TECHNISCHE UNIVERSITÄT MÜNCHEN

Max-Planck-Institut für Biochemie

Genetic code engineering with methionine analogs for synthetic biotechnology

Carlos Guillermo Acevedo Rocha

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Michael Groll

Prüfer der Dissertation:

1. Univ.-Prof. Dr. Nediljko Budisa
2. Univ.-Prof. Dr. Thomas Kiefhaber

Die Dissertation wurde am 24.06.2010 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 23.07.2010 angenommen.

To my lovely family

“To put the world right in order, we must first put the nation in order; to put the nation in order, we must first put the family in order; to put the family in order, we must first cultivate our personal life; we must first set our hearts right.”

Confucius

Parts of this work are planned to be, or have been submitted to publication:

Nehring S*, **Acevedo-Rocha CG***, Hoesl MG*, et al. Lipase congeners with enhanced resistance toward additives. (*contributed equally; in preparation)

Hoesl MG*, **Acevedo-Rocha CG***, Nehring S*, et al. Lipase congeners designed by genetic code engineering. (*contributed equally; submitted)

Poster presentations:

Acevedo-Rocha CG, Merkel L, and Budisa N. “Translation initiation by means of β-Cyclopropylalanine” 2008. International Conference on aminoacyl-tRNA synthetases: From basic mechanisms to systems biology, Veyrier-du-Lac, France

Other scientific publications:

Acevedo-Rocha CG, et al. Synthetic cells and genomes (in preparation)

Acevedo-Rocha CG, Geiermann AS, Budisa N, Merkel L. *In vivo* unnatural translation initiation. (in preparation)

Sharp L and **Acevedo-Rocha CG**. New Frontiers in Science. *Offspring* (Official magazine from the Max Planck PhDnet; in press)

Acevedo-Rocha CG, Yu J y Dobay A. Entgrenzung – transcending boundaries across scientific disciplines. *Arkhaï* 2010; (14)23-40

Acevedo-Rocha CG and Steinbach J. The Art of Science and the Science of Art. *Offspring* 2009; (4)8-13

Acknowledgments

I would like to thank very much the following people:

- Prof. Dr. Nediljko Budisa for his kindness, openness, support, motivation, originality, good sense of humour, priceless freedom, exciting discussions about science, history, politics, and the world. He provided me a solid ground where I could grow as a person and as an open-minded scientist.

- Prof. Dr. Michael Groll and Prof. Dr. Thomas Kiefhaber for allowing me to present them my dissertation.

- Prof. Dr. Thomas Kiefhaber, Prof. Dr. Elena Conti, Dr. Daniel Wilson, and Dr. Birgit Wiltschi for being part of my thesis advisory committee.

- Prof. Dr. Dieter Oesterhelt and his incredible friendly group of people for permitting me to work in his department of “Membrane Biochemistry”.

- Prof. Dr. Garabed Antranikian and Dr. Maryna Royter for allowing me to cooperate and to learn new things about the exciting field of biocatalysis.

- Prof. Dr. Zoya Ignatova for her motivation and input with my thesis work, and for the original scientific cooperations.

- Prof. Dr. Yves Mechulam and Prof. Dr. Peter Schultz for sending kindly some of the plasmids used in this work.

- Prof. Dr. Robert Huber for giving me the great chance to meet his student.

- Dr. Caroline Koehrer for her technical help and nice scientific discussions.

- Dr. Harald Huber, Prof. Dr. Nick Talbot, Dr. Michael Kershaw, Dr. Gaia Travosanis, Mrs. Tatjana Kleine, Mr. Christian Hoffman, and Mrs. Sylvia Reddehase for giving me genomic DNA or cDNA from different organisms.

- Dr. Joerg Tittor, Dr. Gong Zhang, and Mr. Wolfgang Schloegl for their special help with radioactive experiments.

- Mrs. Elisabeth Weyher-Stingl, Mr. Snezan Marinkovic and Mr. Reinhard Mentele for their great expertise and help concerning scientific issues.

- Mrs. Petra Birle, Mrs Tatjana Krywcun, and Mrs. Traudl Wenger for her invaluable technical assistance and help, especially for the cloning work and the purification of alloproteins.

- Dr. Birgit Wiltschi for her great supervision, exciting scientific discussions, and helpful manners to make me think differently about science, as well as her feedback with the thesis and her organized way of working.

- Dr. Yuri Cheburkin for his supervision during my first internship at the Budisa lab, and for his technical aid and advises.

- Dr. Lars Merkel for his cordial support in the lab, synthesis of amino acids, great input with my thesis, and for this great teachings in chemistry and art.

- Prof. Dr. Shouliang Dong for his great support with the synthesis of amino acids.

- Mr. Martin Turk for his great help and expertise related to enzyme modeling.

- Mrs. Pamela David for her exact and great proof-reading quality of my thesis.

- Dr. Maud Larregola, Thomas Steiner, and Michael Hoesl for their great input in my thesis and their cordiality in the last weeks of this voyage.

- My easy-going colleagues Michael Hoesl, Dr. Christina Wolschner, Dr. Sandra Lepthien, Sebastian Nehring, Lena Strube, Anna Skrollan-Geiermann, Dr. Hamid Heidary, Dr. Maud Larregola, and Thomas Steiner for their great support and exciting moments during the PhD.

- My hard-working students Alessandro de Simone, Franziska Benedikter, Harini Venkataraman, Kathy Schmohl, Julia Kobuch, Tian Tian, and Katharina von Roman for their special help and fun in the lab.

- Mr. Konrad Frischeisen for giving me the great opportunity to start a scientific career in Germany, and for his influential discussions on epistemology.

- Dr. Hans-Joerg Schaeffer, Dr. Ingrid Wolf, and Mrs. Maximiliane Reif for their great support and dedication on having an amazing IMPRS office.

- My dearest friends Rosy Elena Yañez García, Ernesto Arturo Márquez Waldthausen, Leonie Waanders, Andreas Kopp, Marisaura López, Konrad Hoenl, Ákos Dobay, and Jerry Yu for their friendship, support, and good moments of joy.

- My wonderful parents Yolanda and Guillermo, as well as my priceless siblings Eyra Angélica, Luis Guillermo, and Braulio for their love, care support, motivation, advises, and guidance during all my life.

- My beloved girlfriend Friederike for her precious love, enormous care, great support, invaluable help, incredible good sense of humour, amazing dedication, and unique patience, among many other great qualities.

Abbreviations

Abbreviations are used according to the International Union of Pure and Applied Chemistry (IUPAC) and the joint commission for biochemical nomenclature (IUPAC-IUB).

A

AA	Amino Acid
AA-AMP	Aminoacyladenylate
AARS	Aminoacyl-tRNA synthetase
*Aa*MetRS	*Aquifex aeolicus* methionyl-tRNA synthetase
Aha	L-Azidohomoalanine
Ala	L-Alanine
Amp	Ampicillin
Anl	L-Azidonorleucine
approx.	approximately
APS	Ammonium persulfate
Arg	L-Arginine
Asn	L-Asparagine
Asp	L-Aspartate or aspartic acid
ATP	Adenosine triphosphate
AUG	Starting codon

B

b*1M	barstar mutant P28A/C41A/C83A containing one AUG codon
b*2M	barstar mutant P28A/C41A/E47M/C83A with two AUG codons
BLA	Basal Lipase Activity
BME	β-Mercaptoethanol

C

CAI	Codon Adaptation Index
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1propanesulfonate
Cpa	L-Cyclopropylalanine
CD	Circular Dichroism
CV	Column Volume
Cys	L-Cysteine

D

Da	Dalton
Dpg	L-Dihomopropargylglycine
DMSO	Dimethyl sulfoxide
Don	L-6-diazo-5-oxo-norleucine
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol

E

E. coli	*Escherichia coli*
*Ec*MetRS	*E. coli* methionyl-tRNA synthetase
EDTA	Ethylenediaminetetraacetic acid
εM	Molar extinction coefficient
ESI-MS	Electrospray ionization mass spectrometry
Eth	L-Ethionine
EtOH	Ethanol

F

fMet	formylmethionine

G

GdnCl	Guanidine hydrochloride
Gln	L-Glutamine
glnS	glutamine-tRNA synthetase promoter

glnS'	glutamine-tRNA synthetase mutated promoter
Glu	L-glutamate or glutamic acid
Gly	L-Glycine

H

h	hour
His	L-Histydine
His·Tag	6x histydine tag
Hpg	L-Homopropargylglycine

I

I	Induced cells
IF	Insoluble Fraction
IEX	Ion Exchange Chromatography
IPTG	Isopropyl-β-D-thiogalactopyranosid
Ile	L-Isoleucine

