

# Structure Investigation of *Fam20C*, a Golgi Casein Kinase

Sharon Grubner  
PRIME 2013

National Taiwan University, Dr. Jung-Hsin Lin

University of California San Diego, Dr. Rommie Amaro

## Abstract

---

This research project focused on molecular dynamics simulation of *Fam20C*, a secreted Golgi casein kinase, which was found to target multiple proteins with the same motif.

Mutations in *Fam20C* lead to osteoclerotic bone dysplasia in humans, also known as Raine syndrome. Raine syndrome is characterized by ectopic calcification which increases bone density and often causes death in neonatal stages. The broad aim of this research project was to investigate the structure and function of this unique secreted kinase, *Fam20C*.

Several tools were used to analyze the structure of *Fam20C* once the 200ns simulation of the unbound form was complete: root mean square deviation and fluctuation (RMSD and RMSF) to study certain regions of the protein, bond lengths throughout the simulation to measure salt bridges in the binding pocket, VolMap VMD plug-in to calculate isovalue and water occupancy, GROMOS clustering, as well as FTMap and FTProd to analyze clusters of binding sites inside and outside of the pocket.

## Introduction

---

Fam20C is part of a secretory pathway protein kinase family that is in charge of phosphorylating casein, and was found to target multiple proteins with the motif Ser-x-Glu/pSer [ref 1]. Several unique features, such as the existence of disulfide bonds [ref 2] shows that Fam20C protein kinase may have diverged early on in evolution and is therefore distinct from the common known protein kinases.

Mutations in FAM20C lead to osteoclerotic bone dysplasia in humans, also known as Raine syndrome. Raine syndrome is characterized by ectopic calcification which increases bone density and often causes death in neonatal stages.

The unique structure and function of Fam20C, which have been unclear until that point, requires design of a distinct inhibitor that would be used as a tool to further study its mechanism and target sites.

## Objective

---

The broad aim of this research project is to investigate the structure and function of a unique secreted Golgi casein kinase, FAM20C. Studying the structure and specifically the ATP binding pocket of Fam20C would allow to better understand its function, better control it for experimental purposes, and use that data for a later drug discovery project.

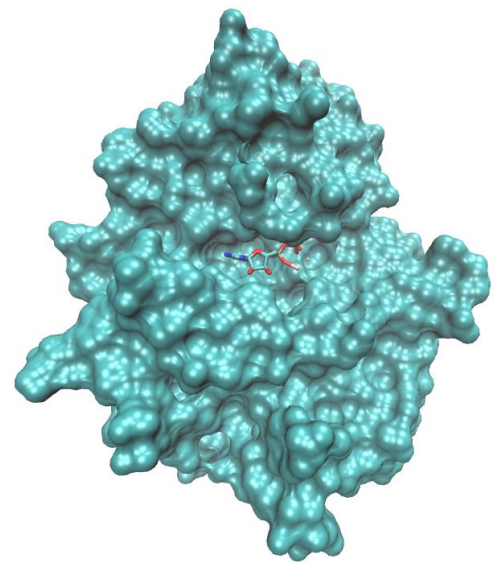


Image 1.

ADP-bound Golgi casein kinase, Fam20C. The kinase is shown in surface representation with ADP bound at its ATP binding pocket-- using the visualization program Visual Molecular Dynamics.

