Fig. 1:

## Amino acid sequence, FLO8 Protein *Komagataella phaffii*, SEQ ID NO:1

>XP_002494151/PP7435_Chr4-0252/PAS_chr4_0711 (Flo8_Komagataella phaffii CBS7435/GS115) LisH domain underlined

MNKPNGSEQQPPSRGMKQESGGPVTSSTTPGTNTGLENSHSMGADMEPDVGATSPRHLLNGYIYDYLVKSNMQNLA
DQFAQETELLETDLTVPMDTPSGYLLEWWMVFWDLFNARLKQRGSQKAHQYIQLNMLRQQQQRTMRNTARVQKVPL
RPHTQSSPSMSQTFIPQQPQQQAQGQQHAQAQAQVQAHQQAQHHAQAQVPVQPQQHQLGGQTQQQQSINTGSPAGP
NAINSRVQHLAQQQMNHLRQQATATTQQPIPQQNIPSNQQGPTGPYPTSPSRRPRLLSNESGASAPSVMTKSQLQG
VPPSQQPHQQQGQQVGPPNQHQGQSSSFYSGMPPQGVVVPHQFNPQQYANMLARQQHVQAQQQVQLQQVQHVQQRQ
QQDQQQHRLSAGSPGHPSFGVFQQPPPMSNHNQVMINQQGETFFDPHSPYAQPNGYPQPQQQQQQQQQQQQQQPQ
QQQQQQQQQKQQPPPPPRQPQRQQAMAMAPLPHSTSAAGTPHSSTTPRFSQPGPVYQQPLPASQPQHSPPSSIQQP
ELVPTPGSQHQQIAQPQSQSQHQQSQQSQSSASKIVGIQEYQKELMMLEKQNKQRHDMACKKGSGHFSNFDPIPEH
TPPEPKFNVNVMLPPQNSAVVTKNTPGTSPGTQTQNTAHSTGNTSAGSTPNNVAPVRKKKEPAKKKAKKATEPPTP
TTPQTPIAARTHQNSTGGIPGNNAATKRRKREPLVDQTVSPNLNEASKSTKTGKISSQTDFTGSDNGFLQDFGDGT
GPPTGTDDMEFDFNSFLNNETGEPNSSTIHFDNVFNWGEGTEAGDL

## Nucleotide sequence, *flo8* gene *Komagataella phaffii*, SEQ ID NO:2

>XM_002494106.1 Komagataella phaffii GS115 Hypothetical protein (PAS_chr4_0711), partial mRNA

ATGAACAAGCCAAACGGGTCTGAACAACAACCACCGTCACGCGGAATGAAGCAAGAGTCAGGAGGCCCAG
TTACTTCATCTACGACGCCGGGTACCAATACTGGCCTAGAAAACTCTCATTCCATGGGGGCGGATATGGA
GCCTGATGTTGGTGCTACCTCTCCTCGCCATCTTCTTAATGGGTACATTTACGATTATTTAGTCAAATCT
AACATGCAAAATTTGGCTGATCAATTTGCCCAAGAGACGGAGCTCTTAGAAACAGACTTGACAGTACCAA
TGGATACGCCTTCAGGCTATCTTCTAGAATGGTGGATGGTATTCTGGGACCTTTTCAATGCCCGCCTAAA
GCAACGGGGTTCACAGAAGGCCCACCAGTATATTCAGTTGAACATGCTACGACAACAGCAACAGAGGACC
ATGCGAAATACAGCCCGTGTTCAAAAAGTCCCGTTGAGGCCACACACCCAATCATCTCCTTCAATGTCAC
AGACTTTTATTCCACAGCAGCCTCAACAGCAAGCACAGGGACAACAGCACGCCCAGGCTCAAGCCCAAGT
GCAAGCACATCAGCAAGCCCAACACCACGCGCAGGCACAAGTGCCAGTGCAACCGCAACAGCACCAGCTA
GGAGGCCAAACTCAACAGCAGCAATCCATTAACACTGGGTCTCCTGCGGGTCCAAATGCTATCAACTCGC
GTGTTCAACACTTAGCACAACAACAGATGAATCACCTTCGCCAGCAGGCGACTGCCACTACGCAACAACC
TATCCCGCAACAGAATATCCCATCAAACCAACAGGGTCCTACAGGCCCTTATCCTACTTCCCCTTCAAGA
AGACCGAGATTACTGTCTAACGAATCGGGTGCAAGTGCACCCTCTGTAATGACAAAGTCACAGCTCCAAG
GAGTCCCTCCCTCACAACAACCACACCAGCAGCAAGGTCAGCAGGTAGGCCCCCCTAATCAACATCAAGG
TCAATCTTCTTCCTTTTATTCGGGCATGCCTCCTCAAGGGGTCGTGGTTCCTCATCAGTTCAATCCTCAG
CAGTATGCCAATATGCTAGCAAGACAACAGCATGTACAAGCTCAACAACAGGTTCAGTTACAACAGGTCC
AACATGTACAACAGAGACAACAGCAAGACCAACAACAACACCGCCTGTCCGCCGGTTCACCGGGGCACCC
TTCATTTGGCGTTTTTCAACAACCTCCTCCGATGTCAAACCATAATCAGGTCATGATCAATCAGCAGGGA
GAAACTTTTTTTGATCCACATTCTCCATATGCTCAACCTAACGGGTACCCCCAGCCACAGCAACAACAAC
AACAACAGCAACAACAACAACAACAGCAGCAACCGCAACAGCAGCAGCAGCAGCAGCAGCAACAGAAGCA
GCAACCACCACCACCACCACGACAGCCTCAGCGCCAACAAGCGATGGCCATGGCTCCTCTGCCTCACTCT
ACTTCTGCCGCCGGTACTCCTCACTCGTCCACCACACCTAGATTCTCGCAACCTGGTCCTGTTTATCAGC
AGCCTTTACCTGCATCTCAACCGCAACATTCTCCGCCTTCTTCTATTCAGCAGCCGGAGCTAGTTCCAAC
TCCAGGGTCACAACATCAGCAAATAGCACAACCACAATCACAGAGCCAACACCAGCAATCGCAACAGTCT
CAATCAAGTGCTTCTAAAATTGTAGGTATACAGGAGTATCAGAAAGAGCTAATGATGCTTGAGAAACAGA
ACAAACAGCGTCATGACATGGCATGTAAGAAGGGAAGCGGGCATTTTTCTAACTTTGATCCAATTCCAGA
GCACACACCGCCCGAACCAAAATTTAATGTGAATGTAATGCTCCCTCCCCAGAACTCTGCAGTGGTCACG
AAGAATACTCCCGGAACTTCACCTGGTACACAAACTCAAAACACTGCACATAGTACTGGTAACACTTCTG
CGGGGTCTACACCAAATAATGTCGCACCTGTACGAAAGAAAAAGGAGCCAGCTAAAAAGAAGGCAAAGAA
AGCTACTGAGCCCCCGACTCCCACTACTCCACAGACTCCAATTGCAGCTAGGACACATCAAAACTCTACA
GGCGGCATTCCTGGTAATAATGCTGCTACTAAGCGACGAAAACGGGAGCCGCTGGTTGATCAAACTGTTT
CACCTAACCTTAACGAAGCTTCCAAGTCAACAAAGACCGGAAAAATTTCATCTCAAACTGACTTTACAGG
TTCTGACAATGGATTCTTACAGGATTTTGGCGATGGAACTGGTCCTCCCACTGGAACCGATGATATGGAA
TTTGATTTTAACAGTTTTCTTAATAACGAAACTGGCGAACCTAATAGTTCAACCATTCATTTTGACAATG
TATTCAATTGGGGAGAAGGTACCGAAGCCGGAGATTTATAG

Fig. 1 (continued)

Amino acid sequence, FLO8 Protein *Komagataella pastoris*, SEQ ID NO:3

>ANZ77695.1 BA75_04606T0 [Komagataella pastoris] LisH domain underlined

MNKPNGSGQQPPSRGMKQDPGGPVTSSTTPGTNTGFENSHSMGADVEPDVGAASPRHILNGYIHDYLVKSNMQNLA
DQFAQESDLLETDLTVPMDTPTGYLLEWWMVFWDLFNARLKQRGSQKAHQYIQLNMLRQQQQRTMRNTARVQKVPL
RPHTQSSPSMSQTFIPQQPQQQAQAQAQQHAQAQAQVQAHQQAQHHAQAQVPMQSQPHQQGGQTQQQQPINTGSPA
GPNAINSRVQHLAQQQMNHLRQQATATTQQPIPQQNIPSNQQGPAGPYPTSPSRRPRLLSNESGASAPSVMTKSQL
QGGPPSQQPHQQQAQQVGPPNQHQGQSSSFYSGMPPQGVVVPHQFNPQQYANMLARQQHVQAQQQVQLQQVPHVQQ
RQQQDQQQHRLSAGSPGHPSFGVFQQPPPMSNHNQVMINQQGETFFDPHSPYAQPNGYPQPQQQQQQQQQQQQQQ
QQQQQQQQKQQQPPPRQPQRQQAMATAPLPHSTSASGTPHTATTPRFSQPGPVYQQPLPASQPQHSPPTSIQQQEP
IPTPGSQHQQIAQPQSQNQHQQPQQPQASASKMVGIQEYQKELMMLEKQNKQRHDMAIKKGSGHFSNFDPIPEHTQ
TEPKFNVNVMLPPQNSAVATKNTPGTSPGTQTQNTAHSTGNTSAGSTPNNVVPVRKKKEPSKKKSKKATEPPTPTT
PQTPIAARAHQNSTGGISGNNAATKRRKREPLVDQTVSPNLNEASKSTKPGKISSQNDFTGSDNGFLQDFGDGTGP
PTGTDDMEFDFNSFLNNETGEPNSSTIHFDNVFNWGEGTEAGDL

