Research of *Ustilago maydis* Cdc14 as a potential target for inhibition to be applied as an antifungal solution for multiple fungal pathogens

Leyna Duong, Sabra Dunakey, and Ketan Jog

From the Purdue University Department of Biochemistry and the Summer Science Program

Running Title: Development of *Ustilago maydis* Cdc14 inhibitor

Keywords: Cdc14, *Ustilago maydis*, fungal pathogens, enzyme inhibitor, dual-specificity kinase

Abstract

Cell division cycle protein 14 (Cdc14) serves an important role in cell division as a required enzyme for mitotic exit and previous research has shown that it is responsible for the pathogenic properties of fungal pathogens such as *Ustilago maydis*. This fungal pathogen is one among many that infect important food crops such as maize, reducing annual yields, calling for a constant need of new pesticides as fungal pathogens develop resistance to current treatments. Previous research done on Cdc14 in different fungal species has shown that without Cdc14 the fungi are noninfectious, which makes it a reasonable target for a fungicide. Multi-sequence alignment has shown that the binding site of Cdc14 is conserved among many species, including *homo sapiens*. The objective of our research was to purify, isolate, and study the Cdc14 enzyme from *Ustilago maydis* to develop an inhibitor that can potentially be effective for a wide range of fungal pathogens. By comparing the sequence to the human Cdc14B enzyme, we were able to generate a model of our enzyme to study its binding site and develop an inhibitor. Our experiment on the substrate selectivity of UmCdc14 confirmed the conservation of the binding site with human Cdc14, which allowed us to make the comparison for the homology model. After testing multiple inhibitors, the most effective one that caused the highest percent of inhibition of the enzyme function was determined to act as a suicide inhibitor. This gave us a starting point when we optimized the molecule and developed a Cdc14 inhibitor that had the potential to stop the parthenogenesis of fungal pathogens and save our essential crop yields.

Introduction

The world corn consumption for 2016-2017 was 40,429 million bushels (1 bushel = 1.25 cubic foot)[(1)]. The United States alone consumed 12,631 million bushels last year[(2)]. With just 1% surplus produced, it is apparent that the demand for corn is cutting it close to the production. The world needs to produce at least 50% more food to feed 9 billion people by 2050[(2)]. About 40% of plant produce in the world is lost to disease each year[(3)]. Crops are afflicted by fungi, bacteria, viruses, and insects that drastically reduce yields. Fungal pathogens constantly evolve to develop resistance to disease-resistant treatments. These crop diseases pose a constant threat to our food supply. Thus, additional treatment options, including novel antifungal targets are constantly needed to limit crop damage from fungal pathogens.

We specifically look at Common smut, a disease that occurs wherever corn (Maize) is grown[(5)]. Susceptibility to biotrophic pathogens is the primary constraint to increasing productivity of Maize. *Ustilago maydis* is one such biotrophic fungal pathogen and the causal agent of corn smut on maize. Just in the U.S, this disease is responsible for significant yield losses of approximately $1.0 billion annually[(12)]. *U. maydis* survives on crop debris and on the soil[( )]. It can infect any tissue of the plant by entering through wounds and forming characteristic smut galls(4).

The fungus can also enter through the silks, causing gall formation on the ear tip(5).
In search of solutions for these fungal infections, Cdc14 has been identified as an attractive target for potential antifungal drug development, for the following reasons: Firstly, the genetic deletion of CDC14 severely retards growth and eliminates pathogenicity of *F. graminearum* and *M. oryzae*[(6)]. It plays a key role in the process of mitosis (cell division). Secondly, this protein has been extensively studied in *Saccharomyces Cerevisiae.* CDK1 activity is inhibited by APC-mediated destruction of cyclin B and activation of the CDK1-antagonizing phosphatase Cdc14, which allows reformation of nuclear envelope and entry to G1(7). Thirdly, CDC14 phosphatase genes are conserved in fungi and animals, but are absent from higher plants(8). Thus, plants are likely to be unaffected by compounds that specifically inhibit Cdc14.

Apart from the aforementioned organism-specific studies, there is also structural evidence for Cdc14 to be a good antifungal drug target: Unlike most serine/threonine phosphatases, Cdc14 is an enzyme whose substrate selectivity appears to be entirely dictated by the structure around its catalytic site, suggesting that the design of highly specific competitive inhibitors should be achievable[(9)]. Cdc14 shows a preference to phosphoserine over phosphothreonine. It also has a high affinity for substrates with Proline at the +1 position and basic residues like Lysine at +3 to +5 positions from the phosphorylated Serine[(10)].

Another advantage of using Cdc14 as a target is that the gene sequence of Cdc14 is highly conserved across species, including humans and various fungal pathogens. Thus, an inhibitor designed to target Cdc14 in *U. maydis* using its structure could potentially work against other pathogens as well. The Cdc14 ortholog in *U.Maydis* hasn't had its crystal structure developed yet. This can be sidestepped by the fact that the very similar human Cdc14 protein has had its crystal structure discovered, and can be used to approximate a model and study our target protein.

In this project we will set up various experiments to determine the structure, function, activity, substrate specificity, inhibitor affinity and inhibition mechanism of the Cdc14 enzyme to be purified and isolated from E.coli, which will be used to mass produce it. The ultimate goal is to design small molecules that would potentially inhibit it.

Results:

Multiple sequence alignment of Cdc14 shows conservation across species

To confirm the fact that UmCdc14 was conserved with other Cdc14 orthologs, so we performed a multiple sequence alignment of our Cdc14 gene sequence obtained from *U. Maydis* against the Cdc14 ortholog sequences from *Homo Sapien, Drosophila, Rhizoctonia Solani, Magnaporthe Oryzae,* and *Aspergillus Niger* (Figure 2 (B)). We purified and extracted UmCdc14 at a 275 uM concentration.

