Physical Sciences

Pak. J. Sci. Ind. Res. 2008 **51**(4) 185-190

# Phylogenetic and Computational Structure-Function Studies of α-1,4-Glucosidase (Maltase) From Baker's Yeast (*Saccharomyces cerevisiae*)

**Mushtaq Hussain[a], Javed A. Qureshi[b], Shagufta A. Shaikh[a] and Liaquat Sultana[a]***

[a]Food and Marine Resources Research Centre, PCSIR Laboratories Complex, Karachi - 75280, Pakistan

[b]HEJ Research Institute of Chemistry, International Centre for Chemical and Biological Sciences, University of Karachi, Karachi - 75270, Pakistan

(received June 25, 2008; revised July 29, 2008; accepted July 30, 2008)

**Abstract.** Study, through blasting, cladogenesis, multiple sequence alignment and protein homology modeling, of α,1,4 glucosidase from *S. cerevisiae* YJM789, showed the presence of different paralogues and orthologues of maltase in different genera of fungi and prokaryotes. The sequences of glucosidases contained 4 characteristic consensus regions. In the tertiary structure (modelled) of Baker's yeast maltase, all the residues of consensus region were congregated in the central region of the folded protein, rendering the formation of catalytic groove. On the basis of the orientation and spatial location of residues in catalytic groove, Asp200 is proposed to be the second substrate-binding site.

**Keywords:** maltase, baker's yeast, protein homology modelling

## Introduction

The functionality of baker's yeast (*Saccharomyces cerevisiae*) is mediated by enzymes, amylase (α-1,4-glycosidase) (Tamaki, 1978) and maltase (α-1,4-glucosidase) (Khan and Eaton, 1971). Genetically, the biosynthesis of maltase is regulated by the set of at least 5 unlinked genes ascribed as *MAL* loci. These include *MAL1* through *MAL4* and *MAL6*, each of which has different function (Charron *et al*., 1989; Michels and Needleman, 1984). For instance *MAL1* codes for a tranporter protein, maltose permease, that causes internalization of maltose into the cell. *MAL2* is a structural gene of maltase while *MAL3* is the positive regulatory protein (Higgins *et al*., 1999). Functionally, maltase from *S. cerevisiae* belongs to group of glucosyl hydrolases (EC 3.2.1.-) which is further segregated into 57 structural families (Henrissat and Bairoch, 1996; Henrissat, 1991). More specifically, the enzyme (maltase) belongs to the family of 13 enzymes (α amylase family), which tend to act on α-1,4- and/or α-1,6-glucosidic linkages present in different carbohydrate moieties (Yamamoto *et al*., 2004; Henrissat, 1991). Primary structures of many of the mentioned proteins are known and have been rationally compared. Resultantly, four highly conserved regions (region I-IV) along with three acidic residues in the same have been reported in most of the compared sequences, suggesting their role in structural and/or catalytic integrity of the molecules (Yamamoto *et al*., 2004; Brayer *et al*., 1995; Machius *et al*., 1995; Svensson, 1988). However, the nature of amino acids responsible for the separate recognition of α-1,4- and α-1,6-glycosidic linkages is still debatable. Similarly, the cognizance of substrate binding site(s) of maltase from baker's yeast *S. cerevisiae* is still based on sequence comparison, site directed mutagenesis, chimeric protein and inhibition studies (Yamamoto *et al*., 2004; Yao *et al*., 2003; Frandsen *et al*., 2002). Moreover, unavailability of any 3D structural data of α-1,4-glucosidase did not help in deciphering structural and functional aspects of the protein.

The present study has been designed to unravel the evolutionary course of the yeast maltase by generating multiple sequence alignment and consequently the phylogenetic tree. In addition to these tools, algorithms of protein homology modeling and tertiary structure analysis have been exploited to explore the structural and functional aspects of the enzyme at the molecular level. Fold recognition has been used to validate the findings in functional terms. At the best of our knowledge this is the first attempt to study α-1,4-glucosidase of *S. cerevisiae* at the tertiary scale level.

