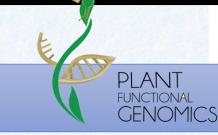
Bioengineering of Saccharomyces cerevisiae with Nepenthes ampullaria neprosin enzyme



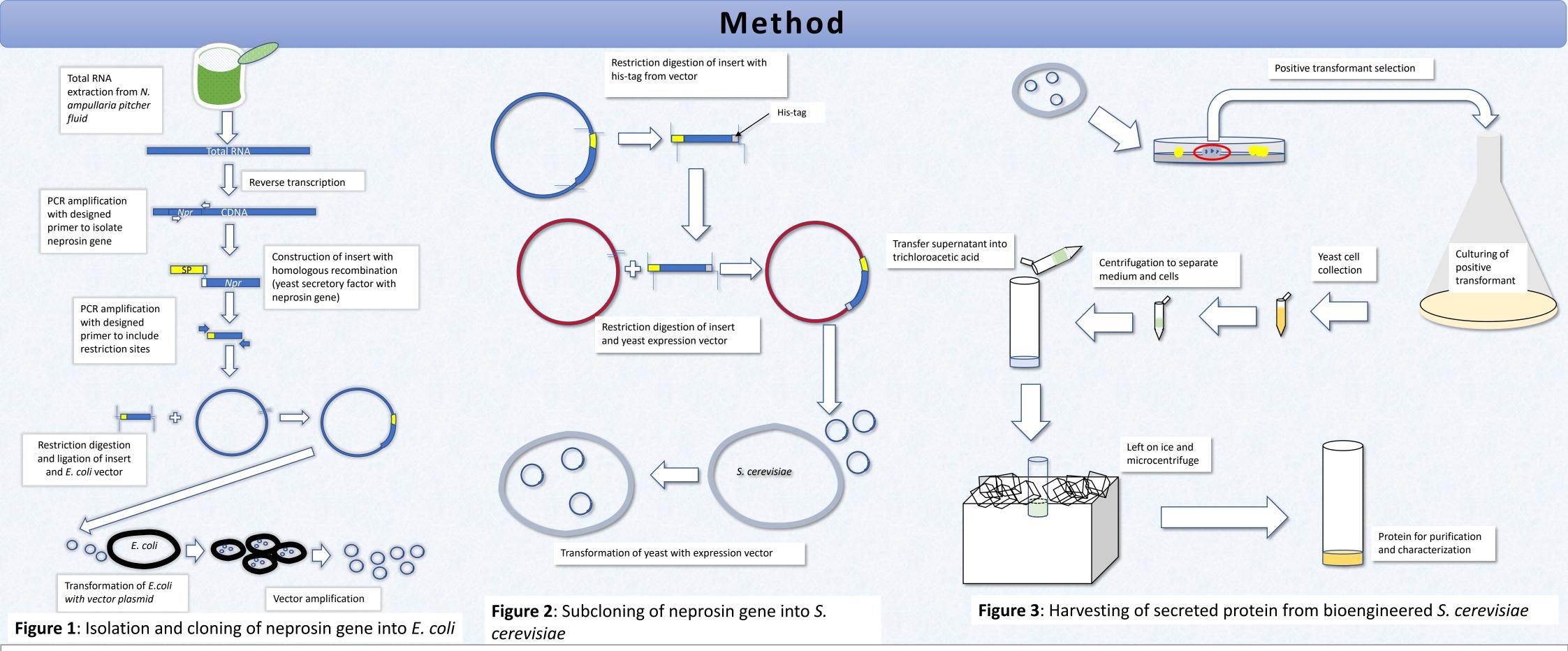
Tiew-Yik Ting¹, Hoe-Han Goh¹* Plant Functional Genomics Group, ¹Institute of Systems Biology, University Kebangsaan Malaysia, UKM Bangi 43600 Selangor, Malaysia. *Corresponding author. Email: <u>gohhh@ukm.edu.my</u>. Funding: FRGS/1/2019/STG05/UKM/02/10

Background

• Protease is a class of enzyme that catalyses proteolysis with wide substrate specificity and biological function. Proteases were secreted into the pitcher fluids of carnivorous plants, such as Nepenthes ampullaria which were discovered in transcriptomics and proteomics study (Zulkapli et al. 2017).

