Volume 294, number 1,2, 109–112 FEBS 10471 December 1991

Striking structural and functional similarities suggest that intestinal sucrase-isomaltase, human lysosomal α-glucosidase and *Schwanniomyces occidentalis* glucoamylase are derived from a common ancestral gene

Hassan Y. Naim[1], Thomas Niermann[2], Ulrich Kleinhans[1], Cornelis P. Hollenberg[1] and Alexander W.M. Strasser[3]

[1]*Institute of Microbiology, Heinrich-Heine-University of Düsseldorf, Universitätsstr. 1, D-4000 Düsseldorf 1, Germany,* [2]*Biocenter of the University of Basel, Klingelbergstr. 70, 4056 Basel, Switzerland* and [3]*Rhein Biotech GmbH, Erkratherstr. 230, D-4000 Düsseldorf 1, Germany*

Received 8 October 1991; revised version received 21 October 1991

Sequence comparison of the primary structure of the yeast *Schwanniomyces occidentalis* glucoamylase (GAM) with GAMs in different microorganisms did not reveal significant similarities. By contrast, striking similarities were, surprisingly, found with 3 mammalian secretory and integral membrane proteins: the 2 subunits of intestinal brush border sucrase-isomaltase and human lysosomal α-glucosidase. The similarities among these proteins are found as clusters of up to 8 amino acids and distributed all over the protein sequences. The major sequence differences are found in the N-terminal regions accounting, probably, for the different cellular locations of these proteins. The high level of similarities between sucrase, isomaltase, *Sch. occidentalis* GAM and human lysosomal α-glucosidase suggest that these proteins are derived from the same ancestral gene. To our knowledge, this is the first report that describes similarities between a yeast secretory protein and mammalian secretory and integral membrane proteins.

Glucoamylase; Sucrase-isomaltase; α-Glucosidase; Ancestral gene

1. INTRODUCTION

The existence of glucoamylases (GAMs) in microorganisms is usually demonstrated by the ability of a microorganism to grow on starch. A large number of yeast species have been characterized by virtue of their ability to grow on starch [1] and of many of these organisms, the gene encoding the GAM-hydrolytic activity has been cloned and characterized [2].

Recently Dohmen et al. [3] published the sequence of the GAM gene (*GAM1*) of the yeast *Schwanniomyces occidentalis*. This gene was isolated from a λ-Charon 4A library using oligodeoxynucleotides as probes. The probes corresponded to several tryptic fragments of a highly purified extracellular enzyme that exhibits GAM activity. In view of this cloning procedure and based on the finding that the recombinant gene revealed GAM activity, it was concluded that the cloned gene encodes an enzyme that can be identified as a GAM. Comparison of the amino acid sequence of this protein with the sequences of GAMs in various microorganisms revealed no significant similarities. This finding was unexpected since all other microbial GAM genes thus far cloned display significant similarities in their primary structure [2]. Since the cloned gene encodes the only known enzyme with a GAM-like function in *Sch. occidentalis*, it was suggested that this protein represents a novel type of a GAM that may have evolved independently.

The fact that *Sch. occidentalis* GAM is an α-glucosidase with specificity towards α-1,6 and α-1,4 linkages prompted us to compare its sequence with other well characterized α-glucosidases, like intestinal sucrase-isomaltase (for a review see [4]) and human lysosomal α-glucosidase [5].

Abbreviations: GAM, glucoamylase; pro-SI, pro-sucrase-isomaltase.

Correspondence address: H.Y. Naim, Institute of Microbiology, Heinrich-Heine-University of Düsseldorf, Universitätsstr. 1, D-4000 Düsseldorf 1, Germany. Fax: (49) (211) 311 53 70.

2. MATERIALS AND METHODS

Sequence comparison of *Sch. occidentalis* GAM with other microbial GAMs was performed via the MIPSX database [6] using the computer programme FASTA [7]. The alignment of sucrase/isomaltase with human lysosomal α-glucosidase as described by Hoefsloot et al. [5] was used to generate a profile with the program PROFILE [8]. *Sch. occidentalis* GAM was aligned to the set of the aligned sequences using PROFILEGAP [8]. Dot Plot analysis was performed using the computer programmes COMPARE and DOTPLOT [9]. The following protein sequences were compared: Rabbit sucrase-isomaltase [10], human isomaltase [11], human lysosomal α-glucosidase [5] and *Sch. occidentalis* GAM [3].

