# Predictive modeling of CDC25 isoforms to provide insight into dual-action phosphatase substrate specificity

Stuti Goel[1] and Arianna Broad[2]
[1]Aragon High School, 900 Alameda de las Pulgas, San Mateo CA, 94404
[2]Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, NY 14850

## Abstract

Cancer is caused by a dysfunctional cell cycle that allows for excess cell division. The cell cycle is regulated by a series of kinases and phosphatases. Phosphatases are important proteins that remove phosphate groups from other proteins to induce a cellular response. CDC25 phosphatase and its isoforms, in particular, are phosphatases that play a key regulatory role in cell proliferation. However, research has shown that these phosphatases can play both an oncogenic and tumor suppressive function depending on the substrate that it is acting upon. This paper analyzes predictive protein interaction models between CDC25 isoforms and an oncogenic substrate, CDK1, and between a tumor suppressive substrate, CHEK2, to find hot spot residues which may have a role in substrate specificity. Identifying these interactions would identify potential targets for cancer therapies that could manipulate CDC25 phosphatases to exhibit tumor suppressive activity and while inhibiting its oncogenic activity.

*Key words: cancer, cell cycle, phosphatase, oncogenic, tumor suppressive, protein modeling, Alpha Fold*

## Introduction

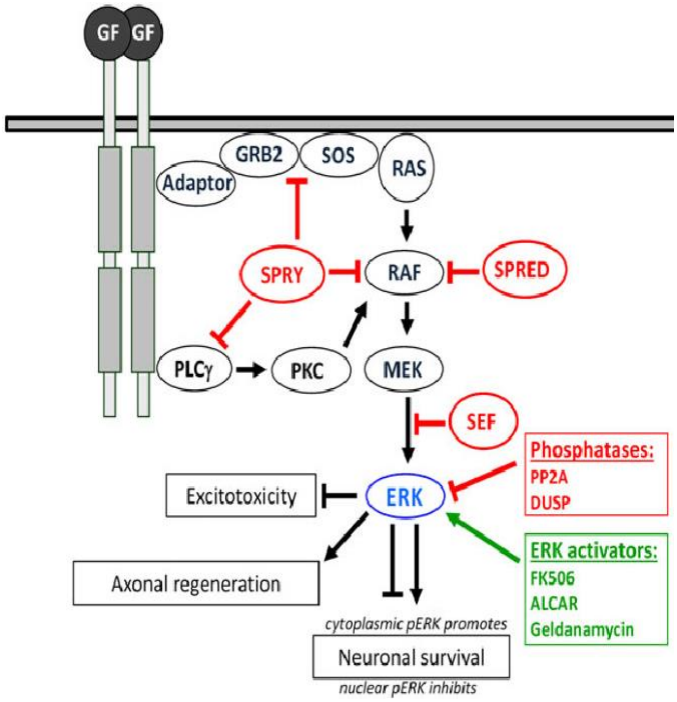


**Figure 1. Ras/Raf Oncogenic Cellular Cascade.** Pathway initiation is started by the binding of growth factors to transmembrane receptors which provokes the activation of a chain of proteins, one of which is Ras. Ras then activates Raf, which initiates another cascade of proteins, eventually activating ERK, a CDK whose activity promotes cell proliferation. The red lines indicate where in the cascade inhibition by various phosphatases can occur. (Hausott and Klimaschewski 2019)

Cancer is a disease characterized by the dysregulation of the cell cycle (Matthews et al 2021). This is caused when there are errors or defects in the key players that regulate the cell cycle, which pushes cells to divide before they are clear of mutations and other aberrations (Matthews et al 2021). This process is regulated by various proteins, namely kinases and phosphatases. A kinase is a protein that phosphorylates a protein, which is the process of adding a high energy phosphate group to a protein, whereas phosphatases are proteins that do the opposite action and remove a phosphate group from a protein (Moura and Conde 2019). Phosphate groups typically activate proteins as they contain three negatively charged oxygen ions, which repel one another creating an abundance of free kinetic energy. The energy from the phosphate group gives proteins energy to perform a function or recruit other proteins. (Cohen 2002). Kinases and phosphatases are extremely important, as they activate or deactivate proteins by adding or removing phosphate groups, which is critical to the makeup of cell signaling pathways. An example of this type of cellular cascade is the Ras/Raf signaling pathway, a well-known oncogenic pathway (Hausott and Klimaschewski 2019).

As Figure 1 demonstrates, the transmembrane receptors receive a growth factor signal from the extracellular space that initiates the signaling pathway. The reception of growth factors triggers a cascade of proteins that eventually activates Ras. Ras, a kinase, phosphorylates Raf, activating it. Raf then phosphorylates (activates) MEK, which consequently phosphorylates (activates) ERK. ERK is responsible for activating certain transcription factors, which encourage the expression of CDKs and cyclins that initiate cell division, making the overexpression or hyperactivation of this signaling pathway oncogenic (Hausott and Klimaschewski 2019). This pathway can also be inhibited at several points of this pathway by various phosphatases, one of such being cell division cycle 25 (CDC25) phosphatases. (Ohouo and Smolka 2012)

CDC25 phosphatases dephosphorylate key regulatory proteins in the cell cycle, and the proteins that do the opposite and phosphorylate these substrates are called cyclin dependent kinases (CDKs). (Smolka et al 2007) CDKs are kinases that are activated by cyclins. Cyclins are a family of proteins specifically produced for cell division processes. Figure 2

demonstrates that as the concentration of cyclin increases, the concentration MPF, a CDK, spikes as well. This results in the cell progressing the cell into mitosis. This implies that the spike in cyclin and CDKs are what promote mitotic activity in the cell (Hunt 2011).

The cyclic nature of CDKs and cyclins is a crucial regulatory process that can be dysregulated in cancer. As mentioned earlier, CDC25 phosphatases play a key role in the dephosphorylation of proteins that play a role in the cell cycle, including CDK and cyclin (Hunt 2011). For example, CDC25 can activate CDKs by releasing them from inhibitory phosphorylation (Shen and Huang 2012). Phosphatase activity can change the outcome of signaling pathways, yet the outcome depends on the role of the substrate that is being dephosphorylated. This change can be tumor suppressive or oncogenic, depending on whether the substrate that CDC25 is enacting upon is oncogenic or a tumor suppressive, respectively. If the phosphatase activity deactivates a tumor suppressive protein, the net effect would be oncogenic, and the inverse would be true if the phosphatase enacted upon an oncogenic protein. (Narla 2018). This makes it crucial to understand CDC25 phosphatases substrate specificity to manipulate it in cancer treatments to serve an anti-cancer role.

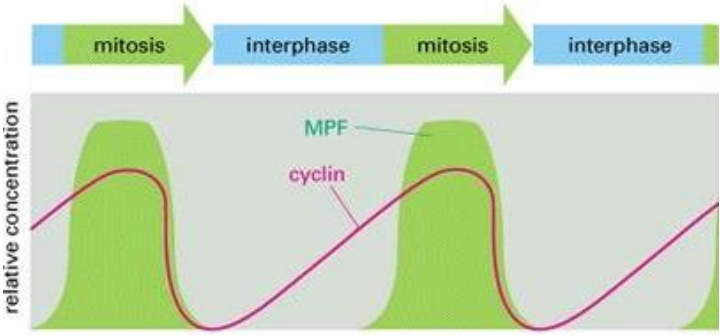


**Figure 2. Cyclic nature of CDKs and cyclin.** Cyclins activate cyclin-dependent kinases (CDKs) by forming CDK-cyclin complexes. The presence and activity of these complexes are what provoke cells to enter mitosis, as the production of CDKs acts as an indicator that the cell is ready to proceed into the next phase of the cell cycle. While cyclins are constitutively expressed, CDKs are expressed specifically to initiate cell division, and as their concentration increases, they form CDK-cyclin complexes, which are only abundant during mitosis. (Hunt 2011)

CDC25 phosphatases have 3 known isoforms in mammalian cells: CDC25A, CDC25B, and CDC25C. They all share similar roles in the regulation of the cell cycle. CDC25 phosphatases activate CDKs by dephosphorylating them, removing the inhibitory phosphorylation. (Boutros et al 2007). Specifically, research has shown that these phosphatases activate CDK1-Cyclin B and encourage promotion into mitosis, as seen in Figure 3 (Moura and Conde 2019). CDC25C directly dephosphorylates CDK1, stimulating its oncogenic activity (Shen and Huang 2012). CDC25B and C function to dephosphorylate CDK1 at inhibitory sites T14/Y15 to encourage cell cycle progression from G2 to M phase (Shen and Huang 2012). CDC25A similarly encourages CDK1 activity to promote continuation of the cell cycle from G1 to S, and also assists with progression from G2 to M. However, CDC25A has a broader role than its other isoforms as it also has a role in promoting proliferation by dephosphorylating CDK4, CDK6 and CDK2 (Shen and Huang 2012). Furthermore, another study showed that the overexpression of CDC25A and CDC25B in mammalian cells pushed cells into mitosis at a faster rate, inducing genetic errors as the cells were forced to divide early by the phosphatases (Boutros et al 2007). It has been demonstrated that the overexpression of CDC25 can lead to increased growth of tumors as it encourages the cell to proceed through the cell cycle even when the cell is not free from anomalies (Timofeev et al 2010).