K

Kan	Kanamycin
kDa	kilo Dalton = 1000 Daltons

L

LB medium	Luria Bertani medium
L	Liter
Leu	L-Leucine
Lys	L-Lysine

M

M	Molar
Met	L-Methionine
min	minute(s)

*Mj*MetRS	*Methanocaldococcus jannaschii* methionyl-tRNA synthetase
Mox	L-Methoxinine
MS	Mass Spectrometry
mRNA	messenger ribonucleic acid
MW	Molecular Weight

N

Nhm	N-hydroxy-methionine
NI	None-Induced cells
Ni-NTA	Nickel nitrilo-acetic acid
Nle	L-Norleucine (L-6-aminohexanoic acid)
NMM	New Minimal Medium
*Np*MetRS	*Natronomonas pharaonis* methionyl-tRNA synthetase

O

OD_{600}	Optical density at 600 nm
Omd	L-aspartic acid β-methyl ester
Ome	L-gutamic acid γ-methyl ester
ON	Overnight

P

*Pa*MetRS	*Pyrobaculum aerophilum* methionyl-tRNA synthetase
*Pab*MetRS	*Pyrococcus abyssi* methionyl-tRNA synthetase
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
Pefabloc	4-(2-aminoethyl)-benzenesulfonyl fluoride hydrochloride
pH_{opt}	pH optimum
Phe	L-Phenylalanine
PMSF	Phenylmethylsulfonyl fluoride

*p*NPP	*p*-nitrophenyl palmitate
PPi	Pyrophosphate
Pro	Proline
pTEc	plasmids for Tandem expression in *E. coli*
PVA	Polyvinylalcohol
Pyr	Pyrrolysine

R

rho T	rho T terminator
RNA	Ribonucleic acid
rpm	rotations per minute
RP-HPLC	Reversed phase high pressure liquid chromatography
RT	Room temperature

S

*Sa*MetRS	*Sulfolobus acidocaldarius* methionyl-tRNA synthetase
Sa-tRNA^{Met1}	S. *acidocaldarius* elongator transfer RNA for methionine No. 1
S. cerevisiae	*Saccharomyces cerevisiae*
*Sc*MetRS	*Saccharomyces cerevisiae* methionyl-tRNA synthetase
Sel	L-Selenomethionine
Ser	L-Serine
SCS	Stop Codon Suppression
SDS	Sodium Dodecyl Sulphate
SF	Soluble Fraction
SPI	Supplementation-based Incorporation
Strep·Tag II	Streptavidine tag II

T

TEMED	*N,N,N',N'*-Tetramethylethyldiamine
*Ta*MetRS	*Thermoplasma acidophylum* methionyl-tRNA synthetase
TFA	Trifluoroacetic acid

Tel	L-Telluromethionine
Tfm	5,5,5-trifluoromethionine
Tfn	6,6,6,-trifluoronorleucine
Thr	L-Threonine
T_{opt}	Temperature optimum
Tris	Tris(hydroxymethyl)aminomethane
Triton	Polyethylene glycol p-(1,1,3,3-tetramethylbutyl)-phenyl ether
tRNA	transfer ribonucleic acid
$tRNA^{fMet}$	Bacterial initiator tRNA for methionine
$tRNA^{Met}$	Bacterial/Eukaryotic/Archaeal elongator tRNA for methionine
$tRNAi^{Met}$	Eukaryotic/Archaeal initiator tRNA for methionine
Trp	L-Tryptophan
TTL	*Thermoanaerobacter thermohydrosulfuricus* Lipase
TTL[Aha]	Methionine residues replaced by azidohomoalanine in TTL
TTL[Nle]	Methionine residues replaced by norleucine in TTL
Tween 20	Polyoxyethylene (20) sorbitan monolaurate
Tween 80	Polyoxyethylen (20) sorbitan monooleat
Tyr	L-Tyrosine

U

UAA	Ochre stop codon
UAG	Amber stop codon
UGA	Opal stop codon

V

Val	L-Valine
v/v	volume per volume

W

WT	Wild Type
w/v	weight per volume

Summary

In this work, the lipase from the thermophilic anaerobic bacterium *Thermoanaerobacter thermohydrosulfuricus* expressed in *E. coli* was used as model protein to engineer the genetic code. Two methionine analogs with opposite polarities, the strictly hydrophobic norleucine and the highly hydrophilic azidohomoalanine, were used to globally replace all methionine residues in lipase. The substitution yielded lipase congeners with remarkable differences in enzyme activity when compared to the parent protein. The novel and emergent features of the lipase congeners are reflected in changes in optimal temperature and pH, thermostability, substrate access and specificity, as well as resistance toward additives of diverse nature.

Additionally, the first efforts toward the development of a method for the position-specific replacements of Met residues in *E. coli* are also presented. To this end, the host methionyl-tRNA synthetase (MetRS) along its cognate initiator $tRNA^{fMet}$ can be used to initiate protein synthesis, while an exogenous MetRS imported with its cognate elongator $tRNA^{Met}$ is left to decode internal positions more efficiently. The fact that *S. acidocaldarius* MetRS is orthogonal in *E. coli,* and that it prefers activating ethionine over azidohomoalanine while *E. coli* MetRS displays an opposite amino acid preference, serves as a starting point to reprogram protein translation *in vivo*. In this context, it is intended that the N-terminus is translated with one Met analog and the internal positions with another one. In the future, this platform could allow engineering the genetic code of important biocatalysts for synthetic biotechnology.

Table of contents

1 Introduction

Biology is changing: During the half-century between the deciphering of the genetic code and the sequencing of the human genome, it became clear that biology is essentially an information science with deoxyribonucleic acid (DNA) as a storage device. Systems biology has latterly emerged fuelled by the generation of enormous amounts of data from large-scale top-down profiling ‘omics’ technologies (genomics, transcriptomics, proteomics, metabolomics, lipidomics, glycomics, interactomics, fluxomics, biomics) and its subsequent analysis using refined computational tools. The volume of these data is indicative of the complexity of biological systems. It is thus necessary to generate tools and methods that will help us to characterize and manipulate any biological molecule as it works in its native habitat. In order to achieve this, the crosstalk between systems biology and the emerging field of synthetic biology will be of primary importance in expanding our understanding of biological phenomena and in finding applications for societal benefit.

1.1 Synthetic biology

Synthetic biologists seek to build well-characterized, nature-inspired complex devices that perform desired tasks. There are two main approaches to engineer artificial biological systems. The best known is the top-down approach where a living system is re-designed for particular purposes. For instance, the genome of *Mycobacterium genitalium* was synthesized and assembled *de novo* through the use of nucleic acid synthesis and recombinant DNA technology.[1] This challenging experiment showed that it is, in principle, feasible to artificially clone and, thus, manipulate a whole genome of an organism. These cells could ultimately be used for numerous applications e.g. efficient production of pharmaceuticals, biofuels, or even extraction of useful genomic information from microorganisms that are difficult to cultivate, given that the majority of microbes are difficult or almost impossible to cultivate. On the other hand, the bottom-up strategy involves the small-scale synthesis of nucleic acids, metabolites, and protocells, which are artificial self-assembling and self-reproducing chemical systems. One example is the concept of 'BioBricks', which are standard DNA parts encoding basic biological functions. There is hope that it will be possible to program living organisms in the same way that a computer scientist can program a computer.[2]

Similarly, living systems have already been designed to stimulate pattern formation[3], disperse biofilms[4], produce drugs[5], or target cancer cells[6]. Furthermore, the subfield of synthetic metabolism, defined as engineering biology at the protein and pathway scales[7], will also play a significant role in the consolidation of synthetic biology. This subfield focuses on the production of valuable and useful compounds for which there are no known natural biochemical synthesis pathways. The design and assembly of synthetic metabolic pathways has much potential for optimizing the current enzyme-catalyzed industrial processes used in the production of fine chemicals. The integration between synthetic pathways and proteins will be of utmost importance for synthetic biology.[8]

The generation of synthetic proteins can be done either by the expansion or the engineering of the genetic code. The expansion of the genetic code refers to methods where DNA mutagenesis is required to consider some termination triplets (e.g. amber) or quadruplets as blank codons for expansion of protein functions. Synthetic amino acids are incorporated into single recombinant proteins by means of a nonsense or frameshift suppression using genetically engineered components of the translational machinery including aminoacyl-tRNA synthetases, transfer RNAs, or ribosomes.

On the other hand, the genetic code engineering relies on the substrate tolerance of cellular uptake and endogenous translation systems, which allows synthetic amino acids to be successfully incorporated into proteins. This methodology allows the residue-specific replacement of a particular amino acid at all positions via sense codon reassignment in a target protein without the need for DNA mutagenesis. Proteins produced in this manner are typically referred to congeneric (from old Latin), meaning ‘born together’ or ‘belonging to the same race or kind’ since they are encoded by the same gene but contain only a small fraction of amino acids exchanged with analogs in a residue-specific manner.

