

CONSERVATION OF ACTIVITY IN FUNGAL CDC14 LEADS TO OPTIMIZATION OF HIGH-SPECIFICITY INHIBITION TARGET DRUG FOR *BOTRYTIS CINEREA*

A. KUMAR¹, H. CAI², AND S. TRUONG³

Biochemistry Department, Purdue University, West Lafayette, IN 47907

ABSTRACT

Fungal pathogens are a global threat to agriculture. Targeting and inhibiting enzymes essential to fungal pathogens is one approach to addressing this threat. The intention of this research is to design and test the effects of hypothetical drug inhibitor molecules on the Cdc14 enzyme ortholog of *Botrytis cinerea*. Cdc14 belongs to a family of highly conserved dual-specificity phosphatases conserved in a wide range of organisms; in general, it is a multi-functional protein that regulates the cell cycle. Multiple areas of high conservation between these species' respective sequences have been documented, though analysis of human paralogs of the Cdc14 family by siRNA overexpression and depletion has conversely suggested that Cdc14 function within the context of the entire cell may not be conserved. Furthermore, higher-order plant species show increasing evolutionary divergence of Cdc14 in the active site, implying that it is a good target for crop preservation by selectively disrupting fungal pathogen cell cycle kinases.

Both computational *in silico* and experimental *in vitro* methods were utilized to identify a competitive inhibitor drug molecule for BcCdc14. Various inhibitor structures were visualized with Molecular Operating Environment (MOE) and AutoDock Vina softwares. Relative efficacies of the candidate molecules in inhibiting enzymatic activity were qualitatively determined *in silico*, while quantified efficacies were determined *in vitro*. Of all candidate inhibitor molecules tested, it was concluded that 5-(4-isopropylbenzylidene)-2,4,6(1H,3H,5H)-pyrimidinetrione (labeled G5), had both the largest percent decrease in computational ligand docking free energy affinity score and the lowest experimentally determined IC₅₀ value. Thus, G5 binds best to BcCdc14, suggesting that an optimal BcCdc14 drug inhibitor would likely assume a similar structure. As BcCdc14 and other previously-characterized fungal Cdc14 display analogous enzymatic activity, such drug inhibitors have potential to act upon a broad spectrum of fungi.

Keywords: *Botrytis cinerea*, Cdc14, inhibition, mitotic exit, dual-specificity phosphatase

1. INTRODUCTION

Enzyme-catalyzed reactions drive all life processes. Cell growth, division, and specialization are all dependent upon enzymatic activity. In understanding the inner workings of specific members of the *Cdc14* enzyme family, we intend to lay the foundation for designing drug molecules that inhibit Cdc14, thereby disrupting normal cellular functions to selectively prevent the spread of *Botrytis cinerea*, a fungal species with pathogenic properties toward plant crop species.

As the human population continues to rapidly increase, protecting our food supply has become increasingly important. Food shortages, compounded with enormous economic losses due to global crop damage, ultimately call for safe, effective, and inexpensive solutions to prevent destructive plant infections and to ensure food security. Current methods of preventing agricultural damage against pathogenic fungi rely on expensive preventative methods as opposed to directly-acting drug molecules. Thus, through advancing human knowledge of the *Botrytis cinerea* Cdc14 (BcCdc14) enzyme,

we hope to address this ominous Malthusian catastrophe.

Cdc14 was recently deemed a key contributor to facilitating plant infection in two widespread fungal pathogens: *Fusarium graminearum*, which infects plants of the *Triticum* genus, and *Magnaporthe oryzae*, which infects plants of the *Zea* genus^[1]. However, although the orthologs of *Cdc14* have been characterized in many quintessential model organisms, many pathogenic fungi Cdc14 have yet to be characterized^[2]. Currently, characterized orthologs of *Cdc14* are known to encode the Cdc14 protein, which functions in regulating the cell cycle as a phosphatase^[3]. Specifically, certain Cdc14 orthologous proteins involved in the FEAR pathway are known to have a wide array of activities; one such function is dephosphorylation of Cdk1-targeting substrates, molecules that regulate the cyclin-dependent kinase essential for ordinary mitotic exit^[4].

To hypothesize the function of the Cdc14 ortholog in *Botrytis cinerea*, orthologs across several species were compared. A heuristic for determining the importance of certain functional groups within a protein relies upon the detection of high conservation within the protein structures of orthologs; i.e., sequence divergence rarely successfully occurs at a conserved protein's important sites. If a certain amino acid sequence is found to be highly conserved, its function is thus also likely to be highly conserved. In the case of enzymes, conservation within

¹ Student, Massachusetts Institute of Technology.

Contact: agnik@mit.edu

² Student, Yale University.

Contact: helen.cai@yale.edu

³ Student, Massachusetts Institute of Technology.

Contact: sdtruong@mit.edu

active sites are of particular interest, as a compromised active site structure can easily result in enzymatic dysfunction. As proteins are characterized by primary, secondary, tertiary, and occasionally quaternary structures, the extent of conservation in a given protein's structure across species can be hypothetically determined by aligning the amino acid sequences of the conserved protein's homologs.

Areas of sequence and structure conservation have already been documented, of which include a characteristic PTP Motif, WPD(A) loops, α 5A/ α 6A loops, a linker α -helix, and Q loops^[5]. To detect for signs of conservation across *Cdc14* orthologs, the multiple sequence alignment tool CLUSTALW was used. Other bioinformatic tools as found on the San Diego Supercomputer (SDSC) Biology Workbench, such as TEXSHADE, allowed for formatting of multiple sequence alignments^[6].

As **Figure 1** shows, amino acid sequences for orthologs of *Cdc14* found with NCBI's BLAST show multiple sites of high conservation. In a majority of eukaryotic species, *Cdc14* displays minimal sequence divergence with the exception of higher-order plant species which are an exception to this pattern. However, as previously acknowledged, although *Cdc14*'s function as a phosphatase may be conserved, its role within the context of the entire cell cycle may not be. Even within *Homo sapiens*, gene duplication has led to the evolution of multiple paralogs of *Cdc14*.

Biochemically speaking, *Homo sapiens* *Cdc14* strongly prefers phosphoserine over other phosphorylated amino acids at cyclin-dependent kinase (Cdk) substrate phosphorylation sites. It has been suggested that this specificity indicates that *Cdc14* plays multiple roles within the cell cycle to regulate cytokinesis in *Homo sapiens*^[3]. Indeed, on the other hand, characterization of *BcCdc14* would lead to a heightened understanding of its function within the cell cycle of *Botrytis cinerea*.

2. METHODS

Preparation and experimentation for this project were divided into several stages: gene cloning and expression, protein purification, enzymatic activity characterization, and identification and *in silico* modeling of hypothetical inhibitor molecules.

Gene cloning and expression: To purify the *Cdc14* protein, a recombinant plasmid containing the c-terminus-truncated *Cdc14* target gene sequence was first produced prior to the commencement of the project. Also contained within the plasmid were an ampicillin resistance coding gene (amp^r), an L-arabinose-inducible promoter, and a sequence coding for six histidines (poly-histidine tag). These recombinant plasmids were then introduced to *E. coli* BL21-AI competent cells. All cells were selectively cultured in agar plates containing ampicillin (100 g/ml) at 30°C to prevent untransformed *E. coli* BL21-AI cells from growing. Several colonies containing the selective marker, amp^r , were then transferred to a flask of 2xYT media. This procedure, termed gene cloning and expression, is outlined in a variety of texts^[7].