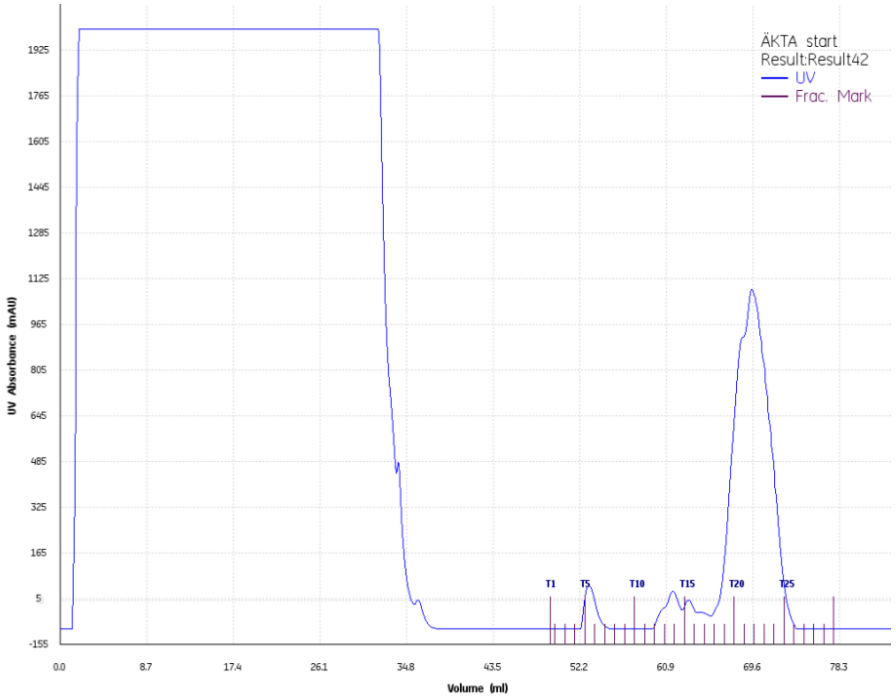


Figure 1 Bacterial transformation and protein purification: (A) The protein was purified using the Akta Method, and extracted from tubes 19-24.

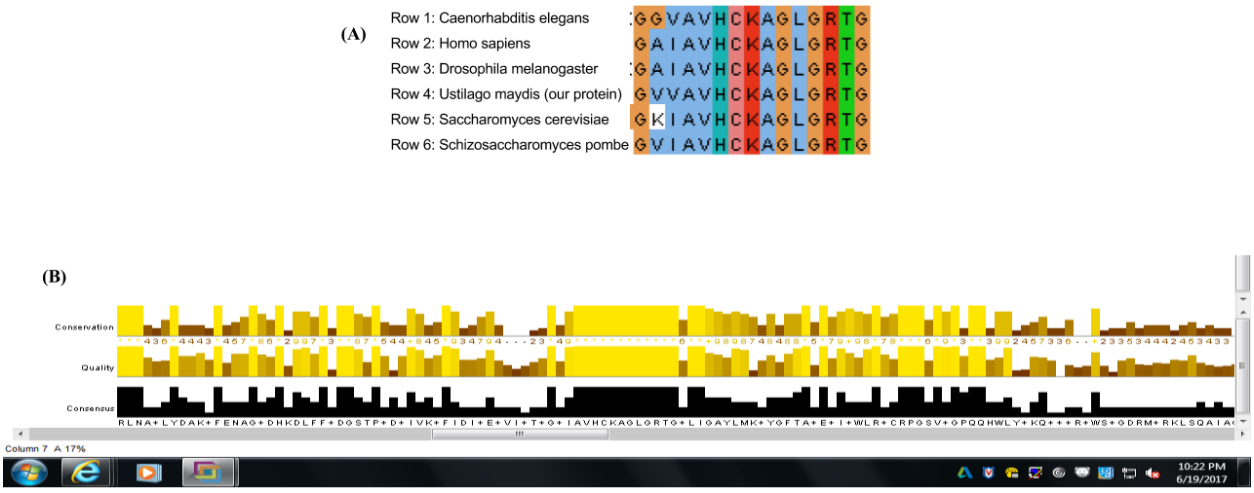


Figure 2 Multiple Sequence Alignment for Cdc14: (A) The conserved active site of CDC14 for various species (B) The sequence alignment for Cdc14 in 6 different species, showing conservation of structure.

Steady State Kinetic Assays show show linearity, and verify function of UmCdc14

We hypothesised that UmCdc14, like its orthologs would be a dual specificity phosphatase, based on the fact that it's active site was conserved. It followed that its activity would be reduced by standard phosphatase inhibitors. A linearity versus time experiment was conducted, measuring the amount of PNP produced from pNPP over time. The steady-state time frame was just over 80 minutes (Figure 3 (B)). Then the steady state assays to measure enzyme activity with and

without the sodium orthovanadate inhibitor were performed. Km of 4.343 mM 1 ,Vmax 0.9 uM/min and a Kcat of 0.2 values were determined (Figure 3 (C)). The pNPP when reacting with UmCdc14, and also showed inhibition from Sodium Orthovanadate, validating that the recorded enzyme activity was indeed from a tyrosine kinase and not from other contamination (Figure 3 (A)). Specific activity was found to be 2.5 (pmol [P]/(pmol [E]*min)). Thus, we confirmed that UmCdc14 belong to the phosphatase family.

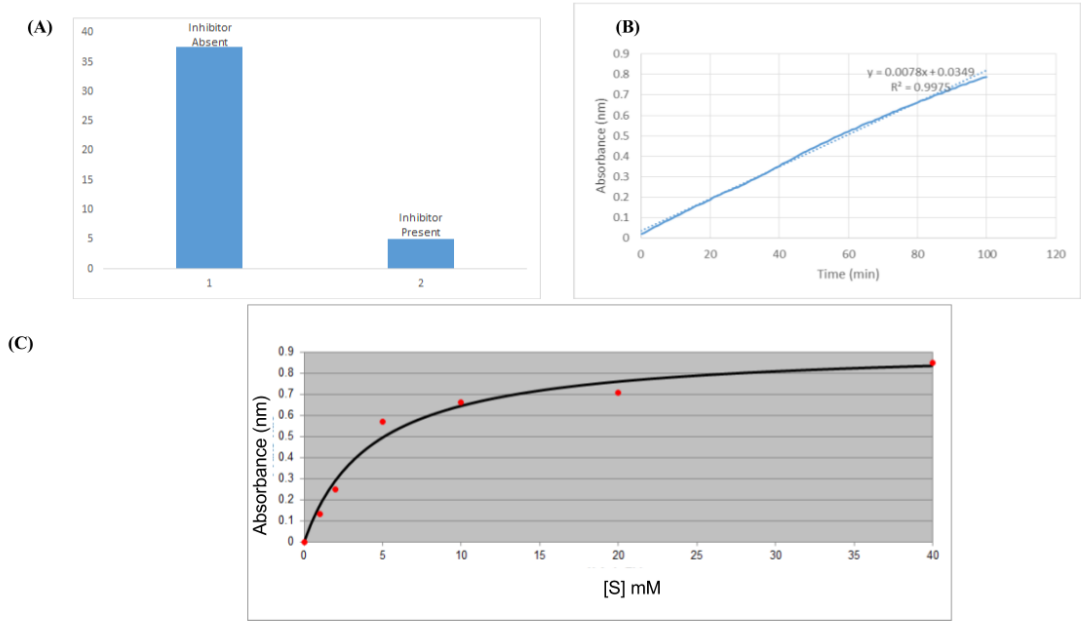
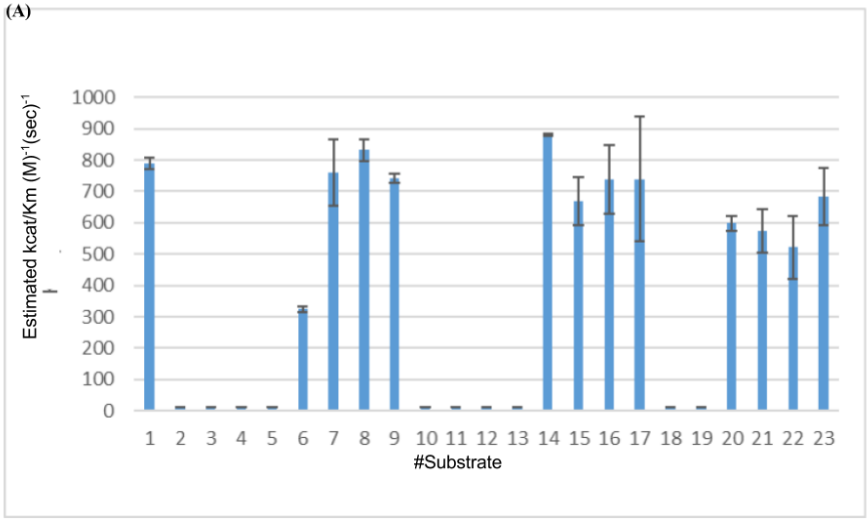


