ICBAA2017-22

VIRTUAL SCREENING FOR GANODERMAL POLYKETIDE SYNTHASES INHIBITOR

Chia Sui Ong[1,*], Hwi Mei Ke[1] and Mui Yun Wong[2,3]

[1]*Faculty of Information Science and Technology, Multimedia University, 75450 Melaka, Malaysia*
csong@mmu.edu.my, kehwimei@yahoo.com
[2]*Department of Plant Protection, Faculty of Agriculture, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia*
[3]*Institute of Plantation Studies, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia*
muiyun@upm.edu.my

Abstract: Polyketide synthases (PKSs) is a huge family of multifunctional enzymes that produce secondary metabolites, known as polyketides, which can be found in different organisms such as fungi, bacteria and plants. Fungal polyketides (PKs) have been reported as essential virulence factors in plant-pathogen interactions. The main purpose of this project is to use in silico virtual screening technique to discover potential inhibitors that can have considerable interactions with the target PKS protein. The PKS gene sequence was identified from the genome library of Ganoderma. The ketosynthase (KS) domain of the PKS protein was taken for structural prediction by homology modelling. The model predicted was then docked against a library of compound in ZINC database for interaction analysis. In consideration of chemical and toxicity properties of ligands found to interact with the target KS domain, two ligands were shortlisted, namely tomatidine and posaconazole. Further verification on the effect of these ligands towards PKS activity will be conducted in future studies.

Keywords**:** *Ganoderma, polyketide synthase, fungal virulence factor, inhibitor, virtual screening*

INTRODUCTION

Polyketides (PKs) in fungi have been reported as one of the important virulence factors employed by a fungal pathogen to establish infection and cause damages in the host (Scharf, Heinekamp, & Brakhage, 2014). A number of polyketides have been proven to contribute to the pathogenicity of a fungal pathogen (Baker et al., 2006; Brown, Busman, & Proctor, 2014; Choquer et al., 2005). Inhibition of the PK biosynthetic gene, polyketide synthase (PKS), offers an approach of antifungal mechanism. PKSs are multi-modular enzymes; the number and combination of domains in each module may vary (Schirmer et al., 2005). One of the core domains, ketosynthase (KS) domain, catalyses chain elongation step and is found to be minimally needed in each module (Lohman et al., 2015). The inhibition of PKS by cerulenin, an antifungal agent, has been reported to be due to the chemical modification of the substrate-binding Cys-204 residue of the KS domain (Child & Shoolingin-Jordan, 1998). Virtual screening (VS) has been widely applied for the *in silico* screening and discovery of small molecules that interact with protein targets (Shoichet, 2004). VS allows screening across a great array of chemical compounds to identify those with good interactions with the target macromolecules (enzyme or protein receptor). The results provide useful hint for drug discovery and optimization (Lionta, Spyrou, Vassilatis, & Cournia, 2014). In this project, VS is applied on the KS domain of PKS. The genome sequence of *Ganoderma lucidum* G.261025-1 (Accession: AGAX01000163.1) was analysed in this project, as it is the closest species available in public

sequence depository to *G. boninense*, an important plant pathogen to oil palm. *G. boninense* is the causal agent for basal stem rot (BSR) disease of oil palm (Paterson, 2007). The disease brings great economic losses to the industry as infected trees suffer from decreased yields (Roslan & Idris, 2012). Effort in discovering novel, specific-acting antifungal agents, as done in this project, may help in controlling the fungal pathogen.

METHODS

Identification of secondary metabolite biosynthetic gene clusters

The genome library of *G. lucidum* (AGAX00000000.1) was searched for secondary metabolite biosynthetic gene clusters using antiSMASH 4.0 (Blin et al., 2017). The genome mining tool enables identification of antibiotics and secondary metabolites gene clusters in the query sequences. The query sequence was submitted to the web server, and the analysis results were returned hours later. Amino acid sequences of the predicted KS domains were retrieved for protein structural modelling.

