Uniwersytet Przyrodniczy w Poznaniu
Wydział Nauk o Żywności i Żywieniu

Maria Gorczyca

Czynniki transkrypcyjne jako narzędzia masowego działania w ulepszaniu cech przemysłowych u drożdży

Transcription factors as tools of massive operation in engineering of industrial traits in yeast

Rozprawa doktorska w dziedzinie nauk ścisłych i przyrodniczych
w dyscyplinie nauk biologicznych
Doctoral dissertation in the field of natural sciences
in the discipline of biological sciences

Promotor:
Prof. dr hab., Ewelina Celińska
Katedra Biotechnologii i Mikrobiologii Żywności, Uniwersytet Przyrodniczy w Poznaniu

Poznań, 2025

Finansowanie

Praca powstała w wyniku realizacji projektu badawczego OPUS o numerze 2021/41/B/NZ9/00086 finansowanego przez Narodowe Centrum Nauki, pod kierownictwem prof. dr hab. Eweliny Celińskiej

NARODOWE CENTRUM NAUKI

Pragnę złożyć serdeczne podziękowania mojej Promotorce, **prof. dr hab. Ewelinie Celińskiej**, za nieocenione wsparcie, przekazaną wiedzę i inspirację na wszystkich etapach mojej ścieżki naukowej – od pracy inżynierskiej i magisterskiej po przygotowanie niniejszej rozprawy doktorskiej. Jestem wdzięczna za stworzenie przestrzeni do rozwoju, motywowanie do podejmowania ambitnych wyzwań oraz dzielenie się pasją badawczą. Jej nieustanna gotowość pomocy, zaangażowanie oraz przykład szacunku i partnerskiego podejścia we współpracy naukowej kształtowały mnie zarówno jako badacza, jak i jako człowieka.

Dziękuję **dr inż. Paulinie Korpys-Woźniak** za przekazaną wiedzę praktyczną i dzielenie się bogatym doświadczeniem w pracy molekularnej z *Yarrowia lipolytica*. Jestem wdzięczna za rozmowy i wspólne poszukiwanie rozwiązań podczas eksperymentów, a także za dzielenie naukowej pasji. Jej pomoc, życzliwość i poczucie humoru sprawiały, że codzienna praca w laboratorium stawała się nie tylko łatwiejsza, lecz także prawdziwie przyjemna.

Szczególne wyrazy wdzięczności kieruję do moich **Rodziców** i **Brata**. Rodzicom dziękuję za nieustanną wiarę we mnie, za inspirowanie do sięgania wyżej i dodawanie odwagi w wyznaczaniu celów. Ogromnie doceniam również stworzenie warunków, które umożliwiły mi podjęcie i kontynuację nauki. Mojemu Bratu dziękuję za przykład cierpliwości, wytrwałości i systematyczności w podejmowanych działaniach.
Dziękuję również **Dziadkom** za wspieranie mojej edukacji i motywowanie do rzetelnej i wytrwałej pracy, a **całej Rodzinie** - za wsparcie i nieustanne kibicowanie moim naukowym przedsięwzięciom.

Jestem ogromnie wdzięczna moim **Przyjaciółkom** za obecność i towarzyszenie mi w czasie najbardziej wymagających etapów pracy badawczej. Wasze wsparcie i rozmowy stanowiły nieocenioną równowagę wobec intensywności pracy naukowej, dodając mi siły w drodze do ukończenia rozprawy.

Na szczególne podziękowania zasługuje także **Szymon** - za wyrozumiałość, niezachwiane wsparcie, i cierpliwość wobec moich naukowych rozterek. Dziękuję za motywację, dodawanie siły, dzielenie się doświadczeniami naukowymi i pomoc w zachowaniu właściwego dystansu do trudności badawczych. Szczególnie cenię wspólną pasję odkrywania i dzielenie doświadczeń z naukowych wypraw oraz szczere cieszenie się z moich sukcesów. I wreszcie – za konsekwentne przekonywanie wszystkich znajomych i rodziny, że *Yarrowia* to najlepsze drożdże jakie istnieją.

Spis treści

Streszczenie

Odporność na stres środowiskowy oraz synteza rekombinowanych białek (rProts) to złożone i silnie ze sobą powiązane cechy biologiczne, zależne od skoordynowanej aktywności wielu genów. Ponieważ czynniki transkrypcyjne (TFs) działają na wyższych poziomach kontroli ekspresji genów, stanowią obiecujące cele inżynierii takich złożonych cech. Ich aktywność podlega wielopoziomowej kontroli – od transkrypcji, poprzez translację aż po zmiany statusu aktywacji białka – zależnych od bodźców zewnętrznych i wewnętrznych. W niniejszej pracy zastosowano dwutorową strategię mającą na celu intensyfikację wybranych cech złożonych w drożdżach *Yarrowia lipolytica*: i) globalną modulację metabolizmu komórkowego poprzez ingerencję w TF oraz ii) precyzyjne sterowanie syntezą rProt w oparciu o specyficzne interakcje TF-promotor, umożliwiające ukierunkowane dostrajanie ekspresji wybranych genów.

Przeprowadzono kompleksowe badania przesiewowe kolekcji szczepów ko-nadekspresjonujących TF z rProt w różnych warunkach środowiskowych. Dla wszystkich testowanych TFs przeprowadzono integrację wielowymiarowych danych w celu identyfikacji wzorców regulacyjnych, natomiast wybrane TFs poddano dalszej charakterystyce przez wprowadzenie delecji. W odrębnej linii badań, analizowano wpływ nadekspresji wybranego TF, w wersjach pełnej i splicingowej, na ekspresję reporterowego genu klonowanego pod różnymi, ale specyficznymi dla tego TF, promotorami. Ponadto, celem wyjaśnienia poczynionych obserwacji, wykorzystano modelowanie komputerowe do przewidywania interakcji TF-DNA promotora i TF-ligand, a dane transkryptomiczne posłużyły do wyjaśnienia mechanizmu regulacji.

W ramach pracy zoptymalizowano protokół hodowli w mikroskali oraz utworzono bazę danych YaliFunTome, gromadzącą wyniki funkcjonalnego skriningu TFs w warunkach stresowych i istotnych dla produkcji rProt. Zidentyfikowano TFs pełniące rolę uniwersalnych „wzmacniaczy" i represorów syntezy rProt, a także TF wspomagający syntezę rProt przy obecności nie preferowanego źródła azotu. Wykazano powiązania niektórych TFs z odpornością na ograniczoną dostępność tlenu, jak również ich rolę w tolerancji na stres osmotyczny i oksydacyjny. TF *EUF1* został scharakteryzowany jako regulator siły kompatybilnego promotora, zwiększając siłę ekspresji, ale kosztem utraty indukowalności układu. Analizy oparte na „recykling

danych” dodatkowo wykazały, że kontrola syntezy rProt przez TFs w wielu przypadkach prawdopodobnie obejmuje mechanizmy posttranskrypcyjne.

Podsumowując, uzyskane wyniki dowodzą, że TFs mogą stanowić wszechstronne narzędzia w inżynierii wybranych cech złożonych *Y. lipolytica*. Mogą działać jako globalne modulatory „przeprogramowując” szlaki metaboliczne bądź jako ukierunkowane regulatory specyficznych sekwencji promotorowych. W ramach pierwszego podejścia zaproponowano ogólne zależności, zgodnie z którymi poprawa asymilacji azotu sprzyjała syntezie rProts, natomiast wzmożony anabolizm węgla działał niekorzystnie. Przedstawione wyniki otwierają nowe perspektywy dla inżynierii opartej na TFs w zakresie syntezy rProts i odporności na stres środowiskowy oraz stanowią podstawę do dalszej funkcjonalnej charakterystyki TFs w tym niekonwencjonalnym gatunku drożdży.

Słowa kluczowe: *Yarrowia lipolytica*, czynniki transkrypcyjne, inżynieria cech złożonych, białka rekombinowane, odporność na stres

Abstract

Resistance to environmental stress and synthesis of recombinant protein (rProt) are complex and tightly interconnected biological traits that depend on the coordinated activity of multiple genes. Because transcription tactors (TFs) act at higher regulatory levels of gene expression, they represent a promising targets for the engineering of such complex traits. Their activity is itself regulated at multiple layers, including transcription, translation, and protein activation status in response to stimuli. In this work, a twofold strategy was adopted to enhance selected complex traits in *Yarrowia lipolytica*: i) global modulation of cellular metabolism through TF perturbation, and ii) precise control of rProt synthesis via TF-promoter interactions.

A comprehensive screening approach was applied, combining co-overexpression library of TFs with rProt under variable environmental conditions. Multidimentional data integration was performed for all tested TFs to uncover regulatory patterns, while selected TFs were further characterized through targeted knock-outs. In a distinct line of investigation, the impact of selected TF overexpression, in spliced and unspliced versions, on the expression of reporter genes driven by different promoters specific to this TF was analysed. In addition, to explain the observed phenomena, computational modeling was employed to predict TF-DNA promoter interactions and TF-ligand binding, while transcriptomic data facilitated elucidation of the underlying regulatory mechanism.

This work established a protocol for reproducible microscale cultivation and introduced the YaliFunTome database, compiling functional screening data of TFs under stress and production-relevant conditions. Several TFs were identified as universal enhancers or suppressors of rProt synthesis, and one was found to promote rProt production specifically in the presence of a non-preferred nitrogen source. Distinct TFs were linked to altered resistance under oxygen limitation, with additional roles demonstrated in oxidative and osmotic stress tolerance. The TF *EUF1* was shown to act as a titrator of its cognate promoter, enhancing expression strength but comprising inducibility. Data-recycling approach highlighted that TF-mediated control of rProt synthesis likely extends beyond transcriptional regulation, implicating post-transcriptional mechanisms.

Together, these results demonstrate TFs as versatile tools for engineering selected complex traits in *Y. lipolytica*. They can act either as global modulators reshaping

metabolic fluxes – where generalized pattern emerged, in which improved nitrogen assimilation proved beneficial for rProt synthesis while enhanced carbon anabolism was detrimental – or as targeted regulators fine-tuning compatible promoters. These findings provide new perspectives for TF-based engineering of rProts and stress resistance and lay the foundation for functional characterization of TFs in this non-conventional yeast species.

Key words: *Yarrowia lipolytica*, transcription factors, global metabolic engineering, recombinant proteins, stress resistance

1. Wstęp

Biotechnologia przemysłowa stanowi jeden z ważnych obszarów współczesnej nauki i gospodarki, oferując narzędzia umożliwiające wytwarzanie związków o wysokiej wartości dodanej w sposób wydajny i zgodny z zasadami zrównoważonego rozwoju. Potrzeba wdrażania takich rozwiązań nabiera szczególnego znaczenia w obliczu systematycznego wzrostu globalnego zapotrzebowania na białka oraz inne związki biologicznie czynne. Jednym z najbardziej wymownych przykładów ilustrujących rolę biotechnologicznej produkcji białek jest insulina. Szacuje się, że w przypadku braku możliwości jej wytwarzania w systemach mikrobiologicznych, dla zaspokojenia światowego zapotrzebowania (ok. 40 ton) konieczne byłoby coroczne pozyskiwanie tego hormonu z około 4 miliardów świń (Kjeldsen i in., 2024). Światowe zapotrzebowanie na białka przemysłowe jest obecnie zapewnione głównie poprzez biotechnologiczną produkcję w mikroorganizmach, z wykorzystaniem mikroorganizmów natywnych lub modyfikowanych genetycznie.

Rekombinowane białka (rProts, ang. *recombinant proteins*) znajdują zastosowanie w rozlicznych gałęziach ludzkiej aktywności, m.in. w profilaktyce, diagnostyce i terapii chorób, produkcji żywności i pasz, w rolnictwie jako biologiczne środki ochrony roślin, w przemyśle chemicznym (np. jako składowe środków piorących), w kosmetyce, w przetwórstwie biomasy odpadowej oraz w produkcji biopaliw. Produkcja mikrobiologiczna przeważa nad ekstrakcją białek z ich natywnych źródeł pochodzenia pod względem łatwości modyfikacji, braku sezonowej zmienności, bezpieczeństwa, a przede wszystkim zapewnia znacznie wyższą efektywność i możliwość skalowania w kierunku zrównoważonej produkcji.

Wiodącymi platformami do produkcji rProts są bakterie (np. *Escherichia coli*) i drożdże modelowe (np. *Saccharomyces cerevisiae*), a także drożdże „niekonwencjonalne", takie jak *Komagataella phaffii* (dawniej *Pichia pastoris*) oraz, zyskujące coraz większe znaczenie w tym obszarze, drożdże *Yarrowia lipolytica*. Każda z tych platform ma określone zalety i ograniczenia. Bakterie charakteryzuje szybkość wzrostu, prostota hodowli i niski koszt prowadzenia procesów, lecz nie zawsze umożliwiają prawidłowe fałdowanie białek eukariotycznych i wprowadzanie modyfikacji posttranslacyjnych. Drożdże modelowe, choć są zdolne do prowadzenia takich zmian, mogą nie być optymalnym wyborem ze względu na niską wydajność

sekrecji (wydzielania poza komórkę) oraz w przypadku białek terapeutycznych, przez skłonność do hiperglikozylacji. Z tego względu coraz większą uwagę zwraca się na organizmy alternatywne, które mogą łączyć w sobie zalety obu systemów i otwierają nowe możliwości inżynieryjne.

1.1 Drożdże *Yarrowia lipolytica* jako platforma do produkcji białek heterologicznych

1.1.1 Uwarunkowania molekularne

Yarrowia lipolytica jest niekonwencjonalnym gatunkiem drożdży, który od kilku dekad stanowi przedmiot intensywnych badań ze względu na liczne cechy predysponujące go do wykorzystania w przemysłowej produkcji rProts. Zainteresowanie tym organizmem jako platformą ekspresyjną pojawiło się już w latach 80. XX wieku, kiedy firma Pfizer Inc. opatentowała pierwsze technologie umożliwiające jego zastosowanie w produkcji białek heterologicznych. Od tego czasu opracowano liczne szczepy zoptymalizowane pod kątem wytwarzania rProts, obejmujące zarówno szczepy nadprodukujące białka homologiczne (np. lipazy; Pigne`de i in., 2000), heterologiczne – pochodzące z obcych organizmów, oraz tzw. szczepy humanizowane – zdolne do wprowadzania wzorów glikozylacji zbliżonych do ludzkich (De Pourcq i in., 2012). Otworzyło to drogę do wykorzystania *Y. lipolytica* w produkcji białek terapeutycznych, co jeszcze do niedawna pozostawało domeną głównie komórek ssaczych. Do tej pory w tym gatunku opisano produkcję ponad 200 różnych rProts (podsumowane w Madzak, 2018), co czyni go jedną z najlepiej scharakteryzowanych niekonwencjonalnych platform ekspresyjnych.

Y. lipolytica wykazuje szereg korzystnych cech biologicznych, szczególnie predysponujących te drożdże do wykorzystania jako „mikrobiologiczna fabryka komórkowa". Charakteryzuje się zdolnością do wzrostu w wysokich gęstościach biomasy, stosunkową łatwością do manipulacji genetycznych oraz szerokim (i nieustannie poszerzanym) zestawem narzędzi genetycznych do wprowadzania precyzyjnych zmian w genomie. Do obecnie dostępnych narzędzi należą: system Gateway (Leplat i in., 2015), GoldenGate (Celińska i in., 2017; Egermeier i in., 2019), technologie edycji genomu oparte na nukleazach TALEN (Rigouin i in., 2017),

YaliBricks (Wong i in., 2017) – modułowy system umożliwiający szybkie konstruowanie i łączenie wielogenowych szlaków metabolicznych. Dalszym osiągnięciem o istotnym znaczeniu jest implementacja technologii CRISPR-Cas9 (Schwartz i in., 2017) oraz systemów na niej bazujących: EasyClone (Holkenbrink i in., 2018), ExpressYali (Jiang i in., 2024). Najnowsze rozwiązania obejmują również systemy umożliwiające indukowalną kontrolę ekspresji genów, co dodatkowo zwiększa skalę możliwości inżynierii (Park i in., 2019a; Trassaert i in., 2017). Ponadto *Y. lipolytica* charakteryzuje się wysoce efektywnym szlakiem sekrecyjnym, bardziej podobnym do szlaku typowego dla grzybów strzępkowych niż dla drożdży modelowych (Celińska i Nicaud, 2019), a także zdolnością do wzrostu na wielu rodzajach substratów, w tym surowcach odpadowych (Rakicka i in., 2015; Sarris i in., 2017), co zwiększa ich atrakcyjność w kontekście gospodarki o obiegu zamkniętym.

1.1.2 Warunki środowiskowe

Przedmiotowy gatunek drożdży wykazuje stosunkowo wysoką odporność na różnorodne czynniki stresowe (Cogo i in., 2020; Kolhe i in., 2021; Kubiak-Szymendera i in., 2022; Lesage i in., 2021; Pomraning i in., 2016; Sekova i in., 2021; Walker i in., 2019; Yang i in., 2015, podsumowane w Celińska, 2022). Czynniki środowiskowe wywierają istotny wpływ na funkcjonowanie komórek-producentów, obejmując tempo wzrostu, aktywność metaboliczną oraz zdolność do syntezy heterologicznych białek u drożdży (Hou i in., 2013; Liu i in., 2013).

Jednym z najważniejszych czynników limitujących procesy biotechnologiczne w *Y. lipolytica* jest niewystarczająca dostępność tlenu, która nierzadko przekracza techniczne możliwości systemów laboratoryjnych i przemysłowych. Oprócz rozwiązań bioprocesowych ukierunkowanych na zwiększenie natlenienia (np. hodowla w bioreaktorach z podwyższonym ciśnieniem; Lopes i in. 2008, 2009), podejmowano również próby ograniczenia zapotrzebowania na tlen poprzez modyfikacje genetyczne. Jednym z ciekawych podejść było wprowadzenie hemoglobiny bakteryjnej, co ostatecznie zwiększało syntezę rProts (Bhave i Chattoo, 2003) i erytrytolu (Mirończuk i in., 2019). Badania Gorczycy i in. (2020) wykazały, że ograniczone natlenienie negatywnie wpływa na wszystkie etapy procesu biosyntezy – od transkrypcji i translacji po sekrecję białek.

Następnym ważnym czynnikiem modulującym zdolności produkcyjne *Y. lipolytica* jest pH (Sassi i in., 2017). Choć w bioreaktorach jego kontrola jest technicznie możliwa, to lokalne wahania kwasowości podłoża występują często (Lara i in., 2006). Wykazano, że bardzo niskie pH (2,0–3,0), mieszczące się w zakresie tolerowanym przez drożdże, choć umożliwia wzrost, prowadzi do tzw. stresu kwasowego (Wu i in., 2020). Z kolei badania Sassi i in. (2017) dowiodły, że spośród czynników takich jak pH i morfologia komórki, to właśnie pH ma kluczowy wpływ na produkcję rProts.

Temperatura oraz osmolalność środowiska stanowią kolejne elementy istotnie kształtujące fenotyp produkcyjny. Optymalizacje warunków hodowli przeprowadzone przez Kubiak i in. (2021) oraz Kubiak-Szymenderę i in. (2022) wykazały, że obniżenie temperatury sprzyja syntezie rProts, głównie poprzez zwiększenie ich sekrecji (Korpys-Woźniak i in., 2021), a tło molekularne tego zjawiska zostało scharakteryzowane dzięki badaniom proteomicznym i transkryptomicznym. Te same analizy wykazały, że hiperosmolalość podłoża nie sprzyja wydajnej produkcji białek sekrecyjnych.

Warto podkreślić, że nadprodukcja dużych i złożonych rProts generuje dodatkowe obciążenie metaboliczne komórek, które znacząco obniża ich odporność na stresy środowiskowe, takie jak niedobór tlenu czy nieoptymalne pH (Gorczyca i in., 2022). W konsekwencji zdolność *Y. lipolytica* do adaptacji na warunki stresowe (stres środowiskowy) należy rozpatrywać w ścisłym związku z mechanizmami molekularnymi odpowiedzi na stres endogenny, wynikający z intensywnej sekrecji i fałdowania białek. Mechanizmy odpowiedzi zarówno na stres środowiskowy, jak i odpowiedzi na stres wewnętrzny były szczegółowo badane u drożdży modelowych (Gasch, 2007; Gasch i in., 2000; Gasch i Werner-Washburne, 2002), co pozwoliło na identyfikację głównych molekularnych regulatorów tych procesów (Craig i in., 1993; Verghese i in., 2012). Wiadomo, że nadmierna produkcja (i sekrecja) rProts często prowadzi do akumulacji nieprawidłowo sfałdowanych polipeptydów i powoduje stres endogenny (Hou i in., 2014; Matsumoto i in., 2005; Mattanovich i in., 2004; Tyo i in., 2012). Prace prowadzone z wykorzystaniem *S. cerevisiae* wykazały ponadto znaczące nakładanie się odpowiedzi szoku cieplnego (stres środowiskowy) i odpowiedzi UPR (ang. *unfolded protein response*) (stres endogenny) (Guyot i in., 2005; Hahn i in., 2004; Hou i in., 2013; Verghese i in., 2012). Sugeruje to, że podobna interakcja odpowiedzi może mieć miejsce również u *Y. lipolytica*, choć zagadnienie to jest dopiero systematycznie wyjaśniane.

1.2 Czynniki transkrypcyjne

Kluczową rolę w integracji sygnałów płynących zarówno ze środowiska zewnętrznego, jak i z wnętrza komórki – wynikających z jej aktywności metabolicznej i fizjologicznej – pełnią kaskady sygnalizacyjne oraz czynniki transkrypcyjne (TFs, ang. *transcription factors*). Te drugie stanowią nadrzędny poziom regulacji ekspresji genów, pozwalając komórce na dostosowanie programu transkrypcyjnego do zmieniających się warunków otoczenia oraz wewnętrznych potrzeb metabolicznych. Inicjacja transkrypcji w komórkach drożdży wymaga obecności wyspecjalizowanych białek regulatorowych. Do tej grupy należą „czynniki transkrypcyjne ogólnej odpowiedzi" (GTFs, ang. *general transcription factors*), niezbędne do rozpoczęcia transkrypcji wszystkich genów kodujących białka. Część z nich wiąże się bezpośrednio z sekwencjami promotorowymi (np. z motywem TATA), jednak większość działa poprzez interakcje białko-białko, stabilizując kompleks inicjacyjny i umożliwiając prawidłowe związanie polimerazy II RNA. Samo działanie GTFs zapewnia jedynie podstawowy poziom transkrypcji, ich działanie ma charakter globalny i niespecyficzny. Wyższa wydajność i specyficzna regulacja ekspresji poszczególnych genów wymagają działania drugiej grupy regulatorów – „specyficznych czynników transkrypcyjnych" (w niniejszej rozprawie określanych ogólnym skrótem TFs). Oddziałując z elementami cis-regulacyjnymi DNA, TFs działają jako aktywatory lub represory transkrypcji, wiążąc się z i modulując skuteczność rekrutacji oraz procesywność maszynerii transkrypcyjnej. Aktywatory zawierają zwykle dwie kluczowe domeny: wiążącą DNA, odpowiedzialną za specyficzność oddziaływania, oraz aktywacyjną, która poprzez interakcje z innymi białkami regulatorowymi lub elementami kompleksu transkrypcyjnego wzmacnia poziom transkrypcji. Represory mogą natomiast hamować ekspresję genów, blokując dostęp aktywatorów do sekwencji DNA, zakłócając interakcje białko-białko lub osłabiając aktywności kompleksu transkrypcyjnego (Urry, 2008).

Wiadomo, że same TFs reagują na różne bodźce zarówno zmianą wzoru ekspresji, jak i modulacją statusu aktywacji (Muriel i Claude, 2014; Peñalva i Arst, 2002). W badaniach funkcjonalnych często wykorzystuje się konstytutywną nadekspresję, która uniezależnia obecność białka TF od natywnej regulacji transkrypcyjnej i zapewnia jego wysoką ilość w komórce. Podejście ekspozycji szczepów na zmienne środowiskowe pozwala modulować status aktywacji TF. Połączenie tych dwóch metod pozwala

dokładniej ocenić rolę poszczególnych TFs w analizowanych procesach biologicznych, które mogą pozostawać niewidoczne przy testach prowadzonych wyłącznie w standardowych warunkach laboratoryjnych.

W przypadku *Y. lipolytica*, badania nad TFs były dotąd prowadzone głównie w kontekście badań podstawowych. Większość doniesień dotyczyła roli pojedynczych TFs w regulacji przejścia dimorficznego; udział w tym procesie potwierdzono m.in. dla: Hoy1 (Torres-Guzman i in., 1997), Mhy1 (Hurtado i Rachubinski, 1999; Wu i in., 2020), Bmh1 (Hurtado i Rachubinski, 2002), TF z rodziny Yap (Morales-Vargas i in., 2012), Znc1 (Martinez-Vazquez i in., 2013) i Msn2 (Pomraning i in., 2018). Rola TFs w regulacji metabolizmu lipidów w *Y. lipolytica* została opisana m.in. dla: Yas1 i Yas2 (Endoh-Yamagami i in., 2007), Yas3 (Hirakawa i in., 2009), Por1 (Poopanitpan i in., 2010), Mig1 (Wang i in., 2013) oraz Gzf2 i Gzf3 (Pomraning i in., 2018). Z kolei gen *MSN4* wskazano jako regulator tolerancji na stres kwasowy (Wu i in., 2020). Praktyczna realizacja koncepcji inżynierii globalną maszynerią transkrypcyjną w *Y. lipolytica* była dotąd bardzo ograniczona i sprowadzała się do zaledwie kilku badań, w tym wysokoprzepustowych przesiewów funkcjonalnych obejmujących ponad 120 szczepów, z których każdy nadekspresjonował jeden TF, w celu identyfikacji globalnych regulatorów zaangażowanych w akumulację lipidów (Leplat i in., 2018; Trébulle i in., 2017). W zakresie produkcji białek, dotychczas udokumentowanym przykładem było jedynie zastosowanie TF Hac1 (Korpys-Woźniak i Celińska, 2023).

W inżynierii metabolicznej procesów biotechnologicznych TFs mogą być wykorzystane w dwóch głównych kierunkach. Po pierwsze, mogą pełnić funkcję narzędzi „globalnej" inżynierii komórkowej, aktywując całe szlaki metaboliczne sprzyjające zwiększonej efektywności procesów produkcyjnych, np. intensywniejszą asymilację związków odżywczych, usprawnione fałdowanie białek czy bardziej wydajny transport metabolitów poza komórkę. Po drugie, ze względu na swój mechanizm działania – specyficzną interakcję z sekwencjami regulatorowymi DNA, przede wszystkim w obrębie promotorów genów – TFs umożliwiają precyzyjną regulację lub zwiększoną aktywację konkretnych genów, w tym promotorów wykorzystywanych w procesach biotechnologicznych do produkcji rProts. Taka właściwość otwiera perspektywy ich zastosowania w biologii syntetycznej, zarówno jako aktywatorów naturalnych promotorów, jak i elementów systemów regulacyjnych o kontrolowanej sile działania. Oba te podejścia zostaną omówione w kolejnych podrozdziałach.

1.2.1 Promotory indukowalne

Ponieważ TFs oddziałują ze specyficznymi sekwencjami regulatorowymi DNA, w tym z promotorami, to właśnie promotory stanowią elementy, przez które TFs realizują swoją funkcję kontrolną. Szczególne znaczenie mają promotory indukowalne, charakteryzujące się regulowaną siłą działania i umożliwiające precyzyjną kontrolę ekspresji genów, co jest wysoce pożądane zarówno w badaniach podstawowych, jak i praktyce przemysłowej. W porównaniu z promotorami konstytutywnymi dają one istotną przewagę – pozwalają na rozdzielenie fazy wzrostu od fazy produkcyjnej, zmniejszając obciążenie metaboliczne komórki i ograniczając ryzyko toksyczności, gdy produkt końcowy jest dla niej niekorzystny.

Biorąc pod uwagę fizjologię drożdży *Y. lipolytica*, która wyjątkowo efektywnie metabolizuje lipidy i białka, można oczekiwać, że w ich genomie obecne będą promotory indukowalne tymi właśnie związkami. Rzeczywiście, pierwszymi zbadanymi przykładami były: i) pXPR2, kontrolujący ekspresję zewnątrzkomórkowej proteazy alkalicznej, indukowany kombinacją takich czynników jak pH, stosunek węgla do azotu i obecność peptonów (Madzak i in., 1999; Ogrydziak, 2013; Roland i in., 1994), oraz ii) pPOX2 i pLIP2 (Juretzek i in., 2000; Sassi i in., 2016), aktywowane w obecności lipidów, m.in. kwasu oleinowego. Ich zastosowanie w praktyce było jednak ograniczone – w przypadku pXPR2 przez złożoność regulacji i niepełną indukowalność, a w przypadku pPOX2 i pLIP2 ze względu na trudności związane z wykorzystaniem hydrofobowych induktorów. Kolejnym krokiem było opracowanie bardziej funkcjonalnych rozwiązań, takich jak: iii) php4d – syntetyczny promotor zależny od fazy wzrostu, wywodzący się z pXPR2 (Madzak i Blanchin-Roland, 2000), oraz iv) silny system indukowalny erytrytolem (Carly i in., 2017; Park i in., 2019a; Park i in., 2019b; Trassaert i in., 2017; Vandermies i in., 2017; Vidal i in., 2023). Zwłaszcza ten ostatni został szczegółowo scharakteryzowany, a jego różne syntetyczne warianty zoptymalizowano pod względem siły i użyteczności. Okazał się on szczególnie wydajny w tle genetycznym szczepu *Δeyk1*, pozbawionego zdolności do asymilacji erytrytolu, dzięki czemu induktor stosowany w fazie indukcji nie jest zużywany. W ostatnich latach zaproponowano również kilka nowych promotorów indukowalnych dla *Y. lipolytica* (Georgiadis i in., 2023; Wang i in., 2023). Autorzy wykorzystali podejście genomowe (Wang i in., 2023) lub transkryptomiczne (Georgiadis i in., 2023) do identyfikacji

elementów regulacyjnych. W pierwszym przypadku analizowano geny aktywowane przez glicerol, co doprowadziło do konstrukcji silnych promotorów syntetycznych, lecz pozbawionych charakteru indukowalnego. W drugim zidentyfikowano promotory związane z metabolizmem węgla i azotu, obejmujące szerokie spektrum aktywności transkrypcyjnej – od słabej, przez średnią, po silną (do 1.6-krotnie wyższą niż „złoty standard" – promotor pTEF), jednak i tutaj pełna indukcyjność nie została osiągnięta.

W świetle powyższych doniesień dalsze poszukiwania silnych i jednocześnie ściśle kontrolowalnych promotorów indukowalnych dla *Y. lipolytica* pozostają w pełni uzasadnione. Dotyczy to nie tylko samych sekwencji promotorowych, lecz także ich elementów regulacyjnych, takich jak TFs, które mogą modulować siłę działania, zapewniać skalowalność – umożliwiać stopniowe dostrajanie aktywności oraz precyzyjną kontrolę w zastosowaniach biotechnologicznych.

1.2.2 Inżynieria cech złożonych

Cechy złożone, warunkowane przez wiele genów i zależne od ich wzajemnych interakcji, obejmują liczne elementy molekularne oraz rozbudowane sieci regulacyjne, przez co rzadko są dobrze opisane w organizmach niemodelowych. Obecnie powszechnie uznaje się, że manifestacja pożądanego fenotypu takiej cechy rzadko zależy od działania pojedynczego genu lub niewielkiej ich grupy. Zazwyczaj konieczne jest precyzyjne dostrojenie wielu zdarzeń molekularnych, aby uzyskać określoną funkcjonalność.

Jednym z efektywnych podejść do ich kształtowania jest adaptacyjna ewolucja w warunkach laboratoryjnych (ALE, ang. *adaptive laboratory evolution*), łącząca losową zmienność genetyczną z długoterminową selekcją korzystnych mutacji w określonych warunkach wzrostu, a następnie może obejmować inżynierię odwrotną (ang. *reverse engineering*). Metoda ta umożliwia nie tylko poznanie podstawowych mechanizmów ewolucji adaptacyjnej, lecz także praktyczne wykorzystanie w inżynierii metabolicznej. W przeciwieństwie do racjonalnego projektowania i modyfikacji wybranych genów, ALE pozwala na ujawnienie korzystnych, często nieintuicyjnych mutacji w wielu genach i regionach regulatorowych jednocześnie. Z tego względu okazała się skuteczna w optymalizacji szczepów produkcyjnych – m. in. poprzez aktywację ukrytych szlaków metabolicznych, poprawę tempa wzrostu, zwiększenie wydajności docelowych metabolitów czy podniesienie tolerancji na stresy typowe dla procesów przemysłowych

(Dragosits i Mattanovich, 2013; Mans i in., 2018; Portnoy i in., 2011; Winkler i Kao, 2014).

Innym podejściem, do inżynierii genetycznej takich cech (pozwalającym na wprowadzenie wielu zmian funkcjonalnych w komórce poprzez modulację pojedynczych regulatorów) jest wykorzystanie globalnych regulatorów – tzn. elementów działających na wyższym poziomie hierarchii molekularnej – np. kaskady sygnałowe czy TFs. Przykładowo, nadekspresja genu *HAC1* – kodującego TF regulujący procesy w obrębie siateczki śródplazmatycznej i przywracający jej homeostazę – poprawia wydajność syntezy rProt w *S. cerevisiae* (Duan i in., 2019), *Komagataella phaffii* (Guerfal i in., 2010) oraz *Y. lipolytica* (Korpys-Woźniak i Celińska, 2023). Produkcja sekrecyjnych rProt była znacząco zwiększona w *S. cerevisiae* w wyniku ko-nadekspresji *HSF1,* kluczowego regulatora odpowiedzi na stres cieplny (Hou i in., 2013). Ciągła aktywacja tej odpowiedzi przez ko-nadekspresję zmutowanego wariantu *HSF1-R206S,* prowadziła do zwiększonej syntezy zarówno białek natywnych, jaki i rekombinowanych. Podobnie ko-nadekspresja *HAP1* (TF odpowiedzialnego za metabolizm tlenowy i aktywację odpowiedzi na stres oksydacyjny) łagodziła negatywne skutki stresu oksydacyjnego wywołanego intensywnym fałdowaniem białek i tym samym zwiększyła zdolność komórek *S. cerevisiae* do produkcji rProt (Martínez i in., 2016). Niedawno wykazano, że syntetyczna aktywacja Msn4 (oraz jego wariantu syntetycznego synMsn4 – samodzielnie lub w kombinacji) powodowała ponad czterokrotne zwiększenie wydajności produkcji rProt (Zahrl i in., 2023).

Podsumowując, efektywna inżynieria cech złożonych wymaga zazwyczaj równoczesnej modulacji wielu genów – podejścia określanego mianem inżynierii „globalnej" (ang. *global metabolic engineering*), które pozwala na precyzyjne dostrojenie działania komórki do określonych warunków lub do realizacji pożądanego procesu biotechnologicznego. Ze względu na fakt, że TFs pełnią rolę nadrzędnych modulatorów, kontrolujących liczne zdarzenia molekularne, ich aktywność bądź dezaktywacja stanowią obiecującą strategię manipulowania tymi cechami. Podejście to może umożliwić wywołanie najbardziej adekwatnej odpowiedzi komórkowej i tym samym zwiększyć efektywność procesów produkcyjnych.

Na tle powyższych rozważań wyraźnie widać, że dalszy rozwój potencjału biotechnologicznego *Y. lipolytica* wymaga wyjścia poza klasyczne modyfikacje pojedynczych genów i optymalizacje warunków hodowli. W centrum zainteresowania niniejszej rozprawy doktorskiej znajdują się TFs – nadrzędne regulatory integrujące sygnały środowiskowe i wewnątrzkomórkowe, koordynujące liczne procesy biologiczne. Ich znaczenie w inżynierii metabolicznej można uchwycić w dwóch odrębnych strategiach. Pierwsza obejmuje inżynierię „globalną", w której modulacja aktywności TFs pozwala przeprogramować całe szlaki metaboliczne, zwiększając odporność komórek na stresy i ich ogólne zdolności produkcyjne. Druga to podejście ukierunkowane, wykorzystujące interakcje TFs ze specyficznymi promotorami w celu regulacji siły działania tychże promotorów, co umożliwia zwiększenie wydajności syntezy rProt. Równocześnie rozprawa ma na celu wzbogacenie wiedzy podstawowej o rolach dotąd nieopisanych u *Y. lipolytica* TFs, dostarczając nowych danych na temat ich funkcji biologicznych. Połączenie tych perspektyw pokazuje, że TFs są jednocześnie narzędziem szerokiego oddziaływania i precyzyjnej kontroli, co uzasadnia ich centralne miejsce w niniejszych badaniach dotyczących inżynierii tego gatunku drożdży.

2. Cel rozprawy i zadania badawcze

Niniejsza rozprawa doktorska ma na celu charakterystykę TFs w drożdżach *Y. lipolytica* pod względem ich przydatności w ulepszaniu pożądanych cech o znaczeniu przemysłowym. W pracy skoncentrowano się na dwóch głównych obszarach, efektywności produkcji rProt oraz odporności na stresy środowiskowe.

Nadrzędnymi celami badań przedstawionych w toku niniejszej rozprawy doktorskiej są:

i) Identyfikacja TFs w niekonwencjonalnym gatunku drożdży *Y. lipolytica* powiązanych ze wzmożoną syntezą heterologicznych białek, oraz charakterystyka warunków środowiska, w których ulegają one aktywacji i zyskują funkcjonalność.

ii) Identyfikacja TFs zaangażowanych w odporność na stresy środowiskowe adekwatne do warunków bioprocesowych.

iii) Ocena możliwości wykorzystania nadekspresji lub delecji TFs w modulowaniu/ulepszaniu powyższych cech przemysłowych u drożdży *Y. lipolytica*.

iv) Wykorzystanie TF jako narzędzia do ukierunkowanej intensyfikacji siły promotora kierującego ekspresją pożądanych genów heterologicznych.

Celami szczegółowymi, będącymi równocześnie głównymi zadaniami badawczymi prezentowanej rozprawy są:

i) Opracowanie wysokoprzepustowego protokołu zminiaturyzowanych hodowli umożliwiających utrzymanie stałych warunków hodowli oraz optymalizacja metodyki dotyczącej poboru i analizy próbek.

ii) Badania przesiewowe (skrining; ang. *screening*) w 72 wybranych warunkach środowiskowych kolekcji 125 szczepów nadekspresjonujących jeden TF wraz z heterologicznym wewnątrzkomórkowym białkiem reporterowym oraz ocena fenotypu drożdży pod względem wzrostu i syntezy białka.

iii) Opracowanie sposobu przedstawiania wielowymiarowych danych będących wynikami hodowli, przy wykorzystaniu modelowania matematycznego i analizy statystycznej.

iv) Wytypowanie TFs, obiecujących pod względem modulacji syntezy rProts i/lub odporności na stresy środowiskowe, do dalszych analiz polegających na ich

nadekspresji lub delecji w szczepach: a) wydzielających białka na zewnątrz komórki (w tym enzymów przemysłowych), lub b) bez wprowadzonych genów heterologicznych.

v) Przedstawienie szczegółowej charakterystyki wybranego TF oraz ocena jego przydatności jako narzędzia „biologii syntetycznej" do regulowania siły kompatybilnego promotora indukowalnego.

Niniejszym podjęto próbę odpowiedzi na następujące pytania badawcze:

i) Czy możliwe jest kształtowanie złożonych cech biologicznych poprzez racjonalną inżynierię genetyczną polegającą na manipulacji TFs pełniącymi rolę nadrzędnych regulatorów wielu procesów molekularnych w *Y. lipolytica*?

ii) Które TFs uczestniczą w utrzymaniu efektywności (ang. *robustness*) wzrostu i aktywności metabolicznej w warunkach stresowych oraz/lub wzmożonej zdolności do produkcji białek heterologicznych u *Y. lipolytica*?

iii) Jakie warunki środowiskowe aktywują fenotypy poszczególnych TFs? Które TFs są aktywowane w warunkach środowiskowych adekwatnych do procesów bioprodukcji?

iv) Czy możliwe jest wykorzystanie TFs jako modulatorów siły transkrypcji specyficznych dla nich promotorów?

2.1 Znaczenie badań dla rozwoju dziedziny i potencjał aplikacyjny

Wyniki prac badawczych przedstawionych w niniejszej rozprawie ubogacają zbiór wiedzy podstawowej dotyczącej metabolizmu *Y. lipolytica*, a jednocześnie mają wysoki potencjał aplikacyjny w przemyśle biotechnologicznym, wykorzystującym te mikroorganizmy jako biokatalizatory. Opracowany protokół wysokoprzepustowych zminiaturyzowanych hodowli będzie przydatny w wielu laboratoriach i działach badawczo-rozwojowych firm, wykorzystujących *Y. lipolytica* jako obiekt badań. Sposób przetwarzania i przedstawienia danych umożliwi ich udostępnienie w formie przeszukiwalnej i dostępnej dla społeczności naukowej.

3. Główne tezy rozprawy doktorskiej

Hipotezą badawczą, stojącą u podstaw niniejszej rozprawy, jest założenie, że modyfikacje TFs mogą być skutecznym narzędziem inżynierii złożonych cech biologicznych u *Y. lipolytica*, obejmujących syntezę heterologicznych białek oraz odporność na stresy środowiskowe. Główne tezy i założenia niniejszej pracy stanowią:

i) Konstytutywna nadekspresja genu kodującego wybrany TF (z kolekcji 125 TFs) pozwala uwolnić jego dostępność od sieci regulacyjnych, którym natywnie podlega.

ii) Sama dostępność (obecność produktu białkowego) TF nie zapewnia jego biologicznej aktywności, stąd dla ujawnienia fenotypu powiązanego z aktywnością badanego TF konieczna jest jego aktywacja; np. na drodze ekspozycji szczepu nadekspresjonującego na określone perturbacje środowiskowe.

iii) Wytypowanie TF przydatnych w modulowaniu złożonych cech biologicznych adekwatnych do warunków bioprocesowych (produkcji białek heterologicznych i/lub odpowiedzi na dany czynnik stresowy), należy przeprowadzić poprzez zapewnienie ich dostępności oraz ich aktywacji poprzez ekspozycję na adekwatne warunki środowiskowe.

iv) Delecja TF o niekorzystnym wpływie na badane cechy może skutkować „odwróconym fenotypem", który w badanym układzie eksperymentalnym prowadzi do poprawy wybranych właściwości.

v) Profil ekspresji TF w szczepach nadprodukujących rProts może stanowić wskaźnik wyboru TFs do inżynierii genetycznej, ukierunkowanej na ulepszenie syntezy rProt.

vi) Wybrane TF mogą być wykorzystywane jako narzędzia zwiększające poziom ekspresji genów znajdujących się pod kontrolą kompatybilnych promotorów.

4. Lista publikacji naukowych składających się na rozprawę doktorską

P1. Gorczyca M., Nicaud JM., Celińska E. (2023)

Transcription factors enhancing synthesis of recombinant proteins and resistance to stress in *Yarrowia lipolytica*

Applied Microbiology and Biotechnology, 107, 4853-4871

IF (5-letni): 4,9

Punkty MEiN: 100

P2. Gorczyca M., Białas W., Nicaud JM., Celińska E. (2024)

'Mother(Nature) knows best' – hijacking nature-designed transcriptional programs for enhancing stress resistance and protein production in *Yarrowia lipolytica*; presentation of YaliFunTome database

Microbial Cell Factories, 23:26

IF (5-letni): 6,1

Punkty MEiN: 100

P3. Celińska E., **Gorczyca M.** (2024)

'Small volume – big problem' – culturing *Yarrowia lipolytica* in high-throughput micro-formats

Microbial Cell Factories, 23:184

IF (5-letni): 6,1

Punkty MEiN: 100

(dane bibliograficzne zestawu danych stanowiącego załącznik do publikacji:

Celińska E., **Gorczyca M**.

Comparison of *Yarrowia lipolytica* growth and rProt synthesis when cultured in different vessels

Mendeley Data, doi: 10.17632/c3trp7jt3n.1)

P4. Gorczyca M., Korpys-Woźniak P., Celińska E. (2024)

An interplay between transcription factors and recombinant protein synthesis in *Yarrowia lipolytica* at transcriptional and functional levels – the global view

International Journal of Molecular Sciences, 25(17), 9450

IF (5-letni): 5,7

Punkty MEiN: 140

P5. Celińska E., Korpys-Woźniak P., **Gorczyca M.**, Nicaud JM. (2024)

Using Euf1 transcription factor as a titrator of erythritol-inducible promoters in *Yarrowia lipolytica*; insight into structure, splicing and regulation mechanism

FEMS Yeast Research, Volume 24, foae027

IF (5-letni): 2,6

Punkty MEiN: 100

Całkowity 5-letni współczynnik IF czasopism powyższych publikacji, zgodny z rokiem opublikowania artykułów naukowych, wynosi 25,4. Całkowita liczba punktów ministerialnych za publikacje według listy czasopism punktowanych MEiN wynosi 540.

5. Osiągnięcia rozprawy doktorskiej

W skład niniejszej rozprawy wchodzi pięć powiązanych tematycznie artykułów naukowych. Dotyczą one wykorzystania inżynierii TFs w drożdżach *Y. lipolytica* celem intensyfikacji syntezy rProts i/lub zwiększenia odporności na wybrane stresy środowiskowe, lub wykorzystania wytypowanych TFs jako narzędzi „biologii syntetycznej". Publikacja **P1** stanowi eksperymentalną walidację głównego założenia pracy i odpowiedzi na pytanie „*Czy możliwe jest kształtowanie złożonych cech biologicznych poprzez racjonalną inżynierię genetyczną polegającą na manipulacji TFs pełniących rolę nadrzędnych regulatorów wielu procesów molekularnych?*". Badanie przeprowadzono na ograniczonym zestawie TFs. Publikacja **P2** wdraża zwalidowany koncept na drodze przeprowadzenia skriningu całej biblioteki 125 TFs w poszukiwaniu aktywatorów wzmożonej produkcji białek oraz odporności na stres, stosując perturbacje środowiskowe jako aktywatory. Trzecia publikacja (**P3**) ma nietypowy charakter – odnosi się do zestawu oryginalnych danych, wygenerowanych na potrzeby oszacowania skali problemu hodowli *Y. lipolytica* w mikroskali, który stanowi podstawę do dalszej dyskusji i przeglądu literatury. W dalszej części praca obejmuje analizę literatury dotyczącej hodowli *Y. lipolytica* w małej skali oraz komentarz dotyczący trudności związanych z miniaturyzacją procesów hodowlanych. Artykuł **P4** rozszerza metodykę badań nad TFs o integrację danych transkryptomicznych z wynikami fenotypowymi i oferuje globalne spojrzenie na metabolizm, umożliwiając poszukiwanie zależności pomiędzy pozornie niepowiązanymi cechami i wynikami. Ostatnia publikacja (**P5**) prezentuje przykład praktycznego wykorzystania jednego z badanych TF jako regulatora siły transkrypcyjnej specyficznych promotorów. W tej pracy podjęto próbę szczegółowej charakterystyki samego białka, jak i poznania mechanizmu jego działania we współdziałaniu z czynnikami modulującymi jego aktywność. Omówienie poszczególnych publikacji wraz z dołączonymi oryginalnymi wersjami artykułów zostało przedstawione w kolejnych podrozdziałach niniejszej rozprawy.

5.1 Ogólne podejście metodyczne

Ze względu na podobieństwo metodyczne publikacji **P1** i **P2** (oraz częściowo **P4**), ogólną metodykę badań zastosowaną w nich przedstawiono w tym podrozdziale. Kluczowym etapem, umożliwiającym precyzyjne różnicowanie fenotypów wynikających z modyfikacji TF, była optymalizacja protokołu hodowli w mikroskali. Jej wstępne założenia wypracowano w ramach badań pilotażowych (**P1**), a następnie wykorzystano w szeroko zakrojonych analizach przesiewowych (**P2**). Istotnym elementem wspólnej metodyki było również opracowanie sposobu prezentacji danych wielowymiarowych oraz ich analiza wspomagana modelowaniem matematycznym, m.in. z wykorzystaniem metod powierzchni odpowiedzi i wizualizacji w postaci map cieplnych. Choć ogólna koncepcja obu prac była zbieżna, drobne różnice wynikające z dalszego udoskonalania metodyki zostały omówione poniżej.

W badaniach wykorzystano zmodyfikowane genetycznie szczepy *Y. lipolytica* prowadzące ko-nadekspresję wewnątrzkomórkowego białka fluorescencyjnego RedStar2 wraz z jednym TF, a także szczepy pozbawione wybranych TF w tle genetycznym szczepu syntetyzującego białko reporterowe. Kolekcję „szczepów ko-nadekspresjonujących" TF i wewnątrzkomórkowy RedStar2 otrzymano dzięki uprzejmości prof. Jean-Marca Nicaud (Uniwersytet Paris-Seclay; INRAE, Francja). Szczepy wchodzące w skład tej kolekcji to pochodne szczepu JMY2566, zmodyfikowanego w *locus URA3* poprzez wbudowanie platformy dokującej zeta oraz konstruktu umożliwiającego konstytutywną nadekspresję białka fluorescencyjnego RedStar2 i równoległą nadekspresję wybranego TF (Leplat i in., 2015, 2018) (**P1**, **P2** i dane **P4**). Wcześniejsze poszukiwania TF w genomie *Y. lipolytica* przeprowadzono metodami bioinformatycznymi, opierając się na analizie najbardziej prawdopodobnych domen strukturalnych i motywów sekwencyjnych wiążących DNA. Powyższą kolekcję wzbogacono o skonstruowane w toku niniejszych badań szczepy z delecją wybranych TF, w których *locus* genu TF zastąpiono kasetą z genem oporności na nourseotrycynę (NATr), flankowaną ~1 kpz regionami homologii po obu stronach kasety (**P1**, **P2**, **P4**).

Optymalizacja protokołu hodowli obejmowała: i) dobór składu podłoża w sposób ograniczający tło fluorescencji oraz zapewniający odpowiednią dostępność źródeł węgla i azotu, ii) wybór buforu pokrywającego wymagany zakres pH i utrzymującego jego stabilność, przy czym kluczowe było zastosowanie jednej substancji buforującej dla

całego zakresu pH, aby nie wprowadzać dodatkowej zmiennej wynikającej z rodzaju użytego związku chemicznego, iii) określenie minimalnej objętości hodowli umożliwiającej powtarzalność między eksperymentami oraz zapewniającej wysokie natlenienie i mieszanie hodowli, iv) zoptymalizowanie zakresu szybkości mieszania zapewniającego warunki wzrostu zgodne z założeniami wariantu eksperymentalnego oraz v) czasu pobierania próbek pozwalającego uchwycić różnice fenotypowe przed momentem, w którym czynnik środowiskowy staje się czynnikiem jednolicie limitującym we wszystkich szczepach.

W **P1** wykorzystano 24-dołkowe płytki (Biologix, Chiny) o pojemności roboczej 1 mL (zgodnie ze wskazaniem producenta) i okrągłych dołkach hodowlanych. W toku dalszej optymalizacji protokołu (**P2**) zastosowano głębokie płytki 24-dołkowe o pojemności roboczej 2.5 mL (zgodnie ze wskazaniem producenta) i kwadratowych dołkach (Duetz, EnzyScreen BV, Holandia), co umożliwiło efektywniejsze mieszanie cieczy hodowlanej oraz poprawę transferu tlenu i masy. Zastosowanie specjalistycznych przykrywek typu „*sandwich*" (kompatybilnych z systemem Duetz) zwiększyło równomierność napowietrzania i ograniczyło parowanie, w przeciwieństwie do standardowych pokryw, które sprzyjają lepszemu natlenieniu jedynie dołków brzeżnych.

Jako celowo wprowadzane zmienne środowiskowe badano zakres wartości pH, dostępność tlenu, temperaturę, osmolalność, oraz rodzaj źródła węgla i azotu, przy czym szczegółowy zestaw i sposób ich implementacji różnił się pomiędzy badaniami. W **P1** zaplanowano 36 wariantów warunków złożonych z kombinacji: i) ograniczonej dostępności tlenu (zmniejszenie szybkości wytrząsania o 25%), ii) hiperosmolalności (dodatek sorbitolu podnoszący osmolalność 3.75-krotnie), iii) trzech poziomów temperatury (22, 28 i 34 °C) i iv) trzech poziomów pH (3,0; 5,0; 7,0). W pracy **P2** wykorzystano natomiast metodologię *Design of Experiments* (DoE) co umożliwiło systematyczne uwzględnienie pięciu zmiennych: dwóch ilościowych (pH i temperatura) oraz trzech kategorycznych (dostępność tlenu – wysoka/niska – zapewniona przez zastosowanie różnych typów przykrywek systemu Duetz, źródło węgla – glukoza/glicerol, źródło azotu – siarczan amonu/hydrolizat kazeiny). Podejście to pozwoliło także precyzyjnie określić liczbę koniecznych powtórzeń eksperymentu, zapewniając wysoką siłę statystyczną uzyskanych modeli. W efekcie zaplanowano 72 warianty środowiskowe będące kombinacją wszystkich zmiennych. Celem wytworzenia zestawu danych zaprezentowanych w powiązaniu z **P3**, badano wzrost *Y. lipolytica* i syntezę heterologicznego białka reporterowego w zoptymalizowanym medium, ale

w 13 różnych formatach (będących kombinacją rodzaju naczynia hodowlanego, jego wypełnienia i szybkości wytrząsania; tak, aby zróżnicować współczynnik transferu masy k_La). W **P4** i **P5** zastosowano standardowe warunki hodowlane, nie wprowadzając dodatkowych czynników stresowych.

Dane dotyczące wzrostu komórek, fluorescencji całkowitej (odpowiadającej poziomowi produkcji rProt) oraz fluorescencji specyficznej (w przeliczeniu na jednostkę biomasy), uzyskane z wykorzystaniem czytnika płytek wielodołkowych zaopatrzonego w stosowne filtry do czytania fluorescencji, poddano modelowaniu matematycznemu w obrębie przestrzeni eksperymentalnej (metodą powierzchni odpowiedzi). Uzyskane modele umożliwiły zarówno ocenę istotności pojedynczych parametrów, jak i ich interakcji oraz ilościowe opisanie procentowego udziału poszczególnych zmiennych środowiskowych w uzyskanej odpowiedzi. Wyniki przedstawiono także w formie map cieplnych, obrazujących zmiany poszczególnych cech fenotypowych względem szczepu kontrolnego nadekspresjonującego białko fluorescencyjne, bez ingerencji w natywne TF.

5.2 **P1**: Czynniki transkrypcyjne zwiększające syntezę rekombinowanych białek i odporność na stresy środowiskowe w drożdżach *Yarrowia lipolytica*

Publikacja **P1**, otwierająca cykl artykułów naukowych składających się na rozprawę doktorską, miała charakter pilotażowy i dotyczyła wybranej grupy genów kodujących TFs. Jej głównym celem było zweryfikowanie słuszności założeń i tez postawionych w początkowym etapie projektu. Badania te stanowiły jednocześnie podstawę do opracowania protokołu wysokoprzepustowych hodowli w mikroskali dostosowanych dla *Y. lipolytica* oraz sposobu prezentacji i interpretowania wielowymiarowych danych fenotypowych.

Celem tego badania było zweryfikowanie potencjalnych powiązań, wybranych pięciu TFs – Hsf1 (*YALI0E13948g*), Gzf1 (*YALI0D20482g*), Crf1 (*YALI0B08206g*), Skn7 (*YALI0D14520g*) oraz TF z rodziny Yap (*YALI0D07744g*), z odpornością na stresy środowiskowe i/lub syntezę rProt w *Y. lipolytica*. Wybrane TFs poddano nadekspresji lub delecji w szczepie wyjściowym – prowadzącym syntezę reporterowego rProt. Szczepy zostały poddane analizie przesiewowej fenotypów w kombinacji różnych czynników środowiskowych, a analiza uzyskanych danych została wsparta modelowaniem matematycznym.

Otrzymane wyniki pokazały, że wzrost oraz synteza rProt w określonych warunkach mogą ulec istotnemu zwiększeniu lub zmniejszeniu w wyniku inżynierii TFs; tym samym potwierdzając słuszność założeń i postawionej tezy. Wskazano czynniki środowiskowe uaktywniające poszczególne TFs, a ich wpływ został opisany matematycznie.

Szczegółowe wyniki dotyczące poszczególnych TFs:

a) Odporność na stres środowiskowy:

i) stres niskiego pH – nadekspresja TF typu Yap łagodzi zahamowanie wzrostu w warunkach wysokiego pH, co sugeruje udział tego TF w procesach zależnych od pH, w tym w dimorficznej transformacji. Wyniki te są zgodne z wcześniejszymi obserwacjami Morales-Vargasa i in. (2012), którzy odnotowali od 3- do 5-krotny wzrost ekspresji tego genu przy przejściu morfologicznym indukowanego zmianą pH z kwaśnego do neutralnego. Białko tego genu zawiera domenę zamka leucynowego, charakterystyczną dla TFs z rodziny Yap w *S. cerevisiae*, znanej z regulacji odpowiedzi na stres środowiskowy. Choć jego dokładny odpowiednik w *Y. lipolytica* nie został

jednoznacznie określony, możliwe że TF typu Yap – YALI0D07744 pełni funkcje zbliżone do Yap1, Yap5, Yap8 lub Yap2 w *S. cerevisiae*, co sugerują obserwacje zależności jego aktywności od pH i dodatkowe eksperymenty wykazujące zwiększoną odporność na stres oksydacyjny.

ii) Stres osmotyczny – delecja *SKN7* i *HSF1* ograniczała wzrost w warunkach stresu osmotycznego, zadanego indywidualnie lub w połączeniu z ograniczoną dostępnością tlenu. Z kolei nadekspresja tych TFs nie poprawiała wzrostu w stresie osmotycznym, pozwalała natomiast utrzymać zdolność do syntezy rProt w tych warunkach. Wyniki te potwierdzają funkcjonalną rolę Skn7 w odpowiedzi osmotycznej i oksydacyjnej, wcześniej postulowaną na podstawie homologii z *S. cerevisiae*, oraz wnioskowaną na podstawie badań nad regulatorem nadrzędnym Hog1 w *Y. lipolytica*. W przypadku Hsf1, mimo braku wcześniejszych doniesień o jego szczególnym udziale w tych procesach w *Y. lipolytica*, jego rola jako globalnego regulatora odpowiedzi na stres jest dobrze znana, a uzyskane wyniki były spodziewane.

b) Zwiększona synteza rProt: geny *GZF1* i *HSF1* jako uniwersalne „wzmacniacze" (ang. *enhancers*) produkcji rProts, niezależnie od zadanych warunków środowiska.

i) Gzf1 – nadekspresja tego TF prowadziła do istotnego zwiększenia syntezy rProt. Jest to nowe odkrycie, ponieważ dotychczas Gzf1 nie był wiązany z produkcją białek. Prawdopodobnie jego rola w tym procesie wynika z udziału w regulacji metabolizmu azotu. Dane Pomraninga i in. (2017) wskazują, że Gzf1 jest najbardziej podobny do aktywatorów genów represjonowanych przez mechanizm represji katabolicznej azotu (NCR, ang. *nitrogen catabolite repression*) u grzybów strzępkowych, a poziom jego ekspresji silnie zależy od dostępności i rodzaju źródła azotu w podłożu. Gzf1 prawdopodobnie uczestniczy w zwiększonym wychwytywaniu i asymilacji azotu, szczególnie w warunkach jego ograniczonej dostępności, co koresponduje z obserwowaną poprawą wydajności syntezy rProts.

ii) Hsf1 – wykazano uniwersalny, pozytywny wpływ nadekspresji *HSF1* na produkcję rProt. Podobny efekt obserwowano wcześniej w *S. cerevisiae* (Hou i in., 2012, 2013, 2014), a uzyskane wyniki potwierdzają jego występowanie także u *Y. lipolytica*. Funkcja Hsf1 jako głównego aktywatora komórkowej zdolności do fałdowania białek i działania chaperonów oraz induktora cytoplazmatycznej formy szlaku UPR wyjaśnia jego rolę we wspomaganiu syntezy rProt.

Badania te dowiodły, że TFs mogą służyć jako efektywne modulatory syntezy rProt oraz zwiększać odporności na wybrane stresy środowiskowe. Co ważne, oprócz walidacji podejścia opartego na inżynierii TF, praca ta dostarczyła także cennej wiedzy o funkcji kilku genów w *Y. lipolytica*, dotychczas nieopisanych funkcjonalnie w tym kontekście. Stanowi to solidną podstawę do szerzej zakrojonych analiz przedstawionych w kolejnej publikacji.

Applied Microbiology and Biotechnology
https://doi.org/10.1007/s00253-023-12607-z

APPLIED GENETICS AND MOLECULAR BIOTECHNOLOGY

Transcription factors enhancing synthesis of recombinant proteins and resistance to stress in *Yarrowia lipolytica*

Maria Gorczyca[1] · **Jean-Marc Nicaud**[2] · **Ewelina Celińska**[1]

Received: 27 March 2023 / Revised: 11 May 2023 / Accepted: 17 May 2023

Abstract

Resistance to environmental stress and synthesis of recombinant proteins (r-Prots) are both complex, strongly interconnected biological traits relying on orchestrated contribution of multiple genes. This, in turn, makes their engineering a challenging task. One of the possible strategies is to modify the operation of transcription factors (TFs) associated with these complex traits. The aim of this study was to examine the potential implications of selected five TFs (*HSF1-YALI0E13948g*, *GZF1-YALI0D20482g*, *CRF1-YALI0B08206g*, *SKN7-YALI0D14520g*, and *YAP-like-YALI0D07744g*) in stress resistance and/or r-Prot synthesis in *Yarrowia lipolytica*. The selected TFs were over-expressed or deleted (OE/KO) in a host strain synthesizing a reporter r-Prot. The strains were subjected to phenotype screening under different environmental conditions (pH, oxygen availability, temperature, and osmolality), and the obtained data processing was assisted by mathematical modeling. The results demonstrated that growth and the r-Prot yields under specific conditions can be significantly increased or decreased due to the TFs' engineering. Environmental factors "awakening" individual TFs were indicated, and their contribution was mathematically described. For example, OE of Yap-like TF was proven to alleviate growth retardation under high pH, while Gzf1 and Hsf1 were shown to serve as universal enhancers of r-Prot production in *Y. lipolytica*. On the other hand, KO of *SKN7* and *HSF1* disabled growth under hyperosmotic stress. This research demonstrates the usefulness of the TFs engineering approach in the manipulation of complex traits and evidences newly identified functions of the studied TFs.

Key points

- *Function and implication in complex traits of 5 TFs in Y. lipolytica were studied.*
- *Gzf1 and Hsf1 are the universal r-Prots synthesis enhancers in Y. lipolytica.*
- *Yap-like TF's activity is pH-dependent; Skn7 and Hsf1 act in osmostress response.*

Keywords Stress response · Environmental stress factor · Heterologous protein · Yeast · Transcription factor · Protein expression platform

Introduction

Recombinant protein (r-Prot) synthesis and stress response are both complex and strongly interconnected biological processes of great industrial importance. Environmental stress factors strongly impact the overall performance of a microbial producer cell, including its growth rate, metabolic activity, and capacity to produce r-Prots. On the other hand, high-level synthesis of r-Prots modulates the cell's ability to cope with unfavorable environmental conditions. Due to the traits' complexity, usually, multiple genes must be fine-tuned to achieve their effective engineering.

The molecular background of the environmental stress response has been extensively studied in the model yeast species – *Saccharomyces cerevisiae* (e.g., Gasch et al. 2000;

✉ Ewelina Celińska
ewelina.celinska@up.poznan.pl

Maria Gorczyca
maria.gorczyca@up.poznan.pl

Jean-Marc Nicaud
jean-marc.nicaud@inrae.fr

[1] Department of Biotechnology and Food Microbiology, Poznan University of Life Sciences, 60-637 Poznań, Poland

[2] Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, 78350 Jouy-en-Josas, France

Published online: 15 June 2023

Gasch and Werner-Washburne 2002; Gasch 2007; Hou et al. 2013; Liu et al. 2013), unrevealing the major molecular players of this process (Craig et al. 1993; Verghese et al. 2012). In addition, it has been evidenced that excessive synthesis of r-Prots leads to endogenous stress, awakening massive cellular response (Mattanovich et al. 2004; Matsumoto et al. 2005; Tyo et al. 2012; Hou et al. 2014; de Ruijter and Frey 2015). Further detailed molecular studies revealed substantial overlap between such environmental and endogenous stress responses (Hahn et al. 2004; Guyot et al. 2005; Verghese et al. 2012; Hou et al. 2013). For example, it was shown that 3% of the *S. cerevisiae* genome was induced by heat shock, out of which 25% was constituted by the proteins involved in translation and polypeptides' secretion (Hahn and Thiele 2004; Hahn et al. 2004). That observation was later supported by evidence that synthetic induction of the heat shock response improved r-Prot synthesis in *S. cerevisiae* (Hou et al. 2013).

Yarrowia lipolytica is a dimorphic yeast species that has gained significant interest as an r-Prots production platform (Nicaud et al. 2002; Groenewald et al. 2014; Madzak 2015, 2018, 2021). Thanks to multiple comprehensive investigations (e.g., Boisramé et al. 1999, 2002; Kabani et al. 2000; Swennen and Beckerich 2007; Babour et al. 2008; Swennen et al. 2010; Madzak and Beckerich 2013; Korpys-Woźniak et al. 2020; Korpys-Woźniak and Celińska 2021), reviewed in Celińska and Nicaud (2019), many mechanistic and functional details of r-Prots synthesis in this species have been revealed. As an industrial workhorse, *Y. lipolytica* is known to efficiently fight against different environmental threats. The molecular background of this resistance/stress response was analyzed in multiple studies (e.g., Yang et al. 2015; Pomraning et al. 2016; Walker et al. 2019; Cogo et al. 2020; Kolhe et al. 2021; Kubiak-Szymendera et al. 2021; Lesage et al. 2021; Sekova et al. 2021), reviewed in Celińska (2022).

Research addressing the mutual impact of r-Prot synthesis and resistance to environmental stress factors in *Y. lipolytica* is still very scarce. Sassi et al. (2017) studied the impact of pH and cell morphology on r-Prot synthesis in *Y. lipolytica* and demonstrated that only the former significantly contributes to the r-Prot yields. Kubiak et al. (2021) optimized conditions of thermal treatment that promote the synthesis of r-Prots in bioreactor cultures, showing that decreased temperature favors the synthesis of r-Prots. Further studies demonstrated that the treatment with decreased temperature specifically relieves secretion of the r-Prots in *Y. lipolytica* (Korpys-Woźniak et al. 2021). The molecular background of that phenomenon was investigated by global proteomics and gene expression analysis (Kubiak-Szymendera et al. 2021). The same report provided data on the impact of hyperosmolality (Osm) on the r-Prot synthesis and ultimately settled that it is not favorable for secretory r-Prots production by *Y. lipolytica* (Kubiak-Szymendera et al. 2021). The mechanisms underlying the globally decreased synthesis of proteins under Osm in *Y. lipolytica* were pointed out in that report. The impact of the key factor limiting *Y. lipolytica* performance in the industry, namely oxygen availability (pO_2), on r-Prots synthesis, was also studied (Gorczyca et al. 2020, 2022). The former study showed that limited pO_2 negatively impacts all – the rate of transcription, translation, and secretion of secretory r-Prots (Gorczyca et al. 2020). It was also established that the r-Prot synthesis rate is not directly related to the rate of biomass growth in *Y. lipolytica*, and the former can be limited when the latter is stably maintained. In the following study, it was demonstrated and quantitatively expressed that the metabolic burden caused by the over-synthesis of two complex r-Prots in *Y. lipolytica* significantly increased substrate consumption, even at a reduced growth rate (Gorczyca et al. 2022). In addition, it was evidenced that the high metabolic burden caused by r-Prots synthesis has a significant impact on the host's stress resistance (pO_2 and pH).

The complex traits can be efficiently engineered by pursuing an adaptive laboratory evolution approach followed by reverse engineering (Portnoy et al. 2011; Dragosits and Mattanovich 2013; Winkler and Kao 2014; Mans et al. 2018) or by applying a directed genetic engineering strategy, but targeting molecular identities operating at a higher level of the molecular events, like signaling cascades or transcription factors (TFs). For example, overexpression of the *HAC1* gene (TF implicated in the regulation of ER (endoplasmic reticulum)-resident events and restoring ER homeostasis) improved r-Prots synthesis in *S. cerevisiae* (Duan et al. 2019), *Pichia pastoris* (Guerfal et al. 2010), and *Y. lipolytica* (Korpys-Woźniak et al. 2021). Titers of different r-Prots were increased by nearly threefold upon overexpression (OE) of different translation initiation factors in *P. pastoris* (Staudacher et al. 2022), and by over fivefold when the polypeptides translocation step was engineered by "pushing" the flux on the cytosolic side and "pulling" on the ER side (Zahrl et al. 2022). Secretory r-Prots production was significantly enhanced in *S. cerevisiae* upon co-overexpression (co-OE) of *HSF1* (heat shock factor) which encodes the key regulator of heat shock response (Hou et al. 2013). Continuous activation of heat shock response by co-OE of a mutant *HSF1*-R206S triggered increased production of native proteins and r-Prots. Co-OE of *HAP1* (responsible for aerobic metabolism and activation of oxidative stress-responsive) mitigated the negative effects of oxidative stress caused by intensive protein folding and hence increased the r-Prots production capacity in *S. cerevisiae* (Martínez et al. 2016). Very recently, synthetic activation of the general stress response TF Msn4 (and its synthetic version synMsn4, alone or in combination) triggered over fourfold enhancement in r-Prot production (Zahrl et al. 2023). Nevertheless, only a limited

number of TFs have been studied in yeast in the context of r-Prots synthesis and the associated stress resistance yet.

While *Y. lipolytica* is a well-recognized r-Prots production platform, TFs implicated in stress response or protein synthesis are still not well described. Most of the literature reports refer to the role of an individual TF in the regulation of another complex trait – dimorphic transition. Involvement in the morphological changes has been proven for TF Msn2 (Pomraning et al. 2018), Yap-like encoded by *YALI0D07744g* (Morales-Vargas et al. 2012), Znc1 (Martinez-Vazquez et al. 2013), Bmh1 (Hurtado and Rachubinski 2002), Mhy1 (Hurtado and Rachubinski 1999; Wu et al. 2019), as well as for Hoy1 (Torres-Guzmán and Domínguez 1997). In addition, TF Msn4 was shown to regulate tolerance to acid-induced stress (Wu et al. 2019). Several reports addressed the issue of TF-driven regulation of lipid accumulation in *Y. lipolytica* (Beopoulos et al. 2009; Nicaud 2012). In this regard, functional studies on individual knockout (KO) genotypes, *Δpor1* (Poopanitpan et al. 2010), *Δmig1* (Wang et al. 2013), *Δyas1* and *Δyas2* (Endoh-Yamagami et al. 2007), *Δyas3* (Hirakawa et al. 2009), and *Δgzf2* and *Δgzf3* (Pomraning et al. 2018), were reported. The impact of constitutive overexpression (OE) of one hundred twenty-five TFs on lipids accumulation in *Y. lipolytica* has been studied through a high-throughput screening approach (Leplat et al. 2018).

The aim of this study was to examine the potential implications of selected five TFs in stress resistance and/or r-Prot synthesis in *Y. lipolytica*. The adopted strategy relied on the co-OE or KO of selected TFs in a host strain synthesizing r-Prot. The strains were subjected to extensive phenotype screening under different environmental conditions, and the data processing was assisted by mathematical modeling. It was presumed that apart from genetic engineering of the TFs-encoding genes, implementation of the environmental perturbations will further manipulate the activation status of the TFs. A practical goal was to evaluate this approach as a rationale engineering strategy to enhance stress resistance and/or r-Prot synthesis in *Y. lipolytica*.

Materials and methods

Microbial strains and basic media

All the strains used in this study are listed in Supplementary Tables S1A (*Y. lipolytica* – final recipient of genetic constructions) and S1B (*Escherichia coli* – cloning and plasmid propagation). The yeast strains are derivatives of the JMY2566 strain, engineered in the *URA3* locus to contain a zeta-docking platform and OE, an intracellular fluorescent (FL) reporter protein (RedStarII) under the p*TEF* promoter (Leplat et al. 2015). JMY2810 strain is a prototrophic derivative of JMY2566 transformed with an empty *URA3* cassette. Strains "OE-TFx" bear an expression cassette encoding one of the selected TFs under the control of the p*TEF* constitutive promoter, integrated at the zeta platform (Leplat et al. 2018). Strains "KO-TFx" were constructed on the platform of the JMY2810 strain by disrupting the indicated locus encoding selected TF with a cassette bearing a *NATr* (nourseothricin) gene flanked with regulatory elements, and approximately 1 kbp of a complementary region on each side of the cassette.

The yeast strains were routinely maintained at 28 °C in rich YPD (g L^{-1}: yeast extract, 5 (BTL, Łódź, Poland); peptone, 10 (BTL); glucose, 20 (POCH, Gliwice, Poland); solidified with agar, 15 (BTL)) or in minimal YNB medium (g L^{-1}: glucose, 10 (POCH); yeast nitrogen base, 1.7 (Sigma-Aldrich, St. Louis, USA); ammonium sulfate, 5 (POCH); solidified with agar, 15 (BTL)). For the selection of recombinant strains bearing *HPHex* (JME4580 plasmid) or *NATr* (KO cassettes) dominant selection marker genes, hygromycin B (Sigma-Aldrich) at 250 mg L^{-1} or nourseothricin (Sigma-Aldrich) at 400 mg L^{-1} was supplemented to YPD medium (liquid or solidified). Liquid cultures were shaken at 220 rpm.

Bacterial strains were routinely maintained at 37 °C (liquid cultures were shaken at 220 rpm) in LB medium (g L^{-1}: Bacto-peptone, 10 (BTL); yeast extract, 5 (BTL); NaCl, 10 (POCH); solidified with agar, 15 (BTL)) supplemented with ampicillin (Sigma-Aldrich) (100 mg L^{-1}) or kanamycin (A&A Biotechnology, Gdynia, Poland) (40 µg L^{-1}), as required.

All the strains were deposited as 15% glycerol stocks at – 80 °C for long-term storage.

Basic molecular biology protocols

Standard molecular biology protocols were used in this study (Sambrook and Russell 2001). All oligonucleotides used here are listed in Supplementary Table S1C. References to backbone plasmids used in this study are given in Supplementary Table S1B. DNA fragments to be cloned were amplified using Phire Hot Start II DNA Polymerase (Thermo Fisher Scientific, Waltham, USA). Colony PCRs were conducted using RUN DNA Taq polymerase (A&A Biotechnology). Plasmid DNA isolation, agarose gel extraction, genomic DNA extraction, and post-reaction purification were all conducted using an appropriate kit from A&A Biotechnology. Restriction enzymes *Bsm*BI and *Bsa*I and T4 DNA ligase were purchased from NEB (New England Biolabs Ltd, Ipswich, USA) and used according to the manufacturer's instructions. *Not*I and *Bgl*II endonucleases were purchased from Thermo Fisher Scientific. The preparation of competent cells and transformation of *E. coli* strains was conducted according to a standard heat shock protocol

(Sambrook and Russell 2001). *Y. lipolytica* strains were transformed using the lithium acetate transformation protocol (Chen et al. 1997).

Construction of Y. lipolytica deletant strains (KO-TFx)

Design and construction of deletion cassettes

The deletion cassette was designed on a GoldenGate scaffold used previously (Celińska et al. 2017; Larroude et al. 2018), limited to three fragments cloning: (i) ARM up, (ii) *NATr*, and (iii) ARM down. ARM up and down are approximately 1 kbp fragments upstream and downstream of the target loci encoding specific TF, flanked with A and B, and C and M overhangs indicated in Celińska et al. (2017). The central fragment, *NATr*, encodes a dominant selection marker gene flanked with regulatory elements and B and C overhangs. The cassettes were assembled using a previous protocol for the GoldenGate reaction (Celińska et al. 2017). White colonies were verified for correctness of the assembly by PCR of adjacent elements and restriction digestion of isolated plasmids. After release from the pSB1A backbone by *Not*I digestion, the cassettes were used for the transformation of *Y. lipolytica* JMY2810.

sgRNA oligonucleotide design and plasmid construction

The sgRNA oligonucleotides were designed using the CRISPR design tool integrated into the Benchling platform (https://benchling.com/). Targeting regions were selected close to the center of the coding sequences to disrupt ORFs (open reading frames) by integration of the deletion cassette. sgRNA oligonucleotides were selected based on their highest efficiency scores and lowest number of off-target sites. The 20-bp-long target sequences were flanked with *Bsm*BI recognition sites and with 4-bp overhangs enabling their correct integration in the JME4580 plasmid, as described previously (Larroude et al. 2020). Prior to cloning into the recipient plasmid, the sgRNA-encoding complementary oligonucleotides were annealed by mixing equimolar amounts of the single-strand oligos, heating at 95 °C for 5 min, and incubation at room temperature for 45 min. Two µL of the diluted oligos (1:100) were mixed with 100 ng of JME4580 plasmid, 2 µL T4 ligase buffer (New England Biolabs), 1 µL *Bsm*BI (New England Biolabs), 0.5 µL concentrated T4 ligase (New England Biolabs), and water up to 20 µL. Such a reaction mixture was subjected to the following cycling profile: (5 min at 55 °C, 5 min at 16 °C) × 40, 5 min at 80 °C. The reaction was then transformed into *E. coli* DH5alpha and transformants selected on LB ampicillin agar plates. White colonies were further used for the DNA constructions propagation, plasmid isolation, and verification of their correctness by PCR and restriction digestion with *Bgl*II.

Co-transformation of the deletion cassettes and sgRNA-encoding episomal plasmids

Y. lipolytica strains deleted in one of the selected TFs were generated by co-transformation of the deletion cassette and a compatible sgRNA-encoding episomal plasmid. Approximately 500 ng of each DNA construction was used per single transformation. The deletant strain selection was conducted according to the previously described methodology (Larroude et al. 2020). Briefly, the co-transformation reactions were inoculated into 9 mL of YPD-hygromycin-nourseothricin liquid medium and cultured at 28 °C for 48 h with shaking at 150 rpm. One mL of such cultures was then transferred into 9 mL of YPD-nourseothricin medium and incubated at 28 °C for 24 h with shaking at 150 rpm to allow plasmid curing (dropping-off JME4580). Finally, the culture was diluted and plated on YPD-nourseothricin agar. Clones appearing after 48 h of incubation at 28 °C were verified for correct integration of the deletion cassettes by PCR and sequencing.

Phenotype analyses

Miniaturized cultures under selected stress factors

The cultures for the OE and KO strains' phenotype testing were performed in 24-well plates (Biologix, Shandong, China) in 1-mL working volume. Medium for the cultures was carefully adjusted to minimize background FL, provide sufficient amounts of carbon and nitrogen to avoid starvation, and maintain the pH stably throughout the culturing time. The basic medium was composed of [g L^{-1}]: yeast nitrogen base, 5.1 (Sigma-Aldrich); $(NH_4)_2SO_4$, 15, (POCH); glucose, 20 (POCH); buffered with 0.2 M maleic acid (Sigma-Aldrich) to pH 5.0 by addition of 20% NaOH (w/w) (POCH). Pre-cultures were developed in the basic medium for 18 h at 28 °C with shaking at 380 rpm in a microplate shaker-incubator (Biosan, Riga, Latvia). The main cultures were inoculated at 5% (v v^{-1}).

The main cultures implementing stress factors were designed to introduce two categoric stress factors: (i) oxygen availability (pO_2) stress, by a change in shaking frequency (~25% reduction in rpm, from 380 to 280 rpm), (ii) hyperosmolality (Osm) stress, by the addition of sorbitol to the medium (360 g L^{-1}, resulting in a 3.75-fold increase in osmolarity from 0.8 to 3 Osm kg^{-1}); and two numeric stress factors: (i) three levels of cultivation temperature (22 °C, 27 °C, or 34 °C); (ii) three pH levels (pH 3.0, 5.0 or 7.0). The pH was stably maintained at the selected levels with 0.1 M maleic acid to pH 3.0 or 0.2 M maleic acid to pH 5.0

or 7.0, regulated by the addition of 20% NaOH (w w^{-1}). Figure 1 illustrates the experimental setup. All the media were filter-sterilized with 0.22-µm syringe filters (Merck-Millipore, Darmstadt, Germany). The cultures were carried out for 72 h, and samples were collected at 0 h, 24 h, 48 h, and 72 h. pH and osmolarity (only high Osm variants) were additionally verified at the final time-point. All the above-mentioned variants were tested in combination, resulting in thirty-six cultivation variants for each tested strain (see Fig. 1). Cultures were performed in biological triplicates for each strain and condition.

Cultures were analyzed for cellular growth and FL from the reporter protein (RedStarII). Cell growth was measured spectrophotometrically immediately after sampling. The samples were diluted in 0.75% NaCl (POCH) to match a linear range of the methods. Absorbance was measured at 600 nm in transparent 96-well plates (Corning®Costar®, Sigma-Aldrich). FL was determined under pre-optimized ex/em settings to maximize discrimination between negative control samples (endogenous FL from cells, without RedStarII) and FL reporter-producing cells and to limit the background FL from cultivation media. FL was measured in black opaque plates (Thermo Fisher Scientific) at ex/em 554/600 nm. Both measurements were done using an Infinite M200 automatic plate reader (Tecan Group Ltd., Männedorf, Switzerland).

