



CHALMERS
UNIVERSITY OF TECHNOLOGY

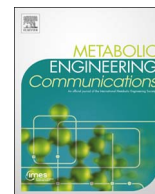
A novel chimaeric flocculation protein enhances flocculation in *Saccharomyces cerevisiae*

Downloaded from: <https://research.chalmers.se>, 2024-07-27 07:33 UTC

Citation for the original published paper (version of record):

Westman, J., Nyman, J., Manara, R. et al (2018). A novel chimaeric flocculation protein enhances flocculation in *Saccharomyces cerevisiae*. *Metabolic Engineering Communications*, 6: 49-55.
<http://dx.doi.org/10.1016/j.meten.2018.04.001>

N.B. When citing this work, cite the original published paper.



A novel chimaeric flocculation protein enhances flocculation in *Saccharomyces cerevisiae*



Johan O. Westman^{a,1}, Jonas Nyman^b, Richard M.A. Manara^b, Valeria Mapelli^{a,2}, Carl Johan Franzén^{a,*}

^a Chalmers University of Technology, Department of Biology and Biological Engineering, Division of Industrial Biotechnology, 412 96 Göteborg, Sweden

^b University of Southampton, Chemistry, University Road, Southampton SO17 1BJ, UK

ARTICLE INFO

Keywords:

Bioethanol

Protein engineering

Cell-cell adhesion

Chimaeric

Saccharomyces cerevisiae

ABSTRACT

Yeast flocculation is the reversible formation of multicellular complexes mediated by lectin-like cell wall proteins binding to neighbouring cells. Strong flocculation can improve the inhibitor tolerance and fermentation performance of yeast cells in second generation bioethanol production. The strength of flocculation increases with the size of the flocculation protein and is strain dependent. However, the large number of internal repeats in the sequence of *FLO1* from *Saccharomyces cerevisiae* S288c makes it difficult to recombinantly express the gene to its full length. In the search for novel flocculation genes resulting in strong flocculation, we discovered a DNA sequence, *FLONF*, that gives NewFlo phenotype flocculation in *S. cerevisiae* CEN.PK 113-7D. The nucleotide sequence of the internal repeats of *FLONF* differed from those of *FLO1*. We hypothesized that a chimaeric flocculation gene made up of a *FLO1* variant derived from *S. cerevisiae* S288c and additional repeats from *FLONF* from *S. cerevisiae* CCUG 53310 would be more stable and easier to amplify by PCR. The constructed gene, *FLOW*, had 22 internal repeats compared to 18 in *FLO1*. Expression of *FLOW* in otherwise non-flocculating strains led to strong flocculation. Despite the length of the gene, the cassette containing *FLOW* could be easily amplified and transformed into yeast strains of different genetic background, leading to strong flocculation in all cases tested. The developed gene can be used as a self-immobilization technique or to obtain rapidly sedimenting cells for application in *e.g.* sequential batches without need for centrifugation.

1. Introduction

Flocculation is a natural way for yeast cells to adhere to each other and increase the survival chance of the collective by formation of multicellular agglomerates. It is used mainly as a rapid and cheap way of separating cells from the broth in fermentation processes (Verstrepen et al., 2003). Additionally, we have shown that in the special case of fermentation of toxic lignocellulose hydrolysates, flocculation can increase the inhibitor tolerance of the cells and result in faster and more complete fermentation (Westman et al., 2014).

Flocculation is caused by lectin-like cell wall proteins, flocculins, that bind to mannoproteins in the cell wall of neighbouring cells (Goossens and Willaert, 2010). In *Saccharomyces cerevisiae*, the flocculation genes are situated in the sub-telomeric regions of the genome (Teunissen and Steensma, 1995). Furthermore, the flocculation genes contain tandem repeats in the middle of the genes. The

number of repeats determines the strength of the flocculation, and to some extent also the sugar specificity of the flocculation proteins (Liu et al., 2007; Verstrepen et al., 2005; Westman et al., 2014). Recombination events can easily alter the number of repeats and the flocculation characteristics can, therefore, change rapidly in response to *e.g.* environmental stress (Verstrepen et al., 2005). For strong flocculation of recombinant flocculating strains, a stable construct with a long flocculation gene, *i.e.* with many tandem repeats, is required. This has proven difficult to achieve due to the nature of the flocculation genes. In PCR and in amplification in *Escherichia coli*, the long sequences of internal repeats can lead to recombination events that shorten the gene by changing the length of the internal region (Teunissen et al., 1993; Westman et al., 2014). It has therefore been difficult to recombinantly express the entire *FLO1* gene (Watari et al., 1994; Westman et al., 2014).

An alternative to recombinant expression of the known *FLO1* is to

* Correspondence to: Department of Biology and Biological Engineering, Division of Industrial Biotechnology, Chalmers University of Technology, Kemivägen 10, 412 96 Gothenburg, Sweden.

E-mail address: franzen@chalmers.se (C.J. Franzén).

¹ Current address: Chr. Hansen A/S, Bøge Allé 10-12, DK-2970 Hørsholm, Denmark.

² Current address: Sacco S.r.l., Via A. Manzoni 29/A, 22071 Cadorago (CO), Italy.

search for novel flocculation genes in strongly flocculating *S. cerevisiae* strains (Zhao et al., 2011). Due to the rapid evolution of flocculation genes it is expected that certain strains can harbour genes resulting in a stronger and more robust flocculation than that obtained by *FLO1* from S288c (Verstrepen et al., 2005).

