Appendices:

Binuclear zinc transcription factors and the regulation of patulin biosynthesis in the filamentous ascomycete *Penicillium expansum*

Appendix A - Binuclear zinc transcription factors with confirmed binding sites in filamentous ascomycetes

Table X: Binuclear zinc transcription factors in filamentous ascomycetes with confirmed binding sites

Name	Function	Binding Motif	Motif Type	Accession Number	References
PoxCxrA	Cellulase in *P. oxalicum*	5'-ATCAGATCCTCAAAGA-3' and 5'-GCTGAGTCCTT-3'	Oligo-nucleotide	KU597419	(Liao et al., 2019)
AflR	Aflatoxin in *Aspergillus nidulans*	5'-TCG-n(5)-CGA-3	Repeat	AAC49195	(Fernandes, Keller and Adams, 1998)
AlcR	Ethanol regulon in *Aspergillus nidulans*	a: CCGCA-n(7)-CCGCA, and TGCGG(N2)CCGCA + (c)?	Oligo-nucleotide	P21228	(Panozzo, Capuano, Fillinger and Felenbok, 1997)
MlcR	Compactin biosynthesis in *Penicillium citrinum*	5'-(A/T)CGG-NGT-n(3-6)-TCGG-3'.	Oligo-nucleotide	Q8J0F2	(Baba et. al, 2009)
CLR-4	Cellulase in *Neurospora crassa*	5'-CGG-n(5)-CGG-3'	Repeat	XP_956425	(Liu et. al, 2018)
NirA	Nitrate and nitrite reductases in *Aspergillus nidulans*	5'-CTCCGHGG-3'.	Oligo-nucleotide	AAA33317	(Strauss, Muro-Pastor and Scazzocchio, 1998)

FarA	Fatty acid catabolism in *Aspergillus nidulans*	5'-CCTCGG-3'	Oligo-nucleotide	ABD51992	(Hynes et al., 2006)
VerZ	Verticillin in *Clonostachys rogersoniana*	5'-(T/C)(C/A)(G/T)G-n(3)-CC(G/T)(A/G)(G/C)-3'	Oligo-nucleotide	A0A1U9YI06	(Guo et al., 2017)
UaY	Purine in *Aspergillus nidulans*	5'-TCGG-N6-CCGA-3'	Repeat	P49413	(Suárez, de Queiroz, Oestreicher and Scazzocchio, 1995)
PrnA	Proline in *Aspergillus nidulans*	5'-CCGG-N-CCGG-3' and 5'-CCGG -n(6-7)-CCGG -3'	Repeat	CAA11374	(Gomez et al., 2002)
XYR1	Cellulolytic and Xylanolytic genes in *Trichoderma reesei*	5'-GGCTAA-3' motif but also with several 5'-GGC(A/T)(3)-3'	Oligo-nucleotide	AAO33577	(Furukawa et al., 2009)
RhaR	L-rhamnose utilization in *Aspergillus nidulans*	5'-CGG-n(11)-CCG-3'	Repeat	CDG06149	(Pardo and Orejas, 2014)
CLR-2	Plant biomass utilization in Neurospora *crassa*	5'-CGG-n(11)-CCG-3'	Repeat	XP_962712	(Craig, Coradetti, Starr and Glass, 2015)
LeuB	BCAA biosynthesis and iron metabolism in *Aspergillus fumigatu*	5'-CCG-n(4)-CGG-3'	Repeat	KEY79403	(Long et al., 2018)

AmyR	Amylotic activity in *A. nidulans*	5'-CGG-n(8)-CGG-3'	Repeat	BAA78564	(Nakamura et al., 2006)

Appendix B - MATLab code for identification and placement of coiled-coil dimerization region using DeepCoil

