Microbiology (2016), **162**, 23–34 DOI 10.1099/mic.0.000211

A novel N-terminal region of the membrane β-hexosyltransferase: its role in secretion of soluble protein by *Pichia pastoris*

Suzanne F. Dagher and José M. Bruno-Bárcena

Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695-7615, USA

Correspondence
José M. Bruno-Bárcena
jbbarcen@ncsu.edu

Received 12 August 2015
Accepted 5 November 2015

The β-hexosyltransferase (BHT) from *Sporobolomyces singularis* is a membrane-bound enzyme that catalyses transgalactosylation reactions to synthesize galacto-oligosaccharides (GOSs). To increase the secretion of the active soluble version of this protein, we examined the uncharacterized novel N-terminal region (amino acids 1–110), which included two predicted endogenous structural domains. The first domain (amino acids 1–22) may act as a classical leader while a non-classical signal was located within the remaining region (amino acids 23–110). A functional analysis of these domains was performed by evaluating the amounts of the rBHT forms secreted by recombinant *P. pastoris* strains carrying combinations of the predicted structural domains and the α mating factor (MFα) from *Saccharomyces cerevisiae* as positive control. Upon replacement of the leader domain (amino acids 1–22) by MFα (*MFα-rBht*$_{(23\text{-}594)}$), protein secretion increased and activity of both soluble and membrane-bound enzymes was improved 53- and 14-fold, respectively. Leader interference was demonstrated when MFα preceded the putative classical rBHT$_{(1\text{-}22)}$ leader (amino acids 1–22), explaining the limited secretion of soluble protein by *P. pastoris* (GS115 : : *MFα-rBht*$_{(1\text{-}594)}$). To validate the role of the N-terminal domains in promoting protein secretion, we tested the domains using a non-secreted protein, the anti-β-galactosidase single-chain variable antibody fragment scFv13R4.The recombinants carrying chimeras of the N-terminal 1–110 regions of rBHT preceding *scFv13R4* correlated with the secretion strength of soluble protein observed with the rBHT recombinants. Finally, soluble bioactive HIS-tagged and non-tagged rBHT (purified to homogeneity) obtained from the most efficient recombinants (GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS and GS115 : : *MFα-rBht*$_{(23\text{-}594)}$) showed comparable activity rates of GOS generation.

INTRODUCTION

There is an increasing interest in the use of enzymes for the production of functional foods, especially in the field of prebiotic production from lactose (Illanes, 2011). A seminal paper by Gorin *et al.* (1964a) showed that the non-conventional yeast *Sporobolomyces singularis* was capable of assimilating lactose and glucose, but not galactose. This physiological feature led to the discovery of the membrane-bound β-hexosyltransferase (BHT) (Phaff & do Carmo-Sousa, 1962; Gorin *et al.*, 1964a, b; Blakely & Mackenzie, 1969; Spencer *et al.*, 2002). BHT is an extracellular membrane-bound enzyme that is able to simultaneously hydrolyse lactose and transfer the galactose monomer to a second lactose molecule, thus generating galacto-oligosaccharides (GOSs), which are considered prebiotics and widely used as functional food additives.

Abbreviations: BHT, β-hexosyltransferase; GOS, galacto-oligosaccharide; qRT-PCR, quantitative real time PCR.

The BHT from *Spo. singularis* also catalyses the hydrolysis of β-glycosidic linkages, such as ONP-Glu and PNP-Glu (Blakely & Mackenzie, 1969), and possesses particularly appealing enzymic capabilities with respect to competing technologies. For instance, this unique membrane-bound enzyme possesses industrial advantages, including enzymic activity at a wide range of pH and temperatures and in the absence of cofactors or additional ions, and GOS formation independent of the initial lactose concentrations (Blakely & Mackenzie, 1969; Gosling *et al.*, 2010; Dagher *et al.*, 2013).

The biotechnological utilization of BHT has been restricted by major rate-limiting steps including secretion levels (limited by its endogenous promoter), and the historical inability to obtain secreted soluble, stable and bioactive BHT. Little is known about the physiology of *Spo. singularis*, a recently accepted GRAS (genetically regarded as safe) organism (Tzortzis & Vulevic, 2009). The majority of available information is associated with the *Bht* gene, an inducible gene repressed by glucose that

codes for this singular protein that also displays a novel N-terminal region that spans 110 aa. When *Spo. singularis* is grown in the presence of an inducer, such as lactose, BHT is expressed and later localized to the cell membrane facing the outside of the cell. Due to its cellular confinement, BHT has been recovered from *Spo. singularis* at very low yields, ranging from 14 % (37.4 units from 265 total units of activity) to 16 % (34 units from 361 total units of activity). Purification attempts have included the release of the membrane-bound BHT using cell-wall lytic enzymes followed by multiple chromatography steps (Cho *et al.*, 2003; Ishikawa *et al.*, 2005). As conventional protein purification protocols give limited BHT recovery, previous studies have assessed alternative strategies to increase enzyme production and expand its technological applications. One study performed directed evolution by mutagenesis and selected an *Spo. singularis* strain lacking glucose repression and consequently able to generate a 10-fold increase in the membrane-bound BHT (Ishikawa *et al.*, 2005). However, there are no reports of soluble enzyme produced by *Spo. singularis*.

Given that there are no rational strategies to increase the expression of membrane proteins in a soluble form, trial-and-error is still the most common approach. Different methods can be used to select the ideal host and the correct decision on which to use appears to be strongly protein-dependent, requiring extensive screening to identify the best organism. Moreover, within that organism, the best cell line or strain and expression system must be determined for a given membrane protein. To arrive at a reasonable choice, we selected and evaluated different systems, initially *Escherichia coli*, and later demonstrated that *P. pastoris* was capable of secreting minor amounts of soluble biologically active recombinant rBHT (Dagher *et al.*, 2013). Although the majority of the enzyme remained associated with the cell membrane, we were able to recover and evaluate the activity of the soluble protein from the broth and compare it with the membrane-bound rBHT, providing evidence for the ability to generate heterologous bioactive soluble protein. Furthermore, the study revealed that *P. pastoris* was a promising host for the production of both soluble and membrane-associated bioactive rBHT, opening for the first time, we believe, the possibility of a straightforward downstream processing protocol (Dagher *et al.*, 2013).

In this study, we evaluated the novel N-terminal region of BHT and its influence on the secretion ratio of soluble versus membrane-associated proteins by *P. pastoris*. Furthermore, the function of the rBHT N-terminal region containing a putative leader domain was validated by generating recombinant chimeras using the non-secreted, hyper-stable single-chain anti-β-galactosidase antibody scFv13R4.

METHODS

Construction of *P. pastoris* GS115 recombinants bearing *rBht* or antibody *scFv13R4*. PCR amplification was performed with Phusion DNA polymerase (New England Biolabs) according to the manufacturer's instructions. Oligonucleotides (primers) were purchased from Integrated DNA Technologies (Table 1). Established protocols were employed for PCR amplifications and cloning (Sambrook & Russell, 2001) into the *P. pastoris* pPIC9 plasmid (Invitrogen Life Technologies) (Table 2). DNA manipulation enzymes (restriction enzymes and T4 DNA ligase) were purchased from New England Biolabs.

The *E. coli* plasmid pJB100 carrying *rBht* (Dagher *et al.*, 2013) was used as a template for PCR amplifications (Table 2). Full-length *rBht* preceded by the α-factor pre-pro leader (*MFα*; 22 aa pre sequence and 67 aa pro sequence) found in the *P. pastoris* expression vector pPIC9 and followed at the 3′ end by the 6 × HIS tag (HIS) was generated using the primers JBB6/JBB5. The digested full PCR product was inserted into pPIC9 at the *Xho*I–*Not*I sites (pJB110, pPIC9-*MFα-rBht*-HIS). The amplicons obtained with the primers JBB7/JBB2 and JBB7/JBB5 were inserted into *Xho*I–*Not*I sites in-frame with *MFα*, replacing the 5′ end sequence ($rBht_{(1\text{-}22)}$) with *MFα* (pJB113, pPIC9-*MFα-*$rBht_{(23\text{-}594)}$) and adding 6 × HIS tag (pJB112, pPIC9-*MFα-*$rBht_{(23\text{-}594)}$-HIS) to the 3′ end, respectively. Removal of *MFα*, at the 5′ end and addition of the 6 × HIS tag to the 3′ end were performed using the primers JBB8/JBB5. The amplicon was later ligated into the pPIC9 plasmid at the *Bam*HI–*Not*I sites (pJB114, pPIC9-*rBht*-HIS). Removal of $rBht_{(1\text{-}110)}$ and addition of the 6 × HIS tag to the 3′ end was performed using JBB10/JBB5, and this was then ligated into pPIC9 *Eco*RI–*Not*I sites (pJB116, pPIC9-MFα-$rBht_{(111\text{-}594)}$-HIS). Removal of both *MFα* and $rBht_{(1\text{-}22)}$ N-terminal sequences was accomplished by inserting the PCR amplicon generated by primers JBB9/JBB5, which added the 3′ 6XHIS tag into the pPIC9 plasmid using the *Bam*HI–*Not*I sites (pJB115, pPIC9-$rBht_{(23\text{-}594)}$-HIS).