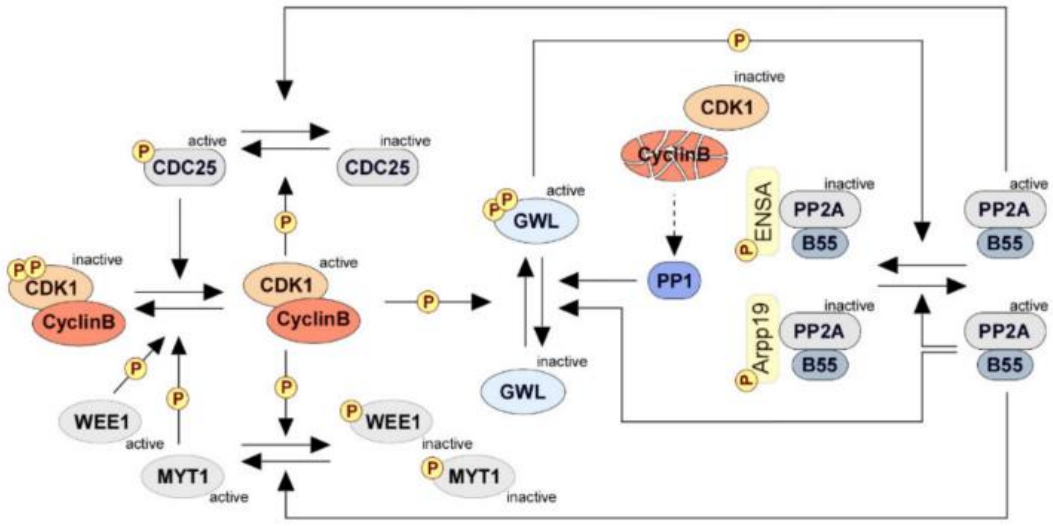


**Figure 3. CDC25 role in an oncogenic pathway** CDC25 dephosphorylates the inhibitory phosphorylation of CDK1-cyclin complexes that rendered it inactive. This dephosphorylation activates the complex. CDK1-cyclin then activates GWL which activates PP2A-B55, creating a positive feedback loop encouraging the activation of CDC25 and other enzymes that can then activate CDK1-cyclin once again. The CDK1-cyclin complex functions to promote mitotic entry, illustrating how CDC25 has an oncogenic effect when it acts upon CDK substrates. (Moura and Conde 2019)

However, CDC25 seems to play a tumor suppressive role at given times as well. For example, in Narla et al 2018, it was demonstrated that depending on cellular context, the phosphatases exhibited both oncogenic and tumor suppressive functions. One example was a signaling cascade with phosphotyrosine15 (pTyr15) as CDC25's substrate, which led to the activation of other tumor suppressive phosphatases (Narla et al 2018). pTyr15 is a tumor suppressive enzyme that is produced during the S and G2 phases in interphase and is negatively regulated by phosphorylation in mitosis (Landrieu et al

2004). CDC25 could have a role in dephosphorylation of pTyr15 in order to promote its tumor suppressive activity that was inhibited by hyperphosphorylation (Landrieu et al 2004). To continue, CDC25 has several inhibitory molecules and proteins. (Fischer 2008) A study looked into CDC25 inhibition as a target for breast cancer research, and revealed that the addition of CHEK2, a CDC25 inhibitor, had a strong resistance to tumor growth. (Liu et al 2018). This indicates that if CDC25 is inhibited by a substrate, the effect overall is tumor suppressive. CHEK2, in particular, inhibits the activity of all 3 isoforms of CDC25 which prevents the activation of CDK-cyclin complexes via CDC25, consequently blocking cell cycle progression (Blasina et al 1999). Other CDC25 inhibitors also produced the same effect. One CDC25 inhibitor called NSC 663284, inhibits the activity of CDC25A and causes cell arrest at both G1 and G2/M phase by blocking CDK2 and CDK1 activation (Liu et al 2018). Another inhibitor is BN82002, which inhibits the phosphatase activity of all CDC25 isoforms, blocks proliferation in many human tumor cell lines, once again demonstrating a tumor suppressive function (Brezak et al. 2004).

Depending on the substrate CDC25 activity can promote either tumor suppressive or oncogenic effects. In order to manipulate CDC25 phosphatases as targets for cancer therapy, we first need to understand its selectivity of substrates, which can aid in the discovery of how to manipulate CDC25 phosphatases to select for tumor suppressive activity. This idea of substrate specificity of CDC25 is an understudied one, yet extremely crucial. Little is known about the differences between the structures of CDC25 isoforms, and understanding its structural landscape will greatly aid in understanding its function and predicting its substrate specificity. CDC25 phosphatases hold a key role in the regulation of the cell-cycle and their expression in cancer cells will allow us to selectively target regions of CDC25 for cancer therapy (Boutros et al 2007).

In this study, we studied the substrate specificity of CDC25 by running protein complex interaction predictions of CDC25 isoforms with tumor suppressive and oncogenic substrates. We compared the protein interaction structures to help find hot spot residues that interact differently depending on tumor suppressive or oncogenic substrates and analyzed this data to also identify regulatory regions in CDC25 isoforms that could be responsible for substrate specificity.

## Methods

*Protein Complex Interaction Predictions*

To predict the interactions between CDC25 isoforms and oncogenic (CDK1) versus tumor suppressive (CHEK2) substrates, Alpha Fold structure predictions were used as the structures in question. Interaction between the two structures was predicted by utilizing the Google collaborative platform ColabFold. Sequences of each respective pair were given to the ColabFold algorithm separated by a colon. ColabFold then would result in one to five protein-protein interaction prediction models, with each model provided with an iPTM score, which is the score that demonstrates how likely this protein-protein interaction was to occur *in vivo*. An iPTM score of 0.8-1.0 means that the protein-protein interaction almost certainly occurs *in vivo*, an iPTM score of 0.6-0.8 means that the protein-protein interaction highly likely occurs *in vivo,* an iPTM score of 0.4-0.6 means that the protein-protein interaction maybe occurs *in vivo*, and an iPTM score of <0.4 means that the protein-protein interaction likely does not occur *in vivo*.

*Hot Spot Residue Identification*

Amino acid bonding pairs between the two proteins were screened for using Chimera. An amino acid bond was only considered to be a putative interaction if the distance between the two terminal atoms was <5Å. Amino acid bonding pairs were then screened for domain placement and categorized as catalytic or regulatory. Catalytic residues are amino acid acids found in a triad, with one acid residue (ex. aspartic acid), one base residue (ex. lysine), and one nucleophile residue (ex. threonine). Regulatory residues are amino acid bonding pairs that form bonds through ionic or covalent interactions in a region distinct from the catalytic triad. For a triad to be catalytic, a proximal threonine or tyrosine residue must be observed on the substrate (CDK1/CHEK2) as the dephosphorylation of threonine or tyrosine residue is coupled with catalytic activity of the CDC25 phosphatase. Identifying a nearby residue that can be dephosphorylated by the catalytic activity of the triad provides evidence for the likelihood of the triad to be the catalytic triad.

## Results

The output from inputting the two putative protein interactors into Google Colabfold is an iPTM score, a quantitative statistical value that measures the likeliness of the two proteins interacting. The iPTM score has a few cut off ranges, iPTM <0.4 = not a likely interaction, iPTM 0.4-0.6 = an interaction is possible, iPTM 0.6-0.8 = very likely interaction, and iPTM >0.8 = an almost certain interaction. After running the models through Colabfold, the program outputs iPTM scores that convey the likeliness of the interaction. The iPTM score evaluating the likeliness of a CDC25 isoform and CDK1 oncogenic interaction ranged from 0.8-0.9, indicating that CDC25 isoforms and CDK1 almost certainly form a protein-protein interaction. The iPTM score evaluating the likeliness of CDC25 isoforms and CHEK2 tumor suppressive interaction ranged from 0.5-0.6, indicating that these two proteins may interact, but if they do it is likely a transient, unstable, or kinetically unfavorable interaction. In figure 4, it is evident that there is an interaction interface at the small alpha helix and at the long alpha helix of CDC25 in

the CDC25-CDK1 interactions (Figure 4A-C). Whereas in CDC25-CHEK2, it is shown that there is an interaction interface at the long alpha helix of CDC25A (Figure 4D), an interaction interface at the short alpha helix of CDC25B (Figure 4E), and there is no proximal interaction interface in CDC25C-CHEK2 interaction (Figure 4F). In total, 10 putative catalytic sites and 11 putative regulatory sites were found amongst the six predicted protein complex models (Table 1).