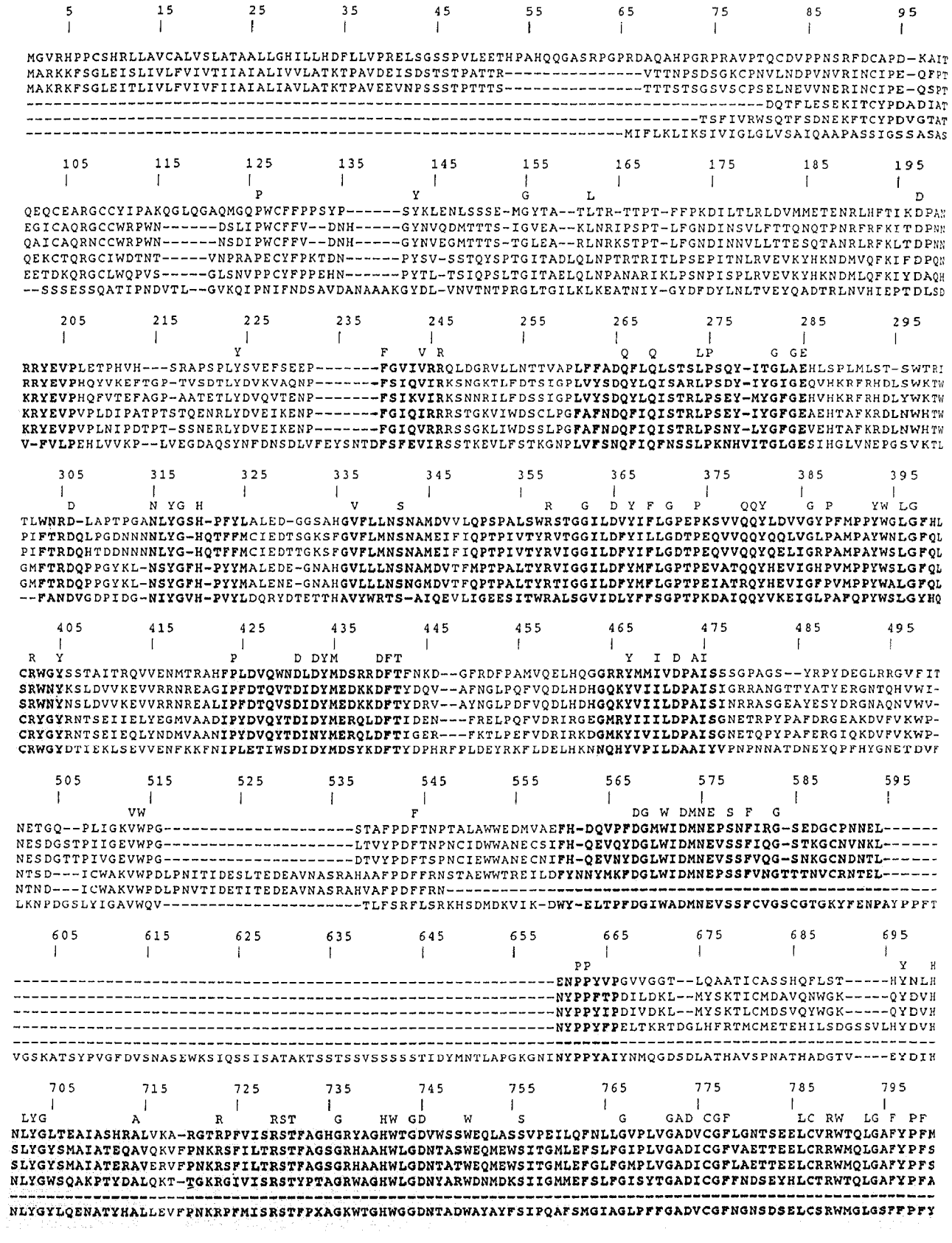

Fig. 1. (a) Amino acid alignment of human lysosomal α-glucosidase (A), human isomaltase (B), rabbit isomaltase (C), rabbit sucrase (D), rat sucrase (E) and *Schwanniomyces occidentalis* glucoamylase (F). Highly conserved domains are shaded. Amino acids which are identical in all sequences are indicated by upper case letters above the stack of aligned sequences.