The final goal of these bottom-up approaches in the framework of synthetic biology is to extend the synthetic capacities of the protein translation machinery. In this way, the coding capacities of the genetic code will also be enhanced either by including reassignment of existing coding units or the introduction of novel ones. In the long-term, these efforts will help in controlling cellular processes and synthetic cellular systems.

To summarize, the long term goal for synthetic biology is that the bottom-up and top-down approaches meet and contribute to produce ‘encapsulated complex systems’ that would be the starting point for a wide range of genetic programming applications (**Figure 1**).

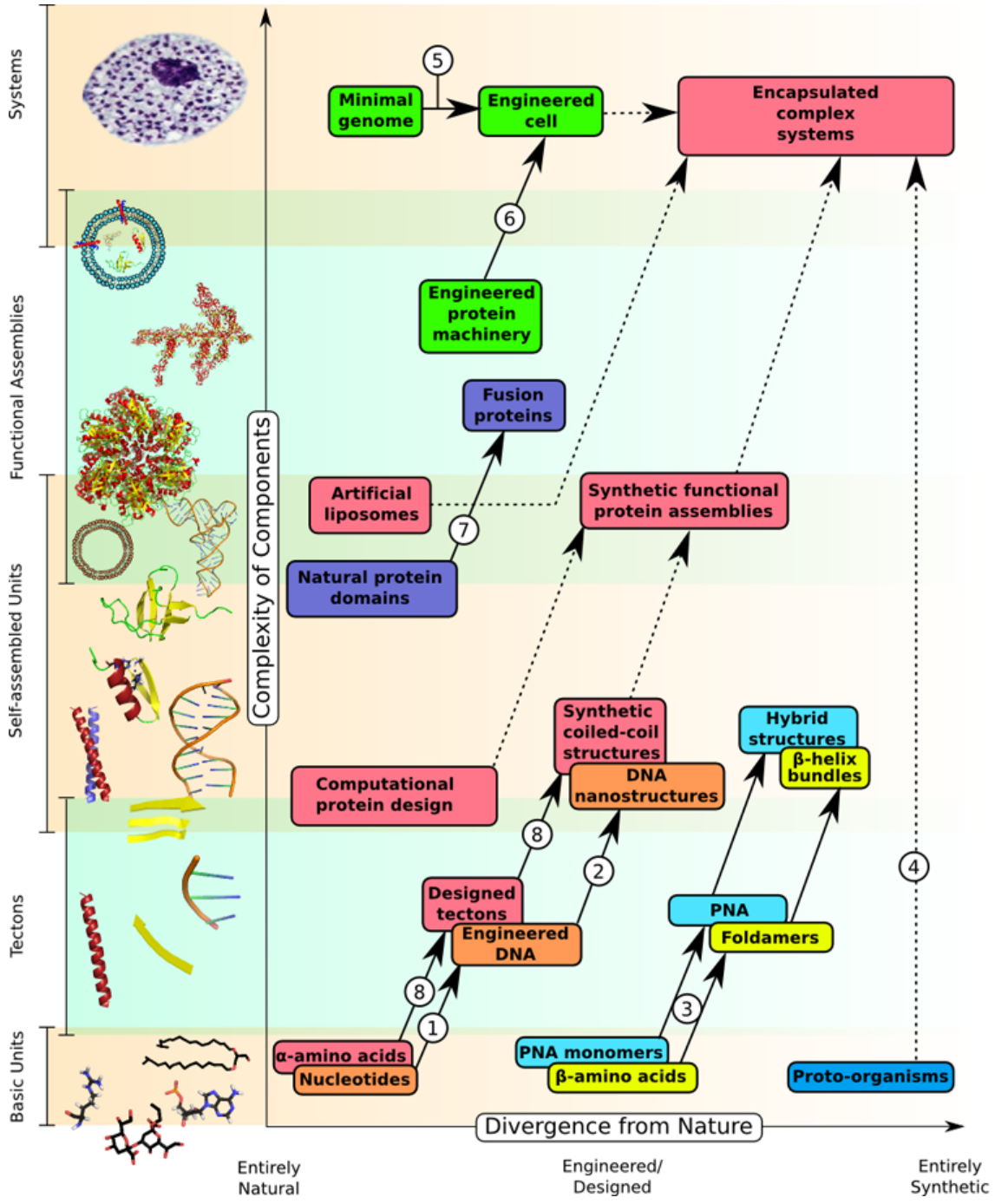


Figure 1. Synthetic biology. Nucleic acids, amino acids or protocells (basic units) of diverse chemistry (entire/partial natural/synthetic) can be linked to form polymers (tectons) whose self-assembly into defined structures will allow functional modules to design a synthetic system that performs a specific task. The efforts in the bottom-up and top-down research approaches (indicated by numbers) to engineer components display different levels of complexity (copyright-free image taken from the Synthetic Components Network at the University of Bristol, UK). PNA = Peptide Nucleic Acid

1.1.1 Genetic code engineering

The genetic code is comprised of 61 codons encoding twenty canonical amino acids. As a rule, each codon is translated unambiguously. Of these 20 amino acids, methionine (Met) and tryptophan (Trp) are encoded by a single codon each, while the rest is encoded by at least two degenerate codons. Theoretically, each of the 61 coding triplets should be decoded by specific transfer RNAs (tRNA), the adapter molecule that translates the nucleotide sequence into the amino acid sequence via anticodon-codon recognition. In practice, this number is usually smaller e.g. there are 46 functional tRNAs in *Escherichia coli*[9] because different tRNAs are capable of reading degenerate (i.e. synonymous) codons that encode only one amino acid type. These tRNAs are termed isoacceptors, and are strictly recognized by their cognate aminoacyl-tRNA synthetase (AARS).

There are generally twenty different AARSs, each one specific for the twenty canonical amino acids. The AARSs recognize their cognate isoacceptor tRNA(s) through several identity elements.[10] For most tRNAs, the determinants are located at the two distal extremities: the amino acid acceptor stem and the anticodon loop. The process of protein synthesis is highly regulated because it involves the precise interplay of a great number of molecules. First, each of the 20 canonical amino acids has to be activated in an ATP-dependent manner by its cognate AARS. Then, the enzyme-bound aminoacyladenylate (AA-AMP) intermediate binds its cognate tRNA(s), and the amino acid is covalently linked to the terminal 2' or 3' hydroxy group of the adenosine in its respective tRNA. Consequently, the charged tRNAs are ready to participate in protein synthesis.

In general, proteins are composed of 'diverse' building blocks, but from the point of view of synthetic biology, the chemical functionalities introduced by the canonical aliphatic, aromatic, basic, acidic, hydroxyl-, and sulfur-containing amino acids are somewhat limited. In fact, to achieve full protein functionality, Nature expands the genetic code by introducing posttranslational modifications (**Figure 2**).[11]