•Proteases from *Nepenthes* pitcher fluid possess unique properties that are desirable in industrial application such as high optimal temperature, high activity in wide pH range, and high stability against chemicals or denaturing agents (Ravee et al. 2018). Neprosin is the first plant prolyl endopeptidase with characteristics distinct from previously characterized prolyl endopeptidases of microbial origin. Neprosin contains two novel functional domains, namely neprosin activation peptide (pfam14365) and neprosin domain (pfam03080).

•A neprosin from Nepenthes × ventrata with a low molecular weight is shown to exhibit proline-cleaving activity towards protein of any size at low pH and enzyme concentration. Thus, it has the potential as a tool in proteomics for whole-proteome profiling and histone mapping (Schräder et al. 2017). Additionally, Rey et al. (2016) has demonstrated the capability of N. × ventrata neprosin in degrading gliadin, a subcomponent of gluten which trigger inflammatory response in celiac disease patients.



Neprosin gene sequences were obtained from the *N. ampullaria* pitcher fluid transcriptome and proteome analyses. Two neprosin genes were cloned into pET-28(b)+ (figure 1) and expressed in the *E. coli* before sub-cloning into yeast expression vector (figure 2). *S. cerevisiae* CEN.PK-2d will be transformed with the yeast expression vector for further analysis (figure 3). A secretory signal such as alpha-mating factor will be engineered for the recombinant neprosin enzyme to be secreted extracellularly into culture medium for harvesting and purification by using His-tag affinity chromatography. Enzyme functional characterization will be conducted to determine the enzyme activity, optimal temperature and pH, as well as substrate and inhibitor specificity of neprosin. Alternatively, a genome engineering approach with CRISPR-Cas will be taken to integrate the gene sequence into the yeast genome for more stable expression.

Result

Phylogenetic analysis of neprosin genes will enable the study on the evolutionary function of the protein in other plant species. This study is expected to produce recombinant neprosin enzymes from *N. ampullaria* in *S. cerevisiae*. Functional characterization will identify optimal temperature, optimal pH, specific inhibitor, and specific substrate for the recombinant neprosin enzymes. Various parameters and optimizations will be tested on the feasibility of the transformed yeast for gluten detoxification in bakery.