## Materials and Methods

**Primary sequences, blasting and multiple sequence alignment.** Amino acid sequence of α-1,4-glycosidase (GenBank accession No. EDN 64908) from *S. cerevisiae* YJM789 was retrieved from NCBI data bank. The programmes FASTA and BLAST (Altschul *et al*., 1997) were subsequently used to retrieve the primary sequences and structural homologues of the protein. Multiple sequence alignment was generated by default parameters of programme CLUSTAL X and manually adjusted wherever found necessary (Thompson *et al*., 1997). The alignment file was analyzed using the programme GENE DOC

*Author for correspondence; E-mail: liaqatsultanapcsir@gmail.com

(Nicholas *et al*., 1997). Evolutionary lineage was calculated using the phylogeny inference programme, Phylip (Felsenstein, 1993).

**Homology modelling.** The atomic coordinates of α-1,6-glucosidase of *Bacillus cereus* (PDBid;1UOK) was obtained from Brookhaven Protein Data Bank (Bernstein *et al*., 1977). The 3D models of yeast maltase were developed using the 1UOK as template (Watanabe *et al*., 1997). Briefly, the programmes SWISS-MODEL (Schwede *et al*., 2003) and 3D-JIGSAW (Bates *et al*., 2001) were used to construct the models with manual input of PDBid.

**Tertiary structure analysis.** The structural coordinates of yeast maltase were viewed using Swiss-Pdb viewer (Guex and Peitsch, 1997) and Weblab. Model. The structure were analyzed for structural and thermodynamic stability using ANOLEA (Melo and Feytman, 1998); Verify 3D (Elsenberg *et al*., 1997); GROMOS (van Gunsteren *et al*., 1996) and Swiss-Pdb viewer and PROCHECK and Whatcheck (Laskowski and Kato, 1980). Fold recognition was performed by 3D-pssm algorithm (Kelley *et al*., 2000).

## Results and Discussion

**Sequence Comparison and Phylogeny.** Holistically, the sequence identity among the sequences retrieved after non-redundant BLASTing of α-1,4-glycosidase from *S. cerevisiae* YJM789 ranges from 58 to 99% with the target protein. However, multiple sequence alignment has shown non-homogenous distribution of sequence homology among the glucosidases of different organisms. Specifically, in comparison to N-terminal, the C-terminal (middle) of the proteins has been found with greater identities and/or similarities among its different organismic versions. Importantly, despite the nonisofunctional nature of the proteins, the characteristic four conserved regions I, II, III and IV (Yamamoto *et al*., 2004; Svensson, 1988) have been found in all homologues of yeast α-1,4-glucosidase at least at partial scale. Precisely, in conserved region I (Asp97-Cys103) and II (Asp200-Gly208), the conservancy has been more pronounced, as discrepancies are only observed limited to substitution of residue with an iso-functional amino acid except for Cys103, which is conserved only in the homologues of yeast strains and replaced by Thr in glucosidases of ascomycota and basidomycota and Ser in some members of firmicutes at the respective position. The conserved region II (Asp200-Gly208) has shown minimum discrepancy in comparison to other three regions, suggesting its potential role in catalysis. In the region, only the last three residues showed some degree of isofunctional variation among the glucosidases of different organisms. Additionally, at the first residue Asp200, Ser has been found in multicellular members of ascomycota like *Aspergillus* and *Ustilago*. It has been earlier stipulated that the acidic amino acids like Asp in the conserved (consensus) regions are crucial for enzymatic activity of maltase (Yamamoto *et al*., 2004; Brayer *et al*., 1995); in consideration of this, it is reasonable to assume that the yeast maltase may be catalytically more active than the maltase from multicellular fungi. Akin to conserved region II, the subsequent region III has shown considerable similarities among glucosidases of different organisms. Absence of any acidic residue suggests that the region has structural importance rather than catalytic. In contrast to region II, region IV has completely conserved last three residues among glucosidases of different organisms. However, the first three residues have exhibited some degree of isofunctional substitutions (Fig. 1). As a whole the considerable degree of homologues of glucosidase suggests the indispensability of the enzymes in the survival of the organisms at the transkingdom level. Moreover, patches of strong identity may also implicate gene duplication followed by the emergence of orthologs and paralogs (Cliften *et al*., 2006).

