Proc. KMB International Symposium
June 24-26, 2003, Muju, Korea

Functional Genomics and Its Impact on the Development of a *Hansenula polymorpha* RB11-Based Expression Platform

Massoud Ramezani-Rad[1], Laura Guengerich[2], Frank Müller II[3], and Gerd Gellissen[3]

[1] Institute for Microbiology, Heinrich-Heine University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany

[2] Zentrum für Präventive Pädiatrie, Johannes-Gutenberg-Universität Mainz, Langenbeckstr. 1, 65301 Mainz

[3] Rhein Biotech GmbH, Eichsfelder Strasse 11, D-40595 Düsseldorf, Germany

Hansenula polymorpha as an Expression Platform

Yeasts constitute an important group of industrial microorganisms with applications as platforms for heterologous gene expression. As such the baker's yeast *Saccharomyces cerevisiae* served as a eukaryotic model organism for basic research and industrial applications [1]. In 1996, it was the first eukaryotic organism for which the complete genome sequence was established [2]. Initial focus on *S. cerevisiae* has been extended to a range of alternative yeast species. As a consequence, the number of fully or partially sequenced budding yeast genomes has continued to grow. Among others, a comparative genomic exploration of 13 hemiascomycetous yeast species was conducted [3]. The methylotrophic yeast *Hansenula polymorpha* (syn. *Pichia angusta*) is a prominent example of the industrially applied non-conventional yeasts [4, 5]. For heterologous gene expression a range of vectors is available (see Fig.1).

The Genome *of Hansenula polymorpha* RB11

Although much was known about the physiology, biochemistry and ultra structure of this yeast (for review see monograph on *H. polymorpha* [4]), little information was available about the genomic structure and function until now [23]. Several groups world wide initiated studies on its genome several years ago. Included in the comparative genome analysis on 13 hemiascomycetous yeasts mentioned above part of the *H. polymorpha (P. angusta)* genome sequence was established using a partial random sequencing strategy with a coverage of 0.3 genome equivalents. Using this approach, about 3 Mbp of sequencing raw data of the *H. polymorpha* genome was yielded [3]. A recently terminated genome analysis aimed at a higher coverage using a BAC-to-BAC approach and resulting in the comprehensive genome analysis of this organism. A first description of the data generated and a detailed description of the employed materials and methods is provided in a recent publication [24]. For sequencing of *H. polymorpha* strain RB11 a BAC library with approximately 17x coverage was constructed in a pBACe3.6 vector according to Osoegawa et al. [25, 26]. Base calling and quality checks were carried out using Phred [27]. Sequences were assembled with Phrap and editing was performed after import into gap4. BAC assemblies and raw data were visualized and edited using the STADEN package (version 4.5; http://www.mrc-lmb.cam.ac.uk/pubseq/staden_home.html developed by Roger Staden et al.). Fully automated annotation was carried out using the ConSequence^{TM} software system provided by QIAGEN (based on Pedant-Pro^{TM} from Biomax Informatics AG) [28].Sequencing was performed by Qiagen, Hilden, the manual annotation by BIOMAX Munich.