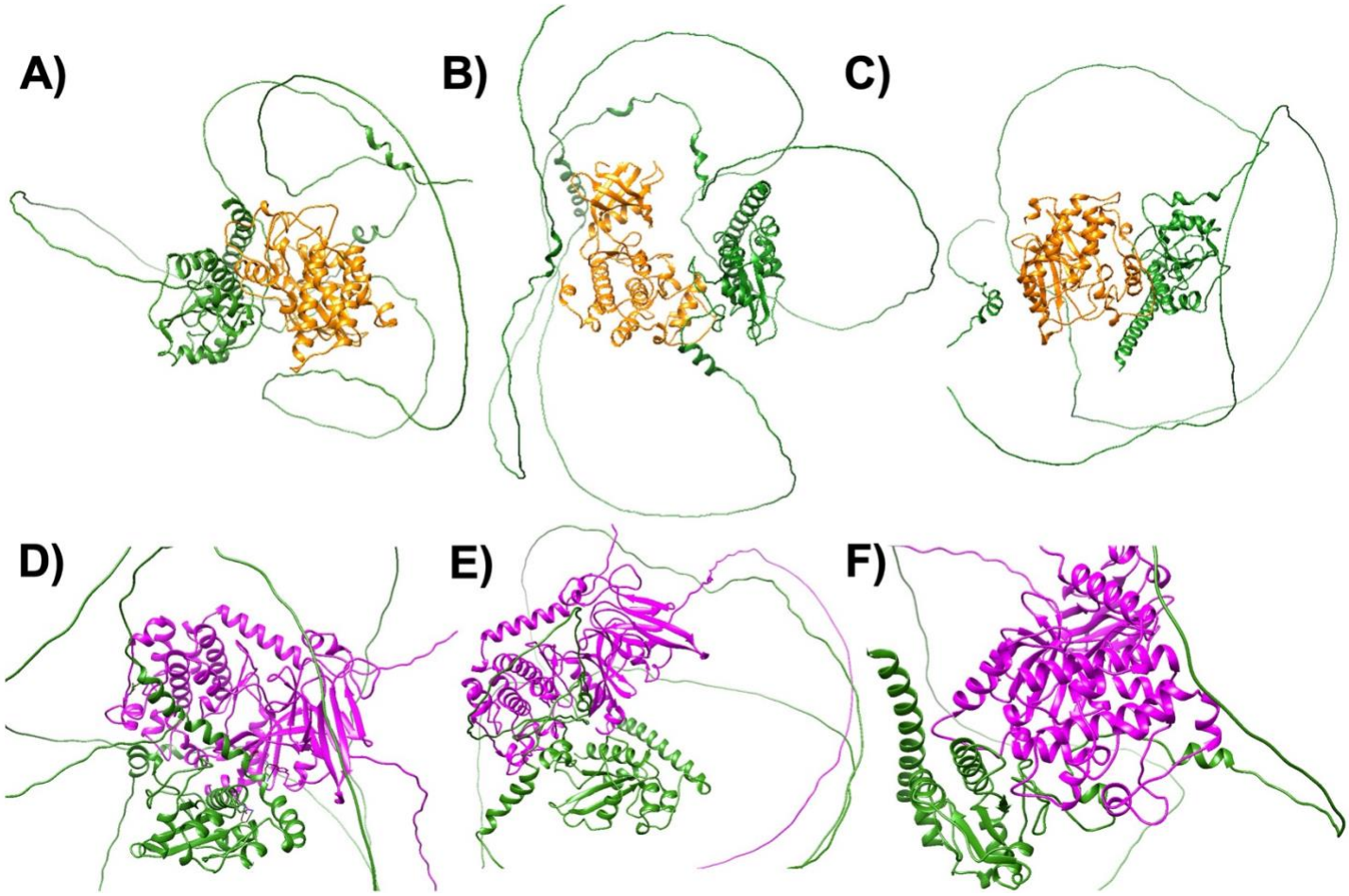


**Figure 4. Images of all predicted protein interactions.** These images show a zoomed-out image of each of the protein interactions that were modeled with Chimera. The interacting interface can be visualized through these images which capture the way they interact. **A-C)** CDC25A-C (green) interacting with CDK1 (orange), an oncogenic substrate. **D-F)** CDC25A-C (green) interacting with CHEK2 (magenta), a tumor suppressive substrate.

*CDC25-CDK1 – Oncogenic Interaction*

When observing the interactions between the CDC25 isoforms and CDK1, there were two goals in mind: (1) identify the catalytic triad interactions, and (2) identify interactions flanking the catalytic triad or at different interfaces to generate inferences on the defining substrate specificity interactions. When visualizing the interaction interface of the predicted CDC25A/B/C-CDK1 interaction models, it was clear there is a conserved triad interaction between all CDC25 isoforms and CDK1. In the CDC25A-CDK1 interaction, Y455 (CDC25A) closely interacts with D211 (CDK1) and R215 (CDK1). In the CDC25B-CDK1 interaction, Y511 (CDC25B) closely interacts with D211 (CDK1) and R215 (CDK1). Lastly, in the CDC25C-CDK1 interaction, Y401 (CDC25C) closely interacts with D211 (CDK1) and R215 (CDK1). Taken together, it is clear that D211 and R215 on CDK1 are two residues involved in the CDC25A/B/C-CDK1 catalytic triad, presenting the acid (D211) and base (R215) required to form a catalytic triad. A catalytic triad is not just composed of an acid and a base, it also requires a nucleophile, which refers to residues that contain hydroxyl groups in their R chain. In the CDC25A/B/C-CDK1 catalytic triad, a tyrosine at varying amino acid positions dependent on the isoform is the clear nucleophile in the catalytic triad. Thus, this CDC25A/B/C-CDK1 interaction prediction within this study successfully identified the residues responsible for catalytic activity in all three isoforms (Figures 5A, 5C, and 5E).

When visualizing the interaction interface of the predicted CDC25A/B/C-CDK1 interaction models, there was initially no residue interactions between CDC25/A/B/C and CDK1 flanking the catalytic site that could be easily identified. However, to computationally resolve the flanking catalytic interactions between CDC25A/B/C and CDK1, we utilized the find H-bond function in Chimera between models to only find interactions between one CDC25A/B/C residue and one CDK1 residue.

| CDC25 Isoform | Interactor | Interaction Type | CDC25 Residue | Interactor Residue |
|---|---|---|---|---|
| CDC25A | CDK1 | Catalytic | Tyr 455 | Asp 211 / Arg 215* |
| CDC25A | CDK1 | Regulatory | Arg 450 | Asp 207 |
| CDC25A | CDK1 | Regulatory | Ile 326 | Arg 218 |
| CDC25B | CDK1 | Catalytic | Tyr 511 | Asp 211 / Arg 215* |
| CDC25B | CDK1 | Regulatory | Arg 506 | Asp 207 |
| CDC25B | CDK1 | Regulatory | Asp 388 | Lys 201 |
| CDC25C | CDK1 | Catalytic | Tyr 401 | Asp 211 / Arg 215* |
| CDC25C | CDK1 | Regulatory | Arg 396 | Asp 207 |
| CDC25C | CDK1 | Regulatory | Gln 399 | Thr 240 |
| CDC25A | CHEK2 | Catalytic (CT1) | Tyr 444 / Glu 409 | Arg 254* |
| CDC25A | CHEK2 | Catalytic (CT2) | Tyr 518 | Asp 347 / Lys 349* |
| CDC25A | CHEK2 | Catalytic (CT3) | Gly 510 | Lys 235 / Tyr 212** |
| CDC25A | CHEK2 | Catalytic (CT4) | Thr 43 / Arg 335 | Glu 429** |
| CDC25A | CHEK2 | Regulatory | Asp 50 | Arg 431 |
| CDC25B | CHEK2 | Catalytic (CT2) | Ser 375 | Asp 347 / Lys 349* |
| CDC25B | CHEK2 | Catalytic (CT4) | Ser 387 / His 389 | Glu 429* |
| CDC25B | CHEK2 | Regulatory | Asp 411 | Lys 255 |
| CDC25B | CHEK2 | Regulatory | Asp 510 | Arg 382 |
| CDC25B | CHEK2 | Regulatory | Glu 464 | Arg 262 |
| CDC25B | CHEK2 | Regulatory | His 159 | Glu 263 |
| CDC25C | CHEK2 | Catalytic (CT2) | Thr 261 | Asp 347 / Lys 349* |

**Table 1. All identified interactions between CDC25 isoforms and CDK1 or CHEK2**. Comprehensive list of all interactions and interacting residues observed in all predicted protein complex interactions. The interaction type of each interaction is also identified as either regulatory or catalytic. Interactions involved in CHEK2 catalytic triads are clearly labeled. Red text indicates conserved catalytic residues amongst all CDC25 isoforms and CDK1. Blue text indicates conserved regulatory residues amongst all CDC25 isoforms and CDK1. Purple text indicates the most likely putative catalytic triad for the CDC25-CHEK2 interaction. * denotes that there is a catalytic triad proximal threonine or tyrosine to couple dephosphorylation to catalytic triad activity. ** denotes there is not a proximal threonine or tyrosine to the catalytic triad.