Figure 3: (A) Standard phosphatase inhibitor Sodium orthovanadate significantly reduces enzyme activity. (B)) Measuring the steady state time for reaction of UmCdc14 with pNPP. (C) Michaelis menten curve for pNPP with UmCdc14 for the Steady State Assays.

Phosphatase Assays determine substrate specificity of UmCdc14 to be conserved between HsCdc14

The next set of experiments determined the substrate specificity of UmCdc14. Literature suggested HsCdc14 to prefer phosphorylated Serine over Threonine, also having higher affinity with a +1 position Proline and basic residues at +3 to +5 positions [10]. Using multiple sequence alignment, we verified that the active site of HsCdc14 was conserved between UmCdc14, and hence predicted the same properties for UmCdc14 (Figure 2 (A)). Accordingly, 24 different different substrates, differing by only a single residue from HT{pS}PIKSIG were selected and their interactions with the enzymes were examined by measuring amount of free phosphatase after reaction was run for 20 minutes (Figure 4 (B)). Three trials were conducted. The catalytic efficiency of each substrate was estimated using catalytic efficiency = [V]/([E][S]). The results confirmed the hypothesis. Substrates without phosphorylated serine or at least one basic residue in the +3 to +5 positions showed catalytic efficiencies under 10, whereas others had catalytic efficiency values of over 500 (Figure 4 (A)). Thus the substrate specificity of UmCdc14 enzyme was established.



(B)

#Index	Sequence
1	HT{pS}PIKSIG
2	HT{pT}PIKSIG
3	HT{pY}PIKSIG
4	HT{pS}AIKSIG
5	HT{pS}PIASIG
6	HT{pS}PIRSIG
7	HT{pS}PIKKIG
8	HT{pS}PKKSIG
9	HT{pS}PIKSKG
10	HT{pS}PKASIG
11	HT{pS}PIAKIG
12	HT{pS}PIASKG
13	HT{pS}PIASIK
14	AT{pS}PIKSIG
15	HA{pS}PIKSIG
16	HT{pS}KIKSIG
17	HT{pS}PEKSIG
18	HT{pS}PIKEIG
19	HT{pS}PIKSEG
20	HT{pS}PIKSIE
21	HT{pS}PIKSIK
22	KT{pS}PIKSIG
23	HK{pS}PIKSIG
24	HT{pS}PIKKRG

Figure 4:Phosphatase assay measuring release of free phosphate using malachite green dye and a spectrophotometer. (A) Catalytic efficiency of substrates with their standard deviations shows the selectivity of UmCdc14 according to composition. (B) Substrate number in (A) corresponds to the amino acid sequence.

Molecular modelling shows crystal structure of UmCdc14 similar to HsCdc14 and indicated conserved docking site

Using the Molecular Operating Environment (MOE), the homology model of our UmCdc14 ortholog was created (Figure 5 (A)). The crystal structure of the 1OHE, which contained the structure of human Cdc14b phosphatase protein with a peptide ligand was used to design it. Using the pre-established peptide ligand, its binding site was defined.We superimposed the human cdc14b phosphatase onto the homology model. Then we docked the peptide ligand into the active site of the homology model of the UmCdc14 using rigid fit, with the same existing peptide to see whether docking gave us the same fit (Figure 5 (B)). The results showed an almost exact overlap, supporting the authenticity of homology model docking. The Phi-Psi plot confirmed that the designed homology model was practical and realistic since almost all residues lay within the allowed regions (Figure 5 (C)).

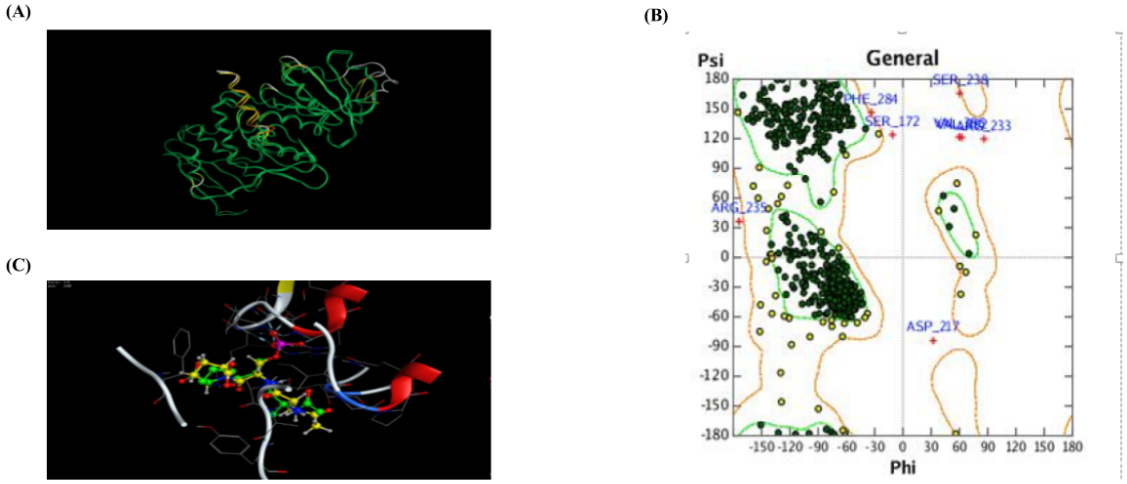
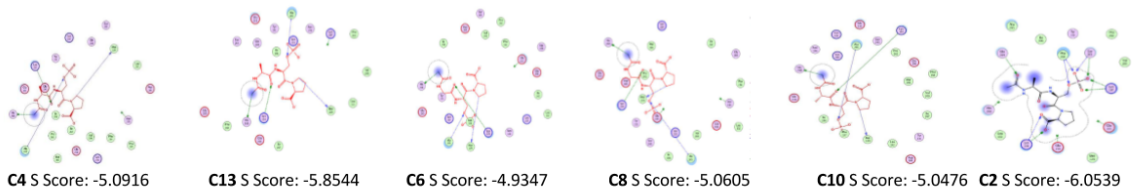


Figure 5: (A) Superposition of the homology model of UmCdc14 with the human Cdc14b shows almost exact resemblance. (B) Phi-Psi plot of UmCdc14 homology model shows under 10 outliers. supporting the authenticity of the model. (C) Re-docking peptide ligand from Human Cdc14b phosphatase into Homology Model of Umcdc14 to shows almost exact overlap, indicating successful docking in the homology model.