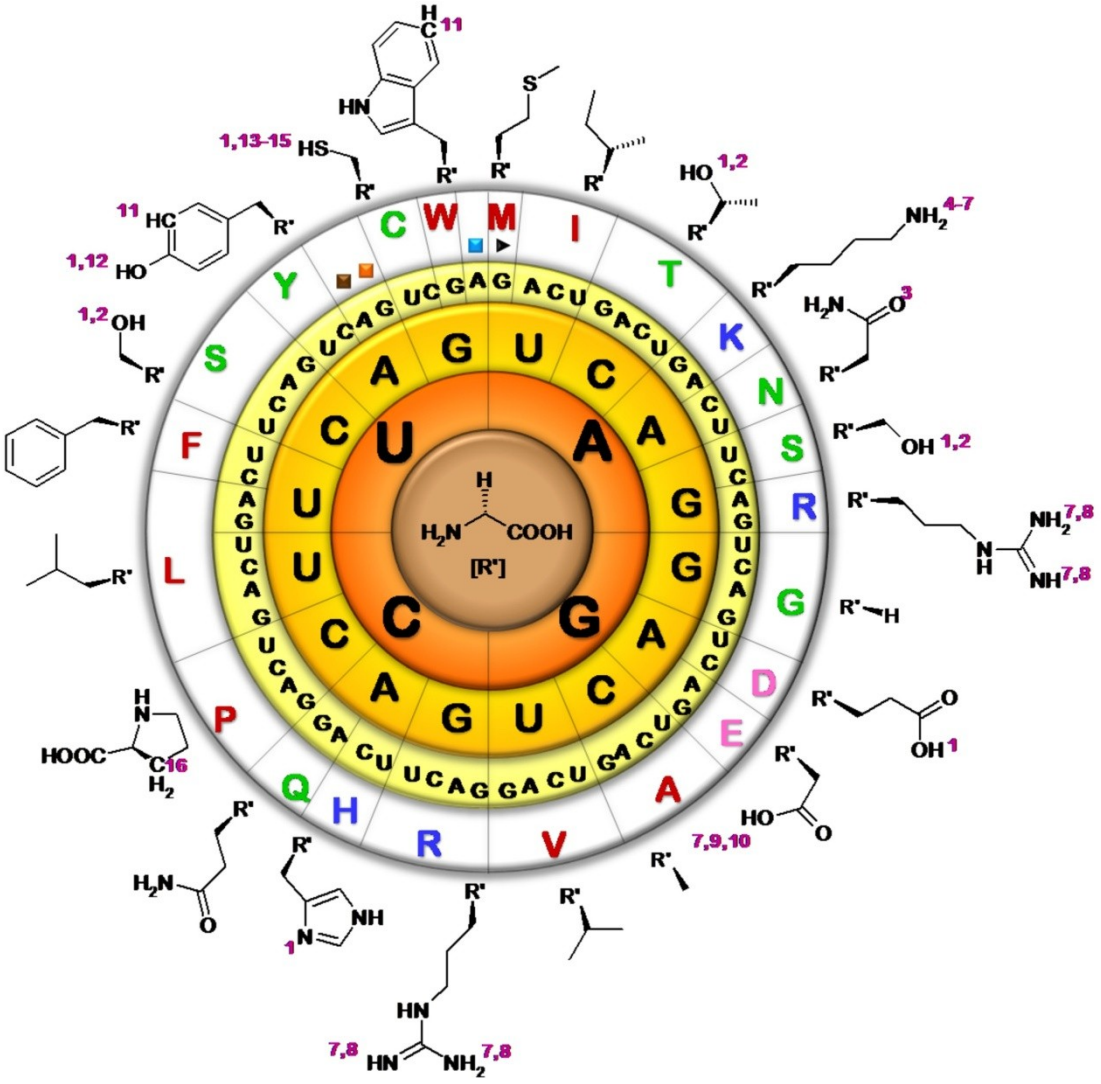


Figure 2. Radial presentation of the universal genetic code in RNA format (previous page). The twenty canonical amino acids are encoded by 61 degenerate codons. M, methionine; I, isoleucine; T, threonine; K, lysine; N, asparagine; S, serine; R, arginine; G, glycine; D, aspartate or aspartic acid; E, glutamate or glutamic acid; A, alanine; V, valine; H, histidine; Q, glutamine; P, proline; L, leucine; F, phenylalanine; Y, tyrosine; C, cysteine; and W, tryptophan. Amino acids are arranged with reference to their physicochemical properties; polar residues are shown in green; nonpolar in red; basic in blue and acidic in pink. Translation initiation and termination are represented with the symbols ► (starting AUG codon) and ■ [stop codon UAA (ochre), UGA (opal), and UAG (amber)], respectively. Post-translational modifications are also indicated with numbers in violet near the specific residue to be modified: 1) Phosphorylation; 2) O-glycosylation; 3) N-glycosylation; 4) Acetylation; 5) Ubiquitation; 6) Biotinylation; 7) Methylation; 8) N-ADP-ribosylation; 9) Polyglycination; 10) Polyglutamylation; 11) Nitration; 12) Sulfation; 13) S-nitrosylation; 14) S-prenylation; 15) S-Acylation; and 16) C-hydroxylation.

Although the AARSs are crucial in the recognition of their cognate amino acids, they are often incapable of distinguishing between similar substrates. This phenomenon is known as substrate tolerance and permits AARSs to recognize and charge noncanonical amino acids onto tRNAs in lieu of the structurally-related canonical counterparts. Thus, it can be said that the fidelity of protein translation basically depends on the tRNA aminoacylation reaction (**Figure 3**).

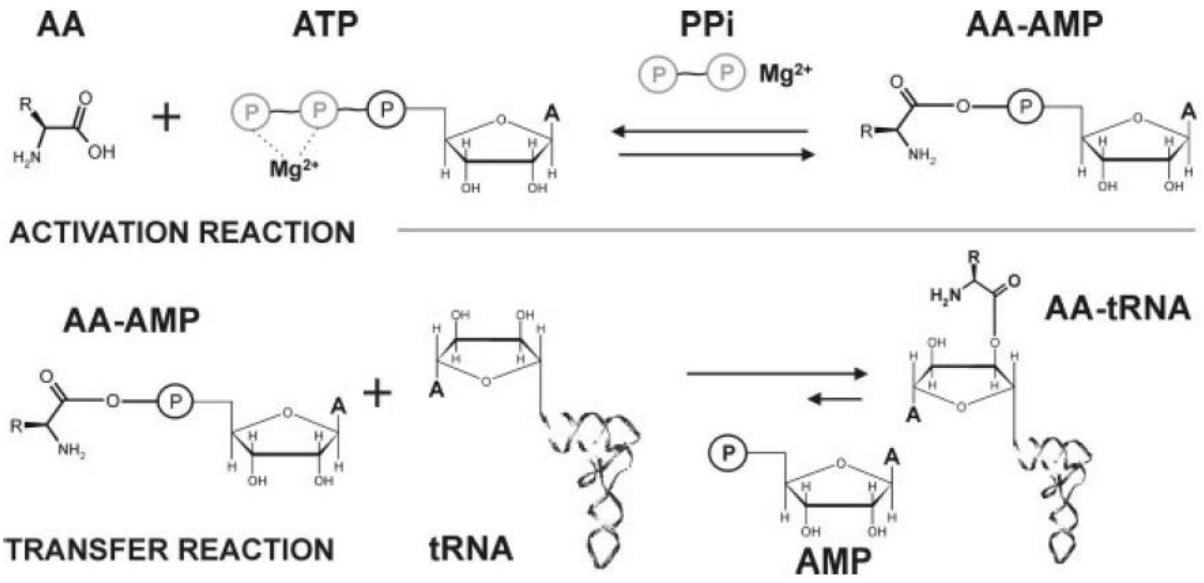


Figure 3. The aminoacylation reaction. In this two-step reaction, the amino acid (AA) is first activated with ATP and Mg^{2+} by the AARS to form an enzyme-bound aminoacyladenylate (AA-AMP) intermediate, accompanied by the release of pyrophosphate (PPi). This complex then binds the cognate tRNA, and the amino acid is covalently linked to the 3'-OH terminal adenosine of tRNA. AMP = adenosine monophosphate (image kindly provided by N. Budisa[12])

Accordingly, the use of noncanonical amino acids, also termed unnatural, noncognate, or nonproteinogenic amino acids provides proteins with novel and even unexpected (emergent) properties due to the inclusion of new side-chains. Consequently, the addition of a vast array of chemical functionalities like halogeno, keto, cyano, azido, nitro, nitroso, and silyl groups, as well as alkenes and alkynes would permit the design of new protein fluorescent probes, photo-switches, redox markers, and also allow the elucidation of protein-protein interactions or enzyme mechanisms, the study of protein folding, or the determination of 3D protein

structures.[13-18] As indicated previously, proteins produced in this way can be called congeners since they originate from the same gene sequence, but they contain only a small fraction of amino acids exchanged with synthetic amino acids.

The engineering of the genetic code permits the production of synthetic proteins through the incorporation of noncanonical amino acids not encoded by the standard genetic code. Supplementation-based Incorporation (SPI) is the predominant methodology for engineering the genetic code. SPI allows the residue-specific incorporation of noncanonical amino acids in the resulting polypeptide sequence by reassigning one of the sense codons. This can be achieved through the use of specific amino acid auxotrophic strains and controlled fermentation conditions.[19] While SPI allows the multiple residue-specific incorporation of noncanonical amino acids in response to sense codons, the stop codon suppression (SCS) method allows the addition of noncanonical amino acids at permissive sites of proteins but in a context-dependent manner. In this way, the incorporation of various useful aromatic analogs or extended aromatic systems is possible via orthogonal pairs e.g. *Methanocaldococcus jannaschii* TyrRS:tRNA^{Tyr} in *E. coli*.[20-22] Similarly, it is possible to incorporate a wide range of pyrrolysine (Pyr) analogs with aliphatic side chains equipped with versatile chemical properties and functionalities into proteins with other pairs such as *Methanosarcina maize* PylRS:tRNA^{Pyl} in response to stop codons in *E. coli*.[23-25]

Both SCS and SPI methods require the efficient cellular uptake of the desired amino acid analog, its efficient activation by the AARS and charging onto the tRNA, ribosome proofreading, and proper folding of the target protein. Nonetheless, the SPI method is more promising than SCS for designing enzymes with economically important applications for several reasons. First, many biological phenomena like enzymatic activity or folding properties are based on the synergistic effects of different amino acids at several positions in the context of folded protein structures. The SCS method is limited in this regard, since it only allows the insertion of only one, or at best, two analogs into the target sequence.[26] Furthermore, the efficiency of synthetic

amino acid incorporation using suppressor-based methodologies is affected by local protein structure, messenger RNA (mRNA) context, and competition with release factors. Third, the design and selection of orthogonal pairs is complex and time-consuming. Finally, the relative low protein yields produced using the SCS method prevents its use in many industrial applications. In contrast, the SPI method is reproducible, efficient, and has the added advantage of requiring a very simple experimental setup.