Phylogenetic tree clearly segregated the glucosidases homologues, present among different organisms, at the taxonomic level. Fundamentally, three nodal branches have been observed, two of them leading to ascomycota while the third bifurcated into basidomycota and prokaryotic bacterial group namely firmicutes. The branch leading to *Saccharomyces* also embarks on other unicellular members of ascomycota (Fig. 2). The assembly of all *Saccharomyces* and other unicellular fungi on the single nodal branch suggests their common ancestry (Cliften *et al*., 2006; Liti and Louis, 2005). BLASTing has revealed the presence of at least 9 homologues (orthologues) of α-1,4-glucosidase in the genome of *S. cerevisiae* YJM789; similarity among them ranges from 80% to 99%. The values suggest occurrence of extensive gene and/or genome duplication events, which rendered the increased number of copies of the understudy gene. Indeed it has been earlier deduced that around 72% of the yeast genome is the result of gene duplication and subsequent substitution (Wong *et al*., 2002; Seoighe and Wolfe, 1999). According to Piskur (2001), genome duplication has great influence on the fermentability of the yeast. The branch length of other unicellular fungi except for *Kluyveromyces* excluded the possibility of lateral gene transfer among *Saccharomyces* and other single cell fungi (Liti and Louis, 2005). However, close proximity of *Kluyveromyces* and *Saccharomyces* homologue of α-1,4-glucosidase may implicate the involvement of horizontal gene transfer. Earlier bifurcation of the nodal branch, leading to unicellular ascomycota into two subbranches (one suggesting evolution of S*accharomyces* and the second including other members of ascomycota),