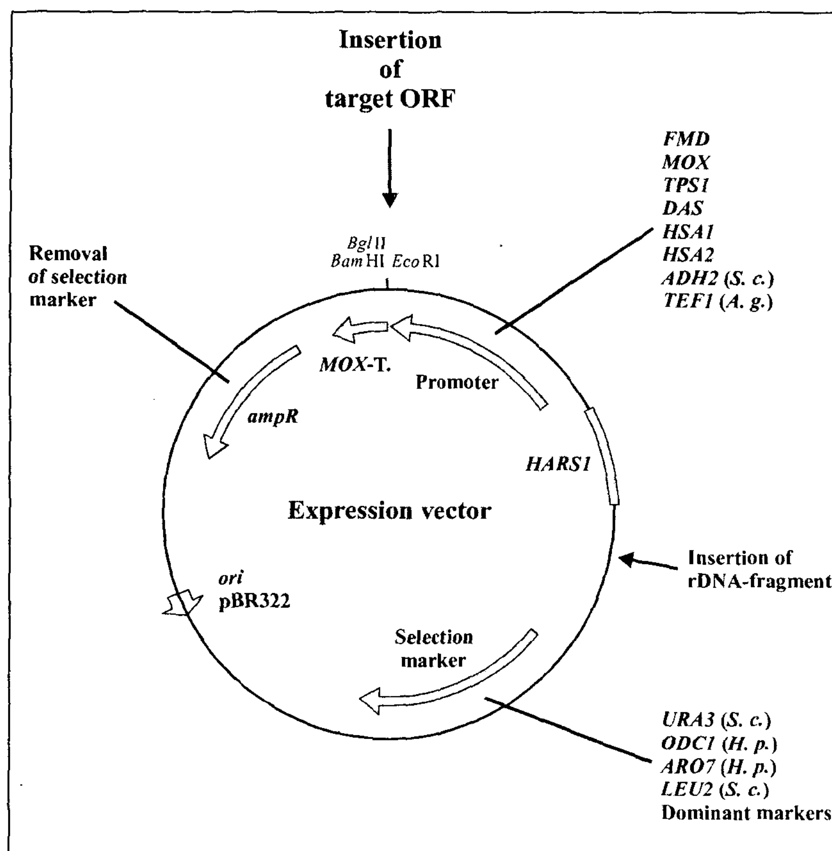

Fig. 1. General design of an *H. polymorpha* integration/expression vector. Abbreviations for taxonomic names are A.g. (*Ashbya gossypii*), H.p. (*Hansenula polymorpha*) and S.c. (*Saccharomyces cerevisiae*) (modified after Suckow and Gellissen [12]).

H. polymorpha is an ubiquitous yeast species occurring naturally in spoiled orange juice, maize meal, in the gut of various insect species and in soil. It grows as white to cream-coloured, butyrous colonies and does not form filaments [6]. *H. polymorpha* isolates are homothallic and reproduction occurs vegetatively by budding. *H. polymorpha* belongs to the fungal family of Saccharomycetaceae, subfamily Saccharomycetoideae [6, 7]. Most research has been performed with three basic strains designated as *H. polymorpha* DL-1, CBS4732 and NCYC495. These strains are of independent origin, have unclear relationship and exhibit different features including different chromosome numbers. Depending on strain and separation conditions between two and seven chromosomes can be distinguished [8, 9]. Strain CBS4732 (syn ATCC34438, NRRL-Y-5445; CCY38-22-2) was originally isolated from soil irrigated with waste water from a distillery in Pernambuco, Brazil [10]. Its *odc1*-derivatives LR9 [11] and RB11 [12] have been developed as hosts for heterologous gene expression [12]. Recombinant compounds produced in these hosts include enzymes like the feed additive phytase [13, 14], anticoagulants like hirudin and saratin [15-17] and an efficient vaccine against hepatitis B infection [18-20]. The significance of *H. polymorpha* in basic research stems largely from studies focused on peroxisome homeostasis [21] and nitrate assimilation [22].

Sequencing resulted in the characterisation of 8.733 million base pairs assembled into 48 contigs. The derived sequence covers over 90 % of the estimated total genome content of 9.5 Mbp located on 6 chromosomes which range in size between 0.9 and 2.2 Mbp [23]. The Pedant-Pro™ Sequence Analysis Suite was used for gene identification. Out of the sequenced 8.73 Mb 5848 ORFs have been extracted for proteins longer than 80 amino acids. 389 ORFs smaller than 100 aa were identified. 4771 ORFs have homologues to known proteins (81.6 %). Calculation of the gene density and protein length, taking into account the gene

numbers, showed an average gene density of 1 gene/1.5 kb and an average protein length of 440 amino acids. 91 introns have been identified by homology to known proteins and confirmed by using GeneWise [29]. 80 tRNAs were identified, corresponding to all 20 amino acids. From approximately 50 rRNA clusters [5], 7 clusters have been fully sequenced. All clusters are completely identical and have a precise length of 5033 bp.

The main functional categories and their distribution in the gene set are manually predicted for energy, 4 %; cellular communication, signal transduction mechanism, 3 %; protein synthesis, 6 %; cell rescue, defense and virulence, 4 %; cellular transport and transport mechanisms, 9 %; cell cycle and DNA processing 9 %; protein fate (folding, modification, destination) 17 %; transcription, 13 % and metabolism, 19 %. A selection of the data obtained from the annotated sequence is provided in the following Tables 1 and 2.