Inhibition Assays determine Compound 4 to be most viable for UmCdc14 inhibition

Unpublished data from a large scale screen for inhibitors of cdc14 in budding yeast was used. 13 inhibitors were selected and sequentially docked into the active site of the homology model, using the rigid fit algorithm.The best 6 models were selected based on the 'S-score' scoring function in MOE (Figure 6 (A)). Subsequently, the 13 inhibitors were tested in-vitro to calculate percent inhibition. The top 6 inhibitors were listed based highest value of inhibition (Figure 6 (B)). With only a ten fold difference between our enzyme concentration and inhibitor concentration, our highest inhibition was approximately 70%. Compound 2, 4, and 6 were the top three inhibitors with the highest percentage of inhibition. They all also belonged to the top 6 inhibitor list from the in-silico docking experiment. So we moved them to the next stage of experiments for IC50 determination. After performing the experiment, it was determined that of the three inhibitors selected, only Compound 4 displayed signs of inhibition of the UmCdc14 enzyme activity (Figure 6 (C)). We obtained a Ki of 24 uM with an IC50 value of 40 uM (Through microsoft excel exponential regression)

(A)



(B)

#Inhibitor	% Inhibition
C2	70.44
C4	64.69
C11	59.38
C6	66.39
C11	57.86
C8	50.4

(C)

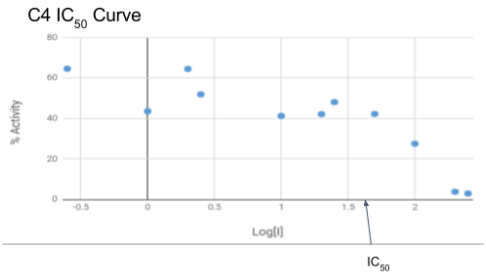


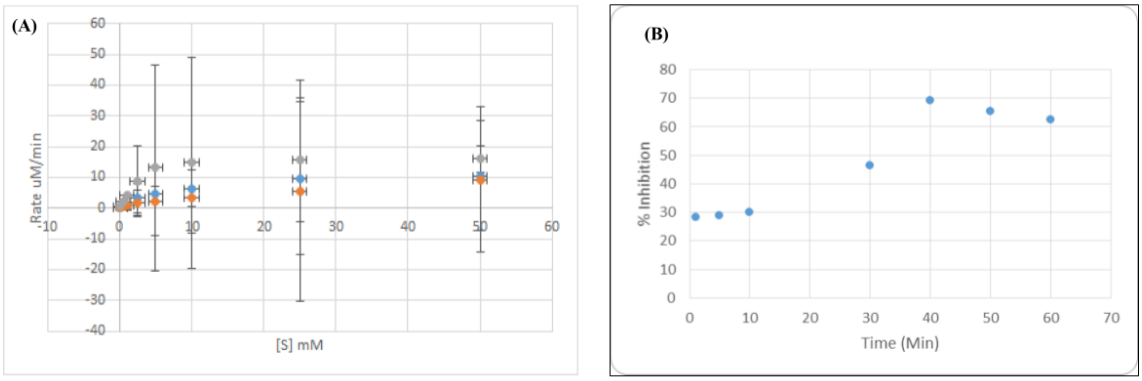
Figure 6: (A) Ligand dockings *(in-silico)* shows predicted molecular interactions between enzyme residues and inhibitor molecules. (B) Results of inhibition assays show 4 inhibitors that also belong to the in-silico screen. (C) IC50 curve of Compound 4 shows significant inhibition

Experiments for determination of mechanism of inhibition indicate suicide inhibition

After choosing Compound 4 as a potential inhibitor against UmCdc14, we examined its mechanism of inhibition. Two Inhibitor concentrations were selected along with a control of no inhibitor. The substrate concentrations were varied over these respective inhibitor concentrations. Based on the IC50 curve data, we had predicted competitive inhibition over non-competitive inhibition. It was noted that the percent activity remained about constant at lower inhibitor concentrations but dropped as inhibitor concentration was significantly raised. However, the assays showed different, if not opposite results. There was a variation in Vmax and the Km, indicating noncompetitive or mixed inhibition (Figure 7 (A)). Another way to explain the data would be suicide inhibition, which meant that inhibitor would potentially covalently bind to the enzyme, taking it out of the equation entirely. This meant that the longer it took to add substrate to the master mix of inhibitor and enzyme, the less enzyme would it have to react with.
This required us to check whether the mechanism of inhibition was reversible or irreversible.

To confirm this hypothesis, we designed another experiment keeping inhibitor concentration constant over varying time to see whether the %inhibition would change. Now in reversible inhibition time doesn't play a factor and the graph would show a flat line, with a zero slope. The resultant graph, however was a peaking curve clearly confirming irreversible inhibition (Figure 7 (B)).

Both of these experiments strongly indicated prevalence of Suicide inhibition in Compound 4 with UmCdc14 enzyme.



	[I] 25 uM	[I] 50 uM	[I] 0 uM
Km	13	22	2
Vmax	11	13	17

Figure 7: (A) Three trials varying substrate concentration at 25uM and 50uM inhibitor concentration indicates noncompetitive inhibition. Data shows varying Vmax and Km (B) irreversible inhibition is confirmed by a variable %inhibition with varying time of inhibitor enzyme reaction.

Modelling hypothetical compound for UmCdc14 inhibition using Compound 4 as model increases binding affinity in-silico

Using the docking file from the in-silico inhibition assays, we re-docked compound 4 into the homology model of the UmCdc14 enzyme. The initial binding affinity of compound 4 was -5.8 (Figure 8(A)). After modifying it to maximize contact with the active site of UmCdc14, we were able to raise the binding affinity to -8.7 (Figure 8(C)). The ligand interactions show increase in the number of hydrogen bonds. The hypothetical compound also fits more completely in the active site (Figure 8(D)). This in-silico design proved to be the best inhibitor.