The *scFv13R4* (anti-β-galactosidase tagged with myc and 6 × HIS) and $rBht_{(1\text{-}594)}$ ORFs found in pPM163R4 (Martineau *et al.*, 1998) and pJB100 (Dagher *et al.*, 2013), respectively, were used as templates to generate chimeric sequences containing *scFv13R4* with the endogenous 6 × HIS tag (scFv13R4-HIS) at the 3′ end and the leader sequences (*MFα*, $rBht_{(1\text{-}22)}$, $rBht_{(1\text{-}110)}$ and $rBht_{(23\text{-}110)}$) at the 5′ end (Tables 1 and 2). A recombinant strain carrying no leader sequence was generated using primers JBB20/JBB18; the amplicon contained *scFv13R4*-HIS and was later inserted into pPIC9 at the *Bam*HI–*Eco*RI sites (pJB122, pPIC9-*scFv13R4*-HIS). The sequence coding for the leader domain $rBht_{(1\text{-}22)}$ was fused to *scFv13R4*-HIS using primers JBB12/JBB18 and pJB114 (pPIC9-$rBht_{(1\text{-}594)}$-HIS) as the template. The product was inserted into pPIC9 at the *Bam*HI–*Eco*RI sites (pJB119, pPIC9-$rBht_{(1\text{-}22)}$-*scFv13R4*-HIS), generating a unique *Sfi*I site between the leader domain and *scFv13R4*-HIS sequences. The *MFα-*$rBht_{(1\text{-}22)}$ chimeric leader sequence was amplified by PCR using pJB110 (pPIC9-*MFα-*$rBht_{(1\text{-}594)}$-HIS) as a template with the primers JBB13/JBB14, and was inserted as a *Bam*HI–*Sfi*I fragment into pJB119 (pPIC9-$rBht_{(1\text{-}22)}$-*scFv13R4*-HIS), thus generating pJB118 (pPIC9-*MFα-*$rBht_{(1\text{-}22)}$-*scFv13R4*-HIS). To produce a recombinant containing *scFv13R4*-HIS preceded by *MFα*, the ORF was PCR-amplified using primers JBB11/JBB18 and inserted into pPIC9 using *Sna*BI–*Eco*RI sites (pJB117, pPIC9-*MFα-scFv13R4*-HIS). The chimeric sequences encoding $rBht_{(1\text{-}110)}$ or $rBht_{(23\text{-}110)}$ followed by *scFv13R4*-HIS were amplified by two rounds of PCR using pPM163R4 (*scFv13R4*-HIS) and the appropriate *rBht* integrative yeast plasmid as a template. Briefly, the primer/template combinations were PCR-amplified and ligated at their *Hin*dIII sites: (JBB15 and JBB17)/pJB114 (pPIC9-$rBht_{(1\text{-}594)}$-HIS) and (JBB19 and JBB18)/pPM163R4 or (JBB16 and JBB17)/pJB115 (pPIC9-$rBht_{(23\text{-}594)}$-HIS) and (JBB19 and JBB18)/pPM163R4. The final two amplicons were generated using the external primers (JBB15 and JBB18) or (JBB16 and JBB18) flanked by *Bam*HI–*Eco*RI sites and later inserted into pPIC9 to generate pJB120 (pPIC9-$rBht_{(1\text{-}110)}$-*scFv13R4*-HIS) and pJB121 (pPIC9-$rBht_{(23\text{-}110)}$-*scFv13R4*-HIS), respectively.

Sequencing of each integration cassette was performed at Eton Bioscience. The yeast integrative plasmids described above were linearized with *Sac*I prior to electro-transformation into *P. pastoris* GS115

Table 1. Primers, antibodies and substrates used in this study

Primer/antibody/substrate	ORF primer/antigen/ abbreviation	Sequence*	Reference/source
Primer			
JBB2	*Not*I-*rBht* Reverse	5′-aaggaaaaaaGCGGCCGCTTACAGATGAT-TACGCCCAAATTG-3′	Dagher *et al.* (2013)
JBB5	*Not*I-*rBht*-6XHIS Reverse	5′-aaggaaaaaaGCGGCCGCTTAGTGGTGG-TGGTGGTGGTGCAGATGATTACGCC-CAAATTG-3′	This study
JBB6	*Xho*I-*MFα*-*rBht* Forward	5′-GAAGAAGGGGTATCTCTCGAGAAAAGAG-AGGCTGAAGCTATG ATGCTGCATGCTGCAC-3′	This study
JBB7	*Xho*I-*MFα*-*rBht*$_{(23\text{-}594)}$ Forward	5′-ccgCTCGAGAAAAGAGAGGCTGAAGCTG-TTACTTATCCGGGAGCC-3′	This study
JBB8	*Bam*HI-*rBht* Forward	5′-cgcggatccaaacgATGATGCTGCATGCTGCAC-3′	This study
JBB9	*Bam*HI-*rBht*$_{(23\text{-}594)}$ Forward	5′-cgcggatccaaacgGTTACTTATCCGGGAGCC-3′	This study
JBB10	*Eco*RI-*rBht*$_{(111\text{-}594)}$ Forward	5′-ccggaattcATGTTTCCAAAGGGGTTTAAG-TTTG-3′	This study
JBB11	*Sna*BI-*MFα*-*scFv13R4* Forward	5′-gaggctgaagctTACGTAATGGCCGAGGTG-CAGCTG-3′	This study
JBB12	*Bam*HI-*rBht*$_{(1\text{-}22)}$-scFv13R4 Forward	5′-cgcggatccaaacgATGATGCTGCATGCTGCA-CTGCTAGTAGCGCTGCCATGTGTTGT-TTTGGCCCGCCCGGCC-GGCGCCATGGCCGAGG TGCAGCTG-3′	This study
JBB13	*Bam*HI-*MFα* Forward	5′-cgcggatccaaacgATGAGATTTC-3′	This study
JBB14	*Sfi*I-*rBht*$_{(1\text{-}22)}$ Reverse	5′-CGCTCCGGCCGGGCGGGCCcgc-3′	This study
JBB15	*Bam*HI-*rBht*$_{(1\text{-}110)}$ Forward	5′-cgcggatccaaacgATGATGCTGCATGCTGCAC-3′	This study
JBB16	*Bam*HI-*rBht*$_{(23\text{-}110)}$ Forward	5′-cgcggatccaaacgATGGTTACTTATCCGGGAGCC-3′	This study
JBB17	*Hin*dIII-*rBht*$_{(1\text{-}110)}$ Reverse	5′-CCCTTTGGAAGCTTCAGTCC-3′	This study
JBB18	*Eco*RI-6XHIS Reverse	5′-ggaattcTTAATGGTGATGATGGTGATG-3′	This study
JBB19	*Hin*dIII-*scFv13R4* Forward	5′-ggactgaagcttATGGCCGAGGTGCAGCTG-3′	This study
JBB20	*Bam*HI-*scFv13R4* Forward	5′-cgcggatccaaacgATGGCCGAGGTGCAGCTG-3′	This study
JBB3	*rBht* Forward internal sequencing	5′-ATCACTATGCCAGCACGCAGTGTA-3′	Dagher *et al.* (2013)
JBB4	*rBht* Reverse internal sequencing	5′-TTTAAAGCCGATTTCACCTGCCGC-3′	Dagher *et al.* (2013)
5′ AOX1	AOX1	5′-GACTGGTTCCAATTGACAAGC-3′	Invitrogen
3′ AOX1	AOX1	5′-GCAAATGGCATTCTGACATCC-3′	Invitrogen
α-Factor	MFα	5′-TACTATTGCCAGCATTGCTGC-3′	Invitrogen
Antibody			
Mouse anti-HIS	6XHIS		Qiagen
Substrate			
*o*NP-β-D-glucopyranoside	ONP-Glu		Sigma
*o*NP-β-D-galactopyranoside	ONP-Gal		Sigma

*Coding regions are capitalized and restriction sites are underlined; *MFα*, α-factor pre-pro sequence.

according to the Invitrogen instruction manual (Invitrogen's *Pichia* expression kit manual, version M) using a Bio-Rad Gene Pulser. The *P. pastoris* GS115 recombinants were selected and confirmed as Mut$^+$ as previously described (Dagher *et al.*, 2013). The selected recombinants were grown in yeast extract–peptone–dextrose medium (YPD) at 26 °C for 24 h and then transferred to buffered glycerol complex medium (BMGY) and incubated further at 26 °C for 2 days. Methanol was then added daily at 0.5 % (v/v) every 24 h for 6 days. After this period of induction, the cells were harvested by centrifugation (5000 ***g*** at 4 °C), washed with 50 mM sodium phosphate (pH 5), and suspended at an OD_{600} of 100 in 50 mM sodium phosphate, pH 5, or 50 mM phosphate–citrate buffer (PC buffer at pH 4).

