Implementing Fluorine as Bioelement through Adaptive Laboratory Evolution of *Escherichia coli*

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

CHRISTIN TREIBER-KLEINKE, geb. Treiber, M.Sc.

from Berlin, Germany

February 2024

Nothing worth having was ever achieved without effort.

- Theodore Roosevelt

The research presented in this doctoral thesis was performed under the supervision of Prof. Dr. Beate Koksch and in collaboration with Prof. Dr. Nediljko Budisa from May 2018 until February 2024 at the Institute of Chemistry and Biochemistry in the Department of Biology, Chemistry, Pharmacy of Freie Universität Berlin.

1st reviewer: Prof. Dr. Beate Koksch

2nd reviewer: Prof. Dr. Nediljko Budisa

Date of disputation: 29.04.2024

Declaration of Independence

Herewith, I certify that I have prepared and written my thesis entitled "Implementing Fluorine as Bioelement through Adaptive Laboratory Evolution of *Escherichia coli*" independently and that I have not used any sources and aids other than those indicated by me. This thesis has not been submitted, accepted, rated as insufficient, or rejected in any other doctorate degree procedure.

Berlin, February 2024   Christin Treiber-Kleinke

Publications, Oral and Poster Presentations

Publications

- **Treiber-Kleinke C,** Berger AA, Adrian L, Budisa N and Koksch B **(2024),** *Escherichia coli* adapts metabolically to 6- and 7-fluoroindole, enabling proteome-wide fluorotryptophan substitution. *Front. Synth. Biol.* 1:1345634. doi: 10.3389/fsybi.2023.1345634

  This publication emerged from this doctoral thesis. (A preprint of the article has been published online on BioRxiv in 2023 and can be found at doi: https://doi.org/10.1101/2023.09.25.559291).

- Tolle I, Oehm S, Hoesl MG, **Treiber-Kleinke C**, Peil L, Bozukova M, Albers S, Adamu Bukari A-R, Semmler T, Rappsilber J, Ignatova Z, Gerstein AC and Budisa N (**2023**), Evolving a mitigation of the stress response pathway to change the basic chemistry of life. *Front. Synth. Biol.* 1:1248065. doi: 10.3389/fsybi.2023.1248065

  (A preprint of the article has been published online on BioRxiv in 2021 and can be found at doi: https://doi.org/10.1101/2021.09.23.461486).

Oral Presentations at Conferences and Workshops

- "Integration von Fluor in lebende Organismen"; 19. Deutscher Fluortag, Schmitten/Germany, 19.09.2022 – 21.09.2022 **(Presentation Award)**

- "Microevolution of *E. coli* towards usage of fluorinated Tryptophan"; Winter Symposium of the CRC 1349 "Fluorine-Specific Interactions: Fundamentals and Functions", Berlin/Germany, 15.03.2022 – 16.03.2022

- "Microevolution of *E. coli* towards fluorinated Trp"; 4th Webex-Seminar of the CRC 1349 "Fluorine-Specific Interactions: Fundamentals and Functions", Berlin/Germany, 10.08.2020

- "Adaptation towards fluorinated tryptophans"; Short Poster Talk at 14th German Peptide Symposium, Cologne/Germany, 18.03.2019 – 21.03.2019

Posters

- "Integration of fluorine into living organisms"; Fluorine Summer School of the CRC 1349 "Fluorine-Specific Interactions: Fundamentals and Functions", Berlin/Germany, 28.08.2023 – 30.08.2023

- "Residue-specific incorporation of artificial fluorinated amino acids into proteins"; 8th Chemical Protein Synthesis Meeting, Berlin/Germany, 16.06.2019 – 19.06.2019

- "Adaptation towards fluorinated tryptophans"; Short Poster Talk at 14th German Peptide Symposium, Cologne, Germany, 18.03.2019 – 21.03.2019

Acknowledgements

Abstract

Fluorine is almost excluded from the biotic world. Although largely ignored by nature, a life based on fluorine is both an interesting concept and an absolutely conceivable scenario. Fluorine-containing building blocks have been extensively used to investigate and modify proteins and their interactions. However, the adaptation of living organisms represents a step forward in exploring and understanding the consequences of global fluorination. Hence, a reliable platform and methodologies for the experimental evolution of novel organisms with fluorine as a bioelement were established, created by mimicking natural selection in the laboratory.

In this study, adaptive laboratory evolution (ALE) was used to force Tryptophan-auxotrophic *Escherichia coli* bacteria to fully adapt metabolically to grow and proliferate indefinitely on 6-fluoroindole (6Fi) and 7-fluoroindole (7Fi) as substitutes for indole (Ind). During long-term cultivation, fluorinated indole precursors were supplied, in situ converted into the corresponding fluorinated amino acids, and 6-fluorotryptophan (6FTrp) and 7-fluorotryptophan (7FTrp) were incorporated into the proteome in response to the UGG codon. As consequence of the imposed selection pressure, the cells relinquished their dependence on canonical Trp and instead acquired the ability to use either of the fluorinated counterparts as an integrated part of their metabolism.

In the course of this process, five independent descendant *E. coli* evolved that have converted their entire lifestyle to exist on these unnatural molecules. 6FTrp adaptation revealed to be superior to 7FTrp adaption. While in 6FTrp even tendencies of an indole-rejecting phenotype were observed, on the contrary, the basic adaptation to 7FTrp required extraordinary perseverance. The effect of extensive fluorination and how living organisms cope with it was investigated through comprehensive analyses of genomic, proteomic, and metabolomic levels. One major strategy found involves mitigation of the stress response system, ignoring detrimental effects caused by global fluorine integration. This study is a further step and establishes a strong foundation for further exploration of the mechanisms underlying fluorine-based life and how a former stressor (fluorinated indole) becomes a vital nutrient.

Kurzzusammenfassung

Fluor ist aus der biotischen Welt nahezu ausgeschlossen. Obwohl es von der Natur weitgehend ignoriert wird, ist ein Leben auf der Grundlage von Fluor sowohl ein interessantes Konzept als auch ein durchaus denkbares Szenario. Fluorhaltige Bausteine wurden bereits ausgiebig zur Untersuchung und Veränderung von Proteinen und ihren Wechselwirkungen verwendet. Die Anpassung von lebenden Organismen stellt jedoch einen Fortschritt bei der Erforschung und dem Verständnis der Folgen der globalen Fluorierung dar. Deshalb wurde mit dieser Arbeit eine zuverlässige Plattform und Methoden für die experimentelle Evolution neuartiger Organismen mit Fluor als Bioelement etabliert, indem die natürliche Selektion im Labor nachgeahmt wurde.

In dieser Studie wurde die adaptive Laborevolution (ALE) eingesetzt, um Tryptophan-auxotrophe *Escherichia coli* Bakterien zu zwingen, sich metabolisch vollständig anzupassen, um auf 6-Fluorindol (6Fi) und 7-Fluorindol (7Fi) als Ersatz für Indol (Ind) zu wachsen und sich unbegrenzt zu vermehren. Während der Langzeitkultivierung wurden fluorierte Indol-Vorstufen zugeführt, in situ in die entsprechenden fluorierten Aminosäuren umgewandelt, und 6-Fluortryptophan (6FTrp) und 7-Fluortryptophan (7FTrp) als Reaktion auf das UGG-Codon in das Proteom eingebaut. Als Folge des auferlegten Selektionsdrucks gaben die Zellen ihre Abhängigkeit von kanonischem Trp auf und erwarben stattdessen die Fähigkeit, die fluorierten Analoga als integrierten Teil ihres Stoffwechsels zu nutzen.

Im Laufe dieses Prozesses entwickelten sich fünf unabhängige Nachkommen von *E. coli*, die ihre gesamte Lebensweise auf diese unnatürlichen Moleküle umgestellt haben. Es zeigte sich dabei, dass die 7FTrp Anpassung deutlich schwieriger war als die Anpassung an 6FTrp. Während bei 6FTrp sogar Tendenzen eines Indol-abweisenden Phänotyps beobachtet wurden, erforderte die grundlegende Anpassung an 7FTrp dagegen außerordentliche Beharrlichkeit und Anpassung der Methoden. Durch umfassende Analysen auf Genom-, Proteom- und Metabolom-Ebene wurde untersucht, wie sich eine umfassende Fluorierung von Trp im Proteom auswirkt und wie lebende Organismen damit umgehen. Eine wichtige Strategie ist die Abschwächung des Stressreaktionssystems, wobei die schädlichen Auswirkungen der globalen Fluorintegration ignoriert werden. Diese Studie ist ein weiterer Schritt und bildet eine solide Grundlage für die weitere Erforschung der Mechanismen, die ein Leben auf der Grundlage von organischen Fluorverbindungen ermöglichen, und der Frage, wie ein ehemaliger Stressfaktor (fluoriertes Indol) zu einem lebenswichtigen Nährstoff wird.

Table of Contents

Abbreviations

6Fi	6-fluoroindole
7Fi	7-fluoroindole
6FW, 6FTrp	6-fluorotryptophan
7FW, 7FTrp	7-fluorotryptophan
aaRS	aminoacyl-tRNA synthetase
Abs.	absorption
ALE	Adaptive Laboratory Evolution
Amp	ampicillin
APS	ammonium persulfate
Ara	arabinose
bp	base pair
cAA(s)	canonical amino acid(s)
CAGO	CRISPR/Cas9-assisted gRNA-free one-step
CDS	coding DNA sequence
CFU	colony forming unit
Cm	chloramphenicol
CoA	coenzyme A
COG	Clusters of Orthologous Groups of proteins
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CV	column volume
Da	Dalton (1.66018×10^{-24} g)
dH$_2$O	distilled water
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotide triphosphates
ds	double-stranded
DTT	dithiothreitol
ECFP	enhanced cyan fluorescent protein
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EGFP	enhanced green fluorescent protein
em	emission
ESI	electrospray ionization
et. al.	et alii
EtBr	ethidium bromide

EtOH	ethanol
fwd	forward
g	multiples of the standard gravity
gDNA	genomic deoxyribonucleic acid
GO	gene ontology
HF	high-fidelity
Ind	indole
InDel	insertion/deletion
IPTG	isopropyl β-D-1-thiogalactopyranoside
Kan	kanamycin
kb	kilobase pairs
Km	Michaelis constant
LB	lysogeny broth
LC	liquid chromatography
LOD	limit of detection
LTEE	long-term evolution experiment
Mb	megabase pairs
MS	mass spectrometry
MQ-H$_2$O	MilliQ-H$_2$O (double distilled water)
ncAA(s)	non-canonical amino acid(s)
NMM	New Minimal Medium
NMR	nuclear magnetic resonance
nt	nucleotide
OD$_{600}$	optical density at λ = 600 nm
ori	origin of replication
p	passage
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
PLP	pyridoxal 5'-phosphate
ppGpp	guanosine 3'-diphosphate 5'-diphosphate
pppGpp	guanosine 3'-diphosphate 5'-triphosphate
PPi	pyrophosphate
R	resistant, resistance
rev	reverse
RNA	ribonucleic acid
RNAP	RNA polymerase
rpm	revolutions per minute

RT	room temperature
SCS	Stop Codon Suppression
SDS	sodium dodecyl sulfate
SEC	size exclusion chromatography
SNP	single nucleotide polymorphism
SOB	Super Optimal Broth
SOC	Super Optimal Broth with Catabolite Repression
SPI	Selective Pressure Incorporation
T	temperature
TAE	Tris/Acetate/EDTA
TEMED	tetramethyl ethylenediamine
TIC	total ion current
Tm	melting temperature [°C]
TOF	time of flight
Trp	L-tryptophan
tRNA	transfer ribonucleic acid
UV	ultraviolet
V	volume
WGS	whole genome sequencing
wt	wild type

Abbreviations of the 20 canonical amino acids are consistent with the one- and three-letter code recommended by the IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN).

[*Eur. J. Biochem.* **1984**, *138*, 9-37]

1   Introduction

The dawn of experimental evolution may has started with the story of William H. Dallinger and Charles Darwin.[1] Darwin already recognized evolution as an ongoing process, but he pointed out that these processes are much too slow to be observed directly. However, in 1878, during his lifetime, he was stunned and disabused from what became the first experimental evolution study by parson-naturalist Dallinger. Impressed by Darwin's theory on natural selection (1859, "On the Origin of Species"[2]), Reverend Dallinger performed long-term cultivation of monads in a controlled thermal environment, thereby successively increasing the temperature from 60 °F to 158 °F to investigate the possibility of induced adaptive changes. Indeed, the monads acquired the ability to survive under elevated temperatures, and interestingly, these resistant forms died in their ancestral medium at 65 °F, which clearly indicates a kind of adaptation. Darwin acknowledged the thermal adaptation with: "results, would no doubt, be extremely curious and valuable".[3] Dallinger thus showed, in the first-ever long-term evolution experiment, that even the simplest organisms are capable of adapting to changing environmental conditions and at the same time, that it was possible to observe the process of evolution in action through the study of fast-replicating microbes, over the course of a human lifetime.

1.1 Natural occurrence of fluorine in living organisms

1.1.1 Fluorinated natural products

Fluorine is the 13th most common element on earth crust and exists primarily in form of fluoride-containing minerals such as fluorite (fluorspar, CaF_2), fluorapatite ($Ca_5(PO_4)_3F$) and cryolite (Na_3AlF_6);[4] in a variant of fluorite (antozonite or stinkspar) indeed also natural occurring elemental difluorine (F_2) was detected.[5] Although fairly present on planet Earth as a living habitat, organofluorine is not a component of living organisms apart from a few exceptions (**Figure 1**). Inorganic fluoride was found in significant quantities in the marine sponge *Halichondria moorei*[6] and in tea plants (*Camellia sinensis*),[7] which are referred to as fluoride hyperaccumulators. In contrast, some tropical and subtropical plants and a small number of microorganisms produce organically bound fluorinated compounds.[8,9] Among them, fluoroacetate dominates because it is the most ubiquitous fluorinated metabolite found in both kingdoms - plants and microorganisms.

Fluoroacetate was the first reported organofluorine compound with biogenic origin, isolated from the highly toxic South African plant *Dichapetalum cymosum* ("gifblaar") as "potassium cymonate" in 1943[10] and one year later identified as potassium salt of monofluoroacetic acid.[11] Its toxicity is based on the interference with the tricarboxylic acid (TCA) cycle. In place of acetate, fluoroacetate converts to fluoroacetyl-CoA, which in turn substitutes for acetyl-CoA. This reaction cascade continues with the formation of fluorocitrate instead of citrate, which blocks the TCA cycle by irreversibly binding the aconitase, leading to severe inhibition of cellular respiration.[12,13] Thus, with the formation of a second fluorinated compound (fluorocitrate), the toxicity of fluoroacetate depends on metabolic activation ("lethal synthesis"[14]).

Altogether, about 40 plant species are known to produce fluoroacetate, mainly from the *Dichapetalum* genus,[9] and the formation of toxic fluoroacetate in plants is believed to be a defense mechanism against herbivores.[15] Seeds of *D. toxicarium* also contained several ω-fluorinated fatty acids (ω-fluorooleic acid ($C_{18:1F}$), ω-fluoropalmitic acid ($C_{16:0F}$), ω-fluoropalmitoleic acid ($C_{16:1F}$), ω-fluorostearic acid ($C_{18:0F}$), ω-fluorolinoleic acid ($C_{18:2F}$), ω-fluoroarachidic acid ($C_{20:0F}$) and ω-fluoroeicosenoic acid ($C_{20:1F}$)),[16] that are probably formed during fatty acid biosynthesis. Accordingly, the promiscuity of the fatty acid synthase allows fluoroacetyl-CoA as an alternative substrate of acetyl-CoA for condensation with malonyl-ACP, although the fluorine is confined to the terminal position.[8,17]
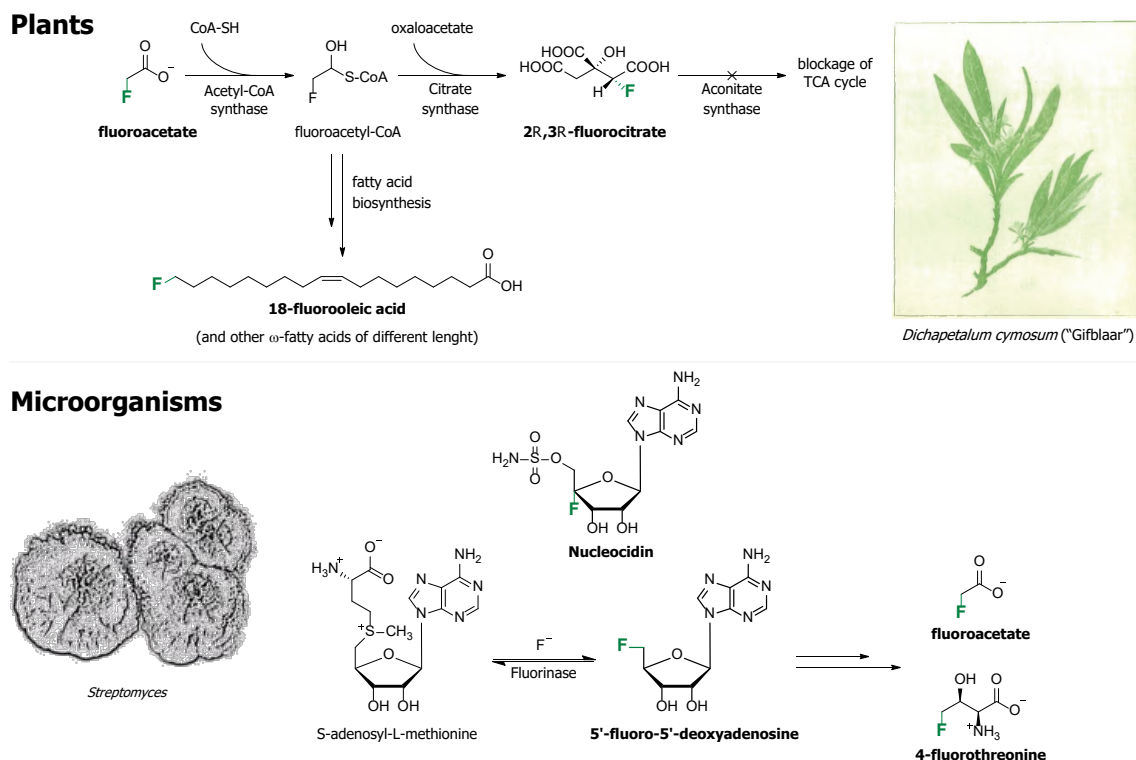
Figure 1 | Fluorinated natural products found in plants and microorganisms. In plants, represented by *Dichapetalum cymosum*,[18] fluoroacetate, its metabolite fluorocitrate, and many derivatives of ω-fatty acids were found. Microorganisms, represented by *Streptomyces* bacteria[19], can also produce fluoroacetate as well as 4-fluorotheonine, nucleocidin and 5′-fluoro-5′-deoxyadenosine (5′-FDA). Remarkably, the synthesis of the common precursor 5′-FDA is catalyzed by fluorinase, the only known enzyme capable of biological C-F bond formation.

Besides plants, a few microorganisms also have the capability to biogenically produce fluorinated compounds. The soil bacterium *Streptomyces calvus* produces the antibiotic nucleocidin (4'-fluoro-5'-*O*-sulfamoyladenosine),[20,21] which is highly effective against *Trypanosoma* parasites but, unfortunately, it turned out to be too toxic for clinical use.[22] Nowadays, 4′-fluoro-nucleosides and -nucleotides have attracted attention as therapeutics, particularly in antiviral programs.[23] Although its biosynthesis pathway is still not completely understood, it is clear that nucleocidin is distinct from all other fluorometabolites because it is not generated from the common precursor 5′-FDA via the fluorinase enzyme.[24,25]

Another example is given by the actinomyces *Streptomyces cattleya*, which is able to synthesize the amino acid 4-fluorothreonine as secondary metabolite, as well as to co-produce the toxin fluoroacetate in the presence of exogenous fluoride.[26] Efforts to elucidate the biosynthetic route culminate apriori with the discovery of a fluorinase enzyme (section **1.1.2**) and later on intense work uncovered the enzymes and intermediates involved in the metabolic pathway of 4-fluorothreonine and fluoroacetate production.[9,27] The biosynthesis starts with the fluorinase-catalyzed formation of 5′-fluoro-5′-deoxyadenosine (5′-FDA). Then, a purine nucleoside phosphorylase catalyzes the phosphorolytic displacement of the adenine base, converting 5'-FDA into 5-fluoro-5-deoxy-D-ribose-1-phosphate (5-FDRP).[28] 5-FDRP

undergoes isomerization to (3R,4S)-5-fluoro-5-deoxy-D-ribulose-1-phosphate (5-FDRulP) and after dihydroxyacetonphosphat (DHAP) cleavage mediated by metal-dependent (Zn²⁺) aldolase, fluoroacetaldehyde forms.[29] With fluoroacetaldehyde as a mutual precursor,[30] the pathway branches. Oxidation of fluoroacetaldehyde yields fluoroacetate in a reaction that is catalyzed by a NADH-dependent aldehyde dehydrogenase (fluoroacetaldehyde dehydrogenase), specifically evolved for the metabolism of fluoroacetate.[31] And the condensation of fluoroacetaldehyde with L-threonine builds the amino acid 4-fluorothreonine. Here, the responsible enzyme, a pyridoxal 5′-phosphate (PLP) dependent transaldolase, was also found to be a distinct, specialized threonine transaldolase.[32] Interestingly, the fluoride uptake and intracellular availability correlate directly with the *in vivo* production of fluorometabolites and can be temporally controlled by *S. cattleya,* which could be understood as a kind of resistance mechanism against the toxic fluoroacetate it produces.[33]

The above-described biotransformations are the only completely reported and proven biosynthesis of natural fluorinated products to date. However, in addition, the formation of a C-F bond and the production of glycosyl fluorides by mutant glycosidase enzymes of *Agrobacterium sp.* and *Cellulomonas fimi* should be mentioned[34] as well as other probably misidentified fluorometabolites reported in the literature, namely fluoroacetone, several 5-fluorouracils, and fluoroarylpropionic acid.[35]

1.1.2   The Fluorinase enzyme

From a chemist's view, the most intriguing fact on the natural occurrence of organic-bound fluorine is how nature could manage the unusual C-F bond formation. Fluorometabolite biosynthesizing *S. cattleya* provides an ideal platform for investigating this phenomenon, and since the early 2000s the knowledge of biological fluorination has developed at a breathtaking pace.

Investigations of cell-free extracts of *S. cattleya* uncovered a fluorination enzyme capable of converting inorganic fluoride into organic fluorine. The fluorinase enzyme is a 5′-fluoro-5′-deoxyadenosine (5′-FDA) synthase (EC 2.5.1.63) that catalyzes the conversion of *S*-adenosyl-L-methionine (SAM) and inorganic fluoride ions (F⁻) into 5′-fluoro-5′-deoxyadenosine (5′-FDA) and L-methionine.[36,37] 5′-FDA is the first product formed in which a C-F bond is generated and is a common precursor in the bacterial biosynthesis of organofluorine metabolites. Upon isolation and characterization of the fluorinase enzyme,[38] kinetic studies have shown that the fluoride ion binds first and has a low affinity (high Km), which likely reflects its high heat of hydration, followed by high-affinity binding of SAM (low Km) which drives its desolvation.[35,39] Fluorinase has a unique quaternary structure and folds into a hexamer composed of two trimers (**Figure 2**). Each monomer (299 amino acids, 31 kDa) features four flexible loops, whereby a 21-residue loop at A94-E114 appears to be a characteristic and essential motif of this

enzyme class.[40,41] Using site-directed mutagenesis studies, structural analyses, and isothermal calorimetry (ITC) experiments, the location of the halogen binding site, key active-site residues, and the molecular mechanism of the fluorinase-catalyzed reaction could be elucidated.[39] The solvated fluoride ion binds the enzyme by forming hydrogen bonds with the amide NH and side chain OH of S158 and possibly the hydroxyl group of T80, thereby partially losing surrounding water molecules. SAM is located at the interface between two monomeric subunits, where D16 seems to be crucial for the correct positioning of the ribose ring (by H-bonding between the 2'- and 3'-hydroxyl groups and the carboxylate side group of aspartate). Specific SAM binding causes displacement of the remaining water molecules of F⁻, which becomes desolvated, thus nucleophilic, and trapped in a reactive position of the fluorinase protein. In the transition state, F⁻ is stabilized by the positive charge (S⁺) of SAM. The fluoride ion thus activated can now attack the C5' of SAM in an S_N2 substitution reaction, displacing methionine. Here, the C-F bond formation proceeds with an inversion of the configuration at the C5' carbon.[42] Finally, methionine is released from the enzyme in which loop A95-Q102, formerly involved in its binding, undergoes a conformational change, and the release of 5'-FDA occurs equally with the T75-R85 loop. The active site constitution and mechanism of action are shown in **Figure 2B**.
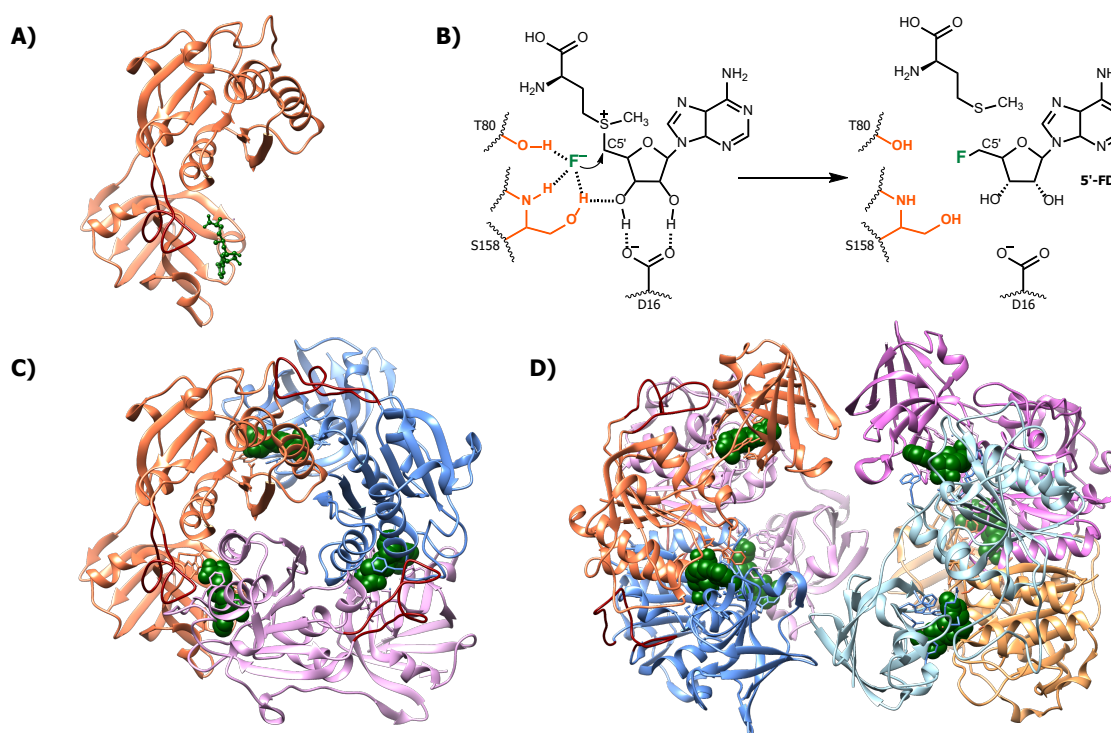
Figure 2 | The structure of the fluorinase enzyme. A) Fluorinase monomer is shown in orange with a characteristic loop colored in red, and SAM ligand is shown as a green ball and stick model. B) Active site constitution and mechanism of action (S_N2 substitution) of the fluorinase enzyme. Functional groups of key amino acids (S158, T80) that interact with fluoride ions (green) are indicated in orange. C) Fluorinase trimer with SAM (green spheres) bound at the subunit interfaces. D) Overall hexameric structure of fluorinase which folds into a dimer of trimers (with bound SAM, green spheres). The 3D structures of the fluorinase (PDB ID: 1RQP [41]) were drawn by UCSF Chimera.[43]

Fluorinase is encoded by the 897 bp *flA* gene that translates into the 299 amino acids long protein monomer.[41] Identifying its genetic basis gave rise to seeking other apparent fluorinase homologs in genome databases. Using genome mining and the characteristic 21-residue loop as diagnostic signature for fluorinases, homologs sharing 79 – 87 % identity have been identified in other bacterial species namely from two terrestrial actinomycetes (*Streptomyces sp.* MA37 and *Actinoplanes sp.* N902-109),[44] the actinomycete pathogen *Nocardia brasiliensis*,[44] from the marine bacterium *Streptomyces xinghaiensis* NRRL B-24674[45] and more recently from halophilic actinomycete *Actinopolyspora mzabensis*.[46] In addition, a putative chlorinase from archaea *Methanosaeta sp.* PtaU1.Bin055 was found to be a nonconventional fluorinase, which lacks the characteristic extra motif found in "conventional" fluorinase enzymes,[47,48] and the fluorinase from *Amycolatopsis sp.* CA-128772 was subjected to site-directed mutagenesis to improve its catalytic activity.[48] Interestingly, the search for fluorinase enzymes produced no matches among plant genomes, although they are known to accumulate fluorometabolites.[39] Just like *S. calvus*, plants have apparently adopted an alternative catalytic strategy for bio-fluorination that is still to be uncovered.[24]

The discovery and investigation of the unusual fluorinase enzyme have revolutionized the field of fluorination and prompted biotechnological prospects for fluorochemical production.[49,50] In terms of chemistry, it is a challenge to replace a C-H bond with a polar-reversed C-F bond, and conventional chemical synthesis methods require complex synthesis routes under often harsh conditions (high temperature and pressure) using precious metals, toxic and contaminating chemical reagents which is also an environmental concern. Therefore, fluorination catalysts bear an enormous biotechnological potential. In an early attempt, the fluorinase gene *flA* was cloned into *Salinospora tropica*, thereby replacing the chlorinase encoding gene *salL* in the biosynthetic gene cluster of salinosporamide A (anticancer metabolite), which led to the production of fluorosalinosporamide.[51] For such heterologous expressions, well-conceived metabolic engineering of the host organism is required, i.e., for instance the consideration of endogenous degradation enzymes or efflux pumps as recently has been reported for direct *in vivo* fluorination using *E. coli*.[52] Other examples of *in vivo* biofluorination include the implementation of synthetic gene circuits for the biosynthesis of fluoronucleotides and fluorosugars in engineered *Pseudomonas putida*,[53] the incorporation of fluoroacetate using engineered polyketide synthase systems,[54] *in vivo* production of fluorinated biopolymers,[55] and fluorinase-mediated ^{18}F labeling of PET (positron emission tomography) tracers for imaging.[56,57] Furthermore, fluorinase was subjected to directed evolution for improved fluorination efficiency for conversion of 5′-chloro-5′-deoxyadenosine (5′-ClDA) into 5′-FDA,[58] and coupled chlorinase-fluorinase system for trans-halogenation of 5′-ClDA to 5′-FDA was developed that profits from the chlorinase-mediated rapid conversion of 5'-ClDA to SAM which is the rate-limiting step for fluorination.[59]

1.1.3   A life based on fluorine

The high abundance of fluorine-containing minerals contrasts with the relative scarcity of organofluorines in nature. That nature has hardly evolved a biochemistry around fluorine, and the accompanying rare occurrence of fluorinated natural products is still not fully understood, but there are several implications.[8,35] For example, most of the fluorine is trapped in biologically unavailable forms such as minerals (e.g., fluorspar, fluorapatite, and cryolite); thus, fluoride ions have a comparatively low abundance in oceans (**Table 1**). Furthermore, fluorine's unique physical and chemical properties contribute to its inaccessibility to biological systems. In particular, the high heat of hydration ($\sim$120 kcal mol^{-1}) of the fluoride ion is a decisive criterion since the resulting easy solvation leads to an extremely weak nucleophilicity and thus considerably restricts the broad field of displacement reactions. However, synthetic chemistry has developed an effective mode of activation using urea-based catalysts that make fluoride soluble through hydrogen bonding (termed hydrogen bonding phase-transfer catalysis, HB PTC).[60]

The participation of halogens in biotransformations is usually mediated by enzymatic action of haloperoxidases,[61,62] which are even harnessed to some extent as catalysts in synthetic chemistry.[63] They catalyze halogenation by facilitating the oxidation of halide ions (Hal$-$) to halonium ions (Hal+), which replace a proton in hydrocarbon substrate under C-Hal bond formation (Hal$^-$ + H$_2$O$_2$ + H$^+$ + R-H $\rightarrow$ R-Hal + 2H$_2$O). However, fluorine is an exception in the set of halogens. The stabilization by hydration and the high electronegativity of fluorine result in a very high redox potential, hampering its oxidation. This, in turn, precludes fluorine from haloperoxidase-catalyzed halogenation because fluoride (E^0 = 2.866 V) cannot be oxidized by hydrogen peroxide (E^0 = 1,776 V[64]). In contrast, the redox potentials of Cl, Br, and I are below that of H$_2$O$_2$, which thermodynamically favors their oxidation to halonium ions according to the Nernst equation. This also explains the presence of about 182 natural products containing iodine, although their abundance in water is even lower than that of fluorine (**Table 1**).[35,65]

Besides haloperoxidases, another type of halogenating enzyme is given by halogenases requiring dioxygen (O$_2$) as a reducible co-substrate.[66,67] The first discovered representative of this enzyme class, the tryptophan 7-halogenase (PrnA) from *Pseudomonas fluorescens*, even permits the regioselective chlorination of tryptophan at the electronically unfavored C7 position of the indole ring to yield 7-chlorotryptophan during the biosynthesis of pyrrolnitrin.[68] Indeed, efficient biocatalytic halogenation on a preparative scale was achieved by using immobilized halogenase with auxiliary enzymes.[69] Another example is given by non-ribosomal peptides that commonly contain groups such as (bis)chlorotyrosine (in the glycopeptide vancomycin) because halogenases often appear within the non-ribosomal peptide synthetase machinery.[70]

All this together could be the reason why organisms have never had a real chance to adapt to this extraordinary element. In contrast, the relative natural abundance of other biogenic organohalides,

especially organochlorides, is notably higher because they were sufficiently present in the prebiotic environment and, therefore, microbes could evolve to metabolize them.[71–73] The latest comprehensive compilation (from 2023) of naturally occurring halogenated products counts more than 8000 compounds.[74] However, especially the marine habitat affords a huge pool of newly discovered organohalogens each year, reported in the annual reviews of marine natural products (whereby only the most recent are cited at this point).[75–78]

Table 1 | Distribution of halides in the environment. The abundance of halides F-, Br-, Cl-, and I- in oceans and sedimentary rocks, their estimated number of naturally occurring organohalogen compounds and selected properties are shown. The redox potential E^0 is given for the halogen reactions with $Hal_2 + 2e^- \rightleftharpoons 2Hal^-$ and values were taken from the electrochemical series of the CRC Handbook of Chemistry and Physics.[64]

halide	no. of natural occurring organohalides	oceans [mg/dm³][74]	sedimentary rocks [mg/kg][74]	redox potential E^0 [V][64]
Fluoride, F-	5 [a]	1.4	270 – 740	+ 2.866
Chloride, Cl-	2200 [b]	19000	10 – 320	+ 1.358
Bromide, Br-	1900 [b]	65	1.6 – 3.0	+ 1.066
Iodide, I-	182 [c]	0.05	0.3	+ 0.536

[a] organofluorines: effective 04/2014;[35]   [b] organochlorines and organobromines: effective 06/2002;[71]
[c] organoiodines: effective 09/2014[65]

Even if largely ignored by nature, a life based on fluorine is both an interesting concept and an absolutely conceivable scenario.[79] Fluorine-containing building blocks have been intensively used for a long time to investigate and modify proteins and their interactions.[80–82] The exchange of single hydrogen for fluorine does not cause great steric disturbances, but the electrostatic properties offer a large and versatile repertoire of other effects, the severity of which rises with increasing degree of fluorination.[83] In general, the fluorine decoration of amino acids affects essential properties such as hydrophobicity, acidity/basicity, and conformation of the specific side chains, which in turn is accompanied by altered properties on stability, folding kinetics, and secondary structure propensities as well as activity of peptides and proteins. However, the impact of fluorine is highly context-dependent and thus mostly unpredictable.[84] Extensively fluorinated molecules exhibit a phenomenon called fluorous effect. This is an unusual phase segregation, where in the presence of highly fluorinated compounds, triphasic mixtures are formed: lipophilic – hydrophilic – fluorous.[85,86] The fluorous effect has been exploited in organic synthesis to extract fluorocarbon molecules and, moreover, attracted attention in protein design to create "Teflon"-like structures.[87–89] However, the fluorous effect is restricted to aliphatic molecules and does not apply to fluorinated aromatics.[90]

Furthermore, beyond the incorporation into distinct proteins, even the proteome-wide incorporation of fluorinated amino acids has been demonstrated, as discussed in section **1.3.3**.

1.2 Tryptophan

1.2.1 Tryptophan: special among the 20 canonical amino acids

Tryptophan (Trp, W) occupies an exceptional status among the 20 canonical amino acids. Its salient structure, comprising a lateral indole moiety composed of a six-membered benzene ring fused to a five-membered pyrrole ring with an integrated NH group (**Figure 3**), endows Trp with unique physicochemical properties.

The heterocyclic indole ring makes Trp the largest and bulkiest proteinogenic amino acid (204.22 g/mol, molecular formula: $C_{11}H_{12}N_2O_2$), and the multifunctional chemical nature of the planar aromatic system enables numerous supramolecular interactions.[91] As aromatic amino acid Trp is considered hydrophobic, even though its amphipathic character complicates unambiguous grading on a hydrophobicity scale.[92,93] The indole ring affords a large electron-rich surface for Van der Waals interactions, and the π-conjugated system engages electrostatic interactions such as π-π-stacking, π-cation interactions, or π-CH interactions.[94-97] Furthermore, the nitrogen electron lone pair of indole's pyrrole subunit contributes to the maintenance of the aromatic system and can function a as hydrogen-bond donor.

Tryptophan | Trp | W
Molecular formula of Tryptophan: $C_{11}H_{12}N_2O_2$
Molar mass: 204.23 g/mol
CAS 77-22-3
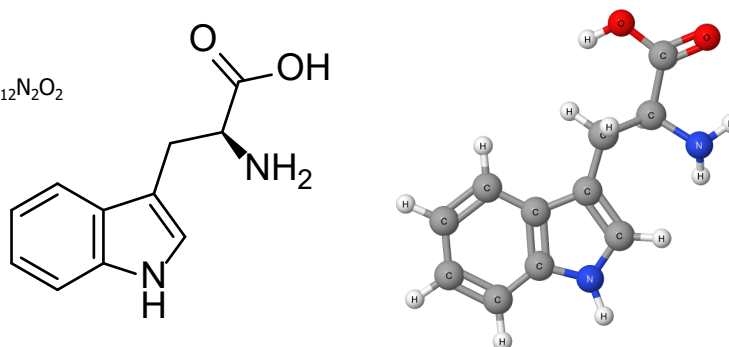
Figure 3 | Structure of *L*-tryptophan. The skeletal formula (left) and ball-and-stick model (right) are shown. The systematic IUPAC name of L-tryptophan is (2*S*)-2-amino-3-(1*H*-indol-3-yl)propanoic acid. The 3D structure was drawn with the JSME Molecular Editor.[98]

The complexity of building the indole side chain is reflected by the high energetic and metabolic costs[99] (e.g., the demand for precursor molecules), but on the other side it is justified by the above-described chemical options offered by this moiety. Apart from its role as a fundamental building block in ribosomal protein biosynthesis, Trp contributes with its manifold properties to protein stability,[100] membrane anchoring,[93] enzyme-substrate binding, receptor–ligand interactions, antigen–antibody recognition,[101] and nucleic acid binding,[102,103] although it infrequently occurs in DNA-binding sites.[104] In addition, it is a major source of UV absorption (λ_{max} = 279 nm) and fluorescence (λ_{em} = 365 nm) in

proteins, and its spectral properties (although highly dependent on the surrounding environment) have been harnessed to study protein dynamics, folding, and ligand binding.[105]

Even in the evolutionary context, tryptophan has unique features. Together with methionine, it is thought to be the latest recruitment to the genetic code,[106,107] and both are encoded by a single nonredundant codon (Trp codon: UGG), breaking the code's degeneracy. Consequently, single nucleotide substitutions of the UGG codon lead in any case to (most likely deleterious) replacement of the amino acid in either the following other amino acids Gly (GGG), Ser (UCG), Cys (UGC, UGU), Leu (UUG), Arg (AGG, CGG) or result even in stop codons (UGA, UAG),[108] which terminate the translation process. Nevertheless, Trp is an attractive target for chemical protein modification by either targeting the natural residue[109] or introducing non-canonical amino acids (ncAAs) by means of genetic code expansion (see section **1.3.1**). In fact, the number of encoding triplets is directly proportional to the abundance of the amino acid in the proteome, having tryptophan with about 20000 UGG codons and about 1 % occurrence in *E. coli*'s proteome the rarest amino acid.[110] Thus, only a limited number of proteins are expected to be affected, and only one sense codon needs to be reassigned.

The side chain moiety of Trp, the indole molecule itself, acts as a versatile signaling molecule for inter-cellular, intracellular, as well as interspecies, and interkingdom communication.[111] As ubiquitous signaling molecule, indole influences myriad aspects of bacterial physiology, including chemotaxis, motility, cell division, the formation of biofilms and bacterial persisters, virulence, plasmid stability, as well as tolerance and resistance to diverse stresses such as antibiotics, acids, and heat.[112,113] However, whether the respective effects are stimulated or suppressed by indole, its mode of action (persistent vs. pulse signaling), and the fact that there is no defined target are controversially discussed.[114] On the one hand, indole signaling occurs at relatively low (< 1 mM) persistent concentrations, but on the other hand, its concentration dramatically increases (approx. 60 mM) during the transition from exponential to stationary growth phase.[115] This phenomenon of rapid indole synthesis is referred to as "indole pulse" and is consequently only experienced by indole-producing species. Indole pulse signaling is associated with long-term survival. Before nutrients are entirely exhausted, glucose depletion induces transiently, highly increased indole production by tryptophanase (TnaA), which in turn inhibits growth and cell division and, therefore, causes the culture to enter the stationary phase earlier at lower density, thereby conserving resources and keeping the population more viable.[116] Inhibition of growth and cell division by indole at higher concentrations (3–5 mM) is attributed to its proton ionophore activity. As such, it allows protons to pass through the cytoplasmic membrane, which reduces its electrochemical potential and prevents the formation of the divisome protein complex responsible for cell division.[117] In its capacity as an ionophore, indole is also involved in regulating the cytoplasmic pH. It was found that the pH is set during the indole pulse and maintained beyond the stationary phase into the next growth cycle until the next pulse (transition from exponential to stationary phase); the authors described this as memory-like manner and referred to as pH cycle.[118] Decoration of indole with halogens (e.g., F) increases the protonophore effect[119] and they can eradicate bacterial persister cells and biofilms.[120]

1.2.2   Tryptophan biosynthesis and metabolism in *E. coli*

Just like the amino acid itself, the tryptophan biosynthesis and, in particular, its regulation is extraordinary. *Escherichia coli* is equipped with seven Trp biosynthetic genes organized within a single transcriptional unit known as Trp operon (**Figure 4A**).[121] It contains five structural genes *trpE*, *trpD*, *trpC*, *trpB*, *trpA* coding the enzymes of the biosynthetic pathway and an upstream regulatory region composed of a promotor/operator at which the transcription is regulated and the TrpL leader peptide (*trpL*), which is a key component of the attenuation system that modulates expression of the Trp operon in response to Trp availability.

1.2.2.1   Regulation of the tryptophan biosynthesis

The tryptophan biosynthesis is an energetically and biologically expensive process; after all, it requires a substantial number of enzymes and four products from other biosynthetic pathways: L-glutamine, 5-phosporibosyl-1-pyrophosphate (PRPP), L-serine and pyridoxal phosphate (PLP). Therefore, it is not surprising that this metabolic pathway is tightly regulated at multiple levels to ensure only Trp production in demand by sensing both the direct intracellular concentration of L-Trp and indirect by the extent of charged tryptophanyl-tRNA^{Trp}.

In its priority regulation mode, the operon is governed by a negative feedback mechanism (**Figure 4A**). Transcription initiation at the Trp promotor is regulated by the tryptophan-activated repressor protein TrpR (DNA-binding transcriptional repressor TrpR, encoded by *trpR*) in response to intracellular levels of tryptophan. In situations of Trp excess, the complexed TrpR-L-Trp changes conformation (allosteric mechanism), which allows binding to the operator, thereby preventing RNA-polymerase (RNAP) binding to the promotor and inhibiting transcription initiation of *trpLEDCBA* genes.[122–124] After transcription initiation, a second regulatory mechanism called "transcription attenuation" also senses the availability of charged versus uncharged tRNA^{Trp} (**Figure 4B-C**).[125,126] Upstream of the structural genes the 14-residue leader peptide TrpL, which contains two adjacent tryptophans, is located. The leader transcript can form different RNA hairpin structures that guide the termination or continuation of the transcription process. The ribosome can only translate the leader segment of the mRNA if tryptophan is present in sufficient quantity (charged tRNA^{Trp}). Formation of the 1:2 hairpin induces transcription pausing. If the Trp level is deficient (uncharged tRNA^{Trp}), the ribosome stalls at a Trp codon, the anti-terminator structure is formed (2:3 hairpin), preventing terminator formation, and transcription continues into the pathway Trp genes (as polycistronic mRNA). Conversely, when there is ample intracellular Trp and Trp-tRNA^{Trp}, the ribosome does not pause at the tandem Trp codons, the terminator structure (3:4 hairpin) is formed which causes the RNAP to be released and thus prematurely terminates the transcription process.[127]

Besides these prevailing mechanisms of repression and attenuation, the Trp synthesis is additionally regulated by L-Trp mediated feedback inhibition of the anthranilate synthase (TrpED complex).[128] Moreover, the Trp biosynthesis is always highly regulated, also among other bacteria. Independent of their genetic arrangement, microorganisms evolved a variety of mechanisms to economize this expensive synthesis pathway.[129]
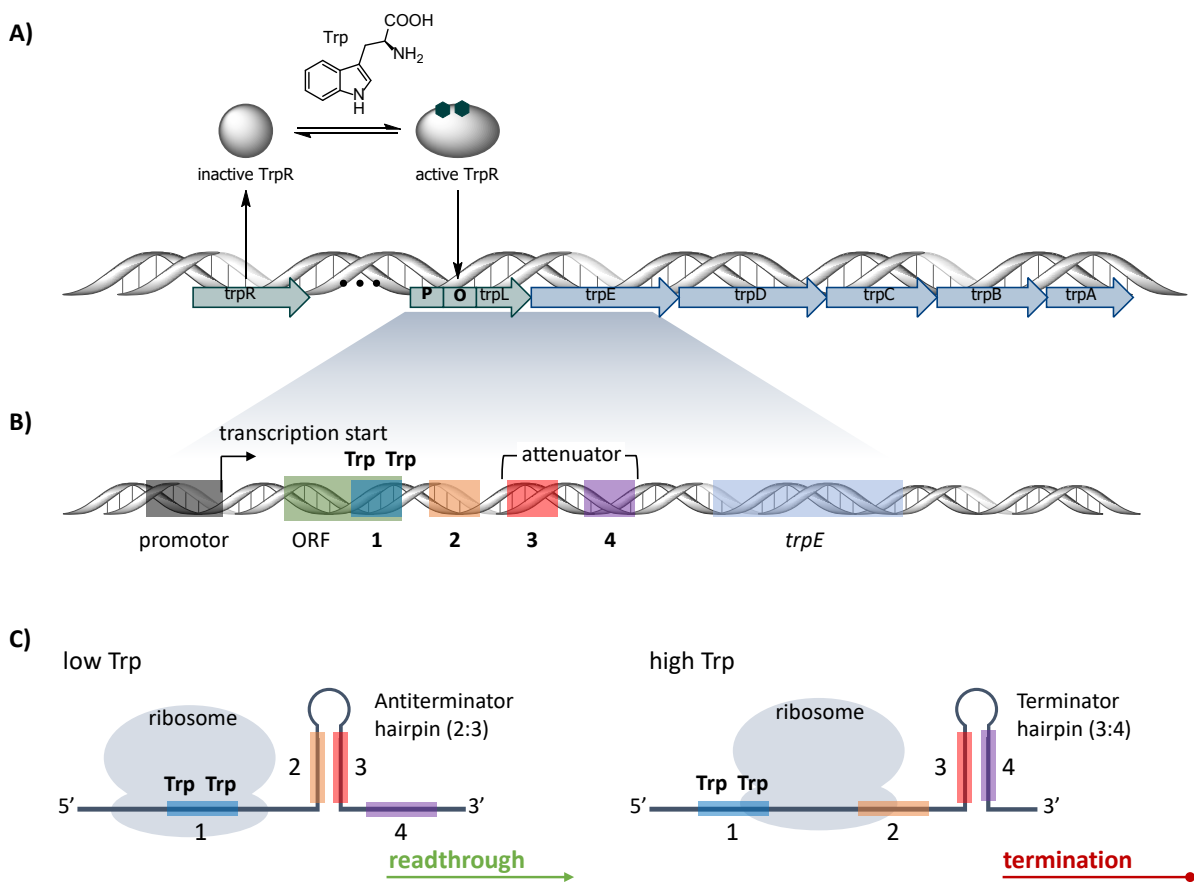
Figure 4 | Regulation of the Trp Biosynthesis. A) Illustration of the Trp operon *trpLEDCBA* and trpR-mediated repressible feedback mechanism. B) Zoom into the leader region containing key regulatory elements required for attenuation control, which are four segments (1, 2, 3, and 4) capable of specifying hairpins 1:2, 2:3, and 3:4 in the leader transcript, a 14-residue ORF (open reading frame) with two adjacent Trp codons (W10, W11) and the Trp attenuator region. C) Transcription attenuation control of Trp operon. At low Trp (and Trp-tRNA^{Trp}) concentrations the antiterminator hairpin (2:3) forms, and readthrough transcription continues downstream. Conversely, at high Trp levels, the terminator hairpin (3:4) forms, and transcription is terminated before the Trp-biosynthetic enzymes are synthesized.

1.2.2.2 Tryptophan biosynthesis

Tryptophan is synthetized in a sequence of five metabolic steps (**Figure 5**);[130,131] beginning with the shikimate pathway in which D-erythrose 4-phosphate and phosphoenolpyruvate are converted into chorismate, which is the common precursor of aromatic amino acids (Trp, Tyr, and Phe).[130,132]

At first, chorismate is converted into anthranilate. The glutamine amidotransferase reaction adds an amine group from glutamine to chorismate to yield anthranilate and glutamate and is catalyzed by anthranilate synthase (TrpDE), an $\alpha_2\beta_2$ heterotetramer composed of anthranilate subunit component I (*trpE*) and component II (*trpD*). Beyond the TrpDE complex, component I (TrpE) can form anthranilate using ammonia as a nitrogen source, but with considerably reduced efficiency. In addition, component II (TrpD) has a second activity termed anthranilate phosphoribosyl transferase. 5-phosporibosyl-1-pyrophosphate (PRPP) provides two carbon atoms for the pyrrole ring of indole, converting anthranilate into *N*-(5'-phosphoribosyl)-anthranilate (PRA), which is the second step of Trp biosynthesis. Then, the fused indole-3-glycerol phosphate synthase/phosphoribosylanthranilate isomerase (*trpC*) executes the third and fourth steps of the pathway. The bifunctional TrpC enzyme has for one phosphoribosylanthranilate isomerase activity and catalyzes the rearrangement into 1-(2-carboxyphenylamino)-1-deoxy-D-ribulose 5-phosphate (CDRP); and secondly the indole-glycerol phosphate synthase activity of TrpC catalyzes the ring closure of CDRP to yield indole-3-glycerol phosphate (InGP). After the indole moiety is formed, in the fifth and last step of the pathway, the Trp synthase (TrpS encoded by *trpBA*) catalyzes the formation of L-tryptophan through a condensation reaction (detailed description in section **1.2.3**).

Besides utilization as a building block for protein biosynthesis (see section **1.2.4**), tryptophan can degrade back to indole (**Figure 5**), where it serves as an important signaling molecule, as discussed above. The degradation is catalyzed by the PLP-dependent tryptophanase (TnaA encoded by *tnaA*).[126,133] This enzyme is part of the Tna operon (*tnaCAB*), which is also regulated by a transcriptional attenuation mechanism in its leader region, TnaC, in response to tryptophan.[126,134,135] TnaB (tryptophan:H+ symporter TnaB encoded by *tnaB*) is a low-affinity tryptophan transporter; it is one of three permeases (Mtr, AroP, and TnaB) that can transport the amino acid tryptophan.[136] In the end, Trp catabolism provides pyruvate processed in the TCA cycle and ammonia used as a precursor in amino acid biosynthesis; however, indole is a dead-end product of Trp since no further degradation process in *E. coli* is known.[137]

The described tryptophan biosynthesis pathway is highly conserved across kingdoms (procaryotic and eucaryotic microorganisms as well as plants), i.e., identical conversion of substrates by homologous enzymes. However, the catalyzing enzymes in fungi and plants are monofunctional, while in bacteria they can exhibit two to three catalytic domains.[138–140] In contrast, for humans and many animals, tryptophan is an essential amino acid; that is, they cannot synthesize tryptophan and must ingest it through their diet.
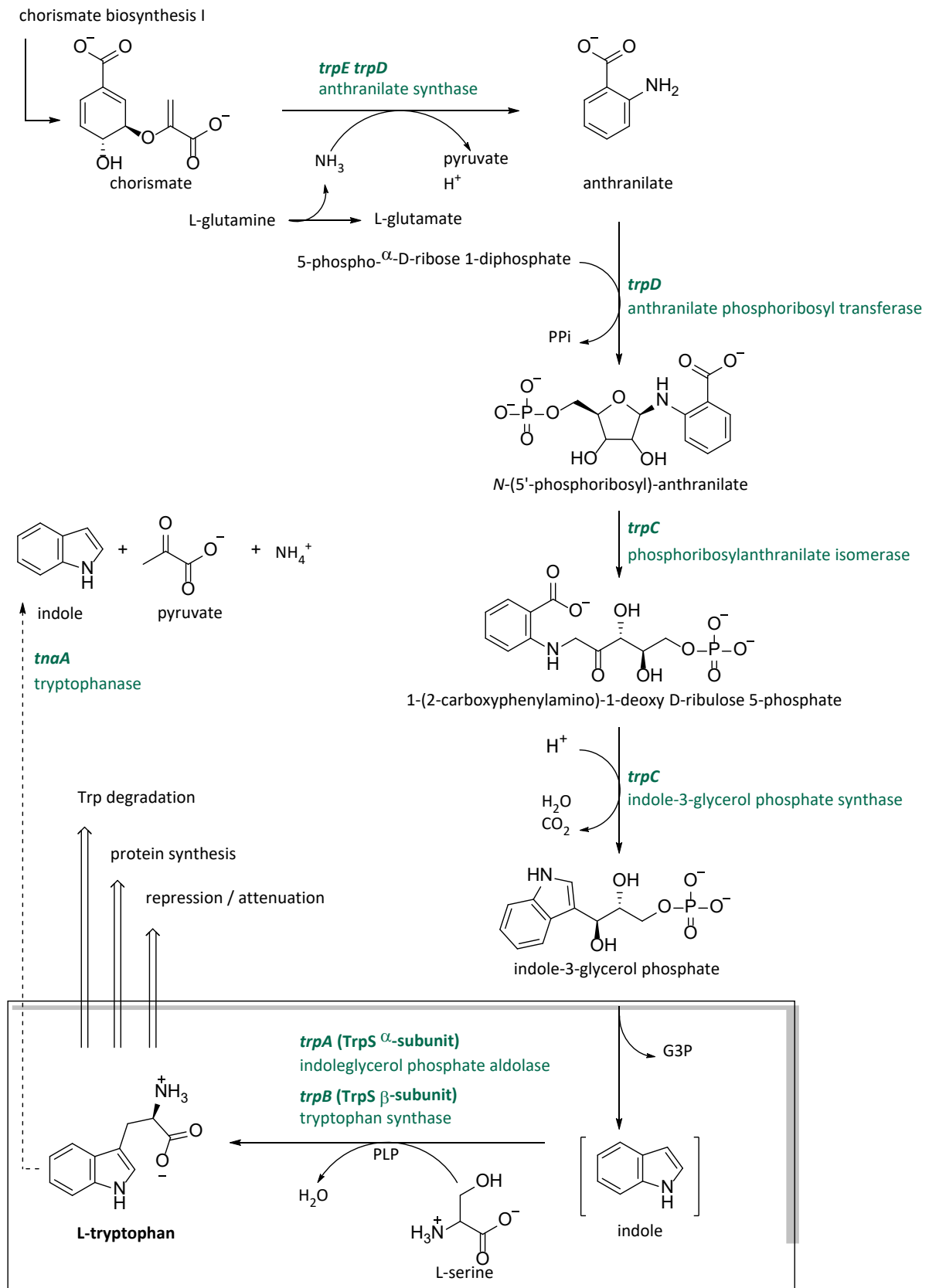
Figure 5 | Tryptophan biosynthesis in *E. coli*. Starting from chorismate, Trp is synthesized in five steps catalyzed by the enzymes (teal) encoded by structural genes of the Trp operon *trpEDCBA*. The amino acid then can undergo three scenarios: tryptophanase (TnaA) catalyzed degradation back to indole, utilization as building for protein biosynthesis, or it induces regulatory mechanisms to transiently stop the synthesis until further demand. The reaction scheme was adapted from the EcoCyc Database.[131]

1.2.3 The Tryptophan Synthase

The tryptophan synthase (EC 4.2.1.20) is a crucial enzyme of the Trp biosynthetic pathway. It catalyzes the terminal steps of the biosynthesis described above and operates in a fairly interesting mechanism of action.

TrpS is a bifunctional heterodimeric complex αββα composed of two α-subunits (TrpA, encoded by *trpA*) and two β-subunits (TrpB, encoded by *trpB*), which exhibit also two distinct catalytic activities.[141] The TrpA polypeptide catalyzes indole formation in a retro-aldol cleavage of indole-3-glycerol phosphate (IGP), termed as α-reaction. Indole then passes through a 25 Å-long tunnel to the β-subunit, where it is condensed with L-serine to form L-tryptophan in the PLP-mediated catalytic cycle, termed as β-reaction (**Figure 6A**).[142–145] The β-reaction can be divided into two stages, where in stage I serine forms with enzyme-bound cofactor PLP (pyridoxal 5'-phosphate) the aminoacrylate intermediate E(A-A), which then in stage II reacts with indole to build tryptophan (**Figure 6C**). The α- and β-subunits form a multienzyme complex with a near linear arrangement, impressively illustrated by Hide *et. al.*[146], that are not only spatial and structural connected by a 25 Å-long intramolecular tunnel[146,147] (**Figure 6B**) but also by allosteric communication.

The sophisticated allosteric regulation is hallmarked by synchronization of the α- and β-reaction, thereby enhancing the effectiveness of the enzyme.[148–151] Briefly, binding of IGP in TrpA initiates a conformational change that activates TrpB and that both increases the affinity of serine for TrpB (lowers Km) and promotes aminoacrylate E(A-A) formation. Conversely, TrpB stimulates the TrpA subunit to induce IGP retro-aldol cleavage by causing the COMM domain (residues $\beta102-\beta189$) to adopt a closed state, thus stabilizing TrpA, leading to a higher K_{cat} of the α-reaction and preventing indole from escaping the enzyme. Instead, it passes through the tunnel to the active site of TrpB, where it attacks the aminoacrylate E(A-A) to form quinonoid intermediates Q_2, then Q_3, and finally the Trp-bound external aldimine, $E(Aex_2)$. The processes occurring between the subunits require the action of monovalent cations (MVC), which are also crucial for regulating β-site catalytic activity.[152] Recently, it was shown that incorporation of the photocaged ncAA *o*-nitrobenzyl-*O*-tyrosine (ONBY) allosteric light-activation of the tryptophan synthase is possible.[153]
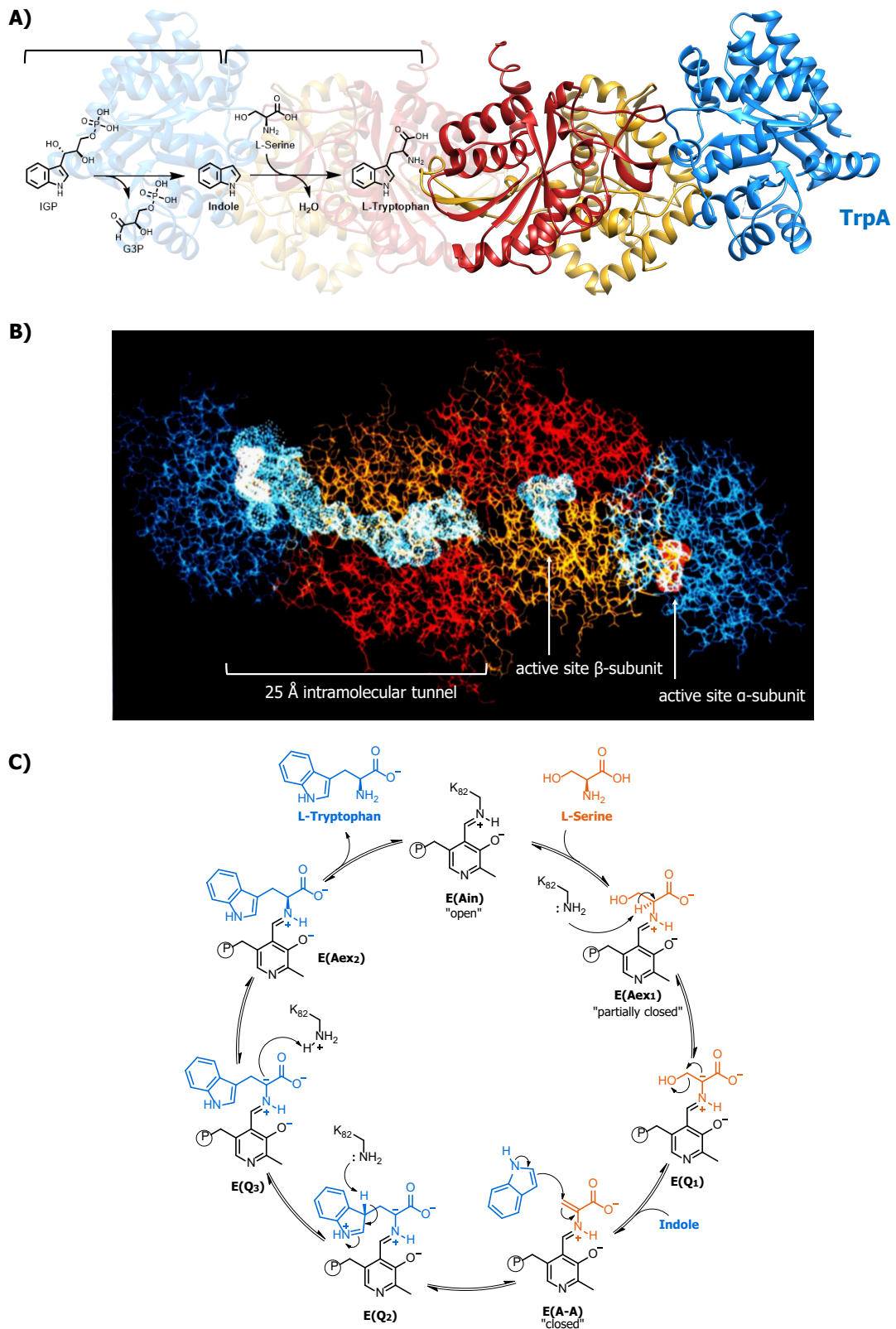
Figure 6| Tryptophan synthase (structure and catalyzed reaction/Trp synthesis). A) Structure of TrpS (αββα complex) in ribbon representation with catalyzed reaction of Trp synthesis. The α-subunit (blue) folds in a TIM barrel conformation, and the β-subunit is composed of two domains (orange and red), with the PLP coenzyme sandwiched at their interface. The four subunits have a nearly linear arrangement, with the two β-subunits at the center flanked by the α-subunits. In the α-reaction indole-3-glycerol phosphate (IGP) is cleaved into glyceraldehyde-3-phosphate (G3P) and indole, catalyzed by TrpA, and in the β-reaction L-tryptophan is formed by condensation of indole with L-serine, catalyzed by TrpB. (The 3D structure of TrpS, PDP ID 1BKS[154] was drawn

by UCSF Chimera.[43]) B) Tryptophan synthase structure in (wire representation) with highlighted intramolecular tunnel (left half) and active sites (right half). The active site of each TrpA is located near the interface with TrpB, whereas the active site of each TrpB is deeply buried near its center. The active centers of each enzyme half are connected by an intramolecular tunnel spanning a distance of 25 Å. (The figure was taken from Hide et.al.[146]). C) Catalytic cycle of the β-reaction. The coenzyme PLP is covalently bound to the ε-nitrogen atom of lysine residue K82 via a protonated Schiff base linkage, termed internal aldimine E(Ain). Transamination with L-serine leads to the external aldimine intermediate E(Aex1), which, after C-α deprotonation and formation of a carbanion, is converted to a quinonoid intermediate E(Q$_1$). Subsequent cleavage of the hydroxyl group from serine leads to the electrophilic aminoacrylate intermediate E(A-A); here, the COMM domain adopts a "fully closed" conformation stabilized by TrpA. Then, indole produced at TrpA and channeled to TrpB, executes irreversible nucleophilic attack on E(A-A) to form the second quinonoid intermediate E(Q$_2$), which deprotonates to the third quinonoid intermediate, E(Q$_3$). Finally, the C-α protonates by K82 to build Trp-bound external aldimine E(Aex$_2$) that releases L-tryptophan by transamination, and the enzyme returns into the E(Ain) resting state.

Owing to the described allosteric crosstalk, the αββα multienzyme complex has significantly higher catalytic activity than the separate subunits,[155] however, in principle, the α-reaction can be completely bypassed for ncAAs synthesis when indole analogs are provided. The group of Frances Arnold has done seminal work in engineering the catalytically relevant TrpB subunit to create a stand-alone enzyme and to further expand its substrate scope.[156] Using directed evolution, the TrpS β-subunit from *Pyrococcus furiosus* (*Pf*TrpB) was freed from allosteric regulation by its TrpA partner. The resulting variant *Pf*TrpB^{0B2} has an 83-fold increased catalytic efficiency compared to native *Pf*Trp^{WT} and needs only six mutations that reproduce the effects otherwise induced by the binding of TrpA.[155,157] The *Pf*TrpB^{0B2} and another engineered variant, *Tm*TrpB^{2F3} (derived from *Thermotoga maritima*),[158] have demonstrated an enormous substrate scope in using indoles substituted with diverse halogen-, nitro-, cyano-, carboxamide-, boronate-, and trifluoromethyl groups.[159,160] Furthermore, the use of either threonine or β-branched serine analogs instead of serine enabled the synthesis of β-branched Trp analogs such as β-MeTrp[161,162] or β-alkyl Trp analogs (β-EtTrp, β-PrTrp)[163] and the use of 3-substituted oxoindoles resulted in Trp analogs with a new stereocenter at the γ-position.[164]

The remarkable substrate range, however, is by no means a peculiarity of the engineered TrpB subunit but a general feature of the TrpS enzyme.[156,165] Its potential as a biocatalyst was recognized very early.[166-168] Since then its (outstanding) promiscuity towards indole analogs is cherished and led to enzymatic syntheses of azido-substituted tryptophans (4-, 5-, 6-, 7-N$_3$-Trp),[169] AzaTrps,[170] methyl- (2-, 4-, 5-, 6-, 7-CH$_3$-Trp),[171] amino- (4-, 6-, 7-NH$_2$-Trp),[172] and nitroTrp (7-NO$_2$-Trp),[173] the haloTrps (4-, 5-, 6-, 7-F; 4-, 5-, 6-, 7-Cl; 5-, 6-, 7-Br; 7-I-Trp)[171-175] and to the sulfur, selenium, and oxygen-containing Trp isosteres thienopyrroles,[176] selenopyrroles,[177,178] furopyrrole.[177] Although substituted tryptophan analogs are also accessible by chemical synthesis,[179,180] they often require enzymatic reactions to resolve racemic product mixtures.[181] In particular, the synthesis of fluorinated Trp from fluoroindoles, 6FTrp and 7FTrp to be precise, is relevant for this study.[168,171]

1.2.4   The tryptophanyl-tRNA synthetase

Once synthesized, L-tryptophan can be subjected to different metabolic fates within the cell; it serves as a precursor for secondary metabolites, as a signaling molecule after degradation back to indole, and in particular, as a building block in protein biosynthesis.

In protein biosynthesis, the information stored as DNA is transcribed (mRNA) and then translated into amino acids, building proteins. For this interpretation of the genetic code, aminoacyl-tRNA synthetases (aaRS, or tRNA-ligase) are pivotal. They catalyze a two-step reaction where the amino acid (AA) is activated (1) and then loaded on the tRNA adapter molecule (2).[182–184]

$$AA + ATP + aaRS \leftrightarrows aaRS \cdot AA\text{-}AMP + PPi \tag{1}$$

$$aaRS \cdot AA\text{-}AMP + tRNA \rightleftarrows aaRS + AA\text{-}tRNA + AMP \tag{2}$$

Thus, the incorporation of Trp by means of ribosomal translation requires the ATP-dependent activation of the amino acid and its conjugation to the cognate tRNA^{Trp} (**Figure 7**); both processes are catalyzed by the tryptophanyl-tRNA synthetase (TrpRS encoded by *trpS*). In this aminoacylation reaction, first, the amino acid and ATP bind the catalytic site of aaRS where α-carboxylate oxygen of the amino acids attacks the α-phosphate group of ATP, which produces aminoacyl-adenylate (AA-AMP). Then, this intermediate is transferred or loaded onto the corresponding tRNA by transesterification, forming aminoacylated tRNA (AA-tRNA). Thereby, the carboxyl group of the activated amino acid is attached to a 2'-OH or 3'-OH of the adenosine residue located at the end of the 3'-CCA sequence of the tRNA acceptor stem. Whether the 2'- or 3'-hydroxyl group is targeted depends on the aaRS class. Based on the active site structure, two classes are distinguished: class I aaRS (Arg, Cys, Gln, Glu, Ile, Leu, Met, Trp, Tyr, Val) aminoacylate at 2'-OH and class II aaRS (Ala, Asn, Asp, Gly, His, Lys, Pro, Phe, Ser, Thr) at 3'-OH of the adenosine ribose. However, the aaRS of the aromatic amino acids Trp, Tyr, and Phe do not follow this pattern; class I TrpRS and TyrRS are dimers and recognize tRNA as do class II aaRSs, and the class II PheRS, in turn, acts like class I enzyme.

As mentioned, the tryptophanyl-tRNA synthetase belongs to class Ic aminoacyl-tRNA synthetases with the ATP binding domain forming a Rossmann dinucleotide-binding fold. TrpRS possesses no editing mechanisms, neither during aminoacyl-AMP formation (pre-transfer editing) nor after attachment of the aminoacyl group to tRNA (post-transfer editing).[183] The crystal structure of the tryptophanyl-RNA synthetase from *Escherichia coli* (*ec*TrpRS) was reported in 2017 by Maltseva *et. al.* on the Protein Data Bank[185] with PDB ID 5V0I, however an attendant publication is still to be published (https://doi.org/10.2210/pdb5V0I/pdb).[186]
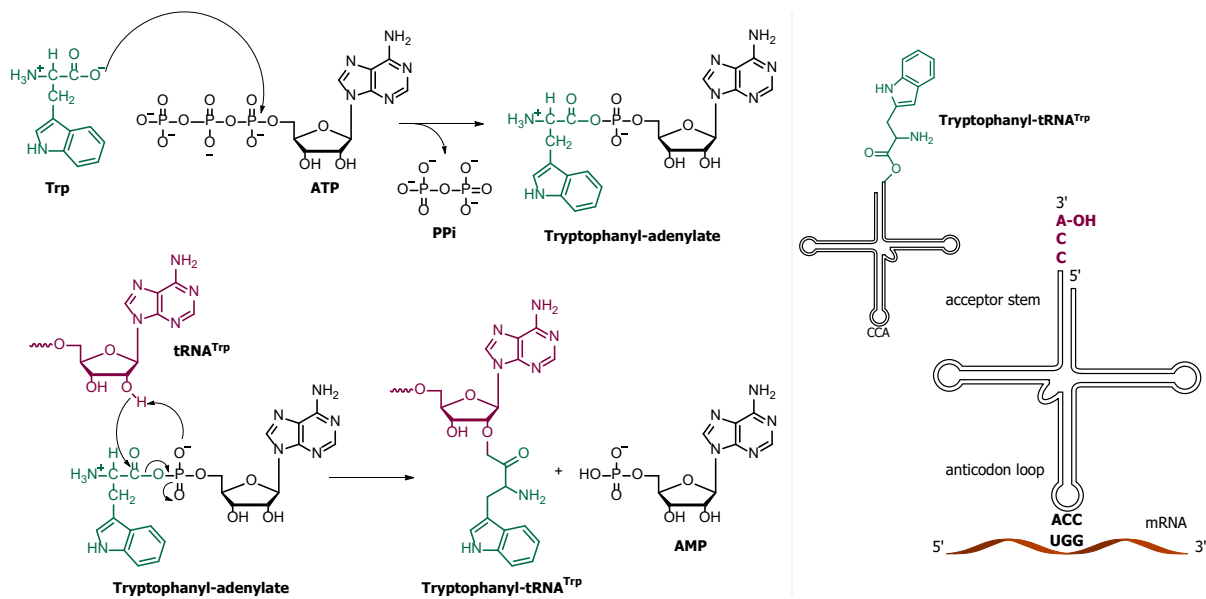
Figure 7 | TrpRS-catalyzed aminoacylation reaction. The amino acid Trp (teal) is activated with adenosine-5'-triphosphate (ATP), forming tryptophanyl-adenylate and releasing pyrophosphate (PPi). Then Trp is transferred to the tRNA (red) and adenosine monophosphate (AMP, 5'-adenylic acid) is released. In the right panel, the charged tRNA (tryptophanyl-tRNA^Trp) and the general constitution of a tRNA with acceptor stem, as well as Trp codon (UGG) and its anticodon (ACC) are shown.

The incorporation of ncAAs requires a certain tolerance in the editing range of the aaRS enzymes and the TrpRS exhibits such a substrate promiscuity. It aminoacylates also tRNA^Trp with D-tryptophan[187] and a wide range of other Trp analogs including, all mono-fluorinated tryptophans (4FTrp, 5FTrp, 6FTrp, and 7FTrp).[188–191] Although the TrpRS can distinguish between Trp and the other 19 canonical amino acids, unlike the IleRS, LeuRS, and ValRS, it does not have an editing domain that ensures high fidelity of tRNA charging.[183] Furthermore, the polyspecificity of the TrpRS might be influenced by the tRNA^Trp anticodon sequence. For example, Fan *et. al.* found efficient incorporation of 2-iodo-L-phenylalanine in response to $tRNA\frac{Trp}{CUA}$ but not $tRNA\frac{Trp}{CAA}$.[192]

1.3   Xenobiology

Life on earth is modularly organized using the core biopolymers nucleic acids (DNA, RNA) and proteins as well as carbohydrates. According to the central dogma of molecular biology, genetic information preserved and replicated as DNA is transcribed into mRNA, and this is used as a template to translate the information into functional proteins.[193] To implement and sustain this blueprint, all organisms, from unicellular to humans, share the same standardized alphabet. Informational polymers DNA/RNA are built from four nucleobases: adenine, guanine (A, G, purines) and cytosine, thymine (C, T, pyrimidines); in RNA instead of T uracil (U) is used. They form $4^3 = 64$ triplet combinations encoding 20 canonical amino acids that give rise to catalytic polymers (proteins). The described building blocks of life are mainly composed of only six chemical elements - carbon, hydrogen, nitrogen, oxygen, phosphorus, and sulfur - out of the 94 (or 84 primordial) naturally existing elements listed in the periodic table. However, nature also makes use of unusual elements, as in the case of fluorine, described in section **1.1** "Natural occurrence of fluorine in living organisms".

Remarkably, these few building blocks and a fairly narrow chemistry enabled the creation of an astonishing biodiversity of species in the history of life. However, considering the diversity of the chemical landscape,[194] the opportunities for alternative life forms appear almost unlimited. And this is where xenobiology comes into play because this scientific discipline aims to implement artificial chemistry by means of non-natural building blocks into living cells.[195-197] In re-design life, xenobiology tackles both the genomic level and the proteomic level.[198,199]

At the informational level, the genetic alphabet was expanded by so-called xeno-nucleic acids (XNAs), which are categorized into three groups: 1) XNAs with modified nucleobases, 2) XNAs with modified phosphodiester backbone, and 3) XNAs with modified sugar moieties.[200,201] They are synthesized either chemically or from a DNA template by use of engineered polymerases,[202,203] which demonstrates that genetic information can even be stored and processed by these artificial XNA polymers.[204-206] Moreover, only a few years after they first proposed the concept of genetic proliferation on XNAs,[207] Marlière *et. al.* even demonstrated the *in vivo* DNA base transliteration by means of almost complete replacement of thymidine with XNA 5-chlorodeoxyuridine using a chemostat system.[208] At the translational level, the genetic code is reprogrammed by incorporation of non-canonical amino acids (ncAAs) through either hijacking the endogenous translation machinery or by harnessing orthogonal translation systems (discussed below in section **1.3.1**).

The integration of foreign functional groups provides valuable tools for the study of protein function and structure. However, it also raised biosafety concerns[209,210] because the universality of the genetic code can enable genetically modified organisms (GMOs) to invade natural ecosystems through, for example, horizontal gene transfer. This scenario is controversially discussed mainly for two reasons: 1) their engineered functions often reduce cellular fitness, and the spread of transgenic material might be unfavorable and, 2) GMOs are becoming increasingly robust as, for instance, through applying ALE

techniques. Thus, the assumption that they might be able to outcompete natural organisms or even dominate an ecosystem is obvious. So far, negative effects on the environment have not been reported, but "the absence of evidence is not the evidence of absence" (precautionary principle or Black Swan argument).[211,212]

The concept of biocontainment is to prevent the unauthorized propagation of GMOs by controlling their proliferation and the release of genetic materials.[213] In doing so, biocontainment pursuits a myriad of different semantic and trophic strategies to constrain the interaction of GMOs with the external world, such as genetic circuit–actuated killing, metabolic and synthetic auxotrophy, horizontal gene-transfer blocking or integration of toxin/antitoxin "kill switches".[212,214,215] For example, trophic containment aims to implement synthetic auxotrophy, i.e., microorganisms that are dependent on xeno-nutrients like ncAAs. Therefore, survival of such GMOs is linked to the availability of synthetic molecules that can only be supplied under laboratory conditions but do not occur in natural ecosystems (**Figure 8D**).[216,217] Yoo *et. al.* used engineered fluoride sensitivity in yeasts, whose growth is massively inhibited at low fluoride concentrations.[218] The greater the biological barrier (i.e., GMOs differing from their wild relatives; genetic firewall), the more likely it is that their desperate will to survive and the intrinsic urge to adapt can be suppressed.[219]
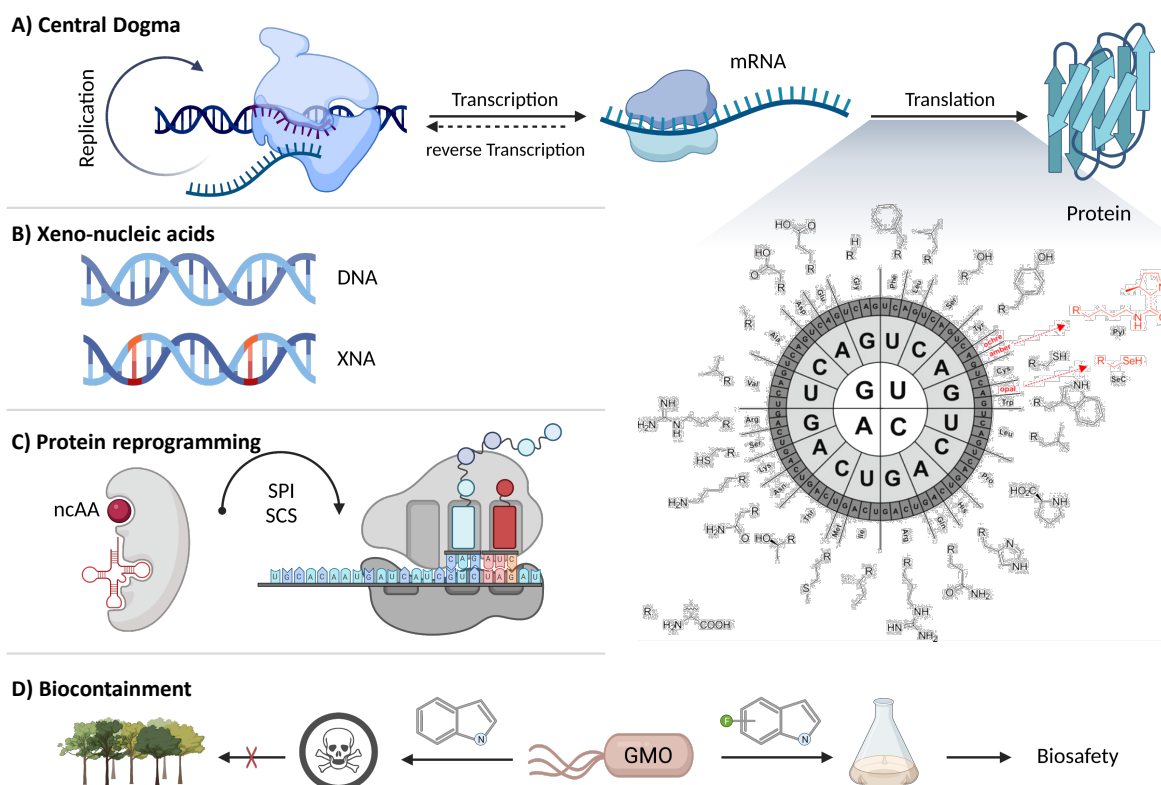
Figure 8 | Aspects of xenobiology. A) The flow of genetic information according to the central dogma of molecular biology is shown. DNA is replicated and transmitted to descendants or transcribed into mRNA. Then, the genetic code is translated into amino acids, building functional proteins. Below a radial representation of the genetic code structure in RNA format is shown. From inside to outside: 64 mRNA triplets (5' → 3') are assigned to the 20

canonical amino acids (cAAs) or to three stop codons (amber, ochre, opal). The natural expansion of the genetic code gave rise to the 21st and 22nd amino acids selenocysteine (SeC) and pyrrolysine (Pyl), depicted in red. The illustration of the genetic code is a courtesy provided by Stefan Oehm.[220] In the context of xenobiology, the genetic code can expanded by either B) introducing xeno-nucleic acids (XNAs) or by C) incorporating non-canonical amino acids (ncAAs) using SPI or SCS. C) Biocontainment concept based on the reliance on synthetic metabolites, i.e., GMOs only survive in laboratory environments supplying certain xeno-nutrients but otherwise die in natural ecosystems that do not provide them. The figure was drawn with BioRender (https://www.biorender.com/).

1.3.1   Genetic code engineering by expanding its amino acid repertoire

In general, synthetic biology employs two design principles to expand the genetic code: "top-down" and "bottom-up",[221] however, the transitions of both strategies are often smooth. The top-down approach is based on an existing organism and aims to simplify its highly complex system until only the minimal equipment of essential components (genetic repertoire) is left over.[222] Such a minimal cell provides a view of the key components and one day could represent a biotechnological production platform that could be retrofitted with special metabolic processes. Using a series of sequential deletions, for example, the genomes of *Escherichia coli* (4.64 Mb) were reduced to genome sizes of 3.62 Mb (77.8 %)[223] and 2.83 Mb (61.1 %, the smallest *E. coli* genome so far)[224] and that of *Bacillus subtilis* (2.68 Mb) to 1.54 Mb (63.5 %).[225] Synthetic genome minimization by means of de novo synthesis enabled the assembly of a synthetic *Mycoplasma mycoides* 1.08 Mb genome[226] and an *E. coli* strain with 61 codons,[227] among others. The "bottom-up" approach creates complex biological systems from scratch using simple chemical building blocks or repurposing individual components for unprecedented new-to-nature functionalities.[228–230] Therefore, bottom-up reconstitutions contribute also to a better understanding of biological systems. In this context, the interaction of individual components of a cell such as membranes (e.g., artificial vesicular systems) and other scaffold modules are investigated in order to produce a kind of hardware of a cell; [231,232] and in this respect, cell-free systems, in particular, have proved useful.[233] This form of forward engineering also includes the synthesis of artificial entities not found in nature by means of implementing synthetic nucleotides and amino acids into living cells (vide infra/supra).

Two main strategies were developed to expand the genetic code by means of reprogramming the protein biosynthesis (**Figure 9**).[234–236] With sense codon reassignment, also referred to as selective pressure incorporation (SPI) method, ncAAs are residue-specific incorporated in response to a certain sense codon.[237] In contrast, stop codon suppression (SCS) uses one of the three non-sense codons (amber, ochre, opal) and an orthogonal pair consisting of aminoacyl-tRNA synthetase/tRNA (aaRS/tRNA) for site-specific incorporation.[238] While SPI relies on the substrate promiscuity of the endogenous translation system and therefore allows only incorporation of isosteric ncAAs with a certain structural similarity to its canonical relatives (termed as analogs),[192] SCS provides a broader

substrate scope given by the opportunity to incorporate chemical more distant ncAAs (surrogates) provided that a suitable orthogonal pair has been found or designed.

Both techniques enabled the incorporation of a large variety of structural different ncAAs into proteins of many organisms and thus allowed the study of protein function by modulation of its activity and structure as well as by serving as spectral or imaging probes.[234,239–241] This includes, for example, ncAAs employing olefins,[242] keto groups, terminal alkynes, or azides[243–246] that facilitate click chemistry/biorthogonal reactions (various types of ligation and alkyne cycloadditions);[247] redox-active ncAAs with electron-withdrawing/-donating groups;[248–250] or ncAAs with unusual elements such fluorine[84,236] and other halogens, selenium,[234,251–254] tellurium[253,255,256] and even the incorporation of multiple ncAAs.[257] Its applications range from usage as spectroscopic probes to study protein biophysics,[258,259] probes to decipher catalytic mechanisms like the radical transport pathway in class Ia ribonucleotide reductase (RNR)[260–263] up to the development of mussel-inspired bioadhesive proteins.[264,265]
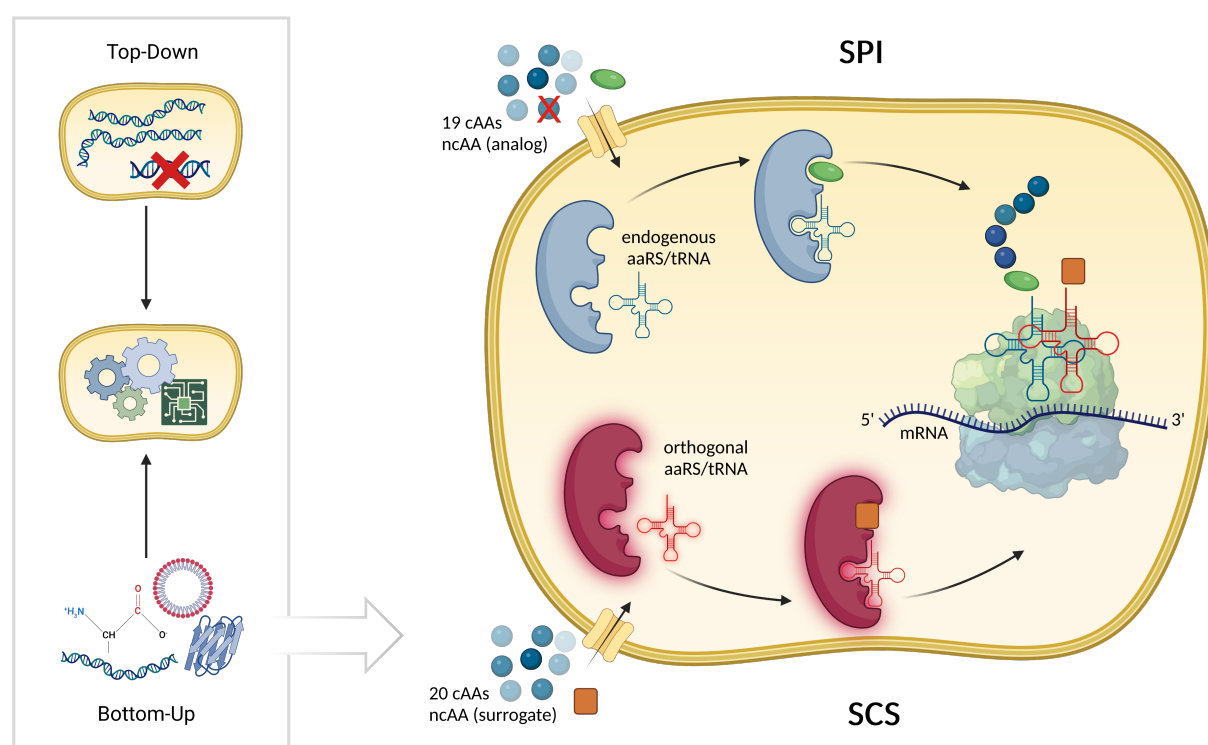
Figure 9 | Strategies for genetic code expansion. Design principles "top-down" and "bottom-up" used in synthetic biology (left). Schematic overview of SPI (selective pressure incorporation) and SCS (stop codon suppression) techniques for *in vivo* incorporating ncAAs (right). SPI uses the endogenous translation system (aaRS/tRNA) to incorporate structurally similar ncAAs in response to a sense codon. SCS makes use of orthogonal aaRS/tRNA pairs to incorporate surrogate ncAAs in response to a non-sense codon. SPI is considered code engineering, as a cAA is exchanged for an ncAA, whereas SCS results in code expansion, as a new ncAAs is added to the existing pool of 20 cAAs. The illustration of SPI and SCS approaches was re-drawn from Isabella Tolle[266]; the figure was drawn with BioRender (https://www.biorender.com/).

1.3.1.1 Sense codon reassignment by Selective Pressure Incorporation (SPI)

The stepwise recognition from speculation to the first efficient incorporation of ncAAs can be traced back to the 1950th, starting with the observation that many analogs of the common 20 amino acids inhibit the growth of bacteria and act as antimetabolites.[267] That was later observed to be a consequence of their incorporation into proteins by the biosynthetic machinery[268,269] and led to early directed experiments[251] as well as to the apace recognition of their potential for biotechnology applications.[252,270]

Selective pressure incorporation (SPI) is a fermentation-based method and experimentally proceeds in two phases.[237] First, an auxotrophic strain is cultured with the complete set of amino acids, including the target amino acid, to produce sufficient healthy cells. Then, the target amino acid is removed from both the medium and the cells (washing, starvation) and cultivation is continued with the ncAA. In the second phase, the expression of a desired protein is also induced. Since the strain is not able to synthesize the target amino acid, its removal exerts a selective pressure that forces the cells to mis-incorporate the ncAA (**Figure 9**), which is due to the substrate promiscuity of the aaRS.[234] Consequently, all positions are replaced by the ncAA.

1.3.1.2 Stop codon suppression (SCS)

The development of stop codon suppression (SCS) took almost four decades longer when the Schultz lab in 1989 first reported the site-specific incorporation of Phe analogs into β-lactamase.[271] At that time, translation was still performed *in vitro* with a supplemented chemically acylated suppressor tRNA that inserted the amino acid in response to a stop codon, whereas SCS later made use of orthogonal translation systems, i.e., aminoacyl-tRNA synthetase/tRNA (aaRS/tRNA) pairs.

The two most commonly used orthogonal systems for codon reassignment derived from archaebacteria and are 1) the tyrosyl-tRNA synthetase/tRNA (TyrRS/tRNA) pair from *Methanococcus jannaschii*[238,272,273] used for incorporation of substituted phenylalanine derivatives and 2) the pyrrolysyl-tRNA synthetase/tRNA (PylRS/tRNA) from *Methanosarcina barkeri* and *Methanosarcina mazei*.[274,275] For site-specific incorporation of ncAAs, the construction of an orthogonal aaRS/tRNA pair with a suitable substrate specificity towards the target ncAA is required (**Figure 9**). Briefly, using directed evolution libraries are created and functional tRNA and aaRS mutants are selected by double-sieve selection. However, since ALE is based on the traditional auxotrophic method, SCS is not discussed in further detail, here, but is reviewed elsewhere.[199,239,276,277]

In the end, however, rational design is limited by the incredible complexity of biological systems. Therefore, scientists began to harness nature's intrinsic capability to improve and to adapt.[278]

1.3.2 Adaptive Laboratory Evolution

Evolution is the driving force of diversity and improvement. Since Darwin's and Dallinger's correspondence[1] at the end of the 18th century and early experiments starting in the mid of 19th century,[279-281] the techniques of experimental evolution have greatly developed, but the principles stay the same.

In its simplest form, evolution experiments are performed by culturing cells in a chosen environment for a prolonged period of time to select those that have acquired the desired genomic or phenotypic trait through adaptation.[282,283] These cells will outcompete their ancestor and dominate the population. Here, it should be noted that selection is strongly dependent on the ALE conditions, i.e., not only on the imposed selection pressure but also on more immediate factors such as growth management. For example, if the culture is kept continuously in the exponential phase, the selection of fast-growing cells rather occurs, whereas a growth setup with alternating conditions (feast-and-famine) favors the selection of cells that have a long-term survival advantage.

In general, two main techniques have been established for conducting long-term propagation: continuous culturing and batch culture (**Figure 10A**). Continuous culturing makes use of bioreactors (chemostats and turbidostats) that allow precise control of environmental conditions, such as pH and oxygenation, and are hallmarked by constant nutrient supply and outflow of random individuals and waste, thereby also keeping the population size and growth rates constant.[284] In contrast, the serial transfer regime relies on repeated inoculation of batch cultures into fresh medium. A hallmark of this technique is the exposure to fluctuating conditions, i.e., population bottlenecks,[285] because during the growth phases, the population density as well as the medium composition and pH changes due to consumption of nutrients and accumulation of metabolic products (anabolism, catabolism). Referring to continuous culturing, batch growth contrasts exposing cells to both a turbidostat-like and a chemostat-like environment. In these alternating "feeding and starvation" regimes or "seasonal environments," fitness gains may result from decreased duration of the lag phase prior to cell division, increased growth rate during the exponential growth phase, or a decreased probability of entering a quiescent, non-reproductive state upon progressively nutrient depletion.[286] No matter which technique is applied, adaptation is a dynamic process (e.g., the occurrence of selective sweeps), and populations always exhibit a certain genetic diversity caused by coexisting competing lineages and clonal interferences.[287]

Another technique to propagate microorganisms in the frame of evolution experiments is given by the colony transfer method (**Figure 10A**).[288,289] Cells are spread on agar plates and one or more arbitrary individuals are selected repeatedly for breeding, thereby deliberately creating population bottlenecks that wipe out genetic diversity so that mutations are random and not linked to fitness effects. This experimental setup is used for unbiased mutation accumulation to determine mutation frequencies and properties of spontaneous mutations.
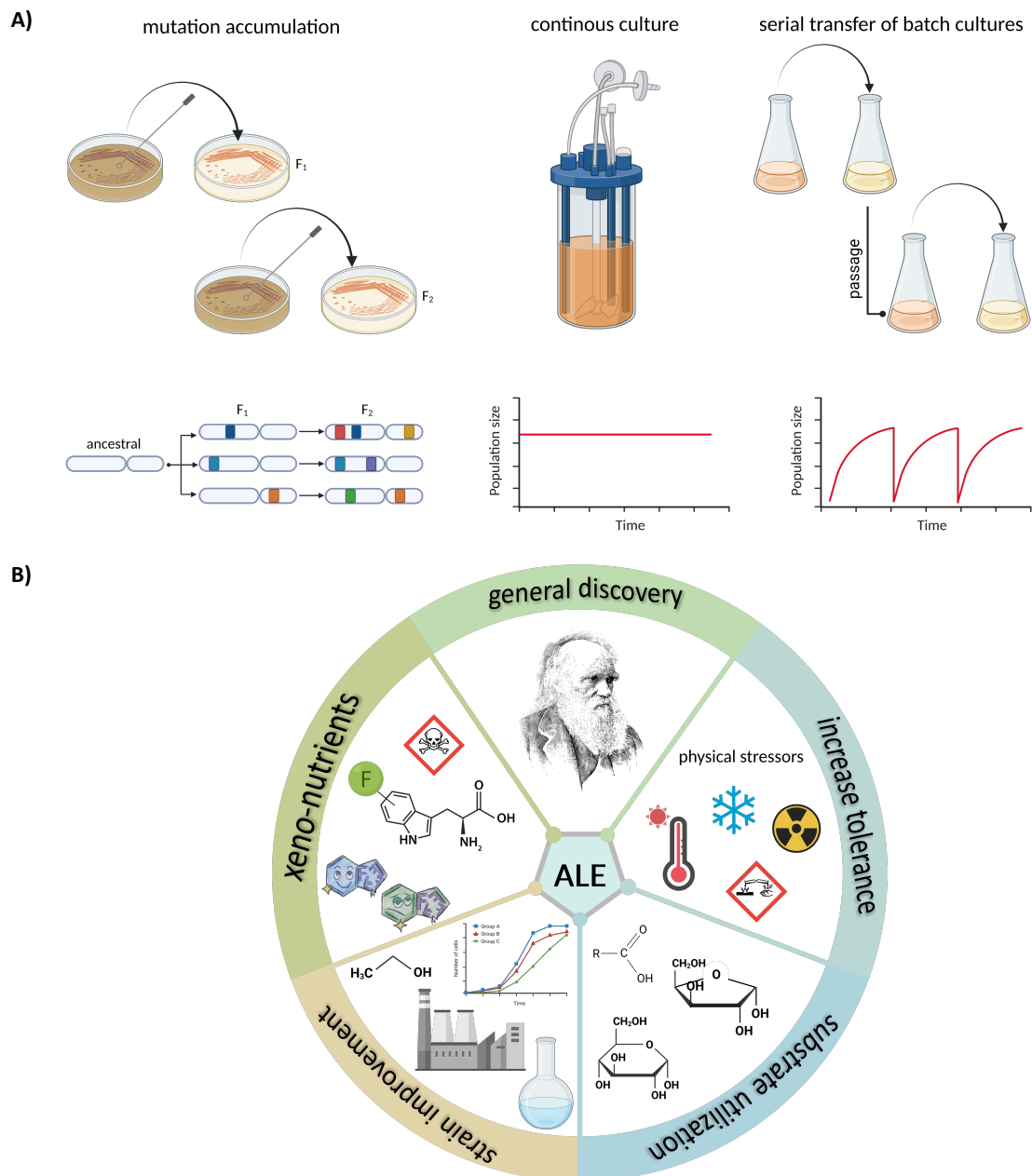
Figure 10| Adaptive Laboratory Evolution: techniques and applications. A) Types of evolution experiments to propagate populations and their experimental traits are illustrated. Colony transfer is used for unbiased mutation accumulation, continuous culturing uses bioreactors to keep the population size constant and in the serial transfer of batch cultures populations experience feeding-and-starvation environments. The illustration is inspired by Barrick *et. al.*[289] B) Categorization of ALE experiments according to their imposed selection pressure or intended application. ALE is used to increase the tolerance of microorganisms to several stressors, to expand its substrate utilization, for strain optimization (e.g., microbial cell factories), for the creation of xeno-nutrient-dependent organisms, and for general discovery purposes. The illustration is inspired by Sandberg *et. al.*[283] and the portrait of Charles Darwin appeared in his "The Origin of Species".[2] The figure was created using BioRender (https://www.biorender.com/).

The probably most popular example of ALE is the long-term evolution experiment (LTEE) by Richard E. Lenski. In 1988, Lenski started the propagation of 12 independent *E. coli* clones in a glucose-limited

medium to study evolution in action, in particular, the evolutionary trajectories and repeatability.[290,291] It took 20 years, about 31500 generations, and a rare mutation for a population to acquire the ability to use citrate as an alternative carbon source.[292] By duplicating the previously silent citrate transporter *citT* and controlling transcription by an aerobically expressed promoter (*rnk*), this rare Cit+ phenotype evolved.[293] Remarkably, Lenski still continues the LTEE to date, and by now, the 12 *E. coli* populations reached over 75000 generations, which exceeds all other reported evolution experiments by far.[294] He did seminal work in the field of experimental evolution and immensely expanded the understanding of evolutionary mechanisms and processes (historical contingency, evolution dynamics, clonal interferences, genetic drift, fitness measures, etc.).[287,289,295–297]

ALE was applied to a multitude of organisms, and even more conditions were used. Depending on the imposed selection pressures and the intended application, ALEs can be categorized as shown in **Figure 10B**. They include increasing tolerance to physical or environmental stress such as temperature,[298–301] including freeze-thaw,[302,303] osmotic pressure,[304–306] pH,[307–309] UV irradiation,[310,311] and multiple stresses;[312] expanded or changed substrate utilization by means of enhanced uptake[313–316] and use of alternative nutrients, in particular carbon sources;[317–319] increase growth performance of fermentative production strains (microbial cell factories) and enhance tolerance of inhibiting byproducts such as alcohols in biofuel or chemical production.[320–324] Another field is ALE experiments on the use of xeno-nutrients (discussed below),[325–328] and of course the knowledge gain factor should not be underestimated such as research into fundamental bioprocesses. Impressive examples of this are ALE experiments that use minimal cells.[329,330] This list of selected examples does not claim to be exhaustive by no means (e.g., Sandberg *et. al.*[283] covered 159 ALEs) but is rather intended to give an impression of the many possibilities of ALE and to illustrate the great interest and frequent use by the scientific community. Approaches have even been developed to accelerate the ALE process, which are based on increasing the mutation rate of the organism in question.[331] These either make use of mutagenic agents[332–335] or introduce defective mutations in genes involved in DNA replication (proofreading deficient polymerase) or repair systems (*mutSLH*).[336–338] Recently, the use of a mutator protein (MutaEco) was described, which increases the mutation rate without affecting cell viability,[339] and the development of mutant libraries for ALE created through CRISPR-enabled trackable genome engineering (CREATE), a multiplex genome engineering technique.[340,341]

Implementing of ALE is only one side of the coin - the interpretation is often much more challenging. Recently, attempts were made to consolidate the knowledge gained from the multitude of ALEs by the generation of the web-based platform ALEdb (https://aledb.org), which lists experimental setups (metadata such as organisms, growth conditions, lineages, etc.), selection pressures and genomic data.[342] Moreover, in this context, ALEdb also provides the possibility for automated key mutation reporting in order to identify mutation trends that could indicate relevant genomic variations for adaptation. However, it is difficult to cover all experiments (to date, 90 publications are listed), and for example, the fluorotryptophan-based ALEs described below are unfortunately not listed.

1.3.3 Encoding Fluorotryptophan

So far, three major attempts have been made to reassign the UGG codon to fluorotryptophan, whose chronology is sketched in **Figure 11**. It started in 1983 when J. Tze-Fei Wong quantitatively replaced Trp residues with 4-fluorotryptophan (4FTrp) in the proteome of *Bacillus subtilis*.[332] He serially mutated the trp-auxotrophic *B. subtilis* QB928 using the mutagenic agent *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine. Only a few rounds of mutagenesis were required to yield *B. subtilis* HR15, which prefers 4FTrp over Trp for growth. Even then, Wong speculated that HR15 might not be the last evolved organism with an alternate genetic code, and he was right. In 2001, Bacher and Ellington revisited the Wong experiment with 4FTrp-tolerant *Escherichia coli* strains[343] and provided the first genomic studies; and found genes *aroP*, *trpS* and *tyrR*, involved in uptake and utilization of Trp to be mutated. Unfortunately, minute amounts of natural Trp were still present in fully adapted cultures (approximately 0.03 %, 90 nM) with serious consequences. It provoked the development of a TrpRS variant with an increased capability of scavenging the smallest amounts of Trp instead of evolving an increased efficiency for 4FTrp, indicated by an improved turnover for Trp and increased discrimination against 4FTrp. Although the evolved strain still preferred Trp and thus does not contain a new genetic code, it exhibits an increased tolerance for isosteres 5-fluorotryptophan (5FTrp) and 6-fluorotryptophan (6FTrp) compared to the parental strain. Bacher and Ellington also performed experiments on adapting of bacteriophage Qβ for growth on 6-fluorotryptophan.[344]

Wong *et. al.* continued work and encouraged *B. subtilis* cells to propagate on Trp, and 4-, 5- and 6-fluorotryptophans.[345] They then suggested oligogenic barriers as a mechanism to preserve invariance of the genetic code, i.e., a small number of essential proteins that malfunction upon the incorporation of ncAAs and thus do not allow expansion of the code; unless it would be remedied by their mutation. Later on, genomic and transcriptomic analysis revealed changes that enabled the displacement of Trp by 4FTrp, such as mutations in genes related to Trp metabolism, stress response as well as transcription and translation system.[346] A non-sense mutation in the *mtrB* gene encoding for the Trp operon RNA-binding attenuation protein (TRAP) confers resistance to FTrp[343,347] and might be causative for the upregulation of the Trp transporters (increased Trp uptake) and Trp biosynthetic genes in HR23 growing on 4FTrp. Furthermore, crucial mutations in *rpoBC* of the RNA polymerase (RNAP) at conserved positions, as well as mutations in sigma factors SigB and SigI, could contribute to the ability of HR23 to propagate on 4FTrp and reject Trp.

The most recent experiment describes the adaptation of *Escherichia coli* to 4FTrp and 5FTrp using adaptive laboratory evolution.[328] Agostini *et. al.* used a Trp-auxotrophic prototype that was gradually forced to use the fluorinated Trp precursors 4Fi and 5Fi for in situ synthesis of the corresponding Trp analogs and their proteome-wide incorporation. Furthermore, they provided insights into cellular processes required for the adaptation using multiomics analysis. While a surprisingly low number of mutations were accumulated during long-term propagation, the changes occurred mainly on regulatory

networks, quality control of protein folding, and membrane integrity (summarized in **Figure 11**). Among others, mutations in *mtr* (high-affinity transporter Trp permease) and *mdtFKO* (multidrug efflux pumps)[348,349] associated with the uptake of Trp and detoxification of xeno-compounds occurred, respectively. Proteomics by means of SILAC (Stable Isotope Labeling by Amino Acids in Cell Culture)[350] revealed changes in the expression levels of protein chaperones and proteases that assist in the folding and degradation of misfolded proteins and the activation of CpxAR-mediated stress response system[351] which is also associated with misfolded proteins. Notably, at the beginning of the ALE, they were upregulated, while their levels decreased back to normal at the end of the ALE, suggesting an adaptation mechanism based on loosening the quality control of protein folding. And finally, the adaptation process was accompanied by membrane rearrangement through increasing production of unsaturated phosphatidylethanolamines, which loosened the packing of adjacent fatty acids and, therefore, increased the fluidity of the lipid bilayer (see section **3.9**).
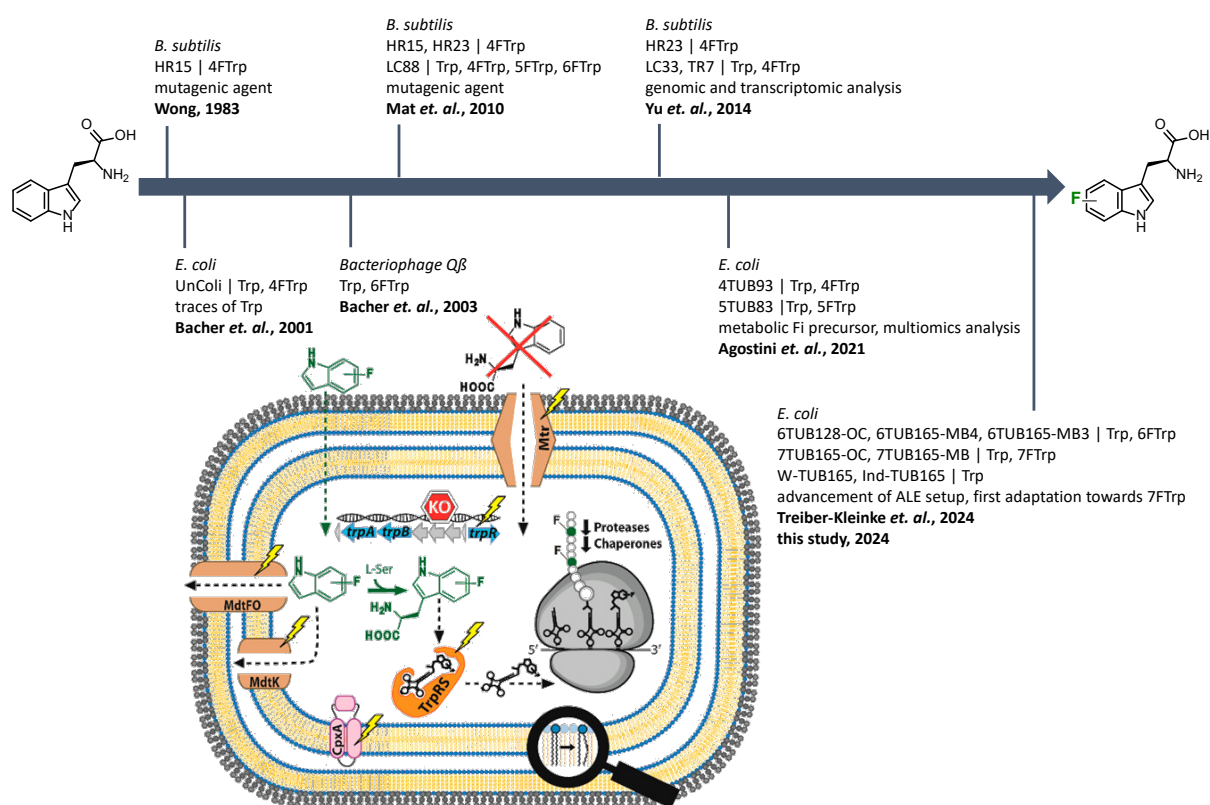
Figure 11 | Milestones in the encoding of fluorotryptophan. The seminal works of the groups of Wong, Bacher and Ellington are shown. Organisms, strain designations, growth substrates, some key points of the studies, as well as the respective publications are depicted. The work of Agostini *et. al.* is highlighted with an illustration that summarizes their ALE setup and results. The figure was taken from the corresponding publication.[328] In addition, reference is made to the adaptation experiments performed in this study and published in Treiber-Kleinke *et. al.*.[352]

The described adaptation experiments contribute to a better understanding of the mechanisms underlying an adaption process and provide valuable indications for further evolution experiments to create organisms with new genetic codes.[353,354]

2   Aim of this study

The flexibility of the genetic code will be probed by recoding the UGG codon towards fluorotryptophan. Based on the studies of Wong *et. al.*,[325] Bacher *et. al.*[343] and particularly of Agostini *et. al.*[328] on the proteome-labeling by fluorotryptophan, this study aims to evolve a tryptophan-auxotrophic *Escherichia coli* strain that is capable of using 6-fluoroindole (6Fi) and 7-fluoroindole (7Fi), respectively, for in situ synthesis of the corresponding fluorotryptophans (6FTrp and 7FTrp) and their global incorporation. It will be investigated whether the effects observed during adaptation to 4-fluoroindole (4Fi) and 5-fluoroindole (5Fi)[328] are generalizable or specific for the position of fluorine substituent.