Nucleotide sequence, FLO8 Protein *Komagataella pastoris*, SEQ ID NO :4

>CP014587.1:1153190-1155604 Komagataella pastoris strain ATCC 28485 chromosome 4 sequence

ATGAACAAGCCAAACGGGTCTGGACAACAACCACCGTCACGCGGAATGAAGCAAGACCCAGGAGGCCCAG
TTACCTCATCTACGACACCCGGTACCAATACTGGCTTTGAAAACTCTCATTCCATGGGAGCGGATGTGGA
GCCTGATGTTGGTGCCGCCTCTCCTCGCCATATTCTTAATGGGTACATTCATGATTACTTAGTCAAATCT
AATATGCAAAATTTGGCCGATCAATTTGCTCAAGAGTCGGATCTCTTAGAAACAGACCTAACGGTACCAA
TGGATACACCTACAGGCTATCTTTTAGAGTGGTGGATGGTATTTTGGGACCTTTTCAATGCCCGCCTAAA
GCAACGAGGTTCACAGAAGGCTCATCAGTATATTCAGTTGAACATGCTACGACAACAACAGCAGAGGACT
ATGCGAAATACAGCCCGTGTTCAGAAAGTCCCGTTGAGACCACACACCCAATCATCTCCTTCAATGTCAC
AGACCTTTATTCCACAGCAGCCTCAACAGCAAGCACAGGCACAGGCACAGCAGCACGCCCAAGCTCAAGC
GCAAGTCCAGGCTCATCAGCAAGCACAGCATCATGCGCAGGCTCAAGTGCCAATGCAATCGCAACCACAT
CAGCAAGGAGGCCAAACTCAACAACAGCAACCCATTAACACTGGGTCTCCTGCGGGGCCAAATGCTATCA
ACTCTCGTGTGCAACACTTAGCGCAACAACAGATGAATCATCTTCGCCAGCAGGCAACTGCCACTACCCA
GCAACCTATCCCGCAACAGAATATTCCATCAAACCAGCAGGGCCCTGCGGGCCCTTATCCTACATCGCCT
TCAAGAAGACCGAGATTACTGTCTAACGAATCGGGTGCAAGTGCACCCTCTGTAATGACGAAGTCACAGC
TCCAAGGAGGTCCTCCATCACAACAACCACACCAACAGCAAGCTCAGCAGGTAGGACCCCCCAATCAACA
TCAAGGCCAGTCCTCTTCCTTTTATTCGGGCATGCCTCCTCAAGGAGTTGTGGTTCCTCATCAGTTCAAT
CCTCAGCAGTATGCCAATATGCTAGCAAGACAACAGCATGTGCAAGCTCAACAACAGGTTCAGTTACAGC
AGGTTCCACATGTGCAACAAAGACAACAGCAAGACCAACAACAACACCGCTTGTCAGCCGGTTCACCAGG
GCATCCTTCATTTGGCGTTTTTCAACAACCTCCTCCGATGTCAAACCATAATCAGGTCATGATCAACCAG
CAGGGAGAAACCTTTTTTGATCCTCATTCACCATATGCTCAACCTAATGGGTACCCCCAGCCACAGCAAC
AACAACAACAGCAGCAACAACAACAACAACAACAACAACAGCAGCAGCAGCAGCAACAACAGAAGCAGCA
ACAGCCACCACCAAGACAGCCTCAGCGCCAACAAGCGATGGCTACGGCTCCTTTGCCTCATTCTACTTCT
GCCTCGGGTACTCCTCACACGGCCACCACACCTAGATTCTCCCAGCCTGGTCCTGTTTATCAGCAGCCTT
TACCTGCATCTCAACCGCAACATTCTCCGCCTACTTCTATTCAGCAACAGGAACCAATTCCAACTCCTGG
GTCACAACATCAGCAAATAGCACAACCGCAATCACAGAATCAACACCAGCAACCGCAGCAACCTCAAGCA
AGTGCTTCTAAAATGGTGGGTATACAGGAGTACCAGAAAGAGTTAATGATGCTTGAGAAGCAGAACAAAC
AACGTCATGACATGGCAATTAAGAAGGGAAGCGGACATTTCTCTAATTTTGATCCAATTCCAGAGCACAC
ACAGACTGAACCAAAATTCAATGTGAACGTAATGCTCCCTCCCCAGAACTCTGCAGTGGCCACGAAGAAT
ACTCCTGGAACTTCTCCTGGTACGCAAACTCAAAACACTGCACACAGTACTGGCAACACTTCTGCTGGGT
CTACACCGAATAATGTTGTACCAGTTCGAAAAAAGAAGGAGCCATCTAAAAAGAAATCAAAGAAGGCTAC
CGAACCTCCAACTCCTACTACTCCACAGACACCAATTGCAGCTAGGGCACATCAAAACTCTACGGGCGGC
ATTTCAGGTAATAATGCTGCTACTAAGCGACGAAAAAGGGAGCCGCTGGTCGACCAAACAGTTTCACCTA
ACCTTAACGAAGCTTCGAAATCAACAAAGCCTGGGAAAATCTCATCTCAAAATGACTTTACAGGTTCTGA
CAATGGATTTTTACAGGATTTTGGCGATGGAACAGGTCCCCCCACTGGAACTGACGATATGGAGTTTGAT
TTTAACAGTTTTCTCAACAACGAAACTGGTGAACCTAATAGTTCAACCATTCATTTTGACAATGTATTTA
ATTGGGGGGAGGGTACTGAAGCCGGGGATCTATAG

Fig. 1 (continued)

FLO8 ortholog of *S. cerevisiae* (strain ATCC 204508 / S288c), SEQ ID NO:5

>sp|P40068|FLO8_YEAST Transcriptional activator FLO8 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=FLO8 PE=5 SV=2
LisH domain underlined

MSYKVNSSYPDSIPPTEQPYMASQYKQDLQSNIAMATNSEQQRQQQQQQQQQQQWINQPTAENSDLKEKMNCKNT
LNEYIFDFLTKSSLKNTAAAFAQDAHLDRDKGQNPVDGPKSKENNGNQNTFSKVVDTPQGFLYEWWQIFWDIFNTS
SSRGGSEFAQQYYQLVLQEQRQEQIYRSLAVHAARLQHDAERRGEYSNEDIDPMHLAAMMLGNPMAPAVQMRNVNM
NPIPIPMVGNPIVNNFSIPPYNNANPTTGATAVAPTAPPSGDFTNVGPTQNRSQNVTGWPVYNYPMQPTTENPVGN
PCNNNTTNNTTNNKSPVNQPKSLKTMHSTDKPNNVPTSKSTRSRSATSKAKGKVKAGLVAKRRRKNNTATVSAGST
NACSPNITTPGSTTSEPAMVGSRVNKTPRSDIATNFRNQAIIFGEEDIYSNSKSSPSLDGASPSALASKQPTKVRK
NTKKASTSAFPVESTNKLGGNSVVTGKKRSPPNTRVSRRKSTPSVILNADATKDENNMLRTFSNTIAPNIHSAPPT
KTANSLPFPGINLGSFNKPAVSSPLSSVTESCFDPESGKIAGKNGPKRAVNSKVSASSPLSIATPRSGDAQKQRSS
KVPGNVVIKPPHGFSTTNLNITLKNSKIITSQNNTVSQELPNGGNILEAQVGNDSRSSKGNRNTLSTPEEKKPSSN
NQGYDFDALKNSSSLLFPNQAYASNNRTPNENSNVADETSASTNSGDNDNTLIQPSSNVGTTLGPQQTSTNENQNV
HSQNLKFGNIGMVEDQGPDYDLNLLDTNENDFNFINWEG

FLO8 ortholog of *S. cerevisiae* (CEN.PK113-7D), SEQ ID NO:6

>EIW10903.1 Flo8p [Saccharomyces cerevisiae CEN.PK113-7D]
LisH domain underlined

MSYKVNSSYPDSIPPTEQPYMASQYKQDLQSNIAMATNSEQQRQQQQWINQPTAENSDLKEKMNCKNTLNEYIFDF
LTKSSLKNTAAAFAQDAHLDRDKGQNPIDGPKSKENNGNQNTFSKVVDTPQGFLYEWWQIFWDIFNTSSSRGGSEF
AQQYYQLVLQEQRQEQIYRSLAAHAARLQHDAERRGEYSNEDIDPMHLAAMMLGNPMAPAVQMRNVNMNPIPIPMV
GNPIVNNFSIPPYNNANPTTGATAVAPTAPPSGDFANVGPTQNRSQNVTGWPVYNYPMQPTTENPVGNPCNNNTTN
NTTNNKSPVNQPKSLKTMHSTDKPNNVPTSKSTRSRSATSKAKGKVKAGLVAKRRRKNNTATVSAGSTNAGSPNIT
TPGSTTSEPAMVGSRVNKTPRSDIATNFRNQAIIFGEEDIYSNSKSSPSLDGASPSALVSKQPTKVRKNTKKASTS
AFPVESANKLGGNSVVTGKKRSPPNTRVLRRKSTPSVILNADATKDENNMLRTFSNTTAPNIHSAPPTKTANSLPF
PGINLGSFNKPAVSSPLSSVTESCFDPESGKIAGKNGPKRAVNSKVSASSPLSIATPPSGDAQEQRSSNVPGNVVI
KPPHGFSTTNLNITLKSSKIITSQNNTAFQELPNGGNILEAQVGNDSRSSKGNRDTLSTPEEKKPSSNNQGYDFDA
LKNSSSLLFPNQAYASNNRTPNENSNVADETSASTNNGDNDNTLIQPSSNVGTTLGPQQTSTNENQNVHSQNLKFG
NIGMVEDQGPDYDLNLLDTNENDFNFINWEG