Small aliquots from the flask media were stored for later analysis. The remaining volume of media was then treated with 20% L-arabinose liquid solution to induce *BcCdc14* transcription. The concentration of *Bc*

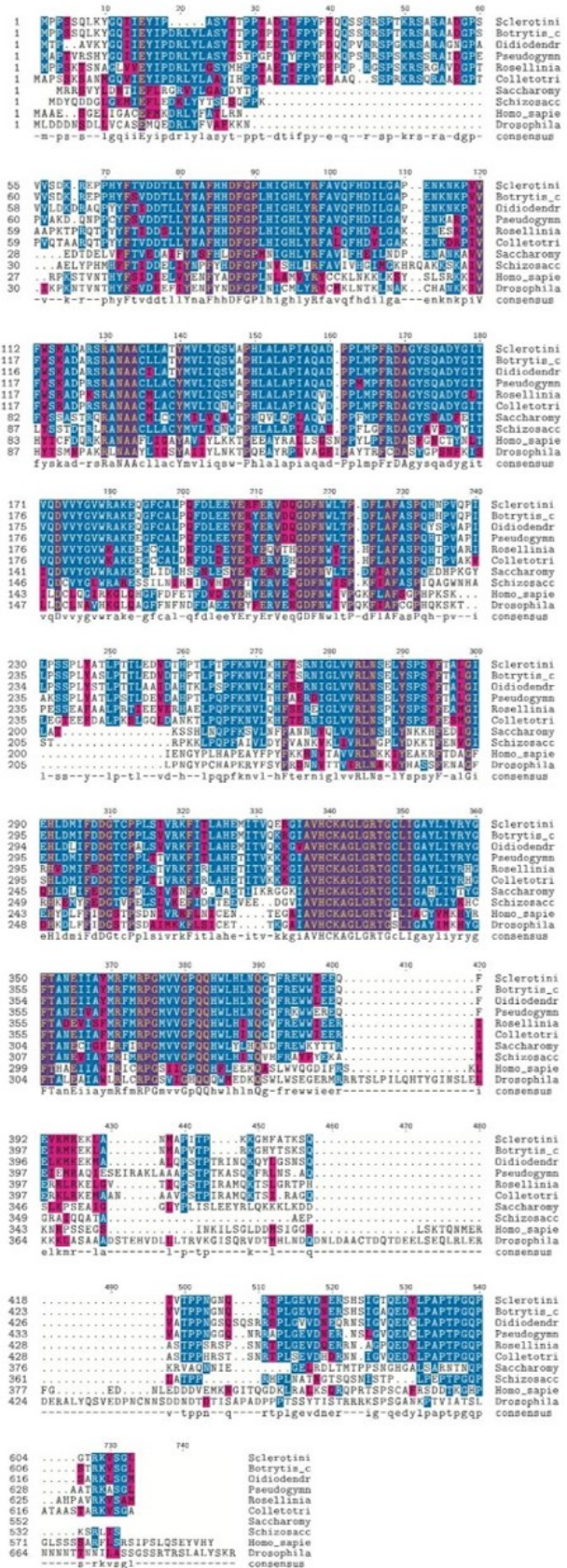


Figure 1. Multiple sequence alignment of *Cdc14* orthologs.

Cdc14, combined with other naturally-produced *E. coli* proteins, was monitored by a standard Bradford Assay protocol^[8]. To determine total protein concentration, a spectrophotometric absorbance curve generated from the media itself was compared against a standard curve for pre-determined concentration of bovine serum albumin. Once the concentration of protein was sufficiently high (36.74 $\mu\text{g}/\text{mL}$), large-volume cultures were spun in a Beckman Coulter J6-MI centrifuge at 6,835 G to isolate suspended *E. coli* cells.

A standard Western Blot protocol was used as an initial screening for *Cdc14* expression. Samples of L-arabinose-induced and uninduced cells from the media were heated to denature all proteins present within the cells, sonicated to shear large and disruptive strands of DNA, and electrophoretically separated based on molecular weight with SDS-PAGE gel. All gels in this project were run at 175 volts for approximately one hour in pH 8.3 running buffer. The resolving gel buffer was composed of 1.5 M Tris-HCl in ultrapure water at pH 8.8. Gels used were composed of a stacking gel (utilizing resolving gel buffer along with 8.2 mL water, 200 μL 10% SDS, 6.5 μL 30% acrylamide:bis-acrylamide solution at a ratio of 29:1, 100 μL 10% APS, and 20 μL TEMED), and a resolving gel (utilizing stacking gel buffer of pH 6.8 along with 6.4 mL ultrapure water, 100 μL 10% SDS, 1 mL 30% acrylamide:bis-acrylamide solution at a ratio of 29:1, 50 μL 10% APS, and 10 μL TEMED). Upon completion, all gels were fixed with a 25% isopropanol and 10% acetic acid solution and then stained with Coomassie blue dye. Digital images were taken for analysis.

Protein purification: Once *BcCdc14* expression was confirmed with a high level of confidence, 1 L of cultured L-arabinose induced cells were resuspended in solutions of 30 mL lysis buffer (25 mM HEPES pH 7.5, 500 mM NaCl, 0.1% Triton x-100, 10 mM imidazole, and 10% glycerol contained in 100 mL water), lysozyme (added 1 mg/mL), leupeptin (added to 10 μM), pepstatin (added to 1 μM), universal nuclease (added to 1 μL per 250 mL culture), and the protease inhibitor phenylmethane sulfonyl fluoride (added to 0.5 mM) to lyse the *E. coli* cells and release all intracellular components, including non-specific proteins and metabolic byproducts.

As a next step to the purification process, the resulting solution was then centrifuged in a Thermo Fisher Scientific Lynx 4000 centrifuge at 107 G to further purify the solution of particles heavier than Cdc14. The resulting supernatant containing BcCdc14 was then purified with a standard nickel affinity chromatography protocol in an Äkta Start protein purification system with a HisTrap nickel column that electrostatically bound to the six-histidine residues DNA-recombinantly engineered into the BcCdc14 protein^[9]. The Äkta Start purification itself was calibrated with nickel buffer A (25 mM HEPES pH 7.5, 500 mM NaCl, and 10% glycerol in 250 mL final volume water) and eluted with nickel buffer B (25 mM HEPES pH 7.5, 500 mM NaCl, 200 mM imidazole, and 10% glycerol in 250 mL final volume water). Elution runoff containing BcCdc14 was collected in 1.5 mL fractions by a Frac30 automatic fraction collector.

Since not all fractions contained equal concentrations of protein by nature of affinity chromatography, relative

protein concentrations of each fraction were evaluated by a Bradford assay. Additionally, purity of the fractions for Cdc14 using bovine serum albumin as a standard was evaluated with a Western Blot. Eight fractions with the highest concentrations of Cdc14 were pooled together and underwent overnight dialysis in 40% glycerol storage buffer solution (25 mM HEPES pH 7.5, 200 mM NaCl, 2 mM EDTA, in 1L final volume water). The final purified solution was then stored at -80°C .

Concentration of BcCdc14 in the purified solution was evaluated by a second Bradford assay and regressively calculated utilizing a standard curve generated with pre-determined concentrations of bovine serum albumin. Purity of Cdc14 solution was qualitatively reassessed with another Western Blot (**Appendix B**). All purified solutions were then stored at -80°C .