Using adaptive laboratory evolution (ALE) by means of a serial inoculation regime and strict selective pressure, the cells will be forced to relinquish their dependence on indole and instead to accept and adapt to the fluorinated counterpart. The previously described ALE concept will be expanded by adaptation of positive controls and advanced through novel amino acid removal strategies, referred to as overall concentration (OC) and metabolic blocks (MB). The adaptation process will be evaluated in detail, and the proteome-wide substitution of Trp by FTrp proved by nano-LC-MS/MS. Subsequently, the resulting adapted lineages will be characterized using comprehensive growth experiments, viability assays, antibiotic susceptibility testing, and investigation of the cell morphology; furthermore, their capability to host protein production will be determined by recombinant expression of GFP.

Finally, to gain insights into how life on fluorine is possible and how a stressor becomes a vital nutrient for bacterial metabolism the different biological information levels will be investigated by genomics and metabolomics. Also, a hypothetical adaptation mechanism involving the *sspA*-mediated stress response system will be examined using CRISPR/Cas9 genetic engineering.

Fluorine as a Bioelement:
the evolution of microbial cells to use fluorinated metabolites as nutrients for growth

native *E.coli*   **A**daptive **L**aboratory **E**volution   fluorine adapted *E.coli*
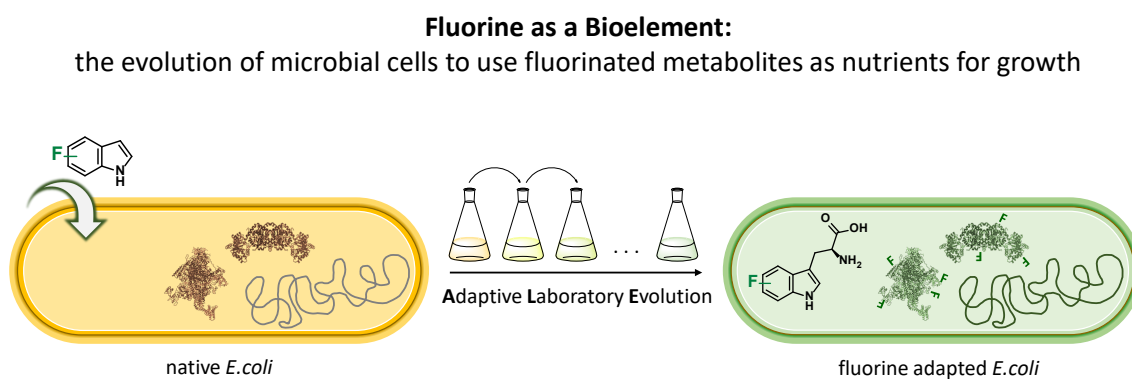
Figure 12 | The main design of Adaptive Laboratory Evolution (ALE) using fluorinated indoles to create fluorine-adapted *Escherichia coli* bacteria. These experiments result in adapted bacterial strains that are able to successfully overcome the negative effects of incorporating fluorinated indoles as indole analogs at Trp positions in cellular proteins.

3 Results and Discussion

3.1 Choice of analogs

For the adaptation experiments, the indole analogs 6-fluoroindole (6Fi) and 7-fluoroindole (7Fi) were chosen (**Figure 13**), which are intracellularly converted into their corresponding tryptophan analogs 6-fluorotryptophan (6FTrp) and 7-fluorotryprophan (7FTrp).

Tryptophan is an ideal candidate for global substitution for several reasons: 1) it is encoded by only one codon (UGG), which eases reassignment; 2) due its low abundance in the proteome minimal perturbation by ncAA replacement is expected; 3) the substrate promiscuity of its processing enzymes TrpS and TrpRS allow easy substitution by ncAAs; and 4) its general molecular structure and properties are attractive for introducing new functionalities; as discussed in section **1.2**.

Fluorination alters the chemical properties, for instance, polarity and lipophilicity. Compared to indole, lipophilicity (logP measured by octanol/water partition) is equally increased in fluorinated indoles (logP: Ind 2.20 < 4Fi 2.58, 5Fi 2.56, 6Fi 2.61, 7Fi 2.66) and they also exhibit a higher polarity (dipole moment, D) except for 7Fi (D: 7Fi 1.74 << Ind 2.24 < 6Fi 3.05 < 4Fi 3.59 < 5Fi 3.93).[328] Likewise, electrostatic interactions of the aromatic system are also influenced by the shift of electron density upon fluorination.[90] All these effects are induced by the unique stereoelectronic properties derived from the unprecedented combination of small size, very low polarizability, and the strongest known inductive effect inherent to fluorine as substituent in organic molecules.
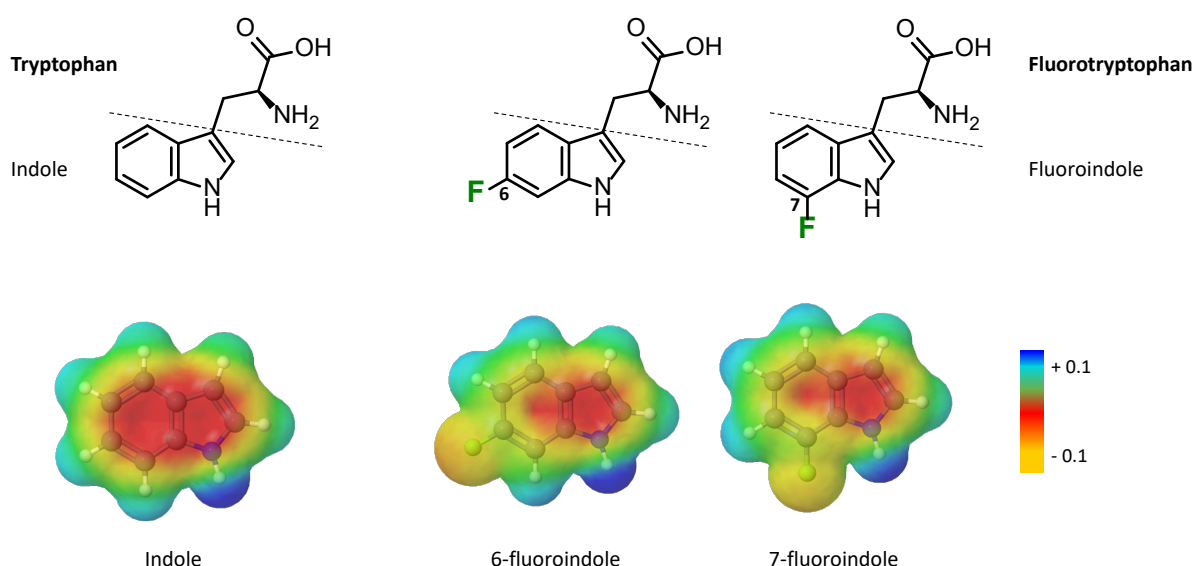
Figure 13 | Structures of Ind|Trp, 6Fi|6FTrp, and 7Fi|7FTrp. The upper panel shows the molecular structure of Trp, 6FTrp and 7FTrp with denoted indole moiety. The lower panel illustrates the changed distribution of electron density in 6Fi and 7Fi caused by fluorine decoration. Surface models representing the molecular electrostatic potential (range -0.1 to +0.1) were generated with the JSME Molecular Editor.[98]

The interest in fluorinated Trp analogs can be traced back to the 1960s and 1970s when their effect on cellular growth and protein synthesis of *E. coli* was investigated.[269,355,356] These phenomenological studies already demonstrated that 4-, 5-, and 6FTrp permit some growth of tryptophan-auxotrophic *E. coli* cells and that mono-fluorinated tryptophans are used for bacterial protein synthesis. In addition, they also described inhibitory effects on wild-type cells and differences derived from the fluorine position; that is, 4FTrp is superior to 5FTrp and 6FTrp in supporting growth and is a less potent growth inhibitor. Later on, fluorotryptophans were used more systematically for proteome-wide incorporation in the context of first evolution experiments, as described in section **1.3.3**.[325,326] However, these early studies exclude the exploration of 7FTrp.

3.1.1 Purity of indole analogs 6Fi and 7Fi

Instead of directly feeding the fluorinated Trp analogs 6FTrp and 7FTrp, respectively, the fluorinated precursors 6Fi and 7Fi were supplied. There are several reasons for this proceeding: 1) indole and its analogs uptake are supposed to diffuse through the cell membrane passively,[357] therefore, the need for transporter-mediated uptake of the ncAAs by the three known tryptophan transporters Mtr, TnaB and AroP[131,136] is circumvented; 2) the metabolic precursors indole as well as 6Fi and 7Fi are readily and enantioselectively converted into the corresponding L-amino acids by the tryptophan synthase (see section **3.1.2**); and 3) indole and its derivatives as small molecules can be obtained commercially at extremely high degrees of purity.

Commercial preparations of 6- and 7-fluoroindole were purchased from Sigma Aldrich and specified with 99.9 % purity (6Fi via infrared spectrum) and 97.3 % purity (7Fi via GC). However, their specific synthesis route and, thus, the information on possible contaminations are proprietary. In literature, fluoroindole syntheses mostly proceed from fluorinated nitrobenzenes using the Bartoli or the Leimgruber-Batcho method,[179,180,358] but also the use of fluorine-containing benzaldehyde as starting material for 6Fi[180] as well as fluorinated anillin for 7Fi[359] are described, and an early report is about Fischer cyclisation of ethyl pyruvate and fluorophenylhydrazone.[360] Thus, the putative synthesis route likely does not include an indole intermediate. However, both compounds (6Fi, 7Fi) were additionally analyzed by GC and HPLC coupled to mass spectrometry to definitely exclude possible contamination with traces of indole (see section **5.7**). This is crucial since minute amounts of indole could narrow the selective pressure and might subvert the adaptation process.[343]

The GC-MS analyses of Ind, 6Fi, and 7Fi are shown in **Figure 14**. With the method used, sufficient separation of the different indoles was achieved; the non-fluorinated indole was used as a positive control and to determine the detection limit (LOD), which was found to be < 10 pm. Indeed, indole was not detected in the 6Fi and 7Fi preparations. Both TIC chromatograms exhibit only one peak that could

be assigned to the respective fluorinated indole based on the mass spectrometric signals identified to be representative of the target substances (using the NIST MS search 2.2 program). Furthermore, attempts to improve the analysis by *N,O*-Bis-(trimethylsilyl)-trifluoracetamid (BSTFA)-derivatization failed.[361]
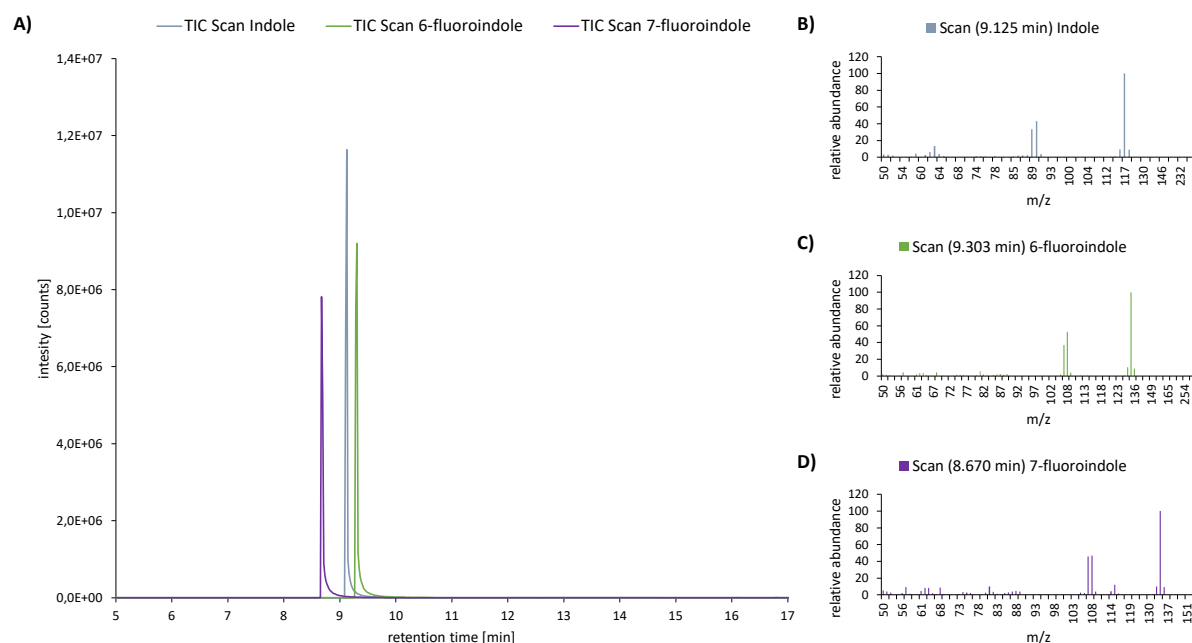
Figure 14 | Analysis of indole, 6-fluoroindole and 7-fluoroindole by GC-MS. A) GC-MS full scan (total ion current chromatogram, TIC) of indole (grey), 6-fluoroindole (green) and 7-fluoroindole (purple). B) Detected ions of TIC from indole (exact mass: 117.06 u) at retention time 9.125 min are m/z 117.1 (100 %) and m/z 90.1 (43 %). C) Detected ions of TIC from 6-fluoroindole (exact mass: 135.05 u) at retention time 9.303 min are m/z 135.1 (100 %) and m/z 108.1 (53 %). D) Detected ions of TIC from 7-fluoroindole (exact mass: 135.05 u) at retention time 8.670 min are m/z 135.1 (100 %) and m/z 108.1 (47 %).

6- and 7-fluoroindole were also analyzed by LC-MS-QQQ (**Figure 15**). Here, chromatograms of 6Fi and 7Fi, as well as mixtures of them with indole, were screened for indole contamination using single ion monitoring (SIM) mode with the mass filter of 118 m/z to detect only indole contaminants. In samples containing only 6Fi or 7Fi, no indole was detected, whereas in the positive controls containing a 1:1 mixture of indole and 6Fi or 7Fi, indole was found. A calibration curve of indole determined the limit of detection to be 1 pmol.
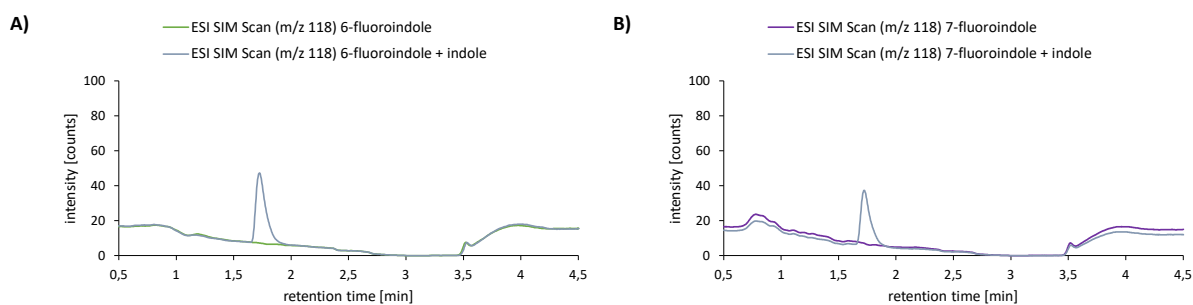
Figure 15| Analysis of 6-fluoroindole and 7-fluoroindole by LC-MS-QQQ. A) ESI SIM Scan of 6Fi (green) and a control mixture of 6Fi + Ind (grey) with mass filter 118 m/z. B) ESI SIM Scan of 7Fi (purple) and a control mixture of 7Fi + Ind (grey) with mass filter 118 m/z.

Since neither the GC-MS nor the LC-MS-QQQ analyses detected indole impurities in the 6Fi and 7Fi preparations, they were considered suitable for ALE. In addition, as part of the determination of proteome-wide FTrp substitution, the growth media NMM itself with and without supplementation of 6Fi or 7Fi, respectively, were also analyzed by HPLC using UV and fluorescence (FL) detection, to ensure the absence of indole in all components used for medium preparation. Here, too, no indole impurities were detected (see **Figure S63** in the appendix).

3.1.2 *In vitro* synthesis of 6- and 7-fluorotryptophan by tryptophan synthase

A key aspect of the described ALE setup is the in situ synthesis of fluorinated tryptophan from fluorinated indole (**Figure 16A**). Although the promiscuity of the tryptophan synthase (TrpS, *trpBA*) toward a remarkably broad variety of indole analogs is well known (see section **1.2.3**),[156,165] the capability of TrpS to convert 6Fi and 7Fi to the corresponding amino acids 6FTrp and 7FTrp has been experimentally validated. For the *in vitro* synthesis, an adapted protocol by Wilcox[168] was followed (see section **5.8**), and the TrpS of *Salmonella typhimurium* was used, which shares high sequence similarity to *E. coli* TrpS (α-subunit 85 % and β-subunit 96.5 %).[362,363] The reaction proceeds via a simple one-pot incubation of the reagents (F)-indole, L-serine, Trp synthase, and cofactor PLP. After purification by ion exchange chromatography, the successful formation of (F)-Trp was confirmed by LC-MS (**Figure 16B-D**), thus fulfilling the prerequisite for a successful forced-feeding regime. The calculated masses match those observed, and both signals (from *in vitro* and commercial samples) overlap.
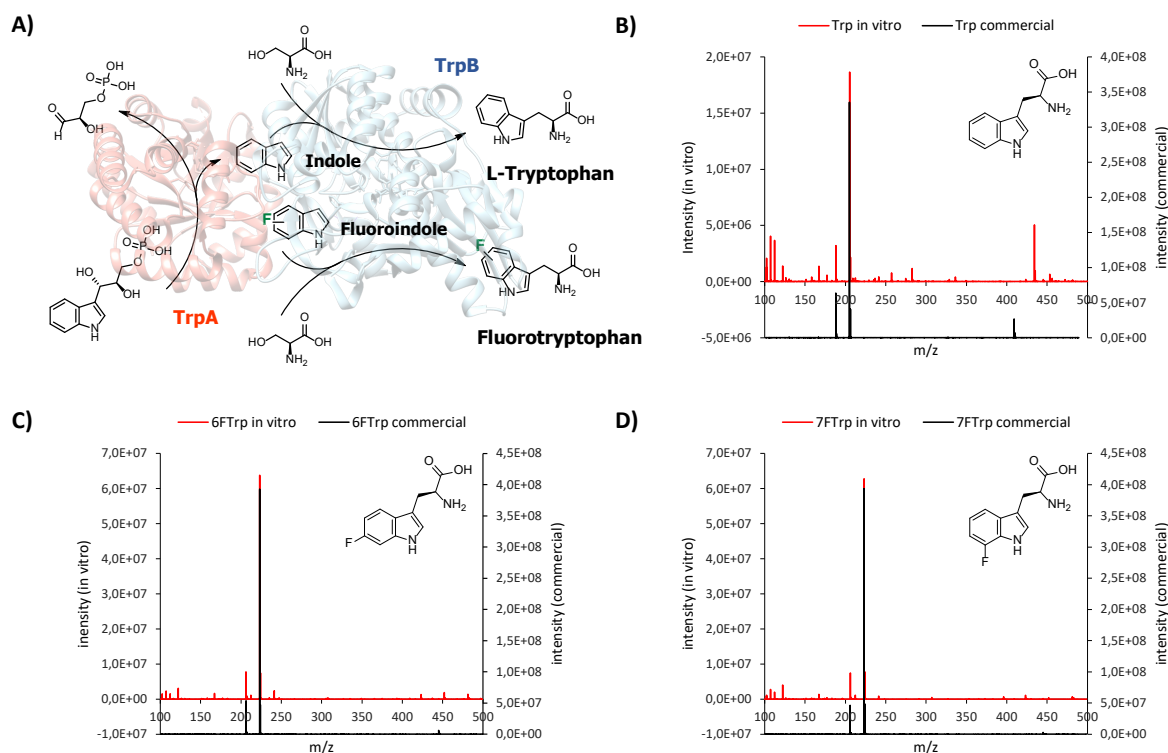
Figure 16 | Enzymatic synthesis of Trp, 6FTrp, and 7FTrp by the tryptophan synthase (TrpS). A) TrpS-catalyzed reaction for intracellular production of Trp or FTrp. In the native reaction indole-3-glycerol phosphate (IGP) is converted to indole and glyceraldehyde (α-reaction of TrpA subunit, orange), followed by condensation of indole and L-serine to build L-tryptophan (β-reaction of TrpB subunit, light blue). The β-reaction also proceeds for in situ synthesis of FTrp from F-indole. B-D) Mass spectrometric analyses (HPLC-ESI-MS) of Trp (B), 6FTrp (C), and 7FTrp (D) synthesized by TrpS. Mass spectra of the *in vitro*-produced tryptophans are colored in red, and those of the purchased (commercial) tryptophans are depicted in black at the second y-axis. Trp [M+H+] calculated: 204.09, Trp [M+H+] obtained: 205.10, FTrp [M+H+] calculated: 222.08, FTrp [M+H+] obtained: 223.09. (The illustration of the TrpS reaction (A) was inspired by Watkins-Dulaney *et. al.*[156] and the 3D structure of the tryptophan synthase (PDB ID: 1KFC [364]) was drawn by UCFS Chimera.[43])

3.2 Experimental configuration

3.2.1 Genetic configuration of the metabolic prototype *E. coli* TUB00

Stringent *Escherichia coli* auxotrophism towards Trp is a prerequisite for successful ALE experiments with fluoroindole to force its utilization. Therefore, the endogenous pathways of Trp synthesis must be knocked out in the genome. In this particular case, Trp auxotrophy was generated by deleting genes of Trp biosynthesis (*trpLEDC*; trp operon) and the Trp degradation pathway (*tnaA*; tryptophanase enzyme), while the genes for tryptophan synthase (TrpS; *trpBA*) and tryptophanyl-tRNA synthetase (TrpRS; *trpS*) remained in the genome. The metabolic prototype strain *E. coli* MG1655 (Δ*trpLEDC* Δ*tnaA*) with clean genomic background was generated by homologous recombination[365] and designated as TUB00 (**Figure 17**). TUB00 serves as ancestral strain in the adaptation experiments. The underlying principle of metabolic processing of F-indole and incorporation of FTrp is based on the substrate promiscuity of the two enzymes TrpS[156,168,171] and TrpRS,[188,332,343] which accept a wide range of indole and tryptophan analogs. Since the metabolic sources of indole (TnaA) and Trp (TrpLEDC) are eliminated, the ALE experiment requires bacteria to accept indole or a fluorinated indole analog and produce the corresponding amino acid via TrpS and L-serine to survive. Thus, depriving the cells of the canonical substrate (indole) while supplying them continually with the fluorinated indole analog exerts selective pressure and brings about the adaptation process.
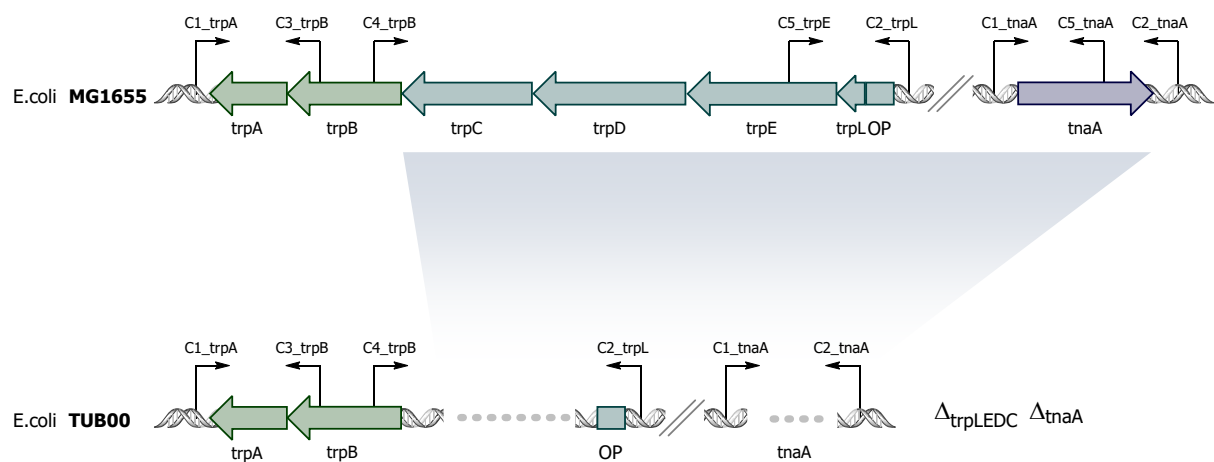
Figure 17 | Genetic configurations of *E. coli* MG1655 and TUB00. The genotypes of the parental strain MG1655 with intact Trp operon and tryptophanase gene, as well as its Trp-auxotrophic derivative TUB00 with the respective deletions, are shown with primer binding in their target genomic regions.

The genotypes of both strains were verified by colony-PCR (section **5.4.1.2**) and DNA sequencing (section **5.4.5**). The analyses confirm the presence of *trpBA* encoding the TrpS enzyme in both strains, as well as the deletion of the genes *trpLEDC* and *tnaA* in TUB00 (**Figure 18A**). According to the primer

binding (**Figure 17**), the following PCR pattern results: 1) detection of shorter amplicons for C4_trpB|C2_trpL (deletion *trpLEDC*) and C1_trpA|C2_trpL (full-length amplification) in TUB00; 2) no band for C5_trpE|C2_trpE since no primer binding due to deleted *trpE* in TUB00; 3) shorter DNA fragment for full-length amplification of the *tnaA* gene (C1_tnaA|C2_tnaA) in TUB00 and 4) no band for C1_tnaA|C5_tnaA because no primer binding C5_tnaA due to *tnaA* deletion in TUB00. This obligatory genetic constitution was checked regularly during the entire adaptation process.
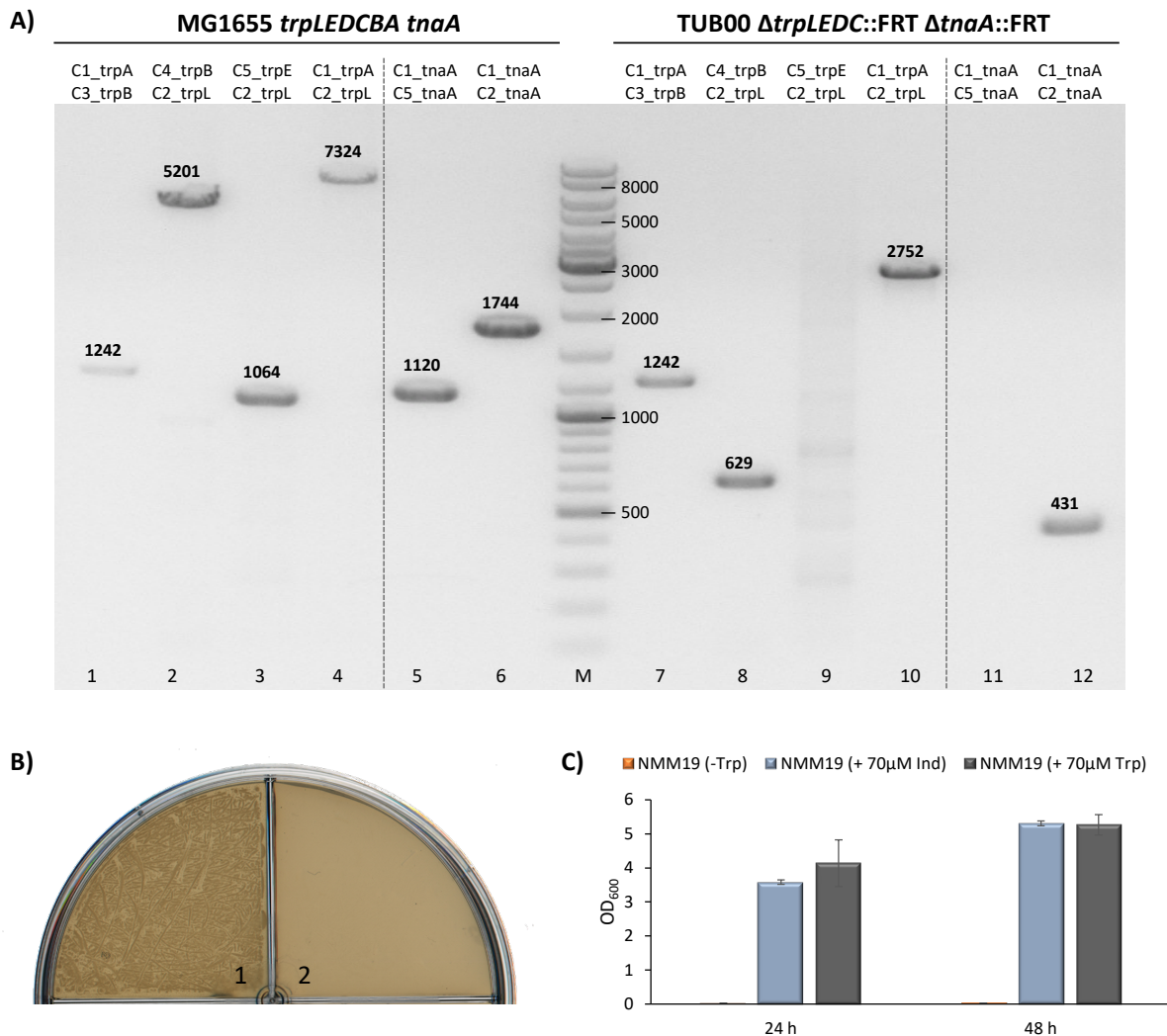
Figure 18| Genomic and phenotypic verification of Trp auxotrophy for the strain TUB00. A) The genotypes of MG1655 and TUB00 are shown by a typical PCR pattern. The fragment sizes are given with the values in silico determined in Geneious Prime 2023.2.1. B) TUB00 plated on NMM19-agar in the presence of 70 μM indole (1) and in the absence of any Trp source (2). C) Optical densities of TUB00 cultures at 24 h and 48 h incubation (at 30 °C, 200 rpm) cultivated in NMM19 without Trp (orange, OD_{600} ≈0.03) and supplemented with either 70 μM indole (blue) or 70 μM Trp (grey).

In addition, also the Trp-auxotrophic phenotype was verified by cultivating TUB00 in NMM19 lacking any Trp source as well as with either indole or tryptophan added as control (**Figure 18B-C**, see section **5.2.1**). Growth could only be observed when indole or Trp was supplemented, thereby confirming auxotrophy generated by the described genetic deletions.

3.2.2 Growth medium

New Minimal Medium (NMM)[366] was used as growth medium in different configurations, depending on the progress of the adaptation process (see **Table S20**, **Table S21**, **Table S22**, **Table S23** in the appendix). The growth medium is decoded by NMM([**x**]-**a-b-c**), where **x** is the concentration of canonical amino acids (cAAs) present in mg/L (only indicated when the concentration was below 50 mg/L); **a** is the number of different cAAs present (at concentration of 50 mg/L); **b** is the concentration of 6Fi or 7Fi in μM; and **c** is the concentration of indole in μM (see section **3.3**, **Figure 22**). NMM is a synthetic medium with a chemically well-defined nutrient content so that the growth during adaptation is not affected by variations in the basic composition, namely sources of carbon, phosphate, nitrogen, vitamins, and trace elements (see **Table 16** for NMM basic composition). The addition of canonical amino acids (NMM19) was always carried out without the target amino acid Trp.

3.2.3 ALE starting conditions

Before starting the adaptation experiments towards indole analog usage, the optimal starting conditions were elucidated. Generally, 10 mL of 0.02 OD_{600} culture were cultivated in 100 ml shaking flasks at 30 °C and 200 rpm. The ancestral strain TUB00 was grown in NMM19-[6Fi]-0 and NMM19-[7Fi]-0, testing different concentrations ranging from 0 – 10000 μM of the target substrates (section **5.2.1**). The growth of TUB00 with 6-fluoroindole resulted in typically shaped growth curves (**Figure 19A**). Concentrations ranging from 5 – 70 μM 6Fi yielded equally high cell densities of about OD_{600} 0.5 – 0.6, and higher concentrations of 100 – 10000 μM resulted in significantly decreased growth. Above 500 μM, only marginal growth could be observed. In 7-fluoroindole supplemented NMM19, the growth curves of TUB00 are clearly more pointed, with a steeper rise (and thus higher growth rates) but also a steeper fall (**Figure 19B**). The rapid OD_{600} drop could be the first indication of a higher toxicity potential of 7Fi and represents the most striking difference between both substrates during these initial growth tests (because culture density on 6Fi remained at a high level). While growth is still moderate at 5 μM 7Fi, TUB00 again reaches similarly high OD_{600} values of 0.5 to 0.6 with concentrations ranging from

30 – 70 µM. Here, too, the growth decreases continuously with increasing 7Fi concentration from 100 µM and from 1000 µM only rudimentary growth is observed.

Both the wide acceptability range of 15 – 70 µM without significant differences in cell density and the relatively high values obtained in NMM19 supplemented exclusively with the fluorinated analogs (without indole) were unexpected. For both substrates, a concentration of 70 µM of 6Fi and 7Fi, respectively, was chosen for the adaptation experiments to set the selection pressure as high as possible.

Figure 19| Elucidation of optimal ALE starting conditions (6Fi, 7Fi). Determination of the growth behavior of TUB00 in different concentrations (0 – 10000 µM) of A) 6-fluoroindole and B) 7-fluoroindole. The curves of the substrate concentration finally used are shown in bold (in each case, growth in 70 µM 6Fi, green or 7Fi, purple). Measurements were performed in triplicates.

The adaptation experiments were started with NMM19-70-0 supplemented exclusively with 70 μM 6Fi or 7Fi, respectively. However, with these initial starting conditions, cells could undergo only a few doublings, and growth stalled after the first inoculation step. A similar behavior was already observed in early studies investigating the growth of *E. coli* on fluorinated tryptophan[355,356] and might be the consequence of a sub-lethal injury caused by the action of fluorinated indoles.[367] This hypothesis is supported by colony-forming unit (CFU) assays (**Figure 20**, section **5.2.1.1**), which show that growth of Fi-treated cells (100 - 1000 μM) restores under non-selective conditions (NMM20) upon removal of the Fi stressors. Thereby, the duration of the Fi treatment (24 h vs. 48 h) also slightly influences the subsequent viability of the cells on NMM20, whereby this effect is more pronounced with 7Fi (compare dilutions 1-3). However, significant differences in the effect of 6Fi and 7Fi were not observed. Concentrations of up to 1000 μM were found to be lethal.
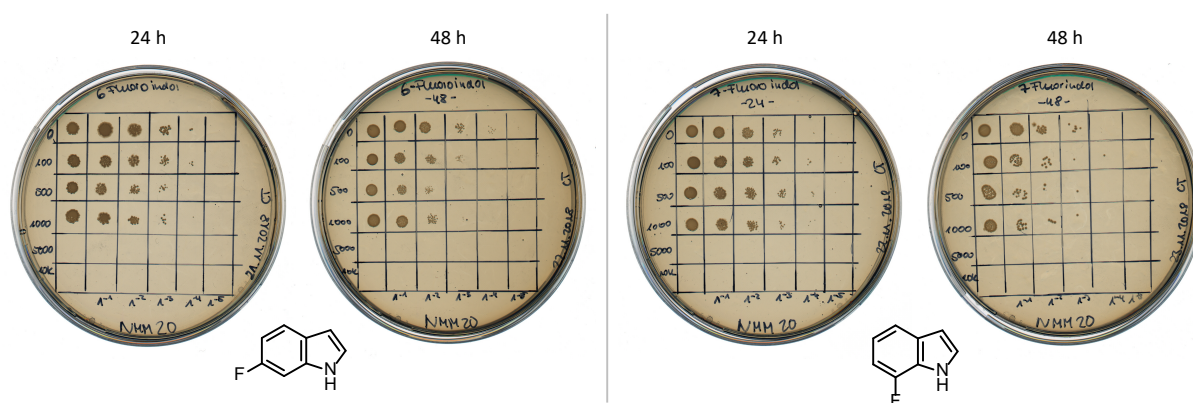
Figure 20 | Fluoroindole-caused sub-lethal injury (CFU-assay). CFU-assays for TUB00 cells treated with either 6Fi (left) and 7Fi (right) for 24 h and 48 h, respectively, are shown. TUB00 grown with different concentrations of 6Fi or 7Fi (100 - 10000 μM, as well as without) were spotted in different dilutions (undiluted, 1-1, 1-2, 1-3, 1-4, and 1-5) on non-selective NMM-agar containing all cAAs in excess (NMM20, 50 mg/L) and incubated for about 96 h at 30 °C. Compared to the positive controls (not treated, 0 μM), the growth of treated cells decreased with increasing Fi concentration.

Consequently, indole had to be added to the growth medium to promote stable growth. Different concentrations of indole ranging from 0 – 30 μM were tested, whereby the cell density increased along with increasing indole supply (**Figure 21**). For indole supply, a concentration of 2.5 μM (OD_{600} = 0.5) was chosen. In combination with 70 μM of the respective fluoroindole, this supported growth sufficiently to reach cell densities of about OD_{600} = 1. In the ALE experiments, analog usage is now encouraged by limiting the indole supply and supplying the desired fluorinated indole (6Fi or 7Fi) in excess.
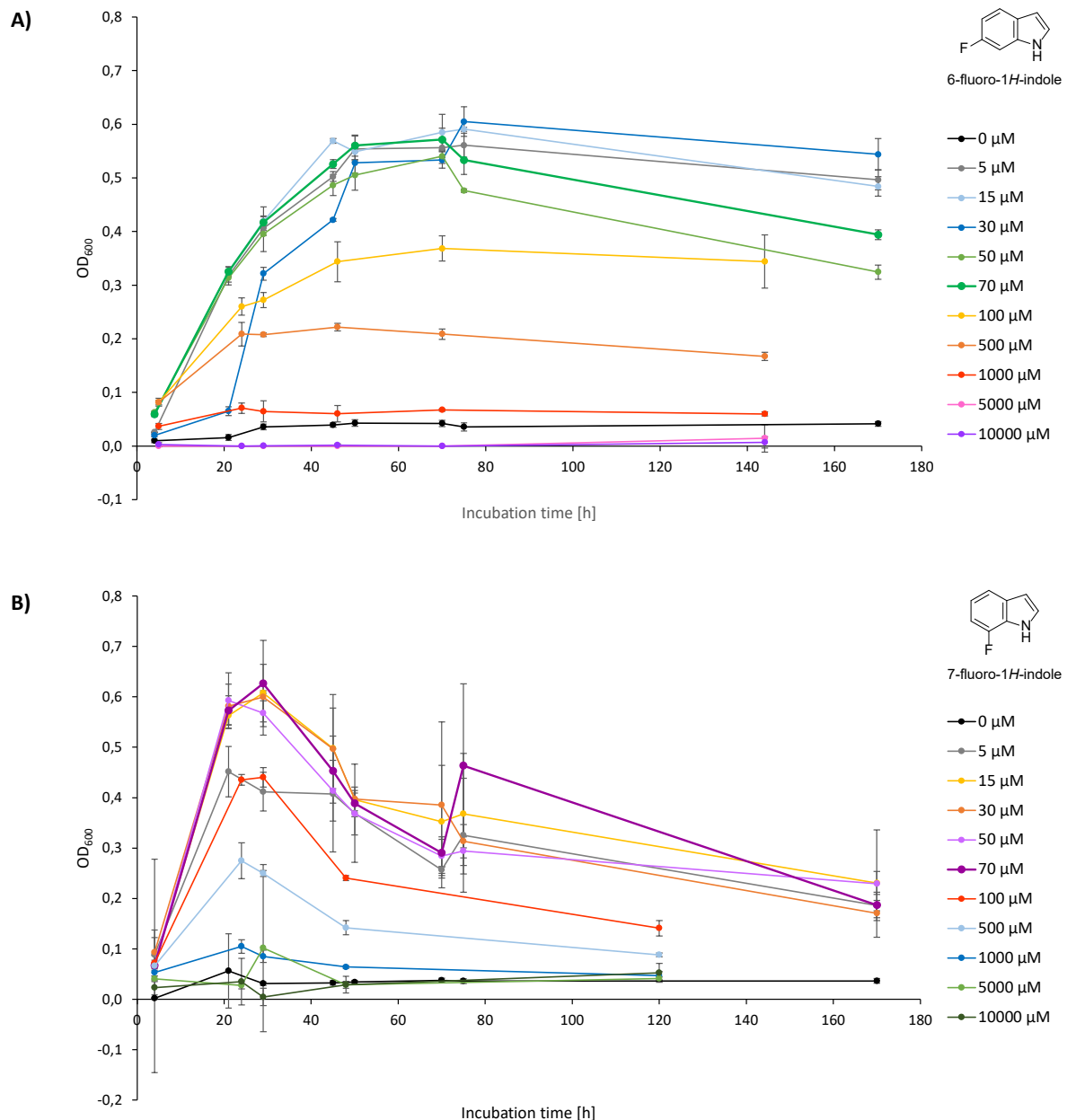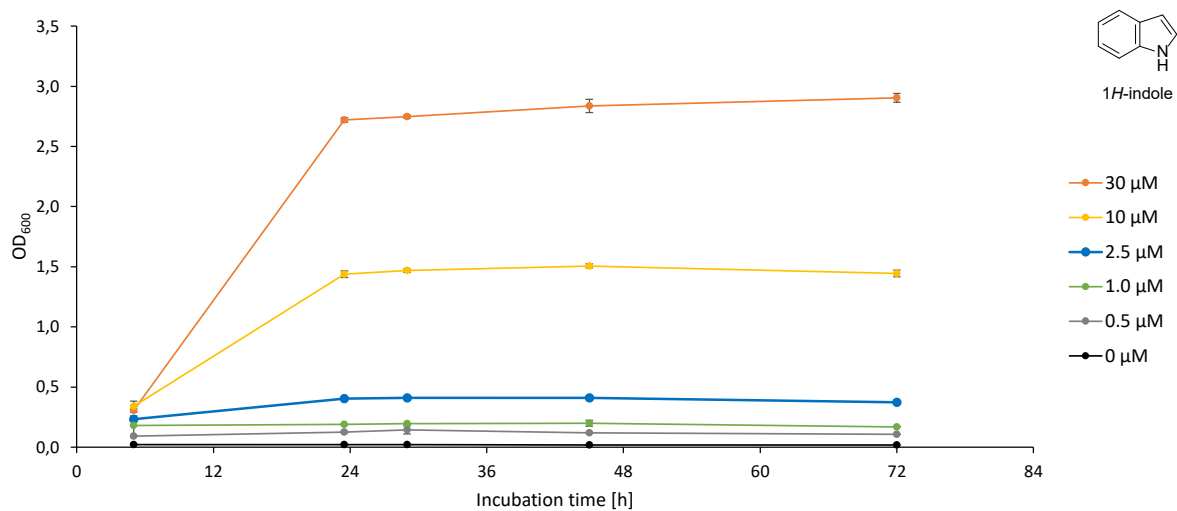
Figure 21 | Elucidation of optimal ALE starting conditions (Ind). Determination of the growth behavior of TUB00 in different concentrations (0 – 30 μM) of indole. The curve of the substrate concentration finally used (30 μM, blue) is shown in bold. Measurements were performed in triplicates.

3.3 Design and concept of adaptation experiments

A conceptual overview of the performed ALE experiments is shown in **Figure 22**. Four parallel experiments were designed: 1) a positive "indole control", 2) a positive "Trp control", 3) adaptation to 6Fi, and 4) adaptation to 7Fi. The indole control served to identify fluorine-induced effects, and the Trp control to investigate the general consequences of long-term cultivation on a minimal medium.[368,369] Liquid cultures were allowed to grow under a given set of conditions (primarily NMM composition) until they reached the early stationary phase, which is marked by a stable, no longer increasing optical density; once this occurred, a constant cell number corresponding to an OD_{600} value of 0.02 was used to inoculate the subsequent culture. These transfer steps or inoculations are defined as a "passage". Unlike the alternative approach of continuous cultivation, this type of serial transfer regime is characterized by alternating "feeding-and-starvation" regimes or "seasonal environments"[370] (i.e., a population bottleneck) but benefits from its flexibility when adjustments to cultivation conditions are required.[284] Furthermore, the serial transfer approach was chosen because this type of experimental management is closest to natural selection in that it mimics the changing conditions in the real environment of natural populations. Consistent inoculation of each new liquid culture with a normalized amount of cells ensures that even poorly growing cultures do not suffer an evolutionary disadvantage,[371] which became relevant, particularly in the case of the 7Fi adaptation. Ultimately, the primary goal was not to select fast-growing populations but those outcompeting by an acquired fitness increase in the presence of fluorinated indole.

Three biological replicates of the starting strain TUB00, which was assumed to be isogenic, were propagated in a synthetic minimal medium (New Minimal Medium, NMM) containing essential nutrients such as phosphate, ammonium, glucose, and vitamins (see **3.2.2**). NMM was differentially supplemented as follows. Indole control populations grew in NMM19-0-70 with a constant concentration of 70 µM indole and a high concentration of 19 of the 20 canonical amino acids (cAAs); that is, all except Trp. Trp control populations grew in NMM19-0-70 without indole but with a constant concentration of 70 µM Trp and a high concentration of the 19 cAAs (see section **3.4.1**). Populations adapted to 6Fi and 7Fi grew in NMM variably enriched with indole, the 19 cAAs, and the respective fluorinated indole (see **Table S20**, **Table S21**, **Table S22**, and **Table S23** for medium compositions). Growth conditions were abbreviated in all cases as NMM([x]a-b-c) as described in section **3.2.2** and **Figure 22**.

The adaptation towards usage of 6Fi and 7Fi was conceived in three phases that are also represented in **Figure 22**. In phase 1, the indole concentration was gradually and stepwise reduced to zero. The initial supply with 2.5 µM indole was necessary to promote cell growth and to facilitate the beginning adaptation process. In phase 2, the 19 cAAs were gradually and stepwise removed from the growth medium according to two distinct strategies (see paragraph below) to exclude possible contamination by traces of indole or tryptophan. They could originate from common amino acid preparations and

would otherwise disturb the adaptation process.[343] Finally, phase 3 represents the period of adaptation in which only the fluorinated indole analog (70 µM) was present as NMM supplement.

The adaptation process was driven by gradually and stepwise increase of the selective pressure, that was to reduce at first the concentration of indole and, after that reached zero, reduce the supply with canonical amino acids, to evolve populations that were exclusively dependent on fluorinated indole. Although this process was to some extent intuitive, a population had to meet the following criteria in order to be subjected to the subsequent growth condition. The population had to have reached early stationary phase and must have proceeded through at least two complete passages under that given condition. Usually, the populations required about 65 h to reach the early stationary phase growth; for that reason, the number of days of growth does not correspond 1:1 with the number of passages. And in most cases, more than two passages were required to observe either stable growth or no further growth improvement. Only after these criteria were met the conditions were tightened to increase the evolutionary pressure on the population.
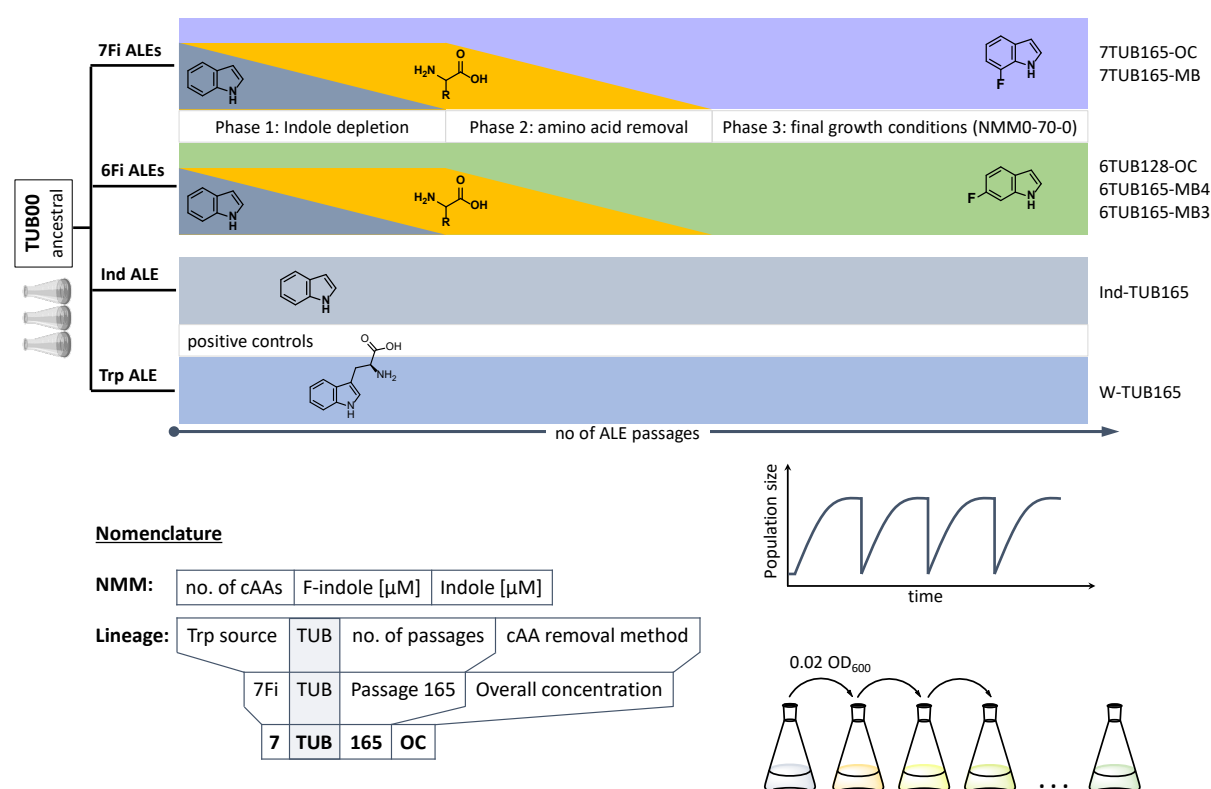
Figure 22 | Conceptual ALE trajectories. Four parallel evolution experiments on populations/lineages over the course of several passages under different growth conditions were conducted: two positive controls (indole and Trp, lower grey bars) and adaptation to 6Fi (green middle cartoon) or 7Fi (purple upper cartoon). Molecular structures of indole, Trp, 6Fi, and 7Fi are shown for each experiment. Increasing passage number (independent variable) is indicated by lowest black arrow; lineage nomenclature follows Z-TUB-X-OC/MB pattern where "Z" indicates the Trp-source (Ind, Trp, 6Fi or 7Fi), "X" the passage number and "OC/MB" the cAA removal method; the nomenclature of growth condition follows NMM(a-b-c) pattern where "a" indicates the number of cAAs, "b" the concentration of fluoroindole [µM] and "c" the concentration of indole [µM]. Phases 1 (indole depletion), 2 (amino acid depletion), and 3 (adjustment in the presence of a high concentration of only 6Fi or 7Fi as supplement)

are indicated. Phase 2 was realized according to overall concentration (OC) or metabolic block (MB) strategies (see text for details). In addition, the growth phases and the serial inoculation strategy are illustrated.

Phase 2 was designed according to two different approaches: so-called "overall concentration (OC)" or "metabolic blocks (MB)". OC and MB are visually represented in **Figure 23B**. In the OC approach, the overall concentration of all 19 amino acids was gradually and stepwise reduced from 50 to 0 mg/L, whereby with decreasing concentration of the added cAA19-mixture, the reduction steps were chosen more conservatively.

In the MB approach, the removal of amino acids was not based on their concentration but rather on their origin in core metabolism (i.e., their metabolic identity). In doing so, the canonical amino acids were assigned to metabolic blocks (**Figure 23A**), considering factors such as their metabolic costs in terms of energy and precursor demand. Metabolic blocks are grouped as follows: group I (Phe, Tyr, Trp) originates from erythrose 4-phosphate (eryP) and phosphoenolpyruvate (pep), group II (Ala, Val, Leu, Ile) from pyruvate (pyr), acetyl-CoA (acCoA) and oxaloacetate (oaa), group III (His) from ribose 5-phosphate (penP), group IV (Ser, Gly, Cys) from 3-phosphoglycerate (3pg), group V (Glu, Gln, Pro, Arg) from α-ketoglutarate (αkg) and group VI (Asp, Asn, Met, Lys, Thr) from oxaloacetate (oaa) and pyruvate (pyr).[99] For example, the first block that was removed comprised the aromatic amino acids tyrosine and phenylalanine, both requiring laborious multistep biosynthesis and originating from the same starting metabolites (eryP and pep). The metabolic blocks are reflected as rows in the MB funnel representation (right) in **Figure 23B**, i.e., the first removal concerned Tyr and Phe from cAA group I (**Figure 23A**).

While the MB approach was successfully applied in previous Trp-based adaptation experiments,[327,328] the OC approach presented here represents an innovation. The idea behind this was that the gradual reduction of all amino acids is more gentle for the cells than the sudden elimination of whole metabolic blocks. In this way, cells are softly encouraged to turn on their amino acid biosynthesis processes instead of being abruptly confronted with ribosome stalling and mistranslation effects that could be caused by sudden amino acid starvation.[372]
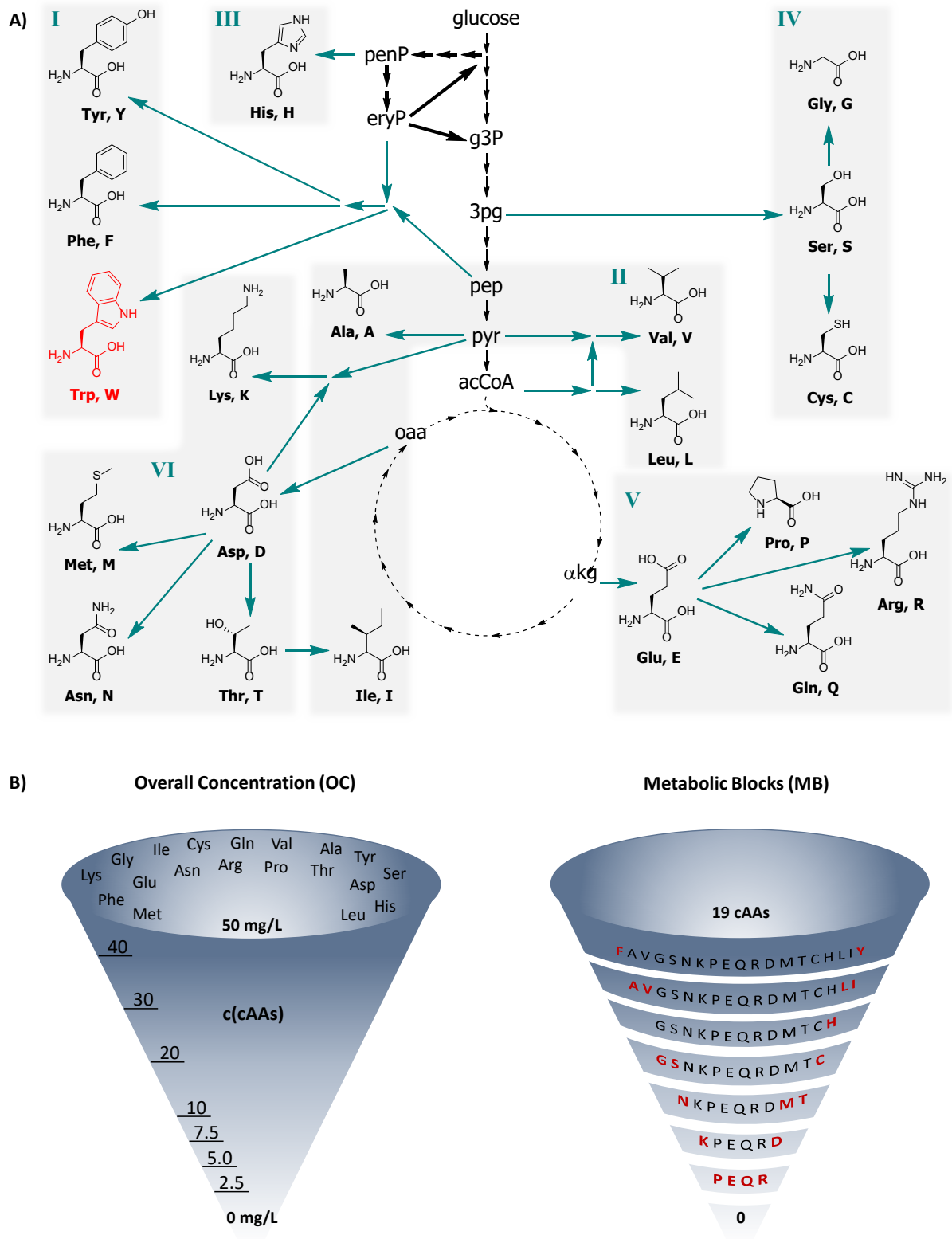
Figure 23 | General concept of cAA removal by OC and MB. A) Amino acid biosynthetic pathways (teal arrows) from the Embden–Meyerhof–Parnas pathway (plain arrows), the tricarboxylic acid (dashed arrows), and pentose phosphate cycles (bold arrows) of central metabolism. Note: tryptophan is shown in red because it is excluded from the cAA removal process. Abbreviations used for metabolites: ribose 5-phosphate (penP), erythrose 4-phosphate (eryP), glyceraldehyde 3-phosphate (g3p), 3-phosphoglycerate (3pg), phosphoenolpyruvate (pep); pyruvate (pyr), acetyl-CoA (acCoA), α-ketoglutarate (αkg), oxaloacetate (oaa). The figure was redrawn from Akashi *et. al.*[99] B) A graphical representation of the differences between OC and MB approaches for amino acid removal in ALE

phase 2 is shown. Overall concentration (OC): mixture of 19 cAAs whose concentration was gradually reduced from 50 mg/L to 0 mg/L. Metabolic blocks (MB): 19 cAAs were removed by identity (shown in red) as blocks or individually (in the case of the 7Fi ALE). This is a representative example (using 6TUB128-OC adaptation for the OC approach and 6TUB165-MB4 adaptation for the MB approach) but does not reflect the experimental details for all ALE experiments (shown below).

3.4 Adaptive Laboratory Evolution Experiments

Figure 24 illustrates an overview of all ALE experiments conducted in this study, including the path the different lineages took. After ALE phase 1 (indole depletion), the 6Fi and 7Fi adapted lineages separate due to the different approaches of amino acid removal (OC and MB) that were applied in ALE phase 2. Consequently, different, independent lineages arose that were separately analyzed and discussed below. Lineage names were assigned according to the pattern Ind-TUBX, W-TUBX, 6TUBX-OC, 6TUBX-MB4, 6TUBX-MB3, 7TUBX-OC, or 7TUBX-MB where "Ind" and "W" indicate the respective positive control, "6TUB" or "7TUB" indicate the respective fluoroindole adaptation, "X" indicates the passage number, and "OC" and "MB" indicate the cAA removal strategy used. The strain designation is also shown in **Figure 22**.
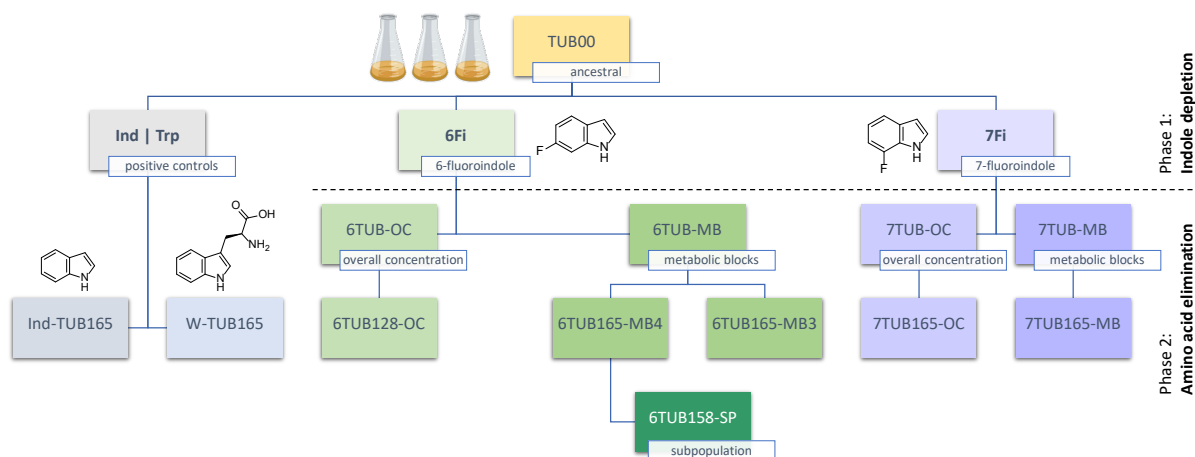
Figure 24| Overview of the evolution experiments and lineages. Eight adapted strains resulted from four independent parallel ALE experiments with TUB00 as metabolic prototype. Indole and Trp positive controls (gray panels) did not participate in the ALE concept (dynamic, progressive adjustment of selective pressure stringency). Adaptation to 6Fi (green panels) and 7Fi (purple panels) yielded either three or two lineages, respectively, depending on the details of amino acid removal (vide infra/supra). In addition, from the 6Fi adaptation, a subpopulation that showed a tendency to prefer 6Fi over Ind evolved (discussed below).

3.4.1 Adaptation of positive controls

A particularly important extension to previous Trp-based evolution experiments was the development of positive controls.[327,328] They allow a direct comparison of the adapted cells not only with the ancestral strain TUB00 but also with an appropriate control experiment at any time point of the evolution, i.e., cells that were likewise propagated under limiting conditions (minimal medium) and in a feeding-starvation environment.

The positive controls Ind-TUBX and W-TUBX (**Figure 25**) were developed by growth of TUB00 in NMM19-0-70 (supplemented with Ind or Trp, respectively), thereby inoculating new passages with OD_{600} = 0.02 whenever the cells reached stationary growth phase. However, the step of amino acid removal was omitted in these ALEs in order to keep the control strains in a general fashion (i.e., the different methods and routes for cAA removal applied in the Fi ALEs). Exactly as the actual Fi ALEs, the control strains were propagated over 165 passages, resulting in the final isolates Ind-TUB165 and W-TUB165.

Figure 25 | Cultivation schemes of the positive controls Ind-TUBX and W-TUBX. Adaptation of *E. coli* towards A) indole and B) tryptophan is shown. The optical density (OD_{600}) is plotted against the incubation time and the number of passages (reinoculation steps); the average OD_{600} is shown as bars, and individual clones are shown as circles. The corresponding counterparts (W-TUBX or Ind-TUBX in grey), as well as 6TUBX-OC (turquoise), 6TUBX-MB (green), and 7TUBX-OC (blue), 7TUBX-MB (yellow) lineages, are depicted as lines for the sake of comparability. The cells grew constantly in NMM19-0-70 supplemented with 19 cAAs (50 mg/L) and either 70 µM Trp or Ind as the target substrate.

The population density of both positive controls remained stable and high at each time point of propagation, and the individual clones showed scarcely any variance among themselves. There is no difference in the growth performance of W-TUBX and Ind-TUBX lineages as they grow to equal high OD_{600} of about 4. However, expectedly, their growth with native substrates (Ind, Trp) is clearly superior (more than twice as high) compared to that of the adapted strains 6TUBX-OC, 6TUBX-MB4/3, and 7TUBX-OC, 7TUBX-MB on the fluorinated indole as indicated by the respective lines in **Figure 25**.

3.4.2   Adaptation towards 6-fluoroindole

Overall, Trp-auxotrophic *E. coli* tolerated 6Fi remarkably well (**Figure 26**, **Figure 27** and **Figure 28**). In phase 1, cells responded sensitively to the challenge of indole depletion, as evidenced by strongly fluctuating OD_{600} values. However, after the populations ceased to be dependent on indole, cell growth remained stable and at a high level.
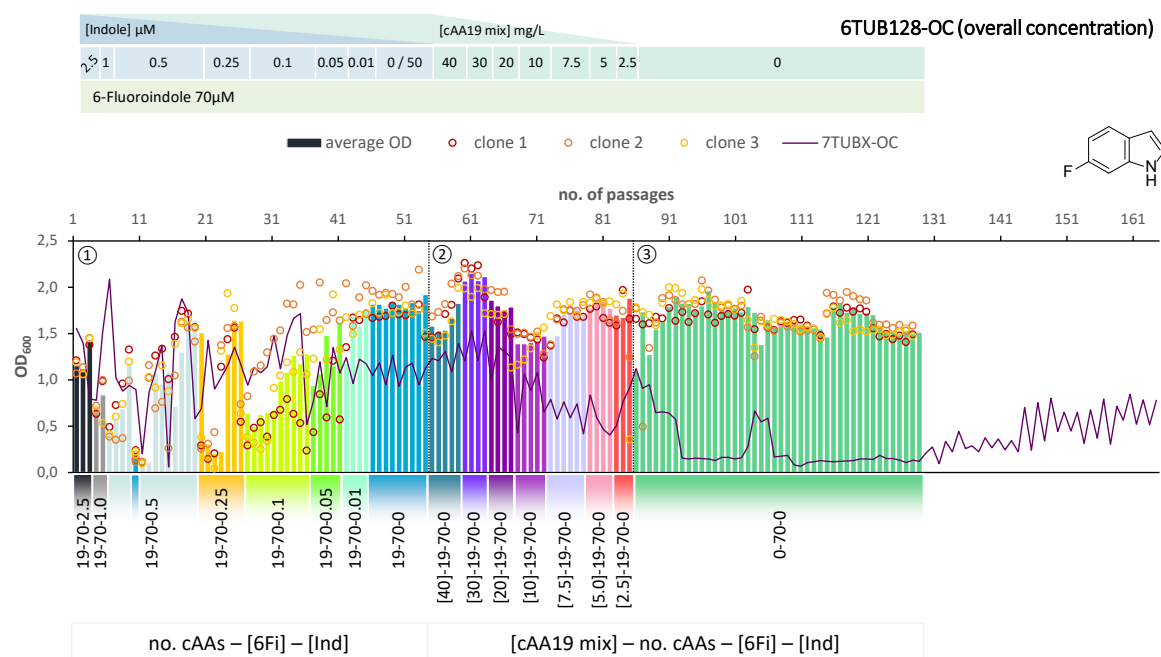
Figure 26 | Cultivation scheme of the 6TUB128-OC lineage. Optical density (OD_{600}) is plotted against the number of passages (reinoculation steps); average OD_{600} is shown as bars, and individual clones are shown as circles. The corresponding 7TUBX-OC counterpart is depicted as purple line. ALE phases ①, ②, ③, and medium composition are illustrated; in the OC approach, the concentration (mg/L) of the cAA19 mix is indicated in squared brackets [cAA conc.]-a-b-c.

Surprisingly, the OC and MB approaches give approximately equal results in terms of adaptation speed and success. Adaptation took place in a similar time frame and the different lineages show equally good growth behavior. The mutual phase 1 lasted 53 passages, whereas the cells adapted via the OC approach (**Figure 26**) took 31 passages in phase 2 and those for the MB approach (MB4 and MB3 cells) took 28 passages. When the amino acid supply in the MB-lineage was almost depleted (9 cAAs left, at passage 73), the removal of different metabolic blocks was tried, which yielded separate lineages 6TUB165-MB4 (**Figure 27**) and 6TUB165-MB3 (**Figure 28**). In 6TUB165-MB4, metabolic blocks were removed in the order "NMT", "KD" and "PEQR" ending up with four cAAs (MB4). In 6TUB165-MB3, the order was "PEQR", "KD" and "NMT" ending up with three cAAs (MB3). Although the cultivation schemes of the MB-lineages show a dramatic decrease in cell densities upon removing the amino acid blocks "AVLI" and "NMT" (MB4) as well as "PEQR" (MB3), even removing different blocks was useful here, and the cells recovered very quickly.

Figure 27 | Cultivation scheme of the 6TUB165-MB4 lineage. Optical density (OD_{600}) is plotted against the number of passages (reinoculation steps); average OD_{600} is shown as bars, and individual clones are shown as circles. The corresponding 7TUBX-MB counterpart is depicted as purple line. ALE phases ①, ②, ③, and medium composition are illustrated; in the MB approach, the amino acid composition is shown where red letters represent removed cAAs.

Figure 28 | Cultivation scheme of the 6TUB165-MB3 lineage. Optical density (OD$_{600}$) is plotted against the number of passages (reinoculation steps); average OD$_{600}$ is shown as bars, and individual clones are shown as circles. The corresponding 7TUBX-MB counterpart is depicted as purple line. ALE phases ①, ②, ③, and medium composition are illustrated; in the MB approach, the amino acid composition is shown where red letters represent removed cAAs.

ALE phase 1, for 6TUBX-OC and the two 6TUBX-MB lineages, comprised 53 passages, which means 301 generations in 128.8 days. The 6TUBX-OC lineage was propagated for a total of 128 passages (780 generations, 317.2 days). These cells completed phase 2 after 31 passages (205 generations, 77.2 days) and grew 44 passages (274 generations, 111.2 days) in the final adaptation medium (NMM0-70-0). In contrast, the two 6TUBX-MB lineages (6TUB165-MB4 and 6TUB165-MB3) grew significantly longer with 165 passages (1024 generations, 410.1 days). They ended phase 2 after 28 passages (179 generations, 71.7 days) and grew twice as long in the final adaptation medium (NMM 0-70-0) with 84 passages (544 generations, 209.6 days). An overview of the described ALE parameters is shown in **Table S24**.

3.4.2.1 Subpopulation screening

When screening for indole and fluoroindole compatibility, it was found that clones of both 6TUBX-MB lineages grew better in the presence of 6Fi than indole. Upon this finding, a simple experiment was designed to validate the growth behavior under different conditions in terms of growth matrix (solid, liquid) and substrates (Ind, 6Fi). For that purpose, population samples of clone 1 of passage 124 of the three 6Fi-ALEs (6TUB124-OC, 6TUB124-MB4, and 6TUB124-MB3) were chosen and spread out for clone isolation. 20 single clones of every lineage were picked, diluted in 20 µL water and 1 µL was spotted on two NMM19 agar plates containing either 6-fluoroindole or indole (19-0-70 (Ind) and 19-70-0 (6Fi)). Relaxed growth conditions by adding cAAs were used to provoke any growth (desired and undesired). Indeed, some clones of the MB lineages were found to reject indole entirely when grown on a solid medium (**Figure 29A**). This behavior was not observed in clones of the 6TUBX-OC lineage; thus, its cultivation was terminated after 128 passages (in contrast to the 165 passages of the 6TUBX-MB lineages).

The ultimate goal of the presented evolution experiments was to create organisms that are intransigently adapted to fluorine, i.e., to evolve cells that are not merely facultative FTrp users. To this end, five promising clones (29, 34, 35, 49, and 60) from the 6TUBX-MB strains were selected, and cultivation was continued under enhanced selection pressure by increasing the 6Fi concentration from 70 µM to 85 µM to 100 µM (subpopulation ALE, **Figure 29B**). The resulting lineages were designated with the suffix "SP", referring to subpopulation. Unfortunately, the previously indicated selectivity for 6Fi did not further increase even after 34 passages under these conditions. Clones derived from the subpopulation ALE as well as clones of the parent lineages (6TUBX-MB4 and 6TUBX-MB3) at different time points (passages 124, 141, and 150) were tested in liquid culture supplemented with either 6Fi or Ind (**Figure 29C**). When grown in a liquid medium, a remarkable difference was also observed in the optical densities achieved either with 6Fi or Ind, with a clear preference for the 6Fi-favoring phenotype. While this tendency was maintained in cells of passage 141 (85 µM 6Fi), it was almost lost with passage 150 (100 µM 6Fi). The increased selection pressure probably worked well up to 85 µM, but further increase to 100 µM did not produce the desired result of an unconditionally 6Fi-dependent phenotype. Further isolation and subpopulation screening did not reveal any new clones, suggesting that this phenotype has been lost. Nevertheless, in-depth analyses (genetic, transcriptomic, metabolomic level) may shed light on possible mechanisms that caused the initial trend.

A)

B) NMM(0-70-0) NMM(0-85-0) NMM(0-100-0)

C) 6-fluoroindole (NMM0-70-0) indole (NMM0-0-70)
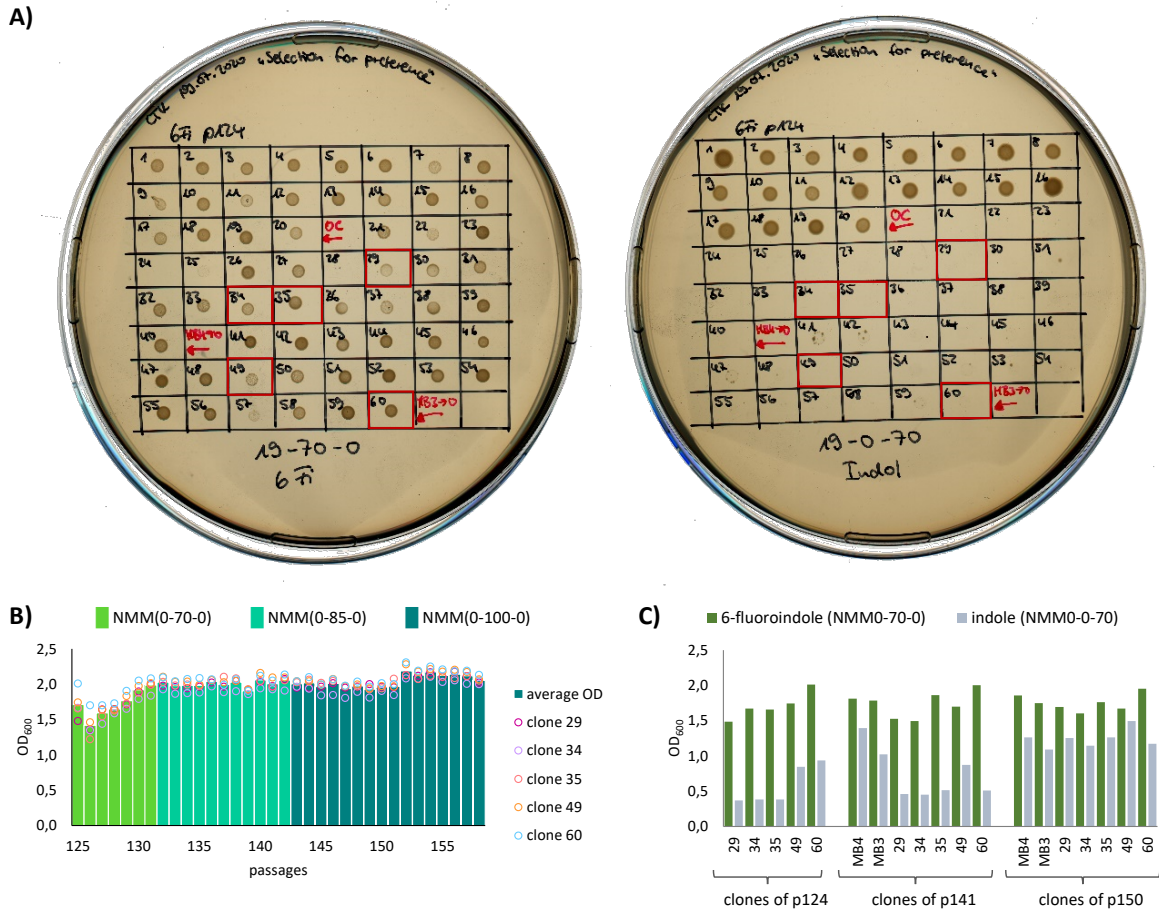
Figure 29 | Subpopulation screening for 6-fluoroindole preferring phenotype. A) Screening for 6Fi preferring clones by plate growth selection. Twenty individual clones of 6TUB124-OC, 6TUB124-MB4, and 6TUB124-MB3 were diluted and spotted on NMM19-70-0 (6Fi) and NMM19-0-70 (Ind) agar plates. The desired phenotype grew on the 6Fi-supplemented plate but not on the indole-enriched plate. Promising clones 29, 34, 35 (from 6TUB124-MB4) and 49, 60 (from 6TUB124-M3) are outlined in red and are labeled with the suffix "SP". B) Of the 60 clones tested, the five promising clones were selected and subjected to an additional adaptation experiment in which the selective pressure (6Fi concentration) was gradually increased from 70 µM to 85 µM up to 100 µM. C) Screening of the subset of promising clones (29, 34, 35, 49, 60) and parent clones of 6TUBX-MB4 and 6TUBX-MB3 in liquid culture supplemented with either 6Fi or Ind (NMM0-70-0 or NMM0-0-70). Different time points of subpopulation ALE were chosen with passage 124 (70 µM 6Fi), passage 141 (85 µM 6Fi), and passage 150 (100 µM 6Fi). However, the parent lineages (MB4, MB3) were kept constantly on 70 µM 6Fi, therefore allowing the effects of an increased 6Fi concentration from the starting point (70 µM) on progress to be distinguished.

3.4.3 Adaptation towards 7-fluoroindole

In contrast to 6Fi, tolerance towards 7Fi appears to be significantly more challenging, as indicated by the low population sizes especially, during the late adaptation process (**Figure 30** and **Figure 31**). Already in phase 1, a high clonal variation in the acceptance to the growth conditions is observed, but in sum similar high OD_{600} values were sustained as in the same phase of the 6Fi adaptation experiments, indicated by 6TUBX-OC/MB trace (green line) in **Figure 30** and **Figure 31**. In phase 2, cell density

decreased continuously in both lineages (OC and MB). It is hypothesized that this resulted from the low-grade toxicity of 7Fi that became increasingly difficult for cells to compensate for while undergoing cAA starvation. In the case of the MB lineage, alternatively, the removal of different amino acids was tried, as well as the conduction of replay experiments in order to investigate if the sudden drop in the cell density is reproducible (see section **3.4.3.1**). However, all such attempts ended with drastic reductions in cell population.

In order to rescue the cells, the concentration of 7Fi had to be reduced (first to 50 µM and then to 30 µM) while temporarily increasing the cAA supply, both starting with passage 84 (displayed in the legends of **Figure 30** and **Figure 31**). In the end, the cells slowly recovered upon the relaxed selection pressure and the concentration of 7Fi could again raised to 70 µM. Still, this process demonstrated that cAAs had to be removed extremely conservatively (compared to 6Fi): in smaller steps in the OC approach and not at all as metabolic blocks in the MB approach, but on an individual basis. Even then, not all clones survived the treatment, and the population sizes were considerably lower compared to the 6Fi adaptations. Taken together, this underlines the great challenge for *E. coli* to accept and adapt to 7Fi. Based on the described hurdles, it is difficult to detect a noteworthy difference between the OC and the MB approaches; however, from an experimental point of view, the OC approach was easier to implement.
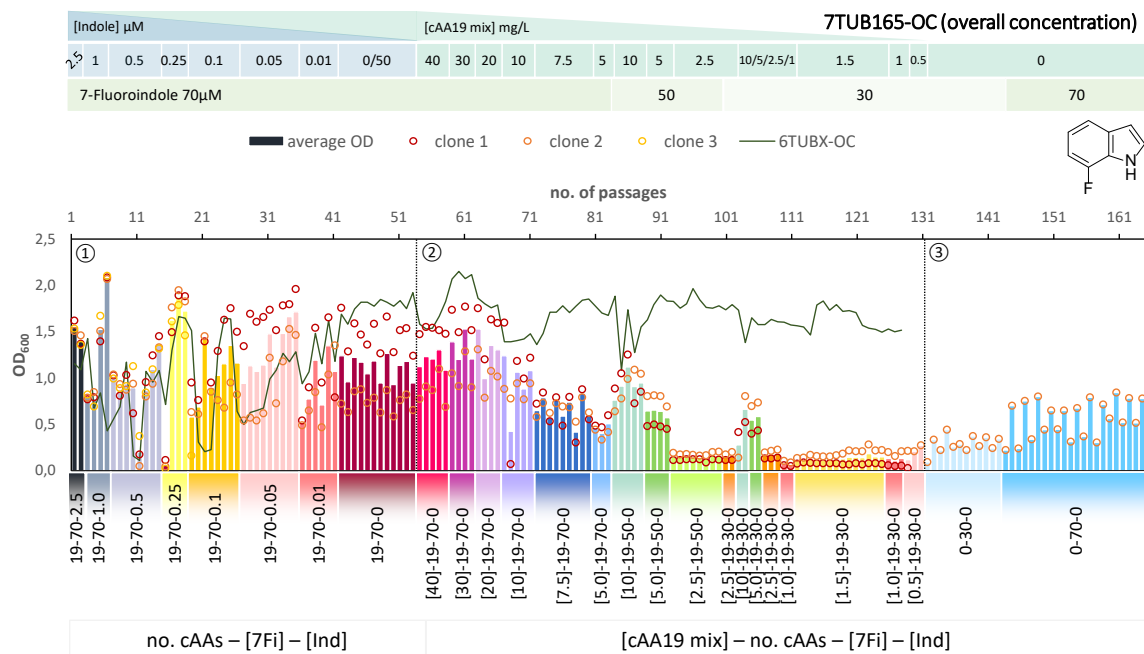
Figure 30 | Cultivation scheme of the 7TUB165-OC lineage. Optical density (OD_{600}) is plotted against the number of passages (reinoculation steps); average OD_{600} is shown as bars, and individual clones are shown as circles. The corresponding 6TUBX-OC counterpart is depicted as green line. ALE phases ①, ②, ③, and medium composition are illustrated; in the OC approach, the concentration (mg/L) of the cAA19 mix is indicated in squared brackets [cAA conc.]-a-b-c. Note the relaxed growth conditions by means of increased amino acid supply and reduced 7Fi concentration, depicted in the legends.

Figure 31 | Cultivation scheme of the 7TUB165-MB lineage. Optical density (OD$_{600}$) is plotted against the number of passages (reinoculation steps); average OD$_{600}$ is shown as bars, and individual clones are shown as circles. The corresponding 6TUBX-MB counterpart (mean of MB4 and MB3) is depicted as green line. ALE phases ①, ②, ③, and medium composition are illustrated; in the MB approach, the amino acid composition is shown where red letters represent removed cAAs. Note the relaxed growth conditions by means of increased amino acid supply and reduced 7Fi concentration, depicted in the legends.

The mutual phase 1 lasted 53 passages (301 generations, 122.1 days). The 7TUB165-OC lineage grew 79 passages (330 generations, 198.4 days) in phase 2, 33 passages (147 generations, 82.5 days) in phase 3 (12 passages (47 generations, 29.9 days) thereof with reduced 7Fi concentration in NMM0-30-0) and took 165 passages (778 generations, 403.0 days) for the entire adaptation process. A very similar trajectory was observed for 7TUB165-MB, whose development required 98 passages (397 generations, 243.9 days) in phase 2, 14 passages (54 generations, 35.0 days) in phase 3 (whereby a transition from 4 passages comprising 11 generations within 10 days in NMM0 30-0 was necessary); and comprises again a total of 165 passages (752 generations, 401.0 days). Since no further significant growth improvement was expected, the adaptation was terminated after only 14 and 33 passages in phase 3 (NMM0-70-0), respectively. An overview of the described ALE parameters is shown in **Table S24**.

3.4.3.1   Recapitulation of the "p72 event" by replay experiments

During the process of cAA removal, the cell densities of 7TUBX-OC and 7TUBX-MB significantly decreased. In the case of the 7TUBX-OC lineage, this behavior started approximately with growth in NMM[7.5]19-70-0 and became unacceptable in NMM[5.0]19-70-0 when the OD_{600} decreased to about 0.5 (7TUB81-OC to 7TUB83-OC). Then, a second and more drastic decrease occurred after increasing the selective pressure from NMM[5.0]19-70-0 to NMM[2.5]19-70-0 in that consequence, the 7Fi concentration had to be further reduced to 30 µM to keep the population alive. However, the 7TUBX-MB lineage suffered even more. Initially, the cells grew well in NMM12-70-0 (cAAs: G, S, C, K, D, N, M, T, P, E, Q, R) for 14 passages to OD_{600} of about 1, but after passage 72 (p72) the cell density dropped dramatically to OD_{600} 0.35 (7TUB72-MB) and to OD_{600} 0.1 in the following passages 73 - 77, without changing the medium composition. After the cells were initially rescued in NMM17-70-0 (addition of cAAs A, V, L, I, H in 7TUB78-MB, see purple bar in **Figure 31**), however, this behavior continued when they were again propagated in NMM12-70-0.

An attempt was made to reproduce this event for the 7TUBX-MB lineage (replay experiments). Two replay experiments were performed with cells from 7TUB66-MB and later also from 7TUB59-MB, which grew well in NMM12-70-0 (**Figure 32**). They were used as starting points, revived, and propagated again under the same conditions (NMM12-70-0) in order to investigate whether their adaptation proceeded parallel or divergent.
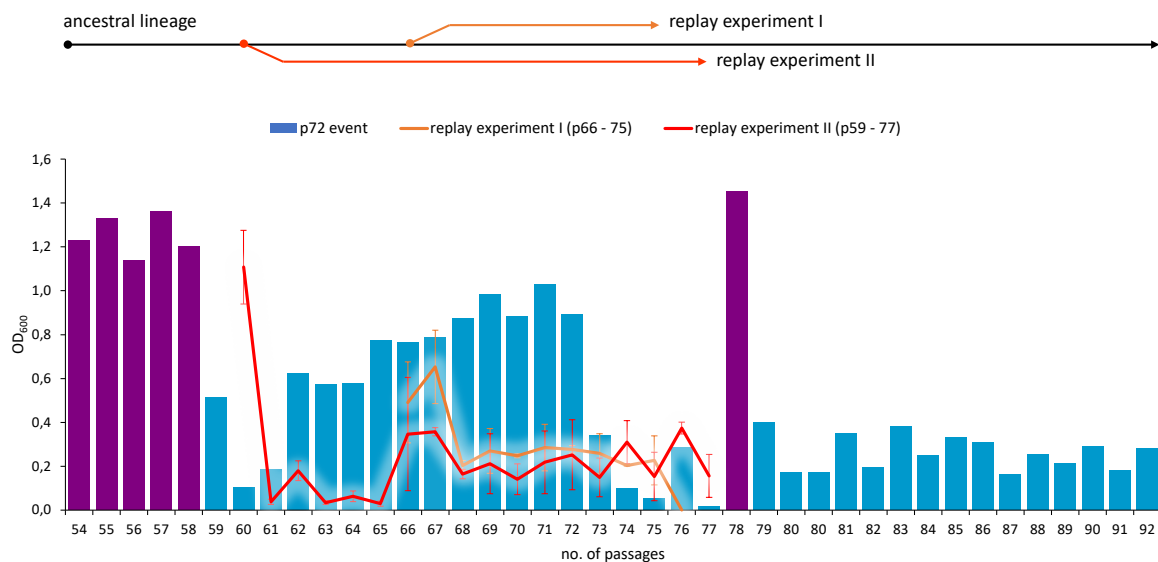
Figure 32 | Replay experiment of the "p72 event" of 7TUBX-MB. A part of the original growth data from 7TUBX-MB are plotted as bars: 7TUB54-MB to 7TUB58-MB as well as 7TUB78-MB grew in NMM17-70-0 (purple) and 7TUB59-MB to 7TUB77-MB as well as 7TUB79-MB to 7TUB92-MB grew in NMM12-70-0 (blue). In replay experiment I (orange), 7TUB66-MB was propagated for 9 passages in NMM12-70-0, and in replay experiment II (red), 7TUB59-MB (revived in NMM17-70-0) was propagated for 18 passages in NMM12-70-0. Growth data are illustrated as lines. Note: 7TUB66-MB propagated in replay experiment II grew just for once in NMM13-70-0 (His added) to restart or else stimulate growth.

In both replay experiments, the cell densities decreased, as observed during the 7Fi ALE, but this might not necessarily associated with passage 72 or a defined number of passages in NMM12-70-0. In contrast to the original growth data, here, there is no recovery phase observed before the cell densities dropped; rather they remained constantly at a low level. Also, neither the creation of more favorable starting conditions by usage of a particularly viable culture (7TUB59-MB in NMM17-70-0) nor the temporary supply of amino acids (7TUB66-MB in NMM13-70-0) could prevent the growth decline as demonstrated in experiment II. In sum, the replay experiments suggest that the "p72 event" did not occur by coincidence but was induced or provoked by an as yet unexplained cause. However, laboratory replay experiments between closely related lineages, as in the case of the presented experiment, often show repeatable outcomes.[373]

Whether evolution is driven by deterministic or contingent forces is still discussed.[373] Deterministic forces would implicate for more repeatable outcomes that are less dependent on past events and thus inevitable, whereas contingency suggests that evolutionary outcomes are contingent on specific events and subject to variations of an organism's history and, therefore, are less repeatable and predictable. Such events can be mutations, even those that initially appear to have no discernible effect but can initiate further changes. For example, Blount *et. al.* described that for the development of a citrate-using variant of *E. coli*, an ordinary mutation was involved, but that its phenotypic manifestation was contingent on prior mutations.[292]

3.4.4   Evaluation of the adaptation process (6Fi vs. 7Fi and OC vs. MB)

Figure 33 provides an overview of the trajectories of all ALE experiments conducted in this study. Overall, parallel serial cultures of seven different lineages were propagated over 165 passages and 752 - 1273 generations in approximately 400 days (**Figure 33A**), depending on the individual binary fission capacity of the different cells (see **Table S24** in appendix for details). The only exception is 6TUB128-OC, whose adaptation was prematurely terminated (see section **3.4.2**). Relinquishment of *E. coli* TUB00 from its indole dependency (ALE phase 1) proceeded very similarly regardless of whether 6Fi or 7Fi were the adaptation targets. Upon increasing the selective pressure and surely also the stress burden when canonical amino acids were removed in phase 2, however, remarkable differences between 6Fi and 7Fi ALEs were observable (yellow bars in **Figure 33A** and **Figure 33B**). The cAA removal in the 7Fi ALEs took longer, was more complex, needed more and discrete reduction steps (as described in section **3.4.3**). This applies to both the adaptation using the OC approach (18 steps) and the MB approach (20 steps), as shown in **Figure 33B**. Consequently, these lineages grew much shorter in their final medium composition (ALE phase 3) and are expected to be less adapted. In contrast, the cAA removal in the 6Fi ALEs proceeded much faster and required not even half the number of steps (7 – 8).
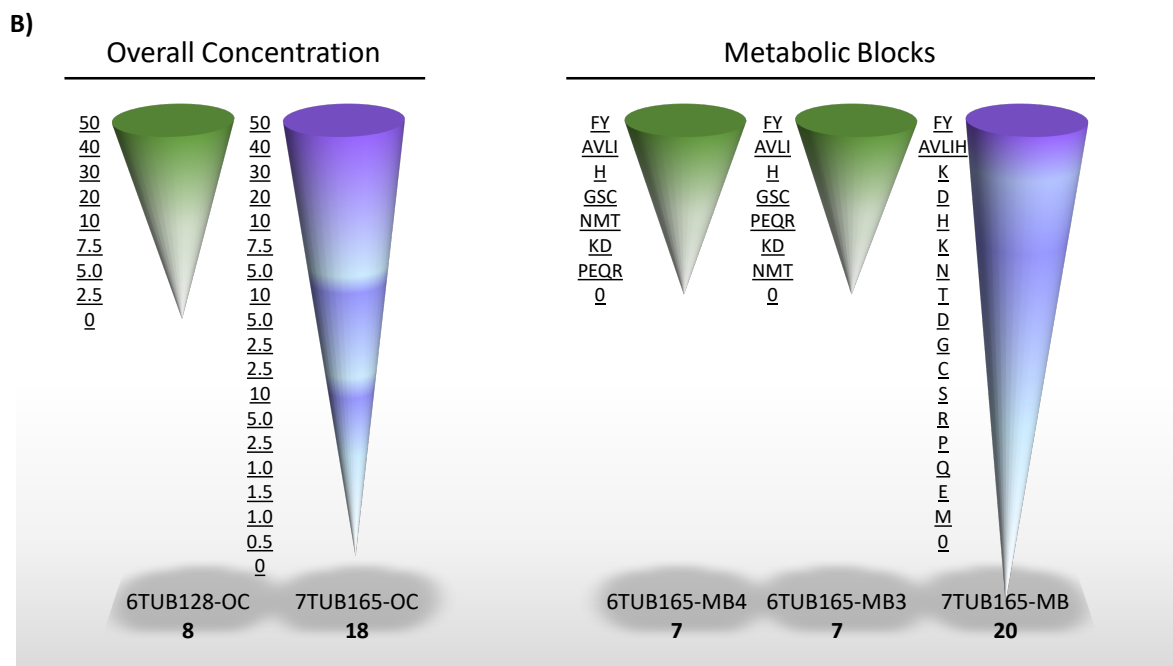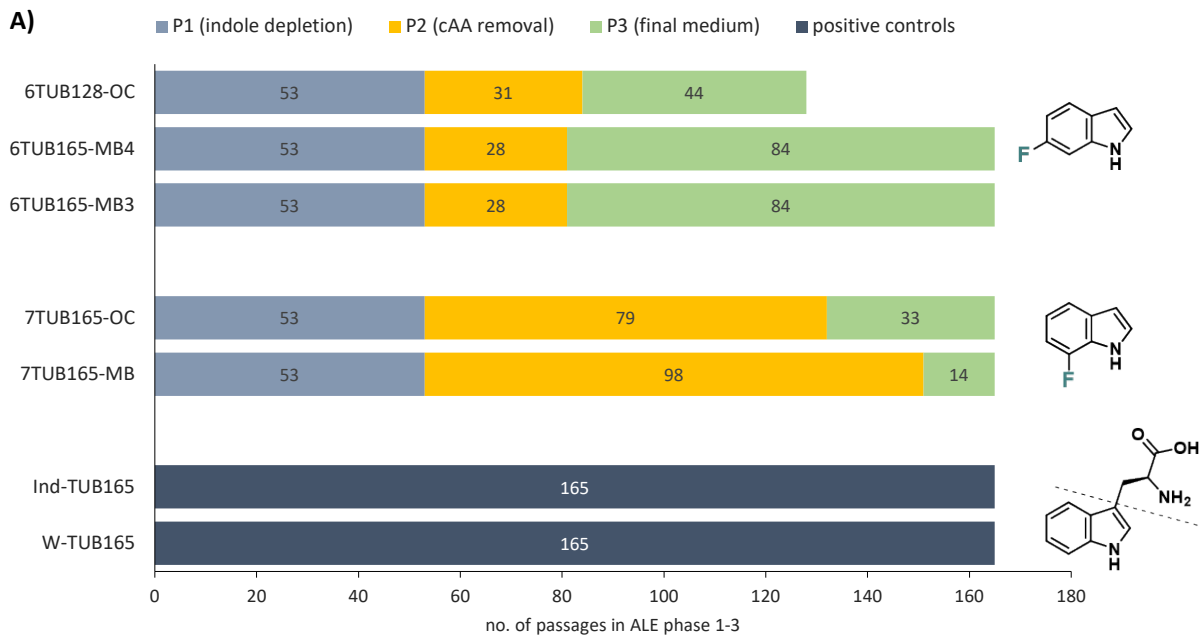
Figure 33 | Overview of the ALE trajectories. A) The extent of the ALE phases is illustrated by means of passages needed to complete each adaptation phase (1, 2) and the number of passages spent under final growth conditions (phase 3). The indole depletion (phase 1, grey) took 53 passages for each ALE. The cAA removal (phase 2, yellow) was accomplished in about 30 passages for the 6Fi ALE as well as 79 and 98 passages, respectively, for the 7Fi ALE. And the 6Fi-adapted cells were grown for 44 or 84 passages in their final medium composition (phase 3, green), whereas the 7Fi-adapted strains were held only 14 or 33 passages under final adaptation conditions. The positive controls were not subjected to the ALE phase concept but cultivated continuously for equally 165 passages in NMM19-0-70 (Ind or Trp). B) Applied routes of the cAA removal for the adapted strains. The 6TUBX strains are colored green, and the 7TUBX strains are colored purple. For the OC approach, the course of the cAA concentration decrease is shown on the left, and for the MB approach the different cAA blocks or single cAAs removed in each step are shown on the right. In the case of 7TUB165-OC and 7TUB165-MB adaptation, the amino acid supply had to be temporally increased to keep the cells viable (indicated by different purple shades).

6Fi-adapted cells showed a fairly good and robust growth behavior, even similar to that of the positive controls indicated by the number of generations per passage (**Table 2**). In contrast, the 7Fi adapted lineages grew fewer generations and also showed greater variance.

Table 2| Number of generations per passage (gen./passage) of adapted lineages.

positive controls		6Fi adapted lineages		7Fi adapted lineages	
lineage	gen./passage	lineage	gen./passage	lineage	gen./passage
Ind-TUBX	7.70 ±0.11	6TUBX-OC	6.04 ±0.77	7TUBX-OC	4.72 ±1.27
W-TUBX	7.71 ±0.11	6TUBX-MB4	6.17 ±0.83	7TUBX-MB	4.56 ±1.32
		6TUBX-MB3	6.17 ±0.85		

3.5   Growth characteristics and cell viability

Growth traits serve as an excellent indicator of adaptation success and enable assessment of the phenotype and bacterial fitness of the adapted cells. To this end, extensive growth experiments of all strains in different media reflecting the conditions along the adaptation trajectory were conducted. In addition, the tolerance of adapted cells to chemically distant analogs and to polyfluorinated peptides was also probed.

3.5.1   Growth behavior of the adapted strains along the evolution trajectory

Isolates of the ancestral strain TUB00 and of the final isolates of W-TUB165, Ind-TUB165, 6TUB128-OC, 6TUB165-MB4, 6TUB165-MB3, 7TUB165-OC and 7TUB165-MB were subjected to comprehensive growth analyses. For this purpose, growth curves in different medium compositions along the adaptation course were recorded (**Figure 34**, section **5.2.5**). These comprised LB (rich, undefined medium), NMM19-0-70 (Ind), NMM19-70-1 (6Fi or 7Fi, Ind), NMM19-70-0 (6Fi or 7Fi) and NMM0-70-0 (6Fi or 7Fi).
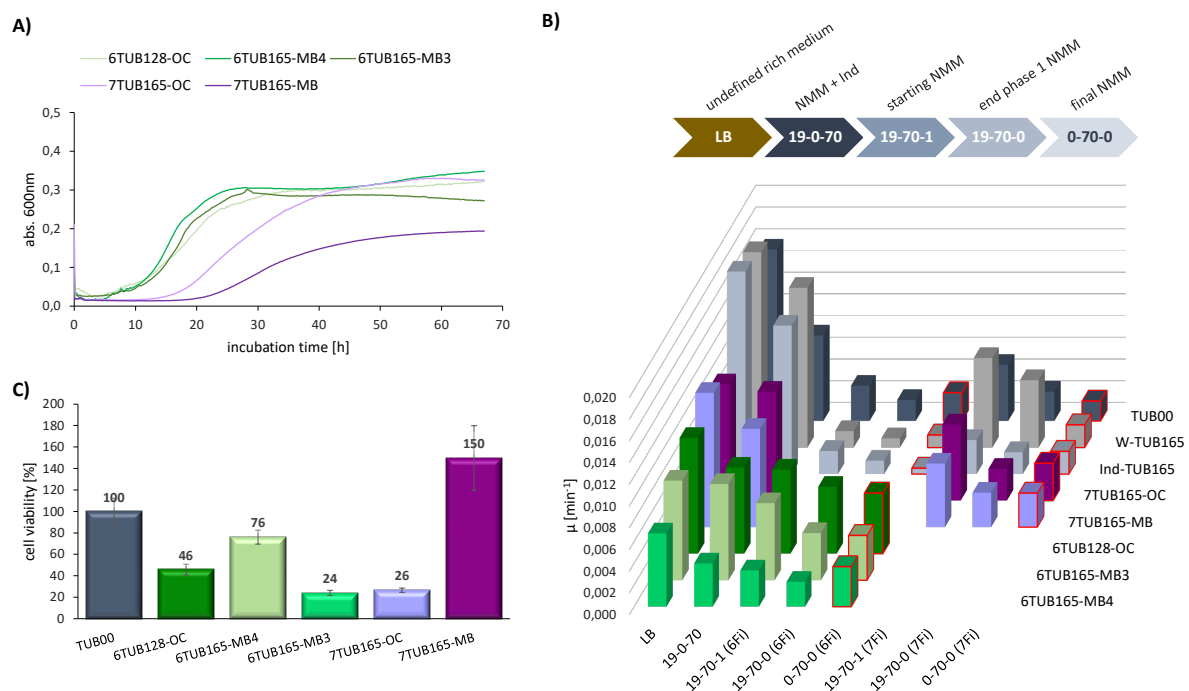
Figure 34| Growth traits of the adapted strains (6TUB128-OC, 6TUB165-MB4, 6TUB165-MB3, 7TUB165-OC, 7TUB165-MB and W-TUB165, Ind-TUB165) in comparison to the ancestral TUB00. A) Entire growth curves of the final isolates of 6Fi-adapted (green) and 7Fi-adapted (purple) strains measured by growth in their final medium composition NMM0-70-0 (70 µM 6Fi or 7Fi, respectively). B) The specific growth rate (µ) was calculated from the slope of the exponential phase using the GrowthRates program[374]; values obtained in the final medium composition are outlined in red, and the characteristics of the different growth media are depicted above. C) Cell viability was determined using the CCK-8 assay (Cell Counting Kit-8); values of fluoroindole-adapted strains were normalized to TUB00 (set 100 %). All measurements were performed at a microscale level in triplicates using a plate reader.

Overall, the impression of the adaptation success gained from the cultivation schemes (sections **3.4.2** and **3.4.3**) can also be seen in the growth curves (**Figure 34A**). The growth performance of the 6Fi-adapted cells is significantly better compared to the 7Fi-adapted cells. The final isolates of the 6Fi-adapted cells exhibit a very similar growth behavior in their final growth medium (NMM0-70-0), although 6TUB165-MB4 appears to be the best-adapted lineage. In contrast, the growth behavior of cells adapted to 7Fi is markedly different. These cells exhibit a significantly increased lag time and decreased growth rate, which is more pronounced in strain 7TUB165-MB, which also has a lower maximum OD_{600} value.

Furthermore, **Figure 34B** provides an overview of the growth of all strains studied along the adaptation course given by the calculation of the specific growth rate[374] (for a comprehensive view, see detailed growth curves in **Figure S65A-H** in the appendix). In general, the growth correlates with the concentration of provided nutrients. Interestingly, the ancestral strain TUB00 tolerates 7Fi better than 6Fi, although 7Fi was considered to be more toxic based on the cultivation schemes (**Figure 30**, **Figure 31**). The growth performance of the positive controls in NMM19-0-70 approaches that in LB (see curves in **Figure S65B-C**), which is comprehensible since these strains were adapted to grow on minimal medium. On the other hand, the fluorinated indoles remain toxic or at least inhibiting to the positive controls. The described effects of the fluorinated indoles might be explained by their chemical nature. While 7Fi is metabolically more inert because of its chemical distance from native indole, 6Fi is more similar to indole, and therefore, it can affect the intracellular metabolism of TUB00 as well as the positive controls more.

Among the 6Fi-adapted cells, the OC lineage exhibits the highest specific growth rate in all tested media. However, as with the MB3 lineage, no further growth improvement is observed after transitioning from NMM19-70-0 to NMM0-70-0. Remarkably, the MB4 lineage shows such improved growth behavior. Although its growth rates are lower compared to the other two lineages, the highest growth rate in 6Fi-supplemented NMM is observed in the final medium NMM0-70-0, which is only surpassed by growth in the ancestral medium. This finding is even more evident in the entire growth curves (see **Figure S65D-F**). This observation strongly suggests that they are approaching a fluoroindole-preferring phenotype, which is also supported by the subpopulation screening results.

Both 7Fi-adapted strains show again a reduced growth rate with a concomitant reduction in nutrient supply (cAAs, Ind). Indole-enriched media enable improved growth in the absence of 7Fi (NMM19-0-70 > NMM19-70-1), which is a clear signal for an incomplete adaptation, as these cells still strongly prefer indole over 7Fi. However, growth performance in 7Fi-supplemented medium with and without the addition of cAAs (NMM19-70-0, NMM0-70-0) is equal, indicating that the removal of canonical amino acids is well accepted, which seems surprising considering the challenging process that was required to adapt to these conditions.

In addition, the picture of bacterial fitness was completed by determining the cell viability using a CCK-8 assay (**Figure 34C**, see section **5.2.6**), which measures the bio-reduction of tetrazolium salts (WST-8).[375] The viability of all 6Fi-adapted lineages correlates very well with the growth parameters, whereas among the 7Fi-adapted strains, a discrepancy was found in 7TUB165-MB. This strain shows a cell viability of 150 % compared to TUB00, although the growth curves and parameters would suggest a considerably lower value.

3.5.2 Growth behavior on chemical distant analogs

Furthermore, the growth of 6Fi-adapted cells on chemically more distant indole analogs was investigated.

According to one theory, the eventual difference between the analog and the original substrate must be substantial for the use of the substitute to stick for a long-term.[219] Schmidt and Kubyshkin developed a metric (Δcode) for genetic code engineering, where polarity indices ($clogD_7$; computed distribution of a substrate between octanol and aqueous buffer with pH 7) were assigned to amino acids to quantify the distances between pairs of genetic codes;[376] for example, UGG translation into Trp in the standard genetic code vs. UGG translation into 6FTrp or 7FTrp in the evolved genetic code presented here. In doing so, their approach used polarity as a distance parameter because the replacement of a polar amino acid with a nonpolar one (and vice versa) leads to major disturbances in the proteome; or, in other words, the higher the difference between the polarities, the higher the contribution to the overall genetic code distance.[377] In this context, they also proposed replacement strategies, including stepwise recoding for the genetic isolation of organisms. According to the low Δcode of FTrp caused by low polarity differences compared to Trp, they hypothesized that the substitution would not or even never lead to complete alienation of the ancestral strains; but instead, a further stepwise alienation of FTrp adapted strains with even more hydrophobic Trp analogs (with increasing Δcode; see **Figure 35A**) could be carried out in order to adapt non-facultative Trp analog using organisms.[376]

Thus, the 6FTrp using strains should exhibit a certain promiscuity to either other mono-fluorinated indoles (4Fi, 5Fi, 7Fi) or other indole analogs substituted at position 6, such as 6Me-Ind, 6OH-Ind, 6Br-Ind and 6CF$_3$-Ind. To test this, 6TUB165-MB4 and 6TUB165-MB3 cells were used since they appeared to be best adapted, and 6TUB002 as control (**Figure 35B**, section **5.2.5.1**). This very early passage of the 6Fi ALE was chosen instead of the ancestral TUB00 because it already experienced the 6Fi imposed stress (see sublethal-injury description in section **3.2.3**).
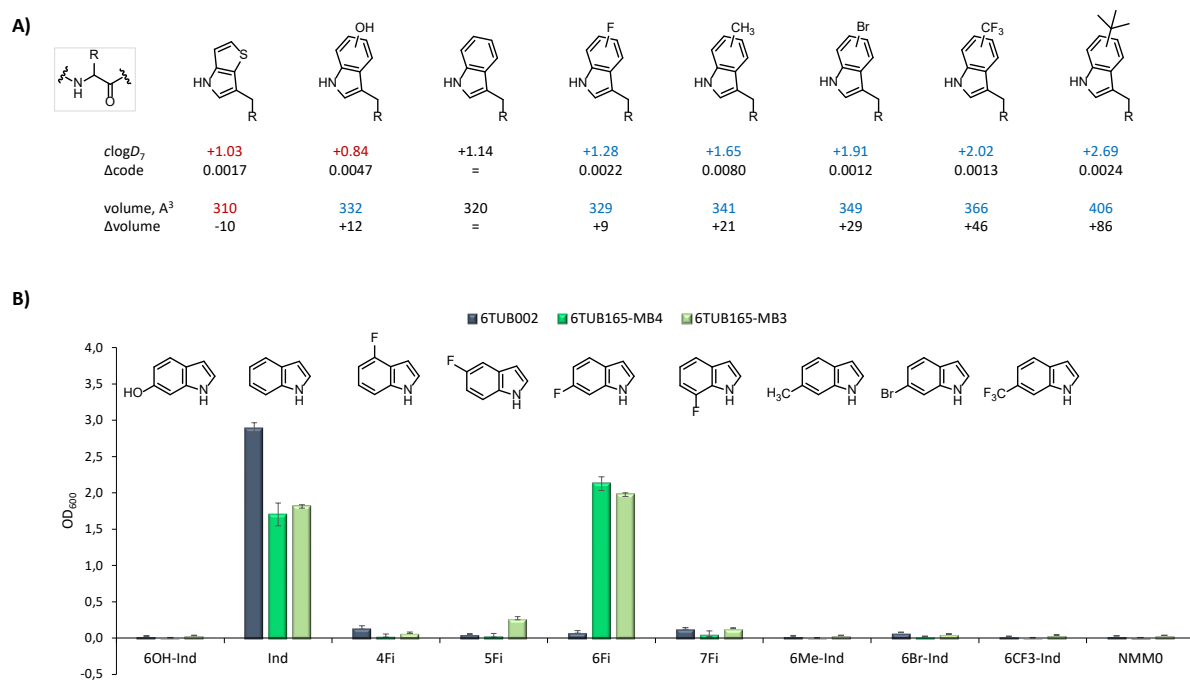
Figure 35 | Tolerance of 6Fi-adapted strains to chemical distant indole analogs. A) Structures and properties of different indole analogs Tpa, hydroxy-Ind, Ind, fluoro-Ind, methyl-Ind, bromo-Ind, trifluoromethyl-Ind, tert-butyl-Ind (from left to right). Indole substituents increase the volume A^3, but they can both increase (F, Br, CH$_3$, CF$_3$, tBu) and decrease (OH) the lipophilicity (clogD$_7$). In contrast, Tpa was used for ALE by Hoesl et. al.[327] decreases both volume and lipophilicity. Properties were calculated by Vladimir Kubyshkin according to Schmidt et. al.[376] B) Growth of 6TUB002, 6TUB165-MB4, and 6TUB165-MB3 with selected chemical distant analogs (structures are shown above, respectively). Cells were cultivated NMM0 supplemented with 70 µM of the respective indole analog; NMM0 was used as a negative control. Growth data correspond to 48 h incubation at 30 °C, 200 rpm.