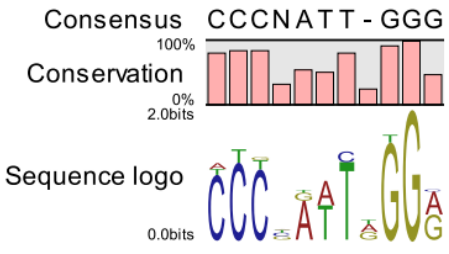
```
% Set input folder

input_folder ='D:\Folder_with_DeepCoil_.out_files';

% Read all *.txt files from input folder
        files = dir(fullfile(input_folder, '*.out'));
% Get full path names for each text file
        file_paths = fullfile({files.folder}, {files.name});

A = zeros(2500,numel(file_paths));
for i = 1 : numel(file_paths)
fileID = fopen(file_paths{i});
C = textscan(fileID,'%s %s %f %f %f','Headerlines',1);
fclose(fileID);

Q = cell2mat(C{2});
J = str2num(Q);

L = size(J);
for p = 1 : L(1,1)
   A(p, i) = J(p,1);
end
end

% Get position of the last Cysteine in the Zn2Cys6 motif for each putative BZTFs in prior
%folder

P = readmatrix("LastCysteineFolder.xlsx");

T = 0.45   % Probability cut-off

for i = 1 : numel(file_paths)
```

```
    R = P(i);
  if any(A(R:2500,i) > T)

    Y = find(A(R:2500,i) > T);
    X = Y(1);
      K(i) = (X-1); % Distance between last Cys residue and first AA with coiled coil prob
%higher than T

 else
    K(i)=0;
  end

end

K = K'          %Listing of distances between last Cys residue and first AA with coiled coil
%prob. higher than T. 0 indicates no coiled coil detected at all.

M=find(K);
Msize = size(M)
B=find(K>0 & K<150); %Number of TFs which have a coiled-coil region under 150 bps
B = size(B) %Gives number of dimerically binding BZTFs, ergo coiled-coil region under
%150bps
```

Appendix C

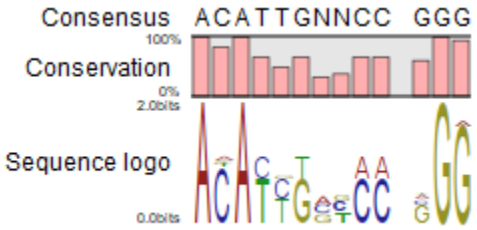
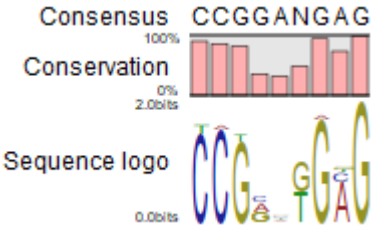
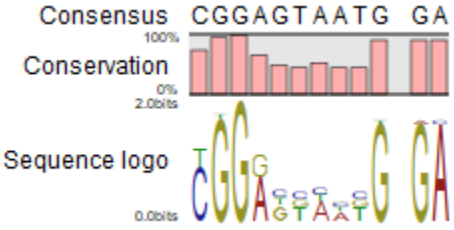
RSAT results Any Repeat

Table X: Overrepresented motifs detected using RSAT (van Helden, André and Collado-Vides, 1998) in the promoter sequences of the patulin biosynthetic gene cluster using *Penicillium digitatum* GCF 000315645.1 PdigPd1 v1 as a background model. Set to detect any direct, everted or inverted repeats with a spacer region between 3 and 11. The sequence logo was constructed using the motif suggested by the tool and putative binding sites where analysed for conservation by alignment of *P. expansum, P. vulpinum* (GCA_002072255.1), *P. paneum* (GCA_000577715.1), *P. antarcticum* (GCA_002072345.1), *P. griseofulvum* (GCA_001561935.1), *P. dipodomyicola* (GCA_015585785.1) and *P. carneum* (GCA_000577495.1).

Sequence	E-value	Present in promoter sequences of the genes	Sequence logo (Number of sites)
5'-CCC-n(4)-GGG	7.0E-01	*patH, patM, patN, patJ, patK*	 (27)
5'-TAT-n(5)-ATA	8.3E-01	-	-
5'-CCG-n(9)-CCG	8.7E-01	*patF, patE*	-

