dal Peraro, M. Protein post-translational modifications: In silico prediction tools and molecular modeling. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 307−319.
(15) Pavšič, M.; Ilc, G.; Vidmar, T.; Plavec, J.; Lenarčič, B. The cytosolic tail of the tumor marker protein Trop2–a structural switch triggered by phosphorylation. *Sci. Rep.* **2015**, *5*, 10324.
(16) Basu, A.; Goldenberg, D. M.; Stein, R. The epithelial/carcinoma antigen EGP-1, recognized by monoclonal antibody RS7-3G11, is phosphorylated on serine 303. *Int. J. Cancer* **1995**, *62*, 472−479.
(17) Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010**, *5*, 725−738.
(18) Wu, S.; Skolnick, J.; Zhang, Y. Ab initio modeling of small proteins by iterative TASSER simulations. *BMC Biol.* **2007**, *5*, 17.
(19) Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinf.* **2008**, *9*, 40.
(20) Colovos, C.; Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **1993**, *2*, 1511−1519.
(21) Maghrabi, A. H. A.; McGuffin, L. J. ModFOLD6: an accurate web server for the global and local quality estimation of 3D protein models. *Nucleic Acids Res.* **2017**, *45*, W416−W421.
(22) Wiederstein, M.; Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **2007**, *35*, W407−W410.
(23) Nishi, H.; Shaytan, A.; Panchenko, A. R. Physicochemical mechanisms of protein regulation by phosphorylation. *Front. Genet.* **2014**, *5*, 270.
(24) Smart, J. L.; McCammon, J. A. Phosphorylation stabilizes the N-termini of alpha-helices. *Biopolymers* **1999**, *49*, 225−233.
(25) Kannan, R.; Cox, E.; Wang, L.; Kuzina, I.; Gu, Q.; Giniger, E. Tyrosine phosphorylation and proteolytic cleavage of Notch are required for non-canonical Notch/Abl signaling in Drosophila axon guidance. *Development* **2018**, *145*, dev151548.
(26) de Vries, S. J.; Bonvin, A. M. J. J. CPORT: a consensus interface predictor and its performance in prediction-driven docking with HADDOCK. *PLoS One* **2011**, *6*, No. e17695.
(27) Tran, M. K.; Kurakula, K.; Koenis, D. S.; de Vries, C. J. M. Protein-protein interactions of the LIM-only protein FHL2 and functional implication of the interactions relevant in cardiovascular disease. *Biochim. Biophys. Acta* **2016**, *1863*, 219−228.
(28) Yang, J.; Zhang, Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* **2015**, *43*, W174−W181.
(29) Hornbeck, P. V.; Kornhauser, J. M.; Tkachev, S.; Zhang, B.; Skrzypek, E.; Murray, B.; Latham, V.; Sullivan, M. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.* **2012**, *40*, D261−D270.
(30) Blom, N.; Sicheritz-Pontén, T.; Gupta, R.; Gammeltoft, S.; Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **2004**, *4*, 1633−1649.
(31) Dou, Y.; Yao, B.; Zhang, C. PhosphoSVM: prediction of phosphorylation sites by integrating various protein sequence attributes with a support vector machine. *Amino Acids* **2014**, *46*, 1459−1469.
(32) Margreitter, C.; Reif, M. M.; Oostenbrink, C. Update on phosphate and charged post-translationally modified amino acid parameters in the GROMOS force field. *J. Comput. Chem.* **2017**, *38*, 714−720.
(33) de Vries, S. J.; van Dijk, M.; Bonvin, A. M. J. J. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **2010**, *5*, 883−897.
(34) Laskowski, R. A.; Chistyakov, V. V.; Thornton, J. M. PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Res.* **2005**, *33*, D266−D268.

(35) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX web server: an online force field. *Nucleic Acids Res.* **2005**, *33*, W382−W388.

https://dx.doi.org/10.1021/acsomega.0c02113
ACS Omega 2020, 5, 30808−30816