Table 1. ***Hansenula polymorpha*** **genome statistics**

Contigs: 48
Total length of contigs: 8,733,442 bp
Average contig length: 182 kb

No. of extracted ORFs:	**5,848**
No. of ORFs < 100 aa:	**389**
Average gene density:	**1 gene / 1.5 kb**
Average gene size (start-stop):	**1.3 kb (1.320 nt)**
Average protein length:	**440 aa**

Table 2. Functional categorization of genes

Functional category	**No. ORFs**	**%**
Metabolism	1114	19
Energy	231	4
Cell growth, Cell division and DNA synthesis	518	9
Transcription	767	13
Protein synthesis	323	6
Protein destination	1014	17
Transport facilitation	423	7
Cellular transport and transport mechanisms	518	9
Control of cellular organization	417	7
Cellular communication / signal transduction	170	3
Cell rescue, defense, and virulence	260	4
Cell fate	282	5
Regulation of/ interaction with cellular environment	184	3

Limitations of the RB11-based expression platform

Despite the most favorable characteristics of the RB11-based platform for application in heterologous gene expression, problems and limitations can be encountered in particular strain and product developments as is similarly and more frequently observed in other yeast systems. These limitations include overglycosylation [12,30], retention within the ER [31], poor secretion, impaired processing [32] and proteolytic degradation [12]. A possible strategy to overcome these limitations is to identify genes and gene

products that may upon disruption or co-expression positively influence the performance of respective strains. This has succesfully been applied in several cases. Among others co-expression of the *S. cerevisiae*-derived *KEX2* gene provided a greatly improved processing of a IFNα-2a pre-pro-sequence in *H. polymorpha* in which production of predominantly N-terminally extended molecules had been observed before [32]. In another example co-expression of a *S. cerevisiae*-derived *CMK2* gene led to an improved secretion and a reduction in overglycosylation of a secreted enzyme (see following Table 3).

Table 3. Improvement of the expression performance of *H. polymorpha* production strains by the co-expression of other genes

GENE EXPRESSED	PROBLEM ENCOUNTERED	COEXPRESSED GENE
IFNα-2a	incorrect pre-pro-cleavage	*KEX2*
Enzyme	Overglycosylation impaired secretion	*CMK2*

The impact of functional genomics on the development of the *H. polymorpha* RB11-based expression platform

Several approaches have been initiated to identify *H. polymorpha* genes that may positively influence the perfomance of particular production strains. Examples include the identification of the *PMR1* gene [33] and glycosylation genes [34]. With the completion of genome sequencing, transcriptome, proteome analyses and other related technologies are feasible enabling a more systematic approach. A summary of current research programs is provided in the following figure and the subsequent section of the article.

In a first approach genes will be identified involved in methanol metabolism, peroxisome homeostasis, protein glycosylation, secretion and cell wall integrity. These tasks are executed within a cooperative effort with partners in Russia, Ukraine, The Netherlands and Germany funded by INTAS (INTAS 2001-0583). For identification of such genes, linear DNA fragments harboring reporter genes are used for random integration thereby generating mutants. By this random integration (RALF) approach certain genes of potential impact for relevant gene expression functions may be disrupted and identified by sequencing the region adjoining the integration site and comparing the deduced sequence with the genome data. Applying a selection of suitable reporter proteins and a range of certain growth conditions the generated strains can be screened for genes having an impact on the functions mentioned above.

For transcriptome analysis a DNA chip is being generated in cooperation with the KRIBB institute in Taejon (Korea). At the time of submission of this article the generation of such a chip was approaching completion comprehensively covering PCR-amplified ORFs of the genome. This tool will enable an in-depth analysis of the transcriptome in correlation to defined physiological conditions.

A third project to be started soon is a comprehensive analysis of the proteome of recombinant *H. polymorpha* production strains in correlation to specific products, secretion efficiency and other characteristics. Extraction of defined proteins from two-dimensional SDS gels and MS analysis of proteolytic fragments will lead to the identification of proteins and their respective genes with a potential impact on the performance of the *H. polymorpha* expression platform.