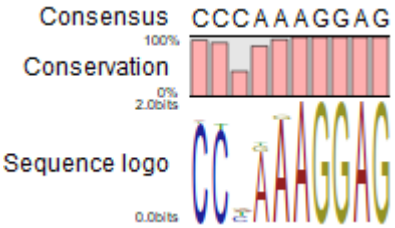
RSAT results Any Dyad

Table X: Overrepresented motifs detected using RSAT (van Helden, André and Collado-Vides, 1998) in the promoter sequences of the patulin biosynthetic gene cluster using *Penicillium digitatum* GCF 000315645.1 PdigPd1 v1 as a background model. Set to detect any dyad with a spacer region between 3 and 11.The sequence logo was constructed using the motif suggested by the tool and putative binding sites where analysed for conservation by alignment of *P. expansum, P. vulpinum* (GCA_002072255.1)*, P. paneum* (GCA_000577715.1)*, P. antarcticum* (GCA_002072345.1)*, P. griseofulvum* (GCA_001561935.1), *P. dipodomyicola* (GCA_015585785.1) and *P. carneum* (GCA_000577495.1).

Sequence	E-value	Present in promoter sequences of the genes	Sequence logo (number of sites used)
5'-ACA-n(7)-GGG	5.3E-03	*patE, patN, patD*	 (18)
5'-CCG-n(3)-GAG	3.0E-01	*patG, patF, patE, patM, patN*	 (24)
5'-CGG-n(6)-GGA	5.6E-01	*patC, patB, patD, patM, patJ, patK*	 (24)
5'-ACC-n(6)-GAG	6.1E-01	*patM*	-

MEME results Oligonucleotide

Table X: Overrepresented motifs detected using MEME (Bailey et al., 2009) in the promoter sequences of the patulin biosynthetic gene cluster using a 0th degree background model based on the promoter sequences themselves. Set to detect any motif between 6 and 10 bps long with any number of repeats per sequence. The sequence logo was constructed using the motif suggested by the tool and putative binding sites where analysed for conservation by alignment of *P. expansum, P. vulpinum* (GCA_002072255.1)*, P. paneum* (GCA_000577715.1)*, P. antarcticum* (GCA_002072345.1)*, P. griseofulvum* (GCA_001561935.1), *P. dipodomyicola* (GCA_015585785.1) and *P. carneum* (GCA_000577495.1).

Sequence	E-value	Present in promoter sequences of the genes	Sequence logo (Number of sites)
5'- CCBRAAGGAG,	9.3E-016	*patH, patG, patF, patE, patC, patB, patM, patN, patJ, patK*	 (53)

Appendix D - Popular Abstract (English)

All living things have metabolism and associated metabolites. Some of these are produced constantly, such as when we break down food, or need to move muscles. There are also metabolites called secondary metabolites (SecMets), which are compounds produced by fungi and have key roles relating to how the organism reacts and interacts with its environment. SecMets are of interest to mankind due to their potential use in the pharmaceutical industry, but also to understand fungal behaviour. A good example of a prominent secondary metabolite is the antibiotic penicillin, which is produced by many fungus from the genus *Penicillium*. *P. expansum* is a species of mold from the genus, but can't produce the antibiotic and is a wide-spread pathogen which infects apples. It causes the apples to rot and also produces the harmful mycotoxin patulin, which is in itself a secondary metabolite and is under a lot of research on ways to prevent it.

Genes which are related to producing a certain compound are in fungi generally grouped together in the genome, in a so-called biosynthetic gene cluster (BGC). This is useful for fungi so that they can be more efficient when producing these compounds. These clusters can then be controlled by different mechanisms, such as widely-impactful global regulator proteins or specific local transcription factors. Transcription factors are proteins which can be used to control the production of a compound in an organism. In fungi these can be of the class binuclear zinc transcription factors (BZTFs). These have a specific characteristic and are only found in fungi. They may work in pairs when modulating metabolism, since they have a specific protein structure which allows them to stick together. They then bind to DNA sites which look palindromic.

This thesis investigated the regulation of patulin production in *P. expansum*, both in regards to the global regulators but also for a specific local transcription factor called patL, which is known to be related to patulin production. The thesis also explores the state of BZTF research in fungi, and discusses if results from one group of fungi may be translated to another.

The results indicate that many BZTFs work exclusively alone, a much larger number than what was thought. This means that they probably do not bind to the palindromic DNA sites. However, the results are inconclusive if the specific protein structure can be used as a prediction method for how the BZTF behaves. The global regulators are very involved with patulin synthesis, and a suggestion for the binding site of the transcription factor patL is made. However, confirming that the protein binds to the binding site is expensive and laborious, and a proposal for new computer methods to predict binding sites is made.