Enzymatic activity characterization: Several pNPP assays were carried out to determine the specific activity of BcCdc14 and to infer interactions of BcCdc14 with naturally-occurring and artificially-synthesized inhibitor molecules. The first pNPP assays sought to measure the K_m and V_{max} values of BcCdc14 with pNPP as a substrate, holding pNPP concentration constant and varying BcCdc14 concentration. With an established optimal enzyme concentration, more pNPP assays were performed to measure enzymatic activity as a function of substrate concentration.

Two series of assays using different phosphopeptides were later done to determine the consensus sequence of the optimal peptide for BcCdc14 dephosphorylation. The assays were performed in a similar fashion to the pNPP assays with the modification of using BIOMOL Green to terminate the reaction as opposed to sodium hydroxide. A total of 20 phosphopeptide substrates were tested (**Figure 9**).

Ten more single point pNPP assays were performed, utilizing ten different hypothetical competitive inhibitors. These inhibitors were modeled and docked in the modeling software MOE prior to the experiment (**Appendix A**). Out of the ten, three inhibitors showing the highest levels of enzyme inhibition were selected to perform subsequent experimental assays for generation of IC_{50} curves. The inhibitor with the lowest IC_{50} value, G5, was selected for further kinetic analysis.

pNPP assays for 75 μM , 50 μM , and 20 μM concentrations of G5 were performed to measure the extent of inhibitory effects. G5's inhibitory efficacy was evaluated by monitoring for changes in K_m and V_{max} values at 2 mM BcCdc14. K_m and V_{max} value changes suggests that G5 is a mixed inhibitor; however, previous literature on analogous inhibitors for *Fusarium graminearum* Cdc14 has contrarily suggested that G5 acts as a competitive inhibitor on BcCdc14. This suggestion comes with the assumption that BcCdc14 and *Fusarium graminearum* Cdc14 respond in a similar fashion upon G5 challenge - thus justifying subsequent *in silico* drug inhibitor design and optimization trials.

In silico optimization of the G5 ligand was performed in MOE. Efficacy and improvement in binding affinity (i.e. inhibitory potential) of the molecule was evaluated by the affinity score, a metric of the free energy of binding.

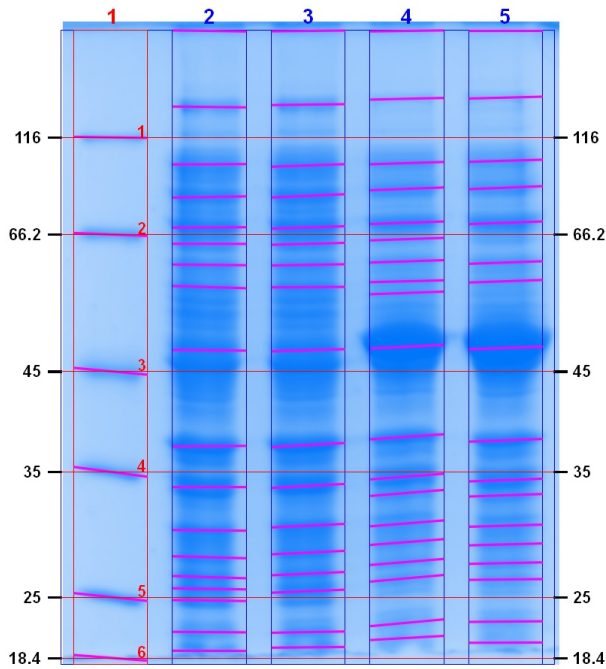


Figure 2. SDS-PAGE gel indicating presence of BcCdc14⁴, with the appearance of a thick, well-defined band at approximately 49.38 kDa. Cell contents from L-arabinose-uninduced cells are in lanes 2 and 3, while cell contents from L-arabinose-induced cells are in lanes 4 and 5. All gels were stained with Coomassie blue.

3. RESULTS

Gene cloning and expression: To express a large amount of target protein, *E. coli* cells with the inserted plasmid were cultured in 2xYT media at 30°C for approximately 36 hours. Sample cell contents from both L-arabinose-induced and L-arabinose-uninduced cells were run through SDS-PAGE with a standard Western Blot protocol.

Statistical Analysis of PS (SAPS) from SDSC Biology Workbench predicts the molecular weight of c-terminus-truncated BcCdc14 to be 49.38 kDa. Relative to the molecular weights of all proteins present in the molecular weight standards (lane 1: β -galactosidase, bovine serum albumin, ovalbumin, lactate dehydrogenase, REase Bsp981, β -lactoglobulin, lysozyme), the location of a thick band around molecular weight 50 kDa suggests that BcCdc14 was indeed expressed.

Protein purification: Lysing all L-arabinose-induced cells yielded solution containing all *E. coli* cell components along with artificially high concentration of BcCdc14. The solution with dissolved cellular components was then purified via nickel affinity chromatography with the Äkta Start system. Monitoring elution of BcCdc14 was made possible with a detection system that measured UV absorbance in the stationary phase nickel column.

Purity of BcCdc14 solution in every other fraction was visually evaluated by SDS-PAGE. Fractions 25-32 were deemed adequately pure for use in experiments and consequently pooled and mixed for overnight dialysis in glycerol solution, increasing concentration of BcCdc14 even further.

The final solution of purified Cdc14 in glycerol was

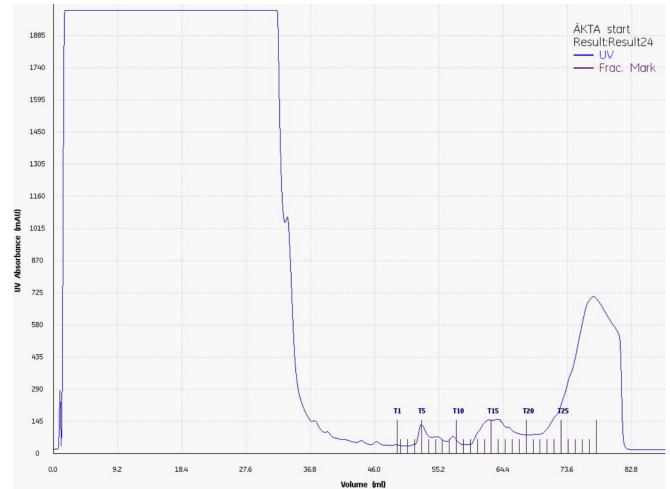


Figure 3. Graph of elution progress against UV absorbance. Within the 30 fractions, maximum elution of BcCdc14 occurred at the end of the fractionation procedure (T30). Highest concentrations of Cdc14 were eluted after T30, as indicated by the width of the peak extending past T30, but were unable to be collected in fractions. Additional fractions labeled F31 and F32 were manually collected instead.

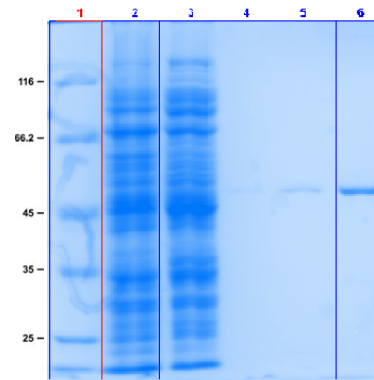


Figure 4. SDS-PAGE gel of purified BcCdc14 solution after processing via Äkta Start system and dialysis. Molecular weight standard is in lane 1. Extracts of pre-purification and post-purification flow are in lanes 2 and 3, respectively. 0.2 μ g BcCdc14, 0.5 μ g BcCdc14, and 2 μ g BcCdc14 are in lanes 4, 5, and 6.

evaluated via a Bradford assay to be approximately 78.514 μ M; despite the initial overexpression and production of BcCdc14, a large amount of BcCdc14 was lost due to delayed elution during the chromatography process. Dilutions of final resulting solution were analyzed with a Western Blot to confirm the concentration and integrity of the final purified protein solution.