1.1.2 Incorporation of methionine analogs into proteins *in vivo*

The AUG sense codon has been reassigned to a large repertoire of Met analogs using the SPI method to study their effect in proteomes or target proteins. These experiments have been performed in their entirety in *E. coli* due to its simplicity as a model organism.

The landmark experiment using Met analogs was documented by Lewine and Tarver in 1951.[27] A few years later, Cowie and Cohen performed a quantitative replacement of Met by selenomethionine (Sel; **Figure 4**) in proteins of a Met auxotrophic *E. coli* strain without impairing its cellular viability.[28] This was rather an exception since it was known that many noncanonical amino acids are bacteriostatic or bactericidal. In order to circumvent this problem and to enable labeling of single proteins, the translation capacity was resolved from metabolic toxicity. This concept could be experimentally verified only after advancements of recombinant DNA technology. In other words, by using heterologous gene expression systems, the production of fully-substituted single target proteins can be controlled with high efficiency.[29,30] Three decades later SeMet emerged as an important tool in X-ray crystallography[31,32] and NMR studies[33] of biological macromolecules. Similar applications can be obtained with the heavy metal-containing Met analog telluromethionine (Tel; **Figure 4**).[30,34]

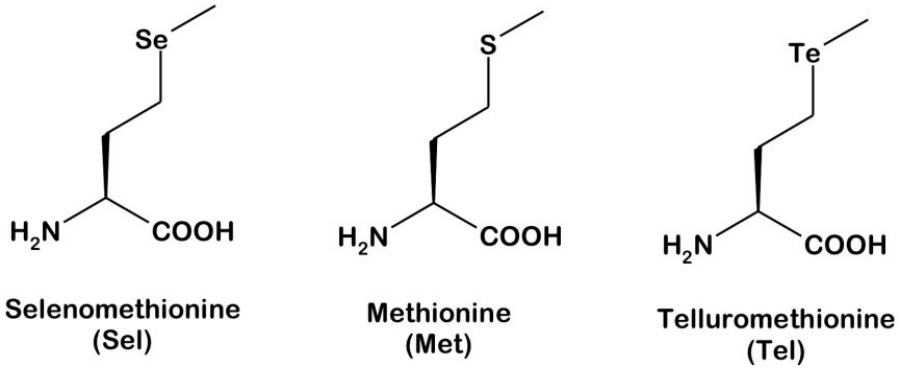


Figure 4. **Methionine analogs useful in X-ray crystallography.** Selenomethionine contains the heavy atom selenium instead of sulfur, whereas telluromethionine has tellurium.

Surprisingly, amino acids containing organic azides and terminal alkynes proved to be Met analogs (**Figure 5**).[35-37] These are currently of particular interest due to the chemical nature of their moieties and their inertness under physiological conditions. These analogs can be used as chemical handles to mimic post-translational modifications using click chemistry.[38-40] The Met analogs azidohomoalanine (Aha) and homopropargylglycine (Hpg), for example, have been used to artificially attach to proteins post-translational modifications like sugars[41] or biotin.[42] These Met analogs are referred to as bioorthogonal chemical reporters, and have been used in a wide-range of applications including cell biology[43], chemical proteomics[44], or selective modification of virus-like particles[45], among others.

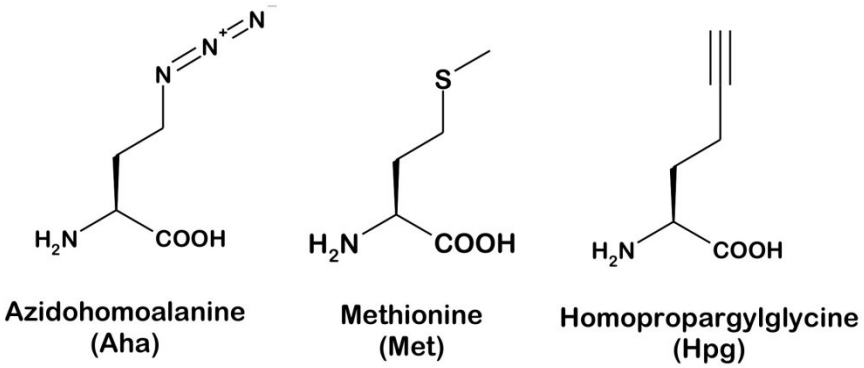


Figure 5. Methionine analogs useful as bioorthogonal reporters. The azide-containing Aha or alkyne-containing Hpg can be used in click chemistry to couple alkyne or azide groups, respectively.

The best approach for unraveling the interactions and relationships responsible for highly specific protein internal architecture and its folding process is by generating non-disruptive isosteric changes in the amino acid side-chains.[46] In the case of Met, these replacements are at the level of single atoms, as in Norleucine (Nle) and methoxinine (Mox). These Met analogs have already been used to design pro- and anti-aggregation congeners of the prion protein as models for investigating oxidative stress in prion-related neurodegenerative diseases.[47] In these substitutions, dramatic differences in physicochemical properties provided by the Met analogs were fully reflected in the related prion protein congeners. Mox and Nle have nearly identical chain lengths, the same number of single bonds and exhibit the same resistance to chemical oxidation; however, their polarities are extremely different. Mox is strictly hydrophilic whereas Nle is strongly hydrophobic (**Figure 6**).

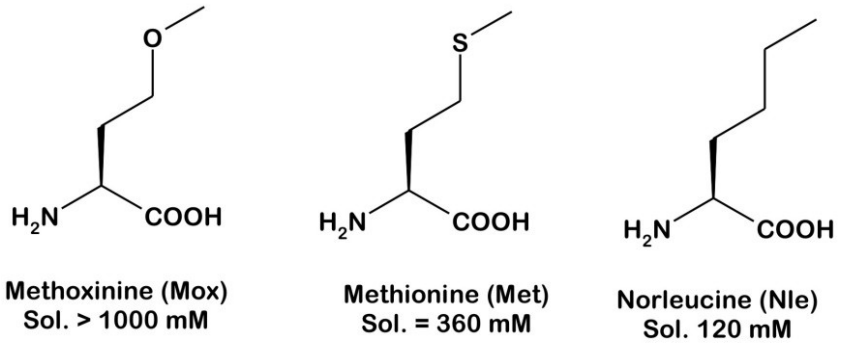


Figure 6. Methionine analogs useful for probing protein hydropathy. Mox is strictly hydrophilic with a water solubility (Sol.) of 1000 mM[47] whereas Nle is neutral and strongly hydrophobic with a water solubility of 120 mM.[48] In comparison, Met solubility in water is 360 mM.[48]

So far, most of the experiments using Met analogs have been for studying effects on protein architecture, folding, or stability. Its effects on enzymatic activity, however, have not been as extensively investigated, even though there are some reports on enzymes whose activities remain at the basal level or were even increased upon incorporation of the analogs. For example, the substitution of Met with Nle or the similar hydrophilic analog ethionine (Eth; **Figure 7**) often does not affect the enzymatic activity of calmodulin.[49] The same holds true when Nle was substituted in nuclease[50] or when trifluoromethionine (Tfm; **Figure 7**) was incorporated into phage lysozyme[51] or

DNA polymerase.[52] The advantage of using fluorinated analogs such as Tfm is that it can be used for the design of teflon-like or 'non-sticky' proteins.[53]

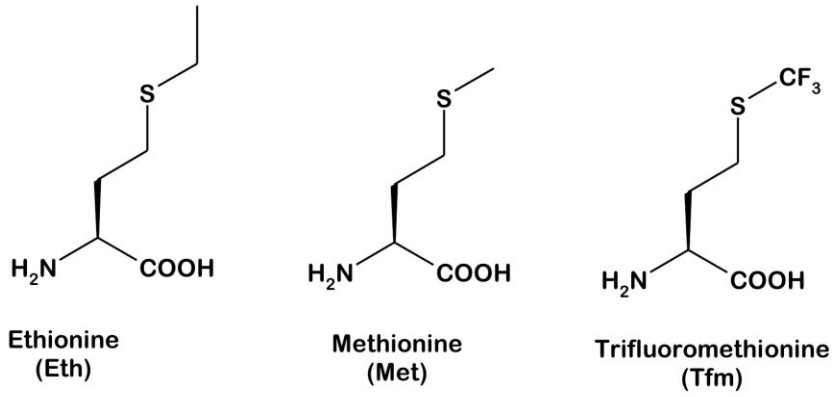


Figure 7. Methionine analogs useful to retain enzymatic activity. Eth and Tfm contain an extra methyl group and three fluorinated atoms, respectively, compared to Met.