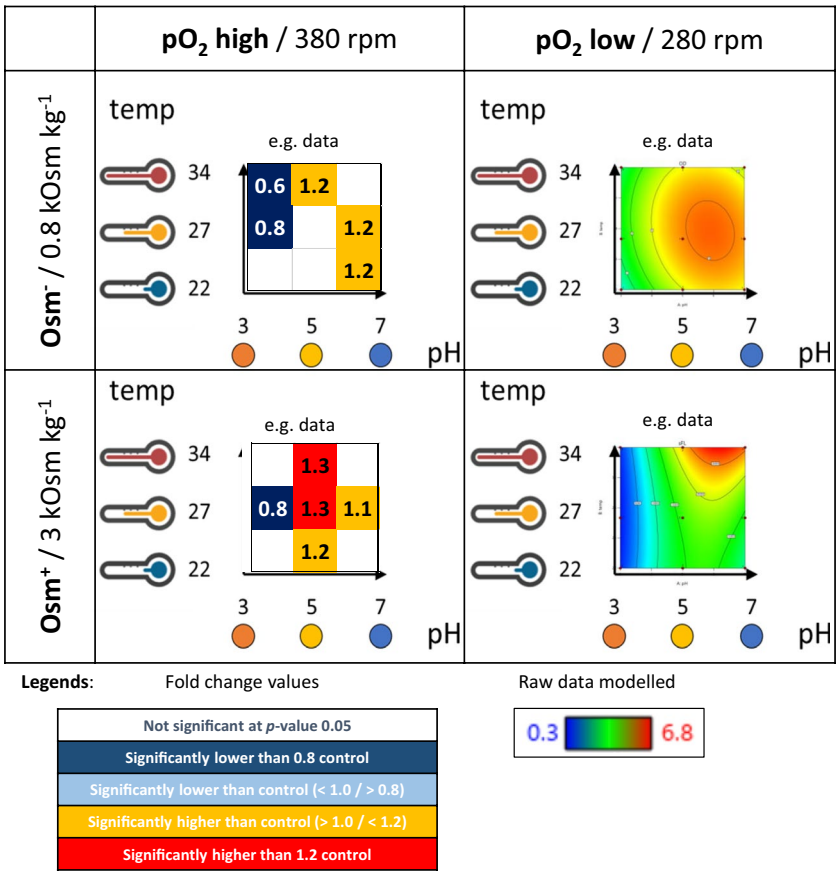


Fig. 1 Schematic representation of experimental design and data presentation scheme of the phenotype testing. A combination of the two categoric factors (pO_2 – oxygen availability at two levels – high 380 rpm/low 280 rpm; Osm – hyperosmolality at two levels – high 3 Osm kg^{-1}/low 0.8 Osm $kg.^{-1}$) results in the division of the design into four subpanels. Each subpanel represents data for a given response under a combination of the two categoric factors at one of the two levels. Data in the subpanels are presented as quadratic response surface (contour plot), where the temperature is plotted on the *y*-axis and pH on the *x*-axis, both ascendingly (from 22 to 34 °C and from pH 3 to pH 7). Contour lines join points of equal value. Results for the control strain are presented as contour plots solely. Data for the TF-engineered strains are presented in the form of a contour plot and a corresponding box plot (the same structure). Box plots indicate fold changes (FC) of a read parameter value for a given TF-engineered strain and the control strain under specific conditions. The two legends show color-coding of the contour plots and box plots. Values in the contour plots are color-coded from green – the lowest, through yellow, to red – the highest. Values in the box plots are color-coded from navy – significantly lower than the control by more than 20%, through blue – significantly lower than the control by < 20%, through yellow – significantly higher than the control by < 20%, to red – significantly higher than the control by 20%. The structure of the scheme of data presentation corresponds to the structure of the data presented in Figs. 2, 3, and 4

Neutral lipid storage analysis

Pre-cultures were developed in 2-mL YNB medium ([g L^{-1}]: yeast nitrogen base, 1.7; $(NH_4)_2SO_4$, 5; glucose, 20) for 23 h under 30 °C with shaking 200 rpm. The main cultures were inoculated at 5%. Two medium variants were used: (i) medium Lip^+ of C/N ratio 150 ([g L^{-1}]: glycerol, 150; $(NH_4)_2SO_4$, 1; KH_2PO_4, 1, $MgCl_2x6H_2O$, 0.5), (ii) medium Lip^- of C/N ratio 4 ([g L^{-1}]: glycerol, 20; $(NH_4)_2SO_4$, 5; KH_2PO_4, 1, $MgCl_2x6H_2O$, 0.5). The main cultures were conducted for 48 h under 27 °C with shaking at 150 rpm. Nile Red (Sigma-Aldrich) reagent was diluted in dimethyl sulfoxide (DMSO) to reach 25 mg mL^{-1} and diluted to 100 µg mL^{-1}. Forty-eight-hour cultures were pelleted, washed in sterile saline solution, diluted tenfold, and stained with Nile Red (working concentration 12.5 µg mL^{-1}; 20 min shaking at 30 °C). Readings were done using black opaque, clear bottom 96-well plates (Thermo Fisher Scientific), using an Infinite M200 automatic plate reader (Tecan) at ex/em 488/585 nm wavelength. Control reactions without cells and without Nile Red were run simultaneously to account for non-specific background FL. Absorbance at 600 nm wavelength was measured to express the results in specific FL (sFL) units.

Data processing and statistical analyses

All the results on cellular growth and FL are mean values ± standard deviation (from three biological replicates and technical duplicate). sFL was calculated by dividing the value for RFU (relative fluorescence units) per OD_{600} value. Fold change (FC) values were calculated by dividing the raw data for OD_{600}, FL, or sFL for the OE/KO strain ($n=3$) by corresponding results for the control strain ($n=3$), in combination, resulting in nine FC values, out of which the mean was calculated. Statistical analyses were performed in Statistica (StatSoft-Tibco, Tulsa, USA), the analysis of variance (ANOVA) with a significance level of $p<0.05$, preceded by Shapiro–Wilk's and Levene's tests to test presumptions and followed by Tukey's HSD multiple comparisons test.

Mathematical models describing growth, FL, and sFL of the analyzed strains under the analyzed conditions were developed using response surface methodology (RSM) in Design Expert software (StatSoft). Raw results (not FC) for growth, FL, and sFL were used in the mathematical modeling. The mode of data transformation, type of model used, fitting, and significance of the model were assessed using default software settings.

FC values were used to present the experimental data as box plots. A twenty % change over the reference strain was considered a considerable change.

Results

Experimental plan

In this study, the experimental design covered a set of five OE strains (OE-YAP-like, OE-SKN7, OE-GZF1, OE-HSF1, OE-CRF1) and three KO strains (KO-SKN7 = *Δskn7*, KO-GZF1 = *Δgzf1*, KO-HSF1 = *Δhsf1*) that were subjected to extensive phenotype testing under thirty-six conditions. The list of engineered TFs with available functional annotations is given in Table 1. The strains were engineered to over-express (OE-)/contain an insertion (KO-) in a gene encoding a single TF (from among those identified in the *Y. lipolytica* genome (Leplat et al. 2018) and an FL reporter protein (RedStarII) to investigate the effects of the TFs manipulation on r-Prots synthesis. Selection of the five TFs was guided by initial high-throughput screens of OE strains conducted in 96-well Micro-Titer Plates (MTP) in different media formulations and under different experimental conditions (data not shown – screens in 96-well plates were prone to high variability and many intrinsic limitations). The following experimental data on the most interesting OE phenotypes guided the choice of the TFs to be deleted in the parental strain (over-expressing the RedStarII reporter).

The set of eight strains engineered in the TFs and a prototrophic control strain were subjected to a pre-designed experimental plan of phenotype testing under variable

Table 1 TFs engineered in this study with the available annotation

TF name	TF gene ID	UniProt	Functional annotation GRYC/UniProt if available
Yap-like TF	*YALI0D07744g*	Q6C9W7	Weakly similar to UniProt|Q9P5L6 *Neurospora crassa* related to AP-1-like transcription factor
Skn7	*YALI0D14520g*	Q6C937	Weakly similar to UniProt|P38889 *Saccharomyces cerevisiae* YHR206W (ortholog of YJR147W) Skn7 transcription factor, response regulator receiver of two components sensing
Gzf1	*YALI0D20482g*	Q6C8D8	Some similarities with UniProt|P78688 *Gibberella fujikuroi* nitrogen regulatory protein AreA
Hsf1	*YALI0E13948g*	Q6C5Z0	Some similarities with UniProt|P10961 *Saccharomyces cerevisiae* YGL073w Hsf1 heat shock transcription factor
Crf1	*YALI0B08206g*	P45815	UniProt|P45815 *Yarrowia lipolytica* YALI0B08206g Crf1 Copper resistance protein UniProt: copper resistance protein Crf1; transcriptional regulator involved in resistance to high copper concentration

pH (numeric factor; "pH"), temperature (numeric factor; "temp"), medium osmolality (categoric factor; "Osm"), and oxygen availability (categoric factor; "pO_2"; see Fig. 1). Ranges and methods of the variables implementation were first carefully adjusted in a series of preliminary studies, covering (i) minimal volume of the culture that enables reasonable repeatability between independent runs, (ii) type of buffer that covers the desired range and stably maintains the desired pH level until the end of the cultures, (iii) range of mixing rate that is sufficient/limiting growth, as desired in a specific culture variant, (iv) medium recipe that will not introduce high background FL noise in the RedStarII ex/em channel, but will support high discrimination between FL + and FL- cells, and (v) sampling time that enables capturing the differences between varying phenotypes, before any environmental factor becomes uniformly limiting to all of the strains (not resulting from the modified genotype).

Growth and FL were monitored in 24-h intervals (time-point data are presented in Supplementary Fig. S1A–C). Raw data were used for mathematical modeling of the response surface illustrating observed and predicted behavior of the read parameters in a 3-dimensional space within the range of the conditions tested (Figs. 2, 3, and 4). The generated mathematical models enable evaluation of each specific parameter's significance (temp, pH, Osm, pO_2) for the investigated response (growth, FL, sFL), as well as the impact of the parameters interactions. These evaluations are given in Table 2. When interpreting the data from Table 2, it is important to note that the values given there correspond to the relative impact of the factor on the response (the higher value – the higher impact); but since the responses reach

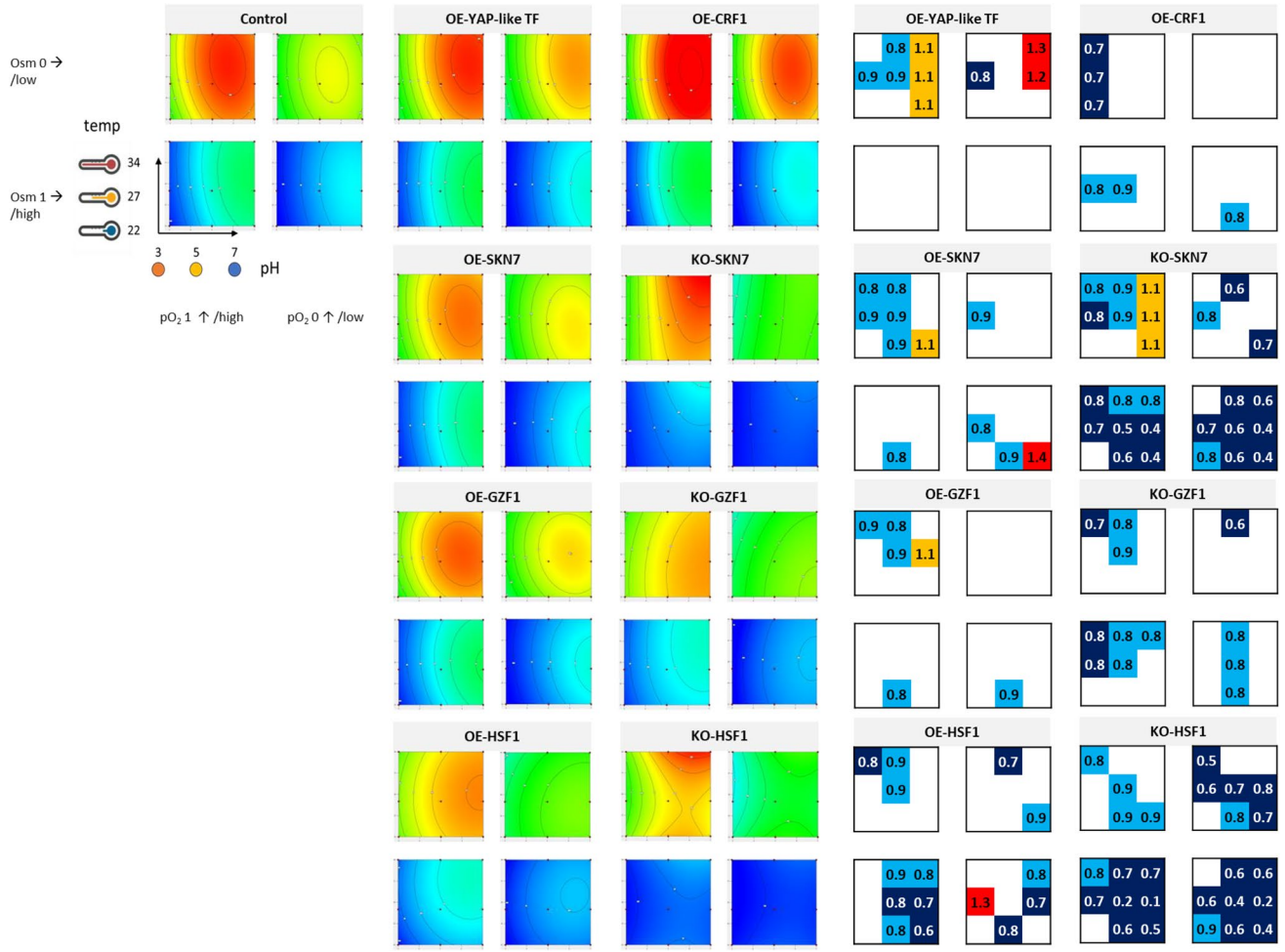


Fig. 2 Growth of TF-engineered *Y. lipolytica* strains and a control strain under a set of the analyzed conditions. Growth is represented in the form of contour plots modeled based on raw experimental data (biological $n = 3$; technical $n = 2$) and in the form of box plots. Values in the contour plots are color-coded, as explained in Fig. 1. Numbers in the box plots correspond to fold changes (FC) of the result read for a given TF-engineered strain over the control strain under specific conditions (mean of $n = 9$). Numbers are indicated only when a result is significantly different compared to the control strain. Color code is explained in the legend of Fig. 1**.** Data presentation and reading is explained schematically in Fig. 1

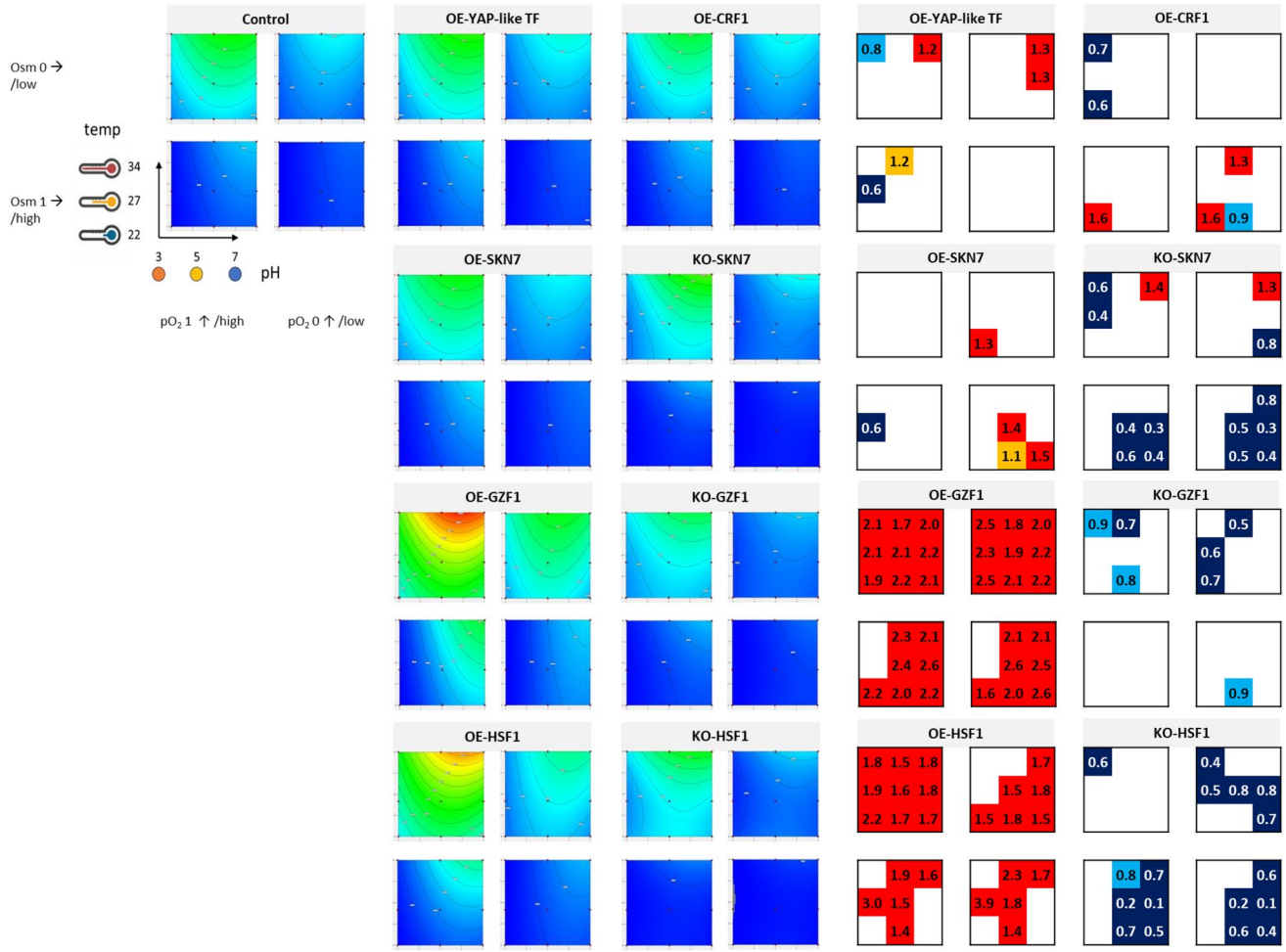


Fig. 3 Total FL denoting total r-Prot synthesis by the TF-engineered *Y. lipolytica* strains and a control strain under a set of the analyzed conditions. FL is represented in the form of contour plots modeled based on raw experimental data (biological $n=3$; technical $n=2$) and in the form of box plots. Values in the contour plots are color-coded, as explained in Fig. 1. Numbers in the box plots correspond to fold changes (FC) of the result read for a given TF-engineered strain over the control strain under specific conditions (mean of $n=9$). Numbers are indicated only when a result is significantly different compared to the control strain. Color code is explained in the legend of Fig. 1. Data presentation and reading is explained schematically in Fig. 1

significantly different ranges for different strain-conditions combination (especially FL), they should not be compared between the strains. Direct comparisons of the responses from the TF-engineered strains over the control strain are shown as box plots (Figs. 2, 3, and 4). Box plots illustrate the prevalence or inferiority of the tested OE-/KO- strains over the control strain, given as FC under specified conditions (Figs. 2, 3, and 4).

Growth of the TFs-engineered Y. lipolytica strains under a set of conditions

Growth of the control strain across the conditions tested is presented in Fig. 2 as contour plots and kinetically in Supplementary Fig. S1A. As expected, the highest growth of the control strain was reached under high pO_2 and no Osm. Under such conditions, a slight preference toward temperature > 25 °C and pH > 4.0 was observed, but the impact of pH was significantly higher than that of temp (0.2951 vs. 0.0405; Table 2). The decrease in pO_2 at the adopted level caused significant growth limitation (impact weight 0.1501). Under Osm, the growth was significantly impeded (the highest negative impact: − 0.5181), especially when combined with low pH (significant negative impact of pH-Osm interaction, − 0.1305; Table 2).

Osm was the most detrimental parameter significantly impacting the growth of all the TF-engineered strains, followed by pH (Table 2, Fig. 2). The latter had the greatest relative impact on OE-YAP-like TF and OE-CRF1 strains' growth (ratio pH:Osm). The modeled peaks of growth for the strains KO-HSF1 and KO-SKN7 were predicted to be at higher temperatures rather than for the remaining strains

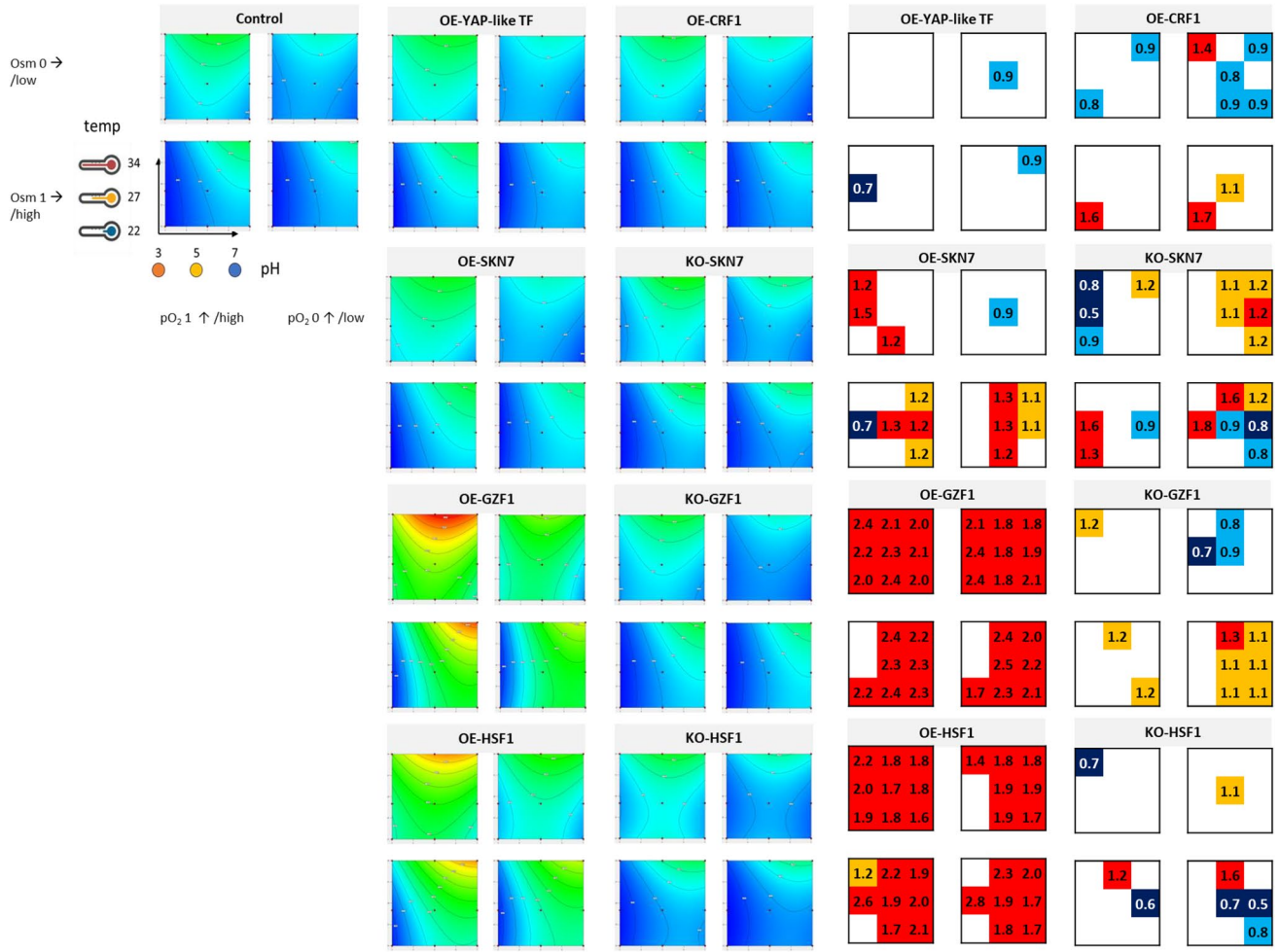


Fig. 4 Normalized FL denoting specific r-Prot synthesis (protein synthesis capacity) by the TF-engineered *Y. lipolytica* strains and a control strain under a set of the analyzed conditions. sFL is represented in the form of contour plots modeled based on raw experimental data (biological $n=3$; technical $n=2$) and in the form of box plots. Values in the contour plots are color-coded, as explained in Fig. 1. Numbers in the box plots correspond to fold changes (FC) of the result read for a given TF-engineered strain over the control strain under specific conditions (mean of $n=9$). Numbers are indicated only when a result is significantly different compared to the control strain. Color code is explained in the legend of Fig. 1. Data presentation and reading is explained schematically in Fig. 1

(Fig. 2), but as estimated by the model, the temp had no significant impact on the growth of KO-HSF1, and only minimal on KO-SKN7 (0.061; Table 2). In contrast, the interaction between the temp and pH gained significance for these two strains and, in addition, positively impacted their growth (+0.0763 and +0.0720), in contrast to what was observed for the control strain (−0.0401; Table 2). But generally, the temp as an individual parameter was the least significant variable impacting growth within the adopted range. Oxygen availability was particularly significant for the growth of KO-GZF1 and OE-HSF1 strains. Interestingly, the interaction between pO_2-Osm was significant only for the TF-engineered strains (with different weights), but not for the control.

Direct comparison of growth of the TF-engineered strains and the control strain (box plots in Fig. 2) indicated that none of the TF's engineering improved growth under Osm, except for two isolated observations for OE-SKN7 (temp 22 °C, pH 7, pO_2 0, Osm 1) and OE-HSF1 (temp 27 °C, pH 3, pO_2 0, Osm 1). Deletion of *SKN7* and *HSF1* had a detrimental impact on the strains' resistance to Osm, irrespectively of inflicted pO_2 stress, pH, or adopted temp. The effect of *GZF1* KO was less uniform, but also was only limiting the growth. In addition, the deletion of *HSF1* severely impacted the growth of the strain when Osm was not implemented, or even, when the conditions were favorable for growth (temp 27 °C, pH 5, pO_2 1, Osm 0; upper left box). In fact, most of the conducted engineering in the TFs, either OE or KO, negatively impacted the growth of the strains under the non-stressed condition. The only exception was the OE of *CRF1*, which, at the same time, exerted a specific negative impact on growth under low pH (all temp, pH 3, pO_2 1, Osm 0;

Table 2 Ranking of variable factors and interactions impacting phenotype of the TF-engineered strains and the control strain

Control	OE-YAP-like TF	OE-SKN7	KO-SKN7	OE-GZF1	KO-GZF1	OE-HSF1	KO-HSF1	OE-CRF1	Strain
Response: **growth**									Model R^2
0.9433	0.9632	0.9754	0.9653	0.9678	0.9650	0.9669	0.9584	0.9719	
Osm -0.5181	**Osm** -0.5122	**Osm** -0.5113	**Osm** -0.5946	**Osm** -0.5172	**Osm** -1.59	**Osm** -1.60	**Osm** -0.6060	**Osm** -0.5134	→ Ranking of factors and interactions →
pH +0.2951	**pH** +0.4002	**pH** +0.3491	**pH** +0.2464	**pH** +0.3268	**pH** +0.7478	**pH** +0.6430	**pH** +0.1954	**pH** +0.3693	
pH² -0.2433	**pH²** -0.2211	**pH²** -0.1934	**pH²** -0.1831	**pH²** -0.1908	**pO₂** +0.5333	**pO₂** +0.5518	**pH²** -0.1936	**pH²** -0.3168	
pO₂ +0.1501	**pO₂** +0.1167	**pH-Osm** +0.1184	**pO₂** +0.1698	**temp²** -0.1259	**pH²** -0.3831	**pH²** -0.4130	**pO₂** +0.1810	**pO₂** +0.1040	
pH-Osm +0.1305	**pH-Osm** +0.0903	**pO₂** +0.1094	**pH-temp** +0.0763	**pH-Osm** +0.1257	**pO₂-Osm** -0.1962	**temp²** -0.2890	**temp²** +0.1235	**temp²** -0.0974	
temp² -0.0822	**temp²** -0.0824	**temp²** -0.0818	**temp** +0.0610	**pO₂** +0.1145	**temp²** -0.1700	**pO₂-Osm** -0.2605	**pO₂-Osm** -0.0855	**pH-Osm** +0.0742	
temp +0.0405	**temp** +0.0412	**pH-temp** -0.0441	**pO₂-Osm** -0.0504	**pH-temp** -0.0426	**temp-Osm** +0.1624	**temp-Osm** +0.1674	**pH-temp** +0.0720	**temp** +0.0498	
pH-temp -0.0401	**pO₂-Osm** +0.0387	**temp-pO₂** +0.0257	**temp-pO₂** +0.0437	**pO₂-Osm** +0.0268	**temp** -0.1033	**temp-pO₂** +0.1215	**temp-pO₂** +0.0374	**pH-pO₂** +0.0433	
pH-pO₂ +0.0226	**pH-temp** -0.0377	**pO₂-Osm** +0.0199	**temp-Osm** +0.0430	**temp** +0.0261	**temp-pO₂** +0.0916	**pH-Osm** -0.0868	**temp** +0.0289	**pO₂-Osm** +0.0424	
temp-pO₂ +0.0220	**temp-Osm** -0.0363	**temp** +0.0053	**pH-pO₂** +0.0375			**temp** +0.0207		**pH-temp** -0.0304	
Response: **FL**									Model R^2
0.9418	0.9623	0.9628	0.9642	0.9590	0.9442	0.9567	0.9535	0.9650	
Osm -46.14	**Osm** -47.39	**Osm** -45.89	**Osm** -51.33	**Osm** -63.38	**Osm** -40.05	**Osm** -58.04	**Osm** -52.03	**Osm** -41.89	→ Ranking of factors and interactions →
pH +34.65	**pH** +42.45	**pH** +38.65	**pH** +36.56	**pH** +51.57	**pH** +36.06	**pH** +42.20	**pH²** -31.85	**pH** +36.33	
pH² -32.83	**pH²** -31.74	**pH²** -29.08	**pH²** -30.84	**pH²** -42.32	**pH²** -24.95	**pH²** -38.65	**pH** +26.12	**pH²** -35.77	
temp +26.27	**temp** +25.21	**temp** +23.99	**temp** +29.11	**temp** +33.75	**pO₂** +23.18	**temp** +35.23	**pO₂** +24.51	**temp** +27.64	
pO₂ +23.13	**pO₂** +22.49	**pH-Osm** +22.36	**pO₂** +21.83	**pO₂** +31.73	**temp** +20.95	**pO₂** +34.73	**temp²** +23.52	**pO₂** +18.07	
pH-Osm +18.67	**pH-Osm** +15.72	**pO₂** +22.18	**pH-temp** +20.59	**pH-Osm** +31.22	**pH-Osm** +14.75	**pH-Osm** +16.18	**temp** +22.88	**pH-Osm** +15.95	
temp-Osm -11.02	**temp-Osm** -10.96	**temp-pO₂** +8.38	**temp²** +14.71	**temp-Osm** -12.12	**pH-temp** +12.44	**pH-temp** +15.80	**pH-temp** +16.01	**temp-Osm** -9.29	
temp-pO₂ +8.53	**pH-temp** +10.09	**temp-Osm** -8.33	**pO₂-Osm** -9.86	**temp-pO₂** +11.98	**pO₂-Osm** -9.69	**temp-pO₂** +14.87	**pO²-Osm** -13.25	**pH-pO²** +8.77	
pH-temp +8.20	**temp-pO₂** +7.55	**pH-temp** +8.04	**pH-pO₂** +9.26	**pO₂-Osm** -9.78	**temp²** +8.60	**pO₂-Osm** -14.36	**temp-pO₂** +7.44	**pH-temp** +7.69	
pO₂-Osm -7.51	**pH-pO₂** +7.03	**pO₂-Osm** -7.48	**temp-pO₂** +8.64	**pH-temp** +9.43	**temp-pO₂** +7.48	**temp-Osm** -4.43	**temp-Osm** -3.74	**temp-pO₂** +5.46	
pH-pO₂ +4.52	**pO₂-Osm** -4.94	**pH-pO₂** +5.50	**temp-Osm** -4.42	**pH-pO₂** +8.51	**temp-Osm** -5.65			**pO₂-Osm** -3.11	
Response: **sFL**									Model R^2
0.8948	0.9143	0.9161	0.9084	0.8949	0.8907	0.9216	0.8866	0.8752	
pH² -14.29	**pH²** -2021.95	**temp** +2188.59	**pH²** -17.30	**pH²** -383.15	**pH** +16.95	**temp** +547.64	**pH²** -17.61	**pH²** -13.59	→ Ranking of factors and interactions →
pH +14.15	**temp** +1936.70	**pH-Osm** +1992.43	**pH** +16.39	**pH-Osm** +328.38	**pH²** -15.17	**pH²** -454.61	**temp** +13.66	**pH-Osm** +13.00	
temp +12.43	**pH** +1793.44	**pH²** -1971.45	**temp** +14.73	**pH** +324.73	**pH-Osm** +13.68	**pH** +438.14	**pH** +12.62	**temp** +12.42	
pH-Osm +12.41	**pH-Osm** +1523.17	**pH** +1933.39	**temp²** +8.42	**temp** +302.97	**temp** +12.64	**pH-Osm** +339.80	**temp²** +12.39	**pH** +12.10	
Osm -6.95	**pH-temp** +1142.34	**pH-temp** +1151.84	**pH-Osm** +8.27	**pO₂** +156.94	**temp²** +6.15	**temp²** +237.63	**pH-Osm** +8.64	**Osm** -5.48	
pH-temp +5.20	**pO₂** +1051.22	**pO₂** +1022.93	**pH-temp** +8.19	**pH-temp** +142.13	**pH-temp** +6.01	**pO₂** +225.62	**Osm** -6.77	**pH-temp** +5.25	
temp² +5.03	**Osm** -991.20	**Osm** -810.29	**Osm** -6.23	**temp²** +139.79	**pO₂** +5.11	**pH-temp** +179.24	**pO₂** +5.44	**pO₂** +3.93	
pO₂ +4.94	**temp²** +714.55	**pO₂-Osm** -706.82	**pO₂** +3.24	**Osm** -134.84	**pO₂-Osm** -3.96	**temp-pO₂** +127.82	**pH-temp** +5.43	**temp²** +3.91	
pO₂-Osm -3.16	**temp-pO₂** +643.63	**temp²** +644.12	**pH-pO₂** +3.04	**pO₂-Osm** -97.68	**Osm** -3.93	**pO₂-Osm** -125.84	**pO₂-Osm** -2.53	**temp-Osm** -3.17	
temp-pO₂ +2.49	**pO₂-Osm** -526.65	**temp-pO₂** +571.75	**pO₂-Osm** -2.76	**temp-pO₂** +86.85	**temp-Osm** -2.43	**Osm** -118.99	**temp-Osm** +1.63	**pO₂-Osm** -2.84	
temp-Osm -1.76	**pH-pO₂** +427.38	**temp-Osm** -344.50		**temp-Osm** -56.65	**temp-pO₂** +1.90			**pH-pO₂** +2.75	
pH-pO₂ +1.10	**temp-Osm** -357.14	**pH-pO₂** +279.62		**pH-pO₂** +50.15					

The impact of the factors and interactions was evaluated for three responses: growth, fluorescence (FL), and FL normalized per biomass (sFL; specific FL). A significant positive/negative impact of a given factor/interaction is shown in red/blue. Values indicate the relative contribution of a given factor/interaction on the response. The ranking is given from the most significant factors with the highest contribution (top) to the least significant or not significant (bottom; white fields). The positive impact is indicated by a positive value and the negative impact "−" as a negative value. All the values are based on developed mathematical models. All the models were significant and well-fitted to the experimental data, as indicated by the R^2 values

upper left box). Interestingly, the OE of the Yap-like TF-encoding gene triggered a uniform improvement in growth under pH 7 (all temp, pH 7, pO_2 1/0, Osm 0; upper boxes). Correspondingly, pH had a high-weighted impact on the two OE strains' growth (Table 2 and mentioned above). A significant, positive impact of higher pH was observed upon deletion of *SKN7*, but solely under high pO_2 provision (all temp, pH 7, pO_2 1, Osm 0; upper left box).

Global amounts of r-Prot synthesized by the TFs-engineered Y. lipolytica strains under a set of conditions

The shape of contour plots illustrating FL from the r-Prot reporter is different from those illustrating growth (Fig. 3 vs. Figure 2), which indicates that growth was decoupled from r-Prot synthesis, even though a constitutive p*TEF* promoter was governing expression of the reporter r-Prot. All the individual variables had a significant impact on FL (Table 2). Under higher pO_2 and no Osm infliction, a clear positive impact of temperature > 27 °C on the FL level was observed, which holds valid for all the strains (temp-pO_2, significant interaction with positive impact for all strains). The combination of Osm and low pO_2 severely limited the synthesis of the reporter r-Prot in all the strains (pO_2-Osm, significant interaction with negative impact), but its relative contribution was the highest for the KO strains (position in the ranking in Table 2). Such an observation suggests that all the TFs studied here in the KO genotype contribute to maintaining r-Prot synthesis under such severe stress. Under the infliction of Osm (the most detrimental individual impact on FL; Table 2), any synthesis of the r-Prot was possible solely under pH > 5. Probably, the effort to fight against Osm and low pH was too high, and the synthesis of r-Prot was switched off (under high stress, this is one of the first biological functions arrested within the cell (Kubiak-Szymendera et al. 2021)). Interestingly, this specific interaction pH-Osm was either the most significant interaction in the ranking (Table 2) for the control strain, OE-YAP-like, OE-SKN7, OE-GZF1, KO-GZF1, OE-HSF1, and OE-CRF1, or not significant at all for KO-SKN7, KO-HSF1, even though the individual impact of Osm and pH on FL was the highest for all the strains. For the majority of the strains, the temp was the third most significant term in the developed models (Table 2), with only two exceptions, KO-GZF1 and KO-HSF1, for which pO_2 was more significant than temp.

Direct comparison of the FL values read for the TF-engineered strains and the control strain (box plots) demonstrated significant, uniform superiority of the OE-GZF1 and OE-HSF1 over the latter in terms of the r-Prot synthesis under most of the conditions (Fig. 3). In contrast, KO-HSF1 and KO-SKN7 showed severe underperformance in terms of FL once Osm was inflicted and pH > 5. This effect was concomitant with limited growth (Fig. 2). No such pH dependency was observed for the KO-GZF1 strain (Fig. 3, Table 2). OE of *CRF1* and *SKN7* caused an increase in the FL upon several specific combinations of the stress factors and Osm (Fig. 3). For example, the OE of *SKN7* enabled maintaining of r-Prot synthesis under Osm and low pO_2 (significant, limiting interaction); on the other hand, *CRF1* OE caused an increase in FL under low temp and pH, and Osm infliction, that sustained irrespectively of the pO_2 factor. It is particularly interesting, considering that this combination of environmental factors is the "most stressful" from among adopted (especially pH 3 and Osm). OE of the Yap-like TF-encoding gene was concomitant with enhanced synthesis of r-Prots under pH > 5, which was associated with increased growth under these conditions (Figs. 2 and 3).

r-Prot synthesis capacity (sFL) of the TFs-engineered Y. lipolytica strains under a set of conditions

Contour plots illustrating calculated sFL values (Fig. 4) resemble those for FL, with several differences. Normalization per biomass highlighted the strains' biological response in terms of r-Prots synthesis under Osm (which was not clear in the FL; Fig. 3) and substantially changed the ranking of the most significant variables (Table 2). The impact of Osm became less important, while the impact of temp gained significance, becoming the most (OE-SKN7, OE-HSF1) or second/third most (control, OE-YAP-like, KO-SKN7, KO-HSF1, OE-CRF1) significant factor. The sFL measure was relatively the least impacted by the temp for *GZF1*-engineered strains (ranking in Table 2), even though the contour plots for OE-GZF1 strain clearly indicate a positive impact of the temperature > 30 °C on sFL (Fig. 4). For these strains, pH either alone or in combination with Osm had the highest impact on sFL value. The pH-Osm followed by pH-temp remained the most significant interactions impacting sFL, as it was observed for FL (Table 2). In terms of sFL, pH-Osm became a significant interaction also for the strains KO-SKN7 and KO-HSF1, which is not true in terms of the FL for these strains (Fig. 3).

Upon direct comparison with the control strain (box plots), it appeared that OE of the *YAP-like* gene did not bring any significant positive impact on the strains' potential toward r-Prot synthesis (Fig. 4). Any increase in the FL was a result of increased growth (Figs. 2 and 3). In contrast, OE of *GZF1* and *HSF1* brought spectacular improvement in the cells' capacity toward r-Prots synthesis. The sFL measures for these strains were uniformly either better or comparable to the control. Surprisingly, under specific conditions (Fig. 4), KO of these genes also

triggered increased sFL measure, especially seen for KO-GZF1 under Osm and KO-HSF1 under Osm (34 °C, pH 5). As in the case of FL measures, sFL readouts were comparatively higher for the OE-CRF1 strain over the control under Osm infliction, low temp, and pH (Figs. 2 and 4).

Normalization of FL per biomass enabled interesting observation of the *SKN7*-engineered strains (Fig. 4). Namely, it appeared that the relative level of sFL was significantly higher for the OE-SKN7 strain vs. the control under Osm (FC > 1.1) for 50% of the culture variants, but only when pH > 5. So, it could be concluded that OE of *SKN7* improves the protein synthesis capacity under Osm and non-stressful pH. KO of *SKN7* promoted the strains' synthesis capacity, especially under no Osm and low pO_2. Maintaining a high level of r-Prot synthesis capacity under Osm proves Skn7's implication in the osmostress response.

Neutral lipids storage capacity of the TFs-engineered Y. lipolytica strains

In an auxiliary experiment, we tested the TF-engineered strains for their capacity to store neutral lipids. To this end, the strains were cultured in two media differing in C/N ratio to promote or prevent extensive lipid accumulation (Lip^+: C/N ratio 150, Lip^-: C/N ratio 4). The results of lipid content expressed in sFL units are presented in Fig. 5. The most striking observation relates to strains KO-GZF1 and KO-HSF1, for which the pattern of lipid storage was inverted, as these strains accumulated significantly more lipids under Lip^- conditions. The same, but to a lesser extent, was observed for the *SKN7*-engineered strain. OE of any – *GZF1, HSF1*, or *CRF1* contributed to an increase in lipid content in the cell, but in these cases – according to the expected pattern (higher in Lip^+, lower in Lip^-).

Discussion

This study aimed to investigate the possibility of engineering complex traits using targeted manipulation with TFs. We presumed that since TFs govern numerous molecular events due to their innate role as master regulators, their activation/disruption will presumably enable manipulation with the complex traits, eliciting the most adequate cellular response. Furthermore, since the TFs respond to different stimuli by both expression pattern and activation status (Peñalva and Arst 2002; Cornet and Gaillardin 2014), we proposed constitutive OE or KO to withdraw the TF-encoding genes from their native transcriptional regulation, combined with systematical challenging the TF-engineered strains with environmental perturbations. It was presumed that such an approach will facilitate a reliable assessment of a given TF's involvement in the analyzed biological processes, which otherwise could be missed.

Technically, the task was indeed very challenging, as we observed that for the strictly aerobic, dimorphic yeast (with a tendency to filament and adhere to specific surfaces under stress), producing high amounts of organic acids (which is inherently associated with aerobic growth), and growing to high densities, optimizing a high-throughput culturing protocol is very demanding. Here presented methodology assures stable maintenance of the culturing conditions in terms of media acidity at the desired level and sufficient culture volume to assure phenotype expression and batch-to-batch repeatability while still maintaining high-throughput character. Equipped with such a protocol and assisted with

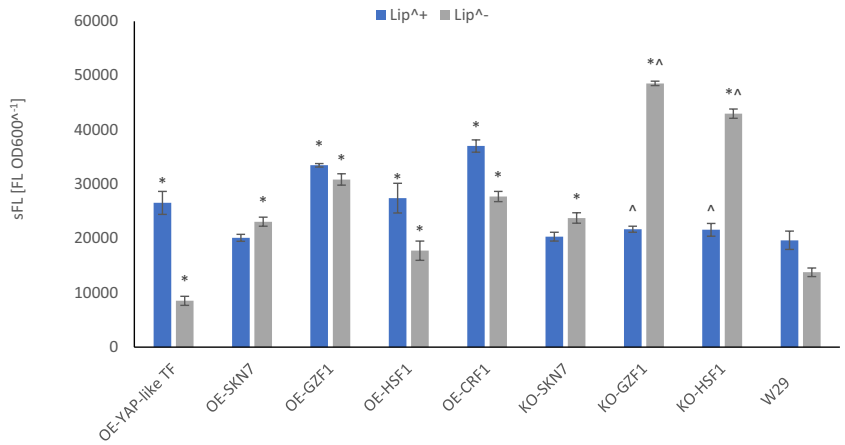


Fig. 5 Neutral lipids storage in the TF-engineered strains and the control strain cultured in a medium promoting lipid accumulation (C/N 150; Lip^+) and a control medium (C/N 4; Lip^-). The cells were stained by Nile Red dye, read at 488/565 nm. Negative control accounting for non-specific FL from the cells and the dye in the polar medium was run simultaneously and considered in calculations. Results are expressed as mean values of sFL units – relative FL units (RFU) divided by absorbance at 600 nm ± SD

the statistical design of experiments, we set for systematic phenotype testing of eight TF-engineered strains and a prototrophic control.

The selection of the TFs to be engineered (Table 1) was based on preliminary studies conducted according to a typical MTP-based protocol (micro-titer plate) used for *Y. lipolytica* (Leplat et al. 2015, 2018; Lazar et al. 2015) (data not shown). The least described TF studied here, the Yap-like TF (*YALI0D07744g*), shows weak similarity to an AP-1-like TF from *Neurospora crassa*. Its structure contains a basic leucine zipper (bZIP) domain that is typical for the yeast activator protein (Yap) from *S. cerevisiae*. The subfamily of Yaps is composed of eight members (Yap1-8) which are involved in environmental stress response (Rodrigues-Pousada et al. 2019). Yap1 is the major regulator of oxidative stress and is also involved in iron metabolism (like Yap5) and detoxification of arsenate (like Yap8). Yap2 is involved in cadmium stress responses, while Yap4 and Yap6 play a role in osmotic stress response. Which of the *S. cerevisiae*'s Yap1-8 is the most plausible counterpart of the *Y. lipolytica*'s Yap-like TF studied here can be only speculative. Some indications come from our auxiliary experiments, showing that OE of the Yap-like TF confers increased resistance to peroxide (like Yap1; 10 mM; assayed in drop test; not shown). In addition, since detoxification of heavy metals is impacted by external pH (via vacuolar V-ATPase activity and proton transport into the vacuole) (Eide et al. 1993; Gharieb and Gadd 1998; Ramsay and Gadd 2006), and the action of the here studied Yap-like TF (*YALI0D07744g*) is clearly pH-dependent, a weak suggestion on its similarity to either Yap1, Yap5, Yap8 or Yap2 is proposed.

Functional studies with the Yap-like (*YALI0D07744g*) TF in *Y. lipolytica* were conducted by Morales-Vargas et al. (2012). The authors observed a 3- to fivefold increase in the *YAP-like* gene transcription level during the dimorphic transition, suggesting its implication in this process. Interestingly, in that study, the dimorphic transition was induced by changing pH from 3.0 to 7.0. In the present study, we observed that the OE-YAP-like strain grew equally well or even better than the control only when pH was > 5 (1.1-fold; Fig. 2). Filamentation is also induced under pH ~ 7.0 (Gorczyca et al. 2020). Hence, based on the current data and the previous studies by Morales-Vargas et al. (2012), we propose that this TF is implicated in pH-dependent biological processes, downstream from the environmental acidity sensing/signaling. In addition, OE of the *YAP-like* gene significantly impacted lipid accumulation by enhancing it by ~ 35% under nitrogen depletion, and decreasing under Lip^- conditions, according to the expected pattern (Fig. 5). Previously conducted high-throughput screens including the OE-YAP-like *Y. lipolytica* strain did not show any impact of the TF-engineering on lipids accumulation (the adopted cut-off point was ± 15% over the control; (Leplat et al. 2018)). While this new observation is indeed interesting, since the capacity to synthesize r-Prot due to *YAP-like* OE was not changed (Fig. 4), this TF was not studied further for KO phenotype.

The molecular function of Skn7 (*YALI0D14520g*) in Osm stress response in *Y. lipolytica* is widely acknowledged, although no direct functional studies have been conducted yet (only studies on its upstream regulator – Hog1 (Rzechonek et al. 2018; Rzechonek et al. 2020)), while *YLSKN7* sequence is only weakly similar to an *S. cerevisiae*'s homolog (< 30%). Nevertheless, recently observed significant (fivefold) up-regulation of *YALI0D14520g* gene expression under Osm (Kubiak-Szymendera et al. 2021) supports its correct identification and its function. In the model yeast species, Skn7 is involved in osmotic and oxidative stress response. It is specifically implicated in securing chaperoning and folding capacity under stress, the onset of the oxidative stress response, and downregulation in protein synthesis. In this study, the OE of *SKN7* limited the host's growth under multiple conditions (Fig. 2), which was the most common under no Osm and high pO_2 (conditions favorable for growth). We are convinced that the growth limitation (in the case of OE-SKN7, but also the other OE- strains) vs. the control strain is a result of the increased metabolic burden caused by the high-level synthesis of two proteins rather than a consequence of a TF's specific action. Our previous comparative study on high- and low-burden *Y. lipolytica* strains supports this notion (Gorczyca et al. 2022). The actual Skn7's involvement in Osm response was highlighted in the KO-SKN7 strain exposed to Osm. Under these conditions, the growth was nearly universally abolished (irrespectively from the other conditions applied). In terms of r-Prots synthesis, Skn7's impact was clearly marked when the FL data were normalized per biomass (sFL). In such a case, OE of *SKN7* enabled maintaining r-Prot synthesis capacity under Osm infliction (when pH was > 5; Fig. 4). The highly variable effect of *SKN7*'s KO on the sFL parameter is noted, but yet not fully clear – It promoted synthesis of r-Prots in the cell challenged with oxygen limitation, but not with Osm (Fig. 4). Manipulation with the *SKN7* gene by its KO or OE had no impact in terms of lipid accumulation capacity (Fig. 5), as both strains responded the same way under the applied conditions (Lip^+/Lip^-). Correspondingly, no effect of the *SKN7* OE on lipids accumulation was revealed in the high-throughput screens conducted previously (Leplat et al. 2018).

It was very interesting to see that the expected effect of *SKN7*'s KO on *Y. lipolytica*'s growth under Osm was indeed corresponding to the one elicited by KO of *HSF1* (*YALI0E13948g*). Such a significant impact of Osm on the KO-HSF1 strain was not expected (although the model confirms the significant impact of Osm on growth in KO-HSF1; Table 2), although it has a clear biological sense, considering the significance of Hsf1 to general stress response

(Craig et al. 1993; Hahn et al. 2004; Verghese et al. 2012). Based on the box plots profile (Fig. 2), it can be concluded that Skn7 and Hsf1 have many overlapping functions, and they are required for growth under the majority of conditions analyzed here. OE of *HSF1* had only a negative impact on the strain's growth, which most probably results from increased metabolic burden (as in the case of OE of *SKN7*, discussed above). Nevertheless, the key observation regarding *HSF1* was that its OE triggered a significant increase in the r-Prot synthesis (either total FL or sFL), nearly irrespectively from the applied conditions. This modification led to the highest relative increase in FL and sFL parameters under Osm infliction (up to a 3- and 3.9-fold increase over the control). Such a promoting effect was also observed in previous studies upon OE of *HSF1* in *S. cerevisiae* (Hou et al. 2012, 2013, 2014) and is now also evidenced in *Y. lipolytica*. As presented in Fig. 5, OE of *HSF1* additionally promoted the accumulation of lipids, which stays in contrast to what was observed previously in the high-throughput screens (Leplat et al. 2018), where a 28% decrease in lipids content was observed for that strain. Differences in the cultivation method and assaying conditions could account for this discrepancy.

In terms of r-Prots production, the most spectacular outcomes were reached by co-OE of *GZF1* (*YALI0D20482g*) and *HSF1* (discussed above) that uniformly improved r-Prot synthesis (either expressed in FL or sFL – Figs. 3 or 4). In the GRYC reference database (http://gryc.inra.fr/; Table 1), Gzf1 shares some similarities with *Gibberella fujikuroi*'s nitrogen regulatory protein AreA. Recent studies by Pomraning et al. (2017) shed some light on Gzf1's function in nitrogen catabolite repression in *Y. lipolytica*. Altogether, six Gzfs are encoded in the *Y. lipolytica* genome. Those authors studied phenotypes of KO strains grown on different nitrogen sources. Structural and phylogenetic analyses showed that from among Gzf1-6, Gzf1 is the most similar to the activators of genes repressed by nitrogen catabolite repression in filamentous ascomycetes. Nonetheless, its KO did not render any aberrant phenotype, in contrast to what was observed for *Δgzf3* and *Δgzf2* strains. On the other hand, the shift in its expression level when grown on ammonium vs. peptone was the highest among *GZF*s. *GZF1* was up-regulated when the strain was cultured in the former medium and highly down-regulated when peptone was provided. Such an observation suggests that *GZF1* is highly responsive to nitrogen levels in the medium and is involved in capturing nitrogen under its limitation (activation of nitrogen assimilation genes).

As observed here, growth of the OE-GZF1 strain was not severely affected, except for occasional growth limitation, probably due to increased metabolic load (Fig. 2). In contrast, r-Prot synthesis capacity was strongly enhanced nearly irrespectively from the culturing conditions (Figs. 3 and 4). It is highly plausible that improved assimilation of nitrogen due to the *GZF1* OE could account for this result. The r-Prot synthesis capacity of the *GZF1*-engineered strains was specifically dependent on pH while less dependent on temp (Table 2). Such a specific pH dependency is related to the biological function of Gzf1 which enhances nitrogen capturing. It is known that the amino acid transporters of yeast are proton-coupled symporters, dependent on external pH. The overall proton motive force driving the internalization of amino acids is driven by the pH gradient between the outside (here pH 3 to 7) and the cytoplasm (slightly alkalic pH 7.5) (Bianchi et al. 2019). In other words, there is a direct link between ambient pH and the process of amino acid internalization. Cellular pools of amino acids impact the operation of the overexpressed *GZF1* (or redundant *GZFs* upon *GZF1*'s KO), which is the reason for the high impact of pH and pH's interactions on OE/KO-GZF1 strains, as observed here (Table 2). Consequently, the numeric factor temp lost importance and was ranked less significant for these strains.

However, as clearly seen in many contour plots (Figs. 3 and 4), r-Prot synthesis is enhanced under higher temperatures, which is a general phenomenon, irrespective of the TFs' engineering. This observation stays in agreement with Arrhenius's formula for the temperature dependence of reaction rates, of course, within the tolerance range of the biological object. On the other hand, it is widely accepted that decreased temperature promotes r-Prot synthesis in multiple yeast hosts, as was also demonstrated for *Y. lipolytica* (Kubiak et al. 2021). In the following study, we explained the previous observation by dissecting r-Prots synthesis which is higher at higher temperatures (within a reasonable range) from their secretion that is promoted under decreased temperatures (Korpys-Woźniak et al. 2021). In addition, the r-Prot model used here and previously are small, fluorescent proteins with limited post-translational modifications (Korpys-Woźniak et al. 2020; Korpys-Woźniak and Celińska 2021). Typically, the temperature decrease is applied to slow down protein synthesis and enable correct folding, which is the limiting step. Since the r-Prot used here (and before) did not impose folding problems, we could observe this temperature dependency, confirmed by the high ranking of the temp factor for the sFL response by many strains in Table 2.

In contrast to what was seen previously in high-throughput screens ((Leplat et al. 2018); no impact on lipids accumulation), in this study, we observed that any genetic manipulation with *GZF1* significantly impacted the lipid accumulation capacity of the host (Fig. 5). Its OE led to enhanced lipid accumulation irrespective of the conditions (C/N) ratio. On the other hand, its disruption led to unexpected phenotype change and promoted lipids storage under the availability of nitrogen sources (conditions not favoring lipids accumulation). While without further detailed biochemical studies, it is impossible to understand how such

an outcome was reached, it is concluded that manipulation with the nitrogen utilization regulator leads to changes in lipid accumulation due to failure to properly regulate/signal nitrogen metabolism or to sense and signal the intracellular nitrogen state. Similar conclusions were reached previously for *GZF2* and *GZF3* in *Y. lipolytica* (Pomraning et al. 2017). Practically, superior accumulation of lipids by the *GZF1*-engineered strain has the potential for biotechnological process development. Altogether, our current observations on *GZF1* in *Y. lipolytica* nicely complement the previously obtained data by Pomraning et al. (2017), where the functional phenotype of *Δgzf1* was not observed.

OE of *CRF1* (*YALI0B08206g*) had only a minor, negative impact on *Y. lipolytica* growth (Fig. 2), which could result from the increased metabolic burden (as discussed above for the other TFs). However, a consistent pattern of uniformly limited growth of OE-CRF1 was observed under pH 3.0 when neither low pO_2 nor Osm stress factors were applied. Crf1 is a transcriptional regulator involved in resistance to high copper concentration, and its role in *Y. lipolytica* was confirmed by functional studies and its localization to the nucleus during growth in copper-supplemented (García et al. 2002). *CRF1* was exploited as a dominant selection marker in the genetic engineering of *Y. lipolytica* (Guo et al. 2011), confirming its function. As demonstrated by García et al. (2002), *CRF1* expression was not induced by the addition of copper to the medium, but its KO resulted in a 4- to fivefold increase in *Y. lipolytica* copper tolerance (but increased sensitivity to cadmium). It was also shown that *SOD1* (superoxide dismutase) and *MTP* (metallothionein) genes are not downstream targets of Crf1, as could be expected. The importance of the external pH in copper detoxification was discussed above for the Yap-like TF and also provides a link between the current observations on pH dependency and the known function of Crf1 in *Y. lipolytica*. In terms of *CRF1*'s OE impact on r-Prot synthesis (FL or sFL), it was interesting to see its very specific promoting effect under low pH, low temp, and Osm infliction (Figs. 3 and 4). These conditions were the harshest for the strains' growth and r-Prot synthesis. These were also the conditions under which the beneficial effect of *HSF1* OE was lost. Therefore, this promoting effect of *CRF1* OE is particularly interesting. On the other hand, the levels of FL and sFL are generally very low under these conditions (see kinetic data in Supplementary Figs. S1B and S1C).

Our auxiliary experiments (drop tests; not shown) suggested that OE of *CRF1* slightly improved the strain's resistance to menadione (0.25 mg mL^{-1}), which is an oxidative stress inducer, but has no effect on the strain's growth upon exposure to Osm, peroxide, or Congo Red. How these functions can be translated into the observed enhanced lipid storage capacity of the OE-CRF1 strain (Fig. 5) remains to be elucidated. One possible explanation is that OE of *CRF1* increases resistance to oxidative stress, and by these lipids (prone to oxidative damage) are accumulated at higher levels, but this statement requires further in-depth studies. Notably, a similar lipid-accumulation-promoting effect (by 85%) was observed in the previous high-throughput screens for the OE-CRF1 strain (Leplat et al. 2018).

In conclusion, our results show that Osm applied as a stress factor primarily limits growth, but its impact on r-Prots synthesis cannot be neglected (shown previously (Kubiak-Szymendera et al. 2021)). Disruption of *SKN7* and *HSF1* impairs survival under Osm, proving their function in osmostress response, but their OE is not alleviating growth under Osm infliction. Nevertheless, the OE of *SKN7* enables maintaining the r-Prot synthesis capacity under Osm. In this regard, we presume that Skn7's action was somewhat similar to what has been recently achieved in *P. pastoris* upon *MSN4* OE (Zahrl et al. 2023). On the other hand, the elevated temperature (within the adopted range) promoted the synthesis of the easy-to-fold non-secretory r-Prots, when decoupled from growth (sFL) since growth was barely affected by the temp factor. The temperature was indeed the key parameter promoting r-Prot synthesis by OE-SKN7 and OE-HSF1, as these two TFs are intrinsically responsive to changes in thermal conditions. In the case of *GZF1*-engineered strains, pH and pH-Osm are significant for r-Prot, as they affect amino acid assimilation, and the impact of the temp factor is overbalanced. Hence, its function in nitrogen assimilation was confirmed and exploited in r-Prot synthesis engineering. OE of the Yap-like TF-encoding gene positively impacts the growth of *Y. lipolytica* under pH > 5, but not its r-Prot synthesis capacity. *Gzf1* and *Hsf1* are proposed universal r-Prot synthesis enhancers. Considering our current findings on *Y. lipolytica*'s TFs, and the results of omics studies by Lubuta et al. (2019) and Hapeta et al. (2020) on global cellular response elicited by different carbon and nitrogen sources in *Y. lipolytica*, it would be interesting to investigate the outcomes of the TFs-genetic engineering in different media formulations. Likewise, knowing the beneficial effects of increased abundance of Msn4 TF on the secretory r-Prot production in *P. pastoris* (Zahrl et al. 2023), it would be of high interest to verify if the here discovered r-Prot synthesis enhancers could also promote the secretion of r-Prots in *Y. lipolytica*.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00253-023-12607-z.

Acknowledgements Graphical abstract was prepared with BioRender.com

Author contribution EC, JMN, and MG conceived and designed the research. MG conducted the majority of experiments, including culturing, sample collection and analysis, data processing, statistical analyses,

and data visualization. EC supervised works and provided funding and resources. EC and MG analyzed data and wrote the manuscript. JMN revised draft. All authors read and approved the manuscript.

Funding This research was funded by the National Science Centre, Poland, grant number 2021/41/B/NZ9/00086.

Data availability All data accompanying this research are presented directly in the manuscript and supplementary materials.

Code availability Not applicable.

Declarations

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest The authors declare no competing interests.

References

Babour A, Kabani M, Boisramé A, Beckerich JM (2008) Characterization of Ire1 in the yeast *Yarrowia lipolytica* reveals an important role for the Sls1 nucleotide exchange factor in unfolded protein response regulation. Curr Genet 53:337–346. https://doi.org/10.1007/s00294-008-0190-1

Beopoulos A, Chardot T, Nicaud JM (2009) *Yarrowia lipolytica*: a model and a tool to understand the mechanisms implicated in lipid accumulation. Biochimie 91:692–696. https://doi.org/10.1016/j.biochi.2009.02.004

Bianchi F, Van't Klooster JS, Ruiz SJ, Poolman B (2019) Regulation of amino acid transport in *Saccharomyces cerevisiae*. Microbiol Mol Biol Rev. 83(4). https://doi.org/10.1128/MMBR.00024-19

Boisramé A, Beckerich JM, Gaillardin C (1999) A mutation in the secretion pathway of the yeast *Yarrowia lipolytica* that displays synthetic lethality in combination with a mutation affecting the signal recognition particle. Mol Gen Genet 261:601–609. https://doi.org/10.1007/s004380050002

Boisramé A, Chasles M, Babour A, Beckerich JM, Gaillardin C (2002) Sbh1p, a subunit of the Sec61 translocon, interacts with the chaperone calnexin in the yeast *Yarrowia lipolytica*. J Cell Sci 115:4947–4956. https://doi.org/10.1242/jcs.00187

Celińska E (2022) "Fight-flight-or-freeze" – how *Yarrowia lipolytica* responds to stress at molecular level? Appl Microbiol Biotechnol 106:3369–3395. https://doi.org/10.1007/s00253-022-11934-x

Celińska E, Nicaud J-M (2019) Filamentous fungi-like secretory pathway strayed in a yeast system: peculiarities of *Yarrowia lipolytica* secretory pathway underlying its extraordinary performance. Appl Microbiol Biotechnol 103:39–52. https://doi.org/10.1007/s00253-018-9450-2

Celińska E, Ledesma-Amaro R, Larroude M, Rossignol T, Pauthenier C, Nicaud JM (2017) Golden Gate Assembly system dedicated to complex pathway manipulation in *Yarrowia lipolytica*. Microb Biotechnol 10:450–455. https://doi.org/10.1111/1751-7915.12605

Chen DC, Beckerich JM, Gaillardin C (1997) One-step transformation of the dimorphic yeast *Yarrowia lipolytica*. Appl Microbiol Biotechnol 48:232–235. https://doi.org/10.1007/s002530051043

Cogo AJD, Façanha AR, Da Silva Teixeira LR, De Souza SB, Da Rocha JG, Figueira FF, Eutrópio FJ, Bertolazi AA, De Rezende CE, Krohling CA, Okorokov LA, Cruz C, Ramos AC, Okorokova-Façanha AL (2020) Plasma membrane H^+pump at a crossroads of acidic and iron stresses in yeast-to-hypha transition. Metallomics 12:2174–2185. https://doi.org/10.1039/d0mt00179a

Cornet M, Gaillardin C (2014) pH signaling in human fungal pathogens: a new target for antifungal strategies. Eukaryot Cell 13:342–352. https://doi.org/10.1128/EC.00313-13

Craig EA, Gambill BD, Nelson RJ (1993) Heat shock proteins: molecular chaperones of protein biogenesis. Microbiol Rev 57:402–414. https://doi.org/10.1016/S0022-1910(02)00176-2

de Ruijter JC, Frey AD (2015) Analysis of antibody production in *Saccharomyces cerevisiae*: effects of ER protein quality control disruption. Appl Microbiol Biotechnol 99:9061–9071. https://doi.org/10.1007/s00253-015-6807-7

Dragosits M, Mattanovich D (2013) Adaptive laboratory evolution – principles and applications for biotechnology. Microb Cell Fact 12:1. https://doi.org/10.1186/1475-2859-12-64

Duan G, Ding L, Wei D, Zhou H, Chu J, Zhang S, Qian J (2019) Screening endogenous signal peptides and protein folding factors to promote the secretory expression of heterologous proteins in *Pichia pastoris*. J Biotechnol 306:193–202. https://doi.org/10.1016/j.jbiotec.2019.06.297

Eide DJ, Bridgham JT, Zhao Z, Mattoon JR (1993) The vacuolar H^+-ATPase of *Saccharomyces cerevisiae* is required for efficient copper detoxification, mitochondrial function, and iron metabolism. MGG Mol Gen Genet 241(3–4):447–456. https://doi.org/10.1007/BF00284699

Endoh-Yamagami S, Hirakawa K, Morioka D, Fukuda R, Ohta A (2007) Basic helix-loop-helix transcription factor heterocomplex of Yas1p and Yas2p regulates cytochrome P450 expression in response to alkanes in the yeast *Yarrowia lipolytica*. Eukaryot Cell 6:734–743. https://doi.org/10.1128/EC.00412-06

García S, Prado M, Dégano R, Domínguez A (2002) A copper-responsive transcription factor, CRF1, mediates copper and cadmium resistance in *Yarrowia lipolytica*. J Biol Chem 277:37359–37368. https://doi.org/10.1074/jbc.M201091200

Gasch AP (2007) Comparative genomics of the environmental stress response in ascomycete fungi. Yeast 24:961–976. https://doi.org/10.1002/yea.1512

Gasch AP, Werner-Washburne M (2002) The genomics of yeast responses to environmental stress and starvation. Funct Integr Genomics 2:181–192. https://doi.org/10.1007/s10142-002-0058-2

Gasch AP, Spellman PTT, Kao CMM, Carmel-Harel O, Eisen MBB, Storz G, Botstein D, Brown POO (2000) Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell 11:4241–4257. https://doi.org/10.1091/mbc.11.12.4241

Gharieb MM, Gadd GM (1998) Evidence for the involvement of vacuolar activity in metal(loid) tolerance: vacuolar-lacking and -defective mutants of *Saccharomyces cerevisiae* display higher sensitivity to chromate, tellurite and selenite. Biometals 11(2):101–106. https://doi.org/10.1023/A:1009221810760

Gorczyca M, Kaźmierczak J, Steels S, Fickers P, Celińska E (2020) Impact of oxygen availability on heterologous gene expression and polypeptide secretion dynamics in *Yarrowia lipolytica*-based protein production platforms. Yeast 37:559–568. https://doi.org/10.1002/yea.3499

Gorczyca M, Kaźmierczak J, Fickers P, Celińska E (2022) Synthesis of secretory proteins in *Yarrowia lipolytica*: effect of combined stress factors and metabolic load. Int J Mol Sci 23:3602. https://doi.org/10.3390/ijms23073602

Groenewald M, Boekhout T, Neuvéglise C, Gaillardin C, Van Dijck PWM, Wyss M (2014) *Yarrowia lipolytica*: safety assessment of an oleaginous yeast with a great industrial potential. Crit Rev Microbiol 40:187–206. https://doi.org/10.3109/1040841X.2013.770386

Guerfal M, Ryckaert S, Jacobs PP, Ameloot P, Van Craenenbroeck K, Derycke R, Callewaert N (2010) The *HAC1* gene from *Pichia pastoris*: characterization and effect of its overexpression on the production of secreted, surface displayed and membrane proteins. Microb Cell Fact 9:1–12. https://doi.org/10.1186/1475-2859-9-49

Guo Y, Feng C, Song H, Wang Z, Ren Q, Wang R (2011) Effect of *POX3* gene disruption using self-cloning *CRF1* cassette in *Yarrowia lipolytica* on the γ-decalactone production. World J Microbiol Biotechnol 27:2807–2812. https://doi.org/10.1007/s11274-011-0758-7

Guyot S, Ferret E, Gervais P (2005) Responses of *Saccharomyces cerevisiae* to thermal stress. Biotechnol Bioeng 92:403–409. https://doi.org/10.1002/bit.20600

Hahn JS, Thiele DJ (2004) Activation of the *Saccharomyces cerevisiae* heat shock transcription factor under glucose starvation conditions by Snf1 protein kinase. J Biol Chem 279:5169–5176. https://doi.org/10.1074/jbc.M311005200

Hahn J, Hu Z, Thiele DJ, Iyer VR (2004) Genome-wide analysis of the biology of stress responses through heat shock transcription factor. Mol Cell Biol 24:5249–5256. https://doi.org/10.1128/MCB.24.12.5249

Hapeta P, Kerkhoven EJ, Lazar Z (2020) Nitrogen as the major factor influencing gene expression in *Yarrowia lipolytica*. Biotechnol Rep 27:e00521. https://doi.org/10.1016/j.btre.2020.e00521

Hirakawa K, Kobayashi S, Inoue T, Endoh-Yamagami S, Fukuda R, Ohta A (2009) Yas3p, an Opi1 family transcription factor, regulates cytochrome P450 expression in response to n-alkanes in *Yarrowia lipolytica*. J Biol Chem 284:7126–7137. https://doi.org/10.1074/jbc.M806864200

Hou J, Tyo KEJ, Liu Z, Petranovic D, Nielsen J (2012) Metabolic engineering of recombinant protein secretion by *Saccharomyces cerevisiae*. FEMS Yeast Res 12:491–510. https://doi.org/10.1111/j.1567-1364.2012.00810.x

Hou J, Österlund T, Liu Z, Petranovic D, Nielsen J, Osterlund T, Liu Z, Petranovic D, Nielsen J (2013) Heat shock response improves heterologous protein secretion in *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 97:3559–3568. https://doi.org/10.1007/s00253-012-4596-9

Hou J, Tang H, Liu Z, Österlund T, Nielsen J, Petranovic D (2014) Management of the endoplasmic reticulum stress by activation of the heat shock response in yeast. FEMS Yeast Res 14:481–494. https://doi.org/10.1111/1567-1364.12125

Hurtado CAR, Rachubinski RA (1999) *Mhy1* encodes a C_2H_2-type zinc finger protein that promotes dimorphic transition in the yeast *Yarrowia lipolytica*. J Bacteriol 181:3051–3057

Hurtado CAR, Rachubinski RA (2002) *YlBMH1* encodes a 14–3-3 protein that promotes filamentous growth in the dimorphic yeast *Yarrowia lipolytica*. Microbiology (reading) 148:3725–3735. https://doi.org/10.1099/00221287-148-11-3725

Kabani M, Beckerich J-M, Gaillardin C (2000) Sls1p stimulates Sec63p-mediated activation of Kar2p in a conformation-dependent manner in the yeast endoplasmic reticulum. Mol Cell Biol 20:6923–6934. https://doi.org/10.1128/mcb.20.18.6923-6934.2000

Kolhe N, Kulkarni A, Zinjarde S, Acharya C (2021) Transcriptome response of the tropical marine yeast *Yarrowia lipolytica* on exposure to uranium. Curr Microbiol 78:2033–2043. https://doi.org/10.1007/s00284-021-02459-z

Korpys-Woźniak P, Celińska E (2021) Global transcriptome profiling reveals genes responding to overproduction of a small secretory, a high cysteine- and a high glycosylation-bearing protein in *Yarrowia lipolytica*. Biotechnol Rep 31:e00646. https://doi.org/10.1016/j.btre.2021.e00646

Korpys-Woźniak P, Kubiak P, Białas W, Celińska E (2020) Impact of overproduced heterologous protein characteristics on physiological response in *Yarrowia lipolytica* steady-state-maintained continuous cultures. Appl Microbiol Biotechnol 104:9785–9800. https://doi.org/10.1007/s00253-020-10937-w

Korpys-Woźniak P, Kubiak P, Celińska E (2021) Secretory helpers for enhanced production of heterologous proteins in *Yarrowia lipolytica*. Biotechnol Rep 32:e00669. https://doi.org/10.1016/j.btre.2021.e00669

Kubiak M, Białas W, Celińska E (2021) Thermal treatment improves a process of crude glycerol valorization for the production of a heterologous enzyme by *Yarrowia lipolytica*. Biotechnol Rep 31:e00648. https://doi.org/10.1016/j.btre.2021.e00648

Kubiak-Szymendera M, Skupień-Rabian B, Jankowska U, Celińska E (2021) Hyperosmolarity adversely impacts recombinant protein synthesis by *Yarrowia lipolytica* – molecular background revealed by quantitative proteomics. Appl Microbiol Biotechnol 106:349–367. https://doi.org/10.1007/s00253-021-11731-y

Larroude M, Celinska E, Back A, Thomas S, Nicaud JM, Ledesma-Amaro R (2018) A synthetic biology approach to transform *Yarrowia lipolytica* into a competitive biotechnological producer of β-carotene. Biotechnol Bioeng 115:464–472. https://doi.org/10.1002/bit.26473

Larroude M, Trabelsi H, Nicaud JM, Rossignol T (2020) A set of *Yarrowia lipolytica* CRISPR/Cas9 vectors for exploiting wild-type strain diversity. Biotechnol Lett 42:773–785. https://doi.org/10.1007/s10529-020-02805-4

Lazar Z, Gamboa-Meléndez H, Le Coq AMC, Neuvéglise C, Nicaud JM (2015) Awakening the endogenous Leloir pathway for efficient galactose utilization by *Yarrowia lipolytica*. Biotechnol Biofuels 8:185. https://doi.org/10.1186/s13068-015-0370-4

Leplat C, Nicaud JM, Rossignol T (2015) High-throughput transformation method for *Yarrowia lipolytica* mutant library screening. FEMS Yeast Res 15:1–9. https://doi.org/10.1093/femsyr/fov052

Leplat C, Nicaud J-MM, Rossignol T (2018) Overexpression screen reveals transcription factors involved in lipid accumulation in *Yarrowia lipolytica*. FEMS Yeast Res 18:1–9. https://doi.org/10.1093/femsyr/foy037

Lesage J, Timoumi A, Cenard S, Lombard E, Lee HLT, Guillouet SE, Gorret N (2021) Accelerostat study in conventional and microfluidic bioreactors to assess the key role of residual glucose in the dimorphic transition of *Yarrowia lipolytica* in response to environmental stimuli. N Biotechnol 64:37–45. https://doi.org/10.1016/j.nbt.2021.05.004

Liu Z, Hou J, Martínez JL, Petranovic D, Nielsen J (2013) Correlation of cell growth and heterologous protein production by *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 97:8955–8962. https://doi.org/10.1007/s00253-013-4715-2

Lubuta P, Workman M, Kerkhoven EJ, Workman CT (2019) Investigating the influence of glycerol on the utilization of glucose in *Yarrowia lipolytica* using RNA-seq-based transcriptomics. G3-GENES GENOM GENET 9:g3.400469.2019. https://doi.org/10.1534/g3.119.400469

Madzak C (2015) *Yarrowia lipolytica*: recent achievements in heterologous protein expression and pathway engineering. Appl Microbiol Biotechnol 99:4559–4577. https://doi.org/10.1007/s00253-015-6624-z

Madzak C (2018) Engineering *Yarrowia lipolytica* for use in biotechnological applications: a review of major achievements and recent innovations. Mol Biotechnol 60:621–635. https://doi.org/10.1007/s12033-018-0093-4

Madzak C (2021) *Yarrowia lipolytica* strains and their biotechnological applications: how natural biodiversity and metabolic engineering could contribute to cell factories improvement. J Fungi 7:548. https://doi.org/10.3390/jof7070548

Madzak C, Beckerich J-MM (2013) Heterologous protein expression and secretion in *Yarrowia lipolytica*. In: Barth G (ed) Microbiology monographs. Springer-Verlag, Berlin Heidelberg, pp 1–76

Mans R, Daran JMG, Pronk JT (2018) Under pressure: evolutionary engineering of yeast strains for improved performance in fuels and chemicals production. Curr Opin Biotechnol 50:47–56. https://doi.org/10.1016/j.copbio.2017.10.011

Martínez JL, Meza E, Petranovic D, Nielsen J (2016) The impact of respiration and oxidative stress response on recombinant α-amylase production by *Saccharomyces cerevisiae*. Metab Eng Commun 3:205–210. https://doi.org/10.1016/j.meteno.2016.06.003

Martinez-Vazquez A, Gonzalez-Hernandez A, Domínguez Á, Rachubinski R, Riquelme M, Cuellar-Mata P, Guzman JCT (2013) Identification of the transcription factor Znc1p, which regulates the yeast-to-hypha transition in the dimorphic yeast *Yarrowia lipolytica*. PLoS One 8:e66790. https://doi.org/10.1371/journal.pone.0066790

Matsumoto R, Akama K, Rakwal R, Iwahashi H (2005) The stress response against denatured proteins in the deletion of cytosolic chaperones *SSA1/2* is different from heat-shock response in *Saccharomyces cerevisiae*. BMC Genomics 6:141. https://doi.org/10.1186/1471-2164-6-141

Mattanovich D, Gasser B, Hohenblum H, Sauer M (2004) Stress in recombinant protein producing yeasts. J Biotechnol 113:121–135. https://doi.org/10.1016/j.jbiotec.2004.04.035

Morales-Vargas AT, Domínguez A, Ruiz-Herrera J (2012) Identification of dimorphism-involved genes of *Yarrowia lipolytica* by means of microarray analysis. Res Microbiol 163:378–387. https://doi.org/10.1016/j.resmic.2012.03.002

Nicaud JM (2012) Yarrowia lipolytica. Yeast 29:409–418. https://doi.org/10.1002/yea.2921

Nicaud JM, Madzak C, Van Den Broek P, Gysler C, Duboc P, Niederberger P, Gaillardin C (2002) Protein expression and secretion in the yeast *Yarrowia lipolytica*. FEMS Yeast Res 2:371–379. Pii S1567–1356(02)00082-X

Peñalva MA, Arst HN (2002) Regulation of gene expression by ambient pH in filamentous fungi and yeasts. Microbiol Mol Biol Rev 66:426–446. https://doi.org/10.1128/MMBR.66.3.426-446.2002

Pomraning KR, Kim YM, Nicora CD, Chu RK, Bredeweg EL, Purvine SO, Hu D, Metz TO, Baker SE (2016) Multi-omics analysis reveals regulators of the response to nitrogen limitation in *Yarrowia lipolytica*. BMC Genomics 17:1–18. https://doi.org/10.1186/s12864-016-2471-2

Pomraning KR, Bredeweg EL, Baker SE (2017) Regulation of nitrogen metabolism by GATA zinc finger transcription factors in *Yarrowia lipolytica*. mSphere 2:1–19. https://doi.org/10.1128/mSphere.00038-17

Pomraning KR, Bredeweg EL, Kerkhoven EJ, Barry K, Haridas S, Hundley H, LaButti K, Lipzen A, Yan M, Magnuson JK, Simmons BA, Grigoriev IV, Nielsen J, Baker SE (2018) Regulation of yeast-to-hyphae transition in *Yarrowia lipolytica*. mSphere 3:1–18. https://doi.org/10.1128/msphere.00541-18

Poopanitpan N, Kobayashi S, Fukuda R, Horiuchi H, Ohta A (2010) An ortholog of *farA* of *Aspergillus nidulans* is implicated in the transcriptional activation of genes involved in fatty acid utilization in the yeast *Yarrowia lipolytica*. Biochem Biophys Res Commun 402:731–735. https://doi.org/10.1016/j.bbrc.2010.10.096

Portnoy VA, Bezdan D, Zengler K (2011) Adaptive laboratory evolution-harnessing the power of biology for metabolic engineering. Curr Opin Biotechnol 22:590–594. https://doi.org/10.1016/j.copbio.2011.03.007

Ramsay LM, Gadd GM (2006) Mutants of *Saccharomyces cerevisiae* defective in vacuolar function confirm a role for the vacuole in toxic metal ion detoxification. FEMS Microbiol Lett 152(2):293–298. https://doi.org/10.1111/j.1574-6968.1997.tb10442.x

Rodrigues-Pousada C, Devaux F, Caetano SM, Pimentel C, da Silva S, Cordeiro AC, Amaral C (2019) Yeast AP-1 like transcription factors Yap and stress response: a current overview. Microbial Cell 6:267–285. https://doi.org/10.15698/mic2019.06.679

Rzechonek DA, Day AM, Quinn J, Mirończuk AM (2018) Influence of ylHog1 MAPK kinase on *Yarrowia lipolytica* stress response and erythritol production. Sci Rep 8:1–12. https://doi.org/10.1038/s41598-018-33168-6

Rzechonek DA, Szczepańczyk M, Wang G, Borodina I, Mirończuk AM (2020) Hog-independent osmoprotection by erythritol in yeast *Yarrowia lipolytica*. Genes (basel) 11:1–15. https://doi.org/10.3390/genes11121424

Sambrook J, Russell D (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, New York

Sassi H, Delvigne F, Kallel H, Fickers P (2017) pH and not cell morphology modulate p*LIP2* induction in the dimorphic yeast *Yarrowia lipolytica*. Curr Microbiol 74:413–417. https://doi.org/10.1007/s00284-017-1207-0

Sekova VY, Kovalyov LI, Kovalyova MA, Gessler NN, Danilova MA, Isakova EP, Deryabina YI (2021) Proteomics readjustment of the *Yarrowia lipolytica* yeast in response to increased temperature and alkaline stress. Microorganisms 9(12):2619. https://doi.org/10.3390/microorganisms9122619

Staudacher J, Rebnegger C, Dohnal T, Landes N, Mattanovich D, Gasser B (2022) Going beyond the limit: increasing global translation activity leads to increased productivity of recombinant secreted proteins in *Pichia pastoris*. Metab Eng 70:181–195. https://doi.org/10.1016/j.ymben.2022.01.010

Swennen D, Beckerich JM (2007) *Yarrowia lipolytica* vesicle-mediated protein transport pathways. BMC Evol Biol 7:219. https://doi.org/10.1186/1471-2148-7-219

Swennen D, Henry C, Beckerich JM (2010) Folding proteome of *Yarrowia lipolytica* targeting with uracil permease mutants. J Proteome Res 9:6169–6179. https://doi.org/10.1021/pr100340p

Torres-Guzmán JC, Domínguez A (1997) *HOY1*, a homeo gene required for hyphal formation in *Yarrowia lipolytica*. Mol Cell Biol 17:6283–6293. https://doi.org/10.1128/mcb.17.11.6283

Tyo KEJ, Liu Z, Petranovic D, Nielsen J (2012) Imbalance of heterologous protein folding and disulfide bond formation rates yields runaway oxidative stress. BMC Biol 10:16. https://doi.org/10.1186/1741-7007-10-16

Verghese J, Abrams J, Wang Y, Morano KA (2012) Biology of the heat shock response and protein chaperones: budding yeast (*Saccharomyces cerevisiae*) as a model system. Microbiol Mol Biol Rev 76:115–158. https://doi.org/10.1128/mmbr.05018-11

Walker C, Ryu S, Trinh CT (2019) Exceptional solvent tolerance in *Yarrowia lipolytica* is enhanced by sterols. Metab Eng 54:83–95. https://doi.org/10.1016/j.ymben.2019.03.003

Wang Z-P, Xu H-M, Wang G-Y, Chi Z, Chi Z-M (2013) Disruption of the *MIG1* gene enhances lipid biosynthesis in the oleaginous yeast *Yarrowia lipolytica* ACA-DC 50109. Biochim Biophys

Acta 1831:675–682. https://doi.org/10.1016/j.bbalip.2012.12.010

Winkler JD, Kao KC (2014) Genomics recent advances in the evolutionary engineering of industrial biocatalysts. Genomics 104:406–411. https://doi.org/10.1016/j.ygeno.2014.09.006

Wu H, Shu T, Mao Y-S, Gao X-D (2019) Characterization of the promoter, downstream target genes and recognition DNA sequence of Mhy1, a key filamentation-promoting transcription factor in the dimorphic yeast *Yarrowia lipolytica*. Curr Genet 66(1):245–261. https://doi.org/10.1007/s00294-019-01018-1

Yang LB, Dai XM, Zheng ZY, Zhu L, Zhan XB, Lin CC (2015) Proteomic analysis of erythritol-producing *Yarrowia lipolytica* from glycerol in response to osmotic pressure. J Microbiol Biotechnol 25:1056–1069. https://doi.org/10.4014/jmb.1412.12026

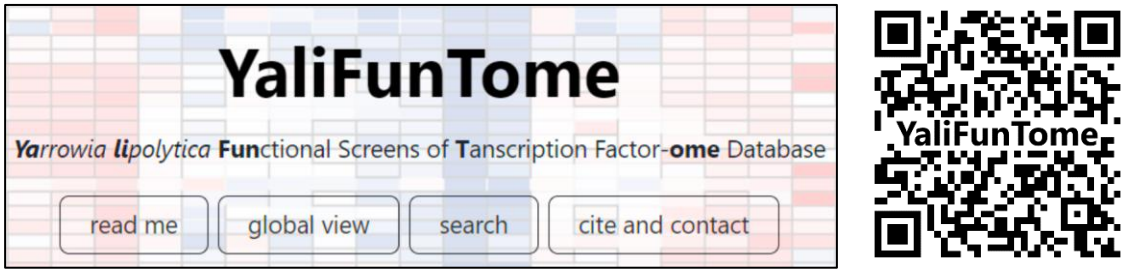
Zahrl RJ, Prielhofer R, Ata Ö, Baumann K, Mattanovich D, Gasser B (2022) Pushing and pulling proteins into the yeast secretory pathway enhances recombinant protein secretion. Metab Eng 74:36–48. https://doi.org/10.1016/j.ymben.2022.08.010

Zahrl RJ, Prielhofer R, Burgard J, Mattanovich D, Gasser B (2023) Synthetic activation of yeast stress response improves secretion of recombinant proteins. N Biotechnol 73:19–28. https://doi.org/10.1016/j.nbt.2023.01.001

5.3 **P2**: “Matka(Natura) wie najlepiej” – wykorzystanie mechanizmów transkrypcyjnych ukształtowanych przez naturę w celu zwiększenia odporności na stres i produkcji białek w drożdżach *Yarrowia lipolytica*; prezentacja bazy danych YaliFunTome

Publikacja **P2** prezentuje wyniki szeroko zakrojonego skriningu ponad 120 szczepów *Y. lipolytica* eksponowanych na 72 kombinacje zmiennych środowiskowych, przy zastosowaniu starannie zoptymalizowanego protokołu hodowli wysokoprzepustowych. Protokół ten, opisany i szczegółowo uzasadniony w artykule (zawierający wytyczne hodowli, uzasadnienie ich wyboru oraz szerszą dyskusję parametrów), umożliwił uzyskanie powtarzalnych wyników oraz wiarygodną ocenę fenotypów w warunkach miniaturyzacji hodowli.