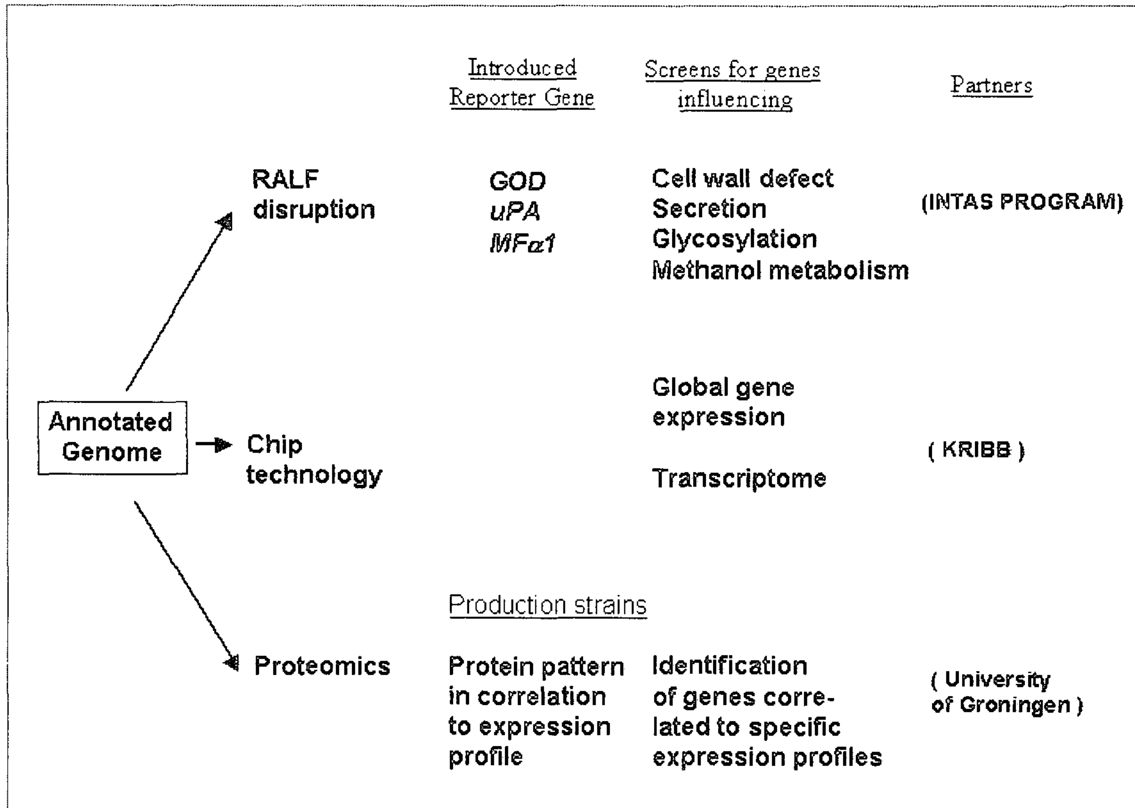

Fig. 2. Approaches to identify genes of a potential impact on the expression of foreign genes in *H. polymorpha.*

Acknowledgments

We thank Drs J. Lauber, H. Wedler, E. Griess from Qiagen and Drs C. Wagner, K. Albermann and J. Hani from BIOMAX for performing sequencing and annotation of the genome. The work is included in the project INTAS 2001-05833 supported by INTAS, Brussels. The project is executed in cooperation with Drs. Agaphonov and Kalebina and their coworkers in Moskau, Dr. Andrei Sibirny and his coworkers in Lviv, Ukraine, Drs. M. Veenhuis and I. v.d.Klei and their coworkers in Groningen (The Netherlands) and Dr. C.P. Hollenberg and his coworkers in Düsseldorf (Germany). The Chip project is executed at the KRIBB under supervision of Dr. H.A. Kang.

References

1. Joseph, R. (1999) Yeasts: production and commercial uses. In: Encyclopedia of food microbiology, Vol 3 (Robinson, R. K., Batt, C. A. and Patel, P.D., Eds.), pp.2335-2341, Academic Press, San Diego, CA.
2. Goffeau, A, Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H. and Oliver, S.G. (1996) Life with 6000 genes. Science 274, 563-567.
3. Feldmann, H. (Ed.) (2000) Génolevures. Genomic exploration of the hemiascomycetous yeasts. FEBS Lett. 487, 1-150.
4. Gellissen, G. (2000) Heterologous protein production in methylotrophic yeasts. Appl. Microbiol. Biotechnol. 54, 741-750.