There are also cases where the catalytic activity of enzymes is enhanced upon Met-substitution with particular analogs. For instance, the complete substitution of 13 Met residues with Nle in cytochrome P450 resulted in a nearly two-fold increase in peroxygenase activity.[54] A similar Met to Nle substitution in adenylate kinase results in a derivative with structural and catalytic properties similar to the original enzyme; nonetheless, it also exhibits much higher resistance to hydrogen peroxide inactivation under denaturing conditions.[55] These studies suggest that noncanonical amino acids could be used not only to retain, but also to enhance the enzymatic activity of proteins. None of these reports, however, systematically explored the impact of using isosteric analogs to manipulate the enzymatic activity of more important biocatalysts like nucleases, isomerases, proteases, amylases, or lipases.

A thorough survey of literature indicates that there are few studies with an intention to improve the catalytic performance of enzymes by incorporating noncanonical amino acids. For example, Tyr analogs have been incorporated into *E. coli* β galactosidase[56], *Pseudomonas* ketosteroid isomerase[57] and *Candida antarctica* lipase B.[58] There are also some reports where Met analogs have been documented.

The first experiment, for instance, was reported by Yoshida in 1959, where Eth as incorporated into *Bacillus subtillus* amylase[59]; the second by Anfisen and Corley in 1969, where Nle was used in *Staphylococcus aureus* nuclease[50]; the third by Cirino *et. al.* in 2003, where Nle was substituted in peroxygenase-active cytochrome P450[60]; the fourth by Walasek and Honek in 2005 for the use of difluoromethionine in alkaline *Pseudomonas aeruginosa* protease[61]; the fifth by Schoffelen *et. al.* in 2008, where Aha is incorporated into *C. antarctica* lipase B; however, the main purpose in the last study was to investigate the potential of the click chemistry reaction[62]; and finally Holzberger *et. al.* in 2010, where Tfm was used in *Thermus aquaticus* DNA polymerase.[52] In all the aforementioned cases the enzymes remained active, but they were not further characterized. Therefore, understanding deeper the role of Met as well as extensively studying the effects of its replacement by synthetic analogs of such important biocatalysts should be very interesting for both academy and industry.

1.2 The role of methionine in proteins

In all known protein structures, Met comprises only 1.5% of all residues, and these are usually located in positions inaccessible to the bulk solvent, with only 15% of the Met being exposed to the surface.[63] There is clearly a link between the nature of Met and the fact that it is normally found inside the protein core. Met is classified as a nonpolar and modestly hydrophobic amino acid. Like cysteine, it contains sulfur; however, the Met sulfur is a part of the relatively inert thioether moiety. Indeed, the Met side-chain is polarizable and flexible, which is often a crucial feature in biological processes such as substrate recognition or packing at hydrophobic interiors of proteins. Furthermore, Met is also implicated in many cellular processes, including cofactor binding, functioning as a methyl donor in DNA methylation, and protein stabilization through its hydrophobic interactions and hydrogen bonding, as well as nonpolar molecular recognition processes.[64]

1.2.1 Methionine as key residue of protein oxidation

Together with cysteine and tryptophan, Met is among the amino acids most prone to oxidation. The Met thioether moiety can be first oxidized to its sulfoxide form, Met(O), to become hydrophilic. The next oxidation step to sulfone is more drastic and rarely occurs in biological systems (**Figure 8**). Many proteins lose activity upon Met oxidation; in contrast, some proteins are functionally activated upon Met oxidation.[65]

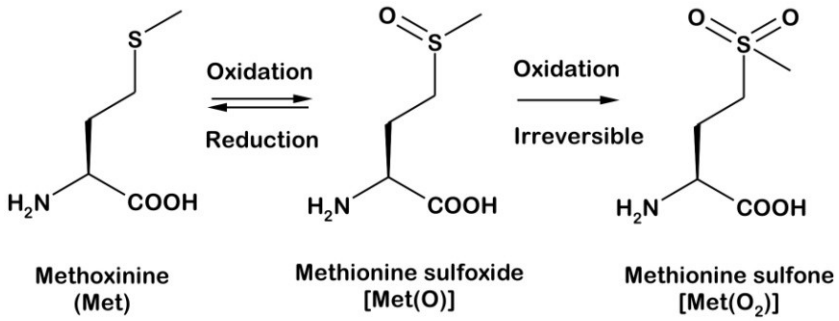


Figure 8. Methionine oxidation. Chemical structures of methionine and its oxidation products methionine sulfoxide and methionine sulfone.

The probability of Met oxidation varies based on its position in the protein structure. For instance, a Met residue near the surface is more easily accessible to oxidation, and would cause the protein to become more hydrophilic. A Met residue buried in a hydrophobic cluster is more protected from oxidation, although some reactive molecular species such as singlet-oxygen diffuse through the protein molecule and could cause oxidation at interior Met residues as well.[66] The oxidation or reduction of Met is therefore an important event that contributes to protein conformational changes and overall hydropathy. The physiological importance of Met oxidation is also gaining much attention because the presence of Met sulfoxide in particular proteins correlates with some pathological conditions like emphysema, arthritis, and cataracts.[67]

1.2.2 The Janus face of protein synthesis

Janus was the Roman god of beginnings and ends. Since he is represented with two

faces looking opposite ways, the term Janus-faced is used to describe someone who is duplicitous. In relation to Met, it is encoded by the AUG codon, but participates in translation with two classes of adaptors, namely the initiator transfer RNA ($tRNA^{fMet}$ in prokaryotes and eukaryotic organelles, or $tRNAi^{Met}$ in eukaryotes and in archaea), which is exclusively used for initiation of protein synthesis, and the elongator transfer RNA ($tRNA^{Met}$), which inserts Met in response to internal AUG codons in a protein sequence. Due to its double function, Met can be considered as the Janus-faced amino acid of protein translation.

In *E. coli*, Met is activated by the methionyl-tRNA synthetase (*Ec*MetRS), whose functional form is a homodimer composed of two 76 kDa monomers. A 64 kDa monomeric *Ec*MetRS fragment generated by a C-terminal trypsin cleavage was used to determine the 3D structure of the free enzyme (**Figure 9**).[68-70] The crystal structures of *Ec*MetRS confirm that the enzyme is organized into an N-terminal domain containing the active site and a C-terminal domain responsible for recognition of its cognate tRNAs. As a metalloenzyme, in addition, *Ec*MetRS contains one tightly bound zinc ion for the correct folding of the enzyme.

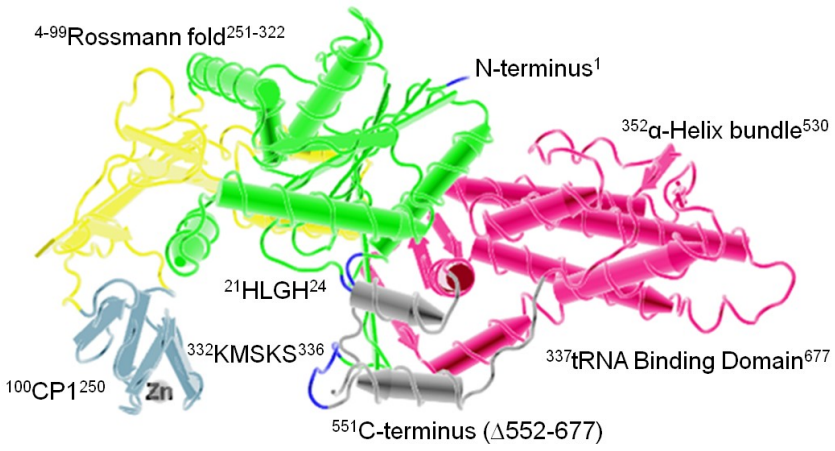


Figure 9. Ribbon diagram of the 3D crystal structure of monomeric *Ec*MetRS. The *Ec*MetRS (residues 1–551 with truncated C-terminus [Δ552-677]; PDB: 1QQT) shows three distinguishable domains: the Rossmann fold (green and yellow), the connective polypeptide-Zn-finger domain (light blue), and the α-helix bundle-anticodon binding domain (red and gray). The Class I Rossmann fold signature sequences HLGH and KMSKS responsible for ATP and Met binding are shown in blue.