Upon resolving the intramodel H-bonds, it was clear there was a conserved arginine – aspartic acid interaction flanking the catalytic site in all three predicted models. In CDC25A-CDK1, this interaction was identified to be between R450 (CDC25A) and D207 (CDK1). In CDC25B-CDK1, this interaction was identified between R506 (CDC25B) and D207 (CDK1). In CDC25C-CDK1, this interaction was identified between R396 (CDC25C) and D207 (CDK1). The interesting thing to note here is how all three conserved arginine – aspartic acid interactions are facilitated by the same residue on CDK1, with the same residue at a different amino acid position on the CDC25 isoform. This conserved arginine – aspartic acid interaction is indicative that this interaction is partially responsible for CDC25 isoform identification by CDK1, thus contributing to substrate specificity. Additionally, this interaction flanking the catalytic site was not only conserved in residue type, but also orientation as in all isoforms this regulatory site was upstream of the catalytic triad (Figures 5B, 5D, and 5F).

The conserved arginine – aspartic acid interaction is only half of the picture though, there must be a variable interaction that occurs to be the “nail in the coffin” to CDK1 when differentiating between CDC25 isoforms. Flanking the catalytic site in the opposite orientation of the conserved flanking regulatory site, is a variable ionic interaction site. In CDC25A, this variable interaction was between I326 (CDC25A) and R218 (CDK1). In CDC25B, this variable interaction was between D388 (CDC25B) and K201 (CDK1). In CDC25C, this variable interaction was between N400 (CDC25C) and T240 (CDK1). Taken together with the conserved flanking regulatory site and the opposition of flanking regulatory site position between CDC25A/C and CDC25B, it is clear that these two sites work cooperatively to aid CDK1 in proper selection of CDC25 isoforms (Figures 5B, 5D, and 5F).

Lastly, we looked for a proximal threonine or tyrosine residue to the CT site as phosphatase activity is coupled with the dephosphorylation of a threonine or tyrosine residue. T199 on CDK1 is a conserved proximal residue for all three isoforms, though there is some uncertainty for CDC25A and CDC25C as T240 is also a viable residue that could fit this function. T199 and T240 both lie on flexible linkers, which means that these residues could lie closer to the catalytic site than it seems in our protein complex interaction predicted models (Figures 5A, 5C, and 5E).

*CDC25-CHEK2 – Tumor Suppressive Interaction*

Upon analyzing the CDC25-CHEK2 protein interaction predictions, we had the same two goals in mind: (1) identify the catalytic triad and dephosphorylated residue on CHEK2, and (2) identify putative regulatory interactions, such as ionic or covalent interactions flanking the catalytic triad. When comparing the three CDC25-CHEK2 predictions, after receiving the iPTM scores for the three predictions and observing the interface of the predictions, it was clear that Alpha Fold was unable to predict the CDC25-CHEK2 interface as effectively in CDC25B/C isoforms comparative to the CDC25A isoform. Therefore,

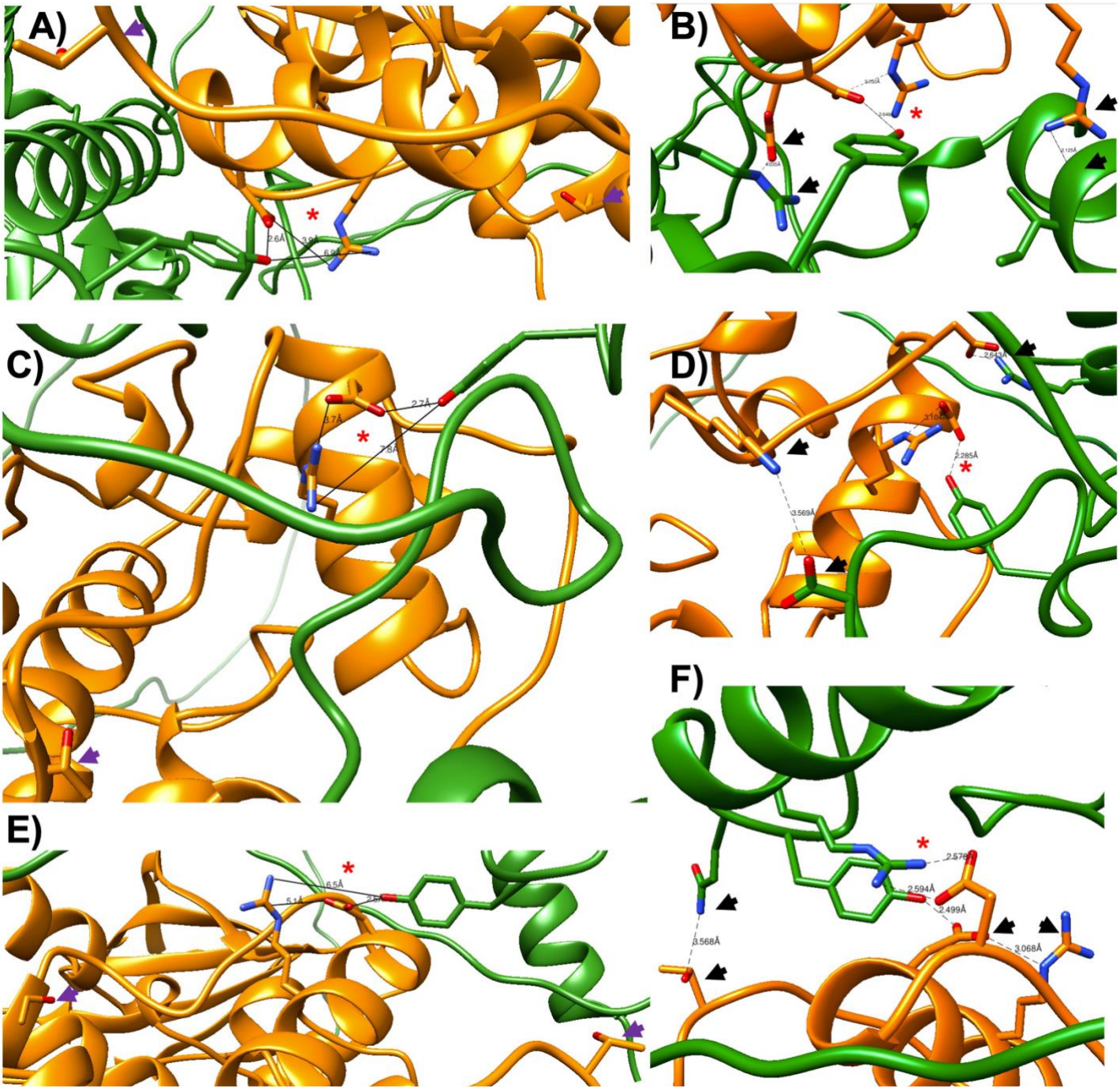


**Figure 5. Residues interactions between CDC25 isoforms and CDK1.** Images A, C, and E depict the catalytic triad formed for each CDC25 (green) isoform (A-C) when interacting with oncogenic substrate CDK1 (orange). Images B, D, and F portray interacting flanking regions that likely contribute to the substrate specificity of CDC25 phosphatases. **A)** possible interacting residues between CDC25A and CDK1. These residues lie in the catalytic domain, indicating the formation of a catalytic triad*. The catalytic triad of CDC25A-CDK1 interaction encompasses two CDK1 residues (D211 and R215)and one CDC25A residue (Y455) in close proximity. These residues lie in catalytic domains, indicating the formation of a catalytic triad. **B)** shows the catalytic triad of CDC25A and CDK1 as well as the two flanking regulatory regions on either side of the catalytic site. The flanking region to the left of the catalytic domain and is upstream of the triad encompass residues R506 (CDC25A) and D207 (CDK1). Opposite of this is another flanking region that is a variable ionic site downstream the catalytic triad between residues I326 (CDC25A) and R218 (CDK1). **C)** possible interacting residues between CDC25B and CDK1. The interacting residues were identified to be two CDK1 residues (D211 and R215)and one CDC25B residue (Y511). Like CDC25A-CDK1 these residues lie within the catalytic domains as well. **D)** depicts the two flanking regions on either side of the catalytic domain. The flanking interaction on the left (upstream) of the catalytic domain is conserved and occurs between residues R506 (CDC25B) and D207 (CDK1). Similarly, downstream another interaction occurs that is an ionic interaction between residues D388 (CDC25B) and K201 (CDK1). **E)** shows a likely catalytic interaction between two CDK1 residues (D211 and R215) and one CDC25C residue (Y401). These residues form a catalytic triad as the three of them interact in close proximity and they're located in the catalytic domains. **F)** depicts two flanking interactions on either side of the catalytic triad formed between CDC25C and CDK1. The flanking interaction upstream the catalytic triad was conserved, identified between R396 (CDC25C) and D207 (CDK1). The other flanking interaction that was identified downstream of the catalytic triad, was an ionic variable interaction between N400 (CDC25C) and T240 (CDK1). Purple arrows denote putative threonines to be dephosphorylated by catalytic triad phosphatase activity. Black arrows emphasize regulatory residue interactions and red asterisk indicates catalytic triad.