FLO8 ortholog of *Yarrowia lipolytica* (CLIB122), SEQ ID NO:7

>XP_503484.1 YALI0E03102p [Yarrowia lipolytica CLIB122]
LisH domain underlined

MPAKSDKELLNAYIYDYLLKHNMHDSARTFGAEAKVVPNVKKEDDKDLPKPLIPIDAPQGFLYEWWALFWDIYSAR
GSKGGGSVPAQQYVQGTMRLRQEHAARAQLQQQHQAQQHAQAQAAAQVQGQAQGQGQGQNPTQGPQPQGHMGMPGQ
GPHQPGGPFMNGNMMFPPGQMRMGQLPQHLQQQGTGVAGANPSDDSSSPGGTSPAKRQRLSPDMGGQSHPEQQGQH
MMGTPNPNNPVFNSQVMQQLKASQGQMPNLQQQQQAQLQQYSNTLSLAQQRAMMNAKGGPNGQPGQPGQPGQPGQP
GQPGQPGQPGQPLPMGYEGMEPGFMGGNGLILNAQQMRQQQAGSGGAAGSGQLNGNSNALHDYQMQLMLLEQQNKK
RLMVARQEQQEGQPRAEGQAAAAGANPGARMSGQFKRPGSSPVVGNIGDGRRVTPKLPNQPSPLIDANRASPLQSN
FNGQGDFNVVVGPSGQMMRMQQPQQQGGPPQQQGQQQGGPGQQQGGPGQQPQQGQQQGQQGGPQGGQHRFDDPQQL
PQSQNGGPQPGSAPGQLPQTQGGPQRPPSRVSQMPPPGVQGGQRTQPSSPGQASNPGSGGSSGSAPGGTTPTQANK
QLKGKKAEPKKRAKKGNQPVTPKAVSESPTPTTPATPSAASNQSLLSKATNFAANNKQAQAAAHAQAQAQAAAQAQ
SQMHMGGMGGSSSGLELDVNGGMGLDNDSSSFLNDFSTGNEGDVGDMDFDFNSFLNTDDNAAGALKFDSGSAFGWG
EGVEAMNE

Fig. 1 (continued)

## FLO8 ortholog of *Ogataea polymorpha*, SEQ ID NO:8

>XP_018209309.1 hypothetical protein OGAPODRAFT_77715 [Ogataea polymorpha]
LisH domain underlined

MSQNATPQTTEGHMSMNDQQSNSSSGTNNTEHPSSSSVSSVPTVSDSTAVNSRELLNAYVYDFILKSGFTATASAF
FKEANIPVIHSEKRPTNSPSSSATGTSDLPASFMTMDAPQGFLYEWWQIFWDVFNARTQRGGTTNATQYYHYVNLK
QKQDHLMSQQAAAVAAASTVMNGNNTSMSGAPVTSAAQDIGVLMPQQQQQQQQAQMAHPPQQVPMAQQQRVNARMQ
QPQQQQSQSMPMHPQAAAQAQMNTLRQQQIAQQAQAAQVAAQAVSQRGSPSKRQRMDAAGNGSATEINQTSSPNIA
MNSQQQQQQQQHPPMPMPQGMMIPNQFGVPQQYAMFAAAQAAQQGQSQQKFNQYMIPNFQNQPQQPQQQMLMHQQM
HQQSVHMQHDQQHPPPPPPNVPQHSQTFSNPNDFFHEMPKQNNPGTARVNDTAIKDYEKQLRLMESQNRRRLDVHR
NVSDSKDPNSPGSAQFTEYSAMLNQMPPQHAAASIMQQRASPATKASPTTKSPANGAPPANGKQKKAPRKARKNSS
SVPATPLTPANNQPTPQTNIPPTPQNTTPQQTPQSSAAAASQVMAGKKGIKRKNRGRAIYPIEYSR

## FLO8 ortholog of *Aspergillus niger* (CBS 513.88), SEQ ID NO:9

### Aspergillus niger

>XP_001395127.2 cAMP-dependent protein kinase pathway protein (Som1)
[Aspergillus niger CBS 513.88]
LisH domain underlined

MNQMNQMNMAGMNPGAGGPVGGVPMINNGSAAPRNEQNINNIPENMINNLNTYIYDYFLKRGYHDCARALVKDESI
KLNTEPPIKTSPGHRRDADVNGVDGDTMMTDGKDGDKLKIPDDLPRPNLPSEGQQSSFLLDWFSLFWDFFWAQRKK
GNSNDVRSYLTHTQNMMRLREQHQNQLLRQQPLMNGQMGQMNIRRNGMVPPNLQKTVLQNNTTGLSQQQLAQMHKN
QQVQMMQQMQREHSDMDMNGHRPQSPASAENAPSPSKRPRLEGGPMNGQQLAPNGRGQAQGIPGQPTPQALLMQNG
LNRAMNPNQFQAFQQSGPAAQQKQMQGMPNGMMNPANVMANPQTEMVSIPEGQVYPINGDYYGANGQMAQVRTGMQ
TPGGQHGNHALQDYQMQLMLLEQQNKRRLMMARQEQDSMARPDGQPQMPGQQLPPGTSPQGSRAGTSPNPNDQMKR
GTPKMPQTGLPGSPSAADAMAQGRGSPASMNFPGGQMPPEMAGPQFFVKNMADGMAAPNGMRPPSSNPAFSTPQMG
QPIQAGANRMPNGGWQPQQGAQGQPMAPQQSPATQPQSTGTPQERNAMPPPQAPPAPGANVGRTQPPSPQTAAPPT
PQQGNKPAPKKKETKDSRKRPKKGAAAAAAAAQANTAATPSSEAEHPPTPTPSTPITPQHPNSFNKTGANATTSAP
QQPTSAPAPPPLVQQPPDQTQQPFNELSIPDASAFNLDFSALENPDILENFDFDTFLNTDADTAGFGFDPNTSYPT
DGVETGAGDGL

Fig. 1 (continued)

SEQ ID NO:10 [pG1-3, pGTH1-D1240]
T motifs: italic and underlined; first and second core regulatory regions: italic, double underlined, spacer core region between first and second core regulatory region: italic; translation initiation site: dotted line; bold: repeated regulatory region;

CAAACATTTGCTCCCCCTAGTCTCCAGGGAAATGTAAAATATACTGCTAATAGAAAACAGTAAGACGCTCAGTTGT
CAGGATAATTACGTTCGACTGTAGTAAAACAGGAATCTGTATTGTTAGAAAGAACGAGAGTTTTTTACGGCGCCGC
CATATTGGGCCGTGTGAAAACAGCTTGAAACCCCACTACTTTCAAAGGTTCTGTTGCTATACACGAACCATGTTTA
ACCAACCTCGCTTTTGACTTGACTGAAGTCATCGGTTAACAATCAAGTACCCTAGTCTGTCTGAATGCTCCTTTCC
ATATTCAGTAGGTGTTTCTTGCACTTTTGCATGCACTGCGGAAGAATTAGCCAATAGCGCGTTTCATATGCGCTTT
TACCCCCTCTTTTGTCAAGCGCAAAATGCCTGTAAGATTTGGTGGGGGTGTGAGCCGTTAGCTGAAGTACAACAGG
CTAATTCCCTGAAAAAACTGCAGCTCAGGGATTCCCACTATTTGGTATTCTGATATGTTTTTCCTGATATGCATCA
AAACTCTAATCTAAAACCTGAATCTCCGC*TATTTTTTTTTTTTTTTT*GATGACCCCGTTTTCGTGACAAATTAATTT
CCAACGGGGTCTTGT**CCGGATAAGAGAATTTTGTTTGATTATCCGTTCGG*ATAAATGGACGCCTGCTCCATATTTT
TCCGGTT*ATTACCCCACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATA**CGGTGGTCTGGATTAATTAA
TACGAGATCTCAGGGATTCCCACTATTTGGTATTCTGATATGTTTTTCCTGATATGCATCAAAACTCTAATCTAAA
ACCTGAATCTCCGC*TATTTTTTTTTTTTTTTT*GATGACCCCGTTTTCGTGACAAATTAATTTCCAACGGGGTCTTGT
**CCGGATAAGAGAATTTTGTTTGATTATCCGTTCGG*ATAAATGGACGCCTGCTCCATATTTTTCCGGTT*ATTACCCC
ACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATAC**GGTGGTCTGGATTAATTAATACGCCAAGTCTTAC
ATTTTGTTGCAGTCTCGTGCGAGTATGTGCAATAATAAACAAGATGAGCCAATTTATTGGATTAGTTGCAGCTTGA
CCCCGCCATAGCTAGGCATAGCCAAGTGCTATGGGTGTTAGATGATGCACTTGGATGCAGTGAGTTTTGGAGTATA
AAAGATCCTTAAAATTCCACCCTT

SEQ ID NO:11 [pG1-D1427, pG1-4 (PG1-D1427-2xTA(T)14]:
Example comprising two T motifs which are (T)n (n=14), each extended by preceding TA (extending the T motif at its 5'-end to become a TA(T)n motif)]
T motifs: italic and underlined; first and second core regulatory regions: italic, double underlined, spacer core region between first and second core regulatory region: italic; translation initiation site: dotted line; bold: repeated regulatory region;