6TUB002 grew well only in NMM0-0-70 (Ind) and in NMM0-70-0 supplemented with monofluorinated indoles (4Fi, 5Fi, 6Fi, 7Fi) as well as 6Br-Ind at least marginal growth was observed. The adapted lineages 6TUB165-MB4 and 6TUB165-MB3 reached high cell densities in NMM supplemented with either Ind or 6Fi, and 6TUB165-MB3 also showed little growth in NMM0-70-0 (5Fi > 7Fi). The growth of the 6TUB002 control only in indole-containing NMM, and the growth of adapted strains in Ind as well as 6Fi containing NMM was expected. However, a certain tolerance of the adapted strains was also expected for 4Fi, 5Fi, and 7Fi since a similar behavior already was reported for FTrp-adapted E. coli[343] and B. subtilis.[345]

However, the tested cells generously rejected all chemical, more distant indole analogs (based on the proposed Δcode metric). Nevertheless, it should be noted that in this test, only the ncAA was fed as a Trp source, and this does not in any way preclude a possible adaptation according to the ALE setup presented here, i.e., gradual reduction of the natural substrate. In the end, adaptation to chemically more distant analogs may be more difficult and require more cellular changes, but conversely, these also ensure that the formerly natural substrate would be rejected.

3.5.3 Growth behavior on polyfluorinated peptides

"Fluorine meets Fluorine." The response of fluorine-adapted organisms to highly fluorinated peptides has never been probed. However, with regard to PFAS bioremediation efforts, it will be very interesting to observe how the cells that evolved in this study respond to such highly fluorinated molecules.

Chowdhary *et. al.* designed and explored a series of self-assembling, amphipathic polyfluorinated peptides consisting of an alternating sequence of lysine (Lys, K) and the non-proteogenic amino acid α-aminobutyric acid (Abu).[378] There are the 16-meric peptides AbuK16 and three derivatives MfeGlyK16, DfeGlyK16, TfeGlyK16 generated by incorporation of the fluorinated amino acids (2*S*)-4-monofluoroethylglycine (MfeGly), (2*S*)-4,4-difluoroethylglycine (DfeGly) and (2*S*)-4,4,4-trifluoroethylglycine (TfeGly) in place of Abu. The peptide motif and the chemical structures of Abu and its derivatives MfeGly, DfeGly, and TfeGly are shown in **Figure 36A**. The high degree of fluorination in these peptides was shown to promote fibrillation, and that they form physical hydrogels. Hydrogels formed by self-assembling peptides can even serve as 3D-cell culture substrates for cell proliferation.[379,380]

The fluorine-adapted strains 6TUB165-MB4 and 7TUB165-MB, as well as TUB00 (as control), were treated with the members of the described peptide library, and their tolerance was assessed by measuring the cell viability using CCK-8 assay (section **5.2.5.2**). On TUB00 (**Figure 36B**), the non-fluorinated peptide AbuK16 exerts a significantly toxic effect (48.4 % ± 1.1), whereas the fluorinated variants are rather tolerated by this cell line. The effect of DfeGlyK16 (75.6 % ± 12.5) is still considered moderate, and those of MfeGlyK16 (105.5 % ± 0.1) and TfeGlyK16 (97.1 % ± 16.3) can even be rated as neutral. On the contrary to the steady values of TUB00, the fluorine-adapted strains experience more sensitivity to the peptide treatment, in particular, implied by the varying cell viability along the peptide dilution series. For 6Fi-adapted cells 6TUB165-MB4 (**Figure 36C**), only a noticeable toxic effect for DfeGlyK16 (77.7 % ± 5.4 to 137.5 % ± 3.7) was observed, which is indicated by the increasing cell viability accompanied by decreasing peptide concentration. Interestingly, an opposite effect was found for MfeGlyK16 (112.4 % ± 12.0 to 69.0 % ± 4.4) and TfeGlyK16 (120.1 % ± 3.3 to 83.7 % ± 2.3), where the viability decreased analogously with decreasing concentration. AbuK16 (93.6 % ± 2.7 to 126.0 % ± 12.3), however, exerts no to light growth-promoting effect on this cell line. However, it must be mentioned that the high viabilities (> 100 %) obtained in this cell line cannot be explained at the moment. In contrast, 7Fi-adapted *E. coli* 7TUB165-MB (**Figure 36D**) are more susceptible to the growth perturbing traits of the polyfluorinated peptides, with the exception of TfeGlyK16. While AbuK16 appears to exert little to no toxic effect (83.9 % ± 7.5 to 113.0 % ± 8.9), more pronounced effects are seen with increasing levels of fluorination, namely with MfeGlyK16 (53.3 % ± 12.1 to 100.2 % ± 7.5) and DfeGlyK16 (57.6 % ± 16.7 to 121.5 % ± 12.6). However, this does not apply to TfeGlyK16 (107.0 % ± 9.3), which, in turn, does not affect cell viability.
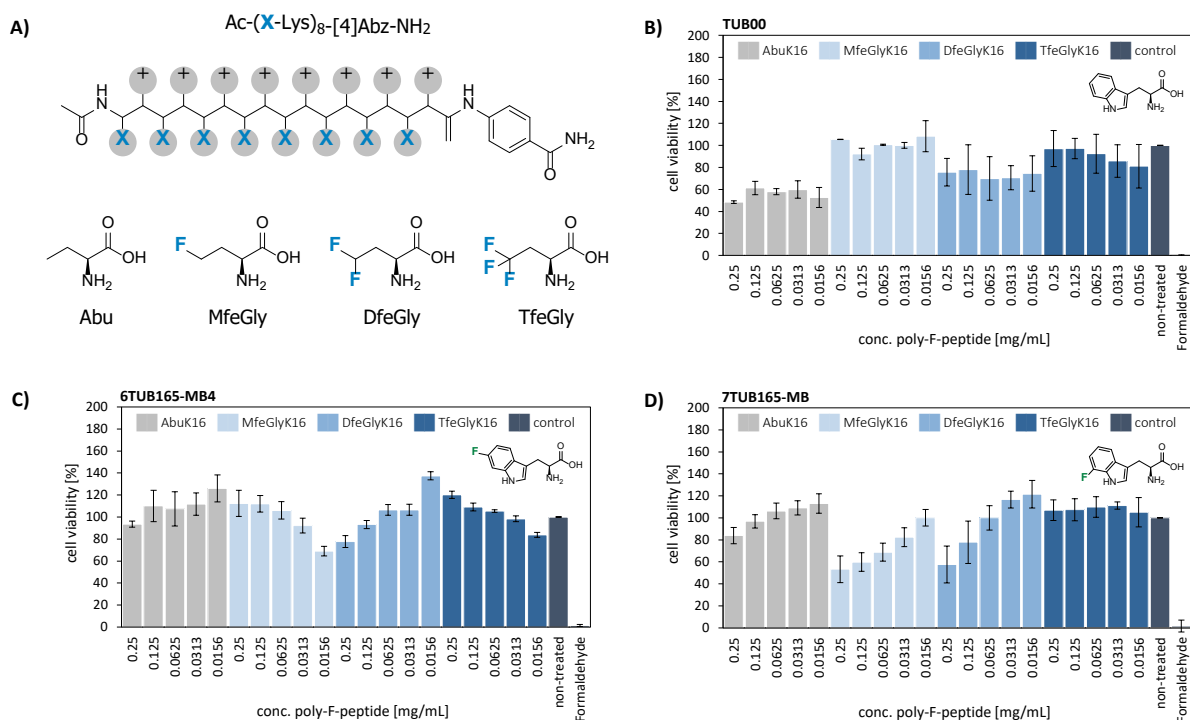
Figure 36| Tolerance of fluorine-adapted strains towards polyfluorinated peptides. A) Peptide motif (AbuK16, MfeGlyK16, DfeGlyK16, TfeGlyK16) and chemical structures of Abu and its fluorinated derivatives MfeGly, DfeGly and TfeGly.[378] B) – D) Cytotoxicity assays of TUB00, 6TUB165-MB4, and 7TUB165-MB. The cell viability was determined after 24 h peptide treatment (in dilution series) of either 5×10^4 cells/mL (TUB00) or 5×10^6 cells/mL (6TUB165-MB4, 7TUB165-MB) and monitored by absorbance measurement at 450 nm. Values were normalized to the viability of respective non-treated cells (positive control), and formaldehyde was used as negative control. Cytotoxicity measurements were performed with three independent biological replicates and repeated three times. Errors are derived from the standard deviation.

Cationic amphipathic peptides are often reported with antimicrobial activity because electrostatic interactions between the cationic side chains of the peptide and the negatively charged bacterial membrane cause disruption of the membrane integrity[381,382] and fluorination of such peptides can influence its antimicrobial activity.[383]

Overall, no trends along the concentration gradient of peptides were observed in TUB00, which may indicate a certain robustness of the cells. In contrast, the fluorine-adapted cells (6TUB165-MB4 and 7TUB165-MB) were more fragile or susceptible to the peptide treatment. Here, tendencies are discernible, although these do not follow any specific or obvious recognizable pattern. As is often the case, fluorine-associated effects are scarcely predictable, particularly in a biological context.[84] Of course, fluorine-caused effects on fluorine-containing cells are no exception. Remarkably, the position of the F-substituent in tryptophan (6F or 7F) seems to influence the response of cells to these peptides, which again illustrates the complexity of the interplay in biological systems.

Furthermore, the cytotoxic effects of AbuK16, MfeGlyK16, DfeGlyK16, and TfeGlyK16 on TUB00, 6TUB165-MB4, and 7TUB165-MB were also estimated by an agar well diffusion assay (**Figure 37**, section **5.3.2**). In doing so, the cells to be tested were inoculated in an agar nutrient medium, and putative antimicrobial agents given by the described peptides, and two control antibiotics were filled into punched holes. After incubation, zones with no growth indicate an inhibiting effect of the respective putative antimicrobial agents. Although the assay did not detect any antimicrobial activity against the peptides, it is interesting to note that, in particular, 7TUB165-MB is more sensitive to the control antibiotics, indicated by greater inhibition zones. Both antibiotics are small, hydrophilic compounds and require porin channels to enter *E. coli*'s outer membrane;[384,385] however, they have different modes of action. While chloramphenicol targets the bacterial ribosome by inhibiting peptide bond formation and thus protein synthesis by blocking the peptidyl transferase,[386] the β-lactam antibiotic ampicillin intervenes in the synthesis of cell wall peptidoglycan (see **3.9**) by binding of penicillin-binding proteins (PBP's).[387] Although genomic analyses (section **3.8**) revealed changes in genes associated with cell membrane and membrane proteins, none of these mutations is shared by both strains. In 7TUB165-MB, mutations were found in the outer membrane porin C (*ompC*; Arg58Cys) and in the phosphatidylglycerophosphate synthase (*pgsA*; Leu11Arg). In 6TUB165-MB4, the genes encoding for lipid A disaccharide synthase (*lpxB*; Ala36Thr), phospho-ethanolamine transferase (*eptB*; Gly307fs), and in lipopolysaccharide core biosynthesis protein (*rfaS*; Ile116Leu and Asn109Tyr) were affected. (Only missense mutations are listed.)
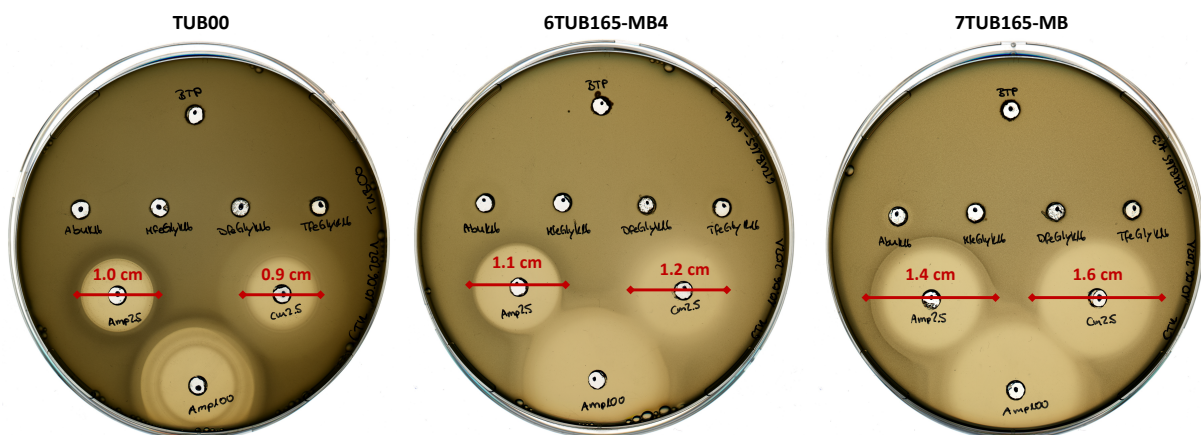
Figure 37 | Agar well diffusion assay for estimation of antimicrobial activity of polyfluorinated peptides. Cells were grown in LB agar and polyfluorinated peptides AbuK16, MfeGlyK16, DfeGlyK16, and TfeGlyK16 (2.5 mg/mL, each), control antibiotics ampicillin (2.5 mg/mL, 100 mg/mL) and chloramphenicol (2.5 mg/mL) and BTP buffer (negative control) were added in punched holes. After 48 h incubation at 30 °C, growth inhibition zones are only observed for the antibiotics Amp and Cm.

3.6 Global substitution of Trp by FTrp

In case of a complete adaptation, every tryptophan in the cells has been replaced by its fluorinated analogs 6FTrp or 7FTrp, after all over 20000 positions. To prove this, the proteomes of the adapted cells were studied in more detail by amino acid analysis and proteomics. Ideally, neither TUB00 nor the positive controls W-TUB165 and Ind-TUB165 should contain any traces of FTrp or Fi because they were never subjected to the fluorinated substrates. Likewise, in the adapted strains 6TUBX and 7TUBX, only FTrp or Fi should be detectable, otherwise the presence of non-fluorinated Trp or indole would hint a contamination that could have been interfering effects on the adaptation process, as described by Bacher and Ellington.[343] With the complete replacement of all Trp processed into proteins, it can also be assumed that Trp present in the metabolism was also replaced. An attempt to analyze Trp and FTrp as small molecules in the metabolic context of 6TUB161-MB4, 7TUB161-MB, and TUB00 yielded at least implications for a complete substitution (see section **3.7.1**).

3.6.1 Determination of Trp and F-Trp content by HPLC (FL/UV)

The content of Trp, 6FTrp, and 7FTrp and its corresponding indole precursors Ind, 6Fi, and 7Fi was determined by the classical amino acid analysis approach. Proteomic samples were hydrolyzed (see section **5.9.1**) and analyzed by HPLC coupled with UV and fluorescence (FL) detection (measurements were performed by Carlo Fasting at Freie Universität Berlin). With the used method, a good separation of these chemically similar target substances Ind, 6Fi, 7Fi, and Trp, 6FTrp, 7FTrp, as well as high sensitivity, particularly with FL detection, could be achieved. However, 7Fi shows quenched fluorescence,[388] and 7FTrp shows decreased fluorescence intensity,[389] making their analysis more difficult since the sensitivity of UV detection was considerably lower. Proper separations and FL detection resulted in the following detection limits (LOD): Trp and 6FTrp 1 pmol, Ind and 6Fi 10 pmol, 7Fi 500 pmol (UV detection) and 7FTrp 100 pmol.

The supernatant of *E. coli* MG1655 wild-type cells grown in LB medium showed a maximum of 0.5 – 1 mM indole in the exponential growth phase and up to 60 mM in the stationary growth phase;[116] and wild-type *E. coli* W3110 produced 0.588 µM Trp (0.12 mg/L).[390] However, TUB00 used in this study is not able to produce indole due to the chromosomal deletions of *trpLEDC* (indole biosynthetic genes) and *tnaA* (Trp degradation gene) but is exclusively dependent on exogenous indole/fluoroindole supply (70 µM). The amino acid analyses carried out revealed Trp concentrations far below the expected values (summarized in **Table 3**), but this was most likely caused by an unsuitable experimental setup, as discussed below in detail.

Table 3 | Determination of Trp, 6FTrp, and 7FTrp in the final ALE isolates.

hydrolyzed sample	Trp [µM], 4.57 min	6FTrp [µM], 7.20 min	7FTrp [µM], 7.15 min
TUB00	0.112 ± 1.90E-03	n.d.	n.d.
W-TUB165	0.174 ± 5.84E-03	n.d.	n.d.
Ind-TUB165	0.138 ± 1.44E-04	n.d.	n.d.
6TUB128-OC	0.214 ± 3.23E-03	4.07 ± 8.59E-03	n.d.
6TUB165-MB4	0.142 ± 5.20E-03	4.89 ± 7.54E-02	n.d.
6TUB165-MB3	0.158 ± 1.10E-02	5.71 ± 2.20E-02	n.d.
7TUB165-OC	0.361 ± 2.88E-03	n.d.	n.d.
7TUB165-MB	0.379 ± 2.89E-03	n.d.	n.d.

In the complex hydrolysate samples, only the detection of tryptophans was possible, but not that of indoles. In the ancestral strain TUB00, as well as in the positive controls W-TUB165 and Ind-TUB165, tryptophan was detected in surprisingly low levels (about 0.1 – 0.2 µM). This low Trp content must be attributed to a flaw in the used hydrolysis protocol, most likely due to oxidative degradation under acidic conditions. The susceptibility of Trp to decompositions is known and can be mitigated by minimizing oxygen exposure, light exposure, high temperatures, and the addition of reducing agents/antioxidants such as thioglycolic acid[391] (used in this hydrolysis protocol), the utilization of functional derivatives of Trp or by hydrolysis under alkaline conditions.[392,393] Unfortunately, neither the addition of thioglycolic acid nor oxygen exclusion appeared to have been sufficient precautions to preserve Trp from decomposition. And there is, in general, still no superior strategy to stabilize tryptophan against degradation.[393] In control measurements of the media, reasonable amounts of Ind and Trp were detected, respectively (NMM0-0 70 (Ind): 63.10 µM and NMM0-0-70 (Trp): 104.70 µM), which supports the assumption of oxidative degradation.

Hydrolyzed samples of the 6Fi-adapted strains 6TUB128-OC, 6TUB165-MB4, and 6TUB165-MB3 revealed 6FTrp concentrations ranging from 4 – 6 µM, suggesting that the fluorinated tryptophan is also affected by oxidative decomposition, albeit to a slightly lesser extent (the NMM0-70-0 (6Fi) control revealed 53.05 µM). The detection of 7FTrp in 7TUB165-OC and 7TUB165-MB was not achieved at all. Unfortunately, in the adapted strains 6TUBX (0.1 - 0.2 µM) and 7TUBX (approx. 0.35 µM), similar or slightly higher values for Trp were also found as in the positive controls; however, its origin from the growth medium was excluded (see **Figure S63**). In samples of NMM0-0-0 and NMM0-70-0 (supplemented with either 6Fi or 7Fi), which were not subjected to the hydrolysis treatment, contaminations by Trp were not detected at all. Likewise, samples of the growth medium NMM0-70-0 (6Fi or 7Fi) that were treated according to the hydrolysis protocol showed no Trp in the case of treated

NMM 0-70-0 (7Fi) or traces of Trp far below the detection limit (1 pmol in FL detection) in case of treated NMM 0-70-0 (6Fi). The additional measurements of hydrolysis-treated NMM were made to exclude any degradation of the pure FTrp components (e.g., defluorination) as a cause of Trp signals.

Furthermore, supernatant samples of different time points (24 h, 48 h, 72 h, 96 h, 120 h) were analyzed for their (F)-Ind and (F)-Trp content in order to get insights into the temporal changes of (F)-Trp secretion and (F)-Ind uptake during cultivation (see **Table S25** in appendix).[116] For TUB00, as expected, a decreased Ind concentration with a concomitant increase of the Trp concentration was observed. In the positive controls, no Indol was detected, but there was a minor increase in the Trp concentration. For 6Fi-adapted strains, 6Fi was only detected at 24 h, and the 6FTrp concentration increased with increasing cultivation time. And for 7Fi-adapted strains only in 7TUB165-OC, a trend was observed that was, decreasing 7Fi concentrations. However, since, in most cases, the measured values were very low (5000 – 200000 uRIU FL peak areas corresponding to values below 1 µM), their courses were considered to be rather unspecific and not really meaningful.

3.6.2 Proteomics

Proteomics refers to the study of the proteome, i.e., the entirety of all proteins of an organism at a given time, in order to investigate the protein's structures, interactions, and functions. One challenge in proteomics is that the proteome is a very dynamic research object because, due to differences in gene expression, its composition changes depending on the current growth phase but also on environmental factors (e.g., nutrition, stress, etc.). This ramified field includes expression proteomics, which studies qualitatively and quantitatively changes in protein expression and reveals differences in the composition of the proteome (e.g., SILAC (stable isotope labeling by amino acids in cell culture[350])), structural proteomics, which concerns the determination of the three-dimensional structure of proteins and allows the investigation of protein-protein interactions as well as the interactions with other biomolecules such as DNA or RNA (e.g., interactomes), and functional proteomics which focuses on the characterization of a selected group of proteins and protein complexes. However, proteomics can also target, for example, post-translational modifications (PTMs), protein localization and compartmentalization, or the identification of protein biomarkers.[394] A standard workflow of proteomics often involves HPLC with tandem mass spectrometry for protein identification and fragment detection (**Figure 38**), but detection with antibodies (immunoassays) is also feasible.

In this study, proteomics was used in addition to amino acid analysis to prove Trp substitution by FTrp. Proteomic samples of all adapted strains and TUB00 were subjected to tryptic digestion, and peptides were analyzed by nano-LC-MS/MS in the laboratory of Lorenz Adrian at Helmholtz Zentrum für Umweltforschung, UFZ Leipzig (section **5.9.3**).

Figure 38| General workflow of proteomics. A) Symbolic picture of proteomics. B) A typical workflow of proteomics comprises the following steps: cell cultivation under specific growth conditions, protein extraction/cell lysis (e.g., by sonication, cell disrupter, or chemical lysis), protein digestion into peptides (e.g., with trypsin), peptide separation by LC-MS and peptide analysis by MS/MS. The figure was drawn with BioRender (https://www.biorender.com/).

In total, 19555 different peptides over all samples were found, i.e., the dataset covered 53 % of the predicted proteins (2299 detected/identified proteins over 4293 predicted proteins based on genomic data of TUB00.fasta). Considering that some proteins are rare, not produced in the respective growth phase, or are only produced under some special conditions, a coverage of 53 % is satisfying. Thereby, a relatively high variance within biological replicates and between each dataset was observed, but this is known for the applied method. A summary of the detected peptides is given in **Table S26**, **Table S27**, and **Table S28** in the appendix. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD048225 (https://proteomecentral.proteomexchange.org/).[395,396]

Interpretation of the Trp-containing peptides revealed unambiguous quantitative replacement in all samples of 6Fi and 7Fi adapted strains within the detection limit of the instrumentation (**Figure 39**). The false discovery rate was set to 1 % at the peptide identification level using the Target Decoy PSM Validator node. The following "false positives" rates (occurrence of FTrp-containing peptides in positive controls and occurrence of Trp-containing peptides in adapted strains; mean of three biological replicates) were determined TUB00 (0.0225 % ±0.0318), Ind-TUB165 (0.0878 % ±0.0098), W-TUB165 (0.0485 % ±0.0485), 6TUB128-OC (0.2309 % ±0.1315), 6TUB165-MB4 (0.1702 % ±0.0555), 6TUB165-MB3 (0.1269 % ±0.1269), 7TUB165-OC (0.3984 % ±0.0822) and 7TUB165-MB (0.5154 % ±0.1173). These data

indicate that despite the careful experimental setup (strict prevention of Trp import in the growth medium and Fi substrates), a certain background Trp incorporation was detected.
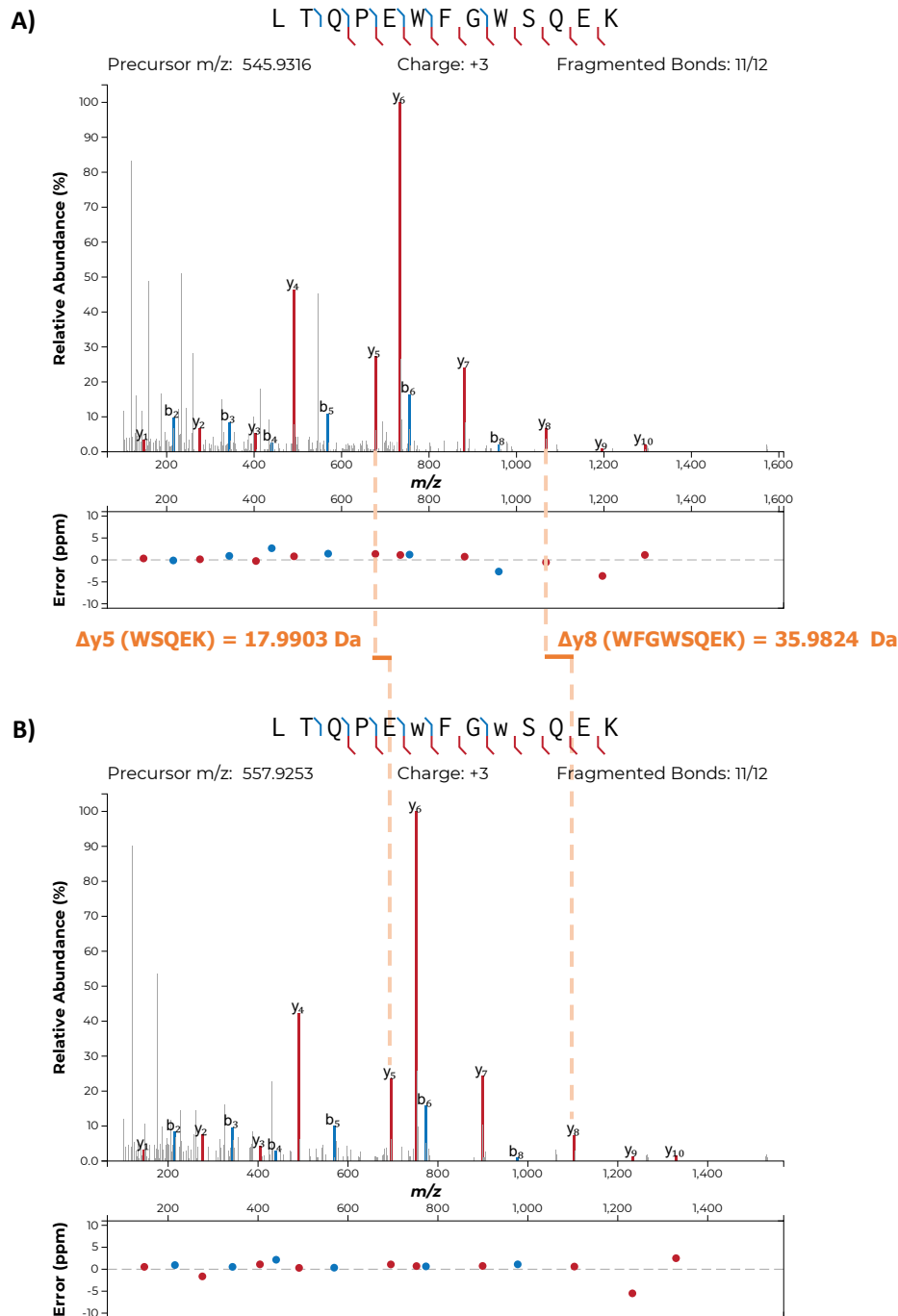
Figure 39 | Representative MS/MS spectra for global Trp to FTrp substitution. The MS/MS spectra of oligopeptide ABC transporter periplasmic binding protein (NP_415759.1) peptide 352-LTQPEWFGWSQEK-366 from ancestral/ positive controls (A) and FW-adapted strains (B) are shown, with and without fluorine modification on W6 and W9, respectively. The selected monoisotopic precursor m/z (z=+3) is given in each diagram, and all detected y-fragment ions are shown in red, and b-fragment ions are shown in blue. Orange brackets visualize the consistent m/z difference of 17.99 Da per fluorine substitution between equivalent y ions; y8 (WFGWSQEK) = 35.9824 Da for W9 and W6 fluorine modifications and y5 (WSQEK) = 17.9903 Da for W9 fluorine modification. The graphic was drawn with http://www.interactivepeptidespectralannotator.com/PeptideAnnotator.html.

3.7   Metabolomics

Metabolomics refers to the systematic study of the so-called metabolome, which is the collection of all soluble low-molecular weight (< 1500 Da) organic molecules in a biological system.[397] Just like the proteome, the metabolome is also very dynamic, as an organism adapts its metabolism to external circumstances, i.e., small molecules are continuously absorbed, transformed, synthesized, degraded, and interact with other molecules. Thus, metabolomics allows the monitoring of metabolic pathways and can be considered a direct readout of the physiological state of an organism. For analyzing the metabolome, in particular, mass spectrometry and NMR spectrometry-based techniques are suitable, which, owing to the wealth of data (high number of small molecules), must be supplemented by statistical methods and often require substantial bioinformatic effort (**Figure 40**).[398]

Subsets of metabolomics are, for example, lipidomics, which is the study of pathways and networks of cellular lipid species,[399] or fluxomics, used to determine the rates of metabolic reactions.[400]
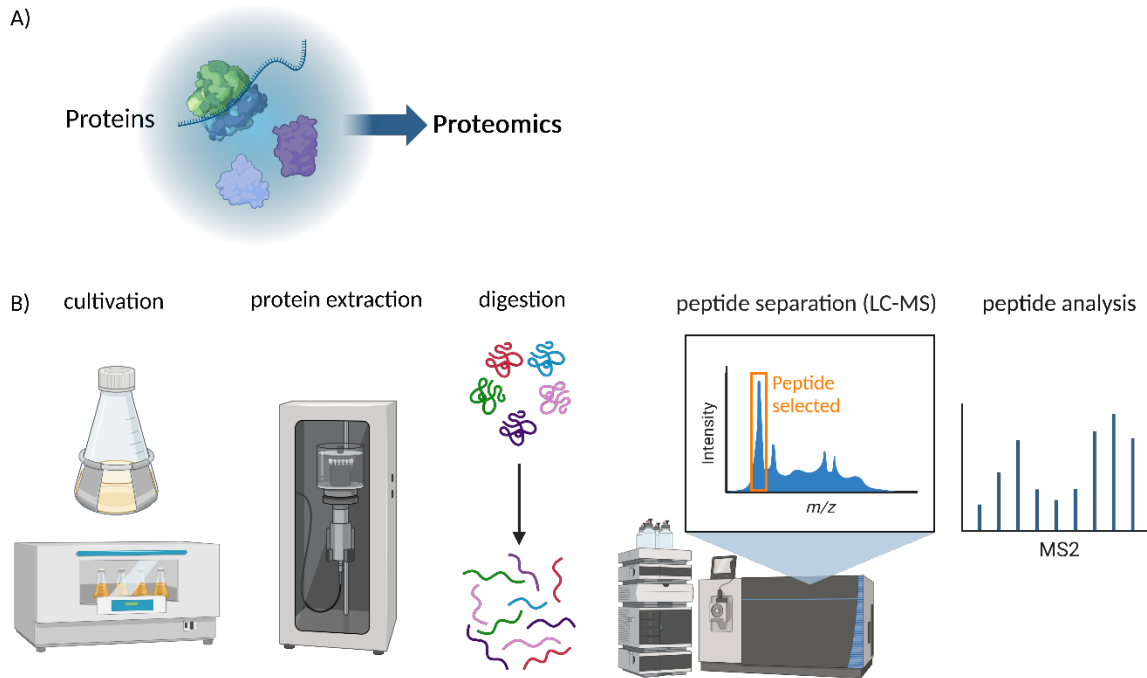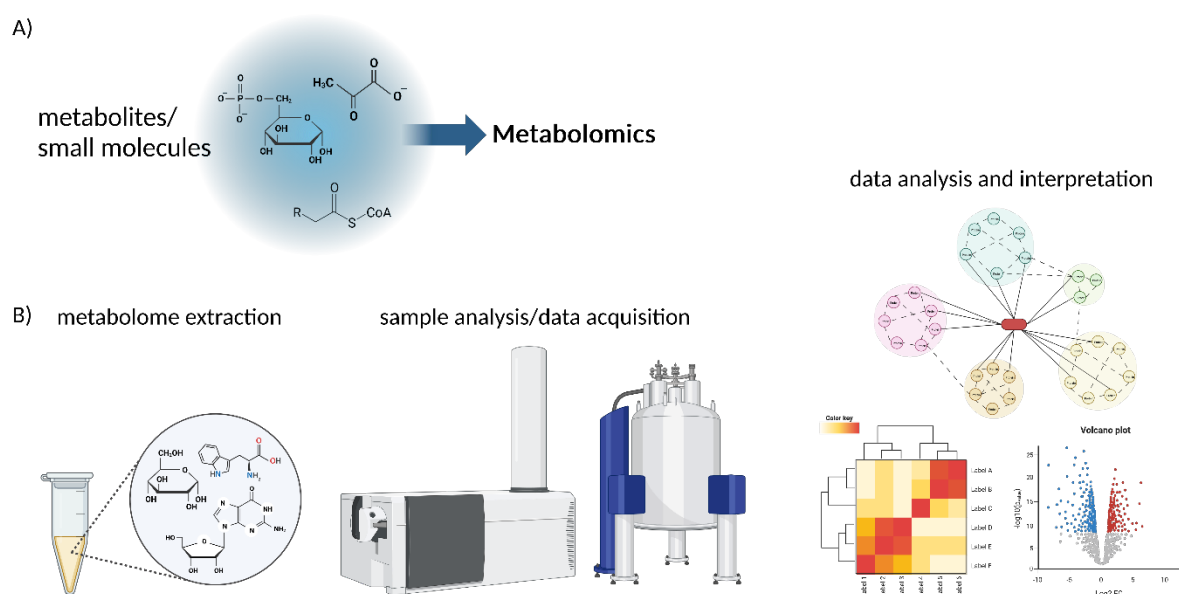

Figure 40 | General workflow of metabolomics. A) Symbolic picture of metabolomics. B) A typical workflow of metabolomics comprises the following steps: cell cultivation under specific growth conditions, metabolome extraction (e.g., hot water extraction), sample analysis/data acquisition using mass spectrometry (left) or NMR spectroscopy (right), and data analysis and interpretation using statistical methods and bioinformatic tools (heat map, volcano plot and network analysis are shown as representative examples). The figure was drawn with BioRender (https://www.biorender.com/).

As part of a mass spectrometry/metabolomics course of the European Training Network (ETN) from the SynCrop project (synthetic circuits for robust orthogonal production), the opportunity arose to examine a small number of samples by metabolomics. For this purpose, the respective MB lineages were chosen in order to nevertheless ensure a comparison between 6Fi- and 7Fi-adapted cells, while the OC lineages must be excluded. In this context, the metabolic composition of TUB00 and isolates at the middle (6TUB53 and 7TUB53) and almost final time points (6TUB161-MB4 and 7TUB161-MB) were examined using untargeted metabolomics; more precisely, flow injection analysis (FIA). Measurements of the metabolic profiles and initial data processing were performed in collaboration with the group of Uwe Sauer (Systems Biology of Metabolism, ETH Zurich).

FIA allows for high-throughput analysis since the chromatography step is omitted and samples are directly fed into the electrospray interface (MS);[401] however, this also substantially increases the risk of matrix effects affecting electrospray ionization, which is why FIA was recommended for initial untargeted metabolome screens where thousands of samples need to be analyzed quickly. Furthermore, due to the hot water extraction method of the metabolome, only polar fractions can be addressed (see section **5.10**), but the authors stated that a coverage of about one-quarter of the total number of compounds listed in a genome-scale reconstruction metabolic model of *E. coli*[402] could be confidently identified. The analysis conducted for this study yielded 179 ion annotations after curated data processing.

3.7.1   Determination of indole, tryptophan, and fluorotryptophan content

FIA was one approach used to determine the content of indole, tryptophan, and fluorotryptophan (**Figure 41**); ions representing fluoroindole could not be annotated. The ancestral TUB00 shows nearly the same intensities for indole and tryptophan. The level of 6-fluorotryptophan is noticeably higher in 6TUB53 than in 6TUB161-MB4. Considering that the measured signals are generated by free amino acids, it can be assumed that 6TUB53 has an excess of the ncAA that was not incorporated into proteins and, hence, accumulated in the cytoplasm. Accordingly, the low level of free 6FTrp in 6TUB161-MB4 indicates its use in protein biosynthesis, i.e., for an increased incorporation efficiency over the course of adaptation, which is a further sign of successful adaptation. In contrast, the levels of free 7-fluorotryptophan in 7TUB53 and 7TUB161-MB are significantly higher, which indicates a generally lower incorporation level and weaker adaptation. Nevertheless, these signals should not be regarded as absolute values or quantification but rather as an overview since intensities of ions corresponding to metabolites were in the range of 10^3 – 10^4 counts, which is close to the detector baseline and a known bearing for the FIA method.[401]
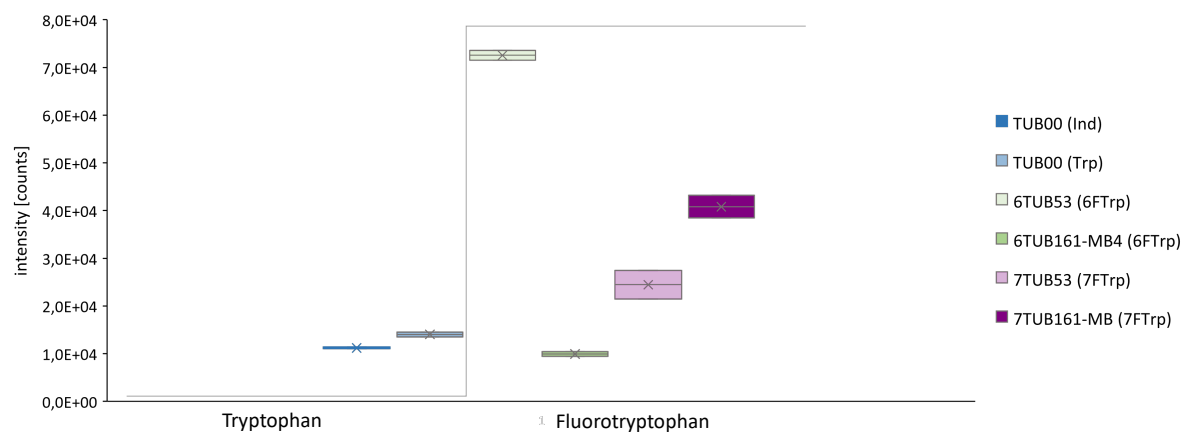
Figure 41 | Metabolite levels of Ind, Trp, and FTrp. The measured features (Ind, Trp, 6FTrp, and 7FTrp) are depicted in the legend, and a boxplot was used to illustrate the variation within the samples, which is especially pronounced in 7TUB53 and 7TUB161-MB.

Particularly relevant is that no biologically significant peaks for Ind/Trp were found in the samples of the adapted strains because this supports the assumption of complete substitution of all Trp positions by FTrp. More precisely, although undesirable signals were found (FTrp in TUB00 that never experienced fluorinated substrates as well as Ind/Trp in the adapted strains), these were in the range of water samples (used for calibration) and were generated by the mass software through gap filling.[403] Gap-filling is an automated algorithm that aims to reduce false missing values by filling these gaps through the re-integration of the peak area where the peak is expected. This applies not only to "missing" features that could not be identified/quantified due to deficient peak detection or poor alignment but also to those that were actually measured with zero intensity. However, by manual curation using a t-test, the FTrp signals in TUB00 and the Trp signals in the adapted strains could be clearly rated as not significant (see volcano plots in **Figure S66**, appendix).

3.7.2   Metabolic profiles of fluorine-adapted strains

The general metabolome composition was evaluated by multivariate statistics of principal component analysis (PCA), which uses a linear combination to reduce the different experimental dimensions to principal components, here visualized as a synchronized 3D plot (**Figure 42**). PCA revealed that the unique chemical identity of the metabolomes of the strains adapted to 6Fi and 7Fi progressively diverged from that of the TUB00 ancestor and that different evolutionary trajectories (indicated by red arrows) were taken depending on the substrate. Thereby, larger deviations are observed for 6TUBX than for 7TUBX, both in terms of the spatial distance between the individual clusters (6Fi in green and 7Fi in purple) and in terms of the groups within the clusters (middle XTUB53 vs final XTUB161-MB isolates).
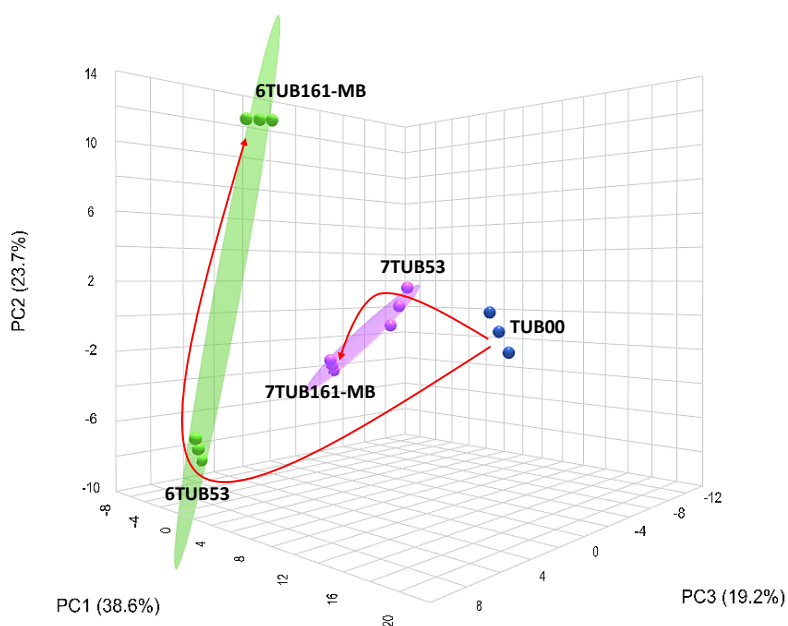
Figure 42| PCA plot of 6Fi and 7Fi adaptations. For the principal component analysis, near-constant variables throughout the experiment were filtered using robust estimate interquantile range (IQR), data were normalized by sum, and auto-scaling was applied. The PCA was created by the MetaboAnalyst 5.0 platform.[404,405]

Figure 43 and **Figure 44** show an overview of the metabolites changed in the isolates 6TUB161-MB4 and 7TUB161-MB, respectively. The data are illustrated as dots visualized on a metabolic network map,[406] which uses the KEGG database.[407-409] The changes are given by fold changes (log2FC) according to a green/red color scheme and the significance by the dot size (-log(p)); molecules are additionally described with KEGG entry ID C0000 in the text.

The metabolic profile of 6TUB161-MB4 (**Figure 43**) shows mainly downregulated metabolites. Among them, changes in nucleotide metabolism (red-colored pathway) are particularly striking since they appear as clusters. Other downregulated metabolites are enriched in the amino acid biosynthetic pathways (orange-colored pathways) of Val, Leu, Ile, Lys and Arg with N-succinyl-2-L-amino-6-oxoheptanedioate (C04462), N-succinyl-L,L-2,6-diaminoheptanedioate (C04421) and L,L-2,6-diaminoheptanedioate (C00666) from lysine biosynthesis strongly reduced with log2FC of – 3.7 to -5.4. In contrast only five metabolites are upregulated (red dots) in this 6Fi-adapted strain as they are phosphatidylethanolamine (C00350, log2FC 3.0), D-tagatose 1,6-bisphosphate (C03785, log2FC 2.2), beta-D-fructose 1,6-bisphosphate (C05378, log2FC 2.2), 5-phospho-alpha-D-ribose 1-diphosphate (PRPP, C00119, log2FC 1.6) and D-fructose 1,6-bisphosphate (C00354, log2FC 2.2). They can be assigned to carbohydrate and lipid metabolisms.

The metabolic profile of 7TUB161-MB (**Figure 44**) shows as well downregulated metabolites of the nucleotide metabolism (red-colored pathway) and similar findings for the amino acid metabolism (orange-colored pathway) with the three lysine-related metabolites C04462, C04421, and C00666 are again most reduced with log2FC -4.9 to -7.4. However, unlike 6TUB161-MB4, this 7Fi-adapted lineage clearly exhibits more upregulated metabolites that are distributed in the carbohydrate metabolism (blue-colored pathways). They are the hexoses alpha-D-galactose (C00984), alpha-D-glucose (C00267), beta-D-glucose (C00221), D-glucose (C00031), D-mannose (C00159), myo-inositol (C00137) with log2FC around 3.8, glycerone (C00184, log2FC 3.1), (S)-lactate (C00186, log2FC 3.1), acetate (C00033, log2FC 2.7), 3-hydroxypropanoate (C01013, log2FC 3.1), hydroquinone (C00530, log2FC 2.7), glycolaldehyde (C00266, log2FC 2.7) and phosphatidylethanolamine (C00350, log2FC 3.0).

Interestingly, the Trp metabolism is not directly affected, i.e., except for indole (C00463) and L-tryptophan (C00078), which are reduced as expected due to the substitution with their fluorinated counterparts, none of the identified metabolites belongs to this pathway (highlighted in teal in **Figure 43** and **Figure 44**). Their levels (log2FC) changed compared to TUB00 as follows: in 6TUB161-MB4 indole -3.1, Trp -1.5 and in 7TUB161-MB indole -3.5, Trp -1.6.

Initially, the metabolome analysis was also thought to shed light on metabolites belonging to the stringent response system (e.g., stress alarmone (p)ppGpp) and those associated with membrane composition (e.g., phosphoglycerol lipids). Both systems were affected in earlier tryptophan-related adaptation experiments. Therefore, it was likely to assume that they would be targeted again in the present ALE.[220,328,410] However, they could not be captured with the used FIA method owing to the intrinsic constraints described above that were missing separation component and restriction to polar metabolite fractions.

The entire dataset of metabolic profiles along the adaptation trajectory encompasses the following comparisons 6TUB53:TUB00 (**Figure S67**), 6TUB161-MB4:6TUB53 (**Figure S68**), 6TUB161-MB4:TUB00 (raw figure, **Figure S69**), and 7TUB53:TUB00 (**Figure S70**), 7TUB161-MB:7TUB53 (**Figure S71**), 7TUB161-MB:TUB00 (raw figure, **Figure S72**); see appendix.
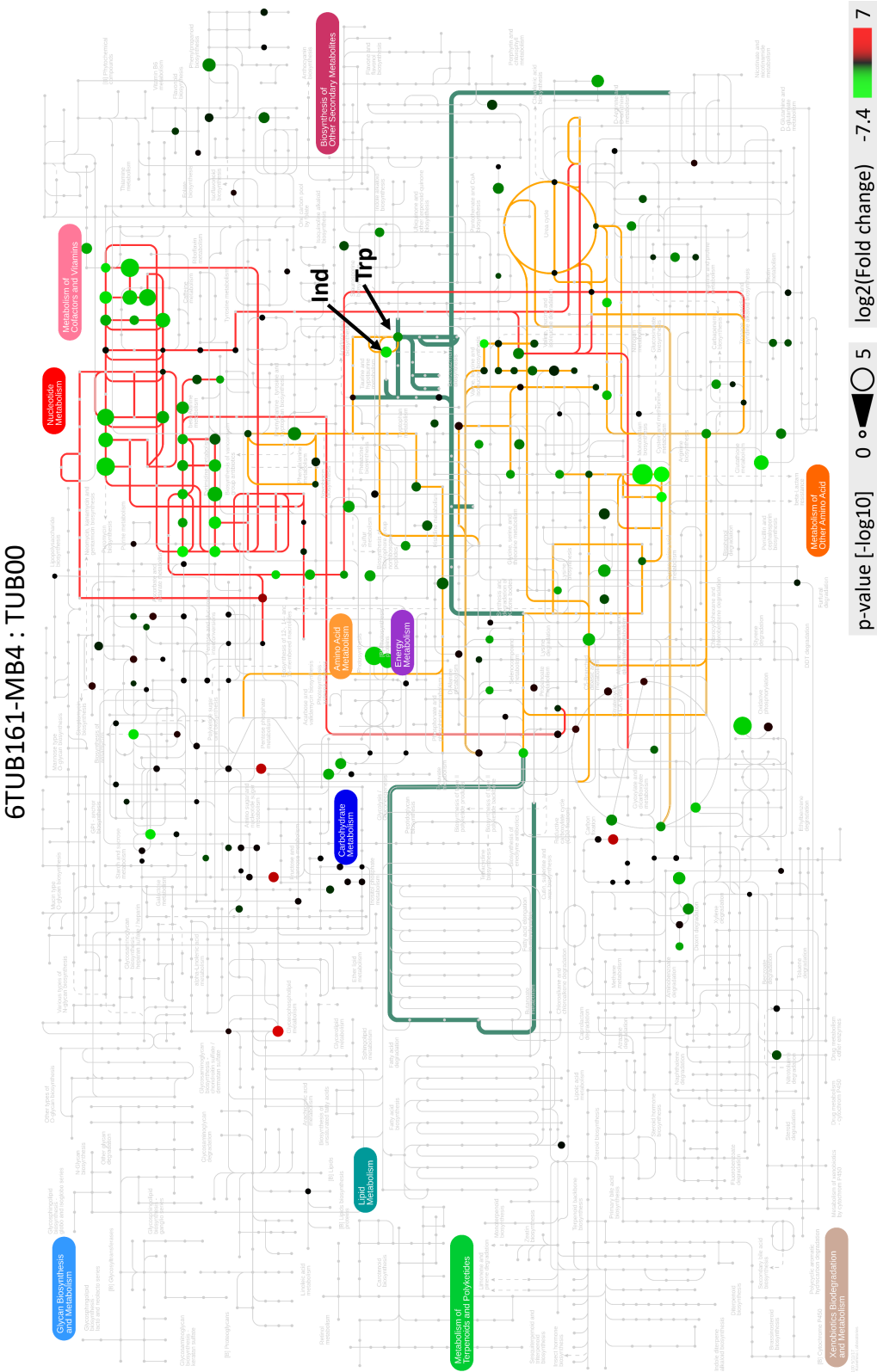
Figure 43 | Metabolic profile of 6TUB161-MB4 in comparison to TUB00. The metabolites indole (Ind), tryptophan (Trp), and pathways with relevant enriched signals are highlighted: the nucleotide metabolism (red), the amino acid metabolisms of Val, Leu, Ile, Lys, and Arg biosynthesis (orange) as well as Trp metabolism (teal, bold). Metabolomics data were visualized as a metabolic network map using the iPath3.0 interactive pathways explorer.[406] Identifiers for highlighted pathways are map00230 (purine metabolism), map00240 (pyrimidine metabolism), map00220 (Arg biosynthesis), map00290 (Val, Leu and Ile biosynthesis), map00300 (Lys biosynthesis), map00380 (Trp metabolism) and map00400 (Phe, Tyr and Trp biosynthesis).
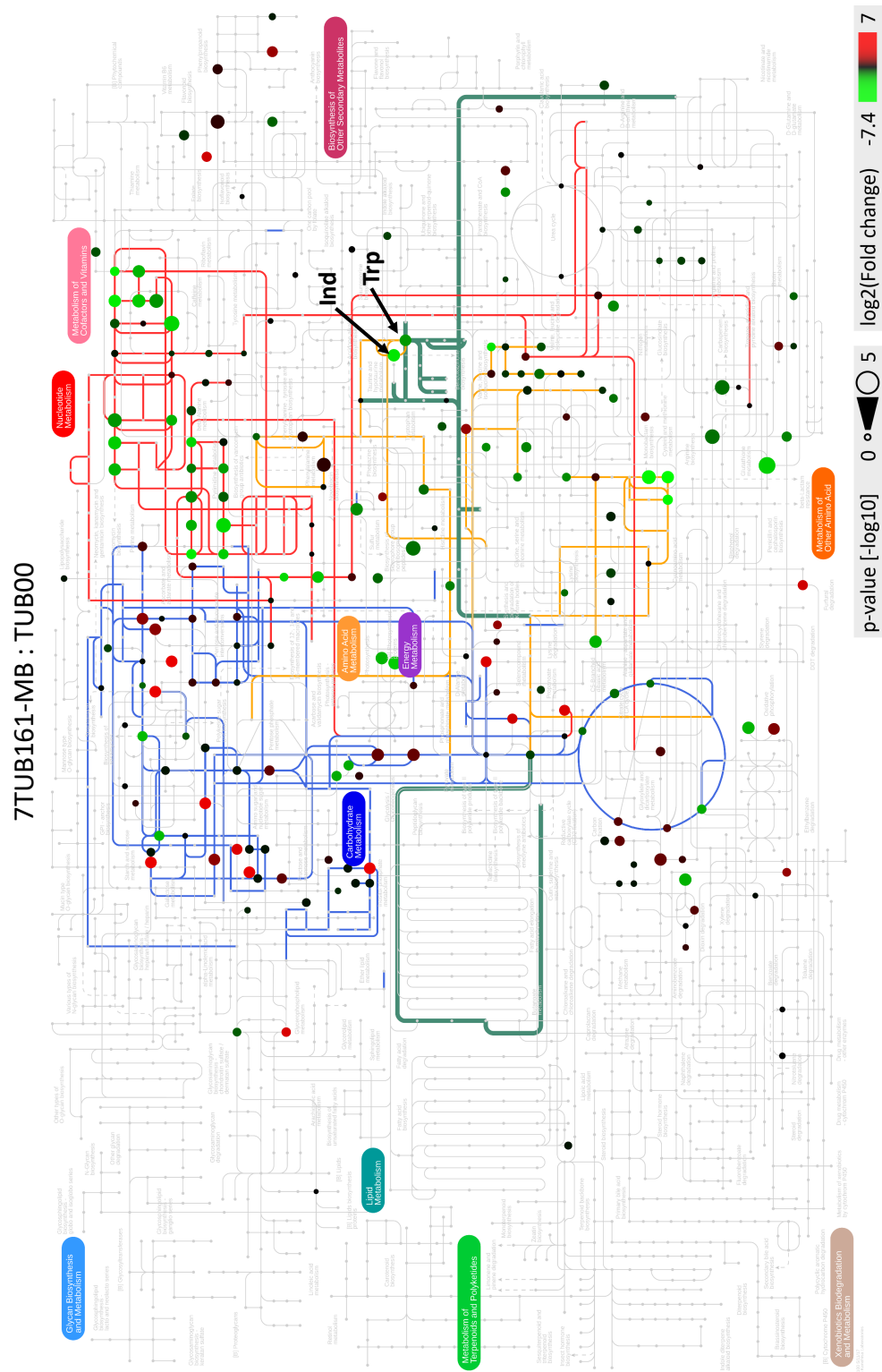
Figure 44 | Metabolic profile of 7TUB161-MB in comparison to TUB00. The metabolites indole (Ind), tryptophan (Trp), and pathways with relevant enriched signals are highlighted: the nucleotide metabolism (red), parts of the carbohydrate metabolism (blue), the amino acid metabolisms of Val, Leu, Ile, and Lys biosynthesis (orange) as well as Trp metabolism (teal, bold). Metabolomics data were visualized as a metabolic network map using the iPath3.0 interactive pathways explorer.[406] Identifiers for highlighted pathways are map00230 (purine metabolism), map00240 (pyrimidine metabolism), map00010 (glycolysis/gluconeogenesis), map00020 (TCA cycle), map00040 (pentose and glucuronate interconversions), map00052 (galactose metabolism), map00562 (inositol phosphate metabolism), map00290 (Val, Leu and Ile biosynthesis), map00300 (Lys biosynthesis), map00380 (Trp metabolism) and map00400 (Phe, Tyr and Trp biosynthesis).

3.8   Genomics

The research branch of genomics refers to the study of the structure, function, evolution, mapping, and editing of an organism's complete set of DNA, the genome. The foundation of this technology was paved by the development of so-called next-generation sequencing (NGS), which allows the DNA sequences of entire genomes to be determined. Well-established high-throughput sequencing (HTS) techniques of the second generation are, for example, 454 pyrosequencing, ion torrent semiconductor sequencing, and Illumina sequencing, but also third-generation methods, e.g., single molecule real-time (SMRT) sequencing, and fourth-generation methods, e.g., nanopore-based DNA sequencing are existent.[411–413] The probably most prevalent sequencing platform is given by Illumina dye sequencing, which is based on sequencing by synthesis. Fragmented DNA is tagged with adapters, immobilized on a flow cell, and amplified by bridge amplification, then sequencing proceeds using fluorescently tagged nucleotides that emit a characteristic fluorescent signal and also act as terminators of synthesis for each reaction, allowing sequence determination.[414] Afterward, the resulting sequence reads are aligned and merged to reconstruct the original sequence (assembly). Finally, in genome annotation, biological information is assigned to the sequence by means of coding or regulatory regions (e.g., genes) either by comparison with a reference genome or by using BLAST[415] (Basic Local Alignment Search Tool; https://blast.ncbi.nlm.nih.gov/Blast.cgi) to search for sequence homologies. A typical genomics workflow is shown in **Figure 45**.
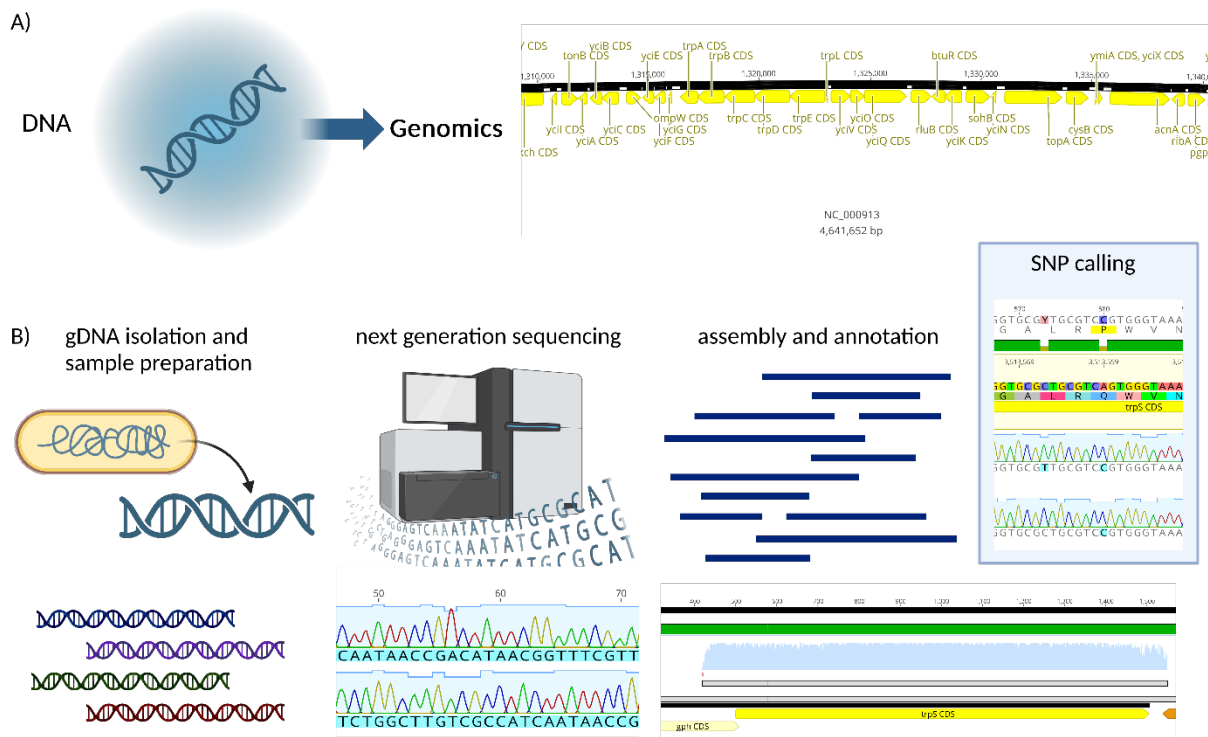
Figure 45 | General workflow of genomics. A) Symbolic picture of genomics (left) and a segment of the *E. coli* genome (right) are shown. B) A typical workflow of genomics comprises the following steps: isolation of

chromosomal DNA (or genomic DNA, gDNA), sample preparation (e.g., library creation), next-generation sequencing, genome assembly, and annotation. The grey box shows a typical application of genomics, namely mutational analysis (e.g., SNP calling). The figure was drawn with BioRender (https://www.biorender.com/); the sequence images are Geneious Prime 2023.2.1 extractions of original data (annotation representation: *trpS* CDS (coding DNA sequence), SNP calling representation: point mutations L25L and Q27P in the *trpS* gene).

A major field of genomics concerns the investigation of genetic variations, such as single nucleotide polymorphisms (SNPs), as focused on in this study. Besides mutational analysis, also the phenomena of pleiotropy, which describes the effect of one gene affecting more than one phenotypic trait (see also antagonistic pleiotropy),[416] and epistasis, which is the effect of one gene on another,[417] are objectives of genomic studies.

Accumulating random genetic mutations is one adaptation strategy of organisms to bring out improved phenotypes specialized to the given environment, thus, microbial diversity. The baseline for mutation rates in wild-type *Escherichia coli* ranges from 0.2×10^{-10} to 5×10^{-10} nucleotides per generation.[418,419] However, this increases dramatically due to defects in the mismatch repair mechanisms, like the MutSLH system.[418,420]

In order to identify changes in the informational polymer DNA that the parental strain TUB00 had to undergo for trophic specialization towards 6Fi and 7Fi, genomic mutations that affected the adapted strains were examined. The ancestral TUB00 and all adapted strains, including lineage samples along the adaptation trajectory (see section **5.11, Table 9**), were subjected to whole genome sequencing (WGS) to investigate the landscape of genomic changes (SNP, single nucleotide polymorphism). The chromosomal DNA (genomic DNA, gDNA) was analyzed on the Illumina sequencing platform[421] in collaboration with Torsten Semmler from the Robert Koch Institute. An overview of the number and type of mutations that occurred in the final ALE isolates is shown in **Figure 46**.

Overall, the SNP landscape shows an equal distribution of mutation types within the different lineages, with missense variants (orange) dominating over synonymous variants (grey), InDel/frameshift (yellow), and non-coding regions (blue). The final isolates of the positive controls exhibit either 8 SNPs (W-TUB165) or 16 SNPs (Ind-TUB165). The fact that the Trp control has the lowest number of mutations is plausible because this strain was cultured in the most native medium, whereas the indole control had first to produce Trp itself. The two lineages, 7TUB165-OC and 7TUB165-MB also showed a relatively low number of mutations with 20 and 21 SNPs, respectively (discussed below in section **3.8.1**). In contrast, the most striking observation from the statistic SNP overview (**Figure 46**) is the unusual high mutation rate in both 6Fi-adapted MB lineages (6TUB165-MB4: 175 SNPs and 6TUB165-MB3: 285 SNPs), which were 9-fold and 15-fold higher than in the others including their 6TUB128-OC relative with 17 SNPs (see section **3.8.2**).

	W-TUB165	Ind-TUB165	6TUB128-OC	6TUB165-MB4	6TUB165-MB3	7TUB165-OC	7TUB165-MB
■ stop lost	1	1	0	0	0	0	1
■ stop gained	0	0	1	3	3	1	1
■ InDel/frameshift	1	2	2	32	41	4	3
■ synonymous	1	2	0	32	48	2	0
■ missense	2	6	11	83	139	11	13
■ non-coding	3	5	3	23	46	2	3
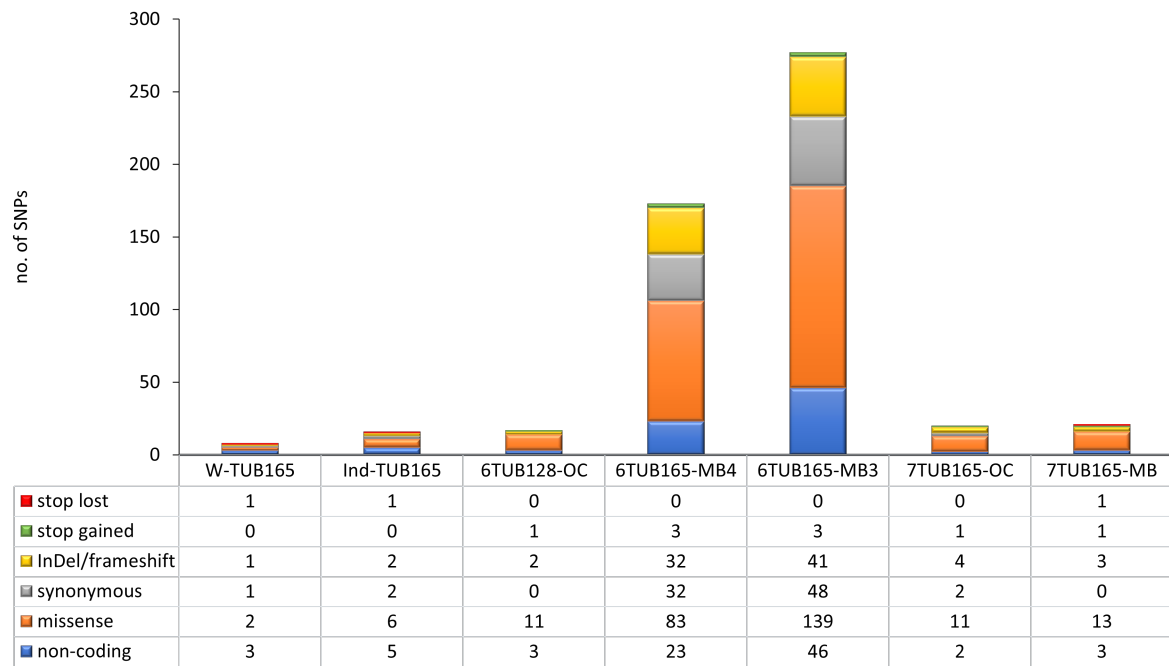
Figure 46 | SNP landscape of the final ALE isolates. The number and types of mutations are shown as stacked bars. Types of SNPs and consequences are: non-coding region means DNA in between genes, all other types are within coding regions and comprise synonymous SNPs (so-called silent mutations) that change the codon but not the amino acid sequence in a protein, missense SNPs that exchange one base resulting in change of the codon and the corresponding amino acid, InDel/frameshift are insertions or deletions of more than one base that results in missense or frameshift mutation, and nonsense SNPs delete a stop codon (stop lost) or create a stop codon (stop gained) which results in premature truncation of the protein. Note: non-coding transcript variants are not listed.

In sum, for both control lineages (W-TUBX and Ind-TUBX), a rather diffuse mutation pattern given by most likely arbitrary SNPs along the adaptation trajectory was observed (the datasets comprised samples of passages (p) 18, 23, 53, 72, 83, 88, 128 and 165; see **Table S29** and **Table S30** in appendix); no crucial targets or permanent mutations were observed during long-term cultivation under canonical conditions. In the W-TUBX lineage, the Gln33Lys variant of the *yibJ* gene persisted in passages 18, 23, 53, 72, and 88 but was absent in the posterior passages. The gene encodes for the putative RHS domain-containing protein YibJ, but no further information regarding the cellular function of this product is available on common databases such as EcoCyc[131] or UniProt.[422] In addition to *yibJ* Gln33Lys in p18 and p53, the Ind-TUBX lineage conspicuously accumulated mutations in transponases, namely the synonymous variant *insA-2* Thr69Thr in passages 23, 83, and 128, as well as missense mutation Phe81Leu of it in p83 and p165, and in passages 88 and 165 synonymous variants of *insAB-1* as well as the missense variant Arg23Ser of it. Transponases are involved in the movement of genes from one position on the chromosome to another, and their encoding genes are the most abundant, most ubiquitous ones found in nature.[423]

3.8.1 Genomic variations in 7Fi ALEs

Samples of 7TUB18, 7TUB53, 7TUB83-OC, 7TUB165-OC, 7TUB72-MB, 7TUB83-MB, and 7TUB165-MB reflecting the ALE trajectory, with special emphasis on the events where the cell density dropped (passages 72 and 83; see section **3.4.3**), were chosen for genomic analysis. The entire dataset of all genes, their function, and the occurring SNPs is deposited in **Table S31** and **Table S32** in the appendix. The genetic variations that emerged during adaptation to 7-fluoroindole are shown in **Figure 47**. Mutations that occurred in the course of both adaptation experiments were analyzed with ClueGO,[424] which is a Cytoscape plug-in to decipher functionally grouped gene ontology (GO) and pathway annotation networks.[425] In this way, the mutation data given by a list of affected genes were systematically organized and placed in a biological context according to their GO terms biological process and molecular function.[426,427] In total, 11 clusters were identified (differently colored in **Figure 47**) that, in turn, could be grouped into four parent domains. Most of the genes belong to protein metabolism by means of amino acid synthesis, transport, and translation. But genes associated with phospholipid synthesis and transport, as well as DNA template transcription and stress response systems, were also affected.
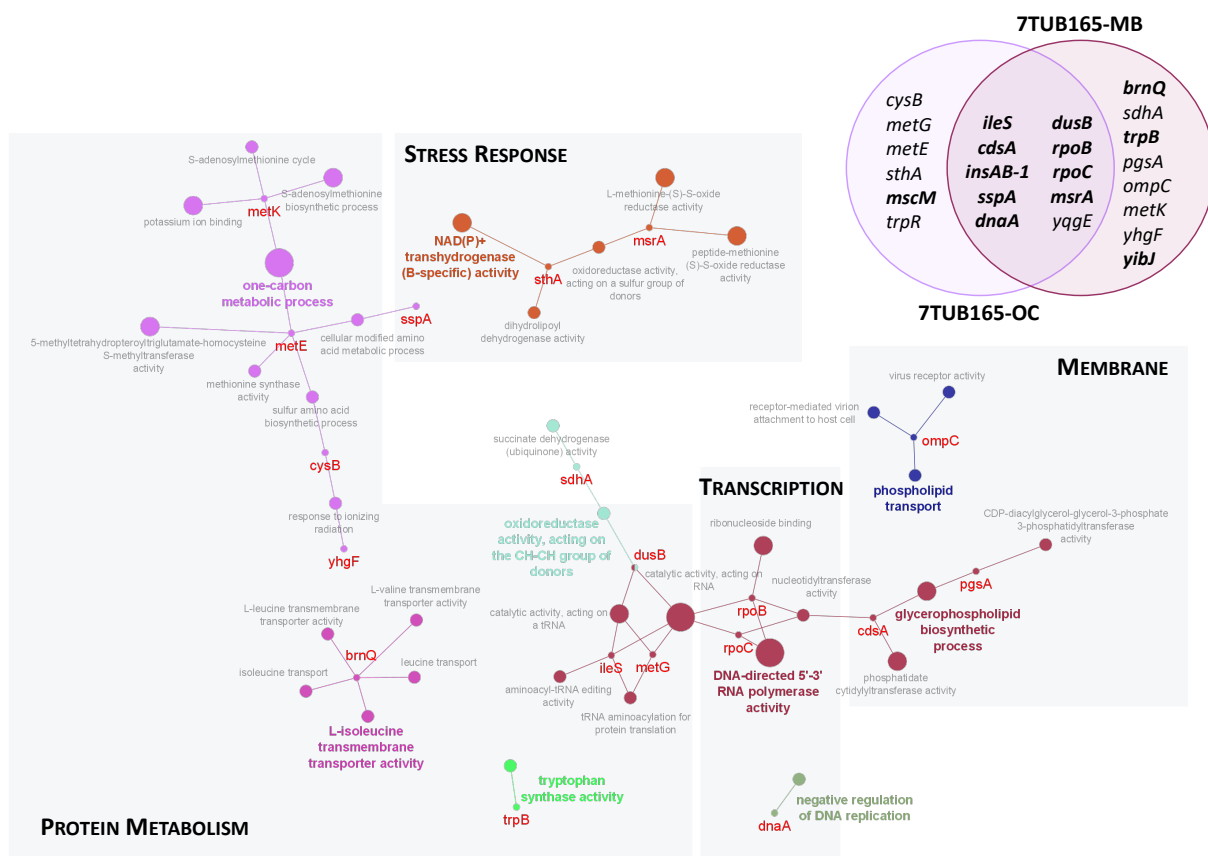
Figure 47| Genetic variations in 7TUB165-OC and 7TUB165-MB. A functional network of the mutations that occurred in 7TUB165-OC and 7TUB165-MB is shown. In the right corner, the Venn diagram illustrates genes shared by both lineages and those unique for the respective lineage; bold genes were fixed in several passages during the ALE trajectory. Annotated genes and their encoded functions are: *metK* (methionine adenosyltransferase), *metE* (cobalamin-independent homocysteine transmethylase), *cysB* (DNA-binding transcriptional dual regulator CysB),

yhgF (RNA-binding protein YhgF), *brnQ* (branched chain amino acid transporter BrnQ), *trpB* (tryptophan synthase, β-subunit dimer), *ileS* (isoleucine—tRNA ligase), *metG* (methionine—tRNA ligase), *dusB* (tRNA-dihydrouridine synthase B), *sdhA* (succinate:quinone oxidoreductase, FAD-binding protein), *sspA* (stringent starvation protein A), *sthA* (soluble pyridine nucleotide transhydrogenase), *msrA* (methionine sulfoxide reductase A), *rpoB* (RNA polymerase subunit β), *rpoC* (RNA polymerase subunit β'), *dnaA* (chromosomal replication initiator protein DnaA), *cdsA* (CDP-diglyceride synthetase), *pgsA* (CDP-diacylglycerol—glycerol-3-phosphate 3-phosphatidyltransferase) and *ompC* (outer membrane porin C).[131] The functional network was visualized with Cytoscape[424] and analyzed with the ClueGO[425] app using the GO database from 04.11.2021.

Comparing the OC and MB lineages, it can be observed that similar domains were affected by genetic changes. These domains were affected to the same extent but with different targets, indicating effects based on the fluorine substituent. For example, genes related to the amino acid metabolism were addressed by both lineages, however, some genes like *trpB* and *brnQ* related to cAA biosynthesis and transport are exclusively targeted by 7TUBX-MB, which could be a result of the more aggressive method for the amino acid removal. The *metE* gene (cobalamin-independent homocysteine transmethylase) in 7TUBX-OC can be compensated by its relative *metH* (cobalamin-dependent methionine synthase), which also catalyzes the last step in the methionine biosynthesis.[428]

Furthermore, two genes (*trpR*, *trpB*) directly involved in the tryptophan metabolism were affected by mutation. The *trpR* gene (DNA-binding transcriptional repressor TrpR) that regulates expression of the Trp regulon in response to intracellular levels of tryptophan was found with frameshift variant Glu39fs in 7TUB83-OC. Since frameshift mutations are usually considered deleterious due to the alteration of the reading frame, it can be assumed that it leads to a dysfunctional gene product, which in this case would mean that the transcription of *trpBA* and thus the production of FTrp from Fi cannot be repressed. This mutation has not been conserved and could possibly be a reason for the difficulties in adapting to 7Fi during mid-ALE. The *trpB* gene was found with Arg175Cys in 7TUB72-MB and 7TUB165-MB, and the mutation was predicted as impactful by mutfunc. The prediction tool mutfunc computes the impact of amino acid and nucleotide variants in a corresponding reference genome based on considerations such as: whether the affected regions are conserved and therefore important for protein structure, or it calculates whether the affected site is part of a catalytic core or protein interface.[429] However, applications like this, or Provean,[430] do not consider the biological context, i.e., a mutation might be detrimental in the natural environment but can bring advantages in a certain foreign environment, such as fluorine-containing nutrition. In the end, the specific effects must be experimentally determined, as demonstrated in section **3.8.3**.

3.8.2 Genomic variations in 6Fi ALEs: emergence of Hypermutators

Of every 6Fi ALE lineage (6TUBX-OC, 6TUBX-MB4, and 6TUBX-MB3), samples of passages 23, 53, 88, 128, and 165 in case of the MB lineages were chosen for genomic analysis (see section **5.11, Table 9**). While the 6TUBX-OC again accumulated only a few (17) SNPs (**Table S33**), remarkably, the lineages 6TUBX-MB4 (**Table S34**) and 6TUBX-MB3 (**Table S35**) have evolved into so-called hypermutators. The functional networks are shown in **Figure 48** and **Figure 49**.

In the 6TUBX-OC lineage, an astonishingly high number of SNPs are conserved, with half of all affected genes comprising *can*, *pcnB*, *yibW* (were not annotated by the algorithm in the network), *fliL*, *sspA*, *trpS*, and *rpoB*. Again, genes associated with the amino acid metabolism are mainly affected. One key change could be given by the missense variant c.80A>C p.Gln27Pro and the synonymous variant c.73C>T p.Leu25Leu in *trpS* encoding for the tryptophanyl-tRNA synthetase (TrpRS), which also occurred in both MB lineages. Although the Leu25Leu mutation is silent, the different codon usage frequency (CUG (48 %) → UUG (13 %) could have regulatory, decelerating effects on the translation process and can even contribute to adaptation.[431] TrpRS was already targeted in previous FTrp-based ALE experiments;[326,328] that and because of its pivotal role in processing tryptophan, making an investigation of the enzyme by Michaelis−Menten kinetic experiments[432] or the enzyme binding affinity of Trp and FTrp by isothermal titration calorimetry (ITC)[433] certainly worthwhile. In this regard, the corresponding mutant genes have already been cloned, and the TrpRS variants were expressed (section **5.6.1**).
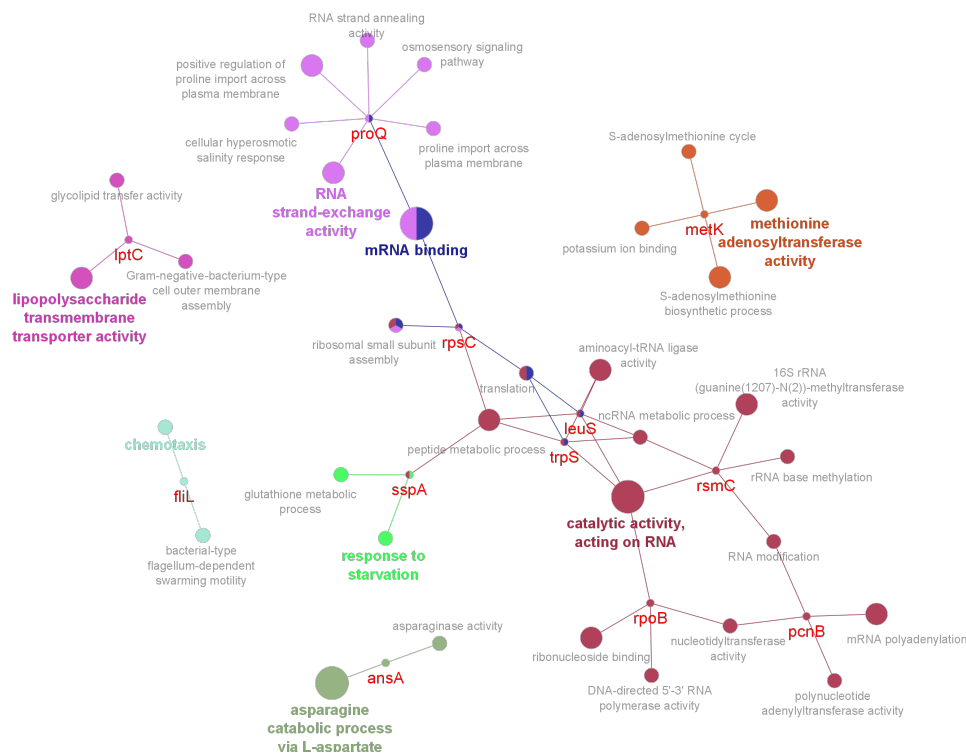
Figure 48 | Functional network of genetic variations in the 6TUBX-OC lineage. Annotated genes and their encoded functions are: *proQ* (RNA chaperone ProQ), *rpsC* (30S ribosomal subunit protein S3), *leuS* (leucine−tRNA ligase), *trpS* (tryptophan−tRNA ligase), *rsmC* (16S rRNA m²G1207 methyltransferase), *pcnB* (poly(A) polymerase I), *rpoB*

(RNA polymerase subunit β), *sspA* (stringent starvation protein A), *ansA* (L-asparaginase 1), *flilL* (flagellar protein FliL), *lptC* (lipopolysaccharide transport system protein LptC) and *metK* (methionine adenosyltransferase).[131] The functional network was visualized with Cytoscape[424] and analyzed with the ClueGO[425] app using the GO database from 04.11.2021.

In contrast, the functional network of the final isolates of 6TUB165-MB4 and 6TUB165-MB3 is much more complex, and there is no clear rationale recognizable attributed to the adaptation process (**Figure 49**). Owing to the hypermutation feature, many of the found genetic variations will be hitchhiking mutations that resulted and persisted due to a disabled DNA repair mechanism (see paragraph below) and that not necessarily direct contribute to the adaptation process.
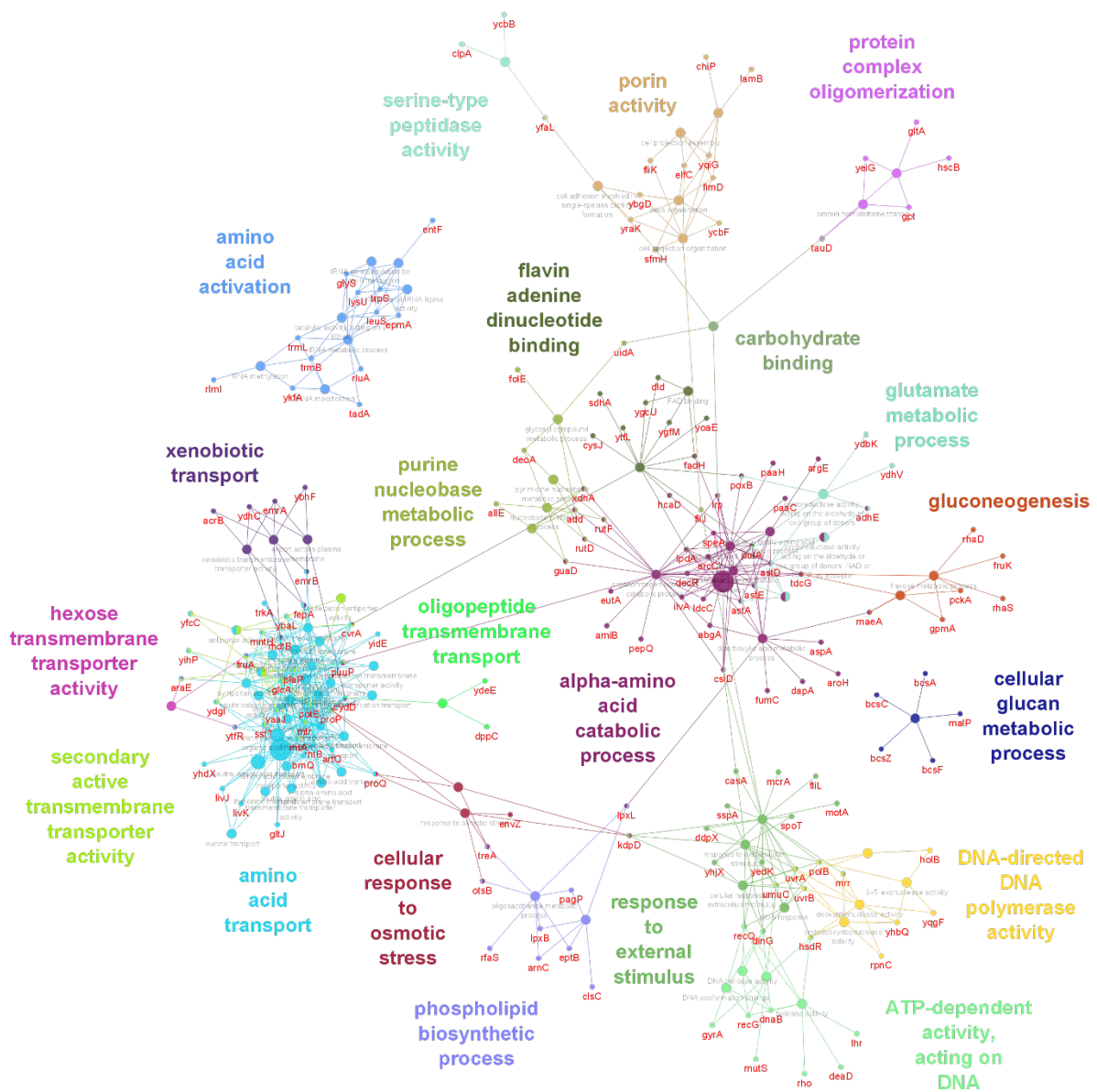
Figure 49 | Functional network of genetic variations in 6TUB165-MB4 and 6TUB165-MB3. The annotated genes belong only to the final isolates of respective lineages and are listed in **Table S34** and **Table S35**. The functional network was visualized with Cytoscape[424] and analyzed with the ClueGO[425] app using the GO database from 04.11.2021.

Hypermutable (or mutator) microorganisms have an increased spontaneous mutation rate (**Figure 50A, D**) due to defects in DNA repair or recombination systems (RecBCD).[434] In particular, the methyl-directed mismatch repair (MMR) system and within it the *mutS* gene is affected, whose product, the DNA mismatch repair protein MutS, is responsible for mismatch recognition. Accordingly, if the *mutS* gene is dysfunctional, DNA damage cannot be detected and repaired. The development of spontaneously emerged mutator strains is also known from experimental adaptation experiments and was found to be a consequence of genetic hitchhiking in association with beneficial, adaptive mutations that defect the MMR system.[435] Nevertheless, the exact mechanism leading to the selection of mutators is still controversially discussed, and the evolution of mutation rates has been extensively reviewed elsewhere.[436-439] Not only the number of mutations, and therefore also an increased probability to acquire beneficial mutations, but also the type of mutations appears to be a relevant factor driving the selection process.[440] Although high mutation rates are generally associated with more likely deleterious mutations than beneficial ones,[441] hypermutability is also considered as mechanism for rapid adaptation to (extreme or life-threatening) stresses, e.g., ethanol or antibiotics.[442-444] However, in the end, the long-term survival of microorganisms facing changing environments and challenges is a balancing act between robustness and evolvability.

Remarkably, the emergence of the hypermutation phenotype occurred independently in the lineages 6TUBX-MB4 and 6TUBX-MB3 sometime between passages 88 and 128, indicated by a jump in the number of observed mutations (**Figure 50A, D**). At this time point the lineages were already long time separated from its parental progenitor 6TUB73-MB. In order to narrow down the exact time of the first occurrence and to get a more accurate picture of the progression and cause of the development of this hypermutator trait, more close-meshed samples of both strains were sequenced (at the beginning of every fifth to seventh passage and towards the end of every 10th passage). The course of the accumulation of mutations in each sample is shown in **Figure 50B, E**. In both lineages, the cause of the hypermutation is an apparent defect in the *mutS* gene, which in both cases must have occurred between passages 88 and 93 or 88 and 95. More specifically, it was the frameshift variant c.1057delG p.Ala353fs in 6TUB93-MB4 and missense variant c.515G>C p.Arg172Pro in 6TUB95-MB3. Frameshift variants are, in general, considered deleterious, and the Arg172Pro variant was predicted to be impactful (by mutfunc[429]) because it occurred at a conserved region. Most of the time, these specific SNPs in *mutS* were conserved, but in the 6TUBX-MB4 lineage, the sample of 6TUB113-MB4 exhibits an additional frameshift variant c.1314_1315delGC p.Leu439fs, sample 6TUB128-MB4 had instead the missense variant c.515G>C p.Arg172Pro, then again the initial Ala353fs SNP in passages 138, 148 and 158, followed by the missense variant c.896T>C p.Val299Ala in the final isolate 6TUB165-MB4. In the 6TUBX-MB3, lineage the Arg172Pro SNP changed temporally to the missense variant c.128A>G p.Asp43Gly in passages 130 and 165.
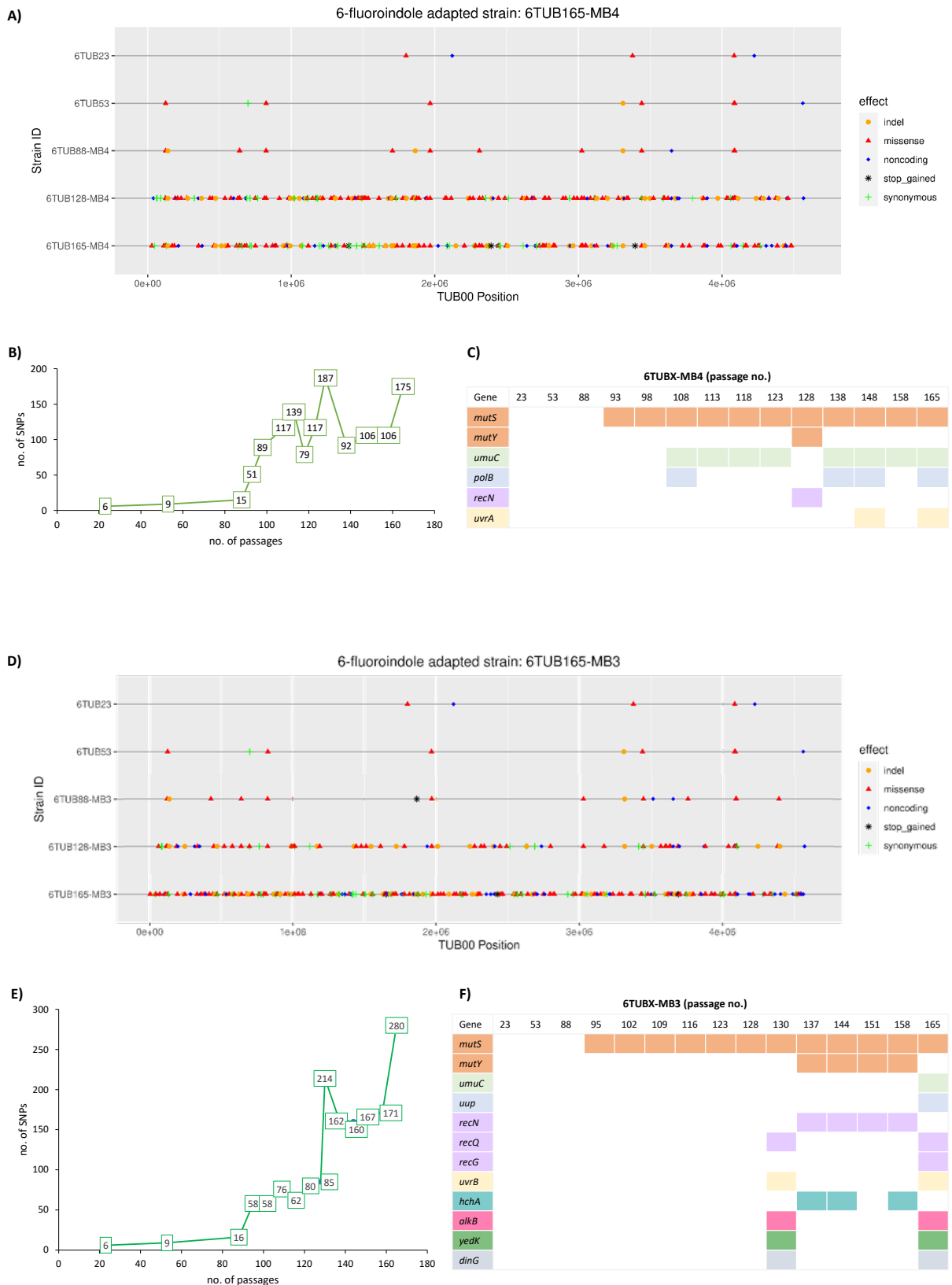
Figure 50 | Emergence of hypermutator lineages 6TUBX-MB4 (A – C) and 6TUBX-MB3 (D – F). A) and D) show the type and position of SNPs that occurred in passages 23, 53, 88, 128, and 165, respectively. B) and E) show the progression of mutation accumulation (number of SNPs), and the affected genes associated with DNA repair are shown in C) and F).

Using the parent GO term "GO:0006281 - DNA repair", which is defined as "the process of restoring DNA after damage" and comprises 98 term members, the genomic datasets were analyzed for further affected genes. Indeed, both hypermutators accumulated SNPs in genes related to this GO (**Figure 50C, F**). In the 6TUBX-MB4 lineages starting with 6TUB108-MB4, they comprised the genes *mutY* (adenine DNA glycosylase is a base excision repair enzyme which catalyzes the removal of adenine when it is paired with the oxidatively damaged form of guanine), *umuC* (DNA polymerase V catalytic protein specialized for translesion synthesis), *polB* (DNA polymerase II for translesion synthesis and nucleotide excision repair), *recN* (DNA repair protein RecN repairs DNA double-strand breaks) and *uvrA* (UvrABC excision nuclease subunit A).[131] In particular, the *umuC* synonymous variant c.1212A>G p.Arg404Arg was conserved in this lineage. Together with *polB* (synonymous variant c.831T>C p.Gly277Gly), they facilitate translesion DNA synthesis, which is an emergency mechanism of normal DNA synthesis triggered by DNA damage. While standard polymerases stop replication when they encounter a damaged nucleotide, these special polymers can continue DNA synthesis despite a damaged template, which ensures genome duplication and, thus, cell survival.[445]

In the 6TUBX-MB3 lineage, initially only the *mutS* missense variant c.515G>C p.Arg172Pro occurred as DNA repair-related mutation until passage 128, with the number of mutations ranging between 58 and 85. Then, in 6TUB130-MB3 additional SNPs occurred in *recQ* (ATP-dependent DNA helicase RecQ that blocks illegitimate recombination), *uvrB* (UvrABC excision nuclease subunit B), *alkB* (DNA oxidative demethylase which catalyzes the non-mutagenic repair of methyl lesions in DNA and RNA), *yedK* (genome maintenance protein) and *dinG* (ATP-dependent DNA helicase DinG implicated in DNA repair).[131] Thereupon, the mutation rate continued to rise sharply to 214, stalled at approximately 160 when only *mutS*, *mutY*, *recN*, and *hchA* were affected and raised again with full repertoire of mutation targets until a maximum of 280.

At this point, only causative mutation targets, namely those concerning DNA repair mechanisms, were discussed. The multitude of other mutations (in total 175 SNPs for 6TUB165-MB4 and 285 SNPs for 6TUB165-MB3) and their consequences cannot be estimated at this stage and without further comprehensive analyses. Their interpretation will turn out to be very complex because phenomena such as epistasis and pleiotropy will complicate this process due to the large number of SNPs alone.

Furthermore, the fact that the two MB lineages, of all things, evolved into hypermutators independently of each other raises questions. Is there a rationale behind the fact that both MB lineages of the 6Fi ALEs were affected, or is this just an incredible coincidence? Why were these lineages affected and not the OC lineage of the 6Fi ALE? Why not the lineages adapted to 7Fi, where the need for a radical adaptation mechanism seems more justified given the difficulties of their adaptation process (since they were undoubtedly closer to death)?

3.8.3 Stress response as adaptation target

One mutation target attracted particular attention; the *sspA* gene was affected by mutation in all adapted strains and, that already during a very early phase of adaptation. In all cases, different frameshift mutations (Val59fs, Met76fs, Lys146fs) occurred by insertion as well as deletions that altered the reading frame and, thus, the amino acid sequence (**Figure 51**). Here, they even led to the formation of stop codons that resulted in partially drastic truncated gene products (e.g., with Val59fs, Met76fs), which accounts likely for the malfunction of the SspA protein. Indeed, SspA (NP_417696) was identified in proteomics with three peptides in both positive controls and the TUB00 ancestral but not in the adapted strains. Its complete absence could have drastic consequences for the cells.
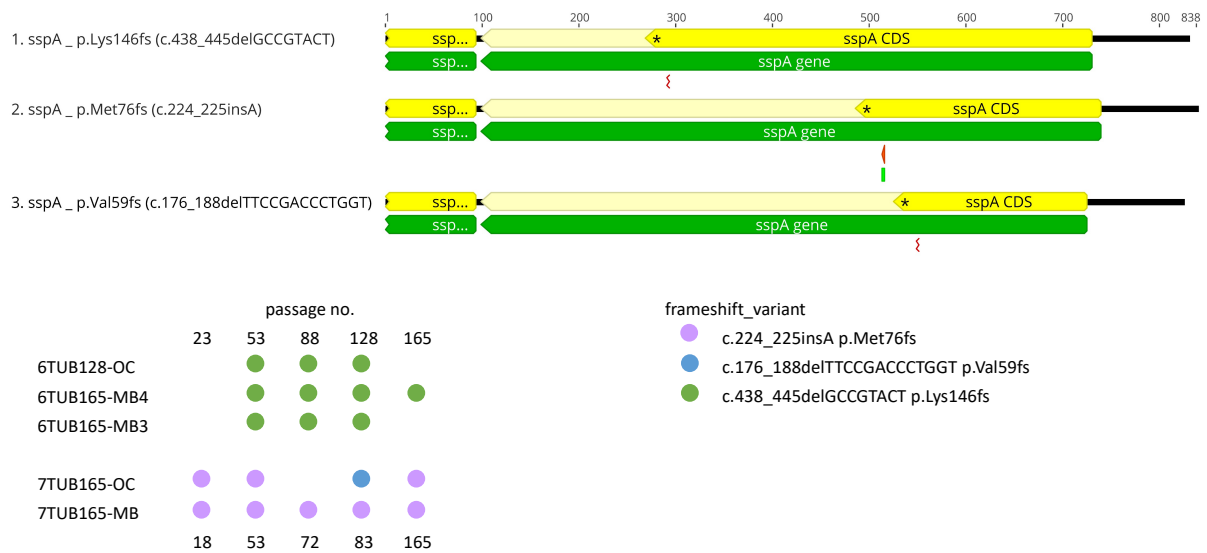
Figure 51 | Illustration of the different frameshift mutations in *sspA* and their occurrence in the respective ALE isolates. Asterisks mark the formation of a stop codon, and the pale-yellow sequence behind indicates the truncated gene. The figure was created using Geneious Prime 2023.2.1.

The *sspA* gene encodes for stringent starvation protein A (SspA) and is part of the stress response system of *E. coli*. Upon depletion of glucose, nitrogen, phosphate, or amino acids and accompanied accumulation of stringent response alarmone (p)ppGpp,[446] *sspA* expression is induced during stationary phase growth. It acts as a regulatory protein that modulates the RNA polymerase (RNAP) activity and is involved in survival during nutrient starvation and prolonged stationary phase.[447,448] Bacterial transcription is initiated by the RNA polymerase holoenzyme, which consists of the RNAP core enzyme and a sigma factor (σ). The primary sigma factor D (σ70, RpoD) mediates transcription initiation of most housekeeping genes required for growth under normal conditions, whereas alternative sigma factors such as sigma S (σ38, RpoS) control transcription of stress-related genes in response to metabolic and environmental signals.[449] Under stress-induced conditions, SspA binds to

the *E. coli* σ[70]-RNAP holoenzyme and inhibits its transcriptional activity by preventing promotor escape, thereby the transcription of housekeeping genes is suppressed. The binding proceeds by forming a bridge between the β'-subunit (RpoC) and the σ[70], whereby the interaction occurs in one part via the zinc-binding domain (ZBD) of RpoC and in the other part via the conserved PHP motif (P84-H85-P86) of SspA with σ[70].[450] The complex is additionally stabilized by the N-terminal helix of the β' subunit (β'NTH) and the C-terminal helix of the ω-subunit of RNAP.[451] Thus, through the SspA-mediated suppressing of σ[70]-RNAP, the competing σ[38]-RNAP can dominate, and the cell enters a kind of survival mode, which is hallmarked, for example, by reduced growth.

Besides regulation of a set of other stress-associated proteins, including the global regulator H-NS (Histone-like nucleoid structuring protein),[452] SspA also has implications in DNA damage response because it affects the expression of *iraD* (anti-adaptor protein IraD).[453] IraD stabilizes RpoS by inhibiting RssB (regulator of RpoS), which facilitates degradation through ClpXP protease.[454] Furthermore, *sspB* (ClpXP protease specificity-enhancing factor), which is part of the *sspAB* operon, also enhances ClpXP peptide proteolysis of incomplete synthesized proteins. [455]

However, since *sspA* was found to be deactivated in the adapted strains, this specific stress response mechanism is impaired. Thus, σ[70]-RNAP continues to drive the regulatory network, and the deleterious effects of FTrp are ignored. In addition, *sspA* inactivation diminishes RpoS stabilization and blocks expression of *sspB*,[448] therefore, ClpXP peptide degradation is negatively affected. The acquisition and fixation of *sspA* frameshift mutations, as well as the drastic consequences, led to the assumption that addressing *sspA* is part of the adaptation mechanism. Therefore, mutations of the *sspA* gene can be seen as antagonistic pleiotropy being beneficial in one and detrimental in the other environment, similar as described for mutations of *rpoS* and *mutS*.[456] Such cells could have higher growth rates and are likely to be selected in the ALE experiments. To probe the influence of dysfunctional SspA, its encoding gene was deleted in the ancestral strain TUB00 by CRISPR/Cas9,[457] and comparative growth experiments were conducted (see section **5.12**). For simplicity and because SspA appears completely absent in the adapted strains, the entire gene was removed instead of introducing the respective frameshift mutation. The reintroduction of ALE mutations into starting strains has already been proven to be a valuable tool for examining their effect.[410,458]

TUB00, the deletion mutant TUB00 Δ*sspA*, as well as chosen adapted strains 6TUB128-OC and 7TUB165-OC, were grown NMM compositions reflecting the phases during the adaptation process. Indeed, it was found that deletion of *sspA* confers a selective growth advantage during cultivation in fluoroindole containing NMM as indicated by the approaching cell densities of SspA deficient TUB00 to those of the respective adapted strains (**Figure 52A - B**). This effect was more pronounced in 6Fi supplemented NMM (6Fi > 7Fi) and marked the more stringent the growth conditions were, as shown in the NMM0-70-0 growth curves. The improved fitness (by means of growth performance) may cause of mitigation the above-described stress response. A similar finding was very recently described by

Tolle *et. al.*[410] who reported mitigation of stress response as an adaptation strategy by testing the impact of *rpoS* (encoding the major RpoS regulon[449]) deletion.
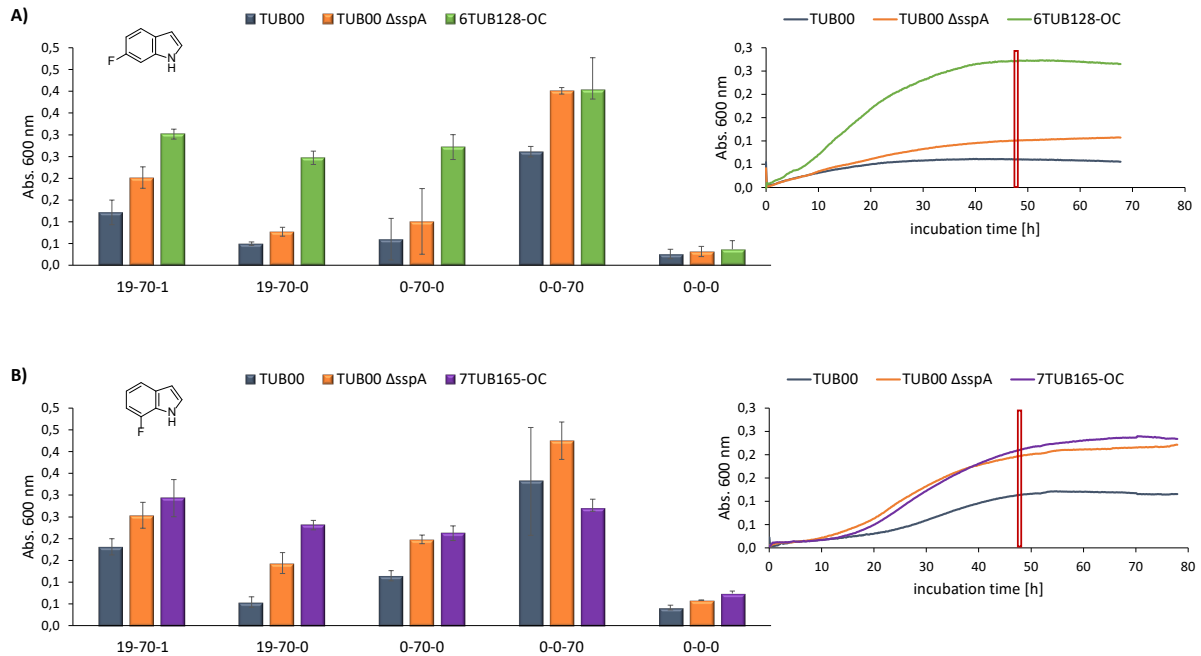
Figure 52 | Evaluation of the effect of dysfunctional *sspA* gene. The growth behavior of TUB00, TUB00 Δ*sspA* and, A) 6TUB128-OC as well as B) 7TUB165-OC in different medium compositions are shown. The bar charts show the growth at 48 h in NMM compositions reflecting the adaptation process (NMM19-70-1: ALE starting conditions, NMM19-70-0: end of ALE phase 1, NMM0-70-0: ALE phase 3, each supplemented with either 6Fi or 7Fi) as well as NMM0-0-70 (Ind) as positive control and NMM0-0-0 as negative control. In addition, the growth curves of the respective strains in NMM0-70-0 (6Fi or 7Fi) are shown. The measurements were performed by Jana Madeleine Göller during her master thesis using biological triplicates and were repeated three times. The entire dataset is deposited in **Figure S74** in the appendix.

Furthermore, SspA is also involved in nucleotide metabolism and can affect nucleoside triphosphate (NTP) synthesis when the NTP pool is altered,[459] which is interesting because nucleotide levels were decreased in the ALE isolates tested in metabolomics (section **3.7.2**).