Łącznie przeprowadzono 26 000 hodowli, a skala zgromadzonych danych skłoniła do przedstawienia ich w formie interaktywnego katalogu – bazy danych YaliFunTome (*Yarrowia lipolytica* Functional Screens of Transcription Factor-ome Database; **Rycina 1**), dostępnej online (https://sparrow.up.poznan.pl/tsdatabase/), aby ułatwić wydobycie sensu biologicznego z danych numerycznych. W artykule opisano szczegółowo sposób korzystania z bazy danych, zakres zgromadzonych informacji oraz możliwości, jakie oferuje. Baza umożliwia przeszukiwanie wyników zarówno na podstawie konkretnego kodu genu (numeru Yali), jak również na podstawie określonych przez użytkownika kombinacji czynników środowiskowych, a także prezentuje zbiorcze mapy cieplne obrazujące globalne wzorce odpowiedzi.



Rycina 1. Główna strony bazy danych YaliFunTome – *Yarrowia lipolytica* Functional Screens of Transcription Factor-ome Database oraz kod QR przekierowujący na powyższą stronę internetową.

Analiza danych pozwoliła na wyróżnienie kilku grup czynników transkrypcyjnych o istotnym znaczeniu dla badanych cech złożonych:

a) Globalne "wzmacniacze" syntezy rProt:

i) Klf1 (*YALI0D05041g*) – nowo zidentyfikowany uniwersalny „wzmacniacz" syntezy białek, którego nadekspresja skutkowała ponad 2-krotnym wzrostem syntezy rProt. Gen ten nie był dotychczas badany w *Y. lipolytica*. W *Schizosaccharomyces pombe* dla jego homologu wykazano bezpośrednie zaangażowanie w regulację genów związanych z odbudową ściany komórkowej (ang. *cell wall renewal*), odpowiedzią na stres oksydacyjny, glikolizą, pobieraniem składników odżywczych, jak również wyciszaniem, glikozylacją i metylacją chromatyny (ang. *RNA-mediated chromatin silencing, glycosidation, and methylation*) (Shimanuki i in., 2013).

ii) Gzf1 – wykazano jego rolę w zwiększonej syntezie rProt, zgodną z wynikami badań pilotażowych z **P1**, również w zmodyfikowanym układzie zmiennych środowiskowych. Co ciekawe, wśród badanych reprezentantów rodziny TFs palca cynkowego wiążących motyw GATA (ang. *GATA-binding zinc finger*), czyli Gzf1, Gzf3 i Gzf4, tylko Gzf1 wykazał ten efekt.

iii) Hsf1 – potwierdzono wyniki wcześniejszych obserwacji (**P1**) wykazując efektywność tego TF dla szerszego spektrum warunków środowiskowych. Zaobserwowano ponadto, że Hsf1 wspiera produkcję rProt, nawet kosztem ograniczonego wzrostu w warunkach podaży mniej preferowanego („niekorzystnego") źródła azotu. Interpretowano to jako efekt zwiększonego zapotrzebowania komórki na zasoby w trakcie intensywnej syntezy białka – ograniczona dostępność azotu mogła prowadzić do zatrzymania cyklu komórkowego i zahamowania wzrostu, podczas gdy syntetycznie wymuszony proces produkcji białek wciąż trwał.

b) Globalne represory syntezy rProt

i) Dep1 (*YALI0F05896g*) – jego nadekspresja silnie obniżała akumulację rProt, podczas gdy delecja powodowała efekt odwrócony, zwiększając produkcję niemal 3-krotnie. Wskazuje to na bezpośrednie zaangażowanie Dep1 w regulację tego procesu. Funkcjonalnie Dep1 był wcześniej wiązany z akumulacją lipidów w *Y. lipolytica* (Leplat i in., 2018), a jego homolog z *Fusarium* sp., z aktywacją biosyntezy fosfolipidów (Zhang i in., 2020).

ii) Azf1 (*YALI0A16841g*) – nadekspresja prowadziła do wyraźnego spadku syntezy rProt, choć genotyp *Δazf1* nie skutkował otrzymaniem fenotypu różnego od szczepu kontrolnego. W *S. cerevisiae* Azf1 jest TF zależnym od źródła węgla:

w obecności glukozy aktywuje geny odpowiedzialne za wzrost, metabolizm węgla i wzrost inwazyjny. U *Y. lipolytica* wskaźnikiem takiego typu wzrostu jest filamentacja, i rzeczywiście nadeskpresja Azf1 skutkowała wzmożonym wydłużaniem komórek. Ponadto zmienna „źródło węgla" miała istotny wpływ na kształtowany fenotyp.

iii) Cat8 (*YALI0C19151g*) – choć jego efekt nie był tak wyraźny jak Azf1 i Dep1, uzyskane fenotypy pozwalały na sklasyfikowanie go jako globalnego represora syntezy rProt. U *S. cerevisiae* Cat8, jako aktywator transkrypcyjny, wiąże się z elementami promotorowymi zależnymi od źródła węgla w warunkach braku glukozy. Jego znana rola w regulacji glukoneogenezy i cyklu glioksylanowego czyni go funkcjonalnie zbliżonym do dwóch pozostałych represorów; wskazując na działanie uniwersalnych zależności pomiędzy metabolizmem węgla a efektywnością syntezy rProt.

c) Regulator związany z syntezą rProt zależną od źródła azotu

i) Hoy1 (*YALI0A18469g*) – szczególnie interesującym odkryciem było wskazanie Hoy1 jako TF sprzyjającego syntezie rProt przy dostępie wyłącznie do nieorganicznego źródła azotu („niekorzystnego" – niepreferowanego w kontekście nadprodukcji rProt). W *Y. lipolytica* Hoy1 bierze udział w tranzycji morfologicznej, a jego nadekspresja promuje tworzenie strzępek (Torres-Guzmán i in., 1997). Analiza sekwencji promotora sugeruje, że gen ten podlega deregulacji w odpowiedzi na sygnały stresowe oraz na bodźce związane z biosyntezą aminokwasów i głodem azotowym. Zależność pomiędzy formowaniem strzępek a rodzajem źródła azotu dla *Y. lipolytica* została uprzednio wyjaśniona w bardzo przekonywujący sposób przez Szabo i Stofaníkovą (2002). Prawdopodobnie, nadekspresja *HOY1* w naszych badaniach, ułatwiała wychwytywanie azotu w warunkach jego niedoboru, co przekładało się na poprawę wydajności syntezy rProt. Odkrycie to ma duży potencjał aplikacyjny, ponieważ może ograniczać koszty produkcji związane z koniecznością stosowania drogich źródeł azotu organicznego.

d) TFs związane z dostępnością tlenu

i) Zidentyfikowano TFs, których nadekspresja wspierała wzrost w warunkach ograniczonego natlenienia. Należą do nich TF036 (*YALI0D20460g*), Jmc2 (*YALI0B14443g*), TF011 (*YALI0B20944g*) i Dal81 (*YALI0D02783g*) poprawiające wzrost w warunkach niedoboru tlenu średnio o 13-23% w porównaniu ze szczepem kontrolnym. Ich funkcje w *Y. lipolytica* nie zostały jeszcze scharakteryzowane, jednak obserwowane efekty fenotypowe sugerują ich potencjalny udział w adaptacji do niedoboru tlenu.

ii) Hap1 (*YALI0F17424g*) – jego nadekspresja przyczyniła się do ukształtowania fenotypu zwiększonej wrażliwości na niedobór tlenu wyrażony zahamowaniem wzrostu o ok. 20%. W *S. cerevisiae* Hap1 odpowiada za regulację transkrypcji w odpowiedzi na poziom tlenu, wykorzystując zdolność wiązania hemu jako wskaźnika natlenienia. Może działać zarówno jako aktywator, jak i represor, regulując geny związane z oddychaniem, cyklem komórkowym i obroną przed reaktywnymi formami tlenu (ROS, ang. *reactive oxygen species*) w warunkach wysokiej dostępności tlenu (Zhang i Hach, 1999). W *Y. lipolytica* mechanizmy detekcji tlenu i sygnalizacji hemowej nie zostały dotąd opisane, jednak uzyskane wyniki wskazują, że homolog Hap1 może odgrywać w nich rolę.

Badania pozwoliły na identyfikację TFs pełniących rolę uniwersalnych „wzmacniaczy" i represorów syntezy białek, TF promującego syntezę rProt przy obecności nieorganicznego źródła azotu, oraz regulatorów adaptacji do ograniczonej dostępności tlenu. Wyniki te stanowią punkt wyjścia do dalszych, pogłębionych analiz, będących w trakcie realizacji (poza zakresem niniejszej rozprawy doktorskiej). Zaproponowano także ogólny wzorzec działania globalnych represorów: aktywacja genów związanych z anabolizmem węgla ogranicza syntezę białek. Stworzona baza danych YaliFunTome udostępnia te dane społeczności naukowej, ułatwiając wyszukiwanie interesujących homologów i fenotypów determinowanych przez TF, a jednocześnie stanowi trwałe repozytorium wiedzy o funkcjonalnej roli regulatorów transkrypcyjnych w *Y. lipolytica*.

Microbial Cell Factories

RESEARCH

Open Access

‘Mother(Nature) knows best’ – hijacking nature-designed transcriptional programs for enhancing stress resistance and protein production in *Yarrowia lipolytica*; presentation of YaliFunTome database

Maria Gorczyca[1], Wojciech Białas[1], Jean-Marc Nicaud[2] and Ewelina Celińska[1*]

Abstract

Background In the era of rationally designed synthetic biology, heterologous metabolites production, and other counter-nature engineering of cellular metabolism, we took a step back and recalled that ‘Mother(-Nature) knows best’. While still aiming at synthetic, non-natural outcomes of generating an ‘over-production phenotype’ we dug into the pre-designed transcriptional programs evolved in our host organism—*Yarrowia lipolytica*, hoping that some of these fine-tuned orchestrated programs could be hijacked and used. Having an interest in the practical outcomes of the research, we targeted industrially-relevant functionalities—stress resistance and enhanced synthesis of proteins, and gauged them over extensive experimental design’s completion.

Results Technically, the problem was addressed by screening a broad library of over 120 *Y. lipolytica* strains under 72 combinations of variables through a carefully pre-optimized high-throughput cultivation protocol, which enabled actual phenotype development. The abundance of the transcription program elicitors—transcription factors (TFs), was secured by their overexpression, while challenging the strains with the multitude of conditions was inflicted to impact their activation stratus. The data were subjected to mathematical modeling to increase their informativeness.

The amount of the gathered data prompted us to present them in the form of a searchable catalog – the YaliFunTome database (https://sparrow.up.poznan.pl/tsdatabase/)—to facilitate the withdrawal of biological sense from numerical data. We succeeded in the identification of TFs that act as omni-boosters of protein synthesis, enhance resistance to limited oxygen availability, and improve protein synthesis capacity under inorganic nitrogen provision.

Conclusions All potential users are invited to browse YaliFunTome in the search for homologous TFs and the TF-driven phenotypes of interest.

Keywords Yeast, Transcription factors, Stress resistance, Protein production, Global metabolic engineering, *Yarrowia* cultivation protocol

*Correspondence:
Ewelina Celińska
ewelina.celinska@up.poznan.pl
Full list of author information is available at the end of the article

Background

Any rationally designed biotechnological production process artificially forces the producer cell to the non-natural 'over-production phenotype' state. It is now well-recognized that such a phenotype is elicited by a concerted action of multiple genes/molecular identities. Hence, designing such a phenotype operating under specific process conditions requires a thorough understanding of the molecular and regulatory mechanisms governing the target bioproduct synthesis; and then—using this knowledge to orchestrate genotype and environment towards the objective functionality. With the current technology progress, reaching the "proof-of-concept" level of microbial cell factory readiness can be relatively rapidly accomplished. However, its fine-tuning to reach actual industrial relevance requires detailed insight into the host's system. Such a goal can be achieved by either a very extensive metabolic engineering strategy (which will inevitably impose a high metabolic burden on the engineered cell) or by hijacking nature-pre-designed programs governed by native global regulators.

The concept of using global regulators for engineering complex traits in yeast (like stress resistance or over-production of proteins) has been already proposed in several different forms. Alper and Stephanopoulos [1] proposed 'global transcription machinery engineering' for optimizing complex target phenotypes, and implemented the idea by conducting the sigma70 subunit of RNA polymerase-based engineering in *Saccharomyces cerevisiae*. Overexpression of the Unfolded Protein Response (UPR)-activating Transcription Factor (TF), Hac1, was the most frequently employed strategy for enhancing recombinant proteins (r-Prots) production. Hac1's direct implication in the regulation of endoplasmic reticulum (ER)-resident events and restoring ER homeostasis made it the most straightforward target in such endeavors. It was demonstrated that overexpression of *HAC1* improves the secretion of r-Prots in *S. cerevisiae* [2], *Pichia pastoris* (renamed to *Komagataella phaffi*) [3, 4], and recently in *Yarrowia lipolytica* [5, 6]. A complementary approach targeted Hsf1 (heat shock factor 1), which is the key regulator of heat shock response and governs the "cytosolic" version of UPR [7, 8]. Hou et al. [9] investigated the regulome of *S. cerevisiae*'s Hsf1 and showed that it is mainly represented by molecular chaperones involved in protein folding, thus preventing the accumulation of misfolded or aggregated proteins. Continuous activation of heat shock response by overexpression of mutant *HSF1*-R206S triggered increased production of native and r-Prots [9]. Global regulators engineering was also employed to attenuate endogenous oxidative stress emerging from the oxidative folding process. Hap1 is a single TF managing aerobic metabolism in *S. cerevisiae*. It is responsible for sensing the oxygen levels via the heme signaling pathway, and activation of the oxidative stress response genes [10]. Overexpression (OE) of *HAP1* in *S. cerevisiae* overproducing a secretory r-Prot led to the mitigation of the negative effects caused by reactive oxygen species accumulation, hence increasing the production capacity of the strain [11]. Very recently, synthetic activation of the general stress response TF Msn4 (and its synthetic version synMsn4, alone or in combination) triggered over fourfold enhancement in r-Prot production [12].

Considering its industrially-relevant metabolic characteristics, its genetic amenability, and its genomic sequence available [13], *Y. lipolytica* has emerged as a robust biotechnological production platform of many added-value bioproducts [14]. Both native and heterologous molecules have been produced with this yeast platform, including organic acids [15], polyols [16], aromas [17, 18], carotenoids [19, 20], usual and unusual fatty acids [21, 22] etc. With the advent of numerous tools and the progress in understanding its molecular background, *Y. lipolytica* has also become a recognized platform for the production of r-Prots [23–25].

To date, the research on TFs in *Y. lipolytica* has been largely conducted in the context of basic studies. Most of the literature reports refer to an individual TF's role in the regulation of dimorphic transition; implication in this phenomenon has been proved for Hoy1 [26], Mhy1 [27, 28], Bmh1 [29], YAP-like YALI0D07744g [30], Znc1 [31], and Msn2 [32]. The role in the regulation of lipid metabolism in *Y. lipolytica* has been documented for several other TFs, Yas1 and Yas2 [33], Yas3 [34], Por1 [35], Mig1 [36], and Gzf2 and Gzf3 [32]. The actual implementation of the 'global transcription machinery engineering' concept in this species has been to date limited to only several reported studies; for example, high-throughput functional screens of over one hundred twenty strains individually over-expressing (OE-ing) TFs that were conducted to fish out global regulators involved in lipids accumulation [37, 38]. Recently, [39] studied the impact of five selected TFs Hsf1, Gzf1, Crf1, Skn7, and YAP-like (YALI0D07744g) on two interconnected complex traits—stress resistance and r-Prots synthesis. OE of Yap-like TF was proven to alleviate growth retardation under high pH, while Gzf1 and Hsf1 were shown to serve as universal enhancers of r-Prot production in *Y. lipolytica*. On the other hand, deletion of *SKN7* and *HSF1* disabled growth under hyperosmotic stress.

As recently pointed out by the inventors of the Yeastract + database [40, 41], the lack of integration of regulatory information in the global metabolic models hinders their predictive ability. Hence, this functionality has been implemented in a recent update of the database [41]. However, due to under-investigation, the information

load for *Y. lipolytica* in this context is rather scarce. Considering how huge potential is held by 'global transcription machinery engineering' we decided to explore this niche and provide experimental evidence on the actual implications of *Y. lipolytica*'s TFs in stress resistance and/or in the synthesis of r-Prots. Knowing that TF's activity is not secured by the sole abundance of the protein, we challenged the strains constitutively OE-ing the TFs with a combination of environmental stress factors. Altogether, 125 *Y. lipolytica* strains individually OE-ing one of the TF-encoding genes were screened under 72 different conditions. We read growth and the level of a reporter r-Prot synthesis, which was used as a gauge for a global protein synthesis capacity—comprising, but not limited to r-Prots. The amount and informativeness of the data we gathered from completing our experimental design inspired us to present them in the form of a searchable catalog—the YaliFunTome database, to facilitate the withdrawal of biological sense from extensive sets of numerical data.

Materials and methods

Data collection and analysis

Biological material

A collection of 125 *Y. lipolytica* strains over-expressing (OE) individually one of the TFs identified in *Y. lipolytica*'s genome, and a reporter protein (RedStar2) was used in this study. The search of the TFs within the *Y. lipolytica* genome and construction of the strains co-over-expressing (co-OE) TF and a reporter r-Prot was done previously [37]. Both genes were cloned under a constitutive promoter pTEF, and were integrated in a zeta platform at *URA3* locus. The collection contained two control strains: (i) a completely prototrophic version of the parental strain (JMY2900); and (ii) complete prototroph OE the reporter r-Prot solely (JMY2810). The former was used as a control of burden imposed by OE of the reporter r-Prot and control of cross-contamination between cultures. The latter was considered the actual "control strain" used for normalization of the data from the TF-OE strains. All the strains were constructed on the background of the JMY2566 strain. More technical details on the cloning strategy can be found in [42].

All the strains were deposited as 15% glycerol stocks at − 80 °C for long-term storage. While running the experimental plan, the strain collection was freshly plated from stocks to YPD-agar medium (g L^{-1}: yeast extract, 5 (BTL, Łódź, Poland); peptone, 10 (BTL); glucose, 20 (POCH, Gliwice, Poland); solidified with agar, 15 (BTL)) every 2 weeks.

Experimental design

The experimental plan was designed using the Design of Experiments (DoE) methodology in DesignExpert software (StatSoft, Tulsa, USA) as a Face-centered Central Composite Design. It implemented five external variables: two numerical (continuous) and three categoric (discrete), as presented in Table 1. Variables were technically implemented as follows: numerical variables: (i) pH (at levels: 3.0, 5.0, or 7.0) adjusted during the media preparation with 20% NaOH (w w^{-1}) and by change of the buffer molarity (pH 3, − 0.1 M maleic acid buffer, pH 5 and 7, − 0.2 M maleic acid buffer); (ii) temperature (22, 28 and 34 °C) by setting thermostat; categoric variables; (iii) oxygen availability (OA) changed by use of different sandwich covers [43]: high OA (2.5 mL min^{-1}

Table 1 Summary of experimental setup designed with DoE (DX, StatSoft)

Design type: face-centered Central composite design	Runs: 104 for each of 125 TF = 13 000 in duplicate = 26,000 readouts on FL (fluorescence) and growth (OD600)

The central variants were repeated five times. The order of culture execution was randomized

Process factors Categoric: A, B, C Numeric: d, e

Factor	Name	Units	Change	Type	Minimum	Maximum	Coded Low	Coded High	Mean	Std. Dev
A	AIR "limiting"	0/1	Easy	Categoric	Low	High			Levels: 2	
B	Carbon source	Glu/Gly	Easy	Categoric	Glucose	Glycerol			Levels: 2	
C	Nitrogen source	i/o	Easy	Categoric	Inorganic: Ammonium Sulfate	Organic: Casamino Acid Hydrolysate			Levels: 2	
d	pH	pH	Hard	Numeric	3	7	− 1 → 3	+ 1 → 7	5	2.83
e	Temperature	°C	Hard	Numeric	22	34	− 1 → 22.00	+ 1 → 34.00	27.1	5.443

headspace air exchange), low OA (0.004 mL min^{-1}); (iv) carbon source—glucose (encoded − 1) or glycerol (encoded + 1); and (v) nitrogen source: inorganic (ammonium sulfate; encoded − 1) or organic (casamino acid hydrolysate; encoded + 1). All the combinations resulted in 72 variants of environmental conditions, 8 of which were considered 'central variants' (according to the heatmaps: these were pH 5 and temp 28 °C with different categoric factors combinations—OA, carbon, and nitrogen). Each environmental variant was conducted in biological duplicate, but the 8 'central variants' were conducted in pentaplicate, resulting in 64 'outer variants' plus 8 'central variants' × 5 = 104 culture runs. Each of the 104 culture runs was conducted in two biological repetitions. The order of culture execution was randomized. Figure 1 schematically illustrates the variables combinations, corresponding to the data presentation scheme.

High-throughput screens

Media preparation

The composition of the culture media was carefully adjusted to minimize background fluorescence (FL), provide sufficient amounts of carbon (C) and nitrogen (N) to avoid starvation, and maintain the pH stably throughout the culturing time. The medium was composed as follows [g L^{-1}]: yeast nitrogen base without amino acids and ammonium sulfate, 5.1 (Sigma-Aldrich); ammonium sulfate or casamino acid hydrolysate, 15, (POCH); glucose or glycerol, 35 (POCH); buffered with 0.1 or 0.2 M maleic acid (POCH). Precultures were developed in media composed of a yeast nitrogen base, 5.1 (Sigma-Aldrich); ammonium sulfate, 15, (POCH); glucose, 25 (POCH); buffered with 0.2 M maleic acid at pH 5.0. All the media were filter-sterilized with 0.22-µm filters (Merck- Millipore, Darmstadt, Germany).

Culture conditions

Since typical multitier plates were found useless for reliable high-throughput screens and phenotype reads of *Y. lipolytica* strains, we used 24-well Duetz-System square plates in 2.5 mL working volume (EnzyScreen BV, Netherlands), agitated at 250 rpm. The cultures were continued for 48 h and samples were collected at the end of the cultivation time (the timing was pre-optimized). Precultures were developed for 18 h at 28 °C, under high OA. The main cultures were inoculated at 4% (v v^{-1}).

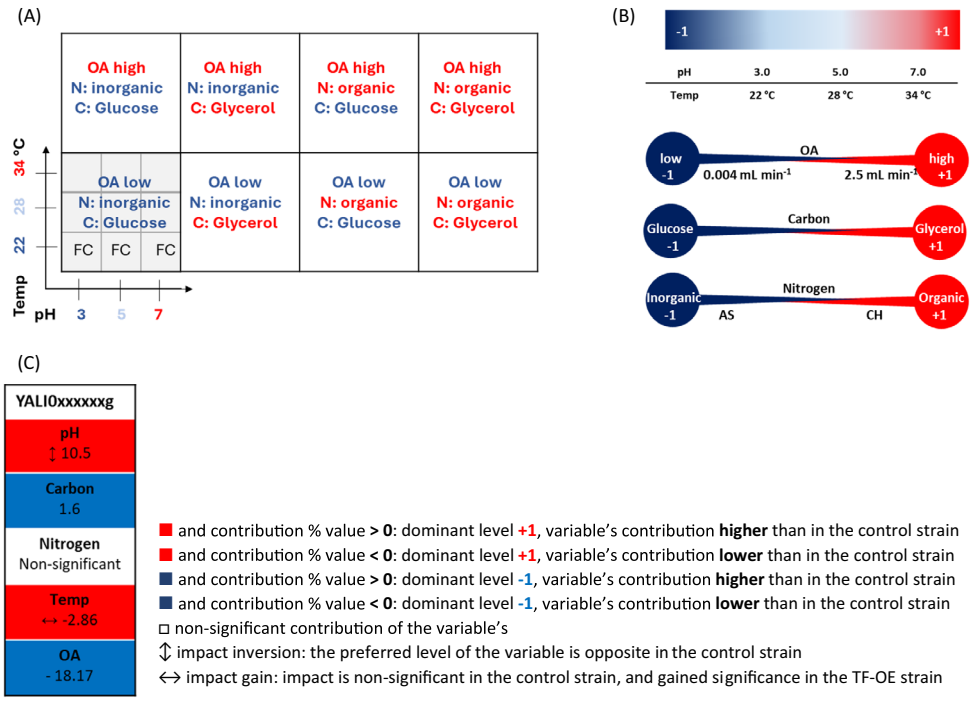


Fig. 1 A system for graphical presentation of the data and variables coding. Graphical presentation of the arrangement of the variables' combinations corresponding to the data presentation scheme (**A**), and the variables' coding system (**B**), exemplary Factor's Contribution ranking with a legend (**C**). The coding system corresponds to Table 1. The rankings are prepared based on Response Surface Models. Each ranking represents the variable's contribution to a response of a given read parameter (growth/r-Prot/normalized r-Prot) elicited by a specific TF-OE strain. CH—casamino acid hydrolysate, AS—ammonium sulfate, OA- oxygen availability, FC–fold change of the response from the TF-OE strain over the control strain

Analytical methods

Samples were analyzed for growth and FL from the reporter protein (RedStar2) following dilution in 0.75% NaCl (POCH) to match a linear range of the methods. Absorbance was measured at 600 nm in transparent 96-well plates (Costar; Merck). FL was determined under at ex/em 550/595 nm in black opaque plates (Thermo Fisher Scientific) Both measurements were done using a Tecan Spark automatic plate reader (Tecan Group Ltd., Mannedorf, Switzerland).

Data processing and statistical analyses

Results for total FL from the reporter r-Prot are expressed in Fluorescence Units (FU); growth is expressed in absorbance units from measuring optical density at 600 nm wavelength. Normalized r-Prot measure was calculated by dividing the FU value per OD600 value giving the specific fluorescence (RFU–relative fluorescence units). Statistical significance of the difference in a given measure between a TF-OE strain and the control strain was assessed by analysis of variance (ANOVA) test, with significance level set at p-value < 0.05 (Statistica, StatSoft-Tibco, Tulsa, USA). Fold change (FC) values express the ratio of a given measure for a TF-OE strain over the control strain.

Mathematical models describing the growth, total r-Prot synthesis, and normalized r-Prot synthesis by the strains within the range covered by the variables were developed using raw data (not FC) response surface methodology (RSM) in DesignExpert software. It resulted in a set of the following items for each of the TF-OE strains: (i) a model equation (including interactions between the variables; equations were not shown); (ii) set of eight 3D contour plots covering all the responses within the range of tested variables (arranged as in Fig. 1; color-coded according to the legend); and (iii) quantitative assessment of the variables contribution to the response (Factor's Contribution Tables; percentage value). In the latter, after manual inspection of the data, the interactions between the variables were ignored (for clarity of data presentation). The significance of each variable was assessed by ANOVA tests, at p-value < 0.05. To compare the contribution of individual variables in the TF-OE strain model and the control, the values obtained for the former were normalized over the latter. The higher/lower contribution was indicated by a positive/negative value of the percentage contribution, while the direction of contribution is color-coded/labeled according to the legend. All of the obtained models were found significant, fitting of data was color-coded.

YaliFunTome database construction

The YaliFunTome database was created using the MariaDB relational database management system (https://mariadb.org/), PHP 7.4.33 (https://www.php.net/), Bootstrap 5.3.1 (https://getbootstrap.com/), HTML 5, CSS 3, JavaScript and jQuery 3.7.0 (https://jquery.com/). The database has been tested on several different web browsers including Google Chrome, Mozilla Firefox and Microsoft Edge.

Results and discussion

Getting the pipeline working

Having a strictly aerobic, filamenting-upon-stress biological object in our hands, it rapidly appeared that the majority of high-throughput cultivation protocols working for the other yeast species are non-operable in this case. In addition, due to the modification of global regulators, the set of 125 strains analyzed here comprised those showing no obvious aberrant phenotype under the majority of conditions, and those severely affected in basic metabolism. Hence, it was necessary to develop a new high-throughput cultivation protocol that will enable reliable examination of all the phenotypes in a streamlined and normalized manner.

First, we defined a set of variables that are relevant to the *Y. lipolytica* r-Prots overproducer phenotype, and which could be affected by the TFs' OE. The primary limitation encountered in *Y. lipolytica* cultures is limited oxygen provision, staying beyond the technical capabilities of laboratory or industrial equipment. Recently, [44] showed that Oxygen Availability (OA) is the key factor determining the rate of r-Prot synthesis. The same conclusions were made for the production of organic acids and polyols [45–47]. Similarly, it has been evidenced that pH is a strong modulator of the r-Prots-producing capacity in this species [48]. While the pH level can be relatively easily stabilized in bioreactor cultivations, local fluctuations in medium acidity are known to occur. Followingly, it was demonstrated that elevated temperature treatment or synthetic induction of heat shock response (by OE of *HSF1* TF) both had a positive impact on the r-Prots synthesis rate in *S. cerevisiae* [9]. In contrast, heat-shock treatment was found detrimental to r-Prots synthesis in *Y. lipolytica*, and only a low-temperature treatment exerted a positive impact in this regard [49]. Still, OE of *HSF1* in *Y. lipolytica* led to enhancement in r-Prot synthesis [6], correspondingly to what was found for *S. cerevisiae*. Therefore, we hypothesize that the OE of some TFs may mitigate the adverse effects of non-optimal OA, pH, and temperature.

In addition, previous investigations into *Y. lipolytica*'s biology clearly showed that the type of carbon source (glucose-glycerol) had a tremendous impact on the evoked regulatory mechanisms [37, 50]. Likewise, the type and concentration of available nitrogen sources were shown to have a fundamental impact on *Y. lipolytica* biology [51], including r-Prots production [52, 53] and dimorphic transition [54, 55].

The abovementioned conditions were then tested in an extensive set of preliminary experiments checking the feasibility and mode of the variables' implementation. Optimization studies were conducted with two control strains—JMY2810 overexpressing a fluorescent reporter r-Prot, and a complete prototroph JMY2900, to estimate the burden imposed by the sole reporter r-Prot OE. The strains were analyzed for growth and, in the case of JMY2810, for protein synthesis (gauged by the reporter r-Prot fluorescence). As the burden imposed by the reporter r-Prot synthesis was not statistically significant in terms of growth, results for JMY2900 were not shown. The results of the preliminary optimization tests indicated that (Fig. 2):

i. Growth and r-Prots synthesis is supported within the pH range of 3 to 7, which can be effectively implemented and stably maintained at a desired level by using maleic buffer; this buffer type fulfills the following requirements: (i) spans the whole range to be tested (dissociation constants (pKa) equal 1.94 and 6.22), using the same chemical compounds; (ii) is not consumed by *Y. lipolytica* (as determined by HPLC analysis); (iii) does not hamper growth, if applied at selected molarities; (iv) does not impose osmotic stress, when used at higher/different molarity (within 250 mOsm kg^{-1}; below the stress level for *Y. lipolytica* [56]); and (v) has sufficient buffering capacity at adopted concentrations (Fig. 2a, b),
ii. The adopted temperature range supports growth and provides differentiation in the r-Prots synthesis level **(**Fig. 2c),
iii. To impose an actual OA stress, but still—support growth and r-Prot synthesis, the headspace volume exchange must be minimized to 0.004 mL min^{-1} **(**Fig. 2d),
iv. Provision of a higher nitrogen source (15 g L^{-1}) is necessary to enable more complete carbon utilization and full phenotype development within a reasonable time frame (48 h) (Fig. 2e),

(See figure on next page.)

Fig. 2 Main findings of the high-throughput cultivation protocol's optimization. **A** Buffer selection—six different chemical compound combinations with buffering properties were tested. The top row in the table represents the composition of the buffer encoded by symbols (in the legend); the buffers' pKa values are marked with these symbols on the grey arrow (target pH values marked with red arrows). Some components have two pKa values making them more suitable for covering the whole span of tested pHs (it is presumed that 1 value away from the pKa value the buffer is still in its range of buffering capacity). Initially set pH, molarity, and the experimentally defined post-culture pH (control strain, main culture medium, 72 h) are given as numbers. 'Buffering capacity' [M L^{-1}], and percentage of maximal growth [%max] at 48h of cultivation (a proxy for growth impairment) are given as 'bubble graphs'. Boxes crossed out indicate that the buffer did not meet requirements in terms of (i) component—utilization by *Y. lipolytica* (based on HPLC data), (ii) final pH—drop in pH > 1, (iii) buffering capacity—buffering capacity level < 10 M L^{-1}, iv) growth—pronounced growth impairment < 40% of maximal growth. Kinetic graphs represent growth and normalized r-Prot for *Y. lipolytica* control strain grown in a maleic-acid buffered medium (3 pH values, 2 molarities). Grey 'bubble plots' represent osmolarity readouts of YPG medium buffered with the indicated maleic acid buffer, and the osmolarity level known to impose severe stress for *Y. lipolytica* cells. **B**, **C** Continuous variables range examination (pH and temperature)—these sections represent pH and temp as numerical, continuous parameters coded from -1 (pH 3.0 or 22 °C) to + 1 (pH 7.0 or 34 °C). Bubble plots represent growth and normalized r-Prot as % values of these parameters read under pH 5.0 or 34 °C for the control strain (48 h). Kinetic graphs represent growth and normalized r-Prot for *Y. lipolytica* control strain in time. **D**, **E**, **G** Discrete variables levels examination (Oxygen availability, nitrogen, and carbon source types and concentrations)—these sections represent OA, N, and C as categoric parameters encoded as -1 (low OA/inorganic N/glucose) or + 1 (high OA/organic N/glycerol) levels. Oxygen availability variation was implemented by the use of Duetz system sandwich covers of different headspace volume exchange rates: 0.004, 0.7, and 2.5 mL min^{-1}. Optimization of Nitrogen source and concentration was conducted with inorganic N (AS–ammonium sulfate) or organic N (CH – casamino acid hydrolysate) substrates at concentrations of 5 or 15 g L^{-1}. Optimization of carbon source concentrations was performed with glucose at concentrations: 20, 30, 40, and 50 g L^{-1}, marked as C20, C30, C40, and C50, respectively—and cultivated on either one of two N sources. Consumed carbon in these cultures was calculated as residual C amount based on HPLC data from 48h of cultivation in relation to its initial load. Kinetic graphs represent growth and normalized r-Prot for *Y. lipolytica* control strain in time. **F** Geometry/volume—this section presents a comparison of two types of 24-well plates with different geometry of wells (square/round) and working volume recommended by the manufacturer (2.5/1 mL). Bubble and time-point plots are constructed as in other sections. Kinetic plots: x-axis time [h] and the y-axis OD_{600} (optical density at 600nm) value or sFL [FL/OD_{600}] (specific fluorescence) for growth and normalized r-Prot, respectively. Shaded areas represent standard deviation (SD) values, based on biological duplicates or triplicates. All bubble plots are calculated in percentage to the highest or optimal variant (variants 100% are framed with black-dashed lines) in 48 h of cultivation. The percentage value is indicated within the bubble and corresponds to the bubble area. If not stated otherwise, cultures were performed in high OA, 28 °C, pH 5.0, and with glucose (20 g L^{-1}) and ammonium sulfate (15 g L^{-1}) as carbon and nitrogen sources in 24-well plates with JMY2810 as the control strain

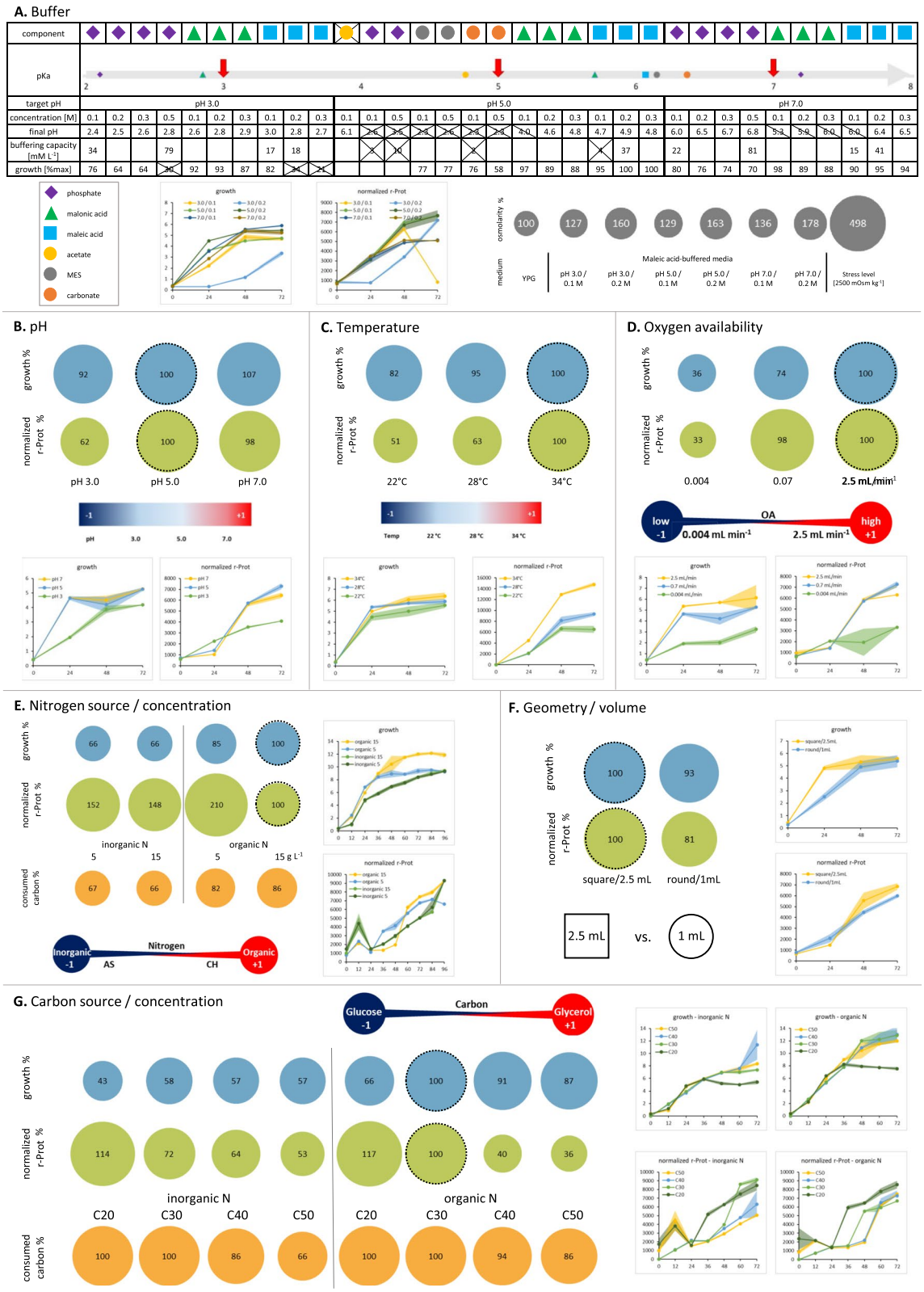


Fig. 2 (See legend on previous page.)

v. Cubic geometry of the cultivation vessels and larger volume of the culture enables faster and more pronounced phenotype development; comparison of readouts from the outer and from the inner wells of 24-well round (1 mL) and 24-well square (2.5 mL) plates, revealed that the well-to-well variation is significant for the former, and insignificant for the latter (p-value < 0.05); the squared wells enabled more homogenous mixing during cultivation and no biomass attachment to the walls of the well; using 24-deep-well square plates enables maintaining high-throughput character of the screen and reliable phenotypes assessment (Fig. 2f), and
vi. Carbon source at > 30 g L^{-1} is sufficient to support growth at the nitrogen concentration of 15 g L^{-1}, and is not fully consumed by the end of the culturing time (does not impose additional, uncontrolled starvation condition) (Fig. 2g).

With the pre-optimized and validated protocol for high-throughput screens of *Y. lipolytica* in our hands, we developed an experimental design that would enable systematic investigations into the effects induced by the variables and their combinations. The design rendered 104 culture runs to be implemented for each of the 125 strains and the controls (Table 1). The levels of the variables were encoded, as depicted in Fig. 1. Acquired data on the strains' growth, synthesis of the reporter r-Prot, or its normalized measure (r-Prot normalized per biomass) were carefully curated, and subjected to statistical significance analysis and mathematical modeling, according to "Data processing and statistical analyses" section.

YaliFunTome enables the withdrawal of biological sense from math

Navigation through YaliFunTome

The amount and informativeness of the data we gathered from completing our experimental design prompted us to present them in the form of a searchable catalog YaliFunTome (*Yarrowia lipolytica* Functional Screens of Transcription Factor-ome Database: https://sparrow.up.poznan.pl/tsdatabase/), to facilitate the withdrawal of biological sense from extensive sets of numerical data. In this, we were inspired by our Colleagues, who put a great effort into preparing and curating publicly open databases, such as: mentioned above Yeastract+, which enables predictions of efficient genetic engineering strategies, and its extension—PathoYeastract, enabling the analysis and prediction of transcription regulatory associations in the pathogenic yeasts [57], y-mtPTM, that provides a comprehensive list of experimentally-validated mitochondrial post-translational modifications [58], or YeastRGB, merging, processing and standardizing graphical data on protein abundance and subcellular localization [59].

Navigation through the YaliFunTome database should start from the 'readme' tab, which contains all the necessary information to guide the user throughout the database. The 'readme' section is divided into four subsections: (i) 'general information' about the project; (ii) 'experimental setup' describing how the project was technically implemented; a must-see-before-starting tab; (iii) 'data presentation', illustrating how to interpret records of the database search; and the fourth tab showing Response Surface Methodology (RSM) models for the; and (iv) 'control strain', which were used as a reference to calculate Fold Change (FC) data for TF-OE-ing strains.

The 'global view ' tab directs the user to three 'global heat maps' with FC values for all the TF-OE-ing strains gathered in collective plots. The three heat maps show the data on each of the three analyzed parameters: 'growth', 'total r-Prot', and 'normalized r-Prot', separately. Data within the table are color-coded according to the legend provided on top of the subpage, and are interactive: (i) while hoovering over a specific color-coded cell, a short note appears summarizing the conditions under which the response was elicited, its numerical value, and whether the response was statistically significant at $p < 0.05$; and (ii) while clicking on a specific 'Yali number', the user is directed to a specific TF-OE strain's subpage.

The 'search' tab links the user to search tools, enabling two modes of querying the database: (i) based on 'Yali number', and (ii) 'Condition-responsive TF'. The former can be used to characterize the functional behavior of a specific TF across an array of conditions, while the latter—to screen for TFs that were responsive to a specific combination of environmental factors, implemented in the experimental design.

The 'cite' tab directs to this article's citation and contact details. More details about querying the database and the retrieved records can be found in "Querying YaliFunTome" and "YaliFunTome record's content" sections.

Querying YaliFunTome

Browsing the YaliFunTome database is enabled in two modes: (i) 'Yali number'; and (ii) 'Condition-responsive TF'.

The 'Yali number' search enables browsing the database based on specific Yali numbers of genes, according to the naming convention from the first release of the *Y. lipolytica* genome [13] (listed in Additional file 2: Table S1). A specific Yali number can be selected from a collapsible drop-down list. The search record directs the

user to three links corresponding to the set of heat maps showing FC in the read parameter for a specific TF-OE-ing strain ('growth', 'total r-Prot', or 'normalized r-Prot'). Clicking on the selected parameter gets the user to a Yali number's specific subpage, which consists of (to be found by scrolling down):

i. A header providing YALI name, number of TF in the experimental design, assigned name (if available), gene identifier, and links to NCBI, RefSeq DNA, RefSeq Peptide, and UniProtKB,
ii. Three heat-maps (for 'growth', 'total r-Prot', or 'normalized r-Prot'), arranged according to Fig. 1a, and factor's contribution tables encoded as presented in Fig. 1b, c. The system of data presentation and variables' coding is reminded at the top of the subpage, for the user's convenience; both the heat maps and the Factor's Contribution tables are interactive, providing the user with detailed information about a specific response point, upon hovering over it.

The 'condition-responsive TF' search mode requires first defining the search constraints. The user can browse the database based on an elicited response in one of three measured parameters ('growth', 'total r-Prot', or 'normalized r-Prot'), which could be: (i) 'up'—higher in the TF-OE-ing strain compared to control (FC > 1.0 in a specified point); (ii)'down' (FC < 1.0); or show (iii) 'inverted phenotype', so the opposite FC response under extreme levels of a variable by a single TF-OE-ing strain (e.g. FC > 1.0 in pH 3.0 and FC < 1.0 in pH 7.0). These response constraints are set in combination with one of the studied variables (pH, temp, carbon source, nitrogen source, or OA). The last drop list specifies the constraining level of the searched variable, or specifies the levels of the variable compared in the 'inverted phenotype' search. The output page returns Yali numbers of the TFs that elicited a statistically significant response, meeting the entered constraints, and an additional 10% level of difference in FC between the TF-OE and the control strains. Yali numbers are sorted by the highest FC values for the 'up' search, by the lowest—'down' search, or by the highest delta of FC values between extreme environmental conditions for the 'inverted phenotype' search. The records specifying TF numbers that were responsive are collapsible—after expanding, the list of specific combinations of variables under which the query conditions were met is listed.

YaliFunTome record's content

The results from the experimental plan may reflect 'growth', 'total r-Prot', or 'normalized r-Prot' by the TF-OE-ing strain and the control strains. The principal mode of data presentation is an FC of a given measure read for the TF-OE-ing strain over the control strain. The data for each TF-OE-ing strain were arranged according to the variables combinations scheme in Fig. 1a, and color-coded to facilitate simple interpretation. The raw data for the three measured parameters were first used to generate RS models. Such models were shown in the 'control strain' tab for the control strain. The same set of models was prepared for all the TF-OE-ing strains (not shown). Modeling the data for TF-OE strains and the control strains enabled the estimation of the response within each point covered by the RS plot (within the adopted ranges of the variables). In addition, the models allowed the assessment of the variables' contribution (individually or as interactions; the latter not shown) to the respective measured parameter. These individual variables' contribution is quantitatively expressed (as percentage values) in the Factor's Contribution ranking tables, provided on each TF-OE-ing strain's subpage (Fig. 1c). Both the heat maps and the factor contribution tables were color-coded according to the provided legends. Details on the data presentation scheme can be found in the 'read me'—> 'data presentation' tab in YaliFunTome database.

YaliFunTome in practice—case-studies

Klf1 (YALI0D05041g) is a newly identified general booster of r-Prots synthesis

By simple inspection of the 'global view' plots for 'growth', 'total r-Prot', or 'normalized r-Prot' it is possible to fish out phenotypes eliciting strong responses nearly irrespectively from the environmental variables.

Such a pronounced phenotype was observed for a strain OE-ing YALI0D05041g gene, encoding a Zinc finger TF Krueppel-like factor 15 (Klf1; TF126). According to NCBI's conserved domain search, *Y. lipolytica*'s Klf1 bears two C2H2 Zinc finger domains at the N-terminus [60], studied this specific gene as one of the model transcripts for the intron retention phenomenon in *Y. lipolytica*. Klf1's mRNA was classified as one of the transcripts displaying intron retention but at a low level. Online translation prediction of the unspliced transcript shows that the intron retention results in premature stop codon occurrence, so it cannot be functional as an alternatively spliced variant, that would render an additional domain, depending on the signals. In yeast, Klf1 was extensively studied in the context of long-term G0 quiescence induced by nitrogen starvation [61]. As shown in those studies on *Schizosaccharomyces pombe*, the expression level of Klf1 in G0 (no nitrogen) was higher than in vegetative growth, when nitrogen was available. At the physiological level, the Δ*klf1* mutants exhibited significantly enlarged cell volume, they were metabolically active but unable to

restore cell division. Prolonged nitrogen starvation of the Δ*klf1* mutants triggered a fatal failure in the cell's morphology maintenance—accumulation of chitin-like material caused severe deformation of the cytoplasm and cell shape. Transcriptomic analyses demonstrated that Klf1 directly regulates genes implicated in cell wall renewal, oxidative stress response, glycolysis, nutrient uptake, RNA-mediated chromatin silencing, glycosidation, and methylation [61]. In higher Eukarya, Klf1 was shown to promote longevity by combating oxidative stress [62]. It was demonstrated that Klf1 relocalizes to the nucleus following a mild ROS pulse. In those studies, the oxidative stress originated from mitochondria, followed by xenobiotics treatment. Transcriptomic analysis revealed that the response elicited by Klf1 is more elaborated rather than simple combating oxidative stress. In the context of our studies, we found out that *KLF1*'s OE in *Y. lipolytica* greatly promotes r-Prots synthesis capacity, nearly irrespectively of the adopted variables' combination (Fig. 3a). The enhancement expressed in FC of (normalized) r-Prots synthesis was more pronounced under high OA (contribution of high OA was positive and higher than in the control strain by 11.08%), frequently reaching ≥ 2.0 FC. The promoting phenotype was enhanced under organic nitrogen provision, higher temperature, and lower pH; and was independent from the carbon source provided. The role of Klf1 has not been studied in *Y. lipolytica* to date. Considering the commonly assigned to Klf1 implication in stress response, and our previous studies on the effects triggered by over-synthesis of r-Prots in *Y. lipolytica* [63], it is tempting to state that the increased abundance of Klf1 in combination with ER-derived oxidative stress (caused by r-Prots over-synthesis) were the two factors enabling the display of the Klf1's promoting phenotype. As shown by our auxiliary experiments (to be published), knock-out in *KLF1*'s loci did not trigger any significant change to r-Prot synthesis capacity in *Y. lipolytica* (under 'optimal' conditions – a central point of the design; Additional file 1: Fig. S1a), which suggests some level of redundancy within this scope with some other factors.

Gzf1 (YALI0D20482g) is the only 'global r-Prots enhancer' from among tested genuine GATA-binding zinc finger family representatives (Gzf1, Gzf3, Gzf4)

Other 'global' r-Prots synthesis enhancers identified through the 'global view' plot inspection comprised Gzf1 (YALI0D20482g; TF037) and Hsf1 (YALI0E13948g; TF068). The former belongs to a six-member family of GATA-binding zinc finger TFs, *i.a.* involved in nitrogen catabolite repression (NCR). Gzf1 was studied in *Y. lipolytica,* along with the other three genuine representatives of the Gzf family, in the context of NCR's role in lipid accumulation regulation [64]. While investigating the molecular background of the nitrogen provision-lipid accumulation interplay, the Authors functionally characterized the Gzfs in *Y. lipolytica*. Insightful sequence-structure analysis demonstrated that Gzf1, so the current 'global r-Prots synthesis enhancer', is highly similar to homologs from *Aspergillus nidulans* and *Neurospora crassa* activated in response to nitrogen starvation, and is likely to operate as an activator of NCR genes, rather than a repressor (like Gzf3; YALI0C22682g). Our current studies demonstrated that upon its OE, Gzf1 greatly enhances the synthesis of r-Prots in *Y. lipolytica* (total and normalized measures; Fig. 3b1). This enhancing phenotype is promoted under high OA (factor's contribution to 'normalized r-Prots' is higher by 9.69% than in the control), organic nitrogen, higher temperature, and lower pH. Factor's contribution analyses demonstrated that the nitrogen source variable did not exert any significant contribution to the growth (and total r-Prots synthesis) in the *GZF1*-OE-ing strain. Considering its putative role as an NCR activator, this could be attributed to enhanced nitrogen scavenging capacity when inorganic nitrogen was provided. However, nitrogen was also a non-significant variable for the control strain's growth. Consistently, in the previous studies, it was observed that *GZF1* was not essential for growth on simple nitrogen sources (as was Gzf2; YALI0F17886g not covered by the YaliFunTome) [64]; even though it was the most upregulated gene from the Gzfs when an organic nitrogen source was changed to inorganic. A similar transcriptional response was also observed for *GZF4* [64]; altogether suggesting that Gzf2 (inducer) and Gzf3 (suppressor) are the overriding nitrogen assimilation regulators, while Gzf1 and Gzf4 serve as 'rapid-reaction force' of this regulation. Our results demonstrate the OE of neither *GZF3* (Fig. 3b2) nor *GZF4* (YALI0E05555g; Fig. 3b3) exert any direct positive effect on r-Prots synthesis in *Y. lipolytica*. While it seems understandable for the NCR's repressor Gzf3, it remains to be clarified for the Gzf4, whose specific role in NCR has not been determined [64]. The YaliFunTome database's 'global r-Prots enhancer', Gzf1, was previously identified as a promoter of r-Prots synthesis in *Y. lipolytica* upon its OE, but under a different array of variables [39]. Our auxiliary studies on *Δgzf1* genotype (Additional file 1: Fig. S1a) showed that it displays no effect on r-Prots synthesis [39], consistent with the results by [64]. Only its OE elicited the promoting phenotype; as was found for Klf1 (mentioned above). Such a consistency supports the generalizable character of these findings.

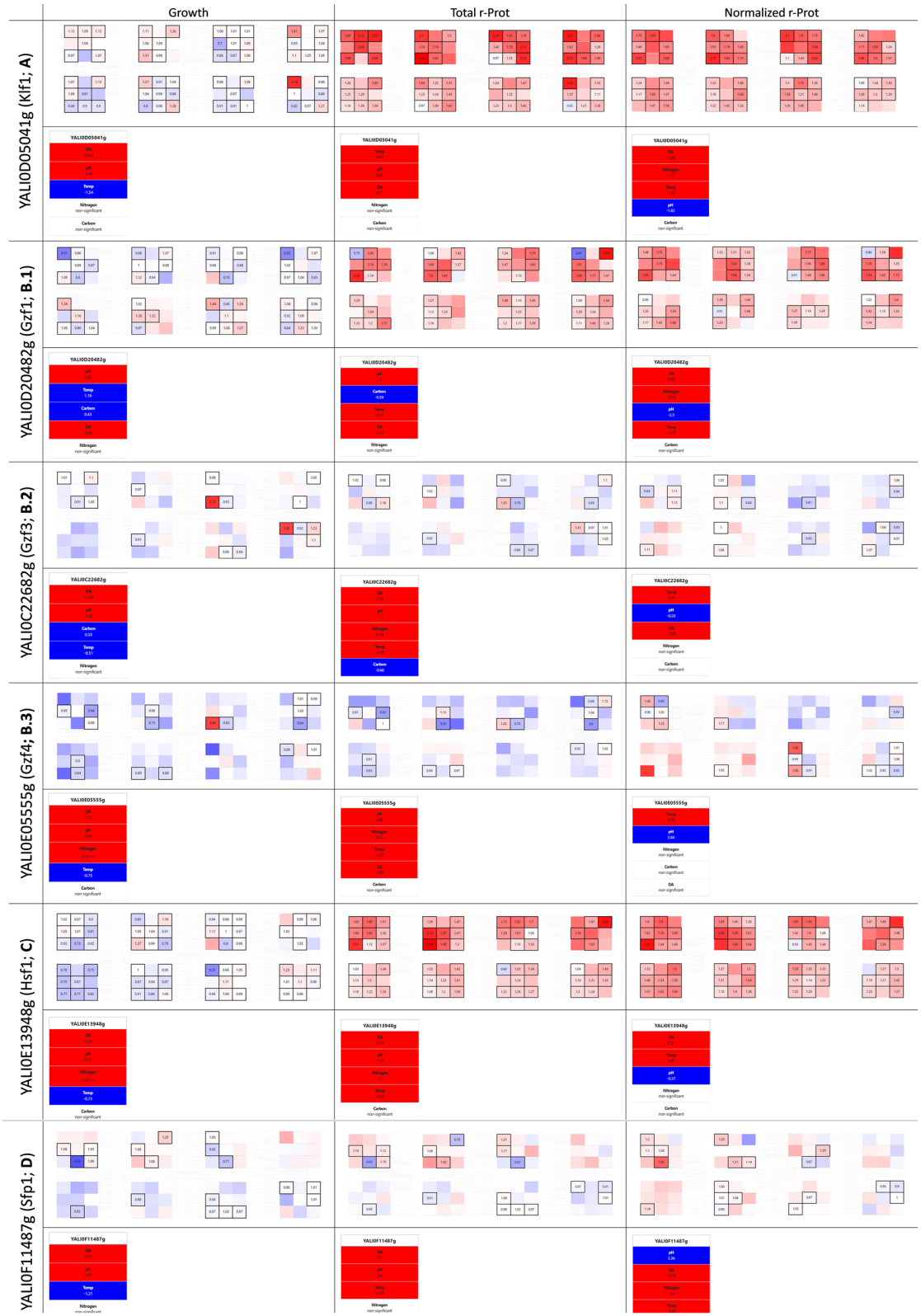


Fig. 3 Selected phenotypes from the YaliFunTome database—the enhancers of r-Prots synthesis. Selected phenotypes are presented according to the data presentation format (Fig. 2). YALI number and assigned name (if available) are given in the first column. The following columns present heat maps and factor's contribution rankings in terms of growth, r-Prots synthesis, and normalized measure of r-Prots synthesis

Hsf1 (YALI0E13948g) specifically promotes protein synthesis at the expense of growth

HSF1 OE elicited a similar promoting response in terms of r-Prots synthesis as Klf1 and Gzf1, however, it acted more selectively in this regard (Fig. 3c). It promoted the synthesis of proteins even at the expense of growth under inorganic nitrogen provision (r-Prot up, when growth is down). To our interpretation, the significantly enhanced r-Prots synthesis under *HSF1* OE leads to a higher demand for resources. Specifically, based on the results of the factor's contribution analysis, we speculate that nitrogen became limiting, as this specific variable gained significance for growth upon *HSF1* OE (not significant for *KLF1*-OE, *GZF1*-OE, and the control strain). The possible shortage in nitrogen supply prompted cell cycle arrest and growth inhibition, while the protein synthesis was still ongoing. Such a type of interplay between intensive r-Prots synthesis and growth was quantitatively expressed in our previous studies [65]. Now, based on the developed mathematical models we show that within the range of tested conditions, nitrogen became the limiting factor under intensive r-Prots synthesis.

A decreased correlation of growth and r-Prot synthesis observed for *HSF1*-OE-ing strain (Pearson coefficient between FC values for 'growth' and 'total r-Prots' $r < 0.45$) was not observed for the other TFs promoting r-Prot synthesis, like Klf1 or Sfp1 (discussed below). For the latter, the FC values for total r-Prots and growth were more positively correlated ($r > 0.58$). The positive correlation between the two parameters indicates that the two biological processes do not compete but rather act synergistically. Considering that the reporter r-Prot used here acted as a gauge of total protein synthesis capacity, such a synergistic action is natural and expected—since high protein synthesis is a prerequisite for robust growth.

The promoting role of Hsf1 in r-Prots synthesis seems straightforward to interpret. Hsf1 serves as the major activator of the cell's folding and chaperoning capacity, and induces the cytoplasmic form of UPR [66]. It also acts as the global regulator of the general stress response, which activates multiple biological processes promoting r-Prots synthesis [7]. The beneficial effect of *HSF1*'s OE on r-Prots synthesis in *Y. lipolytica* was also observed previously, under the infliction of a different set of variables [39].

Ribosome biogenesis-associated Sfp1 (YALI0F11487g)—'a fussy eater'

A strong promoting effect on growth and r-Prots synthesis was expected from the strain OE-ing an *SFP1* gene (YALI0F11487g; Split Finger Protein; TF123). Sfp1 is a stress- and nutrient-sensitive regulator of ribosomal protein gene expression, localized to the nucleus in exponentially dividing cells [67]. Sfp1 was recognized as a downstream effector of the TOR pathway (in part through the PKA pathway), which is inactivated under stress or nutrient limitation; under such conditions, Sfp1 is relocalized through nucleocytoplasmic trafficking to the cytoplasm, and the expression of ribosomal protein genes is down-regulated. In this sense, Sfp1 localization is regulated in a manner opposite to that of the known TOR downstream transcriptional effectors (Msn2/4, Gln3, Gat1, or Rtg1/3) activating stress-response programs. Considering its mode of operation, we expected that an increased abundance of Sfp1 would contribute to significantly enhanced growth and r-Prots synthesis under favorable environmental conditions. Indeed, we observed some enhancement in growth under selected variables combined with high OA; and a corresponding increase in r-Prots synthesis (Fig. 3d). However, the level of FC values either for growth or r-Prots was rather disappointing. Considering the abovementioned pattern of Sfp1 relocalization in response to different stimuli (including the phase of growth), we hypothesize that the sampling time, which was optimized to enable standardized reading of the majority of phenotypes from the collection, was, in this specific case, inadequate. Phenotype reading was conducted at 48 h of culturing—by that time, all the strains reached a stationary phase of growth, but nutrients were not yet limiting. As demonstrated for *S. cerevisiae* [67], the growth rate cessation in the stationary phase is accompanied by cytoplasmic localization of Sfp1, and inactivation of its transcriptional program. Experiments testing this hypothesis are also currently underway.

Azf1 (YALI0A16841g) and Dep1 (YALI0F05896g) act specifically against r-Prots synthesis

Browsing YaliFunTome database revealed the existence of specific silencers of r-Prots synthesis, in opposition to the abovementioned global enhancers. Such global silencers specific to r-Prots synthesis are characterized by not affected growth (mean FC: 0.96–1.01), but severely decreased r-Prots synthesis capacity (mean FC: 0.4–0.49). Among these, we identified Dep1 (YALI0F05896g; Deregulated Expression of Phospholipid biosynthesis), whose OE drastically reduced the accumulation of the reporter r-Prot (Fig. 4a). Previously, Dep1 was found to enhance the accumulation of lipids in *Y. lipolytica* [37], and specifically—activate phospholipid biosynthesis in *Fusarium* sp. [68]. In contrast, [69] identified Dep1 as a repressor of phospholipids synthesis genes (e.g. *INO1, CHO1, OPI3*) in *S. cerevisiae*. Our auxiliary experiments demonstrated that the *Δdep1* strain displays an inverted phenotype in terms of r-Prots synthesis—enhances it by ~ threefold ('optimal' conditions; Additional file 1: Fig.

S1b). Such behavior suggests its direct implication in the regulation of this process.

This kind of direct interaction was not observed for the second identified 'global silencer' of r-Prot synthesis, Azf1 (YALI0A16841g; Asparagine-rich Zinc Finger protein). While its OE triggered a dramatic decrease in r-Prots synthesis (Fig. 4.b), *Δazf1* mutation rendered no aberrant phenotype in this regard ('optimal' conditions; Additional file 1: Fig. S1b). According to SGD (*Saccharomyces* Genome Database), Azf1 is a carbon source-responsive TF, and in the presence of glucose, it activates genes involved in growth, carbon metabolism, and invasive growth. Indeed, we observed that the *AZF1*-OE strain displayed enhanced filamentation when compared to the control strain (Additional file 1: Fig. S2), and that the 'carbon' variable had a significant contribution to the developed phenotype (dominant level − 1, glucose; Fig. 4b).

Notably, for both the 'global silencers' we noticed that the OA variable's direction of contribution to r-Prots synthesis was inverted when compared to the control strain. Typically, high OA was the key variable with the highest promoting impact on the (normalized) r-Prot synthesis; practically meaning that high OA (even if normalized per control strain that also synthetizes r-Prot) still has the highest importance in shaping the response. For Dep1- and Azf1-driven response, it was the low OA, that promoted the limiting phenotype display (the higher OA the lower the 'normalized r-Prot' value). Similarly, such an 'inverted phenotype' in response to OA was observed for a strain OE-ing *CAT8* gene (YALI0C19151g; CATabolite repression; Fig. 4c), whose OE also triggered a decrease in r-Prots synthesis, particularly marked under high OA. Cat8 is a Zinc cluster transcriptional activator, binding to carbon source-responsive elements in the absence of glucose (contrary to Mig1). Although the Cat8-induced phenotype was not as radical as in the case of Azf1 or Dep1 [growth mean FC: 0.96; (normalized) r-Prot mean FC: (0.82) 0.76], it still justified its classification as a 'global silencer'. Its known involvement

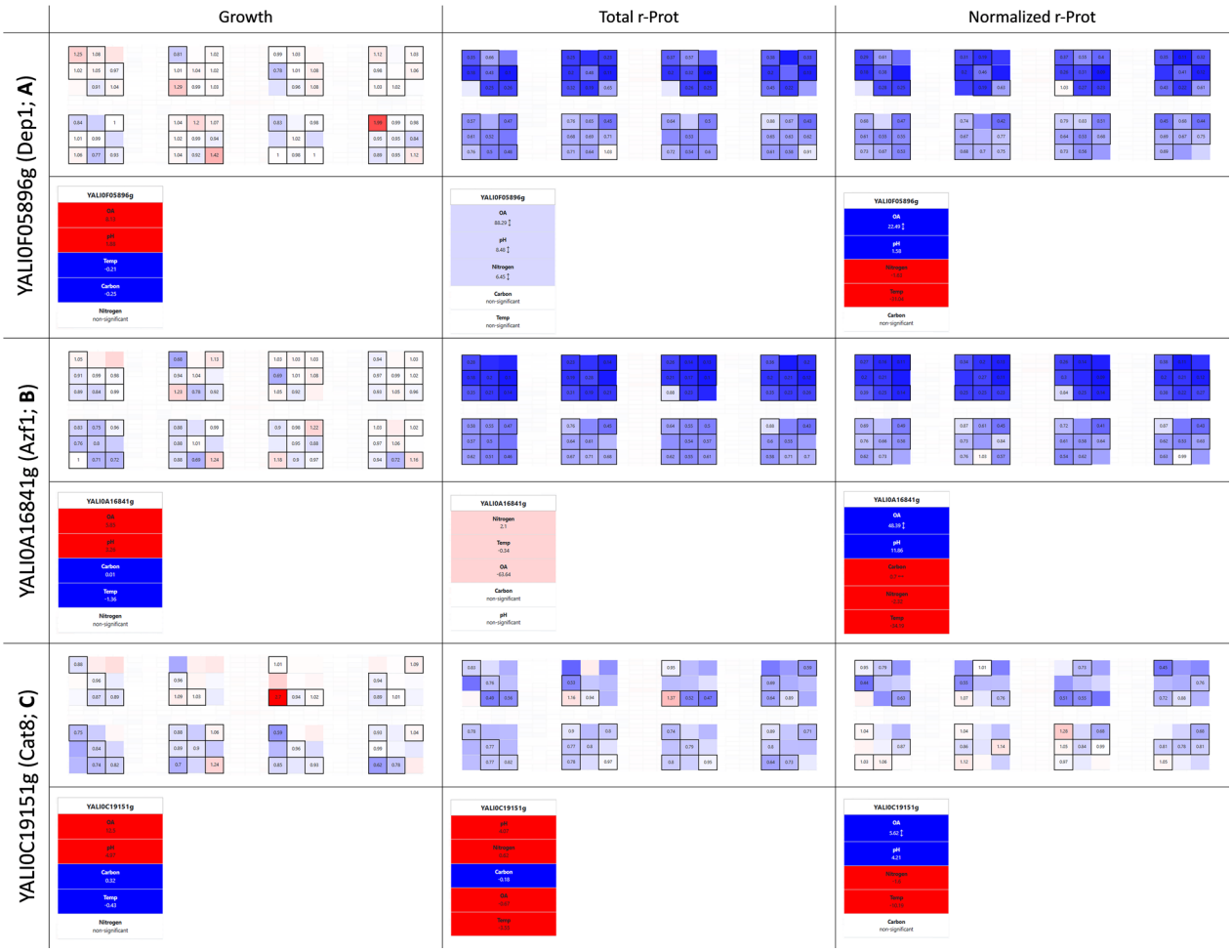


Fig. 4 Selected phenotypes from the YaliFunTome database—the repressors of r-Prots synthesis. Legend as in Fig. 3

in the regulation of gluconeogenesis and the glyoxylate cycle makes it functionally similar to the two others, suggesting the operation of a general pattern: activation of carbon anabolism genes specifically limits protein synthesis in a categoric manner.

Unmaking anoxia a problem for Y. lipolytica

As mentioned above, insufficient OA is the key technical limitation in industrial bioprocesses implementing *Y. lipolytica*. Its strictly aerobic metabolism, relying on the TCA cycle and β-oxidation, forces high oxygen provision for full development of phenotype. Apart from the implementation of technical bioprocessing solutions aiding higher oxygen provision, some efforts to minimize the high requirement for oxygen by genetic manipulations have been made. One such interesting approach comprised the OE of bacterial hemoglobin [47, 70], which ultimately led to the enhancement of r-Prot and erythritol synthesis.

Our methodological approach followed by browsing the YaliFunTome database enabled the identification of TFs that displayed significantly enhanced growth (compared to the control strain) when the OA limitation was inflicted (Fig. 5a–d). These were TF036 (YALI0D20460g; contribution of low OA higher by 14.6% than in the control, calculated by modeling FC data), Jmc2 (YALI0B14443g; TF009; 10.6%), TF011 (YALI0B20944g; 7.3%), and Dal81 (YALI0D02783g; TF118; 2.7%) (Additional file 1: Fig. S3). All the TFs shared the same pattern for averaged responses for enhanced growth (FC: 1.06–1.14), and a slightly inhibitory effect on r-Prots synthesis (FC: 0.91–0.97). The highest FC gain in growth under low OA reached even 60–80% under OE of these TFs. Except for Dal81, all these TFs contributed to significant limitations in total r-Prots synthesis, typically decreased by 10–20%, even under conditions promoting growth (FC for growth up, while FC for r-Prots down). The function for these TFs in *Y. lipolytica* has not been assigned yet. A blastp search (under default settings) showed that TF036 shares some similarities with a positive regulator of lactose-galactose metabolism, Jmc2 is similar to lysine-specific methylase, TF011 is similar solely to the other putative proteins from *Y. lipolytica*, while TF118 shows similarity to Dal81 TF. Recently, Jmc2 has been identified as one of the upregulated genes in *Y. lipolytica* evolved strain exhibiting enhanced growth and lipid accumulation [71]. In that study, it was defined as a representative of JmjC domain family of histone demethylase which promotes global demethylation of H3K4 subunit. According to SGD, Dal81 (Degradation of Allantoin) is a positive regulator of genes in multiple nitrogen degradation pathways; however detailed studies demonstrated that the TF most probably does not bind to the specific motif in the promoter region of many genes involved in allantoin catabolism. The current lack of solid evidence for the TFs' putative functions disallows any speculations on the possible mechanisms by which the enhanced growth under anoxia is achieved. Considering the significance of practical implications of identifying 'anoxia-resistance factors' in *Y. lipolytica*, further functional studies are underway.

In strong contrast to the effects triggered by the TFs enhancing growth under low OA, we identified such that make *Y. lipolytica* significantly more sensitive to limited OA (Fig. 5e). As summarized in SGD, Hap1 (YALI0F17424g; TF120; Heme Activating Protein; previously referred to as CYP1) is known to be involved in the transcriptional regulation in response to OA using its ability to bind heme as a proxy for detecting oxygen levels. It is localized in the nucleus and mitochondrion, and can act as both – activator and repressor of transcription [10]. Under high OA, Hap1 activates the genes involved in respiration, cell cycle progression, and ROS defense, promoting growth and respiratory metabolism in *S. cerevisiae* [10]. While the mechanisms of oxygen sensing and heme signaling in *Y. lipolytica* have not been investigated to date, our current results strongly suggest that the YlHap1 homolog is involved in such regulation (Fig. 5e). Our data demonstrate a very strong limitation of growth upon a combination of *HAP1*-OE and low OA. In previous studies with *S. cerevisiae*, OE of *HAP1* triggered a significant increase in the r-Prots synthesis capacity [11]. Insight into Hap1's regulome revealed that it activates a set of oxidative stress response genes, therefore mitigating the negative effect of ROS accumulation associated with protein folding. In our current studies, the total r-Prot synthesis was not enhanced, however, when normalized per growth, the r-Prots synthesis capacity of *Y. lipolytica* was enhanced by an average value of 14% (Fig. 5e). That increase was mainly observed under low OA, hallmarking that the growth limitation under this condition was at least partly decoupled from r-Prots synthesis (that was enhanced). Interestingly, we observed a highly marked contribution of the pH variable to growth and r-Prot synthesis by the *HAP1*-OE-ing strain. That strain exhibited a strong growth limitation under acidic pH. Hap1's implication in pH response regulation has not been studied to date, in contrast to the evidenced functional operation of the Rim101-governed pathway [72]. Hence, the determination of whether the action of Hap1 is an actual causative factor of the decreased growth under pH 3, remains to be addressed.

What about the golden standards for enhancing stress resistance—Mhy1, Msn4?

In pursuance of stress-resistant yeast phenotypes, engineering one of the general stress response factors is

Fig. 5 Selected phenotypes from the YaliFunTome database—the OA-responsive TFs. Legend as in Fig. 3

frequently applied. These comprise the Msn2/4 zinc finger TFs family that typically binds to cis-acting DNA stress response elements (STRE). In *Y. lipolytica*, two such Msn2/4 family representatives are known and were studied in this context—Mhy1 (YALI0B21582g; TF095) and Msn4 (YALI0C13750g; TF107). *MHY1*'s expression was shown to be dramatically increased in *Y. lipolytica* during the yeast-to-hypha transition, neutral-alkaline pH, and presence of glucose [27, 73]. Mhy1 was indicated as a key positive regulator of both alkaline- and glucose-induced filamentation, inhibited in the presence of glycerol [73]. Hence, in our studies, we expected to observe a strong contribution of these variables to the growth of the *MHY1*-OE-ing strain (Fig. 6a). Indeed, a significant contribution of glucose and a higher pH to the growth of this strain was observed; in addition, to an increased frequency of filament occurrence (Additional file 1: Fig. S2). However, the strain's growth was not promoted (average FC: 0.96), even though some of the inflicted conditions can definitely be considered stressful/promoting Mhy1's action. In high relevance to this observation, [28] evidenced recently that Mhy1 action is not required for increased stress resistance in this species but plays a role in the dimorphic transition. The lack of Mhy1's implication in stress response was confirmed in the following studies [74]; which agrees well with our current observations.

The other representative of the Msn2/4-like TFs family in *Y. lipolytica*, Msn4 was previously shown to serve as a regulator of acid-induced stress [28]. In our studies, this effect upon *MSN4*'s OE was rather moderate (Fig. 6b); although leading to some very high FC in growth (4.25), but in isolated cases. Globally, *MSN4*-OE contributed to minor induction of growth under the adopted conditions (average FC: 1.04). Previously, *Y. lipolytica* strains deleted in *MSN4* loci displayed growth defects when grown at pH 3, while its OE impacted the cells' morphology—reduced cell chain formation in wild-type cells but did not affect cell size [28] (compare Additional file 1: Fig. S2).

For both the Msn2/4-like TFs family, the effect of their OE on r-Prots synthesis was relatively small, and not demonstrating any noticeable pattern. It was rather disappointing considering the recent success of [12], who by OE-ing *MSN4* in *K. phaffi* significantly improved r-Prot synthesis and secretion (from 52% in small-scale cultures, up to > 300% when co-OE with other factors and in bioreactor cultivations).

Counter-intuitive effector acting on r-Prots synthesis in inorganic nitrogen—Hoy1 (YALI0A18469g)

Organic nitrogen, high OA, and mild acidic to neutral pH are the variables expected to promote growth and r-Prot synthesis. Considering that sufficient OA supply is the key technical limitation in the bioprocesses employing *Y. lipolytica*, it was exciting to discover TFs that promote growth under limited oxygen provision ("Unmaking anoxia a problem for *Y. lipolytica*" section). Likewise, considering that the cost of complex, organic nitrogen

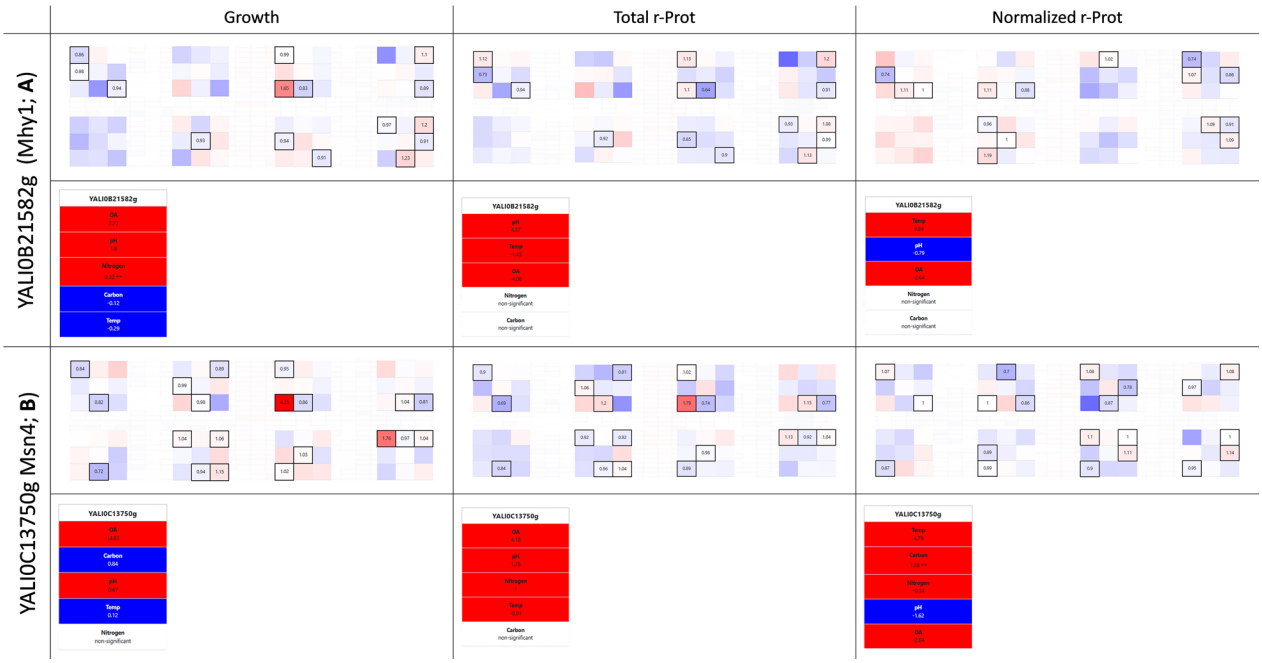


Fig. 6 Selected phenotypes from the YaliFunTome database—the global stress response TFs. Legend as in Fig. 3

source constituted the highest unit production cost in the pilot-scale waste-free process of r-Prot production [49], it would be exciting to identify a factor that would enable enhanced r-Prot synthesis in inorganic nitrogen. Our current data suggest the appearance of such a tendency upon OE of the *HOY1* gene (YALI0A18469g; TF099; Homeobox-containing gene). In *Y. lipolytica* Hoy1 plays a role in the morphological transition and is not essential for growth, but its OE was proved to promote hyphal growth [26]. Based on its promoter sequence architecture, *HOY1* was proposed to be deregulated in response to stress cues, and downstream of signals for amino acid biosynthesis regulation/nitrogen starvation. In fact, this co-occurrence of stress response and nitrogen regulation cis-elements well corresponds with the following findings by Szabo and Štofaníková 2002. It is known that *Y. lipolytica* grows in filamentous morphotype at neutral-alkaline pH, while acidic pH promotes growth in ovoid form. That study evidenced that the pH-dependent morphogenetic shift operates solely in the presence of an organic nitrogen source, and is still operable in *Δrim101* background (the pH sensing and signaling pathway is off). Thus, the Authors concluded that pH affects the formation of hyphae indirectly—by modulation of availability and/or utilization of transportable nitrogen sources, and not by a Rim101-governed, pH-response mechanism. The combination of data from those two studies portrays Hoy1 as a TF whose expression is enhanced under a dimorphic transition that is natively activated in the presence of organic nitrogen.

In our experiments, we synthetically enhanced the abundance of TFs under conditions that natively do not promote their expression. Considering what is known about *Y. lipolytica*'s Hoy1, its expression should be attenuated under inorganic nitrogen source provision. Strikingly, as depicted in (Fig. 7a), its artificially forced presence under these conditions contributed to enhancement in the r-Prots synthesis. Taking into account findings by Szabo and Štofaníková 2002 it is highly plausible that the 'synthetic presence' of Hoy1 facilitated nitrogen capturing under its scarcity; which in turn contributed to enhanced r-Prots synthesis. Indeed, as demonstrated by our factor's contribution analysis, nitrogen was a variable that gained significance in terms of normalized r-Prots synthesis measure for the *HOY1*-OE-ing strain. The importance of the organic nitrogen source presence was decreased when compared to the control strain; and when the models were developed on FC values, the presence of an inorganic nitrogen source (level − 1) was a prerequisite for the enhanced r-Prot synthesis capacity phenotype display (the factor contribution higher by > 7%) (Additional file 1: Fig. S3).

