UNIVERSITI PUTRA MALAYSIA

STRUCTURAL PREDICTION AND CLONING OF SMALL, HALOACID DEHALOGENASE-LIKE PROTEIN OF *glaciozyma antarctica* PI12 IN *pichia pastoris* GS115

SYAIDATUL SYAZANA PAUZI

FBSB 2015 171

STRUCTURAL PREDICTION AND CLONING OF SMALL, HALOACID DEHALOGENASE-LIKE PROTEIN OF *Glaciozyma Antarctica* PI12 IN *Pichia Pastoris* GS115

By

SYAIDATUL SYAZANA BINTI PAUZI

Thesis submitted to the Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Cell and Molecular Biology

June 2015

A special dedication to

My parents, who have always been with me through thick and thin,

My supervisor, Dr. Normi Mohd Yahaya, who has always giving me support,

And to the twenty-six letters of the alphabet,

Especially the letters A, B, and C, which taught me a great deal.

Also to you, whom I haven't met/have only glancing acquaintances/are just crazy about each other/haven't seen each other in much too long/are in some way related/will never meet, but will, I trust, despite that, always think fondly of each other…

This one's for you.

With you know what, and you probably know why…

And finally,

In loving memory

Of sleep,

So suddenly taken away from us,

While finishing this thesis…

Abstract of thesis presented to the Department of Cell and Molecular Biology in fulfilment of the requirement for the Degree of Cell and Molecular Biology

STRUCTURAL PREDICTION AND CLONING OF SMALL, HALOACID DEHALOGENASE-LIKE PROTEIN OF *Glaciozyma Antarctica* PI12 IN *Pichia Pastoris* GS115

By

Syaidatul Syazana Binti Pauzi

June 2015

Chairman: Dr. Normi Mohd Yahaya, PhD

Faculty: Biotechnology and Biomolecular Sciences

ABSTRACT

Glaciozyma antarctica PI12 is a psychrophilic yeast isolated from Antarctic sea ice with temperature ranged between -20°C – 15°C near Casey Research Station and its full genome were sequenced. Previously, a scan on its genome for putative phosphatases has led to the discovery of a hypothetical protein termed, LAN_14_281, which contained haloacid dehalogenase-like hydrolase domain. BLAST analysis on this protein showed that it share 58% similarity to haloacid dehalogenase-like protein. Analysis of the protein using SignalP indicates that there is no signal peptide available in LAN_14_281. The constructed phylogenetic tree of LAN_14_281 showed that the protein shares the common ancestors with haloacid dehalogenase-like superfamily. The ProtParam analysis revealed that the theoretical pI for LAN_14_281 are 4.99 with 34 negatively charge residues and 20 positively charged residues. As the role of haloacid dehalogenase can vary and their mechanisms are not fully established, this present study aimed at predicting the structure

computationally and cloning of LAN_14_281 for expression and characterization. The 3D structure of LAN_14_281 was built via homology modelling by using the SWISS-MODEL software. The alignment of the protein with 50 templates revealed that the structure has the highest similarity with human haloacid dehalogenase-like hydrolase domain containing protein 1a (Hdhd1a) with 46% sequence identity. Comparison of these two protein structure revealed that the RMSD are 0.421 Å. The 675bp gene sequence encoding LAN_14_281 was synthesized with the incorporation of EcoRI restriction enzymes sites at the 5' and XbaI restriction enzymes at the 3' ends. The recombinant plasmid with the genes that codes for LAN_14_281 are cloned into *E.coli* TOP10 Dh5α for propagation purposes.

Keywords: *Glaciozyma antarctica* PI12, hypothetical protein, protein production, *Pichia pastoris*

Abstrak tesis yang dikemukan kepada Jabatan Biologi Sel dan Molekul sebagai memenuhi keperluan untuk ijazah Biologi Sel dan Molekul

RAMALAN STRUKTUR DAN PENGKLONAN PROTEIN KECIL BERCIRIKAN HALOASID DEHALOGENASE TERPILIH DARI *Glaciozyma Antarctica* PI12 KE DALAM *Pichia Pastoris* GS115.