The aim of the current work was to design a flocculation gene resulting in improved flocculation characteristics when expressed in otherwise non-flocculating strains. To this end, a DNA sequence (*FLONF*) was isolated from the constitutively strongly flocculating *S. cerevisiae* CCUG 53310 (Westman et al., 2012), using primers designed for *FLO1* from S288c. Expression in CEN.PK 113-7D did not lead to the expected Flo1 phenotype flocculation, inhibited only by mannose. Instead the resulting strain displayed NewFlo phenotype flocculation that was inhibited by several sugar types. Since the internal repeat region of *FLONF* differed from that of *FLO1*, we hypothesized that a more stable flocculation gene could be created by merging a partial *FLO1* gene with a part of the internal repeat region of *FLONF*. Construction of a gene with a mix of different repeats would make it more stable in the strain, and easier to transfer between different strains to obtain rapidly sedimenting cells for application in e.g. sequential batches without need for centrifugation. With this in mind, we created a chimaeric gene that resulted in strong constitutive flocculation upon expression in otherwise non-flocculating yeast strains.

2. Materials and methods

2.1. Yeast strains and medium

The naturally constitutively flocculating strain *S. cerevisiae* CCUG 53310 was used for isolation of *FLONF*. The recombinant strains used in this study originated from *S. cerevisiae* CEN.PK 113-7D (MATa, MAL2–8C, SUC2) (van Dijken et al., 2000) which was used as reference strain, as well as from KE6–12.A (Tomás-Pejó et al., 2014) and IBB10B05 (Novy et al., 2014). Inoculations of pre-cultures were done by picking a colony from fresh YPD agar plates (10 g/l yeast extract, 20 g/l soy peptone, 20 g/l glucose, and 20 g/l agar), prepared from a frozen glycerol stock.

Aerobic cultures for cell propagation were grown in 250 ml cotton-plugged conical flasks in a shaker (125–200 rpm) at 30 °C. Liquid YPD medium containing 20 g/l of glucose was used for propagation of cells.

2.2. Isolation of a flocculation gene from *S. cerevisiae* CCUG 53310 and construction of a recombinant flocculating yeast strain

Using the primers *FLO1*-FW and *FLO1*-HO-RV (Table 1), giving a region homologous to the *HO* locus in *S. cerevisiae* CEN.PK 113-7D, a gene sequence from *S. cerevisiae* CCUG 53310 was amplified. The *TDH3* promoter (*TDH3p*) was amplified from genomic DNA of *S. cerevisiae* CEN.PK 113-7D by PCR using the primers EcoRV-*TDH3p*-FW and SpeI-*TDH3p*-RV (Table 1) and cloned in the pUG6 vector (Güldener et al., 1996). The resulting vector was used as template for amplification of the KanMX-*TDH3p* cassette using the primers SphI-KAN-FW and SalI-*TDH3p*-RV (Table 1). The cassette was cloned in YIplac211 (Gietz and Akio, 1988) and amplified by PCR using the primers *TDH3p*-*FLO1*-RV and *HO*-KAN-FW (Table 1), giving flanking ends homologous to the *HO* locus in *S. cerevisiae* CEN.PK 113-7D, and to the DNA sequence isolated from *S. cerevisiae* CCUG 53310, respectively. By fusion PCR the KanMX-*TDH3p* cassette and the DNA sequence were merged together and amplified using the primers *HO*-FW and *HO*-RV (Table 1) in two PCR reactions. In the first step the flocculation gene and the KanMX-*TDH3p* cassette were mixed in a single PCR reaction for PCR-based fusion, using the Phusion polymerase (Thermo Scientific, Lafayette, CO, USA). The DNA fragment resulting from the fusion was used as template for the second PCR reaction, with the primers *HO*-FW and *HO*-RV. This yielded a PCR

product with flanking regions homologous to the *HO* locus that was used for homologous recombination in CEN.PK 113-7D using the lithium acetate based transformation method (Gietz and Woods, 2002). Transformants were selected on YPD plates containing 200 µg/ml G418 (Sigma-Aldrich, Steinheim, Germany) and tested for flocculation ability in liquid YPD-medium, shaking overnight. The correct integration into the *HO* locus was confirmed by PCR using the primers SapI-KAN-RV and *HO*-upstream-2 (Table 1). The correctness of the sequences and of the integration was verified by sequencing (Eurofins MWG Operon, Ebersberg, Germany). The strain was named CEN.PK *FLONF*.

2.3. Inverse PCR

The flanking regions of *FLONF* were investigated by inverse PCR (Ochman et al., 1988). Genomic DNA was digested by 10 different restriction enzymes (AatII, BpI, XbaI, HindIII, MluI, NotI, PstI, SalI, BstXI and XhoI) creating sticky ends. The digested DNA was ligated at a DNA concentration of 1 µg/ml, to favour self-ligation. After phenol:chloroform extraction, the DNA was used as template for inverse PCR reactions using the primers IPCR-FW, IPCR-RV1 and IPCR-RV2. The samples from the inverse PCR were analysed on an agarose gel and interesting bands excised, purified and sequenced (Eurofins MWG Operon, Ebersberg, Germany). The sequences were investigated using EMBOSS Needle to assess the similarity to the flanking sequences of the flocculation genes in *S. cerevisiae* S288c.

2.4. Ploidy test

The ploidy of *S. cerevisiae* CCUG 53310 was investigated by flow cytometry of cells stained with Sytox green. Exponentially growing cells were harvested and fixated in cold 70% ethanol. The cells were washed in 10 mM EDTA, pH 8.0, followed by treatment with RNase A, 0.1 mg/ml at 37 °C for 2 h. Sytox green was added to a final concentration of 1 µM and the cells were analysed with a guava easyCyt 8HT flow cytometry system. The haploid strain *S. cerevisiae* BY4741, the diploid *S. cerevisiae* CEN.PK 122 MDS and the tetraploid strain *S. cerevisiae* G26 were used as references.