Detection, quantification and activity analysis of soluble and membrane-bound rBHT-HIS. Production of rBHT-HIS by recombinant strains was verified by analysing methanol-induced cultures and broth. Activity assays were performed using the artificial substrate ONP-Glu (Dagher *et al.*, 2013). Culture supernatants were collected by centrifugation and concentrated by ultrafiltration with an Amicon 15, molecular weight cut-off 30 000. The concentrated supernatants were washed by diluting 30-fold with PC buffer (pH 4) and then reconcentrated by two consecutive ultrafiltration steps. Protein cellular extracts were obtained after vigorous glass bead disruption of cells (OD_{600} of 100) in 200 µl of 1 × SDS Laemmli buffer. The solution was heated at 95 °C with periodic vortexing for 15 s six times followed by

Table 2. Strains and plasmids used in this study

Strain/plasmids	Description or genotype*	Source or reference
E. coli		
XL1-Blue	*recA1 endA1 gyrA*96 *thi-1 hsdR17 supE44 relA1 lac* [F′ *pro*AB *lacI*^{q}ZΔM15 Tn*10* (Tet^{R})]	Stratagene
P. pastoris		
GS115	*his4* (his^{-} mut^{+})	Invitrogen
JB210	GS115 : : *MFα-rBht*$_{(1\text{-}594)}$-HIS (his^{+} mut^{+})	This study
JB212	GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS (his^{+} mut^{+})	This study
JB213	GS115 : : *MFα-rBht*$_{(23\text{-}594)}$ (his^{+} mut^{+})	This study
JB214	GS115 : : *rBht*$_{(1\text{-}594)}$-HIS (his^{+} mut^{+})	This study
JB216	GS115 : : MFα-rBht$_{(111\text{-}594)}$-HIS (his^{+} mut^{+})	This study
JB215	GS115 : : *rBht*$_{(23\text{-}594)}$-HIS (his^{+} mut^{+})	This study
JB217	GS115 : : *MFα-scFv13R4*-HIS (his^{+} mut^{+})	This study
JB218	GS115 : : *MFα-rBht*$_{(1\text{-}22)}$-*scFv13R4*-HIS (his^{+} mut^{+})	This study
JB219	GS115 : : *rBht*$_{(1\text{-}22)}$-*scFv13R4*-HIS (his^{+} mut^{+})	This study
JB220	GS115 : : *rBht*$_{(1\text{-}110)}$-*scFv13R4*-HIS (his^{+} mut^{+})	This study
JB221	GS115 : : *rBht*$_{(23\text{-}110)}$-*scFv13R4*-HIS (his^{+} mut^{+})	This study
JB222	GS115 : : *scFv13R4*-HIS (his^{+} mut^{+})	This study
Plasmids *E. coli*		
pJB100	pGS21a-*rBht*$_{(1\text{-}594)}$	Dagher *et al.* (2013)
pPM163R4	pPM163 containing a mutant anti-β-galactosidase antibody gene *scFv13R4* with C-terminal myc and 6 × HIS	Martineau *et al.* (1998)
Plasmids *P. pastoris*		
pPIC9	*P. pastoris* integrative vector carrying *AOX1* promoter and transcription terminator, *HIS4*, Amp^{r} in *E. coli*, pBR322 ori, α-factor pre-pro leader from *Sac. cerevisiae* (*MFα*)	Invitrogen
pJB110	pPIC9-*MFα-rBht*$_{(1\text{-}594)}$-HIS	This study
pJB112	pPIC9-*MFα-rBht*$_{(23\text{-}594)}$-HIS	This study
pJB113	pPIC9-*MFα-rBht*$_{(23\text{-}594)}$	This study
pJB114	pPIC9-*rBht*$_{(1\text{-}594)}$-HIS	This study
pJB115	pPIC9-*rBht*$_{(23\text{-}594)}$-HIS	This study
pJB116	pPIC9-*MFα*-rBht$_{(111\text{-}594)}$-HIS	This study
pJB117	pPIC9-*MFα-scFv13R4*-HIS	This study
pJB118	pPIC9-*MFα-rBht*$_{(1\text{-}22)}$-*scFv13R4*-HIS	This study
pJB119	pPIC9-*rBht*$_{(1\text{-}22)}$-*scFv13R4*-HIS	This study
pJB120	pPIC9-*rBht*$_{(1\text{-}110)}$-*scFv13R4*-HIS	This study
pJB121	pPIC9-*rBht*$_{(23\text{-}110)}$-*scFv13R4*-HIS	This study
pJB122	pPIC9-*scFv13R4*-HIS	This study

**MFα*, *Sac. cerevisiae* α-factor pre-pro secretion leader found in pPIC9 vector.

removal of cellular debris by centrifugation at 14 000 ***g*** for 10 min at 25 °C. SDS-PAGE (8 % gel) and Western blotting using monoclonal anti-HIS antibody were performed as previously described (Dagher *et al.*, 2013).

Quantitative real time PCR (qRT-PCR). qRT-PCR was carried out with a Fast Real-time 7500 device (Applied Biosystems). The final reaction volume was 20 µl and contained: each primer at a final concentration of 200 nM, 1 × Power SYBR Green (Applied Biosystems) and 2 µl of template (cDNA concentration 20 ng µl^{-1}). Samples and standards were run in triplicate. The following primers were used: *GAPDH* Forward Primer, 5′-CGGTTTCGGACGTATTGGAC-3′; *GAPDH* Reverse Primer, 5′-CTGGAGCAATGAATGGGTCG-3′; r*BHT*$_{(110\text{-}594)}$ Forward Primer, 5′-CCATCCGGGTTTGCTATTGG-3′; r*BHT*$_{(110\text{-}594)}$ Reverse Primer, 5′-TTCAGCCCAACCAAATTCCG-3′; *ScFV13R4* Forward Primer, 5′-CTGACAATCTCTGGGCTCCA-3′; *ScFV13R4* Reverse Primer, 5′-GGCCCCATTCAGATCCTCTT-3′.

The thermal cycling conditions were as follows: one cycle at 50 °C for 20 s and 95 °C for 10 min, followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C. Melting curve analysis was carried out using the continuous method from the 7500 Software (Applied Biosystems) conducted at 60 °C, with increments of 1 °C every 15 s. Data analysis was carried out with 7500 Software (Applied Biosystems). The auto threshold and baseline options were used for the calculations of C_t values per well. The linear equation for the standard curve (i.e. for preparations containing known quantities of DNA) was then used to interpolate the numbers of copies present in unknown samples.

Purification of rBHT-HIS and rBHT. Secreted soluble rBHT-HIS was recovered from the culture medium under native conditions using nickel affinity chromatography according to the manufacturer's instructions (Qiagen). Fractions showing the highest enzymic activity

were desalted using Amersham PD10 G25 columns (Amersham Biosciences 52-1308-00 Edition AO), pooled and concentrated as described above using ultrafiltration. Protein concentrations were determined as described by Bradford (1976) using BSA as the standard. The purity of the protein was assessed by SDS-PAGE and silver staining according the Bio-Rad manual (Bio-Rad's Mini-Protean II Electrophoresis Cell Instruction manual). Non-tagged proteins were purified as described previously (Dagher *et al.*, 2013).

Kinetic studies for rBHT-HIS. Enzyme kinetics and optimum enzymic activity were determined as previously described using the artificial substrate ONP-Glu (Dagher *et al.*, 2013). Kinetic constants were determined from the initial rates of hydrolysis at eight different concentrations (10.4, 5.2, 2.6, 1.3, 0.65, 0.325, 0.1625 and 0.08125 mM) of ONP-Glu at 42 °C in 50 mM PC buffer at pH 4 using 0.5 U rBHT ml^{-1} in an assay volume of 0.25 ml. Duplicate reactions were stopped at 1 min intervals with 0.25 ml of 0.25 M sodium carbonate and the absorbance measured at OD_{405}. Glucose release was quantified by using a *p*-nitrophenol standard curve. An enzyme unit (U) was defined as the amount of enzyme required to release 1 μmol of *p*-nitrophenol min^{-1} under the assay conditions. The molar extinction coefficient at pH 4 of *p*-nitrophenol was $\varepsilon = 0.033\ mM^{-1}\ cm^{-1}$. The V_{max} and K_m were determined by fitting the Michaelis–Menten equation to the initial rates. Plotting and fitting were performed as previously described (Dagher *et al.*, 2013).