5. Gellissen, G. (Ed.) (2002) *Hansenula polymorpha* - biology and applications. Wiley-VCH, Weinheim.
6. Barnett, J.A., Payne, R.W. and Yarrow, D. (2000) Yeasts: characteristics and idendification, 3rd edn. Cambridge University Press, Cambridge, UK.
7. Middelhoven, W.J. (2002) History, habitat, variability, nomenclature and phylogenetic position of *Hansenula polymorpha.* In: *Hansenula polymorpha* - biology and applications (Gellissen, G., Ed.), pp. 1-7. Wiley-VCH, Weinheim.
8. Marri, L., Rossolini, G.M. and Satta, G. (1993) Chromosome polymorphism among strains of *Hansenula polymorpha.* Appl. Environ. Microbiol. 59, 939-941.
9. Lahtchev, K. (2002) Basic genetics of *Hansenula polymorpha* . In: *Hansenula polymorpha* - biology and applications (Gellissen, G., Ed.), pp. 8-20. Wiley-VCH, Weinheim.
10. Morais, J.O.F. and Maia, M.H.D. (1959) Estudos de microorganismos enconcentrados em leitos de despéjos de caldas de destilarias de Pernambuco. II. Uma nova espécie de *Hansenula, H. polymorpha.* Anais de Escola Superior de Qimica, Universidade do Recife 1,15-20.
11. Roggenkamp, R., Hansen, H., Eckart, M., Janowicz, Z. and Hollenberg, C.P. (1986) Transformation of the methylotrophic yeast *Hansenula polymorpha* by autonomous replication and integration vectors. Mol. Gen. Genet. 202, 302-308.
12. Suckow, M. and Gellissen, G. (2002) The expression platform based on *H. polymorpha* strain RB11 and its derivatives - history, status and perspectives. In: *Hansenula polymorpha* - biology and applications (Gellissen, G., Ed.), pp. 105-123. Wiley-VCH, Weinheim.
13. Mayer, A.F., Hellmuth, K., Schlieker, H., Lopez-Ulibarri, R., Oertel, S., Dahlems, U., Strasser, A.W.M. and van Loon, A.P.G.M. (1999) An expression system matures: a highly efficient and cost-effective process for phytase production by recombinant strains of *Hansenula polymorpha.* Biotechnol. Bioeng. 63, 373-381.
14. Papendieck, A., Dahlems, U. and Gellissen, G. (2002) Technical enzyme production and whole-cell biocatalysis: application of *Hansenula polymorpha.* In: *Hansenula polymorpha* - biology and applications (Gellissen, G., Ed.), pp. 255-271. Wiley-VCH, Weinheim.
15. Avgerinos, G.C., Turner, B.G., Gorelick, K.J., Papendieck, A., Weydemann, U. and Gellissen, G. (2001) Production and clinical development of a *Hansenula polymorpha*-derived PEGylated hirudin. Sem. Thromb. Hemostas. 27, 357-371.
16. Barnes, C.S., Krafft, B., Frech, M., Hofmann, U.R., Papendieck, A., Dahlems, U., Gellissen, G. and Hoylaerts, M.F. (2001) Production and charcaterization of saratin, an inhibitor of von Willebrand's factor-dependent platelet adhesion to collagen. Sem. Thromb. Hemostas. 27, 337-347.
17. Bartelsen, O., Barnes, C.S. and Gellissen, G. (2002) Production of anticoagulants in *Hansenula polymorpha.* In: *Hansenula polymorpha* - biology and applications (Gellissen, G., Ed.), pp. 211-228. Wiley-VCH, Weinheim.
18. Janowicz, Z.A., Melber, K., Merckelbach, A., Jacobs, E., Harford, N., Comberbach, M. and Hollenberg, C.P. (1991) Simultaneous expression of the S and L surface antigens of hepatitis B and formation of mixed particles in the methylotrophic yeast, *Hansenula polymorpha.* Yeast 7, 431-433.
19. Schaefer, S., Piontek, M., Ahn, S.-J., Papendieck, A., Janowicz, Z.A. and Gellissen, G. (2001) Recombinant hepatitis B vaccines - characterization of the viral disease and vaccine production in the methylotrophic yeast, *Hansenula polymorpha.* In: Novel therapeutic proteins - selected case studies. (Dembowsky, K. and Stadler, P., Eds.), pp. 245-274. Wiley-VCH, Weinheim.
20. Schaefer, S., Piontek, M., Ahn, S.-J., Papendieck, A., Janowicz, Z.A., Timmermans, I. and Gellissen G. (2002) Recombinant hepatitis B vaccines - disease characterization and vaccine production. In: *Hansenula polymorpha* - biology and applications. (Gellissen, G.. Ed.), pp. 175-210. Wiley-VCH, Weinheim.
21. Van der Klei, I.J. and Veenhuis, M. (2002) *Hansenula polymorpha*: a versatile model organism in peroxisome research. In: *Hansenula polymorpha* - biology and applications (Gellissen, G., Ed.), pp. 76-94. Wiley-VCH, Weinheim.
22. Siverio, J.M. (2002) Biochemistry and genetics of nitrate assimilation. In: *Hansenula polymorpha* - biology and applications. (Gellissen, G.. Ed.), pp. 21-40. Wiley-VCH, Weinheim.