2.5. Construction of *FLOW*

The cassette with flanking regions of the *HO* locus containing KanMX, the *TDH3* promoter and the strongly flocculating *FLO1* variant was amplified in two parts from the genomic DNA of *S. cerevisiae* CEN.PK 113-7D strongly flocculating mutant (Westman et al., 2014). The 5' part of the 'strongly flocculating' *FLO1* (Genbank accession number KM366095) variant cassette was amplified with *HO*-upstream and *FLO1*-N-term-RV (Table 1). This amplicon contained the part of the flocculation gene encoding the carbohydrate binding N-terminal of the protein. The amplicon ended after the first (5' to 3') of the eleven internal repeats in the gene, according to a domain analysis by Pfam (Punta et al., 2012). The 3' part of the 'strongly flocculating' *FLO1* was amplified with *FLO1*-reps-FW and *HO*-downstream (Table 1). This amplicon contained the remaining, 3' part, of the flocculation gene, including the last ten (5' to 3') repeats.

Repeats from the *FLONF* flocculation gene (Genbank accession number KJ716851.1) of *S. cerevisiae* CCUG 53310 were amplified using the primers *FLONF*-reps-FW and *FLONF*-reps-RV (Table 1) from genomic DNA of CEN.PK *FLONF*. This gave regions homologous to the ends created by primers *FLO1*-N-term_RV and *FLO1*-reps_FW in the two fragments amplified from *S. cerevisiae* CEN.PK 113-7D strongly flocculating mutant. The PCR product was purified from an agarose gel.

By fusion PCR, the three fragments were assembled and the entire cassette was amplified using the nested primers *HO*-upstream-2 and *HO*-downstream-2 (Table 1) in two PCR reactions. Phusion polymerase

Table 1
Oligonucleotides used in the study.

Primer name	Sequence 5' → 3' ^a
<i>HO</i> -upstream	CAGTTGAAGACATGTGCGTC
<i>FLO1</i> -N-term-RV	GTACTGGCAGTTGTGGAGTTCTGATGAC
<i>FLO1</i> -reps-FW	GTCAAACTCCAAGTCTAGCACCA
<i>HO</i> -downstream	CAATCAGTCCCGTAACG
<i>FLONF</i> -reps-FW	CATTGTGTCAGAACTCCAACAAGTCCAGTACTACAATAACTACAAGTGGCCATGG
<i>FLONF</i> -reps-RV	GTAGTTATGATGGTGCTAGCAGTTGGAGTTTGGACAATAATGACAGTTTCATCAG
<i>HO</i> -upstream-2	CAGAAAGGGTTCGCAAGTC
<i>HO</i> -downstream-2	GGCGTATTTCTACTCCAGCATTC
<i>EcoRV</i> - <i>TDH3p</i> -FW	ATGATATCCAGTTCCGAGTTTATCATTATC
<i>SpeI</i> - <i>TDH3p</i> -RV	TACTAGTGTGTGTTTATTCGAAA
<i>SphI</i> -KAN-FW	ATAATGCATGCTTCGTACGCTGCAGGTAGACAAC
<i>Sall</i> - <i>TDH3p</i> -RV	TACGTCGACGTGTGTTTATTCGAACTAAG
<i>FLO1</i> -FW	ATGACAATGCCTCATCGCTATATGTTTGGC
<i>FLO1</i> -HO-RV	TTAGCAGATGCGCGCACCTGCGTTGTTACCACAACCTCTTATGAGTTAAATAATTGCCAGCAATAAG
<i>TDH3p</i> - <i>FLO1</i> -RV	CATATAGCGATGAGGCATTGTCATGTGTGTTTATTCGAACT
<i>HO</i> -KAN-FW	TACTTTGAATTGTACTACCGCTGGGCGTTATTAGGTGTGAAACCACGAGCTTCGTACGCTGCAGGTAG
<i>HO</i> -FW	TACTTTGAATTGTACTACCGCTGGGCG
<i>HO</i> -RV	TTAGCAGATGCGCGCACCTGCGTTG
<i>SapI</i> -KAN-RV	CAGCTCTTCCGCTCCTAATAACTTCGTATAG
IPCR-FW	GCACTATGTCGCAACAGCCTC
IPCR-RV1	AGTGGTATTAAGCAATGGAATGAAAGTC
IPCR-RV2	AACCTCAGGACATATCAGTGC
Mut-RV	GGGACTTCCATTCCATTGC
Mut-FW	CAATGGAATGGAAGTCCCCC

^a Boldface type indicates restriction sites.

(Thermo Scientific, Lafayette, CO, USA) was used for all PCR steps. The successful fusion was confirmed by agarose gel analysis. The resulting PCR product, with flanking regions homologous to the *HO* locus, was used for homologous recombination in CEN.PK 113-7D using the lithium acetate based transformation method (Gietz and Woods, 2002). The transformation resulted in a large number of transformants after selection on YPD agar plates with 200 µg/ml G418 (Sigma-Aldrich, Steinheim, Germany). Of these transformants, 30 randomly chosen colonies were picked and tested for flocculation in 3 ml YPD medium each. The flocculation characteristics were investigated by visual inspection after growth over night. The genomic DNA of the 6 most strongly flocculating strains was isolated and the length of the flocculation gene as well as the correct integration in the *HO* locus was investigated by PCR using the primers *HO*-upstream and *HO*-downstream (Table 1). The size of the inserted flocculation genes did indeed vary between the strains, as expected from the size variation obtained after the fusion PCR. The strain with the longest flocculation gene was chosen for further characterization and was named CEN.PK FLOW. The *FLOW* gene was amplified from genomic DNA and sent for sequencing (Eurofins MWG Operon, Ebersberg, Germany).