N-terminal sequencing analysis. Nickel-purified protein was resolved by SDS-PAGE (8 %), electroblotted to a PVDF membrane (Bio-Rad) and stained with Coomassie blue G-250 (Bio-Rad). The bands were excised for sequencing at the Tufts core facility.

GOS synthesis and analysis using soluble and membrane-bound HIS-tagged and non-tagged rBHT. GOS was generated as previously described (Dagher *et al.*, 2013). The transgalactosylation reactions were performed in mixtures containing lactose as indicated in the figure legends. Briefly, the products of the reaction catalysed by rBHT were analysed by HPLC (Shimadzu) using an Alltech IOA-1000 organic acids column (300 × 7.8 mm) coupled to a refractive-index detector. Isocratic conditions were set at 65 °C, 0.4 ml min^{-1} flow rate with 5 mM sulfuric acid as the mobile phase. Standard globotriose (galactosyl-lactose) was obtained from Carbosynth, and lactose, glucose and galactose standards were purchased from Sigma.

Nucleotide sequence accession number. The nucleotide sequence of the *rBht* gene has been previously deposited in the GenBank database under accession number JF29828.

RESULTS

In silico analysis of the BHT

We previously reported the heterologous expression of a bioactive full-length polypeptide (rBHT) by a recombinant strain of *P. pastoris* (GS115 : : *MF*α-HIS-TEV-*rBht*$_{(1\text{-}594)}$) (Dagher *et al.*, 2013). This strain secreted a limited amount of soluble rBHT, which was recovered using traditional purification methods as the N-terminal 6XHIS tag was not detected by Western blot using anti-HIS antibody, and could not be purified using nickel affinity chromatography.

As the secretion/translocation efficiency of rBHT by *P. pastoris* may be influenced by structural elements determining protein solubility or cell-wall association, we performed an *in silico* analysis of the BHT sequence (594 aa). The analysis revealed a novel N-terminal region spanning 110 aa. This region was composed of three putative regions: an N-terminal classical leader domain (amino acids 1–22) followed by a non-classical signal (amino acids 23–75) (Bendtsen *et al.*, 2004) and a region of low complexity (amino acids 72–83) (Wootton, 1994). The C-terminal portion (amino acids 111–594) of the BHT polypeptide had noticeable homology to β-glucosidases as determined by the SMART program (Letunic *et al.*, 2012). This glycosyl hydrolase family I (GH1) region contains a putative catalytic acid/base WFTFNEP (amino acids 270–276) domain followed by a catalytic nucleophile FSEFG domain (amino acids 494–498), and three asparagine residues potentially required for protein *N*-glycosylation (Gupta & Brunak, 2002) (Fig. 1a).

The N-terminal leader domain (amino acids 1–22) could be further divided into the N-region (amino-terminal; amino acids 1–5), H-region (hydrophobic; amino acids 6–17) and C-region (carboxy-terminal; amino acids 18–22) (Fig. 1b). The Phobius web-based program (http://phobius.sbc.su.se/) and SignalP (Petersen *et al.*, 2011) also predicted a classical leader domain and potential cleavage sites between residues 17 and 18 and between 22 and 23. Furthermore, the classical leader domain was predicted to contain five amino acids that contact the membrane within the H-region (Punta *et al.*, 2007) and a charge distribution that may determine the confinement of membrane proteins (Boyd & Beckwith, 1990). These structural features may also indicate that the BHT N-terminal classical leader domain might act as a membrane anchor during protein secretion. The transmembrane region prediction algorithm (http://www.ch.embnet.org/software/TMPRED_form.html) also forecasted stretches of hydrophobic residues from amino acids 1 to 17 and 177 to 199 (typical for integral membrane-spanning proteins) and a non-cytoplasmic region (amino acids 23–594). The regions between residues 3 and 17 and 72 and 83 were predicted to be regions of low complexity by the SEG algorithm within SMART (Fig. 1c). Regions of low complexity often form separate domains within a multidomain protein, rarely have a defined 3D structure and usually coincide with linker sequences (Coeytaux & Poupon, 2005).

The N-terminal domains participate in protein secretion

To investigate the physiological roles of the classical leader and non-classical signal, we examined the influence of each domain on secretion of the C-terminal BHT domain (amino acids 111–594) and the single-chain anti-β-galactosidase antibody (scFv13R4). Recombinant *P. pastoris* strains were generated by chromosomal integration of the appropriate modified gene combinations preceded by the *rBht* N-terminal domains and/or the 9.3 kDa MFα pre-pro sequence. The different versions of *rBht* and the chimeric *scFv13R4* gene combinations were inserted

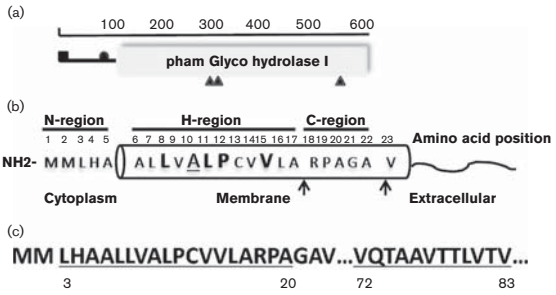

Fig. 1. *In silico* analysis of the BHT sequence. (a) Schematic representation of the BHT polypeptide from *Spo. singularis* determined by the Web-based SMART program (http://smart.embl-heidelberg.de). The leader peptide, represented by a solid square (■), was determined by the SignalP program (http://www.cbs.dtu.dk/services/SignalP/). The segment of low compositional complexity (http://www.cbs.dtu.dk/services/SecretomeP), represented by a solid circle (●), was also determined by the SEG algorithm (http://mendel.imp.ac.at/METHODS/seg.server.html). Hits only found by BLAST are indicated by the glycosyl hydrolase domain (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Solid triangles (▲) indicate the positions of the three putative *N*-glycosylation sites determined by NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/). (b) Schematic representation of the leader peptide (amino acids 1–22) calculated by the RHYTHM transmembrane prediction method (http://proteinformatics.charite.de/rhythm/). The membrane α-helix (amino acids 6–23) was predicted by HMMTOP (http://www.enzim.hu/hmmtop/html/document.html). The amino acids that contact the membrane are in large bold type (amino acids 8, 11, 12 and 15) and the amino acid in contact with the helix (amino acid 10) is in large type and underlined. Also indicated are the positions of predicted cytoplasmic (amino acids 1–5) and extracellular (amino acids 23–594) regions (http://www.ch.embnet.org/software/TMPRED_form.html). Arrows (↑) indicate possible cleavage sites (http://www.cbs.dtu.dk/services/SignalP/ and http://phobius.sbc.su.se/). (c) Schematic representation of amino acids 1–110 containing segments of low compositional complexity (underlined) determined by the SEG program.

downstream of the AOX1 promoter and followed by the C-terminal 6 × HIS tag to assist in protein detection and purification. Additionally, the levels of transcription were analysed by qRT-PCR and no differences in expression levels of the transformants were observed between the individual *P. pastoris* recombinants (Fig. 2a, d).

Our results demonstrate that replacement of the leader domain (amino acids 1–22) with the MFα pre-pro leader (GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS) increased secretion of the soluble protein 20-fold (9.80 μg ml^{-1}) compared with the values shown when full-length rBHT-HIS was preceded by the MFα leader (GS115 : : *MFα-rBht*$_{(1\text{-}594)}$-HIS) (0.49 μg ml^{-1}) (Fig. 2b). Similarly, in the absence of MFα, the leader domain was able to direct secretion of soluble protein (GS115 : : *rBht*$_{(1\text{-}594)}$-HIS) (6.35 μg ml^{-1}). Therefore, relying on its natural leader domain or by substitution, the membrane-bound rBHT can be overexpressed in a soluble form by *P. pastoris*, while leader interference was revealed when both MFα pre-pro leader and BHT leader domains were present (GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS) (Fig. 2b). Additionally, soluble protein was detected in the broth in the absence of both MFα and leader domains (GS115 : : *rBht*$_{(23\text{-}594)}$-HIS) (4.65 μg ml^{-1}), suggesting that both the classical leader and the non-classical signal contain information targeting the protein for secretion. This was also corroborated by the finding that substitution of the 110aa N-terminal region with MFα (GS115 : : *MFα-rBht*$_{(111\text{-}594)}$-HIS) gave no secretion of soluble protein (Fig. 2b).