carboxyl-terminal peptidase (DUF239) [Arabidopsis thaliana] (GAV68921.1)	Reference sequence (1): NvNpr		
NEP-interacting protein 1 [Arabidopsis thaliana] (GAV83246.1)	Identities normalised by aligne	d length.	
NIP1 [Arabidopsis thaliana] (OAP09352.1)	Colored by: identity		
NEP-interacting protein (DUF239) [Arabidopsis thaliana] (NP_181951.3)			
NEP-interacting protein (DUF239) [Arabidopsis thaliana] (NP_030962.1)	cov pid	1[80
 C-terminal peptidase [Nepenthes alata] (BAW35438.1)	1 NvNpr 100.0% 100.0%	MQAKFFTFVILSSVFYFNYPLA <mark>E</mark> ARSIQARLANKPKGTIKTIKGDDGEVVD <mark>CVDIYKQ</mark> PAFDHP	
	2 NaNpr1 98.7% 37.1% 3 NaNpr2 98.7% 38.5%	MAVPFSK-YLVTAAFIVVL <mark>F</mark> LW <mark>VSIS</mark> LTSAARSTA <mark>A</mark> SREKIEVHKH <mark>L</mark> KRLN <mark>K</mark> PAVHSVQSPDGDVIDCVHTSHOPAFDHP MTSQFSGDLMRRAAFGVL <mark>FF</mark> SWVSV <mark>S</mark> LTCAAGF <mark>P</mark> AASRQKLDVAKHLKRLNKPALH <mark>TT</mark> QSPDGDIIDCVHTSROPAFDHP	
C-terminal peptidase [Nepenthes alata] (BAW35437.1)	3 NaNpr2 98.7% 38.5% consensus/100%	misgrsgblmkkaafgvlFfswysvslicaagfpaaskyklbvakhlkklnkpalhiigspjgblibcvhiskopafbip h.shhF.aV.lS.s.hht.shApthpltt+Ltphskssl+olpusDG-llDCVcI.+QPAFDHP	
neprosin [Nepenthes x ventrata]	consensus/90%	h.shhF.aV.lS.s.hht.shApthpltt+LtphsKssl+olpusDG-llDCVcI.+0PAFDHP	
AsIB (DUF239) [Arabidopsis thaliana] (NP_001318652.1)	consensus/80%	h.shh <mark>F.a</mark> V.l <mark>S</mark> .s.hht.sh <mark>A</mark> pthpltt+Ltphs <mark>K</mark> ssl+olpusDG-llDCVcI.+0PAFDHP	
— DUF239 domain-containing protein/DUF4409 domain-containing protein [Cephalotus follicularis] (GAV68921.	1) consensus/70%	h.shh <mark>a</mark> .a <mark>V.lS</mark> .s.hht.sh <mark>A</mark> pthpltt+Ltphs <mark>K</mark> ssl+olpus <mark>DG</mark> -ll <mark>DCV</mark> cI.+OPAFDHP	
— DUF239 domain-containing protein/DUF4409 domain-containing protein [Cephalotus follicularis] (GAV83246.	1) cov pid	01 1	160
carboxyl-terminal peptidase [Arabidopsis thaliana] (NP_001320047.1)	1 NvNpr 100.0% 100.0%	LLKNHTLQMQPSSYASKVGEYNKLEQPWHKNGECPKGSIPIRRQVITGLPVVKKQFPN	100
carboxyl-terminal peptidase [Arabidopsis thaliana] (NP_001329052.1)	2 NaNpr1 98.7% 37.1%	FLRNMTVOMRPSYHPEGLFDENNKIVTTPKERTNPITOLWHLNGRCPEGTVPIRRTKKDDILRASSVKRFGRKPHRSIPE	
 carboxyl-terminal peptidase (DUF239) [Arabidopsis thaliana] (NP_030959.1)	3 NaNpr2 98.7% 38.5%	FL <mark>KNHTIQIRPS</mark> YHPEGLFDEN-KIATEQKERI <mark>N</mark> PIT <mark>OLWHLNGKCPEGTIPIRR</mark> TKKNDILRASSVKRYGRKKHRSV <mark>P</mark> Q	
carboxyl-terminal peptidase (DUF239) [Arabidopsis thaliana] (NP_J01031538.1)	consensus/100%	hL+NhT1QhpPS.aspthtchN.lpQ.WHhNGc <mark>C</mark> PcGolPIRRphhssl.hsp+phPp	