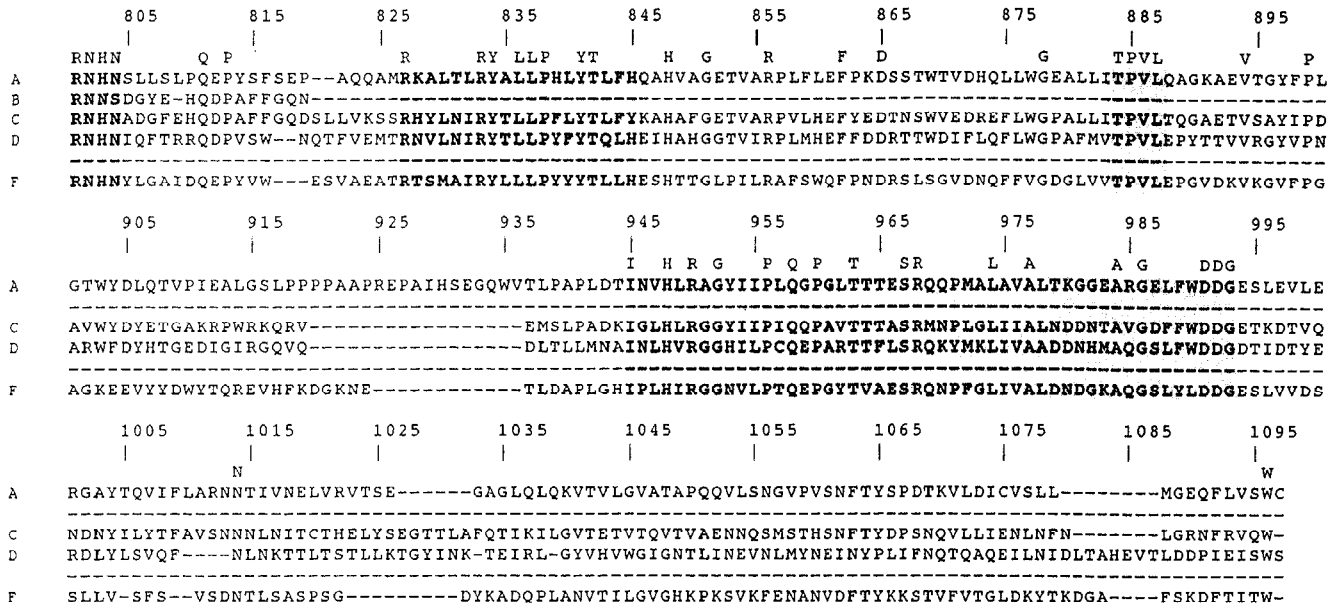

Fig. 1 (b).

3. RESULTS AND DISCUSSION

Sequence comparisons of *Sch. occidentalis* GAM with GAMs in various organisms using FASTA computer program [7] revealed that *Sch. occidentalis* GAM shares no similarities with other GAMs and suggested that this protein is encoded by a novel GAM gene. By virtue of similar hydrolytic functions we performed a further search for structural similarities with the mammalian proteins sucrase and isomaltase, the 2 subunits of the pro-sucrase-isomaltase complex (pro-SI), and human lysosomal α-glucosidase using the computer programs Profile and Profilegap [8]. All these enzymes are capable of hydrolyzing α-1,4 and α-1,6 glycosidic bonds although with different reaction kinetics and at different pH optima. Human lysosomal α-glucosidase and *Sch. occidentalis* GAM hydrolyze disaccharides at acidic pH values (between pH 4–5) [12–14] while sucrase and isomaltase function at neutral pH (for review see [4]).

The results of the computer search revealed that the GAM protein in *Sch. occidentalis* shares several common structural similarities with sucrase, isomaltase and human lysosomal α-glucosidase. Fig. 1 depicts the sequence comparisons of *Sch. occidentalis* GAM (F) with human lysosomal α-glucosidase (A), human isomaltase (B), rabbit isomaltase (C), rabbit sucrase (D) and rat sucrase (E).