To confirm our results, we selected the antibody scFv13R4, a cytoplasmic protein devoid of leader sequences. Diagrams of the antibody scFv13R4 chimeras are shown in Fig. 2(d). When scFv13R4-HIS was expressed by GS115 : : *scFv13R4*-HIS (without a leader domain), it was not detected in the culture broth by SDS-PAGE silver staining or Western blot analysis (data not shown). Secreted levels of soluble scFv13R4-HIS by GS115 : : *rBht*$_{(1\text{-}22)}$-*scFv13R4*-HIS, GS115 : : *rBht*$_{(1\text{-}110)}$-*scFv13R4*-HIS and GS115 : : *rBht*$_{(23\text{-}110)}$-*scFv13R4*-HIS were 16.48, 25.17 and 7.03 μg ml^{-1}, respectively. Preliminary analysis of scFv13R4 chimeras indicated a lack of membrane association (data not shown). The larger amounts of secreted soluble scFv13R4 compared with rBHT may be attributed to the fact that scFv13R4 is a cytoplasmic soluble protein. Likewise, as observed with rBHT$_{(23\text{-}594)}$, secretion driven by MFα (GS115 : : *MFα-scFv13R4*-HIS) provided the highest level of soluble protein (91.02 μg ml^{-1}) while the presence of both leader domains MFα and Bht$_{(1\text{-}22)}$ (GS115 : : *MFα-rBht*$_{(1\text{-}22)}$-*scFv13R4*-HIS) interfered with secretion of soluble protein, thus reducing the levels of protein detected (20.15 μg ml^{-1}). This demonstrated that the leader domain (BHT$_{(1\text{-}22)}$) is involved in the signal peptide-mediated mechanism (classical secretory pathway).

Enzyme activity analysis of the secreted rBHT-HIS

To confirm that the amounts of secreted protein correlated with the enzymic activity, soluble and membrane-bound rBHT-HIS activities were quantified (Fig. 2 and Table 3). The soluble protein secreted by GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS displayed an enzymic activity of 3.7 mU OD^{-1}, approximately sixfold higher than the activity observed when secretion was only driven by the complete N-terminal region (amino acids 1–110) [GS115 : : *rBht*$_{(1\text{-}594)}$-HIS (0.63 mU OD^{-1})] and 53-fold higher than that obtained from the recombinant containing both MFα and *rBht*$_{(1\text{-}110)}$ domains [GS115 : : *MFα-rBht*$_{(1\text{-}594)}$-HIS (0.07 mU OD^{-1})]. Additionally, the recombinant GS115 : : *rBht*$_{(23\text{-}594)}$-HIS (0.26 mU OD^{-1}) secreted a reduced amount of active soluble enzyme (Table 3).

Membrane-bound enzymic activity was tested using resting cells. A 14-fold increase in activity was obtained for

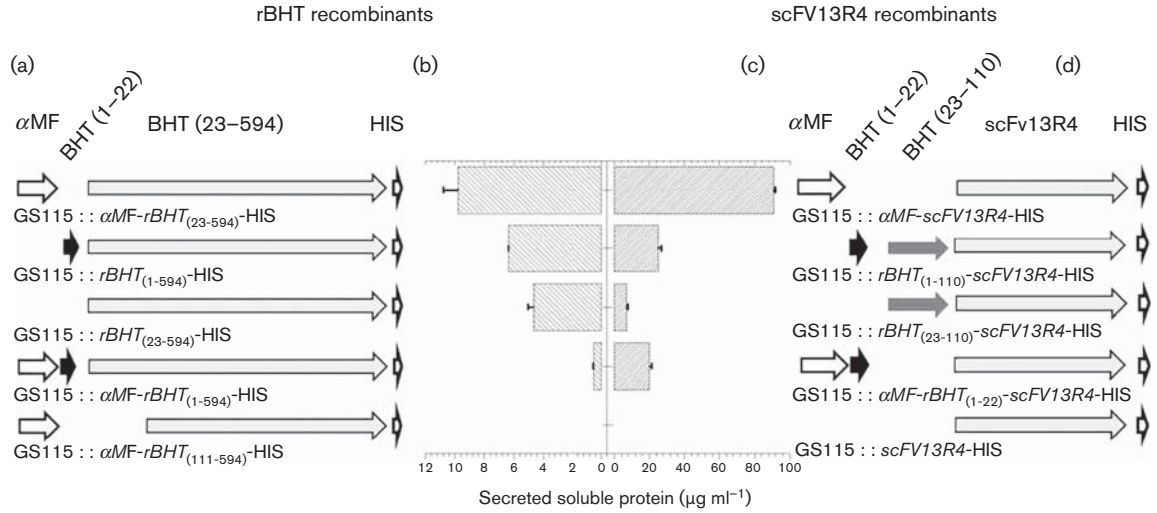

Fig. 2. Comparisons of amounts of secreted soluble proteins (normalized for the final culture OD_{600}) generated by recombinant strains of *P. pastoris* carrying *rBht*-HIS or *scFv13R4*-HIS under the control of AOX1 (b, c). (a, d) Graphic representations of chimeric genes containing combinations of leader domains and ORFs of *rBht* or *scFv13R4*. Protein secretion of the following recombinant strains was compared: (a, b) row 1, GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS; row 2, GS115 : : *rBht*$_{(1\text{-}594)}$-HIS; row 3, GS115 : : *rBht*$_{(23\text{-}594)}$-HIS; row 4, GS115 : : *MFα-rBht*$_{(1\text{-}594)}$-HIS; row 5 GS115 : : *MFα-rBht*$_{(111\text{-}594)}$-HIS. (c, d) Row 1, GS115 : : *MFα*-*scFv13R4*-HIS; row 2, GS115 : : *rBht*$_{(1\text{-}110)}$-*scFv13R4*-HIS; row 3, GS115 : : *rBht*$_{(23\text{-}110)}$-*scFv13R4*-HIS; row 4, GS115 : : *MFα-rBht*$_{(1\text{-}22)}$-*scFv13R4*-HIS; row 5, GS115 : : *scFv13R4*-HIS.

strain GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS (21.52 mU OD^{-1}) and a 1.3-fold increase was observed for strain GS115 : : *rBht*$_{(1\text{-}594)}$-HIS (1.94 mU OD^{-1}) compared with GS115 : : *MFα-rBht*$_{(1\text{-}594)}$-HIS (1.48 mU OD^{-1}). The recombinant GS115 : : *rBht*$_{(23\text{-}594)}$-HIS (0.15 mU OD^{-1}) showed a reduced amount of membrane-bound enzyme, suggesting that this recombinant strain may redirect the protein through a putative non-classical secretion pathway. Moreover, the strain carrying the substitution of the 110 aa N-terminal region by MFα (GS115 : : *MFα-rBht*$_{(111\text{-}594)}$-HIS) showed no soluble or membrane-associated activity. Overall, these results show that both leader sequences were able to increase the amounts of soluble and membrane-bound enzymes. Neither MFα nor the *rBht*$_{(1\text{-}22)}$ leader independently could spur the complete release of the membrane rBHT-HIS fraction, almost certainly due to the presence of the predicted transmembrane region between amino acids 177 and 199, which limits protein mobility.

Western blot analysis of the secreted rBHT-HIS

Western blot analysis using anti-HIS antibody confirmed the presence of soluble protein generated by GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS, GS115 : : *rBht*$_{(1\text{-}594)}$-HIS and GS11 5 : : *rBht*$_{(23\text{-}594)}$-HIS (Fig. 3a). In each case, a prominent rBHT-HIS band corresponding to a molecular mass of approximately 110 kDa was observed. These results are in agreement with previously reported SDS-PAGE data and

Table 3. Enzyme activity (mU OD^{-1}, mean ± sd) of soluble and membrane-bound rBHT-HIS generated by *P. pastoris* recombinant strains

The final cell density (OD_{600}) reached by the recombinant strains after methanol induction was used to normalize the enzymic activities of soluble and membrane-bound BHT. The maximum cell densities obtained were between 54 and 78 OD_{600}. The results are mean values from three measurements.

Enzyme source	Soluble	Membrane-bound	Ratio soluble/membrane-bound
GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS	3.70 ± 0.063	21.52 ± 1.38	0.172
GS115 : : *rBht*$_{(1\text{-}594)}$-HIS	0.63 ± 0.018	1.94 ± 0.02	0.325
GS115 : : *rBht*$_{(23\text{-}594)}$-HIS	0.26 ± 0.003	0.15 ± 0.02	1.606
GS115 : : *MFα-rBht*$_{(1\text{-}594)}$-HIS	0.07 ± 0.003	1.48 ± 0.02	0.046

size exclusion chromatography migration patterns (Dagher *et al.*, 2013).