2.6. Flocculation and hydrophobicity tests

The flocculation tests were performed on yeast cells cultivated in YPD medium to stationary phase (48 h from inoculation) according to a previously reported protocol (Westman et al., 2012) with slight modifications. In short, cells were deflocculated by EDTA (maximum 250 mM), heat-killed and mixed at a concentration of approximately 1×10^8 cells/ml in citrate buffer (50 mM, pH 4.5) containing 4 mM CaCl₂ and various concentrations of the sugars tested, in a total volume of 2 ml placed in a 12 ml round bottom tube. The tubes were placed on an orbital shaker and agitated at 160 rpm at 25 °C, at an angle of approximately 30° for 4 h to ensure equilibrium. A sample of 150 µl was taken from just below the meniscus after the tubes had been left unagitated in a vertical position for 30 s. The sample was dispersed in 850 µl of 100 mM EDTA solution and the cell concentration was measured as OD₆₀₀. The extent of flocculation is presented as the

$$\text{percentage of free cells} = \left(\frac{\text{OD}_{600} \text{ of sample}}{\text{OD}_{600} \text{ of reference without CaCl}_2} \right) * 100.$$

The hydrophobicity of cells was tested by the Microbial Adhesion To Hydrocarbons (MATH) assay as previously described (Westman et al., 2012). The hydrophobicity is reported as the relative difference between the absorbance before and after vortexing with octane: $\text{hydrophobicity} = \left(1 - \frac{\text{OD}_{600} \text{ after vortexing}}{\text{OD}_{600} \text{ before vortexing}} \right) * 100.$

2.7. Site-directed mutagenesis

The site-directed mutagenesis E200G in *FLONF* was performed in three PCR-steps. In the first step the cassette from CEN.PK *FLONF* was amplified in two parts, with the primers Mut-RV and *HO*-upstream-2, and Mut-FW and *HO*-downstream, respectively, introducing the mutation in an overlapping part between the two fragments. The two fragments were fused together in a third PCR step and the resulting gene cassette transformed into CEN.PK 113-7D as described in Section 2.2, creating the strain CEN.PK *FLONF* E200G.

3. Results and discussion

3.1. Isolation of a novel flocculation gene sequence from *S. cerevisiae* CCUG 53310

Strong constitutive flocculation has been shown to be beneficial for inhibitor tolerance resulting in more complete fermentation of toxic lignocellulosic hydrolysates (Westman et al., 2014). However, due to the difficulty of recombinantly expressing the most well-characterized flocculation gene, *FLO1*, in its complete form, new genetic sources are necessary. In the search for flocculation genes conferring robust constitutive expression, naturally strongly flocculating yeast strains are the obvious starting point. This approach has been previously reported and a long *FLO1* variant resulting in strong flocculation upon overexpression was identified (Zhao et al., 2011). Here, we hypothesized that flocculation genes isolated from the inhibitor tolerant, constitutively strongly flocculating *S. cerevisiae* CCUG 53310 (Westman et al., 2012) would aid in the creation of strongly flocculating recombinant strains. Using the primers *FLO1*-FW and *FLO1*-HO-RV (Table 1), designed for *FLO1* from

S. cerevisiae S288c, retrieved from the *Saccharomyces* Genome Database (Cherry et al., 2012) we attempted to isolate novel flocculation genes by PCR.

Surprisingly, the amplified DNA fragment was not the expected *FLO1*, but a previously undescribed gene variant of approximately 3.4 kb (GenBank: KJ716851.1), significantly smaller than the 4.6 kb of *FLO1* in S288c. There were additional DNA fragments amplified in the PCR, one of similar size as *FLO1* in S288c and one even larger (Supplementary Fig. S1A). However, despite numerous attempts it was not possible to isolate the larger fragments.

The 3.4 kb DNA fragment was isolated and sequenced. Analysis of the sequence using Pfam (Punta et al., 2012) revealed a PA14 domain which is responsible for interactions with sugars, and a total of 12 repeats in the central domain, which is also characteristic for flocculins. The gene caused flocculation when transformed into the genome of the otherwise non-flocculating *S. cerevisiae* CEN.PK 113-7D and expressed constitutively under control of the *TDH3* promoter. The strain was named CEN.PK FLONF.

3.2. FLONF is a variant of FLO1 in the tetraploid CCUG 53310

From the results of the PCR amplification of *FLONF*, using primers designed for *FLO1* from S288c, it seemed likely that there are more than one copy of *FLO1* in *S. cerevisiae* CCUG 53310, with differences in length. The ploidy of *S. cerevisiae* CCUG 53310 has not been reported previously. Flow cytometry of cells stained with Sytox Green showed that the strain is tetraploid (Supplementary Fig. S1B). It is thus possible that the strain harbours four alleles of *FLO1*.

Inverse PCR was used to identify the regions flanking *FLONF* in the genome of CCUG 53310. From the sequencing results it was clear that the gene is located in the same site as *FLO1* (Supplementary Table S1). With primers specific for a small part of the *FLO1* sequence in *S. cerevisiae* S288c that is not present in *FLONF*, *FLO1*-FW2 and *FLO1*-RV2 (Table 1), it was shown that the *FLO1* gene is indeed also present in CCUG 53310. It is hence likely that *FLONF* is an allelic variant of *FLO1*. The tetraploid nature with different versions of the same genes makes CCUG 53310 a difficult strain to work with. Although they are most likely present, further attempts to isolate longer flocculation genes from CCUG 53310 were hence abandoned. Nonetheless, it was of interest to investigate the flocculation phenotype conferred by expression of *FLONF*.