Enzymatic activity characterization: The functionality of the protein solutions obtained from fractionation indicates presence of BcCdc14. Since orthologous proteins of Cdc14 have been characterized as phosphatases, a sodium phosphate assay was employed to check for BcCdc14 phosphatase activity. Qualitative characterization of BcCdc14 enzymatic activity was determined with a SpectroVis spectrophotometer. Quantitative characterization of BcCdc14 enzymatic activity was performed using the same spectrophotometry method.

pNPP assays were used to determine BcCdc14's enzyme kinetics parameters. Although initially, the wild-

⁴ BcCdc14 used in the study was truncated and histidine-tagged and is thus not approximately 68.7 kDa as its putative sequence would suggest. Rather, c-terminus-truncated BcCdc14 is approximately 50 kDa.

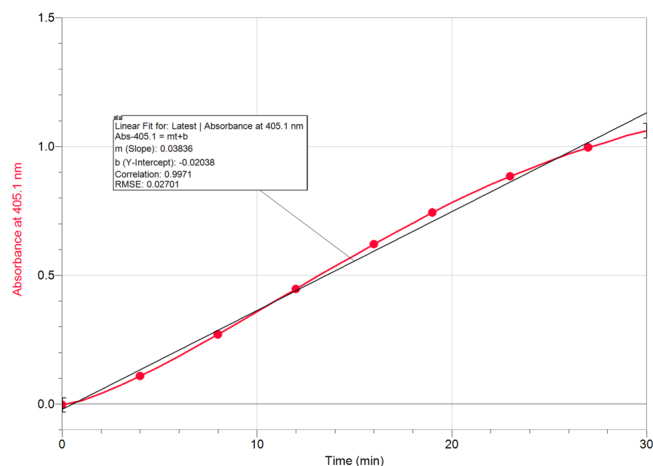


Figure 5. Absorbance curve of enzymatic activity over time. The curve is not perfectly linear, but shows sufficient linearity for linear approximation. The least-squares linear regression curve yielded a root mean squared value of 0.02701, indicating relatively strong correlation.

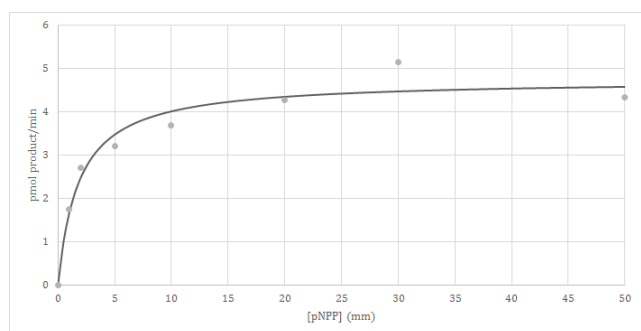


Figure 6. Absorbance results measuring activity as a function of varying substrate concentration. K_m and V_{max} are noted for later analysis. The fit equation was given to be $y = \frac{4.745142x}{1.830239+x}$.

type substrate Cdc6ps7 was used, tight binding between enzyme complex and the wild-type substrate resulted in rapid catalysis, creating a narrow window for data collection and resulting yielding imprecise data. Accordingly, all subsequent assays were performed varying pNPP, instead of Cdc6ps7, and BcCdc14 concentrations.

Upon establishing steady-state parameters, additional pNPP assays were performed to test the efficacy of individual substrates from two distinct series, totalling 20 phosphopeptides. Experimental data from these series were compared with sequences of phosphopeptide substrates to determine substrate parameters for optimal catalysis.

Substrate preferences of BcCdc14 were ultimately determined. Phosphoserine phosphopeptides with a peptide sequence of pSer, +1 proline, and +3 Lysine/Arginine are preferentially dephosphorylated, as also seen in *Homo sapiens* orthologous 1OHD.

The similarity in substrate specificity between BcCdc14 and Cdc14 orthologous proteins suggests that enzymatic binding at the active site is highly conserved, particularly phosphate group binding to amino acid residue Cys344. Based on this reasoning, an ideal competitive inhibitor drug molecule should have residues that bind tightly to Cys344 but not include a phosphate group as to avoid dephosphorylation and to remain

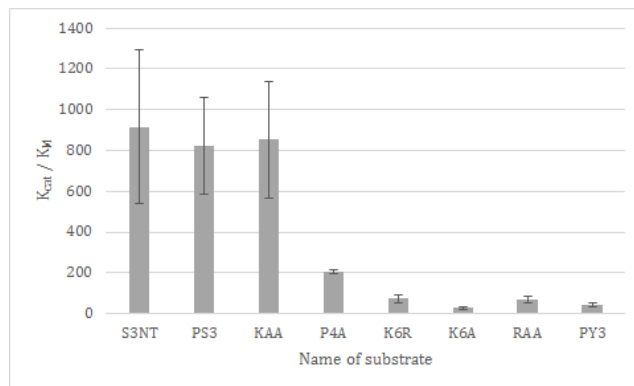


Figure 7. Relative K_{cat}/K_m of each substrate in initial series as determined by single-point assay. We assume that each reaction was linear and is measurable above background signal; substrates with consumption levels below 10% at given parameters (NAA and PT3) were not evaluated with this reasoning.

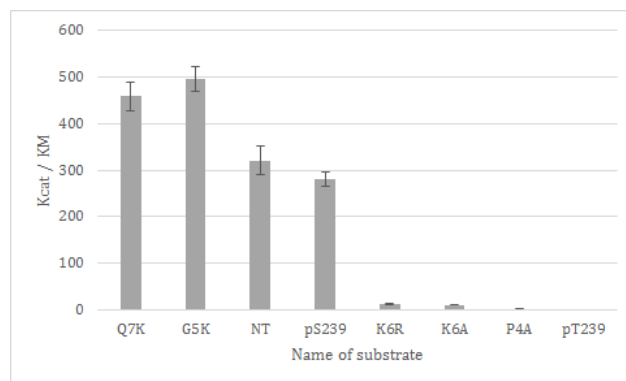


Figure 8. Relative K_{cat}/K_m of each phosphopeptide in second series as determined by single-point assay. Variance for each of the trials is substantially lower. Four substrates clearly give a much greater K_{cat}/K_m value, likely due to presence of both a phosphorylated serine and lysine at +3 position. This evidence suggests that the function, preference, and selectivity of BcCdc14 is conserved, as hypothesized.

Shorthand name	Acm1pS3	Acm1pT3	pY3	P4A	K6A	K6R	S3NT	KAA	RAA	NAA
Amino acid sequence	MlpSPSKKRTI	MlpTSPSKKRTI	MlpYPSKRTI	MlpSASKKRTI	MlpSPSAKRTI	MlpSPSRKRTI	MlpSPSKKRTI	MlpSPSKAATI	MlpSPSRAATI	MlpSPSNAATI
Shorthand name	pS239	pT239	P4A	G5K	K6A	K6R	tT	Q7K		
Amino acid sequence	LLpSPKQFRQ	LLpTPKQFRQ	LLpSAGKQFRQ	LLpSPKKQFRQ	LLpSPCAQFRQ	LLpSPGRQFRQ	LLpSPGRQFRQ	LLpSPGKQFRQ	LLpSPKKQFRQ	

Figure 9. Sequences and shorthand names for all phosphopeptide substrates in both series.

bound to BcCdc14.