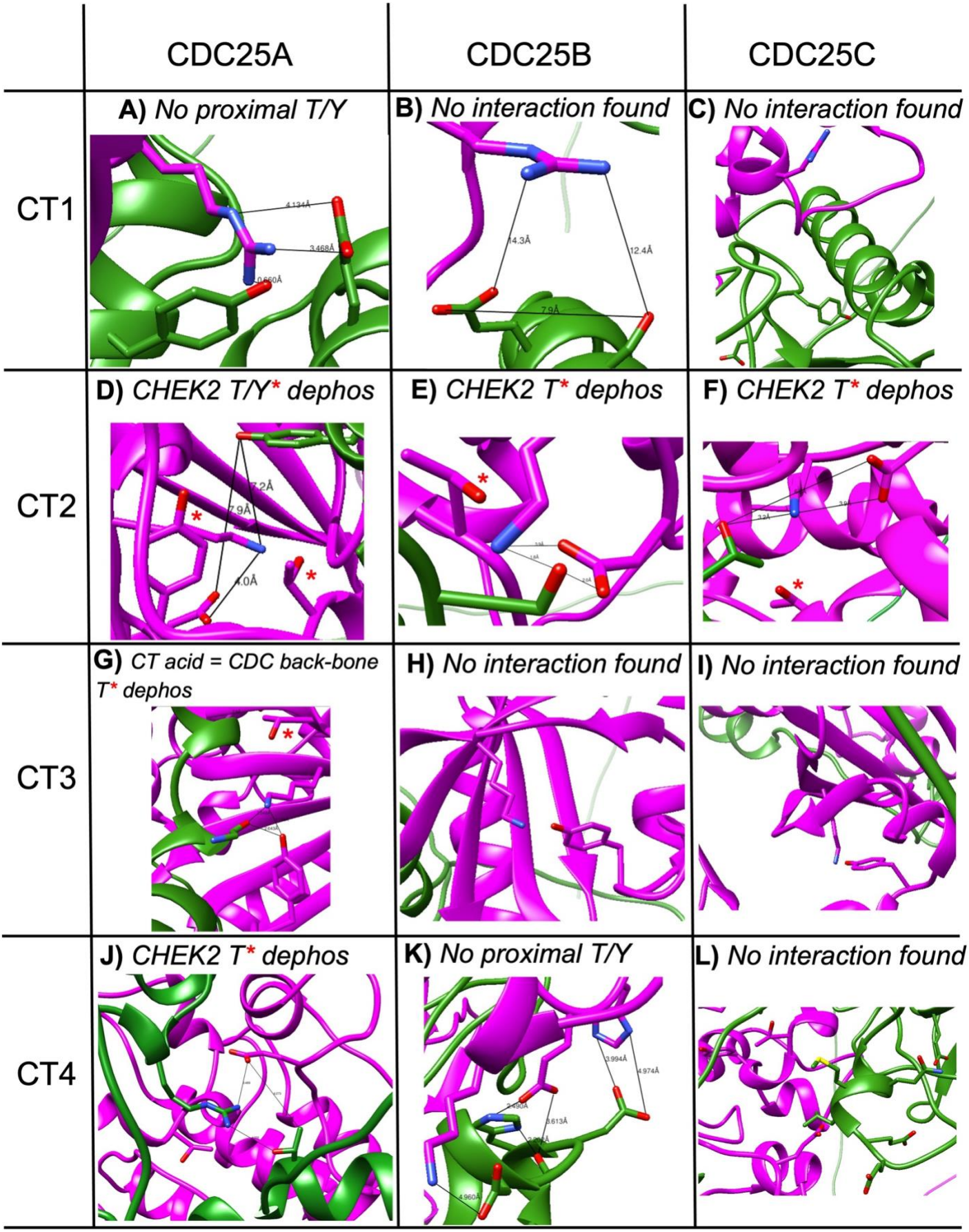


**Figure 6. Residue interactions between CDC25 isoforms and CHEK2.** These images portray all four identified potential catalytic sites between CDC25 (green) isoforms and CHEK2 (magenta). **A)** depicts catalytic triad 1 only observed in CDC25A-CHEK2 interaction. The interacting residues are two CDC25A residues (E409 and Y444) and one CHEK2 residue (R254). No proximal tyrosine or threonine residue was identified. **B)** shows a potential catalytic triad that would likely not interact, showing that catalytic triad 1was not observed in CDC25B. **C)** shows that catalytic triad 1 was not observed in

CDC25C. **D)** illustrates catalytic triad 2 that was identified for the interaction between CDC25A and CHEK2. This catalytic triad occurs between Y518 (CDC25B) and D347 and K349 (CHEK2). Additionally, a proximal threonine residue T367 was identified as was a proximal tyrosine residue Y390 one of which is likely dephosphorylated by the activity of the catalytic triad. **E)** shows catalytic triad 2 for CDC25B-CHEK2. The interacting residues for CHEK2 were conserved (D347 and K349). These two residues interacted with S375 on CDC25B forming a putative catalytic triad. A threonine residue, T367 on CHEK2 was observed in flanking regions close to the catalytic triad. **F)** shows catalytic triad 2 in CDC25C-CHEK2 interaction. This interacting CHEK2 residues were conserved (D347 and K349) and CDC25C nucleophile residue T261. Close to the catalytic triad, a threonine residue, T367 on CHEK2, was identified to be the residue that is dephosphorylated by catalytic activity. **G)** depicts catalytic triad 3 that was observed for CDC25A-CHEK2 interaction. This catalytic triad occurs between G510 from CDC25A and Y212/K235 from CHEK2. These residues are located in the CDC25 backbone, meaning this interaction is unlikely to occur. A dephosphorylation target, T225, was also identified proximal to CT3. **H)** shows the aligned residues of CT3 for CDC25B-CHEK2. These residues do not form a catalytic triad, meaning there is no CT3 interaction for this isoform. **I)** shows the aligned residues of CT3 for CDC25C-CHEK2. No interaction was identified as these residues do not form a catalytic triad, meaning there is no CT3 interaction for this isoform. **J)** illustrates catalytic triad 4 identified for CDC25A-CHEK2 interaction. The interacting residues are T43 and R335 (CDC25A) and E429 (CHEK2). A proximal threonine residue, T387, was identified as a dephosphorylation target. **K)** depicts catalytic triad 4 in CDC25B-CHEK2 interaction in aligned residues with CT4 identified with CDC25A. A catalytic triad was identified for CDC25B between residues S387/H389 (CDC25A) and E429 (CHEK2), but no threonine or tyrosine residue was observed proximal to this triad. **L)** shows the interface of the CDC25C-CHEK2 protein interaction for aligned residues for CT4. The residues aligned with the residues of CT4 do not interact, and no interaction was found.

this was the primary reason why several putative catalytic triads were identified for CDC25A, yet only one or two putative catalytic triads were observed for CDC25C and CDC25B, respectively.

Presented here are four putative catalytic triads formed by CDC25 isoforms and CHEK2 (Figure 6). Catalytic triad one (CT1) was only identified in the CDC25A-CHEK2 interaction and involves two CDC25A residues, E409 and Y444, and one CHEK2 residue, R254. There was no proximal threonine or tyrosine to the catalytic triad, indicating that this was likely not the true catalytic triad (Figure 6A-C). CT3 was also only identified in the CDC25A-CHEK2 interaction. With G510 from CDC25A and Y212/K235 from CHEK2 forming the catalytic triad. There was also a putative dephosphorylative target, T225 that is proximal to CT3, indicating that this may be a candidate for the catalytic triad. However, the triad utilizes the backbone of G510 instead of an R group, which makes this triad unlikely to be the catalytic triad in question (Figure 6G-I). CT4 was identified in both CDC25A and CDC25B. In both isoforms, the interacting CHEK2 residue in the catalytic triad was E429. Where CDC25A interacted with CHEK2 E429 with T43-R335, and CDC25B interacted with CHEK2 E429 with S387-H389. However, it was peculiar to observe that in the CT4 interaction, the CDC25A-CHEK2 interaction has a putative dephosphorylative target, T387, whereas CDC25B-CHEK2 did not have a putative dephosphorylative target. This inconsistent evidence between isoforms leads us to conclude that CT4 is likely not the catalytic triad in question (Figure 6J-L).

Lastly, CT2 was found to be present in all three CDC25 isoforms. The CHEK2 residues involved in the catalytic triad were D347 and K349. In CDC25A, the nucleophile residue was Y518, which is not conserved. In CDC25B, the nucleophile residue was S375. S375 is conserved as the sequence aligned residues are S320 in CDC25A and S265 for CDC25C. And in CDC25C, the nucleophile residue was T261, which was not conserved as the sequence aligned residue is a serine in CDC25B and a leucine in CDC25A. All three isoform catalytic triads had a putative threonine (387) and tyrosine (390) residue flanking the catalytic triad, making CT2 the most convincing putative catalytic triad. However, it was difficult to distinguish the dominant catalytic triad in CDC25A-CHEK2. In addition to the Y518 (CDC25A), D347 (CHEK2), and K349 (CHEK2) catalytic triad flanking T367, there is another putative catalytic triad on the opposite side of T367 which is composed of T43/K522 (CDC25A) and S398 (CHEK2) (Figure 6D-F).