---

## Methods

---

### 1. MD (Molecular Dynamics) simulation

The first step in preparing the Fam20C structure for molecular dynamics simulation was using PDB2PQR in order to approximate protonation at neutral pH of 7.0. The system was set with FF99SB Amber force field, xLEaP was used to solvate the system in a TIP3P explicit water in a cubic unit cell to a distance of 10Å around the kinase, and 19 Na<sup>+</sup> Cl<sup>-</sup> ions were added into the system to reach a final solution of 20mM. The next step before starting MD simulation was energy minimization in four steps which removed any interactions in the system: first only let hydrogens move for 5,000 steps, next also allow water and ions to move for 5,000 steps, third also allow side-chains to move for 10,000 steps, and lastly relax the system and allow everything to move for 25,000 steps. Following the energy minimization steps, the system was equilibrated in four steps of harmonic constraints, starting from 1.0, 0.75, 0.5, 0.25 and lastly allowing the whole system to move. After the minimization and equilibration steps, a 200ns free dynamics simulation was completed with 10<sup>8</sup> steps of 2fs. The crystal structure used in the MD simulation was based on the published structure (pdb ID: 4KQA) with a few mutations that would make it resemble to the human form of the kinase; the active site mutations include: Met171 mutated to Lys, Gln295 mutated to Ala, His388 modified so that the nitrogen close to the bound water will be protonated, a water molecule added in the back of the ATP binding pocket, and glycosylation sugars were removed.

The trajectory DCD files have been concatenated using CATDCD in order to skip every 50 frames of the simulation. The above allowed working with Fam20C simulation more easily while maintaining credibility of the data.

## 2. Clustering

RMSD-based clustering was calculated on Fam20C trajectory, using `g_cluster` algorithm in the Gromacs package, to extract 6 distinct conformation snapshots from the 200ns simulation. The clustering was defined by the active site residues in the ATP binding pocket: Lys171, Gly173-Gln176, Lys178, Val190, Gly192, Glu213, Ala295- Leu298, Arg306, His308, Glu371, Leu386 and Asp387. In order to produce 6 clusters from the simulation, the RMSD cutoff was set to 0.125 nm. The RMSD ranges found were from 0.0294271nm to 0.338968nm and the average RMSD was 0.110407nm.

## 3. Visual Molecular Dynamics (VMD)

VMD was used to visualize Fam20C trajectory as well as calculate various aspect of it using plug-ins, such as: VolMap for water occupancy in the pocket, RMSD Trajectory Tool for bond length and RMSD calculation and alignment, and FTProd to analyze potential binding sites using FTMap product files.