3.3. FLONF confers a NewFlo phenotype to CEN.PK 113-7D

The reason that CEN.PK 113-7D was used for expression of the flocculation gene is that it lacks the major flocculation genes *FLO1*, *FLO5* and *FLO9* (Nijkamp et al., 2012). Flocculation occurring as a result of a gene transformation can thus be unambiguously assigned to the function of the new gene product being expressed. It has also been previously shown by real time PCR that constructs similar to *FLONF* were expressed at equal levels in CEN.PK 113-7D (Westman et al., 2014).

Treating the CEN.PK 113-7D FLONF cells with EDTA effectively deflocculated the cells, showing the expected Ca^{2+} dependence of the flocculation. EDTA acts as a chelating agent removing the Ca^{2+} from the flocs (Stratford, 1989). As the calcium ions were removed the cells were unable to bind to each other. Flocs were rapidly reformed after removal of the EDTA and addition of CaCl_2 to the cells. The amount of CaCl_2 present did not have a large effect on the flocculation of the cells over a threshold concentration of 0.1 mM (Supplementary Fig. S2). At lower concentrations, flocculation did not occur at significant levels.

To investigate whether strains overexpressing *FLONF* would have the ability to confer strong flocculation in 2nd generation bioethanol media, where several sugars are present in high concentration, the flocculation ability in the presence of different sugars at various concentrations (0–1.0 M) was studied for CEN.PK FLONF. With no

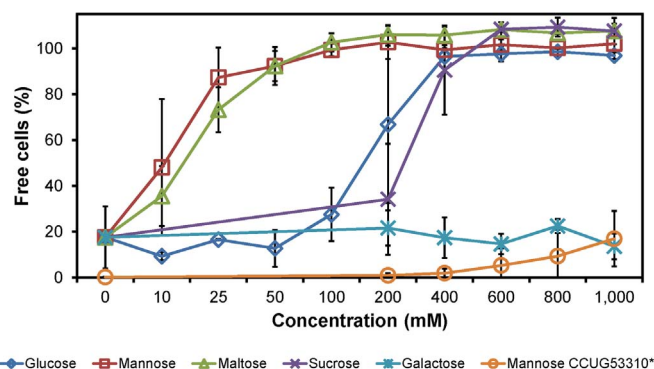


Fig. 1. The effect of sugars on the flocculation of the recombinant strain. The flocculation of the recombinant strain expressing the novel protein, FloNFp, was investigated in the presence of various sugars. Average values of at least duplicate experiments with duplicate technical replicates, shown with \pm one standard deviation, $n \geq 2$. *Data for *S. cerevisiae* CCUG 53310, from (Westman et al., 2012).

sugars present, the fraction of free cells was $18\% \pm 13\%$ -points. This shows that the flocculation conferred by *FLONF* is significantly weaker than that of the native *S. cerevisiae* CCUG 53310 (Westman et al., 2012). Furthermore, the flocculation was strongly inhibited by all sugars except galactose (Fig. 1). This means that the gene gave rise to NewFlo phenotype flocculation (Stratford and Assinder, 1991). The flocculation of the recombinant strain bearing the *FLONF* gene thus differed significantly from that of the strain it originated from. The flocculation of CCUG 53310 was affected only by mannose at concentrations of 0.8 M and above, and not at all by other sugars (Westman et al., 2012). Although we did not further investigate if the isolated *FLONF* gene is actively expressed in CCUG 53310, this indicates that the strain has more than one flocculation gene in its genetic repertoire, as has been previously proposed (Westman et al., 2012). The observation strengthens the conclusion that at least one more allelic variant of *FLO1* is present in CCUG 53310.

Simulations of the N-terminal part of the protein product, FloNFp, were made in order to elucidate potential differences between FloNFp and other members of the Flo-protein family (Supplementary text ST1-ST3; Supplementary Fig. S3; Supplementary Fig S4). In short, the Flo5 subdomain and the Gln 197-Pro 203 sequence (in L3, a loop flanking the carbohydrate binding site) were predicted to be the most flexible parts of the protein, causing these parts to be the least accurately predicted in the homology modelling. The flexibility of L3 allows it to approach and possibly interact with ligands in the active site.

Of special interest due to its position in L3, was a glutamate residue, E200, that replaces a glycine residue found in the corresponding positions in Flo1, Flo5 and Flo9. We hypothesized that the carboxy-group of Glu 200 can interact with the C_6 hydroxy-group of hexoses. Computationally mutating the Gly 198 residue (corresponding to Glu 200 in FloNFp) in the known Flo5p structure 2XJP to a glutamate, and rotating the C_6 -OH hydrogen atom, showed that the distance between these groups (1.7 \AA) should be ideal for a hydrogen bond (Supplementary Fig. S5). This led to the hypothesis that the glutamate residue contributes to the observed change from the Flo1 to the NewFlo flocculation phenotype. However, the E200G mutation and expression of the mutant flocculin in *S. cerevisiae* CEN.PK 113-7D did not cause any change in the flocculation properties (data not shown). Thus, the reason for the NewFlo phenotype must reside in other differences between the proteins.

3.4. Assembly of a chimaeric flocculation gene

Since *FLONF* did not live up to the criteria of conferring constitutive strong flocculation to CEN.PK 113-7D, a new hypothesis was postulated. We hypothesized that insertion of internal repeats isolated from *FLONF* into an S288c *FLO1* variant, would result in robust,

strong, constitutive flocculation of Flo1 phenotype. By inserting repeats with a different sequence than the ones already present in the gene, the risk of recombination events changing the phenotype would, plausibly, be reduced. Furthermore, since the strength of the flocculation is dependent on the length of the flocculation gene (Smukalla et al., 2008; Westman et al., 2014), a chimaeric gene with an increased number of repeats may result in stronger flocculation. The sugar recognition of the flocculation protein is mainly dependent on the N-terminal part (Kobayashi et al., 1998), which in this case would be identical to the S288c Flo1. Hence, a Flo1 phenotype flocculation would most likely be obtained.