We consider the finding of Hoy1's promoting action in the context of r-Prot synthesis from inorganic nitrogen an important discovery of high applicatory potential.

Conclusions

Without extensive insight into the molecular biology of a cell, one can intuitively state that modification of complex traits, like stress resistance or r-Prots synthesis, will be more adequately addressed using 'global transcription machinery engineering'. Such an approach fine-tunes multiple molecular identities according to a nature-designed program to elicit a specific response. It is the investigator's work to define, which of the programs most adequately corresponds with the target cellular functionality, and hijack it.

In our research, we made an effort to screen a repertoire of the transcriptional programs developed over the evolution of *Y. lipolytica* species. A similar high-throughput study into TFs' action was earlier executed for *S. cerevisiae* [67]; where a collection of > 200 strains OE-ing GFP-fused TFs was investigated for the changes in TF's localization as a proxy of its activation. While that former study aimed at providing evidence on the TFs' activation status, in our research we put great effort into finding the conditions that prompt the

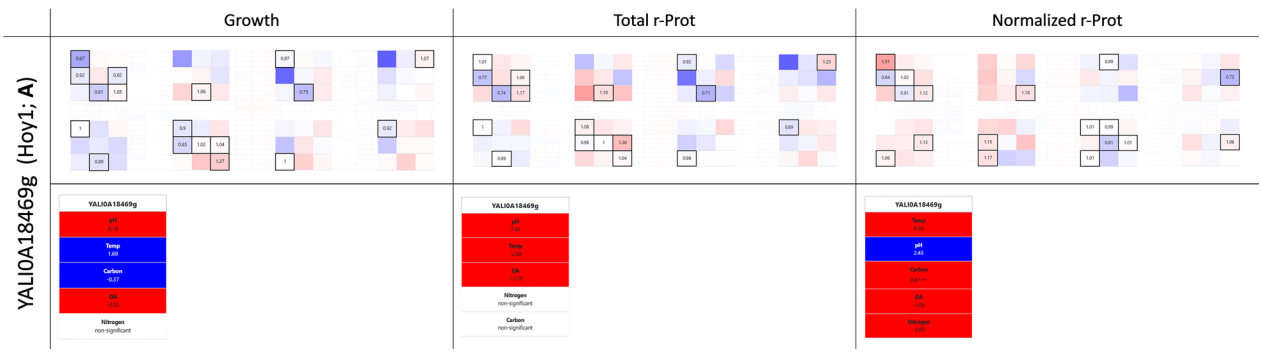


Fig. 7 Selected phenotype from the YaliFunTome database—the nitrogen-responsive TF. Legend as in Fig. 3

TF-driven program's onset, by challenging the strains with an array of variables and their combinations. To enhance the applicatory character of the outcomes, the selection of the variables was kept relevant to the conditions that actually hamper/are relevant to bioprocesses run with *Y. lipolytica*. Likewise, the target functionalities, gauging the relevance of the TFs' programs, were kept practical. Consequently, we developed a streamlined protocol for high-throughput *Y. lipolytica* cultivation that enables reliable assessment of the phenotypes, and provided evidence for the TF-driven programs worth hijacking.

Crucially, we identified the 'omni-booster' of r-Prots synthesis in *Y. lipolytica* – Klf1, which has not been earlier connotated with this functionality. And found no evidence of Msn2/4s' usefulness in enhancing *Y. lipolytica*'s stress resistance. We discovered several TFs that hold a promise of capturing the 'holy grail' of *Y. lipolytica*-based bioprocesses—resistance to fluctuations in OA/growth under oxygen limitation (TF036, Jmc2, TF011, and Dal81), and sustained protein synthesis under inorganic nitrogen supply (Hoy1). These findings are the starting point for the following detailed studies underway.

With the progress in high-throughput cloning, cultivation, and analytics, the processing of large-scale data and their interpretation imposes a challenge. Withdrawal of biological sense from extensive sets of numerical data is sometimes facilitated by arranging these data in queryable and filtrable databases. Completion of our experimental design covering all the *Y. lipolytica* bioprocesses-relevant variables, left us with such a problem, and prompted us to follow the proposed solution. YaliFunTome database was the answer to the arose challenge. We invite all the Yeast Society Members, especially those working with non-*S. cerevisiae* species, to browse the database in search of responses elicited by the TF homologs from *Y. lipolytica*. Our experience shows that the results coming from the non-conventional yeast may be more relevant for many species than those coming from the model.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12934-023-02285-x.

Additional file 1: Figure S1. Recombinant protein production by *Y. lipolytica* strains engineered in genes encoding selected Transcription Factors by their overexpression (OE) or deletion (KO). **Figure S2**. Microscopic images of *Y. lipolytica* strains overexpressing one of the selected Transcription Factors Azf1 (YALI0A16841g), Mhy1 (YALI0B21582g), Msn4 (YALI0C13750g), and the control strain. **Figure S3.** Factor's contribution rankings in terms of growth, r-Prots synthesis, and normalized measure of r-Prots synthesis based on mathematical models developed using FC values readouts, for TFs: TF036 (YALI0D20460g), Jmc2 (YALI0B14443g), TF011 (YALI0B20944g), Dal81 (YALI0D02783g), and Hoy1 (YALI0A18469g). Ranking tables are color-coded according to a convention presented in Fig. 1. Percentage contribution values discussed in the manuscript are bolded.

Additional file 2: Table S1. List of the 125 TFs over-expressed in *Y. lipolytica* strains used in this study.

Acknowledgements
The authors would like to acknowledge the fantastic work done by the data-2biology Ltd. Team in preparing the YaliFunTome database.

Author contributions
EC and JMN conceived the project. EC and MG designed the methodology. MG performed all experiments, analyzed data, and built the mathematical models. EC supervised the research. WB developed the experimental design in DX. EC and MG examined all data and discussed data analyses. EC and MG wrote the manuscript. EC secured funds.

Funding
This research was funded by the National Science Centre, Poland, grant number 2021/41/B/NZ9/00086. Open Access publication costs were covered by a subvention (506.771.09.00B) from the Ministry of Education and Science in Poland received by Poznan University of Life Sciences.

Availability of data and materials
The datasets generated and/or analyzed during the current study are presented directly in the manuscript, online database https://sparrow.up.poznan.pl/tsdatabase/, and as Additional materials.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Biotechnology and Food Microbiology, Poznan University of Life Sciences, 60-637 Poznań, Poland. [2]Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, 78350 Jouy-en-Josas, France.

Received: 6 November 2023 Accepted: 21 December 2023
Published online: 18 January 2024

References

1. Alper H, Stephanopoulos G. Global transcription machinery engineering: a new approach for improving cellular phenotype. Metab Eng. 2007;9:258–67.
2. Duan G, Ding L, Wei D, Zhou H, Chu J, Zhang S, et al. Screening endogenous signal peptides and protein folding factors to promote the secretory expression of heterologous proteins in *Pichia pastoris*. J Biotechnol. 2019. https://doi.org/10.1016/j.jbiotec.2019.06.297.
3. Guerfal M, Ryckaert S, Jacobs PP, Ameloot P, Van Craenenbroeck K, Derycke R, et al. The HAC1 gene from *Pichia pastoris*: characterization and effect of its overexpression on the production of secreted, surface displayed and membrane proteins. Microb Cell Fact. 2010;9:1–12.
4. Liu J, Han Q, Cheng Q, Chen Y, Wang R, Li X, et al. Efficient expression of human lysozyme through the increased gene dosage and co-expression of transcription factor Hac1p in *Pichia pastoris*. Curr Microbiol. 2020;77:846–54.
5. Korpys-Woźniak P, Celińska E. Molecular background of HAC1-driven improvement in the secretion of recombinant protein in *Yarrowia*

lipolytica based on comparative transcriptomics. Biotechnol Rep. 2023;38: e00801.
6. Korpys-Woźniak P, Kubiak P, Celińska E. Secretory helpers for enhanced production of heterologous proteins in *Yarrowia lipolytica*. Biotechnol Rep. 2021;32:e00669.
7. Lindquist S, Craig EA. The heat-shock proteins. Annu Rev Genet. 1988;22:631–77.
8. Sorger PK. Heat shock factor and the heat shock response. Cell. 1991;65:363–6.
9. Hou J, Österlund T, Liu Z, Petranovic D, Nielsen J, Osterlund T, et al. Heat shock response improves heterologous protein secretion in *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol. 2013;97:3559–68.
10. Zhang L, Hach A. Molecular mechanism of heme signaling in yeast: the transcriptional activator Hap1 serves as the key mediator. Cell Mol Life Sci. 1999;56:415–26.
11. Martínez JL, Meza E, Petranovic D, Nielsen J. The impact of respiration and oxidative stress response on recombinant α-amylase production by *Saccharomyces cerevisiae*. Metab Eng Commun. 2016;3:205–10.
12. Zahrl RJ, Prielhofer R, Burgard J, Mattanovich D, Gasser B. Synthetic activation of yeast stress response improves secretion of recombinant proteins. N Biotechnol. 2023;73:19–28.
13. Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, et al. Genome evolution in yeasts. Nature. 2004;430:35–44.
14. Madzak C. *Yarrowia lipolytica* strains and their biotechnological applications: how natural biodiversity and metabolic engineering could contribute to cell factories improvement. Journal of Fungi. 2021;7:548.
15. Papanikolaou S, Muniglia L, Chevalot I, Aggelis G, Marc I. *Yarrowia lipolytica* as a potential producer of citric acid from raw glycerol. J Appl Microbiol. 2002;92:737–44.
16. Rakicka-Pustułka M, Mirończuk AM, Celińska E, Białas W, Rymowicz W. Scale-up of the erythritol production technology: process simulation and techno-economic analysis. J Clean Prod. 2020;257:120533.
17. Celińska E, Borkowska M, Białas W, Kubiak M, Korpys P, Archacka M, et al. Genetic engineering of Ehrlich pathway modulates production of higher alcohols in engineered *Yarrowia lipolytica*. FEMS Yeast Res. 2019;19:1–13.
18. Gomes N, Teixeira JA, Belo I. Fed-batch versus batch cultures of *Yarrowia lipolytica* for γ-decalactone production from methyl ricinoleate. Biotechnol Lett. 2012;34:649–54.
19. Larroude M, Celinska E, Back A, Thomas S, Nicaud JM, Ledesma-amaro R. A synthetic biology approach to transform *Yarrowia lipolytica* into a competitive biotechnological producer of β-carotene. Biotechnol Bioeng. 2018;115:464–72.
20. Matthäus F, Ketelhot M, Gatter M, Barth G. Production of lycopene in the non-carotenoid-producing yeast *Yarrowia lipolytica*. Appl Environ Microbiol. 2014;80:1660–9.
21. Ledesma-Amaro R, Nicaud JM. *Yarrowia lipolytica* as a biotechnological chassis to produce usual and unusual fatty acids. Prog Lipid Res. 2016;61:40–50.
22. Rigouin C, Gueroult M, Croux C, Dubois G, Borsenberger V, Barbe S, et al. Production of medium chain fatty acids by *Yarrowia lipolytica*: combining molecular design and TALEN to engineer the fatty acid synthase. ACS Synth Biol. 2017. https://doi.org/10.1021/acssynbio.7b00034.
23. Madzak C. *Yarrowia lipolytica*: recent achievements in heterologous protein expression and pathway engineering. Appl Microbiol Biotechnol. 2015;99:4559–77.
24. Theron CW, Vandermies M, Telek S, Steels S, Fickers P. Comprehensive comparison of *Yarrowia lipolytica* and *Pichia pastoris* for production of *Candida antarctica* lipase B. Sci Rep. 2020;10:1–9.
25. Celińska E, Nicaud J-M. Filamentous fungi-like secretory pathway strayed in a yeast system: peculiarities of *Yarrowia lipolytica* secretory pathway underlying its extraordinary performance. Appl Microbiol Biotechnol. 2019;103:39–52.
26. Torres-Guzmán JC, Domínguez A. HOY1, a homeo gene required for hyphal formation in *Yarrowia lipolytica*. Mol Cell Biol. 1997;17:6283–93.
27. Hurtado CAR, Rachubinski RA. Mhy1 encodes a c2h2-type zinc finger protein that promotes dimorphic transition in the yeast *Yarrowia lipolytica*. J Bacteriol. 1999;181:3051–7.
28. Wu H, Shu T, Mao Y-S, Gao X-D. Characterization of the promoter, downstream target genes and recognition DNA sequence of Mhy1, a key filamentation-promoting transcription factor in the dimorphic yeast *Yarrowia lipolytica*. Curr Genet. 2019. https://doi.org/10.1007/s00294-019-01018-1.
29. Hurtado CAR, Rachubinski RA. YlBMH1 encodes a 14-3-3 protein that promotes filamentous growth in the dimorphic yeast *Yarrowia lipolytica*. Microbiology (Reading). 2002;148:3725–35.
30. Morales-Vargas AT, Domínguez A, Ruiz-Herrera J. Identification of dimorphism-involved genes of *Yarrowia lipolytica* by means of microarray analysis. Res Microbiol. 2012;163:378–87.
31. Martinez-Vazquez A, Gonzalez-Hernandez A, Domínguez Á, Rachubinski R, Riquelme M, Cuellar-Mata P, et al. Identification of the transcription factor Znc1p, which regulates the yeast-to-hypha transition in the dimorphic yeast *Yarrowia lipolytica*. PLoS ONE. 2013;8: e66790.
32. Pomraning KR, Bredeweg EL, Kerkhoven EJ, Barry K, Haridas S, Hundley H, et al. Regulation of yeast-to-hyphae transition in *Yarrowia lipolytica*. mSphere. 2018;3:1–18.
33. Endoh-Yamagami S, Hirakawa K, Morioka D, Fukuda R, Ohta A. Basic helix-loop-helix transcription factor heterocomplex of Yas1p and Yas2p regulates cytochrome P450 expression in response to alkanes in the yeast *Yarrowia lipolytica*. Eukaryot Cell. 2007;6:734–43.
34. Hirakawa K, Kobayashi S, Inoue T, Endoh-Yamagami S, Fukuda R, Ohta A. Yas3p, an opi1 family transcription factor, regulates Cytochrome P450 expression in response to n-alkanes in *Yarrowia lipolytica*. J Biol Chem. 2009;284:7126–37.
35. Poopanitpan N, Kobayashi S, Fukuda R, Horiuchi H, Ohta A. An ortholog of farA of *Aspergillus nidulans* is implicated in the transcriptional activation of genes involved in fatty acid utilization in the yeast *Yarrowia lipolytica*. Biochem Biophys Res Commun. 2010;402:731–5.
36. Wang Z-P, Xu H-M, Wang G-Y, Chi Z, Chi Z-M. Disruption of the MIG1 gene enhances lipid biosynthesis in the oleaginous yeast *Yarrowia lipolytica* ACA-DC 50109. Biochim Biophys Acta. 2013;1831:675–82.
37. Leplat C, Nicaud J-MM, Rossignol T. Overexpression screen reveals transcription factors involved in lipid accumulation in *Yarrowia lipolytica*. FEMS Yeast Res. 2018;18:1–9.
38. Trébulle P, Nicaud JM, Leplat C, Elati M. Inference and interrogation of a coregulatory network in the context of lipid accumulation in *Yarrowia lipolytica*. NPJ Syst Biol Appl. 2017;3:1–8.
39. Gorczyca M, Nicaud JM, Celińska E. Transcription factors enhancing synthesis of recombinant proteins and resistance to stress in *Yarrowia lipolytica*. Appl Microbiol Biotechnol. 2023. https://doi.org/10.1007/s00253-023-12607-z.
40. Monteiro PT, Oliveira J, Pais P, Antunes M, Palma M, Cavalheiro M, et al. YEASTRACT+: a portal for cross-species comparative genomics of transcription regulation in yeasts. Nucleic Acids Res. 2020;48:D642–9.
41. Teixeira MC, Viana R, Palma M, Oliveira J, Galocha M, Mota MN, et al. YEASTRACT+: a portal for the exploitation of global transcription regulation and metabolic model data in yeast biotechnology and pathogenesis. Nucleic Acids Res. 2023;51:D785–91.
42. Leplat C, Nicaud JM, Rossignol T. High-throughput transformation method for *Yarrowia lipolytica* mutant library screening. FEMS Yeast Res. 2015;15:1–9.
43. Duetz WA. Microtiter plates as mini-bioreactors: miniaturization of fermentation methods. Trends Microbiol. 2007;15:469–75.
44. Gorczyca M, Kaźmierczak J, Steels S, Fickers P, Celińska E. Impact of oxygen availability on heterologous gene expression and polypeptide secretion dynamics in *Yarrowia lipolytica*-based protein production platforms. Yeast. 2020;37:559–68.
45. Rywinska A, Musial I, Rymowicz W, Zarowska B, Boruczkowski T. Effect of agitation and aeration on the citric acid production by *Yarrowia lipolytica* grown on glycerol. Prep Biochem Biotechnol. 2012;42:279–91.
46. Li C, Yang X, Gao S, Wang H, Lin CSK, Sze C, et al. High efficiency succinic acid production from glycerol via in situ fibrous bed bioreactor with an engineered *Yarrowia lipolytica*. Bioresour Technol. 2017;225:9–16.
47. Mirończuk AM, Kosiorowska KE, Biegalska A, Rakicka-Pustułka M, Szczepańczyk M, Dobrowolski A. Heterologous overexpression of bacterial hemoglobin VHb improves erythritol biosynthesis by yeast *Yarrowia lipolytica*. Microb Cell Fact. 2019;18:1–8.
48. Sassi H, Delvigne F, Kallel H, Fickers P. pH and not cell morphology modulate pLIP2 induction in the dimorphic Yeast *Yarrowia lipolytica*. Curr Microbiol. 2017;74:413–7.

49. Kubiak M, Białas W, Celińska E. Thermal treatment improves a process of crude glycerol valorization for the production of a heterologous enzyme by *Yarrowia lipolytica*. Biotechnol Rep. 2021;31:e00648.
50. Lubuta P, Workman M, Kerkhoven EJ, Workman CT. Investigating the influence of glycerol on the utilization of glucose in *Yarrowia lipolytica* using RNA-seq-based transcriptomics. G3 Genes Genomes Genetics. 2019;9:4059.
51. Hapeta P, Kerkhoven EJ, Lazar Z. Nitrogen as the major factor influencing gene expression in *Yarrowia lipolytica*. Biotechnol Rep. 2020;27:e00521.
52. Dulermo R, Brunel F, Dulermo T, Ledesma-Amaro R, Vion J, Trassaert M, et al. Using a vector pool containing variable-strength promoters to optimize protein production in *Yarrowia lipolytica*. Microb Cell Fact. 2017;16:1.
53. Gasmi N, Ayed A, Nicaud JM, Kallel H. Design of an efficient medium for heterologous protein production in *Yarrowia lipolytica*: case of human interferon alpha 2b. Microb Cell Fact. 2011;10:1.
54. Szabo R, Štofaníková V. Presence of organic sources of nitrogen is critical for filament formation and pH-dependent morphogenesis in *Yarrowia lipolytica*. FEMS Microbiol Lett. 2002;206:45–50.
55. Ruiz-Herrera J, Sentandreu R. Different effectors of dimorphism in *Yarrowia lipolytica*. Arch Microbiol. 2002;178:477–83.
56. Kubiak-Szymendera M, Skupień-Rabian B, Jankowska U, Celińska E. Hyperosmolarity adversely impacts recombinant protein synthesis by *Yarrowia lipolytica*—molecular background revealed by quantitative proteomics. Appl Microbiol Biotechnol. 2021;106:349–67.
57. Monteiro PT, Pais P, Costa C, Manna S, Sa-Correia I, Teixeira MC. The PathoYeastract database: an information system for the analysis of gene and genomic transcription regulation in pathogenic yeasts. Nucleic Acids Res. 2017;45:D597-603.
58. Brejová B, Vozáriková V, Agarský I, Derková H, Fedor M, Harmanová D, et al. y-mtPTM: Yeast mitochondrial posttranslational modification database. Genetics. 2023;224:iyad087.
59. Dubreuil B, Sass E, Nadav Y, Heidenreich M, Georgeson JM, Weill U, et al. Yeastrgb: comparing the abundance and localization of yeast proteins across cells and libraries. Nucleic Acids Res. 2019;47:D1245–9.
60. Mekouar M, Blanc-Lenfle I, Ozanne C, Da Silva C, Cruaud C, Wincker P, et al. Detection and analysis of alternative splicing in *Yarrowia lipolytica* reveal structural constraints facilitating nonsense-mediated decay of intron-retaining transcripts. Genome Biol. 2010;11:R65.
61. Shimanuki M, Uehara L, Pluskal T, Yoshida T, Kokubu A, Kawasaki Y, et al. Klf1, a C2H2 zinc finger-transcription factor, is required for cell wall maintenance during long-term quiescence in differentiated G0 phase. PLoS ONE. 2013;8:e78545.
62. Herholz M, Cepeda E, Baumann L, Kukat A, Hermeling J, Maciej S, et al. KLF-1 orchestrates a xenobiotic detoxification program essential for longevity of mitochondrial mutants. Nat Commun. 2019;10:3323.
63. Korpys-Woźniak P, Celińska E. Global transcriptome profiling reveals genes responding to overproduction of a small secretory, a high cysteine- and a high glycosylation-bearing protein in *Yarrowia lipolytica*. Biotechnol Rep. 2021;31:e00646.
64. Pomraning KR, Bredeweg EL, Baker SE. Regulation of nitrogen metabolism by GATA zinc finger transcription factors in *Yarrowia lipolytica*. mSphere. 2017;2:1–19.
65. Gorczyca M, Każmierczak J, Fickers P, Celińska E. Synthesis of secretory proteins in *Yarrowia lipolytica*: effect of combined stress factors and metabolic load. Int J Mol Sci. 2022;23:3602.
66. Masser AE, Kang W, Roy J, Mohanakrishnan Kaimal J, Quintana-Cordero J, Friedländer MR, et al. Cytoplasmic protein misfolding titrates Hsp70 to activate nuclear Hsf1. Elife. 2019;8:e47791.
67. Marion RM, Regev A, Segal E, Barash Y, Koller D, Friedman N, et al. Sfp1 is a stress- and nutrient-sensitive regulator of ribosomal protein gene expression. Proc Natl Acad Sci USA. 2004;101:14315–22.
68. Zhang Y, Wang L, Liang S, Zhang P, Kang R, Zhang M, et al. FpDep1, a component of Rpd3L histone deacetylase complex, is important for vegetative development, ROS accumulation, and pathogenesis in *Fusarium pseudograminearum*. Fungal Genet Biol. 2020;135:103299.
69. Lamping E, Lückl J, Paltauf F, Henry SA, Kohlwein SD. Isolation and characterization of a mutant of *Saccharomyces cerevisiae* with pleiotropic deficiencies in transcriptional activation and repression. Genetics. 1994;137:55–65.
70. Bhave SL, Chattoo BB. Expression of *Vitreoscilla* hemoglobin improves growth and levels of extracellular enzyme in *Yarrowia lipolytica*. Biotechnol Bioeng. 2003;84:658–66.
71. Tsirigka A, Theodosiou E, Patsios SI, Tsoureki A, Andreadelli A, Papa E, et al. Novel evolved *Yarrowia lipolytica* strains for enhanced growth and lipid content under high concentrations of crude glycerol. Microb Cell Fact. 2023;22:1.
72. Lambert M, Blanchin-Roland S, Dé F, Le Louedec R, Lé Pingle AE, Gaillardin C. Genetic analysis of regulatory mutants affecting synthesis of extracellular proteinases in the yeast *Yarrowia lipolytica*: identification of a RIM101/pacC homolog. Mol Cell Biol. 1997. https://doi.org/10.1128/MCB.17.7.3966.
73. Shu T, He X-Y, Chen J-W, Mao Y-S, Gao X-D. The pH-responsive transcription factors YlRim101 and Mhy1 regulate alkaline ph-induced filamentation in the dimorphic yeast *Yarrowia lipolytica*. mSphere. 2021;6:10.
74. Konzock O, Norbeck J. Deletion of MHY1 abolishes hyphae formation in *Yarrowia lipolytica* without negative effects on stress tolerance. PLoS ONE. 2020;15:1–11.

Publisher's Note

5.4 **P3**: “Mała objętość – duży problem”: hodowle drożdży *Yarrowia lipolytica* w wysoko-przepustowych zminiaturyzowanych hodowlach

Inspiracją do przygotowania tej nietypowej w swym formacie publikacji były doświadczenia zdobyte podczas optymalizacji protokołu hodowlanego, które ujawniły, jak trudnym zagadnieniem jest miniaturyzacja hodowli *Y. lipolytica*, która nieuchronnie prowadzi do zaburzeń w manifestacji właściwego fenotypu. Ściśle tlenowy metabolizm, skłonność do formowania strzępek i adhezji do hydrofobowych powierzchni, spontaniczna flokulacja oraz silne zakwaszanie podłoża hodowlanego to cechy, które sprawiają, że transfer protokołów w mikroskali opracowanych dla innych gatunków jest w tym przypadku wyjątkowo trudny. Po serii doświadczeń własnych, oraz Kolegów pracujących z *Y. lipolytica* (Beneyton i in., 2017; Lebrun i in., 2023; Leplat i in., 2018), którzy opisali swe doświadczenia w publikacjach, stało się oczywiste, że bez dodatkowej „spersonalizowanej” dla *Y. lipolytica* optymalizacji, zarówno protokoły oparte na płytkach MTP (ang. *micro-titer plates*), jak i te bazujące na analizie pojedynczych komórek, są bezużyteczne do precyzyjnych badań fenotypów tych drożdży.

Publikacja ta stanowi z jednej strony podsumowanie aktualnego stanu wiedzy w zakresie miniaturyzacji i prowadzenia wysoko przepustowych hodowli *Y. lipolytica*, a z drugiej – prezentację wyników eksperymentu typu ‘*proof-of-concept’*; który stanowi zobrazowanie problemu i uzasadnienie podjęcia tego tematu w części przeglądowej. Omówiono w niej najczęściej występujące trudności i strategie ich przezwyciężania, w tym badania z prowadzeniem hodowli w różnych formatach płytek MTP, w mikro- i pikobioreaktorach, a także podejścia wykorzystujące mikroprzepływy (ang. *microfluidics*), systemy zrobotyzowane (np. BioLector) czy technologię FADS (ang. *fluorescence-assisted droplet sorting*). Ta ostatnia metoda umożliwia analizę metabolitów wydzielanych do środowiska, a nie tylko frakcji wewnątrzkomórkowej, jak w przypadku metody FACS (ang. *fluorescence-assisted cell sorting*). Jako przykład ciekawego rozwiązania problemu filamentacji wskazano strategię inżynierii genetycznej, polegającej na delecji głównego regulatora tego procesu – TF Mhy1 (*YALI0B21582g*). Szczególną uwagę poświęcono mechanistycznym ograniczeniom zminiaturyzowanych hodowli – niewystarczającemu natlenieniu i małym objętościom roboczym – które w przypadku gatunków o wysokich wymaganiach tlenowych, jak *Y. lipolytica*, są szczególnie problematyczne.

Celem eksperymentu ‘*proof-of-concept*’ była ocena powtarzalności wyników uzyskiwanych z hodowli w 13 różnych formatach, oraz wpływu sposobu hodowli na metabolizm *Y. lipolytica*. Jak pokazano, hodowle w 96-dołkowych płytkach okazały się nieadekwatne do porównywania fenotypów *Y. lipolytica*, natomiast 24-dołkowe płytki zapewniały wysoką powtarzalność, efektywne wykorzystanie substratu, intensywny wzrost i produkcję rProts. Wyniki te wskazały, że nawet dla tak „wymagającego” gatunku jak *Y. lipolytica*, format 24-dołkowy może stanowić praktyczną alternatywę dla klasycznych kolb Erlenmayera.

Ostatnia część artykułu **P3** zawiera opis naszego starannie zoptymalizowanego protokołu wysokoprzepustowych hodowli w płytkach kwadratowych MTP, pierwotnie opublikowanego w **P2** (zgodnie z decyzją redaktora, szczegółowe omówienie zagadnień technicznych wykraczało poza zakres jednej publikacji). W **P3**, w rozdziale poświęconym autorskiemu protokołowi szczegółowo omówiono trudności, na które natrafiliśmy, oraz sposoby ich przezwyciężania, co czyni go praktycznym przewodnikiem dla badaczy planujących przesiewowe hodowle *Y. lipolytica*.

Celińska and Gorczyca
Microbial Cell Factories (2024) 23:184
https://doi.org/10.1186/s12934-024-02465-3

Microbial Cell Factories

REVIEW

Open Access

'Small volume—big problem': culturing *Yarrowia lipolytica* in high-throughput micro-formats

Ewelina Celińska[1*] and Maria Gorczyca[1]

Abstract

With the current progress in the 'design' and 'build' stages of the 'design-build-test-learn' cycle, many synthetic biology projects become 'test-limited'. Advances in the parallelization of microbes cultivations are of great aid, however, for many species down-scaling leaves a metabolic footprint. *Yarrowia lipolytica* is one such demanding yeast species, for which scaling-down inevitably leads to perturbations in phenotype development. Strictly aerobic metabolism, propensity for filamentation and adhesion to hydrophobic surfaces, spontaneous flocculation, and high acidification of media are just several characteristics that make the transfer of the micro-scale protocols developed for the other microbial species very challenging in this case. It is well recognized that without additional 'personalized' optimization, either MTP-based or single-cell-based protocols are useless for accurate studies of *Y. lipolytica* phenotypes. This review summarizes the progress in the scaling-down and parallelization of *Y. lipolytica* cultures, highlighting the challenges that occur most frequently and strategies for their overcoming. The problem of *Y. lipolytica* cultures down-scaling is illustrated by calculating the costs of micro-cultivations, and determining the unintentionally introduced, thus uncontrolled, variables. The key research into culturing *Y. lipolytica* in various MTP formats and micro- and pico-bioreactors is discussed. Own recently developed and carefully pre-optimized high-throughput cultivation protocol is presented, alongside the details from the optimization stage. We hope that this work will serve as a practical guide for those working with *Y. lipolytica* high-throughput screens.

Keywords High-throughput screens, Yeast, *Yarrowia lipolytica*, Microfluidics, Micro-titer-plates, Droplet sorting, Square plates, Screening protocol

Introduction

Every synthetic biology project, to avoid bottlenecks, requires fine-tuning the 'design-build-test-learn' cycle at all stages. With the current advent of computational design tools, DNA synthesis capacity, and genome editing technology, a high fraction of synthetic biology projects become 'test-limited' [8]. Advances in the parallelization of cell population cultivations by the use of microtiter plates (MTPs) or macro-plates combined with robotic handling and automatic data acquisition are of great aid [4, 41, 45, 58]. Likewise, the progress in single-cell-based flow cytometry approaches, like fluorescence-assisted cell sorting (FACS) and fluorescence-assisted droplet sorting (FADS), debottlenecks the 'test' stage in specific applications [5, 15, 35, 57, 62].

For some microbial species, the metabolic footprint left by down-scaling to microvolumes is negligible. For example, a good quantitative agreement in the performance of *Escherichia coli* and *Acinetobacter* grown on a microwell scale (1 mL) and in a laboratory stirred-tank

*Correspondence:
Ewelina Celińska
ewelina.celinska@up.poznan.pl
[1] Department of Biotechnology and Food Microbiology, Poznan University of Life Sciences, 60-637 Poznań, Poland

bioreactor (2 L) [21] was reported. Highly similar patterns of growth were also found for several laboratory strains of *Saccharomyces cerevisiae* (including CEN-PK.2) cultured on a 'micro' scale (0.35 mL) and a 'medium' scale of 10 mL [58]. Maximum growth rates and biomass accumulation of *S. cerevisiae* cells grown in MTPs for 'microvinification' were indistinguishable from those observed in self-induced anaerobic flask cultures [42]. Yet, in the majority of cases, scaling-down leaves some metabolic footprint on the cell,making the scaling-up or -down a separate and demanding task in bioprocess development. The problem was highlighted *i.a.* by [51], who performed comparative cultures of engineered *Komagataella phaffi* (formerly–*Pichia pastoris* in deep-well plates and 1 L bioreactor. The small-scale cultures disallowed phenotype development and no differences were seen in volumetric enzyme activity between four differently-engineered strains; while the phenotypes were significantly different in the bioreactor cultivations.

Yarrowia lipolytica, is one of such demanding species, for which scaling-down inevitably leads to perturbations in phenotype development. Strictly aerobic metabolism, propensity for filamentation and adhesion to hydrophobic surfaces, spontaneous flocculation, and high acidification of media are just several characteristics that make the transfer of the micro-scale protocols developed for the other yeast species very challenging. Our own experience and the experience of our Colleagues (e.g. [5]) provide evidence that without additional consideration and optimization, either MTP-based or single-cell-based protocols are useless for accurate studies of *Y. lipolytica* phenotypes.

This review summarizes the research in parallelization of *Y. lipolytica* culturing, highlighting the pitfalls that occur most frequently. We hope that it will serve as a practical guideline for those working with *Y. lipolytica* high-throughput screens.

How small can we go with *Yarrowia*? And what are the costs?

To quantitatively describe this problem, we previously conducted a small 'proof-of-concept' experiment by culturing *Y. lipolytica* strain synthesizing a fluorescent reporter protein (rProt) in different volumes ([11], Table 1). Growth and fluorescence from the intracellular rProt, as well as substrate and metabolite concentration, were analyzed (Fig. 1). For all the thirteen variants we calculated a volumetric mass transfer coefficient (k_La), according to Eq. 1 [43, 44]:

$$k_La = 6.67 \times 10^{-6} \times n^{1.16} \times V_L^{-0.83} \times d_0^{0.38} \times d^{1.92} \quad (1)$$

where, vessel diameter (d), culture volume (V_L), shaking frequency (n), shaking amplitude (d_0).

K_La is a parameter that determines the rate at which a gaseous compound transfers between the gas and the liquid phases. It has two principal components: the mass transfer coefficient (k_L) and the specific exchange surface (a). It is impacted by the geometry of the culturing vessel (volume or maximum diameter), shaking parameters (amplitude and frequency), and the liquid phase properties (volume and viscosity). As demonstrated previously, the oxygen transfer rate depends linearly on the culture volume (V_L), and twice the lower volume V_L contributes to twice the higher oxygen transfer rate [18].

Table 1 Technical parameters of the compared culturing vessels and conditions for mixing the cultures

Code	Topology/type of vessel	Diameter	Total vol	Culture vol	Filling	Shaking frequency	Shaking amplitude	Calc. K_La
	In horizontal plane [mm]		[mL]	[mL]	[%]	[rpm]	[cm]	[/h]
O_96_0.25	○ MTP U-shaped	6.5	0.3	0.2	67	450	0.2	0.25
O_96_0.55	○ MTP U-shaped	6.5	0.3	0.08	27	450	0.2	0.55
O_48_0.29	○ MTP flat-bottom	9.75	1.3	0.5	38	450	0.2	0.29
O_48_0.56	○ MTP flat-bottom	9.75	1.3	0.2	15	450	0.2	0.56
O_24_0.29	○ MTP flat-bottom	15.5	3.3	1	30	380	0.2	0.29
O_24_0.56	○ MTP flat-bottom	15.5	3.3	0.45	14	380	0.2	0.56
□_24_0.55	□ MTP conical-baffled-bottom	17	11	2.5	23	250	1.91	0.55
□_24_1.1	□ MTP conical-baffled-bottom	17	11	1.1	10	250	1.91	1.1
T_0.25	○ tube U-bottom	14	24	2	8	180	1	0.25
T_0.56	○ tube U-bottom	14	24	0.75	3	180	1	0.56
F_0.25	○ Erlenmeyer flask	80	250	150	60	180	2	0.25
F_0.56	○ Erlenmeyer flask	80	250	58	23	180	2	0.56
F_1.12	○ Erlenmeyer flask	80	250	25	10	180	2	1.12

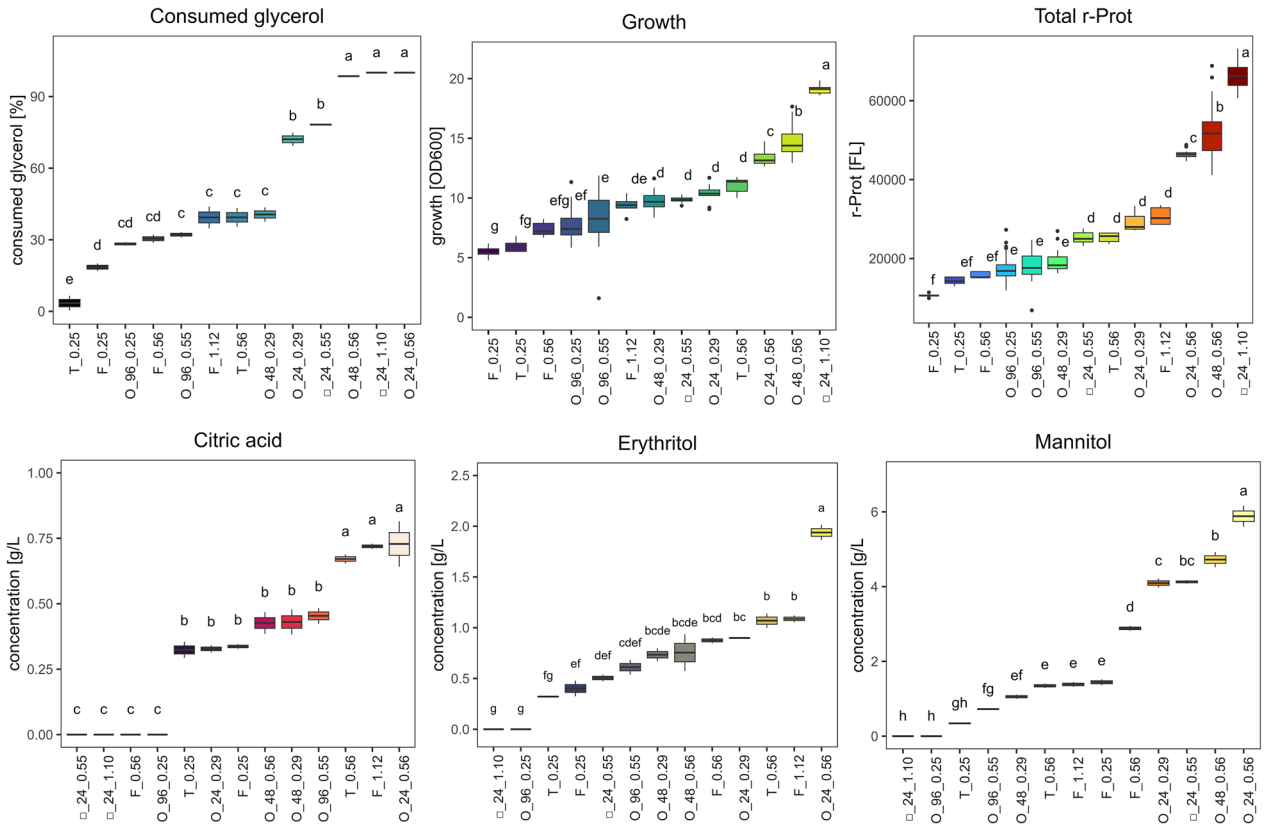


Fig. 1 Glycerol (GLY) consumption [%] metabolites concentration (citric acid, CA, erythritol, ERY, and mannitol, MAN), growth [OD600], and fluorescence from an intracellular reporter protein rProt, determined in the post-culturing of *Y. lipolytica* strain run in different scales and vessels (encoded according to Table 1), at different kLa parameter settings (Mendeley data [11]). *Y. lipolytica* strain used in this experiment was JMY2810 (genotype: *MATa, ura3::pTEF-RedStar2-LEU2-Zeta-URA3ex-pTEF-empty, leu2-270, xpr2-322*; phenotype: ΔAEP, ΔAXP, suc+, ura+, leu+, intracellular RedStarII, Zeta platform). The main cultures were continued for 48 h and samples were collected at the end of the cultivation time. Precultures were developed for 18 h at 28 °C in the 300 mL shake flasks. The main cultures were inoculated at 5% (v/v). Samples were analyzed for growth and fluorescence from the reporter protein (RedStar2) following dilution in 0.75% NaCl (POCH) to match a linear range of the methods. Absorbance was measured at 600 nm in transparent 96-well plates (Costar; Merck). FL was determined under at ex/em 550/595 nm in black opaque plates (Thermo Fisher Scientific). Both measurements were done using a Tecan Spark automatic plate reader (Tecan Group Ltd., Mannedorf, Switzerland). pH level at the end of cultures was determined in the supernatant using indicatory strips (POCh, Poland). Each variant of the flask (F) and tube (T) cultures was conducted in pentaplicate, cultures in 96-well 48-well, and 24-well were conducted in 48, 24, and 12 parallel runs. Values show the average mean from the replicates. Error bars show ± SD from: (i) pentaplicate for flask and tube cultures, (ii) 48, (iii) 24, or (iv) 12 parallel runs of the cultures run in 96-well 48-well, and 24-well, respectively. Letters indicate homogenous groups determined in statistical analysis. Statistical significance of the difference in a given measure was assessed by analysis of variance (ANOVA) test, with a significance level set at p-value < 0.05 (RStudio and Visual Studio Code, Microsoft), and equality of variances was checked with the Levene test. The homogenous groups were calculated using Tukey HSD post-hoc analysis (RStudio with relevant packages)

Parameters such as temperature and composition of the liquid phase, determining solubility and diffusion of oxygen, specifically affect the k_L component. As evidenced earlier [46], minor changes in liquid viscosity caused by the uptake of nutrients and the formation of cellular biomass can be ignored. K_La is commonly used as a proxy of 'aeration rate' in bioprocessing, and a useful tool when re-scaling the process. Equation 1 was developed for standard glass Erlenmeyer flasks with hydrophilic walls, shaking frequencies of 50–500 rpm, relative filling volumes of 4–20% (the relative filling volume is defined as the filling volume divided by the nominal flask volume), shaking diameters of 1.25–10 cm, and nominal flask volume between 50 and 1000 mL [43, 44, 50]. The 'proof-of-concept' experiment [11] covered vessels beyond those indicated (in volume and geometry), so the use of Eq. 1 for modeling kLa in such a wide set of different vessels is burdened with uncertainty. Nevertheless, estimations of the impact of relative culture volume change or culturing system format on the relative change in oxygen availability (expressed in kLa or oxygen transfer velocity, OTV; Table S1) show good agreement between different

studies [11, 16, 18, 26, 29, 59]. For example, for a culture run in O_48-well format, a 40% decrease in volume caused an increase in kLa by 63% in a study by [29], while decreasing the volume by 60% led to an increase in kLa parameter by 93% in the 'proof-of-concept' experiment [11]. Changing the culturing format from O_96-well to O_24-well triggered an increase in the kLa parameter by 9% [16] or 2% [11]. Going down from a test tube culture to O_96-well decreased the kLa level by 46% [59], or 54% [11]. More details on the specific culturing system can be found in Table S1 (Supplementary Material). Still, the kLa values in Table 1 calculated according to Eq. 1 should be perceived as only rough estimations.

As can be inferred from Fig. 1 [11], consumed glycerol (GLY) and growth, but also amounts of rProt, and the principal metabolites (citric acid, CA, erythritol, ERY, and mannitol, MAN) were in high positive correlation (consumed GLY and growth, r = 0.85 across all the conditions). The highest values for growth and GLY consumption were reported for cultures run in 24-well plates, either round (O)- or square (□)-shaped, and in an O-shaped 48-well plate, but in this case—only when k_La was set at 0.56. Comparison of cultures run at different k_La in a specific vessel suggests a straightforward impact of k_La level on GLY consumption, rProt amounts, and the amounts of the major metabolites. Only the case of □_24_1.1 escaped this rule, as no metabolites were found in the post-culturing medium. The lack of metabolites in the post-culturing medium in our experiment could be a consequence of either very robust (□_24_1.1) or very limited (O_96_0.25) metabolism. To our interpretation, in cultures □_24_1.1 *Y. lipolytica* reconsumed CA, ERY, and MAN, which served as an additional carbon source. The highest biomass accumulation in this specific case supports this statement. In other words, the actual amounts of oxygen supplied in □_24_1.1 surpassed those provided in tubes, flasks, 48-well, and 96-well MTPs, enabling oxygenation of all GLY from the medium, as well as the reutilization of own metabolites. This notion complies with a general rule saying that in aerobic metabolism the consumption of the substrate is positively correlated with the oxygen supply. However, when comparing the correlated parameters (e.g. GLY consumption) achieved at the same k_La values but in different vessels, this correlation is no longer valid; for example, at k_La of 0.55- 0.56, GLY was utilized at 30% (O_96_0.55) to 100% (O_24_0.56). Correspondingly, GLY consumption degree could reach from 3.4% (T_0.25) to over 70% (O_24_0.29) at the estimated k_La of ~ 0.27. This notion means that either the calculated k_La values are not adequate measures of oxygen availability in our experimental setup, or that the scaling between selected vessels is subjected to some non-considered 'indirect factors'. Several significant 'indirect factors' affecting such a vessel-to-vessel variation were identified and well-evidenced (discussed hereafter). In our 'proof-of-concept' experiment, the importance of those 'indirect factors' (not covered by Eq. 1) is particularly well seen under the limited oxygen supply. For k_La set at ~ 0.27, the GLY consumption (or growth) could reach one of four (or three) significantly different levels, represented in separate homogenous groups (Fig. 1; [11]). On the other hand, the adopted k_La model [43, 44] does not consider, for example, the head-space volume (partly, but not perfectly expressed as % filling in Table 1). The mass transfer between the head-space and the bulk of the culture is facilitated when compared with the mass transfer across the cover of any kind, acting as a mechanical hindrance. Its direct effect could be reliably assessed in cultures of the same working volume, with the same effective mass transfer area, but differing in the headspace volume (no such condition was studied here).

In our accompanying 'proof-of-concept' experiment the synthesis of metabolites was concomitant with changes in the medium acidity (Fig. 2; [11]). Significant changes in this parameter (> 0.5 unit) were observed in the 24-well MTP cultures and flasks with $k_La > 0.56$. The unusual increase in the pH value in □_24-well cultures is a known phenomenon occurring at the end of *Y. lipolytica*, highlighting cell death [9]. This minor observation on varying pH depending on the culturing vessel, highlights an issue of technical importance. Namely, depending on the growth conditions and aeration rate, the cell population may yield products of different oxidation levels (for

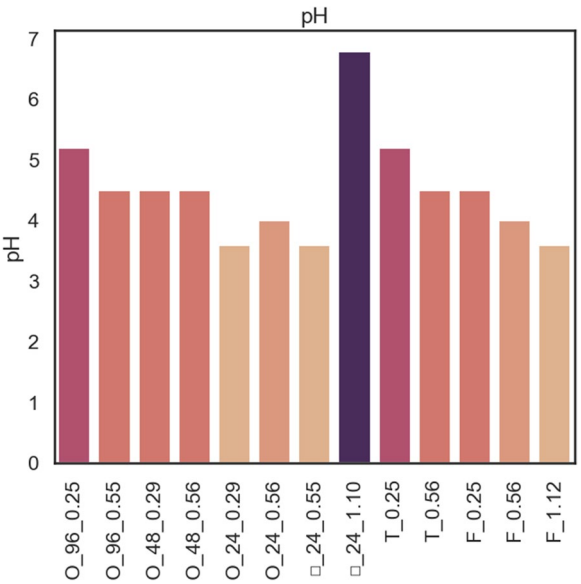


Fig. 2 pH level determined in the culture medium supernatants of *Y. lipolytica* strain run in different scales and vessels (encoded according to Table 1), at different kLa parameter settings (Mendeley data [11])

Y. lipolytica–CO_2 or organic acids/polyols). The variation in the product profile impacts medium acidity, and hence—requires different technical solutions to stabilize this parameter.

To assess reproducibility between repetitions of a specific culture variant, we calculated the percentage of standard deviation within each result for growth and rProt (Fig. 3). The least consistent results were obtained in cultures run in 96-well plates (%SD reached approx. 15–20%), followed by cultures run in test tubes (%SD reached approx. 5–10%). Surprisingly, the least variable were the repetitions of cultures conducted in 24-well plates (and not in the flasks, as presumed). In these cases, the percentage share of SD in the result reached several % (%SD between 2 and 7%). Growth readouts were the most consistent in the cultures run in □_24-well plates (%SD 2–2.7%). The reproducibility of cultures run in these vessels was assessed previously for bacterial cultures (mainly *Pseudomonas* and *Rhodococcus* spp) [18]. For the majority of strains, the duplicates differed by less than 5%. Depending on the strain, the difference between the replicates could reach more (up to 20%).

Collectively, the 'proof-of-concept' experiment showed that going below the 24-well plate with *Y. lipolytica* cultures is burdened with greatly increased well-to-well variability (Fig. 3; the smaller the volume, the bigger the SD). Cultures in 96-well plates seem to be inadequate for *Y. lipolytica* culturing, at least when a comparison of different phenotypes is aimed at, and not just revival of a strain or biomass propagation. Noteworthy, our experiment demonstrated that even for such a 'demanding' species, cultures run in 24-well plates constitute a mature alternative to Erlenmeyer flasks; offering high reproducibility, efficient substrate consumption, growth, and rProt synthesis. This format also greatly helps to make the handling and screening of a large number of strains less time- and material-consuming (the □-plates are reusable). The orbital shaking used for 24-well MTPs was suitable to generate sufficient oxygen provision, even surpassing the ones achieved in Erlenmeyer flasks.

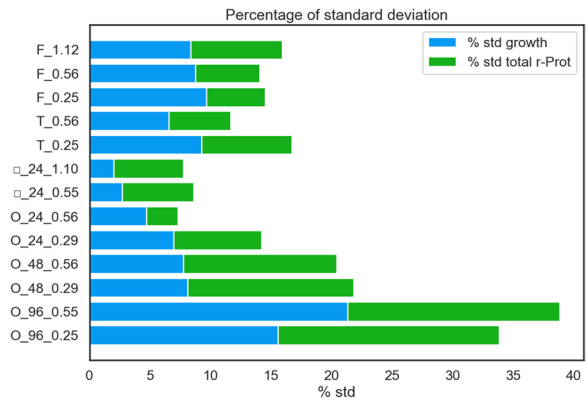


Fig. 3 Variability of data on growth and rProt amounts read in *Y. lipolytica* strain cultures run in different scales and vessels (encoded according to Table 1), at different kLa parameter settings (Mendeley data [11]). Variability is expressed as a share of standard deviation in the mean result obtained from a specific culture variant. Each variant of the flask and tube cultures was conducted in pentaplicate, cultures in 96-well 48-well, and 24-well were conducted in 48, 24, and 12 parallel runs

Mechanistic view on scaling-down and the main limitations for *Y. lipolytica* culturing

Multi-cell population culturing – MTPs

The phenomena taking place upon down-scaling of microbial bioprocesses to MTPs were thoroughly investigated and discussed in a series of works [17–20]. In this review, we focus on volumes that could be useful in high-throughput screens. In such a case, a cell of (average) several μm is cultured either in the liquid bulk of several mm (MTPs) or cm (flasks). As stated by [17] such a dimensional change in the culturing vessel has a negligible impact on the cell's physiology.

The principal limitations of (*Y. lipolytica*) micro-volume culturing in MTPs are insufficient aeration rates and small working volumes [18]. The former becomes specifically troublesome for microbes with high oxygen demands, like *Y. lipolytica*. Typically, limited aeration is overcome by increasing mixing frequency and amplitude, but on the other side—it leads to increased risk of cross-contaminations between the wells. Interestingly, these apparent limitations are also shaped by other 'indirect factors' affecting cellular physiology, and gaining importance upon down-scaling. Awareness of their occurrence is important to develop counteracting measures. These 'indirect factors' are (i) the ratio of the effective gas–liquid exchange area to the culture volume (considered in Eq. 1), (ii) the increased importance of the surface tension, which counteracts the flow and movement of the culture bulk in micro-volumes. The latter is specific to MTPs and has not been observed for larger scales [26]. Since the bulk of the medium in MTPs is relatively small, its inertness is small, and the movement induced by the centrifugal force (shaking amplitude) is overbalanced by adhesion (liquid molecule to the vessel) and cohesion (liquid molecules interaction) forces. Adhesion and cohesion forces remain constant irrespectively from the culture volume–they depend on the properties of the liquid and the vessel's material. So up to a specific shaking intensity, the liquid surface remains horizontal in O-shaped micro-wells of small bulk volumes (no turbulence occurs, and no expansion of the gas–liquid exchange area occurs). Turbulence may be introduced by using square-shaped micro-wells. Then, the corners

of the well act as baffles, greatly expanding the effective gas–liquid exchange area, and amplifying the oxygen transfer rate by a factor of 2 [17, 20]. As demonstrated, the surface area is a more important determinant of the oxygen transfer rates than the frequency of shaking [18].

Another significant problem faced in MTP culturing is that the oxygen diffusion rates to an individual well are typically affected by its position in the MTP, with the central wells being less aerated (Fig. 4; [11]). Our experience shows that the problem is valid for 24-, 48-, and 96-well MTPs equipped in a typical solid plastic cover or sealed with 'breath freely' film. On the other hand, such an effect was not observed for □_24-well MTPs with individual air exchange systems above each well. In these MTPs, at k_La 0.55 and 1.1–the difference between inner and outer (central and peripheral) wells was not significant at $p<0.001$. Such an uncontrollable well-to-well variation leads to misinterpretation of the results coming from a single MTP. Interestingly, our previous studies showed, that upon (anoxia) stress *Y. lipolytica* population splits into more and less metabolically active cells, and the latter subpopulation increased in counts with the time of exposure to stress [23, 24], so the heterogeneity came from cell-to-cell variation within a single well, rather than from global better or worse performance of a population at a specific location in the MTP.

Another recognized challenge of MTPs culturing is that insufficient aeration contributes to the differential profile of metabolites being produced in different wells. Consequently, depending on the strain's biology, difficulties in maintaining stable and comparable pH across the plate may occur; further amplifying the well-to-well variation in MTPs.

And finally, in terms of well-to-well variation illustrated in Fig. 4, the better aeration of the well located at the corners has also a negative side – it contributes to enhanced culture volume loss in those wells. Hence, especially when working with the low-volume 96-well format, peripheral wells are typically used as cell-less humidifiers filled with medium or water, and are lost from the analysis.

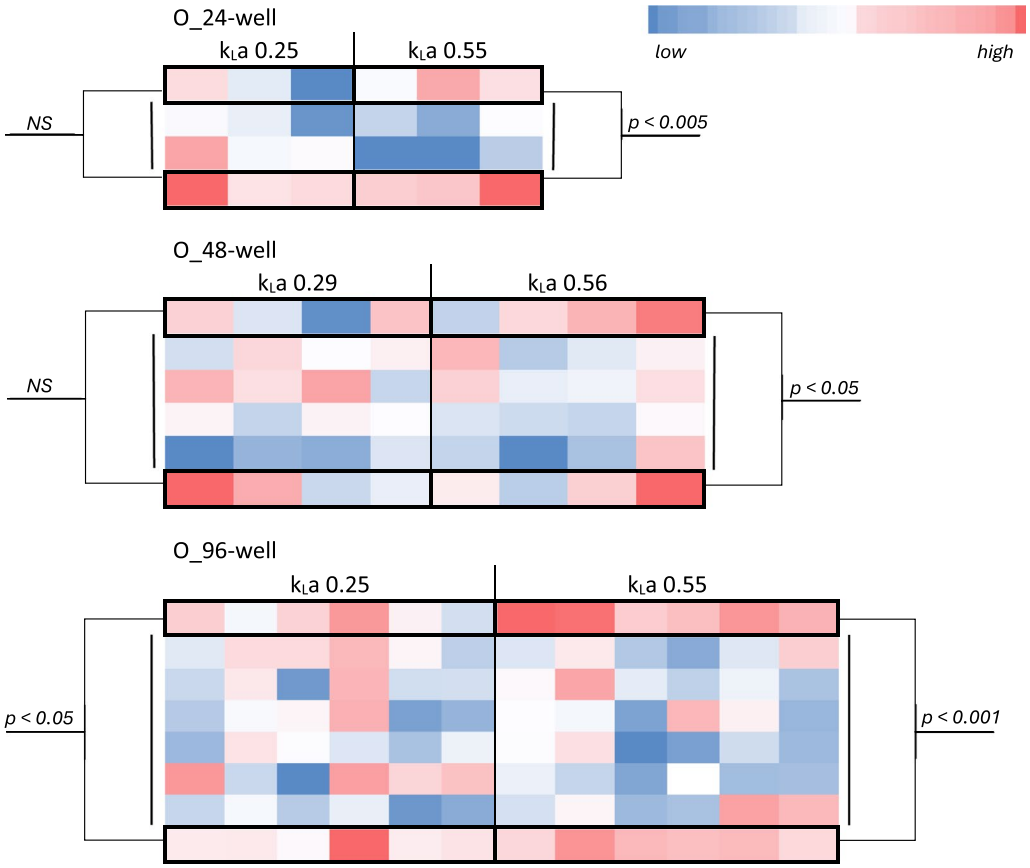


Fig. 4 Exemplary heat-maps showing growth of a *Y. lipolytica* strain in typical O-shaped 24-, 48-, and 96-well MTPs with solid cover. Color spaces correspond to well in the MTPs. Growth was measured as absorbance at OD600 (Mendeley data [11]). The values are color-coded according to a legend. The same strain was grown in all the wells, in the same culture medium. Comparisons of the color scale should be done only within a specific section covering identical culture variants. Inter-plates comparisons of the color are not adequate. Statistical significance was determined via ANOVA analysis (Mendeley data [11])

To overcome these limitations, the inventors of □_24/96-well plates proposed a cover system with holes above each well, but tightly sealing rants with a soft-silicone mat [17, 18]. Such a solution secures individual air exchange from each location on the plate, and limits the impact of the neighboring wells and cross-contaminations. The wells in the central part of a plate are equally well aerated as those in the peripheries. Indeed, in our 'proof-of-concept' experiment, the growth of *Y. lipolytica* in □_24-well plates was not significantly different between central and peripheral wells (p-value < 0.001).

In addition, the implementation of this individual aeration system unified the propensity for volume loss across the plate. Still, thanks to a multi-layer covering system and the selection of the material, the volume losses are negligible (10 µL per day), and cross-contaminations are also eradicated ([18] and own experiments discussed hereafter).

Single-cell-per-vessel approaches

A further step in scaling-down microbial cultures is miniaturization to pico-liter volume droplets, each initially loaded with a single cell. Such an approach differs from what is typically called 'single-cell analysis' by Fluorescence-Assisted Cell Sorting (FACS), which is an analytical method that does not require the droplet-forming step. The pico-liter cultivation followed by the droplet analysis and sorting is called Fluorescence-Assisted Droplet Sorting (FADS) [2].

An individual microbial cell can be entrapped in a 'single-cell vessel' either by using agarose beads or by creating an emulsion. A microbial cell must always be suspended in the 'water' (w) phase. Then, it can be surrounded by a spherical layer of immiscible fluid ('oil' phase (o)) suspended in another w-phase (w/o/w double emulsion), or suspended in the oil phase (w/o emulsion) [61]. W/o/w system is compatible with typical FACS fluidics and optics, so the analytical part can be conducted in a commercial FACS sorter. Analytics of droplets generated as w/o emulsion is typically run in laboratory-constructed microchips with electrodes embedded into the chip itself, due to incompatibility of the oily mobile phase with commercial FACS equipment. As pointed out earlier [57], using the droplet-based system reconstitutes a direct link between phenotype and genotype, which is lost in the case of secreted molecules and FACS-based sorting. In addition, the Authors demonstrated that using an intracellular concentration of a molecule (a measure compatible with FACS as a proxy of the total production rate of a secreted molecule is not adequate, making FADS specifically relevant for such an application.

The droplet generation and analytics have an incomparably higher capacity than any other high-throughput screening technique. With *Y. lipolytica*, the current technologies reach the droplets producing rate of ~ 10^6/h, and sorting rate of over 10^5 to 10^6/h [5, 8, 35]. But the question is whether encapsulation in a droplet is compatible with the biology of the strictly aerobic, filamenting species?

The principal limitation of the actual 'culturing' of a living microbial cell in the pico-reactors is limited volume and mass transfer, which directly implies low provision of oxygen and nutrients, and, increasing with time, accumulation of metabolites. Furthermore, considering that FADS ensures a super-high resolution (at a single-cell level and not the average of a population), the incubation time of all the cells must be sufficient and comparable [14], which requires additional consideration in the continuous flow fluidics system. The other commonly pointed challenges are the investment costs, the need to couple the desired phenotype with some fluorescent product, the risk of the partitioning of hydrophobic products into the mobile phase (or the oil layer in w/o/w emulsions), and the low stability of the double w/o/w emulsions, which are compatible with FACS [6]. Moreover, the frequency of the droplet seeding at 1 cell/droplet is limited by the Poisson distribution [13], resulting in a high number of empty microdroplets. Some of these challenges were ameliorated by using fluorinated oil with good oxygen solubility [35] and fluorescent substrates of the analyzed enzymes [5], designing incubation chamber ensuring first-in-first-out dynamics of the droplets flow [14], pico-injection of the substrate to initially preselected droplets containing the desired 1 cell/ drop seeding rate [5]. Still, additional amendments were required to adapt the droplet-based microfluidic screening system to the use of *Y. lipolytica* [5] (discussed hereafter).

Another interesting ultra-high-throughput approach for yeast culturing was developed by [52, 53]. In that system, a multiplexed microfluidic chip was operated in a perfusion mode, so the cells were maintained in microliter volumes of a continuous stream of medium. In this system, the key limitations of closed pico-liter bioreactors (oxygen and nutrients availability, accumulation of metabolites) are no longer valid. In such a format, the screening system is scaled-down by 2000-fold, and the desired phenotypes can be identified in 2 to fourfold shorter times [52, 53]. Operation in a perfusion mode enabled precise control of the culturing conditions, making room for the process conditions optimization with this system. Notably, the results obtained with *K. phaffi* grown in such a microfluidic perfusion system were validated at larger scales, proving their effectiveness.

A similar system exploiting a 1 mL microfluidic bioreactor, operating in perfusion mode was also used with *Y. lipolytica* by [40]. More details on that system

and 'personalized' adjustments required for *Y. lipolytica* are discussed hereafter ("*Y. lipolytica* in microfluidics" section).

Yarrowia lipolytica 'personalized' scaling-down approaches

MTP format

Considering the high interest in *Y. lipolytica* as a platform species in metabolic engineering and rProt production, urgent needs are placed on developing dedicated testing and screening platforms. However, as mentioned above, several characteristics of this yeast make the task very challenging. The principal limitations encountered in the high-throughput cultivations of *Y. lipolytica* and the state-of-the-art developments in this area are summarized below and in Table 2.

In a series of works [1, 5, 35, 38, 39] a French group transformed *Y. lipolytica* high-throughput screens from typical 96-well MTPs to a versatile and ultra-high-throughput FADS system.

In their initial work, [38] used the typical O-96-well MTPs filled in two-thirds (200 per 300 µL in total) of culture medium buffered with 50 mM phosphate buffer at pH 6.8. To address the problem of evaporation (illustrated in Fig. 4) the peripheral wells of the plate were filled with water. Cultures were conducted in a plate reader Biotek Synergy MX (Biotek Instruments) for 48 h at 28 °C with continuous agitation. In that work, the MTP culturing was used to manage a large library of strains developed from high-throughput transformation protocol, and not much focus was placed on its optimization, and no assessment of the well-to-well variation was provided. A similar high-throughput cultivation protocol was adopted in the following screens, using the same plate format and the reader [39, 54]. [39] noted that the cultures could be continued only up to 24 h, as later the OD600 readouts were > 1.2, above which the correlation between OD and cell density is not linear anymore. We found that point relevant in our experiments with *Y. lipolytica*. In addition, the necessity of shortening the culturing time in the automated plate reader used by the Authors was appropriate due to the limited buffering capacity of 50 mM phosphate buffer, which was probably insufficient for longer cultivations [22]. In the following work, [1, 55] used a specialized MTP shaker-incubator-reader, BioLector (mp2-labs, Baesweiler, Germany), and unique MTPs—48-well Flower-Plates (mp2-labs). The topology of these plates is highly compatible with the strictly aerobic *Y. lipolytica*. Each inner edge of the 'six-petal flower' acts as a baffle, enhancing the oxygen transfer rate substantially (up to > 0.11 mol/L/h). *Y. lipolytica* strains were cultured in either rich or minimal medium buffered with phosphate buffer at pH 6.8, under 28 °C and constant agitation (1200 rpm, shaking diameter = 3 mm, orbital) in a working volume of 800 µL (possible filling volume up to 2400 µL). In the BioLector system, growth (and fluorescence) can be continuously monitored by measuring scattered light intensities, which offers a huge advantage over the formerly used OD600, especially with high-cell-density-growing *Y. lipolytica*. As evidenced in that work, scattered light measurements in *Y. lipolytica* cultures displayed a linear correlation with those of OD600nm up to 115 units. The assessment of dry cellular weight concentration using OD600 as a proxy could reach maximally 58 g/L, while using scattered light—as much as 100 g/L. In [1], the cultures were continued for > 100 h, and pH was stabilized at pH 6.8 with 50 mM phosphate buffer. Our previous analyses [22] (Fig. 5A) showed that even twice stronger phosphate buffer (100 mM) does not possess sufficient buffering capacity to stably maintain pH at ~ 7.0 for 48 h in *Y. lipolytica* cultures (drop to pH 6.0 after 48 h; noteworthy—a different culturing medium was used). Currently, the BioLector system can be configured with an optional microfluidic module, which may be used to regulate pH (2 feeding lines are available). Such functionality is highly relevant for *Y. lipolytica* cultures, which tend to strongly acidify the medium upon growth.

High throughput screening in 48-well plates of an extensive library of *Y. lipolytica* strain was conducted by [60]. The screening was performed in MicroScreen system (Gering Ltd., China) in 1 mL of mineral medium (per 3 mL total volume), with an initial pH of 6.0 (no buffer provided) for up to 36 h. According to the system's provider, the device can ensure shaking up to 1000 rpm (for *Y. lipolytica* the Authors used 800 rpm), but foremost, the plates are equipped with the independent vented cap above each well, which altogether contributes to higher k_La levels. The important drawback of that system is that in growth tracking it uses OD600 measurements with the upper limit of 5.0. Still, it allows to calculate the initial growth rate, which is sufficient to assess the strain's robustness.

Knowing the limitations of *Y. lipolytica* culturing and its high oxygen demands, [48] tested different covering systems in combination with 96-deep-well plates for *Y. lipolytica* screens. The experiment aimed at screening over 1000 *Y. lipolytica* clones in a high-glucose medium (900 µL working volume in 2 mL total well volume) with shaking at 220 rpm for 10 days. Inoculated deep-well plates were either sealed with 'breathing freely film' or covered with paper and stainless steel covers. When stainless steel covers were used, the growth of *Y.*

Table 2 Summary of the main limitations encountered in the high-throughput cultivations formats (O_96-well MTP and the pico-liter volume droplet) and solutions proposed in different studies

Limitations (including those specific to *Y. lipolytica*)	Proposed solutions with references
O_96-well MTP	
Insufficient aeration due to the use of plastic cover and 'breathing freely film'	Stainless steel covers [48], sandwich covers [18]
Small working volumes for sampling	Multiplication at the cost of high-throughputness
Low buffering capacity, insufficient for longer cultivations	BioLector configured with an optional microfluidic module to be used to regulate pH (https://www.m2p-labs.com/bioreactors) Use of an optimized maleate-based buffering system [22]
Volume loss at the corners	Use of peripheral wells as cell-less humidifiers filled with medium or water at the cost of high-throughputness
Measurement limitations due to intensive growth (OD600 > 1.2), loss of linearity between the absorbance readout and cell density	Reading scattered light intensities rather than absorbance [1]
High surface tension/lack of turbulence	Flower-shaped plates from BioLector (https://www.m2p-labs.com/bioreactors) [1, 55] Square plates [17–20]
Position-on-the-plate-dependent oxygen diffusion rates to an individual well → differential profile of metabolites (difficulties in maintaining stable and comparable pH across the plate)	Use of individual air exchange systems above each well: holes above each well, but tightly sealing rants with a soft-silicone mat [17, 18] the independent vented cap above each well [60]
Pico-liter volume droplets	
Low oxygen availability (low nutrient availability and accumulation of metabolites were shown not to be limiting)	Use of fluorinated oil with good oxygen solubility [5, 35, 57] Increase the volume of oxygen-permeable fluorinated oil [35] Use of perfusion-mode microreactor (Air supply by gas diffusion through a sterile filter, online pH control) [40]
Poisson distribution of the droplet seeding—a high number of empty microdroplets	Adjustment of cell suspension density to match the desired seeding rate [5, 35] Removal of cellular aggregates prior to encapsulation by filtering through a 5 µm filter [35] Co-injection of the bioassay components and cells [35] Sub-sorting of the charged droplets from empty ones based on their size (use of a phenomenon of cell-containing droplet shrinkage [35]
Non-equal time of cell incubation prior to reading	For w/o droplet-type: designing incubation chamber ensuring first-in-first-out dynamics of the droplets flow and its incorporation into the microchip design [5, 14] For w/o/w droplet-time: additional incubation step [57]
Risk of the partitioning of hydrophobic products into the mobile phase (or the oil layer in w/o/w emulsions)	Use of synthetic reporters/alternative enzymatic substrates (fluorescent substrates) and its pico-injection to initially pre-selected droplets [5, 35]
Droplet integrity lost due to invasive growth	Use of a *Y. lipolytica* fil- strain (deleted for the Mhy1 gene, rendering the strain with a non-filamenting phenotype) (Hurtado and Rachubinski, 1999; [35])
Lipolytic activity-driven degradation of the oil phase	Use of fluorinated oil that is resistant to extracellular lipases due to fluorination of the aliphatic chains oil [5, 57]
Low stability of the emulsions	Addition of fluorosurfactant [5]
Specific to w/o emulsion (requires laboratory-constructed microchips)	
Investment costs for laboratory-constructed microchips with electrodes embedded into the chip itself	NA
Aggregation and adhesion of *Y. lipolytica* to hydrophobic surfaces	Proposed by [5]: Supplementation of the cell suspension with non-ionic detergent to limit the cells' adhesion, Increase shear stress by adopting tubing with a smaller inner diameter (100 µm) to limit aggregation before encapsulation, Implementation of continuous stirring of the cell suspension during the encapsulation to prevent cell sedimentation, Adjustment of the initial cell density to approach the desired seeding rate of 0.03–0.1 cells/droplet [35] further decreased the inner diameter of the tubing to 50 µm

lipolytica strains was 10- to 20-fold higher, when compared to the cultures sealed with the film, and by only 60% worse than when grown in shake flasks (using the same carbon source load). Control and regulation of pH were not addressed in that research.

Y. lipolytica in microfluidics

The very first implementation of microfluidics technology to *Y. lipolytica* high-throughput screens was reported by [5]. To screen for strains secreting hydrolases, the Authors developed two microfluidics devices: (i) a drop-maker and (ii) an integrated screening device. The former encapsulated *Y. lipolytica* cells into 20 picoliter droplets at nearly 10^7 /h rate, creating a w/o-type emulsion in fluorinated oil. The mobile phase was enriched in fluorosurfactant to stabilize the emulsion. Noteworthy, the selected oil phase was: (i) resistant to extracellular lipases due to fluorination of the aliphatic chains oil [12, 56], and (ii) permeable to oxygen–both characteristics of key importance considering *Y. lipolytica*'s biology. The Authors reliably assessed the limitations of performing *Y. lipolytica* screens in the microfluidics system, and introduced the necessary ameliorations. For example, it was noted that the yeast cells display affinity to hydrophobic substrates, and a tendency to aggregate on the channel walls. To overcome these problems, the Authors implemented several specific amendments: (i) supplemented the cell suspension with non-ionic detergent to limit the cells' adhesion, (ii) increased shear stress by adopting tubing with a smaller inner diameter (100 µm) to limit aggregation before encapsulation. In addition, the system's operation was improved by (iii) the implementation of continuous stirring of the cell suspension during the encapsulation to prevent cell sedimentation, and (iv) adjustment of the initial cell density to approach the desired seeding rate of 0.03–0.1 cells/droplet [5]. Following the droplet formation, the seeded pico-bioreactors were set in operation for 16 h at 28 °C. Over that period, the cells divided and produced the targeted enzymes. The demonstration that *Y. lipolytica* cells could actually replicate inside the droplets is of key importance, proving that the system can be truly regarded as 'cultivation in pico-bioreactors' and not just 'coating and incubation'. It was the first reliable system accompanied by a detailed protocol for managing the pico-cultivation of *Y. lipolytica* for high-throughput screens.

[57] developed a double emulsion w/o/w FADS system to screen for *Y. lipolytica* strains producing and secreting high amounts of riboflavin, and compared such a screening technology with the typical single-cell FACS sorting. In both cases, a two-step cascade of sequential FA(C/D)S enrichment was used to isolate the final 'hyper-producer' populations. The aim was to evaluate if successive enrichment and sorting using either of the two methods would impact the outcome–performance of the best producer. Technically, the double emulsion was obtained by recurring flow-focusing-driven encapsulation: (i) cell suspension in fluorinated oil with amphiphilic block copolymer surfactant, followed by the droplets collection and incubation for 3 days to allow for riboflavin production inside the pico-bioreactors; (ii) entrapment of the w/o emulsion from step (i) in the outer phase water solution. To enrich the initial population in the best producers, two subsequent FACS sorts with increasing stringency were applied. Intermittent incubations (3 days, 30 °C) between the sorts allowed to spontaneously phase separate. It was not shown if *Y. lipolytica* propagated inside the droplets over the 3 days between w/o and w/o/w emulsion generation. The period of 3 days is fourfold higher than the 18 h adopted by [5]. However, the cells could produce the target molecule, when incubated in the w/o emulsion, and be screened for its concentration, meaning that the metabolism was active. The developed FADS technology proved to be significantly more efficient in selecting *Y. lipolytica* producing higher amounts of riboflavin in total, considering the intra- and extracellular levels.

Most recent research implementing microfluidics to *Y. lipolytica* aimed at screening large libraries of strains secreting antibodies in their final, water-soluble format [35]. The Authors developed and optimized a protocol for growing and producing full-length antibodies by *Y. lipolytica* strains in pico-bioreactors, and applied a microfluidics approach to sort and recover target-specific antibody-secreting yeasts. On their way, several interesting aspects were addressed. Foremost, the host strain was deleted for the *Mhy1* gene, rendering the strain with a non-filamenting phenotype, which is highly relevant for any microfluidics approach and important for

(See figure on next page.)

Fig. 5 Optimization of a protocol for high-throughput *Y. lipolytica* cultivation in square-24-well-MTPs. Selected data from [22] are graphically presented here: (**A**) buffering capacity of selected buffers and growth of Y. lipolytica cells in the specifically-buffered 3xYNB medium, (**B**) growth of Y. lipolytica in O- and □-shaped 24-well MTPs, and under different 'sandwich covers' of specified air-exchange volume; (**C**) growth of Y. lipolytica in □-shaped 24-well MTPs under different 'sandwich covers' and at different shaking frequencies; (**D**) optimization of carbon load (C20, C30, C40, and C50) at fixed nitrogen load of 15 g/L (organic—casein hydrolysate; inorganic—ammonium sulfate) based on results on growth, fluorescence, and concentration of the main metabolites (CA, ERY, MAN)

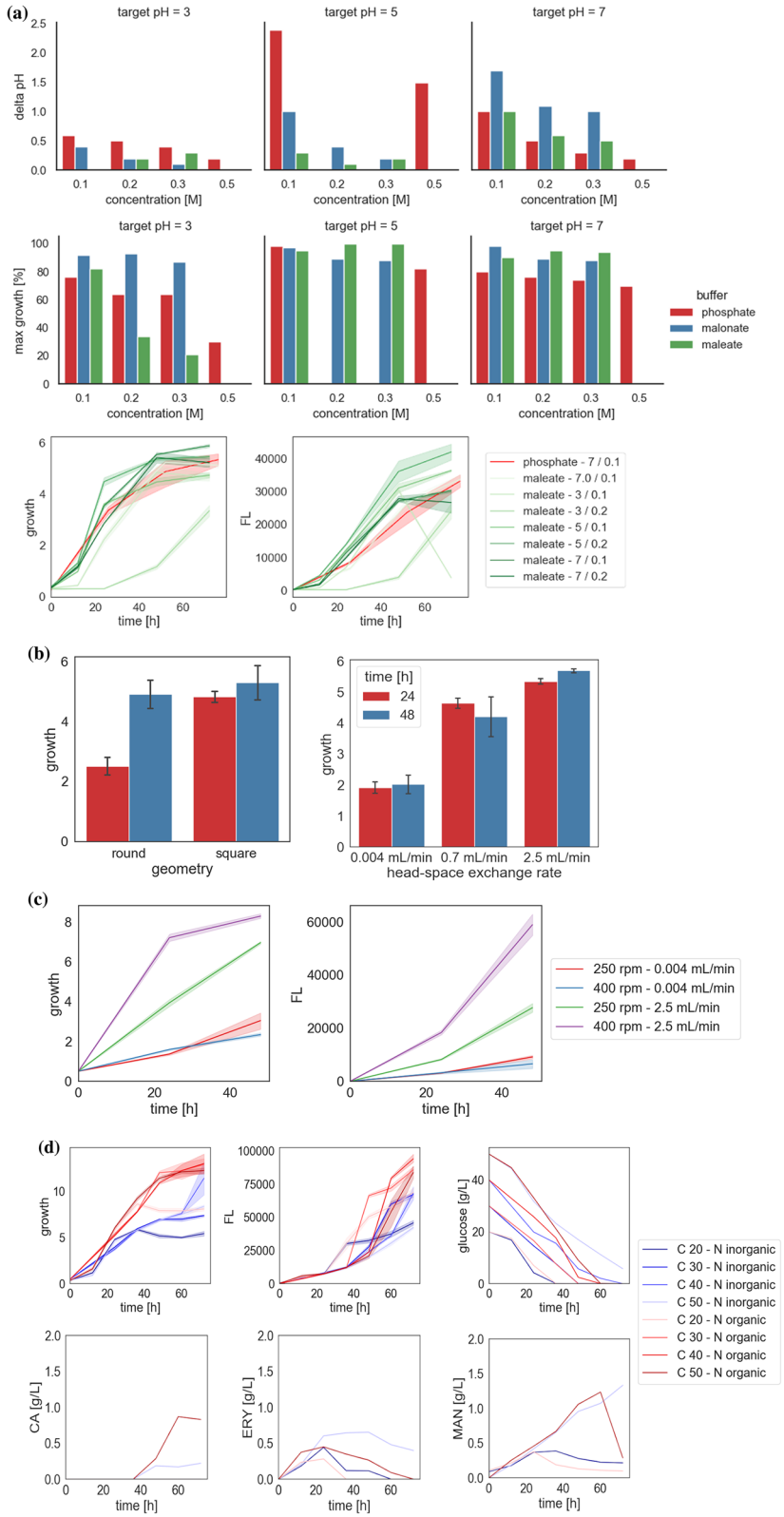


Fig. 5 (See legend on previous page.)

droplet integrity. The chip channel diameter was further decreased to 50 µm (when compared to the solution proposed by [5]), to prevent *Y. lipolytica* adhesion to the walls. The 30 picoliter volume droplets were formed in w/o emulsion, using the hydrofluoroether oil phase. The cellular aggregates were removed prior to encapsulation by filtering through a 5 µm filter. Yeast cell suspension density was adjusted to match the desired seeding rate. Charging the drops with the yeast cells and the bioassay components was conducted by co-injection, followed by overnight growth of the cells and production of the antibodies. As concluded by the Authors, growth and product synthesis in pico-bioreactors were not limited by nutrient availability and metabolite accumulation. The only limiting factor was oxygen availability. To cope with this issue, the volume of oxygen-permeable fluorinated oil was increased, which, as stated, directly increased the oxygen provision. At the sorting stage, the Authors took advantage of the commonly occurring phenomenon of cell-containing droplet shrinkage. In this way, the seeded droplets were sorted from empty ones based on size, without the need for an extra cell tracer. Following the microchip sorting, the droplets containing the target cells were deflected by dielectrophoresis to recover the secreting yeasts. Due to high survival over the sorting process and rapid out-growth after the procedure, *Y. lipolytica* strains could be reused in further screenings—for enrichment studies or selection according to different criteria.

As briefly mentioned above, [40] reported on the use of a 1 mL microfluidic perfusion bioreactor to study fundamental aspects of *Y. lipolytica* biology. Technically, the cultures were conducted in a commercial single-use 1 mL microbioreactor chip (Pharyx Inc., USA). The microreactor was equipped with optical density, oxygen, pH, and temperature probes to ensure precise control of the process parameters. Air was supplied by gas diffusion through a sterile filter, pH was controlled by the automated addition of strongly diluted $NaHCO_3$ (1 mM), while a temperature of 28 °C was by regulated a heating element located at the bottom of the device. The growth chamber comprised three interconnected 500 µL sections, out of which two were occupied to secure both the 1 mL working volume and the mixing. More details on the system can be found in [7, 36, 37]. After several volume exchanges, *Y. lipolytica* culture reached a steady-state highlighted by a stable biomass concentration (~ 5 g of dry cellular weight/L), stable oxygen saturation (above 40%), and, most importantly for that experiment, a stable residual glucose concentration < 1 mg/L. The latter was of key importance, as the Authors investigated the relationship between residual glucose levels and dimorphism in *Y. lipolytica*. Notably, *Y. lipolytica* displayed high viability in such a system (above 97%). By definition, exploring dimorphism in a microfluidics system is a challenging task, due to the high risk of clogging the tubings by filaments. Being aware of this, the Authors first simulated conditions promoting ovoid morphotype (then, *Y. lipolytica* cell is ~ 4.4 µm in diameter). Further 'personalization' of the experiment consisted of adjusting the time of the 'filamentation-inducing state', disallowing full development of the long filaments. Using such a system it was possible to prove that the dimorphic shift is dependent on the residual glucose levels. Very precise determination of the threshold level (0.35–0.37 mg/L) was possible due to the incomparably high resolution of the perfusion microbioreactor, and the excellent technical execution of the experiment. In that experiment, the microfluidics system was not used for different phenotype comparisons, so it was also not assessed for its screening capacity, run-to-run variability, and throughputness. However, the observations done with the microfluidics were validated in a standard 1 L bioreactor, which confirmed the operability of the system and its usefulness for the dimorphic *Y. lipolytica*.