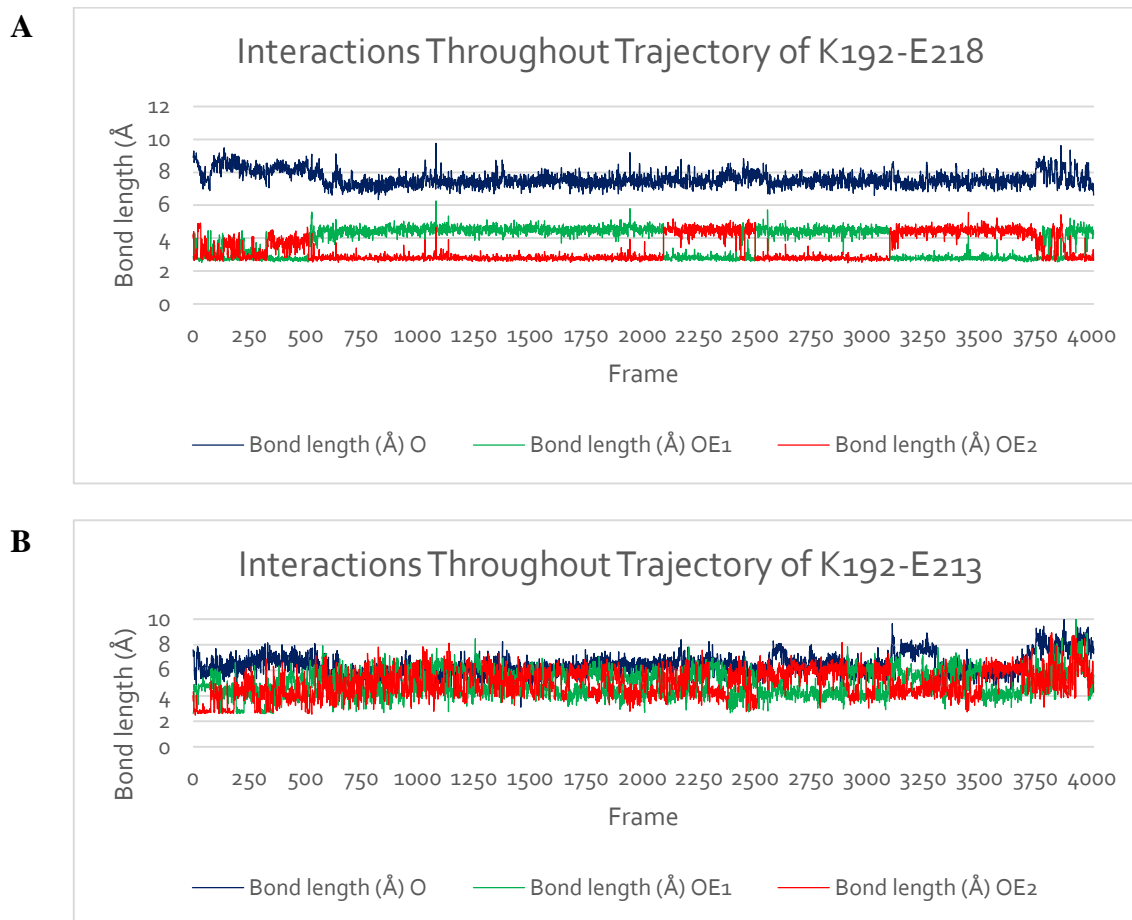
## Results

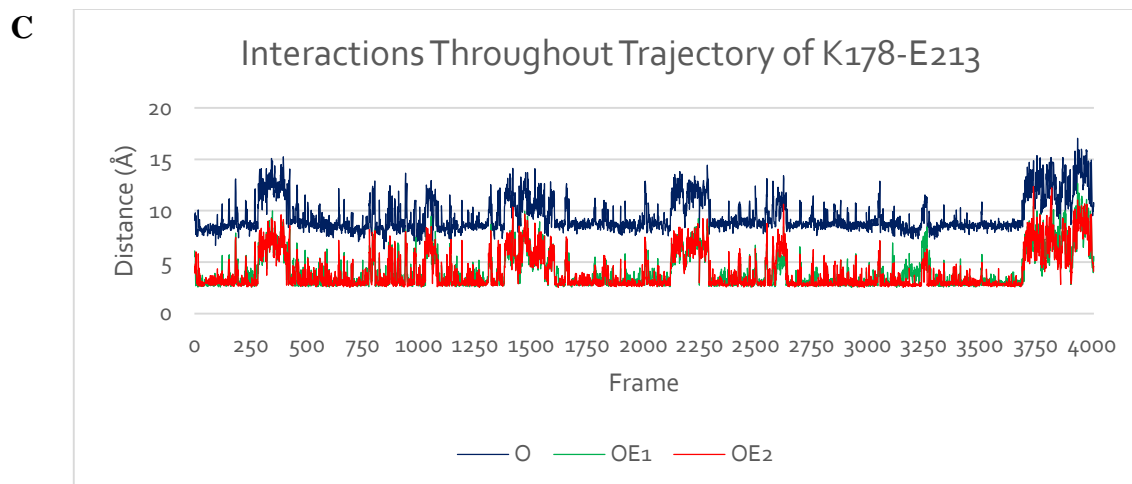
---

Several salt bridges interaction were measured between key residues in Fam20C ATP binding pocket: Lys192-Glu218 [figure 1A], Lys192-Glu213 [figure 1B], and Lys178-Glu213 [figure 1C]. These interactions were selected for inspection due to their significant roles in most types of kinase activity. Each of these bond lengths were measured throughout the 200ns simulation between the nitrogen of Lysine to all 3 oxygens of glutamate. The salt bridge interactions between the conserved residues analyzed ensures the coordination with the ADP phosphate in order to facilitate binding and activation of the kinase. The above analysis shows that Lys192-