The strongest sequence similarities were revealed between human lysosomal α-glucosidase (A), rabbit isomaltase (C), rabbit sucrase (D) and *Sch. occidentalis* GAM (F). In all these cases, almost 35% of amino acid identity was found. In addition, 60% of the non-identical amino acids represent conservative changes. The similarity appears as clusters of identical amino acids of up to 8 residues per cluster and is distributed almost all over these proteins. More than 50% amino acid identity is found in many stretches of 50 amino acid residues.

Particularly interesting is a highly conserved region that all these proteins possess between positions 568–583 of the alignment. The same region was suggested by Hunziker et al. [10] to be the putative active site of rabbit sucrase or isomaltase. In fact, within this region in the sequences of human and rabbit isomaltase on one hand (Fig. 1B and C, respectively) and rabbit sucrase on the other (Fig. 1D) there exists a stretch of 8 amino acids between positions 573–580 that reveal only one single amino acid substitution (V→P or vice versa). The sequence of the *Sch. occidentalis* GAM protein contains also a similar stretch (Fig. 1, F) that is completely identical to that found in human and rabbit isomaltase and could represent the putative active site of the enzyme. A similar stretch of 8 amino acids is also present in the sequence of human lysosomal α-glucosidase (Fig. 1, A), which reveals 2 amino acids differences in comparison to human and rabbit isomaltase and *Sch. occidentalis* GAM (N instead of S and P instead of V) (Fig. 1B,C and F, respectively) and 1 amino acid in comparison to rabbit sucrase (N instead of S) (Fig. 1D). It is important to note that isomaltase and *Sch. occidentalis* GAM are efficient in hydrolyzing α-1,6-linked disaccharides, while the ideal substrates for sucrase and human lysosomal α-glucosidase are α1,4-linked disaccharides. Together, these observations suggest that substituting V for P in the putative active site of these enzymes could be important in providing the corresponding enzyme with a conformation that has higher affinity to one type of disaccharide linkages than to the other.

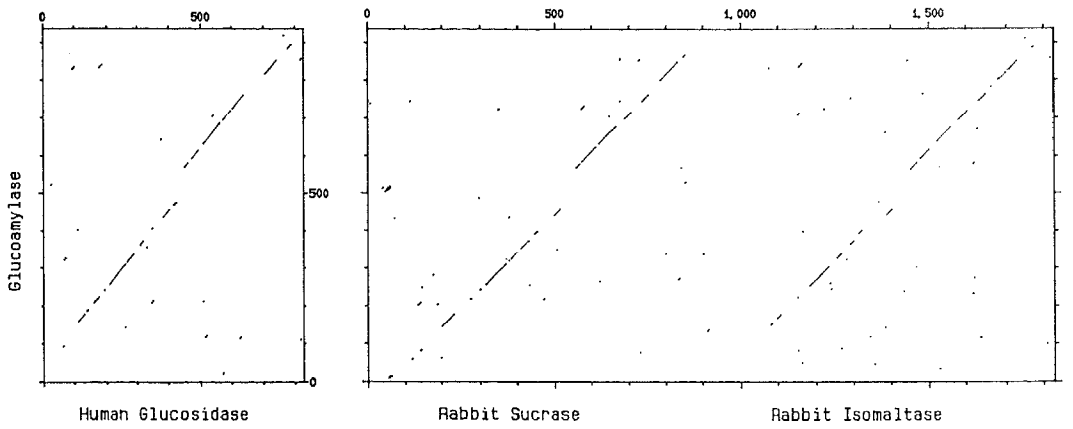

Fig. 2. Comparison of the amino acid sequence of *Schwanniomyces occidentalis* glucoamylase with human lysosomal α-glucosidase and rabbit sucrase-isomaltase using the computer program COMPARE (span: 15 amino acids, stringency: 11) and DOTPLOT for the graphic output (both programs were made by Genetics Computer Group, University of Wisconsin, Sequence Analysis Software Package, Version 6.2, 1990).