Protein structure prediction and evaluation

Four programs were used to construct 3D protein structures of the query sequence: 1) Modeller (difficult modelling)(Šali & Blundell, 1993), 2) I-TASSER (Yang et al., 2014), 3) Geno3D (Combet, Jambon, Deleage, & Geourjon, 2002) and 4) SWISS-MODEL (Biasini et al., 2014). For homology modelling, a template molecule was needed. The template was selected based on the similarity search results, analysed via blastp algorithm, against all the protein entries in Protein Data Bank (PDB)(Berman et al., 2000). The use of multiple modelling tools generated a total of 4 models for subsequent quality evaluation. Evaluation on all the models was done using 3 tools: ProSA (Wiederstein & Sippl, 2007), PROCHECK (Laskowski, MacCarthur, & Thorton, 1993), and Verify3D(Lüthy, Bowie, & Eisenberg, 1992). The model with the best evaluation results was chosen for the ligand-target docking analysis. Energy minimisation (using YASARA Energy Minimization Server)(Krieger et al., 2009) and binding site prediction (using MetaPocket 2.0)(Zhang, Li, Lin, Schroeder, & Huang, 2011) were performed on the selected model.

Docking for ligand interaction analysis

The ligand library, "Target on fungi", was downloaded from ZINC database (Irwin, Sterling, Mysinger, Bolstad, & Coleman, 2012) and prepared for docking with the target KS domain using Open Babel (O'Boyle et al., 2011) in PyRx (Dallakyan & J Olson, 2015) virtual screening software. The ligands went through energy minimisation step and the file was converted to PDBQT format. Structural-based virtual screening (SBVS) was performed using AutoDock Vina (Trott & Olson, 2010) docking software in PyRx. The ligand-target interaction was visualised using Pymol (Delano, 2002). From the docking results, ten highest ranked ligands were shortlisted for further review on their chemical, biological and toxicological details.

RESULTS AND DISCUSSION

Analysis of the genome library of *G. lucidum* by antiSMASH returned 4 clusters of PKS Type I. Domains like KS, AT, KR, DH, and TE were detected. The amino acid sequence of the KS domain detected in contig 5 (Figure 1) was retrieved for subsequent analysis steps. The retrieved sequence has 420 amino acid residues. Similarity search in PDB database suggested a molecule, 4MZ0_A, as the nearest available template to use (e-value 1E-68, score 236, identity 34%, query coverage 99%). This molecule is an X-ray crystallographic structure of KS-AT di-domain of a cyanobacterial PKS, curacin A.

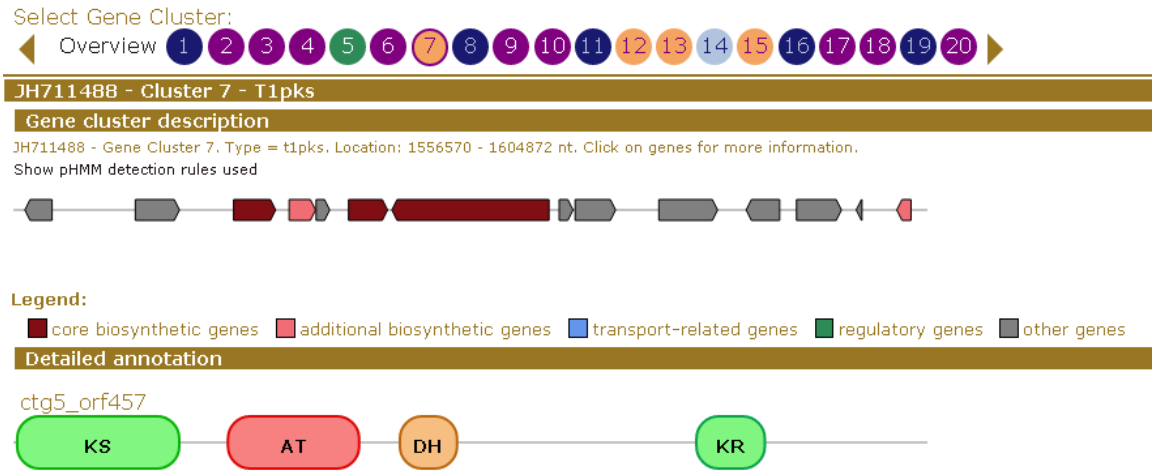

Figure 1: The PKS gene cluster identified using antiSMASH.

Homology modelling was performed using molecule 4MZ0_A as template. As different modelling tools apply different algorithms, and many of these are of heuristic nature, modelling was done using multiple tools. The models generated were numbered as Model 1 to 4 (Model 1: Difficult Modelling module by Modeller; Model 2: CPHmodels; Model 3: I-TASSER; Model 4: SWISS-MODEL, by T-Coffee alignment mode, respectively).