Oleh

SYAIDATUL SYAZANA BT PAUZI

June 2015

Pengerusi : Dr. Normi Mohd Yahaya, PhD
Fakulti : Bioteknologi dan Sains Biomolekul

ABSTRAK

Glaciozyma antarctica PI12 merupakan sejenis yis psikrofilik yang diasingkan daripada lautan Antartik di kawasan suhu dalam linkungan -20°C – 15°C berhampiran dengan Pusat Penyelidikan Casey dan keseluruhan genomnya telah pun dijujukkan. Hasil imbasan genom untuk protein fosfatase telah membawa kepada penemuan protein hipotetikal dikenali sebagai LAN_14_281, yang mengandungi domain bercirikan haloasid dehalogenase hydrolase. Analisis BLAST ke atas protein ini menunjukkan protein tersebut berkongsi 58% persamaan terhadap protein dengan ciri-ciri haloasid dehalogenase. Analisis terhadap protein tersebut menggunakan SignalP menunjukkan protein LAN_14_281 tidak mempunyai sebarang isyarat peptida. Pokok filogenetik yang dibina menunjukkan LAN_14_281 berkongsi keturunan yang sama dengan haloasid dehalogenase-seperti superfamili. Analisis ProtParam menunjukkan titik isoelektrik secara teori bagi LAN_14_281 adalah 4.99 dengan 34 atom bercas negatif dan 20 residu bercas positif. Peranan haloasid dehalogenase boleh berubah dan

mekanisma bagi haloasid dehalogenase masih belum difahami sepenuhnya, oleh itu, kajian ini bertujuan untuk meramal struktur secara pengkomputeran dan pengklonan bagi LAN_14_281 untuk pengekspresan dan pencirian. Sktuktur 3D bagi protein LAN_14_281 dibina menggunakan pendekatan homolog menggunakan SWISS-MODEL. Penjajaran protein LAN_14_281 dengan 50 templat menunjukkan struktur bagi protein LAN_14_281 mempunyai persamaan yang paling tinggi dengan struktur domain bercirikan haloasid dehalogenase hydrolase manusia mengandungi protein 1a (Hdhd1a) dengan 46% persamaan jujukan. Hasil perbandingan di antara kedua-dua protein menunjukkan RMSD bagi protein tersebut ialah 0.421 Å. Gen yang mengkodkan LAN_14_281 dengan 675 pasangan bes disintesis dengan penambahan jujukan enzim pemotongan *EcoRI* di hujung 5' dan jujukan enzim pemotongan *XbaI* di hujung 3'. Rekombinan plasmid dengan gen yang mengkodkan hipotetikal protein LAN_14_281 diklonkan ke dalam *E. coli* TOP10 DH5α untuk tujuan propagasi.

Kata kunci: *Glaciozyma antarctica* PI12, protein hipotetikal, penghasilan protein, *Pichia pastoris*

ACKNOWLEDGEMENTS

In the name of Allah, the most gracious and the most merciful.

Alhamdulillah, I am very grateful to Allah S.W.T with His permission and Blessing, I have completed my final year project entitled “Structural prediction and cloning of small, haloacid dehalogenase-like protein of *Glaciozyma antarctica* PI12 in *Pichia pastoris* gs115”.

I would like to express my utmost gratitude to my supervisor, Dr. Normi Mohd Yahaya for her unending support, guidance, advices and approval for this project. My most sincere thanks to my co-supervisors, Dr. Siti Nornaya Oslan for her guidance and supports. Without them, this would not make possible.

A special thanks also to all the members of Protein Engineering Laboratory at Biotech 3, especially Ang Thiau Fu for their guidance, help and encouragement. A special thanks also to Pik Mun Foong for assisting me with this project. I also wish to thank my course mates and all the master and PhD students for their support and motivation throughout this project. Also a million thanks to my beloved family for always believing in me and encouraging me in preceding my dream and to all individual who had contributed in this project.

Thank you.