Western blot analysis of cell extracts obtained from GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS, GS115 : : *rBht*$_{(1\text{-}594)}$-HIS and GS115 : : *MFα-rBht*$_{(1\text{-}594)}$-HIS revealed a molecular mass slightly above 110 kDa, which may indicate the presence of pre-cleaved enzyme, while GS115 : : *rBht*$_{(23\text{-}594)}$-HIS showed prominent bands between 98 and 64 kDa that may indicate intracellular protein degradation or alternative glycosylation patterns (Fig. 3b). No band was observed for the strain carrying the substitution of the 110aa N-terminal region with MFα (GS115 : : *MFα-rBht*$_{(111\text{-}594)}$-HIS).

HIS-tagged and non-tagged rBHT activity

Additionally, we purified and tested the soluble HIS-tagged and non-tagged enzymes from GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS and GS115 : : *MFα-rBht*$_{(23\text{-}594)}$. The enzymes delivered comparable results and were active over a wide range of temperatures (10–50 °C) and pH values (2.8–6). Maximum activity was observed between pH 3.6 and 5 (91–100 % of maximum activities) followed by a steady decrease down to pH 2.6 (43 % of maximum activity) and up to pH 6.8 (29 % of maximum activity). Likewise, the enzymes' optimum temperature was in the range 40–45 °C (97–100 % maximum activities) but rapidly decreased at temperatures above 50 °C and below 20 °C (<25 % of maximum activity) (data not shown). The enzymes were stable in 50 mM sodium phosphate buffer (pH 5) at 4 °C for at least 6 months and activity was unaffected by storage at −80 °C. The values for the kinetic constants for the enzyme secreted by GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS were obtained from the Hill equation [K_m 0.79 mM and V_{max} 3.97 mmol min^{-1} (mg enzyme)$^{-1}$ at 42 °C, pH 4]. These findings were in agreement with previous reports by our group and others (Gorin *et al.*, 1964a, b; Blakely & Mackenzie, 1969; Shin *et al.*, 1998; Shin & Yang, 1998; Cho *et al.*, 2003; Ishikawa *et al.*, 2005; Sakai *et al.*, 2008; Dagher *et al.*, 2013).

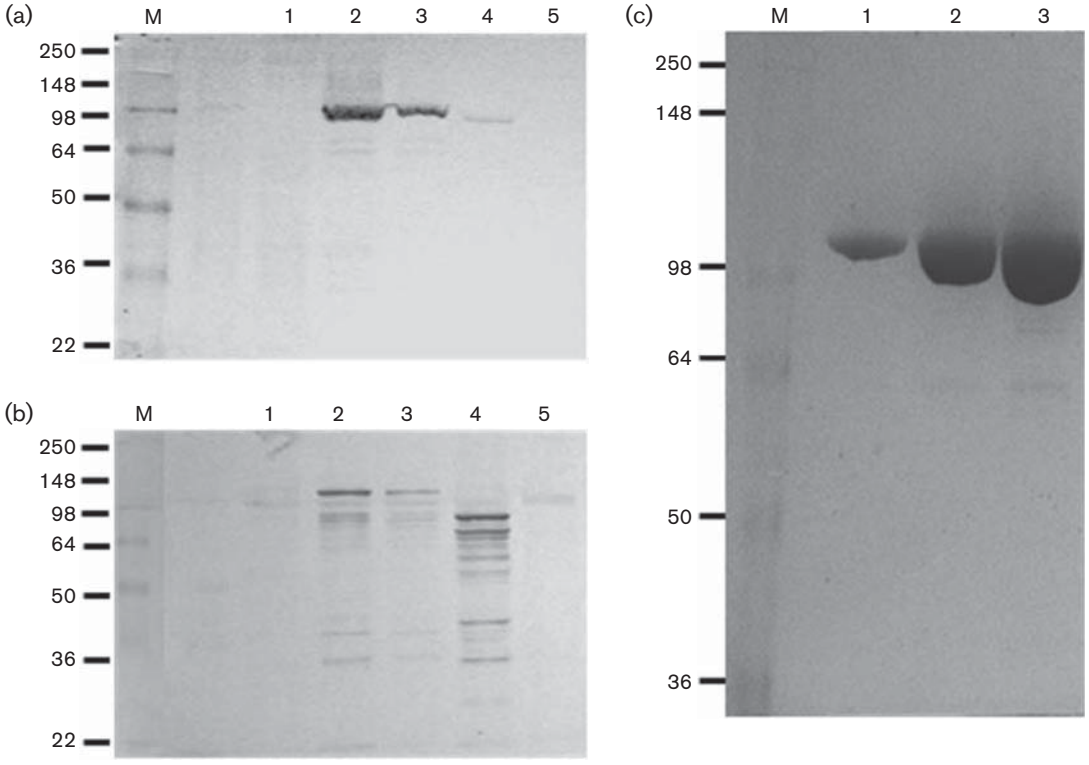

Fig. 3. Western blots exposed with an anti-HIS antiserum after separation by SDS-PAGE (8 %). The figures show soluble, cell-associated, and purified soluble rBHT-HIS secreted by different recombinants of *P. pastoris* GS115. (a) Secreted soluble proteins generated by all recombinants were concentrated 20-fold. (b) Cell-associated proteins were obtained from five OD_{600} cultures disrupted with glass beads in 1 × Laemmli buffer. Lane 1, GS115 : : *MFα-rBht*$_{(1\text{-}594)}$-HIS; lane 2, GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS; lane 3, GS115 : : *rBht*$_{(1\text{-}594)}$-HIS; lane 4, GS115 : : *rBht*$_{(23\text{-}594)}$-HIS; lane 5, GS115 control. (c) Silver stain of increasing concentrations of soluble rBHT-HIS secreted by GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS and purified using nickel affinity chromatography and resolved by SDS-PAGE (8 %). Lane 1, 5 μg; lane 2, 20 μg; lane 3, 50 μg; M, lane containing the molecular mass protein markers. The molecular masses (kDa) of the marker proteins are shown to the left of the panels.

rBHT stability

To examine the long-term stability of the enzyme, induced resting cells were incubated in buffer containing 2 % glucose and the enzymic activity of membrane-bound and soluble rBHT was measured over time. Soluble rBHT-HIS obtained from all recombinants through the classical or the non-classical secretion pathways remained stable over a 1 week testing period and retained more than 95 % of its initial activity. The same stability was observed in resting cells (GS115 : : *MFα-rBht*$_{(1\text{-}594)}$-HIS, GS115 : : *rBht*$_{(1\text{-}594)}$-HIS and GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS) containing the membrane-associated enzyme. However, resting cells of GS115 : : *rBht*$_{(23\text{-}594)}$-HIS showed a reduced enzymic stability, with a decrease in membrane-bound activity within 24 h, suggesting secretion of the enzyme through a non-classical pathway.

Purification and characterization of rBHT-HIS generated by GS115 : : *MFα*-rBht$_{(23\text{-}594)}$-HIS

The secreted soluble rBHT protein generated by GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS was purified using nickel affinity chromatography. Placement of the 6XHIS tag on the C terminus successfully allowed the single-step recovery of more than 73 % of the original enzymic activity and a single polypeptide band of approximately 110 kDa was observed (Fig. 3c). After protein purification from the culture supernatant, 7.24 mg of enzyme was recovered, rendering a specific activity of 18.45 U mg^{-1} at 42 °C and pH 4 (Table 4). Moreover, following the same methodology, we purified the rBHT-HIS secreted by the different recombinant strains and found comparable specific activities for purified HIS-tagged and non-tagged rBHT ranging from 18.45 to 18.65 U mg^{-1}. Determination of the N-terminal sequences of the polypeptide secreted by GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS showed that the entire rBHT$_{(23\text{-}594)}$-HIS protein was present in the broth (VTYPG residues, cleavage at amino acid 22) as well as a product containing two additional N-terminal amino acids (EAVTY residues). Variability in the cleavage of C-terminal MFα amino acids A–E during secretion can be affected by the surrounding amino acid sequence and the tertiary structure (Cereghino & Cregg, 2000). The remaining non-classical sequence did not introduce a new cleavage site.