	consensus/90% consensus/80%	h <mark>L+NhTlO</mark> hp <mark>PS</mark> .aspthtch <mark>N.lpO.WH</mark> hNGc <mark>CP</mark> cGolPIRRphhssl.hsp+phPp hL+NhTlOhp <mark>PS</mark> .aspthtch <mark>N.lpO.WH</mark> hNGc <mark>CP</mark> cGolPIRRphhssl.hsp+phPp	
Neprosin 2 [Nepenthes ampullaria] (ARA95696.1)	consensus/70%	h <mark>L</mark> +NhT1Qhp <mark>PS</mark> .aspthtchN.lpQ.WHhNGc <mark>C</mark> PcGolPIRRphhssl.hsp+phPp	
Neprosin 1 [Nepenthes ampullaria] (ARA95695.1)			
tRNA-splicing ligase (DUF239) [Arabidopsis thaliana] (NP_J01190555.1)	cov pid		240
DUF239 domain-containing protein/DUF4409 domain-containing protein [Cephalotus follicularis] (GAV68607.	1 NvNpr 100.0% 100.0% 2 NaNpr1 98.7% 37.1%	L <mark>KFAPPS-ANTNHQYAVIAYFYGNASLQGANATINIWEPNLKNPNGDFSLTQIWI</mark> SAGS-GSSLNTIEAGWQVYPG <mark>RTGD</mark> PRS <mark>ADPDLVN</mark> ESGHQHAIAYVEG-DKYYGAKATINVWEPQIQQSN-EFSLSQIWVLGGSFGEDLNSIEAGWQVSPDLYGD	
tRNA-splicing ligase (DUF239) [Arabidopsis thaliana] (NP-I75933.1)	3 NaNpr2 98.7% 37.1%	PKSADPDIVNESGHQHAIATVEG-DKYYGAKATINVWEPQIQQPN-EFSLSQIWVLGGSFGEDLVSIEAGWQVSPDLYGD	
DUF239 domain-containing protein/DUF4409 domain-containing protein [Cephalotus follicularis] (GAV69762.		.+.AsPs.sNpstp.tsIAYh.G.sph.GApATIN1WEPplppsNFSLoQIWl.uGS.GpsLNoIEAGWQV.PshhGD	
carboxyl-terminal peptidase (DUF239) [Arabidopsis thaliana] (NP_I97347.1)	consensus/90%	.+.AsPs.sNpstp.tsIAYh.G.sph.GApATIN1WEPplppsNFSLoQIWl.uGS.GpsLNoIEAGWQV.PshhGD	
F20B24.18 [Arabidopsis thaliana] (AAF17666.1)	consensus/80%	.+.AsPs.sNpstp.tsIAYh.G.sph.GApATIN1WEPp1ppsNFSLoQIW1.uGS.GpsLNoIEAGWQV.PshhGD	
	consensus/70%	.+.AsPs.sNpstp.tsIAYh.G.sph.GApATIN1WEPplppsNFSLoQIW1.uGS.GpsLNoIEAGWQV.PshhGD	
F26F24.22 [Arabidopsis thaliana] (AAF87010.1)	cov pid	241 :	320
At1g70550 [Arabidopsis thaliana] (AAO42874.1)	1 NvNpr 100.0% 100.0%	<mark>SQPRFFIYWTADGYTSTGCYDLTCPGFVQTNNYYAIGMALQP-SVYGGQQYELNESIQRDPATGNWWLYLW-GTVVGYWP</mark> NNTRLFTYWTSDAYQATGCYNLLCSGFIQINNEIAMGASISPVSAYRNSQYDISILVWKDPKEGNWWMQFGNNYVLGYWP NNTRLFTYWTSDAYQATGCYNLLCSGFIQINNEIAMGASISPVSAYRNSQYDISILVWKDPKEGNWMMQFGKDYILGYWP	
DUF239 domain-containing protein/DUF4409 domain-containing protein [Cephalotus follicularis] (GAV79524.		NNT <mark>RLFTYWTSDAYQAT<mark>GCY</mark>NLL<mark>C</mark>SGFIQINNEIAMGASISPV<mark>S</mark>AYRNSQYDISILVWK<mark>D</mark>PKEGNWWMQFGNNYVLGYWP</mark>	
DLIE239 domain_containing protein/DLIE/409 domain_containing protein [Conhalotus follicularis] (CAV/58307	1 3 NaNpr2 98.7% 38.5%	NNTILLTYMISDAYQALIGOYNLLCSGIIQINNEIAMGASISPVSAYRNSQYDISILVWKDPKEGNMMQFGKDYILGYMP	