CAAACATTTGCTCCCCCTAGTCTCCAGGGAAATGTAAAATATACTGCTAATAGAAAACAGTAAGACGCTCAGTTGT
CAGGATAATTACGTTCGACTGTAGTAAAACAGGAATCTGTATTGTTAGAAAGAACGAGAGTTTTTTACGGCGCCGC
CATATTGGGCCGTGTGAAAACAGCTTGAAACCCCACTACTTTCAAAGGTTCTGTTGCTATACACGAACCATGTTTA
ACCAACCTCGCTTTTGACTTGACTGAAGTCATCGGTTAACAATCAAGTACCCTAGTCTGTCTGAATGCTCCTTTCC
ATATTCAGTAGGTGTTTCTTGCACTTTTGCATGCACTGCGGAAGAATTAGCCAATAGCGCGTTTCATATGCGCTTT
TACCCCCTCTTTTGTCAAGCGCAAAATGCCTGTAAGATTTGGTGGGGGTGTGAGCCGTTAGCTGAAGTACAACAGG
CTAATTCCCTGAAAAAACTGCAGCTCAGGGATTCCCACTATTTGGTATTCTGATATGTTTTTCCTGATATGCATCA
AAACTCTAATCTAAAACCTGAATCTCCGC*TATTTTTTTTTTTTTTTT*GATGACCCCGTTTTCGTGACAAATTAATTTC
CAACGGGGTCTTGT**CCGGATAAGAGAATTTTGTTTGATTATCCGTTCGG*ATAAATGGACGCCTGCTCCATATTTTT
CCGGTT*ATTACCCCACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATA**CGGTGGTCTGGATTAATTAAT
ACGCCAAGTCTTACATTTTGTTGCAGTCTCGTGCGAGTATGTGCAATAATAAACAAGATGAGCCAATTTATTGGAT
TAGTTGCAGCTTGACCCCGCCATAGCTAGGCATAGCCAAGTGCTATGGGTGTTAGATGATGCACTTGGATGCAGTG
AGTTTTGGAGTATAAAAGATCCTTAAAATTCCACCCTTAGATCTCAGGGATTCCCACTATTTGGTATTCTGATATG
TTTTTCCTGATATGCATCAAAACTCTAATCTAAAACCTGAATCTCCGC*TATTTTTTTTTTTTTTTT*GATGACCCCGTT
TTCGTGACAAATTAATTTCCAACGGGGTCTTGT**CCGGATAAGAGAATTTTGTTTGATTATCCGTTCGG*ATAAATGG
A*CGCCTGCTCCATATTTTTCCGGTT*ATTACCCCACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATA**CG
GTGGTCTGGATTAATTAATACGCCAAGTCTTACATTTTGTTGCAGTCTCGTGCGAGTATGTGCAATAATAAACAAG
ATGAGCCAATTTATTGGATTAGTTGCAGCTTGACCCCGCCATAGCTAGGCATAGCCAAGTGCTATGGGTGTTAGAT
GATGCACTTGGATGCAGTGAGTTTTGGAGTATAAAAGATCCTTAAAATTCCACCCTT

Fig. 1 (continued)

pG1, SEQ ID NO:12

CAAACATTTGCTCCCCCTAGTCTCCAGGGAAATGTAAAATATACTGCTAATAGAAAACAGTAAGACGCTCAGTTGT
CAGGATAATTACGTTCGACTGTAGTAAAACAGGAATCTGTATTGTTAGAAAGAACGAGAGTTTTTTACGGCGCCGC
CATATTGGGCCGTGTGAAAACAGCTTGAAACCCCACTACTTTCAAAGGTTCTGTTGCTATACACGAACCATGTTTA
ACCAACCTCGCTTTTGACTTGACTGAAGTCATCGGTTAACAATCAAGTACCCTAGTCTGTCTGAATGCTCCTTTCC
ATATTCAGTAGGTGTTTCTTGCACTTTTGCATGCACTGCGGAAGAATTAGCCAATAGCGCGTTTCATATGCGCTTT
TACCCCCTCTTTTGTCAAGCGCAAAATGCCTGTAAGATTTGGTGGGGGTGTGAGCCGTTAGCTGAAGTACAACAGG
CTAATTCCCTGAAAAAACTGCAGATAGACTTCAAGATCTCAGGGATTCCCACTATTTGGTATTCTGATATGTTTTT
CCTGATATGCATCAAAACTCTAATCTAAAACCTGAATCTCCGCTATTTTTTTTTTTTTTnnGATGACCCCGTTTTC
GTGACAAATTAATTTCCAACGGGGTCTTGTCCGGATAAGAGAATTTTGTTTGATTATCCGTTCGGATAAATGGACG
CCTGCTCCATATTTTTCCGGTTATTACCCCACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATACGGTG
GTCTGGATTAATTAATACGCCAAGTCTTACATTTTGTTGCAGTCTCGTGCGAGTATGTGCAATAATAAACAAGATG
AGCCAATTTATTGGATTAGTTGCAGCTTGACCCCGCCATAGCTAGGCATAGCCAAGTGCTATGGGTGTTAGATGAT
GCACTTGGATGCAGTGAGTTTTGGAGTATAAAAGATCCTTAAAATTCCACCCTT

wherein n at positions 592 and 593 is T or absent

pG3, SEQ ID NO:13

GTAAATAGCGGCAGCAATCCAGTAACCTTTTCTGAATAGCAGAGCCTTAACTAAAATAATGGCCAGGGTAAAAAAT
TCGAAATTTGACACCAAAAATAAAGACTTGTCGTTATAAGTCTTAACAAAGTCCGCAATTTTGGAGCTAACGGTGG
CGGTTGCTGGGATATTCAATAATGGTAGAATGTTGCTGCGGGTATATGACAGAGCGTGAAACACACTGAACAAGGT
AAATGGAACAACAGCAATTGCAATATGGGGGAGGATAGTCAAGAACAAAGCAGCAATGGCAAAGTACTGAATATTC
TCCAAAGCCAAAAGGTCCAGTGGTTTCAACGACAAAGTCTTGTTGGTATAGCTTTGGAACAAAAGGACACCGAAAG
ACTCGACAGCGCCCACAAATACAGCGTTGTAGAAGAACGAATTGATTGCTCCAGAGCTTCTAATAGTCAGAAGATA
CCCCAAACCTCCGAGCAACGTTAGCACATGACCTAAGAACCAGGCGAAGTGAAGAGTCTGGAATAACGACACCCAG
TCAGTTTTTCCTGAGCTCCTGGTGGGATTGGTAGAAGCATTTGATTTGCTTGGAGTGGTTTTATTTGAAGATGGTG
TTGAAGCCATTGTTGCTAAAGAGTCGGAGTTTTGCTTTTAGGGTTTGTTAAGCAAAGGAGGAAAAACTGCGCCGTT
TGAAGTCCCAGGTAGTTTCGCGTGTGAGGCCAGCCAGGGAAAGCTTCCTTCGGTACTTTTTTTTCTTTTGCAGGTT
CCGGACGGATTAAGCTTCGGGTTATGAGGGGGGCGGTAGCCAATTCCGGACACAATATTGCGTCGCAGCTAGTCAC
CCCGCCATAAATATACGCAGGATTGAGGTAATAACATCGATAGTCTTAGTAATTAATACAATTCAGTGGCGAATTT
GGCAACATGACGTAAGGCCCACTGTTGTCTATAAAAGGGGATGAATTTTCATGTTTTTGAGGCCTCCCGGACAATT
TATTGAACTCAA

pG4, SEQ ID NO:14

TGGACTGTTCAATTTGAAGTCGATGCTGACGATGTCAAGAGAGATGCTCAATTATATTTGTCATTTGCTGGTTACA
CTGGAAACGCTACTTTTGTTGGCGGAAACTCTACCAGTTTGGCCGTCCATGTAAACGATGTCGTTCTGGGCCGTGA
CCGTTTCAACACGAACATAACCAATGACAAATCCACTTACAGGTCTAGTTCATATGGAGGCAATTGGTACCTTACT
TCTTTGGATGTCCCAAGTGGGGCTTTAACGTCTGGTACTAACAATGTCTCGTTTGTCACTACAAACTCCGAGGTAA
ATAAAGGATTCTTGTGGGATTCTCTCAAGTTTGTTTGGAAGTTGTAACAGGTTTATAAGCATATCGTGCGCTTGTC
CACAATTGAATCATTTATTGTTGCGAGATACATGAACAAAGTGTGAACTGGGACCCATTACTACAATTCCCACGCA
ACCGTTGTTTCAAAGCCCATATTTTTTGACAATTGTTTCGTTACACCCCCAGTTTGATGTACATCGCTTGCAATGA
TGTGTGTCCCGGAGTATTTTCCATATTCAGCTTGAATTCGTATACTCAACCAATATCTGGGGGTATACTTTTATGT
AACCTATACAAATCAACTATACTATTTCACCTTTCGACCAATCATCTCCCATCTTGTTAAGTTTTGCTTCCTATAT
CCCTGACCCTGACATCACCCATGATTCCGCTCAACGGTTCTCCTCTACATCGTCCCTCTTTTGGAGAGGGTGTTCA
GTTTGACATTCAAATTACCCCCCGCCATCACGCGCAACCGAGACCGCACCCCCGAATTTTCACAAATTACCCCACA
CCCTATACTCCACCACTATGAGGGTTATTAGAACTGATCACGTATAAATACCACCGCAAGTTCCCAAGGGATCGTG
TTCTTCTTCTCCAATTGCAATCATATTTCTGACTCTTTCTAGTTCAGATTAATTCCTTTACACTTGCTTTTTTCCC
TTACCTTTATCC