However, the *sspA* gene is only one puzzle piece in the complex interplay of regulatory processes, and the introduction of other SNPs (also in combination) by metabolic engineering techniques on a genome scale, such as MAGE (multiplex automated genome engineering)[460,461] will expedite the knowledge gain (principle of design-build-test-learn).[462]

3.9   Integrity of the cell membrane

As a gram-negative bacterium, *Escherichia coli*'s envelope is composed of an inner membrane (IM) and an outer cell membrane (OM), enclosing the periplasmic space with its peptidoglycan cell wall that provides structural integrity to the cell (**Figure 53**).[463] Starting from the inside and proceeding outward, the first layer is given by the inner membrane. This is a classical biological membrane, i.e., a symmetric phospholipid bilayer. These amphiphilic lipids, usually glycerophospholipids, are composed of two fatty acids, a glycerol moiety, a phosphate group, and a variable head group. In *E. coli*, the principal components are zwitterionic phosphatidyl ethanolamine (PE, 75 % of membrane lipids), anionic phosphatidyl glycerol (PG, 20 %), and, to a lesser extent, anionic cardiolipin (CL).[464,465] In the inner membrane, proteins that function in energy production, lipid biosynthesis, protein secretion, and transport are located. The next layer is the periplasm, which contains a thin cell wall composed of peptidoglycan (murein).[466,467] The peptidoglycan sacculus forms a mesh-like scaffold made of linear glycan strands of alternating, β-1,4-linked *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) residues, that are cross-linked by short pentapeptides containing rare D-amino acids (L-ala-D-glu-mDAP (meso-diaminopimelic acid)-D-ala-D-ala). The cell wall serves as a rigid exoskeleton that dictates cell shape[468] and functions as a stress-bearing component to protect cells from turgor pressure of the cytoplasm. Furthermore, the thin monolayer cannot retain the gram stain and, therefore, classifies *E. coli* as a gram-negative bacterium. Surrounding the cell wall, another component that hallmarks gram-negative bacteria is the outer membrane (OM).[469] Anchored via a unique OM lipoprotein, Lpp or Braun's lipoprotein in the cell wall, the outer membrane consists of an asymmetric lipid bilayer. The inner leaflet (facing the periplasm) consists of phospholipids, and the outer leaflet (facing the extracellular space) is composed of glycolipids, principally lipopolysaccharide (LPS). LPS, in turn, is built up of three components: the hydrophobic lipid A (phosphorylated glucosamine disaccharide decorated with multiple fatty acids), a conserved oligosaccharide core, and a variable polysaccharide (a repetitive glycan polymer) called the "O-antigen". Unlike symmetrical biological membranes, which are impermeable to protons, the special architecture of the outer membrane allows diffusion of small, hydrophilic molecules such as sugars and amino acids via channels called porins but provides a permeability barrier to large, sterically hindered, and hydrophobic molecules (e.g., antibiotics). In addition, embedded transmembrane proteins, referred to as outer membrane proteins (OMPs), fulfill diverse functions such as nutrient and ion transport and signaling by receptor proteins.
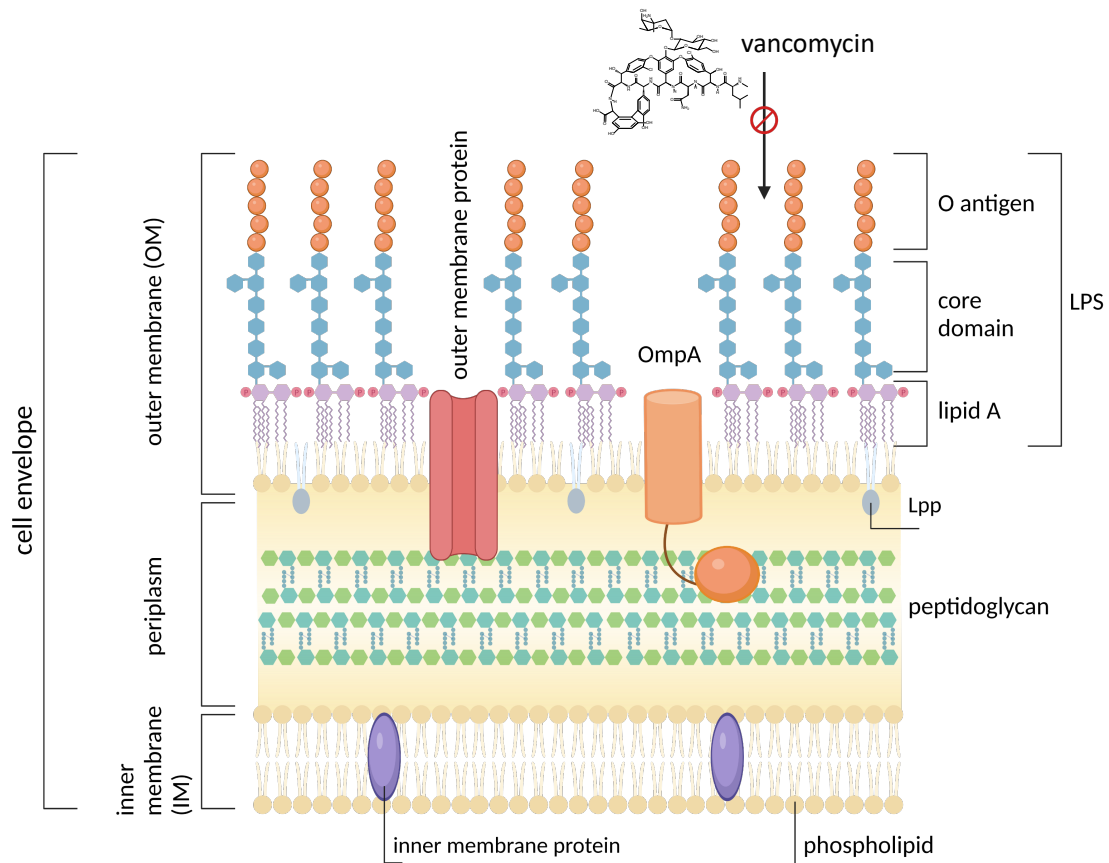
Figure 53 | The cell envelope of gram-negative *E. coli* bacteria. The cell membrane is comprised of an inner membrane (phospholipid bilayer), the periplasm composed of a structure granting peptidoglycan mesh, and the outer membrane is composed of phospholipids and lipopolysaccharides (LPS). Vancomycin antibiotic is not able to penetrate an intact *E. coli* cell membrane. The figure is re-drawn from Sun *et. al.*[469] using BioRender (https://www.biorender.com/).

The cell envelopes of bacteria are complex, dynamic structures that allow them to adapt to and protect against various environments. Thus, the composition of the lipid bilayer determines the fluidity of the membrane by stacking of adjacent fatty acids. The recruitment of saturated hydrocarbon chains confers rigidity to the membrane, while the double bonds of unsaturated fatty acids introduce kinks in the hydrocarbon chains that loosen the packing.[470,471] This basic mechanism appears, for example, when the ambient temperature changes, whereby high temperatures increase fluidity and vice versa, low temperatures stiffen the membrane. Furthermore, in the previous ALE study of Agostini *et. al.*, the membranes of 4Fi and 5Fi adapted strains (4TUB93, 5TUB83) had remodeled due to an increased content of lipids containing unsaturated fatty acids, which at the same time increased their susceptibility to antibiotics.[328] Owing to this finding, the integrity of the cell membranes of TUB00 and the adapted strains were investigated by testing their antibiotic susceptibility towards vancomycin, an antibiotic that is inactive against gram-negative *E. coli* because the penetration intrusion through the (intact) outer membrane is blocked.

3.9.1 Susceptibility and tolerance of evolved strains to vancomycin

Vancomycin is a glycopeptide antibiotic that inhibits the assembly of the structural polymer of the cell wall (peptidoglycan) of gram-positive bacteria by blocking the building blocks that are crucial for cross-linking.[472] Precisely, it complexes the terminal D-ala-D-ala motif of the peptidoglycan units through a network of five hydrogen bonds (**Figure 54**) and thus prevents polymerization (transpeptidation) and synthesis/extension of sugar chains (transglycosylation) within the peptidoglycan layer.[473,474] The lack of cross-linking weakens the cell wall, which can no longer withstand the relatively high osmotic pressure and thus leads to cell death by bursting. In contrast, gram-negative bacteria are resistant towards (hydrophilic) vancomycin because its large size, high molecular weight (~1500 Da), and globular structure prevent it from penetrating the outer membrane.[475-477] However, if the cell envelope is disordered, vancomycin intrusion is not blocked, and therefore, it can be used as an indicator for membrane damage. Indeed, very recently, vancomycin fluorescent probes were used to detect gram-negative outer membrane damage by visualization of outer membrane leakiness.[478]

The susceptibility was assessed by determination of the minimal inhibitory concentration (MIC) using the broth macrodilution method (see section **5.3**). MIC is defined as the lowest concentration of an antimicrobial agent that will inhibit the visible growth of a microorganism. Compared to TUB00 (MIC 200 µg/mL), the positive controls show increased vancomycin sensitivity (MIC 100 µg/mL). This might be a temperature-related effect, both because membrane fluidity is generally affected by temperature and because cold stress increases the susceptibility of *E. coli* to vancomycin.[479] Like strains 4TUB93 and 5TUB83,[328] cells adapted to 7Fi also show significantly increased vancomycin sensitivity (MIC 25 µg/mL), which can likely be attributed to a membrane rearrangement. Interestingly, the strains adapted to 6Fi exhibit the opposite effect: they are highly resistant to vancomycin treatment (MIC 400 µg/mL).

One possible explanation is the accumulation of phosphatidic acid, which is also a normal component of the membrane, but its increase is under suspicion to impede the entry of vancomycin and to confer resistance.[480] In the case of the adaptation to 6Fi, it might be hypothesized that an upregulation of phosphatidic acid production occurs, preventing vancomycin penetration and increasing resistance. Conversely, adaptation to 7Fi may have the opposite effect, with a down-regulation of phosphatidic acid potentially increasing susceptibility to vancomycin. However, *cdsA*, which is responsible for the conversion of phosphatidic acid to CDP-diacylglycerol in phospholipid biosynthesis, was found to be affected by a mutation (Tyr70Ser, predicted as neutral)[429] in 7TUB165-OC, of all places. In the vancomycin "resistant" strains, instead, other genes were affected, which are associated with LPS assembly and thus the membrane barrier: *lptC, lptG, lpoB, lpxB, rlpA,* and *rfaS*, as well as a synonymous mutation in *ddpX* (encodes D-alanyl-D-alanine dipeptidase, associated with vancomycin resistance)[481] found exclusively in 6TUB165-MB4.

In addition, bacteria respond to internal concentrations of antibiotics by regulating proteins involved in membrane permeability, such as porins or efflux pumps.[482] Indeed, in 6TUB165-MB4 and 6TUB165-MB3, several multidrug efflux pumps were found to be affected by mutations, including *mdtB*, *emrB*, *emrA*, *acrB*, *evgS* and *ompF*, *ompC*. However, this finding should be viewed with caution, as many passenger mutations occur in the hypermutator strains. In the end, without further investigations involving metabolic, proteomic, and genomic analysis of the adapted strains, it is challenging to provide a definitive rationale.
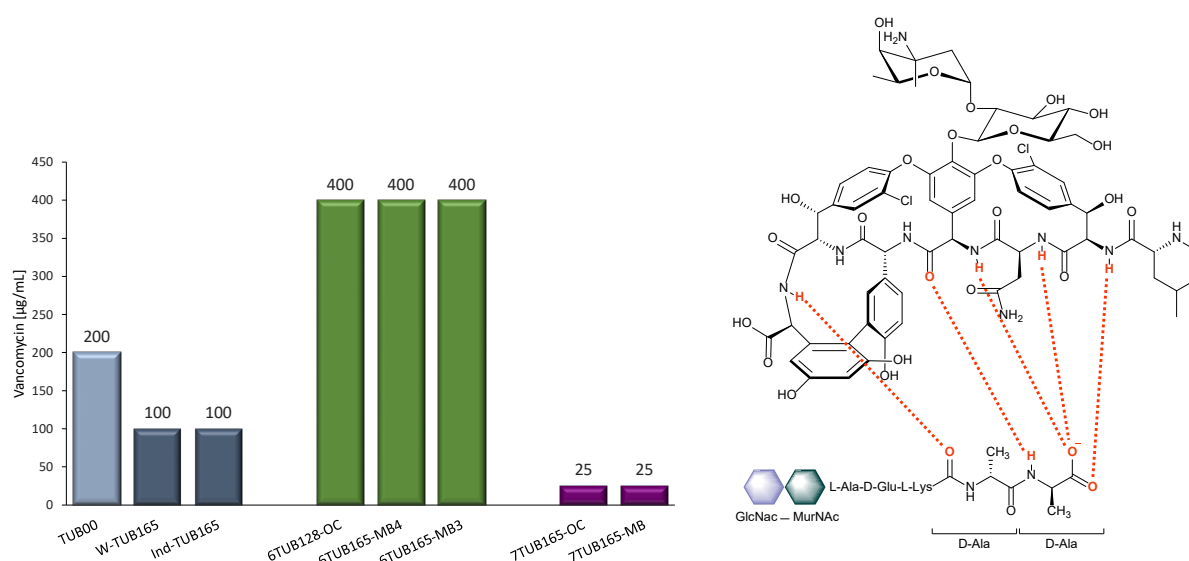
Figure 54 | Susceptibility of evolved strains to vancomycin. The structure and mode of action of vancomycin glycopeptide antibiotic are shown right. It should be noted that in the sequence of the pentapeptide of glycopeptide units in gram-positive bacteria, the middle amino acid (cAA-3) is an L-Lys instead of mDAP. The minimal inhibitory concentrations (MIC) of vancomycin are shown for the ancestral strain TUB00 (light grey), both positive controls W-TUB165 and Ind-TUB165 (grey), 6Fi-adapted strains 6TUB128-OC, 6TUB165-MB4 and 6TUB165-MB3 (green) as well as 7Fi-adapted strains 7TUB165-OC and 7TUB165-MB (purple). Cells were treated with a dilution series of vancomycin (0 – 400 μg/mL), and growth after 24 h incubation at 30 °C, 200 rpm is shown.

3.9.2 Morphology

In the context of investigating membrane integrity, attempts were also made to determine permeability by Nile Red staining. Nile Red (9-(Diethylamino)-5H-benzo[a]phenoxazin-5-one) is a lipophilic lysochrome (fat-soluble) dye that intercalates between membrane lipids. Its fluorescence properties are dependent on its immediate environment, which are $\lambda_{ex} = 515 – 560$ nm and $\lambda_{em} > 590$ nm in a hydrophobic solvent, while fluorescence is quenched upon exposure to polar solvents such as water.[483] Thus, cells with an intact cell envelope will retain the dye and fluoresce red, whereas those with a damaged/leaky membrane will not fluoresce since Nile Red is exposed to the polar buffer.

Cells of TUB00 and all adapted strains were grown in their respective NMM (positive controls in Ind or Trp, 6TUBX in 6Fi, and 7TUBX in 7Fi) with and without the addition of 19 cAAs, respectively, in order to assess the influence of amino acid deprivation-induced stress on the membrane permeability; followed by Nile Red staining (section **5.13**). The determination of membrane permeability in a quantitative sense by spectroscopic measurement of fluorescence intensity failed or did not yield meaningful results. For example, immense differences (up to 8-fold increased intensity) were observed within the positive controls and when comparing them to TUB00. Given that these strains evolved in a native medium, such a difference in membrane properties seems very unlikely. By comparison, the adapted strains showed less drastic differences but also anything but meaningful readings (e.g., high variance of FL intensity between cells of the same lineage that grew in NMM19 or NMM0). The measurements are deposited in the appendix (**Figure S75**).

However, in the course of this experiment, fluorescence micrographs of the stained cells were also taken, providing an interesting view of cell morphology (**Figure 55**). In total, three phenomena were qualitatively identified. 1) except for 6TUB165-MB4 and 6TUB165-MB3, all adapted strains showed remarkable heterogeneity of cell sizes within the same cell sample. 2) cells of 7TUB165-OC and 7TUB165-MB, and in particular those of 6TUB128-OC, exhibit cell elongation, whereby 6TUB128-OC cells in some cases appeared even 2 – 3-fold longer than TUB00 wild-type cells. And 3) cells of 6TUB165-MB3 have a rounder, cocci-like appearance.

Apart from these observations, no differences in cell shape are observed between NMM0 and NMM19 cultured samples. However, using TUB00 as a benchmark, the cells are smaller than typical *E. coli*, which are approximately 5 μm long. This may be attributed to growth in minimal medium, as the cell size of heterotroph *E. coli* correlate positively with nutrient supply (minimal medium vs. rich medium such as LB). Generally, morphology is affected by central carbon metabolism because of its impact on cycle progression, lipid synthesis, and cell wall biogenesis.[484] The elongated cells may have developed as a stress reaction and temporarily stopped cell division as a survival mechanism. In this anomaly growth, known as filamentation, the cells continue to expand, but no cell division (septa formation) occurs. Such cells can have multiple chromosome copies and complete cell division after resolving the stress situation.[485,486]
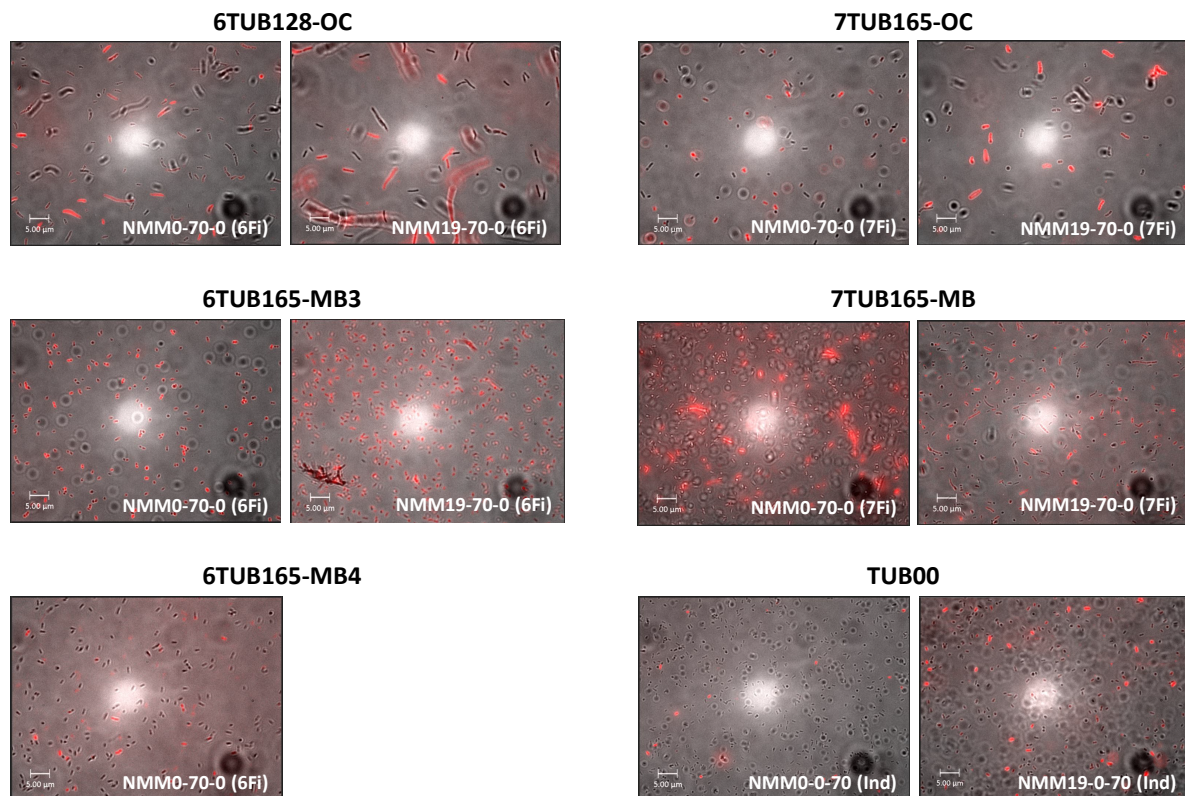
Figure 55 | Morphology of TUB00 and fluoroindole-adapted cells. Overlays of fluorescent micrographs (λ_{ex} = 549 nm) and white light irradiation are shown. Cells with a red appearance retained the Nile Red dye, however, staining seems unspecific and thus cannot be correlated to membrane permeability. For 6TUB165-MB4, only a micrograph for NMM0-70-0 (6Fi) grown cells is available.

Of all the strains examined, 6TUB165-MB4 most closely resembles the typical appearance of rod-shaped *E. coli* cells, as in the ancestral TUB00. Together with the results of the growth studies, this is a further indication of its superior adaptation. In addition, the visual insight through the morphology shows very impressively the differences between individual lineages despite adaptation to the same substrate, in particular for 6TUB165-MB3, 6TUB165-MB4, and 6TUB128-OC. Nevertheless, the micrographs shown are only snapshots that must be confirmed by statistically valid measurements such as imaging flow cytometry (IFC).[487] Indeed, a combination of FACS-based selection (fluorescence-activated cell sorting) with IFC analysis was recently applied to guide experimental evolution.[488]

3.10 Expression of 6FTrp and 7FTrp substituted EGFP and ECFP

The 6Fi- and 7Fi-adapted strains were tested for their robustness in driving recombinant protein expression.[489] Two different variants of the green fluorescent protein (GFP)[490] model protein were used for heterologous protein expression: 1) the enhanced green fluorescent protein EGFP, which harbors one single Trp residue at position 57 in its primary sequence, as well as 2) the enhanced cyan fluorescent protein ECFP, which also contains a Trp in the chromophore at position 66 (besides position 57). The structures of both proteins and the relevant amino acids are shown in **Figure 56**. The GFP variants denoted as "enhanced" differ from wild-type GFP by carrying some mutations that improve folding and chromophore maturation at 37 °C. These mutations are F64L, S65T, and V163A in EGFP and in ECFP beside F64L, S65T, V163A further mutations N146I and M153T in the chromophore cavity (*p*-hydroxybenzylidene imidazolinone formed from residues T65-Y66-G67) compensate for the increased bulk of the indole moiety of Y66W to restore the intrinsic brightness.[490,491]

Cells of the final isolates of all adapted strains (6TUB128-OC, 6TUB165-MB4, 6TUB165-MB3 and 7TUB165-OC, 7TUB165-MB) were transformed with a plasmid bearing the sequence of either His-tagged EGFP (C-terminal EGFP-H6) or ECFP (N-terminal H6-ECFP). Recombinant protein expression was performed under the same conditions as used for adaptation, i.e., cultivation in NMM0-70-0 containing either 6- or 7-fluoroindole at 30 °C, followed by isolation and purification using Ni-NTA affinity chromatography and analysis via LC-ESI-MS (see section **5.6.2**, **Table S36** and **Figure S76**). The ancestral TUB00 was used as a positive control, expressing wt-ECFP and wt-EGFP using NMM0-0-70 (Ind).
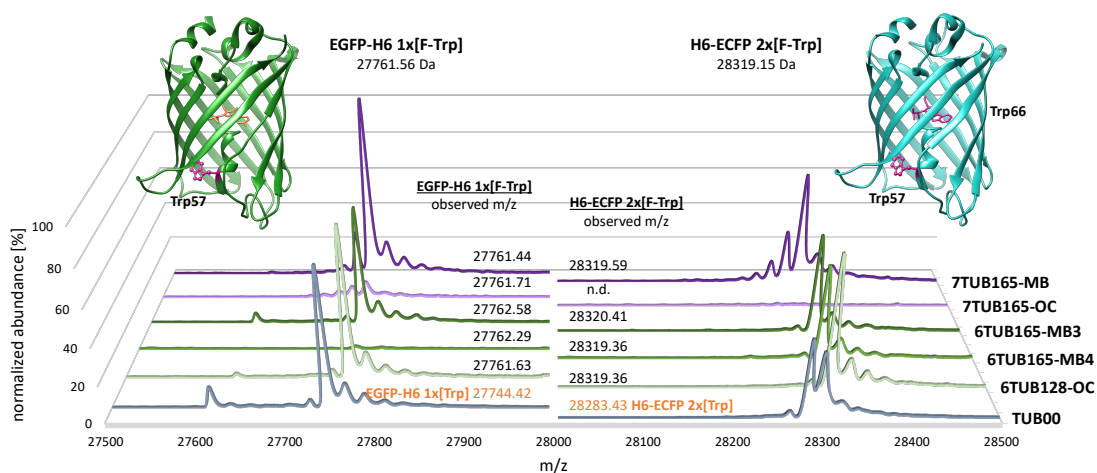
Figure 56 | Global incorporation of 6FTrp and 7FTrp into ECFP and EGFP. Deconvoluted mass spectra of C-terminal His-tagged enhanced green fluorescent protein (EGFP-H6, left) and of N-terminal His-tagged enhanced cyan fluorescent protein (H6-ECFP, right) are shown. MS spectra of expression products of 6Fi-adapted strains (6TUB128-OC, 6TUB165-MB4, and 6TUB165-MB3) are colored in green, and respective spectra of proteins expressed with 7Fi-adapted strains (7TUB165-OC and 7TUB165-MB) are colored in purple. EGFP-H6 and H6-ECFP expressed in TUB00 are colored in grey and have calculated masses of 27743.79 Da and 28283.38 Da, respectively.

The structures of EGFP (PDB ID 2Y0G)[492] and ECFP (PDB ID 2WSN)[493] from *Aequorea victoria* were drawn using UCSF Chimera;[43] Trp-positions are highlighted.

Almost all adapted strains could produce GFP variants in which all Trp residues are quantitatively replaced by fluorinated counterparts (**Figure 56**). The mass spectrometric analyses of 6FTrp- and 7FTrp-substituted EGFP and ECFP exhibit homogeneous mass profiles and prove unambiguous incorporation of the respective fluorinated Trp analogs into the target proteins, indicated by a mass shift of 18 Da per FTrp (which correlates with a H-to-F exchange on the aromatic ring of the indole). This mass shift can also be observed when compared with the wild-type variants of ECFP and EGFP containing native Trp. The expected theoretical values for translated EGFP and ECFP after chromophore maturation[494] are consistent with the observed masses (**Figure 56**). However, the abundance of 6FTrp labeled EGFP expressed with 6TUB165-MB4 was very low, and the 7TUB165-OC strain could either merely produce small amounts of 7FTrp-EGFP or was even unable to afford 7FTrp-ECFP at all. Furthermore, in addition to the dominant mass of the full-length protein, the spectra of substituted EGFPs exhibit a second signal that is assigned to a truncated product after processing of the initial methionine (masses: 27630 Da for FTrp-EGFP and 27613 Da for Trp-EGFP, Δm/z -131 Da). Owing to the fluorine substitution Y66W in the catalytic triage of ECFP, a slight change in the fluorescence profiles is expected.[495] The successful expression confirmed that the protein synthesis was reprogrammed in the adapted strains as Trp was completely replaced by 6- and 7-fluorotryptophan, respectively. Overall, the protein translation machinery of the adapted strains works perfectly with fluorinated analogs. Therefore, these findings can be extrapolated to the whole proteome: there should be no doubt that not all Trp residues in the proteome of the adapted strains are completely replaced by 6- or 7-fluorotryptophan, which was also proved by proteomics (see section **3.6**). In this way, the biological generation of fluorine-containing biomass becomes a reality, and fluorine comes closer and closer to being the element of life.

4   Conclusion and future perspectives

This study aimed to permanently integrate organofluorine into living organisms through the global incorporation of fluorinated tryptophans. By applying adaptive laboratory evolution (ALE), *Escherichia coli* was successfully adapted to the use of 6- and 7-fluoroindole, i.e., the adapted cells are able to convert them into their fluorinated counterparts 6- and 7-fluorotryptophan and to incorporate them into the proteome consequently. In the course of the described adaptation process, the cells relinquished their dependence on canonical Trp and instead acquired the ability to use the fluorinated counterparts as an integrated part of their metabolism. This once again demonstrates the feasibility of assimilation of fluoroindoles and trophic reassignment of the UGG codon by fluorinated Trp.[328,332,343] Furthermore, the present adaptation experiments towards 6Fi and 7Fi convincingly demonstrate the reproducibility of the previous study on the adaptation towards 4Fi and 5Fi[328] as well as of the ALE process itself. Hence, a reliable platform and methodologies for the experimental evolution of novel organisms with fluorine as a bioelement were established, created by mimicking the process of natural selection in the laboratory without external influences like mutagens.[496]

The adaptation process was expanded and further developed by adding, for the first time, two positive controls long-term cultivated on either Trp or indole, as well as a new approach for amino acid removal. The positive controls will facilitate differentiation between general consequences of long-term propagation in minimal medium and those that are caused by the adaptation to different fluorinated metabolites. They can also be consulted for other Trp-based ALEs[327,328] since the same adaptation setup was conserved. The novel OC approach for amino acid removal provides an excellent alternative for the adaptation of challenging substrates and might be more suitable to apply for already stressed cells, as shown in the case of 7-fluoroindole. This straightforward technique is more comfortable in laboratory handling, but in addition, it seems that the metabolic burden to the cells of having to biosynthesize missing amino acids is reduced upon application of this method (compared to the MB approach) since they were stepwise convinced to do so. Nevertheless, both approaches led to the desired adaptation, notwithstanding the different cellular outcomes for instance that lineage 7TUB165-OC clearly exhibits better growth performance than 7TUB165-MB, but that in terms of recombinant protein expression capability, an opposite tendency is observed. Therefore, one approach cannot be single out over the other, but the two techniques are rather to be understood as independent alternative approaches to increase the chance of adaptation.

The conducted ALE experiments resulted in five differently adapted independent strains that grew stably for over 400 days, 165 passages, and about 1000 generations. Based on the adaptation process itself and subsequent characterization experiments, considerable differences in the acceptance for 6Fi and 7Fi were observed. 6Fi is unambiguously better tolerated by *E. coli* because the 6Fi-adapted lineages reach higher optical densities of about $OD_{600} = 2$ and higher specific growth rates. In contrast, 7Fi represents a considerable challenge evident through the intricate nature of the adaptation process and

the compromised growth performance (population sizes of about $OD_{600} = 0.5 - 0.8$ and lower growth rates). These differences are attributed to the physicochemical properties of the constitutional isomers, such as polarity and lipophilicity.[328]

Contrary to 4-, 5- and 6-fluoroindole/Trp, the Ind/Trp isomer fluorinated at position 7 is so far rather underrepresented in studies on biosynthetic incorporation, and only a few have investigated the differences of all four isomers in specific proteins in direct comparison.[495,497–499] Why the 7-fluorinated isomers have not made much of an appearance so far remains speculative because the chemical synthesis and use of both 7Fi[360] and 7FTrp[191,359] were already described decades ago. Moreover, the proteome-wide insertion of 7-fluorotryptophan (by cellular conversion of 7-fluorindole) described here represents the first study of its kind involving this non-canonical amino acid. Due to the described challenges of the 7Fi ALEs, exciting opportunities lie ahead in future studies to unravel the molecular factors responsible for these difficulties, which likely stem from the chemistry and structure of this analog and its impact on cellular metabolism. Potential areas of investigation include its effects on metabolic pathways and the influence on protein folding quality for proteins containing 7-FTrp.

The adaptation experiments towards the usage of 6Fi, on the other hand, showed an astonishing readiness of the cells for this isomer. Here, the most notable outcomes were the emergence of a subpopulation within the MB lineages that had started to become nearly independent of indole and even prefer 6Fi and the emergence of independently developed hypermutator strains. The 9-fold and 15-fold increased mutation rates are a consequence of different defects in the DNA mismatch repair system of *E. coli*. Equally astonishing was the finding that all 6Fi-adapted lineages showed increased resistance to vancomycin, which is unique among all other 4-, 5-,[328] and 7Fi-adapted strains.

In addition, genomics identified central mutation targets, and the involvement of the SspA-mediated stress response system as part of an adaptation mechanism towards the usage of fluorinated indole was confirmed. The chromosomal deletion of the *sspA* gene by CRISPR/Cas9, and therefore deactivation or relaxation of the stress response, conferred a growth advantage that could be partly responsible for driving the adaptation process. This is in accordance with other ALE experiments that revealed the mitigation of RpoS-mediated stringent response system as a driving force.[410]

Proteome-wide substitution of Trp by FTrp was demonstrated for all fluoroindole-based ALEs. However, although *E. coli* had acquired the competence to use the fluorinated substrates as integrated metabolic intermediates, the adapted strains remained (still) facultative fluorotryptophan/tryptophan users. That is, at least for now, ALE led to broader Trp/FTrp analog use rather than changed specialized use. In addition, they appear to be exclusively specified to their adapted indole analog since no growth advantages of the adapted cells over their other mono-fluorinated analogs could be observed, that is, 6TUBX strains on 4-, 5- or 7Fi or other substituted indole analogs.[343,345] The TrpRS could play a key role here, as it is the decision-maker in incorporating the respective amino acid Trp or ncAA FTrp. According to Schmidt *et. al.*[376], the chance of adapting unconditionally using FTrp strains is highest for 7Fi because

it exhibits the highest distance to Trp among the set of mono-fluorinated Trp analogs. Indeed, this adaptation was more challenging probably due to this chemical distance (esp. dipole moment), but on the other side, a stably adapted organism might also have a lower chance to regress and would someday result in a "really" alienated organism. Either the way, the option to continue the evolution of the 6Fi- and 7Fi-adapted strains is given since the experiment can be revived from the comprehensive fossil database (preserved cryo-stocks of nearly every passage).

Furthermore, the fluorine-adapted strains proved to be appropriate expression hosts to readily produce FTrp-substituted proteins and peptides, which opens up the opportunity to go beyond and overcome limitations of automated solid-phase peptide synthesis (SSPS) by harnessing the ribosomal synthesis of adapted cells.[432] However, the application of classical molecular biology methods such as amber stop codon suppression (SCS) and selective pressure incorporation (SPI)[236] also benefit from the abolished dependence on canonical amino acids of the adapted strains, allowing improved and effective synthesis of fully fluorine-labeled proteins.

Although largely ignored by nature, a life based on fluorine is both an interesting concept and an absolutely conceivable scenario.[79] Fluorine-containing building blocks have been extensively used to investigate and modify proteins and their interactions.[80-82] However, the adaptation of living organisms represents a step forward in exploring and understanding the consequences of global fluorination. At this point, the outstanding achievements of a cell to cope with the intricate and myriad consequences associated with the exchange of a hydrogen against a fluorine atom in an essential building block (after all, more than 20000 Trp positions only in the proteome) should be highlighted.[410] Ultimately, this affects the complex interplay of cellular processes that include not only protein interactions through incorporated FTrp, such as π-π stacking, hydrogen bonding, and cation-π interactions,[108] but also all metabolic processes to which Ind as a signaling molecule[114] and Trp or downstream metabolites contribute to. Ultimately, the characterization experiments described in this thesis have once again impressively demonstrated that the effects and consequences of fluorination can hardly be generalized but are highly context dependent.

These adaptation experiments using fluorinated Trp precursor will contribute to a better understanding of the mechanism required to enable life with this extraordinary element. With this study, the set of *Escherichia coli* bacteria adapted to all mono-fluorinated indoles (4-, 5-, 6- and 7Fi) was completed and proved them to be suitable metabolic substrates for the assembly of fluorinated proteomes. Based on this achievement now, the influence of the position of the F-substituent on cellular adaptability, metabolic pathways, regulatory networks, and protein folding quality can be investigated and correlated with the physicochemical properties of the constitutional isomers, such as polarity and lipophilicity.[328] Even a cursory comparison of the cultivation regimes of all four substrates shows parallels in the adaptation behavior of 4Fi and 6Fi, which are well tolerated, while 5Fi and 7Fi present adaptation challenges. Currently, MD simulations on the interaction energies of fluorinated Trp with its most common interaction partners Arg, Lys, Phe, Tyr, and Trp itself are in progress. Not the direct

influence of the position of the F-substituent, but the preference for one of the four isomers could be investigated by the cells themselves using competition experiments. Starting with considering whether the Trp processing enzyme TrpS and TrpRS already show tendencies for one of the substrates, a kind of competition ALE could also be conceivable. The obstacle of discrimination of the four isomers, which have the same mass, could be tackled by the usage of isotope-labeled substrates. This technique is already established in the field of expression proteomics by means of SILAC.[350] And the synthesis of $^{13}C/^{19}F/^2H$ labeled indoles via ruthenium-mediated ring-closing metathesis is readily available.[500] Instead of a strict bias in favor of one Fi variant, a heterogeneous use would be conceivable, depending on the environment of the respective Trp in the protein network.

This study focused on improving protocols and procedures and provided a reliable empirical determination of parameters for any ALE experiment on microbial adaptation to non-canonical amino acids. The procedures established here could become one of the main routes in engineering synthetic cells with a chemically modified makeup of life. Further multi-omic experiments will unravel a rationale behind the adaptation towards xenobiotics and fluorine in particular. Enlightening the responsible mechanisms will push advancements in scientific fields such as synthetic biology and biotechnology, but also across the board of environmental concerns such as bioremediation and biocontainment.[501]

After the integration of fluorinated building blocks into a living organism was achieved, now the time is up to elucidate its fate within the cell, both of the fluorinated amino acid FTrp and of its precursor Fi. Since indole has a variety of functions as a signaling molecule, the question arises to what extent fluorinated indole can take over these functions or change them. For example, do fluoroindole-fed cells experience indole pulse signaling, or can FTrp also serve as a suitable trigger for this phenomenon,[118] as described in section **1.2**. What happens to FTrp besides utilization in protein biosynthesis? It is known that tryptophanase-mediated degradation of Trp is slower for fluorinated variants 4,7-, 5,7-, and 7-FTrp.[359,502] Although the encoding *tnaA* gene is deleted from the genome of the adapted cells anyway, it is reasonable to assume that further enzymes are hindered by the fluorine substituent. Given the number of reactions that Trp can undergo due to its structure,[503] interference from an existing F decoration would not be surprising. One example of the relevance of the F-position is given by the aerobic Trp degradation, referred to as the kynurenine pathway, where the fluorine position highly influences the formation of terminal products and could explain the poor action of 5Fi and 7Fi (e.g., the conversion to 3-hydroxy-L-kynurenine from 7FTrp would be complicated). The kynurenine pathway is conserved in many species, including bacteria,[504] but not in *E. coli*, which only possess the aspartate aminotransferase (AspAT), which exhibits substantial kynurenine aminotransferase and glutamine transaminase K activities.[505] Besides that, why not use the intrinsic properties of FTrp as a spectroscopic probe to investigate the fate (F)Trp within the cells. All these considerations can be started with a simple experiment, namely, to clarify first whether the fluorine substituent can be cleaved off from Trp in the cell.[506]

5   Methods

All cell experiments were conducted according to German genetic engineering laws and German biosafety guidelines in the laboratory (safety level 1). Genetically modified organisms (GMOs) generated in this study are documented in Formblatt Z, in Anlage 424/95 at Technische Universität Berlin, and in Anlage 92/14 at Freie Universität Berlin.

5.1   Microbiological methods

5.1.1   Cultivation and storage of *E. coli* cells

Depending on the intended use, *Escherichia coli* batch cultures were performed in non-limiting liquid and solid media (see section **6.5**) supplemented with appropriate selection markers or other supplements and incubated at 30 – 37 °C with constant shaking at 200 – 250 rpm, if not otherwise stated. Cultivation in the context of adaptation experiments (see section **5.2**) and expression of GFP variants (see section **5.6.2**) was conducted under limiting conditions by using minimal media and other supplements (e.g., ncAAs). Long-term storage of liquid cultures was accomplished at -80 °C in 25 % sterile glycerol (250 µL 80 % glycerol and 550 µL culture).

5.1.2   Production of competent cells

Creating artificial cell competency allows the uptake of exogenous genetic material into a respective recipient (known as artificial transformation; see section **5.1.3**) and is based on a temporally increased permeability of bacterial cell membrane. The two most common methods in molecular biology are treatment with divalent cations or applying an electrical field. In chemical transformation, the bacterial cells are treated with Ca^{2+} or Mg^{2+} cations, followed by short exposure to an elevated temperature (42 °C), known as the heat shock procedure.[507,508] In the other technique, the cells are desalted to avoid short circuits during the process of electroporation (short pulses of high-voltage electric fields).[509] For standard transformations (e. g., simple plasmid uptake), chemically competent cells were used, and for transformations where high efficiency was required (e. g., uptake of large linear DNA fragments) electrocompetent cells were used.

5.1.2.1 Chemically competent cells

50 – 200 mL LB medium (optionally supplemented with appropriate antibiotics) were inoculated 1:100 with an overnight culture. The culture was incubated at 37 °C, 200 rpm, until an optical density at $\lambda = 600$ nm (OD_{600}) of about 0.6 – 0.8 was reached. The cells were chilled and, from here on, strictly handled on ice. After harvesting by centrifugation (4000 g, 10 min, 4 °C), the cells were resuspended in sterile 100 mM ice-cold $MgCl_2$ (10 mL per 50 mL culture) and then incubated on ice for 30 min. Subsequently, the cells were again harvested by centrifugation (4000 g, 10 min, 4 °C), resuspended in a solution of sterile, ice-cold 50 mM $CaCl_2$ + 15 % glycerol (200 µL per 50 mL culture), and stored at -80 °C in 50 µL aliquots.

5.1.2.2 Electrocompetent cells

50 – 200 mL LB medium (optionally supplemented with appropriate antibiotics) were inoculated 1:100 with an overnight culture. The culture was incubated at 37 °C, 200 rpm, shaking until an OD_{600} of about 0.3 – 0.4 was reached. The cells were chilled and, from here on, strictly handled on ice. After harvesting by centrifugation (4000 g, 10 min, 4 °C), the cells were washed three times with ice-cold, sterile 10 % glycerol (10 mL per 50 mL culture). Finally, desalted cells were resuspended in 10 % glycerol (200 µL per 50 mL culture) and stored at -80 °C in 50 µL aliquots.

5.1.3 Bacterial transformation

To introduce foreign DNA (e. g., plasmids or PCR products), the bacteria were transformed either via chemical transformation or via electroporation.

5.1.3.1 Chemical transformation by heat shock

For chemical transformation, 100 – 1000 ng DNA were added to one aliquot (50 µL) of competent cells (thawed on ice) and incubated for 20 min on ice. Afterward, the cells were subjected to heat shock for 90 sec at 42 °C and additionally incubated for 2 min on ice. Then, cells were resuspended in 950 µL SOC and incubated for 60 – 120 min at 37 °C, 750 rpm. Following this regeneration phase, the cell suspension was concentrated (by centrifugation 2000 g, 2 min, RT), and 50 – 100 µL were spread on LB-Agar plates with appropriate selection markers. The plates were incubated either overnight at 37 °C or for 72 h at room temperature.

5.1.3.2 Electroporation

100 – 1000 ng DNA (solved in water or dialyzed to prevent short circuits) were added to one aliquot (50 µL) of electrocompetent cells (thawed on ice). The volume of the DNA solution should not exceed 10 % of the total cell suspension volume. The mixture was placed between electrodes of a precooled electroporation cuvette (1 mm). The electroporation was carried out with a pulse generator (MicroPulser™), and an electrical pulse of 1.8 kV was applied for approximately 5 – 6 ms. Immediately afterward, cells were resuspended in 950 µL SOC and incubated for 60 – 120 min at 37 °C, 750 rpm. Following this regeneration phase, the cell suspension was concentrated (by centrifugation 2000 g, 2 min, RT), and 50 – 100 µL were spread on LB-Agar plates with appropriate selection markers. The plates were incubated either overnight at 37 °C or for 72 h at room temperature.

5.1.4 Isolation of plasmid and genomic DNA

The principle of isolation of plasmid DNA relies on the narrow pH range (about 12.0 – 12.5) where linear DNA selectively denatures, while covalently closed circular DNA (plasmid DNA) remains double-stranded.[510] Plasmid DNA is prepared from bacterial cultures by subjecting the pelleted cells to SDS/alkaline lysis that mediates liberation of the plasmid DNA, while chromosomal DNA, high molecular weight RNA, and proteins precipitate and are separated by centrifugation. After purification over a silica spin column membrane, the plasmid DNA is eluted with either 10 mM Tris-HCl (pH 8.5) or water. For this procedure, the standardized GeneJET™ Plasmid MiniPrep Kit or GeneJET™ Plasmid MidiPrep Kit (Thermo Fisher Scientific) were used according to the manufacturer's instructions.

For isolation of genomic DNA (gDNA), cells were lysed, subjected to RNase digestion, proteins were removed by salt precipitation, and finally, gDNA was concentrated and desalted by isopropanol precipitation.[511] This method avoids the use of hazardous organic solvents phenol and chloroform. For this procedure, the standardized Wizard® Genomic DNA Purification Kit (Promega) was used according to the manufacturer's instructions. Genomic DNA was used as a high-quality template for cloning procedures and for genomic analyses by means of whole genome re-sequencing.

The concentration and purity of plasmid DNA and genomic DNA was determined by UV-Vis spectroscopy (see section **5.4.4**).

5.2 Adaptive laboratory evolution (ALE) experiments

5.2.1 Determination of the ALE starting conditions

6-fluoroindole, 7-fluoroindole, and indole were dissolved in 96 % EtOH and used as a 50 mM stock solution. If not stated otherwise, cells were always cultured at 30 °C, 200 rpm in an Ecotron incubator shaker (HT Infors). For the adaptation experiments, three biological replicates of *Escherichia coli* TUB00 were freshly streaked on LB agar plates, new isolates were generated by cultivation of single clones (each one per replicate) and stored as cryo-stocks at -80 °C. The Trp auxotrophy of TUB00 was confirmed on the genetic level ($\Delta trpLEDC$::FRT $\Delta tnaA$::FRT) by colony-PCR (described in section **5.4.1.2**) and on the phenotypic level by simple cultivation procedures (see section **3.2.1**). For determination of the phenotype, TUB00 was grown in 5 mL NMM19 either supplemented with 70 µM Trp or 70 µM indole or without the addition of Trp/indole at 30 °C, 200 rpm, and OD_{600} was measured after 24 h and 48 h. In addition, 100 µL of a suspension of TUB00 cells, corresponding to OD_{600} = 1, was plated on NMM19-agar in the presence of 70 µM indole or in the absence of any Trp source and incubated at 30 °C for 48 h. The growth was documented with the Scanner ViewPix700 (Biostep, Jahnsdorf/Germany).

Determination of the ALE starting conditions was performed with these three biological replicates of TUB00. They were resuscitated from cryo-stocks and cultured in 5 mL NMM19-0-10 overnight at 30 °C, 200 rpm. Cells were washed twice with 3 mL NMM0 each to remove residual indole, and then 10 mL culture each was inoculated into 100 mL flasks with the respective test media to OD_{600} 0.02. The test concentrations for 6Fi and 7Fi comprised 0, 5, 15, 30, 50, 70, 100, 500, 1000, 5000, and 1000 µM, respectively, in NMM19-[Fi]-0 and that for the indole substrate comprised 0, 0.5, 1.0, 2.5, 10 and 30 µM in NMM19-0-[Ind]. Growth was performed at 30 °C, 200 rpm and monitored by regular measurement of OD_{600}.

5.2.1.1 Colony-forming unit (CFU) assay

In order to estimate the growth-inhibiting or toxic effect of the fluoroindoles (see section **3.2.3**), colony-forming unit (CFU) assays were performed, which provided information about the number of viable cells. This contrasts with optical density (turbidity) measurements, where all cells, including nonviable entities, are considered. The number of CFUs was estimated by surface viable count (Miles and Misra method[512]). Different from the common surface viable count technique, the cells were not spread on agar plates but instead droplets of the different bacterial culture dilutions were spotted on a single plate. Agar plates contained NMM19-agar (mixture of 2x H_2O agar and 2x NMM19) supplemented with all 20 cAAs (50 mg/L). Cells "grown" with 0, 100, 500, 1000, 5000, and 10000 µM 6Fi or 7Fi, respectively, were serial diluted to 1^{-1}, 1^{-2}, 1^{-3}, 1^{-4}, and 1^{-5}, and 3 µL each were spotted. Plates were air-dried and incubated at 30 °C for several days.

5.2.2 Inoculation method of the ALE experiments

Adaptive laboratory evolution (ALE) is a culturing technique used in modern evolutionary biology and biotechnology to propagate microbial cultures in controlled microenvironments in order to provoke evolutionary adaptation. In this study, ALE was conducted by a serial transfer of batch cultures in a defined minimal medium (NMM). Analog usage was encouraged by limiting the indole supply and supplying the desired fluorinated indole (6Fi or 7Fi) in excess. The experiments towards 6Fi and 7Fi adaptation were conceived in three phases: 1) indole depletion, 2) cAA removal and, 3) growth in final medium composition. In ALE phase 1, NMM was supplemented with a limiting amount of indole (2.5 µM), an excess of the respective fluoroindole (70 µM), and an excess of 19 cAAs (50 mg/L), except for Trp. The initial addition of indole was intended to stimulate cell growth and facilitate the gradual adaptation process in this early phase (see section **3.2.3**). The indole concentration was then stepwise reduced until complete depletion in order to liberate the cells from its indole dependency. In ALE phase 2, the 19 cAAs were removed by either the OC or MB approach to exclude possible contaminations with traces of indole or Trp that could originate from common amino acid preparations (see section **3.3**). And in ALE phase 3, the cells were allowed to propagate in their final medium compositions, exclusively supplemented with 70 µM fluoroindole. In contrast, the positive controls adapted to grow on either Ind or Trp in the minimal medium were constantly kept in NMM19-0-70 (Ind or Trp).

Generally, cultivation was performed in triplicates with 10 mL NMM in 100 mL Erlenmeyer flasks covered with aluminum foil at 30 °C and 200 rpm shaking. For repeated inoculation, an $OD_{600} = 0.02$ was used or the entire parental culture volume in cases where its cell density was too low to reach the intended inoculum. The cells were allowed to grow until their early stationary phase was reached, usually for about 55 – 65 h, and then transferred into fresh medium (termed as passage). Whenever the cultures grew stable for at least 2 – 3 passages under a given condition, i.e., no further fitness increase was observed, the NMM was adjusted to increase the selection pressure according to the respective ALE phase. Samples of almost every passage were conserved as cryo-stock for subsequent analyses (see section **5.1.1**).

In total, populations were constantly subjected to the serial inoculation regime for about 400 days from 04/2019 to 10/2020, and this was only interrupted for compelling reasons, that was, in fact, seven times with a long pausing period of about two months during COVID-19 lockdown.

5.2.3 Generation times

Bacterial growth is the proliferation of a bacterium into two daughter cells in a process called binary fission.[513] Each subsequent cell division of the accrued daughter cells results in a doubling of the cell number and thus an exponential increase in the population (i.e., the number of cells is: 2^0, 2^1, 2^2, 2^4 to 2^n, where n is the number of divisions). Mathematically, this is expressed by equation 1 (Eq. 1). The time scale in ALE experiments is usually set in terms of the number of generations, as this includes growth characteristics (i.e., slow vs. fast-growing cells) as opposed to simple cultivation time. The number of generations (n) was calculated according to Eq. 2, where N(0) is the initial cell density at t = 0 and N(t) is cell density at a certain time point (t).

$$N(t) = N(0) \cdot 2^n \hspace{4cm} \text{Eq. 1}$$

$$n = log_2\left(\frac{N(t)}{N(0)}\right) \hspace{3.5cm} \text{Eq. 2}$$

5.2.4 Subpopulation screening

A routine examination of the growth behavior of the strains adapted to 6Fi in both indole and 6-fluoroindole revealed that growth appeared to be reduced in indole compared to 6Fi (see section **3.4.2.1**). Subpopulation screening was used to identify and select single clones that show a preference for 6Fi as a Trp source. These individuals were afterward subjected to another ALE experiment (see below). Therefore, population samples of clone 1 of passage 124 of the three 6Fi ALEs (6TUB124-OC, 6TUB124-MB4, and 6TUB124-MB3) were spread, respectively, on NMM19-70-0 (6Fi) agar plates for clone isolation. NMM agar is composed of a 1:1 mixture of 2x H_2O agar (3 %) and 2x NMM. Then 20 single clones of every lineage were picked, each diluted in 20 μL MQ H_2O, and 1 μL was spotted on two NMM19 agar plates containing either 6-fluoroindole or indole (19-0-70 (Ind) and 19-70-0 (6Fi)). Plates were incubated for 48 h at 30 °C.

Since growth conditions are, in general, more stringent in liquid culture than on plates, this setup was repeated in liquid media. Therefore, from the "19-70-0"-plate again, 20 promising clones out of the 60 were selected and used to inoculate 2 mL NMM0-70-0 (6Fi) precultures, incubated at 30 °C, 200 rpm for 48 h. 1 mL each of the resulting cultures were, washed twice in 800 μL NMM0-0-0 (centrifugation 4000 g, 3 min, RT) to remove residual 6Fi and used to inoculate media (2 mL to OD_{600} 0.02) again containing either 6Fi or indole (NMM0-0-70 (Ind) and NMM0-70-0 (6Fi)). The cultures were incubated at 30 °C, 200 rpm and growth was monitored by frequently measuring the OD_{600}.

Five promising clones (29, 34, 35 derived from 6TUBX-MB4 and 49, 60 derived from 6TUBX-MB3) that showed the highest growth differences (Ind vs. 6Fi supplemented NMM) were preserved as cryo-stocks

and used as ancestral strains for an additional ALE experiment. In this subpopulation ALE, the same conditions were applied as described in section **5.2.2**, which were propagation in 10 mL NMM0-[6Fi]-0 in 100 mL shaking flasks, incubation at 30 °C, 200 rpm, inoculation of fresh NMM with OD_{600} 0.02 whenever populations reached early stationary growth phase and cryo-storage of every population sample. However, in this case, the populations were exposed to increased selective pressure by gradually and stepwise increasing the 6Fi concentration from 70 µM to 85 µM up to 100 µM. The cultivation experiments were repeated with subpopulation cells of passages 141 (6TUB141-SP, 85 µM 6Fi) and 150 (6TUB150-SP, 100 µM 6Fi) as well as with the respective parent lineages 6TUB141-MB4, 6TUB150-MB4 and 6TUB141-MB3, 6TUB150-MB3 that were kept on NMM0-70-0 (6Fi).

5.2.5 Determination of the growth behavior

In order to assess the success of the adaptation experiments comprehensive growth experiments with the final isolates of all adapted strains were conducted. Cells of the positive controls Ind-TUB165 and W-TUB165 were revived from cryo-stocks in NMM19-0-70 (supplemented with either Ind or Trp); 6Fi-adapted cells could be directly resuscitated in NMM0-70-0 (6Fi) from cryo-stock, whereas 7Fi-adapted cells had to be revived in nutrient-rich LB medium, followed by two-times serial inoculation in their final adaptation medium NMM0-70-0 (in order to restore the intracellular Trp-free environment). Subsequently, the cells were washed twice with 1 mL NMM0-0-0 each to remove residual Ind, 6Fi, or 7Fi (2000 g, 3 min, RT), normed to $OD_{600} = 1$ (prevention of differences of inoculation volume), and then 198 µL of the respective medium composition were inoculated to $OD_{600} = 0.02$. The tested media comprised LB (rich, undefined medium), NMM19-0-70 (Ind), NMM19-70-1 (6Fi or 7Fi, Ind), NMM19-70-0 (6Fi or 7Fi) and NMM0-70-0 (6Fi or 7Fi). The growth was monitored in 96-well plates (Greiner, flat bottom clear), sealed with the breathable membrane (Breathe-Easy sealing membrane, Sigma Aldrich) at 30 °C with continuous shaking using a microplate reader (Tecan Infinite M200) that measured the absorbance at 600 nm in 10 min intervals. The measurements were performed at least twice and in triplicates (biological replicates).

The growth experiments were performed at the microscale level for two reasons: 1) compared to the manual recording of growth curves by cultivation in flasks, this procedure is advantageous because repeated disturbances caused by the measurement process are avoided, and 2) very detailed curves are obtained, due to the short measurement intervals. Although growth at this scale is not exactly the same as under ALE conditions, i.e., due to differences in culture volume (200 µL in 96-well plates vs. 10 mL in 100 mL flasks) as well as influencing factors such as agitation, oxygen supply, etc., this approach has proven to be an acceptable approximation. Fundamental growth curves were also measured using the flask approach, and growth parameters were found to be comparable (see **Figure S64** in appendix).

5.2.5.1 Chemical distant indole analogs

The indole analogs 6-hydroxyindole (6OH-Ind), 4-fluoroindole (4Fi), 5-fluoroindole (5Fi), 6-fluoroindole (6Fi), 7-fluoroindole (7Fi), 6-methylindole (6Me-Ind), 6-bromoindole (6Br-Ind), 6-(Trifluoromethyl)indole (6CF$_3$-Ind) and indole (Ind) were dissolved in 96 % EtOH. 6TUB002, 6TUB165-MB4 and 6TUB165-MB3 were revived in NMM19-70-10 (6Fi, Ind) and washed twice with 800 µL NMM0 each to remove residual Ind and 6Fi. Then, 10 mL NMM0-70-0 supplemented with the respective indole analog were inoculated to an OD$_{600}$ of 0.02. Cultivation of biological triplicates was performed at 30 °C, 200 rpm and growth was monitored frequently.

5.2.5.2 Polyfluorinated peptides

The polyfluorinated peptides MfeGlyK16, DfeGlyK16, TfeGlyK16, as well as its non-fluorinated precursor peptide Abu, were a courtesy of Suvrat Chowdhary.[378] They were desalted to remove cytotoxic TFA and provided in 2.5 mg/mL stock solutions.

Cells of the strains 6TUB165-MB4 and 7TUB165-MB were cultured in NMM0-70-0 (6Fi and 7Fi, respectively), and 90 µL/per well were seeded in a 96-well plate at a density of 5 x 10^6 cells/mL. TUB00 was used as a control, cultured in NMM0-0-70 (Ind), and 90 µL/per well were seeded to 5 x 10^4 cells/mL. Then, these cell suspensions were treated with a dilution series of Abu, MfeGlyK16, DfeGlyK16, and TfeGlyK16 (0.25, 0.125, 0.0625, 0.0313, 0.0156 mg/mL). 10 µl of each sample (solved in 50 mM Bis-tris propane buffer (BTP), 150 mM NaCl, pH 7.4) were added and incubated for 24 h at 30° C, 150 rpm. This also included a positive control (treatment with 1 % formaldehyde) and negative controls (non-treated cells and a mixture of respective peptides in BPT and NMM without the addition of cells). For background subtraction, the wells containing no cells but only samples were used. After the 24 h peptide treatment, the cell viability was determined using a CCK-8 Kit (Sigma-Aldrich) according to the manufacturer's instructions (see also section **5.2.6**).[514] 10 µL/per well CCK-8 solution were added, the reaction mixtures were incubated for another 3 h at 30 °C without shaking, and absorbance was measured at 450 nm using a Tecan plate reader. After background subtraction, values were normalized (set to 100 %) to the viability of respective non-treated cells (positive control), and formaldehyde was used as negative control. Cytotoxicity measurements were performed with three independent biological replicates and repeated three times. Errors are derived from the standard deviation.

The effect of polyfluorinated peptides was also estimated by agar well diffusion assay, discussed in section **5.3.2**.

5.2.6 Determination of cell viability

Measuring the optical density (OD_{600}) provides information about the number of cells in a given volume; however, live as well as dead cells contribute to this turbidity measurement; thus, no conclusions about the health of the population can be drawn. The number or percentage of live, healthy cells in a sample is determined using cell viability assays that use indicators of metabolically active cells. Many methods are based on the conversion of an indicator substrate into a colored or fluorescent substance through the metabolic/enzymatic activity of the living cell. Such methods include, for example, tetrazolium reduction, resazurin reduction, protease activity assays, or dehydrogenase activity assays, where absorption or fluorescence are measured in multi-well plates using a plate reader.[515] Other techniques like live/dead staining use fluorescent dyes and detection with microscopy or flow cytometry, or viable cells are simply counted as colonies.[516,517]

In this study, cell viability was assessed using colorimetric redox assays that are based on the use of tetrazolium salts that were reduced by metabolically active cells to formazans. Tetrazolium reduction is considered to be proportional to the number of cells and to their metabolic activity.[515] Among the variety of tetrazolium-based dyes available MTT (2-(4,5-dimethyl-2-thiazolyl)-3,5-diphenyl-2H-tetrazolium bromide) and WST-8 (2-(2-methoxy-4-nitrophenyl)-3-(4-nitrophenyl)-5-(2,4-disulfo-phenyl)-2H-tetrazolium) were tested.[515,518] MTT has a net positive charge that facilitates cellular uptake and is reduced by NAD(P)H-dependent oxidoreductases and dehydrogenases to insoluble formazan (purple, λ_{abs} = 590 nm). However, MTT was observed not to be sensitive enough, therefore, the CCK-8 assay (Cell Counting Kit-8) was used instead. This is based on WST-8 tetrazolium salt, which is not cell permeable due to negative charge, requires electron coupling reagents such as 1-methoxy-5-methylphenazinium methyl sulfate (1-methoxy PMS), and is reduced by a dehydrogenase to water-soluble formazan (orange-yellow, λ_{abs} = 450 nm).[375]
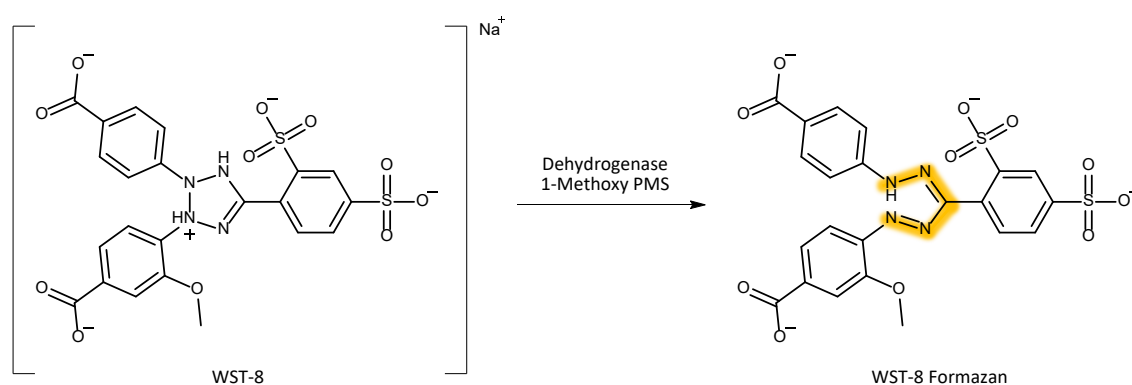

Figure 57 | Tetrazolium (WST-8) reduction to WST-8 formazan.

The viability assay was used as a fitness measurement (see section **3.5.1**) and to estimate the cytotoxic effects of polyfluorinated peptides on cell viability (see section **3.5.3**). For the proliferation assay, 190 µL of 5.0 x 10^7 cells/mL suspension were seeded, added with 10 µL of the CCK-8 solution, and incubated for 4 h at 30 °C without shaking. It is important to use exponential growing cells since, in batch liquid cultures, the cell number is high and still increases in the stationary phase while the metabolic activity is often already decreasing.[515] Then, the absorbance at 450 nm was measured using a microplate reader (Tecan reader M200), which is directly proportional to the number of proliferating cells. As background control medium without cells was used, and the viability of TUB00 was set at 100 % compared to the adapted strains. Measurements were performed in triplicates and repeated at least twice.

5.3 Antibiotic susceptibility testing

In the clinical context, antibiotic susceptibility testing (AST) is used for empirical therapy to specify an appropriate antibiotic effective against the given infection; however, based on bacterial resistance mechanisms such as inactivation of porin channels or modification of antibiotic targets (e.g., cell division, cell wall integrity) AST is also used in research to draw conclusions on the respective cellular component. Conventional antibiotic susceptibility testing methods encompass phenotypic (e.g., agar-based disk diffusion and Etest gradient disk diffusion (epsilometer testing) as well as broth macro and microdilution) and genotypic (e.g., identification of resistance genes by PCR, DNA microarray and DNA chips) techniques; but there are also other emerging approaches such as fluorescent, colorimetric, and electrochemical microfluidic sensors.[519,520] All methods use the minimal inhibitory concentration (MIC) as a determining parameter, which is defined as the lowest concentration of an antimicrobial agent that prevents visible bacterial growth under defined conditions. For determination of the minimal inhibitory concentration, the most commonly used methods are broth dilution and agar dilution, which both involve subjecting the cells to be tested with different concentrations of the putative antimicrobial substance either in a liquid medium or on a nutrient agar medium.[521]

5.3.1 Measurement of minimal inhibitory concentration by broth dilution

The broth dilution method was used to determine the MIC of vancomycin (see section **3.9.1**). Isolates from cryo-stocks of the respective strains were resuscitated either in 2 – 3 mL LB or appropriate NMM, incubated at 30 °C, 200 rpm, and re-inoculated into NMM if necessary. Subsequently, cells were used to inoculate 3 mL NMM in culture tubes (NMM 0-0-70 (Ind, Trp) for TUB00, W-TUB165, Ind-TUB165 and NMM 0-70-0 (6Fi, 7Fi) for adapted strains). To determine the MIC, a cell density corresponding to 0.5 McFarland standard was used[522] and cells were treated with serial dilutions of vancomycin (0, 25,

50, 100, 200, 400 µg/mL; dissolved in MQ-H$_2$O). Cell growth (incubation at 30 °C, 200 rpm) was monitored by measuring the OD$_{600}$ every 24 hours, and MIC was defined as the lowest concentration at which no increase in cell density could be observed. Measurements were performed using biological triplicates.

5.3.2 Agar well diffusion

The susceptibility of TUB00, 6TUB165-M4, and 7TUB165-MB towards polyfluorinated peptides was determined using an agar well diffusion assay (see section **3.5.3**).[521] 50 mL lukewarm LB agar (1.5 %) were inoculated with 1 mL of the respective cell suspension (OD$_{600}$ = 1) and poured into a sterile petri dish (Ø 14 cm) to solidify. Then, holes were punched into the agar using a sterile glass pipette and filled with 20 µL of a solution containing the antimicrobial agent to be tested. The peptide samples AbuK16, MfeGlyK16, DfeGlyK16, and TfeGlyK16 were added at a concentration of 2.5 mg/mL; the antibiotics ampicillin (Amp) and chloramphenicol (Cm) were used as positive controls and added in the same concentration (2.5 mg/L) as well as ampicillin with a preventively increased concentration of 100 mg/L; and 50 mM BTP peptide buffer was used as negative control. The petri dish was then left for at least 1 h at room temperature to allow the test solution to diffuse into the agar before it was incubated for 48 h at 30 °C. Antimicrobial activity is indicated by the inhibition zone (also called halo), where no cell growth is observed.

5.4   Molecular biological methods

5.4.1   Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) is a method for *in vitro* amplification of specific target DNA regions. It is a fundamental technique in genetic research and was developed by Kary Mullis in 1983.[523,524] The PCR technique relies on thermal cycling, undergoing a series of repeating discrete temperature changes.[525] Each cycle constitutes of three thermal steps: (i) denaturation at 95 – 98 °C causes denaturation of double-stranded DNA (dsDNA) by breaking the hydrogen bonds between complementary bases; (ii) annealing (temperatures 3 - 5 °C below Tm of primers) permits hybridization of the primer to complementary parts of each generated single-stranded DNA (ssDNA) template and the DNA polymerase binds to the primer-template hybrid; (iii) during elongation at 72 °C the DNA polymerase synthesizes a complimentary copy to the DNA template in $5' \rightarrow 3'$ direction by adding free deoxynucleotide triphosphates (dNTPs) through condensation of the 5'-phosphate group of the dNTPs with the 3'-hydroxy group at the end of the nascent DNA strand. The cycling is preceded by an initial denaturation step and followed by an additional elongation step at the end for the final product extension. Since with every successive cycle, the original template strand and all newly synthesized strands become template strands, PCR leads to exponential amplification of specific DNA regions (with number of dsDNA copies is given by 2^n, where n is the number of cycles).

PCR reactions were always prepared on ice, and the thermocycler was preheated to avoid unspecific amplification and prevent undesired side products and primer dimers. Annealing temperatures depend on the size and sequence of the primers and were calculated using the NEB Tm Calculator v1.15.0 from New England Biolabs (https://tmcalculator.neb.com/#!/main).

5.4.1.1 Standard-PCR

Standard-PCR was performed using high-fidelity DNA polymerases, either Q5 (New England Biolabs) or Phusion (Thermo Fisher Scientific), providing robust DNA amplification, ideal for cloning procedures and long or difficult amplicons. These polymerases possess increased accuracy due to a $3' \rightarrow 5'$ proofreading exonuclease activity but suffer from possible degradation of single-stranded primers. For GC-rich targets (≥ 65 % GC), amplification was improved by the addition of either 5x Q5 High GC Enhancer or DMSO to prevent secondary structure formation. Thereby, DMSO decreases the melting temperature (Tm) and thus destabilizes weaker hydrogen bonding in unspecific primer hybridization. And for amplification of long products (> 6 kb), the extension time was increased to 40 - 50 sec/kb (e.g., amplification of 7324 bp full-length trp operon *trpLEDCBA*).

Table 4 | Typical PCR reaction and program for standard-PCR using Phusion High-Fidelity DNA Polymerase.

PCR-reaction mixture		PCR-program		
5x Phusion™ HF Buffer	4 µL	Initial Denaturation	98 °C	30 sec
Primer fwd (20 µM)	0.5 µL		**30 Cycles**	
Primer rev (20 µM)	0.5 µL	Denaturation	98 °C	10 sec
DMSO (optional)	0.6 µL	Annealing	45-72 °C	20 sec
dNTPs (10 mM)	0.4 µL	Elongation	72 °C	30 sec/kb
Template DNA	1 µL			
Phusion™ HF DNA Polymerase	0.2 µL	Final Extension	72 °C	10 min
nuclease-free H$_2$O	ad to 20 µL	store	4 °C	forever, ∞

Table 5 | Typical PCR reaction and program for standard-PCR using Q5 High-Fidelity DNA Polymerase.

PCR-reaction mixture		PCR-program		
5x Q5 Reaction Buffer®	5 µL	Initial Denaturation	98 °C	1 min
Primer fwd (20 µM)	0.625 µL		**30 Cycles**	
Primer rev (20 µM)	0.625 µL	Denaturation	98 °C	10 sec
GC-Enhancer® (optional)	5 µL	Annealing	50-72 °C	20 sec
dNTPs (10 mM)	0.5 µL	Elongation	72 °C	30 sec/kb
Template DNA	1 µL			
Q5 High-Fidelity DNA Polymerase	0.25 µL	Final Extension	72 °C	2 min
nuclease-free H$_2$O	ad to 25 µL	store	8 °C	forever, ∞

When primers with overhangs (Tm ≥ 72 °C) were used, a two-step thermocycling protocol with different annealing temperatures was implemented (**Table 6**). Initially, only the binding part of the primer anneals to the template at lower temperatures, and a product that has the overhangs incorporated is generated (5 cycles). Thus, a new template with overhangs is available for the subsequent cycles so that the entire primer can be annealed at higher Tm, and the complete transcript is amplified with the desired modifications, such as restriction sites or mutations (25 cycles). Therefore, to accommodate both annealing temperatures (for only the binding part and the whole primer), a program with two different cycles was written (**Table 6**). The two-step PCR was used to introduce restriction sites (BsaI) in the context of golden gate cloning for TrpRS cloning (see section **5.6.1**) and CRISPR/Cas9 genome editing (see section **5.5.1**).

Table 6 | PCR-program for using primers with overhangs.

Two-step PCR-program		
Initial Denaturation	98 °C	1 min
	5 Cycles	
Denaturation	98 °C	10 sec
Annealing	50-72 °C	20 sec
Elongation	72 °C	30 sec/kb
	25 Cycles	
Denaturation	98 °C	10 sec
Annealing	50-72 °C	20 sec
Elongation	72 °C	30 sec/kb
Final Extension	72 °C	2 min
store	8 °C	forever, ∞

5.4.1.2 Colony-PCR

Colony-PCR was applied in the context of verification of cloning steps or for the frequent validation of the strain identity of TUB00 (i.e., Trp-auxotrophic genotype) during the ALE experiments. A colony of interest was picked from an agar plate and resuspended in 30 µL sterile water, or a sample of liquid culture was used. The cells were lysed by boiling the suspension to 95 °C for 10 min, releasing the DNA from the cells, and this suspension was used as a template for the PCR. Owing to the analytical purpose, the usage of (in-house produced) Taq Polymerase was adequate.

Table 7 | Typical PCR reaction and program for colony-PCR using Taq DNA Polymerase.

PCR-reaction mixture		PCR-program		
10x Taq-Green Buffer®	2 µL	Initial Denaturation	95 °C	5 min
Primer fwd (20 µM)	1 µL		**30 Cycles**	
Primer rev (20 µM)	1 µL			
MgCl$_2$ (50 µM)	0.6 µL	Denaturation	95 °C	30 sec
dNTPs (10 mM)	0.5 µL	Annealing	52-72 °C	30 sec
lysed DNA suspension	1 µL	Elongation	72 °C	60 sec/kb
Taq DNA Polymerase	0.6 µL	Final Extension	72 °C	4 min
nuclease-free H$_2$O	ad to 20 µL	store	8 °C	forever, ∞

5.4.2 Agarose gel electrophoresis

Agarose gel electrophoresis is a method for separation, identification, and purification of nucleic acids (0.5 kb – 20 kb) and was used for analytical as well as preparative purposes in this study. Agarose is a linear polymer (composed of D-galactose and 3,6-anhydro-L-galactopyranose) that gels to a three-dimensional matrix formed of helical agarose molecules in supercoiled bundles by glycosidic bonds. By applying an electric field, DNA fragments migrate through the agarose matrix owing to the negatively charged phosphate backbone depending on their molecular size and the agarose concentration. The rate at which nucleic acids migrate and the separation capacity are affected by electroendosmosis (EEO).[526,527]

Agarose gels were prepared with 1 % agarose and 1xTAE buffer and supplemented with either ethidium bromide (0.5 µg/mL; EtBr) or GelRed;[528,529] dyes that intercalate into the DNA, fluoresce (λ_{em} = 590 nm) under UV radiation (λ_{ex} = 302 nm), and allow visualization. DNA samples (PCR products, restriction digests) were supplemented with 1x DNA loading dye. The loading dye causes DNA to sink to the bottom of the gel pocket (by added glycerol) and contains two electrophoretic color markers that allow to monitor the progress of the gel (bromophenol blue comigrates at 300 - 500 bp and xylene cyanol FF comigrates at 4000 - 5000 bp, depending on the specific composition of the loading dye). For interpretation of the gels, GeneRuler™ 1 kb DNA Ladder (Thermo Fisher Scientific) was used as standard. The electrophoresis was performed by applying 80 - 100 mV for 45 - 90 min, and illumination with UV and documentation by photography was carried out in a gel documentation system (Felix 2050).

In another approach, DNA was stained using Midori green dye (4 – 6 µL/100 mL),[530] and analysis was performed either with ChemiDoc™ MP Imaging System (BioRad) or UVsolo TS Imaging System (Biometra). The spectral properties of Midori green are λ_{ex} = 490 nm (with two secondary peaks at 270 nm and 290 nm) and λ_{em} = 530 nm; thus, filters for GFP/SYBR Green but also for EtBr are suitable, as well as UV light.

5.4.3 DNA purification and gel extraction

After PCR or restriction digests, the DNA was purified using the GeneJET™ PCR Purification Kit according to the manufacturer's instructions. A chaotropic reagent denaturates proteins and promotes DNA binding to a silica membrane, where primers, dNTPs, unincorporated labeled nucleotides, enzymes, and salts from PCR and other reaction mixtures are removed.[531] DNA purification from agarose gels was carried out by using the GeneJET™ Gel Extraction Kit according to the manufacturer's specifications. The desired DNA fragment was excised from the gel, and the treatment with a chaotropic reagent dissolves agarose, denaturates proteins, and promotes DNA binding to the silica membrane, where impurities are removed.[532]

5.4.4   Determination of DNA concentration

The concentration and purity of DNA was determined using spectrophotometric quantification. The aromatic nucleobases of nucleic acids (DNA and RNA) absorb light at $\lambda = 260$ nm. The Lambert-Beer law (Eq. 3) is then used to calculate the DNA concentration. It describes the attenuation of the intensity of the radiation (I_0/I: incident and transmitted intensity) as it passes through a medium with an absorbent substance, depending on the concentration (c [ng/µL]) of the absorbent substance and the path length (d [cm]); the average extinction coefficient ε_λ is given with 0.020 (µg/mL)$^{-1}$ cm^{-1} for dsDNA.

The sample purity is assessed by the ratio of absorbance 260/280 as well as 260/230, where λ_{max} DNA = 260 nm, λ_{max} proteins = 280 nm, and λ_{max} salts = 230 nm is. DNA solution is deemed "pure" at 260/280 values of 1.8 – 2.0 and 260/230 values of 2.0 – 2.2. Lower ratios indicate the presence of contaminations, such as proteins, phenol, or salts, which could originate from buffers of the DNA extraction procedure.

$$\text{Lambert-Beer law} \qquad A_{260} = log\left(\frac{I_0}{I}\right) = \varepsilon_\lambda \cdot c \cdot d \qquad\qquad \text{Eq. 3}$$

5.4.5   Sequencing

DNA sequencing is a method for determining the exact nucleotide sequence of nucleic acids. In the dideoxy-method according to Sanger,[533] also known as chain termination synthesis, the DNA strand is primer-mediated extended by a polymerase analogous to PCR. However, the difference is that in addition to the regular dNTPs, labeled deoxyribonucleotides (ddNTPs), which lack both the 2' and 3' OH groups, are added to the reaction mixture. When these ddNTPs are incorporated into the newly synthesized strand, the extension by the DNA polymerase is no longer possible, as the OH group at the 3'-position is missing for linkage with the phosphate group of the next nucleotide. Because of the discontinued reaction, a mixture of oligonucleotides of different lengths is produced. The identity of the last nucleotide can be decoded by marking each type of ddNTP with a different fluorescent dye. The resulting color pattern shows the nucleotide sequence after the fragments have been separated.

For DNA sequencing, 700 ng DNA and 30 µM primer were used in a 12 µL reaction mixture. Sequencing was handled by Microsynth Seqlab GmbH (Göttingen) in this study.

5.5 Cloning procedures and Genome Engineering

The introduction of genetic variation, either mutations as well as by insertion or deletion of entire genes, is a valuable tool in molecular biology. Molecular cloning is used to assemble recombinant DNA molecules (e.g., insertion of a target gene in a vector) and to direct their replication within host organisms, for example, to perform recombinant protein expression. However, genetic engineering is also possible at the chromosomal level. Common methods include the Datsenko-Wanner approach[365] and P1 phage-mediated unspecific transduction,[534,535] the latter of which involves transferring a knockout from a donor strain (e.g., from the Keio collection,[536] Coli Genetic Stock Center (CGSC)) to a recipient strain. Both need the action of a helper plasmid (e.g., pCP20)[537] encoding the Flp recombinase to excise resistance genes used during the editing process. However, the excision of a resistance gene sequence by the Flp/FRT system requires additional experimental steps, leaves scars, and involves the risk of unintended recombination between these FRT sites at different locations.[538] Modern methods also enable seamless genome editing, for instance, the λ-Red based DIREX (direct and inverted repeat stimulated excision) technique[539] or REXER (Replicon Excision Enhanced Recombination),[540,541] which makes use of CRISPR/Cas9 (described below) and that was even used for retrosynthesis of a synthetic genome with compressed genetic code.[227]

5.5.1 Golden Gate Assembly

Golden Gate Assembly is a cloning procedure that combines two advantages by allowing seamless cloning in a one-pot reaction of simultaneous restriction and ligation.[542,543] The underlying key feature is the usage of type IIs restriction endonucleases, which cut dsDNA outside of their recognition sequences. Owing to the fact that the original cleavage site of the type IIS restriction enzyme is no longer contained in the ligated product, digestion, and ligation can occur simultaneously in one tube. Thus, Golden Gate Assembly allows efficient, fast, directional, and seamless multiplex DNA cloning of up to 50 fragments of different sizes (< 100 bp to > 15 kb) and GC content in a single reaction without unwanted additional bases at the junctions.

Experimentally, the ligation fragments were either mixed in a molar ratio of 1:1:1 (assembly of the CAGO editing cassette, see section **5.5.3**) or in a molar ratio of 1:3 vector to insert (TrpRS cloning, see section **5.6.1**) according to their concentration. For restriction digest type IIs BsaI-HF®v2 (New England Biolabs) employing recognition sequence $\frac{5'\dots GGTCTC(N)_1 \downarrow \dots 3'}{3'\dots CCAGAG(N)_5 \uparrow \dots 5'}$ was used, and for ligation, the T4 ligase (Thermo Scientific) was used; reaction conditions are shown in **Table 8**. The NEBridge Golden Gate assembly tool v2.8.3 from New England Biolabs (https://goldengate.neb.com/#!/) was used for primer design.

Table 8 | Golden Gate Assembly reaction and cycler program.

Assembly reaction		Cycler program		
			25 Cycles	
Fragment 1 DNA/vector	(ng/µL)			
Fragment 2 DNA/vector	(ng/µL)	Restriction	37 °C	3 min
		Ligation	16 °C	4 min
T4 Ligase Buffer	2.5 µL			
BsaI	0.75 µL	Inactivation	80 °C	10 min
T4 Ligase	1.25 µL		10 °C	hold
nucfree H_2O	up to 25 µL			

5.5.2 CRISPR/Cas9

The CRISPR/Cas9 system (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated protein) is an ancient immune system in bacteria to protect against foreign DNA (e.g., from viruses).[544,545] After survival of an infection, Cas proteins (CRISPR-associated proteins that possess nuclease and helicase activity) cut out a part of the foreign invader DNA (called protospacer) at a characteristic short recognition sequence, called PAM (protospacer adjacent motif). The acquired piece is integrated and stored in a CRISPR array of the bacterial chromosome, which serves as a library. There are repeat regions between each spacer in the CRISPR array. The entire CRISPR array is then transcribed into pre-crRNA (CRISPR RNA), tracrRNA (trans-activating crRNA) binds to the palindromic repeats of pre-crRNA, and then recruits Cas9. The CRISPR array is processed by RNase III, resulting in functional CRISPR/Cas9 units (mature crRNAs) containing a single spacer (and tracrRNA). Upon a recurring infection with the same pathogen, its DNA is targeted by the CRISPR/Cas9 effector complex carrying the homologous RNA fragment (spacer), and the foreign DNA is degraded. For this purpose, the viral DNA is unwound, crRNA binds the ssDNA, and Cas9 cuts at the virus-specific PAM recognition sequence; this also prevents self-cleavage. The degradation of its DNA thus neutralizes the virus.

The described CRISPR/Cas9 system of *Streptococcus pyogenes* can be exploited in the laboratory for genome engineering. To this end, scientists have developed a single guide RNA (gRNA or sgRNA) that combines the spacer and the attached tracrRNA in one molecule.[546] For their development, Emmanuelle Charpentier and Jennifer Doudna were awarded the Nobel Prize in Chemistry 2020 "for the development of a method for genome editing". There are numerous CRISPR/Cas systems for a variety of applications and several Cas proteins for targeting a wide range of DNA sequences.[547,548] However, most systems require the design of a new guide RNA (gRNA, a tracrRNA/CRISPR RNA chimera) for each target, which can be tedious despite the availability of several online tools, and Cas9 off-target effects limit the range of target sequences.[549,550] In this study, the CAGO technique

(CRISPR/Cas9-assisted gRNA-free one-step (CAGO) genome editing)[457] was used in which a universal gRNA is encoded on the pCAGO plasmid, and thus no design of specific gRNAs for each target is required. The gRNA targets a universal CRISPR/Cas9 recognition sequence (N20PAM) with minimal sequence homology to the *E. coli* genome to minimize the off-target effects of Cas9. The N20PAM can be incorporated into any editing cassette and is driven by its homologous gRNA provided on the pCAGO plasmid. Also encoded on the pCAGO plasmid is the Cas9 protein for double-strand-break (DSB) and the λ-Red system for homologous recombination. The latter consists of the *gam*, *beta*, and *exo* genes of bacteriophage λ. Gam inhibits the host RecBCD exonuclease V that would otherwise degrade the linear editing cassette, Exo is a dsDNA-dependent exonuclease that generates 3′ ssDNA overhangs, and beta protects newly formed ssDNA and catalyzes homologous recombination with the complementary ssDNA in the chromosome along with the DNA-dependent ATPase RecA.[551]

5.5.3   Chromosomal deletion of *sspA* gene using CAGO

The editing cassette for *sspA* chromosomal deletion comprised three fragments: 1) L_homo (left homology arm flanking 5′ *sspA*), 2) CmR-N20PAM (chloramphenicol resistance cassette and N20PAM recognition sequence) and 3) Ls_R_homo (right homology arm flanking 3′ *sspA* and a 40-nt sequence of L_homo at 3′ end); see **Figure 58**. The fragments of the editing cassette were amplified using Q5 DNA polymerase in a two-step PCR (see section **5.4.1.1**, **Table 6**) and analyzed by agarose gel electrophoresis and sequencing (see section **5.4.2** and section **5.4.5**). Then, the editing cassette was assembled in a one-pot Golden Gate Assembly (see section **5.5.1**), amplified in a preparative scale, and purified by gel extraction (see section **5.4.3**). For the primer design and *in silico* assembly of the editing cassette, Geneious Prime 2023.2.1 and the NEBridge Golden Gate assembly tool v2.8.3 from New England Biolabs were used. The entire genome editing process is illustrated in **Figure 58**.

Transformation of the editing cassette into recipient strain TUB00 pCAGO and induction of λ-Red mediated homologous recombination permits replacement and, thus, deletion of the *sspA* target gene.[551] Therefore, 5 mL LB Amp were inoculated 1:100 with an overnight culture of TUB00 carrying the pCAGO plasmid (**Figure S78**). The culture was incubated at 30 °C, 200 rpm until OD_{600} of around 0.2 was reached, and then the λ-Red was induced by the addition of 1 mM IPTG. The induction prior to transformation ensures sufficient production of λ-Red recombination genes (*exo*, *beta*, *gam*). After further cultivation to OD_{600} of around 0.6, the cells were rendered electrocompetent (see section **5.1.2.2**) and resuspended in 100 µL ice-cold 10 % glycerol. Approximately 600 ng of the editing cassette were transformed via electroporation (see section **5.1.3.2**), and the cells were allowed to recover for 2 hours at 30 °C, 750 rpm. Following the regeneration phase, the concentrated cell suspension (centrifugation: 2000 g, 2 min, RT) was selected overnight at 30 °C (ts replicon) on an LB-Agar plate containing Amp100

(selection marker of the pCAGO plasmid), Cm37 (selection marker of the editing cassette) and 1 % glucose (to suppress the leaky expression of Cas9 from the arabinose promotor). Homologous recombination of the editing cassette was verified by colony-PCR (see section **5.4.1.2**) and sequencing.

Genome editing is completed by Cas9-mediated double-strand break by which the chloramphenicol selection marker is excised and homology sequences that previously have flanked the *sspA* gene are recombined by the λ-Red system. Positive clones were incubated for at least 6 h at 30 °C in 2 mL LB Amp containing 1 mM IPTG for expression of the λ-Red system and 10 mM arabinose for Cas9 expression, followed by selection on LB agar (Amp100) at 30 °C. The removal of CmR and correct *sspA* deletion was verified by colony-PCR and sequencing. Finally, the pCAGO plasmid was cured by incubation at 42 °C, at which the temperature-sensitive replicon (ts) did not function[552] until no more growth on ampicillin could be observed. The resulting *sspA* deficient mutant was designated TUB00 Δ*sspA*, and biological replicates were preserved as cryo-stocks. The genome editing procedure was performed by Jana Madeleine Göller in the course of her master's thesis. Sequencing data and agarose gel analyses of the editing process are shown in **Figure 59**.
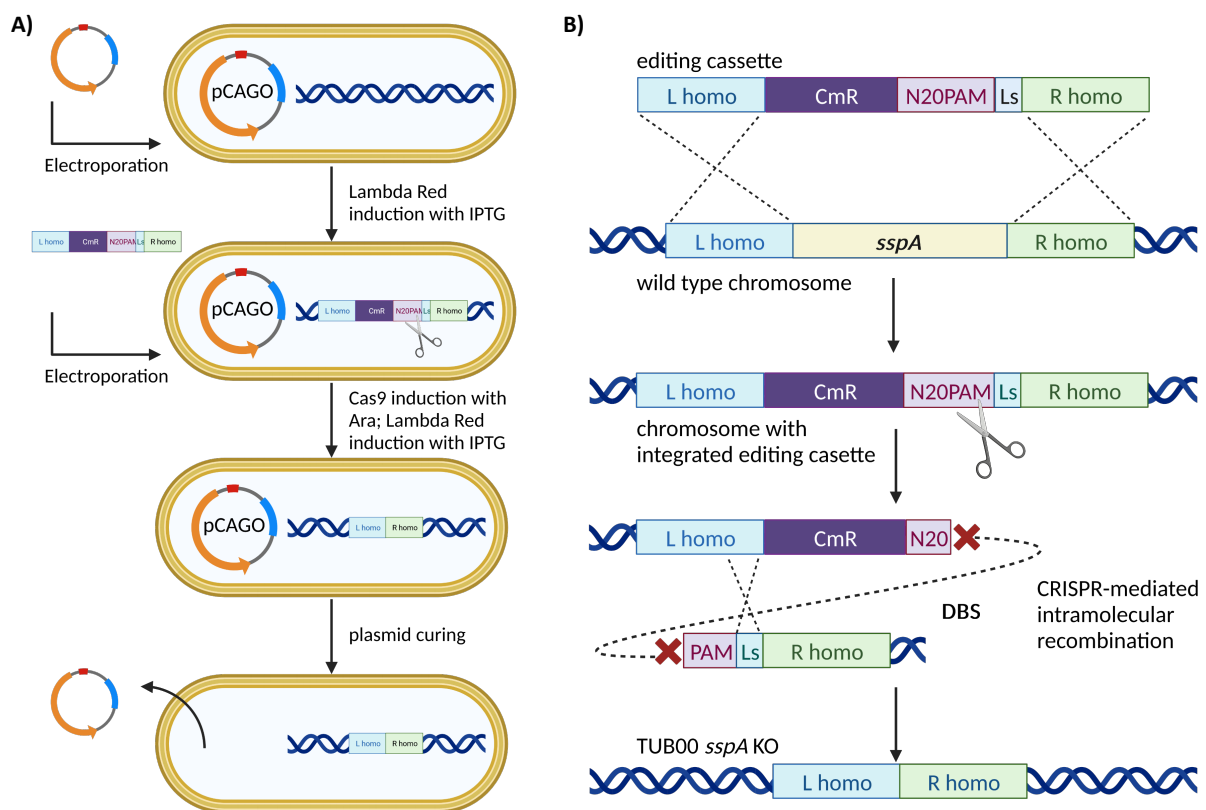

Figure 58 | Genome editing process using CAGO. A) Experimentally steps and B) events on the genomic level of the CAGO (CRISPR/Cas9-assisted gRNA-free one-step genome editing) process are illustrated. The figure is based on Zhao *et. al.*[457] and was created using BioRender (https://www.biorender.com/).
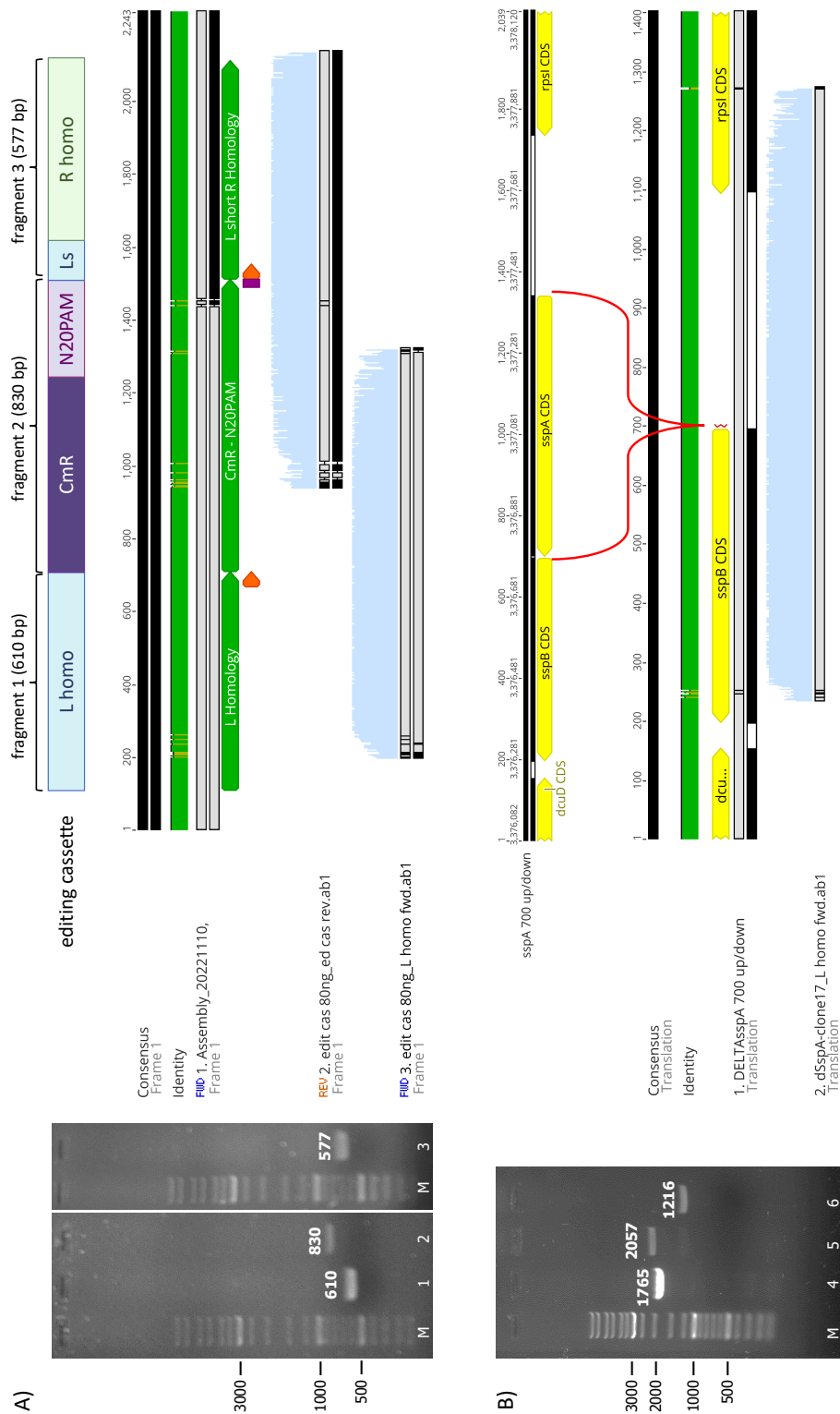
Figure 59 | Verification of the genome editing process for *sspA* deletion. A) The construction of the editing cassette is shown. The editing cassette comprises three fragments: fragment 1 (L homo) has 610 bp, fragment 2 (CmR N20PAM) has 830 bp, and fragment 3 (Ls R homo) has 577 bp (left). The arrangement of the editing cassette and a sequencing analysis of the assembled editing cassette (2057 bp) are shown right. The editing cassette was sequenced forward and reverse due to its length (> 1000 bp). B) The different stages of the *sspA* deletion are shown (left). The signal in lane 4 corresponds to an intact *sspA* gene, lane 5 to a construct where the *sspA* gene is substituted by the editing cassette, and lane 6 to a construct where *sspA* is deleted. An in-silico representation of the *sspA* region before and after the knockout is shown right, as well as sequencing data of the resulting deletion mutant TUB00 Δ*sspA*. The sequencing data are derived by Geneious Prime 2023.2.1.

5.6 Protein expression and purification

Immobilized metal ion affinity chromatography (IMAC) allows the simple purification of histidine-tagged recombinant proteins. The concept of IMAC can be traced back to 1975[553] and is based on the affinity of transition metal ions such as Zn^{2+}, Cu^{2+}, Ni^{2+}, and Co^{2+} to histidine and cysteine.[554] Those metal ions are fixed on chelating ligand-containing resins such as iminodiacetic acid (IDA) or nitrilotriacetic acid (NTA)[555] on agarose beads. The principal mechanism of IMAC-mediated purification relies on the interaction of His-tagged proteins to immobilized metal ions, whereby the NTA ligand coordinates Ni^{2+} with four valences, while two valences are available for interaction with imidazole moiety of the histidine residues.[556] The most commonly used oligo-histidine tag (His-tag) is an extension that consists of six consecutive histidine residues (H6) located either at the N-terminus or C-terminus of the recombinant polypeptide/protein. This His-tag provides six metal-binding sites for the association with the Ni-NTA ligand, leading to a stable binding.[557] During affinity chromatography, the protein mixture to be purified is loaded on a column, and owing to the affinity to the Ni-ligands, the recombinant proteins are immobilized on the resin. Using an imidazole gradient, proteins that are not specifically bound are removed, and the target protein is eluted at last by applying a very high imidazole concentration as it acts competitively against the target protein (more precisely, the imidazole motif of the histidine-tag) and supplants it. Thus, the eluate contains the purified and enriched protein.

5.6.1 Cloning, expression and purification of TrpRS

The TrpRS encoding gene *trpS* was targeted by mutation several times (Q27P and Q27P, L25L) in all 6Fi-adapted strains (**Figure 60**). Since mutation conservation, the pivotal role of *trpS* in processing Trp, and because *trpS* was already targeted in previous Trp-based ALEs, an investigation of the enzyme kinetics will be justified.

Cloning procedure: For this purpose, the TrpRS variants Q27P and Q27P L25L were amplified from one of the affected strains and cloned into the pQE80L expression vector. The vector used was specifically the pQE80L H6 TEV (TTL) plasmid, which already had a His-tag (H6) and a TEV cleavage site (*Tobacco Etch Virus* nuclear-inclusion-a endopeptidase) to remove it after purification; the already encoded TLL gene was cut out and replaced by the corresponding *trpS*. Both the vector and the *trpS* inserts were amplified using Q5-PCR and a two-step program in order to generate overhangs for restriction sites. For the cloning procedure, the Golden Gate Assembly was used as described in section **5.5.1** using a vector : insert ratio of 1 : 3 and the type IIs restriction enzyme BspMI (5'… ACCTGCnnnn^nnnn …3') from NEB. A respective plasmid map (**Figure S77**) is deposited in the appendix.

Then pQE80L H6-TEV-TrpRS Q27P, pQE80L H6-TEV-TrpRS Q27P L25L and pQE80L H6-TEV-TrpRS wt (as positive control) were transformed into electrocompetent *E. coli* NEB10-beta by electroporation (see sections **5.1.2.2** and **5.1.3.2**), for duplication and prepped (see section **5.1.4**). The expression was performed in *E. coli* BL21(DE3) after the respective transformation procedure.
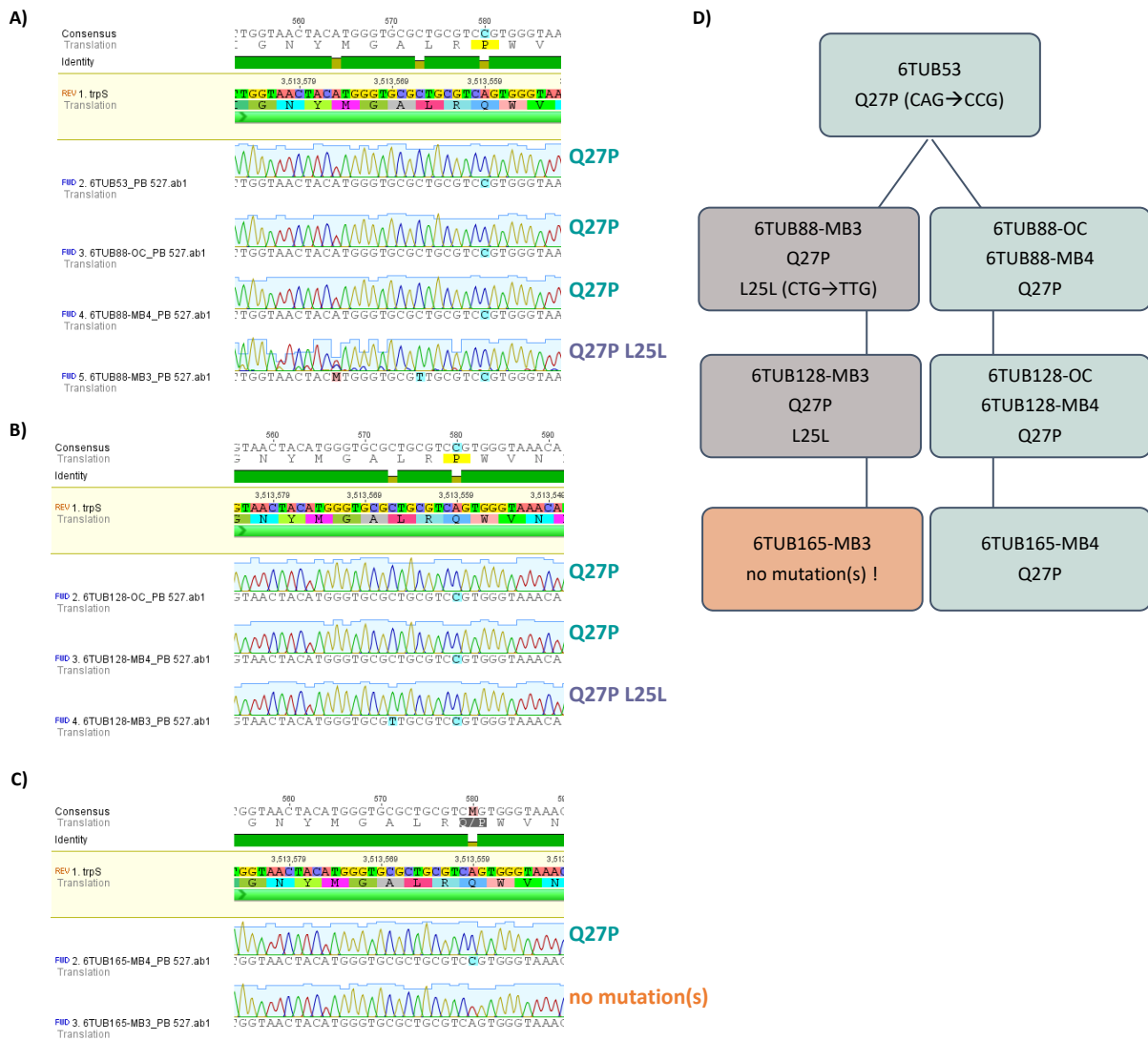

Figure 60| Mutations occurred in the TrpRS encoding *trpS* gene. The *trpS* gene was mutated in A) 6TUB53, 6TUB88-OC, 6TUB88-MB4, 6TUB88-MB3; B) in 6TUB128-OC, 6TUB128-MB4, 6TUB128-MB3; C) in 6TUB165-MB4 but not in 6TUB165-MB3. D) Overview of the specific SNPs in the respective ALE strains. The mutations shown were detected by standard PCR and sequencing (figures are derived by Geneious Prime 2023.2.1); and were additionally found in the WGS analysis.

Expression: For expression of the three H6-TEV-TrpRS variants 1 L LB Amp (100 µg/mL ampicillin) each were inoculated 1:100 with the respective preculture, that were 1) BL21(DE3) pQE80L H6-TEV-TrpRS wt, 2) BL21(DE3) pQE80L H6-TEV-TrpRS Q27P and 3) BL21(DE3) pQ80L H6-TEV-TrpRS Q27P L252L. Three main cultures were incubated at 37 °C, 250 rpm until they reached OD_{600} = 0.6 – 0.7, followed by induction with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and expression for 4 h at 30 °C, 250 rpm. Appropriate samples (1 mL) of the cells were collected before induction and after expression for subsequent SDS-PAGE and stored at -20 °C. After expression, cells were harvested by centrifugation (10000 g, 4 °C, 20 min), and pellets were stored at -80 °C until purification.

Purification: The H6-TEV-TrpRS variants were purified by IMAC using the Äkta pure FPLC system (GE Healthcare Life Sciences). The cell pellets were thawed on ice, if necessary, resuspended in 10 mL/g lysis buffer (50 mM Na-P, 300 mM NaCl, pH 7.8) supplemented with 1.5 mg lysozyme, 135 µg RNase, 135 µg DNase, and 3 mM $MgCl_2$) and incubated on ice for 1 h. This was followed by mechanical cell disruption with the cell disrupter (Julabo IXTUA coupled with a cooling system) using 20 psi pressure. After removal of the cell debris by centrifugation (18000 g, 4 °C, 45 min), the lysate was filtered (Rotilabo PVDF Filter, 0.45 µm pore diameter) and target proteins were purified by using affinity chromatography on 5 mL HisTrap™ FF column (Ni-NTA column from GE Healthcare Life Sciences). Purification was performed as follows: for sample application, the Ni-NTA column was equilibrated with 2 column volumes (CV) buffer A (binding buffer: 50 mM Na-P, 300 mM NaCl, pH 7.8) with 5 mL/min; and sample loading was performed with 4 % buffer B (elution buffer: 50 mM Na-P, 300 mM NaCl, 500 mM imidazole, pH 7.8) with 2.5 - 4 mL/min. For the purification procedure, the column was washed with 20 CV 5 % B at 3 mL/min to remove unbound proteins, and the target proteins were eluted by increasing the imidazole concentration using a linear gradient from 5 % B to 100 % B within 12 CV at 3 mL/min followed by 2 CV 100 % B (**Figure S73** shows a representative chromatogram). Relevant fractions were pooled and dialyzed three times against dialysis buffer (50 mM Na-P, 100 mM NaCl, 10 % glycerol, pH 7.8) to remove imidazole and salt excess. At each step of the purification procedure, an 80 µL sample was taken for SDS-PAGE analysis.

The identity of the H6-TEV-TrpRS was confirmed by LC-ESI-Q-TOF mass spectrometry (Agilent 6530 Accurate-Mass Q-TOF). Protein samples were measured in a concentration of 0.1 mg/mL using the following HPLC parameters: linear gradient from 5 % to 80 % buffer A within 20 min (A: 0.1 % formic acid in MQ-H_2O, B: (0.1 % formic acid in acetonitrile), flow rate 0.3 mL/min, injection volume 5 µL. For the MS spectrum, a range of 38000 – 41000 amu was selected in the total ion current (TIC) plot, and the maximum entropy deconvolution algorithm was applied. The MS spectra are shown in **Figure 61B**.

The His-tag was removed by TEV cleavage, whereby the TEV protease digested specifically at the recognition site ENLYFQ|G between the H6-tag and the TrpRS (**Figure 61A**). TEV cleavage was performed overnight at 4 °C in TEV cleavage buffer (**Table 17**) using a 1:1 ratio TEV protease : H6-TEV-TrpRS with gentle shaking. Afterward, the TrpRS was separated from the cleaved H6-tag and the

protease by Ni-NTA affinity. The identity of the TEV-cleaved TrpRS variants was confirmed by LC-ESI-Q-TOF (**Figure 61C**).

A) MRGSHHHHHHGSRENLYFQ↓GMTKPIVFSGAQPSGELTIGNYMGALRQWVNMQDDYHCIYCIVDQHAITVRQDAQKLRKATLDTLALYL
ACGIDPEKSTIFVQSHVPEHAQLGWALNCYTYFGELSRMTQFKDKSARYAENINAGLFDYPVLMAADILLYQTNLVPVGEDQKQHLELS
RDIAQRFNALYGEIFKVPEPFIPKSGARVMSLLEPTKKMSKSDDNRNNVIGLLEDPKSVVKKIKRAVTDSDEPPVVRYDVQNKAGVSNL
LDILSAVTGQSIPELEKQFEGKMYGHLKGEVADAVSGMLTELQERYHRFRNDEAFLQQVMKDGAEKASAHASRTLKAVYEAIGFVAKP

B)

■ H6-TEV-TrpRS Q27P L252L   ■ H6-TEV-TrpRS Q27P   ■ H6-TEV-TrpRS wt

Mass	H6-TEV-TrpRS wt	H6-TEV-TrpRS Q27P	H6-TEV-TrpRS Q27P L25L
calculated	39844.41	39813.39	39813.39
observed	39845.10	39815.12	39814.56
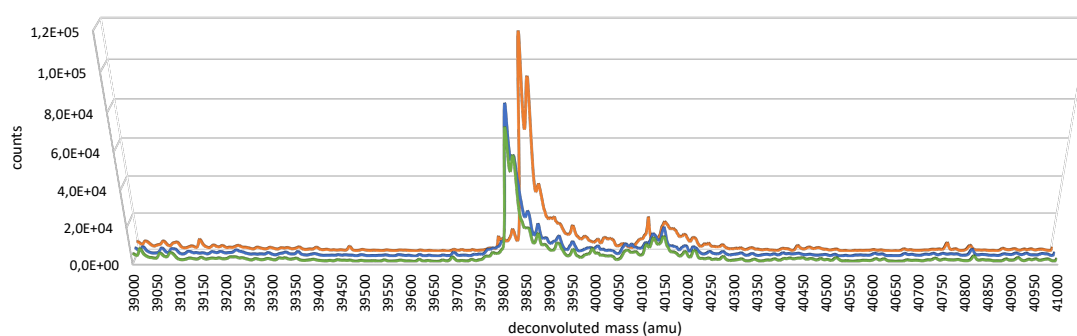
C)

■ TrpRS Q27P L252L   ■ TrpRS Q27P   ■ TrpRS wt

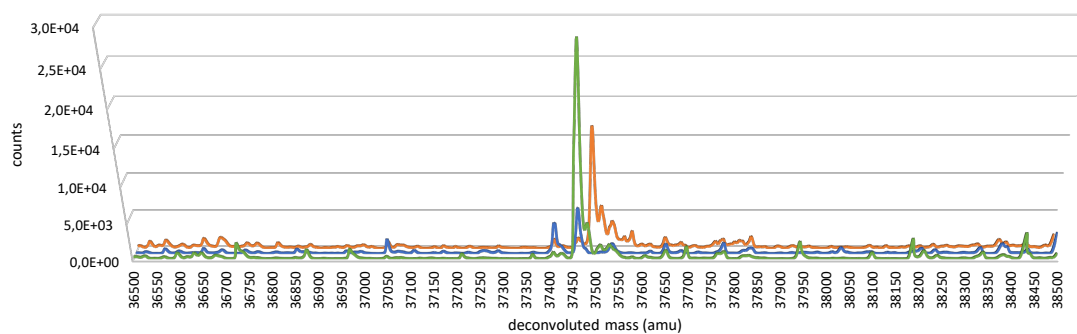| Mass	TrpRS wt	TrpRS Q27P	TrpRS Q27P L25L
calculated	37494.87	37463.86	37463.86
observed	37495.23	37464.23	37462.75

Figure 61 | Mass spectrometric analyses of TrpRS variants. A) A general sequence of H6-TEV-TrpRS is shown, where the His-tag (H6, HHHHHH) is colored orange, the TEV cleavage site (RENLYFQG) in pink, the mutation positions L25 in green (for synonymous variant L25L) and Q27 in blue (for missense variant Q27P). B) Mass spectra of recombinant expressed H6-TEV-TrpRS wt, H6-TEV-TrpRS Q27P, and H6-TEV-TrpRS Q27P L25L. C) Mass spectra of TrpRS wt, TrpRS Q27P, and TrpRS Q27P L252L after TEV cleavage. The calculated and observed masses are indicated in B) and C).

5.6.2 Expression of EGFP and ECFP in ALE strains

Variants of the green fluorescent protein (GFP) were expressed in the final isolates of 6TUB128-OC, 6TUB165-MB4, 6TUB165-MB3 and 7TUB165-OC, 7TUB165-MB, as well as in TUB00 (positive control). These variants are the C-terminal his-tagged enhanced green fluorescent protein (EGFP-H6) and the N-terminal his-tagged enhanced cyan fluorescent protein (H6-ECFP). The strains were rendered electro-competent (by three times washing in 10 % glycerol, see section **5.1.2.2**) and transformed with the respective expression plasmids pQE80L EGFP-H6 or pQE80L H6-ECFP, both endowed with ori ColEI, AmpR (see section **5.1.3.2**). Plasmid maps are deposited in the appendix (**Figure S79**).