Evaluation of the 4 protein models was performed using different evaluation tools. Aspect of assessment and presentation of assessment results are somewhat different from one tool to another. Models were evaluated and given assessment in the form of Z-score, energy value, percent of buried outlier protein atoms, overall quality score, and in Ramachandran plot. To compare and summarise the assessment results, all the 4 models were ranked as 1st to 4th, according to the results by each evaluation tool. Model that obtained the best result in an evaluation tool was rewarded with the highest score, while the worst model gets the lowest score (the 1st ranked model = 4 scores; the 4th ranked model = 1 score). By considering the results by 3 evaluation tools, the final scores fell within the range from 1-12. Table 1 presents the evaluation results ranking and the final scores of the 4 models. Of all the models, Model 1 obtained the highest score, 10, and was selected for docking analysis (Table 1).

Table 1: Protein models evaluation results.

Evaluation Software	Model 1		Model 2		Model 3		Model 4	
	Results	Score	Results	Score	Results	Score	Results	Score
ProSa (Z-score)	-8.04	4	-7.71	1	-7.84	2	-8.01	3
PROCHECK (Ramachandran Plot – most favoured regions [%])	88.2%	3	84.8%	2	75.9%	1	89.5%	4
VERIFY3D (% residues with average 3D-1D score >=0.2)	92.14	3	88.07	2	94.29	4	85.95%	1
		10		**5**		**7**		**8**

Docking analysis of Model 1 with ZINC ligand library "Target on fungi" subset generated a list of compounds that have considerable interaction with the target KS domain. The interaction was presented in the form of docking score / binding energy; the lower the energy, the more stable the binding is (i.e. higher binding affinity). The top ten ligands were shortlisted. Visualisation of the ligand-target interactions revealed the amino acid residue(s) a ligand binds to. All but one ligand bind to the amino acid residues that were predicted as part of the binding pocket. This increases the chance that the ligands shortlisted may have impact on the activity of the target KS domain. Two ligands, ZINC08143640 (Tomatidine), and ZINC03938482 (Posaconazole) are well-known antifungal agents but have yet been studied on effect on *Ganoderma* specifically. The other ligands, ZINC38803626 Nyssoside, ZINC00598970 BIBB, ZINC03917574 Pseudohypericin, ZINC20233084 Xanthylic acid, ZINC00000640 Piritrexim, ZINC18271146, and ZINC36378703 have either not sufficient info to support their antifungal potential, or the antifungal properties evaluation is still underway .

CONCLUSIONS

Structure-based virtual screening enables identification of ligands that have potential to bind to a protein target. The two ligands identified through this project are worth noting. Their impact on the virulence of *Ganoderma* should be tested.

REFERENCES

Baker, S. E., Kroken, S., Inderbitzin, P., Asvarak, T., Li, B.-Y., Shi, L., … Turgeon, B. G. (2006). Two polyketide synthase-encoding genes are required for biosynthesis of the polyketide virulence factor, T-toxin, by Cochliobolus heterostrophus. *Molecular Plant-Microbe Interactions : MPMI*, *19*(2), 139–149. https://doi.org/10.1094/MPMI-19-0139.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., … Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, *28*(1), 235–242. https://doi.org/10.1093/nar/28.1.235.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., … Schwede, T. (2014). SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, *42*(W1), 252–258. https://doi.org/10.1093/nar/gku340.

Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., Medema, M. H. (2017). AntiSMASH 4.0 - improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Research*, *45*(W1), W36–W41. https://doi.org/10.1093/nar/gkx319.

Brown, D. W., Busman, M., & Proctor, R. H. (2014). *Fusarium verticillioides SGE1* Is Required for Full Virulence and Regulates Expression of Protein Effector and Secondary Metabolite Biosynthetic Genes. *Molecular Plant-Microbe Interactions*, *27*(8), 809–823. https://doi.org/10.1094/MPMI-09-13-0281-R.

Child, C. J., & Shoolingin-Jordan, P. M. (1998). Inactivation of the polyketide synthase, 6-methylsalicylic acid synthase, by the specific modification of Cys-204 of the beta-ketoacyl synthase by the fungal mycotoxin cerulenin. *Biochem J*, *330*, 933–937. Retrieved from http://www.biochemj.org/bj/330/0933/bj3300933.htm%5Cnhttp://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9480912%5Cnhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC1219227/pdf/9480912.pdf.

Choquer, M., Dekkers, K. L., Chen, H., Cao, L., Ueng, P. P., Daub, M. E., & Chung, K. (2005). The CTB1 gene encoding a fungal polyketide synthase is required for cercosporin biosynthesis and fungal virulence of Cercospora nicotianae. *Molecular Plant-Microbe Inetractions: MPMI*, *18*(5), 468–476. https://doi.org/10.1094/MPMI-18-0468.