Protocol for high-throughput screens of *Y. lipolytica* strains in square-MTPs

Our recent project required the development of a high-throughput screening protocol to reliably compare phenotypes of *Y. lipolytica* strains, individually overexpressing a single transcription factor per strain [22, 25, 39]. Knowing the key role of transcription factors in modulating cellular biology, we expected that the spectrum of obtained phenotypes (for 125 strains tested) would be very wide. Hence, developing a single protocol, that matches the requirements of all the strains would be a significant challenge. The first trials conducted in the standard O_96-well plates turned out to be unsuccessful due to low repeatability. Even if the strains grew and the reporter rProt was detected, the repeatability between successive runs was low and implementation of different environmental variables did not allow to determine any statistically significant factors affecting the observed outcomes. Considering the number of strains and conditions to be tested (rendering altogether 26,000 cultures), and limited budget (disallowing purchase of BioLector and consumables), we set for the development of a culturing protocol that would enable full development of a variables-driven phenotype, without uncontrolled infliction of unknown environmental factors [22].

From an array of different MTPs available on the market, reusable □_24-well MTPs (www.enzyscreen.com) were chosen. Reusability, and hence limitation of plastic waste, was a huge asset of the system. The Duetz system is well-developed and supported by reliable studies [17–20]. The usefulness of that system was proved for

culturing different bacteria species [18–20], filamentous fungi [32], the model yeast species *S. cerevisiae* [28, 31], and recently also–*Y. lipolytica* [27, 30, 47, 49]. In that project, a system enabling a controlled variable-driven phenotype development was of key importance. Therefore, several key parameters, like pH maintenance, oxygen availability, and sufficient nutrient provision throughout the culturing time were carefully investigated.

Since that research relied on batch cultivations, and the number of cultures run in parallel disallowed manual correction of the cultures' acidity in response to the cells' metabolism, it was necessary to establish a robust buffering system [22]. That experimental plan covered a (problematic) pH range spanning pH 3–5–7. The aim was to have a buffer that enables buffering at the three levels using the same chemical compound,with the final change in the set pH value no higher than 0.5 pH unit. Considering the metabolic characteristics of *Y. lipolytica*, buffers based on citric acid and acetic acid were immediately eliminated, as the components would be consumed [3]. Hence, the investigated set covered: phosphate, malonate, maleate, MES, and carbonate buffers at concentrations from 0.1 to 0.5 M (Fig. 5A; based on data presented in [22]). Phosphate buffer was a good buffering system at 0.3 M concentration but only close to its pKa—pH 3.0 and 7.0 (concentration of the buffer's acid and conjugate base forms are equal). Even at a concentration of 0.5 M, it had no buffering capacity close to pH 5.0. In addition, 0.5 M phosphate buffer at pH 3 strongly limited the growth of *Y. lipolytica* (30% of the maximum read for all the conditions considered; Fig. 5A). Carbonate and MES were considered complementary buffers to the phosphate buffer at pH 5.0. However, at none of the concentrations tested (0.1 and 0.5 M), they secured sufficient buffering capacity for *Y. lipolytica* metabolism–the drop in pH was by more than 1.5 pH unit (from pH 5.0 to pH 3.5 to pH 2.3). The final two tested buffers were based on dicarboxylic acids having two pKa values. The pKa of maleic acid is 1.9 and 6.2, while of malonic acid—2.83 and 5.69 (PubChem data). Since these compounds were not previously used for buffering *Y. lipolytica* cultures, possible consumption of the two chemical compounds by *Y. lipolytica* was tested. The change in their concentration at the end of culturing (HPLC; data not showed) was within a technical error, indicating that *Y. lipolytica* does not utilize the compounds and they can be used as buffering agents.

Considering the target pH levels (3, 5, 7), and that the buffering capacity of a buffer is highest when the pH is within one unit of the pKa value, the maleate buffer would be more adequate to buffer at pH 7, while malonate–at pH 3 and 5. Indeed, malonate at a concentration of > 0.2 M buffered *Y. lipolytica* cultures very well at acidic pH, but turned out to lack the capacity at neutral pH. On the other hand, 0.2 M maleic buffer was very efficient in buffering pH at 5 and 7. At pH 3 its buffering capacity was also sufficient, but it limited *Y. lipolytica* growth when its concentration was > 0.2 M (to 20–30% of the maximum; Fig. 5A). Hence, to meet the requirement of having a single chemical compound in all the buffers, and considering that the buffering capacity of 0.1 M maleate was sufficient at pH 3, an optimal buffering system for *Y. lipolytica* cultures comprised of maleate buffer at 0.1 M for pH 3.0, and at 0.2 M for pH 5.0 and 7.0. The potential toxicity of the maleic buffer was investigated in comparison to phosphate buffer (both pH 7.0, 0.1 M); demonstrating a lack of the limiting effect (Fig. 5A). To ensure that no uncontrolled variable is introduced by using the buffer at different concentrations, osmolality in the medium formulation was experimentally determined. The 0.1 M vs 0.2 M maleate buffer yielded osmolality of 1.27- to 1.78-fold higher than the standard YPG medium [22], which is away from the stress level (nearly fivefold higher [34]). Growth and rProt synthesis by *Y. lipolytica* in media buffered with the developed buffering system was studied, proving its functionality (Fig. 5A). Using such a buffering system, it was established that *Y. lipolytica* grows and produces rProt most efficiently at pH 5.0. Initial growth was significantly slower at pH 3, but finally reached only slightly lower yields at the stationary phase (8% lower biomass than at pH 5). On the other hand, rProt synthesis at the acidic pH was significantly diminished (by nearly 40% vs pH 5). This observation well aligned with previous results conducted in bioreactors with a different reporter rProt, showing that a pH of 3 is a limiting factor of rProt synthesis, especially when combined with some other stress factor [23]. No such differences were observed between growth and rProt synthesis by *Y. lipolytica* in pH 5 and 7.

Another parameter investigated and optimized in that research was oxygen level supply [22]. The aim was to orchestrate the type of consumables, equipment, shaking frequency, and culture volume to generate two distinctly different conditions – 'high' and 'low' oxygen supply. It was noted that when *Y. lipolytica* is cultured in □_24-well vs O_24-well MTPs, according to the manufacturer's suggestions (2.5 mL in the former, 1 mL in the latter), biomass in the stationary phase is not much different (by 7% lower in the O-shaped MTPs; please compare also Fig. 1 and Table 1), but the stationary phase is reached by nearly 24 h faster in the square-shaped MTPs (Fig. 5B–based on data presented in [22]). Considering the scope of that investigation (a strain library screening under multiple conditions), such a shift in time was of high interest. To simulate conditions of clearly divergent oxygen supply, covers with different nominal head-space

exchange rates were tested ('sandwich covers' of 0.004, 0.7, and 2.5 mL/min). Technically, the differences were implemented by differing the type of the silicone mat and the diameter of the openings above each well. According to [18] an average aerobic microbial culture needs an air supply of one working volume per minute (1 vvm; in our case—2.5 mL). If supplied, the oxygen concentration in the headspace is maintained at > 18% (v/v), even at very high oxygen consumption rates (40 mmol oxygen/L/minute). Depending on the cover type, the growth of *Y. lipolytica* nearly linearly decreased along with a decrease in the nominal head-space exchange rates (r = 0.84 across all the pH values; r = 0.91 for pH 5); at 0.004 mL/min reaching only 36% of its value achieved at 2.5 mL/min (Fig. 5B). A comparable decrease (to 33%) was read for the rProt synthesis level expressed in biomass-normalized values, indicating a direct effect of oxygen provision on rProt synthesis. This experiment showed that the 'sandwich cover' system solution offered by (www.enzyscreen.com) works efficiently in simulating different aeration rates in *Y. lipolytica* cultures. Moreover, the difference in oxygen supply driven by the use of 0.07 and 2.5 mL/min sandwich covers was not sufficient, to use them to simulate the distinctly different aeration conditions; hence, the middle variant of 0.07 mL/min was eliminated from further studies.

Selection of the 'sandwich cover' type was done at the same shaking parameters (shaking amplitude of 1.91 cm, and frequency of 250 rpm). To investigate if changes in shaking frequencies would further modulate *Y. lipolytica* growth under the selected 'sandwich covers', shaking frequencies up to 500 rpm were tested (amplitude remained constant) (Fig. 5C). Such an experiment would also answer the question of whether the use of 2.5 mL/min sandwich covers suffice the oxygen demand by *Y. lipolytica* at the used shaking frequency. In that experiment, the risk of cross-well contamination was monitored by a checkered inoculation pattern. As observed, shaking at 500 rpm leads to splashing of the culture on the cover imposing a risk of cross-contamination, so this variant was eliminated. Shaking at a frequency of 400 rpm led to a significantly higher rate of growth and rProt synthesis vs shaking at 250 rpm, but only when 2.5 mL/min sandwich covers were used (no such an effect was observed under the covers 0.004 mL/min). The observed increase in those parameters suggests that the air supply of 1 vvm at 250 rpm is not sufficient for *Y. lipolytica*'s robust metabolism. Such a result is concurrent with previous observations done with *Y. lipolytica* bioreactor cultures, where at least 2 vvm of air had to be supplied to meet the desired 20% oxygen concentration (e.g. [33]). Notably, when shaking at 400 rpm, the level of growth and rProt amount in 24 h of culture was the same, as in 48 h of culturing when shaking at 250 rpm. It is a substantial increase in pace, greatly increasing the throughputness of the system. No effect of shaking frequency on the amounts of biomass and rProts under air-exchange-limiting covers use, indicates the efficiency of the system in securing a controlled head-space exchange rate.

The supply of carbon and nitrogen sources was also subjected to adjustment, but in combination. The aim was to secure sufficient amounts of carbon and nitrogen until the end of culturing, without the risk of introducing a starvation period (which would be an additional, uncontrolled variable affecting the phenotypes). That experiment was conducted with the fastest-growing strain from the library [39] to determine the minimum sufficient amounts of carbon (20 to 50 g/L) and nitrogen (5 to 15 g/L, organic and inorganic, Fig. 5D). Under the adopted culturing conditions, nitrogen load at 5 g/L was insufficient, leading to limited biomass growth and carbon source consumption (by 34% when inorganic nitrogen was used; partly compensated when organic nitrogen was supplied; as shown previously [10]). Using nitrogen at 15 g/L and carbon at > 30 g/L, eliminated the risk of a starvation period occurrence, as by the end of the culture, still some minimal level of the nutrient was present. Profile of CA, ERY, and MAN concentration support that thesis, as within the investigated time (up to 48 h), these compounds were not reconsumed in those culture variants (Fig. 5D).

Further notions from the high-throughput cultivation protocol optimization, more related to the research question than to the technical aspects of the cultivation, are presented in the accompanying article [22] and the open, freely accessible database YaliFunTome (https://sparrow.up.poznan.pl/tsdatabase/) presenting the results obtained with the protocol.

Summary and conclusions

Literature review and our own experience show that high-throughput cultivation protocols require extra consideration when being adjusted to *Y. lipolytica* characteristics. The vast knowledge accumulated on its biology is of great assistance in rationally designing the amendments. Technological progress offers many solutions that adequately respond to challenges from the yeast species biology. The authors of the original research covered by this review provided evidence for the repeatability, high capacity, and operability of their systems for *Y. lipolytica* cultivations. Using such dedicated and validated protocols may greatly aid the cumbersome screening stage of a synthetic biology-based clone library or run parallelized cultures for extensive experimental design.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12934-024-02465-3.

Additional file 1.

Author contributions

MG conducted the 'proof-of-concept' experiment ("How small can we go with *Yarrowia*? And what are the costs?"section), analyzed the data, and prepared the graphics. MG participated in writing and prepared all the graphical elements. EC conceived and designed the work, collected the literature, and wrote the manuscript.

Funding

This research was funded by the National Science Centre, Poland, grant number OPUS 2021/41/B/NZ9/00086.

Availability of data and materials

The datasets generated for the 'proof-of-concept' experiment are available from Mendeley Data [11]. YaliFunTome database, accompanying "Protocol for high-throughput screens of *Y. lipolytica* strains in square-MTPs" section. is publicly available free of charge at: https://sparrow.up.poznan.pl/tsdatabase/

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The Authors declare no conflict of interest.

Received: 15 April 2024 Accepted: 18 June 2024
Published online: 24 June 2024

References

1. Back A, Rossignol T, Krier F, Nicaud JM, Dhulster P. High-throughput fermentation screening for the yeast *Yarrowia lipolytica* with real-time monitoring of biomass and lipid production. Microb Cell Fact. 2016. https://doi.org/10.1186/s12934-016-0546-z.
2. Baret J-C, Miller OJ, Taly V, Ryckelynck M, El-Harrak A, Frenz L, Rick C, Samuels ML, Hutchison JB, Agresti JJ, Link DR, Weitz DA, Griffiths AD. Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity. Lab Chip. 2009;9:1850. https://doi.org/10.1039/b902504a.
3. Barth G, Gaillardin C. *Yarrowia lipolytica*, In: Wolf K. Nonconventional yeasts in biotechnology a handbook. Springer Berlin Heidelberg. 1996, pp. 313–388. https://doi.org/10.1007/978-3-642-79856-6
4. Baryshnikova A, Costanzo M, Kim Y, Ding H, Koh J, Toufighi K, Youn JY, Ou J, San Luis BJ, Bandyopadhyay S, Hibbs M, Hess D, Gingras AC, Bader GD, Troyanskaya OG, Brown GW, Andrews B, Boone C, Myers CL. Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. Nat Methods. 2010;7:1017–24. https://doi.org/10.1038/nmeth.1534.
5. Beneyton T, Thomas S, Griffiths AD, Nicaud JM, Drevelle A, Rossignol T. Droplet-based microfluidic high-throughput screening of heterologous enzymes secreted by the yeast *Yarrowia lipolytica*. Microb Cell Fact. 2017. https://doi.org/10.1186/s12934-017-0629-5.
6. Boitard L, Cottinet D, Kleinschmitt C, Bremond N, Baudry J, Yvert G, Bibette J. Monitoring single-cell bioenergetics via the coarsening of emulsion droplets. Proc Natl Acad Sci. 2012;109:7181–6. https://doi.org/10.1073/pnas.1200894109.
7. Bower DM, Lee KS, Ram RJ, Prather KLJ. Fed-batch microbioreactor platform for scale down and analysis of a plasmid DNA production process. Biotechnol Bioeng. 2012;109:1976–86. https://doi.org/10.1002/bit.24498.
8. Bowman EK, Alper HS. Microdroplet-assisted screening of biomolecule production for metabolic engineering applications. Trends Biotechnol. 2020;38:701–14. https://doi.org/10.1016/j.tibtech.2019.11.002.
9. Celińska E, Białas W, Borkowska M, Grajek W. Cloning, expression, and purification of insect (*Sitophilus oryzae*) alpha-amylase, able to digest granular starch, in *Yarrowia lipolytica* host. Appl Microbiol Biotechnol. 2015;99:2727–39. https://doi.org/10.1007/s00253-014-6314-2.
10. Celińska E, Borkowska M, Białas W. Enhanced production of insect raw-starch-digesting alpha-amylase accompanied by high erythritol synthesis in recombinant *Yarrowia lipolytica* fed-batch cultures at high-cell-densities. Process Biochem. 2017;52:78–85. https://doi.org/10.1016/j.procbio.2016.10.022.
11. Celińska E, Gorczyca M. Comparison of *Yarrowia lipolytica* growth and rProt synthesis when cultured in different vessels. Mendeley Data. 2024. https://doi.org/10.17632/c3trp7jt3n.1.
12. Chowdhury MS, Zheng W, Kumari S, Heyman J, Zhang X, Dey P, Weitz DA, Haag R. Dendronized fluorosurfactant for highly stable water-in-fluorinated oil emulsions with minimal inter-droplet transfer of small molecules. Nat Commun. 2019. https://doi.org/10.1038/s41467-019-12462-5.
13. Clausell-Tormos J, Lieber D, Baret J-C, El-Harrak A, Miller OJ, Frenz L, Blouwolff J, Humphry KJ, Köster S, Duan H, Holtze C, Weitz DA, Griffiths AD, Merten CA. Droplet-based microfluidic platforms for the encapsulation and screening of mammalian cells and multicellular organisms. Chem Biol. 2008;15:427–37. https://doi.org/10.1016/j.chembiol.2008.04.004.
14. Dai J, Kim HS, Guzman AR, Shim W-B, Han A. A large-scale on-chip droplet incubation chamber enables equal microbial culture time. RSC Adv. 2016;6:20516–9. https://doi.org/10.1039/C5RA26505C.
15. Deere D, Shen J, Vesey G, Bell P, Bissinger P, Veal D. Flow cytometry and cell sorting for yeast viability assessment and cell selection. Yeast. 1998;14:147–60. https://doi.org/10.1002/(SICI)1097-0061(19980130)14:2%3c147::AID-YEA207%3e3.0.CO;2-L.
16. Doig SD, Pickering SCR, Lye GJ, Baganz F. Modelling surface aeration rates in shaken microtitre plates using dimensionless groups. Chem Eng Sci. 2005;60:2741–50. https://doi.org/10.1016/j.ces.2004.12.025.
17. Duetz WA. Microtiter plates as mini-bioreactors: miniaturization of fermentation methods. Trends Microbiol. 2007;15:469–75. https://doi.org/10.1016/j.tim.2007.09.004.
18. Duetz WA, Rüedi L, Hermann R, O'Connor K, Büchs J, Witholt B. Methods for intense aeration, growth, storage, and replication of bacterial strains in microtiter plates. Appl Environ Microbiol. 2000;66:2641–6. https://doi.org/10.1128/AEM.66.6.2641-2646.2000.
19. Duetz WA, Witholt B. Oxygen transfer by orbital shaking of square vessels and deepwell microtiter plates of various dimensions. Biochem Eng J. 2004;17:181–5. https://doi.org/10.1016/S1369-703X(03)00177-3.
20. Duetz WA, Witholt B. Effectiveness of orbital shaking for the aeration of suspended bacterial cultures in square-deepwell microtiter plates. Biochem Eng J. 2001;7:113–5. https://doi.org/10.1016/S1369-703X(00)00109-1.
21. Ferreira-Torres C, Micheletti M, Lye GJ. Microscale process evaluation of recombinant biocatalyst libraries: application to Baeyer-Villiger monooxygenase catalysed lactone synthesis. Bioprocess Biosyst Eng. 2005;28:83–93. https://doi.org/10.1007/s00449-005-0422-4.
22. Gorczyca M, Białas W, Nicaud J-M, Celińska E. 'Mother(Nature) knows best'–hijacking nature-designed transcriptional programs for enhancing stress resistance and protein production in *Yarrowia lipolytica*; presentation of YaliFunTome database. Microb Cell Fact. 2024;23:26. https://doi.org/10.1186/s12934-023-02285-x.
23. Gorczyca M, Każmierczak J, Fickers P, Celińska E. Synthesis of secretory proteins in *Yarrowia lipolytica*: effect of combined stress factors and metabolic load. Int J Mol Sci. 2022;23:3602. https://doi.org/10.3390/ijms23073602.
24. Gorczyca M, Każmierczak J, Steels S, Fickers P, Celińska E. Impact of oxygen availability on heterologous geneexpression and polypeptide secretion dynamics in *Yarrowia lipolytica*-based protein production platforms. Yeast. 2020;37:559–68. https://doi.org/10.1002/yea.3499.

25. Gorczyca M, Nicaud J-M, Celińska E. Transcription factors enhancing synthesis of recombinant proteins and resistance to stress in *Yarrowia lipolytica*. Appl Microbiol Biotechnol. 2023;107:4853–71. https://doi.org/10.1007/s00253-023-12607-z.
26. Hermann R, Lehmann M, Büchs J. Characterization of gas-liquid mass transfer phenomena in microtiter plates. Biotechnol Bioeng. 2003;81:178–86. https://doi.org/10.1002/bit.10456.
27. Holkenbrink C, Dam MI, Kildegaard KR, Beder J, Dahlin J, Domenech Belda D, Borodina I. EasyCloneYALI: CRISPR/Cas9-based synthetic toolbox for engineering of the yeast *Yarrowia lipolytica*. Biotechnol J. 2018. https://doi.org/10.1002/biot.201700543.
28. Jessop-Fabre MM, Jakočiūnas T, Stovicek V, Dai Z, Jensen MK, Keasling JD, Borodina I. EasyClone-MarkerFree: a vector toolkit for marker-less integration of genes into Saccharomyces cerevisiae via CRISPR-Cas9. Biotechnol J. 2016;11:1110–7. https://doi.org/10.1002/biot.201600147.
29. Kensy F, Zimmermann HF, Knabben I, Anderlei T, Trauthwein H, Dingerdissen U, Büchs J. Oxygen transfer phenomena in 48-well microtiter plates: determination by optical monitoring of sulfite oxidation and verification by real-time measurement during microbial growth. Biotechnol Bioeng. 2005;89:698–708. https://doi.org/10.1002/bit.20373.
30. Kildegaard KR, Adiego-Pérez B, Doménech Belda D, Khangura JK, Holkenbrink C, Borodina I. Engineering of *Yarrowia lipolytica* for production of astaxanthin. Synth Syst Biotechnol. 2017;2:287–94. https://doi.org/10.1016/j.synbio.2017.10.002.
31. Kildegaard KR, Tramontin LRR, Chekina K, Li M, Goedecke TJ, Kristensen M, Borodina I. CRISPR/Cas9-RNA interference system for combinatorial metabolic engineering of *Saccharomyces cerevisiae*. Yeast. 2019;36:237–47. https://doi.org/10.1002/yea.3390.
32. Kosa G, Vuoristo KS, Horn SJ, Zimmermann B, Afseth NK, Kohler A, Shapaval V. Assessment of the scalability of a microtiter plate system for screening of oleaginous microorganisms. Appl Microbiol Biotechnol. 2018;102:4915–25. https://doi.org/10.1007/s00253-018-8920-x.
33. Kubiak M, Borkowska M, Białas W, Korpys P, Celińska E. Feeding strategy impacts heterologous protein production in *Yarrowia lipolytica* fed-batch cultures—insight into the role of osmolarity. Yeast. 2019;36:305–18. https://doi.org/10.1002/yea.3384.
34. Kubiak-Szymendera M, Skupien-Rabian B, Jankowska U, Celińska E. Hyperosmolarity adversely impacts recombinant protein synthesis by *Yarrowia lipolytica*—molecular background revealed by quantitative proteomics. Appl Microbiol Biotechnol. 2022;106:349–67. https://doi.org/10.1007/s00253-021-11731-y.
35. Lebrun E, Shenshin V, Plaire C, Vigneres V, Pizette T, Dumas B, Nicaud JM, Mottet G. Efficient full-length IgG secretion and sorting from single yeast clones in droplet picoreactors. Lab Chip. 2023;23:3487–500. https://doi.org/10.1039/d3lc00403a.
36. Lee KS, Boccazzi P, Sinskey AJ, Ram RJ. Microfluidic chemostat and turbidostat with flow rate, oxygen, and temperature control for dynamic continuous culture. Lab Chip. 2011;11:1730. https://doi.org/10.1039/c1lc20019d.
37. Lee KS, Ram RJ. Plastic–PDMS bonding for high pressure hydrolytically stable active microfluidics. Lab Chip. 2009;9:1618. https://doi.org/10.1039/b820924c.
38. Leplat C, Nicaud JM, Rossignol T. High-throughput transformation method for *Yarrowia lipolytica* mutant library screening. FEMS Yeast Res. 2015;15:1–9. https://doi.org/10.1093/femsyr/fov052.
39. Leplat C, Nicaud J-MM, Rossignol T. Overexpression screen reveals transcription factors involved in lipid accumulation in *Yarrowia lipolytica*. FEMS Yeast Res. 2018;18:1–9. https://doi.org/10.1093/femsyr/foy037.
40. Lesage J, Timoumi A, Cenard S, Lombard E, Lee HLT, Guillouet SE, Gorret N. Accelerostat study in conventional and microfluidic bioreactors to assess the key role of residual glucose in the dimorphic transition of *Yarrowia lipolytica* in response to environmental stimuli. N Biotechnol. 2021;64:37–45. https://doi.org/10.1016/j.nbt.2021.05.004.
41. Li Z, Vizeacoumar FJ, Bahr S, Li J, Warringer J, Vizeacoumar FS, Min R, VanderSluis B, Bellay J, DeVit M, Fleming JA, Stephens A, Haase J, Lin Z-Y, Baryshnikova A, Lu H, Yan Z, Jin K, Barker S, Datti A, Giaever G, Nislow C, Bulawa C, Myers CL, Costanzo M, Gingras A-C, Zhang Z, Blomberg A, Bloom K, Andrews B, Boone C. Systematic exploration of essential yeast gene function with temperature-sensitive mutants. Nat Biotechnol. 2011;29:361–7. https://doi.org/10.1038/nbt.1832.
42. Liccioli T, Tran TMT, Cozzolino D, Jiranek V, Chambers PJ, Schmidt SA. Microvinification—how small can we go? Appl Microbiol Biotechnol. 2011;89:1621–8. https://doi.org/10.1007/s00253-010-2992-6.
43. Maier U, Büchs J.Gas-Flüssigkeits-Stofftransfer im Schüttelkolben. Aachen, Techn. Hochsch. 2002.
44. Maier U, Losen M, Büchs J. Advances in understanding and modeling the gas-liquid mass transfer in shake flasks, In: Biochemical Engineering Journal. Elsevier. 2004, pp. 155–167. https://doi.org/10.1016/S1369-703X(03)00174-8
45. Marešová L, Sychrová H. Applications of a microplate reader in yeast physiology research. Biotechniques. 2007;43:667–72. https://doi.org/10.2144/000112620.
46. Meier K, Klöckner W, Bonhage B, Antonov E, Regestein L, Büchs J. Correlation for the maximum oxygen transfer capacity in shake flasks for a wide range of operating conditions and for different culture media. Biochem Eng J. 2016;109:228–35. https://doi.org/10.1016/j.bej.2016.01.014.
47. Milne N, Tramontin LRR, Borodina I. A teaching protocol demonstrating the use of EasyClone and CRISPR/Cas9 for metabolic engineering of *Saccharomyces cerevisiae* and *Yarrowia lipolytica*. FEMS Yeast Res. 2020. https://doi.org/10.1093/femsyr/foz062.
48. Qiu X, Xu P, Zhao X, Du G, Zhang J, Li J. Combining genetically-encoded biosensors with high throughput strain screening to maximize erythritol production in *Yarrowia lipolytica*. Metab Eng. 2020;60:66–76. https://doi.org/10.1016/j.ymben.2020.03.006.
49. Sáez-Sáez J, Wang G, Marella ER, Sudarsan S, Cernuda Pastor M, Borodina I. Engineering the oleaginous yeast *Yarrowia lipolytica* for high-level resveratrol production. Metab Eng. 2020;62:51–61. https://doi.org/10.1016/j.ymben.2020.08.009.
50. Seletzky JM, Noak U, Fricke J, Welk E, Eberhard W, Knocke C, Büchs J. Scale-up from shake flasks to fermenters in batch and continuous mode with *Corynebacterium glutamicum* on lactic acid based on oxygen transfer and pH. Biotechnol Bioeng. 2007;98:800–11. https://doi.org/10.1002/bit.21359.
51. Sturmberger L, Menczik P, Moser S, Kotz D, Aleschko M. Scale-dependent effect of helper factor co-expression in Pichia pastoris, In: ICY15 Meets ICYGMB30, Vienna, Austria. 2021, p. B10.
52. Totaro D, Radoman B, Schmelzer B, Rothbauer M, Steiger MG, Mayr T, Sauer M, Ertl P, Mattanovich D. Microscale perfusion-based cultivation for pichia pastoris clone screening enables accelerated and optimized recombinant protein production processes. Biotechnol J. 2021. https://doi.org/10.1002/biot.202000215.
53. Totaro D, Rothbauer M, Steiger MG, Mayr T, Wang HY, Lin YS, Sauer M, Altvater M, Ertl P, Mattanovich D. Downscaling screening cultures in a multifunctional bioreactor array-on-a-chip for speeding up optimization of yeast-based lactic acid bioproduction. Biotechnol Bioeng. 2020;117:2046–57. https://doi.org/10.1002/bit.27338.
54. Trassaert M, Vandermies M, Carly F, Denies O, Thomas S, Fickers P, Nicaud JM. New inducible promoter for gene expression and synthetic biology in *Yarrowia lipolytica*. Microb Cell Fact. 2017. https://doi.org/10.1186/s12934-017-0755-0.
55. Vidal L, Lebrun E, Park YK, Mottet G, Nicaud JM. Bidirectional hybrid erythritol-inducible promoter for synthetic biology in *Yarrowia lipolytica*. Microb Cell Fact. 2023. https://doi.org/10.1186/s12934-023-02020-6.
56. Wackett LP. Why Is the Biodegradation of Polyfluorinated Compounds So Rare? mSphere. 2021. https://doi.org/10.1128/msphere.00721-21.
57. Wagner JM, Liu L, Yuan S-F, Venkataraman MV, Abate AR, Alper HS. A comparative analysis of single cell and droplet-based FACS for improving production phenotypes: riboflavin overproduction in *Yarrowia lipolytica*. Metab Eng. 2018;47:346–56. https://doi.org/10.1016/j.ymben.2018.04.015.
58. Warringer J, Blomberg A. Automated screening in environmental arrays allows analysis of quantitative phenotypic profiles in *Saccharomyces cerevisiae*. Yeast. 2003;20:53–67. https://doi.org/10.1002/yea.931.
59. Wittmann C, Schütz V, John G, Heinzle E. Quantification of oxygen transfer in test tubes by integrated optical sensing. J Microbiol Biotechnol. 2004;14:991–5.
60. Yu S, Zhang G, Liu Q, Zhuang Y, Dai Z, Xia J. Construction and testing of *Yarrowia lipolytica* recombinant protein expression chassis cells based

on the high-throughput screening and secretome. Microb Cell Fact. 2023. https://doi.org/10.1186/s12934-023-02196-x.
61. Zagnoni M, Cooper JM. Droplet microfluidics for high-throughput analysis of cells and particles. Methods Cell Bio. 2011. https://doi.org/10.1016/B978-0-12-374912-3.00002-X.
62. Zhang Q, Wang T, Zhou Q, Zhang P, Gong Y, Gou H, Xu J, Ma B. Development of a facile droplet-based single-cell isolation platform for cultivation and genomic analysis in microorganisms. Sci Rep. 2017. https://doi.org/10.1038/srep41192.

Publisher's Note

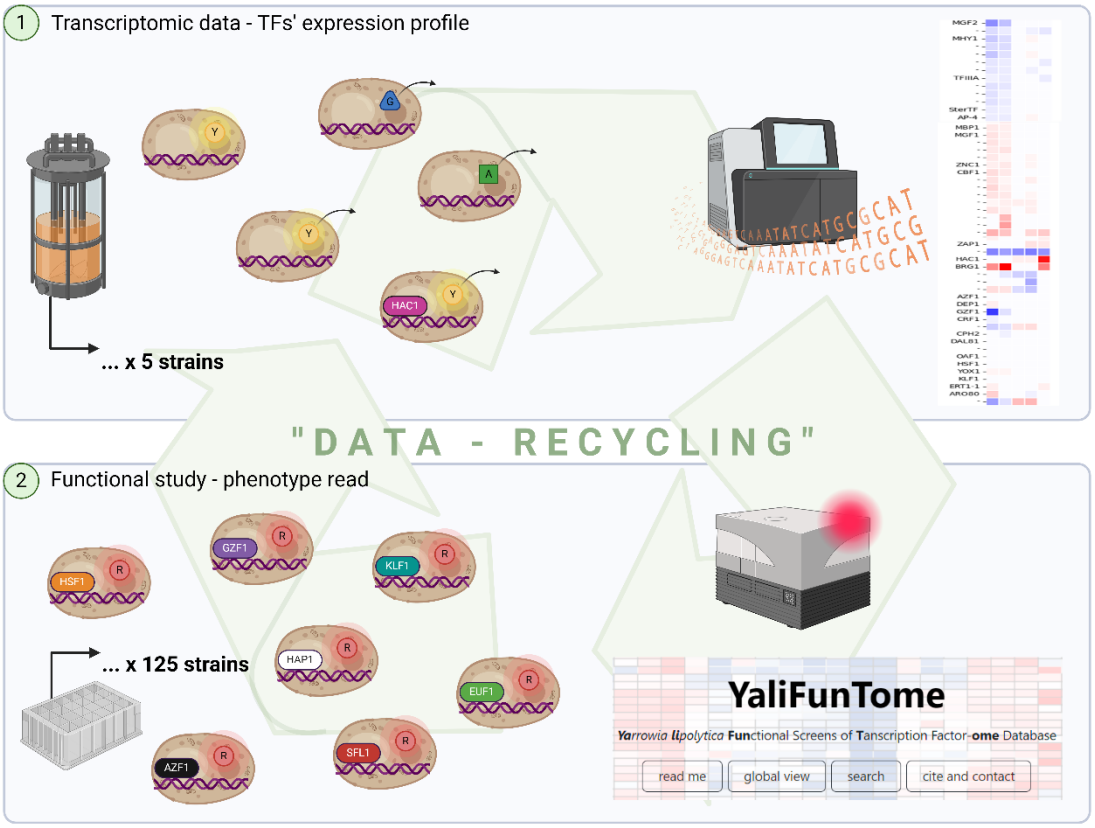
5.5 **P4**: Współoddziaływanie czynników transkrypcyjnych i syntezy rekombinowanych białek w drożdżach *Yarrowia lipolytica* na poziomach transkrypcyjnym i funkcjonalnym – perspektywa globalna

Czwarta publikacja **P4** stanowi rozszerzenie wachlarza metodycznego w badaniach nad rolą TFs w regulacji syntezy rProts w *Y. lipolytica*. We wcześniejszych badaniach prowadzonych przez nasz zespół (Korpys-Woźniak i Celińska, 2021) analizowano transkryptomy szczepów *Y. lipolytica* nadprodukujących rProts o różnych właściwościach biochemicznych. W wyniku przeprowadzonych analiz przedstawiono profile genów deregulowanych na skutek nadprodukcji rProt, w tym zestawy TFs. Z kolei badania prowadzone w toku niniejszej rozprawy (**P1** i **P2**) pozwoliły wskazać funkcjonalny wpływ nadekspresji TFs na zdolność syntezy rProt u *Y. lipolytica*. W ten sposób zgromadzono znaczną ilość danych i zdeponowano ją w publicznych repozytoriach (NCBI-SRA oraz YaliFunTome). W niniejszym badaniu oba zbiory danych biologicznych zostały zintegrowane i wzbogacone o dodatkowe dane eksperymentalne, aby zbadać współoddziaływanie TFs i syntezy rProt u *Y. lipolytica* na poziomie transkrypcyjnym i funkcjonalnym.

Punktem wyjścia była hipoteza, że dane transkryptomiczne mogą służyć jako narzędzie prognostyczne do wyboru TFs potencjalnie zaangażowanych w syntezę rProt i wstępnie wskazywać ich rolę jako aktywatorów/represorów tego procesu biologicznego. Z drugiej strony, celem badań było sprawdzenie zależności odwrotnej – powiązanie fenotypu w odniesieniu do produkcji białek z profilem ekspresji genów kodujących TFs w szczepach produkujących białko.

Koncepcyjnie zastosowane podejście wpisuje się w ideę „recyklingu danych” (ang. *data-recycling*), rozwiniętą wraz z akumulacją „dużych danych biologicznych” (ang. *big biological data*), zgromadzonych zarówno dzięki postępowi metod sekwencjownowania, jak i rozwojowi zautomatyzowanych, wysokoprzepustowych podejść eksperymentalnych, oraz ich deponowaniu i upowszechnianiu w publicznych repozytoriach. Zgodnie z koncepcją „data-recycling” zamiast powielać eksperymenty i wytwarzać kolejne zbiory danych, postuluje się ponowne wykorzystanie już istniejących wyników, co pozwala na minimalizację zużycia zasobów oraz bardziej racjonalne wykorzystanie budżetu publicznego przeznaczonego na rozwój nauki. „Recykling danych” zyskuje coraz większą popularność w biologii. Niekiedy, okazuje

się efektywnym narzędziem w generowaniu nowych hipotez i odkryć (Doughty i Kerkhoven, 2020).



Rycina 2. Graficzne podsumowanie koncepcji prac opisanych w artykule **P3**. Technicznie praca opiera się na zestawieniu danych, zgodnie z koncepcją „data-recycling". Wykorzystano dane 1) transkryptomiczne – profile eksresji TFs w szczepach nadekspresjonujących biochemicznie zróżnicowane rProts, kierowane do wnętrza komórki lub do szlaku sekrecyjnego. Szczepy zawierały: i) α-amylazę z *Sitophilus oryzae*, sekrecyjną, posttranslacyjnie ulegająca tworzeniu licznych mostków disiarczkowych, ii) sekrecyjne białko żółtej fluorescencji (YFP, ang. *yellow fluorescent protein*), podlegające minimalnym zmianom posttranslacyjnym, iii) wewnątrzkomórkowe białko YFP, iv) glukoamylazę z *Thermomyces lanuginosus*, sekrecyjną, posttranslacyjnie silnie glikozylizowana, v) szczep prowadzący ko-nadekspresję sekrecyjnego YFP oraz TF Hac1 (*YALI0B12716g*)); dane zostały wyekstrahowane z bazy danych NCBI SRA; 2) funkcjonalne – profili poziomu syntezy fluorescencyjnego białka w 125 szczepach ko-nadekspresjonujących wewnątrzkomórkowo białko fluorescencyjne RedStar2 oraz pojedynczy gen TF, w „standardowych" warunkach (pozyskane z danych **P2**). Na podstawie tych dwóch cech, TFs pogrupowano metodą k-średnich.

Rycina 2 przedstawia podsumowanie koncepcji i prac opisanych w artykule **P3** w zakresie dotyczącym zestawienia dwóch typów danych (transkryptomicznych i funkcjonalnych). Dla wybranych TF wprowadzono delecję w *locus* kodującym TF, dla lepszego zrozumienia ich rzeczywistego udziału w syntezie rProt.

Szczepy oznaczone jako i) oraz ii) w opisie **Ryciny 2** zostały scharakteryzowane wspólnie jako wykazujące wysoką syntezę i sekrecję białek (HSS, ang. *high synthesis and secretion*). Pozostałe trzy, na podstawie danych transkryptomicznych, odznaczały się nasilonym zjawiskiem stresu UPR oraz ograniczonym poziomem syntezy rProt, i zostały uogólniająco sklasyfikowane do jednej grupy określonej tą nazwą (UPR). W dalszej części omówienia wyników będę odnosić się do obu tych grup, stosując wskazane powyżej akronimy.

Analiza profili ekspresji pozwoliła wyróżnić kilka interesujących podgrup TFs, m.in. takie, które są deregulowane w sposób odwrotny w szczepach o wysokiej (HSS) i niskiej (UPR) produkcji rProt, a także takich, które były deregulowane w podobny sposób w obu grupach. Wbrew założeniu, że odwrócony profil ekspresji lub jednolita tendencja deregulacji powinny wskazywać na bezpośrednie zaangażowanie w badane procesy, nadekspresja tych TFs nie prowadziła do jednoznacznych zmian fenotypowych pod względem syntezy rProt, jak wskazały wyniki z bazy YaliFunTome. Oznacza to, że same dane transkryptomiczne nie mogą być traktowane jako bezpośredni prognostyk udziału danego TF w regulacji produkcji białek. Innym przykładem jest *GZF1*, który jako jedyny z wcześniej rozpoznanych globalnych „wzmacniaczy" i represorów (Klf1, Hsf1, Gzf1, Dep1, Azf1 i Cat8) wykazywał deregulowany profil ekspresji – ulegał silnej represji w szczepach HSS, a jednocześnie jego nadekspresja wywołała wysoki poziom produkcji rProt. Z kolei porównanie profilu ekspresji (brak zmian) i fenotypu szczepu nadekpresjonującego TF *KLF1* (uniwersalny „wzmacniacz"), pozwala stwierdzić, że biologiczna aktywacja TF odbywa się na innym poziomie niż poziom transkrypcji, np. poprzez modyfikacje posttranslacyjne (jak fosforylacja) lub relokalizację w obrębie struktur komórki. Według danych literaturowych Klf1 jest relokalizowany do jądra komórkowego w warunkach stresu oksydacyjnego (Herholz i in., 2019), gdzie zyskuje funkcjonalność; co jest spójne z naszym wnioskiem.

Analizy szczepów z delecjami wybranych TFs dostarczyły dodatkowych informacji o ich zaangażowaniu w proces syntezy rProt. W przypadku *GZF1* i *HSF1*, podobnie jak obserwowano już w **P1**, delecja nie prowadziła do „odwróconego fenotypu" w stosunku do nadekspresji. Może to sugerować istnienie mechanizmów kompensacji,

prawdopodobnie wynikających z częściowej redundancji funkcjonalnej w obrębie rodziny czynników Gzf, bądź pełnienie przez Hsf1 roli globalnego regulatora. Ciekawym rozwiązaniem interpretacyjnym, pozwalającym zrozumieć te obserwacje, jest niedawno przedstawiona koncepcja „genów buforowych" (ang. *buffer genes*) (Tawfeeq i in., 2024). Manipulacja genetyczna w ich *loci* nie musi zawsze skutkować manifestacją fenotypową, ale zmienia wrażliwość danego genomu na dalsze perturbacje genetyczne lub środowiskowe. Dodatkowym przykładem złożoności oddziaływań regulacyjnych jest główna podjednostka kompleksu remodelującego chromatyny *INO80* w *S. cerevisiae*. W wyniku jego nadekspresji lub delecji wykazano klasyczny „odwrócony" fenotyp tolerancji na wysokie stężenia kwasu octowego, jednak analizy transkryptomiczne i epigenomiczne ujawniły, że jedynie 7,7% zmian pokrywało się w obu wariantach (Yuan i in., 2024, 2025). Nadekspresja aktywowała przede wszystkim metabolizm azotu, natomiast delecja prowadziła do szerokich zmian w ekspresji i dostępności chromatyny genów związanych z integralnością ściany komórkowej, co wskazuje na całkowicie odmienne mechanizmy regulacyjne leżące u podstaw fenotypu. Podkreśla to wielopoziomową organizację sieci transkrypcyjnych (TRN, *transcriptional regulatory network*), złożoność regulacji metabolicznej, jak również brak prostej, lustrzanej relacji między nadekspresją a delecją danego regulatora; zarazem sugeruje, że brak odwróconych fenotypów w innych przypadkach – jak obserwowane dla *HSF1* i *GZF1* – może wynikać z różnorodnych mechanizmów kontroli procesów molekularnych.

W wyniku przeprowadzonych analiz w ramach prac eksperymentalnych w **P4**, dokonano ciekawych obserwacji dla genu *DEP1*. Jego delecja skutkowała największym wzrostem syntezy rProt, uzyskanym w niniejszej rozprawie doktorskiej. W przypadku obserwacji na temat innych TF, wybranych na podstawie danych literaturowych jako ogólne regulatory odpowiedzi na stres (Skn7, Hap1, Sfl1 – *YALI0D04785g*), ostrożnie wnioskowano, że w warunkach braku stresu środowiskowego, usunięcie ekspresji globalnego regulatora odpowiedzi stresowej może sprzyjać kosztownemu biologicznie procesowi, jakim jest synteza rekombinowanych białek.

Zaprezentowana w **P4** globalna analiza danych pozwoliła zaobserwować wyraźny wzorzec metaboliczny: aktywacja szlaków związanych z zwiększoną degradacją związków azotowych sprzyjała syntezie rProt (Dal81, Aro80 – *YALI0C18645g*), podczas gdy nasilone sygnalizowanie głodu azotowego (Gcn4 – *YALI0E27742g*) już nie; podobnie indukcja wykorzystania alternatywnego źródła węgla (Adr1 – *YALI0F21923g*, Cat8, Ert1-2 – *YALI0E03410g*) działa hamująco

na produkcję białek. Analogiczne zależności zaobserwowano także w **P2**, gdzie globalne represory (Dep1, Azf1, Cat8) powiązane były z anabolizmem węgla. Wyniki te pozwalają sformułować hipotezę o wyraźnie przeciwnym wpływie metabolizmu azotu i węgla na produkcję białek. Przykładem regulatorów promujących syntezę rProt są Dal81 i Aro80, pierwszy, będący pozytywnym regulatorem genów degradacji wielu związków azotowych, drugi wspierający dostępność prekursorów aminokwasów. Ich ekspresja nie była wyraźnie zmieniona w analizowanych profilach transkrypcyjnych – co sugeruje regulację na poziomie posttranslacyjnym. Wnioski te wskazują, że przyszłe badania nad inżynierią metaboliczną *Y. lipolytica* powinny koncentrować się na regulatorach związanych z metabolizmem azotu jako potencjalnych celach modyfikacji genetycznych.

Integracja danych transkryptomicznych i funkcjonalnych pozwoliła na krytyczną weryfikację hipotezy, zgodnie z którą profil ekspresji TF mógłby stanowić bezpośredni predyktor jego roli w syntezie rProt. Analiza wykazała, że w przypadku TF, korelacja pomiędzy poziomem transkryptu a fenotypem wynikającym z nadekspresji TF była w większości przypadków słaba lub nieistotna. Mimo to badania dostarczyły cennych informacji i pozwoliły wskazać nowe, potencjalne cele inżynierii metabolicznej. Do interesujących kandydatów należą: Zap1 (*YALI0D23749g*) - TF odpowiadający za homeostazę jonów cynku, którego ekspresja była podwyższona w szczepach UPR, a nadekspresja zwiększała poziom syntezy rProt o 19%. TF024 (*YALI0D01353g*) i Brg1 (*YALI0E31757g*) geny o jednolitej represji i indukcji w większości szczepów nadekspresjonujących białka.

International Journal of
Molecular Sciences

MDPI

Article

An Interplay between Transcription Factors and Recombinant Protein Synthesis in *Yarrowia lipolytica* at Transcriptional and Functional Levels—The Global View

Maria Gorczyca, Paulina Korpys-Woźniak and Ewelina Celińska *

Department of Biotechnology and Food Microbiology, Poznan University of Life Sciences, Wojska Polskiego 48, 60-637 Poznan, Poland; maria.gorczyca@up.poznan.pl (M.G.); paulina.korpys@up.poznan.pl (P.K.-W.)
* Correspondence: ewelina.celinska@up.poznan.pl

Abstract: Transcriptional regulatory networks (TRNs) associated with recombinant protein (rProt) synthesis in *Yarrowia lipolytica* are still under-described. Yet, it is foreseen that skillful manipulation with TRNs would enable global fine-tuning of the host strain's metabolism towards a high-level-producing phenotype. Our previous studies investigated the transcriptomes of *Y. lipolytica* strains overproducing biochemically different rProts and the functional impact of transcription factors (TFs) overexpression (OE) on rProt synthesis capacity in this species. Hence, much knowledge has been accumulated and deposited in public repositories. In this study, we combined both biological datasets and enriched them with further experimental data to investigate an interplay between TFs and rProts synthesis in *Y. lipolytica* at transcriptional and functional levels. Technically, the RNAseq datasets were extracted and re-analyzed for the TFs' expression profiles. Of the 140 TFs in *Y. lipolytica*, 87 TF-encoding genes were significantly deregulated in at least one of the strains. The expression profiles were juxtaposed against the rProt amounts from 125 strains co-overexpressing TF and rProt. In addition, several strains bearing knock-outs (KOs) in the TF loci were analyzed to get more insight into their actual involvement in rProt synthesis. Different profiles of the TFs' transcriptional deregulation and the impact of their OE or KO on rProts synthesis were observed, and new engineering targets were pointed.

Keywords: heterologous protein; yeast; protein production; omics data; transcriptional regulation

Citation: Gorczyca, M.; Korpys-Woźniak, P.; Celińska, E. An Interplay between Transcription Factors and Recombinant Protein Synthesis in *Yarrowia lipolytica* at Transcriptional and Functional Levels—The Global View. *Int. J. Mol. Sci.* **2024**, *25*, 9450. https://doi.org/10.3390/ijms25179450

Academic Editor: Clemente Capasso

Received: 31 July 2024
Revised: 26 August 2024
Accepted: 28 August 2024
Published: 30 August 2024

1. Introduction

Transcriptional regulatory networks (TRNs) can be defined as regulatory interactions among transcription factors (TFs) and their target genes [1,2]. A given TRN is responsive to a given stimulus and governs a specific biological process. Skillful manipulation with TRN will presumably lead to global optimization of a microbial cell towards a researcher-defined target. TRNs have not yet been described for *Y. lipolytica*; however, a very insightful metabolic model has been recently released [3], though not yet tested for recombinant protein (rProt) production prediction. In our studies, we are interested in defining TFs, and in further perspective, TRNs, that are involved in enhancing *Yarrowia lipolytica*'s capacity towards high-level production of rProts. The first step in the endeavor is identifying TFs implicated in the process. *Y. lipolytica* has long been used as a microbial platform for rProts production [4–7]. Thanks to recent achievements in developing an erythritol-inducible expression system [8–13], *Y. lipolytica* has become a potent alternative to conventional expression platforms in industrial applications (several companies are implementing this system into their portfolio).

Previously, we have shown that by applying the knowledge-driven genetic engineering of *Y. lipolytica*'s synthesis and secretion mechanisms, the rProt yields can be significantly enhanced [14–16]. By defining the genes deregulated upon production of biochemically-different rProt [16] and using them as engineering 'helper genes' [15], the target rProt

amounts were increased by over 2-fold. Notably, the previous set of the 'helper genes' covered those involved in the translation, folding, and trafficking of polypeptides, as well as one TF, Hac1 (*YALI0B12716*; hereafter, gene signatures will be abbreviated by skipping the '*YALI0*' prefix), which is known to initiate the ER-based unfolded protein response (UPR). Hac1 mediates the activation of hundreds of molecular events, including the increased provision of chaperones and membranes, to concertedly relieve the burdened secretory pathway, e.g., [17,18]. Hence, the Hac1-governed TRN is directly associated with the biological process 'protein synthesis and secretion' and matches our practical aim of enhancing rProts synthesis. Technically, co-overexpression (co-OE) of *HAC1* with a reporter protein initially led to mediocre yield improvement (of ~30%; [15]), but more in-depth studies showed that the effect was indeed bifurcated, causing a nearly 7-fold drop in the retained fraction of rProt but promoting its secretion by nearly 2.5-fold [19].

The following studies on exploiting *Y. lipolytica*'s TRNs for researcher-defined practical aims were focused on enhancing lipid accumulation [20,21], stress resistance, and rProts synthesis [22,23]. In their high-throughput screens, [20] i.a., identified *Brg1* (*E31757*) as a TF promoting total lipid accumulation by ~1.9-fold when OE (the highest increase read from amongst all TFs tested). Gorczyca et al. [22] proved that the increased abundance of Skn1 (*D14520*) enables maintaining the rProt synthesis capacity under heavy osmotic stress, while OE of the Yap-like TF-encoding gene (*D07744*) positively impacts the growth of *Y. lipolytica* under alkaline conditions. OE of *GZF1* (*D20482*) and *HSF1* (*E13948*) brings universal improvement in the rProts synthesis, irrespectively from the environmental conditions. Deletion (knock-out; KO) of *GZF1* and *HSF1* promotes the synthesis of neutral lipids under a low C/N ratio (C/N of 4; not promoting lipid accumulation). Finally, a comprehensive investigation of a library of *Y. lipolytica* strains over-expressing (OE-ing) one of 125 TFs under 72 different environmental conditions revealed multiple interesting phenotypes driven by the increased abundance of the TFs [23], data presented in a searchable database YaliFunTome (https://sparrow.up.poznan.pl/tsdatabase/, accessed throughout June 2024). *Inter alia*, the YaliFunTome database presents the phenotype of another global enhancer of rProt synthesis, Klf1 (*D05041*), as well as TFs acting against this biological process, Azf1 (*A16841*), Dep1 (*F05896*), and Cat8 (*C19151*), all of which are known to be involved in carbon metabolism [24–28]. Notably, several of the TFs studied there promoted growth under limited oxygen availability (*LAC9* (*D20460*), *B14443*, *B20944*, and *DAL81* (*D02783*)), which is the major limiting factor in *Y. lipolytica* growth and rProt synthesis [29–31]. Altogether, those studies proved the validity of the concept of using TFs as tools in driving desired phenotypes in *Y. lipolytica*.

With the current advent of next-generation sequencing technologies, DNA synthesis, and even miniaturization and parallelization of strain culturing and sample analysis, the repositories of 'big biological data' have grown enormously. This is specifically true for the model microorganisms, like *Saccharomyces cerevisiae* or *Escherichia coli*. Considering the extent of the datasets deposited in publicly accessible databases, 'big biological data recycling' has become a very efficient research approach in driving novel hypotheses and findings, also in the field of yeast research [32]. If approached carefully and with adequate methodology [33–37], reusing previous data enables substantial time- and resource-savings. Examples of successful completion of such analyses for yeast are available [3,38,39]. When compared to the model yeast species, the high-throughput datasets for *Y. lipolytica* deposited publicly are still limited. Nevertheless, some transcriptomics data recycling for *Y. lipolytica* has been very recently reported [11,19,39]; where RNAseq datasets were extracted, arranged in different combinations, and re-analyzed to answer new research questions. Those previous studies reported recycling of transcriptomics data only.

In this study, we aimed to combine and reanalyze datasets related to the expression profile of TFs and the functional effect of the TFs OE on rProt synthesis in *Y. lipolytica*. The previously generated datasets (transcription level of TFs and amounts of rProts triggered by TF's OE) were juxtaposed to investigate an interplay between TRNs and rProts synthesis in *Y. lipolytica* at transcriptional and functional levels. Having transcriptomics data on one

hand and functional screening results on the other, we wanted to test a working hypothesis of whether transcriptomics data could be used as a selection-driver of TF involved in a specific biological process, here—rProt synthesis in *Y. lipolytica*, and if a preliminary prognostic of its mode of implication can be inferred. For several TFs, further experimental verification followed the computational analyses.

2. Results and Discussion

2.1. Data Extraction, Juxtaposition, and Global Clustering—Clusters Overview

Transcriptomics data were extracted from the NCBI SRA (Sequence Read Archive) database (PRJNA701856 and PRJNA869113). The data represent the relative expression level of a given TF-encoding gene in a *Y. lipolytica* strain OE-ing one of the biochemically different rProt: (i) scSoA—highly disulfide-bonded secreted alpha-amylase from *Sitophilus oryzae*; (ii) scYFP—minimally modified post-translationally secreted small yellow fluorescence protein; (iii) inYFP—the same protein (YFP) in intracellular form; (iv) scTlG—highly glycosylated secreted glucoamylase from *Thermomyces lanuginosus*; or a strain co-OE-ing scYFP and a TF Hac1 [16,19]. The strains were maintained in steady-state, as indicated previously [14,16]. The two strains, scSoA and scYFP, were characterized by high-level synthesis and secretion of the rProts and hence were assigned names and abbreviated as high-level synthesis and secretion, HSS. The other three, inYFP, TlG, and scYFP-HAC1, produced limited amounts of the rProts, and based on transcriptomics data were shown to be facing UPR, hence they were assigned and abbreviated as UPR strains. The raw data were filtered for genes encoding TFs, according to a previously defined list of putative TFs [20]. Only these records were considered further in this study. Differential expression data, used hereafter, express fold change (FC) values calculated as a ratio of the normalized transcript counts (log2) in the OE-ing strain to a prototrophic control, maintained likewise.

Functional data on the amounts of rProt (intracellular fluorescent protein) synthesized by *Y. lipolytica* strains OE-ing a specific TF were extracted from the YaliFunTome database [23]. The data were read as fluorescence of the strain, expressed as total amounts of rProt or normalized per biomass. Differential phenotype data, used hereafter, denote FC values calculated as a ratio of the total or normalized fluorescence readouts in the OE-ing strain to a prototrophic control, maintained likewise.

All the TFs considered in this analysis, along with assigned or putative functions, are given in Table 1. Both datasets generated previously were expanded due to different sets of TFs scoring significant results in the transcriptomics (87) and functional screens (significant data and tendencies—124). Hence, here analyzed dataset is original in terms of extent and also considering the combination of functional and transcriptomics data.

Table 1. *Y. lipolytica* TFs with putative or assigned function.

Yali Signature	Assigned Name	Putative/Known Function	Reference
B12716	*HAC1*	Transcriptional activator of genes involved in ER-based Unfolded Protein Response (UPR).	[16,19]
E31757	*BRG1*	Biofilm regulator 1.	[20]
D14520	*SKN7*	TF involved in the activation of osmotic and oxidative stress.	[22,40]
D07744	*YAP-like*	TF involved in pH-dependent dimorphic transition and maintaining redox balance.	[22,41]
D20482	*GZF1*	Inducer of the nitrogen catabolite repression (NCR) genes.	[23,42]
E13948	*HSF1*	Heat shock transcription factor.	[22,23]
D05041	*KLF1*	Krueppel-like factor 15, regulates the expression of genes for gluconeogenic and amino acid-degrading enzymes.	[23]
A16841	*AZF1*	Asparagine-rich Zinc Finger protein, regulates carbon metabolism in yeast and cell wall organization.	[20,23]

Table 1. *Cont.*

Yali Signature	Assigned Name	Putative/Known Function	Reference
F05896	*DEP1*	Part of the Rpd3C(L) histone deacetylase complex (HDAC). Transcriptional modulator involved in regulation of structural phospholipid biosynthesis genes.	[20,23]
C19151	*CAT8*	CATabolite repression TF 8. Inducer of gluconeogesis, maintains energy homeostasis, presumed to regulate formate dehydrogenases expression.	[39]
F21923	*ADR1*	Alcohol Dehydrogenase II synthesis Regulator, inducer of genes involved in alternative carbon utilization upon glucose starvation.	[11,39]
D20460	*LAC9*	LACtose regulatory protein, controls induction of the lactose-galactose regulation.	[23]
B14443	*JMC2*	JmjC domain family histone demethylase, promotes global demethylation of H_3K_4.	[23,43]
D02783	*DAL81*	Positive regulator of genes in multiple nitrogen degradation pathways, involved in nitrogen catabolite activation of transcription from RNA polymerase II promoter.	[23]
B19602	*MGF2*	Mycelial growth factor 2. TF of known roles in the dimorphic transition.	[16,41]
B21582	*MHY1*	Mns2/Mns4-like protein, a key regulator of yeast-to hypha dimorphic transition but not stress resistance, regulates both alkaline-pH and glucose-induced filamentation.	[44–46]
D01573	*MGF1*	Mycelial growth factor. Potential driver of the transition between morphological phases.	[16]
B05038	*ZNC1*	Zinc finger transcriptional factor, regulates the yeast-to-hypha transition in the dimorphic yeast.	[47]
F17886	*GZF2*	GATA—binding zinc finger transcription factor 2, inducer of NCR, essential for growth on simple nitrogen sources.	[42]
C22682	*GZF3*	GATA-zinc finger transcription factor 3, repressor of NCR.	[20,42]
E05555	*GZF4*	GATA-zinc finger transcription factor 4. Putative: inducer of NCR.	[42]
E16577	*GZF5*	Non-genuine GATA-zinc finger transcription factor.	[20,42]
E03410	*ERT1-2*	Positive regulator of gluconeogenesis.	[20]
E27742	*GCN4*	General Control Non-derepressible 4 TF. Key transcriptional activator of amino acid biosynthesis genes.	[48]
F17424	*HAP1*	TF responsible for oxygen sensing and signaling.	[49]
D04785	*SFL1*	Repressor of filamentous growth and flocculation.	
F01562	*EUF1*	Transcription factor mediating expression of erythritol synthesis genes.	[11,50,51]
C16863	*SKO1*	*ATF/CREB* family transcription factor, repressor that mediates *HOG* pathway-dependent regulation of osmotic stress response, involved in protection from oxidative damage.	[40]
C14784	*YAS3*	Transcriptional repressor of *ALK* genes, de-repressed on alkanes.	[52]

Table 1. *Cont.*

Yali Signature	Assigned Name	Putative/Known Function	Reference
C13750	*MSN4*	General stress response, regulates tolerance to acid-induced stress. Regulates genes involved in the antioxidant cellular response.	[53]
B13640	*RIM101*	pH-response transcription factor, regulator of alkaline-induced filamentation	[46]
C12364	*NRG1*	*NRG1* repressor of erythritol utilization genes, plays a minor role in repression of filamentation.	[50,54]
D23749	*ZAP1*	*ZAP1* involved in zinc ion homeostasis by zinc-responsive transcriptional regulation.	[20]
E30789		Putative[a]: transcriptional regulator of form adherence 5.	
C07821		Putative[a]: glucose transport transcription regulator *RGT1*-related.	
C18645	*ARO80*	Transcription activator required for the expression of genes involved in the catabolism of aromatic amino acids.	[55]
B08206	*CRF1*	Copper resistance protein transcriptional regulator.	[56]
D14872		Putative [a]: transcriptional regulatory protein *STB4*	
E18304	*ERT1-1*	Transcription activator of gluconeogenesis.	[20]
E07942	*MIG1*	Controls genes involved in beta-oxidation, involved in carbon catabolite repression.	[57]
A18469	*HOY1*	Homeobox protein, a positive regulator of hyphae formation.	[20,58]
B06853	*PUT3*	*PUT3* proline utilization trans-activator.	
B09713	*PPR1*	*PPR1* pyrimidine pathway regulatory protein, de novo biosynthesis.	[54,59]
D13904	*LEU3*	Leucine-responsive transcription regulator, regulates genes involved in branched-chain amino acid biosynthesis and ammonia assimilation.	
D23045	*AHR1*	*AHR1* adhesion and hyphal regulator 1.	
E10681	*WAR1*	Weak Acid Resistance transcription factor.	
E20449	*YOX1*	Homeobox protein *YOX1*. Transcriptional repressor of ECB-dependent genes (early cell box) to the G1/M phase.	
F03157	*MET32*	*MET32* auxiliary transcriptional regulator of sulfonate and sulfur amino acid metabolism, methionine biosynthesis, and sulfate assimilation.	
F05104	*TFIIIA*	*PZF1* general transcription factor IIIA.	
F09361	*U4/U6.U5 component*	U4/U6.U5 tri-snRNP, spliceosomal complex—may play a role in mRNA splicing.	
F09493	*SAP61*	Pre-mRNA-splicing factor sap61, involved in mRNA splicing, associates with cdc5 and the other cwf proteins as part of the spliceosome.	
F16599	*STB4*	Putative transcription factor *STB4*—Sin Three Binding protein, involved in the transcription of transmembrane transporters.	[60]
F18788	*RLM1*	May function as a TF downstream of *MPK1*, at least some RML1 target genes are involved in cell wall biosynthesis.	
E31845	*PRZ1*	Involved in the regulation of calcium ion homeostasis.	

Table 1. *Cont.*

Yali Signature	Assigned Name	Putative/Known Function	Reference
B15818	*SterTF*	Sterol transcription factor, regulating sterol biogenesis.	[61]
D05005	*SEF1*	Zn2-Cys6 transcription factor; regulates iron uptake.	
D12628	*POR1*	Primary oleate regulator 1—transcriptional activator regulating genes involved in fatty acid utilization.	[62]
C02387	*YAS1*	Transcription factor essential for cytochrome p450 induction in response to alcanes, heteromeric Yas1p/Yas2p complex transcription factor.	[63]
B08734	*REI1*	Cytoplasmic pre-60S factor *REI1* involved in maturation of the ribosomal 60S subunit.	
E01606	*OAF3*	Oleate activated transcription factor 3, transcriptional inhibitor with a significantly increased number of target genes in response to oleate.	
C06842	*MCM1*	*MCM1* transcription factor involved in biofilm formation, cell adhesion, and hyphal growth.	
A19778	*MBP1*	MluI-box Binding Protein, involved in regulation of cell cycle progression from G1 to S phase.	[39,53]
D15334	*CPH2*	Transcription factor that positively controls filamentous growth.	
D01463	*CRZ1*	*CRZ1* transcription regulator involved in the regulation of calcium ion homeostasis.	
D13068	*BUD20*	*BUD20* bud site selection protein 20—positioning the proximal bud pole signal; protein required for ribosome assembly.	
B13354	*AP-4*	Transcription factor that activates viral and cellular genes.	
F25861	*RPN4*	*RPN4* transcription factor regulating proteasomal genes.	
D24167	*CBF1*	*CBF1* centromere binding factor 1, required for chromosome stability and chromosomal segregation.	
C12639	*SWI6*	*SWI6* part of a complex involved in cell-cycle-dependent transcription. *SWI4* and *SWI6* are required for formation of the cell-cycle box factor-DNA complex	
E25960	*SWI1*	*SWI/SNF* chromatin-remodeling complex subunit *SWI1*.	
F13321	*OAF1*	Oleate activated transcription factor 1, activates transcription of genes involved in fatty acid beta-oxidation.	
F11487	*STP1*	Nutrient- and stress-responsive activator of ribosome biogenesis genes.	
B05478	*STP3*	Possibly involved in pre-tRNA splicing and in uptake of branched-chain amino acids.	[59]
E15510		Putative [a]: Homeobox protein *YOX1*.	
C22990		Putative [a]: *ASG1* general activator of stress genes.	

Table 1. *Cont.*

Yali Signature	Assigned Name	Putative/Known Function	Reference
D04466		Putative [a]: Regulatory protein cys-3. Turns on the expression of structural genes which encode sulfur-catabolic enzymes.	
D09647		Putative [a]: Arginine metabolism regulation protein II. With *ARG80*, *ARG82* and *MCM1*, coordinates the expression of arginine anabolic and catabolic genes in response to arginine.	
D10681		Putative [a]: Adhesion and hyphal regulator 1.	
D18678		Putative [a]: Respiration factor 2. Transcription factor that regulates expression genes required for glycerol-based growth and respiration.	
E17721		Putative [a]: Phosphatidylinositol N-acetylglucosaminyltransferase subunit P.	
E31669		Putative [a]: Metal-binding activator 1. Copper ion-sensing transcription factor, promotes filamentous and invasive growth.	
F05126		Putative [a]: Phosphorus acquisition-controlling protein.	
F17468		Putative [a]: Multidrug resistance regulator 1. Acts as the central regulator of the *MDR1* efflux pump.	
E11693		Putative [a]: TF required for repression of genes during iron starvation. Represses iron-dependent and mitochondrial-localized activities including respiration, TCA cycle, amino acid metabolism, iron-sulfur-cluster and heme biosynthesis.	
E10131		Putative [a]: Transcription elongation factor 1. Implicated in the maintenance of proper chromatin structure in actively transcribed regions.	
D09757		Putative [a]: AP-1-like transcription factor.	
F05346		Putative [a]: Binds a palindromic promoter element essential for induction of fungal cutinase gene.	
C11858		Putative [a]: DNA-binding transcription factor Moc3.	
B15312		Putative [a]: May act as a co-chaperone for *HSP70*.	
C18667		Putative [a]: Transcriptional activator of the arabinanolytic system.	
E16973		Putative [a]: Metallothionein expression activator.	
F11979		Putative [a]: Retrograde regulation protein 1.	
C19063		Putative [a]: Transcription factor that mediates stress and developmental response.	
F18326, *E05577*, *B08360*, *C20977*, *C13794*		Putative [a]: sterol uptake control.	

[a] assignment of genes' names and genes' functions was manually curated by cross-referencing several databases—anther, KEGG, UniProt, and NCBI and blasting (blastp) the sequences against the databases' collections; The following TFs have no assigned function in the databases mentioned above: *A10637*, *A12925*, *B00660*, *B20944*, *C01375*, *C03564*, *C05995*, *C09009*, *C09482*, *C13178*, *C15202*, *C16390*, *D01353*, *D01419*, *D06952*, *E17215*, *E18161*, *E18656*, *E24277*, *E31383*, *F06072*, *F13695*, *F15037*, *D09757*, *B04510*, *F22649*, *D17988*, *F03135*, *B22176*, *F11011*, *E19965*, *F15543*, *E14971*, *E20251*, *F13761*, *E34925*, *E23518*, *B13200*, *E10087*, *F16852*, *D02475*, *D06193*. The references are given for functional studies in *Y. lipolytica*.

Based on the overall performance profile comprising both the transcriptomics and functional data, the set of TFs was clustered into 10 clusters (Figure 1; the optimal number of clusters was defined using an elbow method; Figure S1).

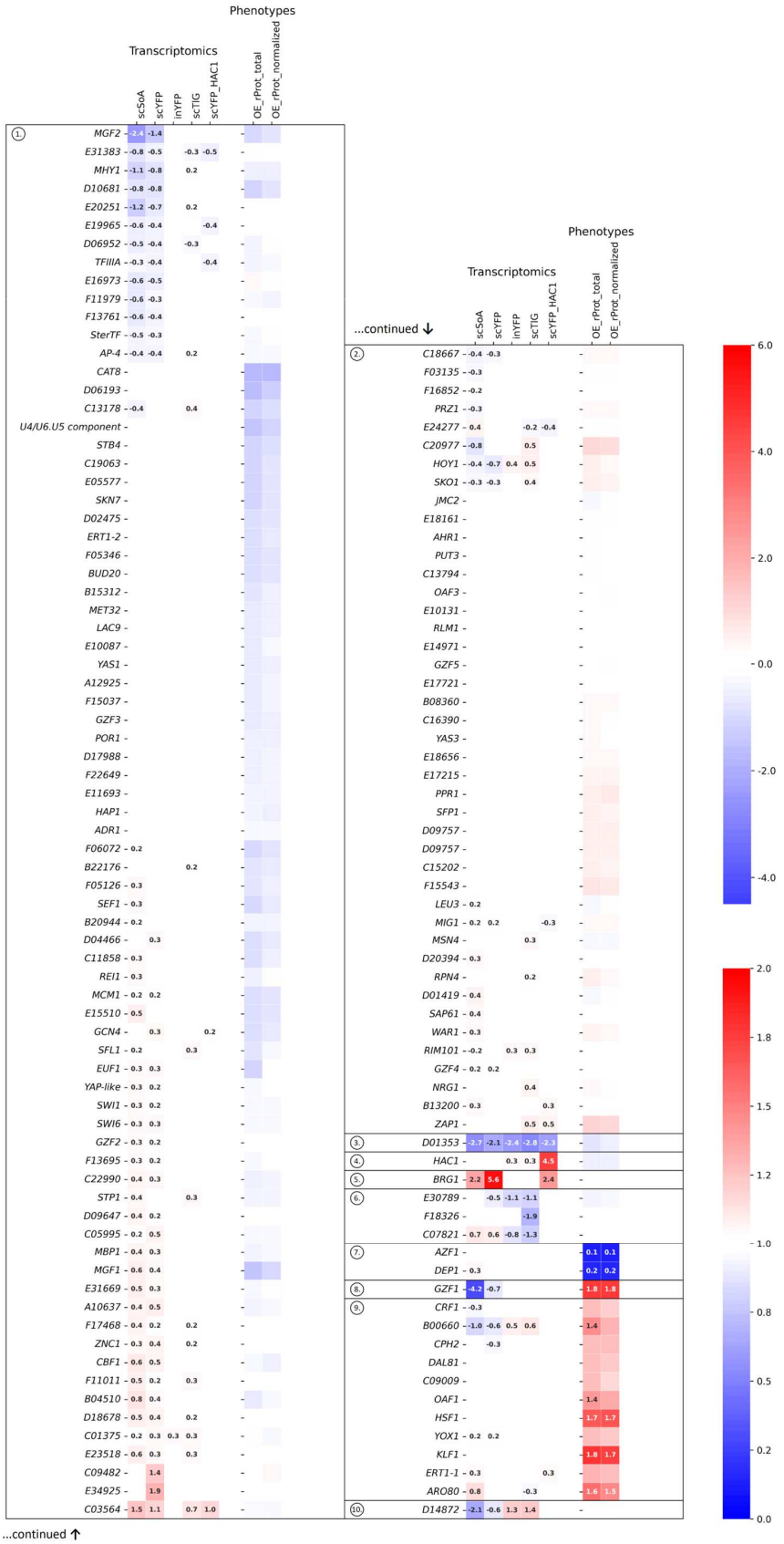


Figure 1. Clustered heatmap of transcriptomics (five sets 'transcriptomics') and functional (two sets 'phenotypes') data for the 140 TFs analyzed. Each row represents a single TF-encoding gene. Data are color-coded according to the legends on the right. Top legend—transcriptomics ('0' and white denotes

no change); bottom legend—phenotype ('1' and white denotes no change). Numbers in the cells denote the fold change value over the control strain. The lack of numbers indicates a lack of statistical significance. Ten clusters were delimited. In the figure, subsequent clusters are separated by a bold horizontal line. The heatmap was split between clusters 1 and 2 for clarity of presentation.

As can be read from Figure 1, cluster 1 is the most abundant and contains the most variable set of TFs. It comprises 76 genes exhibiting moderate up-/down-regulation (especially in the HSS strains) or an unchanged transcriptomic profile upon co-OE of any rProt. Their OE leads to an insignificant but generally negative impact on the rProt production phenotype. This cluster comprises several TFs whose function is well-described, and their role in rProt synthesis has been confirmed. For example, Skn7 (*D14520*) was proved to play a role in maintaining the rProt synthesis capacity under osmotic stress infliction [22]. Here, no transcriptional or functional effects were observed as the conditions were not adequate (no osmotic stress simulation). Cluster 1 is enriched in significantly deregulated TFs of known roles in the dimorphic transition, like *MGF2/B19602*, *MHY1/B21582*, *MGF1/D01573*, and *ZNC1/B05038* [47,53,64,65]. Their OE concertedly contributed to a slight decrease in the rProt synthesis, though they displayed an opposite deregulation pattern (two former were downregulated and the two latter upregulated in the HSS strains). Two representatives of the Gzf-family (genuine GATA—binding zinc finger family) known to be involved in nitrogen catabolite repression (NCR) were assigned to this cluster as well (*GZF2/F17886*—essential for growth on simple nitrogen sources, inducer of NCR; and *GZF3/C22682*—known repressor of NCR; [42]). But the other members of the Gzf family were clustered in either cluster 2 (*GZF4/E05555*—strongly upregulated under inorganic nitrogen [42], hence—putative inducer of NCR; and *GZF5/E16577*—function not investigated) or a self-contained cluster 8 (GZF1/D20482—known inducer of NCR; discussed hereafter). *GZF2* and *GZF4*, although clustered separately, displayed a similar pattern of deregulation (upregulation in HSS strains); likewise, the *GZF3* and *GZF5* genes; however, minor changes to the OE effect contributed to their separation.

Of note, several other global regulators of carbon/nitrogen metabolism were assigned to cluster 1, i.e., *ERT1-2/E03410*, *GCN4/E27742*, *ADR1/F21923*, *CAT8/C19151*, and *E10087*. Strikingly, none of these genes displayed any specific deregulation pattern in the analyzed transcriptomics data, but their OE contributed universally to the inhibition of rProt synthesis (Figures 1 and 2). Furthermore, several of these genes were assigned to a separate functional category upon running a statistical overrepresentation of GO terms test. Their specific function and involvement in rProt synthesis are discussed in Section 2.2. Cluster 1 also contained 4 genes (*HAP1/F17424*, *SKN7/D14520*, *SFL1/D04785*, *EUF1/F01562*) that were either slightly upregulated in the HSS strains or showed insignificant changes in transcriptomic data. OE of these 4 genes uniformly led to a slight, insignificant decrease in the rProt synthesis capacity. Due to their previously documented importance to *Y. lipolytica* phenotype modulation [22,23], they will be subjected to further functional studies (Section 2.4).

Delimited cluster 2 is the second largest cluster, containing 43 TF-encoding genes with low/lack of transcriptional deregulation and a functional phenotype similar to the control strain but with a tendency towards an increase in rProt synthesis upon OE. That cluster contains the two previously mentioned members of the Gzf family involved in the nitrogen metabolism regulation, as well as TFs of known involvement in stress response, like *SKO1/C16863*, *YAS3/C14784*, *MSN4/C13750*, *RIM101/B13640*, or a recently described repressor of erythritol utilization, *NRG1/C12364* [50]. Consistently, all the 'stress response' TFs were upregulated transcriptionally in at least one UPR strain, which was shown to face severe stress. Another interesting gene from this cluster is *ZAP1/D23749* (Zinc-responsive transcriptional regulator 1), involved in zinc ion homeostasis by zinc-responsive transcriptional regulation. It displayed upregulation in the UPR strains, and its OE led to an increase in the rProt synthesis (FC 1.19), which makes it a candidate for a 'helper gene' worth further functional studies.

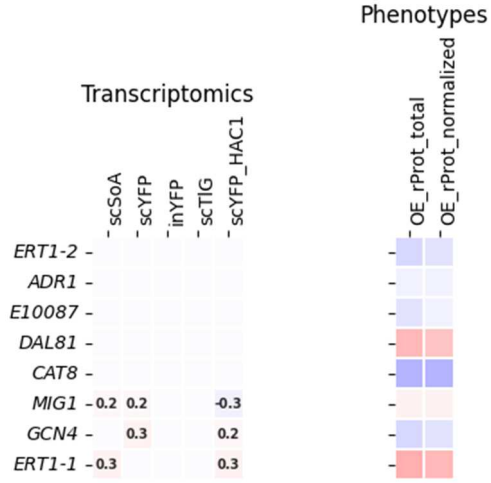


Figure 2. A subgroup of TFs assigned to a 'cellular response to nutrient levels' category. Data color-coding and mode of presentation are according to Figure 1, and a legend is presented there.

A single-gene cluster 3 is represented by the *D01353* gene, which has a unique expression profile. The gene was strongly downregulated upon any rProt OE. Contrary to expectations, its OE had no significant impact on the rProt phenotype. Another single-gene-contained cluster, cluster 4, containing the *HAC1/B12716* gene, was delimited mostly because of its strong upregulation in the scYFP-HAC1 strain, induced synthetically [19]. Due to the peculiar, unexpected outcome (lack of induction of rProts synthesis upon OE), this TF was subjected to further functional studies by investigating the Δ*hac1* phenotype (Section 2.4). Cluster 5 also contains only 1 gene (*BRG1/E31757*) displaying strong upregulation in HSS and scYFP-HAC1 strains, with no significant change in OE phenotype data. Knowing its role in the synthesis of lipids in *Y. lipolytica* [20], it is tempting to state that its expression was enhanced in the HSS strains in response to high demands for membranes intensively exploited in the vesicular transportation of rProts. Representatives of cluster 6 (*E30789*, *F18326*, and *C07821*) showed strong downregulation in the UPR strains but no change in the OE phenotype. TFs assigned to cluster 7, *DEP1/F05896* and *AZF1/A16841*, were characterized by minimal changes in transcriptional profile, but both contributed to a distinctive, very strong limitation of rProt synthesis upon their OE. These two were also selected as the candidates for studying Δ*azf1* and Δ*dep1* phenotypes with high expectations (Section 2.4).

Cluster 8 contained only 1 gene (*GZF1/D20482*), which exhibits a unique performance, namely, strong downregulation in HSS strains and high over-production of rProt when OE. Gzf1 is a member of the discussed above Gzf-family, acting as an activator of NCR genes. It was previously identified as one of the universal enhancers of rProt synthesis [22]. Considering its putative role as an NCR activator, such an effect could be attributed to the enhanced nitrogen scavenging capacity required for the high-level production of rProt. But then its significant downregulation in the HSS strains is not clear, as those strains definitely encountered nitrogen limitation due to their extensive consumption for rProt synthesis. A previous report detailing the impact of Gzf family members on growth and lipid synthesis by *Y. lipolytica* cultivated in the presence of different nitrogen sources sheds some light on the structure of the Gzf TRN [42]. Primarily, they found that expression of *GZF1*, *GZF2*, *GZF4*, and *GZF5* was increased when the cells were grown on ammonium as the sole nitrogen source, while the level of expression of *GZF3* was not significantly affected. Translating into our data, their expression should be upregulated in the HSS, facing nitrogen depletion. It was true only for the *GZF2* and *GZF4* genes, while *GZF5* escaped this mechanism. The highest amplitude of the upregulation due to nitrogen source type was previously noted for the *GZF1*. In our transcriptomics, *GZF1* displayed a tremendous drop in expression in the HSS strains, which is in contrast to our expectations. Interestingly, [42] showed that when *GZF1* was deleted (Δ*gzf1*), no growth defect on any of

the nitrogen sources tested was displayed, suggesting that its function was taken over by the other gene. This could be performed by either *GZF2* or *GZF4*, whose expression was significantly upregulated in the HSS strains. It was also discovered that Δgzf3 (repressor of NCR) led to the loss of regulatability of *GZF1* by the nitrogen source, i.e., *GZF1* expression was elevated irrespectively of the nitrogen provided. The opposite was observed in the Δ*gzf2* background. In our data, *GZF3* was not deregulated under any conditions, which is consistent with the findings by [42]. Though functional *GZF3* and *GZF2* were present in the cell. Hence, *GZF1*, which is of key interest to us as the potent inducer of rProt synthesis, could be regulated by the nitrogen availability. Considering increased demands for nitrogen and the upregulation of *GZF2*, we expected upregulation of *GZF1* expression in HSS strains. The question of why *GZF1* was downregulated in the HSS strains hence remains to be answered. The effect of OE and Δ*gzf1* genotypes on rProt synthesis in *Y. lipolytica* in a direct comparative experiment was investigated and is presented hereafter (Section 2.4).

Representatives of cluster 9 in general exhibited no significant deregulation in transcriptomic data, but their OE led to strong overproduction of rProt. This group of TFs is definitely of the highest practical interest, gathering candidate enhancers of rProts synthesis. Indeed, 2 genes, *KLF1/D05041* and *HSF1/E13948*, were previously reported as global enhancers of rProt synthesis [22,23]. Interestingly, a similar level of rProt synthesis enhancement was observed for TF *ARO80/C18645*, involved in the catabolism of aromatic amino acids [66]. Likewise, *DAL81/D02783,* known for its role in nitrogen turnover, was expected to bring a positive effect to rProt synthesis. It is thus concluded that the positive impact triggered by the overrepresentation of Aro80 and Dal81 is associated with the increased supply of the rProt building blocks. By similarity to *KLF1*'s function, the beneficial effect of *CRF1/B08206* OE may be attributed to its involvement in oxidative stress response.

An interesting expression profile was observed for the gene *D14872* assigned to a self-contained cluster 10. The gene showed a strong and inverted deregulation trend in the HSS (downregulated) and UPR (upregulated) strains, however, yielding no significant impact on phenotype upon OE. Genes displaying inverted deregulation patterns were deemed to play a direct role in the analyzed biological process (rProt synthesis), which turned out to be not the case, at least for *D14872*. They will be combined and analyzed separately in Section 2.3.

2.2. Statistical Overrepresentation Test Delimited Two Major Categories amongst TFs

The total set of 140 TFs was subjected to a statistical overrepresentation of biological processes test [67,68]. Expectedly the majority of output records were assigned to 'regulation of DNA-templated transcription' or 'regulation of RNA biosynthetic process' categories, or related (altogether 31 categories and subcategories comprising a variable number of genes, with a significant enrichment fold from >6 to >17; Table S1).

Interestingly, six genes were assigned to a 'cellular response to nutrient levels' category (altogether 8 categories and subcategories comprising a variable number of genes, with a significant enrichment fold from >6 to >13; Table S1). These were transcriptional activators of gluconeogenesis *ERT1-1/2* (cluster 9 *E18304*/ cluster 1, *E03410*), *DAL81* (*D02783*; cluster 9), *GCN4* (*E27742*; cluster 1), a *MIG1*-related regulatory protein (*E07942*; cluster 2), and unknown TF *E10087* (cluster 1; Figure 2). It was the only significantly enriched category, other than related to DNA transcription/RNA biogenesis.