After identifying the putative catalytic triads, we next looked for ionic and covalent interactions flanking the catalytic site, which may contribute to substrate specificity. In CDC25A-CHEK2, D50 (CDC25A) - R431 (CHEK2) form an ionic interaction (Figure 7A). In CDC25B-CHEK2, D411 (CDC25B) - K255 (CHEK2), H159 (CDC25B) - E263 (CHEK2), E464 (CDC25B) - R262 (CHEK2), and D510 (CDC25B) - R382 (CHEK2) were all observed ionic interactions flanking the catalytic site or positioned in another region at the interface (Figure 7B-D). Furthermore, CDC25B-CHEK2 CT4 had two flanking putative regulatory interactions: D386 (CDC25B) - K437 (CHEK2) and D388 (CDC25B) - H430 (CHEK2) (not shown). There were no regulatory interactions observed for the CDC25C-CHEK2 interaction. Overall, the predictive model for the CDC25C-CHEK2 complex interaction was very poor, scoring an iPTM score of only ~0.5. It was clear by the predictive model that the algorithm was unsure of the interactions in this model and therefore these results are less conclusive and require extensive biochemical studies to follow up on these observations resulted from Alpha Fold to conclude any definitive complex interaction. However, this low score could relate back to the prior evidence suggesting that CHEK2 inhibits CDC25. If this were to be the mechanism of action, there would likely be a poor interaction between the two proteins as inhibition of CDC25 is not its ideal state.

*Comparing and Contrasting CDK and CHEK2 Interactions with CDC25 Isoforms*

In prior literature, a common method to identify the catalytic domain of CDC25 isoforms was through X-ray crystal structures (Fauman et al 1998). However, this method contains limitations as X-ray crystals of one protein cannot capture the

experimental interaction between CDC25 isoforms and their substrates. Furthermore, if an experimentalist was able to

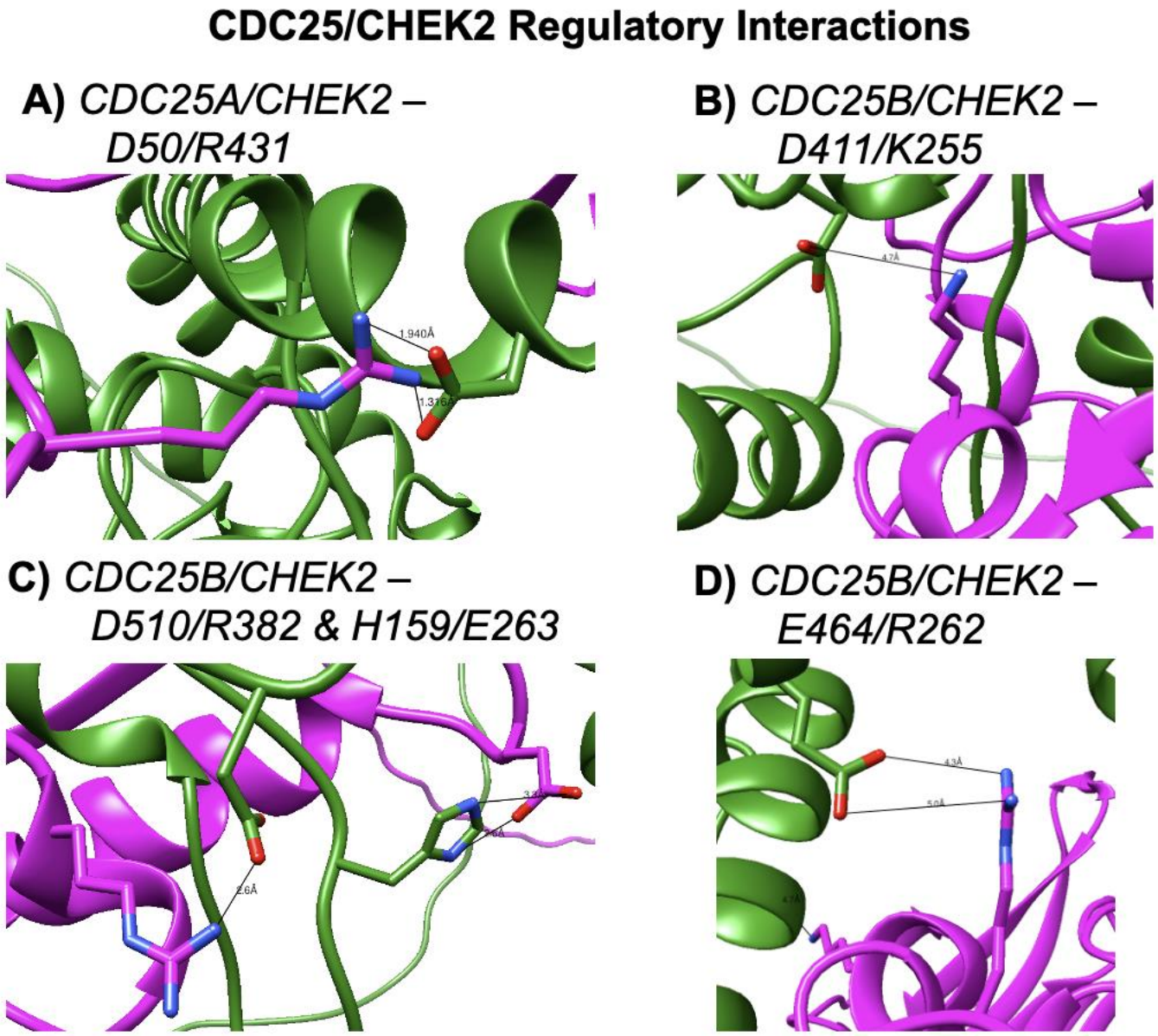


**Figure 7. CDC25-CHEK2 regulatory interaction**. These images identify potential sites of regulation for CDC25 isoforms catalytic activity with CHEK2. These residues lie nearby an identified catalytic triad. **A)** shows an ionic interaction observed in a flanking region proximal to CT2 for CDC25A-CHEK2. This interaction occurs between D50 (CDC25A) and R431 (CHEK2). **B)** depicts a potential regulatory interaction for CDC25B-CHEK2 that was observed close to CT2. This interaction occurred between D411 (CDC25B) and K225 (CHEK2) in an ionic interaction. **C)** shows another potential regulatory interaction for CDC25B-CHEK2. One ionic interaction took place between D510 (CDC25B) and another between R382 (CHEK2) and H159 (CDC25B) and E263 (CHEK2) in a nearby region flanking the catalytic site. **D)** shows an ionic interaction observed in another region of the interface proximal to CT2. The interaction occurs between E464 (CDC25B) and R262 (CHEK2).

obtain a structure of both proteins in complex, it is likely they would only capture one structure, the most energetically stable one, which may not be the interacting structure. (Rudolph 2007) By using protein interaction prediction programs, our results are free from these limitations allowing the visualization of all interaction types whether that interaction may be stable or transient. However, this does not mean the protein complex interaction predictions are free to all limitations. Many of the interactions observed were on flexible linkers or within regions of low certainty, meaning distances between residues may not be exact nor should be interpreted as a conclusive interaction. To corroborate the structural interactions, sequence alignments were performed to determine if the interacting residues are conserved, and if they are conserved, this increases the likelihood that these residues in question are functionally relevant.

Between the two protein model interactions, our results for CDC25-CDK1 isoform complex predictions were far more reliable than the results for CDC25-CHEK2 isoform complex predictions. In the CDC25-CDK1 isoform complex predictions, not only did the high iPTM scores indicate the interactions present have a high certainty, but the interactions present were consistent amongst the isoforms. The conserved regulatory site was always upstream of the catalytic site for all CDC25-CDK1 interactions and always occurred between the same residues, making it highly probable that these interactions are genuine.

To further corroborate these regulatory sites, R450/506/396 (A/B/C) were all aligned and conserved as well (Figure 8). The variable regulatory site, which may be responsible for isoform specificity of CDK1, was partially conserved (Figure 8). I326 identified in CDC25A and N399 identified in CDC25C are all conserved and aligned, whereas D388 is conserved between CDC25A/B but not C (Figure 8). To further corroborate the consistency between these variable regulatory residues, the location of this variable region was conserved being downstream from the catalytic triad. Its similarity in location and variability amongst isoforms suggests that it likely performs a regulative role that distinguished the isoforms from one another. Lastly, there is a structurally and sequentially conserved threonine proximal to the catalytic triad on CDK1 that likely is dephosphorylated upon activation of CDC25 phosphatase activity.

As for CDC25/CHEK model interaction while some conservation was observed, results of CDC25-CHEK2 interactions had more variability between isoforms. For instance, the catalytic triad that we proposed to be most true, CT2, is conserved in all three isoforms with a putative proximal threonine residue as the dephosphorylation substrate. Yet the interacting nucleophile residue was different between isoforms, and the distance of the residues from one another suggests the interaction may not tightly bind or be as stable as the CDC25-CDK1 interaction, adding uncertainty to the conclusions made concerning CDC25-CHEK2 complex interaction. It was also identified that a probable stabilizing region was conserved in CDC25A and CDC25B, but not in CDC25C. While this could be explained by the evolutionary differences in sequences of the isoforms, as A and B are more similar to each other than to C, this could also be an artifact of the low certainty predictive model, considering no regulatory interactions were identified in the CDC25C-CHEK2 predictive model or this could be indicative of an inhibitory interaction, one that is not energetically favorable. Overall, the results presented here concerning CDC25-CDK1 are supported by multiple facets of corroborating evidence and therefore are likely to accurately model the CDC25-CDK1 interaction *in vivo*. This same statement cannot be said for the CDC25-CHEK2 predictive models due to the low certainty of the models statistically, structurally, and sequentially and henceforth, these results need to be further explored before definitive conclusions on the interaction between CDC25 and CHEK2 can be made.