Fig. 1 (continued)

pG6, SEQ ID NO:15

AGACCAGCAGTTTAACTACGCAAATCCACAGGAATTTCTACATCACAATACCAATGGTAATACCACGACGTCAAGG
AATGGAAACGACGACTTGGAGGAAGACTTCGTCAACCTCTTGCGGAGTACCCGAGGCTAAGACAATAAGAAGAAAA
AAAAAAGAAAAGCGGTGGGGGAGGGATTATTAAATAAGGATTATGTAACCCCAGGGTACCGTTCTATACATATTTA
AGGATTATTTAGGACAATCGATGAAATCGGCATCAAACTGGATGGGAGTATAGTGTCCGGATAATCGGATAAATCA
TCTTGCGAGGAGCCGCTTGGTTGGTTGGTGAGAGGAGTGAAATATGTGTCTCCTCACCCAAGAATCGCGATATCAG
CACCCTGTGGGGGACACTATTGGCCTCCCTCCCAAACCTTCGATGTGGTAGTGCTTTATTATATTGATTACATTGA
TTACATAGCTAAACCCTGCCTGGTTGCAAGTTGAGCTCCGAATTCCAATATTAGTAAAATGCCTGCAAGATAACCT
CGGTATGGCGTCCGACCCCGCTTAATTATTTTAACTCCTTTCCAACGAGGACTTCGTAATTTTTGATTAGGGAGTT
GAGAAACGGGGGGTCTTGATACCTCCTCGATTTCAGATCCCACCCCCTCTCAGTCCCAAGTGGGACCCCCCTCGGC
CGTGAAATGCGCGCACTTTAGTTTTTTTCGCATGTAAACGCCGGTGTCCGTCAATTAAAAGTCGCAGACTAGGGTG
AACTTTACCATTTTTGTCGCACTCCGTCTCCTCGGAATAGGGGTGTAGTAATTCTGCAGTAGTGCAATTTTTACCC
CGCCAAGGGGGGGCGAAAAGAGACGACCTCATCACGCATTCTCCAGTCGCTCTCTACGCCTACAGCACCGACGTAG
TTAACTTTCTCCCATATATAAAGCAATTGCCATTCCCCTGAAAACTTTAACCTCTGCTTTTTCTTGATTTTTCCTT
GCCCAAAGAAAAG

pG8, SEQ ID NO:16

CTGCACAACCATTGCCAGTAAGGACGAAGAGAAGGCCCCACTACCCAAAATTCAGGATAACGTCTTCATACCATGC
AGCGACGCCTACAAGACGCTGTCAAGACATGCCAACTTCAACGAAGTGAACTTTAACACATTGATCGGGAAATTGA
CCACCAAGGGAATGCTGGTTGAGGCTGGAAGCGTTGCCAGTGTCCTGAGGGAACTGGACCGAAAGTTTAGTAATGC
ATAAGAGGATATATATAGGAATGCAGTAATAATATTAGTACCCATTAAGTGGGCTAAGCCATTGGAAGGCCGTCTG
ACTGATGGTGGTGTTCTTCTCATTTAGATAGTGCATTTGCAACTACCGTCTGAGATTGAGTTTGATGTGAAGCTCC
AGCGCCAAAACAGTATAAGAACCTTATCTCCGCATTATTGTTCTTGCGTAAAAGTTTGTGTGAAGAAACAGGGGTA
GTTGCGCAGATTAGTTGTAATATGCGCATAGGATGGGTCATTGACTTCTTTCCTCGAAAGAGCCACACCGTTAGCT
AAAAAAGGACGCGCATCTACCCCAAAATAGAATGTGGGGAAATAGGACGCGCAACTTCCTCTCAATCACTGGACGT
CAGAAAAACAAATGCGCAATCGAGTCACCCTCCGTGATACCCTCCGTGATACCCCCTCTCCGTCTATTCTGACAGC
GTCTCCCCATGACGTTTCAATCTACTTAGAAAAGATTTCGTTTTTTTTTCCTTCAATTACACGATCTCATCTTCTG
CAAGGGTCTGGAGGACATCACCAATCTGCGACTCCATAACTTAGTCCTGAGTTTATATTTACGCTTCATCTGATGA
GTAGGAAGAAAAAGTTTCACGAAATTCCCCCGCCAACTTGCCCTTCGGAATAAGCAGCCACTCTCCTTCTGCCCAT
AGTAAGCTTGCGCGAGGCCCCAACTTGGCCAGAAACTTTAAATATGCCAAACAATCTCCCCCAATCTAAGTTCTCC
CTCTTCTAAAAA

first core regulatory region, SEQ ID NO:17

ATAAATGGA

second core regulatory region, SEQ ID NO:18:

CATATTTTTCCGGTT

Fig. 1 (continued)

T motif identified by any of SEQ ID NO:19-34

- TA$(T)_{13}$: SEQ ID NO:19: *TATTTTTTTTTTTT*
- $(T)_{15}$ : SEQ ID NO:20 *TTTTTTTTTTTTTT*
- TA$(T)_{14}$: SEQ ID NO:21: *TATTTTTTTTTTTTT*
- $(T)_{16}$ : SEQ ID NO:22: *TTTTTTTTTTTTTT*
- TA$(T)_{15}$: SEQ ID NO:23: *TATTTTTTTTTTTTTT*
- $(T)_{17}$ : SEQ ID NO:24: *TTTTTTTTTTTTTTT*
- TA$(T)_{16}$: SEQ ID NO:25: *TATTTTTTTTTTTTTTT*
- $(T)_{18}$ : SEQ ID NO:26: *TTTTTTTTTTTTTTTT*
- TA$(T)_{17}$: SEQ ID NO:27: *TATTTTTTTTTTTTTTTT*
- $(T)_{19}$ : SEQ ID NO:28: *TTTTTTTTTTTTTTTTT*
- TA$(T)_{18}$: SEQ ID NO:29: *TATTTTTTTTTTTTTTTTT*
- $(T)_{20}$ : SEQ ID NO:30: *TTTTTTTTTTTTTTTTTT*
- TA$(T)_{19}$: SEQ ID NO:31: *TATTTTTTTTTTTTTTTTTT*
- $(T)_{21}$ : SEQ ID NO:32: *TTTTTTTTTTTTTTTTTTT*
- TA$(T)_{20}$: SEQ ID NO:33: *TATTTTTTTTTTTTTTTTTTT*
- $(T)_{22}$ : SEQ ID NO:34: *TTTTTTTTTTTTTTTTTTTT*

Main regulatory region: (SEQ ID NO:35): ATAAATGGACGCCTGCTCCATATTTTTCCGGTT
(including the first and second core region, SEQ ID NO:17 and 18, underlined)

Spacer core region: (SEQ ID NO:36): CGCCTGCTC

Spacer main region: (SEQ ID NO:37): [incl. a T16 motif, italic and underlined]
**ATTACCCCACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATA**CGGTGGTCTGGATTAATTAATACGA
GATCTCAGGGATTCCCACTATTTGGTATTCTGATATGTTTTTCCTGATATGCATCAAAACTCTAATCTAAAACCTGA
ATCTCCGC*TTTTTTTTTTTTTTTT*GATGACCCCGTTTTCGTGACAAATTAATTTCCAACGGGGTCTTGT**CCGGATAA**
**GAGAATTTTGTTTGATTATCCGTTCGG**

Exemplary translation initiation site (of pG1): (SEQ ID NO:38): TTCCACCCTT

Exemplary translation initiation site (of pAOX1): (SEQ ID NO:39): TTCGAAACG

Fig. 1 (continued)

Exemplary translation initiation site (consensus sequence of eukaryotes):

(SEQ ID NO:40):

gccgccRcc

wherein R at position 7 is A or G, preferably A
[upper-case letters indicate highly conserved bases; R=purine (A or G, with A being more frequent here), see also Kozak M (1987). "An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs". Nucleic Acids Res. 15 (20): 8125–8148.]

Exemplary translation initiation site (e.g. used in pPUZZLE vectors):

(SEQ ID NO:85): CCTGCAGGCC

Exemplary translation initiation site (consensus sequence of *P. pastoris*):

(SEQ ID NO:86):

nnnnnnAnn

wherein
n at positions 1-6 is any nucleic acid, preferably A;
n at positions 8-9 is any nucleic acid, preferably A

pG1a (SEQ ID NO:41)

GGAATCTGTATTGTTAGAAAGAACGAGAGTTTTTTACGGCGCCGCCATATTGGGCCGTGTGAAAACAGCTTGAAAC
CCCACTACTTTCAAAGGTTCTGTTGCTATACACGAACCATGTTTAACCAACCTCGCTTTTGACTTGACTGAAGTCA
TCGGTTAACAATCAAGTACCCTAGTCTGTCTGAATGCTCCTTTCCATATTCAGTAGGTGTTTCTTGCACTTTTGCA
TGCACTGCGGAAGAATTAGCCAATAGCGCGTTTCATATGCGCTTTTACCCCCTCTTTTGTCAAGCGCAAAATGCCT
GTAAGATTTGGTGGGGGTGTGAGCCGTTAGCTGAAGTACAACAGGCTAATTCCCTGAAAAAACTGCAGATAGACTT
CAAGATCTCAGGGATTCCCACTATTTGGTATTCTGATATGTTTTTCCTGATATGCATCAAAACTCTAATCTAAAAC
CTGAATCTCCGCTATTTTTTTTTTTTTTTTGATGACCCCGTTTTCGTGACAAATTAATTTCCAACGGGGTCTTGTC
CGGATAAGAGAATTTTGTTTGATTATCCGTTCGGATAAATGGACGCCTGCTCCATATTTTTCCGGTTATTACCCCA
CCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATACGGTGGTCTGGATTAATTAATACGCCAAGTCTTACA
TTTTGTTGCAGTCTCGTGCGAGTATGTGCAATAATAAACAAGATGAGCCAATTTATTGGATTAGTTGCAGCTTGAC
CCCGCCATAGCTAGGCATAGCCAAGTGCTATGGGTGTTAGATGATGCACTTGGATGCAGTGAGTTTTGGAGTATAA
AAGATCCTTAAAATTCCACCCTT

pG1b (SEQ ID NO:42)