Approval

The thesis was submitted to the Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences and has been accepted as fulfilment of the requirement for the degree of Cell and Molecular Biology. The member of the Supervisory Committee was as follows:

Normi Mohd Yahaya, PhD
Faculty of Biotechnology and Biomolecular Sciences,
Universiti Putra Malaysia

_____
Janna Ong Abdullah, PhD
Head of Department
Cell and Molecular Biology
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
Date:

Declaration

Declaration by undergraduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by the Department of Cell and Molecular Biology.
- written permissions must be obtained from supervisor before thesis is published (in the form of written, printed, or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules, or any other materials
- there is no plagiarism or data falsification/fabrication in the thesis. The thesis has undergone plagiarism detection software. (TURNITIN)

Signature: _____ Date: _____

Name and Matric No: _____

Declaration by supervisor

This is to confirm that:

- the research conducted and the writing of this thesis was under supervision

Signature: _____
Name of Chairman
of Supervisory
Committee: _____

TABLE OF CONTENTS

LIST OF TABLE **PAGE**

LIST OF FIGURES **PAGE**

LIST OF ABBREVIATION

Bp	Base pair
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleoside 5'-triphosphates
g	gravity
h	hour
Kb	Kilobase pair
LB	Luria Bertani
MCS	Multiple cloning site
min	minutes
OD	Optical density
ORF	Open reading frame
Ori	Origin of replication
PCR	polymerase chain reaction
RE	Restriction enzyme
RE	Restriction enzymes
rpm	revolution per minute
SDS	Sodium dodecyl sulphate
sec	seconds
TAE	Tris-acetate/EDTA buffer
TE	Tris EDTA

1.0 INTRODUCTION

Extremophiles are microorganisms that are able to grow and survive in extreme environments. Proteins, mainly enzymes, which are isolated from extremophiles are of interest due to their special characteristics whereby they can remain stable and are able to function effectively even under extreme conditions. An example of extremophiles are psychrophiles. Psychrophiles live at very low temperatures and can only be found in cold environments, such as the Antarctic.

Glaciozyma antarctica PI12 (Fell *et al*., 1969) (re-classification from *Leucosporidium antarcticum* PI12 is the focus in this project. This psychrophilic yeast was isolated from Antarctic sea ice with temperature ranged between -20°C – 15°C near Casey Research Station. It is an obligatory psychrophilic yeast that belongs to the bud-forming Basidiomycota phylum. It has an optimum growth at 4°C, and maximum growth of 22 °C. However, it is capable to adapt its metabolism in continually low temperatures (the temperature of Antarctic waters ranges from -2.2°C in shelf water to 4°C in open waters, the average temperature being -1°C). The optimal growth for *G. antarctica* PI12 is at 12°C. (Hashim *et al*., 2013)

A scan on the full genome sequence of *G. antarctica* PI12 revealed that a hypothetical protein termed LAN_14_281 contains haloacid dehalogenase-

like domain with several conserved residues and motifs similar to haloacid dehalogenase. Haloacid dehalogenase has been reported to be involved in various physiological functions ranging from dehalogenation, phosphoryl transfer to hydrolysis of phosphate ester and phosphonates. Despite its roles in different physiological functions, in depth knowledge on these proteins are poor and this is especially so for those from the uncharacterized, hypothetical protein category. This present study aims to investigate the possible structure and function of LAN_14_281, a haloacid-dehalogenase-like protein from *G. antarctica* PI12.

The specific objectives of this project are to:

1. Predict the 3-dimensional structure of LAN_14_281 hypothetical protein
2. Construct pPICZα::LAN_14_281 recombinant plasmid for expression in *Pichia pastoris*

6.0 REFERENCES

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., 25, 3389-3402.

Anandakumar S, Shanmughavel P (2008) Computational Annotation for Hypothetical Proteins of MycobacteriumTuberculosis. J Comput Sci Syst Biol 1: 050-062. doi:10.4172/jcsb.1000004

Andreeva, A., D. Howorth, et al. (2004). "SCOP database in 2004: refinements integrate structure and sequence family data." Nucleic Acids Res 32(Database issue): D226-9.

Arnold K, Bordoli L, Kopp J, and Schwede T (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics.,22,195-201.