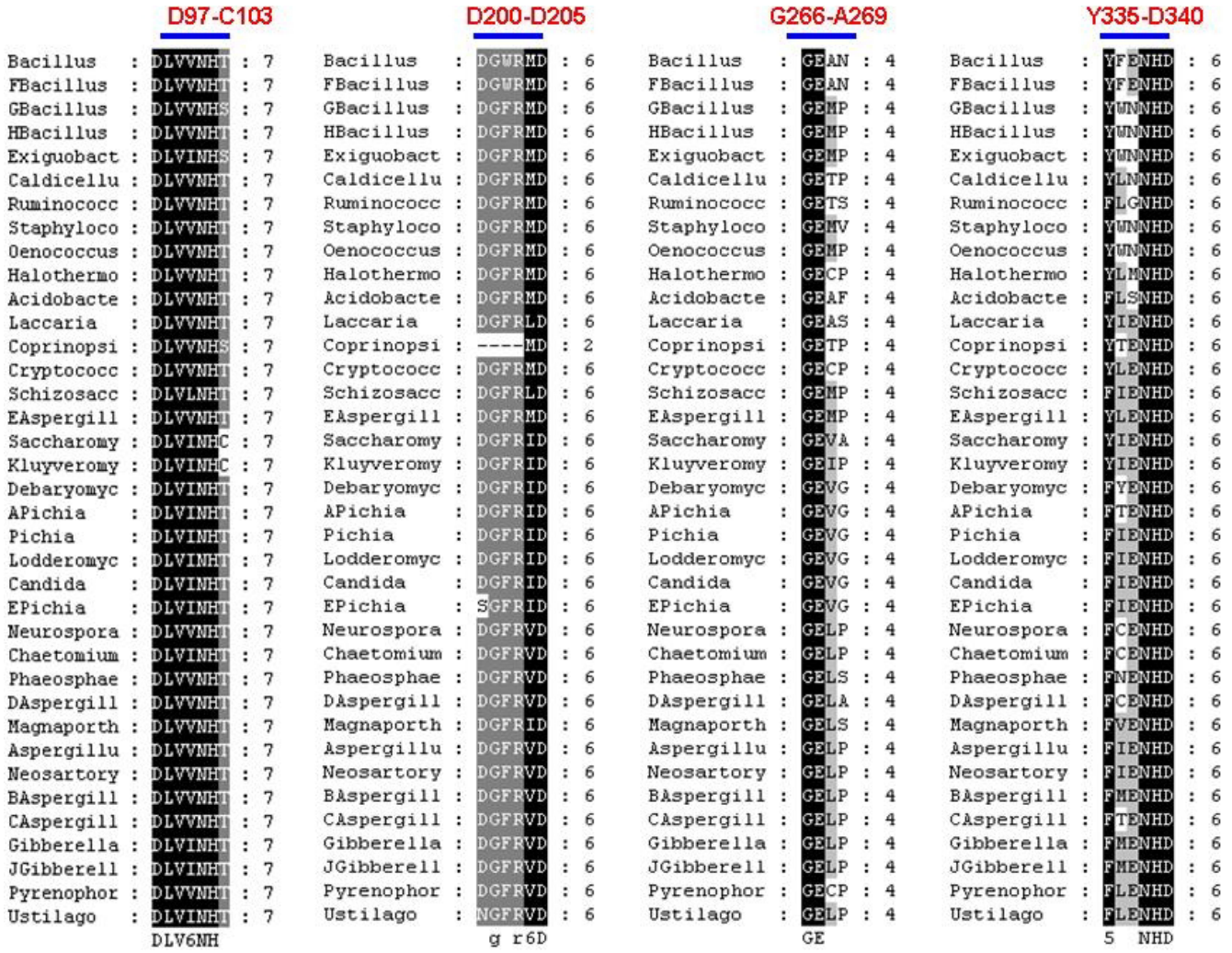


**Fig. 1.** Multiple Sequence Alignment: Selected conserved regions of different α-glucosidases are mentioned and annotated according to their spatial order. Note the presence of two acidic residues (Asp; D) in conserved region II.

may be inferred in terms of gene duplication in the ancestral organisms, followed by vertical gene transfer. However, it is worth mentioning here that almost all the homologues of α-1,4-glucosidase are actually α-1,6-glucosidase in nature suggesting the process of paralogous gene formation in the species distant to yeast (Cliften *et al*., 2006).

**Three-dimensional structures.** Overall structural features of modelled α-1,4-glucosidase from *S. cerevisiae* YJM789 are similar to those found in template α-1,6-glucosidase from *B. cereus*. The fact is validated by the RMS deviation (1.4′Å) of Cα backbone between the modelled and template protein. The α-1,4-glucosidase molecule comprises of three domains namely N-terminal, followed respectively by subdomain and the C-terminal domain as suggested by Watanabe *et al*. (1997). The topology of N-terminal domain is similar to that found in a (β/α)8 barrel or TIM barrel containing proteins (Banner *et al*., 1975). Interestingly, the mentioned region is somewhat sandwiched between the subdomain and C-terminal domain. Such spatial location of N-terminal has also been observed in other α-1,6-glucosidase for instance amylase (Matsuura *et al*., 1984), pig pancreatic amylase (Larson *et al*., 1994) etc. Both N-terminal and subdomain have been found to contain both α-helices and β-pleated sheets with disparaging number with its template and other α-1,6-glucosidases. Intriguingly, the C-terminal of α-1,4-glucosidase is exteriorly localized and mainly contains six antiparallel β-pleated sheets in contrast to 8 such sheets, observed, in α-1,6-glucosidase from *B. cereus* (Watanabe *et al*., 1997). The helices and β-pleated sheets, particularly those present in N-terminal and subdomain

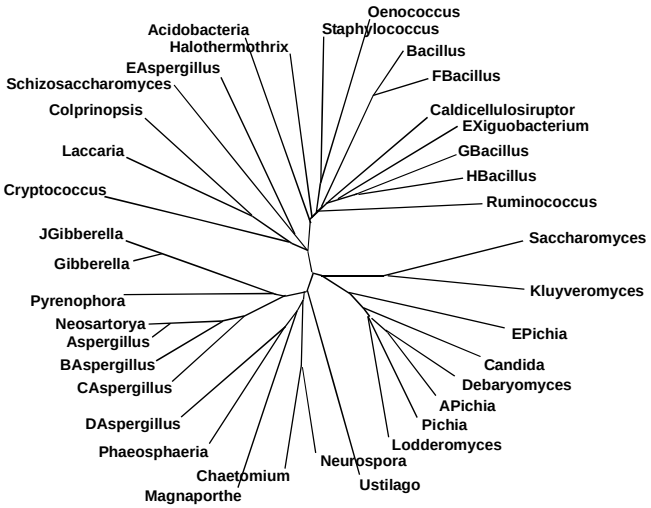


**Fig. 2.** Phylogenetic tree as developed by neighbour joining method. The proximal right and left nodal branches respectively belong to firmicutes and members of basidomycota. The distal right and left nodal branches respectively embark unicellular and multicellular members of ascomycota.

region are intervened by coils and loops of different span (Fig. 3). Despite the presence of 5 Cys residues, no disulphide bridges were noted in the modelled molecule; this is not surprising in the light of intracellular nature of the enzyme. Absence of any transmembrane signature sequence in the molecule has further strengthened this notion. Rationally, the reducing environment inside the cell links the free end of Cys residues with hydrogen, thus halting the formation of any cysteine bridge. Although the absence of cysteine bridges may suggest the vulnerability of the molecule, but contrarily the total free energy ($\Delta G$) of the molecule has been deduced as -18974 KJ/mole, indicating thermodynamic stability of the enzyme in the cytosol. Moreover, the spatial placement of enzyme residues with reference to their $\Phi$ and $\psi$ angles in the Ramachandran plot (Ramachandran and Sasiskharan, 1968) also validates the legitimacy of the protein on the architectural basis.