23. Waschk, D., Klabunde, J., Suckow, M. and Hollenberg, C.P. (2002) Characteristics of the *Hansenula polymorpha* genome. In: *Hansenula polymorpha* - biology and applications (Gellissen, G., Ed.), pp. 95-104. Wiley-VCH, Weinheim.
24. Ramezani-Rad, M., Hollenberg, C.P., Lauber, J., Wedler, H., Griess, E., Wagner, C., Albermann, K., Hani, J., Piontek, M., Dahlems, U. and Gellissen, G (2003) The *Hansenula polymorpha* (strain CBS4732) genome sequencing and analysis. FEMS Yeast Res. (in press).
25. Osoegawa, K., de Jong, P.J., Frengen, E., Ioannou, P. A. (1999) Construction of bacterial artificial chromosome (BAC/PAC) libraries. Current Protocols in Human Genetics. 5.15.1 – 5.15.33.
26. Osoegawa, K, Woon, P.Y., Zhao, B, Frengen, E., Tateno, M., Catanese J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics. 52, 1-8.
27. Ewing, B., Hillier, L., Wendl, M.C. and Green, P. (1998) Base-calling of automated sequencer traces using phred. Genome Res. 8, 175-194.
28. Frishman, D., Albermann, K., Hani, J., Heumann, K., Metanomski, A., Zollner, A. and Mewes, H.W. (2001) Functional and structural genomics using PEDANT. Bioinformatics 17, 44-57.
29. Birney, E and Durbin, R. (2000) Using GeneWise in the *Drosophila* annotation experiment. Genome Res. 10, 547-548.
30. Agaphonov, M.O., Romanos, N.V., Trushina, P.M., Smirnov, V.N. and Ter-Avanesyan, M.D. (2002) Aggregation and retention of human urokinase type plasminogen activator in the yeast endoplasmic reticulum. BMC Mol. Biol. 3, 15-22.
31. Agaphonov, M.O., Deev, A.V., Kim, S.-Y., Sohn, J.-H., Choi, E.-S. and Ter-Avanesyan, M.D. (2003) A novel approach to isolation and functional chracterization of genomic DNA sequences from the methylotrophic yeast *Hansenula polymorpha*. Mol. Biol. 37, 74-80.
32. Müller, F. II, Tieke, A., Waschk, D., Mühle, C., Müller, F I, Seigelchifer, M., Pesce, A., Jenzelewski, V. and Gellissen, G. (2002) Production of IFNα-2a in *Hansenula polymorpha*. Process Biochem. 38, 15-25.
33. Kang, H.A., Kim, J.-Y., Ko, S.-M., Park, C.S., Ryu, D.Y., Sohn, J.-H., Choi, E.-S., Rhee, S.-K. (1998) Cloning and characterization of the *Hansenula polymorpha* homologue of the *Saccharomyces cerevisiae PMR1* gene. Yeast 14, 1233-1240.
34. Guengerich, L., Kang, H.A., Behle, B., Gellissen, G. and Suckow, M (2003) A platform for heterologous gene expression based on the methylotrophic yeast *Hansenula polymorpha*. In: Kück, U. (Ed.), Mycota II, Genetics and Biotechnology, 2nd edition. Springer Berlin (in press).