Inoculated from cryo-stocks, precultures of transformed cells were incubated in NMM0-70-0 (6Fi or 7Fi), supplemented with 100 µg/mL ampicillin (Amp100) and subsequently 150 mL expression culture (NMM0-70-0 (6Fi or 7Fi), 100 µg/mL Amp) were inoculated to an OD_{600} of 0.05. NMM0-0-70 (Ind) was used accordingly for the TUB00 control. The cultures were incubated at 30 °C, 180 rpm, until they reached the exponential growth phase (OD_{600} 0.4 – 0.7). Recombinant protein expression was induced by adding 0.5 mM isopropyl β-*D*-1-thiogalactopyranoside (IPTG) and performed overnight at 30 °C, 180 rpm.

Cells were harvested by centrifugation (20 min, 10000 g, 4 °C) and resuspended in B-PER™ (Bacterial Protein Extraction Reagent, Thermo Fisher Scientific) according to the manufacturer's instructions. The lysis mixture contained B-PER, lysozyme, DNAse, and was incubated for 10 – 15 min at room temperature. After centrifugation of cell debris (5 min, 15000 g, RT), the cell lysate was filtered (Rotilabo PVDF Filter, 0.45 µm pore diameter), and target proteins were purified by using affinity chromatography on a Protino Ni-NTA column (1 mL Fast Flow, Machery-Nagel). The purification was carried out using a single-channel peristaltic pump P-1 (VWR) and sodium phosphate-based purification buffer (50 mM Na-P, 300 mM NaCl, pH 7.8, two-step imidazole gradient: 20 mM and 40 mM); followed by elution with 500 mM imidazole. EGFP- or ECFP-containing fractions were identified, and chromophore maturation was proved by the detection of fluorescence by irradiation with a UV lamp at 356 nm. Hereinafter, respective elution fractions were collected, and imidazole was removed by dialysis (Microsep™ Advanced Centrifugal Devices, MCOW 3kDa, Pall Laboratory) against 50 mM Na-P, 100 mM NaCl (pH 7.8). Finally, protein identity and incorporation of 6- and 7-fluorotryptophan was confirmed by LC-ESI-Q-TOF mass spectrometry (Agilent 6530 Accurate-Mass Q-TOF).

Protein samples were measured in a concentration of 0.1 mg/mL using the following HPLC parameters: linear gradient from 5 % to 80 % buffer A within 20 min (A: 0.1 % formic acid in MQ-H_2O, B: 0.1 % formic acid in acetonitrile), flow rate 0.3 mL/min, injection volume 5 µL. For the MS spectrum, a range of 27000 – 29000 amu was selected in the total ion current (TIC) plot, and the maximum entropy deconvolution algorithm was applied. The entire MS spectra are shown in **Figure S76**, the mass values are shown in **Table S36**, and the isolated yields are shown in **Table S37**.

5.6.3 Polyacrylamide gel electrophoresis (SDS–PAGE)

The SDS-PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis) is used for the analysis of proteins by their electrophoretic separation according to their charge, size, and molecular weight.[558] Thereby, the anionic surfactant SDS (sodium dodecyl sulfate) masks the inherent charges of proteins, resulting in a constant negative charge distribution. The SDS-PAGE process consists of gel production, sample preparation, electrophoresis, protein staining, and analysis of the generated band pattern.

In this study, a discontinuous polyacrylamide-based gel was used for electrophoresis. The gel matrix consists of acrylamide and N, N'-methylene bisacrylamide (ratio of 37.5 : 1) and is formed by radical polymerization, whereby the bisacrylamide permits the crosslinking of the otherwise linear polyacrylamide chains. A 12 % resolving gel is covered with a lower layer of 4 % stacking gel (**Table 17**); added ammonium peroxodisulfate (APS) and the polymerization catalyst tetramethyl ethylene diamine (TEMED) serve as a radical starter for the polymerization.

Samples collected during the protein expression procedure (usually comprising not induced, induced, flowthrough, pellet, lysate, wash steps, and elution fractions) were mixed 5:1 with 5xSDS loading dye and denatured by heating up to 95 °C for 10 min. This treatment causes a breakdown of the secondary and tertiary structures, resulting in linearized proteins and allows separation according to the chain length, proportional to the molecular weight. The sample buffer fulfills several functions: the contained β-mercaptoethanol causes an additional reduction of disulfide bridges, a high concentration of glycerol ensures the sinking of the samples into the gel pockets, and bromophenol blue allows easy handling of the otherwise colorless samples. For electrophoresis an SDS-running buffer (**Table 17**) was used, and 10 µL of the denatured protein samples as well as of a standardized molecular-weight size marker (PageRuler Unstained Protein Ladder, Thermo Fisher Scientific) were loaded into the gel pockets. Thereafter, a voltage of around 80 – 180 V was applied, which caused a migration of negatively charged molecules through the gel in the direction of the anode. Along with a discontinued SDS-PAGE, different voltages are applied to migrate the proteins through the gel. The proteins migrate first into the stacking gel with 80 V, in which they are concentrated, and then they migrate into a separating gel with 150 – 180 V, in which the actual separation takes place. Stacking and separating gels differ by different pore size, ionic strength, and pH values. Small proteins migrate relatively easily through the gel, while larger proteins are more likely to be retained and, thereby, migrate more slowly through the gel. Thus, the gel acts like a sieve, whereby the acrylamide concentration determines the pore size.

After the electrophoretic separation, the protein bands were stained with Coomassie solution for 1 – 4 h. The Coomassie Brilliant blue R-250 dye interacts electrostatically with the amino and carboxyl groups of proteins and forms a deep blue protein-dye complex. The gels were destained either by washing out the dye with water for several days or by using a 10 % acidic acid solution within a few hours. The band pattern then provides information about the size of the proteins to be analyzed.

5.6.4 Determination of the protein concentration

The concentration of purified and dialyzed proteins was determined by measuring the absorbance at 280 nm. Analogous to the quantification of DNA (see section **5.4.4**), the Lambert-Beer law (Eq. 4) can also be used for calculating the protein concentration, where A_{280} = absorbance at 280 nm, I_0/I = incident and transmitted radiation intensity, c = concentration of the absorbent substance in mol/L and d = the path length in cm. The molar extinction coefficient ε_m was calculated with the ProtParam tool using the amino acid sequence of the respective target protein (Expasy Proteomics Server, https://web.expasy.org/protparam/). For wild-type ECFP, an ε_m = 25900 L/mol cm and for wild-type EGFP ε_m = 21890 L/mol cm were computed. The protein concentration can also be determined using the Bradford assay, described in section **5.9.2**.

$$\text{Lambert-Beer law} \qquad A_{280} = log\left(\frac{I_0}{I}\right) = \varepsilon_m \cdot c \cdot d \qquad\qquad \text{Eq. 4}$$

5.7 Analysis of the purity of indole analogs 6Fi and 7Fi

6-fluoroindole, 7-fluoroindole and indole were dissolved 96 % EtOH.

5.7.1 GC–MS analysis

Indole was used as a standard to set the GC parameters and to determine the limit of detection (LOD) of the method. Then equal mixtures of indole and either 6-fluoroindole or 7-fluoroindole were used to develop a method with sufficient separation capacity (retention time indole: 9.125 min, retention time 6Fi: 9.303 min, retention time 7Fi: 8.670 min). Total ion chromatograms (TIC) of 6- and 7-fluoroindole were screened for contaminations, especially for adaptation-disturbing non-fluorinated indole.

The analysis by gas chromatography coupled with mass spectrometry was performed on an Agilent 5977 MSD system with the following parameters: column 5 % phenyl-methylpolysiloxane (Agilent 19091S-433UI), 325 °C: 30 m x 250 µm x 0.25 µm; gradient 50 °C (3 min isothermal), ramp 20 °C/min to 300 °C for 2 min; full-scan (50 – 350 m/z); injector 1 µL with 10:1 split ratio; flow rate 1 mL/min; pressure 7.6522 psi; heater 300 °C; EI mode MS source: 230 °C, MS quad 150 °C. All measurements were performed at least twice.

5.7.2 HPLC-MS-QQQ analysis

Also, in the HPLC analysis, indole was used as a standard to determine the separation parameters and LOD using a dilution series. 6Fi and 7Fi were not detectable with this analysis, however, evidence for the absence of indole was obtained. For this, 6Fi and 7Fi preparations, as well as mixtures of indole and either 6Fi or 7Fi, were analyzed in selected ion monitoring (SIM) mode using a mass filter of 118 m/z.

The analysis by high-performance liquid chromatography (HPLC) coupled with mass spectrometry was performed on an Agilent 6460 Triple Quadrupole LC/MS System with 1290 Infinity HPLC system (Agilent), operated in positive ionization mode. HPLC separation was performed using a Grom-Sil 120 ODS-5ST, 3 µm, 100 x 2.0 mm column (Grace), with flowrate 0.6 mL/min and the following gradient (A: H_2O/B: ACN): isocratic equilibration 5% B for 2.5 min, linear gradient of 5 - 100 % solvent B in 0.5 min, hold at 100 % B for 10 sec following isocratic at 5 % B for 1 min 20 sec. Either ESI full-scan (50 – 350 m/z) or SIM scan (118 m/z) were applied. All measurements were performed at least twice.

5.8 Enzymatic synthesis of 6- and 7-fluorotryptophan

The synthesis of tryptophan from indole corresponds to the last step of the endogenous Trp biosynthesis and is catalyzed by the tryptophan synthase (TrpS). The reaction is a cofactor (PLP) mediated condensation of L-serine and indole to yield Trp. Owing to its substrate promiscuity, TrpS also permits the conversion of fluorinated indole to fluorinated Trp. This reaction is described in the literature for a variety of indole analogs,[156,165] but was also confirmed experimentally for the target substrates 6Fi and 7Fi, as well as Ind, which serves as a positive control.

For the enzymatic synthesis, the TrpS enzyme of *Salmonella typhimurium* was used, a courtesy of Dr. Stefan Oehm who expressed and purified the enzyme using crystallization.[559] The reaction conditions were adapted from the Wilcox protocol.[168] The optimized reaction (1 mL) encompassed 1 mM L-serine, 1 mM indole or 6- or 7-fluoroindole, 0.8 mM pyridoxal-5'-phosphase (PLP), and 0.5 mM Trp synthase in 50 mM potassium phosphate buffer (pH 7.8). The mixture was incubated for 24 h at 37 °C, and the reaction progress was monitored by thin-layer chromatography (TLC)[560] using ninhydrin staining to detect the amine function of the resulting amino acid (**Figure 62**); stationary phase: TLC silica gel 60 F$_{254}$, mobile phase: nBuOH : CH$_3$COOH : H$_2$O (2 : 1 : 1), UV = 254 nm. After purification of the crude product by ion-exchange chromatography (IEX; Dowex 50WX8 50-100 mesh cationic resin, elution by ≈ 12 % NH$_3$), the conversion of the desired Trp, 6FTrp or 7FTrp was confirmed by HPLC-ESI-MS (from Center for Mass Spectrometry of Technische Universität Berlin). HPLC was performed on a 1290 Infinity HPLC system (Agilent) coupled to LTQ Orbitrap XL (Thermo Fisher Scientific) using the following parameters: Grom-Sil-120-ODS-5-ST, 3 µm, 50 x 2.0 mm column (Grace), flowrate 0.3 mL/min, gradient (A: H_2O +

0.1 % HCOOH/B: ACN + 0.1 % HCOOH): linear gradient of 20 – 100 % solvent B in 10 min, hold at 100 % B for 3 min following 20 % B for 5 min, DAD spectrum scan range 205 nm – 850 nm (step 2 nm), UV/Vis channels: A (215 nm, 4 nm bandwidth), B (280 nm, 40 nm bandwidth) and C (350 nm, 100 nm bandwidth). For the ESI full-scan, a mass range of 50 – 500 m/z was applied. As controls, commercially available compounds (purchased from abcr) were used.
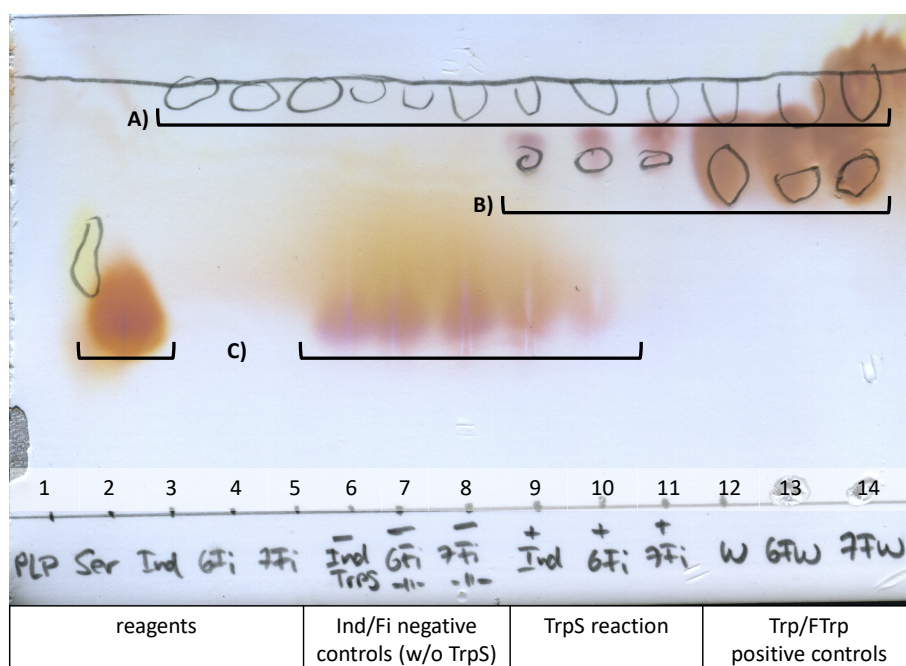

Figure 62 | Evaluation of TrpS reaction progress by TLC and ninhydrin staining. A) The indole moiety of indoles (spots 3 – 11) and tryptophans (spots 12 – 14) was detected by UV at 254 nm, B) the conversion of indoles into tryptophans was detected by UV (254 nm) and by ninhydrin staining, which attests the amino group (spots 9 – 11 for *in vitro* TrpS reaction and spots 12 – 14 for positive controls), and C) L-serine (spot 2 reagent and spots 6 – 11 reaction mixture) was also detected via its amino functionality by ninhydrin staining.

5.9 Determination of proteome-wide FTrp substitution

5.9.1 Culture preparation for amino acid analysis and proteomics

Final isolates of the adapted strains 6TUB128-OC, 6TUB165-MB4, 6TUB165-MB3, 7TUB165-OC, 7TUB165-MB and positive controls Ind-TUB165, W-TUB165 as well as TUB00 were grown in the respective NMM supplemented with either 70 µM indole, 6Fi or 7Fi. Exponential growing cells were harvested by centrifugation (18000 g, 4 °C, 20 min), and pellets were stored at -20 °C. Additionally, supernatant samples were collected from the cultures after 24 h, 48 h, 72 h, 96 h, and 120 h and also stored at -20 °C.

5.9.2 Amino acid analysis by HPLC-UV/FL

For amino acid analysis, the proteome of cellular samples was extracted, and resulting peptides/proteins were hydrolyzed to yield the free amino acids.

Therefore, cell pellets were thawed on ice and lysed using B-PER™ (Bacterial Protein Extraction Reagent, Thermo Fisher Scientific) according to the manufacturer's instructions. The lysis mixture contained B-PER, lysozyme, DNAse, and was incubated for 10 – 15 min at room temperature. After centrifugation of cell debris (15.000 g, 5 min, RT), proteins were precipitated from the cell lysate by adding four volumes of -20 °C acetone (200 µL lysate and 800 µL acetone). The mixtures were vortexed and incubated for at least 1 h at -20 °C, followed by centrifugation (10 – 30 min, 17.000 g, 4 °C) and twice washing of precipitated proteins with each 1000 µL 90 % acetone (-20 °C). Miscible organic solvents such as acetone decrease the dielectric constant of water, which causes proteins in the solution to precipitate. Then, proteins were re-solubilized in 500 µL 8 M urea for 30 min at RT. By treatment with such chaotropic agents' the hydrogen bonding network between water molecules and, therefore, hydrophobic interactions in proteins are disrupted. Residual urea was removed by size exclusion chromatography (SEC) using Illustra NAP-10 columns (GE Healthcare) equilibrated with 150 mM NaCl and 25 mM potassium phosphate (SEC buffer, pH 7.8), according to the manufacturer's instructions. The protein concentration was measured calorimetrically by Bradford assay.[561] The triphenylmethane dye Coomassie Brilliant Blue G-250 (unbound cationic red form, A_{max} 470 nm) forms complexes with cationic and nonpolar side chains of proteins, stabilizing the dye in its blue, unprotonated, anionic sulfonate form and shifting the absorption maximum to 595 nm. Samples were diluted 1:100 in 1x Bradford solution, incubated for 5 min at RT in the dark, and absorption at 595 nm was measured. For quantification, a BSA (bovine serum albumin) calibration (0.2, 0.4, 0.6, 0.8, 1.0 mg/mL) was used. Then, the solutions were transferred into glass ampoules and lyophilized. 200 µL of 2.5 % thioglycolic acid in 6 M HCl_{aq} (grade for amino acid analysis; Sigma Aldrich) was added, and the ampoule was sealed under vacuum. Oxygen exclusion and the addition of thioglycolic acid prevent oxidative destruction of tryptophan during acid hydrolysis.[391] The proteins were hydrolyzed for 24 h at 110 °C. The hydrolysis was also conducted for NMM0-70-0 (6Fi) and NMM0-70-0 (7Fi) in order to investigate whether the harsh condition of the acid treatment, had an effect on the fluorinated indoles, such as fluoride cleavage or other decomposition reactions. Hydrolyzed samples of TUB00, Ind-TUB165, W-TUB165, 6TUB128-OC, 6TUB165-MB4, 6TUB165-MB3, 7TUB165-OC, 7TUB165-MB were subjected to HPLC-UV/FL analyses, which were performed by Carlo Fasting at Freie Universität Berlin. The following parameters were applied: Machery-Nagel Nucleoshell Phenylhexyl column (5 µm, 150 x 4.6 mm) which provided good separation results due to doubling of the retention times caused by aromatic interactions, column pressure 230 bar (solvent A) and 120 bar (solvent B), flow rate 1 mL/min, solvent A: 10 % MeCN (80 % (v/v), H_2O (20 % (v/v) / 90 % AcOH (0.5 M), solvent B: MeCN (80 % (v/v) H_2O (20 % (v/v), 2x Knauer K-1000 pump, UV detection at 280 nm, fluorescence excitation at 280 nm and emission at 350 nm and sensitivity MED/potentiation 4-times.

5.9.3 Peptide analysis by nLC-MS/MS (Proteomics)

E. coli extracts (see section **5.9.1**) were briefly run into an SDS-gel until the dye front had migrated about 5 mm into the gel, and only the 3 mm behind the dye front were used for protein digestions. This step removed all positively charged molecules, small highly mobile negatively charged ions, uncharged molecules, and very large agglomerates that did not enter the gel. Following a protocol described by Kublik *et. al.*,[562] the gel slices were washed with ddH$_2$O, and the proteins were reduced with 10 mM dithiothreitol (DTT) for 30 min at room temperature to break disulfide bridges. To protect the reduced cysteine residues, carbamidylation using 100 mM iodoacetamide was performed for 30 min at room temperature in the dark. Subsequently, an overnight in-gel protein digestion using trypsin (at 37 °C) was initiated to generate peptides. The resulting peptides were extracted from the gel pieces by incubation in 50 % (v/v) acetonitrile and 5 % (v/v) formic acid for 10 min at least two times and desalted using C18-tips through zip-tipping. Extracted peptide samples were then analyzed by nLC-MS/MS on a nanoUPLC system (nanoAcquity, Waters) coupled to an Orbitrap Fusion mass spectrometer (Thermo Scientific) operated in positive mode, as previously described.[563] Briefly, the separation was performed on an Acclaim PepMap100 C18 column (Thermo Scientific) at a flow rate of 0.3 µL/min using a linear gradient of eluent A (0.1 % formic acid in water) and eluent B (80 % (v/v) acetonitrile, 0.08 % formic acid in water); followed by ESI continuous scan between 350 and 2000 m/z by Orbitrap mass analyzer and a resolution of 120,000. For fragmentation, the higher energy collisional dissociation (HCD) mode with a collision energy of 30 % was used. For the identification of peptides, carbamidomethylation of cysteine residues was taken as a fixed modification, oxidation of methionine residues, and fluorination of tryptophan residues were taken as dynamic modifications. Fluorination was calculated as the substitution of a hydrogen atom by fluorine, giving a mass change of m/z +17.991. Protein identification was achieved using Proteome Discoverer (v2.4, Thermo Fisher Scientific) with the *E. coli* fasta genome database and SequestHT as the search engine, setting a false discovery rate threshold of 1 % for peptide identification using the Target Decoy PSM Validator node. The abundance of proteins was quantified through label-free quantification based on intensity values in precursor scans using the Minora node in Proteome Discoverer. The described sample preparation, nLC-M/MS analysis, and data processing were performed by Lorenz Andrian at Helmholtz-Zentrum für Umweltforschung (UFZ Leipzig).

5.10 Metabolome profiling using FIA-QTOF-MS

Isolates of the strains TUB00, 6TUB53, 7TUB53, 6TUB161-MB4, 7TUB161-MB and 6TUB124(35)-SP were cultured in respective medium, that was NMM0-0-70 (Ind) for TUB00, NMM0-70-0 (6Fi) for 6Fi-adapted strains and NMM0-70-0 (7Fi) for 7Fi-adapted strains. The middle-time point strains 6TUB53 and 7TUB53 are adapted to grow on 19-70-0 (6Fi or 7Fi, respectively) and were re-grown in corresponding

NMM without cAA addition (NMM0-70-0 supplemented with either 6Fi or 7Fi). Cells were cultured to their individual mid-exponential phase and placed on ice to quench metabolism. For in situ hot water extraction of intracellular metabolites, 1 mL bacterial cell pellet corresponding to OD_{600} 1 was mixed with 1 mL 80 °C hot MilliQ water, vortexed for 10 sec, incubated for 3 min at 80 °C in an Eppendorf thermomixer at max. speed (1500 rpm); thereby, the mixtures were additionally vortexed for 10 sec once per minute. Afterward, the samples were centrifuged at RT for 3 min at max. speed (about 17000 g) and 500 µL of clean precipitate-free supernatants were collected in safe-lock tubes. Finally, the samples were stored at -80 °C and shipped on dry ice to ETH Zurich for metabolome analysis.

Untargeted metabolomics measurements were performed by Karin Meier at ETH Zurich using FIA-QTOF-MS consisting of an Agilent Series 1100 LC pump coupled to a Gerstel MPS2 autosampler and an Agilent 6550 accurate mass Q-TOF mass spectrometer using Dual Agilent Jet Stream Electrospray Ionization (Agilent Technologies), operated in negative ionization mode. The following parameters were applied: flow rate 150 µL/min, mobile phase consisted of isopropanol/MilliQ water (60:40 v/v) buffered with ammonium carbonate and added with 1 µM of hexakis (1H, 1H, 3H tetrafluoropropoxy) phosphazine (HP-0921; Agilent Technologies) for mass axis correction. Mass spectra were recorded from m/z 50 to m/z 1000 with a frequency of 1.4 spectra/sec for 0.48 min using 4-GHz high-resolution settings. Electrospray settings were as follows: ion spray voltage 3.5 kV negative mode, capillary temperature 325 °C, 5 L/min drying gas flow, nebulizer pressure 30 psi; fragmentor, skimmer and octopole voltages were set to 175, 65, and 750 V, respectively. Data processing and ion annotation were performed by Nicola Zamboni from ETH Zurich, who provided curated data sets.

5.11 Genome Analysis

Whole genome sequencing (WGS) was used to identify and characterize the mutations that drive the adaptation process. This entails determining the entire DNA sequence of an organism's genome. Therefore, strains along the adaptation trajectory (**Table 9**) were cultured in LB medium, and gDNA was extracted using Wizard® Genomic DNA Purification Kit (see section **5.1.4**). DNA concentration was measured using a NanoDrop device and stored, if necessary, at -20 °C. Later on, the genomic dataset was expanded with samples from the hypermutator lineages 6TUBX-MB4 (passages 93, 98, 108, 113, 118, 123, 138, 148 and 158) and 6TUBX-MB3 (passages 95, 102, 109, 116, 123, 130, 137, 144, 151 and 158). All samples fulfilled the quality criteria described in section **5.4.4**, that were ratios of A260/280 = 1.8 – 2.0 and A260/230 = 2.0 – 2.2.

Table 9 | Summary of gDNA samples subjected to whole genome sequencing.

ALE lineage	passage number							
	18	**23**	**53**	**72**	**83**	**88**	**128**	**165**
W-TUBX	X	X	X	X	X	X	X	X
Ind-TUBX	X	X	X	X	X	X	X	X
6TUBX-OC		X	X			X	X	
6TUBX-MB4		X	X			X	X	X
6TUBX-MB3		X	X			X	X	X
7TUBX-OC	X		X		X			X
7TUBX-MB	X		X	X				X

All samples were Illumina paired-end sequenced (2x 150bp). Genomes were assembled using SPAdes (v3.15.3) with the "--isolate" flag. The resulting contig names were shortened using the rename function of BBMap (v38.92). TUB00 (sample 1) was extracted and mapped to the *E. coli* MG1665 reference genome (NC_000913.3)[564] using the MAUVE plugin (v1.1.1) within Geneious Prime (v2020.2.3). Ordered contigs were merged by concatenating them with 100 Ns using the masking function of Geneious. TUB00 was then annotated using Prokka (v1.14.6), with its settings specified for *E. coli* and enabling the annotation of ncRNA ("--rfam"). This resulted in an ordered, annotated reference genome, which was used for further SNP calling. All assemblies were mapped by Snippy (v4.6.0) to the annotated reference genome. SNPs were then called, and their functional impact was assessed by the Snippy pipeline. In addition, Snippy was used to generate a core SNP profile. The resulting SNP profiles were analyzed in R (v4.1.1) and visualized using ggplot2 (v3.3.5). Genomic analysis was performed in collaboration with Torsten Semmler and Silver Wolf at the Robert Koch Institute (Berlin). Torsten Semmler conducted sample preparation by means of generating sequencing libraries and Illumina WGS, and Silver Wolf conducted data processing and SNP calling.

5.12 Growth assay to determine the *sspA* impact

The impact of *sspA* deletion on the adaptation process, that is, conferring *E. coli* a growth advantage, was investigated by growth experiments. The ancestral TUB00, the deletion mutant TUB00 Δ*sspA* and 6Fi-adapted 6TUB128-OC were resuscitated in 4 mL NMM19-0-10 (Ind); and 7Fi-adapted 7TUB165-OC must be resuscitated in 4 mL LB at 30 °C, 180 rpm. Cells of 7TUB165-OC were re-grown (1 : 50 inoculation) in NMM0-70-0 (7Fi) twice to remove residual LB and additionally once in NMM19-0-10 (Ind) to set the same starting conditions as with the other samples.

Afterward, the cells were washed three times with 1 mL NMM0-0-0 each to remove residual Ind, 6Fi, or 7Fi (2000 g, 3 min, RT), normed to OD_{600} = 1 (prevention of differences of inoculation volume), and

then 198 µL of the respective medium composition were inoculated to OD_{600} = 0.02. The tested media comprised NMM19-70-1 (6Fi or 7Fi, Ind), NMM19-70-0 (6Fi or 7Fi), NMM0-70-0 (6Fi or 7Fi), and NMM0-0-70 (Ind) as a positive control as well as NMM0-0-0 as a negative control. The growth was monitored in 96-well plates (Greiner, flat bottom clear), sealed with the breathable membrane (Breathe-Easy sealing membrane, Sigma Aldrich) at 30 °C with continuous shaking using a microplate reader (Tecan Infinite M200) that measured the absorbance at 600 nm in 10 min intervals. The measurements were performed at least twice and in triplicates (biological replicates) by Jana Madeleine Göller. The entire dataset is deposited in **Figure S74** in the appendix.

5.13 Nile Red fluorescence assay and microscopy

Cells of TUB00 and all adapted strains were resuscitated from cryo-stocks and grown in their respective NMM (TUB00 and positive controls in Ind or Trp, 6TUBX in 6Fi, and 7TUBX in 7Fi) with and without the addition of 19 cAAs, respectively. For Nile Red staining, cells were harvested by centrifugation (2000 g, 2 min, 4 °C), washed twice with PBS buffer (0.1 M Na_2HPO_4, pH 7.4), and finally normed to OD_{600} = 1. Subsequently, cell suspensions were mixed 1:500 incubated with Nile Red (10 mM stocks in DMSO, final concentration 20 µM) and incubated for 30 min in black 96-well microplates with transparent bottom, suitable for fluorescence measurements. Quantitative fluorescence reading was performed on a Synergy™ HTX Multi-Mode Microplate Reader, with filter settings: excitation at 530 nm, bandwidth 25 nm, and emission at 620 nm, bandwidth at 40 nm.

For fluorescence microscopy, 8 µL of the cell/Nile Red suspensions were transferred onto an object slide for cell imaging. Images were recorded with a Leica DMI6000B inverted microscope in fluorescence mode and white light using the following settings: objective HCX PL APO 63x/1.40 oil, magn. camera 63x, fluorescence was measured with excitation at 549 nm, emission at 562 nm, and exposure for 250 ms. Recorded images were processed and analyzed with LAS X. Fluorescence measurements and imaging were performed by Jana Madeleine Göller in cooperation with Fereshteh Ghazisaeedi (Freie Universität Berlin). All measurements were performed in biological triplicates.

6 Materials

6.1 Microorganisms

Table 10| Microorganisms used in this study. Microorganisms are of the species *Escherichia coli* (*E. coli*) if it is not stated otherwise.

strain	Genotype	Source and description
MG1655	*E. coli* K-12 F⁻ λ⁻ *ilvG⁻ rfb-50 rph-1*	Defined as the wild-type *E. coli* K-12 strain.[564]
TUB00	*E. coli* K-12 MG1655 Δ*trpLEDC*::FRT Δ*tnaA*::FRT	Dr. Volker Doering, Dr. Michael Hoesl. Ancestor for ALE.
TUB00 Δ*sspA*	*E. coli* K-12 MG1655 Δ*trpLEDC*::FRT Δ*tnaA*::FRT Δ*sspA*	created in this study (section **5.5.3**)
W-TUBX	see **Table S29**	ALE strains created in this study (section **3.4.1**); Trp-adapted positive controls; lineage comprises 165 passages/strains
Ind-TUBX	see **Table S30**	ALE strains created in this study (section **3.4.1**); Ind-adapted positive controls; lineage comprises 165 passages/strains
6TUBX-OC	see **Table S33**	ALE strains created in this study (section **3.4.2**); 6Fi-adapted strains (OC approach); lineage comprises 128 passages/strains
6TUBX-MB4	see **Table S34**	ALE strains created in this study (section **3.4.2**); 6Fi-adapted strains (MB approach); lineage comprises 165 passages/strains
6TUBX-MB3	see **Table S35**	ALE strains created in this study (section **3.4.2**); 6Fi-adapted strains (MB approach); lineage comprises 165 passages/strains
7TUBX-OC	see **Table S31**	ALE strains created in this study (section **3.4.3**); 7Fi-adapted strains (OC approach); lineage comprises 165 passages/strains

7TUBX-MB	see **Table S32**	ALE strains created in this study (section **3.4.3**); 7Fi-adapted strains (MB approach); lineage comprises 165 passages/strains
6TUBX-SP	The sequenced isolate 6TUB124-SP accumulated 142 SNPs. The data are available upon request.	Subpopulation ALE strains created in this study (section **3.4.2.1**); 6Fi-adapted strains (under increased 6Fi concentration); lineage comprises 34 passages/strains
NEB10-beta	*E. coli* B *Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- ϕ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL* (Str^R) *rph spoT1 Δ(mrr-hsdRMS-mcrBC)*	Cloning strain (New England Biolabs).
BL21(DE3)	*E. coli* B F- *[lon] ompT gal* (λ DE3) *[dcm] ΔhsdS*(r^-_B m^-_B); λ DE3 = λ *sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7&nb*	Used for protein expression (New England Biolabs).

6.2   Primers

All oligonucleotides were purchased from Sigma-Aldrich (Taufkirchen, Germany), dissolved in MQ-H$_2$O to a final concentration of 100 µM; and stored at -20 °C. They were used in 10 µM or 20 µM working concentrations. Primers were generally purchased in desalted form, and only primers exceeding about 50 bp in length were purchased as HPLC-purified.

Table 11 | Primers used in this study.

Name	Sequence (5′ – 3′)
Δ*trpLEDC* verification	
C1_trpA	CTCCCGTTGCGTTGCATTG
C2_trpL	CGACCATGACACCACAGCT
C3_trpB	TGTCTTTATCGCCGCGACC
C4_trpB	CGCCGTGCAGCAAATCTTC
C5_trpE	AGTTCGTTCAGGCGAGCAG

Δ*tnaA* verification

C1_tnaA	CTTGATTTGCCCTTCTGTAGC
C2_tnaA	CTTGATCAGTCATGATGCCAC
C5_tnaA	CGCAAAGGGTTCTGCACTC

TrpRS cloning	**verification of *trpS***
ecWRS – fwd (PB527)	TACGGATATAACTACGGCGAGG
ecWRS – rev (PB528)	AAATTGCCTGATGCGCTACG

	Golden Gate Assembly (pQE80L TrpRS)
TrpS insert - fwd	GGCTACACCTGCATAGTTCCAGGGCATGACTAAGCCCATC GTTTTTAG
TrpS insert - rev	GGCTACACCTGCCGACGCTAATTAAGCTTTTATTACGGCTT CGCCACAAAAC
pQE80L backbone – fwd	GGCTACACCTGCCGACTAGCTGAGCTTGGACTCC
pQE80L backbone – rev	GGCTACACCTGCATAGGGAAATACAAGTTTTCTCTGGATC

sspA deletion

L_homo - fwd	GTCGGTTTCGTGGTGAATATGATTGC
L_homo - rev (BsaI)	GGCTACGGTCTCAGCCATCTGTATGGATTTGTCACAGCTAA CACC
CmRN20PAM - fwd (BsaI)	GGCTACGGTCTCATGGCGAAAATGAGACGTTGATCG
CmRN20PAM - rev (BsaI)	GGCTACGGTCTCACTACCTTACTTCGGTTCGATGGACTATT ACGCCCCGCCCTG
Ls_R_homo - fwd (BsaI)	GGCTACGGTCTCAGTAGGGACGACGTGGTGTTAGCTGTGA CAAATCCATACAGAGAAAACCTCCAGGTATAG
L_short - rev	CGACATGGTTGAGAAAC

ECFP/EGFP sequencing

GFP – fwd (PB 1016)	GATGGAGTTCTGAGGTCATTACTGG
GFP – rev (PB 372)	GAGCGGATAACAATTTCACACAGG

6.3 Plasmids

Table 12 | Plasmids used in this study.

plasmid	ori	Resistance marker	origin
pCAGO	pSC101	AmpR, CmR	Zhao *et. al.* [457] (Genscript)
pQE80L H6-TEV-TrpRS wt	ColE1	AmpR	generated in this study
pQE80L H6-TEV-TrpRS Q27P	ColE1	AmpR	generated in this study
pQE80L H6-TEV-TrpRS Q27P L25L	ColE1	AmpR	generated in this study
pQE80L H6-ECFP	ColE1	AmpR	Budisa group (PDB ID 06)
pQE80L EGFP-H6	ColE1	AmpR	Budisa group (PDB ID 156)

6.4 Biomolecular reagents, standards, preparation Kits, and enzymes

Table 13 | Biomolecular reagents and standards used in this study.

Reagents	
dNTP mix	Thermo Fisher Scientific (Waltham, USA)
GeneRuler DNA Ladder Mix	Thermo Fisher Scientific (Waltham, USA)
Ethidium bromide solution 1%	Carl Roth (Karlsruhe, Germany)
GelRed nucleic acid stain	Biotium, Inc. (Fremont, USA)
Midori green dye	NIPPON Genetics EUROPE GmbH (Düren, Germany)
PageRuler Unstained Protein Ladder	Thermo Fisher Scientific (Waltham, USA)
ROTI®Quant	Carl Roth (Karlsruhe, Germany)
Nile Red	Merck KGaA (Darmstadt, Germany)

Table 14 | Kits used in this study.

Preparation Kit	
GeneJET™ PCR Purification Kit	Thermo Fisher Scientific (Waltham, USA)
GeneJET™ Gel Extraction Kit	Thermo Fisher Scientific (Waltham, USA)
GeneJET™ Plasmid Miniprep Kit	Thermo Fisher Scientific (Waltham, USA)
GeneJET™ Plasmid Midiprep Kit	Thermo Fisher Scientific (Waltham, USA)
Wizard® Genomic DNA Purification Kit	Promega GmbH (Walldorf, Germany)
MTT Assay Kit (Cell Proliferation)	Abcam (Cambridge, UK)
Cell Counting Kit - 8	Merck KGaA (Darmstadt, Germany)

Table 15 | Enzymes used in this study.

Enzymes	
Phusion High-Fidelity DNA Polymerase	Thermo Fisher Scientific (Waltham, USA)
Q5® High-Fidelity DNA Polymerase	New England Biolabs (Ipswich, USA)
DreamTaq™ DNA Polymerase	Thermo Fisher Scientific (Waltham, USA)
Taq DNA Polymerase	AG Budisa
T4 DNA Ligase	Thermo Fisher Scientific (Waltham, USA)
Restriction Enzymes	New England Biolabs (Ipswich, USA)
FastDigest Restriction Enzymes	Thermo Fisher Scientific (Waltham, USA)
Lysozyme	Carl Roth (Karlsruhe, Germany)
DNase	Carl Roth (Karlsruhe, Germany)
RNase	Carl Roth (Karlsruhe, Germany)

6.5   Media and Supplements

For standard cultivation of *E. coli* under non-limiting conditions LB medium was used. LB was prepared using dH$_2$O and autoclaved for 20 min at 121 °C, 1.5 bar for sterilization. LB-Agar was prepared by adding 1.5 % agar-agar to LB media prior to autoclaving and used for singling or selection. Super Optimal Broth with Catabolite Repression (SOC) medium was prepared freshly from Super Optimal Broth (SOB)[508] medium by supplementation of 20 mM sterile glucose and used in the recovery step of the transformation of chemical- or electrocompetent cells. In contrast, New Minimal Medium (NMM)[366] is a glucose-based mineral salt medium for cultivation under growth-limiting conditions and was used for long-term propagation of *E. coli* during the Adaptive Laboratory Evolution (ALE) experiments and for controlled heterologous protein expression for the purpose of ncAA incorporation (Selective Pressure Incorporation (SPI) method).[237] When NMM was used, the addition of 19 cAAs (NMM19(-W), except for the target cAA Trp) was optional and represented a relaxation of the growth conditions. The NMM and all stocks were prepared with MQ-H$_2$O and sterilized by filtration or autoclaving. NMM-Agar plates were poured by mixing 2x concentrated NMM with 3 % (w/w) agar-agar in dH$_2$O. Supplements were prepared with dH$_2$O (except for Cm37, which was solved in abs. EtOH), sterilized by filtration (Ø 0.22 µm), and subsequently added to the media. 6-fluoroindole, 7-fluoroindole and indole were dissolved 96 % EtOH.

Table 16 | Media and Supplements used in this study.

Medium composition	Final concentration
Lysogeny Broth (LB)	
Yeast extract	10 g/L
Bacto Tryptone	5 g/L
NaCl	10 g/L
LB-Agar	
LB + 1.5 % agar-agar	15 g/L
Super Optimal Broth (SOB)	
Yeast extract	5 g/L
tryptone	20 g/L
NaCl	10 mM
KCl	2.5 mM
$MgCl_2$	10 mM
$MgSO_4$	10 mM
New Minimal Medium (NMM)	
$(NH_4)_2SO_4$	7.5 mM
NaCl	8.5 mM
KH_2PO_4	22.5 mM
K_2HPO_4	50 mM
$MgSO_4$	1 mM
D-Glucose	20 mM
Ca^{2+}	1 mg/L
Fe^{2+}	1 mg/L
Trace elements (Cu^{2+}, Zn^{2+}, Mn^{2+}, MoO_4^{2-})	10 ng/L
Thiamine	10 mg/L
Biotin	10 mg/L
Canonical amino acids (optional)	50 mg/L
NMM-Agar	
2x NMM + 3.0 % H_2O agar-agar	30 g/L

Supplements

Ampicillin (Amp100)	100 µg/mL
Kanamycin (Kan50)	50 µg/mL
Chloramphenicol (Cm37)	37 µg/L
IPTG	0.1 mM – 1 mM
Arabinose (Ara)	0.002 % - 0.2 % (w/v)

6.6 Buffers and Solutions

Table 17 | Buffers and Solutions used in this study.

	composition
Agarose gel electrophoresis	
50x TAE buffer	2 M Tris
	2 M acetic acid
	10 % (v/v) 0.5 EDTA, pH 8.0
6x DNA loading dye	
Polyacrylamide gel electrophoresis	
Resolving gel	380 mM Tris-HCl, pH 8.8
	15 % acrylamide/bis-acrylamide (37.5:1)
	0.1 % SDS
	0.05 % APS
	0.05 % TEMED
Stacking gel	125 mM Tris-HCl, pH 6.8
	5 % acrylamide/bis-acrylamide (37.5:1)
	0.1 % SDS
	0.05 % APS
	0.05 % TEMED
SDS running buffer	190 mM glycine
	25 mM Tris
	3.5 mM SDS

5x SDS loading dye	80 mM Tris-HCl, pH 8.8
	10 % SDS
	12.5 % glycerol
	4 % (v/v) ß-mercaptoethanol
	0.2 % (v/v) bromophenol blue
Coomassie staining solution	1 g Coomassie Brilliant Blue R-250
	500 mL ethanol
	100 mL acetic acid
	ad 1 L dH$_2$O

Protein purification (IMAC)

Na-P	0.5 M NaH$_2$PO$_4$ · 2 H$_2$O
	0.5 M Na$_2$HPO$_4$ · 7 H$_2$O
Lysis/Binding buffer	50 mM Na-P, pH 7.8
	300 mM NaCl
Wash Buffer	50 mM Na-P, pH 7.8
	300 mM NaCl
	20 – 40 mM Imidazole
Elution Buffer	50 mM Na-P, pH 7.8
	300 mM NaCl
	500 mM Imidazole
Dialysis Buffer	50 mM Na-P, pH 7.8
	100 mM NaCl
	10 % glycerol
Stripping Buffer	20 mM Na-P, pH 7.4
	0.5 M NaCl
	50 mM EDTA
TEV protease reaction buffer	50 mM TRIS-HCl pH 8.0
	150 mM NaCl
	1 mM EDTA
	5 mM DTT

	10 % w/v glycerol

Size exclusion chromatography

SEC buffer	150 mM NaCl
	25 mM $K_xH_xPO_4$ (8.3 mM KH_2PO_4, 16.3 mM K_2HPO_4)

6.7 Databases, web applications and software

Table 18 | Databases, web applications and software used in this study.

Database	
EcoCyc[131]	https://ecocyc.org/
UniProt[422]	https://www.uniprot.org/
RCSB PDB (Protein Data Bank)[185]	https://www.rcsb.org/

Web applications	
MetaboAnalyst[404,405]	https://www.metaboanalyst.ca/
Interactive Pathways Explorer, iPath 3.0[406]	https://pathways.embl.de/
mutfunc[429]	http://mutfunc.com/
biorender	https://www.biorender.com/
NEB Tm Calculator v1.15.0	https://tmcalculator.neb.com/#!/main
NEBridge Golden Gate assembly tool v2.8.3	https://goldengate.neb.com/#!/

Software
Geneious Prime 2023.2.1
ChemDraw Professional 16.0
MassHunter (Agilent)
Chimera[43]
Cytoscape[424]
ClueGO[425]
Microsoft Office 365
Mendeley

6.8 Technical equipment and miscellaneous

Table 19 | Technical equipment and miscellaneous used in this study.

Equipment	
Incubators and shakers	
Ectron	Infors HT (Bottmingen, Schweiz)
Multitron	Infors HT (Bottmingen, Schweiz)
Incubator series B, KB	Binder (Tuttlingen)
Centrifuges	
Centrifuge 5810 R	Eppendorf AG (Hamburg, Deutschland)
Centrifuge 5418 R	Eppendorf AG (Hamburg, Deutschland)
MiniSpin® plus	Eppendorf AG (Hamburg, Deutschland)
Universal 32 R	Hettich Lab Technology (Tuttlingen, Germany)
Heraeus Fresco 17	Thermo Scientific (Waltham, MA, USA)
Thermomixer	
Thermomixer compact	Eppendorf AG (Hamburg, Deutschland)
Thermomixer 5437	Eppendorf AG (Hamburg, Deutschland)
Mixing Block MB-102	Bioer Technology (Binjiang, China)
Thermocycler	
Mastercycler Gradient	Eppendorf AG (Hamburg, Deutschland)
Peqstar 2x Gradient	Peqlab (Erlangen, Deutschland)
Gel electrophoresis	
Horizontal Gel System	Workshop Max Planck Institute of Biochemistry (Martinsried, Germany)
Gel documentation Felix 2050	Biostep (Jahnsdorf, Germany)
Scanner ViewPix 700	Biostep (Jahnsdorf, Germany)
UV-VIS spectroscopy	
Microplate reader Infinite M200	Tecan Group AG (Männedorf, Switzerland)
BioPhotometer plus	Eppendorf AG (Hamburg)
Ultrospec 6300 pro	GE Healthcare Life Sciences (Little Chalfont, UK)

Balances

TE 1502S	Sartorius (Göttingen, Deutschland)
GR-120	A&D (San Jose, CA, USA)
Mettler TE 3600 Deltarange	Mettler Toledo (Gießen, Deutschland)

Miscellaneous

Sonopuls HD 3200	Bandelin (Berlin)
Sonotrodes MS72, KE76	Bandelin (Berlin)
S20 SevenEasy™ pH	Mettler Toledo (Gießen, Germany)
MicroPulser™	Bio-Rad Laboratories GmbH (Hercules, USA)
Magnetic Stirrer/Hot Plate IKA combimag RET	IKA® (Staufen, Germany)
Scotsman AF80 Ice Flaker	Scotsman (Vernon Hills, USA)
Microwave KOR-6305	Daewoo (Butzbach)
Power Supply Unit Power Pack P25 T	Biometra (Jena, Germany)
Power Supply Unit Consort EV 261 and E 143	Sigma-Aldrich (St. Louis, USA)
Microsep™ Advance (MWCO 3kDa)	Pall Laboratory (Dreieich, Germany)
Microsep™ Advance (MWCO 10kDa)	Pall Laboratory (Dreieich, Germany)
Dialysis membrane ZelluTrans, MWCO 3500	Carl Roth (Karlsruhe, Germany)
PD-10 desalting columns (Illustra)	GE Healthcare Life Sciences (München, Germany)
Protino Ni-NTA column, 1 mL	Machery-Nagel
HisTrap FF, 5 mL	GE Healthcare Life Sciences (München, Germany)

7   Bibliography

[1]   Lenski, R. E. Evolution in Action: A 50, 000-Generation Salute to Charles Darwin. *Microbe* **6,** 30–33 (2011).

[2]   Darwin, C. *On the Origin of Species by Means of Natural Selection, Or, The Preservation of Favoured Races in the Struggle for Life*. (London: John Murray, Albemarle Street, 1859).

[3]   Hass, J. W. The Reverend Dr William Henry Dallinger, F.R.S. (1839-1909). *Notes Rec. R. Soc. Lond.* **54,** 53–65 (2000).

[4]   Jaccaud, M., Faron, R., Devilliers, D., Romano, R., Riedel, S. and Pernice, H. in *Ullmann's Encycl. Ind. Chem.* 1–19 (2020). doi:10.1002/14356007.a11

[5]   Schmedt auf der Günne, J., Mangstl, M. and Kraus, F. Occurrence of Difluorine F2 in Nature — In Situ Proof and Quantification by NMR Spectroscopy. *Angew. Chem. Int. Ed.* **51,** 1–4 (2012).

[6]   Gregson, R. P., Baldo, B. A., Thomas, P. G., Quinn, R. J., Bergquist, P. R., Stephens, J. F. and Horne, A. R. Fluorine is a major constituent of the marine sponge Halichondria moorei. *Science (80-.).* **206,** 1108–1109 (1979).

[7]   Peng, C., Xu, X., Ren, Y., Niu, H., Yang, Y., Hou, R., Wan, X. and Cai, H. Fluoride Absorption, Transportation and Tolerance Mechanism in Camellia sinensis, and Its Bioavailability and Health Risk Assessment: a Systematic Review. *J. Sci. Food Agric.* **101,** 379–387 (2021).

[8]   Harper, D. B. and O'Hagan, D. The Fluorinated Natural Products. *Nat. Prod. Rep.* **11,** 123–133 (1994).

[9]   Carvalho, M. F. and Oliveira, R. S. Natural production of fluorinated compounds and biotechnological prospects of the fluorinase enzyme. *Crit. Rev. Biotechnol.* **37,** 880–897 (2017).

[10]   Marais, J. S. C. The Isolation of the Toxic Principle 'Potassium Cymonate' from 'Gifblaar' Dichapetalum cymosum. *Onderstepoort J. Vet. Sci. Anim. Ind.* **18,** 203–206 (1943).

[11]   Marais, J. S. C. Monofluoroacetic Acid, the Toxic Principle of 'Gifblaar' Dichapetalum cymosum. *Onderstepoort J. Verterinar y Sci. Anim. lnrl11sf1 y* **20,** 67–73 (1944).

[12]   Proudfoot, A. T., Bradberry, S. M. and Vale, J. A. Sodium fluoroacetate poisoning. *Toxicol. Rev.* **25,** 213–219 (2006).

[13]   Lauble, A., Kennedy, M. C., Emptage, M. H., Beinert, H. and Stout, C. D. The reaction of fluorocitrate with aconitase and the crystal structure of the enzyme-inhibitor complex. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 13699–13703 (1996).

[14]   Peters, R. A. Croonian Lecture - Lethal synthesis. *Proc. R. Soc. London. Ser. B - Biol. Sci.* **139,** 143–170 (1952).

[15]   Leong, L. E. X., Khan, S., Davis, C. K., Denman, S. E. and McSweeney, C. S. Fluoroacetate in plants - a review of its distribution, toxicity to livestock and microbial detoxification. *J. Anim. Sci. Biotechnol.* **8,** 1–11 (2017).

[16]   Hamilton, J. T. G. and Harper, D. B. Fluoro fatty acids in seed oil of Dichapetalum toxicarium. *Phytochemistry* **44,** 1129–1132 (1997).

[17]   Harper, D. B., O'Hagan, D. and Murphy, C. D. in *Gribble, G. Nat. Prod. organohalogen Compd. Handb. Environ. Chem. Springer, Berlin* **3,** 141–169 (2003).

[18]   Stent, S. M. Giftblaar (Dichapetalum cymosum). *Agric. J. small-holder south africa* **3,** 70–73 (1916).

[19]   Newman, H. Streptomyces colony. *https://www.microbiologyinpictures.com/bacterial%20colonies.html* at <https://www.microbiologyinpictures.com/bacterial colonies.html>

[20]   Thomas, S. O., Singleton, V. L., Lowery, J. A., Sharpe, R. W., Pruess, L. M., Porter, J. N., Mowat, J. H. and Bohonos, N. Nucleocidin, a new antibiotic with activity against Trypanosomes. *Antibiot Annu* 716–721 (1956).

[21]   Morton, G. O., Lancaster, J. E., Van Lear, G. E., Fulmor, W. and Meyer, W. E. The structure of nucleocidin. III (a new structure). *J. Am. Chem. Soc.* **91,** 1535–1537 (1969).

[22]   Florini, J. R., Bird, H. H. and Bell, P. H. Inhibition of protein synthesis in vitro and in vivo by nucleocidin, an antitrypanosomal antibiotic. *J. Biol. Chem.* **241,** 1091–1098 (1966).

[23]   Lowe, P. T. and O'Hagan, D. 4′-Fluoro-nucleosides and nucleotides: from nucleocidin to an emerging class of therapeutics. *Chem. Soc. Rev.* **52,** 248–276 (2023).

[24]   Feng, X., Bello, D., Lowe, P. T., Clark, J. and O'Hagan, D. Two 3′-O-β-glucosylated nucleoside fluorometabolites related to nucleocidin in Streptomyces calvus. *Chem. Sci.* **10,** 9501–9505 (2019).

[25]   Ngivprom, U., Kluaiphanngam, S., Ji, W., Siriwibool, S., Kamkaew, A., Ketudat Cairns, J. R., Zhang, Q. and Lai, R.-Y. Characterization of NucPNP and NucV involved in the early steps of nucleocidin biosynthesis in Streptomyces calvus. *RSC Adv.* **11,** 3510–3515 (2021).

[26]   Sanada, M., Miyano, T. and Iwadare, S. Biosynthesis of fluorothreonine and fluoroacetic acid by the thienamycin producer, Streptomyces cattleya. *J. Antibiot. (Tokyo).* **39,** 259–265 (1986).

[27]   Deng, H., O'Hagan, D. and Schaffrath, C. Fluorometabolite biosynthesis and the fluorinase from Streptomyces cattleya. *Nat. Prod. Rep.* **21,** 773–784 (2004).

[28]   Cobb, S. L., Deng, H., Hamilton, J. T. G., McGlinchey, R. P. and O'Hagan, D. Identification of 5-

fluoro-5-deoxy-D-ribose-1-phosphate as an intermediate in fluorometabolite biosynthesis in Streptomyces cattleya. *Chem. Commun.* **4,** 592–593 (2004).

[29]   Onega, M., McGlinchey, R. P., Deng, H., Hamilton, J. T. G. and O'Hagan, D. The identification of (3R,4S)-5-fluoro-5-deoxy-d-ribulose-1-phosphate as an intermediate in fluorometabolite biosynthesis in Streptomyces cattleya. *Bioorg. Chem.* **35,** 375–385 (2007).

[30]   Moss, S. J., Murphy, C. D., Hamilton, J. T. G., McRobert, W. C., O'Hagan, D., Schaffrath, C. and Harper, D. B. Fluoroacetaldehyde: A precursor of both fluoroacetate and 4-fluorothreonine in Streptomyces cattleya. *Chem. Commun.* 2281–2282 (2000). doi:10.1039/b007261n

[31]   Murphy, C. D., Moss, S. J. and O'Hagan, D. Isolation of an Aldehyde Dehydrogenase Involved in the Oxidation of Fluoroacetaldehyde to Fluoroacetate in Streptomyces cattleya. *Appl. Environ. Microbiol.* **67,** 4919–4921 (2001).

[32]   Murphy, C. D., OHagan, D. and Schaffrath, C. Identification of a PLP-dependent Threonine Transaldolase: A Novel Enzyme Involved in 4-Fluorothreonine Biosynthesis in Streptomyces cattleya. *Angew. Chem. Int. Ed.* **40,** 4479–4481 (2001).

[33]   Walker, M. C., Wen, M., Weeks, A. M. and Chang, M. C. Y. Temporal and fluoride control of secondary metabolism regulates cellular organofluorine biosynthesis. *ACS Chem. Biol.* **7,** 1576–1585 (2012).

[34]   Zechel, D. L., Reid, S. P., Nashiru, O., Mayer, C., Stoll, D., Jakeman, D. L., Warren, R. A. J. and Withers, S. G. Enzymatic Synthesis of Carbon-Fluorine Bonds. *J. Am. Chem. Soc.* **123,** 4350–4351 (2001).

[35]   O'Hagan, D. and Deng, H. Enzymatic fluorination and biotechnological developments of the fluorinase. *Chem. Rev.* **115,** 634–649 (2015).

[36]   O'Hagan, D., Schaffrath, C., Cobb, S. L., Hamilton, J. T. G. and Murphy, C. D. Biosynthesis of an organofluorine molecule. *Nature* **416,** 279–279 (2002).

[37]   Schaffrath, C., Cobb, S. L. and O'Hagan, D. Cell-free biosynthesis of Fluoroacetate and 4-Fluorothreonine in Streptomyces cattleya. *Angew. Chem. Int. Ed.* **41,** 3913–3915 (2002).

[38]   Schaffrath, C., Deng, H. and O'Hagan, D. Isolation and characterisation of 5′-fluorodeoxyadenosine synthase, a fluorination enzyme from Streptomyces cattleya. *FEBS Lett.* **547,** 111–114 (2003).

[39]   Zhu, X., Robinson, D. A., McEwan, A. R., O'Hagan, D. and Naismith, J. H. Mechanism of enzymatic fluorination in Streptomyces cattleya. *J. Am. Chem. Soc.* **129,** 14597–14604 (2007).

[40]   Dong, C., Deng, H., Dorward, M., Schaffrath, C., O'Hagan, D. and Naismith, J. H. Crystallization and X-ray diffraction of 5′-fluoro-5′- deoxyadenosine synthase, a fluorination enzyme from

Streptomyces cattleya. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **59,** 2292–2293 (2003).

[41] Dong, C., Huang, F., Deng, H., Schaffrath, C., Spencer, J. B., O'Hagan, D. and Naismith, J. H. Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature* **427,** 561–565 (2004).

[42] Cadicamo, C. D., Courtieu, J., Deng, H., Meddour, A. and O'Hagan, D. Enzymatic fluorination in Streptomyces cattleya takes place with an inversion of configuration consistent with an SN2 reaction mechanism. *ChemBioChem* **5,** 685–690 (2004).

[43] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

[44] Deng, H., Ma, L., Bandaranayaka, N., Qin, Z., Mann, G., Kyeremeh, K., Yu, Y., Shepherd, T., Naismith, J. H. and O'Hagan, D. Identification of Fluorinases from Streptomyces sp MA37, Norcardia brasiliensis, and Actinoplanes sp N902-109 by Genome Mining. *ChemBioChem* **15,** 364–368 (2014).

[45] Huang, S., Ma, L., Tong, M. H., Yu, Y., O'Hagan, D. and Deng, H. Fluoroacetate biosynthesis from the marine-derived bacterium Streptomyces xinghaiensis NRRL B-24674. *Org. Biomol. Chem.* **12,** 4828–4831 (2014).

[46] Sooklal, S. A., De Koning, C., Brady, D. and Rumbold, K. Identification and characterisation of a fluorinase from Actinopolyspora mzabensis. *Protein Expr. Purif.* **166,** 105508 (2020).

[47] Pardo, I., Bednar, D., Calero, P., Volke, D. C., Damborský, J. and Nikel, P. I. A Nonconventional Archaeal Fluorinase Identified by in Silico Mining for Enhanced Fluorine Biocatalysis. *ACS Catal.* **12,** 6570–6577 (2022).

[48] Feng, X., Cao, Y., Liu, W. and Xian, M. Identification of Two Novel Fluorinases From Amycolatopsis sp. CA-128772 and Methanosaeta sp. PtaU1.Bin055 and a Mutant With Improved Catalytic Efficiency With Native Substrate. *Front. Bioeng. Biotechnol.* **10,** 1–9 (2022).

[49] Wu, L., Maglangit, F. and Deng, H. Fluorine biocatalysis. *Curr. Opin. Chem. Biol.* **55,** 119–126 (2020).

[50] Cheng, X. and Ma, L. Enzymatic synthesis of fluorinated compounds. *Appl. Microbiol. Biotechnol.* **105,** 8033–8058 (2021).

[51] Eustáquio, A. S., O'Hagan, D. and Moore, B. S. Engineering fluorometabolite production: Fluorinase expression in salinispora tropica yields fluorosalinosporamide. *J. Nat. Prod.* **73,** 378–382 (2010).

[52] Markakis, K., Lowe, P. T., Davison-Gates, L., O'Hagan, D., Rosser, S. J. and Elfick, A. An Engineered E. coli Strain for Direct in Vivo Fluorination. *ChemBioChem* **21,** 1856–1860 (2020).

[53] Calero, P., Volke, D. C., Lowe, P. T., Gotfredsen, C. H., O'Hagan, D. and Nikel, P. I. A fluoride-responsive genetic circuit enables in vivo biofluorination in engineered Pseudomonas putida. *Nat. Commun.* **11,** 5045 (2020).

[54] Walker, M. C., Thuronyi, B. W., Charkoudian, L. K., Lowry, B., Khosla, C. and Chang, M. C. Y. Expanding the fluorine chemistry of living systems using engineered polyketide synthase pathways. *Science (80-.).* **341,** 1089–1094 (2013).

[55] Thuronyi, B. W., Privalsky, T. M. and Chang, M. C. Y. Engineered Fluorine Metabolism and Fluoropolymer Production in Living Cells. *Angew. Chemie - Int. Ed.* **56,** 13637–13640 (2017).

[56] Dall'Angelo, S., Zhang, Q., Fleming, I. N., Piras, M., Schweiger, L. F., O'Hagan, D. and Zanda, M. Efficient bioconjugation of 5-fluoro-5-deoxy-ribose (FDR) to RGD peptides for positron emission tomography (PET) imaging of αvβ 3 integrin receptor. *Org. Biomol. Chem.* **11,** 4551–4558 (2013).

[57] Thompson, S., Onega, M., Ashworth, S., Fleming, I. N., Passchier, J. and O'Hagan, D. A two-step fluorinase enzyme mediated 18F labelling of an RGD peptide for positron emission tomography. *Chem. Commun.* **51,** 13542–13545 (2015).

[58] Sun, H., Yeo, W. L., Lim, Y. H., Chew, X., Smith, D. J., Xue, B., Chan, K. P., Robinson, R. C., Robins, E. G., Zhao, H. and Ang, E. L. Directed Evolution of a Fluorinase for Improved Fluorination Efficiency with a Non-native Substrate. *Angew. Chemie* **128,** 14489–14492 (2016).

[59] Sun, H., Zhao, H. and Ang, E. L. A coupled chlorinase-fluorinase system with a high efficiency of trans -halogenation and a shared substrate tolerance. *Chem. Commun.* **54,** 9458–9461 (2018).

[60] Pupo, G., Ibba, F., Ascough, D. M. H., Vicini, A. C., Ricci, P., Christensen, K. E., Pfeifer, L., Morphy, J. R., Brown, J. M., Paton, R. S. and Gouverneur, V. Asymmetric nucleophilic fluorination under hydrogen bonding phase-transfer catalysis. *Science (80-.).* **360,** 638–642 (2018).

[61] Littlechild, J. Haloperoxidases and their role in biotransformation reactions. *Curr. Opin. Chem. Biol.* **3,** 28–34 (1999).

[62] van Pée, K.-H. Biosynthesis of Halogenated Metabolites by Bacteria. *Annu. Rev. Microbiol.* **50,** 375–399 (1996).

[63] Höfler, G. T., But, A. and Hollmann, F. Haloperoxidases as catalysts in organic synthesis. *Org. Biomol. Chem.* **17,** 9267–9274 (2019).

[64] Lide, D. R. in *CRC Handb. Chem. Physics, Internet Version* 1253–1263 (2005). doi:10.1002/jctb.280500215

[65] Wang, L., Zhou, X., Fredimoses, M., Liao, S. and Liu, Y. Naturally occurring organoiodines. *RSC Adv.* **4,** 57350–57376 (2014).

[66]   Neumann, C. S., Fujimori, D. G. and Walsh, C. T. Halogenation Strategies In Natural Product Biosynthesis. *Chem. Biol.* **15,** 99–109 (2008).

[67]   Vaillancourt, F. H., Yeh, E., Vosburg, D. A., Garneau-Tsodikova, S. and Walsh, C. T. Nature's inventory of halogenation catalysts: Oxidative strategies predominate. *Chem. Rev.* **106,** 3364–3378 (2006).

[68]   Keller, S., Wage, T., Hohaus, K., Hölzer, M., Eichhorn, E. and van Pée, K.-H. Purification and Partial Characterization of Tryptophan 7-Halogenase (PrnA) from Pseudomonas fluorescens. *Angew. Chem. Int. Ed.* **39,** 2300–2302 (2000).

[69]   Frese, M. and Sewald, N. Enzymatic halogenation of tryptophan on a gram scale. *Angew. Chem. Int. Ed.* **54,** 298–301 (2015).

[70]   Schmartz, P. C., Zerbe, K., Abou-Hadeed, K. and Robinson, J. A. Bis-chlorination of a hexapeptide-PCP conjugate by the halogenase involved in vancomycin biosynthesis. *Org. Biomol. Chem.* **12,** 5574–5577 (2014).

[71]   Gribble, G. W. The diversity of naturally produced organohalogens. *Handb. Environ. Chem.* **3,** 1–15 (2003).

[72]   Gribble, G. W. The diversity of naturally produced organohalogens. *Chemosphere* **52,** 289–297 (2003).

[73]   Gribble, G. W. A recent survey of naturally occurring organohalogen compounds. *Environ. Chem.* **12,** 396–405 (2015).

[74]   Gribble, G. W. in *Kinghorn, A.D., Falk, H., Gibbons, S., Asakawa, Y., Liu, JK., Dirsch, V.M. Nat. Occur. Organohalogen Compd. Prog. Chem. Org. Nat. Prod. vol 121. Springer, Cham* 546 (2023). doi:org/10.1007/978-3-031-26629-4_1

[75]   Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A. and Prinsep, M. R. Marine natural products. *Nat. Prod. Rep.* **36,** 122–173 (2019).

[76]   Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A. and Prinsep, M. R. Marine natural products. *Nat. Prod. Rep.* **37,** 175–223 (2020).

[77]   Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A. and Prinsep, M. R. Marine natural products. *Nat. Prod. Rep.* **39,** 1122–1171 (2022).

[78]   Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A. and Prinsep, M. R. Marine natural products. *Nat. Prod. Rep.* **40,** 275–325 (2023).

[79]   Budisa, N., Kubyshkin, V. and Schulze-Makuch, D. Fluorine-Rich Planetary Environments as Possible Habitats for Life. *Life* **4,** 374–385 (2014).

[80]   Marsh, E. N. G. and Suzuki, Y. Using 19F NMR to probe biological interactions of proteins and peptides. *ACS Chem. Biol.* **9,** 1242–1250 (2014).

[81]   Berger, A. A., Völler, J. S., Budisa, N. and Koksch, B. Deciphering the Fluorine Code - The Many Hats Fluorine Wears in a Protein Environment. *Acc. Chem. Res.* **50,** 2093–2103 (2017).

[82]   Monkovic, J. M., Gibson, H., Sun, J. W. and Montclare, J. K. Fluorinated Protein and Peptide Materials for Biomedical Applications. *Pharmaceuticals* **15,** 1201 (2022).

[83]   O'Hagan, D. and Rzepa, H. S. Some influences of fluorine in bioorganic chemistry. *Chem. Commun.* 645–652 (1997). doi:10.1039/a604140j

[84]   Salwiczek, M., Nyakatura, E. K., Gerling, U. I. M., Ye, S. and Koksch, B. Fluorinated amino acids: Compatibility with native protein structures and effects on protein–protein interactions. *Chem. Soc. Rev.* **41,** 2135–2171 (2012).

[85]   Buer, B. C. and Marsh, E. N. G. Fluorine: A new element in protein design. *Protein Sci.* **21,** 453–462 (2012).

[86]   Biava, H. and Budisa, N. Evolution of fluorinated enzymes: An emerging trend for biocatalyst stabilization. *Eng. Life Sci.* **14,** 340–351 (2014).

[87]   Marsh, E. N. G. Towards the nonstick egg: Designing fluorous proteins. *Chem. Biol.* **7,** 153–157 (2000).

[88]   Merkel, L., Schauer, M., Antranikian, G. and Budisa, N. Parallel incorporation of different fluorinated amino acids: On the way to 'teflon' proteins. *ChemBioChem* **11,** 1505–1507 (2010).

[89]   Budisa, N., Pipitone, O., Siwanowicz, I., Rubini, M., Pal, P. P., Holak, T. A. and Gelmi, M. L. Efforts towards the design of 'Teflon' proteins: In vivo translation with trifluorinated leucine and methionine analogues. *Chem. Biodivers.* **1,** 1465–1475 (2004).

[90]   Pace, C. J. and Gao, J. Exploring and exploiting polar-π interactions with fluorinated aromatic amino acids. *Acc. Chem. Res.* **46,** 907–915 (2013).

[91]   Samanta, U., Pal, D. and Chakrabarti, P. Environment of tryptophan side chains in proteins. *Proteins Struct. Funct. Genet.* **38,** 288–300 (2000).

[92]   Mant, C. T., Kovacs, J. M., Kim, H. M., Pollock, D. D. and Hodges, R. S. Intrinsic amino acid side-chain hydrophilicity/hydrophobicity coefficients determined by reversed-phase high-performance liquid chromatography of model peptides: comparison with other hydrophilicity/hydrophobicity scales. *Biopolymers* **92,** 573–595 (2009).

[93]   Khemaissa, S., Sagan, S. and Walrant, A. Tryptophan, an Amino-Acid Endowed with Unique Properties and Its Many Roles in Membrane Proteins. *Crystals* **11,** 1032–1045 (2021).

[94] Asensio, J. L., Ardá, A., Canada, F. J. and Jiménez-Barbero, J. Carbohydrate-Aromatic Interactions. *Acc. Chem. Res.* **46,** 946–954 (2013).

[95] Burley, S. K. and Petsko, G. A. Aromatic-Aromatic Interaction: A Mechanism of Protein Structure Stabilization. *Science (80-.).* **229,** 23–28 (1985).

[96] Gallivan, J. P. and Dougherty, D. A. Cation-π interactions in structural biology. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 9459–9464 (1999).

[97] Dougherty, D. A. Cation-π interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp. *Science (80-.).* **271,** 163–168 (1996).

[98] Bienfait, B. and Ertl, P. JSME: A free molecule editor in JavaScript. *J. Cheminform.* **5,** 1–6 (2013).

[99] Akashi, H. and Gojobori, T. Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 3695–3700 (2002).

[100] Santiveri, C. M. and Jiménez, A. M. Tryptophan residues: scarce in proteins but strong stabilizers of β-hairpin peptides. *Biopolymers* **94,** 779–790 (2010).

[101] Lepthien, S., Wiltschi, B., Bolic, B. and Budisa, N. In vivo engineering of proteins with nitrogen-containing tryptophan analogs. *Appl. Microbiol. Biotechnol.* **73,** 740–754 (2006).

[102] Kanei-Ishii, C., Sarai, A., Sawazaki, T., Nakagoshi, H., He, D. N., Ogata, K., Nishimura, Y. and Ishii, S. The tryptophan cluster: A hypothetical structure of the DNA-binding domain of the myb protooncogene product. *J. Biol. Chem.* **265,** 19990–19995 (1990).

[103] Saikumar, P., Murali, R. and Reddy, E. P. Role of tryptophan repeats and flanking amino acids in Myb-DNA interactions. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 8452–8456 (1990).

[104] Luscombe, N. M., Laskowski, R. A. and Thornton, J. M. Amino acid-base interactions: A three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.* **29,** 2860–2874 (2001).

[105] Sinkeldam, R. W., Greco, N. J. and Tor, Y. Fluorescent Analogs of Biomolecular Building Blocks: Design, Properties, and Applications. *Chem. Rev.* **110,** 2579–2619 (2010).

[106] Hartman, H. and Smith, T. F. The evolution of the ribosome and the genetic code. *Life* **4,** 227–249 (2014).

[107] Fournier, G. P. and Alm, E. J. Ancestral Reconstruction of a Pre-LUCA Aminoacyl-tRNA Synthetase Ancestor Supports the Late Addition of Trp to the Genetic Code. *J. Mol. Evol.* **80,** 171–185 (2015).

[108] Barik, S. The uniqueness of tryptophan in biology: Properties, Metabolism, Interactions and Localization in Proteins. *Int. J. Mol. Sci.* **21,** 8776–8798 (2020).

[109]   Hu, J.-J., He, P.-Y. and Li, Y.-M. Chemical modifications of tryptophan residues in peptides and proteins. *J. Pept. Sci.* **27,** e3286 (2020).

[110]   Lajoie, M. J., Kosuri, S., Mosberg, J. A., Gregg, C. J., Zhang, D. and Church, G. M. Probing the limits of genetic recoding in essential genes. *Science (80-.).* **342,** 361–363 (2013).

[111]   Lee, J.-H., Wood, T. K. and Lee, J. Roles of indole as an interspecies and interkingdom signaling molecule. *Trends Microbiol.* **23,** 707–718 (2015).

[112]   Lee, J.-H. and Lee, J. Indole as an intercellular signal in microbial communities. *FEMS Microbiol. Rev.* **34,** 426–444 (2010).

[113]   Kim, J. and Park, W. Indole: a signaling molecule or a mere metabolic byproduct that alters bacterial physiology at a high concentration? *J. Microbiol.* **53,** 421–428 (2015).

[114]   Zarkan, A., Liu, J., Matuszewska, M., Gaimster, H. and Summers, D. K. Local and Universal Action: The Paradoxes of Indole Signalling in Bacteria. *Trends Microbiol.* **28,** 566–577 (2020).

[115]   Gaimster, H., Cama, J., Hernández-Ainsa, S., Keyser, U. F. and Summers, D. K. The indole pulse: A new perspective on indole signalling in Escherichia coli. *PLoS One* **9,** e93168 (2014).

[116]   Gaimster, H. and Summers, D. Regulation of indole signalling during the transition of E. coli from exponential to stationary phase. *PLoS One* **10,** e0136691 (2015).

[117]   Chimerel, C., Field, C. M., Piñero-Fernandez, S., Keyser, U. F. and Summers, D. K. Indole prevents Escherichia coli cell division by modulating membrane potential. *Biochim. Biophys. Acta 1818 - Biomembr.* **1818,** 1590–1594 (2012).

[118]   Zarkan, A., Caño-Muñiz, S., Zhu, J., Al Nahas, K., Cama, J., Keyser, U. F. and Summers, D. K. Indole Pulse Signalling Regulates the Cytoplasmic pH of E. coli in a Memory-Like Manner. *Sci. Rep.* **9,** 3868 (2019).

[119]   Chimerel, C., Murray, A. J., Oldewurtel, E. R., Summers, D. K. and Keyser, U. F. The effect of bacterial signal indole on the electrical properties of lipid membranes. *ChemPhysChem* **14,** 417–423 (2013).

[120]   Lee, J.-H., Kim, Y.-G., Gwon, G., Wood, T. K. and Lee, J. Halogenated indoles eradicate bacterial persister cells and biofilms. *AMB Express* **6,** 123 (2016).

[121]   Yanofsky, C., Platt, T., Crawford, I. P., Nichols, B. P., Christie, G. E., Horowitz, H., VanCleemput, M. and Wu, A. M. The complete nucleotide sequence of the tryptophan operon of Escherichia coli. *Nucleic Acids Res.* **9,** 6647–6668 (1981).

[122]   Gunsalus, R. P. and Yanofsky, C. Nucleotide sequence and expression of Escherichia coli trpR, the structural gene for the trp aporepressor. *Proc. Natl. Acad. Sci. U. S. A.* **77,** 7117–7121 (1980).

Bibliography

[123] Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. and Sigler, P. B. Crystal structure of trp represser/operator complex at atomic resolution. *Nature* **335,** 321–329 (1988).

[124] Somerville, R. The Trp Repressor , A Ligand-Activated Regulatory Protein. *Prog. Nucleic Acid Res. Mol. Biol.* **42,** 1–38 (1992).

[125] Yanofsky, C. The different roles of tryptophan transfer RNA in regulating trp operon expression in E. coli versus B. subtilis. *Trends Genet.* **20,** 367–374 (2004).

[126] Yanofsky, C. RNA-based regulation of genes of tryptophan synthesis and degradation, in bacteria. *RNA* **13,** 1141–1154 (2007).

[127] Turnbough, C. L. Regulation of Bacterial Gene Expression by Transcription Attenuation. *Microbiol. Mol. Biol. Rev.* **83,** e00019-19 (2019).

[128] Ito, J. and Yanofsky, C. Anthranilate Synthetase, an Enzyme Specified by the Tryptophan Operon of Escherichia coli: Comparative Studies on the Complex and the Subunits. *J. Bacteriol.* **97,** 725–733 (1969).

[129] Merino, E., Jensen, R. A. and Yanofsky, C. Evolution of bacterial trp operons and their regulation. *Curr. Opin. Microbiol.* **11,** 78–86 (2008).

[130] Pittard, J. and Yang, J. Biosynthesis of the Aromatic Amino Acids. *EcoSal Plus* **3,** 1–40 (2008).

[131] Keseler, I. M., Gama-Castro, S., Mackie, A., Billington, R., Bonavides-Martínez, C., Caspi, R., Kothari, A., Krummenacker, M., Midford, P. E., Muñiz-Rascado, L., Ong, W. K., Paley, S., Santos-Zavaleta, A., Subhraveti, P., Tierrafría, V. H., Wolfe, A. J., Collado-Vides, J., Paulsen, I. T. and Karp, P. D. The EcoCyc Database in 2021. *Front. Microbiol.* **12,** 1–10 (2021).

[132] Mir, R., Jallu, S. and Singh, T. P. The shikimate pathway: Review of amino acid sequence, function and three-dimensional structures of the enzymes. *Crit. Rev. Microbiol.* **41,** 172–189 (2015).

[133] Snell, E. E. Tryptophanase: Structure, Catalytic Activities, and Mechanism of Action. *Adv. Enzymol. Relat. Areas Mol. Biol.* **42,** 287–333 (1975).

[134] Gong, F. and Yanofsky, C. Analysis of tryptophanase operon expression in vitro. Accumulation of TnaC-peptidyl-tRNA in a release factor 2-depleted S-30 extract prevents Rho factor action, simulating induction. *J. Biol. Chem.* **277,** 17095–17100 (2002).

[135] Li, G. and Young, K. D. Indole production by the tryptophanase TnaA in escherichia coli is determined by the amount of exogenous tryptophan. *Microbiol. (United Kingdom)* **159,** 402–410 (2013).

[136] Yanofsky, C., Horn, V. and Gollnick, P. Physiological studies of tryptophan transport and

tryptophanase operon induction in Escherichia coli. *J. Bacteriol.* **173,** 6009–6017 (1991).

[137] Chant, E. L. and Summers, D. K. Indole signalling contributes to the stable maintenance of Escherichia coli multicopy plasmids. *Mol. Microbiol.* **63,** 35–43 (2007).

[138] Hütter, R., Niederberger, P. and DeMoss, J. A. Tryptophan Biosynthetic Genes in Eukaryotic Microorganisms. *Annu. Rev. Microbiol.* **40,** 55–77 (1986).

[139] Radwanski, E. R. and Last, R. L. Tryptophan Biosynthesis and Metabolism: Biochemical and Molecular Genetics. *Plant Cell* **7,** 921–934 (1995).

[140] Maeda, H. and Dudareva, N. The shikimate pathway and aromatic amino acid biosynthesis in plants. *Annu. Rev. Plant Biol.* **63,** 73–105 (2012).

[141] Crawford, I. P. and Yanofsky, C. On the Separation of the Tryptophan Synthetase of Escherichia coli Into Two Protein Components. *Proc. Natl. Acad. Sci.* **44,** 1161–1170 (1958).

[142] Miles, E. W. in *Adv. Enzymol. Relat. Areas Mol. Biol.* **49,** 127–186 (1979).

[143] Miles, E. W. Tryptophan synthase: A multienzyme complex with an intramolecular tunnel. *Chem. Rec.* **1,** 140–151 (2001).

[144] Raboni, S., Bettati, S. and Mozzarelli, A. Tryptophan synthase: A mine for enzymologists. *Cell. Mol. Life Sci.* **66,** 2391–2403 (2009).

[145] Miles, E. W. The tryptophan synthase α2β2 complex: A model for substrate channeling, allosteric communication, and pyridoxal phosphate catalysis. *J. Biol. Chem.* **288,** 10084–10091 (2013).

[146] Hyde, C. C., Ahmed, S. A., Padlan, E. A., Miles, E. W. and Davies, D. R. Three-dimensional Structure of the Tryptophan Synthase a2b2 Multienzyme Complex from Salmonella typhimurium. *J. Biol. Chem.* **263,** 17857–17871 (1988).

[147] Hilario, E., Caulkins, B. G., Huang, Y.-M. M., You, W., Chang, C.-E. A., Mueller, L. J., Dunn, M. F. and Fan, L. Visualizing the tunnel in tryptophan synthase with crystallography: Insights into a selective filter for accommodating indole and rejecting water. *Biochim. Biophys. Acta - Proteins Proteomics* **1864,** 268–279 (2016).

[148] Ngo, H., Kimmich, N., Harris, R., Niks, D., Blumenstein, L., Kulik, V., Barends, T. R., Schlichting, I. and Dunn, M. F. Allosteric regulation of substrate channeling in tryptophan synthase: modulation of the L-serine reaction in stage I of the β-reaction by α-site ligands. *Biochemistry* **46,** 7740–7753 (2007).

[149] Dunn, M. F., Niks, D., Ngo, H., Barends, T. R. M. and Schlichting, I. Tryptophan synthase: the workings of a channeling nanomachine. *Trends Biochem. Sci.* **33,** 254–264 (2008).

[150] Dunn, M. F. Allosteric regulation of substrate channeling and catalysis in the tryptophan

synthase bienzyme complex. *Arch. Biochem. Biophys.* **519,** 154–166 (2012).

[151] Ghosh, R. K., Hilario, E., Chang, C. en A., Mueller, L. J. and Dunn, M. F. Allosteric regulation of substrate channeling: Salmonella typhimurium tryptophan synthase. *Front. Mol. Biosci.* **9,** 1–23 (2022).

[152] Dierkers, A. T., Niks, D., Schlichting, I. and Dunn, M. F. Tryptophan synthase: structure and function of the monovalent cation site. *Biochemistry* **48,** 10997–11010 (2009).

[153] Kneuttinger, A. C., Zwisele, S., Straub, K., Bruckmann, A., Busch, F., Kinateder, T., Gaim, B., Wysocki, V. H., Merkl, R. and Sterner, R. Light-regulation of tryptophan synthase by combining protein design and enzymology. *Int. J. Mol. Sci.* **20,** 1–24 (2019).

[154] Rhee, S., Parris, K. D., Ahmed, S. A., Miles, E. W. and Davies, D. R. Exchange of K+ or Cs+ for Na+ induces local and long-range changes in the three-dimensional structure of the tryptophan synthase α2β2 complex. *Biochemistry* **35,** 4211–4221 (1996).

[155] Buller, A. R., Brinkmann-Chen, S., Romney, D. K., Herger, M., Murciano-Calles, J. and Arnold, F. H. Directed evolution of the tryptophan synthase β-subunit for stand-alone function recapitulates allosteric activation. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 14599–14604 (2015).

[156] Watkins-Dulaney, E., Straathof, S. and Arnold, F. Tryptophan Synthase: Biocatalyst Extraordinaire. *ChemBioChem* **22,** 5–16 (2021).

[157] Buller, A. R., van Roye, P., Cahn, J. K. B., Scheele, R. A., Herger, M. and Arnold, F. H. Directed Evolution Mimics Allosteric Activation by Stepwise Tuning of the Conformational Ensemble. *J. Am. Chem. Soc.* **140,** 7256–7266 (2018).

[158] Murciano-Calles, J., Romney, D. K., Brinkmann-Chen, S., Buller, A. R. and Arnold, F. H. A Panel of TrpB Biocatalysts Derived from Tryptophan Synthase through the Transfer of Mutations that Mimic Allosteric Activation. *Angew. Chemie* **128,** 11749–11753 (2016).

[159] Romney, D. K., Murciano-Calles, J., Wehrmüller, J. E. and Arnold, F. H. Unlocking Reactivity of TrpB: A General Biocatalytic Platform for Synthesis of Tryptophan Analogues. *J. Am. Chem. Soc.* **139,** 10769–10776 (2017).

[160] Boville, C. E., Romney, D. K., Almhjell, P. J., Sieben, M. and Arnold, F. H. Improved Synthesis of 4-Cyanotryptophan and Other Tryptophan Analogues in Aqueous Solvent Using Variants of TrpB from Thermotoga maritima. *J. Org. Chem.* **83,** 7447–7452 (2018).

[161] Herger, M., van Roye, P., Romney, D. K., Brinkmann-Chen, S., Buller, A. R. and Arnold, F. H. Synthesis of β-Branched Tryptophan Analogues Using an Engineered Subunit of Tryptophan Synthase. *J. Am. Chem. Soc.* **138,** 8388–8391 (2016).

[162] Francis, D., Winn, M., Latham, J., Greaney, M. F. and Micklefield, J. An Engineered Tryptophan

Synthase Opens New Enzymatic Pathways to β-Methyltryptophan and Derivatives. *ChemBioChem* **18,** 382–386 (2017).

[163] Boville, C. E., Scheele, R. A., Koch, P., Brinkmann-Chen, S., Buller, A. R. and Arnold, F. H. Engineered Biosynthesis of β-Alkyl Tryptophan Analogues. *Angew Chem Int Ed* **57,** 14764–14768 (2018).

[164] Dick, M., Sarai, N. S., Martynowycz, M. W., Gonen, T. and Arnold, F. H. Tailoring Tryptophan Synthase TrpB for Selective Quaternary Carbon Bond Formation. *J. Am. Chem. Soc.* **141,** 19817–19822 (2019).

[165] Phillips, R. S. Synthetic applications of tryptophan synthase. *Tetrahedron Asymmetry* **15,** 2787–2792 (2004).

[166] Yanofsky, C. in *Methods Enzymol.* **2,** 233–238 (1955).

[167] Hall, A. N., Lea, D. J. and Rydon, H. N. The behaviour of the Bz-Methylindoles as substrates and inhibitors for Neurospora crassa tryptophan synthase. *Biochem. J.* **84,** 12–16 (1962).

[168] Wilcox, M. The enzymatic synthesis of L-tryptophan analogues. *Anal. Biochem.* **59,** 436–440 (1974).

[169] Saito, A. and Rilling, H. C. Synthesis of Photoreactive Tryptophan Analogs: Enzymatic Conversion of Azidoindoles to Azidotryptophans. *Prep. Biochem.* **11,** 535–546 (1981).

[170] Sloan, M. J. and Phillips, R. S. Enzymatic synthesis of aza-L-tryptophans: The preparation of 5- and 6-Aza-L-tryptophan. *Bioorganic Med. Chem. Lett.* **2,** 1053–1056 (1992).

[171] Goss, R. J. M. and Newill, P. L. A. A convenient enzymatic synthesis of L-halotryptophans. *Chem. Commun.* 4924–4925 (2006). doi:10.1039/b611929h

[172] Winn, M., Roy, A. D., Grüschow, S., Parameswaran, R. S. and Goss, R. J. M. A convenient one-step synthesis of L-aminotryptophans and improved synthesis of 5-fluorotryptophan. *Bioorganic Med. Chem. Lett.* **18,** 4508–4510 (2008).

[173] Smith, D. R. M., Willemse, T., Gkotsi, D. S., Schepens, W., Maes, B. U. W., Ballet, S. and Goss, R. J. M. The First One-Pot Synthesis of L-7-Iodotryptophan from 7-Iodoindole and Serine, and an Improved Synthesis of Other L-7-Halotryptophans. *Org. Biomol. Chem.* **16,** 2622–2625 (2014).

[174] Lee, M. and Phillips, R. S. Enzymatic synthesis of chloro-L-tryptophans. *Bioorganic Med. Chem. Lett.* **2,** 1563–1564 (1992).

[175] Perni, S., Hackett, L., Goss, R. J. M., Simmons, M. J. and Overton, T. W. Optimisation of engineered Escherichia coli biofilms for enzymatic biosynthesis of L-halotryptophans. *AMB Express* **3,** 1–10 (2013).

Bibliography

[176] Phillips, R. S., Cohen, L. A., Annby, U., Wensbo, D. and Gronowitz, S. Enzymatic synthesis of Thia-L-tryptophans. *Bioorganic Med. Chem. Lett.* **5,** 1133–1134 (1995).

[177] Welch, M. and Phillips, R. S. Enzymatic syntheses of 6-(4H-selenolo[3,2-b]pyrrolyl)-L-alanine, 4-(6H- selenolo[2,3-b]pyrrolyl)-L-alanine, and 6-(4H-furo[3,2-b]pyrrolyl)-L-alanine. *Bioorganic Med. Chem. Lett.* **9,** 637–640 (1999).

[178] Boles, J. O., Henderson, J., Hatch, D. and Silks, L. A. 'Pete'. Synthesis and incorporation of [6,7]-selenatryptophan into dihydrofolate reductase. *Biochem. Biophys. Res. Commun.* **298,** 257–261 (2002).

[179] Schlosser, M., Ginanneschi, A. and Leroux, F. In search of simplicity and flexibility: A rational access to twelve fluoroindolecarboxylic acids. *European J. Org. Chem.* 2956–2969 (2006). doi:10.1002/ejoc.200600118

[180] Nosova, E. V., Lipunova, G. N., Charushin, V. N. and Chupakhin, O. N. Fluorine-containing indoles: Synthesis and biological activity. *J. Fluor. Chem.* **212,** 51–106 (2018).

[181] Blaser, G., Sanderson, J. M., Batsanov, A. S. and Howard, J. A. K. The facile synthesis of a series of tryptophan derivatives. *Tetrahedron Lett.* **49,** 2795–2798 (2008).

[182] Ibba, M. and Dieter, S. Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* **69,** 617–650 (2000).

[183] Giegé, R. and Springer, M. Aminoacyl-tRNA Synthetases in the Bacterial World. *EcoSal Plus* **7,** 1–83 (2016).

[184] Gomez, M. A. R. and Ibba, M. Aminoacyl-tRNA synthetases. *RNA* **26,** 910–936 (2020).

[185] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242 (2000).

[186] Maltseva, N., Kim, Y., Mulligan, R., Grimshaw, S. G., Joachimiak, A. and Anderson, W. F. Crystal Structure of Tryptophanyl-tRNA Synthetase from Escherichia coli Complexed with AMP and Tryptophan. doi:10.2210/PDB5V0I/PDB

[187] Soutourina, J., Plateau, P. and Blanquet, S. Metabolism of D-aminoacyl-tRNAs in Escherichia coli and Saccharomyces cerevisiae cells. *J. Biol. Chem.* **275,** 32535–32542 (2000).

[188] Azim, M. K. and Budisa, N. Docking of tryptophan analogs to trytophanyl-tRNA synthetase: Implications for non-canonical amino acid incorporations. *Biol. Chem.* **389,** 1173–1182 (2008).

[189] Budisa, N., Pal, P. P., Alefelder, S., Birle, P., Krywcun, T., Rubini, M., Wenger, W., Bae, J. H. and Steiner, T. Probing the role of tryptophans in Aequorea victoria green fluorescent proteins with an expanded genetic code. *Biol. Chem.* **385,** 191–202 (2004).

[190] Budisa, N. and Pal, P. P. Designing novel spectral classes of proteins with a tryptophan-

expanded genetic code. *Biol. Chem.* **385,** 893–904 (2004).

[191] Nevinsky, G. A., Favorova, O. O., Lavrik, O. I., Petrova, T. D., Kochkina, L. L. and Savchenko, T. I. Fluorinated tryptophans as substrates and inhibitors of the ATP-[32P]PPi exchange reaction catalyzed by tryptophanyl tRNA synthetase. *FEBS Lett.* **43,** 135–138 (1974).

[192] Fan, C., Ho, J. M. L., Chirathivat, N., Söll, D. and Wang, Y.-S. Exploring the substrate range of wild-type aminoacyl-tRNA synthetases. *ChemBioChem* **15,** 1805–1809 (2014).

[193] Crick, F. Central Dogma of Molecular Biology. *Nature* **227,** 561–563 (1970).

[194] Walsh, C. T., Garneau-Tsodikova, S. and Gatto, G. J. Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew. Chemie - Int. Ed.* **44,** 7342–7372 (2005).

[195] Acevedo-Rocha, C. G. and Budisa, N. Xenomicrobiology: a roadmap for genetic code engineering. *Microb. Biotechnol.* **9,** 666–676 (2016).

[196] Kubyshkin, V. and Budisa, N. Synthetic alienation of microbial organisms by using genetic code engineering: Why and how? *Biotechnol. J.* **12,** 1600097 (2017).

[197] Budisa, N., Kubyshkin, V. and Schmidt, M. Xenobiology: A Journey towards Parallel Life Forms. *ChemBioChem* **21,** 2228–2231 (2020).

[198] Gerecht, K., Freund, N., Liu, W., Liu, Y., Fürst, M. J. L. J. and Holliger, P. The Expanded Central Dogma: Genome Resynthesis, Orthogonal Biosystems, Synthetic Genetics. *Annu. Rev. Biophys.* **52,** 413–432 (2023).

[199] de la Torre, D. and Chin, J. W. Reprogramming the genetic code. *Nat. Rev. Genet.* **22,** 169–184 (2021).

[200] Pinheiro, V. B. and Holliger, P. The XNA world: Progress towards replication and evolution of synthetic genetic polymers. *Curr. Opin. Chem. Biol.* **16,** 245–252 (2012).

[201] Maurya, R., Gohil, N., Bhattacharjee, G., Lam, N. L., Alzahrani, K. J. and Singh, V. *Recent development and applications of xeno nucleic acids*. *New Front. Appl. Synth. Biol.* (INC, 2022). doi:10.1016/B978-0-12-824469-2.00019-1

[202] Houlihan, G., Arangundy-Franklin, S. and Holliger, P. Exploring the Chemistry of Genetic Information Storage and Propagation through Polymerase Engineering. *Acc. Chem. Res.* **50,** 1079–1087 (2017).

[203] Sun, L., Ma, X., Zhang, B., Qin, Y., Ma, J., Du, Y. and Chen, T. From polymerase engineering to semi-synthetic life: artificial expansion of the central dogma. *RSC Chem. Biol.* **3,** 1173–1197 (2022).

[204] Pinheiro, V. B., Taylor, A. I., Cozens, C., Abramov, M., Renders, M., Zhang, S., Chaput, J. C., Wengel, J., Peak-Chew, S.-Y., McLaughlin, S. H., Herdewijn, P. and Holliger, P. Synthetic genetic

polymers capable of heredity and evolution. *Science (80-.).* **336,** 341–344 (2012).

[205] Nikoomanzar, A., Chim, N., Yik, E. J. and Chaput, J. C. Engineering polymerases for applications in synthetic biology. *Q. Rev. Biophys.* **53,** e8 (2020).

[206] Medina, E., Yik, E. J., Herdewijn, P. and Chaput, J. C. Functional Comparison of Laboratory-Evolved XNA Polymerases for Synthetic Biology. *ACS Synth. Biol.* **10,** 1429–1437 (2021).

[207] Herdewijn, P. and Marliere, P. Toward Safe Genetically Modified Organisms through the Chemical Diversification of Nucleic Acids. *Chem. Biodivers.* **6,** 791–808 (2011).

[208] Marlière, P., Patrouix, J., Döring, V., Herdewijn, P., Tricot, S., Cruveiller, S., Bouzon, M. and Mutzel, R. Chemical evolution of a bacterium's genome. *Angew. Chemie - Int. Ed.* **50,** 7109–7114 (2011).

[209] Berg, P., Baltimore, D., Brenner, S., Roblin III, R. O. and Singer, M. F. Asilomar conference on recombinant DNA molecules. *Science (80-.).* **188,** 991–994 (1975).

[210] Berg, P. Reflections on Asilomar 2 at Asilomar 3: Twenty-five years later. *Perspect. Biol. Med.* **44,** 183–185 (2001).

[211] Taleb, N. N., Read, R., Douady, R., Norman, J. and Bar-Yam, Y. The Precautionary Principle (with Application to the Genetic Modification of Organisms). *arXiv Prepr. arXiv 1410.5787 2014.* (2014). at <http://arxiv.org/abs/1410.5787>

[212] Schmidt, M. and de Lorenzo, V. Synthetic bugs on the loose: Containment options for deeply engineered (micro)organisms. *Curr. Opin. Biotechnol.* **38,** 90–96 (2016).

[213] Torres, L., Krüger, A., Csibra, E., Gianni, E. and Pinheiro, V. B. Synthetic biology approaches to biological containment: pre-emptively tackling potential risks. *Essays Biochem.* **60,** 393–410 (2016).

[214] Lee, J. W., Chan, C. T. Y., Slomovic, S. and Collins, J. J. Next-generation biocontainment systems for engineered organisms. *Nat. Chem. Biol.* **14,** 530–537 (2018).

[215] Kim, D. and Lee, J. W. Genetic Biocontainment Systems for the Safe Use of Engineered Microorganisms. *Biotechnol. Bioprocess Eng.* **25,** 974–984 (2020).

[216] Mandell, D. J., Lajoie, M. J., Mee, M. T., Takeuchi, R., Kuznetsov, G., Norville, J. E., Gregg, C. J., Stoddard, B. L. and Church, G. M. Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518,** 55–60 (2015).

[217] Rovner, A. J., Haimovich, A. D., Katz, S. R., Li, Z., Grome, M. W., Gassaway, B. M., Amiram, M., Patel, J. R., Gallagher, R. R., Rinehart, J. and Isaacs, F. J. Recoded organisms engineered to depend on synthetic amino acids. *Nature* **518,** 89–93 (2015).

[218] Yoo, J. I., Seppälä, S. and O'Malley, M. A. Engineered fluoride sensitivity enables biocontainment

and selection of genetically-modified yeasts. *Nat. Commun.* **11,** 5459 (2020).

[219] Marliere, P. The farther, the safer: A manifesto for securely navigating synthetic species away from the old living world. *Syst. Synth. Biol.* **3,** 77–84 (2009).

[220] Oehm, S. Adaptation of E. coli towards Tryptophan analogue usage. PhD Thesis, Technische Universität Berlin. (2016).

[221] Ausländer, S., Ausländer, D. and Fussenegger, M. Synthetic Biology — The Synthesis of Biology. *Angew. Chemie - Int. Ed.* **56,** 6396–6419 (2017).

[222] Xu, X., Meier, F., Blount, B. A., Pretorius, I. S., Ellis, T., Paulsen, I. T. and Williams, T. C. Trimming the genomic fat: minimising and re-functionalising genomes using synthetic biology. *Nat. Commun.* **14,** 1984 (2023).

[223] Mizoguchi, H., Sawano, Y., Kato, J. I. and Mori, H. Superpositioning of deletions promotes growth of escherichia coli with a reduced genome. *DNA Res.* **15,** 277–284 (2008).

[224] Iwadate, Y., Honda, H., Sato, H., Hashimoto, M. and Kato, J.-I. Oxidative stress sensitivity of engineered Escherichia coli cells with a reduced genome. *FEMS Microbiol. Lett.* **322,** 25–33 (2011).

[225] Reuß, D. R., Altenbuchner, J., Mäder, U., Rath, H., Ischebeck, T., Sappa, P. K., Thürmer, A., Guérin, C., Nicolas, P., Steil, L., Zhu, B., Feussner, I., Klumpp, S., Daniel, R., Commichau, F. M., Völker, U. and Stülke, J. Large-scale reduction of the Bacillus subtilis genome: Consequences for the transcriptional network, resource allocation, and metabolism. *Genome Res.* **27,** 289–299 (2017).

[226] Hutchison III, C. A., Chuang, R. Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., Gill, J., Kannan, K., Karas, B. J., Ma, L., Pelletier, J. F., Qi, Z.-Q., Richter, R. A., Strychalski, E. A., Sun, L., Suzuki, Y., Tsvetanova, B., Wise, K. S., Smith, H. O., Glass, J. I., Merryman, C., Gibson, D. G. and Venter, J. C. Design and synthesis of a minimal bacterial genome. *Science (80- .).* **351,** aad6253 (2016).

[227] Fredens, J., Wang, K., de la Torre, D., Funke, L. F. H., Robertson, W. E., Christova, Y., Chia, T., Schmied, W. H., Dunkelmann, D. L., Beránek, V., Uttamapinant, C., Llamazares, A. G., Elliott, T. S. and Chin, J. W. Total synthesis of Escherichia coli with a recoded genome. *Nature* **569,** 514–518 (2019).

[228] Schwille, P. Bottom-up synthetic biology: Engineering in a Tinkerer's World. *Science (80-.).* **333,** 1252–1254 (2011).

[229] Schwille, P., Spatz, J., Landfester, K., Bodenschatz, E., Herminghaus, S., Sourjik, V., Erb, T., Bastiaens, P., Lipowsky, R., Hyman, A., Dabrock, P., Baret, J. C., Vidakovic-Koch, T., Bieling, P., Dimova, R., Mutschler, H., Robinson, T., Tang, T.-Y. D., Wegner, S. and Sundmacher, K. MaxSynBio: Avenues Towards Creating Cells from the Bottom Up. *Angew. Chemie - Int. Ed.* **57,**

13382–13392 (2018).

[230] Jia, H. and Schwille, P. Bottom-up synthetic biology: reconstitution in space and time. *Curr. Opin. Biotechnol.* **60,** 179–187 (2019).

[231] Jiang, W., Wu, Z., Gao, Z., Wan, M., Zhou, M., Mao, C. and Shen, J. Artificial Cells: Past, Present and Future. *ACS Nano* **16,** 15705–15733 (2022).

[232] Hirschi, S., Ward, T. R., Meier, W. P., Muller, D. J. and Fotiadis, D. Synthetic Biology: Bottom-Up Assembly of Molecular Systems. *Chem. Rev.* **122,** 16294–16328 (2022).

[233] Laohakunakorn, N., Grasemann, L., Lavickova, B., Michielin, G., Shahein, A., Swank, Z. and Maerkl, S. J. Bottom-Up Construction of Complex Biomolecular Systems With Cell-Free Synthetic Biology. *Front. Bioeng. Biotechnol.* **8,** 1–26 (2020).

[234] Budisa, N. Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire. *Angew. Chemie - Int. Ed.* **43,** 6426–6463 (2004).

[235] Budisa, N. *Engineering the Genetic Code: Expanding the Amino Acid Repertoire for the Design of Novel Proteins*. (Wiley-VCH Verlag GmbH & Co. KGaA, 2005). doi:10.1002/3527607188

[236] Hoesl, M. G. and Budisa, N. Recent advances in genetic code engineering in Escherichia coli. *Curr. Opin. Biotechnol.* **23,** 751–757 (2012).

[237] Minks, C., Alefelder, S., Moroder, L., Huber, R. and Budisa, N. Towards new protein engineering: In vivo building and folding of protein shuttles for drug delivery and targeting by the selective pressure incorporation (SPI) method. *Tetrahedron* **56,** 9431–9442 (2000).

[238] Wang, L., Brock, A., Herberich, B. and Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science (80-.).* **292,** 498–500 (2001).

[239] Liu, C. C. and Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79,** 413–444 (2010).

[240] Dumas, A., Lercher, L., Spicer, C. D. and Davis, B. G. Designing logical codon reassignment-Expanding the chemistry in biology. *Chem. Sci.* **6,** 50–69 (2015).

[241] Agostini, F., Völler, J. S., Koksch, B., Acevedo-Rocha, C. G., Kubyshkin, V. and Budisa, N. Biocatalysis with Unnatural Amino Acids: Enzymology Meets Xenobiology. *Angew. Chemie - Int. Ed.* **56,** 9680–9703 (2017).

[242] Exner, M. P., Köhling, S., Rivollier, J., Gosling, S., Srivastava, P., Palyancheva, Z. I., Herdewijn, P., Heck, M.-P., Rademann, J. and Budisa, N. Incorporation of amino acids with long-chain terminal olefins into proteins. *Molecules* **21,** 287 (2016).

[243] Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L. and Schultz, P. G. Addition of p-

Azido-L-phenylalanine to the Genetic Code of Escherichia coli. *J. Am. Chem. Soc.* **124,** 9026–3027 (2002).

[244] Wangt, L., Zhang, Z., Brock, A. and Schultz, P. G. Addition of the keto functional group to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 56–61 (2003).

[245] Prescher, J. A. and Bertozzi, V. R. Chemistry in living systems. *Nat. Chem. Biol.* **1,** 13–21 (2005).

[246] Johnson, J. A., Lu, Y. Y., Deventer, J. A. Van and Tirrell, D. A. Residue-specific incorporation of non-canonical amin acids into Proteins: recent developments and applications. *Curr Opin Chem Biol* **14,** 774–780 (2011).

[247] Bird, R. E., Lemmel, S. A., Yu, X. and Zhou, Q. A. Bioorthogonal Chemistry and Its Applications. *Bioconjug. Chem.* **32,** 2457–2479 (2021).

[248] Baumann, T., Schmitt, F. J., Pelzer, A., Spiering, V. J., Freiherr von Sass, G. J., Friedrich, T. and Budisa, N. Engineering 'golden' fluorescence by selective pressure incorporation of non-canonical amino acids and protein analysis by mass spectrometry and fluorescence. *J. Vis. Exp.* **2018,** e57017 (2018).

[249] Yu, Y., Zhou, Q., Wang, L., Liu, X., Zhang, W., Hu, M., Dong, J., Li, J., Lv, X., Ouyang, H., Li, H., Gao, F., Gong, W., Lu, Y. and Wang, J. Significant improvement of oxidase activity through the genetic incorporation of a redox-active unnatural amino acid. *Chem. Sci.* **6,** 3881–3885 (2015).

[250] Alfonta, L., Zhang, Z., Uryu, S., Loo, J. A. and Schultz, P. G. Site-Specific Incorporation of a Redox-Active Amino Acid into Proteins. *J. Am. Chem. Soc.* **125,** 14662–14663 (2003).

[251] Cowie, D. B. and Cohen, G. N. Biosynthesis by Escherichia coli of active altered proteins containing selenium instead of sulfur. *BBA - Biochim. Biophys. Acta* **26,** 252–261 (1957).

[252] Hendrickson, W. A., Horton, J. R. and LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): A vehicle for direct determination of three dimensional structure. *EMBO J.* **9,** 1665–1672 (1990).

[253] Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J. and Huber, R. High-level Biosynthetic Substitution of Methionine in Proteins by its Analogs 2-Aminohexanoic Acid, Selenomethionine, Telluromethionine and Ethionine in Escherichia coli. *Eur. J. Biochem.* **230,** 788–796 (1995).

[254] Bae, J. H., Alefelder, S., Kaiser, J. T., Friedrich, R., Moroder, L., Huber, R. and Budisa, N. Incorporation of β-selenolo[3,2-b]pyrrolyl-alanine into proteins for phase determination in protein X-ray crystallography. *J. Mol. Biol.* **309,** 925–936 (2001).

[255] Boles, J. O., Lewinski, K., Kunhkle, M., Odom, J. D., Dunlap, R. B., Lebioda, L. and Hatada, M. Bio-incorporation of telluromethionine into buried residues of dihydrofolate reductase. *Nat.*

Struct. Biol. **1,** 283–284 (1994).

[256]   Budisa, N., Karnbrock, W., Steinbacher, S., Humm, A., Prade, L., Neuefeind, T., Moroder, L. and Huber, R. Bioincorporation of telluromethionine into proteins: A promising new approach for X-ray structure analysis of proteins. *J. Mol. Biol.* **270,** 616–623 (1997).

[257]   Hoesl, M. G. and Budisa, N. In vivo incorporation of multiple noncanonical amino acids into proteins. *Angew. Chemie - Int. Ed.* **50,** 2896–2902 (2011).

[258]   von Freiherr Sass, G. J., Blain-Hartung, M., Baumann, T., Forest, K. T., Hildebrandt, P. and Budisa, N. Orthogonal translation with 5-cyanotryptophan as an infrared probe for local structural information, electrostatics, and hydrogen bonding. *Protein Sci.* **32,** e4705 (2023).

[259]   Baumann, T., Hauf, M., Schildhauer, F., Eberl, K. B., Durkin, P. M., Deniz, E., Löffler, J. G., Acevedo-Rocha, C. G., Jaric, J., Martins, B. M., Dobbek, H., Bredenbeck, J. and Budisa, N. Site-Resolved Observation of Vibrational Energy Transfer Using a Genetically Encoded Ultrafast Heater. *Angew. Chemie - Int. Ed.* **58,** 2899–2903 (2019).

[260]   Minnihan, E. C., Young, D. D., Schultz, P. G. and Stubbe, J. Incorporation of Fluorotyrosines into Ribonucleotide Reductase Using an Evolved, Polyspecific Aminoacyl-tRNA Synthetase. *J. Am. Chem. Soc.* **133,** 15942–15945 (2011).

[261]   Minnihan, E. C., Nocera, D. G. and Stubbe, J. Reversible, Long-Range Radical Transfer in E.coli Class Ia Ribonucleotide Reductase. *Acc* **46,** 2524–2535 (2013).

[262]   Stubbe, J. and Nocera, D. G. Radicals in Biology: Your Life Is in Their Hands. *J. Am. Chem. Soc.* **143,** 13463–13472 (2021).

[263]   Cui, C., Song, D. Y., Drennan, C. L., Stubbe, J. and Nocera, D. G. Radical Transport Facilitated by a Proton Transfer Network at the Subunit Interface of Ribonucleotide Reductase. *J. Am. Chem. Soc.* **145,** 5145–5154 (2023).

[264]   Budisa, N. and Schneider, T. Expanding the DOPA Universe with Genetically Encoded, Mussel-Inspired Bioadhesives for Material Sciences and Medicine. *ChemBioChem* **20,** 2163–2190 (2019).

[265]   Hauf, M., Richter, F., Schneider, T., Faidt, T., Martins, B. M., Baumann, T., Durkin, P., Dobbek, H., Jacobs, K., Möglich, A. and Budisa, N. Photoactivatable Mussel-Based Underwater Adhesive Proteins by an Expanded Genetic Code. *ChemBioChem* **18,** 1819–1823 (2017).

[266]   Tolle, I. Towards the creation of synthetic Escherichia coli via Tryptophan and Methionine substitutions, PhD Thesis, Technische Universität Berlin. (2021).

[267]   Mitchell, H. K. and Niemann, C. The competitive inhibition of the metabolism of α-amino acids by their halogenated analogs. *J. Am. Chem. Soc.* **69,** 1232 (1947).

[268] Levine, M. and Tarver, H. Studies on Ethionine III. Incorporation of Ethionine into rat proteins. *J. Biol. Chem.* **192,** 835–850 (1951).