**Catalytic Site.** Constant conservancy of four regions dispersed among different homologues of $\alpha$-1,4-glucosidase suggests their importance in the structural and functional aspects of the enzymes. In the modelled yeast maltase, it has been observed that despite the substantial difference along the domain placement in the primary structure of the enzyme, on thermodynamically stable folding of the protein, these regions of consensus has been found congregated at the specific central region. (Fig. 4). This conglomeration results in the formation of catalytic groove through which an active residue could possibly bind with the substrate (Fig. 5). Electrostatic space topology suggests that the groove opening is sufficient for the accessibility of the active residue to the disaccharide substrate molecule. Specifically, only the residues of consensus regions I, II and III were observed to be accessible to the substrate molecule. Indeed, as mentioned earlier, that acidic residues (Asp) present in these regions are the key player in mediating the enzyme activity. Several site-directed mutagenesis and enzyme inhibition studies have suggested that Asp205 in the consensus region II is involved in the binding to the substrate molecule (Yamamoto *et al*., 2004). Conversely, it has also been stipulated that single residue may not be

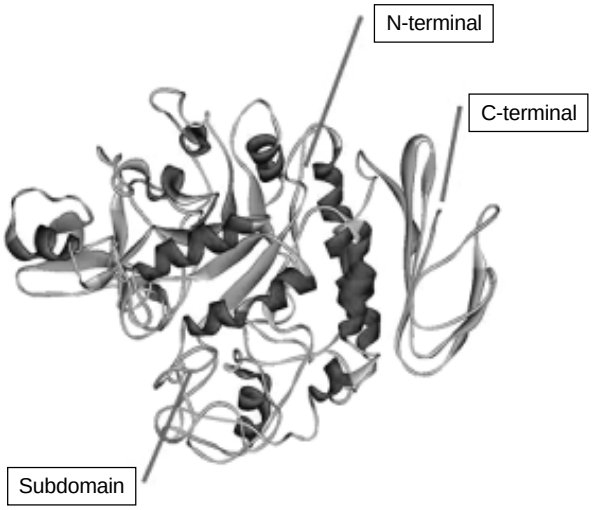


**Fig. 3.** Tertiary structure of $\alpha$-1,4-glucosidase from *S. cerevisiae*. All three domains are annotated accordingly.

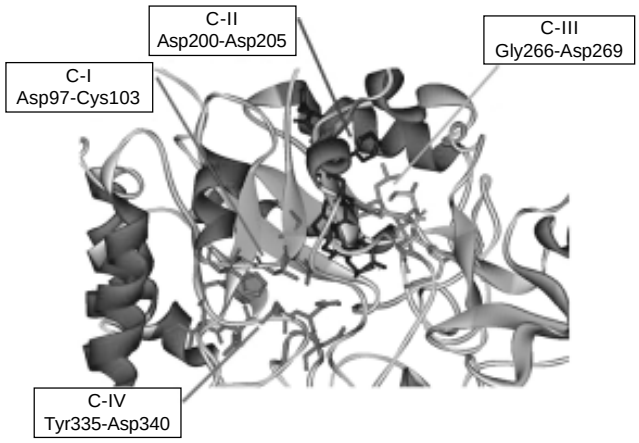


**Fig. 4.** Catalytic region of $\alpha$-1,4-glucosidase from *S. cerevisiae*. Note the accumulation of cuddling of conserved regions residues at the central region of the molecule. Conserved region I, II, III and IV are respectively coloured as green, blue, orange and pink.