Bairoch, A., R. Apweiler, et al. (2005). "The Universal Protein Resource (UniProt)." Nucleic Acids Res 33 Database Issue: D154-159.

Baum, D. (2008) Reading a phylogenetic tree: The meaning of monophyletic groups. Nature Education 1(1):190

Boeckmann B., Bairoch A., Apweiler R., Blatter M.-C., Estreicher A., Gasteiger E., Martin M.J., Michoud K., O'Donovan C., Phan I., Pilbout S. and Schneider M. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Research. 31, 365- 370 [PUBMED]

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I. et al. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res., 31, 365-370.

Boo SY, Wong CMVL, Rodrigues KF, Najimudin N, Murad AMA, Mahadi NM (2013) Thermal stress responses in Antarctic yeast, *Glaciozyma antarctica* PI12, characterized by real-time quantitative PCR. Polar Biol 36(3):1–9

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J. and Schwede, T. (2009). Protein structure homology modelling using SWISS-MODEL Workspace. Nature Protocols, 4,1.

Buzzini P, Branda E, Goretti M, Turchetti B (2012) Psychrophilic yeasts from worldwide glacial habitats: diversity, adaptation strategies and biotechnological potential. FEMS Microbiol Eco l82:217–241

Cocca, E., Ratnayake-Lecamwasam, M., Parker, S. K., Camardella, L., Ciaramella, M., di Prisco, G. & Detrich, H. W. (1995) *Proc. Natl. Acad. Sci. USA* 92**,** 1817–1821.

Cuff J. A. and G. J. Barton (1999) Application of enhanced multiple sequence alignment profiles to improve protein secondary structure prediction. Proteins 40, 502-511.

Daly R, Hearn MT (2005). "Expression of heterologous proteins in Pichia pastoris: a useful experimental tool in protein engineering and production". *J. Mol. Recognit.* 18 (2): 119–38.

Dieffenbach, C.W., Lowe, T.M., Dveksler, G.S (1993) General Concepts for PCR primer design. CSH Press, Genome Research 3:30-37.

Eisenberg, D., R. Luthy, et al. (1997). "VERIFY3D: assessment of protein models with three-dimensional profiles." Methods Enzymol 277: 396-404.

Eyrich, V.A., Marti-Renom, M.A., Przybylski, D., Madhusudhan, M.S., Fiser, A., Pazos, F., Valencia, A., Sali, A. and Rost, B. (2001) EVA: continuous automatic evaluation of protein structure prediction servers. Bioinformatics, 17, 1242-1243.

G. E. Tusnády and I. Simon (1998) Principles Governing Amino Acid Composition of Integral Membrane Proteins: Applications to Topology Prediction. Journal of Molecular Biology 283, 489-506.

Galperin MY, Koonin EV (2004) 'Conserved hypothetical' proteins: prioritization of targets for experimental study. Nucleic Acids Res 32: 5452–5463. doi: 10.1093/nar/gkh885

Galperin,M.Y. (2001) Conserved 'hypothetical' proteins: new hints and new puzzles. Comp. Funct. Genomics, 2, 14–18.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005). pp. 571-607

Gasteiger, E., Jung, E. and Bairoch, A. (2001) Swiss-Prot: connecting biological knowledge via a protein database. Curr. Issues Mol. Biol., 3, 47-55.

Guex N and Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. Electrophoresis 18:2714-2723.

Hashim, N. H. F., Bharudin, I., Nguong, D. L. S., Higa, S., Bakar, F. D. A., Nathan, S., Rabu, A., Kawahara, H., Illias, R. M., Najimudin, N., Mahadi,

N. M., & Murad, A. M. A. (2013). Characterization of Afp1, an antifreeze protein from the psychrophilic yeast *Glaciozyma antarctica* PI12. Extremophiles, 17(1), 63-73.

Henrick K, Thornton JM, PQS: "A protein quarternary file server." Trends Biochem. Sci. 1998;23:358-361.