High-throughput screening of the ChemBridge small molecule library previously performed by another group at Purdue University allowed for the isolation of a set of 10 potential inhibitors. By modeling these 10 potential inhibitor *in silico*, the efficacy of each molecule was predicted and experimentally ranked relative to one another using *in vitro* assays (**Appendix A**).

Initial assays with wild-type Cdc6ps7 substrate failed to generate conclusive data, likely due to the high affinity of enzyme for substrate. However, a general subset of relatively effective inhibitors from the original ten were selected to redo assays with pNPP substrate. These included 2'-hydroxy-1,1'-binaphthalene-3,4-dione (G7), 5-(4-isopropylbenzylidene)-2,4,6-(1H,3H,5H)-pyrimidinetrione (G5), and 5-(5-chloro-2-hydroxy-

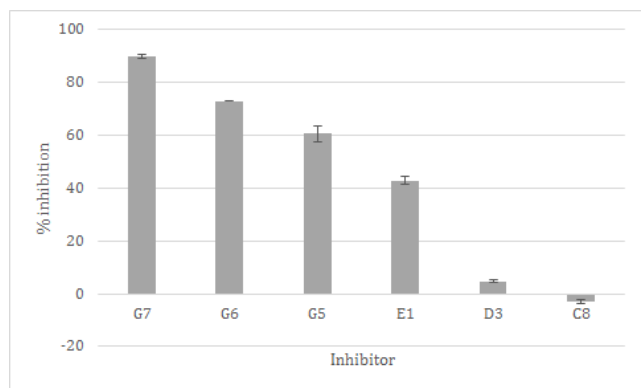


Figure 10. Results from assay performed for comparative effectiveness of inhibitor molecules, performed with pNPP as a substrate. Top three inhibitor picks for further testing are G7, G5, and E1.

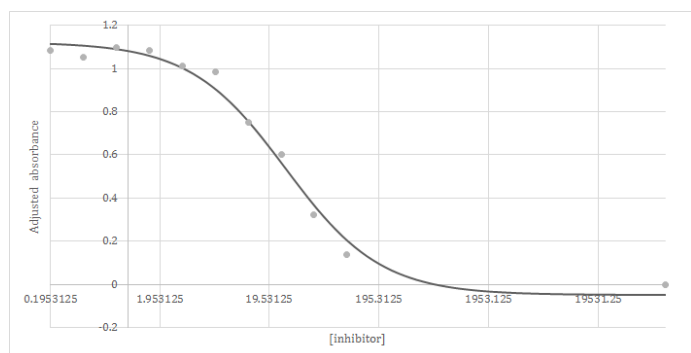


Figure 11. Generated IC₅₀ curve of inhibitor G5. An artificial data point was added at a high inhibitor concentration corresponding to 0 absorbance. The slope of the curve is well-defined. Sigmoid regression without Hill coefficient gives fit with an R² value of 0.9878.

3-nitrobenzylidene)-1,3-dimethyl-2,4,6(1H,3H,5H)-pyrimidinetrione (E1).

Identification and in silico modeling of candidate inhibitor molecules: These three relatively effective inhibitor molecules were experimentally confirmed to have the highest potential for competitive inhibitor optimization. Additional pNPP assays comparing the inhibiting activity of the three inhibitors at varying pNPP concentrations were performed to determine the molecules' IC₅₀ values. G5 had the lowest IC₅₀ value of approximately 27 μM (**Figure 11**). IC₅₀ values of the two other inhibitors, G7 and E1, were calculated to be 16.10161 μM and 25.09859 μM, respectively.

Although G5 did not possess the lowest IC₅₀ value, data collected for G5 reactions appeared the most well-behaved. Correlation of regression lines through collected data points was low for all inhibitors, but significantly higher for that of G5.

Concentrations selected in this IC₅₀ range centered on this value yielded potential concentrations to continue mechanism characterization. Confirmation of a competitive inhibitory mechanism is crucial to further develop an inhibitor that binds at the active site of BcCdc14.

Increasing apparent K_m values and decreasing V_{max} correlated with increasing concentrations of G5 imply a mixed mechanism of inhibition. However, previous studies performed on *Fusarium graminearum* have indicated

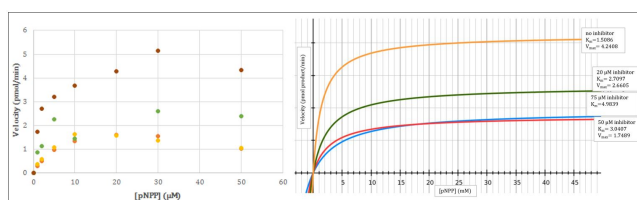


Figure 12. Regression curves plotting reaction velocity for assays completed with varying substrate concentrations with 75 μM inhibitor (blue), 50 μM inhibitor (red), and 20 μM inhibitor (green) added. Standard Michaelis-Menten curve for 2 μM BcCdc14 with no inhibitor is shown in orange. Lowest R² value of the curves shown is 0.9357.

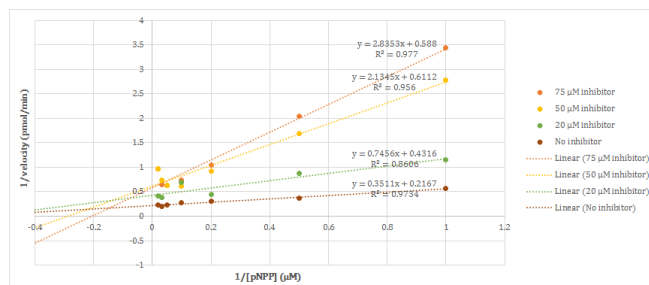


Figure 13. Lineweaver-Burk plots for uninhibited enzymatic activity and enzymatic reactions with varying concentrations of inhibitor. Axis intercepts vary, implying that apparent K_m and V_{max} have both changed, a behavior inconsistent with expectations of a competitive inhibitor.

that the inhibitor molecules used in this experiment behave in a competitive manner. Differences in the inhibitory mechanisms of these molecules may be due the miniscule variances in protein structure. Further work may be required to confirm the inhibiting mechanism of inhibitor molecules for BcCdc14.

Designing an inhibitor docked in the active site first required a homology model of BcCdc14 to be created. This was accomplished in MOE, using the crystal structure of *Homo sapiens* Cdc14b as a template. Optimization and construction of the inhibitor G5 was completed in examining the active site of the BcCdc14 homology model. Efficacy and improvements were measured by the ligand affinity scores generated by MOE.

Considering feasibility of organic synthesis, the final weight of the modified substrate was limited to twice the weight of the original substrate. The molecular weight of the original substrate is 258.28 Da while the modified substrate's molecular weight is 468.57 Da—complying with the molecular weight criteria.

The most significant modifications to the substrate were the reduction of three ketone groups within the center phenyl ring to an overall triol. This allowed for the conversion of three hydrogen bond accepting groups to dual-acting hydrogen bond accepting and donating groups. Another significant modification was the addition of amine groups throughout the rest of the substrate in order to form cation- π , dipole-dipole, and hydrogen bonds with the enzyme's active site.