**Discussion**

Cancer is a disease caused by malfunctions in the cell cycle. One way these malfunctions occur is if there is malfunction in the regulation of the cell cycle, which is tightly controlled by special proteins called kinases and phosphatases. Kinases are responsible for adding phosphate groups to other molecules, while phosphatases remove them. Both effect the cell cycle by changing the shape of the substrate either activating or deactivating the protein, which can induce or inhibit a cellular response. (Matthews et al 2022) In this study, CDC25 phosphatases, which have three mammalian isoforms – CDC25A, CDC25B and CDC25C, are examined. Prior research indicates that when these phosphatases enact upon CDK1, a cyclin dependent kinase, whose normal activity induces cell proliferation, when CDC25 dephosphorylates CDK1, it “undoes” an inhibitory phosphorylation, promoting CDK1 activity, exhibiting an overall oncogenes effect. This role that CDC25 phosphatases play make them attractive candidates for cancer drug targets (Narla et al 2018).

While CDC25 plays an oncogenic role when its substrate is CDK1, with other substrates, such as CHEK2, the phosphatases can play a very different role, even a tumor suppressive one. It has been demonstrated that when interacting with CHEK2, CDC25 phosphatases are inhibited from activating CDK1 and promote tumor growth. CHEK2 proves to be a substrate that discourages tumor growth. (Shen and Huang 2012) Therefore, CDC25’s activity is dependent on the substrate it is enacting on, which is why this paper investigates and discusses the substrate specificity of CDC25 isoforms with both oncogenic and tumor suppressive substrates through the use of protein interaction prediction models.

It is without a doubt that CDK1 and CDC25 isoforms interact. The iPTM score of the models of CDC25 and CDK1 was 0.8, meaning that this is a very probable and stable interaction *in vivo.* Additionally, all three isoforms of CDC25 phosphatase formed a catalytic triad with CDK1 with nearby variable regions that likely serve a regulatory function, making it even more likely that these two proteins interact *in vivo*, as three separate interacting regions were identified. When taking a closer look at the catalytic triad, we noticed that not only was there a catalytic interaction between CDK1 and all three isoforms of CDC25, but also that this triad occurred between the same types of residues, specifically a tyrosine residue on CDC25. This tyrosine residue is of extreme importance at it is responsible for the catalytic activity of these two proteins. Prior research implicates that CDK1 and CDC25 activity is oncogenic, making this catalytic tyrosine residue a possible drug therapy target for deactivating the catalytic activity of the proteins to discourage the oncogenic effect.

While the catalytic triad could prove a feasible drug therapy target, targeting a regulatory region that is responsible for the substrate specificity of CDC25 is a better option to only target the phosphatase when it is interacting with CDK1. Two flanking regulatory regions were identified in all isoform predictions. Due to the canonical regulative nature of these flanking regions, one being conserved and the other being variable, it is therefore likely that these regulatory interactions occur prior to the CT site forming, as the catalytic interaction is dependent on the regulative interaction. The conserved regulatory interaction upstream of the CT is proximal to the CT, implying that it is highly likely to be involved in substrate specificity, therefore suggesting this regulatory interaction can be excellent target for drug therapy. (Fischer 2008) Inhibiting this regulatory interaction would specifically inhibit CDC25-CDK1 interaction and henceforth phosphatase activity, while not

affecting CDC25 interaction with other substrates that bind to CDC25 through other structural interfaces. Furthermore, these conserved regulatory residues can be targeted on all isoforms of CDC25, making this an ideal CDC25-CDK1 inhibition candidate. Recall that the interaction between CDC25 and CDK1 has demonstrated to be oncogenic, therefore, inhibiting this interaction would result in the depletion of the previous oncogenic activity, promoting normal cell cycle activity.

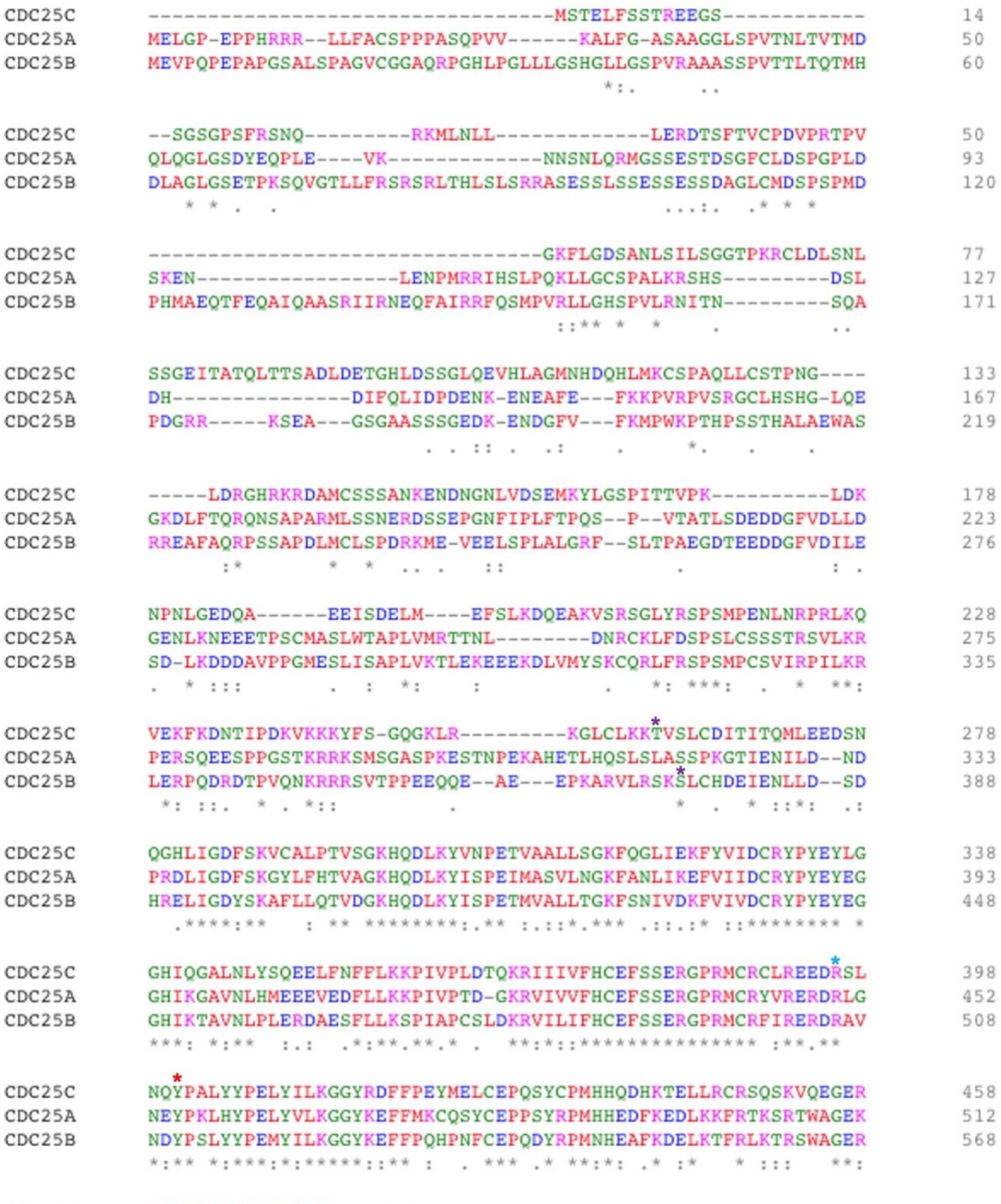


**Figure 8. Sequence alignment between CDC25 isoforms.** CDC25A and CDC25B share higher percent similarity with each other than either of those isoforms with CDC25C. Colored asterisks follow the same legend as the colored text in Table 1. Red asterisk denotes CDC25 participating residue in the CDC25-CDK1 catalytic triad. Blue asterisk denotes conserved regulatory arginine residue on CDC25 that interacts with an aspartate residue on CDK1. Purple asterisks denote the variable nucleophile residue in the CDC25-CHEK2 catalytic triad. CDC25B S375 is the only sequentially conserved CDC25 participating catalytic triad residue; however, the serine is not proximal to the catalytic triad in CDC25A/C, but this could be due to CDC25C S375 being located on a flexible linker, therefore it could lie closer to the catalytic triad than what the current model predicts.