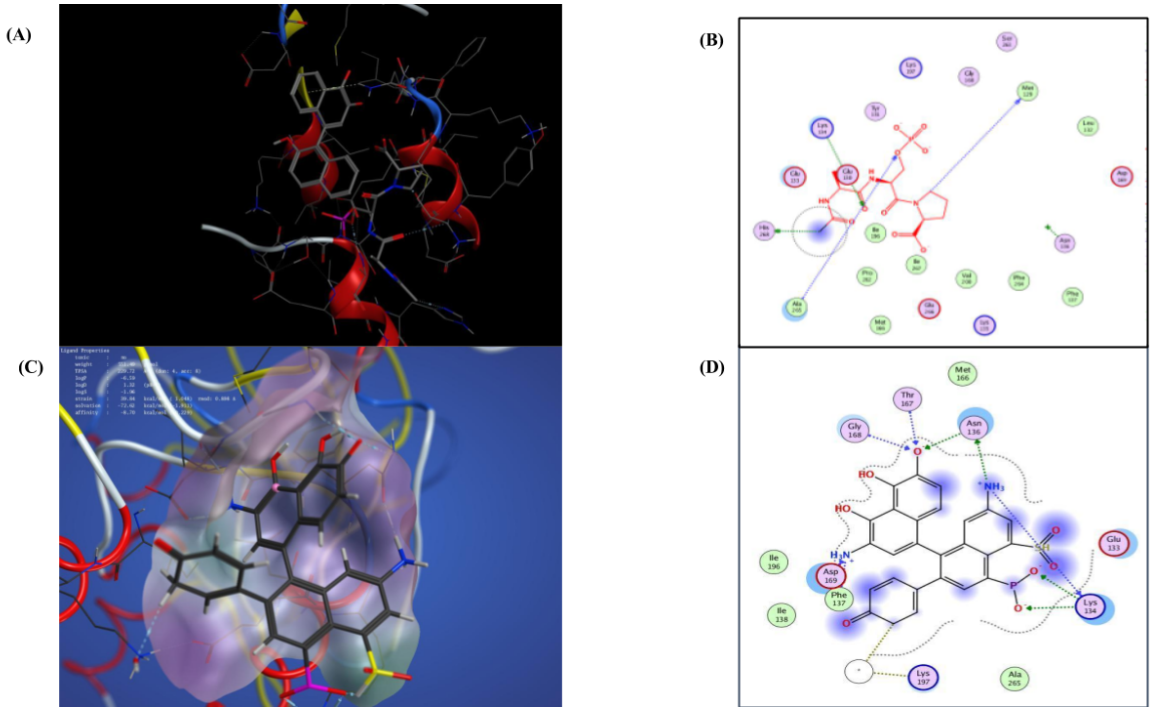


Figure 8: (A) compound 4 bonded to the active site of UmCdc14. (B) the ligand diagram for compound 4 shows few hydrogen bonds and a decent interactions. (C) The hypothetical inhibitor bound to the active site of UmCdc14. (D) Ligand interactions of hypothetical inhibitor with UmCdc14 shows increased affinity and increase in the number of hydrogen bonds.

Discussion:

In order to induce expression of UmCdc14, we introduced L-arabinose to the large cultures of *E. coli*, to allow the cells to produce the enzyme by way of the araBAD promoter on the pET15b vector. We used this vector for multiple reasons, including the fact that it is a standard plasmid which is compatible with the methods used in this study, and because it contains the 6-His tag at the N-terminus of the protein necessary for the Nickel chromatography we intended to perform.

We ran an SDS-PAGE gel to ensure that the protein we had was the correct length. The protein was the size it was expected to be if it were UmCdc14.

The presumed UmCdc14 was tested against the substrate p-nitrophenol-phosphate (pNPP) to prove that it had activity. The compound was dephosphorylated, thus proving that the enzyme was a phosphatase and most likely the phosphatase we intended to study. To ensure that it was the correct phosphatase, we tested 24 different peptide substrates for UmCdc14 (Figure 3B). The best substrate for UmCdc14 is consistent with specificity data found elsewhere in the literature[9]. In particular, the enzyme prefers a phosphorylated serine followed immediately by a proline and at least one basic residue at the +3 position from the phosphorylated residue, just as the others which have been studied. This specificity is consistent with Cdc14, indicating that this is the correct enzyme. The most effective substrate is peptide 6 (Figure 3B). The enzyme was later assessed with mass spectrometry, giving a further affirmation that the enzyme was correct.

PNPP, which, while a poor substrate for UmCdc14 - the K_{cat}/K_m for for pNPP was low, at only 1.05 x 10^{-5} s^{-1} uM^{-1}; the optimal substrate had a K_{cat}/K_m value of 407.5 s^{-1} uM^{-1} - is quite easy to measure, because the product when the phosphate is removed, as UmCdc14 does, p-nitrophenol, absorbs light at 405 nm. This was the main reason for testing with the compound.

Wang et al. found the K_m of *Saccharomyces cerevisiae* Cdc14 to be 10.8 +- .8 mM, where UmCdc14 is slightly lower, at 9.77 mM. This would seem to indicate that UmCdc14 is slightly less efficient than ScCdc14; however, this is not unexpected, as the *UmCDC14* gene is not complete; there is a domain which is not present in the enzyme used in this study.

The IC_{50} value of the optimal inhibitor, numbered 4, was 40 uM. Unfortunately, it was impossible to measure IC_{50} values of the other compounds tested for inhibition due to experimental error.

The expected mechanism of inhibition was competitive inhibition, so the initial mechanism assay was meant to confirm this hypothesis- if it had been competitive, K_m would be increased. Instead, V_{max} was lowered and the K_m showed no significant change, indicating either noncompetitive or suicide inhibition. To test this, time dependence had to be determined. The inhibitor did show more activity after greater amounts of time incubating with the enzyme, indicating suicide inhibition, as the enzyme was permanently disabled because the inhibitor was covalently bound to its active site and could not reach as high a velocity as it had previously.

After successfully creating a homology model of our target enzyme, we executed docking algorithms using the rigid fit model to determine the best possible inhibitors, and then determined the percent inhibition by running assays against pNPP.

An interesting phenomenon observed was that while *in silico* certain inhibitors appeared to bind well to the homology model, their actual abilities to inhibit were significantly lower than would be implied by the mechanism expected, namely competitive inhibition, which would benefit from a higher binding affinity, as calculated by MOE. It is possible that the docking algorithm itself was flawed, and an inaccurate measure of the binding affinity in the first place. In addition, the docking in MOE is not capable of predicting covalent bonding, meaning that, even though compound 4 appeared to have a lower affinity for the binding site, it was able to bind far more effectively than expected. The inaccurate predictions may also be attributed to a faulty homology model, which is possible since the crystal structure of UmCdc14 has not been characterised yet. However, this phenomenon also poses the question of whether there is a certain peculiarity in the active site of UmCdc14 that causes compound 4 to incorporate a suicide inhibition mechanism. The specificity of the substrate helped us identify the necessary binding site required to test our new designed inhibitors *in silico*. The new inhibitor, though hypothetical, can potentially inhibit the enzyme better than existing substrates. Its stability in solution and and its possible capacity to bind to more than one active site, and hence inhibit more than one target, is unknown.