[269] Richmond, M. H. The effect of amino acid analogues on growth and protein synthesis in microorganisms. *Bacteriol. Rev.* **26,** 398–420 (1962).

[270] Hortin, G. and Boime, I. Applications of amino acid analogs for studying co- and posttranslational modifications of proteins. *Methods Enzymol.* **96,** 777–784 (1983).

[271] Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. and Schultz, P. G. A General Method for Site-Specific Incorporation of Unnatural Amino Acids into Proteins. *Science (80-.).* **244,** 182–188 (1989).

[272] Wang, L., Magliery, T. J., Liu, D. R. and Schultz, P. G. A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins. *J. Am. Chem. Soc.* **122,** 5010–5011 (2000).

[273] Wang, L. and Schultz, P. G. A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* **8,** 883–890 (2001).

[274] Baumann, T., Exner, M. and Budisa, N. in *Zhao, H., Zeng, AP. Synth. Biol. – Metab. Eng. Adv. Biochem. Eng.* **162,** (2016).

[275] Wan, W., Tharp, J. M. and Liu, W. R. Pyrrolysyl-tRNA Synthetase: an ordinary enzyme but an outstanding genetic expansion tool. *Biochim. Biophys. Acta* **6,** 1059–1070 (2014).

[276] Wang, L., Xie, J. and Schultz, P. G. Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 225–249 (2006).

[277] Young, T. S. and Schultz, P. G. Beyond the canonical 20 amino acids: Expanding the genetic lexicon. *J. Biol. Chem.* **285,** 11039–11044 (2010).

[278] Portnoy, V. A., Bezdan, D. and Zengler, K. Adaptive laboratory evolution-harnessing the power of biology for metabolic engineering. *Curr. Opin. Biotechnol.* **22,** 590–594 (2011).

[279] Luria, S. E. and Delbrück, M. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* **28,** 491–511 (1943).

[280] Novick, A. and Szilard, L. Experiments with the Chemostat on spontaneous mutations of bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **36,** 708–719 (1950).

[281] Atwood, K. C., Schneider, L. K. and Ryan, F. J. Periodic selection in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **37,** 146–155 (1951).

[282] Dragosits, M. and Mattanovich, D. Adaptive laboratory evolution principles and applications in industrial biotechnology. *Microb. Cell Fact.* **12,** 64 (2013).

[283] Sandberg, T. E., Salazar, M. J., Weng, L. L., Palsson, B. O. and Feist, A. M. The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. *Metab. Eng.* **56,** 1–16 (2019).

[284] Gresham, D. and Dunham, M. J. The enduring utility of continuous culturing in experimental evolution. *Genomics* **104,** 399–405 (2014).

[285] Vasi, F., Travisano, M. and Lenski, R. E. Long-term experimental evolution in Escherichia coli. II.Changes in life-history traits during adaptation to a seasonal environment. *Am. Nat.* **144,** 432–456 (1994).

[286] Lenski, R. E., Mongold, J. A., Sniegowski, P. D., Travisano, M., Vasi, F., Gerrish, P. J. and Schmidt, T. M. Evolution of competitive fitness in experimental populations of E. coli: What makes one genotype a better competitor than another? *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* **73,** 35–47 (1998).

[287] Maddamsetti, R., Lenski, R. E. and Barrick, J. E. Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with Escherichia coli. *Genetics* **200,** 619–631 (2015).

[288] Halligan, D. L. and Keightley, P. D. Spontaneous mutation accumulation studies in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* **40,** 151–172 (2009).

[289] Barrick, J. E. and Lenski, R. E. Genome dynamics during experimental evolution. *Nat. Rev. Genet.* **14,** 827–839 (2013).

[290] Lenski, R. E., Rose, M. R., Simpson, S. C. and Tadler, S. C. Long-Term Experimental Evolution in E.coli. I. Adaptation and Divergence durign 2,00 Generations. *Am. Nat.* **138,** 1315–1341 (1991).

[291] Lenski, R. E. and Travisano, M. Dynamics of adaptation and diversification: A 10,000-generation experiment with bacterial populations. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 6808–6814 (1994).

[292] Blount, Z. D., Borland, C. Z. and Lenski, R. E. Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 7899–7906 (2008).

[293] Blount, Z. D., Barrick, J. E., Davidson, C. J. and Lenski, R. E. Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature* **489,** 513–518 (2012).

[294] Lenski, R. E. Revisiting the Design of the Long-Term Evolution Experiment with Escherichia coli. *J. Mol. Evol.* **91,** 241–253 (2023).

[295] Wiser, M. J. and Lenski, R. E. A comparison of methods to measure fitness in Escherichia coli. *PLoS One* **10,** e0126210 (2015).

[296] Lenski, R. E. Convergence and divergence in a long-term experiment with bacteria. *Am. Nat.* **190,** S57–S68 (2017).

[297] Lenski, R. E. in *Plant Breed. Rev* **24,** 225–265 (2004).

[298] Riehle, M. M., Bennett, A. F., Lenski, R. E. and Long, A. D. Evolutionary changes in heat-inducible gene expression in lines of Escherichia coli adapted to high temperature. *Physiol. Genomics* **14,** 47–58 (2003).

[299] Rudolph, B., Gebendorfer, K. M., Buchner, J. and Winter, J. Evolution of Escherichia coli for growth at high temperatures. *J. Biol. Chem.* **285,** 19029–19034 (2010).

[300] Blaby, I. K., Lyons, B. J., Wroclawska-Hughes, E., Phillips, G. C. F., Pyle, T. P., Chamberlin, S. G., Benner, S. A., Lyons, T. J., de Crécy-Lagard, V. and de Crécy, E. Experimental evolution of a facultative thermophile from a mesophilic ancestor. *Appl. Environ. Microbiol.* **78,** 144–155 (2012).

[301] Deatherage, D. E., Kepner, J. L., Bennett, A. F., Lenski, R. E. and Barrick, J. E. Specificity of genome evolution in experimental populations of Escherichia coli evolved at different temperatures. *Proc. Natl. Acad. Sci. U. S. A.* **114,** E1904–E1912 (2017).

[302] Sleight, S. C. and Lenski, R. E. Adaptation to Freeze-Thaw-Growth Cycles in Escherichia coli. *Physiol. Biochem. Zool.* **80,** 370–385 (2007).

[303] Sleight, S. C., Orlic, C., Schneider, D. and Lenski, R. E. Genetic basis of evolutionary adaptation by Echerichia coli to stressful cycles of freezing, thawing and growth. *Genetics* **180,** 431–443 (2008).

[304] Stoebel, D. M., Hokamp, K., Last, M. S. and Dorman, C. J. Compensatory evolution of gene regulation in response to stress by Escherichia coli lacking RpoS. *PLoS Genet.* **5,** e1000671 (2009).

[305] Dragosits, M., Mozhayskiy, V., Quinones-Soto, S., Park, J. and Tagkopoulos, I. Evolutionary potential, cross-stress behavior and the genetic basis of acquired stress resistance in Escherichia coli. *Mol. Syst. Biol.* **9,** 1–13 (2013).

[306] Winkler, J. D., Garcia, C., Olson, M., Callaway, E. and Kao, K. C. Evolved osmotolerant Escherichia coli mutants frequently exhibit defective N-acetylglucosamine catabolism and point mutations in cell shape-regulating protein MreB. *Appl. Environ. Microbiol.* **80,** 3729–3740 (2014).

[307] Fletcher, E., Feizi, A., Bisschops, M. M. M., Hallström, B. M., Khoomrung, S., Siewers, V. and Nielsen, J. Evolutionary engineering reveals divergent paths when yeast is adapted to different acidic environments. *Metab. Eng.* **39,** 19–28 (2017).

[308] Nguyen-Vo, T. P., Liang, Y., Sankaranarayanan, M., Seol, E., Chun, A. Y., Ashok, S., Chauhan, A. S., Kim, J. R. and Park, S. *Development of 3-hydroxypropionic-acid-tolerant strain of Escherichia coli W and role of minor global regulator yieP. Metab. Eng.* **53,** (Elsevier Inc., 2019).

[309] Salas-Navarrete, P. C., de Oca Miranda, A. I. M., Martínez, A. and Caspeta, L. Evolutionary and reverse engineering to increase Saccharomyces cerevisiae tolerance to acetic acid, acidic pH, and high temperature. *Appl. Microbiol. Biotechnol.* **106,** 383–399 (2022).

[310] Alcántara-Díaz, D., Breña-Valle, M. and Serment-Guerrero, J. Divergent adaptation of Escherichia coli to cyclic ultraviolet light exposures. *Mutagenesis* **19,** 349–354 (2004).

[311] Shibai, A., Takahashi, Y., Ishizawa, Y., Motooka, D., Nakamura, S., Ying, B. W. and Tsuru, S. Mutation accumulation under UV radiation in Escherichia coli. *Sci. Rep.* **7,** 14531 (2017).

[312] Çakar, Z. P., Seker, U. O. S., Tamerler, C., Sonderegger, M. and Sauer, U. Evolutionary engineering of multiple-stress resistant Saccharomyces cerevisiae. *FEMS Yeast Res.* **5,** 569–578 (2005).

[313] Papapetridis, I., Verhoeven, M. D., Wiersma, S. J., Goudriaan, M., Van Maris, A. J. A. and Pronk, J. T. Laboratory evolution for forced glucose-xylose co-consumption enables identification of mutations that improve mixed-sugar fermentation by xylose-fermenting Saccharomyces cerevisiae. *FEMS Yeast Res.* **18,** foy056 (2018).

[314] Cadière, A., Ortiz-Julien, A., Camarasa, C. and Dequin, S. Evolutionary engineered Saccharomyces cerevisiae wine yeast strains with increased in vivo flux through the pentose phosphate pathway. *Metab. Eng.* **13,** 263–271 (2011).

[315] Sanchez, R. G., Karhurmaa, K., Fonseca, C., Nogué, V. S., Almeida, J. R. M., Larsson, C. U., Bengtsson, O., Bettiga, M., Hahn-Hägerdal, B. and Gorwa-Grauslund, M. F. Improved xylose and arabinose utilization by an industrial recombinant Saccharomyces cerevisiae strain using evolutionary engineering. *Biotechnol. Biofuels* **3,** 13 (2010).

[316] Guimarães, P. M. R., François, J., Parrou, J. L., Teixeira, J. A. and Domingues, L. Adaptive evolution of a lactose-consuming Saccharomyces cerevisiae recombinant. *Appl. Environ. Microbiol.* **74,** 1748–1756 (2008).

[317] Guzmán, G. I., Sandberg, T. E., LaCroix, R. A., Nyerges, Á., Papp, H., de Raad, M., King, Z. A., Hefner, Y., Northen, T. R., Notebaart, R. A., Pál, C., Palsson, B. O., Papp, B. and Feist, A. M. Enzyme promiscuity shapes adaptation to novel growth substrates. *Mol. Syst. Biol.* **15,** e8462 (2019).

[318] Rajaraman, E., Agarwal, A., Crigler, J., Seipelt-Thiemann, R., Altman, E. and Eiteman, M. A. Transcriptional analysis and adaptive evolution of Escherichia coli strains growing on acetate. *Appl. Microbiol. Biotechnol.* **100,** 7777–7785 (2016).

[319] Lee, D.-H. and Palsson, B. O. Adaptive evolution of Escherichia coli K-12 MG1655 during growth on a nonnative carbon source, L-l,2-propanediol. *Appl. Environ. Microbiol.* **76,** 4158–4168 (2010).

[320] Luan, G., Cai, Z., Li, Y. and Ma, Y. Genome replication engineering assisted continuous evolution (GREACE) to improve microbial tolerance for biofuels production. *Biotechnol. Biofuels* **6,** 137 (2013).

[321] Brennan, T. C. R., Williams, T. C., Schulz, B. L., Palfreyman, R. W., Krömer, J. O. and Nielsen, L. K. Evolutionary engineering improves tolerance for replacement jet fuels in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **81,** 3316–3325 (2015).

[322] Royce, L. A., Yoon, J. M., Chen, Y., Rickenbach, E., Shanks, J. V. and Jarboe, L. R. Evolution for exogenous octanoic acid tolerance improves carboxylic acid production and membrane integrity. *Metab. Eng.* **29,** 180–188 (2015).

[323] Vilela, L. de F., de Araujo, V. P. G., Paredes, R. de S., Bon, E. P. da S., Torres, F. A. G., Neves, B. C. and Eleutherio, E. C. A. Enhanced xylose fermentation and ethanol production by engineered Saccharomyces cerevisiae strain. *AMB Express* **5,** 16 (2015).

[324] Shepelin, D., Hansen, A. S. L., Lennen, R., Luo, H. and Herrgård, M. J. Selecting the best: Evolutionary engineering of chemical production in microbes. *Genes (Basel).* **9,** 249 (2018).

[325] Tze Fei Wong, J. Membership mutation of the genetic code: Loss of fitness by tryptophan. *Proc. Natl. Acad. Sci. U. S. A.* **80,** 6303–6306 (1983).

[326] Bacher, J. M. and Ellington, A. D. Selection and Characterization of Escherichia coli Variants Capable of Growth on an Otherwise Toxic Tryptophan Analogue. *J. Bacteriol.* **183,** 5414–5425 (2001).

[327] Hoesl, M. G., Oehm, S., Durkin, P., Darmon, E., Peil, L., Aerni, H.-R., Rappsilber, J., Rinehart, J., Leach, D., Söll, D. and Budisa, N. Chemical Evolution of a Bacterial Proteome. *Angew. Chem. Int. Ed.* **54,** 10030–10034 (2015).

[328] Agostini, F., Sinn, L., Petras, D., Schipp, C. J., Kubyshkin, V., Berger, A. A., Dorrestein, P. C., Rappsilber, J., Budisa, N. and Koksch, B. Multiomics Analysis Provides Insight into the Laboratory Evolution of Escherichia coli toward the Metabolic Usage of Fluorinated Indoles. *ACS Cent. Sci.* **7,** 81–92 (2021).

[329] Moger-Reischer, R. Z., Glass, J. I., Wise, K. S., Sun, L., Bittencourt, D., Lynch, M. and Lennon, J. T. Evolution of a minimal cell. *Nature* **620,** 122–127 (2023).

[330] Choe, D., Lee, J. H., Yoo, M., Hwang, S., Sung, B. H., Cho, S., Palsson, B., Kim, S. C. and Cho, B.-K. Adaptive laboratory evolution of a genome-reduced Escherichia coli. *Nat. Commun.* **10,** 935 (2019).

[331] Hirasawa, T. and Maeda, T. Adaptive Laboratory Evolution of Microorganisms: Methodology and Application for Bioproduction. *Microorganisms* **11,** 1–21 (2023).

[332] Wong, J. T.-F. Membership mutation of the genetic code: Loss of fitness by tryptophan. *Proc. Natl. Acad. Sci. U. S. A.* **80,** 6303–6306 (1983).

[333] Ohnishi, J., Mizoguchi, H., Takeno, S. and Ikeda, M. Characterization of mutations induced by N-methyl-N'-nitro-N-nitrosoguanidine in an industrial Corynebacterium glutamicum strain. *Mutat. Res.* **649,** 239–244 (2008).

[334] Mobini-Dehkordi, M., Nahvi, I., Zarkesh-Esfahani, H., Ghaedi, K., Tavassoli, M. and Akada, R. Isolation of a novel mutant strain of Saccharomyces cerevisiae by an ethyl methane sulfonate-induced mutagenesis approach as a high producer of bioethanol. *J. Biosci. Bioeng.* **105,** 403–408 (2008).

[335] Maeda, T., Shibai, A., Yokoi, N., Tarusawa, Y., Kawada, M., Kotani, H. and Furusawa, C. Mutational property of newly identified mutagen L-glutamic acid γ-hydrazide in Escherichia coli. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **823,** 111759 (2021).

[336] Schaaper, R. M. and Radman, M. The extreme mutator effect of Escherichia coli mutD5 results from saturation of mismatch repair by excessive DNA replication errors. *EMBO J.* **8,** 3511–3516 (1989).

[337] Selifonova, O., Valle, F. and Schellenberger, V. Rapid Evolution of Novel Traits in Microorganisms. *Appl. Environ. Microbiol.* **67,** 3645–3649 (2001).

[338] Abe, H., Fujita, Y., Takaoka, Y., Kurita, E., Yano, S., Tanaka, N. and Nakayama, K. ichi. Ethanol-tolerant Saccharomyces cerevisiae strains isolated under selective conditions by over-expression of a proofreading-deficient DNA polymerase δ. *J. Biosci. Bioeng.* **108,** 199–204 (2009).

[339] Eom, G. E., Lee, H. and Kim, S. Development of a genome-targeting mutator for the adaptive evolution of microbial cells. *Nucleic Acids Res.* **50,** e38 (2022).

[340] Garst, A. D., Bassalo, M. C., Pines, G., Lynch, S. A., Halweg-Edwards, A. L., Liu, R., Liang, L., Wang, Z., Zeitoun, R., Alexander, W. G. and Gill, R. T. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat. Biotechnol.* **35,** 48–55 (2017).

[341] Zheng, Y., Kong, S., Luo, S., Chen, C., Cui, Z., Sun, X., Chen, T. and Wang, Z. Improving Furfural Tolerance of Escherichia coli by Integrating Adaptive Laboratory Evolution with CRISPR-Enabled Trackable Genome Engineering (CREATE). *ACS Sustain. Chem. Eng.* **10,** 2318–2330 (2022).

[342] Phaneuf, P. V., Gosting, D., Palsson, B. O. and Feist, A. M. ALEdb 1.0: A database of mutations from adaptive laboratory evolution experimentation. *Nucleic Acids Res.* **47,** D1164–D1171 (2019).

[343] Bacher, J. M. and Ellington, A. D. Selection and Characterization of Escherichia coli Variants

Capable of Growth on an Otherwise Toxic Tryptophan Analogue. *J. Bacteriol.* **183,** 5414–5425 (2001).

[344] Bacher, J. M., Bull, J. J. and Ellington, A. D. Evolution of phage with chemically ambiguous proteomes. *BMC Evol. Biol.* **3,** 24 (2003).

[345] Mat, W.-K., Xue, H. and Wong, J. T.-F. Genetic code mutations: The breaking of a three billion year invariance. *PLoS One* **5,** e12206 (2010).

[346] Yu, A. C.-S., Yim, A. K.-Y., Mat, W.-K., Tong, A. H.-Y., Lok, S., Xue, H., Tsui, S. K.-W., Wong, J. T.-F. and Chan, T.-F. Mutations enabling displacement of tryptophan by 4-fluorotryptophan as a canonical amino acid of the genetic code. *Genome Biol. Evol.* **6,** 629–641 (2014).

[347] Gollnick, P., Ishino, S., Kuroda, M. I., Henner, D. J. and Yanofsky, C. The mtr locus is a two-gene operon required for transcription attenuation in the trp operon of Bacillus subtilis. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 8726–8730 (1990).

[348] Bohnert, J. A., Schuster, S., Fähnrich, E., Trittler, R. and Kern, W. V. Altered spectrum of multidrug resistance associated with a single point mutation in the Escherichia coli RND-type MDR efflux pump YhiV (MdtF). *J. Antimicrob. Chemother.* **59,** 1216–1222 (2007).

[349] Pietsch, F., Bergman, J. M., Brandis, G., Marcusson, L. L., Zorzet, A., Huseby, D. L. and Hughes, D. Ciprofloxacin selects for RNA polymerase mutations with pleiotropic antibiotic resistance effects. *J. Antimicrob. Chemother.* **72,** 75–84 (2017).

[350] Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. and Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. proteomics* **1,** 376–386 (2002).

[351] Pogliano, J., Lynch, A. S., Belin, D., Lin, E. C. C. and Beckwith, J. Regulation of Escherichia coli cell envelope proteins involved in protein folding and degradation by the Cpx two-component system. *Genes Dev.* **11,** 1169–1182 (1997).

[352] Treiber-Kleinke, C., Berger, A. A., Adrian, L., Budisa, N. and Koksch, B. Escherichia coli adapts metabolically to 6- and 7-fluoroindole, enabling proteome-wide fluorotryptophan substitution. *Front. Synth. Biol.* **1,** 1345634 (2024).

[353] Bacher, J. M., Hughes, R. A., Wong, J. T.-F. and Ellington, A. D. Evolving new genetic codes. *Trends Ecol. Evol.* **19,** 69–75 (2004).

[354] Zhang, F. and Ellington, A. D. Hurdling and Hurtling Toward New Genetic Codes. *ACS Cent. Sci.* **7,** 7–10 (2021).

[355] Browne, D. T., Kenyon, G. L. and Hegeman, G. D. Incorporation of monofluorotryptophans into protein during the growth of Escherichia coli. *Biochem. Biophys. Res. Commun.* **39,** 13–19 (1970).

[356] Pratt, E. A. and Ho, C. Incorporation of Fluorotryptophans into Proteins of Escherichia coli. *Biochemistry* **14,** 3035–3040 (1975).

[357] Piñero-Fernandez, S., Chimerel, C., Keyser, U. F. and Summers, D. K. Indole transport across Escherichia coli membranes. *J. Bacteriol.* **193,** 1793–1798 (2011).

[358] Bartoli, G. and Palmieri, G. The reaction of vinyl Grinard reagent with 2-substituted nitroarenes: a new approach to synthesis of 7-substituted indoles. *Tetrahedron Lett.* **30,** 2129–2132 (1989).

[359] Lee, M. and Phillips, R. S. Synthesis and resolution of 7-Fluorotryptophans. *Bioorg. Med. Chem. Lett.* **1,** 477–480 (1991).

[360] Allen, F. L., Brunton, J. C. and Suschitzky, H. Heterocyclic Fluorine Compomds. Part II.* Bz-Monofluoroindoles. *J. Chem. Soc.* 1283–1286 (1955).

[361] Orata, F. in *Adv. Gas Chromatogr. - Prog. Agric. Biomed. Ind. Appl.* 83–108 (2012). doi:10.5772/33098

[362] Nichols, B. P. and Yanofsky, C. Nucleotide sequences of trpA of Salmonella typhimurium and Escherichia coli: An evolutionary comparison. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 5244–5248 (1979).

[363] Crawford, I. P., Nichols, B. P. and Yanofsky, C. Nucleotide sequence of the trpB gene in Escherichia coli and Salmonella typhimurium. *J. Mol. Biol.* **142,** 489–502 (1980).

[364] Kulik, V., Weyand, M., Seidel, R., Niks, D., Arac, D., Dunn, M. F. and Schlichting, I. On the role of αThr183 in the allosteric regulation and catalytic mechanism of tryptophan synthase. *J. Mol. Biol.* **324,** 677–690 (2002).

[365] Datsenko, K. A. and Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 6640–6645 (2000).

[366] Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J. and Huber, R. High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in Escherichia coli. *Eur. J. Biochem.* **230,** 788–796 (1995).

[367] Barer, M. R. *Bacterial Growth, Culturability and Viability. Mol. Med. Microbiol.* (Elsevier Ltd, 2015). doi:10.1016/B978-0-12-397169-2.00010-X

[368] LaCroix, R. A., Sandberg, T. E., O'Brien, E. J., Utrilla, J., Ebrahim, A., Guzman, G. I., Szubin, R., Palsson, B. O. and Feist, A. M. Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of Escherichia coli K-12 MG1655 on glucose minimal medium. *Appl. Environ. Microbiol.* **81,** 17–30 (2014).

[369] Paliy, O. and Gunasekera, T. S. Growth of E. coli BL21 in minimal media with different gluconeogenic carbon sources and salt contents. *Appl. Microbiol. Biotechnol.* **73,** 1169–1172 (2007).

[370] Vasi, F., Travisano, M. and Lenski, R. E. Long-term experimental evolution in Escherichia coli. II. Changes in life-history traits during adaptation to a seasonal environment. *Am. Nat.* **144,** 432–456 (1994).

[371] LaCroix, R. A., Palsson, B. O. and Feist, A. M. A Model for Designing Adaptive Laboratory Evolution Experiments. *Appl. Environ. Microbiol.* **83,** e03115-16 (2017).

[372] Wong, H. E., Huang, C.-J. and Zhang, Z. Amino Acid Misincorporation Propensities Revealed through Systematic Amino Acid Starvation. *Biochemistry* **57,** 6767–6779 (2018).

[373] Blount, Z. D., Lenski, R. E. and Losos, J. B. Contingency and determinism in evolution: Replaying life's tape. *Science (80-.).* **362,** eaam5979 (2018).

[374] Hall, B. G., Acar, H., Nandipati, A. and Barlow, M. Growth rates made easy. *Mol. Biol. Evol.* **31,** 232–238 (2013).

[375] Yang, X., Zhong, Y., Wang, D. and Lu, Z. A simple colorimetric method for viable bacteria detection based on cell counting Kit-8. *Anal. Methods* **13,** 5211–5215 (2021).

[376] Schmidt, M. and Kubyshkin, V. How To Quantify a Genetic Firewall? A Polarity-Based Metric for Genetic Code Engineering. *ChemBioChem* **22,** 1268–1284 (2021).

[377] Schmidt, M. A metric space for semantic containment: Towards the implementation of genetic firewalls. *BioSystems* **185,** 104015 (2019).

[378] Chowdhary, S., Schmidt, R. F., Sahoo, A. K., tom Dieck, T., Hohmann, T., Schade, B., Brademann-Jock, K., Thünemann, A. F., Netz, R. R., Gradzielski, M. and Koksch, B. Rational design of amphiphilic fluorinated peptides: evaluation of self-assembly properties and hydrogel formation. *Nanoscale* **14,** 10176–10189 (2022).

[379] Edwards-Gayle, C. J. C. and Hamley, I. W. Self-assembly of bioactive peptides, peptide conjugates, and peptide mimetic materials. *Org. Biomol. Chem.* **15,** 5867–5876 (2017).

[380] Hellmund, K. S., von Lospichl, B., Böttcher, C., Ludwig, K., Keiderling, U., Noirez, L., Weiß, A., Mikolajczak, D. J., Gradzielski, M. and Koksch, B. Functionalized peptide hydrogels as tunable extracellular matrix mimics for biological applications. *Pept. Sci.* **113,** e24201 (2021).

[381] Kundu, R. Cationic Amphiphilic Peptides: Synthetic Antimicrobial Agents Inspired by Nature. *ChemMedChem* **15,** 1887–1896 (2020).

[382] Huan, Y., Kong, Q., Mou, H. and Yi, H. Antimicrobial Peptides: Classification, Design, Application and Research Progress in Multiple Fields. *Front. Microbiol.* **11,** 582779 (2020).

[383] Chowdhary, S., Pelzer, T., Saathoff, M., Quaas, E., Pendl, J., Fulde, M. and Koksch, B. Fine-tuning the antimicrobial activity of β-hairpin peptides with fluorinated amino acids. *Pept. Sci.* **115,**

e24306 (2023).

[384]   Nikaido, H. Molecular basis of bacterial outer membrane permeability. *Microbiol. Mol. Biol. Rev.* **67,** 593–656 (2003).

[385]   Delcour, A. H. Outer membrane permeability and antibiotic resistance. *Biochim. Biophys. Acta - Proteins Proteomics* **1794,** 808–816 (2009).

[386]   Pongs, O. in *Mech. Action Antibact. Agents* 26–42 (1979). doi:org/10.1007/978-3-642-46403-4_3

[387]   Tipper, D. J. Mode of action of β-lactam antibiotics. *Pharmac. Ther.* **27,** 1–35 (1985).

[388]   Ghosh, P., Nandy, T., Singh, P. C. and Ghosh, D. Substitution enables significant new decay channels for a non-canonical amino acid. *Phys. Chem. Chem. Phys.* **24,** 17695–17702 (2022).

[389]   Nandy, T., Mondal, S. and Singh, P. C. Fluorine induced conformational switching and modulation in photophysical properties of 7-fluorotryptophan: Spectroscopic, quantum chemical calculation and molecular dynamics simulation studies. *J. Photochem. Photobiol.* **3–4,** 100011 (2020).

[390]   Gu, P., Yang, F., Kang, J., Wang, Q. and Qi, Q. One-step of tryptophan attenuator inactivation and promoter swapping to improve the production of L-tryptophan in Escherichia coli. *Microb. Cell Fact.* **11,** 30 (2012).

[391]   Matsubara, H. and Sasaki, R. M. High recovery of tryptophan from acid hydrolysates of proteins. *Biochem. Biophys. Res. Commun.* **35,** 175–181 (1969).

[392]   la Cour, R., Jørgensen, H. and Schjoerring, J. K. Improvement of Tryptophan Analysis by Liquid Chromatography-Single Quadrupole Mass Spectrometry Through the Evaluation of Multiple Parameters. *Front. Chem.* **7,** 797 (2019).

[393]   Bellmaine, S., Schnellbaecher, A. and Zimmer, A. Reactivity and degradation products of tryptophan in solution and proteins. *Free Radic. Biol. Med.* **160,** 696–718 (2020).

[394]   Graves, P. R. and Haystead, T. A. J. Molecular Biologist's Guide to Proteomics. *Microbiol. Mol. Biol. Rev.* **66,** 39–63 (2002).

[395]   Deutsch, E. W., Bandeira, N., Perez-Riverol, Y., Sharma, V., Carver, J. J., Mendoza, L., Kundu, D. J., Wang, S., Bandla, C., Kamatchinathan, S., Hewapathirana, S., Pullman, B. S., Wertz, J., Sun, Z., Kawano, S., Okuda, S., Watanabe, Y., Maclean, B., Maccoss, M. J., Zhu, Y., Ishihama, Y. and Vizcaíno, J. A. The ProteomeXchange consortium at 10 years: 2023 update. *Nucleic Acids Res.* **51,** D1539–D1548 (2023).

[396]   Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D. J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., Walzer, M., Wang, S., Brazma,

A. and Vizcaíno, J. A. The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50,** D543–D552 (2022).

[397] Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., Fung, C., Nikolai, L., Lewis, M., Coutouly, M.-A., Forsythe, I., Tang, P., Shrivastava, S., Jeroncic, K., Stothard, P., Amegbey, G., Block, D., Hau, D. D., Wagner, J., Miniaci, J., Clements, M., Gebremedhin, M., Guo, N., Zhang, Y., Duggan, G. E., MacInnis, G. D., Weljie, A. M., Dowlatabadi, R., Bamforth, F., Clive, D., Greiner, R., Li, L., Marrie, T., Sykes, B. D., Vogel, H. J. and Querengesser, L. HMDB: The human metabolome database. *Nucleic Acids Res.* **35,** D521–D526 (2007).

[398] Misra, B. B. New software tools, databases, and resources in metabolomics: updates from 2020. *Metabolomics* **17,** 49 (2021).

[399] Züllig, T., Trötzmüller, M. and Köfeler, H. C. Lipidomics from sample preparation to data analysis: a primer. *Anal. Bioanal. Chem.* **412,** 2191–2209 (2020).

[400] Emwas, A. H., Szczepski, K., Al-Younis, I., Lachowicz, J. I. and Jaremko, M. Fluxomics - New Metabolomics Approaches to Monitor Metabolic Pathways. *Front. Pharmacol.* **13,** 805782 (2022).

[401] Fuhrer, T., Heer, D., Begemann, B. and Zamboni, N. High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry. *Anal. Chem.* **83,** 7074–7080 (2011).

[402] Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V. and Palsson, B. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* **3,** 1–18 (2007).

[403] Katajamaa, M., Miettinen, J. and Orešič, M. MZmine: Toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* **22,** 634–636 (2006).

[404] Xia, J. and Wishart, D. S. *Metabolomic data processing, analysis, and interpretation using MetaboAnalyst. Curr. Protoc. Bioinforma.* (2011). doi:10.1002/0471250953.bi1410s34

[405] Chong, J., Wishart, D. S. and Xia, J. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Curr. Protoc. Bioinforma.* **68,** e86 (2019).

[406] Darzi, Y., Letunic, I., Bork, P. and Yamada, T. iPath3.0: Interactive pathways explorer v3. *Nucleic Acids Res.* **46,** W510–W513 (2018).

[407] Kanehisa, M. and Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).

[408] Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.*

28, 1947–1951 (2019).

[409] Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. and Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51,** D587–D592 (2023).

[410] Tolle, I., Oehm, S., Hoesl, M. G., Treiber-Kleinke, C., Peil, L., Bozukova, M., Albers, S., Bukari, A.-R. A., Semmler, T., Rappsilber, J., Ignatova, Z., Gerstein, A. and Budisa, N. Evolving a mitigation of the stress response pathway to change the basic chemistry of life. *Front. Synth. Biol.* **1,** 1248065 (2023).

[411] Goodwin, S., McPherson, J. D. and McCombie, W. R. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17,** 333–351 (2016).

[412] Hu, T., Chitnis, N., Monos, D. and Dinh, A. Next-generation sequencing technologies: An overview. *Hum. Immunol.* **82,** 801–811 (2021).

[413] Slatko, B. E., Gardner, A. F. and Ausubel, F. M. Overview of Next-Generation Sequencing Technologies. *Curr. Protoc. Mol. Biol.* **122,** e59 (2018).

[414] Illumina. Illumina Sequencing Technology. *https://www.youtube.com/watch?v=womKfikWlxM* (2013).

[415] Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).

[416] Paaby, A. B. and Rockman, M. V. The many faces of pleiotropy. *Trends Genet.* **29,** 66–73 (2013).

[417] Phillips, P. C. Epistasis - the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* **9,** 855–867 (2008).

[418] Lee, H., Popodi, E., Tang, H. and Foster, P. L. Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **109,** E2774–E2783 (2012).

[419] Drake, J. W. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 7160–7164 (1991).

[420] Kunkel, T. A. and Erie, D. A. DNA mismatch repair. *Annu. Rev. Biochem.* **74,** 681–710 (2005).

[421] Moti, T. B. Illumina Sequencing Technology. *Mix. Res. Int.* **10,** 25–31 (2022).

[422] The Uniprot Consortium. UniProt: the Universal Protein Knowledgebase in 2023 - Google Scholar. *Nucleic Acids Res.* **51,** D523–D531 (2023).

[423] Aziz, R. K., Breitbart, M. and Edwards, R. A. Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res.* **38,** 4207–4217 (2010).

[424] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13,** 2498–2504 (2003).

[425] Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z. and Galon, J. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25,** 1091–1093 (2009).

[426] Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. and Sherlock, G. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25,** 25–29 (2000).

[427] Dessimoz, C. and Skunca, N. *The Gene Ontology Handbook. Methods Mol. Biol. 1446* (2017).

[428] Ferla, M. P. and Patrick, W. M. Bacterial methionine biosynthesis. *Microbiology* **160,** 1571–1584 (2014).

[429] Wagih, O., Galardini, M., Busby, B. P., Memon, D., Typas, A. and Beltrao, P. A resource of variant effect predictions of single nucleotide variants in model organisms. *Mol. Syst. Biol.* **14,** e8430 (2018).

[430] Choi, Y. and Chan, A. P. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31,** 2745–2747 (2015).

[431] Bailey, S. F., Morales, L. A. A. and Kassen, R. Effects of Synonymous Mutations beyond Codon Bias: The Evidence for Adaptive Synonymous Substitutions from Microbial Evolution Experiments. *Genome Biol. Evol.* **13,** 1–18 (2021).

[432] Völler, J. S., Dulic, M., Gerling-Driessen, U. I. M., Biava, H., Baumann, T., Budisa, N., Gruic-Sovulj, I. and Koksch, B. Discovery and Investigation of Natural Editing Function against Artificial Amino Acids in Protein Translation. *ACS Cent. Sci.* **3,** 73–80 (2017).

[433] Wang, Y., Wang, G., Moitessier, N. and Mittermaier, A. K. Enzyme Kinetics by Isothermal Titration Calorimetry: Allostery, Inhibition, and Dynamics. *Front. Mol. Biosci.* **7,** 583826 (2020).

[434] Jolivet-Gougeon, A., Kovacs, B., Le Gall-David, S. L., Le Bars, H., Bousarghin, L., Bonnaure-Mallet, M., Lobe, B., Guillé, F., Soussy, C. J. and Tenke, P. Bacterial hypermutation: Clinical implications. *J. Med. Microbiol.* **60,** 563–573 (2011).

[435] Lenski, R. E., Sniegowski, P. and Gerrish, P. Evolution of high mutation rates in experimental populations of E. coli. *Nature* **387,** 703–705 (1997).

[436] Drake, J. W., Charlesworth, B., Charlesworth, D. and Crow, J. F. Rates spontaneous mutation. *Genetics* **148,** 1667–1686 (1998).

[437] Radman, M., Matic, I. and Taddei, F. Evolution of Evolvability. *Ann. N. Y. Acad. Sci.* **18,** 146–155 (1999).

[438] Sniegowski, P. D., Gerrish, P. J., Johnson, T. and Shaver, A. The evolution of mutation rates: Separating causes from consequences. *BioEssays* **22,** 1057–1066 (2000).

[439] Denamur, E. and Matic, I. Evolution of mutation rates in bacteria. *Mol. Microbiol.* **60,** 820–827 (2006).

[440] Couche, A., Guelfo, J. R. and Blázquez, J. Mutational Spectrum Drives the Rise of Mutator Bacteria. *PLoS Genet.* **9,** e1003167 (2013).

[441] Funchain, P., Yeung, A., Stewart, J. L., Lin, R., Slupska, M. M. and Miller, J. H. The consequences of growth of a mutator strain of Escherichia coli as measured by loss of function among multiple gene targets and loss of fitness. *Genetics* **154,** 959–970 (2000).

[442] Mehta, H. H., Prater, A. G., Beabout, K., Elworth, R. A. L., Karavis, M., Gibbons, H. S. and Shamoo, Y. The essential role of hypermutation in rapid adaptation to antibiotic stress. *Antimicrob. Agents Chemother.* **63,** e00744-19 (2019).

[443] Swings, T., van Den Bergh, B., Wuyts, S., Oeyen, E., Voordeckers, K., Verstrepen, K. J., Fauvart, M., Verstraeten, N. and Michiels, J. Adaptive tuning of mutation rates allows fast response to lethal stress in Escherichia coli. *Elife* **6,** e22939 (2017).

[444] Jayaraman, R. Hypermutation and stress adaptation in bacteria. *J. Genet.* **90,** 383–391 (2011).

[445] Vaisman, A., McDonald, J. P. and Woodgate, R. Translesion DNA Synthesis. *EcoSal Plus* **5,** 1–32 (2012).

[446] Traxler, M. F., Summers, S. M., Nguyen, H. T., Zacharia, V. M., Hightower, G. A., Smith, J. T. and Conway, T. The global, ppGpp-mediated stringent response to amino acid starvation in Escherichia coli. *Mol. Microbiol.* **68,** 1128–1148 (2008).

[447] Ishihama, A. and Saitoh, T. Subunits of RNA polymerase in function and structure. IX. Regulation of RNA polymerase activity by stringent starvation protein (SSP). *J. Mol. Biol.* **129,** 517–530 (1979).

[448] Williams, M. D., Ouyang, T. X. and Flickinger, M. C. Starvation-induced expression of SspA and SspB: the effects of a null mutation in sspA on Escherichia coli protein synthesis and survival during growth and prolonged starvation. *Mol. Microbiol.* **11,** 1029–1043 (1994).

[449] Battesti, A., Majdalani, N. and Gottesman, S. The RpoS-mediated general stress response in Escherichia coli. *Annu. Rev. Microbiol.* **65,** 189–213 (2011).

[450] Wang, F., Shi, J., He, D., Tong, B., Zhang, C., Wen, A., Zhang, Y., Feng, Y. and Lin, W. Structural

basis for transcription inhibition by E. coli SspA. *Nucleic Acids Res.* **48,** 9931–9942 (2020).

[451] Wang, F., Feng, Y., Shang, Z. and Lin, W. A unique binding between SspA and RNAP β′NTH across low-GC Gram-negative bacteria facilitates SspA-mediated transcription regulation. *Biochem. Biophys. Res. Commun.* **583,** 86–92 (2021).

[452] Hansen, A.-M., Qiu, Y., Yeh, N., Blattner, F. R., Durfee, T. and Jin, D. J. SspA is required for acid resistance in stationary phase by downregulation of H-NS in Escherichia coli. *Mol. Microbiol.* **56,** 719–734 (2005).

[453] Sass, T. H., Ferrazzoli, A. E. and Lovett, S. T. DnaA and SspA regulation of the iraD gene of Escherichia coli: an alternative DNA damage response independent of LexA/RecA. *Genetics* **221,** iyac062 (2022).

[454] Merrikh, H., Ferrazzoli, A. E., Bougdour, A., Olivier-Mason, A. and Lovett, S. T. A DNA damage response in Escherichia coli involving the alternative sigma factor, RpoS. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 611–616 (2009).

[455] Levchenko, I., Seidel, M., Sauer, R. T. and Baker, T. A. A specificity-enhancing factor for the clpXP degradation machine. *Science (80-.).* **289,** 2354–2356 (2000).

[456] Ferenci, T. What is driving the acquisition of mutS and rpoS polymorphisms in Escherichia coli? *Trends Microbiol.* **11,** 457–461 (2003).

[457] Zhao, D., Feng, X., Zhu, X., Wu, T., Zhang, X. and Bi, C. CRISPR/Cas9-assisted gRNA-free one-step genome editing with no sequence limitations and improved targeting efficiency. *Sci. Rep.* **7,** 16624 (2017).

[458] Herring, C. D., Raghunathan, A., Honisch, C., Patel, T., Applebee, M. K., Joyce, A. R., Albert, T. J., Blattner, F. R., Van Den Boom, D., Cantor, C. R. and Palsson, B. Comparative genome sequencing of Escherichia coli allows observation of bacterial evolution on a laboratory timescale. *Nat. Genet.* **38,** 1406–1412 (2006).

[459] Shankar, S., Schlictman, D. and Chakrabarty, A. M. Regulation of nucleoside diphosphate kinase and an alternative kinase in Escherichia coli: role of the sspA and rnk genes in nucleoside triphosphate formation. *Mol. Microbiol.* **17,** 935–943 (1995).

[460] Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R. and Church, G. M. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–898 (2009).

[461] Gallagher, R. R., Li, Z., Lewis, A. O. and Isaacs, F. J. Rapid editing and evolution of bacterial genomes using libraries of synthetic DNA. *Nat. Protoc.* **9,** 2301–2316 (2014).

[462] Liu, R., Bassalo, M. C., Zeitoun, R. I. and Gill, R. T. Genome scale engineering techniques for

metabolic engineering. *Metab. Eng.* **32,** 143–154 (2015).

[463] Silhavy, T. J., Kahne, D. and Walker, S. The Bacterial Cell Envelope. *Cold Spring Harb Perspect Biol* **2,** 1–16 (2010).

[464] Raetz, C. R. H. and Dowhan, W. Biosynthesis and function of phospholipids in Escherichia coli. *J. Biol. Chem.* **265,** 1235–1238 (1990).

[465] Sohlenkamp, C. and Geiger, O. Bacterial membrane lipids: Diversity in structures and pathways. *FEMS Microbiol. Rev.* **40,** 133–159 (2016).

[466] Vollmer, W. and Bertsche, U. Murein (peptidoglycan) structure, architecture and biosynthesis in Escherichia coli. *Biochim. Biophys. Acta - Biomembr.* **1778,** 1714–1734 (2008).

[467] Garde, S., Chodisetti, P. K. and Reddy, M. Peptidoglycan: Structure, Synthesis, and Regulation. *EcoSal Plus* **9,** 1–35 (2021).

[468] Gan, L., Chen, S. and Jensen, G. J. Molecular organization of Gram-negative peptidoglycan. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 18953–18957 (2008).

[469] Sun, J., Rutherford, S. T., Silhavy, T. J. and Huang, K. C. Physical properties of the bacterial outer membrane. *Nat. Rev. Microbiol.* **20,** 236–248 (2022).

[470] Mansilla, M. C., Cybulski, L. E., Albanesi, D. and de Mendoza, D. Control of membrane lipid fluidity by molecular thermosensors. *J. Bacteriol.* **186,** 6681–6688 (2004).

[471] Parsons, J. B. and Rock, C. O. Bacterial lipids: Metabolism and membrane homeostasis. *Prog. Lipid Res.* **52,** 249–276 (2013).

[472] Loll, P. J. and Axelsen, P. H. The Structural Biology of Molecular Recognition by Vancomycin. *Annu. Rev. Biophys. Biomol. Struct.* **29,** 265–289 (2000).

[473] Reynolds, P. E. Structure, biochemistry and mechanism of action of glycopeptide antibiotics. *Eur. J. Clin. Microbiol. Infect. Dis.* **8,** 943–950 (1989).

[474] Walsh, C. Molecular mechanisms that confer antibacterial drug resistance. *Nature* **406,** 775–781 (2000).

[475] Nikaido, H. Molecular basis of bacterial outer membrane permeability. *Microbiol. Mol. Biol. Rev.* **67,** 593–656 (2003).

[476] Pagès, J.-M., James, C. E. and Winterhalter, M. The porin and the permeating antibiotic: A selective diffusion barrier in Gram-negative bacteria. *Nat. Rev. Microbiol.* **6,** 893–903 (2008).

[477] Muheim, C., Götzke, H., Eriksson, A. U., Lindberg, S., Lauritsen, I., Nørholm, M. H. H. and Daley, D. O. Increasing the permeability of Escherichia coli using MAC13243. *Sci. Rep.* **7,** 17629

(2017).

[478] Zhang, B., Phetsang, W., Stone, M. R. L., Kc, S., Butler, M. S., Cooper, M. A., Elliott, A. G., Łapińska, U., Voliotis, M., Tsaneva-Atanasova, K., Pagliara, S. and Blaskovich, M. A. T. Synthesis of vancomycin fluorescent probes that retain antimicrobial activity, identify Gram-positive bacteria, and detect Gram-negative outer membrane damage. *Commun. Biol.* **6,** 409 (2023).

[479] Stokes, J. M., French, S., Ovchinnikova, O. G., Bouwman, C., Whitfield, C. and Brown, E. D. Cold Stress Makes Escherichia coli Susceptible to Glycopeptide Antibiotics by Altering Outer Membrane Integrity. *Cell Chem. Biol.* **23,** 267–277 (2016).

[480] Sutterlin, H. A., Zhang, S. and Silhavya, T. J. Accumulation of phosphatidic acid increases vancomycin resistance in Escherichia coli. *J. Bacteriol.* **196,** 3214–3220 (2014).

[481] Lessard, I. A. D. and Walsh, C. T. VanX, a bacterial D-alanyl-D-alanine dipeptidase: Resistance, immunity, or survival function? *Proc. Natl. Acad. Sci. U. S. A.* **96,** 11028–11032 (1999).

[482] Masi, M., Réfregiers, M., Pos, K. M. and Pagès, J.-M. Mechanisms of envelope permeability and antibiotic influx and efflux in Gram-negative bacteria. *Nat. Microbiol.* **2,** 17001 (2017).

[483] Greenspan, P., Mayer, E. P. and Fowler, S. D. Nile red: A selective fluorescent stain for intracellular lipid droplets. *J. Cell Biol.* **100,** 965–973 (1985).

[484] Westfall, C. S. and Levin, P. A. Comprehensive analysis of central carbon metabolism illuminates connections between nutrient availability, growth rate, and cell morphology in Escherichia coli. *PLoS Genet.* **14,** e1007205 (2018).

[485] Wehrens, M., Ershov, D., Rozendaal, R., Walker, N., Schultz, D., Kishony, R., Levin, P. A. and Tans, S. J. Size Laws and Division Ring Dynamics in Filamentous Escherichia coli cells. *Curr. Biol.* **28,** 972–979 (2018).

[486] Jones, T. H., Vail, K. M. and McMullen, L. M. Filament formation by foodborne bacteria under sublethal stress. *Int. J. Food Microbiol.* **165,** 97–110 (2013).

[487] Power, A. L., Barber, D. G., Groenhof, S. R. M., Wagley, S., Liu, P., Parker, D. A. and Love, J. The Application of Imaging Flow Cytometry for Characterisation and Quantification of Bacterial Phenotypes. *Front. Cell. Infect. Microbiol.* **11,** 716592 (2021).

[488] Tian, D., Wang, C., Liu, Y., Zhang, Y., Caliari, A., Lu, H., Xia, Y., Xu, B., Xu, J. and Yomo, T. Cell Sorting-Directed Selection of Bacterial Cells in Bigger Sizes Analyzed by Imaging Flow Cytometry during Experimental Evolution. *Int. J. Mol. Sci.* **24,** 3243 (2023).

[489] Kuthning, A., Durkin, P., Oehm, S., Hoesl, M. G., Budisa, N. and Süssmuth, R. D. Towards Biocontained Cell Factories: An Evolutionarily Adapted Escherichia coli Strain Produces a New-to-nature Bioactive Lantibiotic Containing Thienopyrrole-Alanine. *Sci. Rep.* **6,** 33447 (2016).

[490] Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67,** 509–544 (1998).

[491] Day, R. N. and Davidson, M. W. The fluorescent protein palette: Tools for cellular imaging. *Chem. Soc. Rev.* **38,** 2887–2921 (2009).

[492] Royant, A. and Noirclerc-Savoye, M. Stabilizing role of glutamic acid 222 in the structure of Enhanced Green Fluorescent Protein. *J. Struct. Biol.* **174,** 385–390 (2011).

[493] Lelimousin, M., Noirclerc-Savoye, M., Lazareno-Saez, C., Paetzold, B., Le Vot, S., Chazal, R., Macheboeuf, P., Field, M. J., Bourgeois, D. and Royant, A. Intrinsic dynamics in ECFP and cerulean control fluorescence quantum yield. *Biochemistry* **48,** 10038–10046 (2009).

[494] Lundqvist, M., Thalén, N., Volk, A.-L., Hansen, H. G., von Otter, E., Nygren, P.-Å., Uhlen, M. and Rockberg, J. Chromophore pre-maturation for improved speed and sensitivity of split-GFP monitoring of protein secretion. *Sci. Rep.* **9,** 310 (2019).

[495] Budisa, N., Pal, P. P., Alefelder, S., Birle, P., Krywcun, T., Rubini, M., Wenger, W., Bae, J. H. and Steiner, T. Probing the role of tryptophans in Aequorea victoria green fluorescent proteins with an expanded genetic code. *Biol. Chem.* **385,** 191–202 (2004).

[496] Treiber-Kleinke, C., Berger, A., Adrian, L., Budisa, N. and Koksch, B. Escherichia coli adapts metabolically to 6- and 7-fluoroindole, enabling proteome-wide fluorotryptophan substitution. *bioRxiv* 1–24 (2023). doi:https://doi.org/10.1101/2023.09.25.559291

[497] Tobola, F., Lelimousin, M., Varrot, A., Gillon, E., Darnhofer, B., Blixt, O., Birner-Gruenberger, R., Imberty, A. and Wiltschi, B. Effect of Noncanonical Amino Acids on Protein-Carbohydrate Interactions: Structure, Dynamics, and Carbohydrate Affinity of a Lectin Engineered with Fluorinated Tryptophan Analogs. *ACS Chem. Biol.* **13,** 2211–2219 (2018).

[498] Kenward, C., Shin, K. and Rainey, J. K. Mixed Fluorotryptophan Substitutions at the Same Residue Expand the Versatility of 19F Protein NMR Spectroscopy. *Chem. - A Eur. J.* **24,** 3391–3396 (2018).

[499] Zlatopolskiy, B. D., Zischler, J., Schäfer, D., Urusova, E. A., Guliyev, M., Bannykh, O., Endepols, H. and Neumaier, B. Discovery of 7-[18F]Fluorotryptophan as a Novel Positron Emission Tomography (PET) Probe for the Visualization of Tryptophan Metabolism in Vivo. *J. Med. Chem.* **61,** 189–206 (2018).

[500] Maleckis, A., Herath, I. D. and Otting, G. Synthesis of13C/19F/2H labeled indoles for use as tryptophan precursors for protein NMR spectroscopy. *Org. Biomol. Chem.* **19,** 5133–5147 (2021).

[501] Atashgahi, S., Sánchez-Andrea, I., Heipieper, H. J., van der Meer, J. R., Stams, A. J. M. and Smidt, H. Prospects for harnessing biocide resistance for bioremediation and detoxification. *Science (80-.).* **360,** 743–746 (2018).

[502] Lee, M. and Phillips, R. S. The mechanism of Escherichia coli tryptophan indole-lyase: substituent effects on steady-state and pre-steady-state kinetic parameters for aryl-substituted tryptophan derivatives. *Bioorganic Med. Chem.* **3,** 195–205 (1995).

[503] Alkhalaf, L. M. and Ryan, K. S. Biosynthetic manipulation of tryptophan in bacteria: Pathways and mechanisms. *Chem. Biol.* **22,** 317–328 (2015).

[504] Kurnasov, O., Jablonski, L., Polanuyer, B., Dorrestein, P., Begley, T. and Osterman, A. Aerobic tryptophan degradation pathway in bacteria: Novel kynurenine formamidase. *FEMS Microbiol. Lett.* **227,** 219–227 (2003).

[505] Han, Q., Fang, J. and Li, J. Kynurenine aminotransferase and glutamine transaminase K of Escherichia coli: Identity with aspartate aminotransferase. *Biochem. J.* **360,** 617–623 (2001).

[506] Bygd, M. D., Aukema, K. G., Richman, J. E. and Wackett, L. P. Microwell Fluoride Screen for Chemical, Enzymatic, and Cellular Reactions Reveals Latent Microbial Defluorination Capacity for -CF3 Groups. *Appl. Environ. Microbiol.* **88,** 1–17 (2022).

[507] Bergmans, H. E. N., van Die, I. M. and Hoekstra, W. P. M. Transformation in Escherichia coli: stages in the process. *J. Bacteriol.* **146,** 564–570 (1981).

[508] Hanahan, D. Studies on transformation of Escherichia coli with plasmids. *J. Mol. Biol.* **166,** 557–580 (1983).

[509] Dower, W. J., Miller, J. F. and Ragsdale, C. W. High efficiency transformation of E.coli by high voltage electroporation. *Nucleic Acids Res.* **16,** 6127–6145 (1988).

[510] Birnboim, H. C. and Doly, J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7,** 1513–1523 (1979).

[511] Miller, S. A., Dykes, D. D. and Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16,** 1215 (1988).

[512] Miles, A. A., Misra, S. S. and Irwin, J. O. The estimation of the bactericidal power of blood. *J. Hyg. (Lond).* **38,** 732–749 (1938).

[513] Maier, R. M. and Pepper, I. L. *Bacterial Growth. Environ. Microbiol. Third Ed.* (Elsevier Inc., 2015). doi:10.1016/B978-0-12-394626-3.00003-X

[514] Tominaga, H., Ishiyama, M., Ohseto, F., Sasamoto, K., Hamamoto, T., Suzuki, K. and Watanabe, M. A water-soluble tetrazolium salt useful for colorimetric cell viability assay. *Anal. Commun.* **36,** 47–50 (1999).

[515] Braissant, O., Astasov-Frauenhoffer, M., Waltimo, T. and Bonkat, G. A Review of Methods to Determine Viability, Vitality, and Metabolic Rates in Microbiology. *Front. Microbiol.* **11,** 547458

(2020).

[516]   Park, S. Y. and Kim, C. G. A comparative study of three different viability tests for chemically or thermally inactivated Escherichia coli. *Environ. Eng. Res.* **23,** 282–287 (2018).

[517]   Berney, M., Hammes, F., Bosshard, F., Weilenmann, H.-U. and Egli, T. Assessment and interpretation of bacterial viability by using the LIVE/DEAD BacLight kit in combination with flow cytometry. *Appl. Environ. Microbiol.* **73,** 3283–3290 (2007).

[518]   Berridge, M. V., Herst, P. M. and Tan, A. S. Tetrazolium dyes as tools in cell biology: New insights into their cellular reduction. *Biotechnol. Annu. Rev.* **11,** 127–152 (2005).

[519]   Gajic, I., Kabic, J., Kekic, D., Jovicevic, M., Milenkovic, M., Mitic Culafic, D., Trudic, A., Ranin, L. and Opavski, N. Antimicrobial Susceptibility Testing: A Comprehensive Review of Currently Used Methods. *Antibiotics* **11,** 427 (2022).

[520]   Khan, Z. A., Siddiqui, M. F. and Park, S. Current and Emerging Methods of Antibiotic Susceptibility Testing. *Diagnostics* **9,** 49 (2019).

[521]   Wiegand, I., Hilpert, K. and Hancock, R. E. W. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat. Protoc.* **3,** 163–175 (2008).

[522]   European Committee for Antimicrobial Susceptibility Testing (EUCAST) of the European Society of Clinical Microbiology and Infectious Diseases (ESCMID). Determination of minimum inhibitory concentrations (MICs) of antibacterial agents by broth dilution. *Clin. Microbiol. Infect.* **9,** 1–7 (2003).

[523]   Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* **51,** 263–273 (1986).

[524]   Mullis, K. B. The polymerase chain reaction (Nobel Lecture). *Angew. Chem Int. Ed. Engl.* **33,** 1209–1213 (1994).

[525]   Green, M. R. and Sambrook, J. Polymerase chain reaction. *Cold Spring Harb. Protoc.* **2019,** 436–456 (2019).

[526]   Armstrong, J. A. and Schulz, J. R. Agarose gel electrophoresis. *Curr. Protoc. Essent. Lab. Tech.* **10,** 7.2.1-7.2. (2015).

[527]   Green, M. R. and Sambrook, J. Analysis of DNA by agarose gel electrophoresis. *Cold Spring Harb. Protoc.* **2019,** 6–15 (2019).

[528]   Huang, Q., Baum, L. and Fu, W.-L. Simple and practical staining of DNA with GelRed in agarose

gel electrophoresis. *Clin. Lab.* **56,** 149–152 (2010).

[529] Haines, A. M., Tobe, S. S., Kobus, H. J. and Linacre, A. Properties of nucleic acid staining dyes used in gel electrophoresis. *Electrophoresis* **36,** 941–944 (2015).

[530] Motohashi, K. Development of highly sensitive and low-cost DNA agarose gel electrophoresis detection systems , and evaluation of non-mutagenic and loading dye-type DNA-staining reagents. *PLoS One* **14,** e0222209 (2019).

[531] Boom, R., Sol, C. J. A., Salimans, M. M. M., Jansen, C. L., Wertheim-Van Dillen, P. M. E. and Van Der Noordaa, J. Rapid and simple method for purification of nucleic acids. *J. Clin. Microbiol.* **28,** 495–503 (1990).

[532] Vogelstein, B. and Gillespie, D. Preparative and analytical purification of DNA from agarose. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 615–619 (1979).

[533] Sanger, F., Nicklen, S. and Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467 (1977).

[534] Sternberg, N. and Maurer, R. Bacteriophage-Mediated Generalized Transduction in Escherichia coli and Salmonella typhimurium. *Methods Enzymol.* **204,** 18–43 (1991).

[535] Thomason, L. C., Costantino, N. and Court, D. L. E. coli Genome Manipulation by P1 Transduction. *Curr. Protoc. Mol. Biol.* **79,** 1.17.1-8 (2007).

[536] Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L. and Mori, H. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol. Syst. Biol.* **2,** 1–11 (2006).

[537] Cherepanov, P. P. and Wackernagel, W. Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158,** 9–14 (1995).

[538] Quan, S., Skovgaard, O., McLaughlin, R. E., Buurman, E. T. and Squires, C. L. Markerless Escherichia coli rrn deletion strains for genetic determination of ribosomal binding sites. *G3 Genes, Genomes, Genet.* **5,** 2555–2557 (2015).

[539] Näsvall, J. Direct and Inverted Repeat stimulated excision (DIRex): Simple, single-step, and scar-free mutagenesis of bacterial genes. *PLoS One* **12,** e0184126 (2017).

[540] Wang, K., Fredens, J., Brunner, S. F., Kim, S. H., Chia, T. and Chin, J. W. Defining synonymous codon compression schemes by genome recoding. *Nature* **539,** 59–64 (2016).

[541] Robertson, W. E., Funke, L. F. H., de la Torre, D., Fredens, J., Wang, K. and Chin, J. W. Creating custom synthetic genomes in Escherichia coli with REXER and GENESIS. *Nat. Protoc.* **16,** 2345–

2380 (2021).

[542] Engler, C., Kandzia, R. and Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3,** e3647 (2008).

[543] Weber, E., Engler, C., Gruetzner, R., Werner, S. and Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS One* **6,** e16765 (2011).

[544] Yosef, I., Goren, M. G. and Qimron, U. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res.* **40,** 5569–5576 (2012).

[545] Horvath, P. and Barrangou, R. CRISPR/Cas, the immune system of Bacteria and Archaea. *Science (80-.).* **327,** 167–170 (2010).

[546] Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.).* **337,** 816–821 (2012).

[547] Sauvagère, S. and Siatka, C. CRISPR-Cas: 'The Multipurpose Molecular Tool' for Gene Therapy and Diagnosis. *Genes (Basel).* **14,** 1542 (2023).

[548] Mougiakos, I., Bosma, E. F., de Vos, W. M., van Kranenburg, R. and van der Oost, J. Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit. *Trends Biotechnol.* **34,** 575–587 (2016).

[549] Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S. and Kim, J. S. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24,** 132–141 (2014).

[550] Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K. and Sander, J. D. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31,** 822–826 (2013).

[551] Murphy, K. C. λ Recombination and Recombineering. *EcoSal Plus* **7,** 1–70 (2016).

[552] Manen, D. and Caro, L. The replication of plasmid pSC101. *Mol. Microbiol.* **5,** 233–237 (1991).

[553] Porath, J., Carlsson, J., Olsson, I. and Belfrage, G. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **258,** 598–599 (1975).

[554] Hearon, J. Z. The configuration of cobaltodihistidine and oxy-bis(cobaltodihistidine). *J. Natl. cancer Inst.* **9,** 1–11 (1948).

[555] Hochuli, E., Döbeli, H. and Schacher, A. New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *J. Chromatogr.* **411,** 177–184 (1987).

[556] Block, H., Maertens, B., Spriestersbach, A., Brinker, N., Kubicek, J., Fabis, R., Labahn, J. and Schäfer, F. in *Methods Enzymol.* **463,** 439–473 (2009).

[557] Knecht, S., Ricklin, D., Eberle, A. N. and Ernst, B. Oligohis-tags: Mechanisms of binding to Ni2+-NTA surfaces. *J. Mol. Recognit.* **22,** 270–279 (2009).

[558] Gallagher, S. R. SDS-polyacrylamide gel electrophoresis (SDS-PAGE). *Curr. Protoc. Essent. Lab. Tech.* **2012,** 7.3.1-7.3.28 (2012).

[559] Miles, E. W., Kawasaki, H., Ahmed, S. A., Morita, H., Morita, H. and Nagata, S. The ß Subunit of Tryptophan Synthase. *J. Biol. Chem.* **264,** 6280–6287 (1989).

[560] Cai, L. Thin layer chromatography. *Curr. Protoc. Essent. Lab. Tech.* **8,** 6.3.1-6.3.18 (2014).

[561] Bradford, M. M. A Rapid and Sensitive Method for Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* **72,** 248–254 (1976).

[562] Kublik, A., Deobald, D., Hartwig, S., Schiffmann, C. L., Andrades, A., von Bergen, M., Sawers, R. G. and Adrian, L. Identification of a multi-protein reductive dehalogenase complex in Dehalococcoides mccartyi strain CBDB1 suggests a protein-dependent respiratory electron transport chain obviating quinone involvement. *Environ. Microbiol.* **18,** 3044–3056 (2016).

[563] Seidel, K., Kühnert, J. and Adrian, L. The complexome of Dehalococcoides mccartyi reveals its organohalide respiration-complex is modular. *Front. Microbiol.* **9,** 1130 (2018).

[564] Blattner, F. R., Plunkett III, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. and Shao, Y. The complete genome sequence of Escherichia coli K-12. *Science (80-.).* **277,** 1453–1462 (1997).

[565] Tatusov, R. L., Galperin, M. Y., Natale, D. A. and Koonin, E. V. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28,** 33–36 (2000).

List of Figures

List of Tables

Appendix

Purity analysis of NMM medium controls

Figure S63 | HPLC-UV/FL analysis of NMM controls. The UV (left) and FL (right) chromatograms of A) NMM0-0-0, B) NMM0-70-0 (6Fi) and C) NMM0-70-0 (7Fi) are shown. The potential occurrence of Ind and Trp impurities are indicated as red arrows, respectively. Medium samples were not subjected to the hydrolysis protocol.

NMM composition during adaptation process

Table S20| NMM composition during the adaptation of the 6TUB128-OC lineage. NMM is designated as follows: NMM(a-b-c) where "a" corresponds to the amino acid supply, "b" to the 6-fluoroindole (6Fi) concentration [µM] and "c" to the indole (Ind) concentration [µM]. Canonical amino acids (cAAs), except for Trp, were supplied as mixture (cAA19 mix, containing Tyr, Phe, Cys, Gly, Ser, Leu, Ile, Val, Ala, Asp, Asn, Glu, Gln, His, Met, Pro, Arg, Lys, Thr) in an initial concentration of 50 mg/L, which was gradually reduced in ALE phase 2 as depicted. Below 50 mg/L the cAA concentration is given in square brackets. The passages cultivated in the respective medium composition are provided.

Medium (NMM)	cAA19 mixture [mg/L]	Ind [µM]	6Fi [µM]	Passages
19-70-2.5	50	2.5	70	1-3
19-70-1.0	50	1.0	70	4-5
19-70-0.5	50	0.5	70	6-18
19-70-0.25	50	0.25	70	19-25
19-70-0.10	50	0.10	70	26-35
19-70-0.05	50	0.05	70	36-40
19-70-0.01	50	0.01	70	41-44
19-70-0	50	0	70	45-53
[40]-19-70-0	40	0	70	54-58
[30]-19-70-0	30	0	70	59-62
[20]-19-70-0	20	0	70	63-66
[10]-19-70-0	10	0	70	67-71
[7.5]-19-70-0	7.5	0	70	72-77
[5.0]-19-70-0	5.0	0	70	78-81
[2.5]-19-70-0	2.5	0	70	82-84
0-70-0	0	0	70	85-128

Table S21 | NMM composition during the adaptation of the 6TUB165-MB4 and 6TUB165-MB3 lineages. NMM is designated as follows: NMM(a-b-c) where "a" corresponds to the amino acid supply, "b" to the 6-fluoroindole (6Fi) concentration [µM] and "c" to the indole (Ind) concentration [µM]. Canonical amino acids (cAAs; F, Y, A, V, L, I, H, G, S, C, K, D, N, M, T, P, E, Q, R), except for Trp, were supplied in separate stock solutions (50 mg/L) and removed as metabolic blocks in ALE phase 2. For 6TUB165-MB4 and 6TUB165-MB3 different routes for the cAA elimination were used as depicted. The passages cultivated in the respective medium composition are provided.

Medium (NMM)	cAAs supplied	Ind [µM]	6Fi [µM]	Passages
19-70-2.5	FY AVLI H GSC KDNMT PEQR	2.5	70	1-3
19-70-1.0	FY AVLI H GSC KDNMT PEQR	1.0	70	4-5
19-70-0.5	FY AVLI H GSC KDNMT PEQR	0.5	70	6-18
19-70-0.25	FY AVLI H GSC KDNMT PEQR	0.25	70	19-25
19-70-0.10	FY AVLI H GSC KDNMT PEQR	0.10	70	26-35
19-70-0.05	FY AVLI H GSC KDNMT PEQR	0.05	70	36-40
19-70-0.01	FY AVLI H GSC KDNMT PEQR	0.01	70	41-44
19-70-0	FY AVLI H GSC KDNMT PEQR	0	70	45-53
17-70-0	AVLI H GSC KDNMT PEQR	0	70	54-59
13-70-0	H GSC KDNMT PEQR	0	70	60-63
12-70-0	GSC KDNMT PEQR	0	70	64-67
9-70-0	KDNMT PEQR	0	70	68-72
6TUB165-MB4				
6-70-0	KD PEQR	0	70	73-77
4-70-0	PEQR	0	70	78-81
0-70-0	0	0	70	82-165
6TUB165-MB3				
5-70-0	KDMNT	0	70	73-77
3-70-0	NMT	0	70	78-81
0-70-0	0	0	70	82-165

Table S22| NMM composition during the adaptation of the 7TUB165-OC lineage. NMM is designated as follows: NMM(a-b-c) where "a" corresponds to the amino acid supply, "b" to the 7-fluoroindole (7Fi) concentration [µM] and "c" to the indole (Ind) concentration [µM]. Canonical amino acids (cAAs), except for Trp, were supplied as mixture (cAA19 mix, containing Tyr, Phe, Cys, Gly, Ser, Leu, Ile, Val, Ala, Asp, Asn, Glu, Gln, His, Met, Pro, Arg, Lys, Thr) in an initial concentration of 50 mg/L, which was gradually reduced in ALE phase 2 as depicted. Below 50 mg/L the cAA concentration is given in square brackets. The passages cultivated in the respective medium composition are provided. During this adaptation the medium composition need several times relaxed (reduction of 7Fi concentration, increase of cAA supply) to uphold cell growth.

Medium (NMM)	cAA19 mixture [mg/L]	Ind [µM]	7Fi [µM]	Passages
19-70-2.5	50	2.5	70	1-2
19-70-1.0	50	1.0	70	3-6
19-70-0.5	50	0.5	70	7-14
19-70-0.25	50	0.25	70	15-18
19-70-0.10	50	0.10	70	19-26
19-70-0.05	50	0.05	70	27-35
19-70-0.01	50	0.01	70	36-41
19-70-0	50	0	70	42-53
[40]-19-70-0	40	0	70	54-58
[30]-19-70-0	30	0	70	59-62
[20]-19-70-0	20	0	70	63-66
[10]-19-70-0	10	0	70	67-71
[7.5]-19-70-0	7.5	0	70	72-80
[5.0]-19-70-0	5.0	0	70	81-83
[10]-19-50-0	10	0	50	84-88
[5.0]-19-50-0	5.0	0	50	89-92
[2.5]-19-50-0	2.5	0	50	93-100
[2.5]-19-30-0	2.5	0	30	101-102
[10]-19-30-0	10	0	30	103-104
[5.0]-19-30-0	5.0	0	30	105-106
[2.5]-19-30-0	2.5	0	30	107-109
[1.0]-19-30-0	1.0	0	30	110-111
[1.5]-19-30-0	1.5	0	30	112-125
[1.0]-19-30-0	1.0	0	30	126-128
[0.5]-19-30-0	0.5	0	30	129-131
0-30-0	0	0	30	132-143
0-70-0	0	0	70	144-165

Table S23 | NMM composition during the adaptation of the 7TUB165-MB lineage. NMM is designated as follows: NMM(a-b-c) where "a" corresponds to the amino acid supply, "b" to the 7-fluoroindole (7Fi) concentration [µM] and "c" to the indole (Ind) concentration [µM]. Canonical amino acids (cAAs; F, Y, A, V, L, I, H, G, S, C, K, D, N, M, T, P, E, Q, R), except for Trp, were supplied in separate stock solutions (50 mg/L) and removed as metabolic blocks or else one by one in ALE phase 2. The passages cultivated in the respective medium composition are provided. The medium composition of this adaption also had to be relaxed multiple times by decreasing the 7Fi concentration and increasing the number of amino acids supplied.

Medium (NMM)	cAAs supplied	Ind [µM]	7Fi [µM]	Passages
19-70-2.5	FY AVLI H GSC KDNMT PEQR	2.5	70	1-2
19-70-1.0	FY AVLI H GSC KDNMT PEQR	1.0	70	3-6
19-70-0.5	FY AVLI H GSC KDNMT PEQR	0.5	70	7-14
19-70-0.25	FY AVLI H GSC KDNMT PEQR	0.25	70	15-18
19-70-0.10	FY AVLI H GSC KDNMT PEQR	0.10	70	19-26
19-70-0.05	FY AVLI H GSC KDNMT PEQR	0.05	70	27-35
19-70-0.01	FY AVLI H GSC KDNMT PEQR	0.01	70	36-41
19-70-0	FY AVLI H GSC KDNMT PEQR	0	70	42-53
17-70-0	AVLI H GSC KDNMT PEQR	0	70	54-58
12-70-0	GSC KDNMT PEQR	0	70	59-83
12-50-0	GSC KDNMT PEQR	0	50	84-92
11-50-0	GSC DNMT PEQR	0	50	93-97
10-50-0	GSC NMT PEQR	0	50	98-100
10-30-0	GSC NMT PEQR	0	30	101-102
13-30-0	H GSC KDNMT PEQR	0	30	103-104
12-30-0	GSC KDNMT PEQR	0	30	105-106
11-30-0	GSC DNMT PEQR	0	30	107-108
10-30-0	GSC DMT PEQR	0	30	109-113
9-30-0	GSC DM PEQR	0	30	114-115
8-30-0	GSC M PEQR	0	30	116-117
7-30-0	SC M PEQR	0	30	118-120
6-30-0	S M PEQR	0	30	121-125
5-30-0	M PEQR	0	30	126-128
4-30-0	M PEQ	0	30	129-131
3-30-0	M EQ	0	30	132-133
2-30-0	M E	0	30	134-135
1-30-0	M	0	30	136-150
0-30-0	0	0	30	151-154
0-70-0	0	0	70	155-165

Summary of ALE parameters

Table S24 | Overview of the technical details of the adaption trajectories. Attributes passage number (pass.), number of generations (gen.) and cultivation time in days for every ALE phase (P1-3) and the entire adaption process are shown.

		Entire adaptation			P1 *(indole depletion)*			P2 *(cAA removal)*			P3 *final medium*		
		pass.	gen.	days	pass.	Gen.	days	pass.	gen.	days	pass.	gen.	days
W-TUB165	W	165	1273	413.8									
Ind-TUB165	Ind	165	1270	413.6									
6TUB128-OC	6Fi	128	780	317.2	53	301	128.8	31	205	77.2	44	274	111.2
6TUB165-MB4	6Fi	165	1024	410.1	53	301	128.8	28	179	71.7	84	544	209.6
6TUB165-MB3	6Fi	165	1024	410.1	53	301	128.8	28	179	71.7	84	544	209.6
7TUB165-OC	7Fi	165	778	403.0	53	301	122.1	79	330	198.4	33	147	82.5
7TUB165-MB	7Fi	165	752	401.0	53	301	122.1	98	397	243.9	14	54	35.0

Comparability of growth curves and parameters in macroscale (flask) or microscale (plate reader)
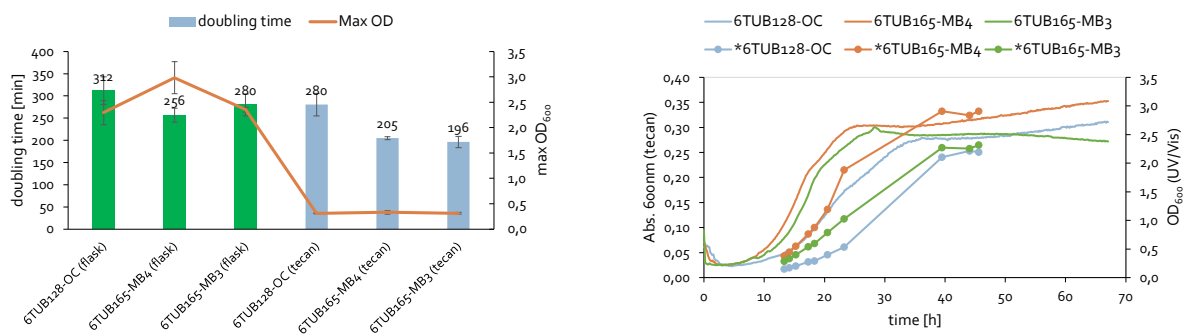
Figure S64 | Comparability of growth curves measured in macroscale and microscale. Comparison of the growth parameters doubling time and maximum cell density determined using either the macroscale growth (10 mL culture in 100 mL shaking flasks) or the microscale growth (200 µL in 96-well plates) are shown left. Representative growth curves obtained with both approaches are shown right. 7TUBX strains were excluded from this as analysis because of their poor growth.

Comprehensive growth analysis

(A) TUB00

(B) W-TUB165

(C) Ind-TUB165
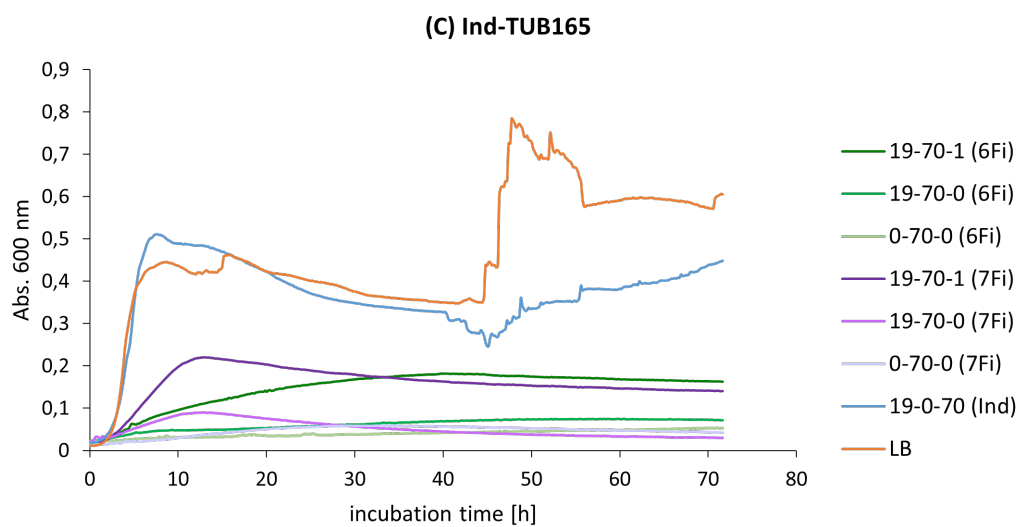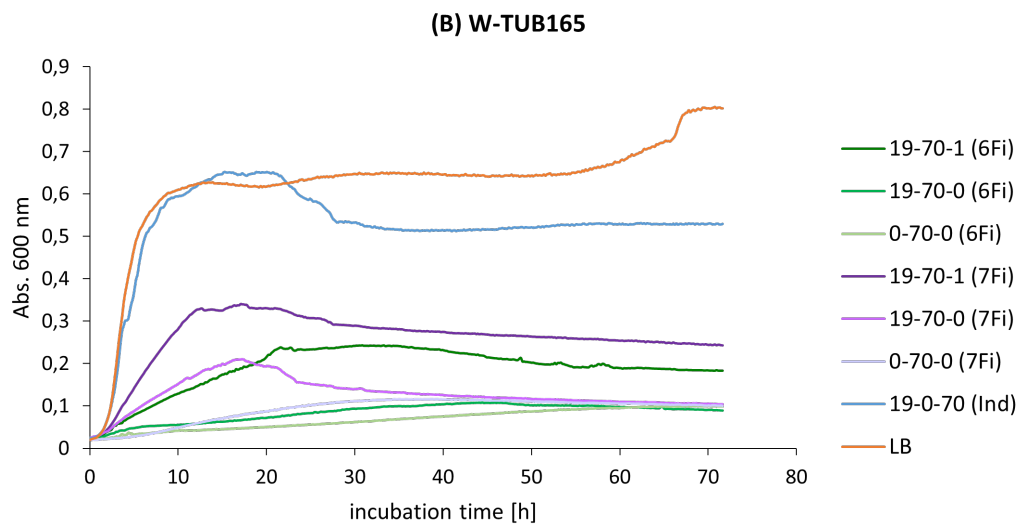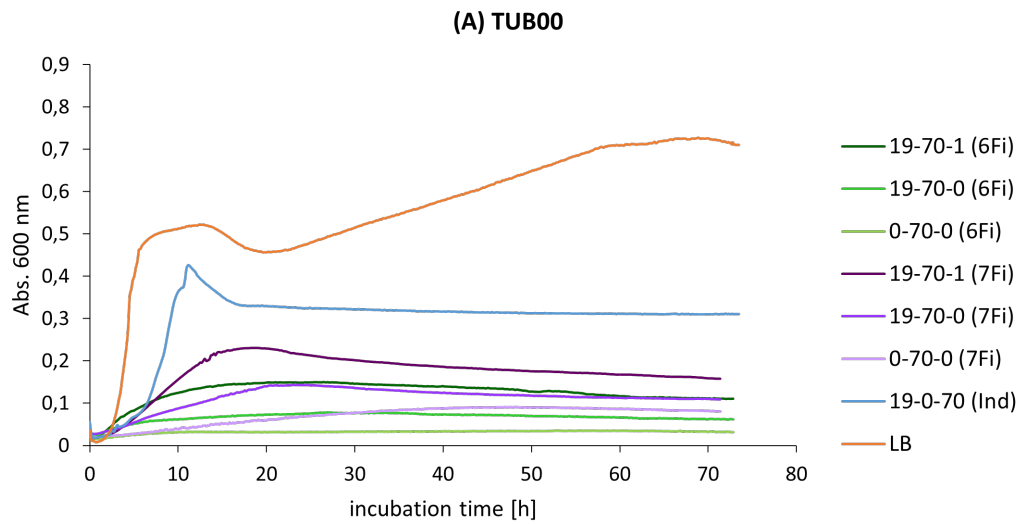
(D) 6TUB128-OC
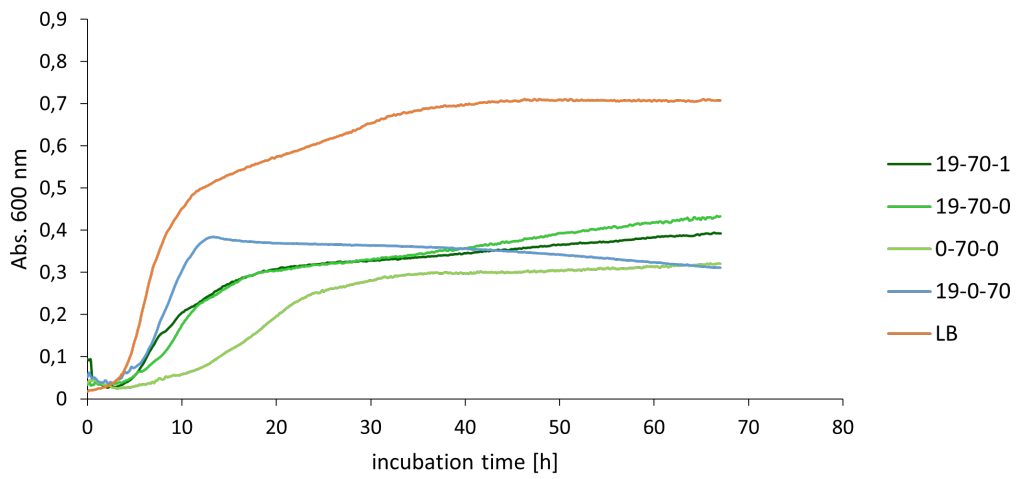
(E) 6TUB165-MB4

(F) 6TUB165-MB3

(G) 7TUB165-OC
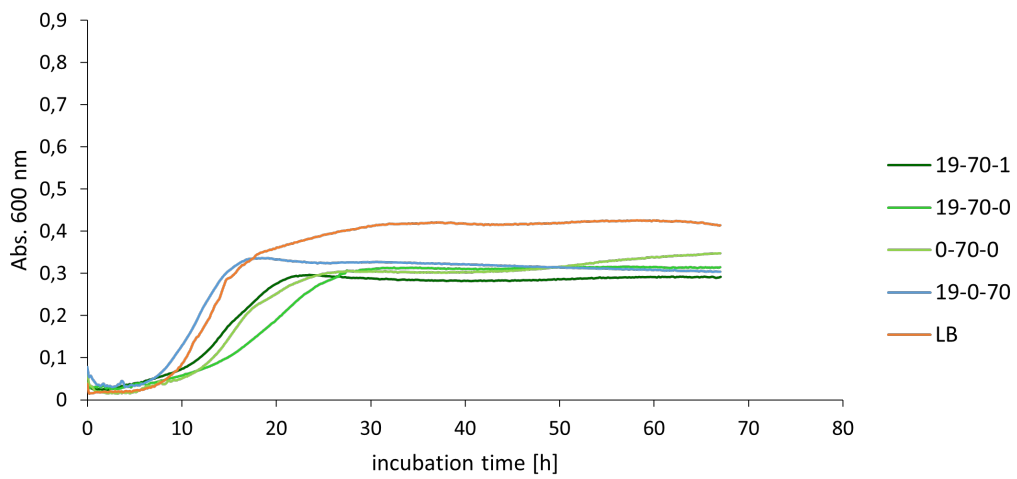
(H) 7TUB165-MB
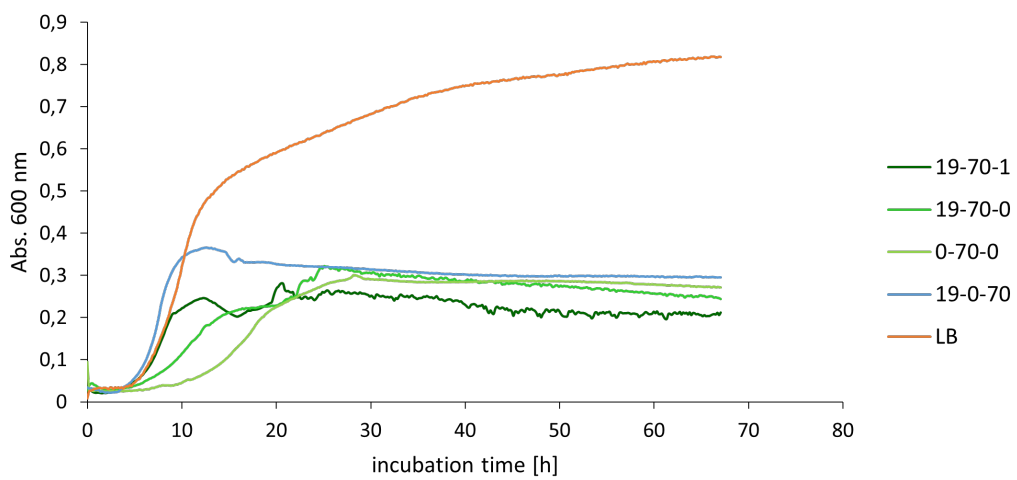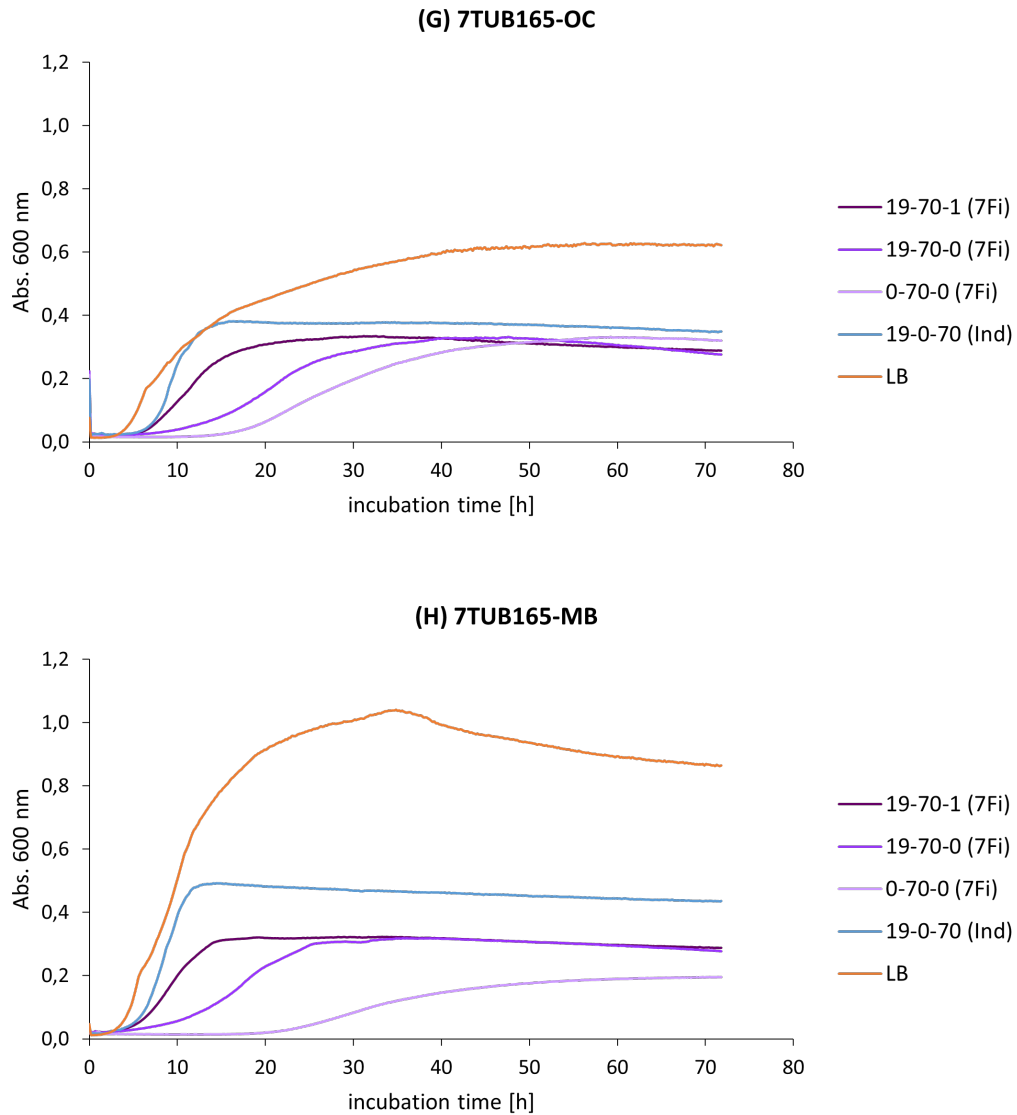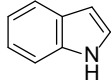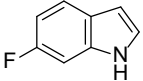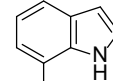
Ancestral (rich)	"6Fi trajectory"	"7Fi trajectory"
— 19-0-70 (Ind)	— 19-70-1 (6Fi)	— 19-70-1 (7Fi)
— LB	— 19-70-0 (6Fi)	— 19-70-0 (7Fi)
	— 0-70-0 (6Fi)	— 0-70-0 (7Fi)

Figure S65 | Growth curves of the ancestral strain and positive controls (A - C), the 6Fi-adapted lineages (D - F) and the 7Fi-adapted lineages (G - H); the growth of final isolates respectively is shown. The absorption at 600 nm is plotted against the incubation time [h] at 30 °C. Growth curves gained in LB are colored orange, in NMM19-0-70 (Ind) are blue, in 6Fi-supplemented NMM are in green shades and in 7Fi-supplemented NMM are colored in purple shades (see legend). The frizzy and instable curve parts obtained in "C" (LB and NMM19-0-70 (Ind)) and "F" (NMM19-70-1 (6Fi)) are rated as negligible, since for estimation of the growth behavior and calculation of the growth parameters only the exponential part of the curve is of interest. All measurements were performed at microscale level in triplicates using a plate reader.

Determination of the (F)-Ind and (F)-Trp content in supernatant samples of final ALE isolates and TUB00

Table S25 | (F)-Ind and (F)-Trp content in supernatant samples of the final ALE isolates and TUB00. The amount of Ind, Trp, 6Fi, 6FTrp, 7Fi and 7FTrp in supernatant samples was determined. Samples were taken after 24h, 48h, 72h, 96h and 120h cultivation and analyzed by HPLC-UV/FL. Values are given as integrated FL peak areas and in case of 7TUB165-X also as UV peak areas. Peak areas correspond to the following concentrations: TUB00: 0.3 – 7.6 µM Ind and 0.7 – 2.5 µM Trp, W-TUB165: < 0.01 µM Trp, Ind-TUB165: 0.01 – 0.2 µM Trp, 6TUB128-OC: 23.9 µM 6Fi and 0.1 – 3.0 µM 6FTrp, 6TUB165-MB4: 0.3 µM 6Fi and 0.2 – 1.2 µM 6FTrp, 6TUB165-MB3: 0.6 µM 6Fi and 1.0 µM 6FTrp. The concentration of 7Fi and 7FTrp could not be determined since no calibration was possible.

	Ind	Trp		6Fi	6FTrp		7Fi	7FTrp
NMM								
0-0-70 (Ind)	13914649		0-70-0 (6Fi)	12904343		0-70-0 (7Fi)	84942 (UV)	
0-0-70 (W)		7286032						
TUB00			**6TUB128-OC**			**7TUB165-OC**		
24 h	1761113	127094	24 h	5830287	4768	24 h	n.d.	1204947
48 h	12610	150022	48 h	n.d.	23814	48 h	51773 (UV)	n.d.
96 h	13188	175048	72 h	n.d.	205928	72 h	36224 (UV)	n.d.
120 h	17349	52341	96 h	n.d.	205375	96 h	39787 (UV)	n.d.
W-TUB165			**6TUB165-MB4**			**7TUB165-MB**		
24 h	n.d.	3226	24 h	102811	8780	24 h	n.d.	n.d.
48 h	n.d.	8849	48 h	n.d.	39957	48 h	29916 (UV)	n.d.
96 h	n.d.	8262	72 h	n.d.	82868	72 h	n.d.	n.d.
120 h	n.d.	9788	96 h	n.d.	76610	96 h	n.d.	n.d.
Ind-TUB165			**6TUB165-MB3**					
24 h	n.d.	4221	24 h	162034	65959			
48 h	n.d.	18033	48 h	n.d.	n.d.			
96 h	n.d.	14710	72 h	n.d.	n.d.			
120 h	n.d.	18437	96 h	n.d.	n.d.			

Proteomics

The following parameters are shown: "no. of proteins" states for the number of proteins corresponding to the observed peptides, "coverage data [%]" states for the percentage of proteins detected in the entire dataset, "coverage fasta [%]" states for the percentage of proteins detected based on genomic data (fasta file for TUB00) that are the number of predicted proteins (all coding sequences that are used to potentially encode a protein) and "average fasta|data [%]" states for the mean of detected proteins in the respective data set. Furthermore, for the determination of FTrp peptides in Trp adapted strains and vice versa, the total number of all peptides and the number of FTrp peptides or else Trp peptides were counted, and the percentage of False/Positive was calculated based on the rational that in Trp-adapted strains no FTrp-containing peptides and in FTrp-adapted strains no Trp-containing peptides without the fluorine modification should be detected. Errors are derived from standard deviation, respectively.

Table S26 | Summary of proteomics data of TUB00, Ind-TUB165 and W-TUB165.

	TUB00			Ind-TUB165			W-TUB165		
protein coverage									
average data|fasta [%]	20,49 ±7.11		10.97 ±3.81	35.73 ±8.06		19.13 ±4.32	37.83 ±11.91		20.26 ±6.38
coverage fasta [%]	6.08	11.46	15.37	25.18	16.84	15.37	11.55	22.57	26.65
coverage data [%]	11.35	21.40	28.71	47.02	31.45	28.71	21.57	42.15	49.76
no. proteins	261	492	660	1081	723	660	496	969	1144

Determination of FTrp peptides in Trp adapted strains

	TUB00			Ind-TUB165			W-TUB165		
average [%]	0.0225 ±0.0318			0.0878 ±0.0098			0.0485 ±0.0485		
% FalsePositives	0.0000	0.0674	0.0000	0.0878	0.0998	0.0758	0.0000	0.0415	0.1039
no. FTrp peptides	0	1	0	4	3	2	0	2	6
no. peptides (tot.)	552	1483	2329	4557	3005	2640	1398	4822	5776

Table S27 | Summary of proteomics data of 6TUB128-OC, 6TUB165-MB4 and 6TUB165-MB3.

	6TUB128-OC			6TUB165-MB			6TUB165-MB3		
protein coverage									
average data|fasta[%]	33.30 ±6.46		17.84 ±3.46	43.77±17.21		23.44 ±9.21	30.69±12.88		14.44± 6.90
coverage fasta [%]	13.74	17.56	22.20	30.65	29.23	10.44	13.44	25.97	9.90
coverage data [%]	25.66	32.80	41.45	57.24	54.59	19.49	25.10	48.50	18.49
no. proteins	590	754	953	1316	1255	448	577	1115	425

Determination of Trp peptides in 6FTrp adapted strains

average [%]	0.2309 ±0.1315			0.1702 ±0.0555			0.1269 ±0.1269		
% FalsePositives	0.3247	0.3231	0.0450	0.2033	0.0921	0.2154	0.0911	0.2111	0.0786
no. Trp peptides	7	10	2	15	5	3	2	13	1
no. peptides (tot.)	2156	3095	4442	7380	5430	1393	2196	6159	1272

Table S28 | Summary of proteomics data of 7TUB165-OC and 7TUB165-MB.

	7TUB165-OC			7TUB165-MB		
protein coverage						
average data|fasta[%]	41.68 ±8.24		22.32 ±4.41	50.54 ±6.65		27.07 ±3.56
coverage fasta [%]	18.54	19.92	28.51	24.32	32.10	24.78
coverage data [%]	34.62	37.19	53.24	45.41	59.94	46.28
no. proteins	796	855	1224	1044	1378	1064

Determination of Trp peptides in 7FTrp adapted strains

average [%]	0.3984 ±0.0822			0.5154 ±0.1173		
% FalsePositives	0,4623	0,2823	0,4505	0,4022	0,6771	0,4669
no. Trp peptides	16	11	33	21	53	25
no. peptides (tot.)	3461	3896	7326	5221	7828	5355

Metabolic levels of indole, tryptophan and fluorotryptophan
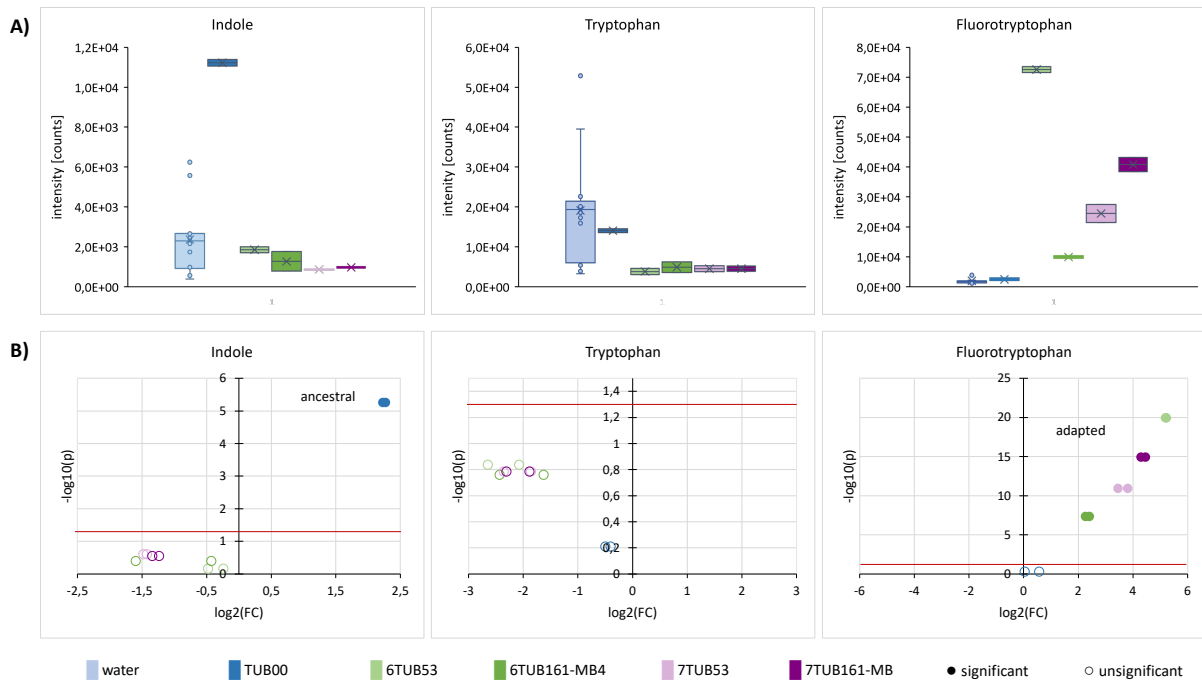
Figure S66| Evaluation of the significance of Ind, Trp and FTrp in TUB00 and the adapted strains 6TUB53, 6TUB161-MB4, 7TUB53 and 7TUB161-MB. A) Boxplots of the features Ind, Trp and FTrp, respectively show signals measured in the water samples and metabolome samples. This illustrates that the target compounds were putatively measured in the water samples, although these signals were generated by automated gap filling algorithm. B) Volcano plots show significance of "unwanted" Ind, Trp and FTrp features in metabolome samples. The p-values were calculated using t-test; p-value <0.05 (-log(p) -1.3) was used as confidence level.