HIS-tag impact on rBHT transferase activity

Soluble proteins generated by the highest-producing strains (GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS and GS115 : : *MFα-rBht*$_{(23\text{-}594)}$) were evaluated to determine if the presence of the HIS-tag may impact GOS synthesis. GOS accumulation was analysed quantitatively by HPLC from reaction mixtures containing an initial lactose contentration of 220 g l^{-1} and 0.5 U rBHT (g lactose)$^{-1}$ at 30 °C. Fig. 4(a) shows equivalent kinetics of GOS accumulation and lactose consumption when the reaction was catalysed by either the HIS-tagged or non-tagged soluble enzymes. In both cases, the maximum GOS production rate was observed during the first 25 h and galactosyl-lactose was the main product. Enzyme competitive glucose inhibition was observed after 125 h and galactosyl-lactose (75 g l^{-1}) accumulation reached an average of 67 % conversion of the 60 % initial lactose utilized. In addition, experiments using resting cells of GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS and GS115 : : *MFα-rBht*$_{(23\text{-}594)}$ harbouring membrane-associated HIS-tagged and non-tagged enzyme showed identical performance of GOS accumulation and lactose consumption over time. Thus, presence of the C-terminal HIS-tag had no impact on the initial reaction rate of galactosyl-lactose formation (1.87 and 1.7 g l^{-1} h^{-1}) (Fig. 4b). As previously reported, glucose was consumed by the resting cells while galactose was used to synthesize GOS [(68 % yield (w/w)], approaching a theoretical maximum yield of 75 % (Dagher *et al.*, 2013).

DISCUSSION

Prebiotics are carbohydrate derivatives marketed as functional foods. They are actively promoted to improve consumer health, and are intended to specifically stimulate the

Table 4. Purification of soluble rBHT-HIS secreted by *P. pastoris* GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS

Enzyme source	Total activity in medium (U l^{-1})*	Total protein in medium (mg)†	Specific activity in medium (U mg^{-1})‡	Ni column (U)§	Ni column (mg)‖	Specific activity Ni column (U mg^{-1})†	Purification (fold)	Recovery (%)¶
GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS	180.67	64.00	2.82	133.54	7.24	18.45	6.54	73.91

*One litre of culture was grown in BMGY broth at 28 °C.

†Protein concentration determined by Bradford assay.

‡Specific activity expressed as the total activity (U) divided by the total soluble protein (mg).

§Total units recovered following nickel chromatography purification.

‖Total soluble protein recovered (mg) following nickel chromatography purification.

¶Yield of total activity recovered following nickel column chromatography divided by total secreted soluble activity.

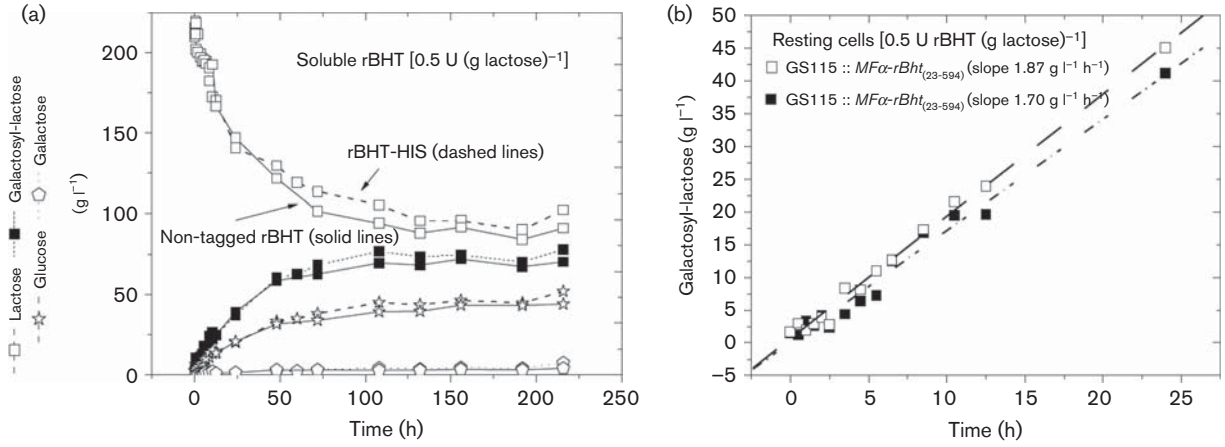

Fig. 4. Enzymic reactions catalysed by HIS-tagged and non-tagged rBHT. The figure shows time-course studies of galactosyl-lactose synthesis using purified soluble protein or resting cells containing membrane-bound enzyme [0.5 U rBHT (g lactose)$^{-1}$]. (a) Synthesis by purified soluble rBHT-HIS secreted by GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS (solid lines) and purified soluble rBHT produced by GS115 : : *MFα-rBht*$_{(23\text{-}594)}$ (dashed lines). (b) The rate of galactosyl-lactose synthesis by resting cells GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS (■) and GS115 : : *MFα-rBht*$_{(23\text{-}594)}$ (□). Assays contained 200 g lactose l^{-1} and either purified soluble enzyme or resting cells of *P. pastoris* in 5 mM sodium phosphate buffer (pH 5.0) and were incubated at 30 °C. Samples were removed periodically and analysed by HPLC. Data represent the means of two independent experiments.

growth of beneficial bacteria in the gut. The fundamental force that drives prebiotic production at industrial scale is the promise of more efficient production processes at lower operating costs. Therefore, the development of enzymic approaches is of practical interest, and genetic modification has been extensively used to modify enzymic activity, to gain deeper understanding of catalytic mechanisms and to increase protein secretion. This is especially important as synthesis of specific carbohydrate derivatives at industrial scale by chemical methods is complex and requires protection and deprotection steps due to the presence of several hydroxyl groups of similar reactivity (Sears & Wong, 2001).

To study the BHT protein domains we chose the *P. pastoris* expression system. This organism is a well-known model used for secretion of soluble proteins; this process is influenced by the nucleotide sequence and occasionally requires codon optimization, as well as consideration of glycosylation patterns, final 3D structure, culture conditions and medium composition (Damasceno *et al.*, 2012). However, membrane proteins have been traditionally difficult to express in a soluble form by *P. pastoris* compared with their soluble counterparts and trial-and-error is still the most common approach taken to successfully over-express them. As we previously reported, the membrane-bound BHT was unexpectedly secreted in minor amounts as soluble protein by *P. pastoris* (Dagher *et al.*, 2013). *In silico* analyses confirmed that this enzyme contained transmembrane domains and a novel N-terminal region whose impact on secretion of soluble enzyme by *P. pastoris* necessitated further investigation. Membrane-bound and soluble proteins destined for secretion are usually preceded by N-terminal leader domains of 20–30 aa and eventually processed by membrane-bound peptidases (von Heijne, 1983). The BHT N-terminal region (amino acids 1–110) lacked significant homology beyond a limited region [30 % identity from amino acids 36 to 82 with a hypothetical protein (*Lachnospiraceae* bacterium V9D3004)]. In contrast, the C-terminal domain (Fig. 1a) was 47 % identical to the corresponding region of the glycoside hydrolase family 1 proteins. The novel N-terminal region contained two domains predicted to function as a classical leader domain (BHT$_{(1\text{-}22)}$) and a non-classical signal (BHT$_{(23\text{-}110)}$), allowing BHT to perform its function at the cellular membrane (as predicted by RHYTHM; Fig. 1). Additionally, the distinctive distribution of charged amino acids within the leader domain plays an important role in facilitating protein localization and further defining orientation (Boyd & Beckwith, 1990).

Therefore, this report builds on the above-mentioned unexpected evidence showing that the membrane-bound rBHT can be produced as a soluble active enzyme by *P. pastoris*. Furthermore, we present functional evidence to demonstrate that the original protein contains a leader domain (BHT$_{(1\text{-}22)}$) involved in the signal peptide-mediated mechanism (classical secretory pathway). Both leader domains (BHT$_{(1\text{-}22)}$ and *MFα*) were individually able to increase protein secretion of membrane-associated and soluble bioactive rBHT. Greatest secretion of soluble (53-fold increase) and membrane-bound (14-fold increase) rBHT was by the recombinant GS115 : : *MFα-rBht*$_{(23\text{-}594)}$-HIS, along with improved classical

secretion ratio values (soluble/membrane) and amount of total secreted protein. Additionally, the data revealed a significant reduction in enzyme secretion following removal of both the leader domains ($BHT_{(1-22)}$) and MFα. Expression of this truncated gene still generated low levels of secreted soluble protein, which confirmed that the unique 110aa N-terminal region contains a dual function by which the leader domain $BHT_{(1-22)}$ acts as an efficient secretion leader (classical secretion pathway) and the predicted domain within $BHT_{(23-110)}$ may operate as a less efficient alternative secretion signal (non-classical secretion pathway). Western blot analyses of intracellular $BHT_{(23-110)}$ confirmed the presence of multiple bands below the maximal mass of 110 kDa, suggesting increased sensitivity to proteolysis during the cytoplasmic transit (Fig. 3). Therefore, the leader sequences may protect the protein by redirecting the protein to the endoplasmic reticulum, which keeps the protein away from proteases found outside the secretory pathway.