Combet, C., Jambon, M., Deleage, G., & Geourjon, C. (2002). Geno3D: automatic comparative molecular modelling of protein. *Bioinformatics*, *18*(1), 213–214. https://doi.org/10.1093/bioinformatics/18.1.213.

Dallakyan, S., & J Olson, A. (2015). *Small-Molecule Library Screening by Docking with PyRx*. *Methods in molecular biology (Clifton, N.J.)* (Vol. 1263). https://doi.org/10.1007/978-1-4939-2269-7_19.

Delano, W. L. (2002). The PyMOL Molecular Graphics System. Retrieved from citeulike-article-id:2816763.

Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., & Coleman, R. G. (2012). ZINC: A free tool to discover chemistry for biology. *Journal of Chemical Information and Modeling*, *52*(7), 1757–1768. https://doi.org/10.1021/ci3001277.

Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., … Karplus, K. (2009). Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins: Structure, Function and Bioinformatics*, *77*(SUPPL. 9), 114–122. https://doi.org/10.1002/prot.22570.

Laskowski, R. A., MacCarthur, M. W., & Thorton, J. M. (1993). Computer Programs. *Journal of Applied Crystallography*, (26), 283–291. https://doi.org/10.1016/B978-0-12-803188-9.09994-0.

Lionta, E., Spyrou, G., Vassilatis, D. K., & Cournia, Z. (2014). Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Current Topics in Medicinal Chemistry*, *14*(16), 1923–1938. https://doi.org/10.2174/1568026614666140929124445.

Lohman, J. R., Ma, M., Osipiuk, J., Nocek, B., Kim, Y., Chang, C., … Shen, B. (2015). Structural and evolutionary relationships of "AT-less" type I polyketide synthase ketosynthases. *Proceedings of the National Academy of Sciences*, *112*(41), 12693–12698. https://doi.org/10.1073/pnas.1515460112.

Lüthy, R., Bowie, J. U., & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature*, *356*(6364), 83–85. https://doi.org/10.1038/356083a0.

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An Open chemical toolbox. *Journal of Cheminformatics*, *3*(10), 1–14. https://doi.org/10.1186/1758-2946-3-33.

Paterson, R. R. M. (2007). Ganoderma disease of oil palm-A white rot perspective necessary for integrated control. *Crop Protection*, *26*(9), 1369–1376. https://doi.org/10.1016/j.cropro.2006.11.009.

Roslan, A., & Idris, A. S. (2012). Economic Impact of Ganoderma Incidence on Malaysian Oil Palm Plantation – A Case Study in Johor. *Oil Palm Industry Economic Journal*, *12*(1), 24–30.

Šali, A., & Blundell, T. L. (1993). Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology*, *234*(3), 779–815. https://doi.org/10.1006/jmbi.1993.1626.

Scharf, D. H., Heinekamp, T., & Brakhage, A. A. (2014). Human and Plant Fungal Pathogens: The Role of Secondary Metabolites. *PLoS Pathogens*, *10*(1), 10–12. https://doi.org/10.1371/journal.ppat.1003859.

Schirmer, A., Gadkari, R., Reeves, C. D., Ibrahim, F., Delong, E. F., Richard, C., & Hutchinson, C. R. (2005). Metagenomic Analysis Reveals Diverse Polyketide Synthase Gene Clusters in Microorganisms Associated with the Marine Sponge Discodermia dissoluta Metagenomic Analysis Reveals Diverse Polyketide Synthase Gene Clusters in Microorganisms Associated with the M. *Applied and Environmental Microbiology*, *71*(8), 4840–4849. https://doi.org/10.1128/AEM.71.8.4840.

Shoichet, B. K. (2004). Virtual screening of chemical libraries. *Nature*, *432*(7019), 862–865. https://doi.org/10.1038/nature03197.

Trott, O., & Olson, A. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Computational Chemistry*, *31*(2), 455–461. https://doi.org/10.1002/jcc.21334.AutoDock.

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in

three-dimensional structures of proteins. *Nucleic Acids Research*, *35*(SUPPL.2), 407–410. https://doi.org/10.1093/nar/gkm290.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2014). The I-TASSER Suite: protein structure and function prediction. *Nature Methods*, *12*(1), 7–8. https://doi.org/10.1038/nmeth.3213.

Zhang, Z., Li, Y., Lin, B., Schroeder, M., & Huang, B. (2011). Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics*, *27*(15), 2083–2088. https://doi.org/10.1093/bioinformatics/btr331.