Appendix E - Popular Abstract (Swedish)

Alla levande varelser har metabolism och då även metaboliter, som är ämnen som produceras och konsumeras under metabolismen. En del av dessa produceras konstant, som när vi bryter ner mat vi äter eller när vi ska röra på oss. Det finns också metaboliter som vi inte har, men finns i till exempel svampar. De kallar vi sekundära metaboliter. De är viktiga för hur svamparna beter sig och interagerar med sin omvärld. Vi har ett intresse av dessa ämnen då de kan vara användbara för oss i medicinskt syfte. Ett bra exempel på en sekundär metabolit är antibiotikan penicillin, som produceras av många svampar från släktet *Penicillium*. *P. expansum* är en mögelsvamp från släktet, men kan inte producera antibiotikan. Istället är den en utbredd patogen av äpplen. Den får äpplena att ruttna och producerar i dem det giftiga ämnet patulin. Patulin är ett annat exempel på en sekundär metabolit, och forskning pågår för att få nya sätt att bekämpa ämnet.

Gener som är involverade i processen med att producera en viss kemisk förening grupperas i svampgenom i ett så kallat biosyntetiskt genkluster (BGC). Detta är användbart för svampar då de kan vara effektivare när de producerar dessa föreningar. Dessa kluster kan sedan styras av olika mekanismer, såsom de brett verksamma globala regulatoriska proteinerna eller mer specifika lokala transkriptionsfaktorer. Transkriptionsfaktorer är proteiner som kan användas för att kontrollera produktionen av en förening i en organism. I svampar kan dessa vara av klassen binukleära zink-transkriptionsfaktorer (BZTF). Dessa har en specifik egenskap och finns bara i svampar. De binder ofta i par när de modulerar metabolismen, eftersom de har en specifik proteinstruktur som gör att de kan hålla ihop. De binder sedan till palindromiska DNA-sekvenser.

Detta examensarbete undersökte regleringen av produktion av patulin i *P. expansum*, både när det gäller de globala regulatoriska proteiner och för en specifik lokal transkriptionsfaktor som kallas patL, som är relaterad till patulinproduktionen. Arbetet undersöker också situationen för forskning på BZTFs och diskuterar om forskningsresultat från en grupp svampar kan översättas till en annan.

Resultaten visar att många BZTF:er arbetar inte i par, ett mycket större antal än vad man tidigare trodde. Detta innebär att de förmodligen inte binder till de palindromiska DNA-sekvenserna om de binder själva. Resultaten är dock ofullständiga om den specifika proteinstrukturen kan användas som en förutsägelsemetod för hur en given BZTF beter sig. De globala regulatorerna är mycket involverade i produktionen av patulin, och det görs ett förslag för bindningsekvesen för transkriptionsfaktorn patL kanske ser ut. Att bekräfta att proteinet faktiskt binder till just den sekvensen är dock dyrt och mödosamt, och ett förslag om nya datormetoder för att förutsäga bindningsställen görs.

References

Baba, S., Kinoshita, H., Hosobuchi, M. and Nihira, T., 2009. *MlcR, a zinc cluster activator protein, is able to bind to a single (A/T)CGG site of cognate asymmetric motifs in the ML-236B (compactin) biosynthetic gene cluster*. Molecular Genetics and Genomics, 281(6).

Bailey, T., Boden, M., Buske, F., Frith, M., Grant, C., Clementi, L., Ren, J., Li, W. and Noble, W., 2009. *MEME SUITE: tools for motif discovery and searching*. Nucleic Acids Research, 37(Web Server), pp.W202-W208.

Craig, J., Coradetti, S., Starr, T. and Glass, N., 2015. *Direct Target Network of the Neurospora crassa Plant Cell Wall Deconstruction Regulators CLR-1, CLR-2, and XLR-1*. mBio, 6(5).

Fernandes, M., Keller, N. and Adams, T., 1998. *Sequence-specific binding by Aspergillus nidulans AflR, a C6 zinc cluster protein regulating mycotoxin biosynthesis*. Molecular Microbiology, 28(6), pp.1355-1365.