The plot in Fig. 2, made with DOTPLOT [9], demonstrates the high level of similarity between *Sch. occidentalis* GAM, human lysosomal α-glucosidase, rabbit isomaltase and rabbit sucrase. It should be noted that in this plot only the identical amino acids are considered.

The most significant sequence differences between *Sch. occidentalis* GAM, pro-SI and human lysosomal α-glucosidase are located in the amino terminal domains. These differences could account for the different biosynthetic pathways and cellular location of these enzymes. The sucrase–isomaltase complex is synthesized by small intestinal epithelial cells with a non-cleavable signal sequence that also contains the membrane anchoring domain [10]. By contrast, the N-terminus of human lysosomal α-glucosidase comprises a signal peptide that is cleaved off, generating a soluble glycoprotein whose final destination is an intracellular organelle, the lysosomes [5]. The *Sch. occidentalis* GAM molecule seems to be synthesized with a cleavable signal sequence in a fashion similar to human lysosomal α-glucosidase.

Elucidation of the primary structure of rabbit sucrase and isomaltase by cDNA cloning has revealed a high degree of similarity between the sucrase and the isomaltase subunits (41% amino acid identity) and subsequently provided a strong support to the hypothesis that sucrase/isomaltase was generated by duplication of the same gene [4,10]. Moreover, in view of the strong similarities between sucrase, isomaltase and human lysosomal α-glucosidase it was more recently [5] suggested that these molecules could have a common ancestral gene. Likewise, by virtue of the striking similarities of the *Sch. occidentalis* GAM protein with sucrase, isomaltase and human lysosomal α-glucosidase it is reasonable to propose that the *Sch. occidentalis* GAM gene has also evolved from the same ancestral gene.

The evolution of these genes has probably occurred prior to the duplication of the isomaltase gene because the human lysosomal α-glucosidase and *Sch. occidentalis* GAM genes did not undergo gene duplication and persisted as a single gene. Later, each molecule should have acquired its unique structural features (like distinct N-termini) as a consequence of adaptation to its own function.

Finally, it is interesting to note that the sucrase of *Sch. occidentalis* (encoded by the invertase gene) has no similarities with its GAM or the mammalian sucrase, indicating that the *Sch. occidentalis* GAM probably did not evolve by gene duplication of the sucrase gene.

Acknowledgements: This work was supported by grants from the Bundesministerium für Forschung und Technologie (BMFT), Bonn, Germany (to H.Y.N. and C.P.H.)

REFERENCES

[1] Lodder, J. and Krieger-van Rij, N.J.W. (1970) The Yeasts, A Taxonomic Study, Elsevier, Amsterdam.
[2] Itoh, T., Ohtsuki, I., Yamashita, I. and Fukui, S. (1987) J. Bacteriol. 169, 4171–4176.
[3] Dohmen, R.J., Strasser, A.W.M., Dahlems, U.M. and Hollenberg, C.P. (1990) Gene 95, 111–121.
[4] Semenza, G. (1986) Annu. Rev. Cell Biol. 2, 255–313.
[5] Hoefsloot, L.H., Hoogeveen. Westerveld, M., Kroos, M.A., van Beeumen, J., Reuser, A.J.J. and Oostra, B.A. (1988) EMBO J. 7, 1697–1704.
[6] Mewes, H.W. (1990) MIPS, Max-Planck Institute for Protein Sequences.
[7] Pearson, W.R. and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444–2448.
[8] Gribskov, M., McLachlan, A.D. and Eisenberg, D. (1987) Proc. Natl. Acad. Sci. USA 84, 4355–4358.
[9] Genetics Computer Group, University of Wisconsin, Sequence Analysis Software Package, Version 6.2 (1990).
[10] Hunziker, W., Spiess, M., Semenza, G. and Lodish, H. (1986) Cell 46, 227–234.
[11] Green, F., Edwards, Y., Hauri, H.-P., Povey, S., Ho, M.W., Pinto, M. and Swallow, D. (1987) Gene 57, 101–110.
[12] Jeffrey, P.L., Brown, D.H. and Brown, B. (1970) Biochemistry 9, 1403–1415.
[13] Palmer, T.N. (1971) Biochem. J. 124, 701–711.
[14] Koster, J.F. and Slee, R.G. (1977) Biochim. Biophys. Acta 482, 89–97.