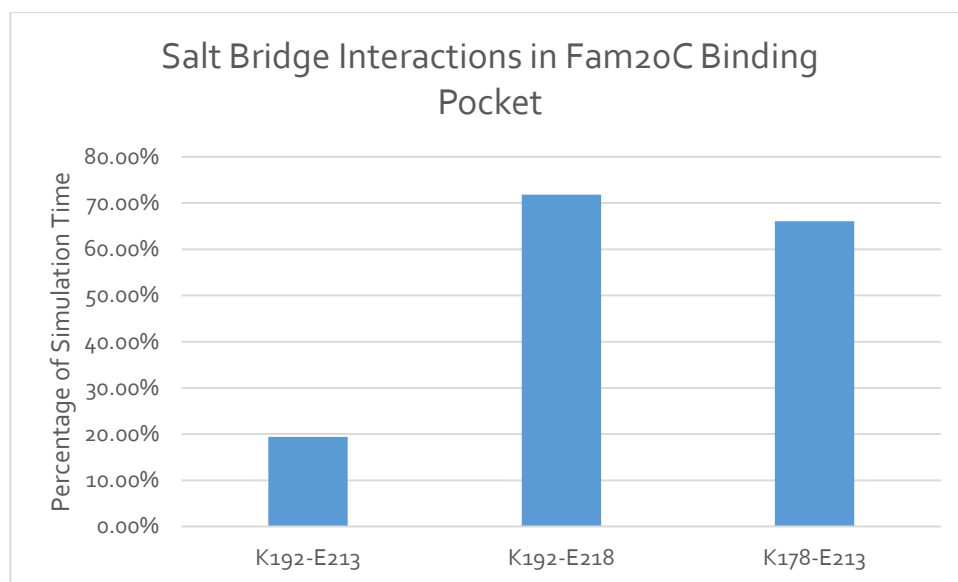
Glu218 salt bridge interaction, of less than  $4\text{\AA}$ , exists 71.8% of the simulation time, while Lys178-Glu213 salt bridge interactions exists 66.1% of the time [figure 2]. These results are interesting since they show that alternate conformation is possible between these pairs and could be exploited for drug design. The above observation also means that the binding pocket will have fluctuations in its electrostatics depending on which conformation the salt bridge is.

**Figure 1.**





**Figure 2.**



*Interactions between p-loop and the C-terminal lobe*

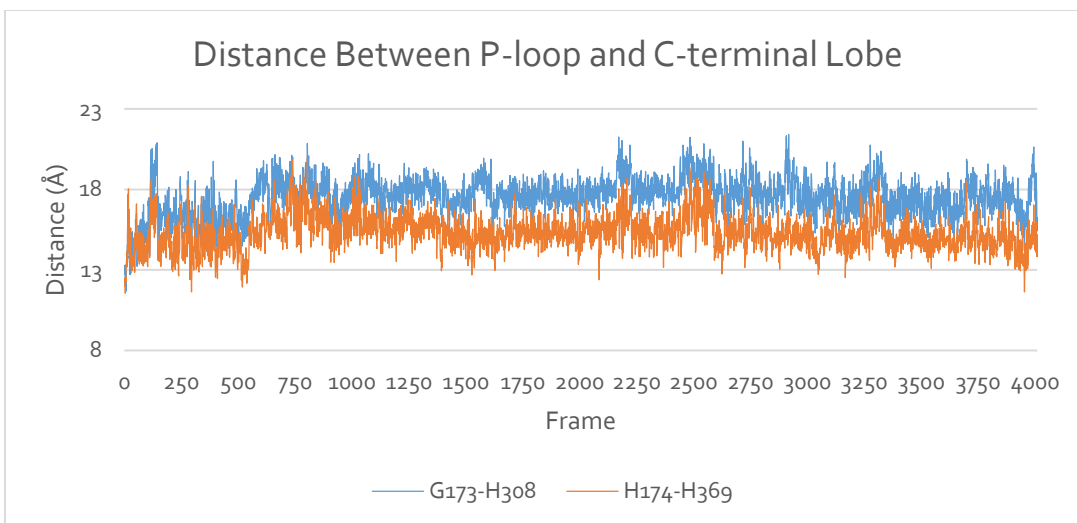
The opening and closing of the ATP binding pocket was measured by the distances between the alpha carbons of the pairs Gly173-His308 and His174-His369 [figure 3]. The distance between the two loops fluctuates between 11.6Å and 21.1Å throughout the 200ns simulation.

