Fast and Precise Discovery of New Enzymes

Yuhong Huang[2,3,4], Peter Kamp Busk[1,2], Morten Nedergaard Grell[1], Hai Zhao[3], Lene Lange[1,2]

[1]Barentzymes A/S, Sykehusveien 23, Tromsø, Norway
[2]Department of Biotechnology, Chemistry and Environmental Engineering, Aalborg University Copenhagen, Denmark
[3]Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, Sichuan 610041, PR China
[4]University of the Chinese Academy of Sciences, Beijing 100049, PR China

Background

BIOLOGICAL PRETREATMENT: Agricultural wastes conversion offers new logistic and scientific challenges, one of these is the biomass recalcitrance for the cell wall. Due to the heterogeneity and complexity of plant cell walls, there is a need to discover and develop more efficient and cost effective enzymes for depolymerization of agricultural wastes to valuable products.

PEPTIDE PATTERN RECOGNITION: A novel method, peptide pattern recognition (PPR) was applied for discovery of new β-glucosidases. PPR is a non-alignment based sequence analysis method, which can simultaneously compare multiple sequences at a time and find the characteristic features. This technology makes it easier to predict the potential function of novel enzymes.

THIS WORK: The aim of this work was to discover new β-glucosidases and other plant cell wall degrading enzyme in the *Mucor circinelloides* genome. One of the newly discovered genes, predicted to encode a β-glucosidase (GH3), was cloned and heterologously expressed in *Pichia pastoris*. Not only could PPR pinpoint genes belonging to different GH families but it did also predict the enzymatic function of the genes.

CONCLUSION: Our new bioinformatic approach was a fast and precise method for discovery of a new β-glucosidases from *Mucor circinelloides*.

Peptide Pattern Recognition, PPR

- A new tool for predicting of function from sequence
- Elucidating biological role of the genes
- Going specifically for the enzymes of interest
- Opens for a short cut to finding the genes with the highest potentials

PPR, the principle:

The principle of PPR is to find a limited number of n-mer peptides that are highly conserved in the input. If PPR is unable to find a common set of n-mers for all the proteins provided as input, it will separate the input into groups of proteins that can be defined by n-mer peptides.

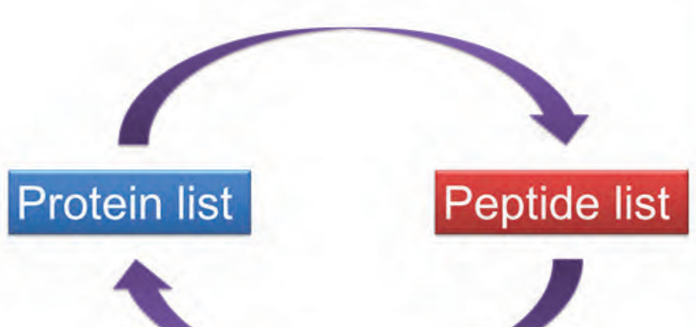
Hat-words (2 points):
Hat
Cap
Bowler
Beanie
Sombrero

Hat-function-words (1 points):
Wear
Would
Head
Want
Like
Put-on
Take-off

tall a would Black I Hat like = 1 hat-word, 2 hat-function-words, Score = 4

he wants his beanie = 1 hat-word, 1 hat-function-words, Score = 3

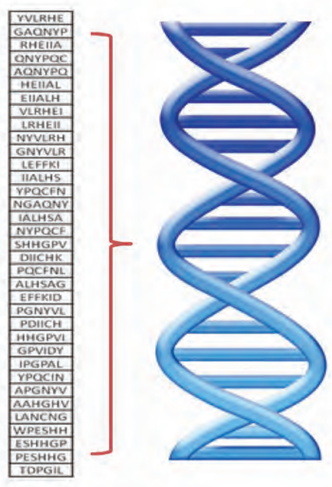
his he wants boots = 0 hat-word, 1 hat-function-words, Score = 1