CCATATTCAGTAGGTGTTTCTTGCACTTTTGCATGCACTGCGGAAGAATTAGCCAATAGCGCGTTTCATATGCGCT
TTTACCCCCTCTTTTGTCAAGCGCAAAATGCCTGTAAGATTTGGTGGGGGTGTGAGCCGTTAGCTGAAGTACAACA
GGCTAATTCCCTGAAAAAACTGCAGATAGACTTCAAGATCTCAGGGATTCCCACTATTTGGTATTCTGATATGTTT
TTCCTGATATGCATCAAAACTCTAATCTAAAACCTGAATCTCCGCTATTTTTTTTTTTTTTTTGATGACCCCGTTT
TCGTGACAAATTAATTTCCAACGGGGTCTTGTCCGGATAAGAGAATTTTGTTTGATTATCCGTTCGGATAAATGGA
CGCCTGCTCCATATTTTTCCGGTTATTACCCCACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATACGG
TGGTCTGGATTAATTAATACGCCAAGTCTTACATTTTGTTGCAGTCTCGTGCGAGTATGTGCAATAATAAACAAGA
TGAGCCAATTTATTGGATTAGTTGCAGCTTGACCCCGCCATAGCTAGGCATAGCCAAGTGCTATGGGTGTTAGATG
ATGCACTTGGATGCAGTGAGTTTTGGAGTATAAAAGATCCTTAAAATTCCACCCTT

Fig. 1 (continued)

pG1c (SEQ ID NO:43)

CTGCAGATAGACTTCAAGATCTCAGGGATTCCCACTATTTGGTATTCTGATATGTTTTTCCTGATATGCATCAAAA
CTCTAATCTAAAACCTGAATCTCCGCTATTTTTTTTTTTTTTTTGATGACCCCGTTTTCGTGACAAATTAATTTCC
AACGGGGTCTTGTCCGGATAAGAGAATTTTGTTTGATTATCCGTTCGGATAAATGGACGCCTGCTCCATATTTTTC
CGGTTATTACCCCACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATACGGTGGTCTGGATTAATTAATA
CGCCAAGTCTTACATTTTGTTGCAGTCTCGTGCGAGTATGTGCAATAATAAACAAGATGAGCCAATTTATTGGATT
AGTTGCAGCTTGACCCCGCCATAGCTAGGCATAGCCAAGTGCTATGGGTGTTAGATGATGCACTTGGATGCAGTGA
GTTTTGGAGTATAAAAGATCCTTAAAATTCCACCCTT

pG1d (SEQ ID NO:44)

GACCCCGTTTTCGTGACAAATTAATTTCCAACGGGGTCTTGTCCGGATAAGAGAATTTTGTTTGATTATCCGTTCG
GATAAATGGACGCCTGCTCCATATTTTTCCGGTTATTACCCCACCTGGAAGTGCCCAGAATTTTCCGGGGATTACG
GATAATACGGTGGTCTGGATTAATTAATACGCCAAGTCTTACATTTTGTTGCAGTCTCGTGCGAGTATGTGCAATA
ATAAACAAGATGAGCCAATTTATTGGATTAGTTGCAGCTTGACCCCGCCATAGCTAGGCATAGCCAAGTGCTATGG
GTGTTAGATGATGCACTTGGATGCAGTGAGTTTTGGAGTATAAAAGATCCTTAAAATTCCACCCTT

pG1e (SEQ ID NO:45)

CCGGATAAGAGAATTTTGTTTGATTATCCGTTCGGATAAATGGACGCCTGCTCCATATTTTTCCGGTTATTACCCC
ACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATACGGTGGTCTGGATTAATTAATACGCCAAGTCTTAC
ATTTTGTTGCAGTCTCGTGCGAGTATGTGCAATAATAAACAAGATGAGCCAATTTATTGGATTAGTTGCAGCTTGA
CCCCGCCATAGCTAGGCATAGCCAAGTGCTATGGGTGTTAGATGATGCACTTGGATGCAGTGAGTTTTGGAGTATA
AAAGATCCTTAAAATTCCACCCTT

Native pGAP promoter sequence of *P. pastoris* (GS115): SEQ ID NO:46

CTGCTACTCTGGTCCCAAGTGAACCACCTTTTGGACCCTATTGACCGGACCTTAACTTGCCAAACCTAAACGCTTA
ATGCCTCAGACGTTTTAATGCCTCTCAACACCTCCAAGGTTGCTTTCTTGAGCATGCCTACTAGGAACTTTAACGA
ACTGTGGGGTTGCAGACAGTTTCAGGCGTGTCCCGACCAATATGGCCTACTAGACTCTCTGAAAAATCACAGTTTT
CCAGTAGTTCCGATCAAATTACCATCGAAATGGTCCCATAAACGGACATTTGACATCCGTTCCTGAATTATAGTCT
TCCACCGTGGATCATGGTGTTCCTTTTTTTCCCAAAGAATATCAGCATCCCTTAACTACGTTAGGTCAGTGATGAC
AATGGACCAAATTGTTGCAAGGTTTTTCTTTTTCTTTCATCGGCACATTTCAGCCTCACATGCGACTATTATCGAT
CAATGAAATCCATCAAGATTGAAATCTTAAAATTGCCCCTTTCACTTGACAGGATCCTTTTTTTGTAGAAATGTCTT
GGTGTCCTCGTCCAATCAGGTAGCCATCTCTGAAATATCTGGCTCCGTTGCAACTCCGAACGACCTGCTGGCAACG
TAAAATTCTCCGGGGGTAAAACTTAAATGTGGAGTAATGGAACCAGAAACGTCTCTTCCCTTCTCTCTCCTTCCACC
GCCCGTTACCGTCCCTAGGAAATTTTACTCTGCTGGAGAGCTTCTTCTACGGCCCCCTTGCAGCAATGCTCTTCCC
AGCATTACGTTGCGGGTAAAACGGAGGTCGTGTACCCGACCTAGCAGCCCAGGGATGGAAAAGTCCCGGCCGTCGC
TGGCAATAATAGCGGGCGGACGCATGTCATGAGATTATTGGAAACCACCAGAATCGAATATAAAAGGCGAACACCT
TTCCCAATTTTGGTTTCTCCTGACCCAAAGACTTTAAATTTAATTTATTTGTCCCTATTTCAATCAATTGAACAAC
TATCAAAACACA

Fig. 1 (continued)

Rgt1 of *P. pastoris* (PAS_chr1-3_0233): SEQ ID NO:47

MIPTIDPKDPELVSEDTAQSASARKRSKVSRACDECRRKKIKCDATFLANSNTLLKPCTNCYKYNCSCSFTRVPLK
RGPSKGFARDGSGYERRRSSSVHSVSSSQSVTSPVPSHASLPIPPANPVSLPRLNVPGDGLLSPKAVPPTNLFWKV
PYELPSFSDRRSSVASADSFRRPSIYQSDSEDDFYSATGSQRNSISQAPRQRNLSPALSVSSTSSLNNRIKSLNMV
ASTLESNIHNYYSQGFNSSLPILPLDERILSTLLSNVSNGSSSASWDAIRSPILELFDKSILMLLRSYESQFNFND
LLDHVTEMQSIYPRIRSHLLSDELLKLIFLMSGVLTDYALILTGQPYSTGLSITVSVFNDWKTYENVQRVLVINRA
GSLDLDYDSLPFLFARCYLSLATLDLIYSLSFSSPRLISSFANLPILDIVQKCGITKDAKLDETPLPVLDQFLNCF
LPGDTYPTALNTLKTGLVLLDFTNNRSTTLRFPFINIHDDNHMTGLSHLLSNVSDFMSQFTEVHSDSKDSQLLFLR
CIWAFWEIGSVLSELIDHFISSSANSQVGDKDASFFYEHQLKVTTLLGTFSNIASAFLTSSTTAASHPPPSISPFH
IISMVESFKMVQFLNKLIASFISLNEKLEKRELEDELSKCKEELNNLNERFQAVSSVQTLPVVHVLFRDLVFSSNR
LDTQRDRASSVVSATTTTSTATTTATTKKSSFGNLLHSDEENILPTVIDWCKEQKHSAEMFLNKNDLNGWLY