To confirm the physiological function of these N-terminal domains, we generated new chimeras by adding rBHT domains to the antibody scFv13R4 protein. The scFv13R4 antibody is an example of a non-secreted, hyper-stable, single-chain protein that is independent of disulfide bridge formation for binding activity (Martineau *et al.*, 1998). As such, scFv13R4 has been heterologously produced in *E. coli*, *Sac. cerevisiae* and Chinese hamster ovary cells (Visintin *et al.*, 1999; Grage & Rehm, 2008; Bach *et al.*, 2001). The $BHT_{(1-110)}$ containing the classical leader followed by putative non-classical signal, the $BHT_{(1-22)}$ enclosing the putative classical leader, the $BHT_{(23-110)}$ enclosing the putative non-classical signal and the MFα leader were placed in-frame at the amino-terminal position followed by scFv13R4. Analyses of these new recombinants confirmed that the leader domains function as a secretory signal by directing the extracellular secretion of soluble antibody by all of the BHT fusion partners.

Purification and analysis of the secreted bioactive soluble proteins followed by SDS-PAGE separation, and subsequent silver staining, demonstrated that most are present as a single band. The molecular mass of rBHT was approximately 110 kDa (confirmed by size exclusion chromatography) and the specific activity of 18.45 U mg^{-1} did not deviate between soluble rBHT enzymes in this study. The enzyme displayed similar enzyme activity, thermostability, reusability and storage stability (Dagher *et al.*, 2013).

Our results confirm that leader sequences have a strong impact on secretion levels of soluble rBHT and scFv13R4, as is probably the case for other proteins. Notably, the new leader domain (22 aa) has been shown herein to be a new sequence capable of directing secretion of heterologous proteins. This leader domain adds a new feature that can be built into intracellular enzymes that otherwise need to be extracted by cellular disruption using mechanical means or through permeabilization with chemical treatments (Panesar *et al.*, 2006). More importantly, this work will allow for future structural and functional analyses to identify features that contribute to transglycosylation activity and substrate specificity.

ACKNOWLEDGEMENTS

A patent application describing the heterologous expression and use of soluble rBHT has been submitted. Special thanks for the qRT-PCR evaluations go to the UNC Microbiome Core Facility (supported in part by the National Institutes of Health grant number NIH P30 DK34987). We thank Alicia Cox for her technical help and Dr M. Andrea Azcárate-Peril for critical reading of the manuscript.

REFERENCES

Bach, H., Mazor, Y., Shaky, S., Shoham-Lev, A., Berdichevsky, Y., Gutnick, D. L. & Benhar, I. (2001). *Escherichia coli* maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies. *J Mol Biol* **312**, 79–93.

Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G. & Brunak, S. (2004). Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel* **17**, 349–356.

Blakely, J. A. & MacKenzie, S. L. (1969). Purification and properties of a β-hexosidase from *Sporobolomyces singularis*. *Can J Biochem* **47**, 1021–1025.

Boyd, D. & Beckwith, J. (1990). The role of charged amino acids in the localization of secreted and membrane proteins. *Cell* **62**, 1031–1033.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal Biochem* **72**, 248–254.

Cereghino, J. L. & Cregg, J. M. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol Rev* **24**, 45–66.

Cho, Y. J., Shin, H. J. & Bucke, C. (2003). Purification and biochemical properties of a galactooligosaccharide producing β-galactosidase from *Bullera singularis*. *Biotechnol Lett* **25**, 2107–2111.

Coeytaux, K. & Poupon, A. (2005). Prediction of unfolded segments in a protein sequence based on amino acid composition. *Bioinformatics* **21**, 1891–1900.

Dagher, S. F., Azcarate-Peril, M. A. & Bruno-Bárcena, J. M. (2013). Heterologous expression of a bioactive β-hexosyltransferase, an enzyme producer of prebiotics, from *Sporobolomyces singularis*. *Appl Environ Microbiol* **79**, 1241–1249.

Damasceno, L. M., Huang, C. J. & Batt, C. A. (2012). Protein secretion in *Pichia pastoris* and advances in protein production. *Appl Microbiol Biotechnol* **93**, 31–39.

Gorin, P. A. J., Phaff, H. J. & Spencer, J. F. T. (1964a). The structures of galactosyl-lactose and galactobiosyl-lactose produced from lactose by *Sporobolomyces singularis*. *Can J Chem* **42**, 1341–1344.

Gorin, P. A. J., Spencer, J. F. T. & Phaff, H. J. (1964b). The synthesis of β-galacto-β-gluco-pyranosyl disaccharides by *Sporobolomyces singularis*. *Can J Chem* **42**, 2307–2317.

Gosling, A., Stevens, G. W., Barber, A. R., Kentish, S. E. & Gras, S. L. (2010). Recent advances refining galactooligosaccharide production from lactose. *Food Chem* **121**, 307–318.

Grage, K. & Rehm, B. H. A. (2008). *In vivo* production of scFv-displaying biopolymer beads using a self-assembly-promoting fusion partner. *Bioconjug Chem* **19**, 254–262.

Gupta, R. & Brunak, S. (2002). Prediction of glycosylation across the human proteome and the correlation to protein function. In *Pacific Symposium on Biocomputing*, vol. 7, pp. 310–322. Edited by R. B. Altman, A. K. Dunker, L. Hunter, K. Lauderdale & T. E. Klein. Singapore: World Scientific Publishing.

Illanes, A. (2011). Whey upgrading by enzyme biocatalysis. *Electron J Biotechnol* **14**, 1–42.

Ishikawa, E., Sakai, T., Ikemura, H., Matsumoto, K. & Abe, H. (2005). Identification, cloning, and characterization of a *Sporobolomyces singularis* β-galactosidase-like enzyme involved in galacto-oligosaccharide production. *J Biosci Bioeng* **99**, 331–339.

Letunic, I., Doerks, T. & Bork, P. (2012). SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* **40** (D1), D302–D305.

Martineau, P., Jones, P. & Winter, G. (1998). Expression of an antibody fragment at high levels in the bacterial cytoplasm. *J Mol Biol* **280**, 117–127.

Panesar, P. S., Panesar, R., Singh, R. S., Kennedy, J. F. & Kumar, H. (2006). Microbial production, immobilization and applications of β-D-galactosidase. *J Chem Technol Biotechnol* **81**, 530–543.

Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**, 785–786.

Phaff, H. J. & do Carmo-Sousa, L. D. (1962). Four new species of yeast isolated from insect frass in bark of *Tsuga heterophylla* (Raf.) Sargent. *Antonie van Leeuwenhoek* **28**, 193–207.

Punta, M., Forrest, L. R., Bigelow, H., Kernytsky, A., Liu, J. & Rost, B. (2007). Membrane protein prediction methods. *Methods* **41**, 460–474.

Sakai, T., Tsuji, H., Shibata, S., Hayakawa, K. & Matsumoto, K. (2008). Repeated-batch production of galactooligosaccharides from lactose at high concentration by using alginate-immobilized cells of *Sporobolomyces singularis* YIT 10047. *J Gen Appl Microbiol* **54**, 285–293.

Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning: a Laboratory Manual*, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sears, P. & Wong, C. H. (2001). Toward automated synthesis of oligosaccharides and glycoproteins. *Science* **291**, 2344–2350.

Shin, H. J. & Yang, J. W. (1998). Enzymatic production of galactooligosaccharide by *Bullera singularis* β-galactosidase. *J Microbiol Biotechnol* **8**, 484–489.

Shin, H. J., Park, J. M. & Yang, J. W. (1998). Continuous production of galacto-oligosaccharides from lactose by *Bullera singularis* β-galactosidase immobilized in chitosan beads. *Process Biochem* **33**, 787–792.

Spencer, J. F. T., Ragout de Spencer, A. L. & Laluce, C. (2002). Non-conventional yeasts. *Appl Microbiol Biotechnol* **58**, 147–156.

Tzortzis, G. & Vulevic, J. (2009). Galacto-oligosaccharide prebiotics. In *Prebiotics and Probiotics Science and Technology*, pp. 207–244. Edited by D. Charalampopoulos & R. Rastall. New York: Springer.

Visintin, M., Tse, E., Axelson, H., Rabbitts, T. H. & Cattaneo, A. (1999). Selection of antibodies for intracellular function using a two-hybrid *in vivo* system. *Proc Natl Acad Sci U S A* **96**, 11723–11728.

von Heijne, G. (1983). Patterns of amino acids near signal-sequence cleavage sites. *Eur J Biochem* **133**, 17–21.

Wootton, J. C. (1994). Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput Chem* **18**, 269–285.

Edited by: M. Diethard