Mutually dependent

PPR for data mining to discover new sequences

-----PPR for data mining in *Mucor circinelloides* genome (Data provided by Dr. Torres Martínez and JGI)

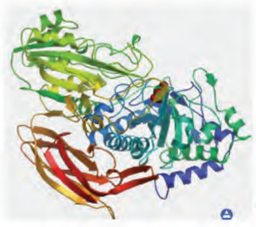


HoTPeP (Homology To Peptide Pattern)
-----PPR data mining in *Mucor circinelloides* genome

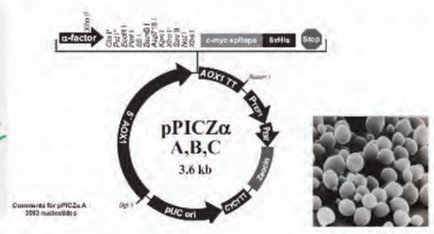
Gene	EC	Known enzyme activity	GH	score	hits	new gene
>3802|2|	3.2.1.21	beta-glucosidase	3	3,55	6	yes
>48|1|rev	3.2.1.58	exo-beta-1,3-glucanase	5	13,58	24	yes
>234|2|	3.2.1.45	beta-1,3-glucanase	5	3,22	9	yes
>2961|7|rev	3.2.1.4	endo-beta-1,4-glucanase	9	5,17	8	yes
>1569|5|	3.2.1.39	beta 1,3-glucanase	16	9,22	17	yes
>2910|6|	3.2.1.6	1,3(4)-beta-glucanase	16	6,83	14	yes
>4543|5|	2.4.1.207	cell wall glucanosyltransferase	16	6,7	12	yes
>153|8|rev	3.2.1.24	AMS1-alpha-mannosidase	38	11,99	17	yes
>2433|2|	3.2.1.4	endo-beta-D-1,4-glucanase	45	33,34	66	no
>3030|8|	3.2.1.113	alpha-mannosidase	47	15,92	28	yes
>3956|2|		alpha-1,2-mannosidase	125	9,6	18	yes

Table 1 Enzyme discoveries with Peptide pattern recognition (PPR) in the *Mucor circinelloides* genome

Expression of the predicted β-glucosidase (GH3)



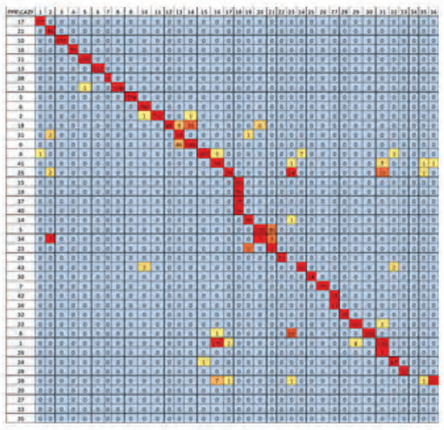
β-glucosidase



EasySelect™ *Pichia* Expression Kit

Of the three GH3 encoding genes found, one was predicted by PPR to encode a β-glucosidase. We expressed this gene in *Pichia pastoris* and found that the purified recombinant protein has β-glucosidase specific activity (1.73 U/mg)

PPR for predicting protein function



Correlation between PPR subfamilies and CAZy subfamilies of 5457 GH13 proteins

-----PPR analysis of 8138 GH13

PPR correctly predicted the function of 91 % (185 of 204 enzymes) of the functionally characterized GH13 proteins.

References

1. Busk PK, Lange L, 2013. Function-based classification of carbohydrate-active enzymes by recognition of short, conserved peptide motifs. Appl Environ Microbiol. 79(11):3380-91
2. Huang Y, Busk PK, Grell MN, Zhao H, Lange L, 2014. Identification of a β-glucosidase from the *Mucor circinelloides* genome by Peptide Pattern Recognition . Enzyme and Microbial Technology. *In press*.