Now with the knowledge of Cdc14 specificity and structure, we modified the structure of inhibitor 4 to a compound with stronger affinity to the active site. Due to a lack of time and resources, we were limited to the design of the molecule. There has not been time or resources to synthesize and test this optimized inhibitor; however, it gains a phenyl group with a double-bonded oxygen, two OH^{-1} groups on adjacent rings, and a O^{-1}. It also gains a SHO_2 group and a PO_2^{-2} group (Figure 8).

A fungicide based on an inhibitor of UmCdc14 would not harm plant life, as higher plants lack Cdc14 (11). However, because Cdc14 has a conserved active site, and humans do have Cdc14, it may be necessary to design an inhibitor which will specifically not interact with HsCdc14. This will take some time and will need to be another study.

Thus this project has various follow-up studies, which can subsequently spawn multiple others, each strongly contributing toward effective antifungal drug target identification and inhibitor design and synthesis. Further studies should include establishing the crystal structure for UmCdc14 enzyme and thus improve *in silico* experiments. Potential inhibitors should be synthesised to establish their efficacy by running *in vitro* trials.

Experimental Procedures:

Bacterial Transformation- For the bacterial transformation, we had *UMCDC14* synthesized and the plasmids created by a company called Genscript. The plasmid used is pET15b and the coding sequence of for the catalytic domain of our Cdc14 gene (Supplemental data 1) was cloned into the sites created by the restriction enzymes Ndel and BamHI to create an in-frame fusion at the 5' end with the 6-His tag

sequence. We performed a standard transformation with E.Coli BL21AI cells.

After the overnight incubation, we prepared the liquid overnight culture in 2xYT media (10.00g yeast extract, 16.00g tryptophan, 5.00g NaCl) from the transformation plates. Two 14 mL plastic round-bottom cell culture tubes were prepared by adding 5 mL of media and 5 uL of ampicillin from a stock solution of 100 mg/ml to each. This was to assure that the bacteria that had survived harbored the plasmid, which included an antibiotic resistance gene. Using a sterilized loop, a single colony was scraped and swirled into the media. The cultures were then set into a 37 degree Celsius shaker set at 225 rpm to be incubated overnight. The next day, small cultures were transferred into 1L of 2xYT medium separated into two 2L flasks.

A program called Logger Pro, with the set to read at a wavelength of 600 nm, was used to determine the absorbance of each culture to ensure that the bacteria were proliferating. We calibrated the spectrometer with 1 mL of the 2xYT media pipetted into a cuvette. After calibration, the absorbance of 1 mL of each flask was collected. The large cell culture flasks were placed back into the shaker set the same as before.

Protein Induction- 5 mL of L-arabinose was added to each culture flasks with ampicillin and placed in a shaking incubator at 37 degree celsius and 225 rpm overnight. Arabinose is used because our plasmid contains an arabinose-inducible *araBAD* promoter (*PBAD*) before our gene which when arabinose is added induces the transcription and ultimately leads to the production of our protein.

Protein Purification by Nickel Chromatography- The contents of the flasks were incubated further and their absorbances checked to assure the bacteria continued to multiply. Then the cultures were combined into a 1L centrifuge bottle which was placed into a large Beckman J-6 centrifuge set for spinning at 3000 rpm for 20 minutes. The supernatant was poured out of the 1L centrifuge bottle and it was then stored in the -80 degree celsius freezer with the cell pellets on the bottom. To prepare the protein for purification, we first thawed the pellet of cells that on ice to keep them from breaking down the protein. While this thawed, the following solution was made:30 mL Lysis buffer (used to break down cell culture), Lysozyme to 1 mg/ml, Pepstatin to 1 µM (protease inhibitor of serine, cysteine, and threonine proteases), Leupeptin to 10 µM (protease inhibitor of aspartyl proteases), and 4 uL Universal nuclease.

These were meant to lyse the cells so that the proteins were accessible for purification. The solution was then added to the 1L container with the cells. The solution was mixed with a rubber policeman until the cells were broken up and everything was allowed to thaw for 30 minutes. 150 uL of phenylmethylsulfonyl fluoride (PMSF), a protease inhibitor, was added after this incubation, then the entire solution was added to 40 mL centrifuge tube to be centrifuged at 35,000 g for 30 minutes at 4 degrees. Once centrifugation was finished, the tube was inverted quickly enough to remove the supernatant without disturbing the pellet. 25 uL of the supernatant was saved for the Bradford protein concentration assay and kept on ice. 75 uL was added to another microcentrifuge tubes which 25 uL of 4x SDS was added, heated for 2 minutes at 95 degrees and frozen at -20 degrees Celsius, ready for SDS analysis to determine concentration and purity. We used an Akta start for nickel chromatography. Akta start setup consisted of clearing all lines and attaching the HisTrap column. Buffers and cell extracts were kept on ice. The same procedure was used to make the flow through as for the extract sample. The rest of the flow through was saved in the refrigerator until it was confirmed that the protein was recovered. A UV spectrum was produced from the Akta and was used to determine which fractions to collect.

Bradford Protein Assay and Further Purification with Dialysis-

We ran a Bradford assay to determine the concentration of our protein aliquots. We calculated the concentration from the equation given by the standard curve, y=0.0006x +0.00890.

The fractions were pooled into dialysis tubing. The tubing was clamped and submerged in 1 L of storage buffer in a large beaker with a stir bar at 4 degrees overnight.

Reagents of storage buffer (1L in H2O): 25 mM HEPES pH7.5 (used 0.5 M stock made previously), 300 mM NaCl, 2 mM EDTA (from 0.5 M stock provided by instructors, pH 8.0), 0.1% 2-mercaptoethanol, 40% (v/v) glycerol.

After the overnight dialysis, the volume went from 15 mL to 2.5 mL. 250 uL of protein was pipetted into 10 microfuge tubes. One tube was stored in the -20 degrees celsius freezer that was used for the Bradford assay that was performed after this, and the rest were stored in the -80 degrees freezer for long term storage. We ran another Bradford assay which utilized a 20 fold serial dilution of protein and a spectrometer with the logger pro program set at absorbance v. concentration at 595 nm. and the measured absorbance was 0.336. The standard curve was used to calculate the concentration, which was multiplied by 20 to account for the 20 fold dilution, then divided by 10 uL, the amount of each dilution added to the cuvette.