The chimaeric flocculation gene *FLOW* was created as described in Section 2.5. However, isolation and amplification of the expected repeats from *FLONF* was not achieved. Since repeated sequences easily bind to each other during the PCR cycles, a mixed product is often obtained, with different sizes and sequences. This was the case in the current project. The repeats from *FLONF* could not be amplified into a single product from the genomic DNA of CEN.PK *FLONF* despite multiple attempts. Rather, mixed products were obtained. From the best attempt, a broad band between approximately 1.2–1.5 kbp was purified from an agarose gel and subsequently merged with the flanking parts derived from the ‘strongly flocculating’ *FLO1* variant. However, a single product was not obtained here either, indicating that many sizes of the repeat units were complementary to the end parts of the 5' and 3' parts of the cassette. Thus, after transformation into CEN.PK 113-7D, several colonies were picked and tested for flocculation and gene length by colony PCR. The strongest flocculating mutant, corresponding to the longest gene length, was designated CEN.PK *FLOW*.

The *FLOW* gene was isolated from CEN.PK *FLOW* and sequenced. However, due to the highly repetitive sequence in the inserted part of the gene, it was not possible to obtain a totally unambiguous sequence by primer walking. The parts of the cassette flanking the repeats from *FLONF* were identical to those previously reported (Westman et al., 2014). The repeated region from *FLONF* was, as expected from the results of the gel electrophoresis (Section 2.5), not identical to that planned in the original design. It was nonetheless clear that the region originated from *FLONF*. Restriction digestion with *Spe1*, *Sca1* and *Acc1* and comparing the sizes of the obtained fragments with the sequencing results helped to determine the correct sequence of the gene (Supplementary Fig. S6). All bases in the final sequence were identified using at least two primers. The gene, *FLOW*, has been deposited in GenBank with accession number KT264162. The size of the flocculation gene was 5151 bp, i.e. significantly larger than the full 4614 bp *FLO1* in S288c and larger than the planned size of 4881 bp.

Protein domain analysis using Pfam (Punta et al., 2012) revealed that 11 repeat units had been added to the internal region of the protein, rather than the planned 9 repeats, giving *FLOW* 22 repeats. This is the same number of extra repeats reported in a long variant of *FLO1* isolated by Zhao et al. (Zhao et al., 2011). The S288c *FLO1* has 18 repeats. An extended analysis of the identified repeats was performed by aligning the repeats identified by Pfam for Flo1 strongly flocculating variant, FloNF and Flow with ClustalW and constructing a phylogenetic tree to visualize the similarities between the repeats (Supplementary Fig. S7A). In this way it was possible to easily identify which repeats that had been inserted into Flow (Supplementary Fig. S7B). The two most N-terminal repeats of FloNF had not been transferred, while the following part seemed to have been inserted twice, but with the loss of the internal repeat FloNF #5 or #6. One repeat, Flow #6, was unique for the sequence of *FLOW*. These changes were likely caused by errors that occurred during the PCR cycles, creating the 11 repeats fragment in Flow from the 9 repeats *FLONF* template.

3.5. Flocculation of CEN.PK *FLOW*

The flocculation of CEN.PK *FLOW* was clearly stronger than that of CEN.PK 113-7D transformed with the strongly flocculating *FLO1* variant, reported in (Westman et al., 2014). As expected, expression of *FLOW* yielded a Flo1 phenotype flocculation, inhibited only by mannose (Fig. 2A). However, the inhibition was even weaker than for the *FLO1* variant it originated from. At a concentration of 1 M mannose there were only $14.6 \pm 2.5\%$ -points planktonic cells, compared to $60.7 \pm 3.3\%$ -points for the strongly flocculating mutant (Fig. 2A). As a comparison, the strongest flocculating strain described by Zhao et al. had 15% and 28% planktonic cells at 0.5 and 1 M mannose, respectively, assessed in a slightly different way (Zhao et al., 2011).

The cell wall hydrophobicity, assayed as previously described (Westman et al., 2012), was also clearly higher for the cells expressing the longer, chimaeric, flocculation protein (Fig. 1B). A hydrophobicity of $81.1 \pm 1.8\%$ -points was measured for CEN.PK *FLOW*, whereas a hydrophobicity of $55.3 \pm 5.0\%$ -points was reported for the ‘Strongly flocculating’ CEN.PK variant (Westman et al., 2014).

The cassette containing the *FLOW* gene could easily be amplified from the genomic DNA of CEN.PK *FLOW* and transformed into other *S. cerevisiae* strains of different genetic background, i.e. KE6–12.A (Tomás-Pejó et al., 2014) and IBB10B05 (Novy et al., 2014). Both of these strains are genetically engineered xylose-fermenting *S. cerevisiae* developed for second generation bioethanol production. The flocculation strength was improved by expression of the *FLOW* gene in both