Cat8-1 of *P. pastoris* (PAS_chr2-1_0757): SEQ ID NO:48

MMPEEQVTSPQRKHQKSKAKTIRAPGSSIERVAQACDRCRSKKTRCDGKRPQCSQCAAVGFECKISDKLSRRAFPR
GYTETLEERIRELEFENKKLHKLIDLKNEQVEIKNRIDQESTLTNENLTLLNKEQEVSHSGNIHHHADGEPCNCAN
SVSARPVSIAGSVDIDTTDLSDEDDSLYSAASYNAKYHQTGTSGPEMVRLSQRYSSGNFNDPLSFEQSNAPGAAAA
ISIQNKMRTQTFVNLANLVAMSIPRTTEETLFIASLLAKICNVHGFQSKAPILTAKSIALLKDKYNYGNDEVFANI
TLKNVNFNKLTSQQSQQFFQSLNLPNQVNLDLFITTFFNTWNNFIPIINRHIFMSSYIKFNKSRETMFTDNSMFGN
EKFGEILLLITTMVMLSQERNNNREAVPSSSYKKDSTPHPHRPDASSQSNVEILQYYDHLIHEFIKSNISDDCSLP
TLESLSLQLLYCLAIGDLTTSYELRGKIITMGQQLRLHRCPSAVLGTNGSKVSQMQQGERRILFWCIYILDTFSAL
ILGVPRLLKDYEIECALPFSNESNNANVKGSIENTTNTVIINNIKLSLAGKVSDCALAVMRYSKVLGNILDSIFQR
SSINNPSVVSKSTNITEETCLLHEHALDLWRRELSPHINVDLDKSPGGVEYERLSDNQLTILFLYYHAKILIYLPL
MANESSQSRSSASYINIQQSTTSILAIANTLATKERNFYFLPLPVNLSREKVRLAFLSAKGSLEYARGGALFQESK
ILLASVINELKIETSIGMLGCLSVPCMEAVDNAMEQIMALPGKVSSVNGSNSEMKRSSSKRKSSPLRQDISGDERK
SHNIEVSDSRTPSVQSSLYPQPNQMHHPNIIKSENNEQMIPENDTPGAINDIFTSHSPPGTVTSMKEEDLPIKVPI
LLQTQQRQIYNNNPNHSLFSQQPGTQVLSGQQMPGPSSTDQQFKRITTPDGLDSLMMQDFGVDASLGLPMLDFDFN
FDFENVQNNYSQSNVSPPNSESVPSSIQGTHSNDPKDSQVSAGSLFGL

Cat8-2 of *P. pastoris* (PAS_chr4_0540): SEQ ID NO:49

MKENQASNKFNLIKNPITGKPRISQACDRCRIKKIKCDGTLPSCTNCSKIGFVCKISDRLTRSSFPKGYTKNLEQK
LIDMELDRNRLMLELNRIKKEGFDGTNNNIAMASSVSSSENLKSDDSSECQSVTVSLSSTSGPSLSPEPKQDDFRF
RVGMDGSFVLNQFLQSPLMDYIKSLNVLQFNGCANFDQSFNDDPLVLNKYHMNLNRFLNLIFYKLLLPLIHRNSNT
LNEKFAEDNNSLDSLIWKFFTNYNKLIPILEFDSFYKDYLQFIHKYYSNNQVFVDGFRKYFEFSEFEQCFIVKLIL
ILKFTLPVIHDTSVPSEIYRLISMDSLQRLFGNIDFLKPSTDKVSILLLVLHYMVLYESPKSLLDTQDEAQKYDEF
IGNLLSTAVHHITSLRLHIDPRKLQFPRPLPSNGNRLRIKLSWCYKLISKLFRVIYNIDNDSLYSLDDSHLPELQS
ISILHEELDVTIQFNNLLNLIPNNFHSLRDKQSLSKIKTQLLEWHKNFNTEFVEHFNLNDTDSDELSAEKINVLRS
KLISLNRLNCYNSYFQLVIELQLKENLDSVVSGIFGLSNEMLIDNKSSTELLNTLQQTPIIHQSSILVSLCYRIQT
GNLQDEICSILVNNYEKLLQCNDAGLPIKILPQLVHYFKGKISTNLSNSAAHEDLMNMFTLNDNLSTTTTDLDSFI
IPPKRKQDQTLPIGTKRSKSASTSSVISSDDCSLFSNSLSVPTTFSGSSISVGMDNPPSSLFGSYKRPSSIVKQEP
TINPRSNGTNTDSNLFDTFNDSIKGSLNNGLKKLKDIRCNSVVERSHSSQRNDFLMDQEDSITKETINFSELFTCG
TPTASQSIDRSPKSLLLNDLAIAPDTLVIKPDAEDLDRLKNKIRSVKSTVH

LisH domain of the FLO8 protein of *K. phaffii* (SEQ ID NO:50)

LLNGYIYDYLVKSNMQNLADQFAQET

LisH domain of the FLO8 protein of *K. pastoris* (SEQ ID NO:51)

LNGYIHDYLVKSNMQNLADQFAQES
BLASTp: 96% identity to SEQ ID NO :50 (23/24 aa)

Fig. 1 (continued)

LisH domain of the FLO8 ortholog of *S. cerevisiae* (strain ATCC 204508 / S288c) (SEQ ID NO:52)

LNEYIFDFLTKSSLKNTAAAFAQDA
BLASTp: 54% identity to SEQ ID NO :50 (13/24 aa)

LisH domain of the FLO8 ortholog of *S. cerevisiae* (CEN.PK113-7D), (SEQ ID NO:53)

LNEYIFDFLTKSSLKNTAAAFAQDA
BLASTp: 54% identity to SEQ ID NO :50 (13/24 aa)

LisH domain of the FLO8 ortholog of *Yarrowia lipolytica* (CLIB122), (SEQ ID NO:54)

ELLNAYIYDYLLKHNMHDSARTFGAEA
BLASTp: 71% identity to SEQ ID NO :50 (12/17 aa)

LisH domain of the FLO8 ortholog of *Ogataea polymorpha*, (SEQ ID NO:55)

ELLNAYVYDFILKSGFTATASAFFKEA
BLASTp: 62% identity to SEQ ID NO :50 (8/13 aa)

LisH domain of the FLO8 ortholog of *Aspergillus niger* (CBS 513.88) (SEQ ID NO:56)

NNLNTYIYDYFLKRGYHDCARALVKD
BLASTp: 73% identity to SEQ ID NO :50 (8/11 aa)

Sequence scR (SEQ ID NO:57):

CAGGAACAACTAATGGAGTCTGGGGGTGGTTTGGTTACCCTGGGTGGTTCTCTTAAGCTTTCATGTAAGGCCTCTG
GTATTGATTTTTCGCACTACGGTATCTCCTGGGTTAGACAAGCTCCTGGAAAAGGTCTGGAATGGATCGCTTACAT
TTACCCAAATTACGGTTCTGTTGACTATGCCTCCTGGGTCAATGGTAGGTTCACTATTTCCCTTGACAACGCTCAG
AACACGGTATTCCTACAGATGATCTCCCTAACCGCTGCTGATACTGCAACCTACTTCTGTGCTCGTGACAGAGGTT
ACTACTCTGGCTCTCGTGGAACTAGACTTGACTTATGGGGACAAGGTACTCTCGTTACCATCTCTAGTGGTGGAGG
TGGTTCTGGAGGAGGAGGTTCCGGCGGAGGTGGTAGCGAGCTGGTCATGACTCAAACCCCTCCATCCCTATCTGCA
TCAGTCGGTGAAACCGTTAGAATTAGATGCCTTGCATCTGAGTTCTTGTTCAACGGTGTGTCCTGGTATCAACAAA
AGCCTGGTAAGCCTCCAAAGTTTCTCATTTCTGGTGCCTCAAACCTCGAATCTGGAGTGCCACCAAGATTTTCCGG
ATCTGGCTCTGGTACTGACTACACTCTGACAATTGGTGGTGTTCAAGCTGAGGATGTTGCTACCTACTATTGTCTC
GGTGGTTACTCAGGATCTTCCGGCCTAACTTTCGGTGCCGGTACAAACGTCGAGATCAAAGGTGGACATCACCACC
ACCATCACTAATAG

Sequence vHH (SEQ ID NO:58):

CAGGTTCAGCTGCAGGAGTCCGGTGGTGGTCTGGTTCAAGCCGGTGGTTCATTAAGATTGTCCTGTGCTGCCTCTG
GTAGAACTTTCACTTCTTTCGCAATGGGTTGGTTTAGACAAGCACCTGGAAAAGAGAGAGAGTTTGTTGCTTCTAT
CTCCAGATCCGGTACTTTAACTAGATACGCTGACTCTGCCAAGGGTAGATTCACTATTTCTGTTGACAACGCCAAG
AACACTGTTTCTTTGCAAATGGACAACCTTAACCCAGATGACACCGCAGTCTATTACTGTGCCGCTGACTTGCACA
GACCATACGGTCCAGGAACCCAAAGATCCGATGAGTACGATTCTTGGGGTCAGGGAACTCAAGTCACTGTCTCTTC
AGGTGGTGGATCTGGTGGTGGAGGTTCAGGTGGTGGAGGATCCGGTGGTGGTGGTTCTGGTGGTGGTGGATCTGGT
GGAGGTGAAGTTCAACTTGTCGAATCCGGTGGTGCACTTGTCCAACCTGGTGGATCTCTTAGACTTTCTTGTGCCG
CCTCCGGTTTTCCTGTTAACCGTTACTCTATGCGTTGGTACAGACAAGCCCCTGGAAAAGAACGTGAATGGGTTGC
CGGAATGTCCTCAGCTGGTGACAGATCCTCCTACGAAGATTCTGTGAAGGGACGTTTCACCATCTCCAGAGATGAC
GCCCGTAACACCGTTTACCTTCAAATGAACTCCCTTAAGCCTGAGGATACTGCCGTCTACTATTGTAACGTGAATG
TCGGATTTGAATACTGGGGACAGGGAACCCAAGTTACTGTCTCTTCCGGTGGACATCACCACCACCATCACTAATA
G

Fig. 1 (continued)

Signal sequence (SEQ ID NO:59):

MKXSTNLILAIAAASXVVSA, wherein
X at position 3 is either F or L; and
X at position 16 is either A or T.