AT5G50150 [Arabidopsis thaliana] (BAH57282.1)	 Doi 259 domain-containing protein/Doi 4409 domain-containing protein [Cephalotus loincularis] (GAV50507.1)	
	AT5G50150 [Arabidopsis thaliana] (BAH57282.1)	6

Figure 4: Phylogenetic tree based on amino acid sequence of BLASTP results of *Nepenthes ampullaria* against *Arabidopsis thaliana* (Plaza database) and all pitcher plant species (NCBI database).

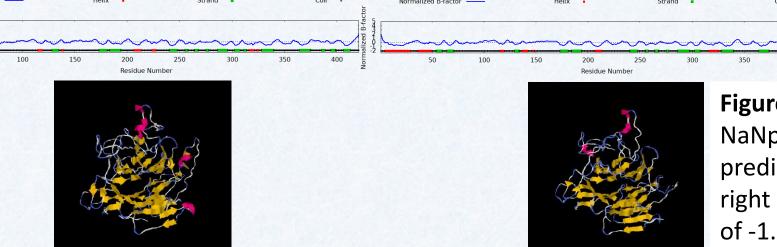


Figure 6: Protein structure prediction of NaNprs using I-TASSER. Left image is predicted 3D structure of NaNpr1 whereas right image is that of NaNpr2, with C-score of -1.99 and -2.05 respectively.

Conclusion

Bioengineering of baker's yeast *S. cerevisiae* with heterologous neprosin holds great promises for applications in the food industry, with ease of use in baking to produce gluten-free flour-based products.

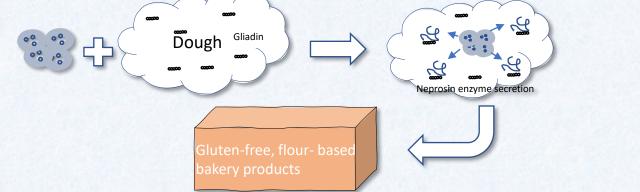


Figure 7: Illustration of application of the bioengineered baker yeast in bakery industry

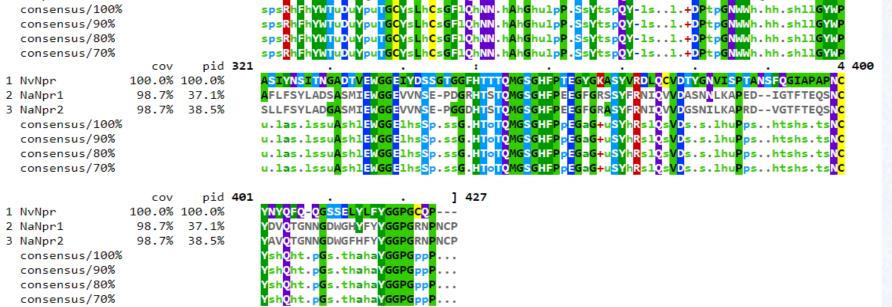


Figure 5: Result of multiple sequence alignment of *N. ampullaria* neprosin proteins (NaNpr) to *N.* \times *ventrata* neprosin (NvNpr). NaNprs and NvNpr have 98.7% consensus, NaNpr1 has 37.1% percentage identity while NaNpr2 has 38.5% percentage identity when compared to NvNpr.

Reference

Ravee, R., Salleh, F.I.M. and Goh, H.H. (2018) Discovery of digestive enzymes in carnivorous plants with focus on proteases. PeerJ, 6, p.e4914

Rey, M., Yang, M., Lee, L., Zhang, Y., Sheff, J.G., Sensen, C.W., Mrazek, H., Halada, P., Man, P., McCarville, J.L. and Verdu, E.F. (2016) Addressing proteolytic efficiency in enzymatic degradation therapy for celiac disease. *Scientific reports*, *6*, p.30980

Schräder, C.U., Lee, L., Rey, M., Sarpe, V., Man, P., Sharma, S., Zabrouskov, V., Larsen, B. and Schriemer, D.C. (2017) Neprosin, a selective prolyl endoprotease for bottom-up proteomics and histone mapping. *Molecular & Cellular Proteomics*, *16*(6), 1162-1171.

Zulkapli, M.M.I., Rosli, M.A.F., Salleh, F.I.M., Noor, N.M., Aizat, W.M. & Goh, H.H. (2017) Iso-Seq analysis of *Nepenthes ampullaria*, *Nepenthes rafflesiana* and *Nepenthes× hookeriana* for hybridisation study in pitcher plants. *Genomics data*, 12, 130-131.