Furukawa, T., Shida, Y., Kitagami, N., Mori, K., Kato, M., Kobayashi, T., Okada, H., Ogasawara, W. and Morikawa, Y., 2009. Id*entification of specific binding sites for XYR1, a transcriptional activator of cellulolytic and xylanolytic genes in Trichoderma reesei*. Fungal Genetics and Biology, 46(8), pp.564-574.

Gómez, D., Cubero, B., Cecchetto, G. and Scazzocchio, C., 2002. *PrnA, a Zn2Cys6 activator with a unique DNA recognition mode, requires inducer for in vivo binding*. Molecular Microbiology, 44(2), pp.585-597.

Guo, Z., Hao, T., Wang, Y., Pan, Y., Ren, F., Liu, X., Che, Y. and Liu, G., 2017. *VerZ, a Zn(II)2Cys6 DNA-binding protein, regulates the biosynthesis of verticillin in Clonostachys rogersoniana*. Microbiology, 163(11), pp.1654-1663.

Hynes, M., Murray, S., Duncan, A., Khew, G. and Davis, M., 2006. *Regulatory Genes Controlling Fatty Acid Catabolism and Peroxisomal Functions in the Filamentous Fungus Aspergillus nidulans*. Eukaryotic Cell, 5(5), pp.794-805.

Liao, L., Li, C., Zhang, F., Yan, Y., Luo, X., Zhao, S. and Feng, J., 2019. *How an essential Zn2Cys6 transcription factor PoxCxrA regulates cellulase gene expression in ascomycete fungi?*. Biotechnology for Biofuels, 12(1).

Liu, Q., Li, J., Gao, R., Li, J., Ma, G. and Tian, C., 2018. *CLR-4, a novel conserved transcription factor for cellulase gene expression in ascomycete fungi*. Molecular Microbiology, 111(2), pp.373-394

Long, N., Orasch, T., Zhang, S., Gao, L., Xu, X., Hortschansky, P., Ye, J., Zhang, F., Xu, K., Gsaller, F., Straßburger, M., Binder, U., Heinekamp, T., Brakhage, A., Haas, H. and Lu, L., 2018. *The Zn2Cys6-type transcription factor LeuB cross-links regulation of leucine biosynthesis and iron acquisition in Aspergillus fumigatus*. PLOS Genetics, 14(10), p.e1007762.

Nakamura, T., Maeda, Y., Tanoue, N., Makita, T., Kato, M. and Kobayashi, T., 2006. *Expression Profile of Amylolytic Genes in Aspergillus nidulans*. Bioscience, Biotechnology, and Biochemistry, 70(10), pp.2363-2370.

Panozzo, C., Capuano, V., Fillinger, S. and Felenbok, B., 1997. T*he Zinc Binuclear Cluster Activator AlcR Is Able to Bind to Single Sites but Requires Multiple Repeated Sites for Synergistic Activation of the alcA Gene in Aspergillus nidulans*. Journal of Biological Chemistry, 272(36), pp.22859-22865.

Pardo, E. and Orejas, M., 2014. *The Aspergillus nidulans Zn(II)2Cys6 transcription factor AN5673/RhaR mediates L-rhamnose utilization and the production of α-L-rhamnosidases*. Microbial Cell Factories, 13(1).

Strauss, J., Muro-Pastor, M. and Scazzocchio, C., 1998. *The Regulator of Nitrate Assimilation in Ascomycetes Is a Dimer Which Binds a Nonrepeated, Asymmetrical Sequence*. Molecular and Cellular Biology, 18(3), pp.1339-1348.

Suárez, T., de Queiroz, M., Oestreicher, N. and Scazzocchio, C., 1995. *The sequence and binding specificity of UaY, the specific regulator of the purine utilization pathway in Aspergillus nidulans, suggest an evolutionary relationship with the PPR1 protein of Saccharomyces cerevisiae*. The EMBO Journal, 14(7), pp.1453-1467.

van Helden, J., André, B. and Collado-Vides, J., 1998. *Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies*. Edited by G. von Heijne. Journal of Molecular Biology, 281(5), pp.827-842.