The other flanking region varied between isoforms in terms of which residues were interacting. While all 3 of the isoforms had a second flanking ionic interaction, they all interacted in different ways. Due to its proximity to the catalytic triad, it is highly likely that this second variable region also has a regulatory function and contributes to specificity of CDC25 substrates. However, the variety between the isoforms could imply that there is substrate specificity between the isoforms that is being regulated by this secondary flanking region for CDK1 to distinguish between CDC25 isoforms. However, to determine which of these two flanking regions of the catalytic triad should be targeted to ameliorate the oncogenic effects in the CDC25-CDK1 interaction in cancer cells, more research is required that examines the specific oncogenic effects of each of the CDC25 isoforms to determine if all isoforms should be inhibited by cancer treatments, or if specific ones have a more potent oncogenic effect and therefore should be predominantly inhibited over other less oncogenic isoforms.

When looking at residue interactions between CDC25 and CHEK2, it is important to note that the iPTM value of these models were 0.5, meaning that these two proteins do not form stable interactions *in vivo.* That is why the substrate specificity interaction is ambiguous for CHEK2. Due to its low interactivity score, CDC25 and CHEK2 do NOT form a stable interaction, yet a transient or energetically unfavorable interaction, which are weak interactions that form and break easily. (Perkins et al 2010) Elsewise, this could be due to the inhibitory nature of CDC25 observed when CHEK2 acts upon it.

Within the CDC25-CHEK2 interactions, we were able to identify 4 possible catalytic triads amongst the isoforms. Out of these, the most promising was CT2 (Y/T/S [CDC25] - D347/K349 [CHEK2]) as this was the only triad that was conserved between the isoforms. Additionally, there is an evident threonine residue that lies nearby CT2 that is oriented towards the interaction, suggesting that this threonine residue likely is the residue being dephosphorylated coupled to the catalytic triad interaction. However, all of the putative catalytic triads had a distance between the interacting residues that was greater than 5 angstroms, which is why it is likely that the CT forms a transient interaction rather than a stable one, or there is a stabilizing interaction that occurs prior to the catalytic interaction.

Another interaction that was observed in the CDC25-CHEK2 interactions that could play a role in stabilizing the interacting interfaces is a conserved interaction of E429 on CHEK2 that was observed with CDC25A and CDC25B (not shown). While this interaction has the potential to be a catalytic triad, there is no nearby tyrosine or threonine residue that can be dephosphorylated by the interaction. It is a very stable interaction in comparison to the other interactions observed between the interfaces as all interacting residues were less than 4 angstroms apart. This interaction had three very stable points of contact, but since it is not catalytic, this likely means that it is important for finding interface stability and contributes to the overall stability of the interaction between the proteins. This stabilizing region was only identified on CDC25A and B, meaning that the catalytic interaction for CDC25C-CHEK2 could be transient as it lacks this stabilization, or simply the CDC25C-CHEK2 model was too uncertain to predict the interface. Alternatively, this could also be explained by the fact that when looking at sequence alignments between CDC25 isoforms, the greatest similarity scores were between CDC25A and CDC25B. This proposes that CDC25C is evolutionarily different than CDC25A and B which is why CDC25C lacks the conserved stabilizing region that was observed in the other isoforms. While it is plausible that CDC25A and CDC25B form more stable interactions with CHEK2 it cannot be said with certainty due to the low iPTM of all the models.

From what is known about CDC25 and its isoforms, depending on the substrate CDC25 isoforms can exhibit oncogenic or tumor suppressive effects. By comparing protein interaction structures and finding hot spot residues that interact differently depending on the substrate, we were able to extrapolate certain residues that could be responsible for CDC25 phosphatases regulatory activity and substrate specificity. These protein interactions prove that CDC25 isoforms interact with these substrates and suggest potential residues that could be responsible for substrate specificity. It is yet to be corroborated if these interactions produce the corresponding tumor suppressive and oncogenic effects *in vivo* as detailed in prior research into this protein. Additionally, since CHEK2 and CDC25 had a low iPTM score, researching alternative tumor suppressive substrates of CDC25 that likely interact *in vivo* is important to further understand the substrate specificity of CDC25 and its tumor suppressive role.

The role of phosphatases and kinases in the regulation of the cell cycle is indisputable. Understanding the way that these proteins regulate the cell cycle can get us closer to combating cancer. Our results, corroborated with other papers, prove that CDC25 phosphatases contain a substrate specific mechanism that allows it to interact with different substrates and consequently produce different cellular effects. However, we cannot simply inhibit CDC25 activity entirely as cell proliferation is a required cellular process, and inhibiting CDC25 function would terminate mitosis completely. Additionally, certain substrates of CDC25 induce a tumor suppressive effect that cancer therapies should encourage, not inhibit. The numerous functions of CDC25 make it a difficult target on its own, hence the need to understand the mechanisms through which it selectively interacts with different substrates to be able to target CDC25 but only in a way to inhibit is oncogenic effects. Inhibiting CDC25 activity, while has its benefits in cancerous cells, would be catastrophic in normal cells as all mitotic activity would cease. In normal cells, this would induce a permanent cell arrest. If damaged, these cells would be unable to repair themselves if CDC25 is inhibited. However, small molecule inhibitors targeting cancer cells specifically, could target the CDC25 residues responsible for selecting CDK1 as its substrate and inhibit that activity, whilst preserving the residue

mechanisms that allow CDC25 to bind to tumor suppressive substrates. More research would need to be done to determine how to manipulate small molecule inhibitors to satisfy this desired function.

**Acknowledgements**

The authors would like to acknowledge Lumiere Education for funding this research. Furthermore, the authors would like to acknowledge the Weill Institute of Cell and Molecular Biology at Cornell University for the utilization of their resources in the production of this research project.

**References**

Blasina et al. (1999) A human homologue of the checkpoint kinase CDS1 directly inhibits cdc25 phosphatase. *Cell Biol*. 9(1):1-10. doi: 10.1016/s0960-9822(99)80041-4.

Boutros et al. (2007) Cdc25 phosphatases in cancer cells: Key players? good targets? *Nat Rev Cancer.* 7, 495–507 doi: 10.1038/nrc2169.

Brezak et al. (2004) A novel synthetic inhibitor of cdc25 phosphatases: BN82002. *Cancer Res.* 64(9):3320-5. doi: 10.1158/0008-5472.can-03-3984.

Cohen. (2002) The origins of protein phosphorylation. *Nat Cell Biol.* 4, E127-130. Doi: 10.1038/ncb0502-e127.

Fauman et al. (1998) Crystal structure of the catalytic domain of the human cell cycle control phosphatase, CDC25A. *Cell.* 93(4):617–625 doi: 10.1016/s0092-8674(00)81190-3.

Fischer. (2008) Chapter 11: Cell Cycle Inhibitors in Cancer: Current Status and Future Directions. *Cancer Drug Design and Discovery, Elsevier Inc*. 253-83.

Hausott and Klimaschewski. (2019). Promotion of peripheral nerve regeneration by stimulation of the ERK pathway. *Anat Rec.* 302(8):1261-1267. doi: 10.1002/ar.24126.

Hunt (2011). Getting In and Out of Mitosis*. *Rambam Maimonides Med J*. 2(3):e0051. doi: 10.5041/RMMJ.10051.

Landrieu et al. (2004) A small cdc25 dual-specificity tyrosine-phosphatase isoform in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A.* 101(36):13380–13385. doi: 10.1073/pnas.0405248101.

Liu et al. (2018) Identification of CDC25 as a common therapeutic target for triple-negative breast cancer. *Cell Rep.* 23(1):112–126. doi: 10.1016/j.celrep.2018.03.039.

Matthews et al. (2022) Cell cycle control in cancer. *Nat Rev Mol Cell Biol*. 23(1):74-88. doi: 10.1038/s41580-021-00404-3.

Moura and Conde. (2019) Phosphatases in Mitosis: Roles and Regulation. *Biomolecules.* 9(2):55 doi: 10.3390/biom9020055.

Narla et al. (2018) The impact of phosphatases on proliferative and survival signaling in cancer. *Cellular Mol Life Sci.* 75(15):2695–2718. doi: 10.1007/s00018-018-2826-8.

Ohouo and Smolka. (2012) The many roads to checkpoint activation. *Cell Cycle.* 11(24):4495. doi: 10.4161/cc.22933.

Perkins et al. (2010) Transient Protein-Protein Interactions: Structural, Functional, and Network Properties. *Structure Rev.* 18(10):1233-1243. doi: 10.1016/j.str.2010.08.007

Rudolph. (2007) Cdc25 phosphatases: structure, specificity, and mechanism. *Biochemistry* 46(12):3595–3604. doi: 10.1021/bi700026j

Shen and Huang. (2012) The role of cdc25a in the regulation of cell proliferation and apoptosis. *Anticancer Agents Med Chem*. 12(6):631-9. doi: 10.2174/187152012800617678.

Smolka et al. (2007) Proteome-wide identification of in vivo targets of DNA damage checkpoint kinases. *Proc Natl Acad Sci U S A.* 104(25):10364–10369 doi: 10.1073/pnas.0701622104.

Timofeev et al. (2010) Cdc25 phosphatases are required for timely assembly of CDK1-cyclin B at the G2/M transition. *J Biol Chem.* 285(22):16978-90. doi: 10.1074/jbc.M109.096552.