Hiller, K., Schobert, M., Hundertmark, C., Jahn, D. & Münch, R. (2003) JVirGel: calculation of virtual two-dimensional protein gels. Nucleic Acids Res. 31, 3862-3865 [PUBMED]

Hooft, R. W., G. Vriend, et al. (1996). "Errors in protein structures." Nature 381(6580): 272.

Huang, X., and Miller, M. (1991) A time-efficient, linear-space local similarity algorithm. Adv. Appl. Math. 12,337-367.

Hutchinson, E. G. and J. M. Thornton (1996). "PROMOTIF - - a program to identify and analyze structural motifs in proteins." Protein Sci 5(2): 212-20.

Jones, D. T. (1999). "Protein secondary structure prediction based on position-specific scoring matrices." J Mol Biol 292(2): 195-202.

Jones, D. T. and J. J. Ward (2003). "Prediction of disordered regions in proteins from position specific score matrices." Proteins 53 Suppl 6: 573-578.

Jones, D.T. , Taylor, W.R. & Thornton, J.M. "A model recognition approach to the prediction of all-helical membrane protein structure and topology." Biochemistry 33, 3038-3049

Kabsch, W. and C. Sander (1983). "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features." Biopolymers 22: 2577-2637.

Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009). The SWISS-MODEL Repository and associated resources.

Kim D. S., Bouchalkha A., Jacob, J. F. Zhou, J. J. Song, and J. F. Klem Phys. Rev. Lett. 72, 1572 – Published 7 March 1994

Koonin, E.V. and Galperin, M.Y. (2004) 'Conserved hypothetical' proteins: prioritization of targets for experimental study. Nucleic. Acids Res., 32:5452-5463

Kopp J, and Schwede T (2006). The SWISS-MODEL Repository: new features and functionalities. Nucleic Acids ResNucleic Acids Research 34, D315-D318.

Krieger E, Koraimann G, Vriend G (May 2002). "Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field". *Proteins* 47 (3): 393–402

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. and Higgins D.G. (2007)

Laskowski R A, Chistyakov V V, Thornton J M (2005). PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. Nucleic Acids Res., 33, D266-D268.

Laskowski R A, MacArthur M W, Moss D S, Thornton J M (1993). PROCHECK - a program to check the stereochemical quality of protein structures. J. App. Cryst., 26, 283-291.

Laskowski R A, Rullmannn J A, MacArthur M W, Kaptein R, Thornton J M (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR, 8, 477-486. i

Laskowski, R. A., V. V. Chistyakov, et al. (2005). "PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids." Nucleic Acids Res 33(Database issue): D266-268

Lee JH, Lee SG, Do H, Park JC, Kim E, Choe YH, Han SJ, Kim HJ (2013) Optimization of the pilot-scale production of an ice-binding protein by fed-batch culture of *Pichia pastoris*. Appl Microbiol Biotechnol 97:3383–3393

Lee JH, Park AK, Do H, Park KS, Moh SH, Chi YM, Kim HJ (2012a) Structural basis for antifreeze activity of ice-binding protein from arctic yeast. J Biol Chem 287:11460–11468

Lee JK, Park KS, Park S, Park H, Song YH, Kang SH, Kim HJ (2010) An extracellular ice binding glycoprotein from an Arctic psychrophilic yeast. Cryobiology 60:222–228

Lee SG, Koh HY, Lee JH, Kang SH, Kim HJ (2012b) Cryopreservative effects of the recombinant ice-binding protein from the arctic yeast Leucosporidium sp. on red blood cells. Appl Biochem Biotechnol 167:824–834

Li, J., Xia, F., Li, W.X. (2003). Coactivation of STAT and Ras is required for germ cell proliferation and invasive migration in Drosophila. Dev. Cell 5(5): 787--798.

Lo Conte, L., Brenner, S.E., Hubbard, T.J.P., Chothia, C. and Murzin, A.G. (2002) SCOP database in 2002: refinements accommodate structural genomics. Nucleic Acids Res., 30, 264-267.

Lukas Käll, Anders Krogh and Erik L. L. Sonnhammer. A Combined Transmembrane Topology and Signal Peptide Prediction Method. Journal of Molecular Biology 338(5):1027-1036, May 2004.