To calculate the micromolar concentration of the protein, we used our stock protein which we found to be 10.93 ug/uL by Bradford Assay and used a spectrometer to measure the absorbances. The spectrometer logger pro was opened and set to absorbance v. concentration at a wavelength at 595 nm. It was blanked with a cuvette that had 990 uL Bradford and 10 uL distilled water and a cuvette with 10 uL of protein and 990 uL of Bradford solution was used to collect the absorbance. We repeated the collection for three trials so that we achieved a set of absorbances that correlated with concentrations corresponding with the standard curve The concentration was converted into uM from ug/uL using the molecular weight obtained from ExPASy.

Enzyme Assays and pNPP Assay- The enzyme assay was set up with a set labeled "no inhibitor" that contained reaction buffer, 0.5 uM enzyme, 30 mM of pNPP and a set labeled "with inhibitor" that contained reaction buffer, 1 mM sodium orthovanadate, 0.5 uM enzyme, and 30 mM pNPP. The blank contained reaction buffer and 30 mM pNPP. The data of the "no inhibitor" and "with inhibitor" was triplicated and there was one blank. Each solution for the enzyme assay was made according to the volumes stated in the table. We first added the reaction buffer to all the microfuge tubes, then the pNPP, para-Nitrophenylphosphate, sodium orthovanadate for the "with inhibitor" and lastly enzyme. Sodium orthovanadate, a phosphatase inhibitor, was used to ensure activity was from the foreign protein. Bacteria phosphatases could be still functioning since they would not be affected by this inhibitor. Since the specific activity of the enzyme decreased significantly in the presence of the inhibitor, sodium orthovanadate, from a product concentration of 37.56 uM to 5.04 uM, we concluded that the enzyme was indeed the foreign protein we targeted. Once we added enzyme to all the microfuge tubes, we set a timer and placed all the tubes in the incubator at 30 degree Celsius for 30 minutes. After 30 minutes the microfuge tubes were taken out of the incubator and the reactions were stopped by adding 100 uL of 5 N NaOH in all of the microfuge tubes in the same order as we added enzyme. A spectrometer and logger pro was used to collect the absorbances of each solution, which was set to absorbance v. concentration in uM and at a wavelength of 405 nm. To calculate specific activity, the average concentration of the product, determined from the absorbances of the no inhibitor set using the standard curve, was divided by the amount of time the reaction was incubated and the concentration of enzyme.

Assay to determine Km, Kcat, and Vmax- We ran an assay to test linearity over time using a microfuge tube with 400 uL of reaction buffer, 50 uL of 10x enzyme, and 50 uL of 300 uM pNPP and a blank with reaction buffer replacing the enzyme. The final concentrations of substrate were 1 mM, 2 mM, 5 mM, 10 mM, 20 mM, 30 mM, and 40 mM. Blanks were also made with buffer and the respective dilutions of substrate for each concentration. First buffer was added, then enzyme, which was kept on ice, and the substrate was added last to start the reaction. After adding the substrate to the respective dilutions according to the table, the tubes were incubated for 30 minutes. We used a microplate reader set at 405 nm to obtain absorbance values. Absorbance was converted to concentration of product using a standard curve created previously. The concentrations calculated were then divided by 30 minutes which was how long the reaction was incubated for to obtain the rate. The data was averaged and entered in the website mycurvefit.com to get the Vmax, Km, and Kcat.

Substrate Selectivity Assay- The substrate selectivity assay was set up with .8 uM of enzyme, 100 uM of substrate, and reaction buffer for each of the 24 substrates. The order of reagents was buffer, then enzyme, then substrate. They were then incubated for 20 minutes at 30 degrees Celsius. 100 uL of Biomol Green dye was added

to stop the reaction once the incubation period ended and then the tubes were incubated at room temperature for another 20 minutes. The absorbance of each of the peptide sequences was collected using a spectrometer and logger pro set at wavelength 450 nm. The peptide sequences that showed some absorbance were used in a repetition of the protocol. Absorbance was converted to concentration using the linear fit equation from the sodium phosphate standard curve previously established, and the rate was calculated by dividing by 20 minutes. Kcat/km was calculated by the equation Kcat/km = V/([E]*[S]). Excel was used to make the bar graph representing the Kcat/Km of some of the peptide sequences that had shown absorbances in the first trial with standard deviations as error bars.

Molecular modeling and optimization- All molecular modeling, which included the generation of a homology model, docking inhibitors, and adjusting molecules for optimization was done on the program Molecular Operating Environment (MOE). To create a homology model of our protein, similarity between the sequence of our protein and other known sequences, that 3D structures existed for, were determined and potential templates were identified. After selecting the potential template with the closest value, a homology model was created by the program. The homology model was then compared to the known crystal structure of human Cdc14b (pdb 1OHE) so that the binding site of the homology model can be defined. Multiple sequence alignments show that Cdc14 in Ustilago maydis and a number of other species including homo sapiens share the conserved amino acid sequence of the catalytic site which means the human Cdc14 crystal structure can be used to define the site on our Cdc14. The multi sequence alignment is shown in supplemental data 2. The quality of the model is sufficient as seen in the Ramachandran diagram because of the low number of amino acids in the area that had unfavorable angles. The pdb file of 1OHE, the crystal structure of human Cdc14b, included the crystal structure and a tripeptide sequence XASP which was used to define the binding site in our homology model after the two structures were aligned and superimposed. We built some of the inhibitors in MOE using the builder functionality and other structural models of the inhibitors was obtained by using the SMILES code. We used the docking features set at rigid receptor to dock each of the 13 inhibitors we planned on testing and compared the s-score to predict which inhibitor would be the most efficient.

For the optimization of compound 4, we opened a file of the compound 4 docked into our homology model on the program MOE along with a peptide structure that was from the homology modeling. The inhibitor we chose was selected as the ligand. Binding affinity in Kcal/mol and ligand efficiency was obtained from the Ligand properties option. We made structural modifications based on the interaction diagram, hydrophobic, electrostatic surface maps, and R-vectors that were generated by the program. The builder functionality was used to add molecules at the R-vectors and change existing atoms. The structure's conformational energy was minimized to obtain a new value of the binding affinity.

Inhibitor Assay and IC50- To run the inhibitor assay, for each of the 13 inhibitors, we used 5mM concentration of pNPP, 2 uM enzyme, reaction buffer, and 5 uM of each inhibitor. The positive control had reaction buffer, 2 uM enzyme, and 5mM substrate and the negative control had reaction buffer and 5 mM substrate. The blank had reaction buffer, 2 uM enzyme, and 5 uM of one of the inhibitors. Reaction buffer, enzyme, inhibitor, then substrate was added in that order with the substrate starting the reaction. The microfuge tubes were incubated for 20 minutes at 30 degrees celsius. After the incubation period, 50 uL of 5N NaOH was added to stop the reaction. 250 uL of the each tube was pipetted onto a 96 well microplate and a microplate reader was used to collect absorbance values. Three trials were ran to account for any errors and contribute to more precise data. After the absorbance values had the blank absorbance value subtracted from it, the value was divided by the positive control which is 2.39 (averaging the triplication of positive control)

then multiplying this gives you the percentage of enzyme activity. 100 minus the percent enzyme activity gives us the percent inhibition. Percent inhibition for each of the 13 inhibitors was calculated and averaged by the three sets. The percent inhibition of the 13 inhibitors was ranked from highest to lowest and from that we determined the top three inhibitors as compound 2, 6, and 4.