When looking at the ADP bound structure from the Fam20C crystal structure it is evident that the average distance between the alpha carbons of the P-loop and the C-terminal lobe is at

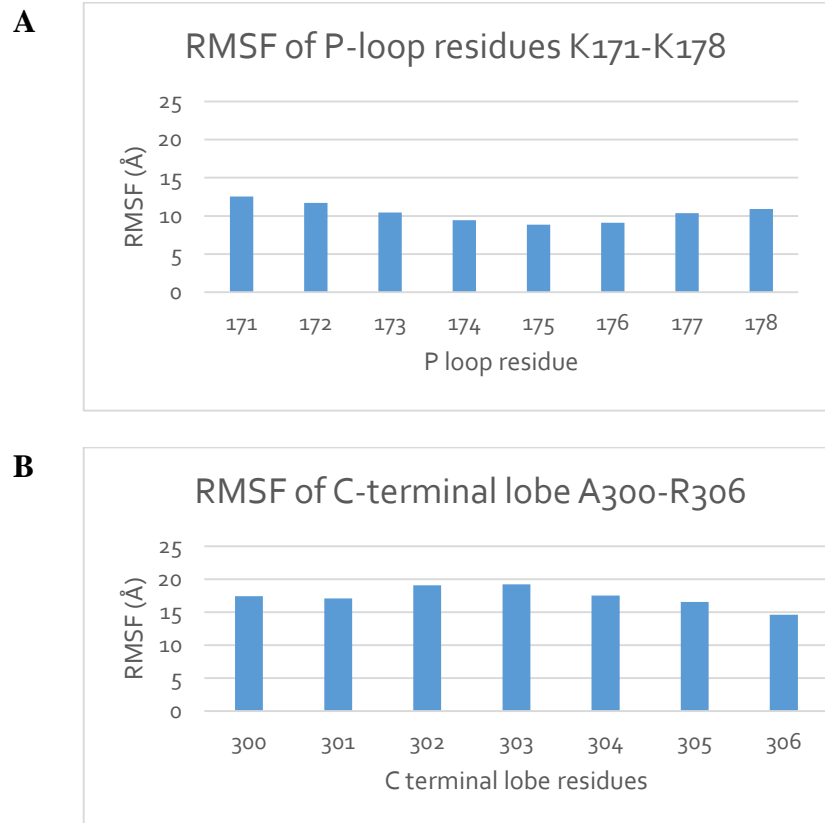
average 14Å. Comparing this value with the inactive apo-form of Fam20C shows that the unbound form is often in and out of position, which can be exploited into design of inhibitors.

To observe the fluctuation in the ATP binding pocket's two major loops- the P loop and the C terminal lobe- the RMSF values were calculated separately for both of the alpha carbons of these loops. It is shown in figure 4A and 4B that the fluctuation of the C-terminal lobe residues is overall higher than the P-loop residues' fluctuation. Flexibility around the ATP binding pocket indicates it is accessible and allows maximization of the number of contacts between the ligand and the binding pocket residues during binding.