The majority of TFs assigned to that category showed no transcriptional deregulation due to rProt OE. Also, the induced phenotypical changes were rather tendencies. Though the patterns behind them are interesting to track. *ERT1* acts as a positive regulator of gluconeogenesis and fermentable carbon utilization but mechanically is a repressor of transcription by RNA polymerase II by a nonfermentable carbon source. An increased abundance of *ERT1-1* (*E18304*) was recorded in transcriptomes of scSoA and scYFP-HAC1; when overexpressed (OEd), it triggered enhanced rProt synthesis. The other homolog, *ERT1-2* (*E03410*), did not show any changes at the transcriptional level in the analyzed

transcriptomes, and its OE led to a slight decrease in rProt amounts. Another gene from this category, *DAL81/D02783*, is a positive regulator of genes in multiple nitrogen degradation pathways, involved in nitrogen catabolite activation of transcription from the RNA polymerase II promoter. Even though previously we reported on enhanced protein degradation and deregulation of nitrogen catabolism processes in the UPR strains [16], and enhanced demand for nitrogenous compounds was expected for the HSS strains, its expression level was not changed. However, functional screens revealed its minor but positive effect on rProts synthesis upon OE, which is biologically well-understood. The lack of transcriptional response but a visible effect at the functional level suggests its regulation at the posttranslational level. Another TF from this functional category, *MIG1/E07942*, is known to be involved in glucose repression by negative regulation of transcription by RNA polymerase II by glucose. Mig1 (together with Mig2) also induces filamentation under glucose starvation stress. The steady-state maintained cells did not face the main carbon source starvation, as its residual levels were very high. *MIG1* expression was only slightly upregulated in the HSS strains, expected to be in very high demand for carbon building blocks and energy. Its OE led to a minor positive effect on rProts synthesis by *Y. lipolytica*. However, due to its known fundamental role in shaping cellular metabolism, the Δ*mig1* genotype was constructed, and its effect on rProt synthesis was tested (Section 2.4).

All the remaining TFs assigned to the 'cellular response to nutrient levels' category (based on enrichment tests or added manually—*CAT8* and *ADR1*) collectively contributed to slightly decreased rProts synthesis upon their OE. Among them, *GCN4* (*E27742*) is a master regulator involved in multiple biological processes, with a primary role in TRN controlling amino acid metabolism. It was also shown to be required for increased biosynthesis of translation precursors such as ribosomal proteins, amino acids, and purines, depending on the external stimuli. Upon amino acid starvation, it plays a key role in the transcriptional induction of almost all genes involved in amino acid biosynthesis (19 per 20 pathways). Notably, it was reported that OE of *GCN4* leads to a reduction in protein synthesis capacity via negative regulation of ribosomal protein gene transcription [48]. Consistently with its assigned role, its OE triggered a slight decrease in the rProt synthesis levels in *Y. lipolytica*. It is known that its levels are strictly regulated at translational/protein and mRNA stability levels (it is constitutively expressed at a low basal level), so the generally observed lack of transcriptional deregulation upon rProts overproduction is not surprising, though the inflicted conditions seem to be relevant stimuli.

TFs *CAT8/C19151* and *ADR1/F21923* are known to be involved in the regulation of carbon catabolite repression, but in the opposite direction than Mig1. They share the common lack of deregulation upon any rProt OE and the negative effect on rProt synthesis upon OE. Cat8 is an inducer of gluconeogenesis, binding to cis-regulatory elements upon glucose starvation. Adr1 induces transcription of genes involved in alternative carbon utilization upon glucose starvation. Its fundamental role in controlling the Euf1-dependent erythritol utilization cluster in *Y. lipolytica* has been recently reported [11].

Altogether, the above-discussed transcriptional-functional profiles of the genes assigned to a specific functional category 'cellular response to nutrient levels' suggest that enhanced degradation of nitrogenous compounds is beneficial for rProts synthesis (Dal81), but enhanced signaling of nitrogen starvation is not (Gcn4). Likewise, the induction of alternative carbon utilization is detrimental.

2.3. Usefulness of Specific Deregulation Patterns as Selectors of TFs Involved in Specific Biological Processes

To gain more insight into the relationship between the transcriptional profile and the functional data, the dataset was further filtered to extract and group together TFs that: (i) exhibited inverted transcriptional deregulation patterns in HSS and UPR strains (Figure 3); (ii) showed a uniform deregulation profile, irrespective of the inflicted perturbation (type of rProt) (Figure 4). The phenotypic effect of their OE is shown together with their expression profile in Figures 3 and 4.

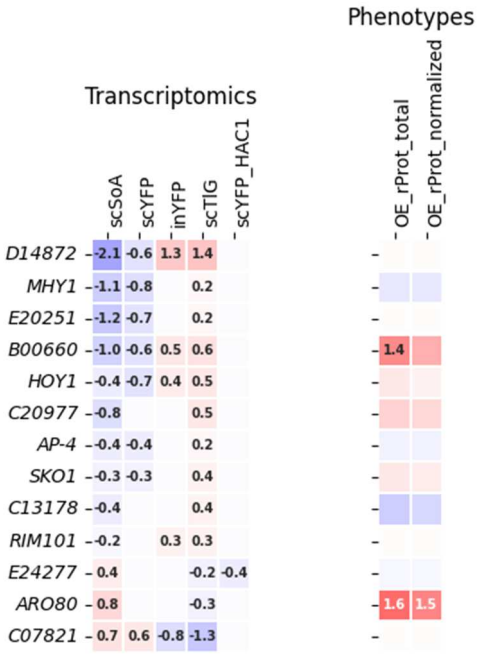


Figure 3. A subgroup of TFs displaying an inverted deregulation pattern in the transcriptomic data of HSS and UPR strains. Data color-coding and mode of presentation are according to Figure 1 and a legend presented there.

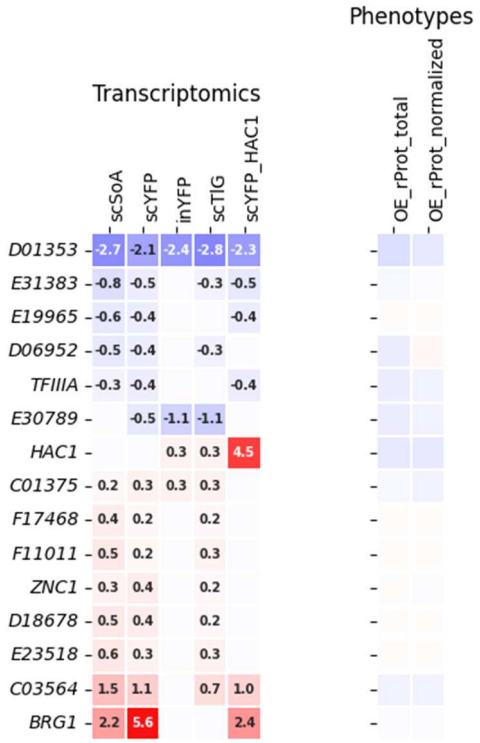


Figure 4. A subgroup of TFs displaying a uniform deregulation pattern in the transcriptomic data (at least 3 out of 5 datasets). Data color-coding and mode of presentation are according to Figure 1, and a legend is presented there. The correlation matrix for these data is shown in Figure S2.

The inverted deregulation pattern of the TFs (grouped in Figure 3) suggests that they are directly transcriptionally responsive to the conditions faced and differing the two groups of strains ('HSS' and 'UPR'). Considering the previous results, we presumed that the amounts of rProts and the associated background biological processes (oxidative

stress and UPR) were the delimiting factors [16,19], ergo, these TFs are directly involved in these processes. Indeed, *HOY1/A18469*, *MHY1/B21582*, and *SKO1/C16863* are known for their role in stress response, and these were consistently upregulated in the UPR strains. In this context, both the transcriptional response and functional effect on rProts synthesis upon OE of TF *B00660* are very interesting, considering its putative role in sterol uptake [20]. As evidenced [61], the enhanced sterols residence in the plasma membrane is important for enhanced stress resistance. The here presented dataset suggests that the overrepresentation of TF *B00660* (either natively due to UPR or synthetically by OE) contributes to generally enhanced rProts synthesis, presumably associated with enhanced stress resistance. Unfortunately, many of the remaining genes exhibiting such inverted deregulation patterns are of unknown function, and hence the mechanism behind the phenotypes cannot be deduced.

Generally, two distinct expression profiles could be distinguished for the TFs grouped in Figure 3, those upregulated in the UPR-encountering strains (inYFP, TlG, scYFP-HAC1) and downregulated in HSS (scSoA, scYFP) (majority of the examples), and those upregulated in the former and downregulated in the latter (three TFs: *E24277*, *C18645*, and *C07821*). If the expression profile could be directly translated into the functional outcomes, then two distinct functional behaviors would be expected. However, that was not the case. For *D14872*, exhibiting significant downregulation in the HSS and significant upregulation in the UPR-facing strains, no functional effect of its OE on rProt synthesis was observed. On the other hand, TFs *B00660* and *HOY1*, while showing the same transcriptional pattern, their OE triggered enhanced rProt synthesis in the host strain (significant or just a tendency). Corresponding enhancement in rProt synthesis was observed for TF *C18645*, whose expression was significantly enhanced in HSS and decreased in the UPR strains. But foremost, the majority of TFs for which the inverted deregulation pattern was observed (so its direct implication in the rProt synthesis was presumed) did not render any significant functional phenotype when OE (only tendencies; Figure 3). Such an outcome implies that our working hypothesis stating that transcriptomics data for the TFs displaying inverted deregulation patterns could be used as a selection-driver of TF involved in a specific biological process, here—rProt synthesis in *Y. lipolytica*, should be partly rejected (for TFs displaying such an expression profile). Furthermore, considering that the TFs exhibiting an inverted deregulation pattern exert similar functional outcomes when OE (*B00660* and *C18645*) implies that the transcriptomics pattern cannot be used as a preliminary prognostic of the TF's mode of implication (inducer/repressor) in the biological process under study.

Another dataset prepared to test our working hypothesis comprised TFs that were responsive to enhanced synthesis of any rProt, irrespective of the biochemical character of the rProt and the background biological processes awakened (Figure 4). These comprised TFs globally up- or down-regulated across all the transcriptomes tested (in at least 3 out of 5). As can be read from Figure 4, global responsiveness to the over-synthesis of rProt is not an optimal prognostic of direct involvement of a TF in rProt synthesis. Indeed, the majority of TFs, when OE, did not exert any significant effect on rProt synthesis. Nevertheless, some significant correlation between the TFs transcriptomics profile and the rProt amounts upon TF's OE was noted (scSoA, r = 0.6, inYFP, r = 0.56, scTlG, r = 0.63; significant at $p < 0.05$; Figure S2). It suggests that the TFs that were downregulated in the transcriptomes contribute to a decrease in the rProts synthesis when OEd, and that TFs upregulated promoted rProt synthesis when OEd; so such data could be used as a careful hypothesis driver on the implication of a given TF in the biological process under study (rProt synthesis in this case).

2.4. Direct Comparison of Phenotypes Elicited by Y. lipolytica Strains Bearing OE or KO of Selected TFs

In the next step, 11 TFs representing different clusters from Figure 1 were further investigated through direct comparative phenotype reading of *Y. lipolytica* strains bearing their OE and KO, along with rProt OE. Altogether 11 TFs were studied using this

approach, selected based on the observed transcriptional-functional profiles (Figure 1) and literature data. These were representatives of the most abundant cluster 1 (*HAP1/F17424*, *SKN7/D14520*, *SFL1/D04785*, *EUF1/F01562*), cluster 2 (*MIG1/E07942*), cluster 4 (*HAC1/B12716*), cluster 7 (*DEP1/F05896*, *AZF1/A16841*), cluster 8 (*GZF1/D20482*), and cluster 9 (*KLF1/D05041*, *HSF1/E13948*). After genotype construction, selected subclones were cultivated under 'standard conditions' (no stress factor inflicted; Sections 4.1 and 4.2). Results on the total rProt amounts, or normalized per biomass, are shown in Figure 5.

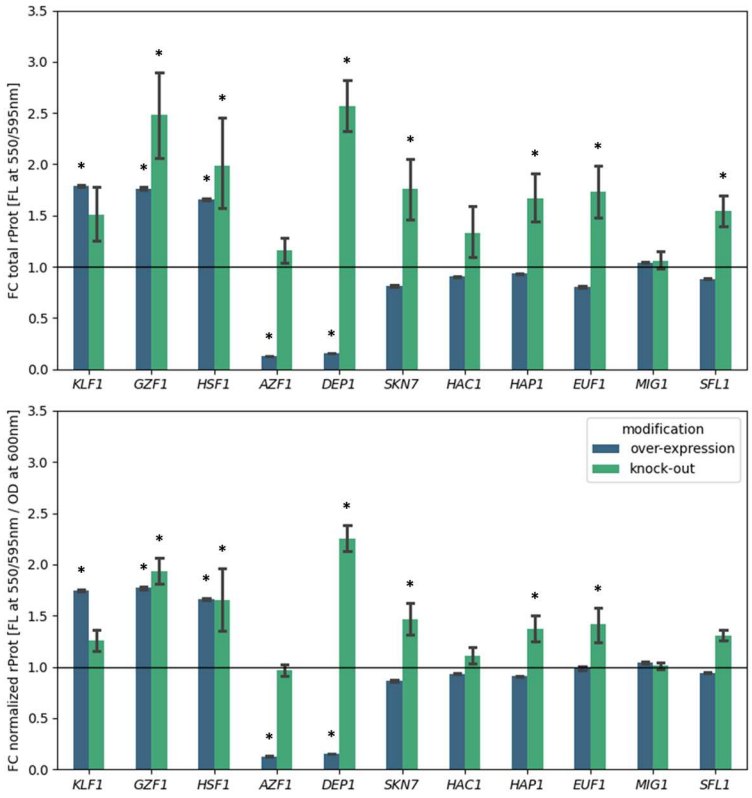


Figure 5. Change in the rProt synthesis capacity investigated in direct comparative functional studies of strains bearing OE (navy) and KO (green) of a given TF. Top chart–fold change (FC) in the total amounts of rProt synthesized by the engineered strains over the control strain, read as fluorescence from the reporter protein. Bottom chart–fold change (FC) in the total amounts of rProt synthesized by the engineered strains over the control strain, read as fluorescence from the reporter protein and normalized per accumulated biomass (read as absorbance at 600 nm). The horizontal line at the level of 1.0 indicates the level of rProt by the control strain. Bars indicate mean values from 2 to 6 subclones cultivated in technical duplicate ± SD. Asterisks (*) indicate a statistically significant difference between the modified strain and the control strain at $p < 0.05$.

Based on the transcriptional-functional performance patterns (Figure 1), the representatives of cluster 1 (*HAP1/F17424*, *SKN7/D14520*, *SFL1/D04785*, *EUF1/F01562*) and cluster 2 (*MIG1/E07942*) were subjected to further studies with rather low expectations (no transcriptional deregulation, no effect of the TF's OE on rProt synthesis). Their selection was based solely on literature data and their important role in shaping the *Y. lipolytica* phenotype [11,22,23,40,49–51,57,69–71]. Strikingly, the deletion of TFs from cluster 1 led to moderate (55 to 76%) but significant ($p < 0.05$) improvement in the rProt synthesis. Suggesting that the observed tendencies in the functional screens (Figure 1) are indeed valid observations. No significant effect (nor even a tendency) was observed when *MIG1* was either OE or KO (also selected based on literature data and not experimental evidence provided here [57,72,73]). It is presumed that the OE/KO *MIG1* strains should be tested under a spectrum of different nutrient types and levels to observe a valid phenotype.

Hap1 is known to play a fundamental role in managing aerobic metabolism in *S. cerevisiae*. It is responsible for sensing the oxygen levels via the heme signaling pathway, and activation of the oxidative stress response genes [49]. Notably, its OE had a beneficial effect on rProt synthesis in *S. cerevisiae* [74]. It was suggested that the effect was caused by diminishing oxidative stress responses awakened by intensive rProt synthesis. The contradictory effects observed here with *Y. lipolytica* may result from a different metabolism of the two species regarding oxygen demands.

Skn7 is a TF involved in protein secretion and activation of oxidative and osmotic stress responses [75,76]. Previously, its KO in *Y. lipolytica* had a detrimental impact on the strains' resistance to osmotic stress, irrespective of inflicted anaerobiosis stress, pH, or adopted temperature, leading to a severe underperformance in terms of rProt synthesis once osmotic stress was inflicted [22]. On the other hand, its OE enabled the maintenance of rProt synthesis under high osmolarity and low aeration [22]. Under the current experimental setup (no stress factor, high aeration), the Δ*skn7* phenotype performs better than the control in terms of rProt synthesis. We presume that the adopted conditions contributed to that outcome. A corresponding effect was observed when TF *SFL1/D04785* was deleted (Figure 5). Sfl1 (Suppressor Gene for Flocculation) acts as a repressor of filamentous growth and flocculation, antagonizing *FLO* gene action; it is also known to activate the expression of stress-responsive genes [69–71]. The beneficial effect of Δ*euf1* is also not yet clear. Euf1 is a master regulator of the 'erythritol utilization cluster' [11,50,51] known to govern the expression of several genes involved in erythritol utilization once glucose or glycerol are depleted. Considering the effects observed after KO of TFs from cluster 1, the only explanation for the moment is a slight relief of the transcriptional–translational machinery due to the limited expression of several genes or inadequate conditions enabling a relevant phenotype expression. Collectively, it is tempting to state that once stress conditions are not inflicted, deletion of the global stress-response regulator is beneficial for the costly biological process—rProt synthesis.

The observed Δ*hac1* and *HAC1*-OE phenotypes escaped our expectations, yielding no significant changes. Hac1 mediates the deregulation of hundreds of genes involved in polypeptide formation, folding, and maturation, but also in lipid synthesis, membrane expansion, and many others [17,18,77–79]. Hence, along with Hsf1, it is the most frequent target of genetic modifications for enhanced rProt synthesis in yeast [15,19,80–85]. As presented in Figure 1, the transcriptional profile of *HAC1* depends on the type of rProt being synthesized due to the background molecular mechanisms. Generally, previously conducted estimations of the rProt amounts synthesized when *HAC1*'s expression was elevated or not showed that the yields of rProts were higher when *HAC1* expression was silenced [14,16]. In the following studies, we attempted to explain this observation. Investigation of Hac1 regulome [19] indicated that elevated Hac1 presence induces massive vacuolar proteolysis. In contrast, in other experiments, we observed a ~30% increase in the rProt synthesis (small, non-modified post-translationally protein, targeted for secretion; scYFP) [15]. The following studies, conducted with cells maintained in the steady state, proved that the effect was indeed bifurcated, causing a nearly 7-fold drop in the retained fraction of rProt but promoting its secretion by nearly 2.5-fold [19]. The protein used in the functional screens (results presented in Figures 1 and 5) has a similar biochemical characteristic to that former but is destined for intracellular retention. We presume that it is the cause of the surprising lack of effect from Δ*hac1* and *HAC1*-OE strains.

Comparable functional outcomes regarding rProt amounts were expected from strains deleted for *AZF1/A16841* and *DEP1/F05896* (cluster 7), as both the OE strains displayed a significant decrease in rProt synthesis capacity (Figure 1). However, a directly inverted phenotype was observed solely for the *DEP1*-deletant (Δ*dep1*), with no effect from the Δ*azf1* strain (at least in terms of rProt synthesis) (Figure 5). Notably, the Δdep1 rendered the highest increase in total rProt or normalized rProt measures, making this modification the most successful example presented here. Dep1 (Deregulated Expression of Phospholipid biosynthesis) was previously found to enhance the accumulation of lipids in *Y. lipolytica* [20]

and specifically—activate phospholipid biosynthesis in *Fusarium* sp. [26]. In contrast, [86] identified Dep1 as a repressor of phospholipid synthesis genes (e.g., *INO1*, *CHO1*, and *OPI3*) in *S. cerevisiae*. It seems that Dep1 from *Y. lipolytica* operates in a similar way as in Fusarium rather than the model yeast species. To our interpretation, an interplay between globally promoted carbon metabolism and enhanced nitrogen scavenging (like in the case of Gzf1/Gzf2+Gzf4 discussed above) may be the molecular mechanism behind the Δ*dep1* phenotype.

For the second representative of cluster 7, *AZF1/A16841* (asparagine-rich zinc finger protein), such a directly inverted phenotype regarding rProt synthesis was not observed. Azf1 is a carbohydrate-sensing TF, and in the presence of glucose, it activates genes involved in growth, carbon metabolism, and filamentation in *S. cerevisiae* and *O. polymorpha* [25,28]. In *Y. lipolytica,* the Δ*azf1* mutation rendered no aberrant phenotype in terms of rProt synthesis. This observation is consistent with our previous notions about this phenotype investigated under different cultivation conditions [16,19]. As in the case of Mig1, we presume that alternating carbon sources would enable a relevant phenotype display (but maybe not necessarily related to rProt synthesis).

An unexpected effect was observed upon direct comparison of the strains with either OE or KO of the 'universal rProt synthesis enhancers', *KLF1/D05041*, *GZF1/D20482*, and *HSF1/E13948*. Surprisingly, the rProts synthesis capacity was enhanced regardless of whether the genes were OE or KO. Gzf1 was selected for further phenotype studies due to its surprising expression pattern in the HSS strains and highly positive effect on rProt synthesis. Considering the data presented in Figure 1 (decreased expression in the HSS producing a high amount of rProts and enhanced production of rProts when *GZF1* was OE), the data presented in Figure 5 stay in peculiar agreement—if *GZF1* is downregulated (HSS strains), rProts synthesis is high, and when *GZF1* is upregulated (OE)—rProt synthesis is also high. Previous analysis of OE/KO *GZF1* phenotype under various environmental stress factors infliction [22] demonstrated that OE of *GZF1* led to very high and universal enhancement in rProt synthesis. The effect of *GZF1* KO was less uniform—under specific conditions, growth was limited, but some indications for enhanced rProt synthesis capacity were revealed, but only under specific conditions not applied here (low oxygen and hyperosmolality) [22,23]. The primary difference between those previous and current studies relates to cultivation parameters (type of vessel, aeration, buffering system), and foremost—the type of nitrogen source used, which is a relevant factor affecting *GZF1* activity (previously—ammonium sulfate; at present—a 1:1 mixture of glutamic acid and ammonium sulfate). Since that former study, our protocol for *Y. lipolytica* has been greatly improved [31]. We presume that this new cultivation system allows for the actual genotype-driven phenotype development, and the type of nitrogen source used accounts for the observed discrepancies in the Δ*gzf1* phenotype.

Representatives of cluster 9, *KLF1/D05041* and *HSF1/E13948*, did not display any significant transcriptional deregulation profile in the analyzed transcriptomes, but their OE led to significant enhancement in rProts synthesis capacity (Figure 1). When deleted, Δ*klf1* showed no changes vs. the control strain in terms of rProts production. The same effect was observed previously, using a slightly different cultivation protocol [23]. Klf1 was shown to be responsive to external nitrogen levels in *Schizosaccharomyces pombe*, and its regulome was implicated in cell wall renewal, oxidative stress response, glycolysis, nutrient uptake, RNA-mediated chromatin silencing, glycosidation, and methylation [87]. It was demonstrated to specifically react to oxidative stress via relocalization to the nucleus [88]. That would explain why the rProt synthesis-promoting effect of *KLF1* OE could be seen, even though its expression remained at the control level (Figure 1).

HSF1/E13948, as the key regulator of global stress response, is the most intensively studied TF in terms of its effect on rProt synthesis in yeast. rProts production was significantly enhanced in *S. cerevisiae* upon OE of *HSF1* in its native or a constitutively active form, *HSF1*-R206S [89–91]. Previously, we demonstrated a universal, promoting effect of wild-type *HSF1* OE on rProt synthesis in *Y. lipolytica* [22,23]. Genotype Δ*hsf1* displayed

severely impaired growth under both 'stressful' and 'optimal' growth conditions. The effect was associated with limited production of rProts, but mainly when harsh stress was inflicted. The results here obtained were hence surprising and difficult to explain even with the differences in the growth conditions. Definitely, more studies focused on this aspect must be conducted to get an insight into the molecular mechanisms behind that observation, specifically, whether the other global regulators of stress response were not induced in the Δ*hsf1 Y. lipolytica* strain.

3. Summary and Conclusions

Integration of multidimensional, essentially diverse biological data imposes a challenge, as the biological processes are rarely linear. However, with the current accumulation of biological knowledge, such a data-recycling approach is reasonable and, if followed with adequate methodology, may drive novel hypotheses and interesting findings that otherwise could be missed. The approach has become popular with the model organisms, and now, considering how much information is stored in public repositories, it is feasible with nonconventional species like *Y. lipolytica*.

In our recent studies, we investigated the possibility of using TFs as tools for global optimization of rProts production in *Y. lipolytica*. Here, we recycled and integrated datasets related to the expression profile of TFs and the functional effect of their OE on rProt synthesis. Specifically, we were testing a hypothesis of whether transcriptomics data could be used as a selection driver of TF involved in this particular biological process.

The conducted analyses showed that many of the TFs with previously documented implications in rProt synthesis did not respond at the transcriptional level (e.g., Skn7, Hap1). We presume that it was due to conditions under which the transcriptomics samples were collected, which were not activating a given TF's TRN. This statement is corroborated by the results of further experimental studies. The deletion of TFs from cluster 1 (showing no transcriptional deregulation upon rProt synthesis) led to significant improvement in the rProt synthesis capacity. Likewise, representatives of cluster 7, showing no transcriptional response to rProts synthesis, were shown to strongly decrease this process when OE. Specifically, the deletion of *DEP1* rendered the highest increase in rProt production, demonstrating on one hand its direct implication in this process and on the other—making this modification the most successful example presented here.

The direct correlation between the transcriptional profiles and functional effects of the TF's OE was either not significant or weak. It was found true only for the TFs showing a uniform pattern of deregulation. In that specific case, TFs that were downregulated in the transcriptomes contributed to a decrease in the rProt synthesis when OEd, and TFs upregulated—promoted rProt synthesis when OEd. Hence, such data could be used as a careful hypothesis driver on the implication of a given TF in the biological process under study (rProt synthesis in this case). In contrast, examples of TFs showing different deregulation patterns but the same functional effect (e.g., TFs of known roles in the dimorphic transition, or *B00660* vs. *C07821* or *C18645*) and the opposite behavior (similar transcriptional pattern but different functional effect; e.g., *HOY1* vs. *MHY1*) were predominant. The intrinsic pleiotropic activity of TFs is the key explanation for this observation (e.g., Gcn4, which is an activator and a repressor).

Nevertheless, the presented data integration led to a new hypothesis and pointed to new directions for experimental studies. For example, the juxtaposition of transcriptional profiles and functional OE results highlighted the putative importance of TF *ZAP1/D23749* for rProt synthesis in *Y. lipolytica*. Likewise, *C18645*, a TF of yet undefined function, was clustered together with the most potent rProt enhancers Hsf1 and Klf1, identified previously. A new research question that remains to be answered through further insightful experimentation is why *GZF1*, an efficient enhancer of rProt synthesis, was indeed transcriptionally downregulated in the strains producing rProts in high quantities. Though it is known to be an activator of NCR. The analysis of transcriptional profiles and functional outcomes of the genes involved in nitrogen metabolism in *Y. lipolytica* implied that the

enhanced degradation of nitrogenous compounds is beneficial for rProts synthesis (e.g., Dal81, Aro80), but enhanced signaling of nitrogen starvation is not (Gcn4). Likewise, the induction of alternative carbon utilization is detrimental (e.g., Adr1, Cat8, Ert1-2).

Altogether, this investigation demonstrates the feasibility and potential of biological data recycling and integration for *Y. lipolytica*. It appears that the amount of knowledge accumulated is sufficient, and such an approach leads to novel findings that were not discovered when the datasets were analyzed individually.

4. Materials and Methods

4.1. Microbial Strains Used for Transcriptomics and Functional OE Screens Data Acquisition

Strains used for transcriptomics and functional OE screen data acquisition were constructed previously [14,16,19,20,23,92]. Briefly, the 'transcriptomics' strains are all derivatives of Po1h (Genotype: *MatA, ura3-302, xpr2-322, axp1-2, leu2-270::LEU2*, Phenotype: Ura-, Δ*AEP*, Δ*AXP*, Suc+). The control strain (Po1h_Ura3) was transformed with a solo *URA3* marker cassette to generate a prototroph. Recombinant *Y. lipolytica* strains were transformed with Golden Gate Assembly [93,94] cassettes bearing a single transcription unit each with one target gene: alpha-amylase—SoA, glucoamylase—TlG, fluorescent YFP (intracellular in-/secretory sc-), expressed individually under the control of a synthetic hybrid promoter 4UASpTEF, or co-OEing scYFP and *HAC1* under the control of the pTEF promoter [14,16,19]. Secretory enzymatic reporters were transcriptionally fused with a signal peptide sequence native to exo-1,3-beta-glucanase (*B03564*) [95]. Strains for the 'phenotype' screens were constructed by [20,92]. Briefly, a collection of 125 *Y. lipolytica* strains OEing individually one of the TFs and a reporter protein (RedStar2) was created in the background of the JMY2566 (*MATa, ura3::pTEF-RedStar2-LEU2ex-Zeta, leu2-270, xpr2-322*) strain. Both genes were cloned under a constitutive promoter pTEF and were integrated in a zeta platform at the *URA3* locus. Strain JMY2810 (complete prototroph OEing the reporter r-Prot solely; *MATa, ura3::pTEF-RedStar2-LEU2ex-Zeta-URA3ex-pTEF, leu2-270, xpr2-322*) was used as the control for data normalization and calculating FC values. The methodology of the above indicated strain cultivation strategy, sample acquisition, and data analysis was detailed previously [14,16,23,31,96].

4.2. Microbial Strains Used for Functional Studies—KO in a Specific TF loci

Microbial strains used for functional studies were partly constructed previously (Gzf1, Hsf1, Skn7, Klf1, Azf1, Dep1) [22,23], and partly were constructed originally in this research (Mig1, Euf1, Sfl1, Hap1, Hac1). All the strains were (re)cultured under the improved cultivation protocol [23,31] to gain a direct comparison of the phenotypes.

Construction of the KO Strains—Deletion Cassette and CRISPR-Cas9 Targeting

Standard molecular biology protocols were followed in this study [97]. The deletant strains (KO-TF) were constructed in the background of the JMY2810 strain by disrupting the indicated TF locus. The deletion cassettes were designed using a GoldenGate scaffold indicated in [93] limited to three fragments cloning: (i) ARM up, (ii) NATr (nourseothricin), and (iii) ARM down, flanked with A and B and C and M overhangs. DNA fragments to be cloned were amplified using Phire Hot Start II DNA Polymerase (Thermo Fisher Scientific, Waltham, MA, USA). The cassettes were assembled using a previous protocol for the GoldenGate reaction [93]. White colonies were verified for correctness of the assembly by PCR of adjacent elements and restriction digestion of isolated plasmids. After release from the pSB1A backbone by *Not*I digestion, the cassettes were used for the transformation of *Y. lipolytica* JMY2810.

To increase targeted integration, the CRISPR-Cas9 vector JME4580 [98,99] with the TF-targeting sgRNA oligonucleotide was co-transformed along with the disruption cassette. The methodology followed the previously described protocol [22]. The sgRNA oligonucleotides were designed using the CRISPR design tool from the Benchling platform (https://benchling.com/ (accessed on 31 July 2024Day)). Targeting regions were selected

close to the center of the coding sequences. sgRNA oligonucleotides with the highest efficiency scores and lowest number of off-target sites were selected. The 20 bp long target sequences were flanked with *BsmB*I recognition sites and with 4-bp overhangs, enabling their correct integration in a plasmid [99]. Before cloning, the sgRNA oligonucleotides were annealed and then ligated to the JME4580 plasmid using the GoldenGate thermal profile, using *BsmB*I and T4 DNA ligase from New England Biolabs (NEB Ltd., Ipswich, MA, USA). The reaction was then transformed into *E. coli DH5alpha,* and transformants were selected on LB ampicillin agar plates. White clones were screened for correctness by PCR and restriction digestion with *Bgl*II (Thermo Fisher Scientific, Waltham, MA, USA). Correct constructs were propagated, plasmids were isolated (Plasmid Mini, A&A Biotechnology, Gdynia, Poland), and used for transformation. *Y. lipolytica* strain was transformed using a standard lithium acetate method [100]. Transformants were selected as instructed by [99], so the co-transformation reactions were inoculated into 9 mL of YPD-hygromycin-nourseothricin liquid medium and cultured at 28 °C for 48 h with shaking at 150 rpm. The following selection conditions were used: hygromycin B at 400 μg L^{-1}, or nourseothricin (both from Sigma-Aldrich-Merck-Millipore, Darmstadt, Germany) at 250 μg L^{-1}, supplemented to YPD medium (liquid or solidified). One mL of such cultures was then transferred into 9 mL of YPD-nourseothricin medium and incubated at 28 °C for 24 h with shaking at 150 rpm to allow plasmid curing (dropping-off JME4580). Finally, the culture was diluted and plated on YPD-nourseothricin agar. Clones appearing after 48 h of incubation at 28 °C were verified for correct integration of the deletion cassettes by PCR and sequencing. All the strains were deposited as 30% glycerol stocks at −80 °C for long-term storage.

4.3. Functional Studies—Direct Comparative Study of OE and KO Strains

4.3.1. Cultivation Conditions

The yeast strains were revived from glycerol stocks and then routinely maintained at 28 °C in rich YPD (g L-1: yeast extract, 5 (BTL, Łódź, Poland); peptone, 10 (BTL); glucose, 20 (POCH, Gliwice, Poland); solidified with agar, 15 (BTL)) or in minimal YNB medium (g L-1: glucose, 10 (POCH); yeast nitrogen base, 1.7 (Sigma-Aldrich, St. Louis, MI, USA); ammonium sulfate, 5 (POCH); solidified with agar, 15 (BTL)). Two to six subclones were cultivated in biological duplicates according to a high-throughput cultivation protocol developed in our laboratory [23].

The cultivation medium and conditions were as follows [g L-1]: yeast nitrogen base without amino acids and ammonium sulfate, 5.1 (Sigma-Aldrich); ammonium sulfate and glutamic acid, both at 7.5 (POCH); glucose, 35 (POCH); buffered with 0.2 M maleic acid at pH 5.0 (POCH). Precultures were developed in media composed of a yeast nitrogen base, 5.1 (Sigma-Aldrich); ammonium sulfate, 15, (POCH); glucose, 25 (POCH); buffered with 0.2 M maleic acid at pH 5.0. All the media were filter-sterilized with 0.22-μm filters (Merck-Millipore, Darmstadt, Germany).

4.3.2. Samples Analysis

Samples were withdrawn at 48 h and assessed for optical density (OD) and fluorescence (FL). Before reading the FL from the reporter protein (RedStar2), samples were diluted in 0.75% NaCl (POCH) to match a linear range of the methods. Absorbance was measured at 600 nm in transparent 96-well plates (Costar; Merck-Millipore, Darmstadt, Germany). FL was determined under ex/em wavelengths 550/595 nm in black opaque plates (Thermo Fisher Scientific). Both measurements were performed using a Tecan Spark automatic plate reader (Tecan Group Ltd., Mannedorf, Switzerland).

Fold change values were calculated by dividing the raw readouts for the TF-engineered strain by the result read for the control strain (overexpressing only in RedStar2 protein, with no TF modification).

4.4. Omics and Functional Data Acquisition

Transcriptomics data were extracted from the NCBI SRA database (PRJNA701856 and PRJNA869113). Data were filtered for gene identifiers of the TF-encoding genes based on [20]. For the non-significantly deregulated genes, a default value of 0 was set.

Phenotype OE screen data were extracted from the YaliFunTome database (https://sparrow.up.poznan.pl/tsdatabase/, accessed throughout June 2024) and accompanying publication [23]. The phenotype data were chosen from a variant of high oxygen availability, pH 5, temperature 28 °C, glucose as a carbon source, and the average of ammonium sulfate (AS) and casamino-acid hydrolysate (CH) as nitrogen sources.

4.5. Data Processing and Mathematical Analysis

The results for each TF gene are presented in the form of a heatmap. K-means clustering was performed on these data, with the number of clusters = 10 (determined by an 'elbow method', Supplementary Material—Figure S1), random state = 42, scaled with StandardScaler (sklearn package version 1.3.2). The genes are sorted according to clusters, to which they have been assigned. Clusters are separated with bold horizontal lines (Figure 1). All the analyses were performed in Visual Studio Code (version 1.92.2) in the Python programming language (version 3.11.9) with corresponding data-processing and machine-learning packages.

The correlation matrix for the TFs displaying a universal deregulation pattern in the transcriptomics datasets (presented in Figure S2) was calculated based on the TFs' transcriptional profile in the five strains and on TF-co-OEing (each in two parameters: total FC rProt synthesis and normalized to biomass FC rProt synthesis). The statistical significance of each correlation was calculated with the Pearson r-coefficient at a p-value < 0.05.

A statistical overrepresentation test was run using Panther [67,68]. Analysis Type: PANTHER Overrepresentation Test (Released 20240226). Annotation Version and Release Date: GO Ontology Database DOI: 10.5281/zenodo.10536401 Released 2024-01-17. Analyzed List—YALI signatures of 139 TFs under study (Table 1). Reference List—The reference gene list for the test. *Y. lipolytica* (all genes in database, 6448), 138 user-defined IDs were mapped to the genome. Annotation Dataset—GO biological process complete—Complete GO biological process annotations, including both manually curated and electronic annotations. Electronic annotations were generated by computer algorithms based on sequence similarity. This is the test that was performed using Fisher's exact test with FDR correction.

Statistical significance of the differences observed between OE and KO strains vs. the control strain (JMY2810) was assessed by the ANOVA analysis of variance and post hoc Tukey test at a significance level of $p < 0.05$ (Visual Studio Code, version 1.92.2).

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms25179450/s1, Figure S1. Graphic representation of running an "Elbow method" used for determination of the optimal number of clusters in k-means clustering of transcriptomic and phenotype data for TFs analyzed in this study. Figure S2. The correlation matrix demonstrates the similarity of the transcriptional profiles and the functional screen profiles between the analyzed samples for the TFs displaying a uniform deregulation pattern in the transcriptomic data. Table S1. Results of Statistical Overrepresentation Test.

Author Contributions: Conceptualization, M.G. and E.C.; methodology, M.G.; software, M.G.; investigation, M.G.; resources, P.K.-W. (transcriptomics data), M.G. (functional screens data) and E.C. (KO strains and laboratory reagents); data curation, M.G. and E.C.; writing—original draft preparation, M.G. and P.K.-W. (Table 1), E.C.; writing—review and editing, M.G. and E.C.; visualization, M.G.; supervision, E.C.; project administration, E.C.; funding acquisition, E.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Centre in Poland, grant number 2021/41/B/NZ9/00086.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The primary transcriptomics datasets used in this study are openly available in the NCBI SRA database at https://www.ncbi.nlm.nih.gov/sra/, under reference numbers PRJNA701856 and PRJNA869113. The primary functional screen datasets used in this study are openly available in the YaliFunTome database at https://sparrow.up.poznan.pl/tsdatabase/.

Acknowledgments: Graphical Abstract was Created with BioRender.com Premium.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. He, B.; Tan, K. Understanding Transcriptional Regulatory Networks using Computational Models. *Curr. Opin. Genet. Dev.* **2016**, *37*, 101–108. [CrossRef] [PubMed]
2. Su, K.; Katebi, A.; Kohar, V.; Clauss, B.; Gordin, D.; Qin, Z.S.; Karuturi, R.K.M.; Li, S.; Lu, M. NetAct: A Computational Platform to Construct Core Transcription Factor Regulatory Networks using Gene Activity. *Genome Biol.* **2022**, *23*, 270. [CrossRef]
3. Czajka, J.J.; Oyetunde, T.; Tang, Y.J. Integrated Knowledge Mining, Genome-Scale Modeling, and Machine Learning for Predicting *Yarrowia lipolytica* Bioproduction. *Metab. Eng.* **2021**, *67*, 227–236. [CrossRef] [PubMed]
4. Madzak, C. Engineering *Yarrowia lipolytica* for Use in Biotechnological Applications: A Review of Major Achievements and Recent Innovations. *Mol. Biotechnol.* **2018**, *60*, 621–635. [CrossRef]
5. Madzak, C. *Yarrowia lipolytica* Strains and Their Biotechnological Applications: How Natural Biodiversity and Metabolic Engineering Could Contribute to Cell Factories Improvement. *J. Fungi* **2021**, *7*, 548. [CrossRef] [PubMed]
6. Celińska, E.; Nicaud, J.-M. Filamentous Fungi-like Secretory Pathway Strayed in a Yeast System: Peculiarities of *Yarrowia lipolytica* Secretory Pathway Underlying Its Extraordinary Performance. *Appl. Microbiol. Biotechnol.* **2019**, *103*, 39–52. [CrossRef]
7. Theron, C.W.; Vandermies, M.; Telek, S.; Steels, S.; Fickers, P. Comprehensive Comparison of *Yarrowia lipolytica* and *Pichia Pastoris* for Production of Candida Antarctica Lipase B. *Sci. Rep.* **2020**, *10*, 1741. [CrossRef]
8. Vandermies, M.; Denies, O.; Nicaud, J.-M.; Fickers, P. EYK1 Encoding Erythrulose Kinase as a Catabolic Selectable Marker for Genome Editing in the Non-Conventional Yeast *Yarrowia lipolytica*. *J. Microbiol. Methods* **2017**, *139*, 161–164. [CrossRef]
9. Vidal, L.; Lebrun, E.; Park, Y.K.; Mottet, G.; Nicaud, J.M. Bidirectional Hybrid Erythritol-Inducible Promoter for Synthetic Biology in *Yarrowia lipolytica*. *Microb. Cell Fact.* **2023**, *22*, 7. [CrossRef]
10. Park, Y.-K.; Korpys, P.; Kubiak, M.; Celińska, E.; Soudier, P.; Trébulle, P.; Larroude, M.; Rossignol, T.; Nicaud, J. Engineering the Architecture of Erythritol-Inducible Promoters for Regulated and Enhanced Gene Expression in *Yarrowia lipolytica*. *FEMS Yeast Res.* **2019**, *19*, foy105. [CrossRef]
11. Celińska, E.; Korpys-Woźniak, P.; Gorczyca, M.; Nicaud, J. Using Euf1 transcription factor as a titrator of erythritol-inducible promoters in *Yarrowia lipolytica*; insight into the structure, splicing, and regulation mechanism. *FEMS Yeast Res.* **2024**, foae027. [CrossRef] [PubMed]
12. Park, Y.K.; Vandermies, M.; Soudier, P.; Telek, S.; Thomas, S.; Nicaud, J.M.; Fickers, P. Efficient Expression Vectors and Host Strain for the Production of Recombinant Proteins by *Yarrowia lipolytica* in Process Conditions. *Microb. Cell Fact.* **2019**, *18*, 167. [CrossRef] [PubMed]
13. Trassaert, M.; Vandermies, M.; Carly, F.; Denies, O.; Thomas, S.; Fickers, P.; Nicaud, J.M. New Inducible Promoter for Gene Expression and Synthetic Biology in *Yarrowia lipolytica*. *Microb. Cell Fact.* **2017**, *16*, 141. [CrossRef] [PubMed]
14. Korpys-Woźniak, P.; Kubiak, P.; Białas, W.; Celińska, E. Impact of Overproduced Heterologous Protein Characteristics on Physiological Response in *Yarrowia lipolytica* Steady-State-Maintained Continuous Cultures. *Appl. Microbiol. Biotechnol.* **2020**, *104*, 9785–9800. [CrossRef] [PubMed]
15. Korpys-Woźniak, P.; Kubiak, P.; Celińska, E. Secretory Helpers for Enhanced Production of Heterologous Proteins in *Yarrowia lipolytica*. *Biotechnol. Rep.* **2021**, *32*, e00669. [CrossRef]
16. Korpys-Woźniak, P.; Celińska, E. Global Transcriptome Profiling Reveals Genes Responding to Overproduction of a Small Secretory, a High Cysteine- and a High Glycosylation-Bearing Protein in *Yarrowia lipolytica*. *Biotechnol. Rep.* **2021**, *31*, e00646. [CrossRef]
17. Oh, M.H.; Cheon, S.A.; Kang, H.A.; Kim, J.Y. Functional Characterization of the Unconventional Splicing of *Yarrowia lipolytica HAC1* MRNA Induced by Unfolded Protein Response. *Yeast* **2010**, *27*, 443–452. [CrossRef]
18. Guerfal, M.; Ryckaert, S.; Jacobs, P.P.; Ameloot, P.; Van Craenenbroeck, K.; Derycke, R.; Callewaert, N. The HAC1 Gene from *Pichia Pastoris*: Characterization and Effect of Its Overexpression on the Production of Secreted, Surface Displayed and Membrane Proteins. *Microb. Cell Fact.* **2010**, *9*, 49. [CrossRef] [PubMed]
19. Korpys-Woźniak, P.; Celińska, E. Molecular Background of *HAC1*-Driven Improvement in the Secretion of Recombinant Protein in *Yarrowia lipolytica* Based on Comparative Transcriptomics. *Biotechnol. Rep.* **2023**, *38*, e00801. [CrossRef] [PubMed]
20. Leplat, C.; Nicaud, J.-M.M.; Rossignol, T. Overexpression Screen Reveals Transcription Factors Involved in Lipid Accumulation in *Yarrowia lipolytica*. *FEMS Yeast Res.* **2018**, *18*, foy037. [CrossRef]
21. Trébulle, P.; Nicaud, J.M.; Leplat, C.; Elati, M. Inference and Interrogation of a Coregulatory Network in the Context of Lipid Accumulation in *Yarrowia lipolytica*. *NPJ Syst. Biol. Appl.* **2017**, *3*, 21. [CrossRef]

22. Gorczyca, M.; Nicaud, J.-M.; Celińska, E. Transcription Factors Enhancing Synthesis of Recombinant Proteins and Resistance to Stress in *Yarrowia lipolytica*. *Appl. Microbiol. Biotechnol.* **2023**, *107*, 4853–4871. [CrossRef] [PubMed]
23. Gorczyca, M.; Białas, W.; Nicaud, J.-M.; Celińska, E. 'Mother(Nature) Knows Best'—Hijacking Nature-Designed Transcriptional Programs for Enhancing Stress Resistance and Protein Production in *Yarrowia lipolytica*; Presentation of YaliFunTome Database. *Microb. Cell Fact.* **2024**, *23*, 26. [CrossRef]
24. Ruchala, J.; Kurylenko, O.O.; Soontorngun, N.; Dmytruk, K.V.; Sibirny, A.A. Transcriptional Activator Cat8 Is Involved in Regulation of Xylose Alcoholic Fermentation in the Thermotolerant Yeast *Ogataea (Hansenula) Polymorpha*. *Microb. Cell Fact.* **2017**, *16*, 36. [CrossRef]
25. Semkiv, M.V.; Ruchala, J.; Tsaruk, A.Y.; Zazulya, A.Z.; Vasylyshyn, R.V.; Dmytruk, O.V.; Zuo, M.X.; Kang, Y.; Dmytruk, K.V.; Sibirny, A.A. The Role of Hexose Transporter-like Sensor Hxs1 and Transcription Activator Involved in Carbohydrate Sensing Azf1 in Xylose and Glucose Fermentation in the Thermotolerant Yeast *Ogataea Polymorpha*. *Microb. Cell Fact.* **2022**, *21*, 162. [CrossRef] [PubMed]
26. Zhang, Y.; Wang, L.; Liang, S.; Zhang, P.; Kang, R.; Zhang, M.; Wang, M.; Chen, L.; Yuan, H.; Ding, S.; et al. FpDep1, a Component of Rpd3L Histone Deacetylase Complex, Is Important for Vegetative Development, ROS Accumulation, and Pathogenesis in *Fusarium Pseudograminearum*. *Fungal Genet. Biol.* **2020**, *135*, 103299. [CrossRef] [PubMed]
27. Bonander, N.; Ferndahl, C.; Mostad, P.; Wilks, M.D.B.; Chang, C.; Showe, L.; Gustafsson, L.; Larsson, C.; Bill, R.M. Transcriptome Analysis of a Respiratory *Saccharomycescerevisiae* Strain Suggests the Expression of Its Phenotype Is Glucose Insensitive and Predominantly Controlled by Hap4, Cat8 and Mig1. *BMC Genom.* **2008**, *9*, 365. [CrossRef] [PubMed]
28. Stein, T.; Kricke, J.; Becher, D.; Lisowsky, T. Azf1p Is a Nuclear-Localized Zinc-Finger Protein That Is Preferentially Expressed under Non-Fermentative Growth Conditions in *Saccharomyces cerevisiae*. *Curr. Genet.* **1998**, *34*, 287–296. [CrossRef]
29. Gorczyca, M.; Każmierczak, J.; Fickers, P.; Celińska, E. Synthesis of Secretory Proteins in *Yarrowia lipolytica*: Effect of Combined Stress Factors and Metabolic Load. *Int. J. Mol. Sci.* **2022**, *23*, 3602. [CrossRef]
30. Gorczyca, M.; Kaźmierczak, J.; Steels, S.; Fickers, P.; Celińska, E. Impact of Oxygen Availability on Heterologous Geneexpression and Polypeptide Secretion Dynamics in *Yarrowia lipolytica*-based Protein Production Platforms. *Yeast* **2020**, *37*, 559–568. [CrossRef]
31. Celińska, E.; Gorczyca, M. 'Small Volume—Big Problem': Culturing *Yarrowia lipolytica* in High-Throughput Micro-Formats. *Microb. Cell Fact.* **2024**, *23*, 184. [CrossRef] [PubMed]
32. Doughty, T.; Kerkhoven, E. Extracting Novel Hypotheses and Findings from RNA-Seq Data. *FEMS Yeast Res.* **2020**, *20*, foaa007. [CrossRef]
33. Li, J.; Singh, U.; Arendsee, Z.; Wurtele, E.S. Landscape of the Dark Transcriptome Revealed Through Re-Mining Massive RNA-Seq Data. *Front. Genet.* **2021**, *12*, 722981. [CrossRef] [PubMed]
34. Sazegari, S.; Niazi, A.; Zinati, Z.; Eskandari, M.H. Mining Transcriptomic Data to Identify *Saccharomyces cerevisiae* Signatures Related to Improved and Repressed Ethanol Production under Fermentation. *PLoS ONE* **2022**, *17*, e0259476. [CrossRef]
35. Abid, D.; Brent, M.R. NetProphet 3: A Machine Learning Framework for Transcription Factor Network Mapping and Multi-Omics Integration. *Bioinformatics* **2023**, *39*, btad038. [CrossRef]
36. Sastry, A.V.; Poudel, S.; Rychel, K.; Yoo, R.; Lamoureux, C.R.; Chauhan, S.; Haiman, Z.B.; Al Bulushi, T.; Seif, Y.; Palsson, B.O. Mining All Publicly Available Expression Data to Compute Dynamic Microbial Transcriptional Regulatory Networks. *bioRxiv* **2021**. [CrossRef]
37. Rychel, K.; Decker, K.; Sastry, A.V.; Phaneuf, P.V.; Poudel, S.; Palsson, B.O. IModulonDB: A Knowledgebase of Microbial Transcriptional Regulation Derived from Machine Learning. *Nucleic Acids Res.* **2021**, *49*, D112–D120. [CrossRef]
38. Lee, S.-I.; Batzoglou, S. Application of Independent Component Analysis to Microarrays. *Genome Biol.* **2003**, *4*, R76. [CrossRef] [PubMed]
39. Kerssemakers, A.; Krishnan, J.; Rychel, K.; Zielinski, D.; Palsson, B.; Sudarsan, S. *Deciphering the TRN of Yarrowia lipolytica Using Machine Learning*; Technical University of Denmark: Lyngby, Denmark, 2023.
40. Kubiak-Szymendera, M.; Skupien-Rabian, B.; Jankowska, U.; Celińska, E. Hyperosmolarity Adversely Impacts Recombinant Protein Synthesis by *Yarrowia lipolytica*—Molecular Background Revealed by Quantitative Proteomics. *Appl. Microbiol. Biotechnol.* **2022**, *106*, 349–367. [CrossRef]
41. Morales-Vargas, A.T.; Domínguez, A.; Ruiz-Herrera, J. Identification of Dimorphism-Involved Genes of *Yarrowia lipolytica* by Means of Microarray Analysis. *Res. Microbiol.* **2012**, *163*, 378–387. [CrossRef]
42. Pomraning, K.R.; Bredeweg, E.L.; Baker, S.E. Regulation of Nitrogen Metabolism by GATA Zinc Finger Transcription Factors in *Yarrowia lipolytica*. *mSphere* **2017**, *2*, 10-1128. [CrossRef] [PubMed]
43. Tsirigka, A.; Theodosiou, E.; Patsios, S.I.; Tsoureki, A.; Andreadelli, A.; Papa, E.; Aggeli, A.; Karabelas, A.J.; Makris, A.M. Novel Evolved *Yarrowia lipolytica* Strains for Enhanced Growth and Lipid Content under High Concentrations of Crude Glycerol. *Microb. Cell Fact.* **2023**, *22*, 62. [CrossRef]
44. Wang, G.; Li, D.; Miao, Z.; Zhang, S.; Liang, W.; Liu, L. Comparative Transcriptome Analysis Reveals Multiple Functions for Mhy1p in Lipid Biosynthesis in the Oleaginous Yeast *Yarrowia lipolytica*. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* **2018**, *1863*, 81–90. [CrossRef] [PubMed]
45. Konzock, O.; Norbeck, J. Deletion of *MHY1* Abolishes Hyphae Formation in *Yarrowia lipolytica* without Negative Effects on Stress Tolerance. *PLoS ONE* **2020**, *15*, e0231161. [CrossRef] [PubMed]

46. Shu, T.; He, X.-Y.; Chen, J.-W.; Mao, Y.-S.; Gao, X.-D. The PH-Responsive Transcription Factors YlRim101 and Mhy1 Regulate Alkaline PH-Induced Filamentation in the Dimorphic Yeast *Yarrowia lipolytica*. *mSphere* **2021**, *6*, 10-1128. [CrossRef] [PubMed]
47. Martinez-Vazquez, A.; Gonzalez-Hernandez, A.; Domínguez, Á.; Rachubinski, R.; Riquelme, M.; Cuellar-Mata, P.; Guzman, J.C.T. Identification of the Transcription Factor Znc1p, Which Regulates the Yeast-to-Hypha Transition in the Dimorphic Yeast *Yarrowia lipolytica*. *PLoS ONE* **2013**, *8*, e66790. [CrossRef]
48. Mittal, N.; Guimaraes, J.C.; Gross, T.; Schmidt, A.; Vina-Vilaseca, A.; Nedialkova, D.D.; Aeschimann, F.; Leidel, S.A.; Spang, A.; Zavolan, M. The Gcn4 Transcription Factor Reduces Protein Synthesis Capacity and Extends Yeast Lifespan. *Nat. Commun.* **2017**, *8*, 457. [CrossRef]
49. Zhang, L.; Hach, A. Molecular Mechanism of Heme Signaling in Yeast: The Transcriptional Activator Hap1 Serves as the Key Mediator. *Cell Mol. Life Sci.* **1999**, *56*, 415–426. [CrossRef]
50. Rzechonek, D.A.; Szczepańczyk, M.; Borodina, I.; Neuvéglise, C.; Mirończuk, A.M. Transcriptome Analysis Reveals Multiple Targets of Erythritol-Related Transcription Factor EUF1 in Unconventional Yeast *Yarrowia lipolytica*. *Microb. Cell Fact.* **2024**, *23*, 77. [CrossRef]
51. Rzechonek, D.A.; Neuvéglise, C.; Devillers, H.; Rymowicz, W.; Mirończuk, A.M. *EUF1*-A Newly Identified Gene Involved in Erythritol Utilization in *Yarrowia lipolytica*. *Sci. Rep.* **2017**, *7*, 12507. [CrossRef]
52. Hirakawa, K.; Kobayashi, S.; Inoue, T.; Endoh-Yamagami, S.; Fukuda, R.; Ohta, A. Yas3p, an Opi1 Family Transcription Factor, Regulates Cytochrome P450 Expression in Response to n-Alkanes in *Yarrowia lipolytica*. *J. Biol. Chem.* **2009**, *284*, 7126–7137. [CrossRef]
53. Pomraning, K.R.; Bredeweg, E.L.; Kerkhoven, E.J.; Barry, K.; Haridas, S.; Hundley, H.; LaButti, K.; Lipzen, A.; Yan, M.; Magnuson, J.K.; et al. Regulation of Yeast-to-Hyphae Transition in *Yarrowia lipolytica*. *mSphere* **2018**, *3*, e00541-18. [CrossRef]
54. Mao, Y.S.; Chen, J.W.; Wang, Z.H.; Xu, M.Y.; Gao, X.D. Roles of the Transcriptional Regulators Fts1, YlNrg1, YlTup1, and YlSsn6 in the Repression of the Yeast-to-Filament Transition in the Dimorphic Yeast *Yarrowia lipolytica*. *Mol. Microbiol.* **2023**, *119*, 126–142. [CrossRef]
55. Celińska, E.; Borkowska, M.; Białas, W.; Kubiak, M.; Korpys, P.; Archacka, M.; Ledesma-Amaro, R.; Nicaud, J.M. Genetic Engineering of Ehrlich Pathway Modulates Production of Higher Alcohols in Engineered *Yarrowia lipolytica*. *FEMS Yeast Res.* **2019**, *19*, foy122. [CrossRef]
56. García, S.; Prado, M.; Dégano, R.; Domínguez, A. A Copper-Responsive Transcription Factor, *CRF1*, Mediates Copper and Cadmium Resistance in *Yarrowia lipolytica*. *J. Biol. Chem.* **2002**, *277*, 37359–37368. [CrossRef]
57. Wang, Z.-P.; Xu, H.-M.; Wang, G.-Y.; Chi, Z.; Chi, Z.-M. Disruption of the MIG1 Gene Enhances Lipid Biosynthesis in the Oleaginous Yeast *Yarrowia lipolytica* ACA-DC 50109. *Biochim. Biophys. Acta* **2013**, *1831*, 675–682. [CrossRef] [PubMed]
58. Torres-Guzma'n, J.C.; Guzma´n, G.; Domi´nguez, A.; Domi´nguez, D. HOY1, a Homeo Gene Required for Hyphal Formation in Yarrowia lipolytica. *Mol. Cell. Biol.* **1997**, *17*, 6283–6293. [CrossRef] [PubMed]
59. Chen, J.-W.; Mao, Y.-S.; Yan, L.-Q.; Gao, X.-D. The Zinc Finger Transcription Factor Fts2 Represses the Yeast-to-Filament Transition in the Dimorphic Yeast *Yarrowia lipolytica*. *mSphere* **2022**, *7*, e0045022. [CrossRef] [PubMed]
60. González-Lozano, K.J.; Aréchiga-Carvajal, E.T.; Jiménez-Salas, Z.; Valdez-Rodríguez, D.M.; León-Ramírez, C.G.; Ruiz-Herrera, J.; Adame-Rodríguez, J.M.; López-Cabanillas-Lomelí, M.; Campos-Góngora, E. Identification and Characterization of Dmct: A Cation Transporter in *Yarrowia lipolytica* Involved in Metal Tolerance. *J. Fungi* **2023**, *9*, 600. [CrossRef] [PubMed]
61. Walker, C.; Ryu, S.; Trinh, C.T. Exceptional Solvent Tolerance in *Yarrowia lipolytica* Is Enhanced by Sterols. *Metab. Eng.* **2019**, *54*, 83–95. [CrossRef]
62. Poopanitpan, N.; Kobayashi, S.; Fukuda, R.; Horiuchi, H.; Ohta, A. An Ortholog of FarA of Aspergillus Nidulans Is Implicated in the Transcriptional Activation of Genes Involved in Fatty Acid Utilization in the Yeast *Yarrowia lipolytica*. *Biochem. Biophys. Res. Commun.* **2010**, *402*, 731–735. [CrossRef] [PubMed]
63. Endoh-Yamagami, S.; Hirakawa, K.; Morioka, D.; Fukuda, R.; Ohta, A. Basic Helix-Loop-Helix Transcription Factor Heterocomplex of Yas1p and Yas2p Regulates Cytochrome P450 Expression in Response to Alkanes in the Yeast *Yarrowia lipolytica*. *Eukaryot. Cell* **2007**, *6*, 734–743. [CrossRef]
64. Hurtado, C.A.R.; Rachubinski, R.A. Mhy1 Encodes a C2h2-Type Zinc Finger Protein That Promotes Dimorphic Transition in the Yeast *Yarrowia lipolytica*. *J. Bacteriol.* **1999**, *181*, 3051–3057. [CrossRef] [PubMed]
65. Hurtado, C.A.R.; Rachubinski, R.A. YlBMH1 Encodes a 14-3-3 Protein That Promotes Filamentous Growth in the Dimorphic Yeast *Yarrowia lipolytica*. *Microbiology* **2002**, *148*, 3725–3735. [CrossRef]
66. Celińska, E.; Olkowicz, M.; Grajek, W. L-Phenylalanine Catabolism and 2-Phenylethano Synthesis in *Yarrowia lipolytica*-Mapping Molecular Identities through Whole-Proteome Quantitative Mass Spectrometry Analysis. *FEMS Yeast Res.* **2015**, *15*, fov041. [CrossRef] [PubMed]
67. Mi, H.; Muruganujan, A.; Huang, X.; Ebert, D.; Mills, C.; Guo, X.; Thomas, P. Protocol Update for Large-Scale Genome and Gene Function Analysis with PANTHER Classification System (v.14.0). *Nat. Protoc.* **2019**, *14*, 703–721. [CrossRef]
68. Mi, H.; Ebert, D.; Muruganujan, A.; Mills, C.; Albou, L.-P.; Mushayamaha, T.; Thomas, P.D. PANTHER Version 16: A Revised Family Classification, Tree-Based Classification Tool, Enhancer Regions and Extensive API. *Nucleic Acids Res.* **2021**, *49*, D394–D403. [CrossRef]
69. Mogi, R.; Watanabe, J. Identification of SFL1 as a Positive Regulator for Flor Formation in *Zygosaccharomyces Rouxii*. *Biosci. Biotechnol. Biochem.* **2020**, *84*, 1291–1298. [CrossRef]

70. Ansanay Galeote, V.; Alexandre, H.; Bach, B.; Delobel, P.; Dequin, S.; Blondin, B. Sfl1p Acts as an Activator of the *HSP30* Gene in *Saccharomyces cerevisiae*. *Curr. Genet.* **2007**, *52*, 55–63. [CrossRef]
71. Atsushi, F.; Yoshiko, K.; Satoru, K.; Yoshio, M.; Shinichi, M.; Harumi, K. Domains of the SFL1 Protein of Yeasts Are Homologous to Myc Oncoproteins or Yeast Heat-Shock Transcription Factor. *Gene* **1989**, *85*, 321–328. [CrossRef]
72. De Vit, M.J.; Waddle, J.A.; Johnston, M. Regulated Nuclear Translocation of the Mig1 Glucose Repressor. *Mol. Biol. Cell* **1997**, *8*, 1603–1618. [CrossRef]
73. Treitel, M.A.; Carlson, M. Repression by SSN6-TUP1 Is Directed by MIG1, a Repressor/Activator Protein (Transcription/Yeast/Zinc-Finger Protein/Glucose Repression). *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 3132–3136. [CrossRef]
74. Martínez, J.L.; Meza, E.; Petranovic, D.; Nielsen, J. The Impact of Respiration and Oxidative Stress Response on Recombinant α-Amylase Production by *Saccharomyces cerevisiae*. *Metab. Eng. Commun.* **2016**, *3*, 205–210. [CrossRef]
75. Rep, M.; Proft, M.; Remize, F.; Tamás, M.; Serrano, R.; Thevelein, J.M.; Hohmann, S. The *Saccharomyces cerevisiae* Sko1p Transcription Factor Mediates *HOG* Pathway-Dependent Osmotic Regulation of a Set of Genes Encoding Enzymes Implicated in Protection from Oxidative Damage. *Mol. Microbiol.* **2001**, *40*, 1067–1083. [CrossRef]
76. Karlgren, S.; Pettersson, N.; Nordlander, B.; Mathai, J.C.; Brodsky, J.L.; Zeidel, M.L.; Bill, R.M.; Hohmann, S. Conditional Osmotic Stress in Yeast: A System to Study Transport through Aquaglyceroporins and Osmostress Signaling. *J. Biol. Chem.* **2004**, *280*, 7186–7193. [CrossRef]
77. Callewaer, N.; Vervecken, W.; De Pourq, K.; Geysens, S.; Guerfal, M. *Yarrowia lipolytica* and *Pichia Pastoris* Hac1 Nucleic Acids. U.S. Patent US8026083B2, 27 September 2011.
78. Hooks, K.B.; Griffiths-Jones, S. Conserved RNA Structures in the Non-Canonical Hac1/Xbp1 Intron. *RNA Biol.* **2011**, *8*, 552–556. [CrossRef]
79. Whyteside, G.; Nor, R.M.; Alcocer, M.J.C.; Archer, D.B. Activation of the Unfolded Protein Response in *Pichia Pastoris* Requires Splicing of a *HAC1* MRNA Intron and Retention of the C-Terminal Tail of Hac1p. *FEBS Lett.* **2011**, *585*, 1037–1041. [CrossRef]
80. Graf, A.; Dragosits, M.; Gasser, B.; Mattanovich, D. Yeast Systems Biotechnology for the Production of Heterologous Proteins. *FEMS Yeast Res.* **2009**, *9*, 335–348. [CrossRef]
81. Zahrl, R.J.; Gasser, B.; Mattanovich, D.; Ferrer, P. Detection and Elimination of Cellular Bottlenecks in Protein-Producing Yeasts. *Methods Mol. Biol.* **2019**, *1923*, 75–95. [CrossRef]
82. Gasser, B.; Sauer, M.; Maurer, M.; Stadlmayr, G.; Mattanovich, D. Transcriptomics-Based Identification of Novel Factors Enhancing Heterologous Protein Secretion in Yeasts. *Appl. Environ. Microbiol.* **2007**, *73*, 6499–6507. [CrossRef]
83. Zahrl, R.J.; Prielhofer, R.; Burgard, J.; Mattanovich, D.; Gasser, B. Synthetic Activation of Yeast Stress Response Improves Secretion of Recombinant Proteins. *N. Biotechnol.* **2023**, *73*, 19–28. [CrossRef] [PubMed]
84. Gasser, B.; Maurer, M.; Rautio, J.; Sauer, M.; Bhattacharyya, A.; Saloheimo, M.; Penttilä, M.; Mattanovich, D. Monitoring of Transcriptional Regulation in *Pichia Pastoris* under Protein Production Conditions. *BMC Genom.* **2007**, *8*, 179. [CrossRef] [PubMed]
85. Zahrl, R.J.; Peña, D.A.; Mattanovich, D.; Gasser, B. Systems Biotechnology for Protein Production in *Pichia Pastoris*. *FEMS Yeast Res.* **2017**, *17*, fox068. [CrossRef] [PubMed]
86. Lamping, E.; Lückl, J.; Paltauf, F.; Henry, S.A.; Kohlwein, S.D. Isolation and Characterization of a Mutant of *Saccharomyces cerevisiae* with Pleiotropic Deficiencies in Transcriptional Activation and Repression. *Genetics* **1994**, *137*, 55–65. [CrossRef]
87. Shimanuki, M.; Uehara, L.; Pluskal, T.; Yoshida, T.; Kokubu, A.; Kawasaki, Y.; Yanagida, M. Klf1, a C2H2 Zinc Finger-Transcription Factor, Is Required for Cell Wall Maintenance during Long-Term Quiescence in Differentiated G0 Phase. *PLoS ONE* **2013**, *8*, e78545. [CrossRef]
88. Herholz, M.; Cepeda, E.; Baumann, L.; Kukat, A.; Hermeling, J.; Maciej, S.; Szczepanowska, K.; Pavlenko, V.; Frommolt, P.; Trifunovic, A. KLF-1 Orchestrates a Xenobiotic Detoxification Program Essential for Longevity of Mitochondrial Mutants. *Nat. Commun.* **2019**, *10*, 3323. [CrossRef] [PubMed]
89. Hou, J.; Österlund, T.; Liu, Z.; Petranovic, D.; Nielsen, J.; Osterlund, T.; Liu, Z.; Petranovic, D.; Nielsen, J. Heat Shock Response Improves Heterologous Protein Secretion in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 3559–3568. [CrossRef]
90. Hou, J.; Tyo, K.E.J.; Liu, Z.; Petranovic, D.; Nielsen, J. Metabolic Engineering of Recombinant Protein Secretion by *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **2012**, *12*, 491–510. [CrossRef]
91. Hou, J.; Tang, H.; Liu, Z.; Österlund, T.; Nielsen, J.; Petranovic, D. Management of the Endoplasmic Reticulum Stress by Activation of the Heat Shock Response in Yeast. *FEMS Yeast Res.* **2014**, *14*, 481–494. [CrossRef]
92. Leplat, C.; Nicaud, J.M.; Rossignol, T. High-Throughput Transformation Method for *Yarrowia lipolytica* Mutant Library Screening. *FEMS Yeast Res.* **2015**, *15*, fov052. [CrossRef]
93. Celińska, E.; Ledesma-Amaro, R.; Larroude, M.; Rossignol, T.; Pauthenier, C.; Nicaud, J.M. Golden Gate Assembly System Dedicated to Complex Pathway Manipulation in *Yarrowia lipolytica*. *Microb. Biotechnol.* **2017**, *10*, 450–455. [CrossRef]
94. Korpys-Woźniak, P.; Kubiak, M.; Borkowska, M.; Celińska, E. Construction and Assembly of Standardized Biobricks for Synthetic Pathways Engineering in Yeasts. In *Synthetic Biology of Yeasts: Tools and Applications*; Darvishi Harzevili, F., Ed.; Springer International Publishing: Cham, Switzerland, 2022; pp. 27–62. ISBN 978-3-030-89680-5.
95. Celińska, E.; Borkowska, M.; Białas, W.; Korpys, P.; Nicaud, J.M. Robust Signal Peptides for Protein Secretion in *Yarrowia lipolytica*: Identification and Characterization of Novel Secretory Tags. *Appl. Microbiol. Biotechnol.* **2018**, *102*, 5221–5233. [CrossRef] [PubMed]

96. Borkowska, M.; Białas, W.; Kubiak, M.; Celińska, E. Rapid Micro-Assays for Amylolytic Activities Determination: Customization and Validation of the Tests. *Appl. Microbiol. Biotechnol.* **2019**, *103*, 2367–2379. [CrossRef] [PubMed]
97. Sambrook, J.; Russell, D. *Molecular Cloning: A Laboratory Manual*, 3rd ed.; Cold Spring Harbor Laboratory Press: New York, NY, USA, 2001; ISBN 978-0879695-76-7.
98. Larroude, M.; Park, Y.K.; Soudier, P.; Kubiak, M.; Nicaud, J.M.; Rossignol, T. A Modular Golden Gate Toolkit for *Yarrowia lipolytica* Synthetic Biology. *Microb. Biotechnol.* **2019**, *12*, 1249–1259. [CrossRef]
99. Larroude, M.; Trabelsi, H.; Nicaud, J.M.; Rossignol, T. A Set of *Yarrowia lipolytica* CRISPR/Cas9 Vectors for Exploiting Wild-Type Strain Diversity. *Biotechnol. Lett.* **2020**, *42*, 773–785. [CrossRef]
100. Chen, D.C.; Beckerich, J.M.; Gaillardin, C. One-Step Transformation of the Dimorphic Yeast *Yarrowia lipolytica*. *Appl. Microbiol. Biotechnol.* **1997**, *48*, 232–235. [CrossRef] [PubMed]

5.6 **P5**: Wykorzystanie czynnika transkrypcyjnego Euf1 jako regulatora siły promotorów indukowanych erytrytolem w drożdżach *Yarrowia lipolytica*; wgląd w strukturę, splicing i mechanizm regulacji

Jednym z pytań postawionych w rozdziale 2. niniejszej rozprawy oraz powiązanych celów, wyróżnionych w rozdziale 3, było zapytanie o możliwość wykorzystania TF jako modulatorów siły transkrypcji specyficznych dla nich promotorów. Postawiony cel szczegółowy zakładał przedstawienie szczegółowej charakterystyki wybranego TF oraz ocenę jego przydatności jako narzędzia „biologii syntetycznej" do regulowania siły kompatybilnego promotora indukowalnego.

Promotory indukowalne oraz o regulowanej sile (ang. *titratable*) stanowią ważne narzędzia „biologii syntetycznej", umożliwiające precyzyjne sterowanie ekspresją genów w produkcji biotechnologicznej. W przypadku *Y. lipolytica* szczególne znaczenie mają promotory indukowalne erytrytolem (np. pEYK1, pEYD1), które wcześniej zoptymalizowano pod kątem architektury sekwencji (Park i in., 2019a; Trassaert i in., 2017; Vidal i in., 2023). Wskazano, że dodawanie kolejnych motywów aktywujących (motyw A) zwiększa ich siłę, jednak tylko do pewnego poziomu – dalsza intensyfikacja ekspresji nie była możliwa (Park i in., 2019a). Co więcej, maksymalny poziom białka reporterowego jaki można było uzyskać z promotorów indukowalnych erytrytolem był nadal niższy niż ten uzyskiwany z promotora konstytutywnego pTEF. Stąd postawiono hipotezę o niewystarczającej dostępności głównego regulatora tego systemu – TF Euf1 (*YALI0F01562g*), kontrolującego grupę genów odpowiedzialnych za utylizację erytrytolu. Uznano, że niewystarczająca dostępność aktywnego biologicznie TF Euf1 uniemożliwia dalsze zwiększenie ilości produkowanego rProt przy wykorzystaniu ekspresji z tego indukowanego promotora.

Celem praktycznym postawionym w **P5** było przesunięcie dotychczasowych granic siły syntetycznych promotorów indukowanych erytrytolem wykorzystując kompatybilny TF Euf1 jako regulator siły promotora. Głównym pytaniem badawczym było, czy zwiększenie dostępności Euf1 umożliwi dalsze zwiększanie siły promotora pEYK1. Ponadto, ze cel postawiono sobie scharakteryzowanie mechanizmu działania tego regulatora, w tym zbadanie wpływu induktorów chemicznych, dostępności źródeł węgla i splicingu na jego funkcję. Jak bowiem uprzednio zaobserwowano (Rzechonek i in., 2017) gen kodujący TF Euf1 posiada niewielki intron (78 pz) w regionie

5’ genu. Na podstawie podobieństwa do TF Hac1 (Korpys-Woźniak i Celińska, 2023), założono, że *EUF1* także jest poddawany aktywującemu splicingowi.

Badania prowadzono w tle genetycznym szczepu niezdolnego do utylizacji erytrytolu (*Δeyk1*), co eliminowało wpływ jego metabolizmu na uzyskane wyniki. Testowano dwa warianty genu kodującego Euf1 – pełną (full*EUF1*) oraz splicingową (pozbawioną wspomnianego intronu; sp*EUF1*) – obie pod kontrolą konstytutywnego promotora pTEF. Jako białko reporterowe zastosowano wewnątrzkomórkowe YFP, którego ekspresja była regulowana promotorami pEYK1 z różną liczbą motywów aktywujących A, oraz z jednym motywem B. W rozszerzonych analizach promotorów wykorzystano pEUF800, pEYK300, pTEF, wraz z dwoma reporterami: YFP i RedStar2. Oceniono także wpływ warunków środowiskowych, obejmujących obecność lub brak źródła węgla oraz różnych substancji osmotycznie czynnych (sorbitolu i erytrytolu) w różnych stężeniach, w tym powodujących stres osmotyczny. Ponadto przeprowadzono analizę zmian profili ekspresji genów zależnych od Euf1, a także modelowanie struktury przestrzennej pełnej i splicingowej wersji Euf1 wraz z oceną siły wiązania ligandów do przewidywanych struktur. Kluczowe wyniki i wnioski streszczono w następujących punktach:

i) Euf1 jako regulator siły promotora pEYK1 – Nadekspresja pełnej wersji *EUF1* nie miała wyraźnego wpływu na aktywność promotora pEYK1. Natomiast wersja splicingowa znacząco zwiększała ekspresję genu reporterowego – w wybranych wariantach ponad trzykrotnie. Osiągnięto jednak górny limit siły promotora, którego przekroczenie nie było możliwe nawet przy zwiększonej dostępności spEuf1, co sugeruje istnienie wewnętrznych ograniczeń regulacyjnych. Jednocześnie zaobserwowano wyraźną utratę indukowalności promotora na skutek stosowania spEuf1, obserwowaną jako wysoką aktywność również w warunkach braku erytrytolu. W praktyce eliminuje to możliwość użycia tego układu jako „indukowalnego”.

ii) Wpływ dostępności węgla i induktorów chemicznych na ekspresję promotorów zależnych od Euf1 – Brak źródła węgla, glukozy lub glicerolu, okazał się silnym induktorem transkrypcji z promotora pEYK1, zarówno w obecności erytrytolu, jak i sorbitolu (w określonych stężeniach). Oznacza to, że sorbitol może być traktowany jako alternatywny induktor chemiczny dla systemów opartych na promotorach pEYK w *Y. lipolytica*, bez konieczności stosowania tła genetycznego *Δeyk1* (lub innego; ponieważ sorbitol nie jest utylizowany przez *Y. lipolytica* (Kubiak i in., 2019). Wyniki te wskazują, że promotory zależne of Euf1 są jednocześnie indukowane erytrytolem i silnie

represjonowane przez glukozę/glicerol. Z punktu widzenia praktyki bioprocesowej uzyskane obserwacje mają istotne znaczenie. Po pierwsze, w systemach ekspresyjnych opartych na pEYK1 korzystne może być wprowadzenie etapu końcowego „głodzenia" komórek, aby w pełni wykorzystać potencjał promotora. Po drugie, wartościowym kierunkiem inżynieryjnym byłoby usunięcie motywu B z syntetycznych promotorów pEYK1, co pozwoliłoby wyeliminować mechanizm represji katabolicznej przez glukozę/glicerol. Jest to zasadne, zważywszy że motyw A stanowi element kluczowy do indukcji promotora przez erytrytol lub erytrytulozę (Trassaert i in., 2017), podczas gdy motyw B pełni rolę elementu represyjnego.

iii) Wycinanie intronów i modelowanie struktur Euf1 – Analiza transkryptomiczna wykazała, że splicing *EUF1* jest promowany w obecności erytrytolu, lecz nie jest od niego zależny – oba warianty transkryptu współistnieją w komórce także bez induktora. Stąd pojawiło się pytanie, w jaki sposób splicing wpływa na strukturę białka, jego powinowactwo do DNA oraz czy możliwe jest wskazanie domeny wiążącej erytrytol. Wyniki modelowania strukturalnego przemawiają za dimeryzacją Euf1 jako formą o wyższym prawdopodobieństwie, a także ujawniają wyraźne różnice między wersjami białka. Splicing stabilizował dimer w stanie kompetentnym do wiązania z DNA, zwiększając powinowactwo i siłę interakcji z sekwencją motywu A. W przypadku pełnej, nie-splicingowanej wersji Euf1, domeny typu palca cynkowego były przestrzennie oddalone i zbliżały się dopiero po dokowaniu DNA. Jednocześnie fullEuf1 zachowywał dodatkową kieszeń wiążącą w swojej strukturze, potencjalnie odpowiedzialną za wiązanie z cząsteczką regulatorową. Sugerujemy, że domena ta odgrywa istotną rolę w regulacji funkcjonalnej fullEuf1. Wariant splicingowy, ze względu na zmiany strukturalne w tej kieszeni, prawdopodobnie utracił możliwość takiej regulacji, pozostając w stanie konstytutywnej aktywności. Poznanie dokładnego mechanizmu regulacji wymaga dalszych szczegółowych analiz.

iv) Geny zależne od Euf1 i rola TF Adr1 – Gen *EUF1* wykazuje podstawowy poziom ekspresji nawet bez obecności erytrytolu jako induktora chemicznego. Dodatkowa aktywacja jego ekspresji następuje, gdy i) erytrytol jest obecny w podłożu oraz ii) inne źródła węgla zostaną wyczerpane. Różnicowa analiza ekspresji genów pomiędzy hodowlami suplementowanymi erytrytolem a prowadzonymi wyłącznie na glukozie ujawniła deregulację nie tylko genów związanych z utylizacją erytrytolu, ale również TF *ADR1* – znanego w *S. cerevisiae* jako induktor genów represjonowanych przez glukozę i aktywator utylizacji alternatywnych źródeł węgla. Wyniki sugerują, że

Adr1 może pełnić rolę współregulatora promotora pEYK1, wiążąc się z motywem B i blokując transkrypcję inicjowaną na skutek przyłączenia Euf1 w obecności innych źródeł węgla. Model tej regulacji przedstawiono graficznie w **P5**.

v) Na podstawie uzyskanych wyników zaproponowano model regulacji promotorów indukowalnych erytrytolem i represjonowanych przez glukozę/glicerol, oparty na współdziałaniu spEuf1 i Adr1, w którym: i) spEuf1 wiąże się z motywem A, natomiast Adr1 wykazuje powinowactwo do motywu B, ii) fullEuf1 prawdopodobnie zachowuje zdolność do dodatkowej regulacji przez niezidentyfikowany ligand poprzez specyficzną kieszeń wiążącą, natomiast spEuf1 tę regulowalność traci, pozostając w stanie konstytutywnej zdolności do wiązania z DNA, iii) transkrypcja i splicing *EUF1* są wzmacniane, ale nie zależą od obecności erytrytolu jako chemicznego induktora, iv) Adr1 wiąże się do motywu B w obecności glukozy lub glicerolu, stanowiąc mechaniczną blokadę transkrypcji genów zależnych od Euf1.

Podsumowując, w pracy szczegółowo opisano mechanizm działania promotorów zależnych od erytrytolu, jak również wskazano jego powiązania z dwoma TFs – Euf1 i Adr1. Wyniki te mają istotne znaczenie praktyczne: zaproponowano strategie bioprocesowe umożliwiające lepsze wykorzystanie tych promotorów na poziomie hodowli, a dalsza optymalizacja ich architektury oraz pogłębione badania nad współoddziaływaniem Euf1 i Adr1 otwierają drogę do tworzenia bardziej kontrolowalnych i uniwersalnych narzędzi regulacyjnych w *Y. lipolytica*.