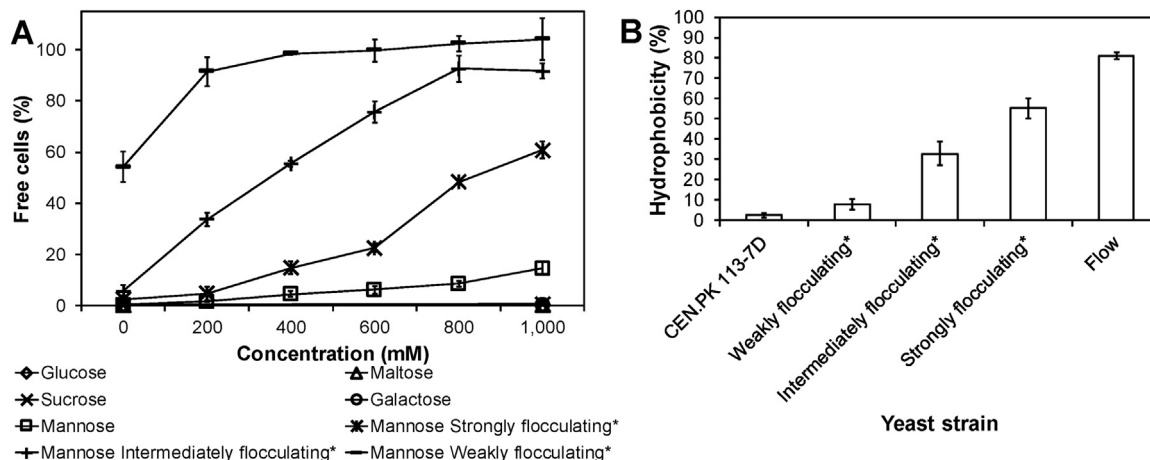


Fig. 2. Flocculation and cell wall hydrophobicity of CEN.PK *FLOW*. A) The flocculation of CEN.PK 113-7D transformed with the cassette containing the new flocculation gene, *FLOW*, was weakly inhibited only by mannose. B) Expression of *FLOW* resulted in a large increase in the cell wall hydrophobicity. For comparison, data from strains expressing previously reported variants of *FLO1* is shown. Error bars show standard deviation, $n \geq 2$. (*Data taken from (Westman et al., 2014).).



Fig. 3. Phenotypic appearance of KE6-12.A variants in shake flasks. A) KE6-12.A – non-flocculating, B) KE6-12.A expressing the strongly flocculating *FLOI* variant described in (Westman et al., 2014) and C) KE6-12.A expressing the chimaeric *FLOW* gene. Scale bars indicate ~10 mm.

cases, compared to expression of the ‘strongly flocculating’ *FLOI* variant, as shown in Fig. 3 for the KE6-12. A strain. The IBB10B05 strain behaved similarly. In-depth characterizations of the resulting strains are described in separate publications (Novy et al., 2017; Westman et al., 2017). The transferability of the cassette to other yeast strains clearly showed the robustness of the construct despite its considerable length.

The developed cassette can be used to obtain yeast strains that are easy to separate from the medium for use in *e.g.* sequential batches. Strong flocculation can also be a significant benefit for slowly growing cells. Large volumes of broth often have to be centrifuged to obtain a substantial amount of such cells. Separation by sedimentation has the potential to save significant amounts of both time and energy in these cases. Flocculation is also an efficient way of self-immobilization, enabling continuous fermentations with retention of cells without the need for filters (Zhao and Bai, 2009). Lastly, strong flocculation resulting in dense cell flocs, as those observed in shake flasks for strains harbouring the *FLOW* gene, leads to enhanced robustness by increasing the tolerance to inhibitors present in lignocellulosic hydrolysates (Westman et al., 2014).

4. Conclusions

We constructed a cassette containing a gene conferring strong flocculation when expressed in *S. cerevisiae* cells. The flocculation is only weakly inhibited by high concentrations of mannose and not at all by other carbohydrates tested. The gene can be easily transformed into different *S. cerevisiae* strains and used to *e.g.* facilitate an easier cell separation or to increase the robustness of the cell population.

Acknowledgements

We thank our students Lina Lawenius, Marcus Hanaeus, John Hellgren, Jonathan Aldridge, Alan Dazay and Christoffer Dall for sequencing the flanking regions of *FLONF* in the genome of CCUG 53310 and constructing the mutant *FLONF* during their BSc thesis project. We thank Cecilia Geijer and Lina Lindahl for help with the flow cytometry measurements.

Funding: This work was supported by the Swedish Energy Agency [grants P13872-2 and P37353-1], the Swedish Research Council [grant 2009-4514], and the Chalmers Energy Initiative (<http://www.chalmers.se/en/areas-of-advance/energy/cei/>).

Conflict of interest

The authors declare no financial nor commercial conflict of interest.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.meten.2018.04.001](https://doi.org/10.1016/j.meten.2018.04.001).