Leader sequence (SEQ ID NO:60):

MKXSTNLILAIAAASXVVSAAPVAPAEEAANHLHKR, wherein
X at position 3 is either F or L
X at position 16 is either A or T

Signal sequence, aMF, *S. cerevisiae* (SEQ ID NO:61):

MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAVLPFSNSTNNGLLFINTTIASIAAK
EEGVSLEKR.

Helper factor 1, *Komagataella pastoris* (SEQ ID NO:62):

MSSDAVEQLE NFQLIKFDRF DPSTQSTIRI ARSPKPIPVK VVIVGDGGCG
KTCLLNVFAT GTFPEAYVPT IIENVVITLV TPTGQIAAVT LWDTAGQEEY
DRLRPLSYSD VDVVLLCYSI DNLSTFHNVA DKWYPEVAHF CPNTPIILVG
TKSDMRRHQK SQPHFVSPQD SSQLARQMGA VMNIECSAKE VSNVNIVFDA
AVSYCLSNSR PKTRGDNDNN RSNRRLSRAK RASMFIRGKD VSSTSGNSRE
ELVEYDQDGL AIIPDRKKRK CSII

Helper factor 2, *Komagataella pastoris* (SEQ ID NO:63):

MLNKLFIAIL IVITAVIGET TTSSTTASLS ESPTLVWVTG TDASGRLATT
QSAYTQQFSQ LYSSIASPSS GSIGLGTIQG TVGIVRTYET ITLAS

Helper factor 3, *Komagataella pastoris* (SEQ ID NO:64):

MSTAIPGGQR TLAKRRAANL DKKQDEPTSA RSAGAGGSSS TMLKLYTDEA
QGLKVDPLIV LVLAVGFIFS VIGLHVVAKL TGKLIN

Helper factor 4, *Komagataella pastoris* (SEQ ID NO:65):

MTPRSHIFFD ISINNQPAGR IIFELFNDIV PKTAENFRAL STGEKGIGKS
GKPLHYKGST FHRIIKDFMV QGGDFTNGNG TGGESIYGEK FEDENFQLTH
DKPFLLSMAN AGPGTNGSQF FITTVPTPHL DNKHVVFGKV IAGKATVRKI
ERNSEGEAPI EPVVIEDCGE LPEDADLTIS DETGDKYEEV LKDNENIDID
DFEQVYQAIT EIKELGTKYF KNGDTKIAFE KYQKAANYLL EYIPSDLSEE
QSSKLELLKT SVFSNVALAG LKVSKFKDTI KYATLVIEDE SADAKAKSKG
YYRRGSAYSS LKDEDSAISD FQKALELSPG DPAISQSLQR TTKARKDRLA
KEKAALSKFF E

Helper factor 5, *Komagataella pastoris* (SEQ ID NO:66):

MTNWKAILTP AQYQVLRLGG TERPYTGQYV NFKKNGTYLC SGCQTPLYKS
GTKFDSSCGW PAFYEALPGA VKRIEDNSLG MRRIEIRCSK CDGHLGHVFE
GEGFDTPTDS RHCVNSISLK FQGEEEN

Fig. 1 (continued)

Helper factor 6, *Komagataella pastoris* (SEQ ID NO:67):

MSHLLLRDSF WGRTIYHLSK HRYFSFPEEK DGFIAPEKYY LNMDQVSIHA
ESEKNIVEGL VDTSNSSLEE VKTTRVIVDW DEYDQKENPQ NWSSLLKCFV
VFEVGILTVA VYMGSAIYTP GIEDIMRDLN VSRTVATLPL TLFVIGYAVG
PMIFSPMSEH PAIGRTTIYV WTLFIFAILQ IPTALTTNIA GFCILRFIGG
FFASPALATG PASVGDVIAI PHLPVGLGLW SICAVCGPSL GPLFGAIFSQ
LVSWRWCFWF LLITSGTLFI VLGFTLPETY VPTLLYRKAR RLRALTKNEL
IISKGELDIQ DRTAKEVLIE CLWRPVDISF RDPVVLMINL YISMVYSIWY
IWFEAFPIVF LEIYGFSLIG MGASFAGILI GVLICSACYC YACHVTFARR
IIANETIHPE FFVPGAIIGG CIMPTGIFIL GWTATKSVHW IVPIIGSGLF
AAGGYLIFQT LFNYLAMSFP RYMASAFAGN DLFRSFSASV FPLFGHALYA
NLGSEKFPVG WGSSVLGFIT VAMIAIPVTF MRYGPRLRAN SRYAGP

Helper factor 7, *Komagataella pastoris* (SEQ ID NO:68):

MTDYVTSKRP DNVLNWTSIH VSSWIGETIP EIDPSLLQNF LEHDIAGDVL
PYLKSEDLKE IGINELKHRI SIKKNIHELL VSNEKHIDTS ILSDTATELG
TLILTNKFIT QMANRKNVVD DSTHHSNNRR LTEQFNKLRK DLLPIFKWIK
ETQPLPTPEN THFANMGSVP ASPVEHTSGE STLSNPSLST INAGEGVNSA
VAGQSLGRKP TLSSRRQSHA LSPTGEHLNV SSSSPSTGNF ETLNGERPNL
RSASSGSQEH TENELLKPLR VKADEPCYKV IQNAMKRHGL SVDDWRKYAL
VICYGDEERV LGLHEKPGSI FKELKDQKQN PAIMLRQIDT NNDDQNHIET
PGGRL

Helper factor 8, *Komagataella pastoris* (SEQ ID NO:69):

MTTNGQKRQK TRKPLLINAF VMGCAGLQNP GLWKHPKDSS HRFNQIDHWT
YLAKLAEKGK FNALFIADVL GGYDVYKGPE NLATPAVAGA QWPVTEPSAV
VSAMAAVTTN LAFGVTFSTI SEAPYHFARR LSTLDHLTKG RIGWNVVSSY
LESAARNLLN GEKLDEHDQR YLKAEEYIQI VYELLLSSWR DDAVVLDKKA
GVYTDPTRFR KINFEGKFFK VPGPHIVDPT PQRLPVILQA GTSKVGKEFA
AKHAEIVFVI SFSPDDLKPK IAEVRQLAKE KFGRNHDDIK FVALATPVIG
ATHELAEEKY QELLSYGDIE GAQALFGGWT GIDLSQYGED EELGNVSSNA
MRGAVQNWTK AIPNEKRWTR KVIAKQITVG GLGPAFVGTP EEIADELEHW
SDHAGLDGFN FTYAVNPLSF EEIVEDLIPV LQRRGLAQKE YPNPETGSTF
RKNLFGTDFV PSTHPAYNLR WRAGVSKEEF EKSLNATTNW YSSFARSGAL
GELHNTCRIL YLQIVKYKYR LRVRSEGNSI PFAKMTKENE AKRQKTSQPK
AKKQLIINAF MSGSSGNQSP GLWSYPGDKS TEYTTLDYWV ELAQKLEKAK
FHSIFIADVL GGYDVYNGPG NYSAAAKSGA QFPMIEPSAA VTAMAAATKS
ITFGVTFSTI SEAPYHFARR LGTLDLLTNG RVGWNIVSSY LDSAARNLLN
GEPLPLHADR YKRAEEFLQV VYRLFLSSWR DDAYKLDKKT RTFADPKLIR
TIDHVGEFFN VPGPQFLPPT PQRLPLILQA GTSKVGMDYA AKHAEVVFLA
SFDPESLQEK IKTVRDIAET KYNRPRDSIK FLILITVVIA DTHEDAVKRY
EDLASYADLE GAQALFSGWT GIDIGKYGED EPLEHVESNA IKSHVKNWTK
FKDNKPRARK DIAKQIGVGG SGPLLVGSVQ EIADELERWA EVSDLDGFNF
AYADYPQTFD DIIEKLLPEL NKRGVFWDDY KIPGGTFRES VFGRKFVDKD
HPAYDLRWRS DQTREEFEKK LAELEKK

Helper factor 9, *Komagataella pastoris* (SEQ ID NO:70):

MRFSNVVLTA IAAAGVQADE ALYTVFYNDV TENAQEYLSY IQANTAAGFT
DLLSLYTELA TYTDDSYTSI FTEEDFPASE LSSFVVNLPW YSSRIEPQVA
AAETGESEEE SETGESEEES ETGEETETET GSESESESES ETSATGTGTG
TSASESAETE TSTDAAVSID HPKSTLLMGL TAAVVSITFG VFAL

Fig. 1 (continued)

Helper factor 10, *Komagataella pastoris* (SEQ ID NO:71):

MSSFRVLDLV KPFTPFLPEV ISPERKVPFQ QKLMWTGVTL LIFLVMSEIP
LYGITSSDSS DPLFWLRMML ASNRGTLMEL GISPIVTSGM VFQLLQGIQI
LDVNMENKAD RELFQTAQKV FAILLSIGQA TVYVLTGMYG PPGELGVGVC
LLLVLQLVFA GIVVILLDEL LQKGYGLGSG ISLFMATNIC EQIFWKTFAP
TTVNRGRGKE FEGAFISFFH LILTKKDKKR ALLESFYRDN APNMFQVIAT
LVVFFTVVYL QGFRLEIPVK STRQRGPYGT YPIRLFYTSN MPIMLQSALT
SNIFIISQML YSHFPDNAFV KLIGTWEAQP GSAQLFAASG LAYYMQPPMS
LSQALLDPIK TVVYVVFVLT TCAIFSKTWI EISGSSPRDV AKQFKDQGLV
IAGHRDATVY KELKKIIPTA AAFGGATIGA LSVVSDLLGT LGSGTSILLA
VTTIYGYYEL AVKEGGFSKG GPSGFVDL