**Figure 3.**



**Figure 4.**



### *Hinging*

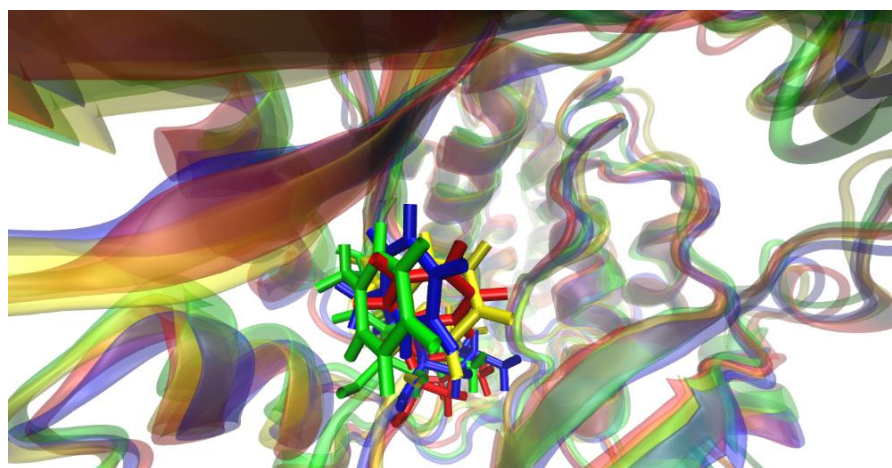
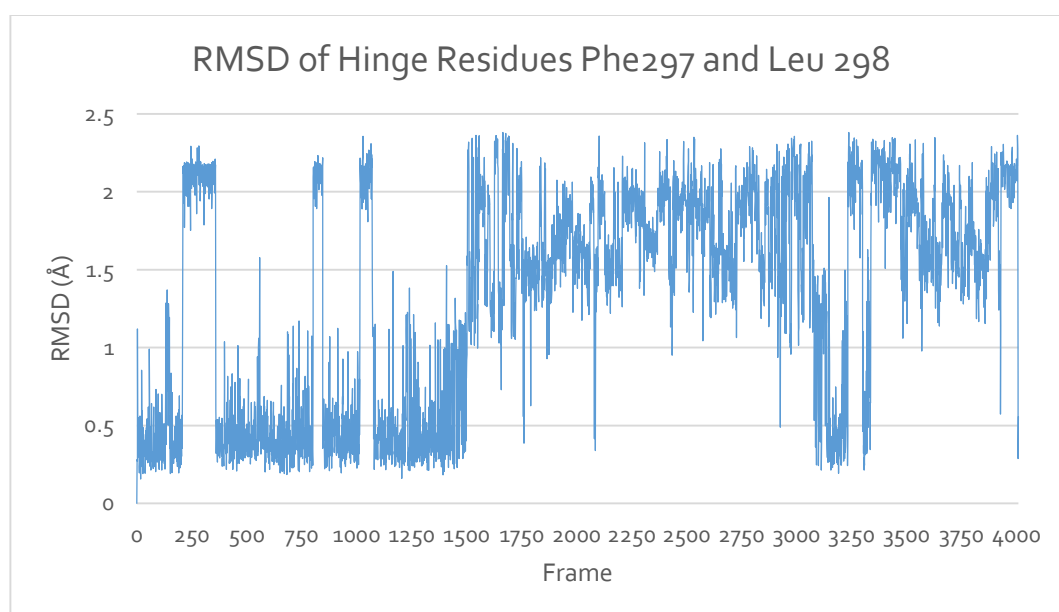
Phe297 ring typically sits over adenine of ATP when it binds to a region called “hinge”, and acts as a gate mechanism for the ATP binding pocket. The hinge residues that form hydrogen bonds to adenine are the backbone NH of Leu298 and the carbonyl oxygen of Phe297. The Phe297 ring seems to be moving quite a lot and shows flexibility as the binding pocket hinge [figure 6]. The movement of the hinge’s ring allows to observe how flexible it is for the drug design purposes, allows to estimate the size of the pocket, how it changes and what can potentially bind to it, as well as the free energy contributions to binding that would need to be compensated. Since RMSD was calculated over only 2 residues, the plot fluctuates greatly but the overall RMSD value does not exceed 2.5Å [figure 5]. Comparing this plot with other 3 additional sets of RMSD



over two residues (not shown in figure) shows another evidence that these two residues are indeed the hinge.