Lupas A. N., M. Van Dyke and J. Stock (1991) Predicting Coiled Coils from Protein Sequences. Science 252, 1162-1164

Menne, K. M. L., Hermjakob, H. and Apweiler, R. (2000) A comparison of signal sequence prediction methods using a test set of signal peptides. Bioinformatics 16, 741-742.

Merritt, Ethan A. and Bacon, David J. (1997). Raster3D: Photorealistic Molecular Graphics, Methods in Enzymology 277, 505-524.

Michael J. Hartshorn (2002) AstexViewer: A visualisation aid for structure-based drug design Journal of Computer Aided Molecular Design16 , 871-881.

Minion, F. C., C. Adams, and T. Hsu. 2000. R1 region of P97 mediates adherence of *Mycoplasma hyopneumoniae* to swine cilia. Infect. Immun.68**:**3056-3060

Mulder, N. J., R. Apweiler, et al. (2005). "InterPro, progress and status in 2005." Nucleic Acids Res 33 Database Issue: D201-205.

Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Barrell, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P. et al. (2003) The InterPro Database, 2003 brings increased coverage and new features. Nucleic Acids Res., 31, 315-318.

Nakamura, Y., Gojobori, T. and Ikemura, T. (2000) Nucl. Acids Res. 28, 292. Codon usage tabulated from the international DNA sequence databases: status for the year 2000.Use the "Insert Citation" button to add citations to this document.

Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. 10, 1-6 [PUBMED]

Nielsen, H., Engelbrecht, J., von Heijne, G. and Brunak, S. (1996) Defining a Similarity Threshold for a Functional Protein Sequence Pattern: The Signal Peptide Cleavage Site. Proteins 24 , 165-177 [PUBMED]

Ohi H, Miura M, Hiramatsu R, Ohmura T: The positive and negative cis-acting elements for methanol regulation in the *Pichia pastoris* AOX2 gene. Mol Gen Genet 1994, 243:489-499.

Orengo, C.A., Pearl, F.M.G., Bray, J.E., Todd, A.E., Martin, A.C., Lo, C.L. and Thornton, J.M. (1999) The CATH Database provides insights into protein structure/function relationships. Nucleic Acids Res., 27, 275-279.

Ouali M. and R. D. King (2000) Cascaded multiple classifiers for secondary structure prediction. Protein Science 9, 1162-1176

Pavlova K, Rusinova-Videva S, Kuncheva M, Kratchanova M, Gocheva M, Dimitrova S (2011) Synthesis and characterization of an exopolysaccharide by antarctic yeast strain Cryptococcus laurentii AL 100. Appl Biochem Biotechnol 163:1038–1052Use the "Insert Citation" button to add citations to this document.

Pearl, F., A. Todd, et al. (2005). "The CATH Domain Structure Database and related resources Gene3D and DHS provide comprehensive domain family information for genome analysis." Nucleic Acids Res 33 Database Issue: D247-51.

Per J. Kraulis (1991) MOLSCRIPT: A Program to Produce Both Detailed and Schematic Plots of Protein Structures, Journal of Applied Crystallography 24, 946-950.

Preumont, A., Rzem, R., Vertommen, D., & Van Schaftingen, E. (2010). HDHD1, which is often deleted in X-linked ichthyosis, encodes a pseudouridine-5'-phosphatase. *The Biochemical Journal*, *431*(2), 237–44.

Rapoport T.A., Jungnickel B. and Kutay U. (1996) Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. Annual review of biochemistry 65, 271-303 [PUBMED]

Rapoport, TA., Jungnickel, B. & Kutay, U. (1996) Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. Annu. Rev. Biochem. 61, 271-303 [PUBMED]

Rayapureddi, J.P., Kattamuri, C., Steinmetz, B.D., Frankfort, B.J., Ostrin, E.J., Mardon, G., Hegde, R.S. (2003). Eyes absent represents a class of protein tyrosine phosphatases. Nature 426(6964): 295--298.