Figure 17 (actually Fig. 8.3)

4. DISCUSSION

From our work with BcCdc14 and previous literature on Cdc14, we conclude that Cdc14 is a strong candidate for fungal growth inhibition through inhibitor drug

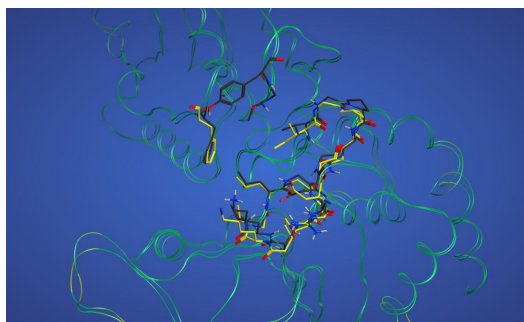


Figure 14. Homology model of BcCdc14 as generated by MOE, with *Homo sapiens* Cdc14 (1OHD) as the template structure. Additional information is available in **Appendix A**.

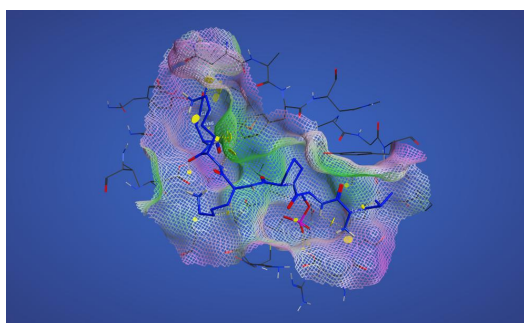


Figure 15. Model of BcCdc14 with optimal phosphopeptide ligand (sequence Leu-pSer-Pro-Lys-Lys-Gln) docked in active site, as predicted by AutoDock. Hydrophilic surfaces are colored in pink, while hydrophobic surfaces are colored in green.

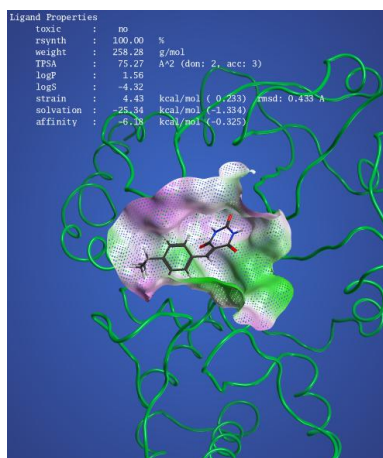


Figure 16. Original model of unmodified G5 ligand. Affinity score as given by MOE is -6.18 kcal/mol.

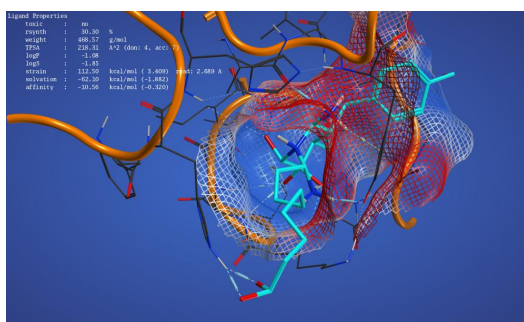


Figure 17. Modified and optimized G5 ligand. Affinity score as given by MOE is -10.52 kcal/mol, a 1.7-fold increase in affinity score.

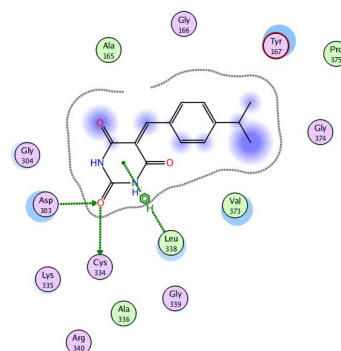


Figure 18. 2D interaction diagram of the original, unmodified G5 ligand. Hydrogen bonds are indicated by dashed lines.

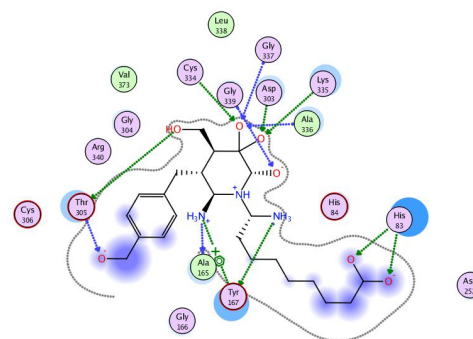


Figure 19. 2D interaction diagram of optimized ligand with hypothetical BcCdc14 homology model.

molecules. The ramifications of such applications are evident in agriculture: inhibiting the activity of Cdc14 in fungal pathogens would interfere with the pathogens' growth and thus prevent the destruction of plant crops. Selective killing of fungal pathogens is achievable as plant crop species' *Cdc14* orthologs are significantly dissimilar to that of fungal pathogen Cdc14, both mechanistically and structurally.

The optimal substrate prediction of BcCdc14 was searched and compared against other similar substrates found in fungal organisms. The results of this search yielded enzymes with potentially similar activity as BcCdc14, particularly in *Saccharomyces cerevisiae*. Tyrosine protein kinase SWE1, found in *Saccharomyces cerevisiae*, regulates mitotic transition through G2 and M stages by modulating Hsp90 activity in chaperoning^[10]. Based on high sequence similarity between the substrates of these two enzymes, BcCdc14 may have a similar function at this stage in the mitotic cycle.

EXG1 is similarly involved in cell replication, monitoring cell wall beta-glucan construction in *Saccharomyces cerevisiae*^[10]. Predicted analogous activity in BcCdc14 indicates that inhibition of BcCdc14 would prevent cell wall construction during fungal cytokinesis, killing the organism. These two hypothesized functions of BcCdc14 in the cell division cycle make BcCdc14 an excellent target for selective destruction.

Several complications were encountered in the duration of the project. Initial experimental assays with the substrate Cdc6ps7 yielded poor results, as raw absorbance



Figure 20. Sequence alignment of BcCdc14 model and *Homo sapiens* Cdc14b. Active site regions show high conservation. However, it should be noted that the *Homo sapiens* Cdc14b sequence is shorter than the *BcCdc14* and thus leaves room for errors within the structural prediction model.

readings giving degree of inhibition remained far too low and were unable to be differentiated from background noise. The low precision and accuracy of spectrophotometry equipment also contributed to poor readings. The relatively high binding affinity of Cdc6ps7 to BcCdc14 may have been the cause for insufficient levels of inhibitor activity. To resolve the issue, assays were performed with pNPP, an artificial substrate with a much lower affinity for BcCdc14.

Prior to *in vitro* trials, *in silico* modeling with MOE and AutoDock provided predictions for the relative inhibitory efficacy of the inhibitor molecules. However, given limited computational power and time, predictions were not wholly accurate. Fortunately, prediction was limited to the qualitative level and thus did not significantly affect data analysis.

5. APPENDIX A: MODELING AND COMPUTATIONAL METHODS

In this project, modeling softwares such as MOE and AutoDock were used extensively to computationally predict relative ligand affinity. To this end, a homology model for BcCdc14 was initially constructed by utilizing the "Homology Model" function in MOE. The known crystal structure of human Cdc14b phosphatase (1OHD) was used to construct a sequence alignment and to create a hypothetical protein structure model of BcCdc14^[11].