Looking at the crystal structure of the ADP bound Fam20C form, it was suspected that these two residues act as the hinge of the binding pocket. After looking at the abovementioned data it seems as the assumption was correct.

**Figure 5.**



**Figure 6.** Frames 500, 1000, 2500 and 4000 aligned by CA Phe297 and Leu 298 highlighted (500 green, 1000 yellow, 2500 red, 4000 blue).

## *Clustering*

Clustering using `g_cluster` allows to find frames throughout the trajectory with unique protein conformations in order to analyze ligand binding at different states. The 200ns trajectory was aligned by the protein's alpha carbons using RMSD Trajectory Tool VMD plugin and clustered by the active site residues: Lys171, Gly173- Gln176, Lys178, Val190, Gly192, Glu213, Ala295- Leu298, Arg306, His308, Glu371, Leu386 and Asp387.

Using the RMSD cutoff of 0.125nm, the following 6 clusters were found: the most populated cluster of middle frame 3003 with 3811 members, frame 86 with 134 members, frame 323 with 29 members, frame 2 with 13 members, frame 2547 with 12 members, and frame 517 with 11 members. All six of these clusters were sent to FTMap server for further analysis [ref 3].

FTMap is a server that analyzes the clusters found by `g_cluster` and determines 'hot spots'; certain regions on the protein binding surface, of drug targets that bind small ligands with higher affinity than other regions of the binding site. Regions that are identified as 'hot spots' and overlap with other clusters are defined as 'consensus' sites.

FTMAP analyzes one snapshot at once and ranks the clusters based on average free energy that incorporate accurate calculations such as of electrostatics, and repulsive and attractive van der Waals interactions.

By using FTProd it is possible to confirm the exact sites of binding in the binding pocket, as well as to find new potential binding sites on the protein surface outside of the pocket [ref 4]. In addition, it is also possible to characterize the different binding sites at different point of the simulation based on the ligands bound. The 6 pdb structures output from FTMap were uploaded to VMD and aligned by the protein. Next these structures were clustered by Average Link

method with a cutoff of 9Å. Finding consensus sites that appear in all 6 structures, shows they are likely to correspond to binding sites that are able to bind ligands throughout the entire simulation. If these sites exist only in one or two structures, it means they are transient and not frequent throughout the simulation.

Looking at the binding sites around the hinge, a polar permanent interaction was found at the hinge as the highest rank in 4/6 of the frames, and ranked high in the other two frames. Polar interaction with the active site residues Gly173- Gly174 and Glu325- Gln327 from both sides simultaneously was also ranked second at 4/6 of the frames, while ranked first in frames 323 and 3003. At the salt bridge Lys192-Glu218 a relatively transient polar interaction was found in 3/6 frames. An additional analyzed binding site was around the P-loop, however a bit more transient, was observed in 4/6 frames; polar interaction around residues Val177- Lys178 was in 3/4 frames while the other interactions were relatively transient and not significant for analysis. The third and last dominant binding sites was opposite to the hinge, at the ADP pocket: polar interaction was found with Arg306 in 4/6 of the frames, and a transient polar interaction was seen around residue Phe209 in 2/6 of the frames.

#### *Water occupancy*

The ADP binding pocket was also analyzed by looking at water occupancy. Water occupancy at the ATP binding pocket was analyzed using VolMap plug-in in VMD by changing the isovalue range. Water occupancy analysis was done since areas with high water occupancy usually mean excellent candidates for drug binding sites, since high water occupancy implies water ordering which is entropically unfavorable. By looking at the different binding pocket residues it was found that there was water around the hinge residues Phe297 and Leu298 55% of the simulation

time, around the P-loop and more specifically residue Lys171 60-65% of the simulation time, around Arg367 70% of the time, and around residues 294 and 295 80% of the simulation time. Moreover, water was found around the salt bridge Lys178-Glu213 65% of the simulation time, and around the salt bridge Lys192-Glu218 80% of the time, which corresponds to the finding of it being a relatively permanent salt bridge throughout the simulation with 71.8%.