FEMS Yeast Research, 2024, **24**, foae027
DOI: 10.1093/femsyr/foae027
Advance access publication date: 21 August 2024
Research Article

Using Euf1 transcription factor as a titrator of erythritol-inducible promoters in *Yarrowia lipolytica*; insight into the structure, splicing, and regulation mechanism

Ewelina Celińska[1,*], Paulina Korpys-Woźniak[1], Maria Gorczyca[1], Jean-Marc Nicaud[2]

[1]Department of Biotechnology and Food Microbiology, Poznan University of Life Sciences, ul. Wojska Polskiego 48, 60-637 Poznań, Poland
[2]Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, 78350 Jouy-en-Josas, France
*Corresponding author. Department of Biotechnology and Food Microbiology, Poznan University of Life Sciences, ul. Wojska Polskiego 48, 60-637 Poznań, Poland. E-mail: ewelina.celinska@up.poznan.pl
Editor: [Zengyi Shao]

Abstract

Controllable regulatory elements, like inducible, titratable promoters, are highly desired in synthetic biology toolboxes. A set of previously developed erythritol-inducible promoters along with an engineered *Yarrowia lipolytica* host strain were shown to be a very potent expression platform. In this study, we push the previously encountered limits of the synthetic promoters' titratability (by the number of upstream motifs) by using a compatible transcription factor, Euf1, as the promoter titrator. Overexpression of spliced *EUF1* turned out to be very efficient in promoting expression from the compatible promoter, however, the erythritol-inducible character of the promoter was then lost. Analysis of the *EUF1*'s splicing pattern suggests that the intron removal is promoted in the presence of erythritol, but is not dependent on it. The 3D structures of spliced versus unspliced Euf1 were modeled, and ligand-binding strength was calculated and compared. Furthermore, the *EUF1*-dependent expression profile under different chemical stimulants was investigated. Depletion of carbon source was identified as the significant factor upregulating the expression from the Euf1-dependent promoter (2–10-fold). Considering these findings and transcriptomics data, a new mechanism of the Euf1-regulated promoter action is proposed, involving a 'catabolite repression' transcription factor—Adr1, both acting on the same ERY-inducible promoter.

Keywords: recombinant protein; heterologous expression; synthetic promoter; inducible expression; yeast; splicing regulation

Introduction

The current demand for custom-made yeast cell factories forces the need to expand the portfolio of tailored, preferably modular, synthetic biology tools. In this regard, inducible, titratable, tightly regulated promoters are in high demand. Knowing the physiology of *Yarrowia lipolytica* species, able to utilize lipids and proteins very efficiently, one could expect to find lipid- or organic nitrogen-inducible promoters present in its genome. Indeed, the first explored inducible promoters were: (i) pXPR2 (native for inducible alkaline extracellular protease) induced under a combination of pH, carbon:nitrogen, and peptones (Blanchin-Roland et al. 1994, Madzak et al. 1999, Ogrydziak 2013), and (ii) pPOX2 (Juretzek et al. 2000), activated in the presence of lipids i.a.oleic acid. Both these inducible promoters were later engineered to further customize their performance (Madzak et al. 2000, Shabbir Hussain et al. 2016). Still, due to the complexity of regulation and leakage (pXPR2), or difficulties with the hydrophobic inducer handling (pPOX2), their common use as tightly regulatable and titratable regulatory elements was limited.

Further research into the physiology and genomics of *Y. lipolytica* facilitated the finding of novel inducible promoters. Foremost, the pioneering studies on the characterization of the 'erythritol utilization gene cluster' (Carly et al. 2017, 2018, Rzechonek et al. 2017, Mirończuk et al. 2018) yielded the necessary molecular background. In that series of papers, the key genes deregulated in response to erythritol (ERY) supply were identified, including ERY dehydrogenase (*EYD1*, YALI0F01650g) (Carly et al. 2018), erythrulose kinase (*EYK1*, YALI0F01606g) (Carly et al. 2017), and a transcription factor (TF)—*EUF1* (YALI0F01562g) governing the expression of the former two (Rzechonek et al. 2017, Mirończuk et al. 2018). Expression of all three genes was shown to be induced under ERY supply: for *EUF1* 10-fold upregulation was reported (Rzechonek et al. 2017, Mirończuk et al. 2018), for *EYK1*–41-fold (Carly et al. 2017), and 46-fold upregulation for *EYD1* (Carly et al. 2018); holding a promise of very potent, inducible promoters governing their expression.

In-depth studies on the promoters' architecture enabled the identification of the key DNA motives underlying their ERY-inducible/regulatable character. (Trassaert et al. 2017) studied a 300-bp fragment upstream from the *EYK1* gene, and demonstrated that this element (pEYK300) governs ERY-inducible expression, quenched on glucose (GLUC) and glycerol (GLY) media. Two functional sequences were identified within pEYK300–motif A (GGAAAGCCGCY) and motif B (CNTGCATWATCCGAYGAC), followed by the TATA box. Interestingly, further studies, covering an expanded fragment pEYK450 showed that this regulatory element operates in two directions (Vidal et al. 2023) (is a

Received 27 March 2024; **revised** 25 June 2024; **accepted** 20 August 2024

bidirectional promoter), and the centrally located motives AB are flanked with two TATA boxes, operating divergently (towards *EYK1* and *EYL1* genes). Following studies by Park et al. (2019), revealed a similar architecture of the *EYD1* promoter, encompassing motif A (ANTTNNNTTTCCNNATNNGG), a motif B (CGGNNCTNNATTGAGAANNC), and a putative TATA box (GATATAWA).

In the next step, these motives and different core promoters were shuffled to create a set of tuneable, titratable ERY-inducible promoters of various strengths and stringency of expression induction (Trassaert et al. 2017, Park et al. 2019, Vidal et al. 2023); including also—pEYK300 (300 bp of pEYK1, one A motif and one B motif), pEYK3AB (triple A motif), and pEYK5AB (quintuple A motif), used in this study. Their strength and titratability, by both the chemical inducer concentration and the number of upstream functional motives, were studied in detail. Strikingly, the transcriptional activity of the pEYK300-based promoters increased with the increasing number of the A motives, from 1 to 4, which was found to be the key element responsible for ERY-inducibility. Motif B was suggested to be more involved in GLUC repression. However, a further increase in the A motives number did not allow for further promoter titration (Park et al. 2019). Since the overall higher titers of a reporter protein could be reached in that study (when pTEF core was used), the most credible early hypothesis was that insufficient abundance of the ERY-inducible TF limited further titration. Hence, knowing from previous studies on *EUF1* (Rzechonek et al. 2017, 2024) (discussed hereafter), and its direct effect on *EYK1* expression (Mirończuk et al. 2018), our primary aim was to test if the ERY-inducible promoter could be further titrated by the increased abundance of a compatible TF—*EUF1*.

Following the former, pioneering study (Rzechonek et al. 2017), a very recent investigation into the *EUF1*'s regulome (Rzechonek et al. 2024) shed some new light on its operation mechanism. Apart from supporting previous notions on the 'erythritol utilization cluster' (*EYI1*, *EYI2*, *EYK1*, and *EYD1*–all upregulated in the presence of operable *EUF1*), it was demonstrated that sharp induction (>10-fold) of these genes requires three conditions to co-occur—(i) *EUF1* must be operable, (ii) ERY must be present, and (iii) GLY must be absent from the culture medium. The latter suggests the existence of some kind of 'GLY catabolite repression' inflicted in the ERY-metabolism regulation in *Y. lipolytica* (the genes did not reach high upregulation until GLY was present). The requirement for ERY presence suggests that some other phenomena, activated in response to ERY, must occur. Moreover, having the sets of genes deregulated in response to *EUF1* action, the authors run regulatory sequences mining in the search for conservative motives (presumably—*EUF1*-binding sites). The identified motives—the most common ATGCA/reverse TGCAT, followed by CGGAT/reverse ATCCG, and CGGCTT/reverse AAGCCG—were previously identified as elements of motif B (CNTGCATWATCCGAYGAC) and the central fragment of motif A (GGAAAGCCGCY). The previous studies identified the least frequent motif/motif A as the one required for promoter induction by ERY and erythrulose (Trassaert et al. 2017), and the most common motif/motif B as the GLUC repression element; which at the moment seems to stay in contrast to what has been inferred from the transcriptomics analysis. The question raised by the authors was whether *EUF1*'s biological function goes beyond orchestrating ERY utilization, as the scope of its regulome suggests a more extensive impact.

In our current research, we are interested in harnessing TFs as tools in engineering biotechnologically relevant traits in *Y. lipolytica*, like stress resistance and recombinant protein (rProt) synthesis (Gorczyca et al. 2023, Korpys-Woźniak and Celińska 2023, Gorczyca et al. 2024). From our high-throughput screens, we have learned that constitutive overexpression of *EUF1* (along with a reporter rProt) contributes to significant growth cessation (YaliFunTome database: https://sparrow.up.poznan.pl/tsdatabase/?page=gene&name=TF054), and that the addition of GLUC and inorganic nitrogen source significantly contributed to shaping this phenotype (no ERY was added). In this study, we aimed to test *EUF1* as the titrator of pEYK300-based promoters for rProt synthesis, using its native and spliced forms.

Materials and methods

Yeast strains and basic handling conditions

All the strains used in this study are listed in Table S1. *Yarrowia lipolytica* strains were constructed in the background of a double auxotrophic strain, deleted for major lipases, bearing zeta platform, and foremost—unable to utilize ERY– JMY7126 (here ECY_49) (Park et al. 2019). Complementation of auxotrophy was done by either transformation with a DNA construction or a solo cassette, as indicated in Table S1. If not reverted, auxotrophic strains were cultured in either rich or lysine-supplemented media.

The strains were routinely maintained as glycerol stocks in a −80°C ultrafreezer and freshly plated on YPD (yeast extract, peptone, dextrose) agar plates [(g/l): yeast extract, 10 (Biomaxima, Lublin, Poland), bactopeptone, 20 (BTL, Łódź, Poland), glucose, 20 POCh, Gliwice, Poland), and agar, 15 (BTL)] before experiments.

Molecular biology protocols

Standard molecular biology protocols were used in this study (Barth and Gaillardin 1996, Sambrook and Russell 2001). Restriction digestion of DNA fragments was done using *BamH*I, *Avr*II, *Cla*I, and *Not*I enzymes purchased from Thermo Fisher Scientific (Waltham, USA). DNA fragment amplification was run using Phire DNA polymerase (Thermo Fisher Scientific) and primers listed in Table S2. All the amplicons to be used as DNA construction parts were initially cloned in pCR Blunt II TOPO vector (Thermo Scientific, Waltham, USA) and restriction digested/sequenced (Genomed, Warsaw, Poland).

Cloning in JMP62 vectors was conducted using 200 U T4 DNA ligase (New England Biolabs, Ipswich, USA). DNA plasmids isolation, DNA fragment gel-extraction, and purification of DNA fragments were all conducted using an appropriate kit from A&A Biotechnology (Gdynia, Poland). All the reactions and protocols were used per the manufacturer's recommendations.

Escherichia coli JM109 transformations were done according to a heat-shock method. Following transformation, the bacterial strains were plated on LB (Luria–Bertani) medium [(g/l): yeast extract (BTL), 5; bactopeptone (BTL), 10; NaCl (POCh), 5], supplemented with kanamycin or ampicillin [Merck-Millipore, Darmstadt, Germany; at 40 or 100 (μg/ml), respectively] and agar [Biomaxima; 15 (g/l)]. *Yarrowia lipolytica* was transformed using the lithium acetate heat-shock protocol (Barth and Gaillardin 1996). Yeast transformants were plated on YNB (yeast nitrogen base) medium [g/l: YNB (Sigma-Aldrich/Merck-Millipore), 1.7; (NH4)2SO4 (PoCh), 5; glucose (PoCh), 20; supplemented with lysine at 0.8 g/l, when required]. Clones were selected after 24 h at 37°C for *E. coli*, and 48 h at 28°C for *Y. lipolytica*.

Culture conditions and samples analysis

Batch cultivations—media composition and conditions

The precultures and the main cultures were each time conducted in the same medium, composed as follows (g/l): YNB, 5.1; glucose,

50; ammonium sulfate and glutamic acid, 7.5 each; buffered with 0.2 M maleate buffer at pH 5.0 (Gorczyca et al. 2024). Alternatively, where indicated, YPD medium was used as the main culturing medium (g/l): yeast extract, 10; bactopeptone, 20; and glucose, 20. The precultures and the main cultures were run in 24-well DuetzSystem square plates in 2.5 ml working volume (EnzyScreen BV, Netherlands), agitated at 250 rpm, at 28°C. The precultures were developed for 18 h at 28°C. The main cultures were inoculated at 10% (v/v) and continued up to 60 h. Samples were collected, and the measurements of fluorescence (FL) and OD600 were done immediately, after applying appropriate dilutions (5–20-fold, in a sterile saline solution).

As indicated, the basic medium YNB could be supplemented with ERY at 5 g/l, to induce expression from the ERY-inducible promoters, or with osmoactive compounds sorbitol (SORB) and ERY (all purchased from POCh or Merck-Millipore) at higher concentrations, to induce an osmotic stress response. As assessed using an osmometer and previously described protocol (Kubiak-Szymendera et al. 2022), the basic medium's osmolality was 1 Osm/kg. The osmoactive compounds were added in a concentration to reach an osmolarity of 3 or 1.2 Osm/kg, considered 'high' and 'low' osmolarity, respectively. The amounts of the osmoactive compounds for 'high' osmolarity concentration were as follows (g/l): SORB, 360 and ERY, 244. To reach the 'low' osmolarity, the loads were 10-fold lower.

Analytical methods

Samples were analysed for growth and FL from the reporter protein (inYFP or RedStar) following dilution in 0.75% NaCl (POCh) to match a linear range of the methods. Absorbance was measured at 600 nm in transparent 96-well plates (Costar; Merck). FL was determined at ex/em 510/550 and 550/595 nm for inYFP and RedStar, respectively, in black opaque plates (Thermo Fisher Scientific). Both measurements were done using a Tecan Spark automatic plate reader (Tecan Group Ltd., Mannedorf, Switzerland).

Determination of GLUC, ERY, SORB, citric acid, and MAN concentrations in the clarified postculturing medium was done by HPLC according to a previously described methodology (Korpys-Woźniak et al. 2020). Briefly, the analysis was run using an Agilent Technologies 1200 series chromatograph (Agilent Technologies, Santa Clara, CA, USA) equipped with a refractive-index detector (G1362A) and a Rezex ROA-Organic Acid H + column (Phenomenex, Torrance, CA, USA). Operating conditions were as follows: 0.005 N H2SO4 as eluent at a flow rate of 0.6 (ml/min); the column temperature was set at 40°C. External standards (Sigma Aldrich) were used for the identification and quantification of the peak areas in chromatograms (analysed using ChemStation for LC 3D software; Agilent).

Transcriptomics data analysis

Transcriptomics datasets reanalysis

The transcriptomics datasets reanalysed in this study were originally deposited in SRA BioProject—NCBI (nih.gov) database under Bioproject number ID 893604 (PRJNA893604) by Rzechonek et al. (2024). The following BioSamples were considered in this reanalysis: BioSample: SAMN31428143; sample name: MK1_gluc_24; SRA: SRS15523360 versus BioSample: SAMN31428139; sample name: MK1_ery_24; SRA: SRS15523356; and BioSample: SAMN31428144; sample name: MK1_gluc_32; SRA: SRS15523361 versus BioSample: SAMN31428140; sample name: MK1_ery_32; SRA: SRS15523357. The two comparisons correspond to differential transcriptomes of *Y. lipolytica* strain MK1 (operable ERY pathway; functional *EUF1*) under ERY provision (YNB + 50 g/l GLY + 50 g/l erythritol—samples 'ERY') versus its lack (YNB + 50 g/l GLY + 50 g/l GLUC—samples 'GLUC') in the culture medium, at 24 and 32 h of culturing.

The datasets were reanalysed (Genomed). *De novo* intron mapping was conducted according to the previously described methodology (Korpys-Woźniak and Celińska 2023). Briefly, splice site detection was carried out using the SGSeq package (Goldstein et al. 2016) in the R programming language. Mapping of exons and introns was conducted in the IGV program (Robinson et al. 2011). Coverage depth was calculated by SamTools (Li et al. 2009).

The data were also analysed to identify differentially expressed genes (DEGs). Generally, this analysis was conducted according to the previously described protocol (Korpys-Woźniak and Celińska 2021). Briefly, the genes were mapped to a reference genome *Y. lipolytica* GCF_000002525.2 using Hisat2 software (Kim et al. 2019). Reads for a specific gene were counted using HTseq (Anders et al. 2015). Differential analysis of sequence read count data was conducted using edgeR (Chen et al. 2023). In addition, since only a single biological replication (single run per condition) was made available in Bioproject number ID 893604, extra restrictions to statistical analysis were implemented to improve the significance of the observations. The considered dispersion of biological variation was bcv = 0.4. Only genes showing logFC (FC–fold change between normalized counts) > |3|, at *P*-value and FDR < .05 were considered significant.

Detection of cis-regulatory elements within pEUF800

Cis-regulatory elements within pEUF800 were detected using MatrixCatch (Deyneko et al. 2013), using the default setting and, *P*-value correction: Bonferroni step-down, *P*-value threshold = .05.

Statistical analysis

Statistical analysis was performed using Statistica 13 software (Tibco, CA, USA). After confirming homogeneity of variance (ANOVA), Tukey's HSD/RIR test was used to identify statistically homogenous groups of data at $P < .05$. Error bars indicate ± SD of biological triplicate, each analysed in technical duplicate.

Euf1 3D structure modelling and ligand docking

The Euf1 protein structure modelling was done in AlphaFold3 (Abramson et al. 2024), including basic 3D structure modelling, oligomerization, and DNA (motif A, 10 bp) binding. The accuracy of the entire structure was assessed using TM (template modelling) score and ipTM for the dimers (assess the accuracy of the predicted relative positions of the subunits within the complex). Structures were visualized and analysed using PyMol open-source version (Copyright (c) Schrodinger, LLC. Published under a BSD-like license. Version 3.0.3). Structures for further analyses were processed using Pdb tools (Rodrigues et al. 2018). Functional domain prediction was done using the InterPro tool (Paysan-Lafosse et al. 2023). Binding pockets were predicted, scored, and visualized using P2Rank (Krivák and Hoksza 2018, Jendele et al. 2019, Jakubec et al. 2022). The binding pockets prediction accuracy was assessed using: the P2Rank score, probability of pocket, and average conservation. The docking of ligand molecule (ERY/DNA) was executed and scored in Haddock (DNA and ERY) and Prodigy (PROtein binDIng enerGY prediction; only ERY) (Kurkcuoglu et al. 2018, Vangone et al. 2019, Honorato et al. 2021, 2024). Results were expressed in either Haddock score or Gibbs free energy (G). The former metric combines various energy

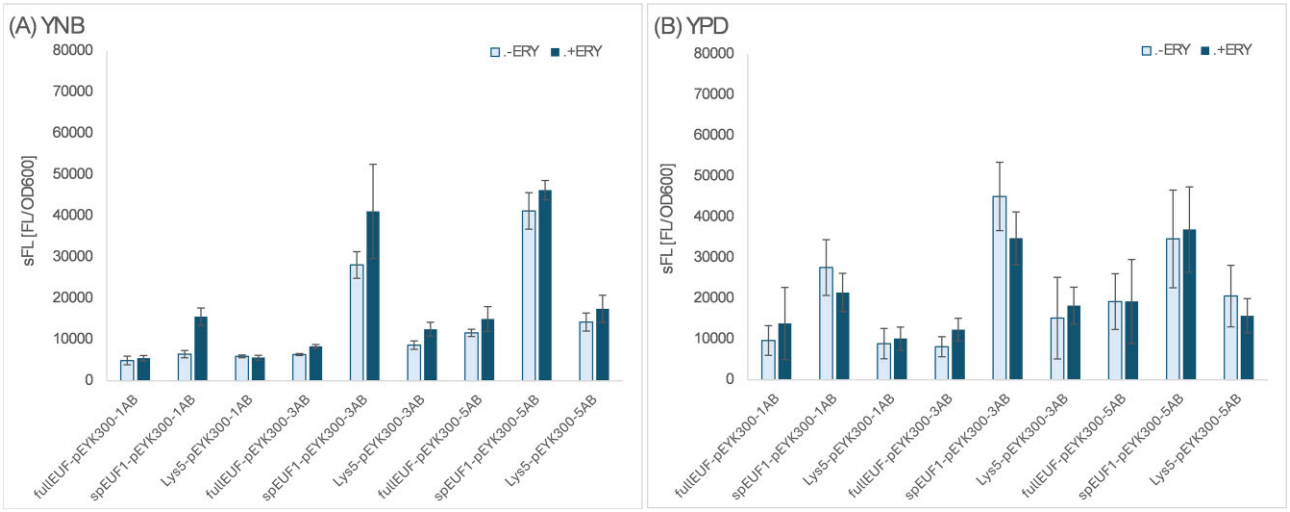


Figure 1. Titration of ERY-inducible/Euf1-dependent promoters (pEYK300-1AB/3AB/5AB) under co-overexpression of full (fullEUF1) or spliced (spEUF1) TF and ERY provision (ERY+; dark blue, solid) or absence (ERY−; light blue, framed) in minimal (YNB; A) or rich (YPD; B). Variant 'Lys5−' indicates prototrophic control without the TF co-overexpression. Y-axis: protein synthesis capacity defined as raw fluorescence readouts normalized per biomass, expressed as sFL. The type of promoter regulating expression is indicated in the x-axis. Bars indicate mean values from at least four biological replicates (four different subclones), each read in technical duplicate ± SD.

contributions to provide a measure of the quality of the docked complexes (both ERY and DNA), designed to be an indicator of the overall strength and stability of the interaction between the protein and DNA. It combines multiple relevant energy terms, i.e. attractive or repulsive forces between atoms due to van der Waals interactions, electrostatic interactions between charged groups in the protein and DNA—ionic interactions/electrostatic energy, and the energy associated with the removal of solvent (water) molecules from the interaction interface—desolvation energy, combines enthalpy and entropy into a single value. Results of Prodigy analysis (ERY binding) are expressed as changes in Gibbs free energy, ΔG, expressed in kcal/mol. Results $\Delta G < -9$ kcal/mol indicate very high affinity, between −7 and −9 kcal/mol: high affinity, between −5 and −7 kcal/mol: moderate affinity, and > −5 kcal/mol: low affinity.

Results and discussion

EUF1 as a titrator of ERY-induced promoters

To answer the principal question of this study, on further titratability of pEYK300-based ERY-inducible promoters using a compatible TF, the *EUF1* gene was cloned and overexpressed under the control of a constitutive promoter pTEF. This way, its load per cell was increased. The effect of the increased abundance of *EUF1* on the transcriptional activity of pEYK300-1AB, pEYK300-3AB, and pEYK300-5AB was investigated; which was gauged by FL from the reporter protein cloned under the pEYK300-based ERY-inducible promoters. Transformed strains were cultured with and without the inducer (ERY+/ERY−) in the basic (YNB) and rich medium (YPD). Results on normalized FL (sFL–specific FL) are shown in Fig. 1 and Table 1. Foremost, we observed that overexpression of unspliced (full) EUF1 yielded no significant differences when compared to the control variants ('Lys5-pEYK300-xAB'; FC below or close to 1.0 in Table 1). On the other hand, the increased load of spEuf1 significantly enhanced transcription from the pEYK-based promoters, reaching over 3-fold higher rProt amounts, in specific cases (Table 1, Fig. 1). Notably, the results obtained in the YNB medium were more consistent than those from the YPD medium. In the former, a clear effect of spEUF1 overexpression could be seen, and a low background level from the control and fullEUF1 variants. This differentiation of the effects reached in different media highlights the importance of nutrient availability for the investigated phenomena; which were depleted from the YNB variant at the time of sample collection (HPLC results, not shown; elaborated on and experimentally verified in more detail hereafter).

Considering the better controllable variant (YNB), in all the analysed cases (overexpression of sp/fullEUF1 or its basic, native expression in Lys5 prototrophs) the amounts of rProt cloned under pEYK300-1/3/5AB increased with the increasing number of A motif (pEYK300 < pEYK300-3AB < pEYK300-5AB) (Table 1). This observation is consistent with previous notions by (Trassaert et al. 2017, Park et al. 2019). That effect was independent of the introduced genetic modification (sp/full EUF1 or sole auxotrophy complementation). The variant with 3A motives rendered from ∼30% to over 4-fold more of the sFL units than the basic variant (1A). The incorporation of two further A motives (the 5A variant) triggered ∼12% to ∼80% improvement over the 3A. This notion corresponds to the previously stated problem, that further promoter titration by engineering its architecture is not linear and some upper limit was reached. This study shows that even upon enhanced availability of the main titrator, Euf1, this effect cannot be overcome. It could be speculated that intrinsic limitations make the transcription from pEYK-5AB saturated, and further titration cannot be achieved. Our observation stays in sharp contrast to the results shown by Blazeck et al. (2011) demonstrating an unlimited, linear increase in the transcriptional activity of the promoters with 1 to (at least) 24 upstream activating sequences.

Some leakage of the pEYK300-3AB/5AB promoter was observed in the absence of ERY, especially in the case of the pEYK300-5AB promoter (light blue bars in Fig. 1). Strikingly, the promoter leakage was unexpectedly high when *spEUF1* was overexpressed. This surprising observation was 'titratable'—higher sFL amounts were reached with increasing promoter strength, in the absence of the chemical inducer ($r = 0.99$), indicating that now the number of A

Table 1. Normalized amounts of the reporter protein (RedStar) expressed under ERY-inducible/Euf1-dependent promoters (pEYK300-1AB/3AB/5AB) under co-overexpression of full (fullEUF1) or spliced (spEUF1) TF and ERY provision (Inducer+) or absence (Inducer−) in minimal (YNB) or rich (YPD) medium. The sFL values were calculated based on fluorescence readouts FL (sFL—specific FL) normalized per biomass. FC—fold change values for sFL read for corresponding strains and conditions, differing in the indicated measures: number of 'A' motives 3AB versus 1AB, or 5AB versus 3AB, or in the type of co-overexpressed TF EUF1–spliced (spEuf1) or its lack ('control').

Strain	Inducer ±	Medium	sFL	FC 3AB versus 1AB	FC 5AB versus 3AB	FC sp/fullEUF1 versus control
fullEUF-pEYK300-1AB	ERY−	YNB	4922.3			0.83
spEUF1-pEYK300-1AB	ERY−	YNB	6458.2			1.09
Lys5-pEYK300-1AB	ERY−	YNB	5941.4			
fullEUF-pEYK300-3AB	ERY−	YNB	6371.6	1.29		0.74
spEUF1-pEYK300-3AB	ERY−	YNB	28080.8	4.35		3.25
Lys5-pEYK300-3AB	ERY−	YNB	8640.7	1.45		
fullEUF-pEYK300-5AB	ERY−	YNB	11605.1		1.82	0.82
spEUF1-pEYK300-5AB	ERY−	YNB	41197.2		1.47	2.90
Lys5-pEYK300-5AB	ERY−	YNB	14225.1		1.65	
fullEUF-pEYK300-1AB	ERY+	YNB	5466.2			0.97
spEUF1-pEYK300-1AB	ERY+	YNB	15565.7			2.77
Lys5-pEYK300-1AB	ERY+	YNB	5620.0			
fullEUF-pEYK300-3AB	ERY+	YNB	8356.5	1.53		0.67
spEUF1-pEYK300-3AB	ERY+	YNB	41084.5	2.64		3.29
Lys5-pEYK300-3AB	ERY+	YNB	12499.3	2.22		
fullEUF-pEYK300-5AB	ERY+	YNB	14970.8		1.79	0.86
spEUF1-pEYK300-5AB	ERY+	YNB	46213.7		1.12	2.65
Lys5-pEYK300-5AB	ERY+	YNB	17470.9		1.40	
fullEUF-pEYK300-1AB	ERY−	YPD	9652.6			1.09
spEUF1-pEYK300-1AB	ERY−	YPD	27578.9			3.11
Lys5-pEYK300-1AB	ERY−	YPD	8873.1			
fullEUF-pEYK300-3AB	ERY−	YPD	8132.6	0.84		0.54
spEUF1-pEYK300-3AB	ERY−	YPD	45050.9	1.63		2.97
Lys5-pEYK300-3AB	ERY−	YPD	15153.4	1.71		
fullEUF-pEYK300-5AB	ERY−	YPD	19186.0		2.36	0.93
spEUF1-pEYK300-5AB	ERY−	YPD	34610.0		0.77	1.68
Lys5-pEYK300-5AB	ERY−	YPD	20563.4		1.36	
fullEUF-pEYK300-1AB	ERY+	YPD	13823.1			1.37
spEUF1-pEYK300-1AB	ERY+	YPD	21437.9			2.12
Lys5-pEYK300-1AB	ERY+	YPD	10117.9			
fullEUF-pEYK300-3AB	ERY+	YPD	12318.4	0.89		0.68
spEUF1-pEYK300-3AB	ERY+	YPD	34777.2	1.62		1.91
Lys5-pEYK300-3AB	ERY+	YPD	18208.3	1.80		
fullEUF-pEYK300-5AB	ERY+	YPD	19264.7		1.56	1.23
spEUF1-pEYK300-5AB	ERY+	YPD	36960.9		1.06	2.35
Lys5-pEYK300-5AB	ERY+	YPD	15710.5		0.86	

motives became limiting. Nevertheless, the new ERY-less system with *spEUF1* as the inducer although it meets the requirements of being titratable by the number of A motives, is not inducible, so it is not useful as a controllable regulatory part for synthetic biology.

Altogether the above experiment showed that overexpression of *spEUF1* allows for further titration of the pEYK300-based promoters; and indeed, some upper limit of the ERY-induced rProt synthesis is reached above 3A motives in the promoter structure. Strikingly, an increased abundance of the spliced variant of *EUF1*, sp*EUF1*, is sufficient to activate expression from pEYK300-3AB/5AB promoters in the absence of the inducing chemical compound. Such a phenomenon suggests some association between the inducer and the splicing event. The important questions raised after these experiments were whether ERY is the actual factor inducing the splicing event, and what ERY-responsive molecular scissors catalyse this phenomenon, if any. But before investigating these, the peculiar effect of high-nutrient availability had to be addressed.

EUF1-dependent expression in relation to carbon source load and under different chemical inducers

To investigate in more detail the observed differences in the pEYK-based promoters' behavior depending on the nutrient availability (YNB versus YPD media, Fig. 1A and B), and instructed by previous studies (Rzechonek et al. 2024) suggesting the importance of carbon source load on ERY-inducible phenomena, we designed an experiment in which the pEYK300-1AB promoter's activity was tested in the presence or absence of carbon source (GLUC), and under supplementation with different concentrations of ERY as well as SORB. The former was used as a specific inducer, while SORB, to inflict osmotic stress. It was interesting for us to investigate if the EUF1-dependent expression is specific to ERY or is responsive to osmostress as well; since hyperosmolality is commonly used as an inducer of ERY synthesis (Yang et al. 2015, Rakicka-Pustułka et al. 2020). Two different fluorescent reporter proteins were used to gauge the pEYK300 transcriptional activity (RedStar and inYFP). In addition, promoter pEUF800 (800 bp

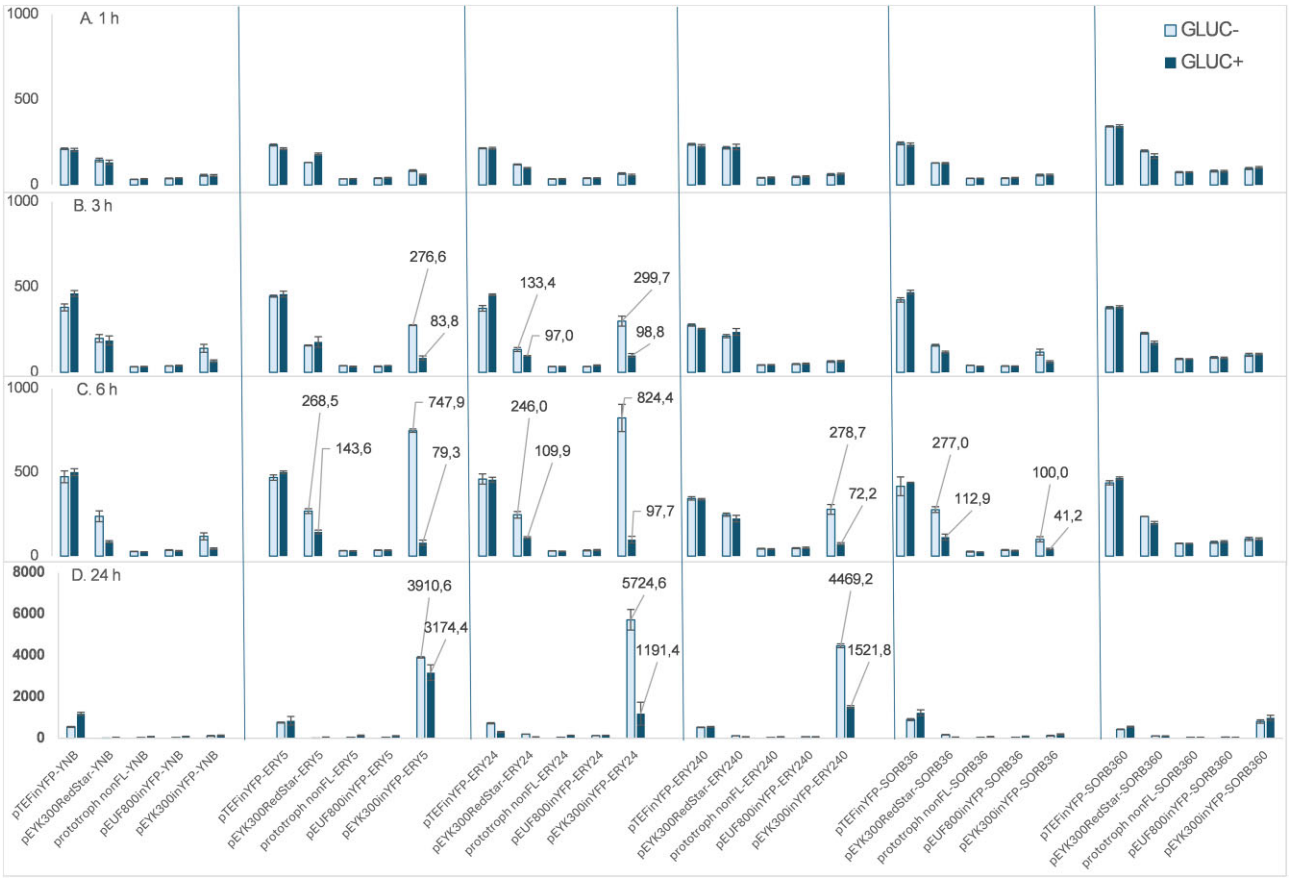


Figure 2. Normalized amounts of a reporter protein (inYFP/RedStar) expressed under different promoters (pTEF, pEYK300, and pEUF800) synthesized after 1 h (A), 3 h (B), 6 h (C), and 24 h (D) of incubation in the presence (GLUC+; dark blue, solid) or absence (GLUC−; light blue, framed) of glucose, in media supplemented with ERY at 5 g/l (ERY5), 24 g/l (ERY24), and 240 g/l (ERY240), or SORB at 36 g/l (SORB36), and 360 g/l (SORB360)—panels from left to right; indicated in the x-axis. Data are mean values from at least two (typically six) biological replicates (different subclones), each read in technical duplicate ± SD.

upstream from *EUF1*'s ATG) was considered in this analysis, as we identified the key A and B motives in its sequence (sequence analysis is presented hereafter in this section). Of technical note relevant to this experiment, SORB was not consumed, nor produced; ERY was not consumed (due to Δ*eyk1* genotype) and not synthesized above 1 g/l (HPLC data not shown). The strains were initially grown in the YNB basic medium, and after reaching the stationary phase of growth they were washed, and inoculated in media with or without a carbon source, in the presence or absence of ERY and SORB at different concentrations. The results of this experiment in terms of specific rProt synthesis are shown in Fig. 2 and Table 2.

Data presented in Table 2, showing FC of sFL read in media devoid of glucose over those with glucose at 20 g/l, confirmed our hypothesis. Indeed lack of carbon source acts as a strong inducer of expression from pEYK300 in the presence of ERY (all concentrations tested) and SORB (36 g/l). The increase in sFL was gradual along the incubation time, reaching a very high level of 8–10-fold higher sFL at 6 h of incubation (ERY5 and ERY24). It decreased afterwards at 24 h, because of glucose consumption in the control variant (with initial glucose at 20 g/l). The results of this experiment should be considered in industrial practice exploiting pEYK300-based *Y. lipolytica* expression platforms, by adopting the final starvation stage. Another relevant point would be removal of the motif B from the pEYK-based synthetic promoters to eliminate the 'glucose/glycerol catabolite repression' mechanism.

The highest, genotype-based induction of inYFP/RedStar synthesis was observed for pEYK300-driven expression under exposure to ERY at 5 and 24 g/l and, surprisingly, SORB at 36 g/l (Fig. 2C, Table 2), meaning that SORB can be considered an alternative chemical inducer of the pEYK-based systems in *Y. lipolytica*, without the need for Δ*eyk1* background. Yet, the concentration must be subjected to optimization. In general, higher concentrations of the chemical inducers (240 and 360 g/l) contributed to a decrease in the rProt levels in this strain. To our interpretation, the profile of sFL levels across increasing concentrations of the inducers resulted from the combination of the following: (i) under severe osmotic stress transcriptional activity of the promoters is quenched, as the cell promotes the osmolyte synthesis (and not consumption) to survive and (ii) under these conditions, the global cellular stress response drives silencing of rProts synthesis via multiple ways (experimentally defined at transcriptional and proteomics level previously; Kubiak-Szymendera et al. 2022). In this context, we would like to explain that the uniform enhancement in the fluorescence under the SORB360 condition (Fig. 2) is a result of enhanced background fluorescence rather than genotype-driven observation. Even though the excitation/emission wavelengths were carefully optimized for inYFP and RedStar readouts in our basic medium, under exposure to SORB360, some compounds

Table 2. FC in the fluorescence from a reporter protein (inYFP or RedStar expressed from pTEF, pEYK300, or pEUF800 promoters) read in media supplemented with chemical inducers (ERY/SORB at different concentrations: 5, 24, 240, 36, and 360 g/l, as indicated in the conditions' name) without glucose (GLUC−) over the readout in the control medium with glucose (GLUC+) at 20 g/l. Values were colour-coded for easier evaluation (blue—the lowest; red—the highest).

Strain	1 h	3 h	6 h	24 h
pTEFinYFP-YNB	1.03	0.82	0.95	0.47
pEYK300RedStar-YNB	1.10	1.07	2.80	0.11
prototroph nonFL-YNB	0.93	0.99	1.07	0.44
pEUF800inYFP-YNB	0.95	0.95	1.14	0.45
pEYK300inYFP-YNB	1.02	2.15	2.70	0.89
pTEFinYFP-ERY5	1.10	0.97	0.94	0.90
pEYK300RedStar-ERY5	0.73	0.88	1.87	0.21
prototroph nonFL-ERY5	0.98	1.12	1.07	0.36
pEUF800inYFP-ERY5	0.95	0.89	1.02	0.43
pEYK300inYFP-ERY5	1.45	3.30	9.43	1.23
pTEFinYFP-ERY24	1.00	0.83	1.01	2.40
pEYK300RedStar-ERY24	1.21	1.38	2.24	3.21
prototroph nonFL-ERY24	0.98	1.00	1.10	0.33
pEUF800inYFP-ERY24	0.98	0.83	0.93	0.92
pEYK300inYFP-ERY24	1.15	3.03	8.44	4.81
pTEFinYFP-ERY240	1.04	1.09	1.01	0.96
pEYK300RedStar-ERY240	0.98	0.90	1.11	1.76
prototroph nonFL-ERY240	0.95	0.97	1.05	0.52
pEUF800inYFP-ERY240	0.94	0.94	0.97	1.01
pEYK300inYFP-ERY240	0.95	0.96	3.86	2.94
pTEFinYFP-SORB36	1.04	0.90	0.95	0.73
pEYK300RedStar-SORB36	1.01	1.33	2.45	2.70
prototroph nonFL-SORB36	1.01	1.15	1.09	0.59
pEUF800inYFP-SORB36	0.94	1.00	1.09	0.48
pEYK300inYFP-SORB36	0.98	1.92	2.43	0.69
pTEFinYFP-SORB360	1.00	0.99	0.94	0.77
pEYK300RedStar-SORB360	1.18	1.32	1.20	1.07
prototroph nonFL-SORB360	1.00	1.00	1.01	1.03
pEUF800inYFP-SORB360	1.01	1.03	0.95	1.26
pEYK300inYFP-SORB360	0.94	0.96	1.01	0.82

interfering with the fluorescence readout at the wavelengths used are synthesized. It is known, that exposure to hyperosmolarity induces synthesis of i.a. proline or betaine, which could contribute to that interference (Thomas et al. 1994).

Trying to understand the unexpected lack of significant induction of pEUF800 under ERY provision, we analysed the promoter's sequence in search of known ERY-related motifs. The whole intergenic space between ATGs of *EUF1* (YALI0F01562g) and, located upstream, YALI0F01540 g is slightly more than 3100 bp. Previously, for ERY-inducible genes (*EYK1*, *EYL1*, and *EYD1*), typically 300–500 bp upstream from a gene's ATG were considered sufficient to be used as a synthetic promoter (Trassaert et al. 2017, Park et al. 2019, Vidal et al. 2023). Within this pEUF800 region, a motif with a consensus sequence GGGAAAGTTTT was found seven times; and that was the most frequent cis-binding element within that region. Considering the cis-elements earlier identified as associated with ERY induction (box A, AAGCCG, or box B, ATGCA/ATCCG; Trassaert et al. 2017, Park et al. 2019, Rzechonek et al. 2024), a single motif A was found at position −276 upstream from ATG, while an element of box B (ATCCG) was found in its proximity—at position −230 (and beyond the selected fragment −2804). On the other hand, another piece of the core element of motif B (ATGCA that was found in the promoter region of all *EUF1*-regulated genes; Rzechonek et al. 2024), was identified in the promoter region of the other gene, sharing the intergenic region, YALI0F01540 g (−2206 and −3194 bp upstream from *EUF1*'s ATG), but not in pEUF800. A core consensus sequence of STRE motif (AGGGG; Estruch 2000) was identified at position −969, and its reverse (CCCCT) twice, at −563 and −1017. It is known that the STRE is functional in both orientations, and the core sequence is strictly conserved (with no restrictions on the neighbouring). The presence of a STRE-like sequence in the promoter does not imply the functionality of this element. It is known that the basic stress-induced activation is then observed, however, additional copies enhance that effect more than in an additive manner. Three copies of STRE in the pEUF800 suggest its strong upregulation in response to environmental stress (like SORB360). Interestingly, a single copy of the STRE motif was also found in the promoter region of pEYK300, approximately 100 bp upstream ATG, which could explain its upregulation in response to SORB36.

In summary, although pEUF800 contains all the necessary DNA motives to be regulated in response to ERY and, according to the just demonstrated glucose-repression mechanism, it remained transcriptionally inactive. We speculate that maybe the length of the promoter was for some reason inadequate, or some important regulatory element is contained in the *EUF1*'s intron, localized at the 5′ terminus of the gene. The intron was not cloned in our experiment, as immediately downstream from the pEUF800, the reporter gene was cloned.

In our auxiliary experiment (Fig. S1) on the expression level of native *EUF1* from its native promoter, we observed very high (>10-fold) upregulation of expression under SORB240 condition;

demonstrating the operability of the promoter, and its inducible character. Surprisingly, no such tremendous increase was observed under high ERY provision (to the same osmolality level) (Fig. S1; more data and discussion on *EUF1*'s expression under ERY supplementation will be given hereafter based on transcriptomics data). Previous studies have shown that ERY indeed has a very distinctive role in *Y. lipolytica*'s response to osmostress, as it provides resistance to hyperosmolality that is independent of the HOG pathway (Rzechonek et al. 2020). No other osmoprotectant (proline and trehalose) could account for such an effect. In our previous studies, the expression of *EUF1* was also slightly upregulated (logFC ~0.3; significant at $P < .05$) in response to oxidative stress, when ERY was not present in the culture medium (Korpys-Woźniak and Celińska 2021).

In summary, the above experiment demonstrated the key role of carbon source load on transcriptional activity from the EUF1-dependent, ERY-inducible promoters. The experimental data and the promoter sequence structure suggest that the ERY-inducible genes are, at the same time, GLUC/GLY-repressible genes. Further insight and a functional-mechanistic model of the promoter's activity are provided below.

Splicing pattern of *EUF1*

To answer the question on putative 'ERY-induced splicing' of *EUF1* transcript, we set out for chemostat cultures under ERY + and ERY− conditions to analyse the associated transcriptomes and map the splicing events within the *EUF1* transcript. In parallel, the other authors published datasets of transcriptomics analyses of *Y. lipolytica* strains maintained in batch bioreactor cultures in the presence of ERY or GLUC run in GLY-based medium (Rzechonek et al. 2024). RNAseq was conducted on samples collected at 24 and 32 h of the batch culture in a GLY-based medium (50 g/l of GLY) supplemented with ERY at 50 g/l (ERY24 and ERY32) and GLUC at 50 g/l (GLUC24 and GLUC32). The comparisons presented there, could not answer our questions, but the available raw datasets could be reanalysed to match our experimental plan.

Here, the splicing events in *EUF1* were mapped *de novo*, based on the predicted junction sites and transcript coverage depth. The transcript coverage depth analysis mapped the junction sites corresponding to those identified previously using bioinformatics prediction (gene start: coordinate: 243 493, followed by 110 bp exon 1; junction site coordinates: 243 604 followed by 78 bp intron; junction site coordinates: 243 683 followed by 2744 bp exon 2; gene end coordinate: 246 427). The following analysis aimed at determining the ratio between the two forms (spliced/unspliced) in respective samples: GLUC 24 h of culturing, ERY 24 h, GLUC 32 h, and ERY 32 h. The results are shown in Fig. 3 and Table S3.

The highest fraction of spliced *EUF1* transcripts was detected in samples from the stationary phase of growth when ERY was contained in the medium at ~50 g/l, while GLY was completely depleted [please refer to Fig. 2 in Rzechonek et al. (2024)]. Considering the normalized amounts of the transcript, the calculated 'count' of the *spEUF1* transcript was 4.42 units in that sample. In the previous time point (end of exponential growth phase) the ERY concentration was the same, but GLY was still present at 20 g/l (Rzechonek et al. 2024). In that sample, the *EUF1* splicing was more frequent, considering the percentage of spliced versus unspliced transcript, but the global counts of *spEUF1* mRNA were nearly halved (2.37 units). The difference comes from i.a. the difference in the expression level of *EUF1*, which was 1.69-fold higher in the stationary phase of growth when ERY was the sole carbon source

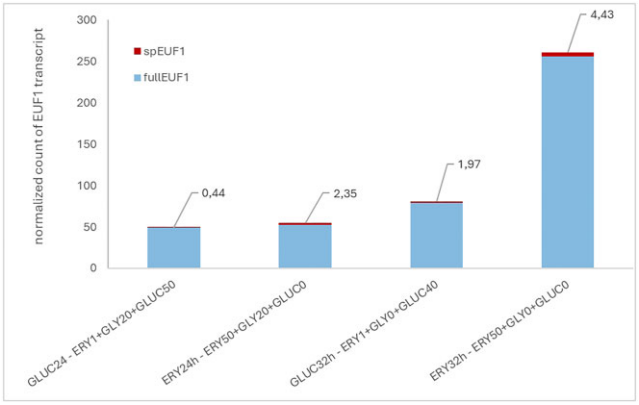


Figure 3. Splicing frequency of *EUF1* in the transcriptomes of *Y. lipolytica* cells in batch cultivations in the presence of glucose or erythritol at two different time points (differing in the carbon source concentration). Datasets published by Rzechonek et al. (2024). Raw datasets were reanalysed. GLUC, glucose, ERY, erythritol, GLYC, glycerol, time of culturing indicated in the sample name (—24 h and 32 h), the concentration of a given chemical compound is marked as a number following a specific metabolite, e.g. ERY50—erythritol at 50 g/l. Data labels indicate a normalized count of spliced forms of EUF1 transcript.

(logFC; induction is NS at $P < .05$; Table S3). To our understanding, the splicing machinery could operate at an equal pace, but the number of mRNA moieties was higher, so the splicing event could become the limiting step. The *EUF1* gene was also expressed in GLUC samples when ERY concentration in the medium was close to 0–1 g/l [values read from Fig. 2. in Rzechonek et al. (2024)]. In our previous studies, we also observed its expression in the absence of ERY (Korpys-Woźniak and Celińska 2021, 2023). Contrary to our presumptions, *EUF1* splicing occurred also in the GLUC32 sample (ERY at 1 g/l and GLUC close to 40 g/l); in that sample, 2.44% of *EUF1* transcripts were spliced. The expression level and splicing frequency were the lowest in GLUC24 samples; still, both processes were operable also in the absence of ERY and high concentration of GLUC and GLY (Fig. 3).

Altogether these findings suggest that *EUF1* is expressed at some basic level without ERY as a chemical inducer (more data provided hereafter). Some upregulation in expression takes place when: (i) ERY is present and (ii) other carbon sources are depleted. The *EUF1* transcript's splicing frequency is higher in the presence of ERY, but the process is also active in the absence of ERY. So, the splicing event cannot be considered as 'ERY-induced'. Considering calculations of the *spEUF1* transcript counts, it can be seen that the splicing events are promoted in the presence of ERY, but they are not strictly dependent on its presence.

Effect of *EUF1* splicing on its 3D structure and ligand-binding affinity

If (as evidenced above) *EUF1* is expressed and spliced either in the presence or absence of ERY and carbon source (GLUC/GLY) and the two forms of transcript coexist in the cell, it is highly probable that both forms are also translated (the intron lacks a stop codon) and hence full and spliced protein structures are present in the cell. So, we were interested in how the splicing event affects the protein structure, its DNA (motif A)-binding affinity, and if the ERY-binding domain could be identified. Therefore, we modelled full and spliced Euf1, docked the DNA and ERY ligands, and assessed the energy of binding (Fig. 4, Table 3; Supplementary Material Files). The modelled monomeric

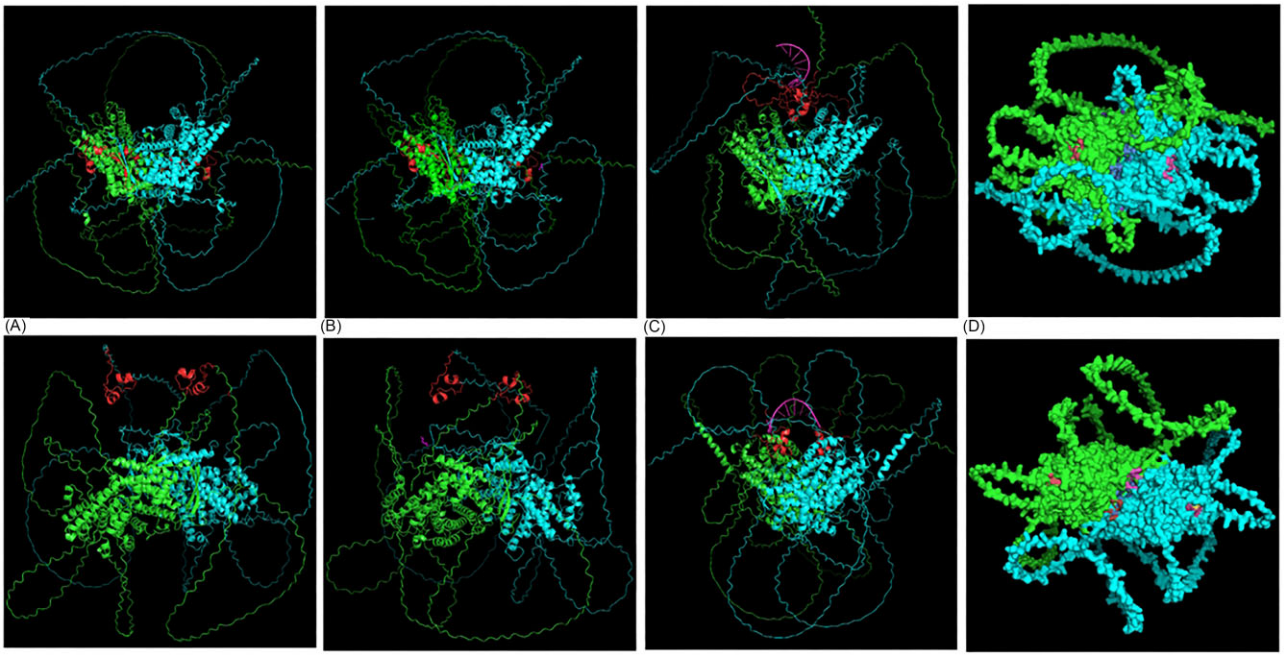
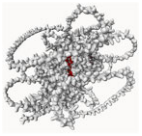
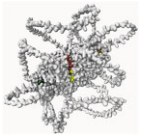


Figure 4. The 3D structure of fullEuf1 (upper row) and spEuf1 (bottom row) TF. The Euf1 structure is modelled as a dimer (monomers—light green and turquoise), Zn2Cys6 domains are shown in red. Docking of ERY (purple; panel B) to one of the two monomers, and DNA to double Zn-fingers (purple; panel C) is shown by a white arrow. The most conservative binding pocket is shown in panel D. in blue/red at the interface of the two monomers' interaction.

Table 3. The energy of interaction between sp/fullEuf1-modelled protein and ligands: ERY and ssDNA (10 bp of motif A). Prediction of binding pockets in sp/fullEuf1-modelled protein and corresponding scoring and parameters of the predicted binding pockets.

	Energy of interaction			
	fullEuf1-dimer		**spEuf1-dimer**	
	+ssDNA	**+ERY**	**+ssDNA**	**+ERY**
HADDOCK score*	**88.7 ± 4.8**	**−15.7 ± 0.8**	**6.0 ± 8.8**	**−12.0 ± 0.3**
Van der Waals energy	−99.2 ± 0.6	−10.6 ± 0.8	−81.0 ± 3.9	−7.7 ± 0.8
Electrostatic energy	−156.9 ± 26.1	0.1 ± 0.7	−34.6 ± 11.4	−0.7 ± 0.3
Desolvation energy	32.8 ± 4.4	−5.0 ± 0.2	19.1 ± 0.9	−4.2 ± 1.0
Sum of energies	−223.3	−15.5	−96.5	−12.6
Buried surface area	1651.7 ± 95.8	251.1 ± 4.7	1274.1 ± 55.5	200.6 ± 4.8
Interaction domain	Targeted at Zn-finger	Disordered	Targeted at Zn-finger	Disordered
Progidy ΔG**		−5.37		−5.26
	Binding pockets prediction			
P2Rank***	**fullEuf1-dimer**		**spEuf1-dimer**	
Rank	1		1	
Score	17.01		8.9	
Probability	0.792		0.523	
Number of residues	32		25	
Average conservation score	1.773		1.579	

*HADDOCK score is an overall energy term that combines various energy contributions to provide a measure of the quality of the docked complexes, designed to be an indicator of the overall strength and stability of the interaction between the protein and DNA. In general—lower HADDOCK scores indicate better binding affinity. It combines multiple relevant energy terms, i.e. attractive or repulsive forces between atoms due to van der Waals interactions, electrostatic interactions between charged groups in the protein and DNA—ionic interactions/electrostatic energy, and the energy associated with the removal of solvent (water) molecules from the interaction interface—desolvation energy. BSA refers to the surface area of the protein and DNA that becomes inaccessible to solvent (water) upon complex formation.
**Score expressed in changes in Gibbs free energy, ΔG, expressed in kcal/mol. ΔG between −5 and −7 kcal/mol indicates moderate affinity. Prodigy enables scoring protein–small ligand interaction. DNA is not accepted.
***P2Rank score. Average conservation (scale 0–4; the range of the per-residue of information content with higher values corresponding to higher conservation).

structures (fullEuf1 and spEuf1) received TM scores of 0.49 (Fig. S2, upper panels). Based on the primary sequence of the polypeptides we knew that Euf1 belongs to Zn(II)2Cys6 (Zn2C6) proteins. These are classified in class III (C6) zinc finger proteins, containing a DNA-binding domain that consists of six Cys residues bound to two Zn atoms. The following InterPro analysis enabled identification of the following three functional domains: (i) Zn2Cys6 DNA-binding domain (fullEuf1: 25–85 AA residues; spEuf1: 24–67 AA), (ii) Fungal TF domain (fullEuf1: 305–584 AA residues; spEuf1: 279–558 AA), and (iii) disordered regions (fullEuf1: 1–27, 112–220,

606–646, 757–789, 833–887, and 914–960 AA; spEuf1: 1–17, 86–194, 580–620, 739–763, 809–861, and 888–934 AA). Since the Zn2C6 proteins very frequently operate as dimers, after the initial modelling of monomers, we modelled homodimers as well (Fig. S2). The dimeric structures received TM scores of 0.52–0.54 for fullEuf1 and spEuf1, respectively (Fig. S2), indicating better ranked structures as dimers than monomers. The dimeric structures possess the double Zn-fingers, creating an elegant space for DNA binding, particularly visible for spEuf1 in (Fig. 4).

Interestingly, substantial differences in the 3D structures between fullEuf1 and spEuf1 dimers were noted (Fig. 4, Table 3; Supplementary Material Files). In the former, the Zn2C6 domains are localized at high distances, on the opposite sites of the dimer; only after DNA motif A docking, do the Zn fingers come together in spatial proximity. In contrast, the 3D structure of spEuf1, even devoid of DNA ligand, presents the Zn2C6 domains immediately next to each other, resembling a 'DNA-binding competent state'. Only a small spatial change happens when docking the DNA motif A to spEUF1. Furthermore, regarding the strength and stability of the interaction between the TF and DNA, the dimer of spEuf1 received much better scores in this matter than the one involving fullEuf1 (Table 3; Haddock score 88.7 ± 4.8 versus 6.0 ± 8.8 for full and spliced Euf1; the lower Haddock score, the better binding affinity); suggesting the indeed spEuf1 has better propensity for DNA binding.

In terms of ERY binding, it was predicted to dock on the surface of a disordered region in both dimers, fullEuf1 and spEuf1. While it is well-documented that the disordered regions possess an important regulatory role, here, ERY binding in such a region was weak and random (Table 3); rather forced by the user, than nature. The energy of interaction estimated by Prodigy was ΔG between −5.26 and −5.37 kcal/mol, indicating moderate/low affinity. Moreover, ERY docking was predicted to localize at the disordered regions, even though several binding pockets were predicted within sp/fullEuf1 (Fig. 4). Another interesting difference between structures fullEuf1 and spEuf1 was revealed regarding the parameters of the most-highly ranked predicted binding pockets (by P2Rank; Table 3). For both dimeric structures, the primary binding pocket was localized on the opposite sites of the 3D structure in relation to Zn-fingers, however, the overall score, probability, and span were all higher for the fullEuf1 structure, when compared to its spliced variant. Notably, this domain becomes substantially modified after Euf1 splicing—it is predicted with significantly lower probability, is smaller, and less conserved. Considering the very good ratings of the binding pocket predicted for fullEuf1, and the results presented in Fig. 1, we carefully suggest that this binding domain plays an important role in functional regulation of fullEuf1, by binding a regulatory molecule. Due to structural changes of the binding pocket in spEuf1, the spliced variant may lose its regulatable character, as shown in Fig. 1. Further research into the identification of a potential interaction partner is foreseen. Preliminary computational assessment of a potential interaction with another TF inflicted in pEYK1 promoter regulation, Adr1 (discussed hereafter), suggest that these two proteins do not interact with each other via the binding pocket identified (Fig. S3).

Altogether these data suggest that Euf1 operates as a dimer, forming a double Zn-finger domain. When in its spliced variant, it is in a constitutive 'DNA-binding competent state'. ERY ligand most probably does not directly interact with the Euf1 dimer, either if formed by fullEuf1 or spEuf1; no domain for such an interaction was predicted. On the other hand, an interesting change to the predicted binding pocket (localized on the opposite side of the 3D structure to Zn fingers) happens when comparing the full and spliced Euf1. Based on functional studies, and the predicted structural changes we propose that this binding pocket plays an important role in functional regulation of fullEuf1, which is lost in spEuf1.

DEGs under spEUF1 abundance

To explain the unexpected lack of pEUF800-driven induction of inYFP synthesis in the presence of ERY (Fig. 2), we dug deeper into the previously obtained transcriptomics data (Rzechonek et al. 2024) in search of the *EUF1* gene expression profile. In the current analysis, we compared samples ERY24 versus GLUC24 and ERY32 versus GLUC32 (such comparisons were not presented previously). To our surprise, *EUF1*'s expression level (logFC) in the ERY medium (50 g/l) was only 0.14-fold (logFC) higher at 24 h and 1.69-fold (logFC) higher at 32 h, and the increase was not statistically significant (NS at P and FDR < .05; Table S3). Just to double-check that observation, we conducted an auxiliary RTqPCR gene expression analysis of *EUF1* gene in the cultures with high (244 g/l) and low (24 g/l) ERY load versus growth in sole GLUC medium. The results are shown in Fig. S1(B). Notably, at the sample collection time-point, the main carbon source (GLUC) was still present in the culture medium, as in ERY24 sample from Rzechonek et al. (2024). Our RTqPCR and (Rzechonek et al. 2024) transcriptomics assessment of *EUF1* expression level well aligned, showing a lack of significant induction of *EUF1* expression in the presence of ERY, at least until the alternative carbon source is still present in the culture media.

Such an observation well corresponds with the data presented in Table 2 and Fig. 2, where pEUF800-driven expression of inYFP was nonsignificant under ERY supplementation. It seems that the *EUF1* expression level is only slightly induced under ERY provision, but its activation is a result of both—this slight induction and the splicing event (Fig. 1). In pioneering studies on *EUF1* (Rzechonek et al. 2017), it was demonstrated the expression of *EUF1* is over 10-fold higher in the presence of ERY (50 g/l), which seems to stay in contrast to what was found in our current research and the transcriptomics (Rzechonek et al. 2024). However, minor technical differences in the experimental setup could account for the apparent discrepancy. Foremost, the previously observed high upregulation was observed in *Y. lipolytica* cultures grown in ERY as the sole carbon source. The only sample corresponding to that former condition is ERY32, however, GLY depletion could have happened just before the sample collection, and the condition of sole ERY provision could have just been initiated. Secondly, the RTqPCR primers designed in that previous study flanked the 78 bp intron at the 5′ end rendering two amplicons—178 bp versus 256 bp (Fig. S1C) and the target gene counts were normalized per *ACT1* (actin) gene. We used primers attached close to polyA tail (within 300 bp from the stop codon), and normalized the target transcript's counts on *SEC61*, as the internal calibrator, according to our previous results (Borkowska et al. 2020). All these small details could collectively contribute to the observed differences.

Finally, we looked for DEGs that exhibited differential expression under ERY addition versus GLUC variants at the corresponding time points. Only seven genes met our criteria and were considered significantly upregulated in ERY24 samples versus GLUC24. Three of them were the known ERY-utilization cluster members—*EYK1*—up by 6.99-fold, *EYD1*—5.14-fold, and *EYI1*—4.35-fold upregulation. The other upregulated DEGs were three members of the unknown family of secreted proteins YALI0A02783g, YALI0A13497g, and YALI0B06160g. A blastp

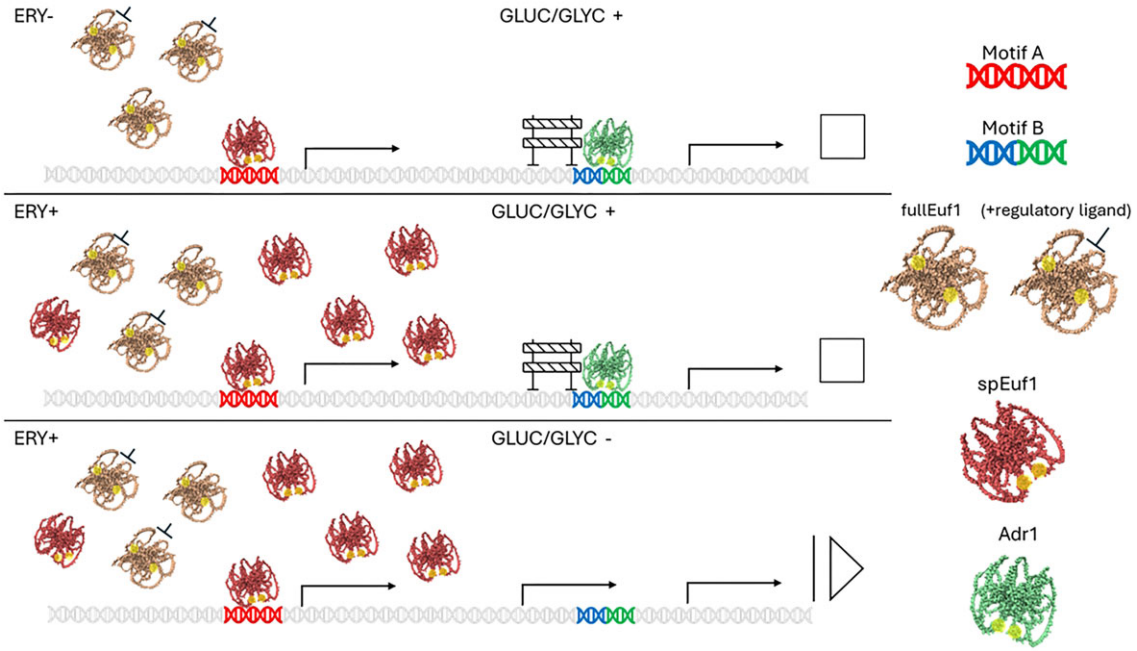


Figure 5. The model of ERY-inducible/carbon-source-repressible/Euf1-driven regulation of gene expression as an interplay between spEuf1 and Adr1 TFs within the ERY-inducible promoters. Functionally important motives and TFs are indicated in the legend and colour-coded. A regulation under three different conditions is presented: (i) top: ERY absent, carbon source (glucose, GLUC or glycerol, GLYC) is present—no substantial expression from the promoter, only leakage; (ii) ERY is present, carbon source is present—no substantial expression from the promoter, only leakage; (iii) ERY is present and the carbon source is exhausted—spEuf1-driven expression from the promoter is on. An additional regulation of the fullEuf1 via the predicted binding pocket is executed by an unidentified ligand (either inhibitor or activator).

similarity search indicated that the former is involved in the cell wall integrity pathway (Kamei et al. 2016). The last DEG upregulated at 24 h was identified as a TF YALI0F21923g, highly similar to *Saccharomyces cerevisiae*'s Adr1, a carbon-source-responsive zinc-finger TF, induces transcription of GLUC-repressed genes and promotes utilization of the alternative carbon sources (in *S. cerevisiae*—ethanol, GLY, and fatty acid). Here, the transcript of YALI0F21923 g was overrepresented in sample ERY24, where a mixture of GLY (20 g/l) and ERY (~50 g/l) was contained in the medium, versus control—GLUC24 (mix of GLY ~25 g/l and GLUC ~50 g/l). In the following time-point (ERY ~50 g/l, no GLUC, no GLYC), its expression was elevated in ERY32 versus GLUC32, but not significantly different (logFC 2.76, but NS at *P* and FDR < .05). So, either the gene's expression is repressed in the presence of GLUC, or—it is induced by ERY. Our previous studies involving YALI0F21923 g revealed that its overexpression supports efficient biomass formation and carbon utilization, but contributes to limited synthesis of recombinant protein (Gorczyca et al. 2024) (YaliFunTome database: TF116). It would be interesting to experimentally verify if this TF physically interacts with motif B, postulated to be involved in catabolite repression (Trassaert et al. 2017).

In the following time-point, altogether 41 DEGs were identified—35 upregulated and 6 downregulated (Table S3). Again, the ERY-utilization cluster genes *EYK1*, *EYD1*, and *EYI1* (5.9-, 5.6-, and 5.1-fold upregulation) were found amongst the most upregulated genes. But still, even though the other carbon source was nearly completely depleted, the *EUF1* gene upregulation level remained relatively low (1.64-fold; not statistically significant). A full list with assigned biological functions is provided in Table S3.

To answer the second question raised regarding *EUF1*'s operation in *Y. lipolytica* (about the 'molecular scissors' managing the *EUF1*'s splicing) we searched the deregulated genes for such that could be involved in the 78-bp intron removal. We expected the gene to be upregulated in ERY presence. None of the significantly deregulated genes had assigned or putative functions involving RNA splicing. As the significance criteria applied in this analysis were very stringent, we looked also at the remaining genes, identified but nonsignificantly deregulated. Directed by the pioneering study on splicing in *Y. lipolytica* (Mekouar et al. 2010), we specifically searched for U1 RNAs (YALI0B20936r and YALI0B14567r), and elements of mechanism degrading the intron-retaining transcripts UPF1/2 (YALI0E24629g and YALI0D23881g). In none of the ERY samples (24 or 32) the genes were significantly deregulated. The UPFs were slightly upregulated in ERY32 (~0.5 logFC), as was U1 RNA's component YALI0B20936r (to logFC 0.31), suggesting that the splicing-related mechanisms were intensified.

Altogether reanalysis of the transcriptomics data under here desired configuration showed that *EUF1* expression is not immediately induced once ERY is provided, but requires an extra requirement to be fulfilled, namely the lack of another carbon source. In fact, its transcription is active, even in the absence of the compatible, chemical inducer; likewise, its splicing. Interestingly, apart from the typical downstream targets of Euf1 belonging to the 'erythritol utilization cluster', another TF, Adr1, was deregulated in the ERY/GLUC differential transcriptomes. We postulate that it may play a key role in the GLUC-dependent regulation of pEYK promoter, demonstrated in section 3.2. A putative mechanism of the two TFs' interaction within the pEYK structure is provided below.

Summary and conclusions

In summary, this study shows that while overexpression of *spEUF1* enabled higher transcriptional activity of the synthetic pEYK-based promoter, it practically eliminated its inducible character.

Contrary to expectations, higher loads of *spEUF1* could not further significantly titrate (by more than 50%) the pEYK300-5AB promoter, surpassing the rProts yields reached with pEYK300-3AB. It is plausible that the metabolic burden imposed by high-level overexpression of the reporter and the titrator contributed to overall lower protein yields. Such overloading-based limitations have been shown and investigated previously in our studies (Gorczyca et al. 2022). Furthermore, we demonstrated that native expression of *EUF1* is not significantly induced by ERY—an increase in the expression level is observed, but is not statistically significant, as demonstrated by RTqPCR and RNAseq analyses. Likewise, the *EUF1* splicing event, while promoted in the ERY presence, was also active when the chemical inducer was absent from the experimental system. No ERY-induced spliceosome element was identified in the omics data, suggesting either their constant presence in the cells or induction of the *EUF1*-specific splicing event by some other factor than ERY.

The most striking observation of this study is that *EUF1*-dependent, ERY-inducible genes are at the same time GLUC/GLY-repressed; and that lack of a carbon source (either GLUC or GLY) is a strong inducer of ERY-dependent expression in *Y. lipolytica*. Based on (i) our current results (Fig. 2), (ii) the promoter sequence of the ERY/*EUF1*-dependent deregulated genes (Trassaert et al. 2017, Rzechonek et al. 2024), and (iii) the reanalysed omics data we propose a mechanism for the ERY-inducible-GLUC/GLY-repressible genes regulation, involving combined activity of spEuf1 and Adr1 TFs. The mechanism involves two previously identified cis-regulatory motives, motif A localized upstream from motif B, both identified as important (Trassaert et al. 2017) and common (Rzechonek et al. 2024) for the ERY/*EUF1*-deregulated genes. The latter suggests a uniform character of the following model of regulation for the genes with motives A and B in their promoter regions. The model's presumptions are: (i) spEuf1 exhibits affinity to a DNA motif A, while Adr1–affinity to motif B; (ii) fullEuf1 may also interact through an identified binding pocket with some inhibitor/activator; this binding pocket was predicted with significantly lower probability for spEuf1, and its structure was also changed (Table 3); it is thus probable that spEuf1 lost its regulatability via this domain; (iii) expression and splicing of *EUF1* are promoted but not reliant on ERY presence; (iv) Adr1 is in a DNA-binding state when GLUC/GLY is present; and (v) when bound to motif B (another carbon source is present), Adr1 constitutes a mechanical hindrance disallowing transcription governed by Euf1. Schematically the model is presented in Fig. 5.

Acknowledgements

We would like to acknowledge Szymon Turek from Genomed for his work with RNAseq data reanalysis. Martyna Przybylak is acknowledged for helping with chromatography analysis. Graphical abstract was Created with BioRender.com.

Author contribution

EC conceived the study, designed all the experiments, constructed the strains, conducted part of the experiments, analysed and interpreted the data, secured funding, and designed and wrote the manuscript, and revision. PKW conducted a significant part of the experiments. MG conducted a part of the experiments, ran all the bioinformatic protein structure modelings, and studied the protein–ligand interactions. JMN inspired the promoter-titration experiment and provided the background strains for that experiment.

Supplementary data

Supplementary data is available at *FEMSYR Journal* online.

Conflict of interest: None declared.

Funding

This study was financially supported by the National Science Centre, Poland, grant number 2021/41/B/NZ9/00086.

References

Abramson J, Adler J, Dunger J *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 2024;**630**:493–500.

Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;**31**:166–9.

Barth G, Gaillardin C. *Yarrowia lipolytica*. In: Wolf K (ed.), *Nonconventional Yeasts in Biotechnology a Handbook*. Berlin, Heidelberg: Springer, 1996, 313–88.

Blanchin-Roland S, Cordero Otero RR, Gaillardin C. Two upstream activation sequences control the expression of the XPR2 gene in the yeast *Yarrowia lipolytica*. *Mol Cell Biol* 1994;**14**:327–38.

Blazeck J, Liu L, Redden H *et al.* Tuning gene expression in *Yarrowia lipolytica* by a hybrid promoter approach. *Appl Environ Microbiol* 2011;**77**:7905–14.

Borkowska M, Białas W, Celińska E. A new set of reference genes for comparative gene expression analyses in *Yarrowia lipolytica*. *FEMS Yeast Res* 2020;**20**. https://doi.org/10.1093/femsyr/foaa059.

Carly F, Gamboa-Melendez H, Vandermies M *et al.* Identification and characterization of EYK1, a key gene for erythritol catabolism in *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* 2017;**101**:6587–96.

Carly F, Steels S, Telek S *et al.* Identification and characterization of EYD1, encoding an erythritol dehydrogenase in *Yarrowia lipolytica* and its application to bioconvert erythritol into erythrulose. *Bioresour Technol* 2018;**247**:963–9.

Chen Y, McCarthy D, Baldoni P *et al.* edgeR: differential analysis of sequence read count data User's Guide. Bioconductor, 2023.

Deyneko IV, Kel AE, Kel-Margoulis OV *et al.* MatrixCatch—a novel tool for the recognition of composite regulatory elements in promoters. *BMC Bioinf* 2013;**14**. https://doi.org/10.1186/1471-2105-14-241.

Estruch F. Stress-controlled transcription factors, stress-induced genes and stress tolerance in budding yeast. *FEMS Microbiol Rev* 2000;**24**:469–86.

Goldstein LD, Cao Y, Pau G *et al.* Prediction and quantification of splice events from RNA-seq data. *PLoS One* 2016;**11**:e0156132.

Gorczyca M, Białas W, Nicaud J-M *et al.* 'Mother(Nature) knows best'—hijacking nature-designed transcriptional programs for enhancing stress resistance and protein production in *Yarrowia lipolytica*; presentation of YaliFunTome database. *Microb Cell Fact* 2024;**23**:26.

Gorczyca M, Kaźmierczak J, Fickers P *et al.* Synthesis of secretory proteins in *Yarrowia lipolytica*: effect of combined stress factors and metabolic load. *Int J Mol Sci* 2022;**23**:3602.

Gorczyca M, Nicaud J-M, Celińska E. Transcription factors enhancing synthesis of recombinant proteins and resistance to

stress in *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* 2023;**107**:4853–71.
Honorato RV, Koukos PI, Jiménez-García B *et al.* Structural biology in the clouds: the WeNMR-EOSC ecosystem. *Front Mol Biosci* 2021;**8**. https://doi.org/10.3389/fmolb.2021.729513.
Honorato RV, Trellet ME, Jiménez-García B *et al.* The HADDOCK2.4 web server for integrative modeling of biomolecular complexes. *Nat Protoc* 2024. https://doi.org/10.1038/s41596-024-01011-0.
Jakubec D, Skoda P, Krivak R *et al.* PrankWeb 3: accelerated ligand-binding site predictions for experimental and modelled protein structures. *Nucleic Acids Res* 2022;**50**:W593–7.
Jendele L, Krivak R, Skoda P *et al.* PrankWeb: a web server for ligand binding site prediction and visualization. *Nucleic Acids Res* 2019;**47**:W345–9.
Juretzek T, Wang HJ, Nicaud JM *et al.* Comparison of promoters suitable for regulated overexpression of β-galactosidase in the alkane-utilizing yeast *Yarrowia hpolytica*. *Biotechnol Bioprocess Eng* 2000;**5**:320–6.
Kamei M, Yamashita K, Takahashi M *et al.* Involvement of MAK-1 and MAK-2 MAP kinases in cell wall integrity in *Neurospora crassa*. *Biosci Biotechnol Biochem* 2016;**80**:1843–52.
Kim D, Paggi JM, Park C *et al.* Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 2019;**37**:907–15.
Korpys-Woźniak P, Celińska E. Global transcriptome profiling reveals genes responding to overproduction of a small secretory, a high cysteine- and a high glycosylation-bearing protein in *Yarrowia lipolytica*. *Biotechnol Rep* 2021;**31**:e00646.
Korpys-Woźniak P, Celińska E. Molecular background of HAC1-driven improvement in the secretion of recombinant protein in *Yarrowia lipolytica* based on comparative transcriptomics. *Biotechnol Rep* 2023;**38**:e00801.
Korpys-Woźniak P, Kubiak P, Białas W *et al.* Impact of overproduced heterologous protein characteristics on physiological response in *Yarrowia lipolytica* steady-state-maintained continuous cultures. *Appl Microbiol Biotechnol* 2020;**104**:9785–800.
Krivák R, Hoksza D. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *J Cheminform* 2018;**10**:39.
Kubiak-Szymendera M, Skupien-Rabian B, Jankowska U *et al.* Hyperosmolarity adversely impacts recombinant protein synthesis by *Yarrowia lipolytica*—molecular background revealed by quantitative proteomics. *Appl Microbiol Biotechnol* 2022;**106**:349–67.
Kurkcuoglu Z, Koukos PI, Citro N *et al.* Performance of HADDOCK and a simple contact-based protein–ligand binding affinity predictor in the D3R Grand Challenge 2. *J Comput Aided Mol Des* 2018;**32**:175–85.
Li H, Handsaker B, Wysoker A *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;**25**:2078–9.
Madzak C, Blanchin-Roland S, Cordero Otero RR *et al.* Functional analysis of upstream regulating regions from the *Yarrowia lipolytica* XPR2 promoter. *Microbiology* 1999;**145**:75–87.
Madzak C, Tréton B, Blanchin-Roland S. Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast *Yarrowia lipolytica*. *J Mol Microbiol Biotechnol* 2000;**2**:207–16.
Mekouar M, Blanc-Lenfle I, Ozanne C *et al.* Detection and analysis of alternative splicing in *Yarrowia lipolytica* reveal structural constraints facilitating nonsense-mediated decay of intron-retaining transcripts. *Genome Biol* 2010;**11**. https://doi.org/10.1186/gb-2010-11-6-r65.
Mirończuk AM, Biegalska A, Zugaj K *et al.* A role of a newly identified isomerase from *Yarrowia lipolytica* in erythritol catabolism. *Front Microbiol* 2018;**9**. https://doi.org/10.3389/fmicb.2018.01122.
Ogrydziak D. Acid and alkaline extracellular proteases of *Yarrowia lipolytica*. In: Barth G (ed.), *Yarrowia Lipolytica, Microbiology Monographs* 25. Berlin, Heidelberg: Springer, 2013, 77–97.
Park Y-K, Korpys P, Kubiak M *et al.* Engineering the architecture of erythritol-inducible promoters for regulated and enhanced gene expression in *Yarrowia lipolytica*. *FEMS Yeast Res* 2019;**19**:1.
Paysan-Lafosse T, Blum M, Chuguransky S *et al.* InterPro in 2022. *Nucleic Acids Res* 2023;**51**:D418–27.
Rakicka-Pustułka M, Mirończuk AM, Celińska E *et al.* Scale-up of the erythritol production technology—process simulation and techno-economic analysis. *J Clean Prod* 2020;**257**:120533.
Robinson JT, Thorvaldsdóttir H, Winckler W *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011;**29**:24–6.
Rodrigues JPGLM, Teixeira JMC, Trellet M *et al.* pdb-tools: a swiss army knife for molecular structures. *F1000Res* 2018;**7**:1961.
Rzechonek DA, Neuvéglise C, Devillers H *et al.* EUF1—a newly identified gene involved in erythritol utilization in *Yarrowia lipolytica*. *Sci Rep* 2017;**7**. https://doi.org/10.1038/s41598-017-12715-7.
Rzechonek DA, Szczepańczyk M, Borodina I *et al.* Transcriptome analysis reveals multiple targets of erythritol-related transcription factor EUF1 in unconventional yeast *Yarrowia lipolytica*. *Microb Cell Fact* 2024;**23**:77.
Rzechonek DA, Szczepańczyk M, Wang G *et al.* Hog-independent osmoprotection by erythritol in yeast *Yarrowia lipolytica*. *Genes* 2020;**11**:1–15.
Sambrook J, Russell D. *Molecular Cloning: A Laboratory Manual*. 3rd edn. New York: Cold Spring Harbor Laboratory Press, 2001.
Shabbir Hussain M, Gambill L, Smith S *et al.* Engineering promoter architecture in oleaginous yeast *Yarrowia lipolytica*. *ACS Synth Biol* 2016;**5**:213–23.
Thomas KC, Hynes SH, Ingledew WM. Effects of particulate materials and osmoprotectants on very-high-gravity ethanolic fermentation by *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 1994;**60**:1519–24.
Trassaert M, Vandermies M, Carly F *et al.* New inducible promoter for gene expression and synthetic biology in *Yarrowia lipolytica*. *Microb Cell Fact* 2017;**16**. https://doi.org/10.1186/s12934-017-0755-0.
Vangone A, Schaarschmidt J, Koukos P *et al.* Large-scale prediction of binding affinity in protein–small ligand complexes: the PRODIGY-LIG web server. *Bioinformatics* 2019;**35**:1585–7.
Vidal L, Lebrun E, Park YK *et al.* Bidirectional hybrid erythritol-inducible promoter for synthetic biology in *Yarrowia lipolytica*. *Microb Cell Fact* 2023;**22**. https://doi.org/10.1186/s12934-023-02020-6.
Yang LB, Dai XM, Zheng ZY *et al.* Proteomic analysis of erythritol-producing *Yarrowia lipolytica* from glycerol in response to osmotic pressure. *J Microbiol Biotechnol* 2015;**25**:1056–69.

Received 27 March 2024; **revised** 25 June 2024; **accepted** 20 August 2024

6. Podsumowanie i wnioski końcowe

Najnowsze podejścia w biologii syntetycznej i biotechnologii przemysłowej opierają się na koncepcji cyklu *design-build-test-learn* (DBTL; projektowanie-budowanie-testowanie-uczenie), który zakłada iteracyjne doskonalenie konstrukcji biologicznych. Każda kolejna iteracja pozwala na udoskonalenie projektów dzięki wiedzy zdobytej w poprzednich etapach, aż do opracowania optymalnych rozwiązań biotechnologicznych. Wyniki przedstawione w niniejszej rozprawie można osadzić właśnie w tej ramie metodologicznej – jako poszczególne etapy cyklu DBTL, z których każdy wnosił wkład w rozwój narzędzi i wiedzy dotyczącej roli TFs w inżynierii *Y. lipolytica* (**Rycina 3**).