*Ec*MetRS methionylates the 2'-hydroxyl group of the terminal adenosine in both the initiator tRNA^{fMet} and elongator tRNA^{Met}, which have as common identity elements the A^{73} base, the base pairs G^{2}C^{71}, C^{3}G^{70} and the CAU anticodon (**Figure 10**).

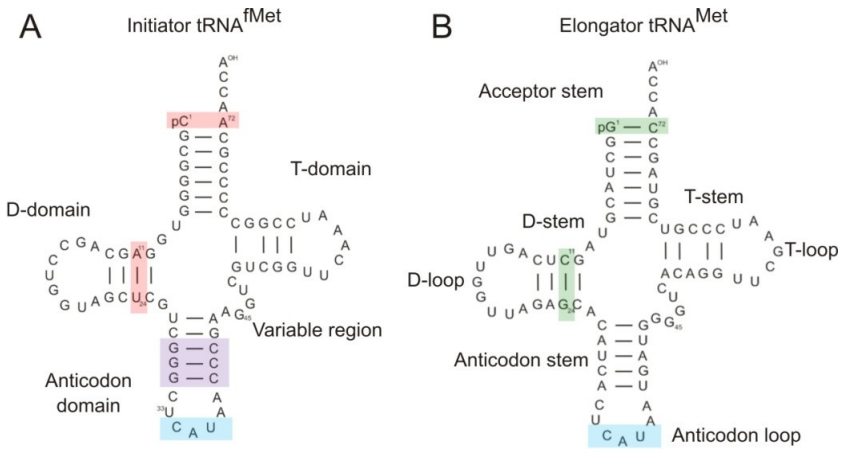


Figure 10. *E. coli* tRNAs^{Met}. Secondary structure of A) initiator tRNA^{fMet} and B) elongator tRNA^{Met} from *E. coli*.[71,72] A tRNA molecule of canonical cloverleaf structure consists of four domains: The dihydrouridine (D), thymidine (T), anticodon and acceptor arm. Each domain has a stem and loop region. A variable arm occurs between the anticodon and T domains. The required identity elements for formylation (red), initiation (purple), elongation (green) and aminoacylation (blue) are shown.

Interestingly, there are approximately five times more internal AUG codons than initiator codons, but twice as much initiator tRNA^{fMet} than elongator tRNA^{Met} in *E. coli*.[73] This high intracellular concentration of tRNA^{fMet}, along with the participation of three initiation factors (IF1, IF2, and IF3), GTP, and the mRNA at the ribosome reflects the complexity of translation initiation.[74] In this event, the crucial step is when IF1 directs the formylmethionyl-tRNA^{fMet} (fMet-tRNA^{fMet}) to the peptidyl-site (P-site) of the small ribosomal subunit so that it could interact with the AUG initiation codon on the mRNA and form a stable complex with IF2 and IF3. Intriguingly, it is believed that Met starts translation because it is the most expensive amino acid to synthesize.[75] Additionally, 60% of the N-terminal Met residues in *E. coli* are subject to removal to provide proteins with further biological activity.[76] In this way, the cell would recycle a pool of Met and save energetic costs by coupling translation either to the intermediary metabolism or to its chemical environment.

In translation elongation, aminoacylated elongator tRNAs (aa-tRNAs), GTP, and three elongation factors (EF1A or EF-Tu, EF1B or EF-Ts, and EF2 or EF-G) are involved. The EFs transfer the aa-tRNA that corresponds to the next codon on the mRNA to the aminoacyl-site (A-site) of the ribosome. A peptide bond is then formed between the first and second amino acids, and the deacylated tRNA moves from the P-site to the exit-site (E-site), while the A-site-aa-tRNA translocates into the P-site with aid of the EFs. This cycle is repeated with the ribosome sliding along the mRNA as it is decoded into a polypeptide chain, which is synthesized at a rate of approximately 12 amino acids per second.[77] If an internal AUG is found, a Met will be incorporated by the elongator $tRNA^{Met}$ in the growing polypeptide chain. Importantly, the initiator $tRNA^{fMet}$ and elongator $tRNA^{Met}$ isoacceptors have remarkable differences in their primary structure (**Table 1)**, which serves as the basis for their accurate recognition during protein initiation or elongation.

Table 1. Identity elements that differentiate the *E. coli* initiator $tRNA^{fMet}$ and elongator $tRNA^{Met}$.

Features	$tRNA^{fMet}$	$tRNA^{Met}$
Initial ribosomal binding site	P-site	A-site
Mismatch 1-72 in the acceptor stem	+	-
Purine11:Pyrimidine24 base pair in the dihydrouridine stem	+	-
3 consecutive paired guanosines and cytosines in the anticodon stem	+	-
Interaction of the anticodon loop (CAU) and stem $(G\text{-}C)_3$ with the IF3	+	-
Recognition of fMet by IF2	+	-
Recognition by EF1A (EF-Tu)	-	+

+ = presence; - = absence

Figure 11 shows the scenario when the AUG triplet is decoded both at initiation and elongation of protein synthesis. Finally, translation is terminated when the ribosome encounters a stop codon on the mRNA and the protein is released.

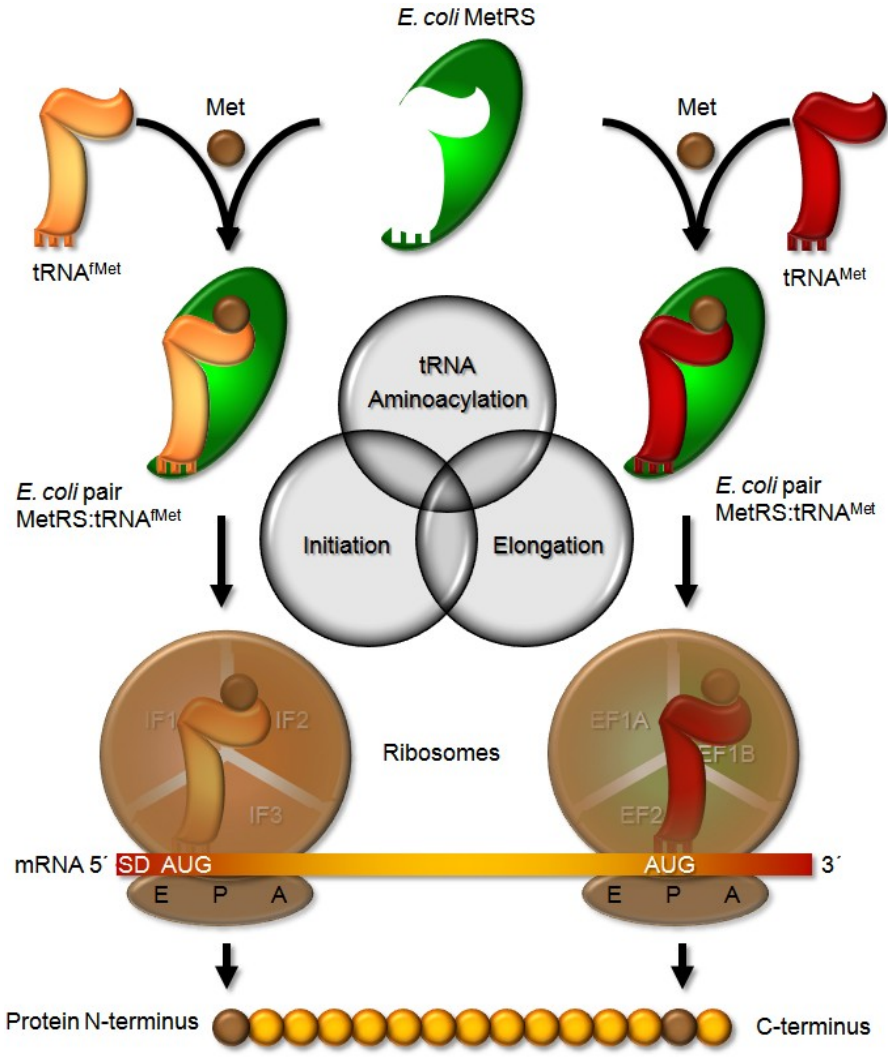


Figure 11. The Janus face of protein synthesis. In *E. coli,* the MetRS (green) normally activates methionine (brown) and charges it onto both the initiator $tRNA^{fMet}$ (orange) and elongator $tRNA^{Met}$ (dark red). After tRNA aminoacylation and translation, the AUG sense codons present in a protein sequence will have an unambiguous meaning: Methionine. SD = Shine-Dalgarno sequence

1.3 Industrial biotechnology

Industrial or white biotechnology uses enzymes or organisms to sustainably produce chemicals, materials, and fuels from renewable sources. Driven by the global market economy, this field promises the optimization of processes at lower operating costs and capital expenditures. The achievements and promises of enzyme engineering, metabolic engineering, synthetic biology, and systems biology (including 'omics' and *in silico* approaches), are expected to speed up the development of white biotechnology as well. The engineering of strains or enzymes in combination with different procedures for the efficient production of biodiesel, for instance, could enable a paradigm shift from fossil fuel-based to bio-based production of value-added chemicals.[78] Last but not least, industrial biotechnology processes are nowadays very strong associated with reduced energy consumption, waste generation, and greenhouse gas emissions.