Final inhibitor concentrations used for the IC50 assay was 500 uM, 300 uM, 250 uM, 100 uM, 50 uM, 40 uM, 25 uM, 10 uM, 5 uM, and 2.5 uM. This was done for each of the top three inhibitors. After this was all set up, the tubes were incubated at 30 degrees for 20 minutes and 50 uL of 5N NaOH was added to end the reaction. The percent inhibition for compound 2, compound 4, and compound 6 were calculated in the same way as in the inhibitor assay and from that it was determined that the best inhibitor was compound 4. We took the log of the concentration of inhibitor and graphed it against percent activity of compound 4 to obtain the IC50 curve which we determined the IC50 value to be 40 uM using the website IC50.tk.

Mechanism Assay- Now that we decided that compound 4 is the best inhibitor, we designed an assay protocol to determine the mechanism of the inhibitor. A master mix was created so that both inhibitor concentrations had the same amount of other reagents as far as it possible. The final inhibitor concentrations used were 25 uM 50 uM, and 0 uM. Knowing IC50 is 40 uM, we chose 25 uM and 50 uM of inhibitor, a point above and below, to show a difference in activity. We had two inhibitor concentrations and picked 7 substrate concentrations (50, 25, 10, 5, 2.5, 1, 0.5 ,0.1 uM). The master mixes already have the buffer, enzyme, and inhibitor which is added to each set of respective microfuge tubes, then substrate was added to each tube to start the reaction. The tubes were incubated for 20 minutes at 30 degrees Celsius and then the reaction was stopped by the addition of 50 uL of 5N NaOH. 250 uL of each tube was pipetted onto a microplate and we used a microplate reader to collect and record the absorbance of each concentration at 405 nm.

To test for suicide inhibition, a stock concentration of substrate of 37.5 uM and stock concentration of enzyme at 275 uM was used. 3040 uL of buffer and 160 uL of enzyme was added into the master mix and 160 uL of the master mix was added to microfuge tubes with 20 uL of inhibitor concentration of 25 uM. A control set with inhibitor replaced by buffer was set up to be certain that the enzyme is not degrading on its own. The tubes were left on the lab bench for ten minutes, then another reaction was set up in the same manner in ten minute intervals. This was repeated until the first reaction had been waiting for 60 minutes. 20uL of substrate was added to every tube and then was placed in a 30 degree incubator for 20 minutes. After 20 minutes, 50 uL 5M NaOH was pipetted into the tubes to end the reaction and a pipette was used to take 250uL out of the tubes to place into a 96 well microplate, and having microplate reader set to read at 405 nm collect absorbance values. Percent inhibition was calculated in the same way as the mechanism assay procedure.

References:

1. United States. Department of Agriculture. Economics, et al. *World agricultural supply and demand estimates*. The Department, 2017.
2. Grain: World Markets and Trade (2017) [online] https://apps.fas.usda.gov/psdonline/circulars/grain.pdf (Accessed July 24, 2017).
3. Tomlinson, I. (2013). Doubling food production to feed the 9 billion: A critical perspective on a key discourse of food security in the UK. *Journal of Rural Studies*, 29, pp.81-90.
4. Christensen, J.J., 1963. Corn smut caused by Ustilago maydis. *Monographs. American Phytopathology Society*, *2*.
5. Snetselaar, K.M. and Mims, C.W., 1994. Light and electron microscopy of Ustilago maydis hyphae in maize. *Mycological Research*, *98*(3), pp.347-355.
6. Li, C., Melesse, M., Zhang, S., Hao, C., Wang, C., Zhang, H., Hall, M. and Xu, J. (2015). FgCDC14 regulates cytokinesis, morphogenesis, and pathogenesis in

Fusarium graminearum. *Molecular Microbiology*, 98(4), pp.770-786.
7. Visintin, R., Craig, K., Hwang, E., Prinz, S., Tyers, M. and Amon, A. (1998). The Phosphatase Cdc14 Triggers Mitotic Exit by Reversal of Cdk-Dependent Phosphorylation. *Molecular Cell*, 2(6), pp.709-718.
8. Kerk, D., Templeton, G., and Moorhead, G. (2007) 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 PHYSIOLOGY* 146, 351-367
9. Gray, C. (2003). The structure of the cell cycle protein Cdc14 reveals a proline-directed protein phosphatase. *The EMBO Journal*, 22(14), pp.3524-3535.
10. Bremmer, S., Hall, H., Martinez, J., Eissler, C., Hinrichsen, T., Rossie, S., Parker, L., Hall, M. and Charbonneau, H. (2011). Cdc14 Phosphatases Preferentially Dephosphorylate a Subset of Cyclin-dependent kinase (Cdk) Sites Containing Phosphoserine. *Journal of Biological Chemistry*, 287(3), pp.1662-1669.
11. Kerk, D., Templeton, G., and Moorhead, G.B. (2008) 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 146: 351–367
12. Calabria, I., Baro, B., Rodriguez-Rodriguez, J., Russiñol, N. and Queralt, E. (2012). Zds1 regulates PP2A Cdc55 activity and Cdc14 activation during mitotic exit through its Zds_C motif. *Journal of Cell Science*, 125(12), pp.2875-2884.
13. Amon, A., Visintin, R. and Hwang, E. (1999). *Nature*, 398(6730), pp.818-823.
14. Chavan, S. and Smith, S. (2014). A Rapid and Efficient Method for Assessing Pathogenicity of *Ustilago maydis* on Maize and Teosinte Lines. *Journal of Visualized Experiments*, (83).
15. Powers, Brendan L., et al. "Measuring Activity and Specificity of Protein Phosphatases." *Cell Cycle Oscillators: Methods and Protocols* (2016): 221-235.
16. Rhind, N., and Russell, P. (2012) Signaling Pathways that Regulate Cell Division. *Cold Spring Harbor Perspectives in Biology* 4, a005942-a005942 (function of cdc14)

Acknowledgements:
We would like to thank the Summer Science Program and the Department of Biochemistry at Purdue University for making all of this possible. We also thank Dr. Mark Hall and Dr. Stefan Paula for running the program. A special thanks to our TAs, GiHun Choi, Rochelle Camden, John Whitney, and Ioana Plesca, and to Mr. Einhorn for helping us in our work.

Author Contributions:
L.D: abstract and experimental procedures
S.D: discussion and references
K.J: introduction and results
Data collection and overall formation of paper was equally distributed with edits and added parts from the other two authors for each part.

Supplemental Data