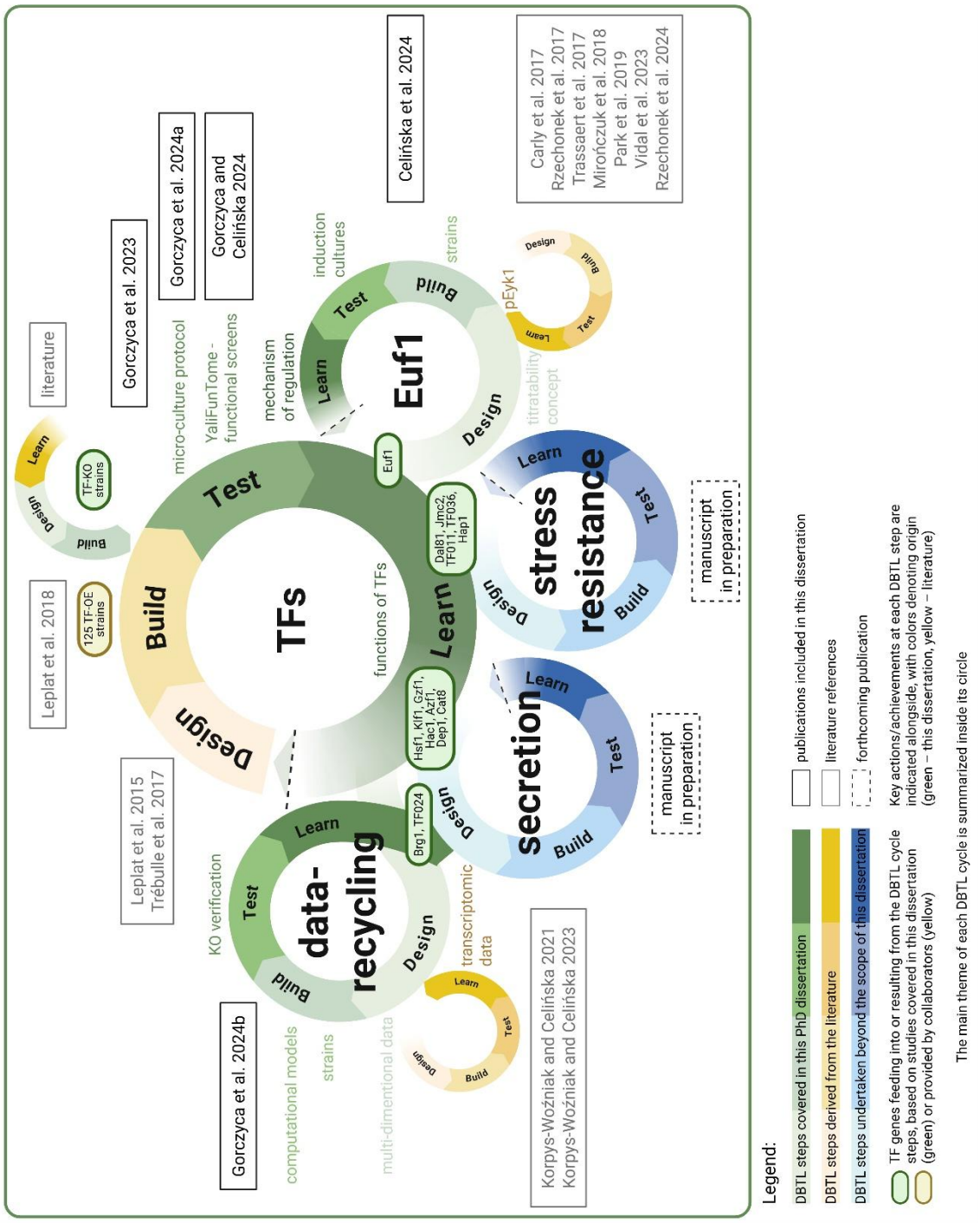
Cykl „*TFs*" – Pierwszymi etapami było zaprojektowanie (D) oraz budowa (B) obszernej biblioteki szczepów *Y. lipolytica* prowadzących indywidualną nadekspresję pojedynczych TFs i genu reporterowego (Leplat i in., 2015, 2018; Trébulle i in., 2017). Kolekcja ta została skonstruowana w laboratorium prof. Jean-Marca Nicaud i została użyczona na potrzeby realizacji pierwszego cyklu DBTL. Kolejny etap (T) – systematyczny skrining fenotypów – został całkowicie zrealizowany w ramach niniejszej rozprawy. To z kolei wymagało opracowania nowego protokołu zminiaturyzowanych hodowli *Y. lipolytica*, co było znaczącym osiągnięciem metodycznym tej pracy (**P2** i **P3**) oraz elementem umożliwiającym rzetelną realizację etapu „T". Jak wspomniano uprzednio, artykuł **P1** przedstawia badania pilotażowe, testujące hipotezę o możliwości wykorzystania TFs jako narzędzi do inżynierii cech złożonych. Jednocześnie zawiera elementy „D" i „B" w postaci nowoskonstruowanych szczepów niosących delecję w *loci* TFs, (wybranych na podstawie danych literaturowych) oraz elementy „T" – pierwsze próby doboru warunków skriningu wysokoprzepustowego.

Cykl „*data-recycling*" – Artykuł **P4** można traktować jako odrębną, pełną iterację cyklu DBTL. W tym konkretnym wypadku, skorzystano z wiedzy zgromadzonej w bazie danych YaliFunTome (**P2**), powstałej w wyniku pierwszej iteracji, ale też danych transkryptomicznych zgromadzonych w publicznych repozytoriach (Korpys-Woźniak i Celińska, 2021, 2023). Badania te wpisują się w koncepcję „recyklingu danych", czyli ponownego wykorzystania istniejących danych w celu formułowania nowych hipotez badawczych. Integracja różnorodnych zbiorów danych biologicznych jest wyzwaniem, gdyż procesy biologiczne rzadko mają charakter liniowy.

Jednak przy obecnym nagromadzeniu wiedzy biologicznej takie podejście staje się zasadne i – przy zastosowaniu odpowiedniej metodologii – pozwala odkrywać zjawiska, które mogłyby być niezauważone. Strategia ta, dotychczas rozpowszechniona w badaniach organizmów modelowych, okazuje się możliwa do wdrożenia także u organizmów niekonwencjonalnych, jak *Y. lipolytica*.

Cykl „*Euf1*" – Trzecią iteracją cyklu DBTL są badania zaprezentowane w **P5**. Badania nad TF Euf1, przeprowadzone w pierwszych etapach rozprawy (**P2**, **P4**), oraz dane literaturowe (seria publikacji o pEYK1 i TF Euf1 – Carly i in., 2017; Park i in., 2019a; Rzechonek i in., 2017, 2018, 2024; Trassaert i in., 2017; Vidal i in., 2023), stały się punktem wyjścia do „projektowania" i przeprowadzania kolejnych eksperymentów. Publikacja ta przedstawia pełną charakterystykę TF Euf1 i jego wpływu na promotor pEYK1: od projektowania „D" i konstruowania szczepów „B", przez ich testowanie „T", aż po wyciągnięcie końcowych wniosków „L" dotyczących roli tego TF i jego zastosowania jako narzędzia do modulowania siły promotorów indukowalnych.

Cykle DBTL wychodzące poza zakres niniejszej rozprawy: „*secretion*" i „*stress resistance*" – Skumulowana wiedza z etapu „L" pierwszej iteracji („*TFs*") oraz iteracji opartej na „*data-recycling*" (**P4**) posłużyła jako podstawa kolejnego cyklu DBTL koncentrującego się wokół zjawiska sekrecji białek heterologicznych. Działania w obrębie tego cyklu obejmowały m.in. skonstruowanie szczepów prowadzących sekrecję rProts, poszerzenie zestawu białek reporterowych o enzymy przemysłowe oraz zastosowanie innych strategii modyfikacji TFs. Równolegle, w oparciu o wnioski z pierwszej iteracji („*TFs*"), zainicjowano także odrębny cykl dotyczący odporności na stresy środowiskowe. Prace te stanowią dalsze rozwinięcie ścieżki „globalnej" inżynierii cech złożonych, nie są ujęte w niniejszej rozprawie i zostały zarysowane w rozdziale „Dalsze kierunki badań".



Rycina 3. Podsumowanie dokonań niniejszej rozprawy doktorskiej w kontekście metodyki cyklu *design-build-test-learn*. Szczegółowy opis znaczenia poszczególnych publikacji w odniesieniu do etapów cyklu przedstawiono w tekście rozdziału. Legenda na rycinie wyjaśnia zastosowane symbole i kolorystykę.

Podsumowując, w ramach niniejszej rozprawy:

1) Potwierdzono możliwość modulowania cech złożonych w drożdżach *Y. lipolytica* poprzez zastosowanie inżynierii TFs. Jest to podejście wciąż niewystarczająco ugruntowane w literaturze, a dowody uzyskane w toku badań mogą inspirować analogiczne strategie inżynieryjne w innych organizmach. Ze względu na kluczową rolę TFs w regulacji procesów komórkowych oraz konserwatywność wielu mechanizmów, uzyskane wyniki mogą stanowić punkt wyjścia do badań w innych gatunkach.

2) Opracowano ogólnodostępny protokół hodowli *Y. lipolytica* w małej skali, umożliwiający wysokoprzepustowe i wiarygodne testowanie fenotypów drożdży. Protokół stanowiący istotne narzędzie dla naukowców pracujących z tym gatunkiem i o potencjale ułatwienia badań przesiewowych w przyszłości.

3) Dostarczono nowej wiedzy na temat funkcji dotychczas nieopisanych TFs oraz nowe ujęcia funkcji TFs już znanych, lecz badanych w innym kontekście. Prowadzi to do identyfikacji TFs zaangażowanych w określone procesy biologiczne i wyznacza kierunki dalszych badań podstawowych nad mechanizmami ich działania w *Y. lipolytica* i pokrewnych gatunkach. W tym obszarze, szczególnie interesujące są wnioski dotyczące:

a) Syntezy białek heterologicznych w *Y. lipolytica*:

i) Nadekspresja TFs Gzf1, Hsf1, Klf1 wpływa na poprawę wydajności syntezy rProt niezależnie od warunków środowiskowych. Z tego względu zaklasyfikowano je do grupy uniwersalnych „wzmacniaczy" produkcji rProt. Zaangażowanie w procesy biologiczne „ogólna odpowiedź na stres", „odpowiedź na stres oksydacyjny", oraz „regulacja metabolizmu azotu", stoją u podstaw ich promującej roli. Delecja tych TFs nie powoduje uzyskania „odwróconego" fenotypu. Uważa się, że mogą pełnić rolę tzw. „genów buforowych" (Tawfeeq i in., 2024). Wniosek ten wymaga przeprowadzenia dalszych pogłębionych badań.

ii) Nadekspresja TF Hoy1 prowadzi do intensyfikacji syntezy rProt w warunkach podaży nieoptymalnego źródła azotu. Ta obserwacja nabiera szczególnego znaczenia w kontekście wcześniejszych badań naszego zespołu, w których wymodelowany bioproces produkcji rekombinowanej α-amylazy w skali pilotowej wykazał, że organiczne źródło azotu stanowi główny czynnik kosztotwórczy (Kubiak i in., 2021).

iii) TFs Dal81, Aro80 pod wpływem nadekspresji poprawiają syntezę rProt w standardowych warunkach. Uważa się, że ich zaangażowanie w metabolizm azotu warunkuje ich promującą funkcję.

iv) Delecja TF Dep1, regulatora związanego z syntezą lipidów, powoduje znaczny wzrost syntezy rProt w standardowych warunkach hodowli (2.6-krotny – najwyższy efekt uzyskany w badaniach tej rozprawy, wyrażony jako krotność zmiany). Co szczególnie istotne – jego nadekspresja powodowała wystąpienie odwrotnego fenotypu, rzadko obserwowanego w przypadku modyfikacji (nadekspresji/delecji) TFs. Zbliżony fenotyp (ograniczenie syntezy rProt, gdy TF był nadekspresjonowany) zaobserwowano dla TF Azf1, zaangażowanego w intensyfikację metabolizmu węgla i wzrostu przy dostępności preferowanego źródła węgla. W tym przypadku jednak fenotyp odwrotny nie był obserwowany.

v) Obserwacje poczynione dla TFs Klf1, Gzf1, Dal81, Aro80 oraz Dep1, Azf1, Cat8, Adr1, Ert1-2 wskazują na ogólną zależność pomiędzy „głównym strumieniem metabolizmu" (ang. *metabolic flux*) a poziomem syntezy rProt: aktywacja genów związanych z metabolizmem azotu sprzyjała jej intensyfikacji, natomiast genów zaangażowanych w metabolizm węgla – prowadziła do jej ograniczenia.

vi) Delecja TFs pełniących rolę globalnych regulatorów odpowiedzi na stres: Hsf1, Skn7, Hap1, Sfl1 powoduje poprawę syntezy rProt w warunkach optymalnych, tj. bez wprowadzenia stresu środowiskowego. Efekt ten można tłumaczyć częściowym „odciążeniem" aparatu transkrypcyjno-translacyjnego – w sytuacji, gdy stres nie występuje, ograniczenie ekspresji genów związanych z odpowiedzią obronną pozwala komórce skierować zasoby na kosztowny proces, jakim jest biosynteza rProts.

vii) Nadekspresja Hsf1 i Skn7 umożliwiają utrzymanie poziomu syntezy rProt w warunkach stresu hiperosmotycznego oraz ograniczonej dostępności tlenu. Potwierdza to udział Skn7 w odpowiedzi osmotycznej i oksydacyjnej, a dla Hsf1 wpisuje się w jego znaną funkcję jako globalnego regulatora odpowiedzi na stres.

b) Roli TFs w kształtowaniu odporności na stresy środowiskowe w *Y. lipolytica*:

i) TFs poprawiające odporność na stres ograniczonej dostępności tlenu pod wpływem ich nadekspresji: TF036, Jmc2, TF011 i Dal81. Najsilniejszy efekt (1.8-krotne zwiększenie wzrostu) uzyskano dla Jmc2 w warunkach pH 7,0, 28°C,

glicerolu i organicznego źródła azotu. Choć ich funkcje w *Y. lipolytica* pozostają nieokreślone, fenotypy te wskazują na możliwy udział w adaptacji do fluktuacji dostępności tlenu.

ii) Nadekspresja Hap1 powoduje spadek tolerancji na niedobór tlenu, co wskazuje, że w *Y. lipolytica* regulator ten może pełnić funkcję powiązaną z detekcją poziomu natlenienia i utrzymaniem równowagi metabolicznej w warunkach zmiennej dostępności tlenu.

iii) Skn7 i Hsf1 są zaangażowane w adaptację do stresu osmotyczny i oksydacyjnego. Ich delecja powoduje pogorszenie wzrostu w tych warunkach, co wskazuje, że są one niezbędne do utrzymania podstawowej kondycji fizjologicznej komórek w warunkach ograniczonej dostępności tlenu i wysokiego ciśnienia osmotycznego.

4) W niniejszej rozprawie dokonano szczegółowej charakterystyki jednego z TFs – Euf1, a na bazie zdobytej wiedzy zaproponowano nowe narzędzie do zwiększania ekspresji genów podlegających regulacji określonego promotora w *Y. lipolytica* oraz dalszych kierunków jego inżynierii. Do kluczowych osiągnięć w tym zakresie należą:

a) Nadekspresja splicingowanej wersji TF Euf1 zapewnia zwiększenie siły transkrypcyjnej promotora pEYK1 przy zastosowaniu większej liczby motywów aktywujących. Jednakże, w tych warunkach układ traci indukowalność zależną od chemicznego induktora. Zmiany w strukturze i powinowactwie do motywu A najprawdopodobniej przyczyniają się do utraty tej właściwości.

b) Na podstawie profili transkrypcyjnych, znanej struktury promotora, oraz wymodelowanej struktury TF Euf1 w postaci pełnej oraz splicingowanej zaproponowano mechanizm kontroli regulacji promotora pEYK1 zależny od współoddziaływania TF Euf1 i Adr1 i od obecności chemicznego induktora oraz źródła węgla.

c) Elementem praktycznym wynikającym z pogłębionych analiz TF Euf1 jest zaproponowana strategia bioprocesowa bazująca na wykorzystaniu promotora pEYK1. Proponuje się zastosowanie etapu „głodzenia” na koniec bioprocesu – tj. warunków, w których źródła węgla są wyczerpane w pożywce – celem maksymalizacji uzysku rProt. Alternatywnie, w świetle zaproponowanego mechanizmu współdziałania TFs Euf1 i Adr1 w obrębie promotora pEYK1, zasadne wydaje się skonstruowanie jego wariantu syntetycznego pozbawionego motywu B, co pozwoliłoby na maksymalizację

syntezy rProt bez konieczności stosowania etapu głodzenia. Ponadto, rekomenduje się także stosowanie alternatywnego induktora, sorbitolu, który natywnie nie jest utylizowany przez *Y. lipolytica* (Kubiak i in., 2019) i nie wymaga stosowania tła genetycznego *Δeyk1*.

6.1 Dalsze kierunki badań

Wyniki uzyskane w ramach niniejszej rozprawy pozwoliły wytypować grupę TFs szczególnie obiecujących w kontekście dalszych badań nad poprawą syntezy rProts oraz zwiększaniem odporności *Y. lipolytica* na brak wystarczającej dostępności tlenu w pożywce. Ich charakterystyka otwiera drogę do kolejnych iteracji cyklu *design-build-test-learn*, ukierunkowanych na weryfikację ich funkcjonalności w rozszerzonym kontekście (**Rycina 3**).

Pierwszy kierunek obejmuje badania nad TFs, które wstępnie wykazały wpływ na modulację syntezy rProts. Do dalszych analiz – zarówno w wariantach nadekspresyjnych, jak i delecyjnych – wytypowano geny istotnie zmieniające poziom syntezy rProts według wyników z bazy danych YaliFunTome oraz wykazujące ciekawą charakterystykę w profilu ekspresji. Badania te są prowadzone w szczepach nadekspresjonujących geny kodujące białka sekrecyjne (w tym enzymy przemysłowe).

Drugim kierunkiem jest pogłębiona analiza TFs, które mogą znacząco wpływać na odporność komórek na ograniczoną dostępność tlenu. Aby wyeliminować wpływ syntezy heterologicznych białek na obserwowany efekt, badane są szczepy, w których nadekspresji lub delecji podlega wyłącznie pojedynczy TF.

Powyższe prace nie stanowią przedmiotu niniejszej rozprawy doktorskiej, jednakże większość eksperymentów została już wykonana, a ich wyniki są obecnie przygotowywane do publikacji. Tym samym, badania opisane w tym podrozdziale można traktować jako kontynuację i domknięcie pełnego cyklu DBTL zapoczątkowanego w **P1**, **P2** i **P4** stanowiącego fundament do dalszych zastosowań biotechnologicznych w zakresie inżynierii *Y. lipolytica*.

7. Wykaz skrótów stosowanych w rozprawie i publikacjach

rProt / **rProts** – białko rekombinowane, białka rekombinowane (ang. *recombinant protein*, *recombinant proteins*)

TF / **TFs** – czynnik transkrypcyjny, czynniki transkrypcyjne (ang. *transcription factor*, *transcription factors*)

UPR – odpowiedź na niesfałdowane białka (ang. *unfolded protein response*)

GTF / **GTFs** – ogólny czynnik transkrypcyjny, ogólne czynniki transkrypcyjne (ang. *general transcription factor*, *general transcription factors*)

ALE – adaptacyjna ewolucja w warunkach laboratoryjnych (ang. *adaptive laboratory evolution*)

pz / **bp** – pary zasad (ang. *base pairs*)

DoE – planowane, projektowanie eksperymentów (ang. *Design of Experiments*)

k_La – objętościowy współczynnik wnikania masy (ang. *volumetric mass transfer coefficient*)

NCR – represja kataboliczna azotu (ang. *nitrogen catabolite repression*)

OE / **co-OE** – nadekspresja, ko-nadekspresja (ang. *overexpression*, *co-overexpression*)

KO – delecja (ang. *knock-out*)

Osm – osmolalność (ang. *osmolality*)

pO_2, **OA** – dostępność tlenu (ang. *oxygen availability*)

ER – siateczka śródplazmatyczna (ang. *endoplasmatic reticulum*)

OD_{600} – gęstość optyczna, absorbancja przy długości fali 600 nm (ang. *optical density*, *absorbance at 600 nm*)

FL / **sFL** – fluorescencja, fluorescencja specyficzna (w przeliczeniu na jednostkę biomasy) (ang. *fluorescence*, *specific fluorescence*)

FU / **RFU** – jednostka fluorescencji, jednostka specyficznej fluorescencji (ang. *fluorescence unit*, *relative fluorescence unit*)

ORF – otwarta ramka odczytu (ang. *open reading frame*)

FC – krotność zmiany (ang. *fold change*)

RSM – metoda powierzchni odpowiedzi (ang. *response surface methodology*)

MTP – płytka mikrotitracyjna (ang. *micro-titer plate*)

Lip^+ / **Lip^-** – warunki stosunku C/N (węgla do azotu) promujące / hamujące akumulację lipidów

bZIP – motyw zamka leucynowego (ang. *basic leucine zipper*)

ROS – reaktywne formy tlenu (ang. *reactive oxygen species*)

CH – hydrolizat kazeiny (ang. *casamino acid hydrolysate*)

AS – siarczan amonu (ang. *ammonium sulfate*)

pKa – stała dysocjacji (ang. *dissociacion constant*)

SD – odchylenie standardowe (ang. *standard deviation*)

SGD – *Saccharomyces* Genome Database

FADS – cytometria przepływowa z sortowaniem kropli aktywowanym fluorescencją (ang. *fluorescence-assisted droplet sorting*)

FACS – cytometria przepływowa z sortowaniem komórek aktywowanym fluorescencją (ang. *fluorescence-assisted cell sorting*)

d – średnica naczynia (ang. *vessel diameter*)

$\mathbf{V_L}$ – objętość hodowli (ang. *culture volume*)

n – liczba powtórzeń / częstotliwość wytrząsania (ang. *shaking frequency*)

$\mathbf{d_0}$ – amplituda wytrząsania (ang. *shaking amplitude*)

GLY – glicerol (ang. *glycerol*)

GLUC – glukoza (ang. *glucose*)

CA – kwas cytrynowy (ang. *citric acid*)

ERY – erytrytol (ang. *erythritol*)

MAN – mannitol (ang. *mannitol*)

NCBI SRA – baza danych National Center for Biotechnology Information – Sequence Read Archive

YFP / **inYFP** / **scYFP** – białko żółtej fluorescencji / wewnątrzkomórkowe białko żółtej fluorescencji / sekrecyjne białko żółtej fluorescencji (ang. *yellow fluorescent protein* / *intracellular yellow fluorescent protein* / *secretory yellow fluorescent protein*)

scSoA – sekrecyjna α-amylaza z wołka ryżowego *Sitophilus oryzae* (ang. *secretory α-amylase from Sitophilus oryzae*)

scTlG – sekrecyjna glukoamylaza z grzyba strzępkowego *Thermomyces lanuginosus* (ang. *secretory glucoamylase from Thermomyces lanuginosus*)

HSS – wysoka synteza i sekrecja (ang. *high synthesis and secretion*)

TRN – sieć regulacji transkrypcyjnej (ang. *transcriptional regulatory network*)

fullEuf1 – pełna wersja genu Euf1 (ang. *unspliced Euf1*)

spEuf1 – wersja genu Euf1 z usuniętym intronem (ang. *spliced Euf1*)

SORB – sorbitol (ang. *sorbitol*)

G – entalpia swobodna (ang. *Gibbs free energy*)

DBTL – cykl projektowania-budowania-testowania-uczenia w biologii syntetycznej (ang. *design-build-test-learn cycle*)

8. Literatura

Beneyton, T., Thomas, S., Griffiths, A. D., Nicaud, JM., Drevelle, A., & Rossignol, T. (2017). Droplet-based microfluidic high-throughput screening of heterologous enzymes secreted by the yeast *Yarrowia lipolytica*. *Microbial Cell Factories*, *16*(1). doi.org/10.1186/s12934-017-0629-5

Bhave, S. L., & Chattoo, B. B. (2003). Expression of *Vitreoscilla* hemoglobin improves growth and levels of extracellular enzyme in *Yarrowia lipolytica*. *Biotechnology and Bioengineering*, *84*(6), 658–666. doi.org/https://doi.org/10.1002/bit.10817

Carly, F., Gamboa-Melendez, H., Vandermies, M., Damblon, C., Nicaud, JM., & Fickers, P. (2017). Identification and characterization of *EYK1*, a key gene for erythritol catabolism in *Yarrowia lipolytica*. *Applied Microbiology and Biotechnology*, *101*(17), 6587–6596. doi.org/10.1007/s00253-017-8361-y

Celińska, E. (2022). "Fight-flight-or-freeze" – how *Yarrowia lipolytica* responds to stress at molecular level? *Applied Microbiology and Biotechnology*, 106, 3369–3395. doi.org/10.1007/s00253-022-11934-x

Celińska, E., Ledesma-Amaro, R., Larroude, M., Rossignol, T., Pauthenier, C., & Nicaud, JM. (2017). Golden Gate Assembly system dedicated to complex pathway manipulation in *Yarrowia lipolytica*. *Microbial Biotechnology*, *10*(2), 450–455. doi.org/10.1111/1751-7915.12605

Celińska, E., & Nicaud, JM. (2019). Filamentous fungi-like secretory pathway strayed in a yeast system: peculiarities of *Yarrowia lipolytica* secretory pathway underlying its extraordinary performance. *Applied Microbiology and Biotechnology*, 103, 39–52. doi.org/10.1007/s00253-018-9450-2

Cogo, A. J. D., Façanha, A. R., Da Silva Teixeira, L. R., De Souza, S. B., Da Rocha, J. G., Figueira, F. F., Eutrópio, F. J., Bertolazi, A. A., De Rezende, C. E., Krohling, C. A., Okorokov, L. A., Cruz, C., Ramos, A. C., & Okorokova-Façanha, A. L. (2020). Plasma membrane H^+pump at a crossroads of acidic and iron stresses in yeast-to-hypha transition. *Metallomics*, *12*(12), 2174–2185. doi.org/10.1039/d0mt00179a

Craig, E. A., Gambill, B. D., & Nelson, R. J. (1993). Heat shock proteins: molecular chaperones of protein biogenesis. *Microbiological Reviews*, *57*(2), 402-414, doi.org/10.1128/mr.57.2.402-414.1993

De Pourcq, K., Vervecken, W., Dewerte, I., Valevska, A., Van Hecke, A., & Callewaert, N. (2012). Engineering the yeast *Yarrowia lipolytica* for the production of therapeutic proteins homogeneously glycosylated with $Man_8GlcNAc_2$ and $Man_5GlcNAc_2$. *Microbial Cell Factories*, 11, 53. doi.org/10.1186/1475-2859-11-53

Doughty, T., & Kerkhoven, E. (2020). Extracting novel hypotheses and findings from RNA-seq data. *FEMS Yeast Research*, *20*(2). doi.org/10.1093/femsyr/foaa007

Dragosits, M., & Mattanovich, D. (2013). Adaptive laboratory evolution - principles and applications for biotechnology. *Microbial Cell Factories*, 12, 64. doi.org/10.1186/1475-2859-12-64

Duan, G., Ding, L., Wei, D., Zhou, H., Chu, J., Zhang, S., & Qian, J. (2019). Screening endogenous signal peptides and protein folding factors to promote the secretory expression of heterologous proteins in *Pichia pastoris*. *Journal of Biotechnology*, 306, 193–202. doi.org/10.1016/j.jbiotec.2019.06.297

Egermeier, M., Sauer, M., & Marx, H. (2019). Golden Gate-based metabolic engineering strategy for wild-type strains of *Yarrowia lipolytica*. *FEMS Microbiology Letters*, *366*(4), fnz022. doi.org/10.1093/femsle/fnz022

Endoh-Yamagami, S., Hirakawa, K., Morioka, D., Fukuda, R., & Ohta, A. (2007). Basic helix-loop-helix transcription factor heterocomplex of Yas1p and Yas2p regulates cytochrome P450 expression in response to alkanes in the yeast *Yarrowia lipolytica*. *Eukaryotic Cell*, *6*(4), 734–743. doi.org/10.1128/EC.00412-06

Gasch, A. P. (2007). Comparative genomics of the environmental stress response in ascomycete fungi. *Yeast*, *24*(11), 961–976. doi.org/10.1002/yea.1512

Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D., & Brown, P. O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Molecular Biology of the Cell*, *11*(12), 4241-4257. doi.org/10.1091/mbc.11.12.4241

Gasch, A. P., & Werner-Washburne, M. (2002). The genomics of yeast responses to environmental stress and starvation. *Functional and Integrative Genomics*, 2, 181–192. doi.org/10.1007/s10142-002-0058-2

Georgiadis, I., Tsiligkaki, C., Patavou, V., Orfanidou, M., Tsoureki, A., Andreadelli, A., Theodosiou, E., & Makris, A. M. (2023). Identification and construction of strong promoters in *Yarrowia lipolytica* suitable for glycerol-based bioprocesses. *Microorganisms*, *11*(5), 1152. doi.org/10.3390/microorganisms11051152

Gorczyca, M., Kaźmierczak, J., Fickers, P., & Celińska, E. (2022). Synthesis of secretory proteins in *Yarrowia lipolytica*: effect of combined stress factors and metabolic load. *International Journal of Molecular Sciences*, *23*(7), 3602. doi.org/10.3390/ijms23073602

Gorczyca, M., Kaźmierczak, J., Steels, S., Fickers, P., & Celińska, E. (2020). Impact of oxygen availability on heterologous gene expression and polypeptide secretion dynamics in *Yarrowia lipolytica*-based protein production platforms. *Yeast*, *37* (9–10), 559–568. doi.org/10.1002/yea.3499

Guerfal, M., Ryckaert, S., Jacobs, P. P., Ameloot, P., Craenenbroeck, K. Van, Derycke, R., & Callewaert, N. (2010). The *HAC1* gene from *Pichia pastoris*: characterization and effect of its overexpression on the production of secreted, surface displayed and membrane proteins. *Microbial Cell Factories*, 9, 49. doi.org/10.1186/1475-2859-9-49

Guyot, S., Ferret, E., & Gervais, P. (2005). Responses of *Saccharomyces cerevisiae* to thermal stress. *Biotechnology and Bioengineering*, *92*(4), 403–409. doi.org/https://doi.org/10.1002/bit.20600

Hahn, J.-S., Hu, Z., Thiele, D. J., & Iyer, V. R. (2004). Genome-wide analysis of the biology of stress responses through heat shock transcription factor. *Molecular and Cellular Biology*, *24*(12), 5249–5256. doi.org/10.1128/mcb.24.12.5249-5256.2004

Herholz, M., Cepeda, E., Baumann, L., Kukat, A., Hermeling, J., Maciej, S., Szczepanowska, K., Pavlenko, V., Frommolt, P., & Trifunovic, A. (2019). *KLF-1* orchestrates a xenobiotic detoxification program essential for longevity of mitochondrial mutants. *Nature Communications*, *10*(1), 3323. doi.org/10.1038/s41467-019-11275-w

Hirakawa, K., Kobayashi, S., Inoue, T., Endoh-Yamagami, S., Fukuda, R., & Ohta, A. (2009). Yas3p, an opi1 family transcription factor, regulates Cytochrome P450 expression in response to n-alkanes in *Yarrowia lipolytica*. *Journal of Biological Chemistry*, *284*(11), 7126–7137. doi.org/10.1074/jbc.M806864200

Holkenbrink, C., Dam, M. I., Kildegaard, K. R., Beder, J., Dahlin, J., Doménech Belda, D., & Borodina, I. (2018). EasyCloneYALI: CRISPR/Cas9-based synthetic toolbox for engineering of the yeast *Yarrowia lipolytica*. *Biotechnology Journal*, *13*(9), 1700543. doi.org/10.1002/biot.201700543

Hou, J., Österlund, T., Liu, Z., Petranovic, D., & Nielsen, J. (2013). Heat shock response improves heterologous protein secretion in *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology*, *97*(8), 3559–3568. doi.org/10.1007/s00253-012-4596-9

Hou, J., Tang, H., Liu, Z., Österlund, T., Nielsen, J., & Petranovic, D. (2014). Management of the endoplasmic reticulum stress by activation of the heat shock response in yeast. *FEMS Yeast Research*, *14*(3), 481–494. doi.org/10.1111/1567-1364.12125

Hou, J., Tyo, K. E. J., Liu, Z., Petranovic, D., & Nielsen, J. (2012). Metabolic engineering of recombinant protein secretion by *Saccharomyces cerevisiae*. *FEMS Yeast Research*, *12*(5), 491–510. doi.org/10.1111/j.1567-1364.2012.00810.x

Hurtado, C. A. R., & Rachubinski, R. A. (1999). *MHY1* encodes a C_2H_2-type zinc finger protein that promotes dimorphic transition in the yeast *Yarrowia lipolytica*. *Journal of Bacteriology*, *181*(10), 3051-3057. doi.org/10.1128/JB.181.10.3051-3057.1999

Hurtado, C. A. R., & Rachubinski, R. A. (2002). *YlBMH1* encodes a 14-3-3 protein that promotes filamentous growth in the dimorphic yeast *Yarrowia lipolytica*. *Microbiology*, 148(11), 3725-3735. doi.org/10.1099/00221287-148-11-3725

Jiang, W., Wang, S., Avila, P., Jørgensen, T. S., Yang, Z., & Borodina, I. (2024). Combinatorial iterative method for metabolic engineering of *Yarrowia lipolytica*: application for betanin biosynthesis. *Metabolic Engineering*, 86, 78-88. doi.org/10.1016/j.ymben.2024.09.003

Juretzek, T., Wang, H.-J., Nicaud, JM., Mauersberger, S., & Barth, G. (2000). Comparison of promoters suitable for regulated overexpression of β-galactosidase in the alkane-utilizing yeast *Yarrowia lipolytica*. *Biotechnol. Bioprocess Eng*, 5, 320-326. doi.org/10.1007/BF02942206

Kjeldsen, T., Andersen, A. S., Hubálek, F., Johansson, E., Kreiner, F. F., Schluckebier, G., & Kurtzhals, P. (2024). Molecular engineering of insulin for recombinant expression in yeast. *Trends in Biotechnology*, *42*(4), 464–478. doi.org/10.1016/j.tibtech.2023.09.012

Kolhe, N., Kulkarni, A., Zinjarde, S., & Acharya, C. (2021). Transcriptome response of the tropical marine yeast *Yarrowia lipolytica* on exposure to uranium. *Current Microbiology*, *78*(5), 2033–2043. doi.org/10.1007/s00284-021-02459-z

Korpys-Woźniak, P., & Celińska, E. (2021). Global transcriptome profiling reveals genes responding to overproduction of a small secretory, a high cysteine- and a high glycosylation-bearing protein in *Yarrowia lipolytica*. *Biotechnology Reports,* 31, e00646. doi.org/10.1016/j.btre.2021.e00646

Korpys-Woźniak, P., & Celińska, E. (2023). Molecular background of *HAC1*-driven improvement in the secretion of recombinant protein in *Yarrowia lipolytica* based on comparative transcriptomics. *Biotechnology Reports*, 38, e00801. doi.org/10.1016/j.btre.2023.e00801

Korpys-Woźniak, P., Kubiak, P., & Celińska, E. (2021). Secretory helpers for enhanced production of heterologous proteins in *Yarrowia lipolytica*. *Biotechnology Reports,* 32, e00669. doi.org/10.1016/j.btre.2021.e00669

Kubiak, M., Białas, W., & Celińska, E. (2021). Thermal treatment improves a process of crude glycerol valorization for the production of a heterologous enzyme by *Yarrowia lipolytica*. *Biotechnology Reports*, 31, e00648. doi.org/10.1016/j.btre.2021.e00648

Kubiak, M., Borkowska, M., Białas, W., Korpys, P., & Celińska, E. (2019). Feeding strategy impacts heterologous protein production in *Yarrowia lipolytica* fed-batch cultures—Insight into the role of osmolarity. *Yeast*, *36*(5), 305–318. doi.org/10.1002/yea.3384

Kubiak-Szymendera, M., Skupien-Rabian, B., Jankowska, U., & Celińska, E. (2022). Hyperosmolarity adversely impacts recombinant protein synthesis by *Yarrowia lipolytica*—molecular background revealed by quantitative proteomics. *Applied Microbiology and Biotechnology*, *106*(1), 349–367. doi.org/10.1007/s00253-021-11731-y

Lara, A. R., Galindo, E., Ramírez, O. T., & Palomares, L. A. (2006). Living with heterogeneities in bioreactors understanding the effects of environmental gradients on cells. *Molecular Biotechnology,* 34, 355–381. doi.org/10.1385/MB:34:3:355

Lebrun, E., Shenshin, V., Plaire, C., Vigneres, V., Pizette, T., Dumas, B., Nicaud, JM., & Mottet, G. (2023). Efficient full-length IgG secretion and sorting from single yeast clones in droplet picoreactors. *Lab on a Chip*, *23*(15), 3487–3500. doi.org/10.1039/d3lc00403a

Leplat, C., Nicaud, JM., & Rossignol, T. (2015). High-throughput transformation method for *Yarrowia lipolytica* mutant library screening. *FEMS Yeast Research*, *15*(6), fov052. doi.org/10.1093/femsyr/fov052

Leplat, C., Nicaud, JM., & Rossignol, T. (2018). Overexpression screen reveals transcription factors involved in lipid accumulation in *Yarrowia lipolytica*. *FEMS Yeast Research*, *18*(5), foy037. doi.org/10.1093/femsyr/foy037

Lesage, J., Timoumi, A., Cenard, S., Lombard, E., Lee, H. L. T., Guillouet, S. E., & Gorret, N. (2021). Accelerostat study in conventional and microfluidic bioreactors to assess the key role of residual glucose in the dimorphic transition of *Yarrowia lipolytica* in response to environmental stimuli. *New Biotechnology*, *64*, 37–45. doi.org/10.1016/j.nbt.2021.05.004

Liu, Z., Hou, J., Martínez, J. L., Petranovic, D., & Nielsen, J. (2013). Correlation of cell growth and heterologous protein production by *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology*, *97*(20), 8955–8962. doi.org/10.1007/s00253-013-4715-2

Lopes, M., Gomes, N., Gonçalves, C., Coelho, M. A. Z., Mota, M., & Belo, I. (2008). *Yarrowia lipolytica* lipase production enhanced by increased air pressure. *Letters in Applied Microbiology*, *46*(2), 255–260. doi.org/10.1111/j.1472-765X.2007.02299.x

Lopes, M., Gomes, N., Mota, M., & Belo, I. (2009). *Yarrowia lipolytica* growth under increased air pressure: Influence on enzyme production. *Applied Biochemistry and Biotechnology*, *159*(1), 46–53. doi.org/10.1007/s12010-008-8359-0

Madzak, C., & Blanchin-Roland, S. (2000). Strong hybrid promoters and integrative expression/ secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast *Yarrowia lipolytica* metabolic engineering. *Article in Journal of Molecular Microbiology and Biotechnology*, *2*(2), 207–216.

Madzak, C., Blanchin-Roland, S., Cordero Otero, R. R., & Gaillardin, C. (1999). Functional analysis of upstream regulating regions from the *Yarrowia lipolytica XPR2* promoter. *Microbiology*, *145*(1), 75–87. doi.org/10.1099/13500872-145-1-75

Mans, R., Daran, J. M. G., & Pronk, J. T. (2018). Under pressure: evolutionary engineering of yeast strains for improved performance in fuels and chemicals production. *Current Opinion in Biotechnology*, *50*, 47–56. doi.org/10.1016/j.copbio.2017.10.011

Martínez, J. L., Meza, E., Petranovic, D., & Nielsen, J. (2016). The impact of respiration and oxidative stress response on recombinant α-amylase production by *Saccharomyces cerevisiae*. *Metabolic Engineering Communications*, *3*, 205–210. doi.org/10.1016/j.meteno.2016.06.003

Martinez-Vazquez, A., Gonzalez-Hernandez, A., Domínguez, Á., Rachubinski, R., Riquelme, M., Cuellar-Mata, P., & Guzman, J. C. T. (2013). Identification of the transcription factor Znc1p, which regulates the yeast-to-hypha transition in the dimorphic yeast *Yarrowia lipolytica*. *PLoS ONE*, *8*(6), e66790. doi.org/10.1371/journal.pone.0066790

Matsumoto, R., Akama, K., Rakwal, R., & Iwahashi, H. (2005). The stress response against denatured proteins in the deletion of cytosolic chaperones *SSA1/2* is different from heat-shock response in *Saccharomyces cerevisiae*. *BMC Genomics*, *6*, 141. doi.org/10.1186/1471-2164-6-141

Mattanovich, D., Gasser, B., Hohenblum, H., & Sauer, M. (2004). Stress in recombinant protein producing yeasts. *Journal of Biotechnology*, *113*(1–3), 121–135. doi.org/10.1016/j.jbiotec.2004.04.035

Mirończuk, A. M., Kosiorowska, K. E., Biegalska, A., Rakicka-Pustułka, M., Szczepańczyk, M., & Dobrowolski, A. (2019). Heterologous overexpression of bacterial hemoglobin *VHb* improves erythritol biosynthesis by yeast *Yarrowia lipolytica*. *Microbial Cell Factories*, *18*(1), 176. doi.org/10.1186/s12934-019-1231-9

Morales-Vargas, A. T., Domínguez, A., & Ruiz-Herrera, J. (2012). Identification of dimorphism-involved genes of *Yarrowia lipolytica* by means of microarray analysis. *Research in Microbiology*, *163*(5), 378–387. doi.org/10.1016/j.resmic.2012.03.002

Muriel, C., & Claude, G. (2014). pH signaling in human fungal pathogens: a new target for antifungal strategies. *Eukaryotic Cell*, *13*(3), 342–352. doi.org/10.1128/EC.00313-13

Ogrydziak D. (2013). Acid and alkaline extracellular proteases of *Yarrowia lipolytica*. W G. Barth (Red.), *Yarrowia Lipolytica, Microbiology Monographs 25* (s. 77–97). Heidelberg: Springer.

Park, Y. K., Korpys, P., Kubiak, M., Celińska, E., Soudier, P., Trébulle, P., Larroude, M., Rossignol, T., & Nicaud, JM. (2019a). Engineering the architecture of erythritol-inducible promoters for regulated and enhanced gene expression in *Yarrowia lipolytica*. *FEMS Yeast Research*, *19*(1), foy105. doi.org/10.1093/femsyr/foy105

Park, Y. K., Vandermies, M., Soudier, P., Telek, S., Thomas, S., Nicaud, JM., & Fickers, P. (2019b). Efficient expression vectors and host strain for the production of recombinant proteins by *Yarrowia lipolytica* in process conditions. *Microbial Cell Factories*, *18*(1), 167. doi.org/10.1186/s12934-019-1218-6

Peñalva, M. A., & Arst, H. N. (2002). Regulation of gene expression by ambient pH in filamentous fungi and yeasts. *Microbiology and Molecular Biology Reviews*, *66*(3), 426–446. doi.org/10.1128/mmbr.66.3.426-446.2002

Pigne`de, G., Pigne`de, P., Wang, H.-J., Fudalej, F., Seman, M., Gaillardin, C., & Nicaud, JM. (2000). Autocloning and amplification of *LIP2* in *Yarrowia lipolytica*. *Applied and Environmental Microbiology*, *66*(8), 3283–3289. doi.org/10.1128/AEM.66.8.3283-3289.2000

Pomraning, K. R., Bredeweg, E. L., & Baker, S. E. (2017). Regulation of nitrogen metabolism by GATA zinc finger transcription factors in *Yarrowia lipolytica*. *mSphere*, *2*(1). doi.org/10.1128/msphere.00038-17

Pomraning, K. R., Bredeweg, E. L., Kerkhoven, E. J., Barry, K., Haridas, S., Hundley, H., LaButti, K., Lipzen, A., Yan, M., Magnuson, J. K., Simmons, B. A., Grigoriev, I. V., Nielsen, J., & Baker, S. E. (2018). Regulation of yeast-to-hyphae transition in *Yarrowia lipolytica*. *mSphere*, *3*(6). doi.org/10.1128/msphere.00541-18

Pomraning, K. R., Kim, Y. M., Nicora, C. D., Chu, R. K., Bredeweg, E. L., Purvine, S. O., Hu, D., Metz, T. O., & Baker, S. E. (2016). Multi-omics analysis reveals regulators of the response to nitrogen limitation in *Yarrowia lipolytica*. *BMC Genomics*, *17*, 138. doi.org/10.1186/s12864-016-2471-2

Poopanitpan, N., Kobayashi, S., Fukuda, R., Horiuchi, H., & Ohta, A. (2010). An ortholog of farA of *Aspergillus nidulans* is implicated in the transcriptional activation of genes involved in fatty acid utilization in the yeast *Yarrowia lipolytica*. *Biochemical and Biophysical Research Communications*, *402*(4), 731–735. doi.org/10.1016/j.bbrc.2010.10.096

Portnoy, V. A., Bezdan, D., & Zengler, K. (2011). Adaptive laboratory evolution – harnessing the power of biology for metabolic engineering. *Current Opinion in Biotechnology*, *22*(4), 590–594. doi.org/10.1016/J.COPBIO.2011.03.007

Rakicka, M., Lazar, Z., Dulermo, T., Fickers, P., & Nicaud, JM. (2015). Lipid production by the oleaginous yeast *Yarrowia lipolytica* using industrial by-products under different culture conditions. *Biotechnology for Biofuels*, *8*, 104. doi.org/10.1186/s13068-015-0286-z

Rigouin, C., Gueroult, M., Croux, C., Dubois, G., Borsenberger, V., Barbe, S., Marty, A., Daboussi, F., André, I., & Bordes, F. (2017). Production of medium chain fatty acids by *Yarrowia lipolytica*: combining molecular design and TALEN to engineer the fatty acid synthase. *ACS Synthetic Biology*, *6*(10), 1870–1879. doi.org/10.1021/acssynbio.7b00034

Roland, S. B., Roman, R., Otero, C., & Gaillardin, C. (1994). Two upstream activation sequences control the expression of the *XPR2* gene in the yeast *Yarrowia lipolytica*. *Molecular and Cellular Biology*, *14*(1), 327–338. doi.org/10.1128/mcb.14.1.327-338.1994

Rzechonek, D. A., Day, A. M., Quinn, J., & Mirończuk, A. M. (2018). Influence of ylHog1 MAPK kinase on *Yarrowia lipolytica* stress response and erythritol production. *Scientific Reports*, *8*, 14735. doi.org/10.1038/s41598-018-33168-6

Rzechonek, D. A., Neuvéglise, C., Devillers, H., Rymowicz, W., & Mirończuk, A. M. (2017). *EUF1*-A newly identified gene involved in erythritol utilization in *Yarrowia lipolytica*. *Scientific Reports*, *7*, 12507. doi.org/10.1038/s41598-017-12715-7

Rzechonek, D. A., Szczepańczyk, M., Borodina, I., Neuvéglise, C., & Mirończuk, A. M. (2024). Transcriptome analysis reveals multiple targets of erythritol-related transcription factor *EUF1* in unconventional yeast *Yarrowia lipolytica*. *Microbial Cell Factories*, *23*, 77. doi.org/10.1186/s12934-024-02354-9

Sarris, D., Stoforos, N. G., Mallouchos, A., Kookos, I. K., Koutinas, A. A., Aggelis, G., & Papanikolaou, S. (2017). Production of added-value metabolites by *Yarrowia lipolytica* growing in olive mill wastewater-based media under aseptic and non-aseptic conditions. *Engineering in Life Sciences*, *17*(6), 695–709. doi.org/10.1002/elsc.201600225

Sassi, H., Delvigne, F., Kallel, H., & Fickers, P. (2017). pH and not cell morphology modulate pLIP2 induction in the dimorphic yeast *Yarrowia lipolytica*. *Current Microbiology*, *74*, 413–417. doi.org/10.1007/s00284-017-1207-0

Sassi, H., Delvigne, F., Kar, T., Nicaud, JM., Coq, A. M. C. Le, Steels, S., & Fickers, P. (2016). Deciphering how *LIP2* and *POX2* promoters can optimally regulate recombinant protein production in the yeast *Yarrowia lipolytica*. *Microbial Cell Factories*, *15*, 159. doi.org/10.1186/s12934-016-0558-8

Schwartz, C., Frogue, K., Ramesh, A., Misa, J., & Wheeldon, I. (2017). CRISPRi repression of nonhomologous end-joining for enhanced genome engineering via homologous recombination in *Yarrowia lipolytica*. *Biotechnology and Bioengineering*, *114*(12), 2896–2906. doi.org/10.1002/bit.26404

Sekova, V. Y., Kovalyov, L. I., Kovalyova, M. A., Gessler, N. N., Danilova, M. A., Isakova, E. P., & Deryabina, Y. I. (2021). Proteomics readjustment of the *Yarrowia lipolytica* yeast in response to increased temperature and alkaline stress. *Microorganisms*, *9*(12), 2619. doi.org/10.3390/microorganisms9122619

Shimanuki, M., Uehara, L., Pluskal, T., Yoshida, T., Kokubu, A., Kawasaki, Y., & Yanagida, M. (2013). Klf1, a C_2H_2 zinc finger-transcription factor, is required for cell wall maintenance during long-term quiescence in differentiated G0 phase. *PLoS ONE*, *8*(10), e78545. doi.org/10.1371/journal.pone.0078545

Szabo, R., & Stofaníková, V. (2002). Presence of organic sources of nitrogen is critical for filament formation and pH-dependent morphogenesis in *Yarrowia lipolytica*. *FEMS Microbiology Letters*, *206*(1), 45–50. doi.org/10.1111/j.1574-6968.2002.tb10984.x

Tawfeeq, M. T., Voordeckers, K., van den Berg, P., Govers, S. K., Michiels, J., & Verstrepen, K. J. (2024). Mutational robustness and the role of buffer genes in evolvability. *EMBO Journal*, *43*, 2294–2307. doi.org/10.1038/s44318-024-00109-1

Torres-Guzma´n, J. C., Guzma´n, G., Domi´nguez, A., & Domi´nguez, D. (1997). HOY1, a homeo gene required for hyphal formation in *Yarrowia lipolytica*. *Molecular and Cellular Biology*, *17*(11), 6283–6293. doi.org/10.1128/MCB.17.11.6283

Trassaert, M., Vandermies, M., Carly, F., Denies, O., Thomas, S., Fickers, P., & Nicaud, JM. (2017). New inducible promoter for gene expression and synthetic biology in *Yarrowia lipolytica*. *Microbial Cell Factories*, *16*, 141. doi.org/10.1186/s12934-017-0755-0

Trébulle, P., Nicaud, JM., Leplat, C., & Elati, M. (2017). Inference and interrogation of a coregulatory network in the context of lipid accumulation in *Yarrowia lipolytica*. *npj Systems Biology and Applications*, *3*, 21. doi.org/10.1038/s41540-017-0024-1

Tyo, K. E. J., Liu, Z., Petranovic, D., & Nielsen, J. (2012). Imbalance of heterologous protein folding and disulfide bond formation rates yields runaway oxidative stress. *BMC Biology*, *10*, 16. doi.org/10.1186/1741-7007-10-16

Urry, L. A. (2008). Gene expression. W N. A. Campbell & J. B. Reece (Red.), *Biology* (8. wyd., s. 351–380). Pearson Education.

Vandermies, M., Denies, O., Nicaud, JM., & Fickers, P. (2017). *EYK1* encoding erythrulose kinase as a catabolic selectable marker for genome editing in the non-conventional yeast *Yarrowia lipolytica*. *Journal of Microbiological Methods*, *139*, 161–164. doi.org/10.1016/J.MIMET.2017.05.012

Verghese, J., Abrams, J., Wang, Y., & Morano, K. A. (2012). Biology of the heat shock response and protein chaperones: budding yeast (*Saccharomyces cerevisiae*) as a model system. *Microbiology and Molecular Biology Reviews*, *76*(2), 115–158. doi.org/10.1128/mmbr.05018-11

Vidal, L., Lebrun, E., Park, Y. K., Mottet, G., & Nicaud, JM. (2023). Bidirectional hybrid erythritol-inducible promoter for synthetic biology in *Yarrowia lipolytica*. *Microbial Cell Factories*, *22*, 7. doi.org/10.1186/s12934-023-02020-6

Walker, C., Ryu, S., & Trinh, C. T. (2019). Exceptional solvent tolerance in *Yarrowia lipolytica* is enhanced by sterols. *Metabolic Engineering*, *54*, 83–95. doi.org/10.1016/j.ymben.2019.03.003

Wang, C., Lin, M., Yang, Z., Lu, X., Liu, Y., Lu, H., Zhu, J., Sun, X., & Gu, Y. (2023). Characterization of the endogenous promoters in *Yarrowia lipolytica* for the biomanufacturing applications. *Process Biochemistry*, *124*, 245–252. doi.org/10.1016/j.procbio.2022.11.023

Wang, Z. P., Xu, H. M., Wang, G. Y., Chi, Z., & Chi, Z. M. (2013). Disruption of the *MIG1* gene enhances lipid biosynthesis in the oleaginous yeast *Yarrowia lipolytica* ACA-DC 50109. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, *1831*(4), 675–682. doi.org/10.1016/j.bbalip.2012.12.010

Winkler, J. D., & Kao, K. C. (2014). Recent advances in the evolutionary engineering of industrial biocatalysts. *Genomics*, *104*(6), 406–411. doi.org/10.1016/j.ygeno.2014.09.006

Wong, L., Engel, J., Jin, E., Holdridge, B., & Xu, P. (2017). YaliBricks, a versatile genetic toolkit for streamlined and rapid pathway engineering in *Yarrowia lipolytica*. *Metabolic Engineering Communications*, *5*, 68–77. doi.org/10.1016/j.meteno.2017.09.001

Wu, H., Shu, T., Mao, Y. S., & Gao, X. D. (2020). Characterization of the promoter, downstream target genes and recognition DNA sequence of Mhy1, a key filamentation-promoting transcription factor in the dimorphic yeast *Yarrowia lipolytica*. *Current Genetics*, *66*(1), 245–261. doi.org/10.1007/s00294-019-01018-1

Yang, L. B., Dai, X. M., Zheng, Z. Y., Zhu, L., Zhan, X. B., & Lin, C. C. (2015). Proteomic analysis of erythritol-producing Y*arrowia lipolytica* from glycerol in response to osmotic pressure. *Journal of Microbiology and Biotechnology*, *25*(7), 1056–1069. doi.org/10.4014/jmb.1412.12026

Yuan, B., Wang, W. Bin, Wang, X. Q., Liu, C. G., Hasunuma, T., Kondo, A., & Zhao, X. Q. (2024). The chromatin remodeler Ino80 regulates yeast stress tolerance and cell metabolism through modulating nitrogen catabolite repression. *International Journal of Biological Macromolecules*, *258*. doi.org/10.1016/j.ijbiomac.2023.129041

Yuan, B., Zhu, Y. F., Li, K., & Zhao, X. Q. (2025). Chromatin regulation of acetic acid stress tolerance by Ino80 in budding yeast *Saccharomyces cerevisiae*. *Journal of Agricultural and Food Chemistry*, *73*(5), 2951–2960. doi.org/10.1021/acs.jafc.4c09994

Zahrl, R. J., Prielhofer, R., Burgard, J., Mattanovich, D., & Gasser, B. (2023). Synthetic activation of yeast stress response improves secretion of recombinant proteins. *New Biotechnology*, *73*, 19–28. doi.org/10.1016/j.nbt.2023.01.001

Zhang, L., & Hach, A. (1999). Molecular mechanism of heme signaling in yeast: the transcriptional activator Hap1 serves as the key mediator. *Cellular and molecular life sciences: CMLS*, 56(5-6), 415–426. doi.org/10.1007/s000180050442

Zhang, Y., Wang, L., Liang, S., Zhang, P., Kang, R., Zhang, M., Wang, M., Chen, L., Yuan, H., Ding, S., & Li, H. (2020). FpDep1, a component of Rpd3L histone deacetylase complex, is important for vegetative development, ROS accumulation, and pathogenesis in *Fusarium pseudograminearum*. *Fungal Genetics and Biology*, *135*, 103299. doi.org/10.1016/j.fgb.2019.103299

9. Aktywność naukowa i osiągnięcia doktorantki

9.1 Udział w realizacji projektów badawczych

- Główny wykonawca w grancie Narodowego Centrum Nauki OPUS 21 „Czynniki transkrypcyjne jako narzędzia masowego działania w ulepszaniu cech przemysłowych u drożdży", numer 2021/41/B/NZ9/00086, pod kierownictwem prof. dr hab. Eweliny Celińskiej

- Praca magisterska „Wpływ wybranych warunków stresowych na wzrost biomasy oraz ekspresję heterologicznych genów w *Yarrowia lipolytica*" wykonana w Katedrze Biotechnologii i Mikrobiologii Żywności Uniwersytetu Przyrodniczego w Poznaniu oraz w jednostce badawczej TERRA, Agro-Bio Tech Uniwersytetu w Liège, w Belgii, pod kierownictwem prof. dr hab. Eweliny Celińskiej

- Wolontariat w projekcie MNiSW Iuventus Plus V, nr IP2015 011074 „Produkcja heterologicznych enzymów amylolitycznych w rekombinowanych komórkach drożdży i ich aplikacja w procesach biotechnologicznych" (08.2018 – 06.2020, z pięciomiesięcznym okresem przerwy na naukowy staż zagraniczny), pod kierownictwem prof. dr hab. Eweliny Celińskiej

- Praca inżynierska "Ocena zdolności do wzrostu z nieskleikowanej skrobi rekombinowanych szczepów *Yarrowia lipolytica* o nabytej zdolności amylolityczej" wykonana w Katedrze Biotechnologii i Mikrobiologii Żywności Uniwersytetu Przyrodniczego w Poznaniu, pod kierownictwem prof. dr hab. Eweliny Celińskiej

9.2 Staże naukowe

- Staż naukowy w ramach grantu EMBO Scientific Exchange Grant przeprowadzony w Centrum Biosustain NovoNordisk na Uniwerystecie Technicznym w Danii, Kongens Lyngby, Dania (01.2025 – 06.2025) z dodatkowym finansowaniem z programu Erasmus+ Praktyki, pod kierownictwem prof. Iriny Borodiny

- Staż w ramach projektu „Najlepsi z Natury 2.0" zrealizowany w firmie Sanprobi, Szczecin (07.2021 – 09.2021)

- Staż naukowy w ramach projektów „Studiujesz?-Praktykuj!" oraz Erasmus+ Praktyki przeprowadzony w jednostce badawczej TERRA, Agro-Bio Tech Gembloux, podlegającej pod Uniwersytet w Liège w Belgii (08.2019 – 10.2019) pod kierownictwem prof. Patricka Fickersa

9.3 Publikacje naukowe, niewchodzące w skład rozprawy doktorskiej

- **Gorczyca M**. (2025)
  DNA jako klucz do nowych perspektyw w badaniach geologicznych
  W Belzyt S., Marszałek H., Modelska M., Ziemniak G. (Red.) Przewodnik geologiczno-kulturowy rejsu badawczo-dydaktycznego wzdłuż wybrzeży zachodniego Spitsbergenu „GeoArctic Expedition 2025. Sailing Through Time II".
  Instytut Nauk Geologicznych Uniwersytetu Wrocławskiego (s 38-42),
  ISBN 978-83-907933-4-4

- **Gorczyca M**., Kaźmierczak J., Fickers P., Celińska E. (2022)
  Synthesis of secretory proteins in Yarrowia lipolytica: Effect of combined stress factors and metabolic load
  International Journal of Molecular Sciences, 23, 3602
  IF (5-letni): 6.2
  Punkty MEiN: 140

- **Gorczyca M**., Kaźmierczak J., Steels S., Fickers P., Celińska E. (2020)
  Impact of oxygen availability on heterologous gene expression and polypeptide secretion dynamics in Yarrowia lipolytica-based protein production platforms.
  Yeast 37(9-10):559-568
  IF (5-letni): 3.2
  Punkty MEiN: 100

- Celińska E., Borkowska M., Korpys-Woźniak P., Kubiak M., Nicaud JM., Kubiak P., **Gorczyca M**., Białas W. (2020)
  Optimization of Yarrowia lipolytica-based consolidated biocatalyst through synthetic biology approach: transcription units and signal peptides shuffling
  Applied Microbiology and Biotechnology, 1-15
  IF (5-letni): 4.7
  Punkty MEiN: 100

9.4 Prezentacja wyników badań na konferencjach naukowych

9.4.1 Prezentacje ustne

- E. Celińska, **M. Gorczyca**. „'Small Volume – Big Problem': Culturing *Yarrowia lipolytica* in high-throughput micro-formats". Prezentacja ustna podczas 38th International Specialized Symposium on Yeasts 'Yeast – The Omni-Tool' 1–5.09.2025, Warszawa, Polska

- **M. Gorczyca**, P. Korpys-Woźniak, E. Celińska. „It's complicated" – an interplay between transcription factors and recombinant protein synthesis in *Yarrowia lipolytica* at transcriptional and functional levels. Prezentacja ustna podczas 16th International Congress on Yeasts (International Commission on Yeasts) 29.09–3.10.2024, Kapsztad, Republika Południowej Afryki

- E. Celińska, P. Korpys-Woźniak, **M. Gorczyca**, D. Onesime, L. Vidal, JM. Nicaud. „Multifaceted approach for developing a synthetic protein-overproducer phenotype in *Yarrowia lipolytica*". Prezentacja ustna podczas 16th International Congress on Yeasts (International Commission on Yeasts) 29.09–3.10.2024, Kapsztad, Republika Południowej Afryki

- **M. Gorczyca**, W. Białas, JM. Nicaud, E. Celińska. "'Mother(Nature) knows best' - hijacking nature-designed transcriptional programs for enhancing stress resistance and protein production in *Yarrowia lipolytica*; presentation of YaliFunTome database". Prezentacja ustna podczas 37th International Specialized Symposium on Yeast 'Yeast Biotech 2.0' 27.11–1.12.2023, Adelaide, Australia

- P. Korpys-Woźniak, **M. Gorczyca**, JM. Nicaud, E. Celińska. „One-for-all: engineering recombinant protein synthesis in *Yarrowia lipolytica* by the use of transcription factors and secretory helpers". Zaproszony prelegent podczas Kongresu FEMS 2023 (Federation of European Microbiological Societies) 9-13.07.2023, Hamburg, Niemcy

- **M. Gorczyca**, JM. Nicaud, E. Celińska. „Transcription factors as tools for enhancing recombinant protein synthesis and resistance to stress in *Yarrowia lipolytica*". Prezentacja ustna podczas 47th Annual Conference on Yeasts 16-19.05.2023, Smolenice, Słowacja

- **M. Gorczyca**, W. Białas, JM Nicaud, E. Celińska. „Co-expression of selected transcription factors modulates synthesis of heterologous proteins in *Yarrowia lipolytica* under stress conditions". Prezentacja ustna podczas 1st Polish Yeast Conference, 22-24.06.2022, Rzeszów, Polska

- **M. Gorczyca**, J. Kaźmierczak, E. Celińska. „Optymalizacja procesów biotechnologicznej produkcji a żywotność dimorficznych drożdży *Yarrowia lipolytica*". Prezentacja ustna podczas III Ogólnopolskiej Konferencji Studenckiej „Biotechnologia Niejedno Ma Imię", 21.11.2020, Poznań, Polska

- J. Kaźmierczak, **M. Gorczyca**, M. Borkowska, E. Celińska. „Stabilność białek reporterowych – niedoceniany a kluczowy aspekt badań". Prezentacja ustna podczas III Ogólnopolskiej Konferencji Studenckiej „Biotechnologia Niejedno Ma Imię", 21.11.2020, Poznań, Polska

- **M. Gorczyca**, J. Kaźmierczak, E. Celińska, P. Fickers. „Warunki indukujące przemianę morfologiczną komórek drożdży *Yarrowia lipolytica*". Prezentacja ustna podczas II Ogólnopolskiej Konferencji Studenckiej „Biotechnologia Niejedno Ma Imię", 23-24.11.2019, Poznań, Polska

- E. Celińska, M. Borkowska, P. Korpys, M. Kubiak, W. Białas, **M. Gorczyca**, JM Nicaud. "Serious study on secretory protein synthesis and optimization of multi-gene expression constructs for *Yarrowia lipolytica*". Prezentacja ustna podczas 46th Annual Conference on Yeasts, 7-10.05.2019, Smolenice, Słowacja

9.4.2 Postery i krótkie prezentacje ustne

- **M. Gorczyca**, JM Nicaud, E. Celińska. „Transcription factors as tools of massive operation in engineering of industrial traits in *Yarrowia lipolytica*". Poster i krótka prezentacja ustna podczas 38th International Specialized Symposium on Yeasts 'Yeast – The Omni-Tool' 1-5.09.2025, Warszawa, Polska

- P. Korpys-Woźniak, **M. Gorczyca**, E. Celińska. „Omics-driven engineering of protein synthesis and secretion in *Yarrowia lipolytica*". Poster podczas 16th International Congress on Yeasts (International Commission on Yeasts) 29.09-3.10.2024, Kapsztad, Republika Południowej Afryki

- **M. Gorczyca**, E. Celińska. "Expression profile of selected transcription factors under internal and external stress conditions in yeast *Yarrowia lipolytica*". Poster podczas Kongresu FEMS 2023 (Federation of European Microbiological Societies) 9-13.07.2023, Hamburg, Niemcy

- **M. Gorczyca**, W. Białas, JM. Nicaud, E. Celińska. „Co-overexpression of transcription factors modulates r-Prot synthesis under a wide range of environmental conditions in *Yarrowia lipolytica*". Poster i krótka prezentacja ustna podczas 47th Annual Conference on Yeasts 16-19.05.2023, Smolenice, Słowacja

- **M. Gorczyca**, J. Kaźmierczak, S. Steels, M. Borkowska, P. Fickers, E. Celińska. „Insight into heterologous protein synthesis and physiology of *Yarrowia lipolytica* under exposure to combination of selected stress factors". Poster podczas 15th International Congress on Yeast and 30th International Conference on Yeast Genetics and Molecular Biology 'The Spirit of Yeast', 23-27.08.2021, Wiedeń, Austria

9.5 Nagrody i wyróżnienia

- Nagroda Polskiego Towarzystwa Biochemicznego za najlepszą prezentację ustną podczas 38th International Specialized Symposium on Yeasts: „Yeast – The Omni-Tool", 1-5.09.2025, Warszawa, Polska

- Grant EMBO Scientific Exchange na trzymiesięczną wizytę naukową Centrum Biosustain NovoNordisk na Uniwerystecie Technicznym w Danii, Kongens Lyngby (08.2024)

- Nagroda FEMS Yeast Research za najlepszy poster podczas konferencji 47th Annual Conference on Yeasts 16-19.05.2023, Smolenice, Słowacja

- Nagroda Komitetu Organizacyjnego za najlepszy poster podczas konferencji 47th Annual Conference on Yeasts 16-19.05.2023, Smolenice, Słowacja

- Grant FEMS dla Młodych Naukowców na uczestnictwo w konferencji – 1st Polish Yeast Conference, 22-24.06.2022, Rzeszów, Polska

- Stypendium Ministra Edukacji i Nauki za znaczące osiągnięcia naukowe na rok akademicki 2020/21

- Znalezienie się wśród najlepszych absolwentów studiów pierwszego stopnia Uniwersytetu Przyrodniczego w Poznaniu w roku akademickim 2019/20 oraz najlepszych absolwentów studiów drugiego stopnia w roku akademickim 2020/21

- Nagroda za najlepszą prezentację ustną podczas II Ogólnopolskiej Konferencji „Biotechnologia Niejedno Ma Imię" wygłoszoną w sesji Biopaliwa, 23-24.11.2019, Poznań, Polska

- Stypendia Rektora dla najlepszych studentów w latach: 2017/18, 2018/19, 2019/20, 2020/21

9.6 Dodatkowa działalność naukowa

- Uczestnictwo w rejsie naukowo-dydaktycznym wzdłuż wybrzeży zachodniego Spitsbergenu „GeoArctic Expedition 2025. Sailing Through Time II" i koordynacja zadań biologicznych wyprawy, we współpracy z Instytutem Nauk Geologicznych Uniwersytetu Wrocławskiego oraz Instytutem Oceanologii Polskiej Akademii Nauk w Sopocie

- Recenzja artykułu naukowego w międzynarodowym czasopiśmie *World Journal of Microbiology and Biotechnology* (Springer Nature Group)

- Studia podyplomowe w Polsko-Japońskiej Akademii Technik Komputerowych w Warszawie na kierunku Bioinformatyka (10.2023 – 08.2024)

- Certyfikat Cambridge: Certificate of Proficiency in English (poziom C2), 08.2022

- Działalność w Studenckim Kole Naukowym „OPERON" (rok akademicki 2017/18)

10. Oświadczenia współautorów publikacji

Mgr inż. Maria Gorczyca
Katedra Biotechnologii i Mikrobiologii Żywności
Uniwersytet Przyrodniczy w Poznaniu
60-637 Poznań, Polska

Oświadczenie o współautorstwie

Niniejszym oświadczam, że w pracy **Maria Gorczyca**, Jean-Marc Nicaud, Ewelina Celińska (2023). Transcription factors enhancing synthesis of recombinant proteins and resistance to stress in *Yarrowia lipolytica*. *Applied Microbiology and Biotechnology*, 107, 4853–4871 mój indywidualny udział w jej powstaniu polegał na autorstwie koncepcji oraz przygotowaniu projektu badań wraz z Jean-Marc Nicaud i Eweliną Celińską, przeprowadzeniu większości eksperymentów, w tym hodowli, pobieraniu i analizie próbek, opracowaniu danych, analizach statystycznych i wizualizacji wyników, a także na interpretacji wyników i opracowaniu manuskryptu wraz z Eweliną Celińską.

Oświadczam, że w pracy **Maria Gorczyca**, Wojciech Białas, Jean-Marc Nicaud, Ewelina Celińska (2024). 'Mother(Nature) knows best' – hijacking nature-designed transcriptional programs for enhancing stress resistance and protein production in *Yarrowia lipolytica*; presentation of YaliFunTome database. *Microbial Cell Factories*, 23:26 mój indywidualny udział w jej powstaniu polegał na opracowaniu metodologii, analizie i interpretacji wyników, opracowaniu manuskryptu wraz z Eweliną Celińską, przeprowadzeniu wszystkich eksperymentów, analizie danych oraz stworzeniu modeli matematycznych.

Oświadczam, że w pracy Ewelina Celińska, **Maria Gorczyca** (2024). 'Small volume—big problem': culturing *Yarrowia lipolytica* in high-throughput micro-formats. *Microbial Cell Factories*, 23:184 mój indywidualny udział w jej powstaniu polegał na przeprowadzeniu eksperymentu typu '*proof-of-concept*' (sekcja *How small can we go with Yarrowia? And what are the costs?*), analizie danych oraz wizualizacji wyników, przygotowaniu wszystkich elementów graficznych oraz uczestnictwie w pisaniu manuskryptu.

Oświadczam, że w pracy **Maria Gorczyca**, Paulina Korpys-Woźniak, Ewelina Celińska (2024). An Interplay between Transcription Factors and Recombinant Protein Synthesis in *Yarrowia lipolytica* at Transcriptional and Functional Levels—The Global View. *International Journal of Molecular Sciences*, 25(17), 9450 mój indywidualny udział w jej powstaniu polegał

na współautorstwie koncepcji badań, analizie i interpretacji wyników wraz z Eweliną Celińską, opracowaniu manuskryptu wraz z Eweliną Celińską i Pauliną Korpys-Woźniak, opracowaniu metodologii badań, przeprowadzeniu analiz komputerowych, przeprowadzeniu badań laboratoryjnych, zapewnieniu zasobów (dane fenotypowe) oraz wizualizacji wyników.

Oświadczam, że w pracy Ewelina Celińska, Paulina Korpys-Woźniak, **Maria Gorczyca**, Jean-Marc Nicaud (2024). Using Euf1 transcription factor as a titrator of erythritol-inducible promoters in *Yarrowia lipolytica*; insight into the structure, splicing, and regulation mechanism. *FEMS Yeast Research*, 24 mój indywidualny udział w jej powstaniu polegał na przeprowadzeniu części eksperymentów, wykonaniu analiz bioinformatycznych, w tym modeli struktury białka oraz przewidywań interakcji białko–ligand.

Data Podpis

16.09.2025 Maria Gorczyca

Prof. dr hab. Ewelina Celińska
Katedra Biotechnologii i Mikrobiologii Żywności
Uniwersytet Przyrodniczy w Poznaniu
60-637 Poznań, Polska

Oświadczenie o współautorstwie

Niniejszym oświadczam, że jestem współautorem publikacji wchodzących w skład rozprawy doktorskiej mgr inż. Marii Gorczycy:

Maria Gorczyca, Jean-Marc Nicaud, **Ewelina Celińska** (2023). Transcription factors enhancing synthesis of recombinant proteins and resistance to stress in *Yarrowia lipolytica*. Applied Microbiology and Biotechnology, 107, 4853–4871

Maria Gorczyca, Wojciech Białas, Jean-Marc Nicaud, **Ewelina Celińska** (2024). 'Mother(Nature) knows best' – hijacking nature-designed transcriptional programs for enhancing stress resistance and protein production in *Yarrowia lipolytica*; presentation of YaliFunTome database. Microbial Cell Factories, 23:26

Ewelina Celińska, Maria Gorczyca (2024). 'Small volume—big problem': culturing *Yarrowia lipolytica* in high-throughput micro-formats. *Microbial Cell Factories*, 23:184

Maria Gorczyca, Paulina Korpys-Woźniak, **Ewelina Celińska** (2024). An Interplay between Transcription Factors and Recombinant Protein Synthesis in *Yarrowia lipolytica* at Transcriptional and Functional Levels—The Global View. *International Journal of Molecular Sciences*, 25(17), 9450

Ewelina Celińska, Paulina Korpys-Woźniak, Maria Gorczyca, Jean-Marc Nicaud (2024). Using Euf1 transcription factor as a titrator of erythritol-inducible promoters in *Yarrowia lipolytica*; insight into the structure, splicing, and regulation mechanism. FEMS Yeast Research, 24

Oświadczam, że w pracy Maria Gorczyca, Jean-Marc Nicaud, Ewelina Celińska (2023). Transcription factors enhancing synthesis of recombinant proteins and resistance to stress in *Yarrowia lipolytica*. *Applied Microbiology and Biotechnology*, 107, 4853–4871 mój indywidualny udział w jej powstaniu polegał na autorstwie koncepcji oraz przygotowaniu projektu badań wraz z Jean-Marc Nicaud i Marią Gorczycą, nadzorowaniu przebiegu prac, zapewnieniu finansowania, interpretacji wyników i opracowaniu manuskryptu wraz z Marią Gorczycą.

Oświadczam, że w pracy Maria Gorczyca, Wojciech Białas, Jean-Marc Nicaud, Ewelina Celińska (2024). 'Mother(Nature) knows best' – hijacking nature-designed transcriptional programs for enhancing stress resistance and protein production in *Yarrowia lipolytica*; presentation of YaliFunTome database. *Microbial Cell Factories,* 23:26 mój indywidualny udział w jej powstaniu polegał na autorstwie koncepcji wraz z Jean-Marc Nicaud; opracowaniu metodologii, analizie i interpretacji wyników, opracowaniu manuskryptu wraz z Marią Gorczycą; nadzorowaniu przebiegu prac oraz zapewnieniu środków finansowych.

Oświadczam, że w pracy Ewelina Celińska, Maria Gorczyca (2024). 'Small volume—big problem': culturing *Yarrowia lipolytica* in high-throughput micro-formats. *Microbial Cell Factories,* 23:184 mój indywidualny udział w jej powstaniu polegał na przygotowaniu koncepcji i projektu pracy, opracowaniu krytycznego przeglądu literatury oraz opracowaniu manuskryptu.

Oświadczam, że w pracy Maria Gorczyca, Paulina Korpys-Woźniak, Ewelina Celińska (2024). An Interplay between Transcription Factors and Recombinant Protein Synthesis in *Yarrowia lipolytica* at Transcriptional and Functional Levels—The Global View. *International Journal of Molecular Sciences,* 25(17), 9450 mój indywidualny udział w jej powstaniu polegał na autorstwie koncepcji badań wraz z Marią Gorczycą, zapewnieniu zasobów badawczych, w tym konstrukcji szczepów z delecjami oraz zapewnieniu odczynników laboratoryjnych, analizie i interpretacji wyników wraz z Marią Gorczycą, opracowaniu manuskryptu wraz z Marią Gorczycą i Pauliną Korpys-Woźniak, nadzorowaniu przebiegu prac, zarządzaniu projektem i pozyskaniu finansowania.

Oświadczam, że w pracy Ewelina Celińska, Paulina Korpys-Woźniak, Maria Gorczyca, Jean-Marc Nicaud (2024). Using Euf1 transcription factor as a titrator of erythritol-inducible promoters in *Yarrowia lipolytica*; insight into the structure, splicing, and regulation mechanism. *FEMS Yeast Research,* 24 mój indywidualny udział w jej powstaniu polegał na opracowaniu koncepcji, zaprojektowaniu wszystkich eksperymentów, skonstruowaniu szczepów, wykonaniu części eksperymentów, analizie i interpretacji wyników oraz opracowaniu manuskryptu.

Data

08/09/2025

Podpis

Jean-Marc Nicaud
Université Paris-Saclay, INRAE,
AgroParisTech, Micalis Institute,
78350 Jouy-en-Josas, France

Statment of Co-authorship

I hereby certify that I am a co-author of publications being a part of a series of publications included in the doctoral dissertation of MSc Eng. Maria Gorczyca:

Maria Gorczyca, **Jean-Marc Nicaud**, Ewelina Celińska (2023). Transcription factors enhancing synthesis of recombinant proteins and resistance to stress in *Yarrowia lipolytica*. *Applied Microbiology and Biotechnology*, Volume 107, pages 4853–4871

Maria Gorczyca, Wojciech Białas, **Jean-Marc Nicaud**, Ewelina Celińska (2024). 'Mother(Nature) knows best' – hijacking nature-designed transcriptional programs for enhancing stress resistance and protein production in *Yarrowia lipolytica*; presentation of YaliFunTome database. *Microbial Cell Factories*, Volume 23, Article number: 26

Ewelina Celińska, Paulina Korpys-Woźniak, Maria Gorczyca, **Jean-Marc Nicaud** (2024). Using Eufl transcription factor as a titrator of erythritol-inducible promoters in *Yarrowia lipolytica*; insight into the structure, splicing, and regulation mechanism. FEMS Yeast Research, Volume 24

I declare that my participation in an article entitled: Transcription factors enhancing synthesis of recombinant proteins and resistance to stress in *Yarrowia lipolytica*, published in: *Applied Microbiology and Biotechnology*, Volume 107, pages 4853–4871, by: Maria Gorczyca, Jean-Marc Nicaud, Ewelina Celińska, consisted of conceptualization and design of the research along with Ewelina Celińska and Maria Gorczyca, as well as revision of the draft and approval of the manuscript.

1/2

In an article entitled: 'Mother(Nature) knows best' – hijacking nature-designed transcriptional programs for enhancing stress resistance and protein production in *Yarrowia lipolytica*; presentation of YaliFunTome database, published in: *Microbial Cell Factories*, Volume 23, Article number: 26, by: Maria Gorczyca, Wojciech Białas, **Jean-Marc Nicaud**, Ewelina Celińska, I declare my individual participation of: conceptualization of the project along with Ewelina Celińska, and manuscript revision and approval.

I certify that my participation in an article entitled: Using Eufl transcription factor as a titrator of erythritol-inducible promoters in *Yarrowia lipolytica*; insight into the structure, splicing, and regulation mechanism, published in: FEMS Yeast Research, Volume 24, by: Ewelina Celińska, Paulina Korpys-Woźniak, Maria Gorczyca, Jean-Marc Nicaud, consisted of inspiring the promoter-titration experiment and providing the background strains for the experiment.

Date

07 września 2025

Signature

Jean-Marc NICAUD

Dr inż. Paulina Korpys-Woźniak
Katedra Biotechnologii i Mikrobiologii Żywności
Uniwersytet Przyrodniczy w Poznaniu
60-637 Poznań, Polska

Oświadczenie o współautorstwie

Niniejszym oświadczam, że jestem współautorem publikacji wchodzących w skład rozprawy doktorskiej mgr. inż. Marii Gorczycy:

Maria Gorczyca, **Paulina Korpys-Woźniak**, Ewelina Celińska (2024). An Interplay between Transcription Factors and Recombinant Protein Synthesis in *Yarrowia lipolytica* at Transcriptional and Functional Levels—The Global View. *International Journal of Molecular Sciences*, 25(17), 9450

Ewelina Celińska, **Paulina Korpys-Woźniak**, Maria Gorczyca, Jean-Marc Nicaud (2024). Using Euf1 transcription factor as a titrator of erythritol-inducible promoters in *Yarrowia lipolytica*; insight into the structure, splicing, and regulation mechanism. FEMS Yeast Research, 24

Oświadczam, że w pracy Maria Gorczyca, Paulina Korpys-Woźniak, Ewelina Celińska (2024). An Interplay between Transcription Factors and Recombinant Protein Synthesis in *Yarrowia lipolytica* at Transcriptional and Functional Levels—The Global View. *International Journal of Molecular Sciences*, 25(17), 9450 mój indywidualny udział w jej powstaniu polegał na zapewnieniu zasobów (dane transkryptomiczne) oraz opracowaniu manuskryptu wraz z Marią Gorczycą i Eweliną Celińską.

Oświadczam, że w pracy Ewelina Celińska, Paulina Korpys-Woźniak, Maria Gorczyca, Jean-Marc Nicaud (2024). Using Euf1 transcription factor as a titrator of erythritol-inducible promoters in *Yarrowia lipolytica*; insight into the structure, splicing, and regulation mechanism. *FEMS Yeast Research*, 24 mój indywidualny udział w jej powstaniu polegał na przeprowadzeniu znaczącej części eksperymentów laboratoryjnych.

Data

18.09.2025 r.

Podpis

Paulina Korpys-Woźniak

Prof. UPP dr hab. Wojciech Białas
Katedra Biotechnologii i Mikrobiologii Żywności
Uniwersytet Przyrodniczy w Poznaniu
60-637 Poznań, Polska

Oświadczenie o współautorstwie

Niniejszym oświadczam, że jestem współautorem publikacji wchodzącej w skład rozprawy doktorskiej mgr inż. Marii Gorczycy:

Maria Gorczyca, **Wojciech Białas**, Jean-Marc Nicaud, Ewelina Celińska (2024). 'Mother(Nature) knows best' – hijacking nature-designed transcriptional programs for enhancing stress resistance and protein production in *Yarrowia lipolytica*; presentation of YaliFunTome database. Microbial Cell Factories, 23:26

Oświadczam, że w powyższej pracy mój indywidualny udział w jej powstaniu polegał na opracowaniu planu eksperymentalnego w oprogramowaniu Design Expert.

Data

.....

Podpis

Signed by /
Podpisano przez:
Wojciech Białas
Date / Data:
2025-09-17 16:34

.....