Metabolic profiles
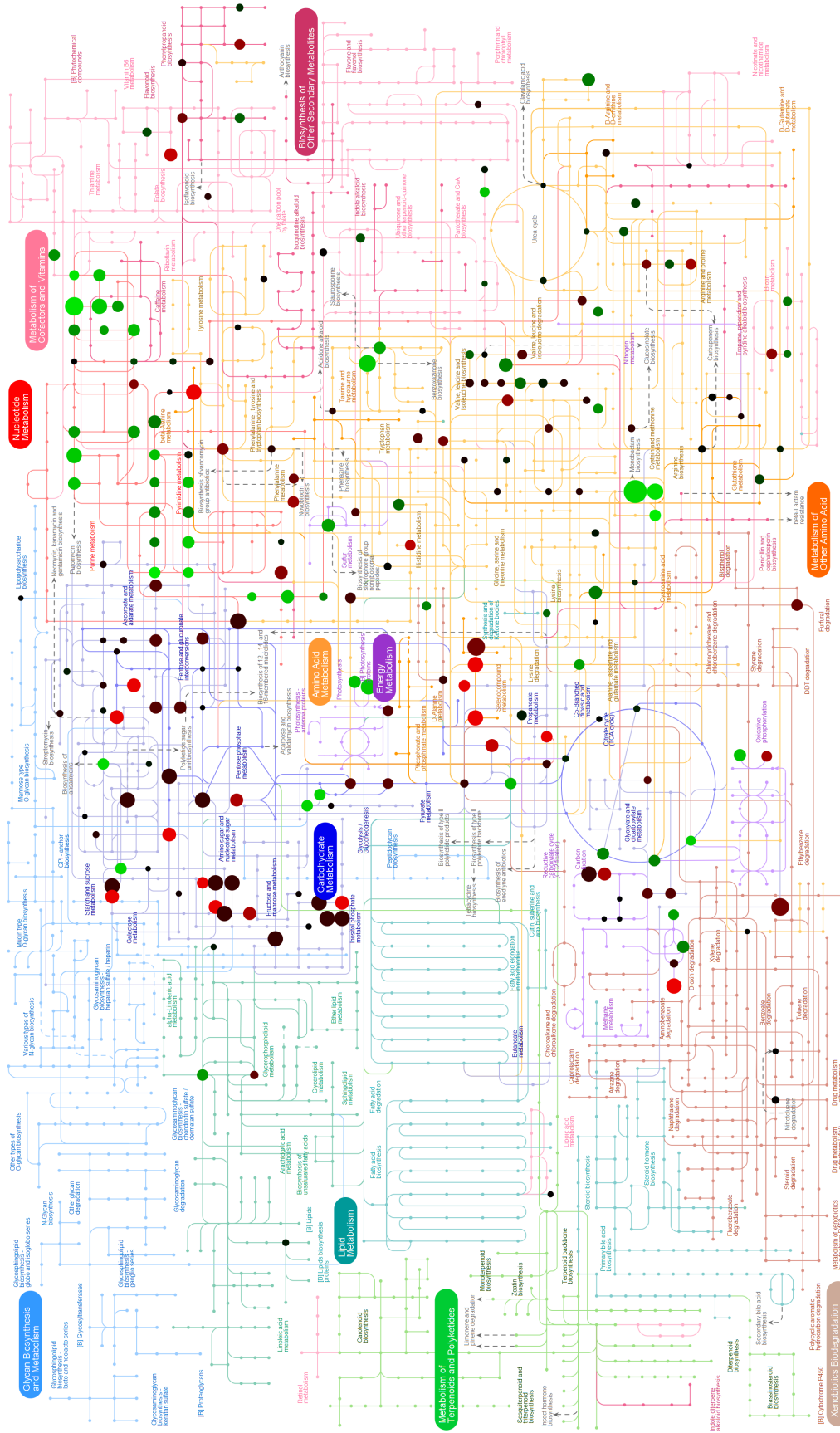
Figure S67| Metabolic profile of 6TUB53 : TUB00.
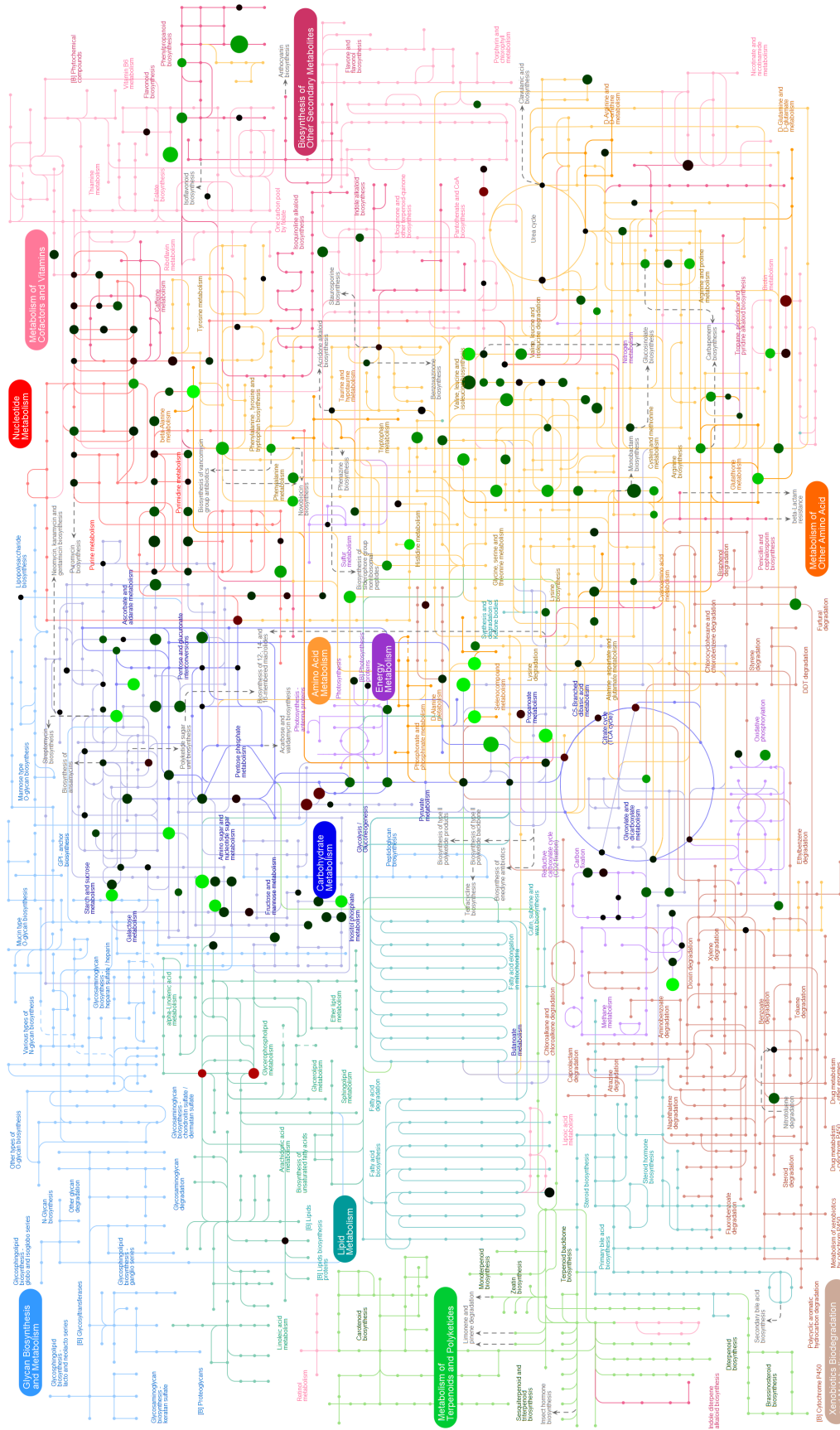
Figure S68| Metabolic profile of 6TUB161-MB4 : 6TUB53.
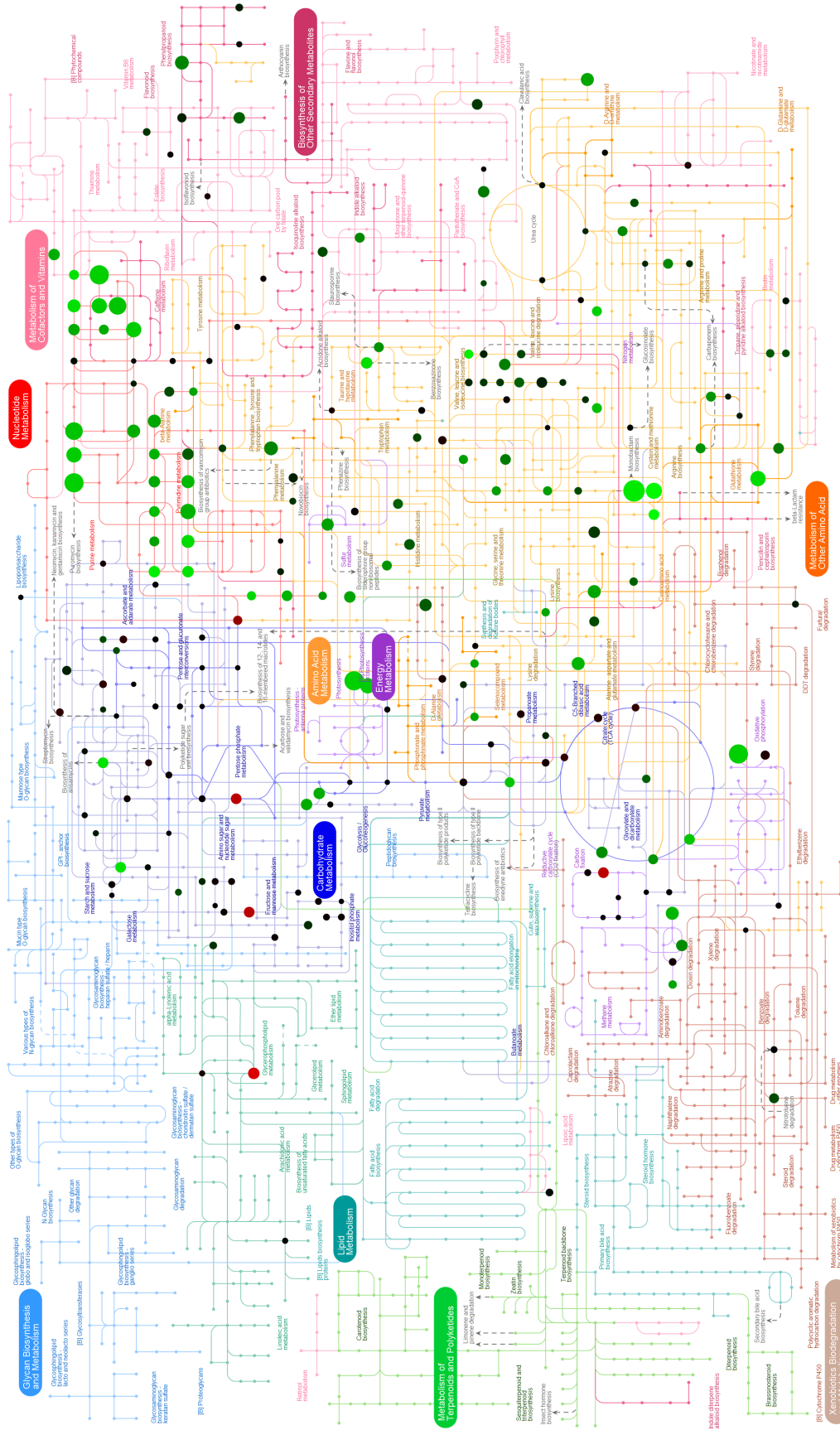
Figure S69| Metabolic profile of 6TUB161-MB4 : TUB00 (raw data).
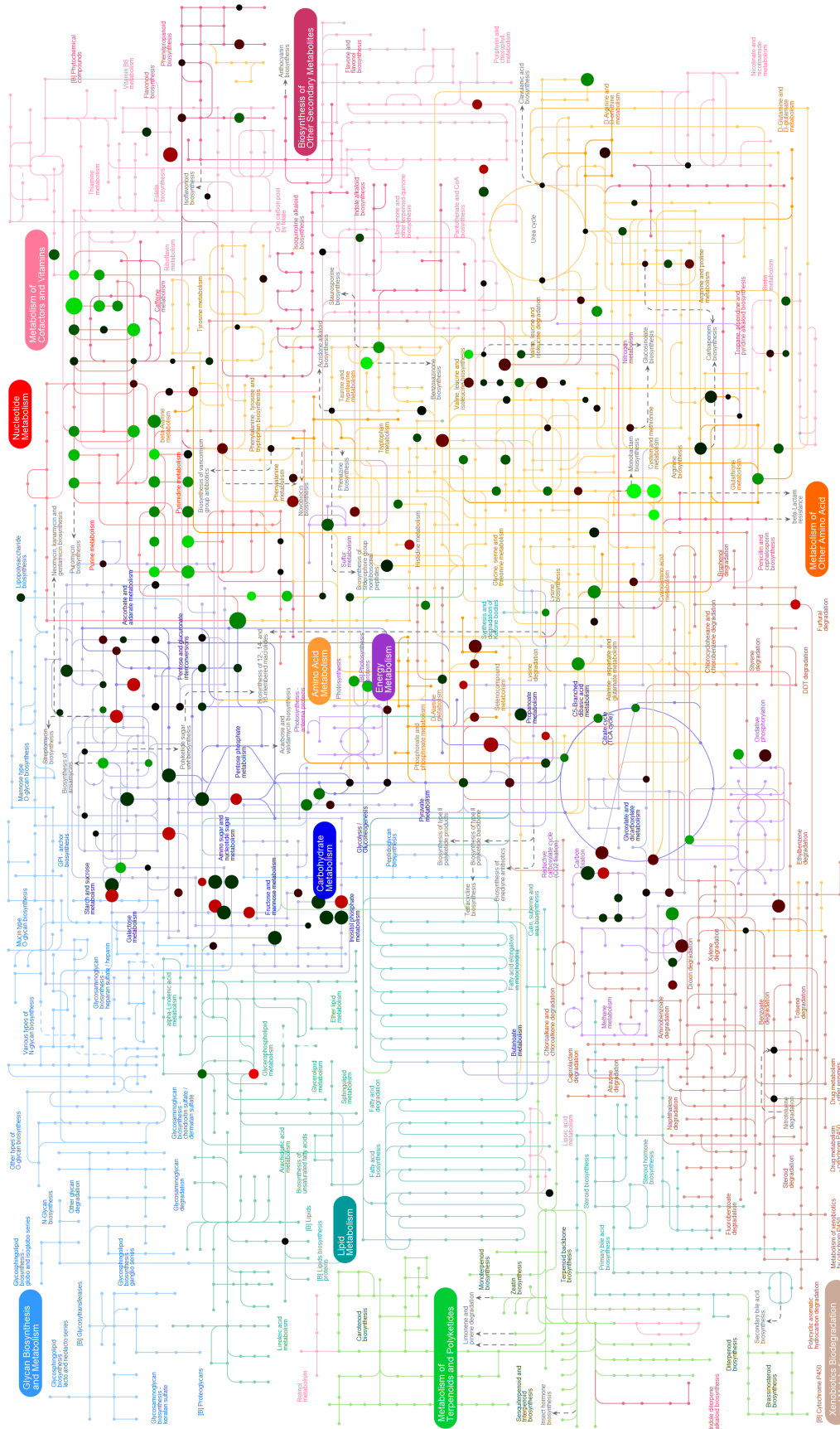
Figure S70| Metabolic profile of 7TUB53 : TUB00.
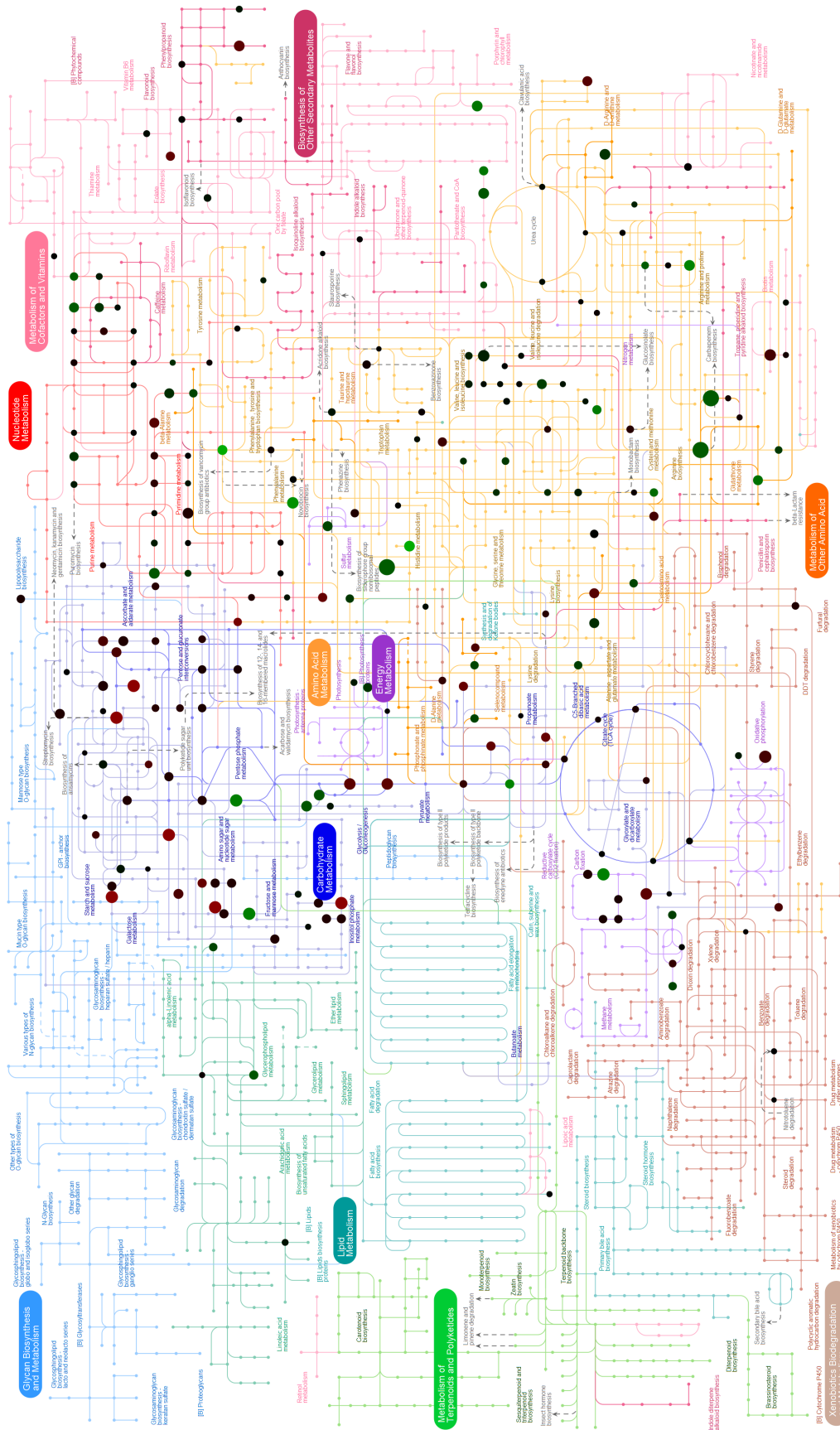
Figure S71 | Metabolic profile of 7TUB161-MB : 7TUB53.
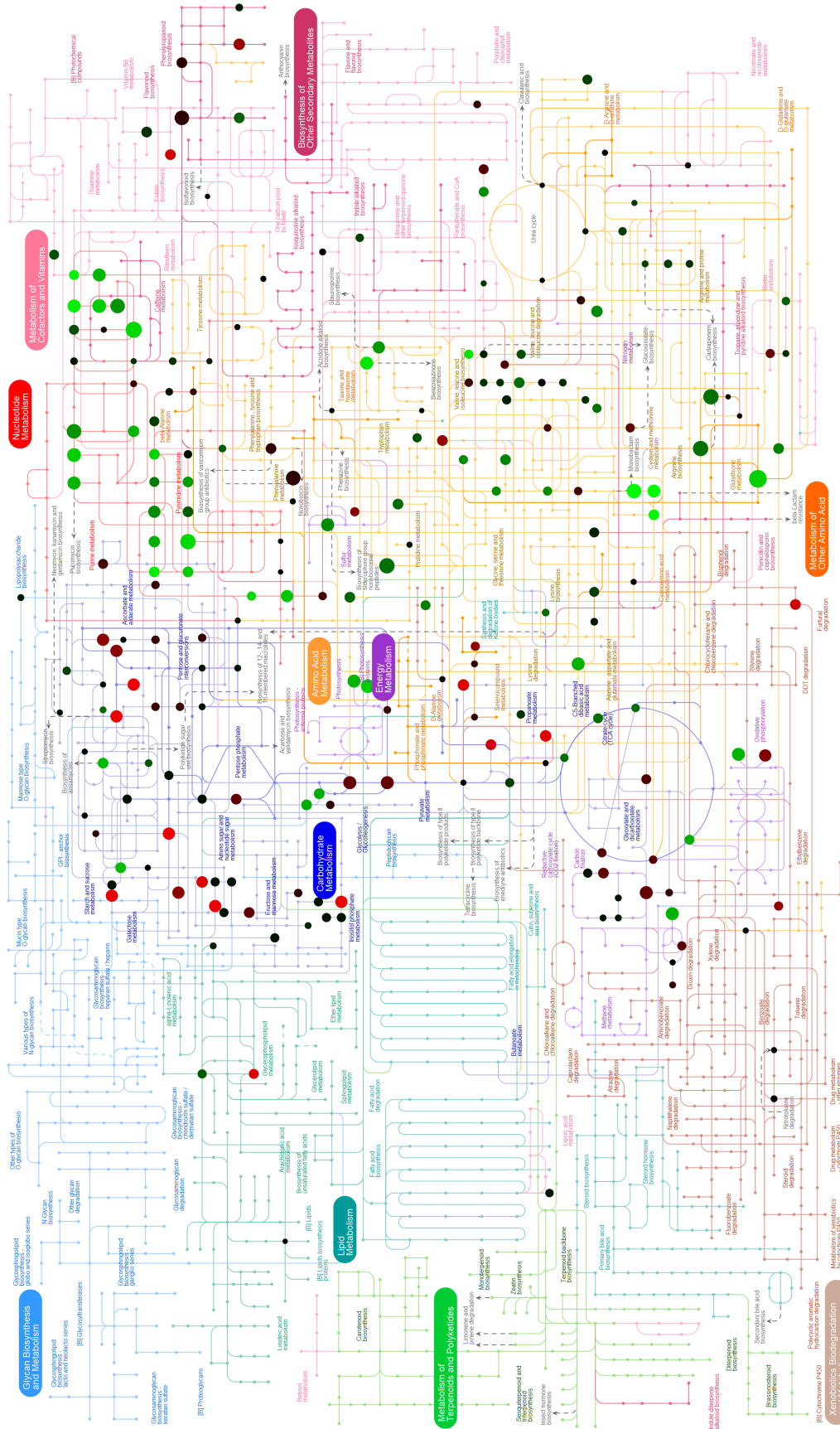
Figure S72| Metabolic profile of 7TUB161-MB : TUB00 (raw data).

Genomics

The Tables S29 – S35 contain the sequencing information of all strains discussed in this study. Only coding regions are considered, information on hypothetical proteins where no gene was assigned to date, intragenetic variants (misc_RNA features) or non-coding transcript variants (tRNA) are excluded from these tables. For the lineages 6TUBX-MB4 and 6TUBX-MB3 reduced information on the SNP data are deposited because these datasets are too big with 175 and 285 SNPs, respectively. The entire dataset including information of the exact gene position in the *E. coli* genome and the encoded gene product are of course available upon request.

For every gene, the COG (Clusters of Orthologous Groups of proteins) is given, when available.[565] This categorization is an attempt to classify encoded proteins by the following functional categories: J- translation, including ribosome structure and biogenesis; L- replication, recombination and repair; K- transcription; O- molecular chaperones and related functions; M- cell wall structure and biogenesis and outer membrane; N- secretion, motility and chemotaxis; T- signal transduction; P- inorganic ion transport and metabolism; C- energy production and conversion; G- carbohydrate metabolism and transport; E- amino acid metabolism and transport; F- nucleotide metabolism and transport; H- coenzyme metabolism; I- lipid metabolism; D- cell division and chromosome partitioning; R- general functional prediction only; and S- no functional prediction.

Table S29 | Mutations (SNPs) found in the coding sequences of the W-TUBX lineages, relative to the *Escherichia coli* MG1655 NC_000913.3 reference genome. The COG (Clusters of Orthologous Groups of proteins) when available, the exact position in the *E. coli* genome (POS), the affected gene (GENE), the encoded product (PRODUCT) and the specific mutation/SNP (EFFECT) for genomic samples of WTUB18, WTUB23, W-TUB53, W-TUB72, W-TUB83, W-TUB88, W-TUB128 and W-TUB165 are shown. Rows colored in grey indicate persistent mutations.

COG	POS	GENE	PRODUCT	W-TUB18	W-TUB23	W-TUB53	W-TUB72	W-TUB83	W-TUB88	W-TUB128	W-TUB165
							EFFECT				
X	1013588	*insA-2*	IS1 protein InsA		missense_variant c.243_249delCCGCCATinsACGTCAC p.Phe81Leu						missense_variant c.243_249delCCGCCATinsACGTCAC p.Phe81Leu
X	1013683	*insAB-1*	InsAB' transposase		synonymous_variant						synonymous_variant

					c.18_21delATACinsCTAT p.8						c.18_21delATACinsCTAT p.8
X	1013734	*insAB-1*	InsAB' transposase		missense_variant c.69G>T p.Arg23Ser						missense_variant c.69G>T p.Arg23Ser
P	1474528	*yncD*	putative TonB-dependent receptor						missense_variant c.439G>T p.Ala147Ser		
	2024101	*yeeW*	CP4-44 prophage					missense_variant c.94A>C p.Ile32Leu			
O, E	2329816	*yfbL*	putative peptidase			missense_variant c.644C>A p.Ser215Tyr					
K	2803670	*rpoS*	RNA polymerase, sigma S (sigma 38) factor				stop_gained c.910C>T p.Gln304*				
T	2905207	*ptsP*	PTS system, enzyme I, transcriptional regulator (with NPR and NTR proteins)								frameshift_variant c.206dupC p.Phe70fs
J	3248691	*infB*	protein chain initiation factor IF2				missense_variant c.1828C>T p.Arg610Cys				
J	3249895	*infB*	protein chain initiation factor IF2							disruptive_inframe_insertion c.603_623dupCGTTGCTGAAGAAGCACGTCG p.Arg208_Met209insValAlaGluGluAlaArgArg	
J	3344692	*dusB*	tRNA-dihydrouridine synthase B							disruptive_inframe_insertion c.936_944dupGTTGGAGGC p.Ala315_Tyr316insLeuGluAla	
E	3589899	*gadA*	glutamate decarboxylase A subunit		synonymous_variant c.642C>T p.Thr214Thr						

COG	POS	GENE	PRODUCT	Ind-TUB18	Ind-TUB23	Ind-TUB53	Ind-TUB72	Ind-TUB83	Ind-TUB88	Ind-TUB128	Ind-TUB165
R	3688467	*yibJ*	putative Rhs-family protein	missense_variant c.97C>A p.Gln33Lys	missense_variant c.97C>A p.Gln33Lys	missense_variant c.97C>A p.Gln33Lys	missense_variant c.97C>A p.Gln33Lys		missense_variant c.97C>A p.Gln33Lys		

Table S30| Mutations (SNPs) found in the coding sequences of the Ind-TUBX lineages, relative to the *Escherichia coli* MG1655 NC_000913.3 reference genome. The COG (Clusters of Orthologous Groups of proteins) when available, the exact position in the *E. coli* genome (POS), the affected gene (GENE), the encoded product (PRODUCT) and the specific mutation/SNP (EFFECT) for genomic samples of Ind-TUB18, Ind-TUB23, Ind-TUB53, Ind-TUB72, Ind-TUB83, Ind-TUB88, Ind-TUB128 and Ind-TUB165 are shown. Rows colored in grey indicate persistent mutations.

COG	POS	GENE	PRODUCT	Ind-TUB18	Ind-TUB23	Ind-TUB53	Ind-TUB72	Ind-TUB83	Ind-TUB88	Ind-TUB128	Ind-TUB165
							EFFECT				
T	388931	*phoR*	Phosphate regulon sensor protein PhoR						missense_variant c.391C>A p.Leu131Ile		
T	389292	*phoR*	Phosphate regulon sensor protein PhoR						missense_variant c.752A>T p.Gln251Leu		
X	1013552	*insA-2*	IS1 protein InsA		synonymous_variant c.207A>C p.Thr69Thr			synonymous_variant c.207A>C p.Thr69Thr		synonymous_variant c.207A>C p.Thr69Thr	
X	1013588	*insA-2*	IS1 protein InsA						missense_variant c.243_249delCCGCCATinsACGTCAC p.Phe81Leu		missense_variant c.243_249delCCGCCATinsACGTCAC p.Phe81Leu
X	1013683	*insAB-1*	InsAB' transposase						synonymous_variant c.18_21delATACinsCTAT p.8		synonymous_variant c.18_21delATACinsCTAT p.8
X	1013734	*insAB-1*	InsAB' transposase						missense_variant c.69G>T p.Arg23Ser		missense_variant c.69G>T p.Arg23Ser
N	1912243	*flhA_2*	flagellar biosynthesis protein FlhA			synonymous_variant c.1338C>A p.Ala446Ala					

	2695196	*yfjH*	CP4-57 prophage								frameshift_variant c.946delA p.Arg316fs
K	2804125	*rpoS*	RNA polymerase, sigma S (sigma 38) factor				disruptive_inframe_deletion c.449_454delTTCGCC p.Ile150_Gln152delinsLys				
K	2804560	*rpoS*	RNA polymerase, sigma S (sigma 38) factor			frameshift_variant c.12_19delTACGCTGA p.Asn4fs					
T	2904049	*ptsP*	PTS system, enzyme I, transcriptional regulator (with NPR and NTR proteins)								frameshift_variant c.1343_1364delAAGAAAAACTGGGCAGCCGTAT p.Glu448fs
T	2905365	*ptsP*	PTS system, enzyme I, transcriptional regulator (with NPR and NTR proteins)							frameshift_variant c.48delA p.Arg17fs	
E, R	2955027	*ygfK*	putative oxidoreductase, Fe-S subunit								missense_variant c.2441A>C p.Glu814Ala
G	3397913	*chiA*	endochitinase						synonymous_variant c.21T>G p.Thr7Thr		
J	3536488	*yhhQ*	hypothetical protein							missense_variant c.26G>T p.Arg9Leu	
E	3589899	*gadA*	glutamate decarboxylase A subunit								
R	3688467	*yibJ*	putative Rhs-family protein	missense_variant c.97C>A p.Gln33Lys		missense_variant c.97C>A p.Gln33Lys					
G	3707662	*gpmM*	2,3-bisphosphoglycerate-independent phosphoglycerate mutase							missense_variant c.1259C>T p.Ala420Val	

COG	POS	GENE	PRODUCT	7TUB18	7TUB53	7TUB83-OC	7TUB165-OC
E, R	3712135	*tdh*	threonine dehydrogenase				missense_variant c.355G>T p.Gly119Cys
E, R	3712201	*tdh*	threonine dehydrogenase				missense_variant c.289C>T p.Arg97Cys
J	3720600	*rfaZ*	protein involved in KdoIII attachment during lipopolysaccharide core biosynthesis			synonymous_variant c.741T>A p.Ser247Ser	
M	3726117	*rfaS*	lipopolysaccharide core biosynthesis protein				missense_variant c.144T>A p.Asp48Glu

Table S31 | Mutations (SNPs) found in the coding sequences of the 7TUBX-OC lineages, relative to the *Escherichia coli* MG1655 NC_000913.3 reference genome. The COG (Clusters of Orthologous Groups of proteins) when available, the exact position in the *E. coli* genome (POS), the affected gene (GENE), the encoded product (PRODUCT) and the specific mutation/SNP (EFFECT) for genomic samples of 7TUB18, 7TUB53, 7TUB83-OC and 7TUB165-OC are shown. Rows colored in grey indicate persistent mutations.

COG	POS	GENE	PRODUCT	7TUB18	7TUB53	7TUB83-OC	7TUB165-OC
					EFFECT		
J	6471	*ileS*	isoleucyl-tRNA synthetase			missense_variant c.1250C>T p.Ala417Val	
J	6825	*ileS*	isoleucyl-tRNA synthetase		missense_variant c.1604T>G p.Phe535Cys		missense_variant c.1604T>G p.Phe535Cys
I	178739	*cdsA*	CDP-diglyceride synthetase		missense_variant c.209A>C p.Tyr70Ser		missense_variant c.209A>C p.Tyr70Ser
O	429636	*lon*	DNA-binding, ATP-dependent protease La			missense_variant c.97C>G p.Arg33Gly	
O	888319	*clpA*	ClpAXP			stop_gained c.1543G>T p.Glu515*	
X	1013588	*insA-2*	IS1 protein InsA			missense_variant c.243_249delCCGCCATinsACGTCAC p.Phe81Leu	
X	1013683	*insAB-1*	InsAB' transposase			synonymous_variant c.18_21delATACinsCTAT p.8	synonymous_variant c.18_21delATACinsCTAT p.8

X	1013734	*insAB-1*	InsAB' transposase			missense_variant c.69G>T p.Arg23Ser	missense_variant c.69G>T p.Arg23Ser
K	1291223	*cysB*	CysB-O-acetyl-L-serine DNA-binding transcriptional dual regulator				missense_variant c.724G>A p.Val242Ile
J	2139068	*metG*	methionyl-tRNA synthetase				missense_variant c.656T>G p.Met219Arg
K	3030022	*yqgE*	hypothetical protein				missense_variant c.631G>A p.Ala211Thr
S	3127695	*yqiK*	putative membrane protein			missense_variant c.328G>A p.Glu110Lys	
O	3310772	*sspA*	stringent starvation protein A	frameshift_variant c.224_225insA p.Met76fs	frameshift_variant c.224_225insA p.Met76fs		frameshift_variant c.224_225insA p.Met76fs
O	3310808	*sspA*	stringent starvation protein A			frameshift_variant c.176_188delTTCCGACCCTGGT p.Val59fs	
J	3344690	*dusB*	tRNA-dihydrouridine synthase B		conservative_inframe_deletion c.928_936delCTGGAGGCG p.Leu310_Ala312del		conservative_inframe_deletion c.928_936delCTGGAGGCG p.Leu310_Ala312del
R	3688467	*yibJ*	putative Rhs-family protein	missense_variant c.97C>A p.Gln33Lys			
L	3804051	*dnaA*	chromosomal replication initiator protein DnaA			missense_variant c.823A>C p.Ile275Leu	
L	3804101	*dnaA*	chromosomal replication initiator protein DnaA		missense_variant c.773T>G p.Leu258Arg		missense_variant c.773T>G p.Leu258Arg
E	3927906	*metE*	cobalamin-independent homocysteine transmethylase				missense_variant c.58A>C p.Lys20Gln
C	4068977	*sthA*	pyridine nucleotide transhydrogenase				stop_gained c.1376T>G p.Leu459*
K	4085163	*rpoB*	RNA polymerase, beta subunit	missense_variant c.547T>A p.Trp183Arg			
K	4092476	*rpoC*	RNA polymerase, beta' subunit				missense_variant c.3755A>T p.His1252Leu
M	4284310	*mscM*	mechanosensitive channel of miniconductance McsM monomer				missense_variant c.3044G>C p.Arg1015Pro
O	4339906	*msrA*	methionine sulfoxide reductase A		missense_variant c.310T>G p.Ser104Ala		missense_variant c.310T>G p.Ser104Ala
K	4528822	*trpR*	TrpR-Tryptophan DNA-binding transcriptional repressor			frameshift_variant c.114_123delCGAAGCGTTG p.Glu39fs	

Table S32| Mutations (SNPs) found in the coding sequences of the 7TUBX-MB lineages, relative to the *Escherichia coli* MG1655 NC_000913.3 reference genome. The COG (Clusters of Orthologous Groups of proteins) when available, the exact position in the *E. coli* genome (POS), the affected gene (GENE), the encoded product (PRODUCT) and the specific mutation/SNP (EFFECT) for genomic samples of 7TUB18, 7TUB53, 7TUB72-MB, 7TUB83-MB and 7TUB165-MB are shown. Rows colored in grey indicate persistent mutations.

COG	POS	GENE	PRODUCT	7TUB18	7TUB53	7TUB72-MB	7TUB83-MB	7TUB165-MB
						EFFECT		
J	5417	ileS	isoleucyl-tRNA synthetase			missense_variant c.196A>C p.Ile66Leu		missense_variant c.196A>C p.Ile66Leu
J	6825	ileS	isoleucyl-tRNA synthetase		missense_variant c.1604T>G p.Phe535Cys	missense_variant c.1604T>G p.Phe535Cys	missense_variant c.1604T>G p.Phe535Cys	missense_variant c.1604T>G p.Phe535Cys
I	178739	cdsA	CDP-diglyceride synthetase		missense_variant c.209A>C p.Tyr70Ser		missense_variant c.209A>C p.Tyr70Ser	
E	390385	brnQ	branched chain amino acid transporter BrnQ				missense_variant c.143T>G p.Phe48Cys	
E	391070	brnQ	branched chain amino acid transporter BrnQ			frameshift_variant c.836_855dupGTAGCTTCCTGCTGGCGGCG p.Leu286fs		frameshift_variant c.836_855dupGTAGCTTCCTGCTGGCGGCG p.Leu286fs
C	720241	sdhA	succinate dehydrogenase flavoprotein					stop_gained c.346C>T p.Gln116*
X	1013734	insAB-1	InsAB' transposase					missense_variant c.69G>T p.Arg23Ser
E	1279110	trpB	tryptophan synthase, beta subunit			missense_variant c.523C>T p.Arg175Cys		missense_variant c.523C>T p.Arg175Cys
I	1940802	pgsA	phosphatidylglycerophosphate synthase					missense_variant c.32T>G p.Leu11Arg
M	2255891	ompC_2	outer membrane porin C					missense_variant c.172C>T p.Arg58Cys
H	3024386	metK	methionine adenosyltransferase					stop_lost&splice_region_variant c.1154A>C p.Ter385Serext*?
K	3029591	yqgE	hypothetical protein					missense_variant c.200T>C p.Leu67Pro
O	3310772	sspA	stringent starvation protein A	frameshift_variant c.224_225insA p.Met76fs	frameshift_variant c.224_225insA p.Met76fs	frameshift_variant c.224_225insA p.Met76fs	frameshift_variant c.224_225insA p.Met76fs	frameshift_variant c.224_225insA p.Met76fs
J	3344690	dusB	tRNA-dihydrouridine synthase B		conservative_inframe_deletion c.928_936delCTGGAGGCG p.Leu310_Ala312del		conservative_inframe_deletion c.928_936delCTGGAGGCG p.Leu310_Ala312del	

COG	POS	GENE	PRODUCT					EFFECT
K	3465508	*yhgF*	putative transcriptional accessory protein					missense_variant c.359A>C p.Glu120Ala
	3631705		tRNA-Pro					non_coding_transcript_variant
R	3688467	*yibJ*	putative Rhs-family protein	missense_variant c.97C>A p.Gln33Lys			missense_variant c.97C>A p.Gln33Lys	missense_variant c.97C>A p.Gln33Lys
L	3804101	*dnaA*	chromosomal replication initiator protein DnaA		missense_variant c.773T>G p.Leu258Arg	missense_variant c.773T>G p.Leu258Arg	missense_variant c.773T>G p.Leu258Arg	missense_variant c.773T>G p.Leu258Arg
K	4085163	*rpoB*	RNA polymerase, beta subunit	missense_variant c.547T>A p.Trp183Arg				
K	4086657	*rpoB*	RNA polymerase, beta subunit					missense_variant c.2041A>C p.Met681Leu
K	4092865	*rpoC*	RNA polymerase, beta' subunit			frameshift_variant c.4146_4147dupGC p.Gln1383fs		frameshift_variant c.4146_4147dupGC p.Gln1383fs
O	4339906	*msrA*	methionine sulfoxide reductase A		missense_variant c.310T>G p.Ser104Ala	missense_variant c.310T>G p.Ser104Ala	missense_variant c.310T>G p.Ser104Ala	missense_variant c.310T>G p.Ser104Ala

Table S33 | Mutations (SNPs) found in the coding sequences of the 6TUBX-OC lineages, relative to the *Escherichia coli* MG1655 NC_000913.3 reference genome. The COG (Clusters of Orthologous Groups of proteins) when available, the exact position in the *E. coli* genome (POS), the affected gene (GENE), the encoded product (PRODUCT) and the specific mutation/SNP (EFFECT) for genomic samples of 6TUB23, 6TUB53, 6TUB88-OC and 6TUB128-OC are shown. Rows colored in grey indicate persistent mutations.

COG	POS	GENE	PRODUCT	6TUB23	6TUB53	6TUB88-OC	6TUB128-OC
						EFFECT	
P	125250	*can*	carbonic anhydrase 2 monomer		missense_variant c.167A>G p.Glu56Gly	missense_variant c.167A>G p.Glu56Gly	missense_variant c.167A>G p.Glu56Gly
J	140880	*pcnB*	poly(A) polymerase I			frameshift_variant c.959dupA p.Lys321fs	frameshift_variant c.959dupA p.Lys321fs
J	641226	*leuS*	leucyl-tRNA synthetase				missense_variant c.563T>G p.Ile188Ser
C	825128	*ybiW*	putative pyruvate formate lyase		missense_variant c.829A>T p.Ile277Phe	missense_variant c.829A>T p.Ile277Phe	missense_variant c.829A>T p.Ile277Phe
E, J	1799812	*ansA*	asparaginase I	missense_variant c.423T>G p.Asn141Lys			
T	1863785	*proQ*	RNA chaperone, involved in posttranscriptional control of ProP levels				stop_gained c.280C>T p.Gln94*

N	1967688	*fliL*	flagellar biosynthesis		missense_variant c.55C>A p.Leu19Met	missense_variant c.55C>A p.Leu19Met	missense_variant c.55C>A p.Leu19Met
	2991928	*ygfB*	hypothetical protein				missense_variant c.44A>C p.Gln15Pro
H	3023365	*metK*	methionine adenosyltransferase			missense_variant c.133T>G p.Tyr45Asp	
H	3024033	*metK*	methionine adenosyltransferase				missense_variant c.801T>A p.Asp267Glu
M	3277811	*lptC*	LptABCFG ABC transporter				missense_variant c.472T>C p.Ser158Pro
O	3310551	*sspA*	stringent starvation protein A		frameshift_variant c.438_445delGCCGTACT p.Lys146fs	frameshift_variant c.438_445delGCCGTACT p.Lys146fs	frameshift_variant c.438_445delGCCGTACT p.Lys146fs
J	3377593	*rpsC*	30S ribosomal subunit protein S3	missense_variant c.371T>G p.Leu124Arg			
J	3441324	*trpS*	tryptophanyl-tRNA synthetase		missense_variant c.80A>C p.Gln27Pro	missense_variant c.80A>C p.Gln27Pro	missense_variant c.80A>C p.Gln27Pro
R	3688467	*yibJ*	putative Rhs-family protein				missense_variant c.97C>A p.Gln33Lys
K	4085059	*rpoB*	RNA polymerase, beta subunit	missense_variant c.443A>T p.Gln148Leu	missense_variant c.443A>T p.Gln148Leu	missense_variant c.443A>T p.Gln148Leu	missense_variant c.443A>T p.Gln148Leu
K	4088611	*rpoB*	RNA polymerase, beta subunit		missense_variant c.3995C>T p.Ser1332Leu	missense_variant c.3995C>T p.Ser1332Leu	missense_variant c.3995C>T p.Ser1332Leu
J	4502810	*rsmC*	16S rRNA methyltransferase			missense_variant c.812T>C p.Phe271Ser	

Table S34| Mutations (SNPs) found in the coding sequences of the 6TUBX-MB4 lineages, relative to the *Escherichia coli* MG1655 NC_000913.3 reference genome. The affected genes and the specific mutation/SNP for genomic samples of 6TUB23, 6TUB53, 6TUB88-MB4, 6TUB93-MB4, 6TUB98-MB4, 6TUB108-MB4, 6TUB113-MB4, 6TUB118-MB4, 6TUB123-MB4, 6TUB128-MB4, 6TUB138-MB4, 6TUB148-MB4, 6TUB158-MB4 and 6TUB165-MB4 are shown.

	AFFECTED GENE (SNP)
6TUB23	*ansA* (missense_variant c.423T>G p.Asn141Lys), *rpsC* (missense_variant c.371T>G p.Leu124Arg), *rpoB* (missense_variant c.443A>T p.Gln148Leu)
6TUB53	*can* (missense_variant c.167A>G p.Glu56Gly), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *fliL* (missense_variant c.55C>A p.Leu19Met), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu)

6TUB88-MB4	*can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *leuS* (missense_variant c.1900G>C p.Val634Leu), *leuS* (missense_variant c.467T>G p.Val156Gly), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *pykF* (missense_variant c.928A>G p.Met310Val), *proQ* (frameshift_variant c.145dupG p.Glu49fs), *fliL* (missense_variant c.55C>A p.Leu19Met), *arnC* (missense_variant c.563C>G p.Ala188Gly), *metK* (missense_variant c.801T>A p.Asp267Glu), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu)
6TUB93-MB4	*pcnB* (frameshift_variant c.959dupA p.Lys321fs), *glnD* (missense_variant c.892C>T p.Arg298Cys), *citC* (missense_variant c.430G>A p.Gly144Arg), *leuS* (missense_variant c.1457A>G p.Glu486Gly), *leuS* (missense_variant c.467T>G p.Val156Gly), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *ybjE* (missense_variant c.235G>A p.Ala79Thr), *insAB-1* (missense_variant c.69G>T p.Arg23Ser), *ymdC* (frameshift_variant c.126delT p.Pro43fs), *plsX* (missense_variant c.512C>T p.Ala171Val), *pspF* (synonymous_variant c.562T>C p.Leu188Leu), *narZ* (synonymous_variant c.996C>T p.Arg332Arg), *rsxB* (missense_variant c.63T>G p.Ile21Met), *ydhV* (frameshift_variant c.710delC p.Pro237fs), *yedW_6* (frameshift_variant c.307delG p.Ala103fs), *fliL* (missense_variant c.55C>A p.Leu19Met), *arnC* (missense_variant c.563C>G p.Ala188Gly), *hyfC* (synonymous_variant c.594A>G p.Gly198Gly), *bamB* (frameshift_variant c.632dupG p.Asp212fs), *mutS* (frameshift_variant c.1057delG p.Ala353fs), *metK* (missense_variant c.19A>C p.Thr7Pro), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *yhiM* (synonymous_variant c.673C>T p.Leu225Leu), *gadA* (synonymous_variant c.642C>T p.Thr214Thr), *yibJ* (missense_variant c.97C>A p.Gln33Lys), *yihP* (missense_variant c.677C>T p.Ala226Val), *cdh* (synonymous_variant c.600A>G p.Val200Val), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *lamB* (synonymous_variant c.1110T>C p.Ile370Ile), *sdsQ* (missense_variant c.1385C>T p.Ala462Val), *proP* (missense_variant c.167T>C p.Phe56Ser), *adiA* (missense_variant c.782A>G p.Gln261Arg), *aspA* (synonymous_variant c.309T>C p.Gly103Gly), *thrC* (missense_variant c.173T>A p.Ile58Asn)
6TUB98-MB4	*caiF* (synonymous_variant c.312G>A p.Thr104Thr), *can* (missense_variant c.167A>G p.Glu56Gly), *can* (missense_variant c.73T>G p.Phe25Val), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *glnD* (missense_variant c.892C>T p.Arg298Cys), *ybaL* (frameshift_variant c.570dupG p.Ile191fs), *appY* (missense_variant c.410_412delTTAinsATC p.PheThr137TyrPro), *appY* (missense_variant c.421A>T p.Ile141Phe), *nfrB* (missense_variant c.1487A>G p.Glu496Gly), *fepA* (missense_variant c.827G>A p.Ser276Asn), *citC* (missense_variant c.430G>A p.Gly144Arg), *dpiB* (missense_variant c.409A>G p.Thr137Ala), *leuS* (missense_variant c.1457A>G p.Glu486Gly), *leuS* (missense_variant c.467T>G p.Val156Gly), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *ybjE* (missense_variant c.235G>A p.Ala79Thr), *ymdC* (frameshift_variant c.126delT p.Pro43fs), *plsX* (missense_variant c.512C>T p.Ala171Val), *potC* (synonymous_variant c.624A>G p.Gly208Gly), *ribA* (missense_variant c.344T>C p.Val115Ala), *yciW* (frameshift_variant c.156delC p.Phe53fs), *pspF* (synonymous_variant c.562T>C p.Leu188Leu), *paaF* (synonymous_variant c.9A>G p.Glu3Glu), *mdoD* (missense_variant c.476G>A p.Arg159His), *narZ* (synonymous_variant c.996C>T p.Arg332Arg), *marC* (synonymous_variant c.550C>T p.Leu184Leu), *rsxB* (missense_variant c.63T>G p.Ile21Met), *ydhV* (frameshift_variant c.710delC p.Pro237fs), *yedW_6* (frameshift_variant c.307delG p.Ala103fs), *fliL* (missense_variant c.55C>A p.Leu19Met), *mtfA* (synonymous_variant c.219C>T p.Cys73Cys), *yegE* (missense_variant c.3257T>C p.Val1086Ala), *apbE* (missense_variant c.613T>C p.Ser205Pro), *atoB* (frameshift_variant c.858delC p.Ala287fs), *arnC* (missense_variant c.563C>G p.Ala188Gly), *nuoF* (missense_variant c.774T>G p.Asn258Lys), *hyfC* (synonymous_variant c.594A>G p.Gly198Gly), *bamB* (frameshift_variant c.632dupG p.Asp212fs), *hcaT* (frameshift_variant c.649delG p.Ala217fs), *mutS* (frameshift_variant c.1057delG p.Ala353fs), *cysJ* (synonymous_variant c.522A>G p.Glu174Glu), *rlmD* (missense_variant c.617G>A p.Arg206His), *xdhA* (frameshift_variant c.999dupG p.Asn334fs), *metK* (missense_variant c.19A>C p.Thr7Pro), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *ftsY* (synonymous_variant c.324A>G p.Glu108Glu), *yhiM* (synonymous_variant c.673C>T p.Leu225Leu), *gadW* (missense_variant c.86G>A p.Ser29Asn), *yhjA* (missense_variant c.196T>C p.Ser66Pro), *yhjJ* (synonymous_variant c.471C>T p.Asn157Asn), *yibJ* (missense_variant c.97C>A p.Gln33Lys), *yihP* (missense_variant c.677C>T p.Ala226Val), *cdh* (synonymous_variant c.600A>G p.Val200Val), rpoB (missense_variant c.443A>T p.Gln148Leu), rpoB (missense_variant c.3995C>T p.Ser1332Leu), lamB (synonymous_variant c.1110T>C p.Ile370Ile), *sdsQ* (missense_variant c.1385C>T p.Ala462Val), *proP* (missense_variant c.167T>C p.Phe56Ser), *adiA* (missense_variant c.782A>G p.Gln261Arg), *aspA* (synonymous_variant c.309T>C p.Gly103Gly), *fecA* (synonymous_variant c.246C>T p.Gly82Gly), *thrC* (missense_variant c.173T>A p.Ile58Asn)
6TUB108-MB4	*caiA* (frameshift_variant c.517delG p.Ala173fs), *caiT* (missense_variant c.970T>G p.Trp324Gly), *polB* (missense_variant c.2030A>G p.Glu677Gly), *murG* (missense_variant c.56T>C p.Val19Ala), *hofB* (missense_variant c.1265A>G p.Glu422Gly), *can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *glnD* (missense_variant c.892C>T p.Arg298Cys), *yagN* (stop_lost&splice_region_variant c.441A>G p.Ter147Trpext*?), *yaiP* (missense_variant c.1004T>C p.Val335Ala), *dnaX* (synonymous_variant c.111T>C p.Ile37Ile), *ybaL* (frameshift_variant c.173delC p.Pro58fs), *allE* (frameshift_variant c.672delA p.Gly225fs), *citC* (missense_variant c.430G>A p.Gly144Arg), *leuS* (missense_variant c.1457A>G p.Glu486Gly), *leuS* (missense_variant c.467T>G p.Val156Gly), *gltJ* (missense_variant c.236T>C p.Val79Ala), *kdpD* (missense_variant c.1186C>T p.Arg396Cys), *ybhK*

	(missense_variant c.884A>G p.Glu295Gly), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *dacC* (missense_variant c.1150T>C p.Trp384Arg), *ybjS* (missense_variant c.169T>C p.Ser57Pro), *cydD* (missense_variant c.188T>C p.Leu63Pro), *pqiA* (synonymous_variant c.603C>T p.Ala201Ala), *rlmI* (frameshift_variant c.504delC p.Val169fs), *ymdC* (frameshift_variant c.126delT p.Pro43fs), *lpxL* (synonymous_variant c.699C>T p.Gly233Gly), *holB* (missense_variant c.794T>C p.Val265Ala), *potC* (missense_variant c.139T>C p.Tyr47His), *umuC* (synonymous_variant c.1212A>G p.Arg404Arg), *kdsA* (missense_variant c.815A>G p.Asp272Gly), *yciC* (synonymous_variant c.483A>G p.Ala161Ala), *yciW* (missense_variant c.760T>A p.Leu254Ile), *sapC* (missense_variant c.557A>G p.Glu186Gly), *pspF* (synonymous_variant c.562T>C p.Leu188Leu), *abgR* (synonymous_variant c.465A>G p.Gly155Gly), *ydcR* (missense_variant c.1154A>G p.Tyr385Cys), *narZ* (synonymous_variant c.996C>T p.Arg332Arg), *rsxB* (missense_variant c.63T>G p.Ile21Met), *rsxD* (frameshift_variant c.274dupC p.Leu92fs), *ydhV* (frameshift_variant c.710delC p.Pro237fs), *pykF* (frameshift_variant c.902_903insG p.Ile302fs), *ydiK* (frameshift_variant c.405dupG p.Thr136fs), *proQ* (frameshift_variant c.254delG p.Gly85fs), *fliL* (missense_variant c.55C>A p.Leu19Met), *hisC* (synonymous_variant c.531T>C p.Gly177Gly), *asmA* (synonymous_variant c.1216C>T p.Leu406Leu), *fruK* (missense_variant c.25A>G p.Thr9Ala), *bcr* (missense_variant c.1168G>A p.Ala390Thr), *atoB* (frameshift_variant c.858delC p.Ala287fs), *gyrA* (synonymous_variant c.1647T>C p.Val549Val), *arnC* (missense_variant c.563C>G p.Ala188Gly), *yfbK* (missense_variant c.1117T>C p.Tyr373His), *flk* (frameshift_variant c.289delC p.Arg97fs), *ypdA_2* (missense_variant c.334A>G p.Ile112Val), *fryC* (synonymous_variant c.561A>G p.Leu187Leu), *hemF* (missense_variant c.824A>G p.Glu275Gly), *ascF* (missense_variant c.265G>A p.Ala89Thr), *mutS* (frameshift_variant c.1057delG p.Ala353fs), *mltA* (synonymous_variant c.444T>C p.Gly148Gly), *guaD* (frameshift_variant c.939delA p.Ala314fs), *speA* (missense_variant c.1921T>C p.Phe641Leu), *metK* (missense_variant c.19A>C p.Thr7Pro), *yqiI* (missense_variant c.869G>A p.Gly290Asp), *yhaC* (missense_variant c.866A>C p.Asn289Thr), *mtr* (frameshift_variant c.1011dupG p.Leu338fs), *ispB* (missense_variant c.705A>C p.Lys235Asn), *gltB* (missense_variant c.976A>G p.Met326Val), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *yhjA* (missense_variant c.97C>T p.Arg33Cys), *bcsC* (missense_variant c.3271A>G p.Ser1091Gly), *eptB* (frameshift_variant c.920dupG p.Ala308fs), *yidL* (synonymous_variant c.649T>C p.Leu217Leu), *yihP* (missense_variant c.677C>T p.Ala226Val), *fdoH* (missense_variant c.115T>C p.Cys39Arg), *rhaB* (frameshift_variant c.896dupG p.Trp301fs), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *lamB* (synonymous_variant c.1110T>C p.Ile370Ile), *proP* (missense_variant c.167T>C p.Phe56Ser), *cadC* (missense_variant c.643T>C p.Tyr215His), *cadC* (synonymous_variant c.432G>A p.Thr144Thr), *aspA* (synonymous_variant c.309T>C p.Gly103Gly), *msrA* (synonymous_variant c.165T>C p.Gly55Gly)
6TUB113-MB4	*can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *glnD* (missense_variant c.892C>T p.Arg298Cys), *ykfM* (synonymous_variant c.189C>G p.Ser63Ser), *ykgM* (missense_variant c.197C>T p.Ala66Val), *yaiP* (missense_variant c.1004T>C p.Val335Ala), *ybaL* (frameshift_variant c.173delC p.Pro58fs), *ushA* (frameshift_variant c.788dupA p.Asn263fs), *allE* (frameshift_variant c.672delA p.Gly225fs), *ybcO* (frameshift_variant c.277dupG p.Val93fs), *fes* (synonymous_variant c.327C>T p.Arg109Arg), *citC* (missense_variant c.430G>A p.Gly144Arg), *leuS* (missense_variant c.574T>G p.Phe192Val), *leuS* (missense_variant c.467T>G p.Val156Gly), *rihA* (missense_variant c.544A>G p.Ile182Val), gltJ (missense_variant c.236T>C p.Val79Ala), fur (synonymous_variant c.357C>T p.Gly119Gly), kdpD (missense_variant c.1186C>T p.Arg396Cys), abrB (synonymous_variant c.213T>C p.Gly71Gly), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *ybjS* (missense_variant c.169T>C p.Ser57Pro), *cydD* (missense_variant c.188T>C p.Leu63Pro), *pqiA* (missense_variant c.308T>C p.Phe103Ser), *pqiA* (synonymous_variant c.603C>T p.Ala201Ala), *rlmI* (frameshift_variant c.504delC p.Val169fs), *rutA* (stop_gained c.464G>A p.Trp155*), *csgG* (missense_variant c.614T>G p.Val205Gly), *ymdC* (frameshift_variant c.126delT p.Pro43fs), *lpxL* (synonymous_variant c.699C>T p.Gly233Gly), *pyrC* (missense_variant c.298A>G p.Thr100Ala), holB (missense_variant c.794T>C p.Val265Ala), umuC (synonymous_variant c.1212A>G p.Arg404Arg), yciC (synonymous_variant c.483A>G p.Ala161Ala), *trpA* (missense_variant c.532T>C p.Ser178Pro), *pspF* (synonymous_variant c.562T>C p.Leu188Leu), *yncD* (missense_variant c.1384A>G p.Asn462Asp), *narZ* (synonymous_variant c.996C>T p.Arg332Arg), *maeA* (missense_variant c.889A>G p.Lys297Glu), *ynfE* (missense_variant c.1712T>C p.Val571Ala), *rsxB* (missense_variant c.63T>G p.Ile21Met), *rsxD* (frameshift_variant c.273_274dupCC p.Leu92fs), *ydhV* (frameshift_variant c.710delC p.Pro237fs), *pykF* (frameshift_variant c.902_903insG p.Ile302fs), *aroH* (missense_variant c.574A>C p.Thr192Pro), *ydjG* (synonymous_variant c.669A>G p.Gly223Gly), *yeaM* (frameshift_variant c.523delG p.Ala175fs), *yeaN* (missense_variant c.709A>G p.Ile237Val), *proQ* (frameshift_variant c.254delG p.Gly85fs), *fliL* (missense_variant c.55C>A p.Leu19Met), *asmA* (synonymous_variant c.1216C>T p.Leu406Leu), *mdtC* (missense_variant c.2122T>C p.Tyr708His), *fruK* (missense_variant c.25A>G p.Thr9Ala), *gyrA* (synonymous_variant c.1647T>C p.Val549Val), *ais* (missense_variant c.560A>G p.Glu187Gly), *arnC* (missense_variant c.563C>G p.Ala188Gly), yfbK (missense_variant c.1117T>C p.Tyr373His), yfbR (missense_variant c.83A>G p.Glu28Gly), flk (frameshift_variant c.289delC p.Arg97fs), *mnmC* (frameshift_variant c.897delC p.Ala300fs), *ypdA_2* (missense_variant c.334A>G p.Ile112Val), *ypdE* (synonymous_variant c.882C>T p.Gly294Gly), *fryC* (synonymous_variant c.561A>G p.Leu187Leu), *hemF* (missense_variant c.824A>G p.Glu275Gly), *hyfD* (frameshift_variant c.1124dupG p.Cys377fs), *pbpC* (frameshift_variant c.709delC p.Arg237fs), *yfjW* (synonymous_variant c.1512A>G p.Ala504Ala), *nrdE* (synonymous_variant c.1650C>T p.Gly550Gly), *ascF* (missense_variant c.265G>A p.Ala89Thr), *mutS* (frameshift_variant c.1057delG p.Ala353fs), *mutS* (frameshift_variant c.1314_1315delGC p.Leu439fs), *fucA* (missense_variant c.622A>C p.Thr208Pro), *guaD* (frameshift_variant c.939delA p.Ala314fs), *speA* (missense_variant c.1921T>C p.Phe641Leu), *metK* (missense_variant c.19A>C p.Thr7Pro), *hybB* (missense_variant c.1013A>G p.Asn338Ser), *tolC* (synonymous_variant c.1416A>G p.Ala472Ala), *yqiI* (missense_variant c.869G>A p.Gly290Asp), *mtr* (frameshift_variant c.1011dupG p.Leu338fs), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *argD* (missense_variant c.559A>G p.Thr187Ala), *cysG* (missense_variant c.182T>C p.Val61Ala), *trpS* (missense_variant c.80A>C p.Gln27Pro),

	envZ (frameshift_variant c.721dupG p.Val241fs), *bcsC* (missense_variant c.3271A>G p.Ser1091Gly), *xylB* (frameshift_variant c.1261dupG p.Asp421fs), *rfaL* (synonymous_variant c.246C>G p.Gly82Gly), *rfaL* (missense_variant c.258A>T p.Leu86Phe), *yihP* (missense_variant c.677C>T p.Ala226Val), *fdoH* (missense_variant c.115T>C p.Cys39Arg), *fieF* (missense_variant c.461A>G p.Tyr154Cys), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *lamB* (synonymous_variant c.1110T>C p.Ile370Ile), zur (synonymous_variant c.435A>G p.Ala145Ala), *proP* (missense_variant c.167T>C p.Phe56Ser), *cadC* (missense_variant c.643T>C p.Tyr215His), *dcuA* (frameshift_variant c.64delA p.Ile22fs), *aspA* (synonymous_variant c.309T>C p.Gly103Gly), *priB* (missense_variant c.4A>G p.Thr2Ala), *ytfF* (frameshift_variant c.685delC p.Arg229fs), *yjhR* (frameshift_variant c.274dupG p.Ala92fs)
6TUB118-MB4	*can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *glnD* (missense_variant c.892C>T p.Arg298Cys), *yaiP* (missense_variant c.1004T>C p.Val335Ala), *ybaL* (frameshift_variant c.173delC p.Pro58fs), *allE* (frameshift_variant c.672delA p.Gly225fs), *citC* (missense_variant c.430G>A p.Gly144Arg), *leuS* (missense_variant c.467T>G p.Val156Gly), *gltJ* (missense_variant c.236T>C p.Val79Ala), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *ybjS* (missense_variant c.169T>C p.Ser57Pro), *cydD* (missense_variant c.188T>C p.Leu63Pro), *pqiA* (synonymous_variant c.603C>T p.Ala201Ala), *rlmI* (frameshift_variant c.504delC p.Val169fs), *ymdC* (frameshift_variant c.126delT p.Pro43fs), *lpxL* (synonymous_variant c.699C>T p.Gly233Gly), *holB* (missense_variant c.794T>C p.Val265Ala), *umuC* (synonymous_variant c.1212A>G p.Arg404Arg), *yciC* (synonymous_variant c.483A>G p.Ala161Ala), *pspF* (synonymous_variant c.562T>C p.Leu188Leu), *narZ* (synonymous_variant c.996C>T p.Arg332Arg), *rsxB* (missense_variant c.63T>G p.Ile21Met), *rsxD* (frameshift_variant c.274dupC p.Leu92fs), *ydhV* (frameshift_variant c.710delC p.Pro237fs), *pykF* (frameshift_variant c.902_903insG p.Ile302fs), *proQ* (frameshift_variant c.254delG p.Gly85fs), *fliL* (missense_variant c.55C>A p.Leu19Met), *asmA* (synonymous_variant c.1216C>T p.Leu406Leu), *gyrA* (synonymous_variant c.1647T>C p.Val549Val), *arnC* (missense_variant c.563C>G p.Ala188Gly), *elaD* (frameshift_variant c.109delA p.Thr37fs), *yfbK* (missense_variant c.1117T>C p.Tyr373His), *flk* (frameshift_variant c.289delC p.Arg97fs), *fryC* (synonymous_variant c.561A>G p.Leu187Leu), *hemF* (missense_variant c.824A>G p.Glu275Gly), *ascF* (missense_variant c.265G>A p.Ala89Thr), *mutS* (frameshift_variant c.1057delG p.Ala353fs), *speA* (missense_variant c.1921T>C p.Phe641Leu), *metK* (missense_variant c.19A>C p.Thr7Pro), *yqiI* (missense_variant c.869G>A p.Gly290Asp), *sstT* (frameshift_variant c.1041delG p.Ser348fs), *mtr* (frameshift_variant c.1011dupG p.Leu338fs), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *bcsC* (missense_variant c.3271A>G p.Ser1091Gly), *yibJ* (missense_variant c.97C>A p.Gln33Lys), *htrL* (stop_gained c.5_8delTCTTinsACTA p.ValLeu2*), *yihP* (missense_variant c.677C>T p.Ala226Val), *fdoH* (missense_variant c.115T>C p.Cys39Arg), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *lamB* (synonymous_variant c.1110T>C p.Ile370Ile), *proP* (missense_variant c.167T>C p.Phe56Ser), *cadC* (missense_variant c.643T>C p.Tyr215His), *aspA* (synonymous_variant c.309T>C p.Gly103Gly)
6TUB123-MB4	*can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *glnD* (missense_variant c.892C>T p.Arg298Cys), *lpxA* (synonymous_variant c.777T>C p.Gly259Gly), *yaiP* (missense_variant c.1004T>C p.Val335Ala), *brnQ* (frameshift_variant c.284delC p.Pro95fs), *dxs* (synonymous_variant c.504C>T p.Gly168Gly), *cyoE* (synonymous_variant c.468T>C p.Gly156Gly), *ybaL* (frameshift_variant c.173delC p.Pro58fs), *allE* (frameshift_variant c.672delA p.Gly225fs), *ybcJ* (stop_lost&splice_region_variant c.213A>T p.Ter71Cysext*?), *citC* (missense_variant c.430G>A p.Gly144Arg), *leuS* (missense_variant c.1289C>T p.Ala430Val), *leuS* (missense_variant c.752A>G p.Asp251Gly), *leuS* (missense_variant c.467T>G p.Val156Gly), *gltJ* (missense_variant c.236T>C p.Val79Ala), *kdpD* (missense_variant c.1186C>T p.Arg396Cys), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *rimO* (synonymous_variant c.1263T>G p.Gly421Gly), *ybjS* (missense_variant c.169T>C p.Ser57Pro), *clpA* (missense_variant c.2129A>G p.Asp710Gly), *cydD* (missense_variant c.188T>C p.Leu63Pro), *focA* (synonymous_variant c.306T>C p.Ser102Ser), *pqiA* (synonymous_variant c.603C>T p.Ala201Ala), *rlmI* (frameshift_variant c.504delC p.Val169fs), *insAB-1* (missense_variant c.69G>T p.Arg23Ser), *torA* (frameshift_variant c.1661delT p.Val554fs), *ymdC* (frameshift_variant c.126delT p.Pro43fs), *lpxL* (synonymous_variant c.699C>T p.Gly233Gly), *holB* (missense_variant c.794T>C p.Val265Ala), *ymfS* (intragenic_variant n.1171488delA), *umuC* (synonymous_variant c.1212A>G p.Arg404Arg), *yciC* (synonymous_variant c.483A>G p.Ala161Ala), *pspF* (synonymous_variant c.562T>C p.Leu188Leu), *narZ* (synonymous_variant c.996C>T p.Arg332Arg), *rsxB* (missense_variant c.63T>G p.Ile21Met), *rsxD* (frameshift_variant c.274dupC p.Leu92fs), *rsxE* (synonymous_variant c.246T>C p.Ala82Ala), *ydhV* (frameshift_variant c.710delC p.Pro237fs), *pykF* (frameshift_variant c.902_903insG p.Ile302fs), *astD* (missense_variant c.374A>G p.Asp125Gly), *yeaM* (frameshift_variant c.523delG p.Ala175fs), *proQ* (frameshift_variant c.254delG p.Gly85fs), *fliL* (missense_variant c.55C>A p.Leu19Met), *asmA* (synonymous_variant c.1216C>T p.Leu406Leu), *fruK* (missense_variant c.25A>G p.Thr9Ala), *gyrA* (synonymous_variant c.1647T>C p.Val549Val), *arnC* (missense_variant c.563C>G p.Ala188Gly), *yfbK* (missense_variant c.1117T>C p.Tyr373His), *flk* (frameshift_variant c.289delC p.Arg97fs), *ypdA_2* (missense_variant c.334A>G p.Ile112Val), *fryC* (synonymous_variant c.561A>G p.Leu187Leu), *hemF* (missense_variant c.824A>G p.Glu275Gly), *hyfB* (frameshift_variant c.200delC p.Pro67fs), *hyfD* (frameshift_variant c.94dupG p.Val32fs), *hyfD* (synonymous_variant c.384A>G p.Gly128Gly), *pepB* (synonymous_variant c.861T>C p.Gly287Gly), *srmB* (missense_variant c.1207G>A p.Ala403Thr), *gabD* (frameshift_variant c.1059_1060delCG p.Val354fs), *hypF* (missense_variant c.1408A>G p.Thr470Ala), *ascF* (missense_variant c.265G>A p.Ala89Thr), *ascB* (synonymous_variant c.717C>T p.Gly239Gly), *mutS* (frameshift_variant c.1057delG p.Ala353fs), *speA* (missense_variant c.1921T>C p.Phe641Leu), *metK* (missense_variant c.19A>C p.Thr7Pro), *yqiI* (missense_variant c.869G>A p.Gly290Asp), *mtr* (frameshift_variant c.1011dupG

	p.Leu338fs), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *gspD* (missense_variant c.224A>G p.Tyr75Cys), *trpS* (missense_variant c.80A>C p.Gln27Pro), *yhgE* (missense_variant c.526A>G p.Thr176Ala), *torR_3* (missense_variant c.410T>C p.Val137Ala), *bcsC* (missense_variant c.3271A>G p.Ser1091Gly), *yihP* (missense_variant c.677C>T p.Ala226Val), *fdoH* (missense_variant c.115T>C p.Cys39Arg), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *lamB* (synonymous_variant c.1110T>C p.Ile370Ile), *proP* (missense_variant c.167T>C p.Phe56Ser), *cadC* (missense_variant c.643T>C p.Tyr215His), *aspA* (synonymous_variant c.309T>C p.Gly103Gly), *creC* (synonymous_variant c.255C>T p.Gly85Gly)
6TUB128-MB4	*leuD* (synonymous_variant c.267G>A p.Ala89Ala), *leuC* (synonymous_variant c.1053C>T p.Arg351Arg), *leuA* (synonymous_variant c.945A>C p.Thr315Thr), *secM* (synonymous_variant c.531G>A p.Thr177Thr), *cueO* (synonymous_variant c.1413T>C p.Asp471Asp), *can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *dnaE* (missense_variant c.3233A>G p.Glu1078Gly), *yafE* (missense_variant c.311A>G p.His104Arg), *pepD* (missense_variant c.125A>G p.Glu42Gly), *paoC* (missense_variant c.380C>T p.Ala127Val), *ecpE* (frameshift_variant c.259dupC p.Arg87fs), *prpD* (synonymous_variant c.1122A>G p.Lys374Lys), *adrA* (frameshift_variant c.421delC p.Arg141fs), *ppiD* (missense_variant c.1207G>A p.Ala403Thr), *hha* (missense_variant c.143A>G p.Asp48Gly), *ybaL* (frameshift_variant c.570dupG p.Ile191fs), *ybcK* (missense_variant c.622C>T p.Arg208Cys), *fes* (missense_variant c.311T>G p.Val104Gly), *leuS* (missense_variant c.2290T>C p.Phe764Leu), *leuS* (missense_variant c.467T>G p.Val156Gly), *kdpD* (synonymous_variant c.330A>G p.Glu110Glu), *ybgO* (synonymous_variant c.336A>G p.Gly112Gly), *aroG* (missense_variant c.490G>A p.Ala164Thr), *ybhI* (synonymous_variant c.1176A>G p.Leu392Leu), *rhtA* (missense_variant c.29T>C p.Val10Ala), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *focA* (missense_variant c.524T>C p.Val175Ala), *matP* (missense_variant c.380A>G p.Glu127Gly), *rlmI* (missense_variant c.1018A>G p.Thr340Ala), *rlmI* (missense_variant c.928T>C p.Cys310Arg), *rlmI* (frameshift_variant c.223delT p.Ser75fs), *torS* (frameshift_variant c.2623dupC p.Leu875fs), *torC* (synonymous_variant c.30A>G p.Arg10Arg), *ycdT* (frameshift_variant c.483delG p.Pro162fs), *flgK* (frameshift_variant c.43delG p.Ala15fs), *fabD* (missense_variant c.835A>G p.Lys279Glu), *holB* (synonymous_variant c.993T>G p.Val331Val), *pepT* (missense_variant c.1172A>G p.Glu391Gly), *ymgG* (synonymous_variant c.162T>C p.Gly54Gly), *cvrA* (missense_variant c.1034T>C p.Val345Ala), *dhaR* (frameshift_variant c.739delG p.Ala247fs), *ychF* (missense_variant c.278A>C p.Asn93Thr), *trpA* (missense_variant c.643T>C p.Ser215Pro), *acnA* (missense_variant c.185A>G p.Asp62Gly), *yciT* (missense_variant c.262G>A p.Glu88Lys), *rnb* (missense_variant c.1682A>G p.Asp561Gly), *yciW* (frameshift_variant c.156delC p.Phe53fs), *ydbH* (missense_variant c.940C>T p.Leu314Phe), *ydbC* (frameshift_variant c.95delC p.Pro32fs), *gapA1* (missense_variant c.17T>C p.Val6Ala), *ydcT* (missense_variant c.68T>C p.Val23Ala), *ydcU* (missense_variant c.128T>G p.Val43Gly), *ansP* (missense_variant c.899T>C p.Val300Ala), *narV* (disruptive_inframe_deletion c.390_392delCAT p.Ile130del), *narY* (missense_variant c.1060T>C p.Tyr354His), *narU* (synonymous_variant c.234A>G p.Leu78Leu), *fdnI* (missense_variant c.541A>G p.Lys181Glu), *ddpC* (missense_variant c.428T>C p.Val143Ala), *yddW* (missense_variant c.323T>C p.Phe108Ser), *ydeP* (missense_variant c.893A>G p.Asp298Gly), *ynfF* (missense_variant c.2122A>T p.Ile708Phe), *nemA* (frameshift_variant c.836dupG p.Glu280fs), *mdtK* (missense_variant c.1049T>C p.Val350Ala), *ydhX* (frameshift_variant c.322delG p.Ala108fs), *ydiP* (missense_variant c.242T>C p.Val81Ala), *ydiS* (synonymous_variant c.855T>C p.Gly285Gly), *astD* (missense_variant c.964A>G p.Ile322Val), *ynjI* (missense_variant c.605A>G p.Glu202Gly), *ydjE* (frameshift_variant c.376delG p.Ala126fs), *pabB* (missense_variant c.1118T>C p.Val373Ala), *yoaE* (missense_variant c.17A>C p.Asp6Ala), *kdgR* (missense_variant c.310A>G p.Thr104Ala), *proQ* (frameshift_variant c.63delT p.His23fs), *rsmF* (synonymous_variant c.723T>C p.Gly241Gly), *yebA* (missense_variant c.1255A>G p.Thr419Ala), *yecC* (missense_variant c.467C>A p.Ala156Glu), *fliL* (missense_variant c.55C>A p.Leu19Met), *fcl* (missense_variant c.176T>C p.Val59Ala), *wzc* (missense_variant c.725T>C p.Ile242Thr), *yegS* (missense_variant c.103A>G p.Ile35Val), *yejO* (frameshift_variant c.346delG p.Ala116fs), *napA* (missense_variant c.1817T>C p.Val606Ala), *atoC* (missense_variant c.1031A>G p.Asp344Gly), *arnC* (missense_variant c.563C>G p.Ala188Gly), *yfbP* (missense_variant c.569A>T p.Asp190Val), *yfbS* (synonymous_variant c.1182A>G p.Lys394Lys), *ypdA_2* (missense_variant c.1388A>G p.Asn463Ser), *murP* (missense_variant c.43A>G p.Thr15Ala), *eutE* (synonymous_variant c.913C>T p.Leu305Leu), *csiE* (missense_variant c.1039A>G p.Asn347Asp), *hmp* (frameshift_variant c.110dupC p.Glu38fs), *yfhL* (missense_variant c.73T>C p.Ser25Pro), *pheA* (missense_variant c.986A>G p.Glu329Gly), *recN* (missense_variant c.1616T>C p.Val539Ala), *srlE* (missense_variant c.434A>G p.Asp145Gly), *norR* (missense_variant c.257T>C p.Val86Ala), *mutS* (missense_variant c.515G>C p.Arg172Pro), *ptrA* (missense_variant c.2741T>C p.Val914Ala), *aas* (missense_variant c.1540A>G p.Thr514Ala), *xdhA* (synonymous_variant c.1323C>T p.Gly441Gly), *uacT* (missense_variant c.839A>G p.Glu280Gly), *ygfB* (missense_variant c.455A>G p.Glu152Gly), *speA* (missense_variant c.409T>C p.Ser137Pro), *metK* (missense_variant c.349G>A p.Ala117Thr), *mutY* (missense_variant c.469A>G p.Lys157Glu), *yghZ* (missense_variant c.499A>G p.Ile167Val), *parE* (synonymous_variant c.1137T>G p.Arg379Arg), *ygiC* (missense_variant c.191A>G p.Glu64Gly), *yqiK* (frameshift_variant c.153delA p.Val52fs), *fadH* (missense_variant c.1382T>A p.Ile461Asn), *yhaB* (frameshift_variant c.451dupC p.Leu151fs), *lptA* (missense_variant c.107T>A p.Ile36Asn), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *yhdH* (missense_variant c.698A>G p.Tyr233Cys), *yhdU* (missense_variant c.19T>G p.Trp7Gly), *trpS* (missense_variant c.80A>C p.Gln27Pro), *trpS* (synonymous_variant c.73C>T p.Leu25Leu), *envZ* (frameshift_variant c.721dupG p.Val241fs), *glgX* (frameshift_variant c.249delG p.His84fs), *livH* (missense_variant c.499G>A p.Ala167Thr), *gadA* (synonymous_variant c.1179C>T p.Tyr393Tyr), *gadA* (synonymous_variant c.699C>T p.Asp233Asp), *gadA* (synonymous_variant c.642C>T p.Thr214Thr), *yhjB* (missense_variant c.341T>C p.Val114Ala), *yhjG* (missense_variant c.2035A>C p.Thr679Pro), *yhjV* (missense_variant c.868A>G p.Ile290Val), *yiaT* (frameshift_variant c.418dupG p.Glu140fs), *selA* (frameshift_variant c.1123delC p.Ala376fs), *dgoR* (synonymous_variant c.666A>C p.Thr222Thr), *yifB* (frameshift_variant c.193dupA p.Ile65fs), *ilvY* (missense_variant c.506T>G p.Ile169Ser), *yihL* (frameshift_variant c.397dupG p.Ala133fs), *cytR* (missense_variant c.361A>C p.Thr121Pro), *metF* (missense_variant c.118A>C p.Thr40Pro), *ppc* (synonymous_variant

	c.1023A>G p.Lys341Lys), *murI* (missense_variant c.173T>C p.Phe58Ser), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *rpoC* (missense_variant c.287A>G p.Lys96Arg), *thiS* (missense_variant c.68A>T p.Glu23Val), *yjaB* (frameshift_variant c.210delG p.Gln71fs), *adiA* (frameshift_variant c.291delA p.Ala98fs), *yeiP_2* (missense_variant c.532A>C p.Thr178Pro), *mscM* (frameshift_variant c.49dupG p.Ala17fs), *ytfB* (missense_variant c.385C>T p.Arg129Cys), *yjgL* (missense_variant c.1781A>G p.His594Arg), *idnD* (frameshift_variant c.699delA p.Gly234fs), *idnK* (frameshift_variant c.261delA p.Lys87fs), *yjiL* (missense_variant c.461A>C p.His154Pro)
6TUB138-MB4	*polB* (synonymous_variant c.831T>C p.Gly277Gly), *can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *glnD* (missense_variant c.892C>T p.Arg298Cys), *yaiP* (missense_variant c.1004T>C p.Val335Ala), *ybaL* (frameshift_variant c.173delC p.Pro58fs), *allE* (frameshift_variant c.672delA p.Gly225fs), *ybcO* (frameshift_variant c.277dupG p.Val93fs), *citC* (missense_variant c.430G>A p.Gly144Arg), *leuS* (missense_variant c.467T>G p.Val156Gly), *gltJ* (missense_variant c.236T>C p.Val79Ala), *kdpD* (missense_variant c.1186C>T p.Arg396Cys), *abrB* (synonymous_variant c.213T>C p.Gly71Gly), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *ybjS* (missense_variant c.169T>C p.Ser57Pro), *cydD* (missense_variant c.188T>C p.Leu63Pro), *pqiA* (synonymous_variant c.603C>T p.Ala201Ala), *rlmI* (frameshift_variant c.504delC p.Val169fs), *ymdC* (frameshift_variant c.126delT p.Pro43fs), *lpxL* (synonymous_variant c.699C>T p.Gly233Gly), *holB* (missense_variant c.794T>C p.Val265Ala), *umuC* (synonymous_variant c.1212A>G p.Arg404Arg), *yciC* (synonymous_variant c.483A>G p.Ala161Ala), *yciW* (frameshift_variant c.156delC p.Phe53fs), *pspF* (synonymous_variant c.562T>C p.Leu188Leu), *narZ* (synonymous_variant c.996C>T p.Arg332Arg), *maeA* (missense_variant c.889A>G p.Lys297Glu), *marA* (missense_variant c.277A>C p.Thr93Pro), *rsxB* (missense_variant c.63T>G p.Ile21Met), *rsxD* (frameshift_variant c.274dupC p.Leu92fs), *ydhV* (frameshift_variant c.710delC p.Pro237fs), *pykF* (frameshift_variant c.902_903insG p.Ile302fs), *aroH* (missense_variant c.574A>C p.Thr192Pro), *yeaN* (missense_variant c.709A>G p.Ile237Val), *proQ* (frameshift_variant c.254delG p.Gly85fs), *fliL* (missense_variant c.55C>A p.Leu19Met), *asmA* (synonymous_variant c.1216C>T p.Leu406Leu), *fruK* (missense_variant c.25A>G p.Thr9Ala), *gyrA* (synonymous_variant c.1647T>C p.Val549Val), *ais* (missense_variant c.560A>G p.Glu187Gly), *arnC* (missense_variant c.563C>G p.Ala188Gly), *yfbK* (missense_variant c.1117T>C p.Tyr373His), *flk* (frameshift_variant c.289delC p.Arg97fs), *ypdA_2* (missense_variant c.334A>G p.Ile112Val), *fryC* (synonymous_variant c.561A>G p.Leu187Leu), *hemF* (missense_variant c.824A>G p.Glu275Gly), *ascF* (missense_variant c.265G>A p.Ala89Thr), *mutS* (frameshift_variant c.1057delG p.Ala353fs), *guaD* (frameshift_variant c.939delA p.Ala314fs), *speA* (missense_variant c.1921T>C p.Phe641Leu), *metK* (missense_variant c.19A>C p.Thr7Pro), *yqiI* (missense_variant c.869G>A p.Gly290Asp), *mtr* (frameshift_variant c.1011dupG p.Leu338fs), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *envZ* (frameshift_variant c.721dupG p.Val241fs), *bcsC* (missense_variant c.3271A>G p.Ser1091Gly), *yibJ* (missense_variant c.97C>A p.Gln33Lys), *yihP* (missense_variant c.677C>T p.Ala226Val), *fdoH* (missense_variant c.115T>C p.Cys39Arg), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), lamB (synonymous_variant c.1110T>C p.Ile370Ile), *proP* (missense_variant c.167T>C p.Phe56Ser), *cadC* (missense_variant c.643T>C p.Tyr215His), *aspA* (synonymous_variant c.309T>C p.Gly103Gly), *yjhR* (frameshift_variant c.274dupG p.Ala92fs)
6TUB148-MB4	*polB* (synonymous_variant c.831T>C p.Gly277Gly), *can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *glnD* (missense_variant c.892C>T p.Arg298Cys), *ykfM* (missense_variant c.249_252delGAAGinsCAAC p.Lys84Asn), *ykfM* (synonymous_variant c.240C>A p.Ile80Ile), *ykfM* (missense_variant c.232T>A p.Phe78Ile), *ykfM* (synonymous_variant c.219G>C p.Thr73Thr), *ykfM* (missense_variant c.198_204delTATTCATinsAATTAAA p.AsnIleHis66LysIleLys), *ykfM* (missense_variant c.188_190delCCTinsACA p.SerPhe63TyrIle), *yaiP* (missense_variant c.1004T>C p.Val335Ala), *ybaL* (frameshift_variant c.173delC p.Pro58fs), *allE* (frameshift_variant c.672delA p.Gly225fs), *ybcO* (frameshift_variant c.277dupG p.Val93fs), *citC* (missense_variant c.430G>A p.Gly144Arg), *leuS* (missense_variant c.467T>G p.Val156Gly), *gltJ* (missense_variant c.236T>C p.Val79Ala), *kdpD* (missense_variant c.1186C>T p.Arg396Cys), *abrB* (synonymous_variant c.213T>C p.Gly71Gly), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *ybjS* (missense_variant c.169T>C p.Ser57Pro), *cydD* (missense_variant c.188T>C p.Leu63Pro), *pqiA* (synonymous_variant c.603C>T p.Ala201Ala), *rlmI* (frameshift_variant c.504delC p.Val169fs), *ymdC* (frameshift_variant c.126delT p.Pro43fs), *lpxL* (synonymous_variant c.699C>T p.Gly233Gly), *holB* (missense_variant c.794T>C p.Val265Ala), *umuC* (synonymous_variant c.1212A>G p.Arg404Arg), *yciC* (synonymous_variant c.483A>G p.Ala161Ala), *yciW* (frameshift_variant c.156delC p.Phe53fs), *pspF* (synonymous_variant c.562T>C p.Leu188Leu), *narZ* (synonymous_variant c.996C>T p.Arg332Arg), *maeA* (missense_variant c.889A>G p.Lys297Glu), *marA* (missense_variant c.277A>C p.Thr93Pro), *rsxB* (missense_variant c.63T>G p.Ile21Met), *rsxD* (frameshift_variant c.274dupC p.Leu92fs), *ydhV* (frameshift_variant c.710delC p.Pro237fs), *pykF* (frameshift_variant c.902_903insG p.Ile302fs), *aroH* (missense_variant c.574A>C p.Thr192Pro), *yeaN* (missense_variant c.709A>G p.Ile237Val), *proQ* (frameshift_variant c.254delG p.Gly85fs), *fliL* (missense_variant c.55C>A p.Leu19Met), *asmA* (synonymous_variant c.1216C>T p.Leu406Leu), *fruK* (missense_variant c.25A>G p.Thr9Ala), *yejO* (frameshift_variant c.346delG p.Ala116fs), *gyrA* (synonymous_variant c.1647T>C p.Val549Val), *ais* (missense_variant c.560A>G p.Glu187Gly), *arnC* (missense_variant c.563C>G p.Ala188Gly), *yfbK* (missense_variant c.1117T>C p.Tyr373His), *flk* (frameshift_variant c.289delC p.Arg97fs), *ypdA_2* (missense_variant c.334A>G p.Ile112Val), *fryC* (synonymous_variant c.561A>G p.Leu187Leu), *hemF* (missense_variant c.824A>G p.Glu275Gly), *ascF* (missense_variant c.265G>A p.Ala89Thr), *mutS* (frameshift_variant c.1057delG p.Ala353fs), *guaD* (frameshift_variant c.939delA p.Ala314fs), *speA* (missense_variant c.1921T>C p.Phe641Leu), *metK* (missense_variant c.19A>C p.Thr7Pro), *yqiI* (missense_variant c.869G>A p.Gly290Asp), *mtr* (frameshift_variant c.1011dupG p.Leu338fs), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *envZ* (frameshift_variant c.721dupG

	p.Val241fs), *bcsC* (missense_variant c.3271A>G p.Ser1091Gly), *yibJ* (missense_variant c.97C>A p.Gln33Lys), *yihP* (missense_variant c.677C>T p.Ala226Val), *fdoH* (missense_variant c.115T>C p.Cys39Arg), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *arpA_2* (stop_gained c.247G>T p.Glu83*), *lamB* (synonymous_variant c.1110T>C p.Ile370Ile), *uvrA* (missense_variant c.2287T>C p.Cys763Arg), *proP* (missense_variant c.167T>C p.Phe56Ser), *cadC* (missense_variant c.643T>C p.Tyr215His), *aspA* (synonymous_variant c.309T>C p.Gly103Gly), *yjhR* (frameshift_variant c.274dupG p.Ala92fs), *fimD* (missense_variant c.1376T>C p.Val459Ala)
6TUB158-MB4	*can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *glnD* (missense_variant c.892C>T p.Arg298Cys), *yaiP* (missense_variant c.1004T>C p.Val335Ala), *ybaL* (frameshift_variant c.173delC p.Pro58fs), *allE* (frameshift_variant c.672delA p.Gly225fs), *ybcO* (frameshift_variant c.277dupG p.Val93fs), *citC* (missense_variant c.430G>A p.Gly144Arg), *leuS* (missense_variant c.467T>G p.Val156Gly), *gltJ* (missense_variant c.236T>C p.Val79Ala), *kdpD* (missense_variant c.1186C>T p.Arg396Cys), *abrB* (synonymous_variant c.213T>C p.Gly71Gly), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *ybjS* (missense_variant c.169T>C p.Ser57Pro), *cydD* (missense_variant c.188T>C p.Leu63Pro), *ycbB* (frameshift_variant c.804delG p.Pro269fs), *pqiA* (synonymous_variant c.603C>T p.Ala201Ala), *rlmI* (frameshift_variant c.504delC p.Val169fs), *insA-2* (missense_variant c.243_249delCCGCCATinsACGTCAC p.Phe81Leu), *insAB-1* (synonymous_variant c.18_21delATACinsCTAT p.8), *insAB-1* (missense_variant c.69G>T p.Arg23Ser), *ymdC* (frameshift_variant c.126delT p.Pro43fs), *lpxL* (synonymous_variant c.699C>T p.Gly233Gly), *holB* (missense_variant c.794T>C p.Val265Ala), *umuC* (synonymous_variant c.1212A>G p.Arg404Arg), *yciC* (synonymous_variant c.483A>G p.Ala161Ala), *yciW* (frameshift_variant c.156delC p.Phe53fs), *pspF* (synonymous_variant c.562T>C p.Leu188Leu), *narZ* (synonymous_variant c.996C>T p.Arg332Arg), *maeA* (missense_variant c.889A>G p.Lys297Glu), *ydeK* (frameshift_variant c.3962delG p.Gly1321fs), *marA* (missense_variant c.277A>C p.Thr93Pro), *ydfO* (missense_variant c.354T>G p.Asp118Glu), *gnsB* (missense_variant c.97G>C p.Val33Leu), *rsxB* (missense_variant c.63T>G p.Ile21Met), *rsxD* (frameshift_variant c.274dupC p.Leu92fs), *ydhV* (frameshift_variant c.710delC p.Pro237fs), *pykF* (frameshift_variant c.902_903insG p.Ile302fs), *ydiK* (frameshift_variant c.405dupG p.Thr136fs), *aroH* (missense_variant c.574A>C p.Thr192Pro), *yeaN* (missense_variant c.709A>G p.Ile237Val), *proQ* (frameshift_variant c.254delG p.Gly85fs), *fliL* (missense_variant c.55C>A p.Leu19Met), *asmA* (synonymous_variant c.1216C>T p.Leu406Leu), *fruK* (missense_variant c.25A>G p.Thr9Ala), *yejO* (frameshift_variant c.346delG p.Ala116fs), *gyrA* (synonymous_variant c.1647T>C p.Val549Val), *ais* (missense_variant c.560A>G p.Glu187Gly), *arnC* (missense_variant c.563C>G p.Ala188Gly), *yfbK* (missense_variant c.1117T>C p.Tyr373His), *flk* (frameshift_variant c.289delC p.Arg97fs), *ypdA_2* (missense_variant c.334A>G p.Ile112Val), *fryC* (synonymous_variant c.561A>G p.Leu187Leu), *hemF* (missense_variant c.824A>G p.Glu275Gly), *ascF* (missense_variant c.265G>A p.Ala89Thr), *mutS* (frameshift_variant c.1057delG p.Ala353fs), *guaD* (frameshift_variant c.939delA p.Ala314fs), *speA* (missense_variant c.1921T>C p.Phe641Leu), *metK* (missense_variant c.19A>C p.Thr7Pro), *yqiI* (missense_variant c.869G>A p.Gly290Asp), *mtr* (frameshift_variant c.1011dupG p.Leu338fs), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *envZ* (frameshift_variant c.721dupG p.Val241fs), *gadA* (synonymous_variant c.642C>T p.Thr214Thr), *bcsC* (missense_variant c.3271A>G p.Ser1091Gly), *yibJ* (missense_variant c.97C>A p.Gln33Lys), *yihP* (missense_variant c.677C>T p.Ala226Val), *fdoH* (missense_variant c.115T>C p.Cys39Arg), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *lamB* (synonymous_variant c.1110T>C p.Ile370Ile), *proP* (missense_variant c.167T>C p.Phe56Ser), *cadC* (missense_variant c.643T>C p.Tyr215His), *aspA* (synonymous_variant c.309T>C p.Gly103Gly), *yjhR* (frameshift_variant c.274dupG p.Ala92fs)
6TUB165-MB4	*yaaU* (missense_variant c.1088T>C p.Phe363Ser), *polB* (synonymous_variant c.831T>C p.Gly277Gly), *yacH* (missense_variant c.1405T>C p.Ser469Pro), *can* (missense_variant c.167A>G p.Glu56Gly), *yadD* (synonymous_variant c.213C>T p.Gly71Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *glnD* (missense_variant c.892C>T p.Arg298Cys), *lpxB* (missense_variant c.106G>A p.Ala36Thr), *yaiP* (missense_variant c.1004T>C p.Val335Ala), *ybaL* (frameshift_variant c.173delC p.Pro58fs), *allE* (frameshift_variant c.672delA p.Gly225fs), *ybcO* (frameshift_variant c.277dupG p.Val93fs), *nfrA* (missense_variant c.257A>G p.Tyr86Cys), *nfrB* (missense_variant c.1631C>T p.Ala544Val), *citC* (missense_variant c.430G>A p.Gly144Arg), *rlpA* (synonymous_variant c.747T>C p.Gly249Gly), *leuS* (missense_variant c.1265T>C p.Val422Ala), *leuS* (missense_variant c.707T>C p.Val236Ala), *leuS* (missense_variant c.467T>G p.Val156Gly), *gltJ* (missense_variant c.236T>C p.Val79Ala), *lnt* (synonymous_variant c.492A>G p.Ala164Ala), *chiP* (missense_variant c.1070T>C p.Val357Ala), *kdpD* (missense_variant c.1186C>T p.Arg396Cys), *abrB* (synonymous_variant c.213T>C p.Gly71Gly), *sdhA* (synonymous_variant c.987T>G p.Gly329Gly), *gpmA* (missense_variant c.484A>G p.Ile162Val), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *ybjS* (missense_variant c.169T>C p.Ser57Pro), *clpA* (missense_variant c.76A>C p.Thr26Pro), *cydD* (missense_variant c.188T>C p.Leu63Pro), *ycbB* (frameshift_variant c.804delG p.Pro269fs), *elfC* (missense_variant c.1942A>G p.Lys648Glu), *ycbF* (missense_variant c.293A>G p.Asp98Gly), *pqiA* (synonymous_variant c.603C>T p.Ala201Ala), *rlmI* (frameshift_variant c.504delC p.Val169fs), *ymdC* (frameshift_variant c.126delT p.Pro43fs), *lpxL* (synonymous_variant c.699C>T p.Gly233Gly), *holB* (missense_variant c.794T>C p.Val265Ala), *umuC* (synonymous_variant c.1212A>G p.Arg404Arg), *treA* (synonymous_variant c.1362A>G p.Ala454Ala), *adhE* (synonymous_variant c.1632T>G p.Gly544Gly), *yciC* (synonymous_variant c.483A>G p.Ala161Ala), *trpA* (missense_variant c.304T>C p.Tyr102His), *yciW* (frameshift_variant c.156delC p.Phe53fs), *pspF* (synonymous_variant c.562T>C p.Leu188Leu), *feaR* (stop_gained c.445C>T p.Gln149*), *paaC* (synonymous_variant c.561T>C p.Asp187Asp), *ydcM* (synonymous_variant c.240T>C p.Phe80Phe), *narZ* (synonymous_variant c.996C>T p.Arg332Arg), *maeA* (missense_variant c.889A>G p.Lys297Glu), *ddpX* (synonymous_variant c.285T>C p.Gly95Gly), *ydeK* (frameshift_variant c.3962delG p.Gly1321fs), *marA* (missense_variant c.277A>C p.Thr93Pro), *ydeE* (frameshift_variant c.309dupT

p.Ala104fs), *dmsD* (synonymous_variant c.468T>C p.Phe156Phe), *rsxB* (missense_variant c.63T>G p.Ile21Met), *rsxD* (frameshift_variant c.273_274dupCC p.Leu92fs), *ydhX* (frameshift_variant c.322delG p.Ala108fs), *ydhV* (frameshift_variant c.710delC p.Pro237fs), *pykF* (frameshift_variant c.902_903insG p.Ile302fs), *ydiK* (frameshift_variant c.405dupG p.Thr136fs), *aroH* (missense_variant c.574A>C p.Thr192Pro), *astE* (missense_variant c.908T>C p.Phe303Ser), *yeaN* (missense_variant c.709A>G p.Ile237Val), *yoaE* (frameshift_variant c.1406delG p.Gly469fs), *proQ* (frameshift_variant c.254delG p.Gly85fs), *yebA* (missense_variant c.56A>G p.His19Arg), *motA* (missense_variant c.446A>G p.Glu149Gly), *fliL* (missense_variant c.55C>A p.Leu19Met), *asmA* (synonymous_variant c.1216C>T p.Leu406Leu), *mdtB* (synonymous_variant c.1284A>G p.Lys428Lys), *fruK* (missense_variant c.25A>G p.Thr9Ala), *gyrA* (synonymous_variant c.1647T>C p.Val549Val), *yfaL* (missense_variant c.3311A>C p.Tyr1104Ser), *ais* (missense_variant c.560A>G p.Glu187Gly), *arnC* (missense_variant c.563C>G p.Ala188Gly), *yfbK* (missense_variant c.1117T>C p.Tyr373His), *yfcC* (frameshift_variant c.1403delG p.Gly468fs), *cvpA* (synonymous_variant c.315T>C p.Gly105Gly), *flk* (frameshift_variant c.289delC p.Arg97fs), *ypdA_2* (missense_variant c.334A>G p.Ile112Val), *fryC* (synonymous_variant c.561A>G p.Leu187Leu), *mntH* (synonymous_variant c.69A>G p.Gly23Gly), *hemF* (missense_variant c.824A>G p.Glu275Gly), *eutA* (frameshift_variant c.1057delG p.Ala353fs), *hcaD* (synonymous_variant c.600T>C p.Gly200Gly), *yfjS* (synonymous_variant c.12A>G p.Lys4Lys), *csiD* (missense_variant c.608T>C p.Phe203Ser), *stpA* (synonymous_variant c.135A>G p.Glu45Glu), *emrB* (missense_variant c.1469T>C p.Val490Ala), *pncC* (missense_variant c.314T>C p.Ile105Thr), *ascF* (missense_variant c.265G>A p.Ala89Thr), *mutS* (missense_variant c.896T>C p.Val299Ala), *mutS* (frameshift_variant c.1057delG p.Ala353fs), *cysI* (missense_variant c.665T>C p.Ile222Thr), *cysJ* (missense_variant c.326A>G p.Glu109Gly), *cysJ* (missense_variant c.167C>T p.Ala56Val), *ygcU* (missense_variant c.808A>C p.Thr270Pro), *xdhA* (frameshift_variant c.999dupG p.Asn334fs), *guaD* (frameshift_variant c.939delA p.Ala314fs), *speA* (missense_variant c.1921T>C p.Phe641Leu), *metK* (missense_variant c.19A>C p.Thr7Pro), *yqgF* (missense_variant c.122A>G p.Asp41Gly), *yghE* (missense_variant c.362A>G p.Tyr121Cys), *yghJ* (missense_variant c.2542T>C p.Ser848Pro), *yqiI* (missense_variant c.869G>A p.Gly290Asp), *yraK* (missense_variant c.413T>G p.Val138Gly), *mtr* (frameshift_variant c.1011dupG p.Leu338fs), *deaD* (synonymous_variant c.1203T>C p.Asp401Asp), *murA* (synonymous_variant c.459T>C p.Gly153Gly), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *bfr* (stop_gained c.105G>A p.Trp35*), *trpS* (missense_variant c.80A>C p.Gln27Pro), *yhgE* (missense_variant c.803T>C p.Val268Ala), *envZ* (frameshift_variant c.721dupG p.Val241fs), *bcsC* (missense_variant c.3271A>G p.Ser1091Gly), *bcsZ* (missense_variant c.734T>C p.Ile245Thr), *dppC* (missense_variant c.773T>C p.Val258Ala), *eptB* (frameshift_variant c.920delG p.Gly307fs), *rfaS* (missense_variant c.346A>T p.Ile116Leu), *rfaS* (missense_variant c.325A>T p.Asn109Tyr), *ilvB_1* (missense_variant c.1294G>A p.Ala432Thr), *ilvA* (missense_variant c.1468A>C p.Thr490Pro), *rho* (missense_variant c.10A>G p.Thr4Ala), *yihP* (missense_variant c.677C>T p.Ala226Val), *fdoH* (missense_variant c.115T>C p.Cys39Arg), *fdoG_1* (missense_variant c.2066A>G p.His689Arg), *argE* (synonymous_variant c.966A>G p.Gly322Gly), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *lamB* (synonymous_variant c.1110T>C p.Ile370Ile), *dnaB* (missense_variant c.856A>G p.Thr286Ala), *uvrA* (missense_variant c.2287T>C p.Cys763Arg), *proP* (missense_variant c.167T>C p.Phe56Ser), *lysU* (missense_variant c.1358A>G p.Glu453Gly), *cadC* (missense_variant c.643T>C p.Tyr215His), *aspA* (synonymous_variant c.309T>C p.Gly103Gly), *lptG* (missense_variant c.292A>G p.Thr98Ala), *yjhR* (frameshift_variant c.274dupG p.Ala92fs), *fimD* (missense_variant c.1376T>C p.Val459Ala), *hsdR* (missense_variant c.2843A>G p.Glu948Gly), *mrr* (missense_variant c.863T>C p.Val288Ala)

Table S35| Mutations (SNPs) found in the coding sequences of the 6TUBX-MB3 lineages, relative to the *Escherichia coli* MG1655 NC_000913.3 reference genome. The affected genes and the specific mutation/SNP for genomic samples of 6TUB23, 6TUB53, 6TUB88-MB3, 6TUB95-MB3, 6TUB102-MB3, 6TUB109-MB3, 6TUB116-MB3, 6TUB123-MB3, 6TUB128-MB3, 6TUB130-MB3, 6TUB137-MB3, 6TUB144-MB3, 6TUB151-MB3, 6TUB158-MB3 and 6TUB165-MB3 are shown.