## Conclusions

---

Fam20C, Golgi casein kinase, beyond having atypical kinase structure, has several distinct characteristics that require a design of a distinct inhibitor. Previously designed inhibitors for other kinases cannot inhibit and bind to the unique structure of the studied kinase. In most cases inhibitors compete with ATP binding in order to inhibit the activation of the kinase. The required inhibitor would have to mimic these interactions of ATP in the pocket- in terms of the interactions with the key residues that have salt bridges as well as the interaction with the hinge.

Future directions of this research project include virtual screening of small molecule libraries in order to predict sets of compounds that may be new inhibitors. Once these compounds are defined, they can be docked onto Fam20C kinase using AutoDock to confirm their binding affinities. A list of compounds can be confirmed and ranked for experimental testing to produce an inhibitor tool for the studied Golgi casein kinase.

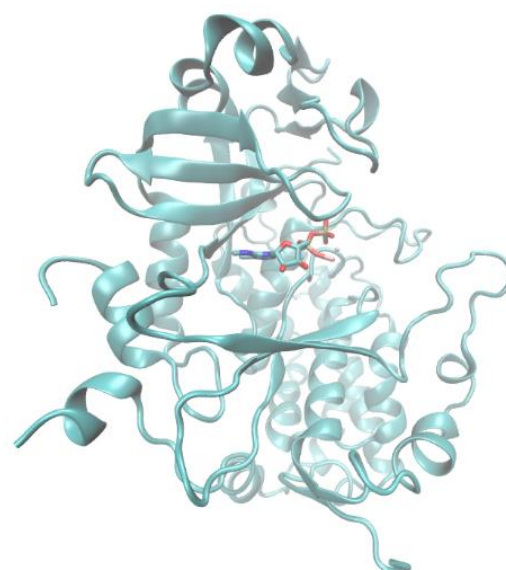


Image 2.

ADP-bound Golgi casein kinase, Fam20C. The kinase is shown in ribbon representation with ADP bound at its ATP binding pocket--using the visualization program Visual Molecular Dynamics.

## Acknowledgments

---

I am grateful for the support of the PRIME team Dr. Gabriele Wienhausen, Dr. Peter Arzberger and Teri Simas. I thank Dr. Rommie Amaro, Dr. Victoria Feher and Lane Votapka for their help throughout the research project in Taiwan as well as in San Diego. Special thanks to Dr. Jung-Hsin Lin for being a great host at his lab at National Taiwan University and assisting as much as possible. Finally, I am grateful for Julia Brown Research Scholarship which allowed me to travel to Taiwan and have a life changing experience.

## References

---

1. Tagliabracci Vincent S., et al. (2012) Secreted Kinase Phosphorylates Extracellular Proteins That Regulate Biomineralization. *Science* 1 June 2012: Vol. 336 no. 6085 pp. 1150-1153.
2. Xiao Junyu, et al. (2013) Crystal Structure of the Golgi Casein Kinase. *Proc Natl Acad Sci U S A*. 2013 Jun 25;110(26):10574-9.  
<http://www.ncbi.nlm.nih.gov/pubmed/23754375>
3. Brenke Ryan, et al. (2009) Fragment-based identification of druggable ‘hot spots’ of proteins using Fourier domain correlation techniques. *Bioinformatics* (2009) 25 (5): 621-627.  
<http://bioinformatics.oxfordjournals.org/content/25/5/621.full>
4. Votapka L, Amaro RE (2012) Multistructural hot spot characterization with FTProd. *Bioinformatics*. 2013 Feb 1;29(3):393-4.  
<http://www.ncbi.nlm.nih.gov/pubmed/23202744>