For areas of no alignment between *BcCdc14* and *Homo sapiens* Cdc14b, MOE runs a loop prediction search, iterating through possible structures of loops approximately the same length of the gap, and isolates the segment loop structure, giving an optimal energy minimization score. Several criteria may be used in the matching and selection of a segment match, including basic amino acid sequence similarity and alignment, conformational similarity as defined by atomic coordinates within the greater structure, and close-distance compatibility with the target structure^[12]. In particular, standard segment matching generates results and likely matches based on random and averaged initial estimates.

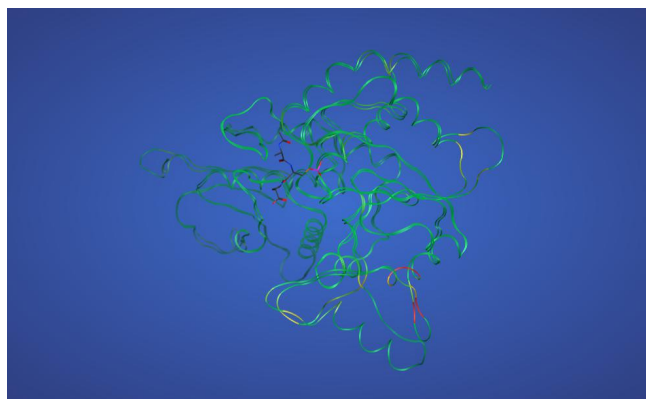


Figure 21. Resulting homology model of *BcCdc14* superimposed upon original template of *Homo sapiens* Cdc14b. Ribbon structures are colored by RMSD value; areas of strong consensus are colored in green. Unaligned loop structures carry more variance and so carry a higher RMSD value. A sample ligand from the *Homo sapiens* template Cdc14b structure points to the active site of both enzymes.

Thus, a large enough sample size of initial conditions and estimates, drawn from already-known structures, will result in a fairly accurate model of a hypothetical structures. While the computational power of MOE itself is not sufficient to generate models with exact matching, it should be noted that not all of the domains of a proteins structure must be conserved to guarantee functional enzymatic activity. Indeed, Cdc14 has a highly variable c-terminus region yet preserves function. Given the strong conservation of the active site amino acid sequences between the *BcCdc14* sequence and human Cdc14b sequence, it is reasonable to hypothesize that MOE models the active site sufficiently accurately.

The unmodeled BcCdc14 sequence and previously-modeled *Homo sapiens* Cdc14b sequences were aligned and superimposed, and found to have a root-mean-squared deviation value of 1.04. A homology model algorithm was run through MOE, creating structurally identical sections where matching sequences of similar amino acid sequence are found, and selecting optimal loop searches to fill areas of dissimilar alignment.

Once the hypothetical homology model of the enzyme was constructed, *in silico* modeling began. As this project focused on the docking mechanism and effects of inhibitor ligands, modeling primarily began with ten small ligand compounds. High-throughput screening conducted *in vitro* by another group had previously identified these compounds as likely targets for enzymatic inhibition. Models were manually constructed in MOE, based on the known chemical structures (Table A1).

All constructed models of ligands and the homology model were electrostatically minimized using MOE's "Minimize" function, utilizing the Amber12 force field to optimize atomic interactions. The "Protonate 3D" model was executed to add hydrogen atoms to the homology model, a crucial step to account for all possible hydrogen bonding and van der Waals forces in the active site.

MOE's included docking feature provided a very convenient and efficient manner of electronically simulating the docking of small ligands; despite lacking enough computing power to dock large ligands that are structurally variable and retain freedom of movement in the active site,













Shorthand reference	G6	G6	E5	G3	G5	D3
Name	sodium 3,4-dioxo-3,4-dihydro-1-naphthalenesulfonate	2-[(2,5-dimethyl-1H-pyridyl-yl)imino]-1,3-thiazolidin-4-one	2-amino-4-(2-chlorophenyl)-7-hydroxy-4H-chromeno[3,2-c]carbazole	N-[[1-(4-(6-methylphenoxy)butyl)benzimidazol-2-yl]methyl]furan-2-carboxamide	5-(4-isopropylbenzylidene)-2,4,6-(1H,3H,5H)-pyrimidinetrione	2-[5-(2-bromophenyl)-2-furyl]acrylic acid
Structure						
Shorthand reference	D9	G8	G7	E1	F5	A9
Name	3-[4-(methyl-3-(1-pyrrolidinylthio)phenyl)acrylic acid	3-(3,4-dethoxy-5-iodophenyl)acrylic acid	2-hydroxy-1,1'-bi[2-naphthalene]-3,4-dione	5-(5-chloro-2-hydroxy-2-aziridinylidene)-1,3-dimethyl-2,4,6-(1H,3H,5H)-pyrimidinetrione	4-[[4-(2-thienyl)-1,3-thiazol-2-yl]amino]benzoic acid hydrobromide	2-amino-4-(2,3-dichlorophenyl)-7-hydroxy-4H-chromeno[3,2-c]carbazole
Structure						

Figure 22. A list of twelve inhibitor molecules ligands constructed in preparation for docking simulations.

MOE is adequate for computationally modeling small ligands that have relatively few rotational possibilities. All simulations were thus completed with the "rigid receptor" option to reduce computational time.

Initially, 10 iterative poses were docked in MOE to generate 10 independent docking poses. Each pose was scored by the free energy of the ligand docking in the receptor, given by summing all intermolecular van der Waals and electrostatic interactions between the two complexes^[13].

The given scores were used to hypothetically rank the experimental efficacy of each ligand. No further experimentation or computations were completed with this data, with the exception of the most effective inhibitor ligand, determined experimentally.

The experiments described above determined that the ligand G5 held the lowest IC₅₀ value and was thus a good candidate for inhibitor molecule optimization. As this project will be unable to reach the synthesis and testing of this new ligand, all modeling and simulations of the hypothetical inhibitor were designed electronically (**Figures 18 and 19**).

The top-scoring ligand docking pose of G5 was kept. Docking this pose in the homology model BcCdc14 allowed for a platform for adding additional atomic structures to the G5 ligand to optimize ligand-receptor affinity. Optimization of G5 yielded an increase from -6.18 kcal/mol to -10.52 kcal/mol in docking score. This was accomplished by addition and removal of functional groups, such as hydroxyl and amine groups, to the structure of the ligand. For feasibility purposes in organic synthesis of G5, the molecular weight was increased no more than two fold, from 258.28 Da to 468.57 Da.

In addition to MOE, AutoDock Vina was also utilized. As MOE is computationally unable to provide accurate docking scores for larger, more flexible ligands, this powerful freeware was utilized to hypothetically dock larger ligands in the active site of the homology model. These larger ligands were manually constructed models of both the experimentally-determined most effective phosphopeptide substrate (LpSPKKN) and hypothetically-determined most effective phosphopeptide substrate (GpSPKKN).

The simulations in AutoDock Vina yielded several possible ligand poses in the active site. As the catalytic Cysteine of the active site is known to bind with phosphorylated serine, only models which had this particular Cysteine residue were kept. Modeling these docked ligands aided in illustrating the efficacy of enzyme binding, as well as the importance of the phosphorylated

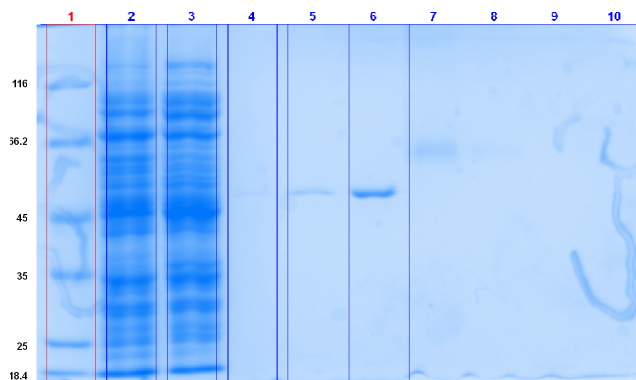


Figure 23. SDS-PAGE image analysis with molecular weight standards in lane 1, pre- and post- purification samples in lanes 2 and 3, BcCdc14 dilutions in lanes 4, 5, and 6, and bovine serum albumin dilutions in lanes 7, 8, 9, and 10.

amino acid coupled with basic residues at key +2 to +5 regions in hypothetical peptide substrates.