Rezende AM, Folador EL, Resende DdM, Ruiz JC (2012) Computational Prediction of Protein-Protein Interactions in *Leishmania* Predicted Proteomes. PLoS ONE 7(12): e51304. doi:10.1371/journal.pone.0051304

Rost B. (2001) Protein secondary structure prediction continues to rise. Journal of Structural Biology 134, 204-218.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Plainview, NY).

Sayle, R.A. and Milner-White, E.J. (1995) RASMOL: biomolecular graphics for all. Trends Biochem. Sci., 20, 374.

Schreiber M. and Brown C. (2002) Compensation for nucleotide bias in a genome by representation as a discrete channel with noise.Bioinformatics 18 , 507-512 [PUBMED]

Schwede, T., J. Kopp, et al. (2003). "SWISS-MODEL: An automated protein homology-modeling server." Nucleic Acids Res 31(13): 3381-3385.

Schwede, T., Kopp, J., Guex, N. and Peitsch, M.C. (2003) SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res., 31,3381-3385.

Shahbaaz M, Md. ImtaiyazHassan, Ahmad F (2013) Functional Annotation of Conserved Hypothetical Proteins from *Haemophilus influenzae* Rd KW20. PLoS ONE 8(12): e84263. doi:10.1371/journal.pone.0084263

Sidell, B. D.Vayda, M. E., & Small, D. J. (1997) in *VI SCAR Antarctic Communities*, eds. Battaglia, B., Valencia, J. & Walton, D. W. H. (Cambridge Univ. Press, Cambridge, U.K.), in press

Siew N and Fischer D (2003) Analysis of singleton ORFans in fully sequenced microbial genomes. Proteins 53(2):241-51

Söding J. (2005) "Protein homology detection by HMM-HMM comparison." Bioinformatics 21, 951-960. doi:10.1093/bioinformatics/bti125.

Somero, G. N. and Weinstein, R. B. and (1998). Effects of temperature on mitochondrial function in the Antarctic fish *Trematomus bernacchii. J. Comp. Physiol. B* 168, 190- 196.

Turchetti B, Thomas-Hall SR, Connell LB, Branda E, Buzzini P, Theelen B, Mu¨ller WH & Boekhout T (2011) Psychrophilic yeasts from Antarctica and European glaciers: description of Glaciozyma gen. nov., Glaciozyma martinii sp. nov. and Glaciozyma watsonii sp. nov. Extremophiles 15: 573–586.

Van Gunsteren, W. F., S. R. Billeter, et al. (1996). Biomolecular Simulations: The GROMOS96 Manual and User Guide. Zürich, VdF Hochschulverlag ETHZ.

Velankar, S., P. McNeil, et al. (2005). "E-MSD: an integrated data resource for bioinformatics." Nucleic Acids Res 33 Database Issue: D262-265.

Von Heijne, G. (1985) Signal sequences. The limits of variation. Journal of molecular biology 184 , 99-105 [PUBMED]

Westbrook, J., Feng, Z., Chen, L., Yang, H. and Berman, H.M. (2003) The Protein Data Bank and structural genomics. Nucleic Acids Res., 31, 489-491.

Westbrook, J., Z. Feng, et al. (2003). "The Protein Data Bank and structural genomics." Nucleic Acids Res 31(1): 489-491.

Wheeler, D. L., T. Barrett, et al. (2005). "Database resources of the National Center for Biotechnology Information." Nucleic Acids Res 33 Database Issue: D39-45.

Zdobnov, E. M. and R. Apweiler (2001). "InterProScan - an integration platform for the signature-recognition methods in InterPro." Bioinformatics 17(9): 847-848.

Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "A greedy algorithm for aligning DNA sequences", J Comput Biol 2000; 7(1-2):203-14

Zheng, N. and Gierasch, L..M. (1996) Signal sequences: the same yet different. Cell 86, 849-852 [PUBMED]

Zsuzsanna Dosztá, Veronika Csizmók, Péter Tompa and István Simon (2005) The Pairwise Energy Content Estimated from Amino Acid Composition Discriminates between Folded and Intrinsically Unstructured Proteins J. Mol. Biol. 347, 827-839.