	AFFECTED GENE (SNP)
6TUB23	*ansA* (missense_variant c.423T>G p.Asn141Lys), *rpsC* (missense_variant c.371T>G p.Leu124Arg), *rpoB* (missense_variant c.443A>T p.Gln148Leu)
6TUB53	*can* (missense_variant c.167A>G p.Glu56Gly), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *fliL* (missense_variant c.55C>A p.Leu19Met), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu)

6TUB88-MB3	*can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *lon* (missense_variant c.995A>G p.Asp332Gly), *leuS* (missense_variant c.563T>G p.Ile188Ser), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *proQ* (stop_gained c.280C>T p.Gln94*), *fliL* (missense_variant c.55C>A p.Leu19Met), *metK* (missense_variant c.801T>A p.Asp267Glu), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *lptF* (missense_variant c.349T>C p.Ser117Pro)
6TUB95-MB3	*secM* (synonymous_variant c.531G>A p.Thr177Thr), *can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *dnaE* (missense_variant c.3233A>G p.Glu1078Gly), *hha* (missense_variant c.143A>G p.Asp48Gly), *ybaL* (frameshift_variant c.570dupG p.Ile191fs), *fes* (missense_variant c.311T>G p.Val104Gly), *leuS* (missense_variant c.467T>G p.Val156Gly), *ybhI* (synonymous_variant c.1176A>G p.Leu392Leu), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *rlmI* (missense_variant c.1018A>G p.Thr340Ala), *insAB-1* (missense_variant c.69G>T p.Arg23Ser), *holB* (synonymous_variant c.993T>G p.Val331Val), *yciW* (frameshift_variant c.156delC p.Phe53fs), *ydbC* (frameshift_variant c.95delC p.Pro32fs), *gapA1* (missense_variant c.17T>C p.Val6Ala), *ansP* (missense_variant c.899T>C p.Val300Ala), *ddpC* (missense_variant c.428T>C p.Val143Ala), *ynfF* (missense_variant c.2122A>T p.Ile708Phe), *astD* (missense_variant c.964A>G p.Ile322Val), *fliL* (missense_variant c.55C>A p.Leu19Met), *yejO* (frameshift_variant c.346delG p.Ala116fs), *atoC* (missense_variant c.1031A>G p.Asp344Gly), *arnC* (missense_variant c.563C>G p.Ala188Gly), *ypdA_2* (missense_variant c.1388A>G p.Asn463Ser), *murP* (missense_variant c.43A>G p.Thr15Ala), *eutE* (synonymous_variant c.913C>T p.Leu305Leu), *hmp* (frameshift_variant c.110dupC p.Glu38fs), *mutS* (missense_variant c.515G>C p.Arg172Pro), *metK* (missense_variant c.349G>A p.Ala117Thr), *fadH* (missense_variant c.1382T>A p.Ile461Asn), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *trpS* (synonymous_variant c.73C>T p.Leu25Leu), *glgX* (frameshift_variant c.249delG p.His84fs), *yhjB* (missense_variant c.341T>C p.Val114Ala), *yhjG* (missense_variant c.2035A>C p.Thr679Pro), *yhjV* (missense_variant c.868A>G p.Ile290Val), *yibJ* (missense_variant c.97C>A p.Gln33Lys), *ilvY* (missense_variant c.506T>G p.Ile169Ser), *cytR* (missense_variant c.361A>C p.Thr121Pro), *murI* (missense_variant c.173T>C p.Phe58Ser), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *adiA* (frameshift_variant c.291delA p.Ala98fs), *yjgL* (missense_variant c.1781A>G p.His594Arg), *idnD* (frameshift_variant c.699delA p.Gly234fs)
6TUB102-MB3	*secM* (synonymous_variant c.531G>A p.Thr177Thr), *can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *dnaE* (missense_variant c.3233A>G p.Glu1078Gly), *hha* (missense_variant c.143A>G p.Asp48Gly), *ybaL* (frameshift_variant c.570dupG p.Ile191fs), *fes* (missense_variant c.311T>G p.Val104Gly), *leuS* (missense_variant c.467T>G p.Val156Gly), *ybhI* (synonymous_variant c.1176A>G p.Leu392Leu), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *rlmI* (missense_variant c.1018A>G p.Thr340Ala), *holB* (synonymous_variant c.993T>G p.Val331Val), *yciW* (frameshift_variant c.156delC p.Phe53fs), *ydbC* (frameshift_variant c.95delC p.Pro32fs), *gapA1* (missense_variant c.17T>C p.Val6Ala), *ansP* (missense_variant c.899T>C p.Val300Ala), *ddpC* (missense_variant c.428T>C p.Val143Ala), *ynfF* (missense_variant c.2122A>T p.Ile708Phe), *astD* (missense_variant c.964A>G p.Ile322Val), *fliL* (missense_variant c.55C>A p.Leu19Met), *yejO* (frameshift_variant c.346delG p.Ala116fs), *atoC* (missense_variant c.1031A>G p.Asp344Gly), *arnC* (missense_variant c.563C>G p.Ala188Gly), *ypdA_2* (missense_variant c.1388A>G p.Asn463Ser), *murP* (missense_variant c.43A>G p.Thr15Ala), *eutE* (synonymous_variant c.913C>T p.Leu305Leu), *hmp* (frameshift_variant c.110dupC p.Glu38fs), *mutS* (missense_variant c.515G>C p.Arg172Pro), *metK* (missense_variant c.349G>A p.Ala117Thr), *fadH* (missense_variant c.1382T>A p.Ile461Asn), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *trpS* (synonymous_variant c.73C>T p.Leu25Leu), *glgX* (frameshift_variant c.249delG p.His84fs), *gadA* (synonymous_variant c.642C>T p.Thr214Thr), *yhjB* (missense_variant c.341T>C p.Val114Ala), *yhjG* (missense_variant c.2035A>C p.Thr679Pro), *yhjV* (missense_variant c.868A>G p.Ile290Val), *ilvY* (missense_variant c.506T>G p.Ile169Ser), *cytR* (missense_variant c.361A>C p.Thr121Pro), *murI* (missense_variant c.173T>C p.Phe58Ser), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *adiA* (frameshift_variant c.291delA p.Ala98fs), *yjgL* (missense_variant c.1781A>G p.His594Arg), *idnD* (frameshift_variant c.699delA p.Gly234fs)
6TUB109-MB3	*secM* (synonymous_variant c.531G>A p.Thr177Thr), *aroP* (missense_variant c.1156G>A p.Ala386Thr), *can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *dnaE* (missense_variant c.3233A>G p.Glu1078Gly), *yaiW* (missense_variant c.257A>G p.Glu86Gly), *adrA* (missense_variant c.638A>G p.Asp213Gly), *hha* (missense_variant c.143A>G p.Asp48Gly), *ybaL* (frameshift_variant c.570dupG p.Ile191fs), *fes* (missense_variant c.311T>G p.Val104Gly), *leuS* (missense_variant c.467T>G p.Val156Gly), *ybhI* (synonymous_variant c.1176A>G p.Leu392Leu), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *rlmI* (missense_variant c.1018A>G p.Thr340Ala), *gfcE* (synonymous_variant c.90G>A p.Ser30Ser), *holB* (synonymous_variant c.993T>G p.Val331Val), *dhaR* (frameshift_variant c.739delG p.Ala247fs), *yciW* (frameshift_variant c.156delC p.Phe53fs), *insH-5* (synonymous_variant c.675A>G p.Gly225Gly), *insH-5* (synonymous_variant c.630T>G p.Leu210Leu), *ydbC* (frameshift_variant c.95delC p.Pro32fs), *gapA1* (missense_variant c.17T>C p.Val6Ala), *ansP* (missense_variant c.899T>C p.Val300Ala), *ddpC* (missense_variant c.428T>C p.Val143Ala), *ynfF* (missense_variant c.2122A>T

	p.Ile708Phe), *astD* (missense_variant c.964A>G p.Ile322Val), *fliL* (missense_variant c.55C>A p.Leu19Met), *wzc* (missense_variant c.725T>C p.Ile242Thr), *yejO* (frameshift_variant c.345_346delGG p.Ala116fs), *atoC* (missense_variant c.1031A>G p.Asp344Gly), *glpB* (missense_variant c.176A>C p.Asp59Ala), *arnC* (missense_variant c.563C>G p.Ala188Gly), *ypdA_2* (missense_variant c.1388A>G p.Asn463Ser), *murP* (missense_variant c.43A>G p.Thr15Ala), *eutE* (synonymous_variant c.913C>T p.Leu305Leu), *eutS* (missense_variant c.42A>C p.Lys14Asn), *hmp* (frameshift_variant c.110dupC p.Glu38fs), *mutS* (missense_variant c.515G>C p.Arg172Pro), *metK* (missense_variant c.349G>A p.Ala117Thr), *rpoD* (missense_variant c.379A>T p.Ile127Phe), *fadH* (missense_variant c.1382T>A p.Ile461Asn), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *trpS* (synonymous_variant c.73C>T p.Leu25Leu), *yrfF* (missense_variant c.1355A>G p.Glu452Gly), *glgX* (frameshift_variant c.249delG p.His84fs), *rbbA* (missense_variant c.2255T>C p.Val752Ala), *yhjB* (missense_variant c.341T>C p.Val114Ala), *yhjG* (missense_variant c.2035A>C p.Thr679Pro), *yhjV* (missense_variant c.868A>G p.Ile290Val), *yiaV* (missense_variant c.715A>G p.Thr239Ala), *yibJ* (missense_variant c.97C>A p.Gln33Lys), *rfaL* (missense_variant c.237T>G p.Phe79Leu), *ilvY* (missense_variant c.506T>G p.Ile169Ser), *cytR* (missense_variant c.361A>C p.Thr121Pro), *murI* (missense_variant c.173T>C p.Phe58Ser), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), adiA (frameshift_variant c.291delA p.Ala98fs), *yjgL* (missense_variant c.1781A>G p.His594Arg), *lptF* (missense_variant c.911T>C p.Leu304Pro), *idnD* (frameshift_variant c.699delA p.Gly234fs)
6TUB116-MB3	*secM* (synonymous_variant c.531G>A p.Thr177Thr), *can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *dnaE* (missense_variant c.3233A>G p.Glu1078Gly), *hha* (missense_variant c.143A>G p.Asp48Gly), *ybaL* (frameshift_variant c.570dupG p.Ile191fs), *fes* (missense_variant c.311T>G p.Val104Gly), *leuS* (missense_variant c.467T>G p.Val156Gly), *ybhI* (synonymous_variant c.1176A>G p.Leu392Leu), ybiW (missense_variant c.829A>T p.Ile277Phe), rlmI (missense_variant c.1018A>G p.Thr340Ala), insA-2 (missense_variant c.243_249delCCGCCATinsACGTCAC p.Phe81Leu), *holB* (synonymous_variant c.993T>G p.Val331Val), *dhaR* (frameshift_variant c.739delG p.Ala247fs), *yciW* (frameshift_variant c.156delC p.Phe53fs), *ydbC* (frameshift_variant c.95delC p.Pro32fs), *gapA1* (missense_variant c.17T>C p.Val6Ala), *ansP* (missense_variant c.899T>C p.Val300Ala), *ddpC* (missense_variant c.428T>C p.Val143Ala), *ynfF* (missense_variant c.2122A>T p.Ile708Phe), *astD* (missense_variant c.964A>G p.Ile322Val), *fliL* (missense_variant c.55C>A p.Leu19Met), *wzc* (missense_variant c.725T>C p.Ile242Thr), *yejO* (frameshift_variant c.345_346delGG p.Ala116fs), *atoC* (missense_variant c.1031A>G p.Asp344Gly), *arnC* (missense_variant c.563C>G p.Ala188Gly), *ypdA_2* (missense_variant c.1388A>G p.Asn463Ser), *murP* (missense_variant c.43A>G p.Thr15Ala), *eutE* (synonymous_variant c.913C>T p.Leu305Leu), *hmp* (frameshift_variant c.110dupC p.Glu38fs), *mutS* (missense_variant c.515G>C p.Arg172Pro), *metK* (missense_variant c.349G>A p.Ala117Thr), *fadH* (missense_variant c.1382T>A p.Ile461Asn), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *trpS* (synonymous_variant c.73C>T p.Leu25Leu), *glgX* (frameshift_variant c.249delG p.His84fs), *yhjB* (missense_variant c.341T>C p.Val114Ala), *yhjG* (missense_variant c.2035A>C p.Thr679Pro), *yhjV* (missense_variant c.868A>G p.Ile290Val), *yiaT* (frameshift_variant c.418dupG p.Glu140fs), *yibJ* (missense_variant c.97C>A p.Gln33Lys), *ilvY* (missense_variant c.506T>G p.Ile169Ser), *cytR* (missense_variant c.361A>C p.Thr121Pro), *murI* (missense_variant c.173T>C p.Phe58Ser), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *adiA* (frameshift_variant c.291delA p.Ala98fs), *ytfB* (missense_variant c.385C>T p.Arg129Cys), *yjgL* (missense_variant c.1781A>G p.His594Arg), *idnD* (frameshift_variant c.699delA p.Gly234fs)
6TUB123-MB3	*secM* (synonymous_variant c.531G>A p.Thr177Thr), *can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *dnaE* (missense_variant c.3233A>G p.Glu1078Gly), *ykfM* (missense_variant c.203_206delATGTinsTTGA p.HisVal68LeuGlu), *ykfM* (missense_variant c.197A>T p.Asn66Ile), *hha* (missense_variant c.143A>G p.Asp48Gly), *ybaL* (frameshift_variant c.570dupG p.Ile191fs), *fes* (missense_variant c.311T>G p.Val104Gly), *leuS* (missense_variant c.467T>G p.Val156Gly), *ybhI* (synonymous_variant c.1176A>G p.Leu392Leu), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *hcp* (synonymous_variant c.201A>G p.Pro67Pro), *matP* (missense_variant c.380A>G p.Glu127Gly), *rlmI* (missense_variant c.1018A>G p.Thr340Ala), *torA* (missense_variant c.1093A>G p.Asn365Asp), *ycdT* (frameshift_variant c.483delG p.Pro162fs), *holB* (synonymous_variant c.993T>G p.Val331Val), *dhaR* (frameshift_variant c.739delG p.Ala247fs), *yciT* (missense_variant c.262G>A p.Glu88Lys), *yciW* (frameshift_variant c.156delC p.Phe53fs), *ydbH* (missense_variant c.940C>T p.Leu314Phe), *ydbC* (frameshift_variant c.95delC p.Pro32fs), *gapA1* (missense_variant c.17T>C p.Val6Ala), *ansP* (missense_variant c.899T>C p.Val300Ala), *ynfF* (missense_variant c.2122A>T p.Ile708Phe), *astD* (missense_variant c.964A>G p.Ile322Val), *dmlR_1* (missense_variant c.395A>G p.Asp132Gly), *proQ* (frameshift_variant c.148delA p.Met50fs), *yecC* (missense_variant c.467C>A p.Ala156Glu), *fliL* (missense_variant c.55C>A p.Leu19Met), *wzc* (missense_variant c.725T>C p.Ile242Thr), *yejO* (frameshift_variant c.346delG p.Ala116fs), *atoC* (missense_variant c.1031A>G p.Asp344Gly), *arnC* (missense_variant c.563C>G p.Ala188Gly), *ypdA_2* (missense_variant c.1388A>G p.Asn463Ser), *murP* (missense_variant c.43A>G p.Thr15Ala), *eutE* (synonymous_variant c.913C>T p.Leu305Leu), *hmp* (frameshift_variant c.110dupC p.Glu38fs), *mutS* (missense_variant c.515G>C p.Arg172Pro), *metK* (missense_variant c.349G>A p.Ala117Thr), *fadH* (missense_variant c.1382T>A p.Ile461Asn), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *trpS* (synonymous_variant c.73C>T p.Leu25Leu), *glgX* (frameshift_variant c.249delG p.His84fs), *yhjB* (missense_variant c.341T>C p.Val114Ala), *yhjG* (missense_variant c.2035A>C p.Thr679Pro), *yhjV* (missense_variant c.868A>G p.Ile290Val), *yiaT* (frameshift_variant c.418dupG p.Glu140fs), *yibJ* (missense_variant c.97C>A p.Gln33Lys), *ilvY* (missense_variant c.506T>G p.Ile169Ser), *cytR* (missense_variant c.361A>C p.Thr121Pro), *murI* (missense_variant c.173T>C p.Phe58Ser), *rpoB*

	(missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *thiF* (missense_variant c.160A>G p.Thr54Ala), *ytfB* (missense_variant c.385C>T p.Arg129Cys), *yjgL* (missense_variant c.1781A>G p.His594Arg), *idnD* (frameshift_variant c.699delA p.Gly234fs)
6TUB128-MB3	*leuA* (missense_variant c.394T>C p.Tyr132His), *ddlB* (synonymous_variant c.30T>G p.Gly10Gly), *secM* (synonymous_variant c.531G>A p.Thr177Thr), *cueO* (missense_variant c.1083G>A p.Met361Ile), *can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *dnaE* (missense_variant c.3233A>G p.Glu1078Gly), *lacZ* (missense_variant c.1718A>T p.Asp573Val), *hha* (missense_variant c.143A>G p.Asp48Gly), *ybaL* (frameshift_variant c.570dupG p.Ile191fs), *lpxH* (missense_variant c.683T>C p.Val228Ala), *fepA* (missense_variant c.947A>C p.Asn316Thr), *fes* (missense_variant c.311T>G p.Val104Gly), *leuS* (missense_variant c.1265T>C p.Val422Ala), *leuS* (missense_variant c.467T>G p.Val156Gly), *ybhI* (synonymous_variant c.1176A>G p.Leu392Leu), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *mgsA* (frameshift_variant c.209dupG p.Asp71fs), *rlmI* (missense_variant c.1018A>G p.Thr340Ala), *etk* (missense_variant c.476A>G p.Glu159Gly), *insA-2* (missense_variant c.243_249delCCGCCATinsACGTCAC p.Phe81Leu), *insAB-1* (synonymous_variant c.18_21delATACinsCTAT p.8), *insAB-1* (missense_variant c.69G>T p.Arg23Ser), *holB* (synonymous_variant c.993T>G p.Val331Val), *minC* (missense_variant c.80T>C p.Val27Ala), *ydbC* (frameshift_variant c.95delC p.Pro32fs), *gapA1* (missense_variant c.17T>C p.Val6Ala), *ansP* (missense_variant c.899T>C p.Val300Ala), *ddpC* (missense_variant c.428T>C p.Val143Ala), *ydeK* (frameshift_variant c.3962delG p.Gly1321fs), *ynfF* (missense_variant c.2122A>T p.Ile708Phe), *ydiM* (frameshift_variant c.1171dupA p.Thr391fs), *astD* (missense_variant c.964A>G p.Ile322Val), *fliL* (missense_variant c.55C>A p.Leu19Met), *cbl* (frameshift_variant c.609delG p.Arg204fs), *yejO* (frameshift_variant c.346delG p.Ala116fs), *atoC* (missense_variant c.1031A>G p.Asp344Gly), *arnC* (missense_variant c.563C>G p.Ala188Gly), *yfbP* (missense_variant c.173A>G p.Glu58Gly), *ypdA_2* (missense_variant c.1388A>G p.Asn463Ser), *murP* (missense_variant c.43A>G p.Thr15Ala), *eutE* (synonymous_variant c.913C>T p.Leu305Leu), *glyA* (synonymous_variant c.411T>A p.Gly137Gly), *hmp* (frameshift_variant c.110dupC p.Glu38fs), *ffh* (synonymous_variant c.570T>C p.Asp190Asp), *mutS* (missense_variant c.515G>C p.Arg172Pro), *metK* (missense_variant c.349G>A p.Ala117Thr), *fadH* (missense_variant c.1382T>A p.Ile461Asn), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *kefG* (synonymous_variant c.306A>G p.Gly102Gly), *trpS* (missense_variant c.80A>C p.Gln27Pro), *trpS* (synonymous_variant c.73C>T p.Leu25Leu), *glgX* (frameshift_variant c.249delG p.His84fs), *pitA* (missense_variant c.869T>C p.Leu290Ser), *yhjB* (missense_variant c.341T>C p.Val114Ala), *yhjG* (missense_variant c.2035A>C p.Thr679Pro), *yhjV* (missense_variant c.868A>G p.Ile290Val), *dppA* (missense_variant c.1106A>G p.Glu369Gly), *selB* (missense_variant c.1411A>G p.Ile471Val), *ilvY* (missense_variant c.506T>G p.Ile169Ser), *cytR* (missense_variant c.361A>C p.Thr121Pro), *murI* (missense_variant c.173T>C p.Phe58Ser), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *thiH* (synonymous_variant c.891A>G p.Ser297Ser), *adiA* (frameshift_variant c.291delA p.Ala98fs), *yjgL* (missense_variant c.1781A>G p.His594Arg), *idnD* (frameshift_variant c.699delA p.Gly234fs)
6TUB130-MB3	*nhaR* (missense_variant c.653T>A p.Ile218Asn), *ribF* (missense_variant c.182A>G p.Glu61Gly), *djlA* (frameshift_variant c.19delA p.Ile7fs), *djlA* (synonymous_variant c.552T>C p.Gly184Gly), *ilvH* (missense_variant c.107C>T p.Ala36Val), *pdhR* (missense_variant c.314A>C p.His105Pro), *lpd* (missense_variant c.586G>A p.Ala196Thr), *can* (missense_variant c.545G>T p.Gly182Val), *ldcC* (missense_variant c.1403C>T p.Ala468Val), *ykfM* (missense_variant c.239T>A p.Ile80Asn), *ykfM* (synonymous_variant c.219G>T p.Thr73Thr), *frsA* (frameshift_variant c.304delG p.Ala102fs), *ykfA* (missense_variant c.341G>A p.Arg114His), *cynR* (missense_variant c.307A>G p.Ser103Gly), *tauD* (missense_variant c.511C>T p.His171Tyr), *rdgC* (missense_variant c.803A>C p.Asp268Ala), *brnQ* (missense_variant c.586A>G p.Thr196Ala), *panE* (missense_variant c.290A>G p.His97Arg), *ybaO* (missense_variant c.452T>C p.Ile151Thr), *acrB* (frameshift_variant c.72dupG p.Leu25fs), *ybaL* (synonymous_variant c.267T>C p.Ile89Ile), *ybcF* (missense_variant c.461T>C p.Val154Ala), *sfmH* (frameshift_variant c.332dupC p.Gln112fs), *entF* (synonymous_variant c.3153T>C p.Arg1051Arg), *entC* (missense_variant c.677G>A p.Arg226His), *leuS* (missense_variant c.283A>G p.Thr95Ala), *chiP* (missense_variant c.587A>G p.Asn196Ser), *ybfF* (missense_variant c.578C>T p.Ala193Val), *potE* (frameshift_variant c.1146delC p.Ser383fs), *ybfC* (frameshift_variant c.121delT p.Ser41fs), *gltA* (missense_variant c.815C>T p.Ala272Val), *modF* (synonymous_variant c.747A>C p.Ser249Ser), *uvrB* (synonymous_variant c.816A>G p.Lys272Lys), *ybhF* (synonymous_variant c.112T>C p.Leu38Leu), *dinG* (missense_variant c.1136A>G p.His379Arg), *artQ* (frameshift_variant c.117delA p.Lys39fs), *poxB* (frameshift_variant c.643delG p.Ala215fs), *ycaD* (synonymous_variant c.999A>G p.Gly333Gly), *ycbL* (missense_variant c.113T>C p.Ile38Thr), *elfC* (missense_variant c.2180A>G p.Asp727Gly), *uup* (frameshift_variant c.1630dupA p.Thr544fs), *rlmI* (frameshift_variant c.223delT p.Ser75fs), *gfcD* (missense_variant c.772A>G p.Asn258Asp), *rutF* (missense_variant c.100T>C p.Phe34Leu), *rutD* (missense_variant c.664G>A p.Ala222Thr), *putA* (missense_variant c.661T>C p.Ser221Pro), *csgA* (missense_variant c.412A>G p.Thr138Ala), *ymdC* (missense_variant c.740A>T p.Glu247Val), *thiK* (missense_variant c.509T>C p.Val170Ala), *mcrA* (synonymous_variant c.663A>G p.Gly221Gly), *cvrA* (missense_variant c.1211T>C p.Val404Ala), *adhE* (missense_variant c.394A>C p.Asn132His), *oppD* (synonymous_variant c.975T>C p.Arg325Arg), *trpA* (missense_variant c.703T>C p.Ser235Pro), *yciK* (missense_variant c.751T>C p.Ser251Pro), *puuP* (frameshift_variant c.566delG p.Gly189fs), *pspF* (synonymous_variant c.816T>C p.Leu272Leu), *pspC* (synonymous_variant c.12T>C p.Ile4Ile), *abgA* (synonymous_variant c.195A>G p.Leu65Leu), *ldhA* (missense_variant c.613T>A p.Cys205Ser), *paaH* (synonymous_variant c.117A>C p.Ser39Ser), *narZ* (missense_variant c.1069A>G p.Ile357Val), *maeA* (synonymous_variant c.1632G>A p.Leu544Leu), *ydeK* (frameshift_variant c.3838dupG p.Glu1280fs), *lsrB* (missense_variant c.770G>T p.Arg257Leu), *ydgI* (frameshift_variant c.842delG p.Gly281fs), *fumC* (frameshift_variant c.1052dupG p.Ala352fs), *uidA* (missense_variant c.1139T>C

	p.Val380Ala), *add* (stop_gained c.624G>A p.Trp208*), *rsxA* (missense_variant c.342T>G p.Cys114Trp), *ydhC* (synonymous_variant c.309A>C p.Val103Val), *ydhV* (missense_variant c.1958A>G p.Asp653Gly), *ydiI* (missense_variant c.169G>A p.Ala57Thr), *astD* (synonymous_variant c.21T>C p.Gly7Gly), *astA* (missense_variant c.611C>T p.Ala204Val), *yeaE* (missense_variant c.151A>C p.Thr51Pro), *proQ* (frameshift_variant c.67dupC p.His23fs), *otsB* (synonymous_variant c.279A>G p.Thr93Thr), *yedK* (synonymous_variant c.360T>C p.Arg120Arg), *fliJ* (frameshift_variant c.315delA p.Lys105fs), *fliK* (frameshift_variant c.145delC p.Gln49fs), *cobT* (missense_variant c.199T>C p.Cys67Arg), *plaP* (missense_variant c.1310T>C p.Val437Ala), *wbbH* (missense_variant c.632T>C p.Val211Ala), *yegE* (frameshift_variant c.419dupG p.Val141fs), *yegV* (missense_variant c.445T>C p.Ser149Pro), *yehF* (missense_variant c.668A>T p.Asp223Val), *dld* (missense_variant c.364A>G p.Thr122Ala), *yohK* (synonymous_variant c.516T>G p.Arg172Arg), *folE* (synonymous_variant c.57A>G p.Gly19Gly), *yeiG* (missense_variant c.596A>G p.Asp199Gly), *fruA* (missense_variant c.101T>C p.Ile34Thr), *alkB* (frameshift_variant c.270delC p.Ala91fs), *yfaL* (frameshift_variant c.1363delG p.Ala455fs), *glpT_1* (missense_variant c.1178A>G p.Tyr393Cys), *yfdF* (missense_variant c.232G>T p.Asp78Tyr), *xapB* (missense_variant c.601G>A p.Ala201Thr), *dapA* (missense_variant c.827A>G p.Asp276Gly), *yfgD* (missense_variant c.53A>G p.Glu18Gly), *ppx* (synonymous_variant c.441A>T p.Gly147Gly), *yfhM* (missense_variant c.2739T>A p.Asp913Glu), *hscB* (synonymous_variant c.291G>A p.Glu97Glu), *tadA* (missense_variant c.490T>C p.Ser164Pro), *smpB* (missense_variant c.355A>G p.Lys119Glu), *yfjV* (missense_variant c.1077T>A p.Asp359Glu), *yfjW* (missense_variant c.1099T>A p.Tyr367Asn), *mutS* (missense_variant c.128A>G p.Asp43Gly), *ygbJ* (missense_variant c.46T>C p.Ser16Pro), *casA* (missense_variant c.1321T>C p.Ser441Pro), *cysI* (missense_variant c.1165G>A p.Asp389Asn), *araE* (synonymous_variant c.300A>G p.Val100Val), *ygfM* (synonymous_variant c.435C>T p.Cys145Cys), *prfB_2* (missense_variant c.494C>T p.Ala165Val), *cmtA* (missense_variant c.1277A>G p.Glu426Gly), *speA* (missense_variant c.866A>G p.Glu289Gly), *metK* (missense_variant c.677A>G p.Asn226Ser), *glcA* (synonymous_variant c.855T>C p.Gly285Gly), *ysgA_1* (frameshift_variant c.340delC p.Arg114fs), *yqiG* (missense_variant c.355A>G p.Thr119Ala), *rpoD* (missense_variant c.1340C>T p.Ala447Val), *fadH* (missense_variant c.272T>C p.Val91Ala), *sstT* (frameshift_variant c.1041dupG p.Ser348fs), *tdcG* (synonymous_variant c.1158A>G p.Gly386Gly), *lptC* (missense_variant c.448A>C p.Thr150Pro), *gltF* (missense_variant c.740T>C p.Val247Ala), *yhdX* (missense_variant c.728T>C p.Val243Ala), *trkA* (missense_variant c.1151T>C p.Val384Ala), *frlR* (missense_variant c.83T>C p.Val28Ala), *php* (missense_variant c.457C>T p.Arg153Cys), *pck* (synonymous_variant c.687T>C p.Ala229Ala), *malP* (missense_variant c.1487T>C p.Phe496Ser), *livJ* (synonymous_variant c.510A>G p.Lys170Lys), *yhjH* (frameshift_variant c.326delC p.Pro109fs), *bcsA* (missense_variant c.836T>C p.Val279Ala), *bcsF* (missense_variant c.167A>G p.Glu56Gly), *yhjX* (missense_variant c.385T>C p.Phe129Leu), *glyS* (missense_variant c.485A>G p.His162Arg), *yiaJ* (frameshift_variant c.27dupA p.Glu10fs), *yibJ* (missense_variant c.97C>A p.Gln33Lys), *yibH* (stop_gained c.367C>T p.Gln123*), *spoT* (missense_variant c.770T>C p.Ile257Thr), *fimZ_7* (frameshift_variant c.283delG p.Ala95fs), *dgoR* (missense_variant c.556G>T p.Ala186Ser), *gpp* (missense_variant c.1063C>T p.Leu355Phe), *rho* (missense_variant c.191T>C p.Phe64Ser), *recQ* (missense_variant c.265G>A p.Ala89Thr), *pepQ* (missense_variant c.985A>G p.Thr329Ala), *yihV* (missense_variant c.758A>G p.Glu253Gly), *frvX* (frameshift_variant c.634delG p.Ala212fs), *rhaD* (missense_variant c.770T>C p.Ile257Thr), *rhaS* (synonymous_variant c.538C>T p.Leu180Leu), *rplL* (missense_variant c.288A>T p.Lys96Asn), *rpoC* (missense_variant c.524A>G p.Glu175Gly), *rpoC* (missense_variant c.548A>G p.Glu183Gly), *rpoC* (synonymous_variant c.3648G>A p.Ala1216Ala), *amiB* (missense_variant c.682A>G p.Thr228Ala), *ytfE* (missense_variant c.14A>G p.Asp5Gly), *ytfL* (missense_variant c.1205T>C p.Phe402Ser), *ytfR* (synonymous_variant c.888A>T p.Gly296Gly), *nanM* (frameshift_variant c.551dupT p.Phe185fs), *deoA* (missense_variant c.33A>C p.Lys11Asn), *deoB* (synonymous_variant c.216T>A p.Ile72Ile), *yjjK* (synonymous_variant c.601C>T p.Leu201Leu), *trpR* (missense_variant c.173A>C p.Gln58Pro), *yaaJ* (missense_variant c.1301T>C p.Val434Ala)
6TUB137-MB3	*leuC* (synonymous_variant c.1053C>T p.Arg351Arg), *secM* (synonymous_variant c.531G>A p.Thr177Thr), *can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *mtn* (missense_variant c.599A>G p.Asp200Gly), *dnaE* (missense_variant c.3233A>G p.Glu1078Gly), *yaiW* (missense_variant c.499T>G p.Ser167Ala), *ppiD* (missense_variant c.1207G>A p.Ala403Thr), *hha* (missense_variant c.143A>G p.Asp48Gly), *ybaL* (frameshift_variant c.570dupG p.Ile191fs), *fes* (missense_variant c.311T>G p.Val104Gly), *entC* (synonymous_variant c.243T>C p.Asp81Asp), *ybeM_1* (synonymous_variant c.171A>G p.Ala57Ala), *leuS* (missense_variant c.467T>G p.Val156Gly), *leuS* (missense_variant c.457A>C p.Thr153Pro), *potE* (frameshift_variant c.469dupG p.Val157fs), *kdpD* (frameshift_variant c.894delA p.Lys298fs), *aroG* (missense_variant c.490G>A p.Ala164Thr), *ybhI* (synonymous_variant c.1176A>G p.Leu392Leu), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *ybiY* (frameshift_variant c.42delC p.Gly15fs), *deoR* (missense_variant c.177A>C p.Glu59Asp), *hcp* (synonymous_variant c.201A>G p.Pro67Pro), *clpA* (missense_variant c.1070T>C p.Ile357Thr), *focA* (missense_variant c.524T>C p.Val175Ala), *ssuC* (missense_variant c.724G>A p.Ala242Thr), *elfG* (synonymous_variant c.942A>G p.Leu314Leu), *matP* (missense_variant c.380A>G p.Glu127Gly), *rlmI* (missense_variant c.1018A>G p.Thr340Ala), *torA* (missense_variant c.1093A>G p.Asn365Asp), *ycdT* (frameshift_variant c.483delG p.Pro162fs), *flgK* (missense_variant c.835T>C p.Ser279Pro), *holB* (synonymous_variant c.993T>G p.Val331Val), *dhaR* (frameshift_variant c.739delG p.Ala247fs), *ychF* (missense_variant c.278A>C p.Asn93Thr), *prmC* (synonymous_variant c.198A>G p.Glu66Glu), *yciT* (missense_variant c.262G>A p.Glu88Lys), *yciW* (frameshift_variant c.156delC p.Phe53fs), *insH-5* (synonymous_variant c.456T>G p.Ser152Ser), *insH-5* (synonymous_variant c.426G>C p.Val142Val), *insH-5* (synonymous_variant c.420T>C p.Thr140Thr), *ydbH* (missense_variant c.940C>T p.Leu314Phe), *ydbC* (frameshift_variant c.95delC p.Pro32fs), *gapA1* (missense_variant c.17T>C p.Val6Ala), *ydcT* (missense_variant c.68T>C p.Val23Ala), *ansP* (missense_variant c.899T>C p.Val300Ala), *narU* (synonymous_variant c.234A>G p.Leu78Leu), *ddpC* (frameshift_variant c.531dupG p.Gln178fs), *ddpC* (missense_variant c.428T>C p.Val143Ala), *gadC* (frameshift_variant c.552delC p.Val185fs), *ynfF* (missense_variant c.2122A>T p.Ile708Phe), *uidR* (missense_variant c.127A>G p.Thr43Ala), *astD* (missense_variant c.964A>G p.Ile322Val), *dmlR_1*

	(missense_variant c.395A>G p.Asp132Gly), *yebQ* (missense_variant c.1048T>C p.Ser350Pro), *yebQ* (missense_variant c.1103T>C p.Ile368Thr), *prc* (missense_variant c.374A>G p.Glu125Gly), *proQ* (frameshift_variant c.148delA p.Met50fs), *yecC* (missense_variant c.467C>A p.Ala156Glu), *fliL* (missense_variant c.55C>A p.Leu19Met), *hchA* (synonymous_variant c.453T>C p.Val151Val), *erfK* (synonymous_variant c.562T>C p.Leu188Leu), *fcl* (missense_variant c.176T>C p.Val59Ala), *wzc* (missense_variant c.725T>C p.Ile242Thr), *yegH* (missense_variant c.448A>G p.Met150Val), *gatD* (missense_variant c.479A>G p.Glu160Gly), *ypdB_1* (synonymous_variant c.255A>G p.Glu85Glu), *yejO* (frameshift_variant c.346delG p.Ala116fs), *atoC* (missense_variant c.1031A>G p.Asp344Gly), *arnC* (missense_variant c.563C>G p.Ala188Gly), *menD* (frameshift_variant c.488dupG p.Val164fs), *hisP* (missense_variant c.130A>G p.Lys44Glu), *ypdA_2* (missense_variant c.1388A>G p.Asn463Ser), *murP* (missense_variant c.43A>G p.Thr15Ala), *eutE* (synonymous_variant c.913C>T p.Leu305Leu), *yfhM* (synonymous_variant c.996T>C p.Thr332Thr), *hmp* (frameshift_variant c.110dupC p.Glu38fs), *recN* (missense_variant c.1616T>C p.Val539Ala), *mutS* (missense_variant c.515G>C p.Arg172Pro), *ptrA* (missense_variant c.2741T>C p.Val914Ala), *ygeK* (frameshift_variant c.7delG p.Ala3fs), *xdhA* (synonymous_variant c.1323C>T p.Gly441Gly), *ygfB* (missense_variant c.455A>G p.Glu152Gly), *speA* (missense_variant c.409T>C p.Ser137Pro), *metK* (missense_variant c.349G>A p.Ala117Thr), *mutY* (missense_variant c.469A>G p.Lys157Glu), *hybD* (synonymous_variant c.216A>G p.Gly72Gly), *fadH* (missense_variant c.1382T>A p.Ile461Asn), *gltD* (missense_variant c.310T>C p.Cys104Arg), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *trpS* (synonymous_variant c.73C>T p.Leu25Leu), *envZ* (frameshift_variant c.721dupG p.Val241fs), *glgX* (frameshift_variant c.249delG p.His84fs), *livH* (missense_variant c.499G>A p.Ala167Thr), *zntA* (synonymous_variant c.1332T>C p.Gly444Gly), *yhiI* (missense_variant c.266A>G p.Glu89Gly), *yhjB* (missense_variant c.341T>C p.Val114Ala), *yhjG* (missense_variant c.2035A>C p.Thr679Pro), *bcsC* (frameshift_variant c.2715delG p.Pro906fs), *yhjV* (missense_variant c.868A>G p.Ile290Val), *yiaJ* (synonymous_variant c.21A>G p.Gly7Gly), *yiaT* (frameshift_variant c.418dupG p.Glu140fs), *gpsA* (synonymous_variant c.987T>G p.Gly329Gly), *ilvY* (missense_variant c.506T>G p.Ile169Ser), *yihL* (frameshift_variant c.397dupG p.Ala133fs), *yihT* (synonymous_variant c.240T>C p.Ile80Ile), *cytR* (missense_variant c.361A>C p.Thr121Pro), *metF* (missense_variant c.118A>C p.Thr40Pro), *fsaB* (missense_variant c.265A>G p.Thr89Ala), *murI* (missense_variant c.173T>C p.Phe58Ser), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *thiF* (missense_variant c.160A>G p.Thr54Ala), *alsA* (missense_variant c.1396T>C p.Ser466Pro), *adiA* (frameshift_variant&synonymous_variant c.287_288delAAinsG p.Glu96fs), *yeiP_2* (missense_variant c.532A>C p.Thr178Pro), *blc* (missense_variant c.269A>G p.Glu90Gly), *hflK* (missense_variant c.995A>G p.Glu332Gly), *ytfB* (missense_variant c.385C>T p.Arg129Cys), *yjgL* (missense_variant c.1781A>G p.His594Arg), *idnD* (frameshift_variant c.699delA p.Gly234fs), *idnD* (frameshift_variant c.580delG p.Ala194fs), *idnK* (frameshift_variant c.261delA p.Lys87fs), *mrr* (synonymous_variant c.9T>G p.Val3Val), *yjjV* (synonymous_variant c.240G>A p.Gln80Gln), *ytjB* (synonymous_variant c.402T>C p.Gly134Gly)
6TUB144-MB3	*leuC* (synonymous_variant c.1053C>T p.Arg351Arg), *secM* (synonymous_variant c.531G>A p.Thr177Thr), *can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *mtn* (missense_variant c.599A>G p.Asp200Gly), *dnaE* (missense_variant c.3233A>G p.Glu1078Gly), *yaiW* (missense_variant c.499T>G p.Ser167Ala), *ppiD* (missense_variant c.1207G>A p.Ala403Thr), *hha* (missense_variant c.143A>G p.Asp48Gly), *ybaL* (frameshift_variant c.570dupG p.Ile191fs), *fes* (missense_variant c.311T>G p.Val104Gly), *entC* (synonymous_variant c.243T>C p.Asp81Asp), *ybeM_1* (synonymous_variant c.171A>G p.Ala57Ala), *leuS* (missense_variant c.467T>G p.Val156Gly), *leuS* (missense_variant c.457A>C p.Thr153Pro), *potE* (frameshift_variant c.469dupG p.Val157fs), *kdpD* (frameshift_variant c.894delA p.Lys298fs), *aroG* (missense_variant c.490G>A p.Ala164Thr), *ybhI* (synonymous_variant c.1176A>G p.Leu392Leu), *ybiV* (missense_variant c.829A>T p.Ile277Phe), *ybiY* (frameshift_variant c.42delC p.Gly15fs), *deoR* (missense_variant c.177A>C p.Glu59Asp), *hcp* (synonymous_variant c.201A>G p.Pro67Pro), *clpA* (missense_variant c.1070T>C p.Ile357Thr), *focA* (missense_variant c.524T>C p.Val175Ala), *ssuC* (missense_variant c.724G>A p.Ala242Thr), *elfG* (synonymous_variant c.942A>G p.Leu314Leu), *matP* (missense_variant c.380A>G p.Glu127Gly), *rlmI* (missense_variant c.1018A>G p.Thr340Ala), *insA-2* (missense_variant c.243_249delCCGCCATinsACGTCAC p.Phe81Leu), *insAB-1* (synonymous_variant c.18_21delATACinsCTAT p.8), *insAB-1* (missense_variant c.69G>T p.Arg23Ser), *torA* (missense_variant c.1093A>G p.Asn365Asp), *ycdT* (frameshift_variant c.483delG p.Pro162fs), *flgK* (missense_variant c.835T>C p.Ser279Pro), *holB* (synonymous_variant c.993T>G p.Val331Val), *dhaR* (frameshift_variant c.739delG p.Ala247fs), *ychF* (missense_variant c.278A>C p.Asn93Thr), *prmC* (synonymous_variant c.198A>G p.Glu66Glu), *yciT* (missense_variant c.262G>A p.Glu88Lys), *yciW* (frameshift_variant c.156delC p.Phe53fs), *ydbH* (missense_variant c.940C>T p.Leu314Phe), *ydbC* (frameshift_variant c.95delC p.Pro32fs), *gapA1* (missense_variant c.17T>C p.Val6Ala), *ydcT* (missense_variant c.68T>C p.Val23Ala), *ansP* (missense_variant c.899T>C p.Val300Ala), *narU* (synonymous_variant c.234A>G p.Leu78Leu), *ddpC* (frameshift_variant c.531dupG p.Gln178fs), *ddpC* (missense_variant c.428T>C p.Val143Ala), *gadC* (frameshift_variant c.552delC p.Val185fs), *ynfF* (missense_variant c.2122A>T p.Ile708Phe), *uidR* (missense_variant c.127A>G p.Thr43Ala), *astD* (missense_variant c.964A>G p.Ile322Val), *dmlR_1* (missense_variant c.395A>G p.Asp132Gly), *yebQ* (missense_variant c.1048T>C p.Ser350Pro), *yebQ* (missense_variant c.1103T>C p.Ile368Thr), *prc* (missense_variant c.374A>G p.Glu125Gly), *proQ* (frameshift_variant c.148delA p.Met50fs), *yecC* (missense_variant c.467C>A p.Ala156Glu), *fliL* (missense_variant c.55C>A p.Leu19Met), *hchA* (synonymous_variant c.453T>C p.Val151Val), *erfK* (synonymous_variant c.562T>C p.Leu188Leu), *fcl* (missense_variant c.176T>C p.Val59Ala), *wzc* (missense_variant c.725T>C p.Ile242Thr), *yegH* (missense_variant c.448A>G p.Met150Val), *gatD* (missense_variant c.479A>G p.Glu160Gly), *ypdB_1* (synonymous_variant c.255A>G p.Glu85Glu), *yejO* (frameshift_variant c.346delG p.Ala116fs), *atoC* (missense_variant c.1031A>G p.Asp344Gly), *arnC* (missense_variant c.563C>G p.Ala188Gly), *menD* (frameshift_variant c.488dupG p.Val164fs), *hisP* (missense_variant c.130A>G p.Lys44Glu), *ypdA_2* (missense_variant c.1388A>G p.Asn463Ser), *murP* (missense_variant c.43A>G p.Thr15Ala), *eutE* (synonymous_variant c.913C>T p.Leu305Leu), *yfhM* (synonymous_variant c.996T>C p.Thr332Thr), *hmp* (frameshift_variant c.110dupC p.Glu38fs), *recN*

	(missense_variant c.1616T>C p.Val539Ala), *mutS* (missense_variant c.515G>C p.Arg172Pro), *ptrA* (missense_variant c.2741T>C p.Val914Ala), *ygeK* (frameshift_variant c.7delG p.Ala3fs), *xdhA* (synonymous_variant c.1323C>T p.Gly441Gly), *ygfB* (missense_variant c.455A>G p.Glu152Gly), *speA* (missense_variant c.409T>C p.Ser137Pro), *metK* (missense_variant c.349G>A p.Ala117Thr), *mutY* (missense_variant c.469A>G p.Lys157Glu), *hybD* (synonymous_variant c.216A>G p.Gly72Gly), *fadH* (missense_variant c.1382T>A p.Ile461Asn), *gltD* (missense_variant c.310T>C p.Cys104Arg), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *trpS* (synonymous_variant c.73C>T p.Leu25Leu), *envZ* (frameshift_variant c.721dupG p.Val241fs), *glgX* (frameshift_variant c.249delG p.His84fs), *livH* (missense_variant c.499G>A p.Ala167Thr), *zntA* (synonymous_variant c.1332T>C p.Gly444Gly), *yhiI* (missense_variant c.266A>G p.Glu89Gly), *yhjB* (missense_variant c.341T>C p.Val114Ala), *yhjG* (missense_variant c.2035A>C p.Thr679Pro), *bcsC* (frameshift_variant c.2715delG p.Pro906fs), *yhjV* (missense_variant c.868A>G p.Ile290Val), *yiaJ* (synonymous_variant c.21A>G p.Gly7Gly), *yiaT* (frameshift_variant c.418dupG p.Glu140fs), *gpsA* (synonymous_variant c.987T>G p.Gly329Gly), *ilvY* (missense_variant c.506T>G p.Ile169Ser), *yihL* (frameshift_variant c.397dupG p.Ala133fs), *yihT* (synonymous_variant c.240T>C p.Ile80Ile), *cytR* (missense_variant c.361A>C p.Thr121Pro), *metF* (missense_variant c.118A>C p.Thr40Pro), *fsaB* (missense_variant c.265A>G p.Thr89Ala), *murI* (missense_variant c.173T>C p.Phe58Ser), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *thiF* (missense_variant c.160A>G p.Thr54Ala), *alsA* (missense_variant c.1396T>C p.Ser466Pro), *adiA* (frameshift_variant&synonymous_variant c.287_288delAAinsG p.Glu96fs), *yeiP_2* (missense_variant c.532A>C p.Thr178Pro), *blc* (missense_variant c.269A>G p.Glu90Gly), *hflK* (missense_variant c.995A>G p.Glu332Gly), *ytfB* (missense_variant c.385C>T p.Arg129Cys), *yjgL* (missense_variant c.1781A>G p.His594Arg), *idnD* (frameshift_variant c.699delA p.Gly234fs), *idnD* (frameshift_variant c.580delG p.Ala194fs), *idnK* (frameshift_variant c.261delA p.Lys87fs), *mrr* (synonymous_variant c.9T>G p.Val3Val), *yjjV* (synonymous_variant c.240G>A p.Gln80Gln), *ytjB* (synonymous_variant c.402T>C p.Gly134Gly)
6TUB151-MB3	*leuC* (synonymous_variant c.1053C>T p.Arg351Arg), *secM* (synonymous_variant c.531G>A p.Thr177Thr), *can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *mtn* (missense_variant c.599A>G p.Asp200Gly), *dnaE* (missense_variant c.3233A>G p.Glu1078Gly), *paoC* (frameshift_variant c.849dupC p.Ala284fs), *yaiW* (missense_variant c.499T>G p.Ser167Ala), *ppiD* (missense_variant c.1207G>A p.Ala403Thr), *hha* (missense_variant c.143A>G p.Asp48Gly), *ybaL* (frameshift_variant c.570dupG p.Ile191fs), *ybcK* (missense_variant c.1055A>T p.Asn352Ile), *fes* (missense_variant c.311T>G p.Val104Gly), *entC* (synonymous_variant c.243T>C p.Asp81Asp), *ybeM_1* (synonymous_variant c.171A>G p.Ala57Ala), *leuS* (missense_variant c.467T>G p.Val156Gly), *leuS* (missense_variant c.457A>C p.Thr153Pro), *potE* (frameshift_variant c.469dupG p.Val157fs), *kdpD* (frameshift_variant c.894delA p.Lys298fs), *aroG* (missense_variant c.490G>A p.Ala164Thr), *ybhI* (synonymous_variant c.1176A>G p.Leu392Leu), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *ybiY* (frameshift_variant c.42delC p.Gly15fs), *deoR* (missense_variant c.177A>C p.Glu59Asp), *hcp* (synonymous_variant c.201A>G p.Pro67Pro), *clpA* (missense_variant c.1070T>C p.Ile357Thr), *focA* (missense_variant c.524T>C p.Val175Ala), *ssuC* (missense_variant c.724G>A p.Ala242Thr), *elfG* (synonymous_variant c.942A>G p.Leu314Leu), *matP* (missense_variant c.380A>G p.Glu127Gly), *rlmI* (missense_variant c.1018A>G p.Thr340Ala), *insAB-1* (missense_variant c.69G>T p.Arg23Ser), *torA* (missense_variant c.1093A>G p.Asn365Asp), *ycdT* (frameshift_variant c.483delG p.Pro162fs), *flgK* (missense_variant c.835T>C p.Ser279Pro), *holB* (synonymous_variant c.993T>G p.Val331Val), *dhaR* (frameshift_variant c.739delG p.Ala247fs), *ychF* (missense_variant c.278A>C p.Asn93Thr), *prmC* (synonymous_variant c.198A>G p.Glu66Glu), *yciT* (missense_variant c.262G>A p.Glu88Lys), *yciW* (frameshift_variant c.156delC p.Phe53fs), *ydbH* (missense_variant c.940C>T p.Leu314Phe), *ydbC* (frameshift_variant c.95delC p.Pro32fs), *gapA1* (missense_variant c.17T>C p.Val6Ala), *ydcT* (missense_variant c.68T>C p.Val23Ala), *ansP* (missense_variant c.899T>C p.Val300Ala), *narU* (synonymous_variant c.234A>G p.Leu78Leu), *ddpC* (frameshift_variant c.531dupG p.Gln178fs), *ddpC* (missense_variant c.428T>C p.Val143Ala), *gadC* (frameshift_variant c.552delC p.Val185fs), *ynfF* (missense_variant c.2122A>T p.Ile708Phe), *uidR* (missense_variant c.127A>G p.Thr43Ala), *astD* (missense_variant c.964A>G p.Ile322Val), *dmlR_1* (missense_variant c.395A>G p.Asp132Gly), *yebQ* (missense_variant c.1048T>C p.Ser350Pro), *yebQ* (missense_variant c.1103T>C p.Ile368Thr), *prc* (missense_variant c.374A>G p.Glu125Gly), *proQ* (frameshift_variant c.148delA p.Met50fs), *yecC* (missense_variant c.467C>A p.Ala156Glu), *fliL* (missense_variant c.55C>A p.Leu19Met), *hchA* (synonymous_variant c.453T>C p.Val151Val), *erfK* (synonymous_variant c.562T>C p.Leu188Leu), *fcl* (missense_variant c.176T>C p.Val59Ala), *wzc* (missense_variant c.725T>C p.Ile242Thr), *yegH* (missense_variant c.448A>G p.Met150Val), *gatD* (missense_variant c.479A>G p.Glu160Gly), *ypdB_1* (synonymous_variant c.255A>G p.Glu85Glu), *yejO* (frameshift_variant c.346delG p.Ala116fs), *atoC* (missense_variant c.1031A>G p.Asp344Gly), *arnC* (missense_variant c.563C>G p.Ala188Gly), *menD* (frameshift_variant c.488dupG p.Val164fs), *hisP* (missense_variant c.130A>G p.Lys44Glu), *ypdA_2* (missense_variant c.1388A>G p.Asn463Ser), *murP* (missense_variant c.43A>G p.Thr15Ala), *eutE* (synonymous_variant c.913C>T p.Leu305Leu), *yfhM* (synonymous_variant c.996T>C p.Thr332Thr), *hmp* (frameshift_variant c.110dupC p.Glu38fs), *recN* (missense_variant c.1616T>C p.Val539Ala), *mutS* (missense_variant c.515G>C p.Arg172Pro), *ptrA* (missense_variant c.2741T>C p.Val914Ala), *ygeK* (frameshift_variant c.7delG p.Ala3fs), *xdhA* (synonymous_variant c.1323C>T p.Gly441Gly), *ygfB* (missense_variant c.455A>G p.Glu152Gly), *speA* (missense_variant c.409T>C p.Ser137Pro), *metK* (missense_variant c.349G>A p.Ala117Thr), *mutY* (missense_variant c.469A>G p.Lys157Glu), *hybD* (synonymous_variant c.216A>G p.Gly72Gly), *fadH* (missense_variant c.1382T>A p.Ile461Asn), *gltD* (missense_variant c.310T>C p.Cys104Arg), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *trpS* (synonymous_variant c.73C>T p.Leu25Leu), *envZ* (frameshift_variant c.721dupG p.Val241fs), *glgX* (frameshift_variant c.249delG p.His84fs), *gntU* (missense_variant c.164A>G p.Glu55Gly), *livH* (missense_variant c.499G>A p.Ala167Thr), *zntA* (synonymous_variant c.1332T>C p.Gly444Gly), *yhiI* (missense_variant c.266A>G p.Glu89Gly), *yhjB* (missense_variant c.341T>C p.Val114Ala), *yhjG* (missense_variant c.2035A>C

	p.Thr679Pro), *bcsC* (frameshift_variant c.2715delG p.Pro906fs), *yhjV* (missense_variant c.868A>G p.Ile290Val), *yiaJ* (synonymous_variant c.21A>G p.Gly7Gly), *yiaT* (frameshift_variant c.418dupG p.Glu140fs), *yibJ* (missense_variant c.97C>A p.Gln33Lys), *gpsA* (synonymous_variant c.987T>G p.Gly329Gly), *ilvY* (missense_variant c.506T>G p.Ile169Ser), *yihL* (frameshift_variant c.397dupG p.Ala133fs), *yihT* (synonymous_variant c.240T>C p.Ile80Ile), *cytR* (missense_variant c.361A>C p.Thr121Pro), *metF* (missense_variant c.118A>C p.Thr40Pro), *fsaB* (missense_variant c.265A>G p.Thr89Ala), *murI* (missense_variant c.173T>C p.Phe58Ser), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *thiF* (missense_variant c.160A>G p.Thr54Ala), *alsA* (missense_variant c.1396T>C p.Ser466Pro), *adiA* (frameshift_variant&synonymous_variant c.287_288delAAinsG p.Glu96fs), *yeiP_2* (missense_variant c.532A>C p.Thr178Pro), *blc* (missense_variant c.269A>G p.Glu90Gly), *hflK* (missense_variant c.995A>G p.Glu332Gly), *ytfB* (missense_variant c.385C>T p.Arg129Cys), *yjgL* (missense_variant c.1781A>G p.His594Arg), *idnD* (frameshift_variant c.699delA p.Gly234fs), *idnD* (frameshift_variant c.580delG p.Ala194fs), *idnK* (frameshift_variant c.261delA p.Lys87fs), *mrr* (synonymous_variant c.9T>G p.Val3Val), *yjjV* (synonymous_variant c.240G>A p.Gln80Gln), *ytjB* (synonymous_variant c.402T>C p.Gly134Gly)
6TUB158-MB3	*leuC* (synonymous_variant c.1053C>T p.Arg351Arg), *murF* (missense_variant c.992A>G p.Asp331Gly), *secM* (synonymous_variant c.531G>A p.Thr177Thr), *can* (missense_variant c.167A>G p.Glu56Gly), *pcnB* (frameshift_variant c.959dupA p.Lys321fs), *dnaE* (missense_variant c.3233A>G p.Glu1078Gly), *prpD* (missense_variant c.974A>G p.His325Arg), *yaiW* (missense_variant c.499T>G p.Ser167Ala), *ppiD* (missense_variant c.1207G>A p.Ala403Thr), *hha* (missense_variant c.143A>G p.Asp48Gly), *ybaL* (frameshift_variant c.570dupG p.Ile191fs), *ybbN* (missense_variant c.380A>G p.Glu127Gly), *fes* (missense_variant c.311T>G p.Val104Gly), *ybeM_1* (synonymous_variant c.171A>G p.Ala57Ala), *leuS* (missense_variant c.467T>G p.Val156Gly), *kdpD* (frameshift_variant c.894delA p.Lys298fs), *ybgO* (synonymous_variant c.627T>C p.Gly209Gly), *aroG* (missense_variant c.490G>A p.Ala164Thr), *gpmA* (missense_variant c.484A>G p.Ile162Val), *ybhI* (synonymous_variant c.1176A>G p.Leu392Leu), *ybiW* (missense_variant c.829A>T p.Ile277Phe), *ybiY* (frameshift_variant c.42delC p.Gly15fs), *potG* (frameshift_variant c.955delG p.Glu319fs), *clpA* (missense_variant c.1070T>C p.Ile357Thr), *focA* (missense_variant c.524T>C p.Val175Ala), *ycbB* (frameshift_variant c.804delG p.Pro269fs), *ssuC* (missense_variant c.724G>A p.Ala242Thr), *ssuC* (missense_variant c.338T>C p.Ile113Thr), *matP* (missense_variant c.380A>G p.Glu127Gly), *rlmI* (missense_variant c.1018A>G p.Thr340Ala), *insA-2* (missense_variant c.243_249delCCGCCATinsACGTCAC p.Phe81Leu), *insAB-1* (synonymous_variant c.18_21delATACinsCTAT p.8), *insAB-1* (missense_variant c.69G>T p.Arg23Ser), *ycdT* (frameshift_variant c.483delG p.Pro162fs), *yceF* (frameshift_variant c.608delA p.Asn203fs), *holB* (synonymous_variant c.993T>G p.Val331Val), *dhaR* (frameshift_variant c.739delG p.Ala247fs), *ychF* (missense_variant c.278A>C p.Asn93Thr), *prmC* (synonymous_variant c.198A>G p.Glu66Glu), *yciT* (missense_variant c.262G>A p.Glu88Lys), *yciW* (frameshift_variant c.156delC p.Phe53fs), *insH-5* (synonymous_variant c.630T>G p.Leu210Leu), *ydbH* (missense_variant c.940C>T p.Leu314Phe), *ydbC* (frameshift_variant c.95delC p.Pro32fs), *gapA1* (missense_variant c.17T>C p.Val6Ala), *ydcT* (missense_variant c.68T>C p.Val23Ala), *ydcW* (synonymous_variant c.639T>C p.Gly213Gly), *ansP* (missense_variant c.899T>C p.Val300Ala), *narU* (synonymous_variant c.234A>G p.Leu78Leu), *ddpC* (frameshift_variant c.531dupG p.Gln178fs), *ddpC* (missense_variant c.428T>C p.Val143Ala), *gadC* (frameshift_variant c.552delC p.Val185fs), *ydfI* (missense_variant c.425A>G p.His142Arg), *ynfF* (missense_variant c.2122A>T p.Ile708Phe), *astD* (missense_variant c.964A>G p.Ile322Val), *gapA_2* (missense_variant c.69A>C p.Lys23Asn), *dmlR_1* (missense_variant c.395A>G p.Asp132Gly), *yebQ* (missense_variant c.1048T>C p.Ser350Pro), *yebQ* (missense_variant c.1103T>C p.Ile368Thr), *prc* (missense_variant c.374A>G p.Glu125Gly), *proQ* (frameshift_variant c.148delA p.Met50fs), *znuC* (missense_variant c.70T>C p.Ser24Pro), *yecC* (missense_variant c.467C>A p.Ala156Glu), *fliL* (missense_variant c.55C>A p.Leu19Met), *hchA* (synonymous_variant c.453T>C p.Val151Val), *zinT* (missense_variant c.558G>T p.Gln186His), *cobU* (frameshift_variant c.20delG p.Gly7fs), *cpsB* (synonymous_variant c.459A>G p.Leu153Leu), *fcl* (missense_variant c.176T>C p.Val59Ala), *wzc* (missense_variant c.725T>C p.Ile242Thr), *yegH* (missense_variant c.448A>G p.Met150Val), *gatD* (missense_variant c.479A>G p.Glu160Gly), *yejO* (frameshift_variant c.346delG p.Ala116fs), *ccmC* (missense_variant c.416A>C p.Tyr139Ser), *atoC* (missense_variant c.1031A>G p.Asp344Gly), *atoB* (frameshift_variant c.341dupG p.Met115fs), *arnC* (missense_variant c.136A>G p.Ile46Val), *arnC* (missense_variant c.563C>G p.Ala188Gly), *menD* (frameshift_variant c.488dupG p.Val164fs), *hisJ* (frameshift_variant c.752delA p.Lys251fs), *ypdA_2* (missense_variant c.1388A>G p.Asn463Ser), *murP* (missense_variant c.43A>G p.Thr15Ala), *eutE* (synonymous_variant c.913C>T p.Leu305Leu), *hmp* (frameshift_variant c.110dupC p.Glu38fs), *recN* (missense_variant c.1616T>C p.Val539Ala), *mutS* (missense_variant c.515G>C p.Arg172Pro), *ptrA* (missense_variant c.2741T>C p.Val914Ala), *ygeK* (frameshift_variant c.7delG p.Ala3fs), *xdhA* (synonymous_variant c.1323C>T p.Gly441Gly), *ygfB* (missense_variant c.455A>G p.Glu152Gly), *speA* (missense_variant c.409T>C p.Ser137Pro), *metK* (missense_variant c.349G>A p.Ala117Thr), *trmI* (frameshift_variant c.518delG p.Gly173fs), *mutY* (missense_variant c.469A>G p.Lys157Glu), *hybD* (synonymous_variant c.216A>G p.Gly72Gly), *rpoD* (missense_variant c.260T>C p.Val87Ala), *fadH* (missense_variant c.1382T>A p.Ile461Asn), *gltD* (missense_variant c.310T>C p.Cys104Arg), *sspA* (frameshift_variant c.438_445delGCCGTACT p.Lys146fs), *trpS* (missense_variant c.80A>C p.Gln27Pro), *trpS* (synonymous_variant c.73C>T p.Leu25Leu), *envZ* (frameshift_variant c.721dupG p.Val241fs), *glgX* (frameshift_variant c.249delG p.His84fs), *livH* (missense_variant c.499G>A p.Ala167Thr), *zntA* (synonymous_variant c.1332T>C p.Gly444Gly), *yhiI* (missense_variant c.266A>G p.Glu89Gly), *yhjB* (missense_variant c.341T>C p.Val114Ala), *yhjG* (missense_variant c.2035A>C p.Thr679Pro), *bcsC* (frameshift_variant c.2715delG p.Pro906fs), *yhjV* (missense_variant c.868A>G p.Ile290Val), *yiaT* (frameshift_variant c.418dupG p.Glu140fs), *yibJ* (missense_variant c.97C>A p.Gln33Lys), *gpsA* (synonymous_variant c.987T>G p.Gly329Gly), *mdtL* (missense_variant c.95A>G p.Asp32Gly), *kup* (frameshift_variant c.246dupG p.Ile83fs), *ilvY* (missense_variant c.506T>G p.Ile169Ser), *aslA* (synonymous_variant c.1503A>G p.Gly501Gly), *ubiE* (missense_variant c.43A>G p.Thr15Ala), *yihL* (frameshift_variant c.397dupG p.Ala133fs), *cytR* (missense_variant c.361A>C p.Thr121Pro), *metF* (missense_variant c.118A>C p.Thr40Pro), *fsaB*

	(missense_variant c.265A>G p.Thr89Ala), *murI* (missense_variant c.173T>C p.Phe58Ser), *rpoB* (missense_variant c.443A>T p.Gln148Leu), *rpoB* (missense_variant c.3995C>T p.Ser1332Leu), *thiF* (missense_variant c.160A>G p.Thr54Ala), *alsA* (missense_variant c.1396T>C p.Ser466Pro), *adiA* (frameshift_variant c.291delA p.Ala98fs), *yeiP_2* (missense_variant c.532A>C p.Thr178Pro), *blc* (missense_variant c.269A>G p.Glu90Gly), *hflK* (missense_variant c.995A>G p.Glu332Gly), *ytfB* (missense_variant c.385C>T p.Arg129Cys), *yjgL* (missense_variant c.1781A>G p.His594Arg), *idnD* (frameshift_variant c.699delA p.Gly234fs), *idnK* (frameshift_variant c.261delA p.Lys87fs), *fimD* (synonymous_variant c.909T>C p.Gly303Gly), *creC* (frameshift_variant c.1048delA p.Ile350fs)
6TUB165-MB3	*nhaR* (missense_variant c.653T>A p.Ile218Asn), *ribF* (missense_variant c.182A>G p.Glu61Gly), *djlA* (frameshift_variant c.19delA p.Ile7fs), *djlA* (synonymous_variant c.552T>C p.Gly184Gly), *rluA* (missense_variant c.19A>G p.Asn7Asp), *ilvH* (missense_variant c.107C>T p.Ala36Val), *cra* (missense_variant c.852T>G p.Cys284Trp), *pdhR* (missense_variant c.314A>C p.His105Pro), *lpd* (missense_variant c.586G>A p.Ala196Thr), *can* (missense_variant c.545G>T p.Gly182Val), *panB* (synonymous_variant c.420A>G p.Thr140Thr), *ldcC* (missense_variant c.1403C>T p.Ala468Val), *frsA* (frameshift_variant c.304delG p.Ala102fs), *ykfA* (missense_variant c.341G>A p.Arg114His), *cynR* (missense_variant c.307A>G p.Ser103Gly), *tauD* (missense_variant c.511C>T p.His171Tyr), *rdgC* (missense_variant c.803A>C p.Asp268Ala), *brnQ* (missense_variant c.586A>G p.Thr196Ala), *panE* (missense_variant c.290A>G p.His97Arg), *ybaO* (missense_variant c.452T>C p.Ile151Thr), *acrB* (frameshift_variant c.72dupG p.Leu25fs), *hemH* (missense_variant c.880A>G p.Lys294Glu), *ybaL* (synonymous_variant c.267T>C p.Ile89Ile), *ybcF* (missense_variant c.461T>C p.Val154Ala), *sfmH* (frameshift_variant c.332dupC p.Gln112fs), *ybcH* (missense_variant c.760A>G p.Thr254Ala), *fepA* (frameshift_variant c.1062delA p.Ala355fs), *entF* (synonymous_variant c.3153T>C p.Arg1051Arg), *entC* (missense_variant c.677G>A p.Arg226His), *pagP* (frameshift_variant c.460delG p.Ala154fs), *leuS* (missense_variant c.283A>G p.Thr95Ala), *chiP* (missense_variant c.587A>G p.Asn196Ser), *ybfF* (missense_variant c.578C>T p.Ala193Val), *potE* (frameshift_variant c.1146delC p.Ser383fs), *ybfC* (frameshift_variant c.121delT p.Ser41fs), *ybgD* (missense_variant c.479T>C p.Phe160Ser), *gltA* (missense_variant c.815C>T p.Ala272Val), *modF* (synonymous_variant c.747A>C p.Ser249Ser), *uvrB* (synonymous_variant c.816A>G p.Lys272Lys), *ybhF* (synonymous_variant c.112T>C p.Leu38Leu), *dinG* (missense_variant c.1136A>G p.His379Arg), *rhtA* (missense_variant c.427T>C p.Ser143Pro), *artQ* (frameshift_variant c.117delA p.Lys39fs), *poxB* (frameshift_variant c.643delG p.Ala215fs), *clpA* (missense_variant c.1501A>G p.Lys501Glu), *lrp* (missense_variant c.428A>C p.Asp143Ala), *ycaD* (synonymous_variant c.999A>G p.Gly333Gly), *ycbL* (missense_variant c.113T>C p.Ile38Thr), *elfC* (missense_variant c.2180A>G p.Asp727Gly), *uup* (frameshift_variant c.1630dupA p.Thr544fs), *rlmI* (frameshift_variant c.504delC p.Val169fs), *rlmI* (frameshift_variant c.223delT p.Ser75fs), *gfcD* (missense_variant c.772A>G p.Asn258Asp), *insA-2* (synonymous_variant c.72T>C p.Thr24Thr), *insA-2* (missense_variant c.243_249delCCGCCATinsACGTCAC p.Phe81Leu), *insAB-1* (missense_variant c.69G>T p.Arg23Ser), *rutF* (missense_variant c.100T>C p.Phe34Leu), *rutD* (missense_variant c.664G>A p.Ala222Thr), *putA* (missense_variant c.661T>C p.Ser221Pro), *csgA* (missense_variant c.412A>G p.Thr138Ala), *ymdC* (missense_variant c.740A>T p.Glu247Val), *lpoB* (synonymous_variant c.291G>T p.Val97Val), *thiK* (missense_variant c.509T>C p.Val170Ala), *mcrA* (synonymous_variant c.663A>G p.Gly221Gly), *ymgD* (missense_variant c.197A>G p.Asp66Gly), *umuC* (missense_variant c.814G>A p.Ala272Thr), *cvrA* (missense_variant c.1211T>C p.Val404Ala), *adhE* (missense_variant c.394A>C p.Asn132His), *oppD* (synonymous_variant c.975T>C p.Arg325Arg), *trpA* (missense_variant c.703T>C p.Ser235Pro), *yciK* (missense_variant c.751T>C p.Ser251Pro), *puuP* (frameshift_variant c.566delG p.Gly189fs), *pspF* (synonymous_variant c.816T>C p.Leu272Leu), *pspC* (synonymous_variant c.12T>C p.Ile4Ile), *abgA* (synonymous_variant c.195A>G p.Leu65Leu), *ydbK* (missense_variant c.2068A>G p.Thr690Ala), *ldhA* (missense_variant c.613T>A p.Cys205Ser), *paaH* (synonymous_variant c.117A>C p.Ser39Ser), *narY* (missense_variant c.376C>A p.Arg126Ser), *narZ* (synonymous_variant c.2302C>T p.Leu768Leu), *narZ* (missense_variant c.1069A>G p.Ile357Val), *maeA* (synonymous_variant c.1632G>A p.Leu544Leu), *ydeK* (frameshift_variant c.3962dupG p.Thr1322fs), *ydeK* (frameshift_variant c.3837_3838dupGG p.Glu1280fs), *lsrB* (missense_variant c.770G>T p.Arg257Leu), *ydgI* (frameshift_variant c.842delG p.Gly281fs), *fumC* (frameshift_variant c.1052dupG p.Ala352fs), *uidA* (missense_variant c.1139T>C p.Val380Ala), *add* (stop_gained c.624G>A p.Trp208*), *rsxA* (missense_variant c.342T>G p.Cys114Trp), *lhr* (frameshift_variant c.1286delT p.Phe429fs), *ydhC* (synonymous_variant c.309A>C p.Val103Val), *ydhV* (missense_variant c.1958A>G p.Asp653Gly), *ydiI* (missense_variant c.169G>A p.Ala57Thr), *ydiK* (missense_variant c.127A>C p.Thr43Pro), *astD* (synonymous_variant c.21T>C p.Gly7Gly), *astA* (missense_variant c.611C>T p.Ala204Val), *yeaE* (missense_variant c.151A>C p.Thr51Pro), *proQ* (frameshift_variant c.67dupC p.His23fs), *yecM* (frameshift_variant c.312delA p.Lys104fs), *otsB* (synonymous_variant c.279A>G p.Thr93Thr), *yedK* (synonymous_variant c.360T>C p.Arg120Arg), *fliJ* (frameshift_variant c.315delA p.Lys105fs), *fliK* (frameshift_variant c.145delC p.Gln49fs), *cobT* (missense_variant c.199T>C p.Cys67Arg), *plaP* (missense_variant c.1310T>C p.Val437Ala), *wbbH* (missense_variant c.632T>C p.Val211Ala), *yegE* (frameshift_variant c.419dupG p.Val141fs), *yegV* (missense_variant c.445T>C p.Ser149Pro), *yehF* (missense_variant c.668A>T p.Asp223Val), *dld* (missense_variant c.364A>G p.Thr122Ala), *yohK* (synonymous_variant c.516T>G p.Arg172Arg), *folE* (synonymous_variant c.57A>G p.Gly19Gly), *yeiG* (missense_variant c.596A>G p.Asp199Gly), *fruA* (missense_variant c.101T>C p.Ile34Thr), *alkB* (frameshift_variant c.270delC p.Ala91fs), *yfaL* (missense_variant c.2150A>T p.Asp717Val), *yfaL* (frameshift_variant c.1921dupG p.Val641fs), *yfaL* (frameshift_variant c.1363delG p.Ala455fs), *glpT_1* (missense_variant c.1178A>G p.Tyr393Cys), *evgS* (frameshift_variant c.942dupC p.Tyr315fs), *evgS* (stop_gained c.2203G>T p.Glu735*), *xapB* (missense_variant c.701T>C p.Phe234Ser), *xapB* (missense_variant c.601G>A p.Ala201Thr), *xapB* (missense_variant c.474T>G p.Ser158Arg), *dapA* (missense_variant c.827A>G p.Asp276Gly), *hyfB* (synonymous_variant c.1410G>A p.Val470Val), *hyfG* (frameshift_variant c.920delT p.Phe307fs), *yfgD* (missense_variant c.53A>G p.Glu18Gly), *ppx* (synonymous_variant c.441A>T p.Gly147Gly), *yfhM* (missense_variant c.2739T>A p.Asp913Glu), *hscB* (synonymous_variant c.291G>A p.Glu97Glu), *tadA* (missense_variant c.490T>C p.Ser164Pro), *smpB* (missense_variant c.355A>G p.Lys119Glu), *emrA* (missense_variant c.761T>C p.Val254Ala), *mutS* (missense_variant c.128A>G

p.Asp43Gly), *ygbJ* (missense_variant c.46T>C p.Ser16Pro), *casA* (missense_variant c.1321T>C p.Ser441Pro), *cysI* (missense_variant c.1165G>A p.Asp389Asn), *araE* (synonymous_variant c.300A>G p.Val100Val), *ygfM* (missense_variant c.262G>T p.Ala88Ser), *ygfM* (synonymous_variant c.435C>T p.Cys145Cys), *prfB_2* (missense_variant c.494C>T p.Ala165Val), *cmtA* (missense_variant c.1277A>G p.Glu426Gly), *speA* (missense_variant c.866A>G p.Glu289Gly), *metK* (missense_variant c.677A>G p.Asn226Ser), *trmI* (frameshift_variant c.518delG p.Gly173fs), *glcA* (synonymous_variant c.855T>C p.Gly285Gly), *ysgA_1* (frameshift_variant c.340delC p.Arg114fs), *yedW_8* (missense_variant c.230T>C p.Ile77Thr), *yqiG* (missense_variant c.355A>G p.Thr119Ala), *rpoD* (missense_variant c.1340C>T p.Ala447Val), *fadH* (missense_variant c.272T>C p.Val91Ala), *sstT* (frameshift_variant c.1041dupG p.Ser348fs), *tdcG* (synonymous_variant c.1158A>G p.Gly386Gly), *yhbQ* (missense_variant c.127A>G p.Lys43Glu), *lptC* (missense_variant c.448A>C p.Thr150Pro), *gltF* (missense_variant c.740T>C p.Val247Ala), *yhdX* (missense_variant c.728T>C p.Val243Ala), *trkA* (missense_variant c.1151T>C p.Val384Ala), *frlR* (missense_variant c.83T>C p.Val28Ala), *php* (missense_variant c.457C>T p.Arg153Cys), *pck* (synonymous_variant c.687T>C p.Ala229Ala), *malP* (missense_variant c.1487T>C p.Phe496Ser), *livK* (missense_variant c.949A>G p.Thr317Ala), *livJ* (synonymous_variant c.510A>G p.Lys170Lys), *yhjH* (frameshift_variant c.326delC p.Pro109fs), *bcsA* (missense_variant c.836T>C p.Val279Ala), *bcsF* (missense_variant c.167A>G p.Glu56Gly), yhjX (missense_variant c.385T>C p.Phe129Leu), glyS (missense_variant c.485A>G p.His162Arg), yiaJ (frameshift_variant c.27dupA p.Glu10fs), *yibJ* (missense_variant c.97C>A p.Gln33Lys), *yibH* (stop_gained c.367C>T p.Gln123*), *trmL* (missense_variant c.196T>C p.Phe66Leu), *spoT* (missense_variant c.770T>C p.Ile257Thr), *recG* (synonymous_variant c.996G>A p.Gln332Gln), *fimZ_7* (frameshift_variant c.283delG p.Ala95fs), *yidE* (frameshift_variant c.1627dupC p.Gln543fs), *dgoR* (missense_variant c.556G>T p.Ala186Ser), *mioC* (missense_variant c.232C>T p.Pro78Ser), *gpp* (missense_variant c.1063C>T p.Leu355Phe), *rho* (missense_variant c.191T>C p.Phe64Ser), *recQ* (missense_variant c.265G>A p.Ala89Thr), *rhtB* (missense_variant c.136C>T p.Leu46Phe), *pepQ* (missense_variant c.985A>G p.Thr329Ala), *yihW* (missense_variant c.758A>G p.Glu253Gly), *yihY* (synonymous_variant c.30G>A p.Arg10Arg), *yihY* (synonymous_variant c.183T>C p.Phe61Phe), *frvX* (frameshift_variant c.634delG p.Ala212fs), *frvA* (missense_variant c.254A>G p.Asp85Gly), *rhaD* (missense_variant c.770T>C p.Ile257Thr), *rhaS* (synonymous_variant c.538C>T p.Leu180Leu), *ptsA* (missense_variant c.52C>T p.Arg18Cys), *rplL* (missense_variant c.288A>T p.Lys96Asn), *rpoC* (missense_variant c.524A>G p.Glu175Gly), *rpoC* (missense_variant c.548A>G p.Glu183Gly), *rpoC* (missense_variant c.2315A>G p.Tyr772Cys), *rpoC* (synonymous_variant c.3648G>A p.Ala1216Ala), *epmA* (synonymous_variant c.783C>T p.Arg261Arg), *amiB* (missense_variant c.682A>G p.Thr228Ala), *ytfE* (missense_variant c.14A>G p.Asp5Gly), *ytfL* (missense_variant c.1205T>C p.Phe402Ser), *ytfR* (synonymous_variant c.888A>T p.Gly296Gly), *nanM* (frameshift_variant c.551dupT p.Phe185fs), *deoA* (missense_variant c.33A>C p.Lys11Asn), *deoB* (synonymous_variant c.216T>A p.Ile72Ile), *yjjK* (synonymous_variant c.601C>T p.Leu201Leu), *trpR* (missense_variant c.173A>C p.Gln58Pro), *yaaJ* (missense_variant c.1301T>C p.Val434Ala)

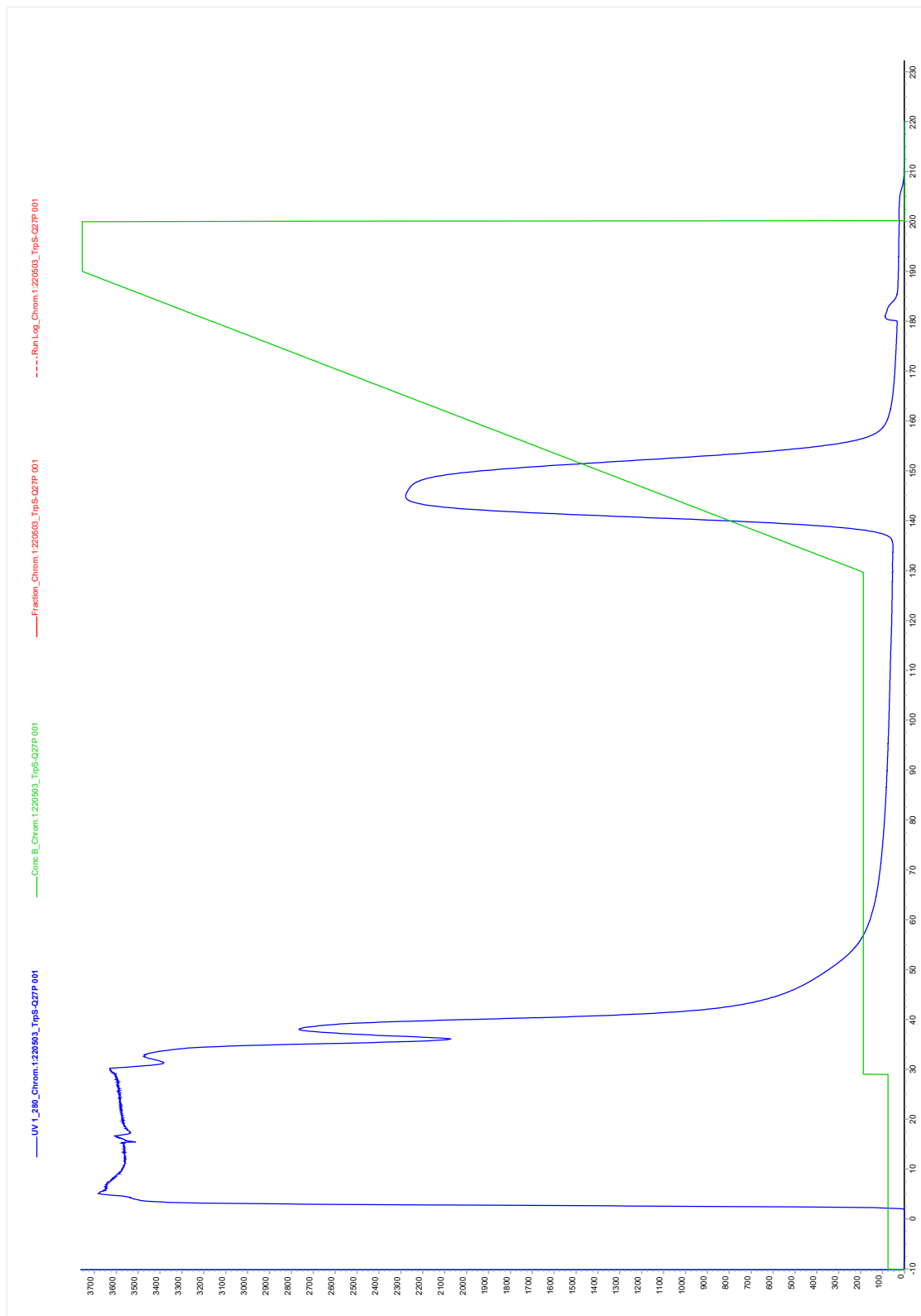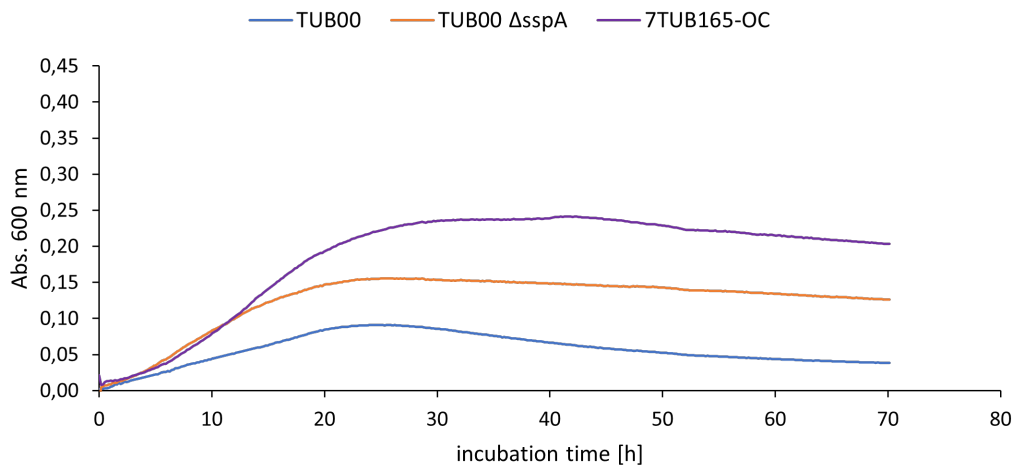**TrpRS purification and LC-ESI-MS analyses**

Figure S73 | Representative Äkta chromatogram for TrpRS purification. The Äkta FPLC readout for the purification of the TrpRS Q27P variant is shown. The UV 280 nm detection signal, the purification gradient and purification steps as well as the fractioning are indicated.
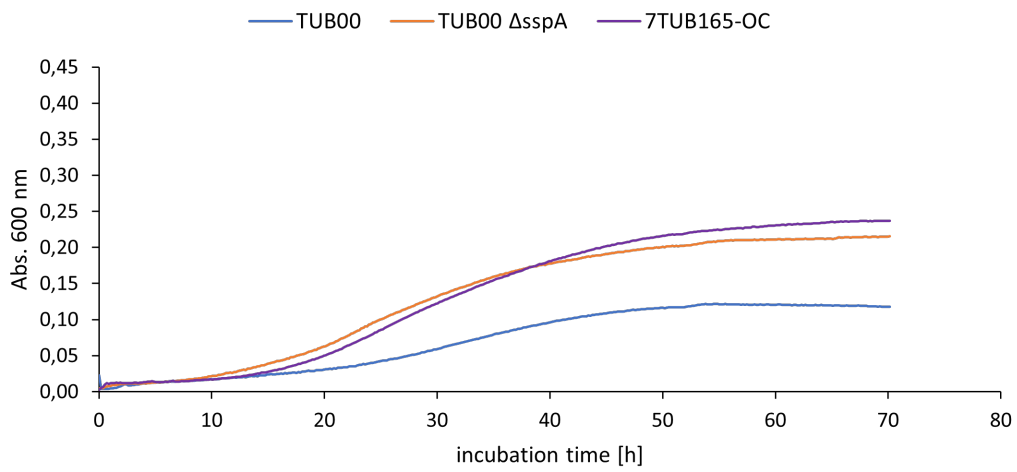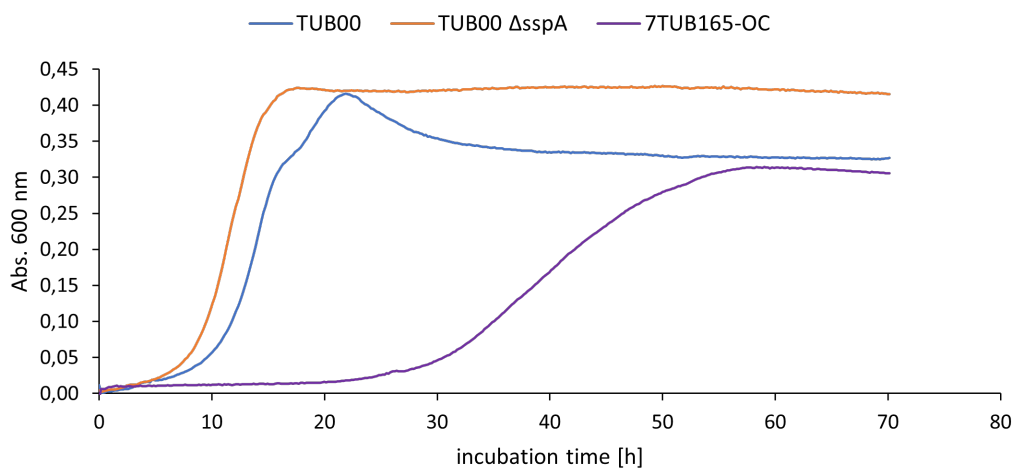
Growth assay to determine the *sspA* impact

(A) 19-70-1 (6Fi, Ind)

(B) 19-70-0 (6Fi)

(C) 0-70-0 (6Fi)

(D) 0-0-70 (Ind)

(E) 0-0-0

(F) 19-70-1 (7Fi, Ind)
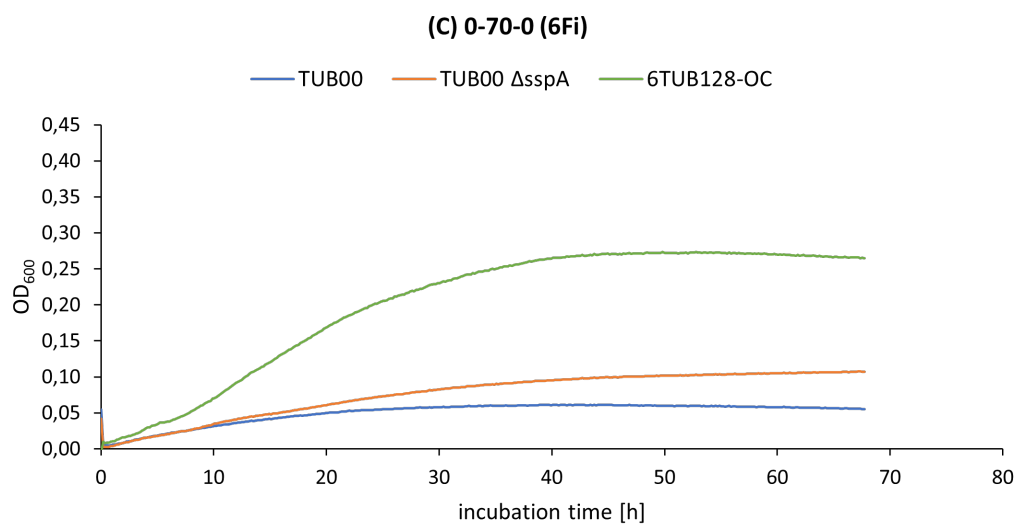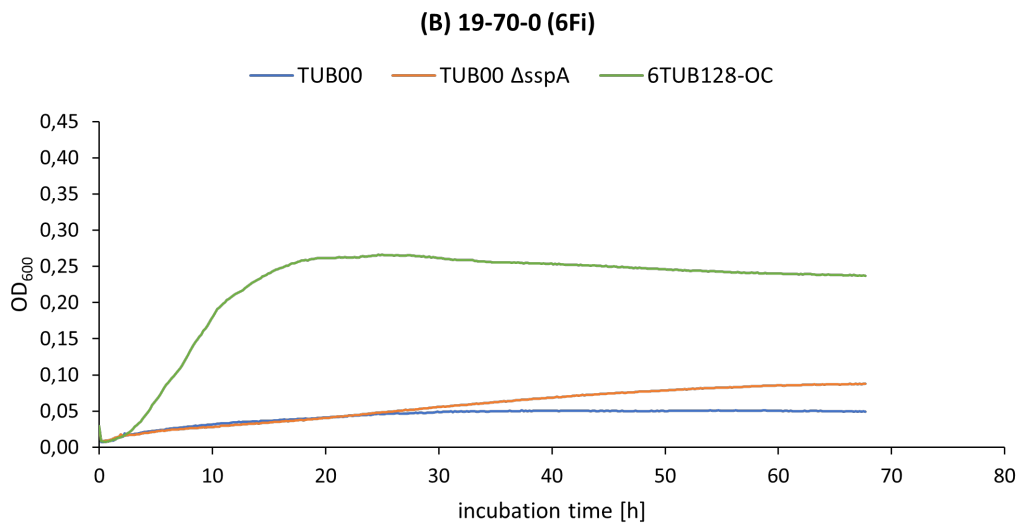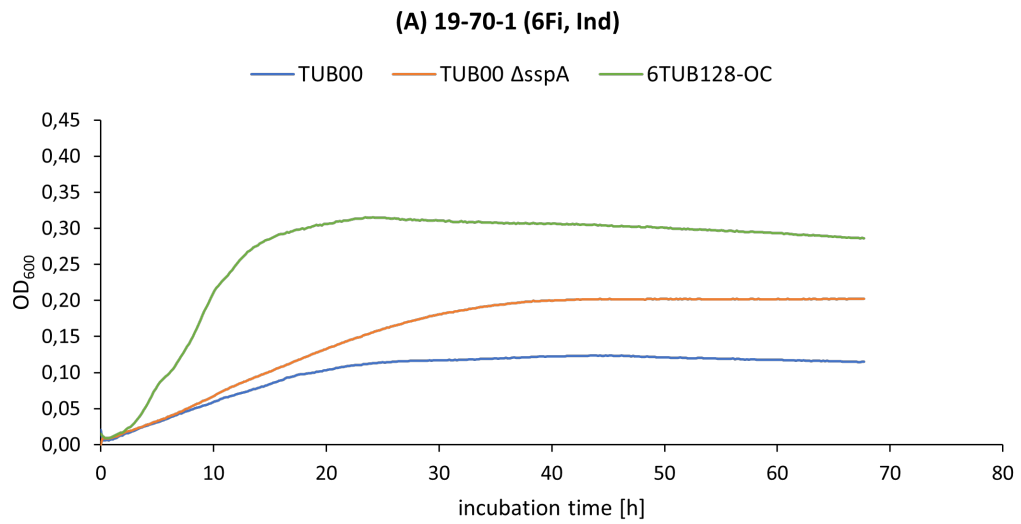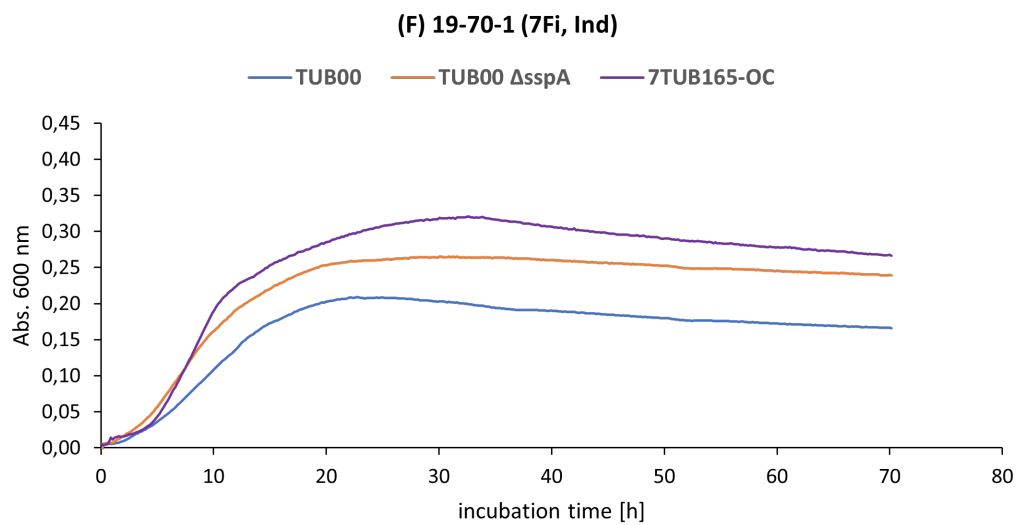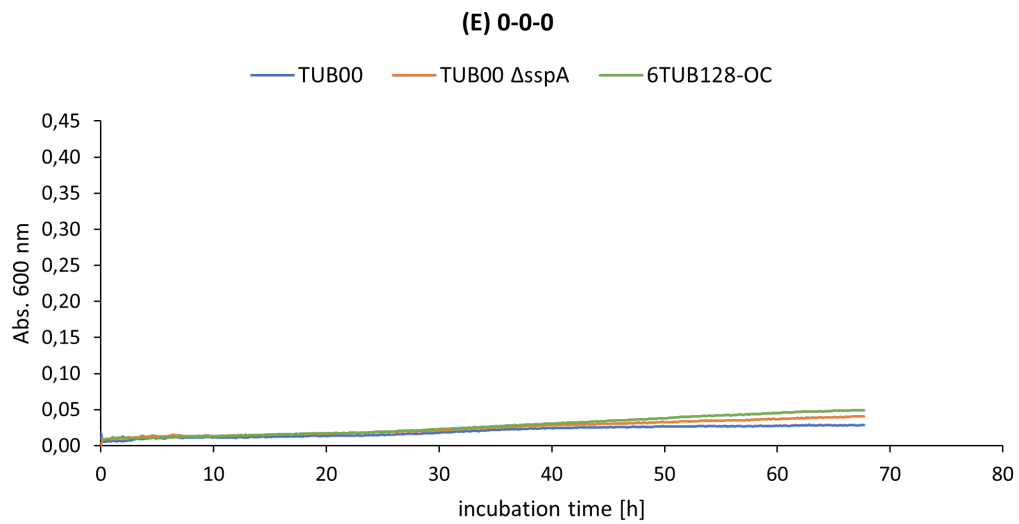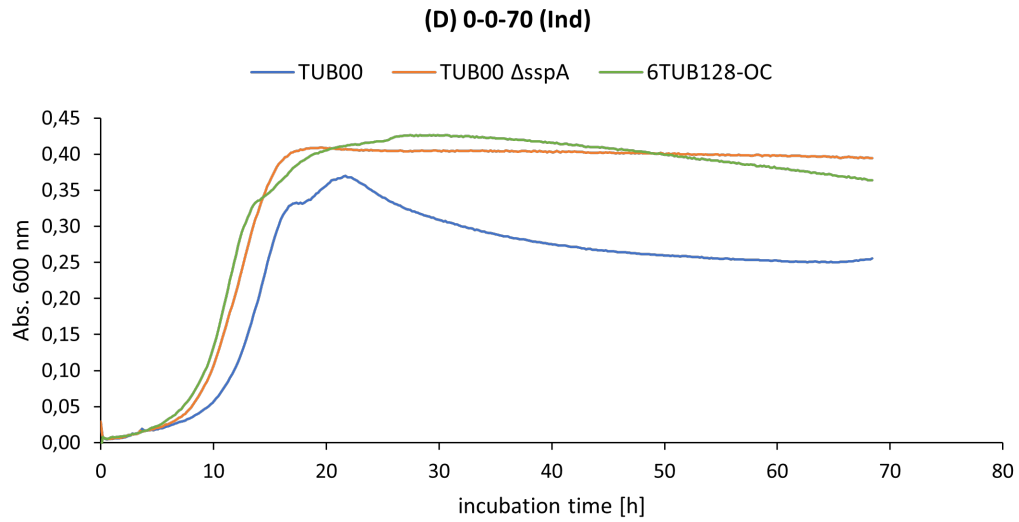
(G) 19-70-0 (7Fi)

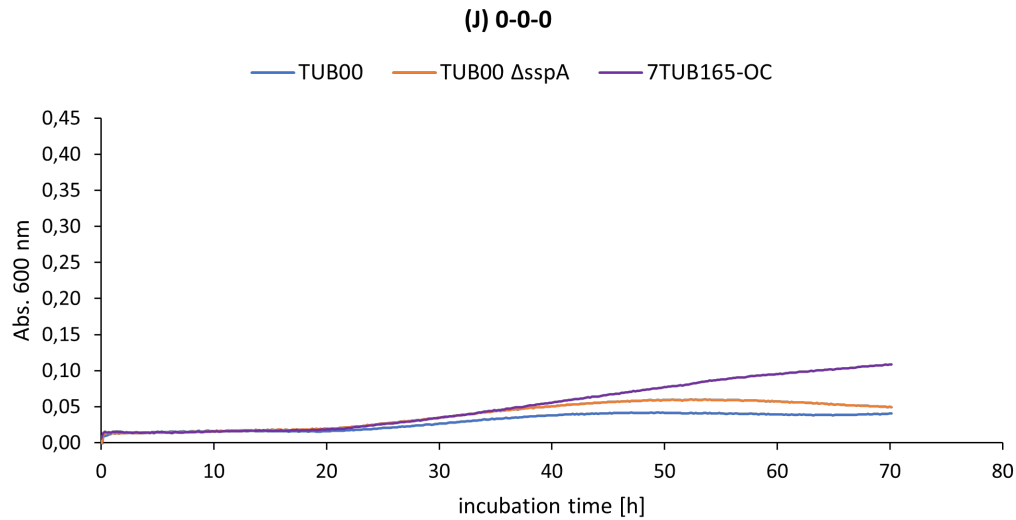## (H) 0-70-0 (7Fi)

(I) 0-0-70 (Ind)

(J) 0-0-0

Figure S74| Growth curves of the *sspA* growth assay. In A – E, the curves evaluating the *sspA* impact in 6Fi-supplemented NMM and in F – J those for 7Fi-supplemented medium are shown. The absorption at 600 nm is plotted against the incubation time [h] at 30 °C. Growth curves of TUB00 are colored in blue, curves of the deletion mutant TUB00 Δ*sspA* are orange, curves of 6TUB128-OC in green and curves of 7TUB165-OC in purple. All measurements were performed at microscale level in triplicates using a plate reader.
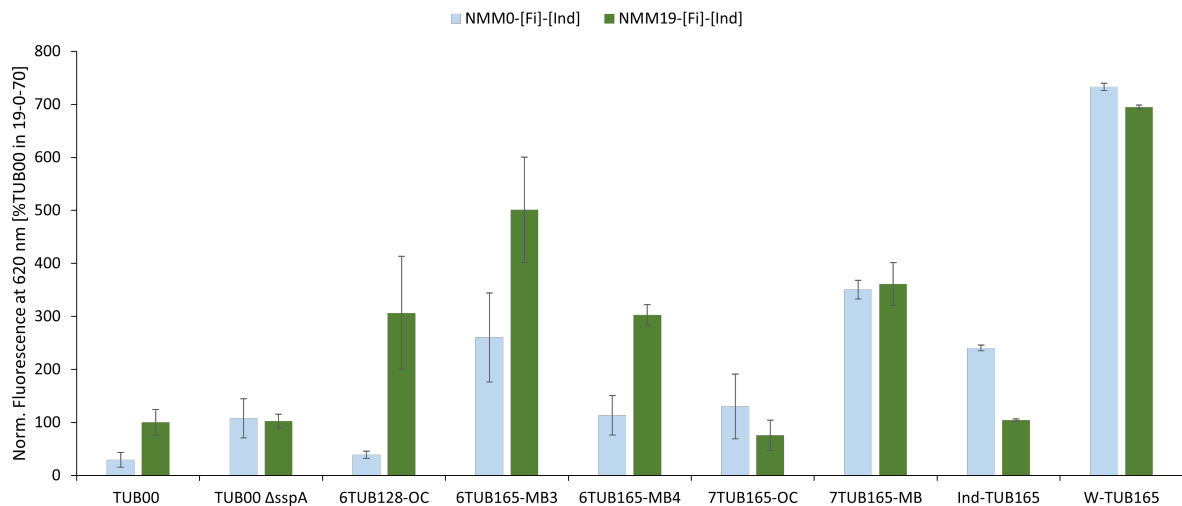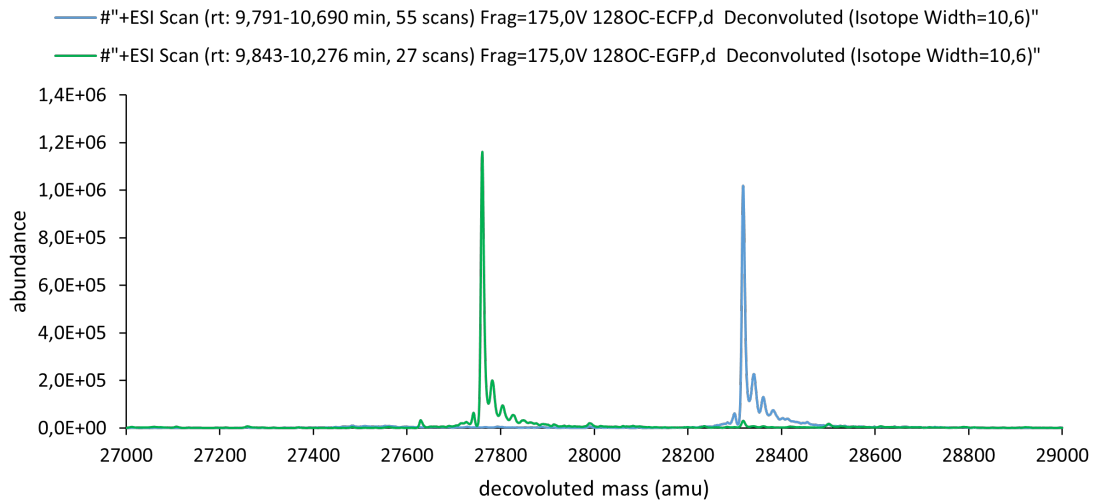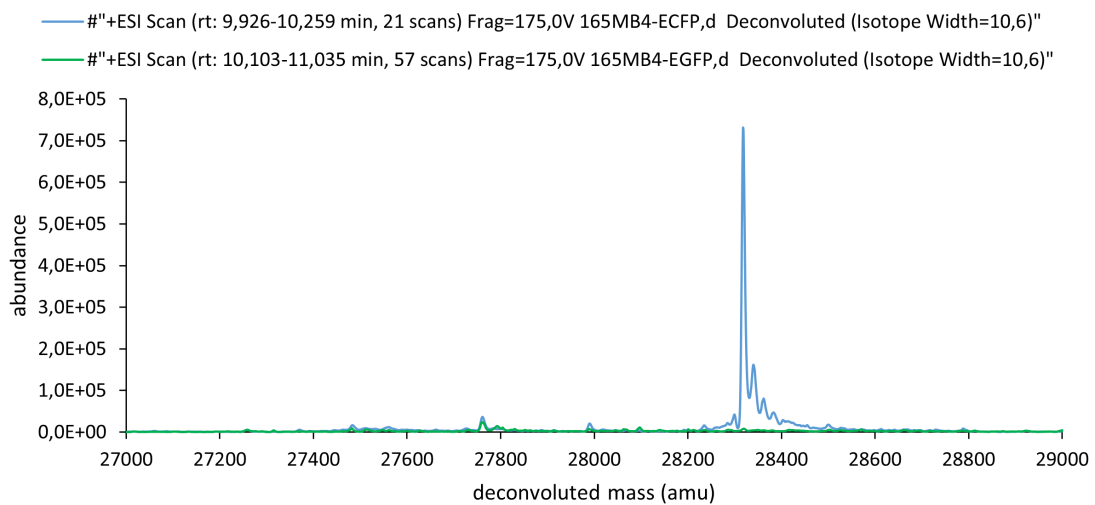
Nile Red fluorescence assay

Figure S75| Determination of membrane permeability. Quantitative fluorescence assay of Nile Red stained cells is shown.

Expression of ECFP and EGFP in ALE strains
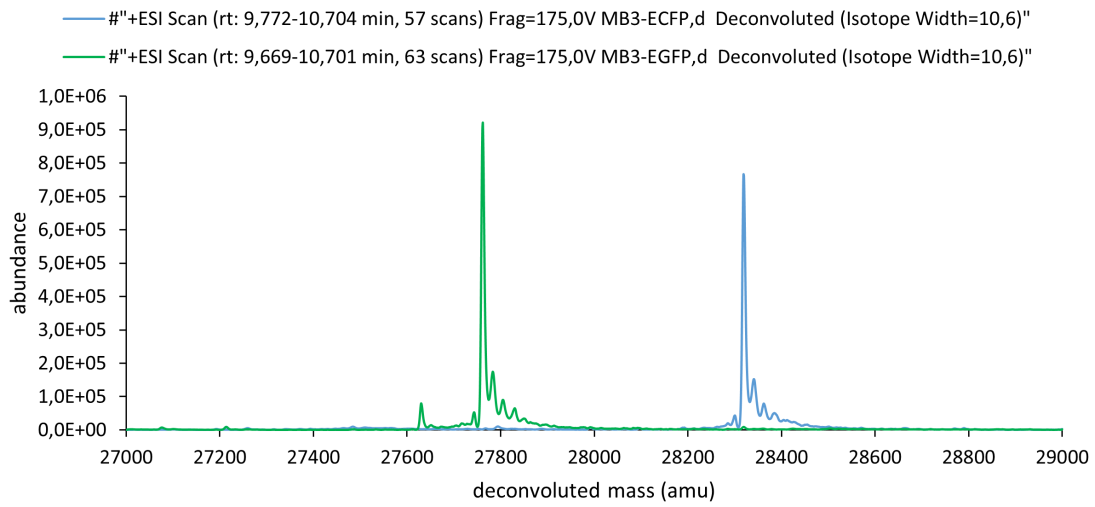
(A) 6TUB128-OC expressed 2x6FW-ECFP and 1x6FW-EGFP

—— #"+ESI Scan (rt: 9,791-10,690 min, 55 scans) Frag=175,0V 128OC-ECFP,d Deconvoluted (Isotope Width=10,6)"

—— #"+ESI Scan (rt: 9,843-10,276 min, 27 scans) Frag=175,0V 128OC-EGFP,d Deconvoluted (Isotope Width=10,6)"

(B) 6TUB165-MB4 expressed 2x6FW-ECFP and 1x6FW-EGFP

—— #"+ESI Scan (rt: 9,926-10,259 min, 21 scans) Frag=175,0V 165MB4-ECFP,d Deconvoluted (Isotope Width=10,6)"

—— #"+ESI Scan (rt: 10,103-11,035 min, 57 scans) Frag=175,0V 165MB4-EGFP,d Deconvoluted (Isotope Width=10,6)"

(C) 6TUB165-MB3 expressed 2x6FW-ECFP and 1x6FW-EGFP

—— #"+ESI Scan (rt: 9,772-10,704 min, 57 scans) Frag=175,0V MB3-ECFP,d Deconvoluted (Isotope Width=10,6)"

—— #"+ESI Scan (rt: 9,669-10,701 min, 63 scans) Frag=175,0V MB3-EGFP,d Deconvoluted (Isotope Width=10,6)"

(D) 7TUB165-OC expressed 2x7FW-ECFP and 1x7FW-EGFP

—— #"+ESI Scan (rt: 9,904-11,103 min, 73 scans) Frag=175,0V 7OC-ECFP,d Deconvoluted (Isotope Width=10,6)"

—— #"+ESI Scan (rt: 9,831-11,046 min, 74 scans) Frag=175,0V 7OC-EGFP,d Deconvoluted (Isotope Width=10,6)"
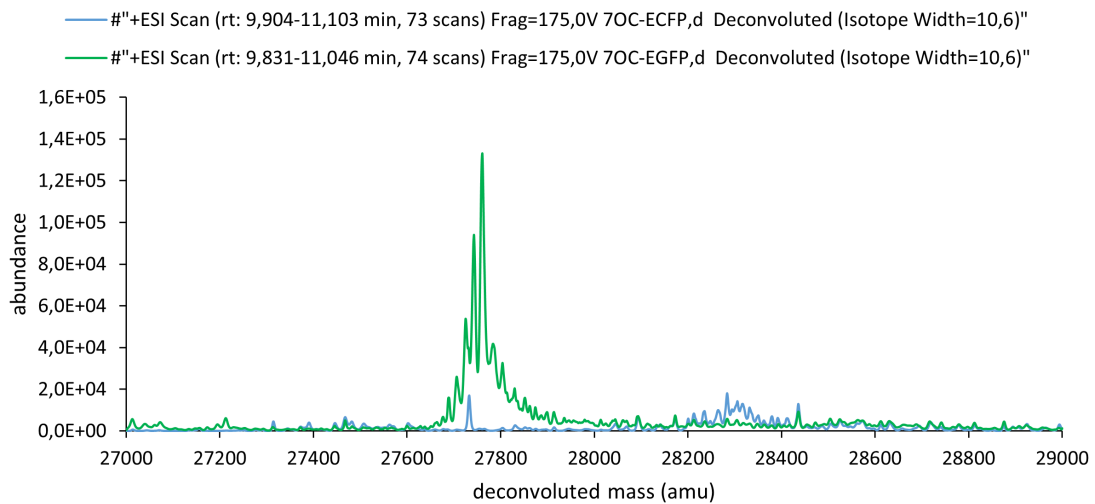
(E) 7TUB165-MB expressed 2x7FW-ECFP and 1x7FW-EGFP

—— #"+ESI Scan (rt: 9,890-10,889 min, 61 scans) Frag=175,0V 7MB-ECFP,d Deconvoluted (Isotope Width=10,6)"

—— #"+ESI Scan (rt: 9,913-10,246 min, 21 scans) Frag=175,0V 7MB-EGFP,d Deconvoluted (Isotope Width=10,6)"
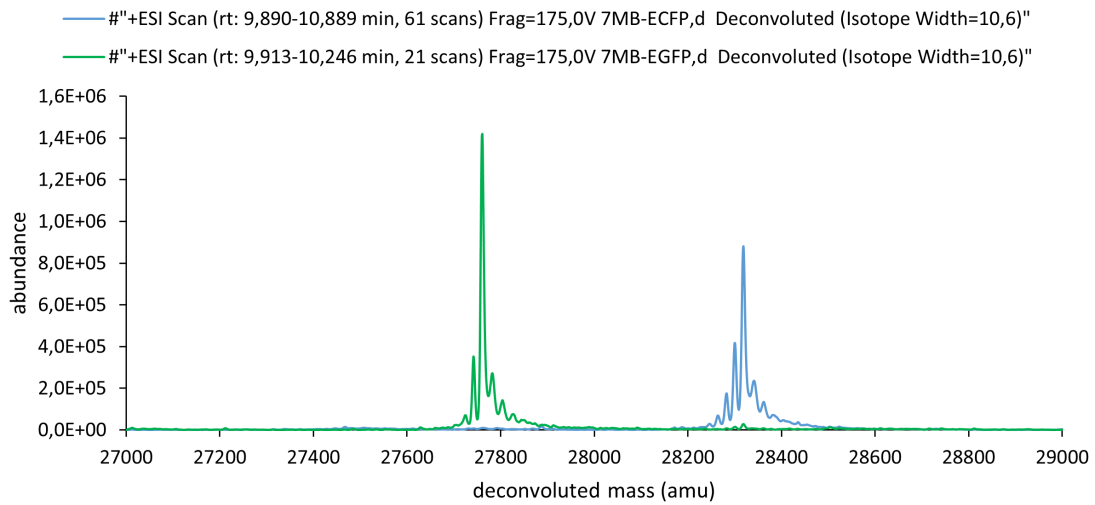
(F) TUB00 expressed 2xW-ECFP and 1xW-EGFP

—— #"+ESI Scan (rt: 9,784-10,583 min, 49 scans) Frag=175,0V TUB00-W-ECFP,d Deconvoluted (Isotope Width=10,6)"

—— #"+ESI Scan (rt: 9,702-10,418 min, 44 scans) Frag=175,0V TUB00-W-EGFP,d Deconvoluted (Isotope Width=10,6)"
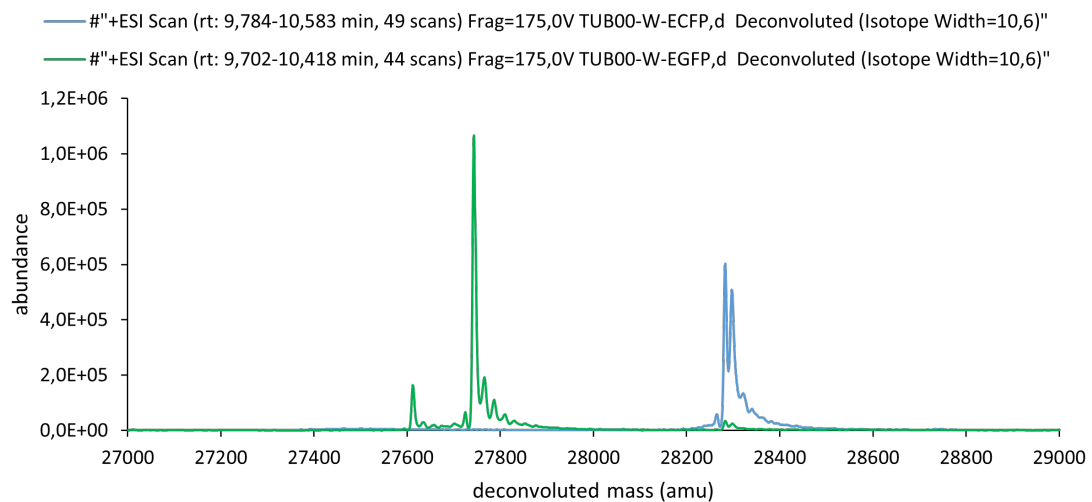
Figure S76| Mass spectrometric analyses of 6FTrp- and 7FTrp-substituted EGFP and ECFP. A) – C) Deconvoluted mass spectra of 6FTrp-substituted ECFP and EGFP expressed in 6Fi-adapted strains, D) - E) deconvoluted mass spectra of 7FTrp-substituted ECFP and EGFP expressed in 7Fi-adapted strains and F) deconvoluted mass spectra of wild-type ECFP and EGFP expressed in TUB00. The ESI scans of EGFP variants are colored in green and of ECFP in blue.

Table S36| Summary of MS analyses of ECFP and EGFP expression in ALE strains. Theoretical and observed mass values as well as normalized abundance of EGFP-H6 and H6-ECFP expressed in 6Fi- and 7Fi-adapted strains, after chromophore maturation.

	EGFP (theoretical mass: 27761.56 Da)		ECFP (theoretical mass: 28319.15 Da)	
	observed mass [Da]	norm. abundance [%]	observed mass [Da]	norm. abundance [%]
6TUB128-OC	27761.63	81.75	28319.36	71.72
6TUB165-MB4	27762.29	1.73	28319.36	51.54
6TUB165-MB3	27762.58	64.92	28320.41	54.01
7TUB165-OC	27761.71	9.38	n.d.	n.d.
7TUB165-MB	27761.44	100.00	28319.59	62.10
TUB00	27744.42	75.15	28283.43	42.51

Table S37| Isolated yields of FTrp-substituted EGFP and ECFP. The protein concentration after purification and dialysis was determined by absorption measurement at 280 nm. EGFP and ECFP expressed in TUB00 are not FTrp substituted.

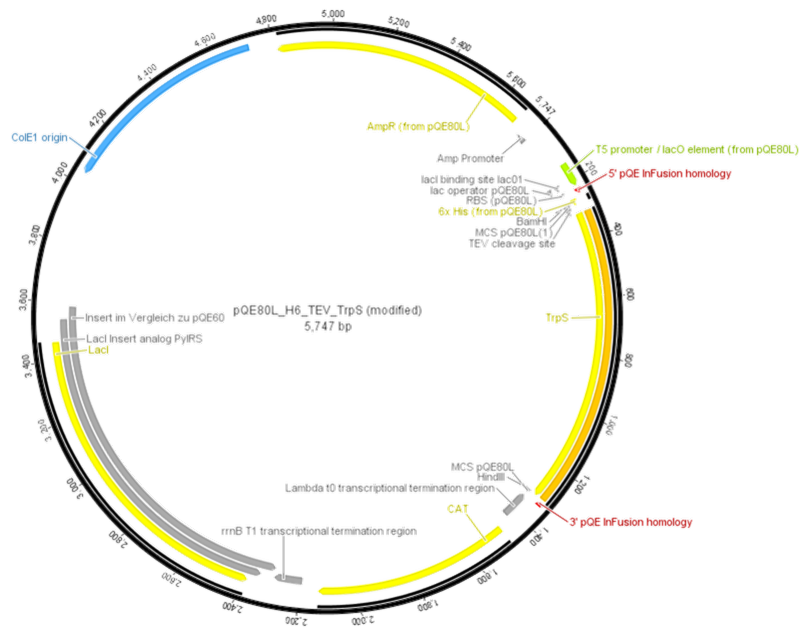|	isolated yield [mg/mL]	
	EGFP	**ECFP**
6TUB128-OC	0.473	0.308
6TUB165-MB4	0.219	0.692
6TUB165-MB3	0.923	0.668
7TUB165-OC	0.133	0.290
7TUB165-MB	0.191	0.495
TUB00	0.480	0.526

Plasmids

Figure S77 | Plasmid map of pQE80L H6 TEV TrpRS. The plasmid carries an ampicillin resistance gene, a chloramphenicol resistance gene (*cat*), ColE1 ORI and the *trpS* gene encoding the TrpRS is under control of a T5 promotor. The plasmid map is representative for TrpRS wt, TrpRS Q27P and TrpRS Q27P L25L plasmids.
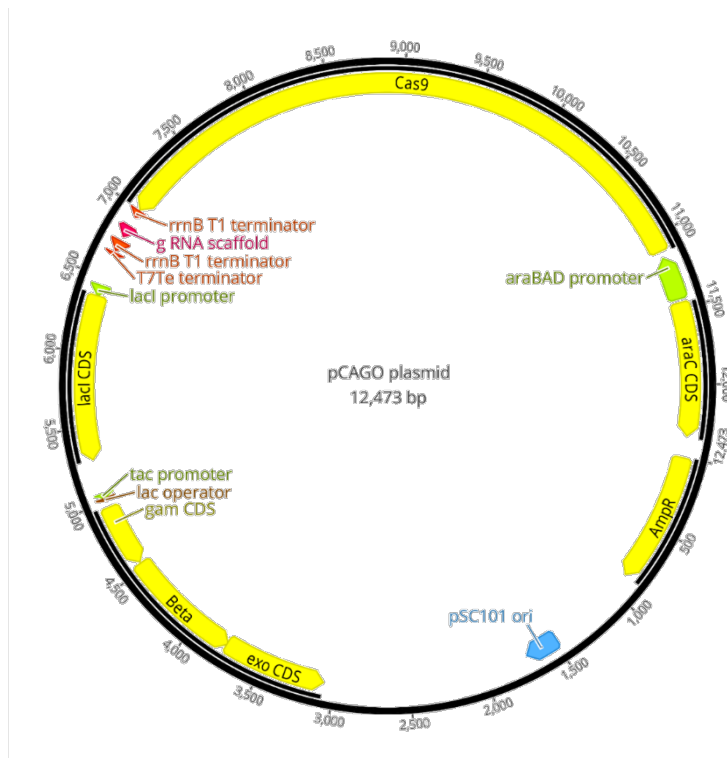

Figure S78 | Plasmid map of pCAGO used for CRISPR/Cas9 genome editing. pCAGO encodes for Cas9 protein under araBAD control, lambda-Red genes for homologous recombination under control of tac promotor, constituently expressed gRNA and an ampicillin resistance gene.
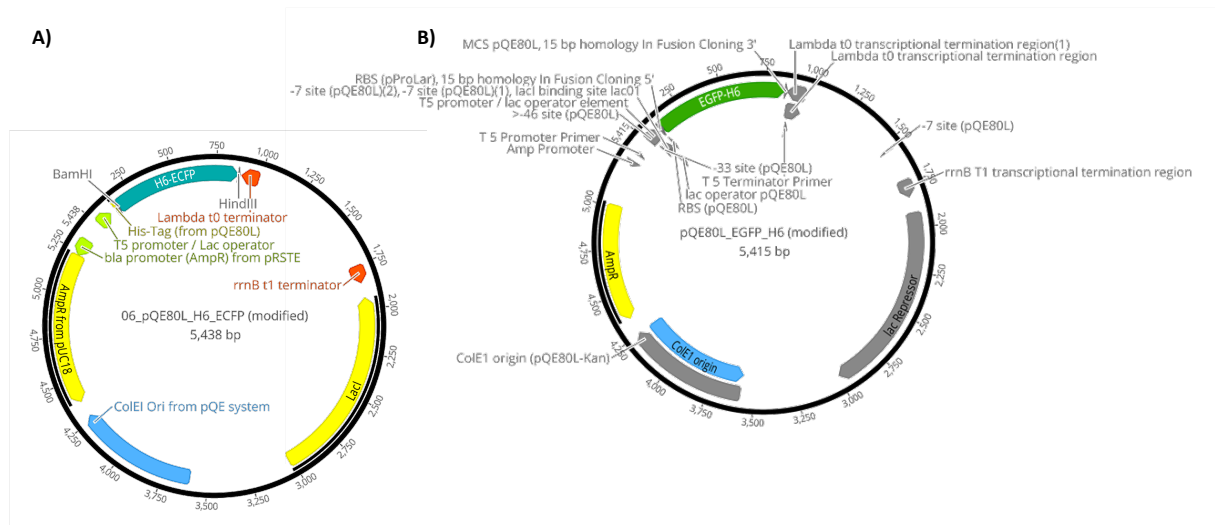
Figure S79 | Plasmid maps of GFPs used for protein expression. A) pQE80L H6-ECFP and B) pQE80L EGFP-H6 are shown, both carry an ampicillin resistance gene, ColE1 ORI and the gene encoding ECFP and EGFP are under control of a T5 promotor, respectively.