6. APPENDIX B: DETERMINING PURITY AND CONCENTRATION

The purity and concentration of the working stock of BcCdc14 were evaluated with two methods: by Bradford assay and SDS-PAGE.

Evaluation by Bradford assay was completed by creating samples of protein stock and staining with Bradford dye. Absorption value read from the spectrophotometer over several trials were 0.195, 0.211, and 0.23 for a 0.5 % dilution of protein stock, and 0.352, 0.334, 0.367 for a 0.8 % / dilution of protein stock.

Evaluations were completed with the following standard curve, generated with bovine serum albumin:

$$\text{Concentration} = \frac{\text{absorbance} - 0.02055}{\frac{0.05318}{\mu\text{L of solution}}}$$

Calculations for concentrations yielded an average of 3.87291 $\mu\text{g}/\mu\text{l}$ of protein, or:

$$\left(\frac{3.87291\mu\text{g}}{1\mu\text{L}}\right)\left(\frac{1000\mu\text{L}}{1\text{mL}}\right)\left(\frac{1000\text{mL}}{1\text{L}}\right)\left(\frac{1\mu\text{M}}{0.049375\text{g}}\right)\left(\frac{1\mu\text{g}}{10^6\mu\text{g}}\right),$$

which gives 78.514 μM stock solution.

Results from this method of calculating concentration are considered reliable. Spectrophotometric readings were in a linear range (absorbance readings of 0.2 to 1.0) characterized by the range of the bovine serum albumin standard curve (up to absorbance readings of 0.55). Dilutions and calculations of both the BcCdc14 protein stock and bovine serum albumin were completed repeatedly by individual team members.

The concentration of the gel was also evaluated by SDS-PAGE. A gel was run under identical conditions, with identical dilutions of estimated concentration of BcCdc14 protein stock and known concentration of bovine serum albumin.

As an initial visual assessment, bands of BcCdc14 are somewhat defined and compact, while the highest concentration (approximately 1 $\mu\text{g}/\mu\text{L}$, lane 6) are more defined and dense. However, bands with varying concentrations of bovine serum albumin (0.25 $\mu\text{g}/\mu\text{L}$, 0.05 $\mu\text{g}/\mu\text{L}$, 0.025 $\mu\text{g}/\mu\text{L}$, 0.01 $\mu\text{g}/\mu\text{L}$, and 0.005 $\mu\text{g}/\mu\text{L}$) were

much too faint to be analyzed.

ImageLab analysis determined the amount of particles detected in each band, visually calculating percent composition of particles in each band. Comparing percent composition between bands of known concentration (bovine serum albumin dilutions) determined the concentration of BcCdc14 to be approximately 30 μ M.

This value is too low and significantly differs from values obtained from the Bradford assays. A number of factors may have contributed to this. Firstly, purity of the bovine serum albumin must be considered; although identical samples were used for the Bradford assay and SDS-PAGE gel, dilutions and pipetting error associated with adding small volumes of solution into gels may have greatly influenced concentration yield, which can be remedied by increasing dilution concentrations. Nevertheless, the Bradford assays are a much more reliable predictor of concentration. The imaging capabilities of the camera used must also be considered; as the camera may not have been powerful enough to detect faintly-stained particles, many calculation errors can occur.

Another group within the Biochemistry department of Purdue University performing parallel experiments and procedures on *Ustilago maydis* Cdc14 experienced many of the same issues in analyzing the SDS-PAGE gel.

7. ACKNOWLEDGEMENTS

The research team would like to thank Dr. Mark Hall and Dr. Stefan Paula for their devotion to the Summer Science Program. Our appreciation for your dedication and patience cannot be expressed.

We would also like to thank Purdue University, the Summer Science Program Board of Trustees, the Gordon and Betty Moore Foundation, and everybody who has made this experience possible.

Thank you to those who believe in the advancement of education and plentiful opportunities for all. Your work has been ineffably impactful for countless individuals.

8. REFERENCES

- [1] Dean, R., Van Kan, J.A., Pretorius, Z.A., Hammond-Kosack, K.E., Di Pietro, A., Spanu, P.D., Rudd, J.J., Dickman, M., Kahmann, R., Ellis, J., and Foster, G.D., (2012). The Top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol.* 13(4): 414-30.
- [2] Kerk D, Templeton G, Moorhead GB. Evolutionary radiation pattern of novel protein phosphatases revealed by analysis of protein data from the completely sequenced genomes of humans, green algae, and higher plants. *Plant Physiol.* 2008;146(2):351-67.
- [3] Bremmer SC, Hall H, Martinez JS, et al. Cdc14 phosphatases preferentially dephosphorylate a subset of cyclin-dependent kinase (Cdk) sites containing phosphoserine. *J Biol Chem.* 2012;287(3):1662-9.
- [4] Yellman CM, Roeder GS. Cdc14 Early Anaphase Release, FEAR, Is Limited to the Nucleus and Dispensable for Efficient Mitotic Exit. *PLoS ONE.* 2015;10(6):e0128604.
- [5] Gray CH, Good VM, Tonks NK, Barford D. The structure of the cell cycle protein Cdc14 reveals a proline-directed protein phosphatase. *EMBO J.* 2003;22(14):3524-35.
- [6] Subramaniam, S. (1998) The Biology Workbench—a seamless database and analysis environment for the biologist. *Proteins*, 32, 1-2.
- [7] Griffiths AJF, Gelbart WM, Miller JH, et al. *Modern Genetic Analysis*. New York: W. H. Freeman; 1999. Cloning a Specific Gene. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK21450/>
- [8] Ernst O, Zor T. Linearization of the Bradford protein assay. *J Vis Exp.* 2010;(38)
- [9] Robichon C, Luo J, Causey TB, Benner JS, Samuelson JC. Engineering *Escherichia coli* BL21(DE3) derivative strains to minimize *E. coli* protein contamination after purification by immobilized metal affinity chromatography. *Appl Environ Microbiol.* 2011;77(13):4634-46.
- [10] Raspelli E, Cassani C, Chirolì E, Fraschini R. Budding yeast Swe1 is involved in the control of mitotic spindle elongation and is regulated by Cdc14 phosphatase during mitosis. *J Biol Chem.* 2015;290(1):1-12.
- [11] Ochoa R, Watowich SJ, Flrez A, Mesa CV, Robledo SM, Muskus C. Drug search for leishmaniasis: a virtual screening approach by grid computing. *J Comput Aided Mol Des.* 2016;
- [12] Levitt M. Accurate modeling of protein conformation by automatic segment matching. *J Mol Biol.* 1992;226(2):507-33.
- [13] Jain AN. Scoring functions for protein-ligand docking. *Curr Protein Pept Sci.* 2006;7(5):407-20.