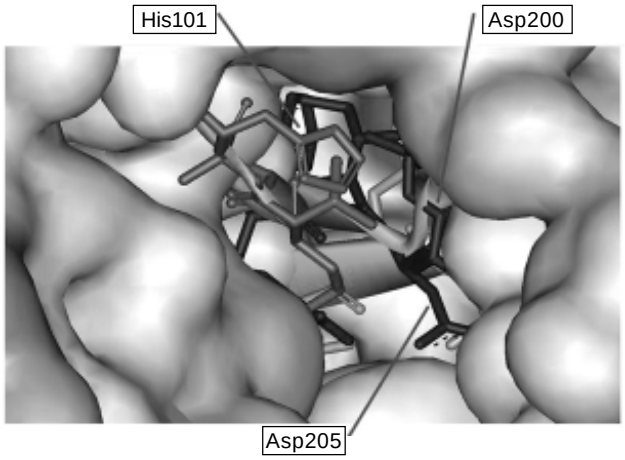


**Fig. 5.** Catalytic groove: Electrostatic surface established by the assemblage of conserved region residues. Importantly, a groove has an opening for the access of substrate molecule to active residue. The active residues are labeled accordingly.

sufficient to bind with the two glucose moieties of the substrate, thereby suggesting possibility of the presence of second substrate site (Yao *et al*., 2003). However, hitherto no such information has been extracted. In the current studies, it was observed that in addition to Asp205 of conserved region II, Asp200 of the same region also exhibited orientation as that of Asp205 and was accessible to the substrate-binding site through catalytic groove. Here, it is plausible to infer that Asp200 may act as a second binding site to the maltose and subsequently cleaves it into 2 glucose moieties (Fig. 5). Furthermore, it is also anticipated that the immense negative charge, developed by the presence of the two Asp residues at the catalytic groove, may get counter balanced by the His101 present in the consensus region I.

## Conclusion

Sequence comparison and phylogenetic inference implicates that glucosidases are important and perhaps indispensable protein for the normal metabolism in yeast and related organisms. Multiple sequence comparison has indicated that at least four consensus regions are always present with mostly isofunctional discrepancies in all homologues of glucosidases of yeast. It is likely that among unicellular fungi, the enzyme and its paralogues are evolved by the gene and/or genome duplication followed by vertical gene transfer from the ancestral organisms. However, the possibility of lateral gene transfer between *Saccharomyces* and *Kluyveromyces* could not be overruled. Structurally, in contrast to similarities in the holistic comparison of model and template proteins, the yeast maltase appears less structurally defined than the other α-1,6-glucosidases. Importantly, all the conserved domain residues have been found congregated at the same site, forming a catalytic groove in the tertiary structure of the compound. But it is inferred that only Asp200 and Asp205 are involved in binding with substrate. It is suggested here that experimental studies like site-directed mutagenesis, enzyme inhibition and chimeric protein formation may explicate the details of the proposed second substrate binding site present in yeast α-1,4-glucosidase.

## References

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

Banner, D.W., Bloomer, A.C., Petsko, G.A., Philips, D.C., Pogson, C.I., Wilson, I.A., Corran, P.H., Furth, A.J., Milman, J.D., Offord, R.E., Priddle, J.D., Waley, S.G. 1975. Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 angstrom resolution using amino acid sequence data. *Nature* **255:** 609- 614.

Bates, P.A., Kelley, L.A., MacCallum, R.M., Sternberg, M.J. 2001. Enhancement of protein modelling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins*, Suppl. **5:** 39-46.

Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F. Jr., Brice, M.D., Rodgers J.R., Kennard, O., Shimanuchi, T., Tasumi, M. 1977. The Protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112:** 535-542.

Brayer, G.D., Luo, Y., Withers, S.G. 1995. The structure of human pancreatic alpha-amylase at 1.8Å resolution and comparisons with related enzymes. *Protein Sci.* **4:** 1730-1742.

Charron, M.J., Read, E., Haut, S.R., Michels, C.A. 1989. Molecular evolution of the telomere-associated *MAL* loci of *Saccharomyces*. *Genetics* **122:** 307-316.

Cliften, P.F., Fulton, R.S., Wilson, R.K., Johnston, M. 2006. After the duplication: Gene loss and adaptation in *Saccharomyces* genomes. *Genetics* **172:** 863-872.

Elsenberg, D., Luthy, R., Bowie, J.U. 1997. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol*. **277:** 396-404.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.57C, distributed by the author. Department of Genetics, University of Washington, Seattle, USA.

Frandsen, T.P., Palcic, M.M., Svensson, B. 2002. Substrate recognition by three family of 13 yeast α-glucosidases. *Eur. J. Biochem.* **269:** 728-734.

Guex, N., Peitsch, M.C. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein

modeling. *Electrophoresis* **18:** 2714-2723.

Henrissat, B., Bairoch, A. 1996. Updating the sequence based classification of glucosyl hydrolases. *Biochem. J.* **316:** 695-696.