References

- Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., Fisk, D.G., Hirschman, J.E., Hitz, B.C., Karra, K., Krieger, C.J., Miyasato, S.R., Nash, R.S., Park, J., Skrzypek, M.S., Simison, M., Weng, S., Wong, E.D., 2012. *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* 40, D700–D705.
- Gietz, R.D., Akio, S., 1988. New yeast-*Escherichia coli* shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites. *Gene* 74, 527–534.
- Gietz, R.D., Woods, R.A., 2002. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* 350, 87–96.
- Goossens, K., Willaert, R., 2010. Flocculation protein structure and cell-cell adhesion mechanism in *Saccharomyces cerevisiae*. *Biotechnol. Lett.* 32, 1571–1585.
- Güldener, U., Heck, S., Fiedler, T., Beinbauer, J., Hegemann, J.H., 1996. A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res.* 24, 2519–2524.
- Kobayashi, O., Hayashi, N., Kuroki, R., Sone, H., 1998. Region of Flo1 proteins responsible for sugar recognition. *J. Bacteriol.* 180, 6503–6510.
- Liu, N., Wang, D., Wang, Z.Y., He, X.P., Zhang, B., 2007. Genetic basis of flocculation phenotype conversion in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 7, 1362–1370.
- Nijkamp, J., van den Broek, M., Datema, E., de Kok, S., Bosman, L., Luttik, M., Daran-Lapujade, P., Vongsangnak, W., Nielsen, J., Heijne, W., Klaassen, P., Paddon, C., Platt, D., Kotter, P., van Ham, R., Reinders, M., Pronk, J., de Ridder, D., Daran, J.-M., 2012. De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN.PK113-7D, a model for modern industrial biotechnology. *Microb. Cell Fact.* 11, 36.
- Novy, V., Krahulec, S., Wegleiter, M., Müller, G., Longus, K., Klimacek, M., Nidetzky, B., 2014. Process intensification through microbial strain evolution: mixed glucose-xylose fermentation in wheat straw hydrolysates by three generations of recombinant *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 7, 49.
- Novy, V., Wang, R., Westman, J.O., Franzén, C.J., Nidetzky, B., 2017. *Saccharomyces cerevisiae* strain comparison in glucose-xylose fermentations on defined substrates and in high-gravity SSCF: convergence in strain performance despite differences in genetic and evolutionary engineering history. *Biotechnol. Biofuels* 10, 205.
- Ochman, H., Gerber, A.S., Hartl, D.L., 1988. Genetic applications of an inverse polymerase chain reaction. *Genetics* 120, 621–623.
- Punta, M., Coghill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L.L., Eddy, S.R., Bateman, A., Finn, R.D., 2012. The Pfam protein families database. *Nucleic Acids Res.* 40, D290–D301.
- Smukalla, S., Caldara, M., Pochet, N., Beauvais, A., Guadagnini, S., Yan, C., Vincens, M.D., Jansen, A., Prevost, M.C., Latgé, J.-P., Fink, G.R., Foster, K.R., Verstrepen, K.J., 2008. *FLOI* is a variable green beard gene that drives biofilm-like cooperation in budding yeast. *Cell* 135, 726–737.
- Stratford, M., 1989. Yeast flocculation: calcium specificity. *Yeast* 5, 487–496.
- Stratford, M., Assinder, S., 1991. Yeast flocculation: Flo1 and NewFlo phenotypes and receptor structure. *Yeast* 7, 559–574.
- Teunissen, A.W., van den Berg, J.A., Steensma, H.Y., 1993. Physical localization of the flocculation gene *FLOI* on chromosome I of *Saccharomyces cerevisiae*. *Yeast* 9, 1–10.
- Teunissen, A.W.R.H., Steensma, H.Y., 1995. The dominant flocculation genes of *Saccharomyces cerevisiae* constitute a new subtelomeric gene family. *Yeast* 11, 1001–1013.
- Tomás-Peño, E., Bonander, N., Olsson, L., 2014. Industrial yeasts strains for biorefinery solutions: constructing and selecting efficient barcoded xylose fermenting strains for ethanol. *Biofuels, Bioprod. Bioref.* 8, 626–634.
- van Dijken, J.P., Bauer, J., Brambilla, L., Duboc, P., Francois, J.M., Gancedo, C., Giuseppe, M.L.F., Heijnen, J.J., Hoare, M., Lange, H.C., Madden, E.A., Niederberger, P., Nielsen, J., Parrou, J.L., Petit, T., Porro, D., Reuss, M., van Riel, N., Rizzi, M., Steensma, H.Y., Verrips, C.T., Vindelov, J., Pronk, J.T., 2000. An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzym. Microb. Tech.* 26, 706–714.
- Verstrepen, K.J., Derdelinckx, G., Verachtert, H., Delvaux, F.R., 2003. Yeast flocculation: what brewers should know. *Appl. Microbiol. Biotechnol.* 61, 197–205.
- Verstrepen, K.J., Jansen, A., Lewitter, F., Fink, G.R., 2005. Intragenic tandem repeats generate functional variability. *Nat. Genet.* 37, 986–990.
- Watarai, J., Takata, Y., Ogawa, M., Sahara, H., Koshino, S., Onnala, M.-L., Airaksinen, U., Jaatinen, R., Penttilä, M., Keränen, S., 1994. Molecular cloning and analysis of the yeast flocculation gene *FLOI*. *Yeast* 10, 211–225.

- Westman, J.O., Taherzadeh, M.J., Franzén, C.J., 2012. Inhibitor tolerance and flocculation of a yeast strain suitable for 2nd generation bioethanol production. *Electron. J. Biotechnol.* 15 (3). <http://dx.doi.org/10.2225/vol15-issue3-fulltext-8>.
- Westman, J.O., Mapelli, V., Taherzadeh, M.J., Franzén, C.J., 2014. Flocculation causes inhibitor tolerance in *Saccharomyces cerevisiae* for second-generation bioethanol production. *Appl. Environ. Microbiol.* 80, 6908–6918.
- Westman, J.O., Wang, R., Novy, V., Franzén, C.J., 2017. Sustaining fermentation in high-gravity ethanol production by feeding yeast to a temperature-profiled multifeed simultaneous saccharification and co-fermentation of wheat straw. *Biotechnol. Biofuels* 10, 213.
- Zhao, X.Q., Bai, F.W., 2009. Yeast flocculation: new story in fuel ethanol production. *Biotechnol. Adv.* 27, 849–856.
- Zhao, X.-Q., Li, Q., He, L.-Y., Li, F., Que, W.-W., Bai, F.-W., 2011. Exploration of a natural reservoir of flocculating genes from various *Saccharomyces cerevisiae* strains and improved ethanol fermentation using stable genetically engineered flocculating yeast strains. *Process Biochem.* 47, 1612–1619.