Henrissat, B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280:** 309-316.

Higgins, V.J., Braidwood, M., Bell, P., Bissinger, P., Dawes, I.W., Attfield, P.V. 1999. Genetic evidence that high non-induced maltase and maltose permease activities, governed by *MALx3*-encoded transcriptional regulators, determine efficiency of gas production by baker's yeast in unsugared dough. *Appl. Environ. Microbiol.* **65:** 680-685.

Kelley, L.A., MacCalluum, R.M., Sternberg, M.J. 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* **299:** 499-520.

Khan, N.A., Eaton, N.R. 1971. Genetic control of maltase formation in yeast. I. Strains producing high and low basal levels of enzyme. *Mol. Gen. Genet.* **112:** 317-322.

Larson, S.B., Greenwood, A., Cascio, D., Day, J., McPherson, A. 1994. Refined molecular structure of pig pancreatic α-amylase at 2.1Å resolution. *J. Mol. Biol.* **235:** 1560-1584.

Laskowski, M.Jr., Kato, I. 1980. Protein inhibitors of proteinases. *Annu. Rev. Biochem*. **49:** 593-626.

Liti, G., Louis, E.J. 2005. Yeast evolution and comparative genomics. *Annu. Rev. Microbiol.* **59:** 135-153.

Machius, M., Wiegand, G., Huber, R. 1995. Crystal structure of calcium depleted *Bacillus licheniformis* α-amylase at 2.2Å resolution. *J. Mol. Biol.* **246:** 545-559.

Matsuura, A.Y., Kusunoki, M., Harada, W., Kakudoi, M. 1984. Structure and possible catalytic residues of Taka-amylase A. *J. Biochem.* **95:** 697-702.

Melo, F., Feytmans, E. 1998. Assessing protein structures with a non-local atomic interaction energy. *J. Mol. Biol.* **277:** 1141-1152.

Michels, C.A., Needleman, R.B. 1984. The dispersed, repeated family of *MAL* loci in *Saccharomyces* spp. *J. Bacteriol.* **157:** 949-952.

Nicholas, K.B., Nicholas H.B. Jr., Deerfield II, D.W. 1997. Gene Doc: Analysis and visualization of genetic variation. *EMBNEW* **4:** 14.

Piskur, J. 2001. Origin of the duplicated regions in the yeast genomes. *Trends Genet.* **17:** 302-303.

Ramchandran, G.N., Sasiskharan, V. 1968. Conformation of polypeptides and proteins. *Advan. Prot. Chem.* **23:** 283-437.

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

Seoighe, C., Wolfe, K.H. 1999. Yeast genome evolution in the post-genome era. *Curr. Opin. Microbiol*. **2:** 548-554.

Svensson, B. 1988. Regional distant sequence homology between amylases, alpha-glucosidases and trans-glucano-sylases. *FEBS Lett*. **230:** 72-76.

Tamaki, H. 1978. Genetic studies of ability to ferment starch in *Saccharomyces*: gene polymorphism. *Mol. Gen. Genet.* **164:** 205-209.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G. 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. **25:** 4876-4882.

van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Huenenberger, P.H., Krueger, P., Mark, A.E., Scott, W.R.P., Tironi, I.G. (eds.), 1996. *Biomolecular Simulations: The GROMOS96 Manual and User Guide*. VdF: Hochschulverlag AG an der ETH, Zürich, Switzerland.

Watanabe, K., Hata, Y., Kizaki, H., Katsube, Y., Suzuki, Y. 1997. The refined crystal structure of *Bacillus cereus* oligo-1,6-glucosidase at 2.0Å resolution: Structural characterization of proline substitution sites for protein thermo stabilization. *J. Mol. Biol.* **269:** 142-153.

Wong, S., Butler, G., Wolfe, K.H. 2002. Gene order evolution and paleopolyploidy in hemiascomycete yeast. *Proc. Natl. Acad. Sci. USA.* **99:** 9272-9277.

Yao, X., Mauldin, R., Byers, L. 2003. Multiple sugar binding sites in alpha-glucosidase. *Biochim. Biophys. Acta. Proteins & Proteomics* **1645:** 22-29.

Yamamoto, K., Nakayama, A., Yamamoto, Y., Tabata, S. 2004. Val216 decides the substrate specificity of α-glucosidase in *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **271:** 3414-3420.