1.3.1 The search for the ideal biocatalyst

One of the most important tools for industrial biotechnology is protein engineering because enzymes isolated directly from nature rarely exhibit the ideal combination of traits and activities required for industrial use. Biocatalysts must be optimized to function in nonnative environments, such as extreme process conditions including exposure to solvents or additives over prolonged periods of time. In the early stages, developing suitable biocatalysts was generally limited to enzymes found in nature.[79]

With the advent of recombinant DNA technology, enzyme optimization via rational and directed evolutionary approaches became the dominant route for the generation of sequence diversity.[80] However, the task of identifying beneficial diversity is nontrivial and mutations that span the entire structure of the enzyme usually cause detrimental effects. Directed evolution or random mutagenesis, for example, is like looking for a needle in a haystack, since billions of variants must be generated at the

DNA level, then functionally screened to bring a desired property.[81] Moreover, complete protein sequences cannot be easily permutated because the generated libraries are limited in size. Finally, rational approaches often require structural information, which is not always available.

To fully realize the potential of customized enzymes in industrial applications, it is imperative to tailor catalyst properties optimal for both a given reaction and in the context of an industrial process.[82] To this end, another alternative approach in finding the ideal biocatalyst could be by exploring the emergent effects generated by global replacements using the SPI-based residue-specific incorporation of noncanonical amino acids. Indeed, this method successfully combines the selectivity of *in vivo* co-translational sequence modifications with classical chemical approaches. The classical chemical approaches themselves are limited since modifications directed to the amino acid side-chains of biocatalysts are nonspecific and thus, it can result in heterogeneous mixtures of enzyme products.[83]

1.3.2 The most versatile biocatalyst: Lipase

Lipases (triacylglycerol acylhydrolase, EC 3.1.1.3) posses the most common natural protein fold, the canonical α/β hydrolase fold, composed of a β sheet (and not a barrel) of eight beta-sheets connected by α-helices.[84] The canonical α/β hydrolase fold is characterized by the presence of a catalytic triad, formed of a nucleophile (Ser, Cys, or Asp), a strictly conserved His, and an acidic residue Asp or Glu in the case of lipases (**Figure 12A**). To promote catalysis, structural rearrangements around the active site are necessary. This process is characterized by the movement or 'opening' of a helical lid domain that covers the active site cleft of the enzyme. This flexible lid is attached to the enzyme core of many lipases and it is well conserved among organisms (**Figure 12B**). Hence, lipases have both an open and a closed conformation. However, in the aqueous phase, the equilibrium is shifted toward the inactive, closed form.

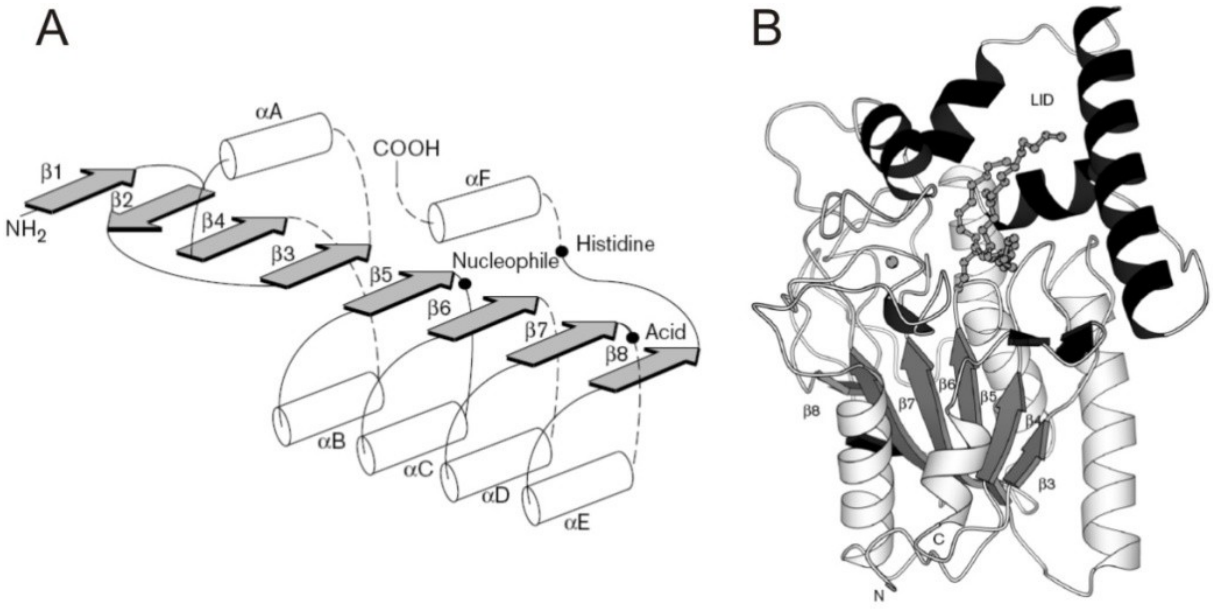


Figure 12. Structural features of lipases. A) Secondary structure diagram of the canonical α/β hydrolase fold: α-helices and β-strands are represented by white cylinders and gray arrows, respectively. The catalytic triad is indicated by black dots. The dashed lines indicate the location of possible insertions. B) Open conformation of *P. aeruginosa* lipase: α-helices, β-strands, and coils are represented by helical ribbons, arrows and ropes, respectively. α-helices and β-strands part of the hydrolase fold are shown in white and dark gray, respectively. Other secondary structure elements are in black. A phosphonate inhibitor covalently bound to the nucleophile and the calcium ion is represented in ball-and stick format. (Images taken from Marco Nardini[85])

Upon interfacial activation, which is defined as a sharp increase in lipase activity when acting at the lipid-water interface of micellar or emulsified substrates[86], the lid moves away, fully exposing the active site and activating the enzyme. Some lipases are likewise subject to an unusual feature called thermal activation; this refers to the phenomenon where enzymes are activated above 40 °C, as it is the case is with lipases from *Pseudomonas aeruginosa*[87-89], *P. glumae*[90], or *Candida Antarctica*[91] all of which nevertheless have an amphiphilic lid covering the active site.

Lipases hydrolyze triacylglycerols into glycerol and fatty acids, or catalyse the reverse reaction under low water conditions (**Figure 13**).

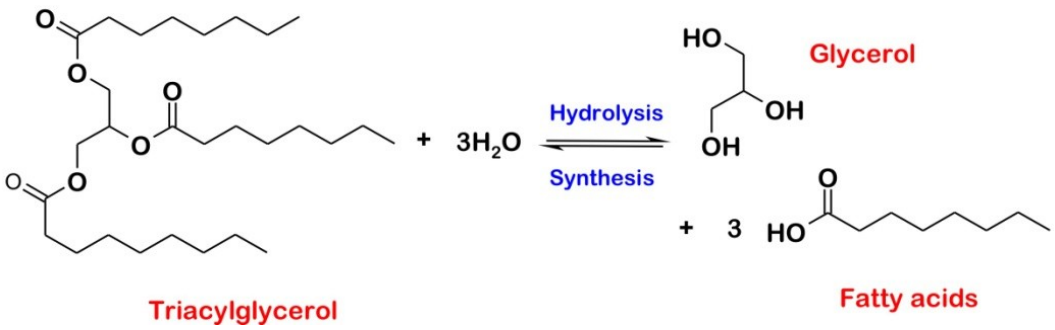


Figure 13. Lipase reactions. Lipases catalyze the hydrolysis or synthesis of triacylglycerols.

Lipolysis occurs exclusively at the lipid-water interface where the concentration of molecules directly determines the catalysis rate. However, the kinetics of a lipase reaction does not follow the classical Michaelis-Menten model, since it is valid only for soluble enzymes and substrates. Instead, when the solubility limit of the substrate is exceeded, there is a sharp increase in lipase activity as the substrate forms an emulsion. Esterase activity, in contrast, is a function of substrate concentration as described by Michaelis-Menten kinetics, with the maximal reaction rate being reached long before the solution becomes substrate-saturated.[89] Accordingly, a model for describing lipase kinetics has been proposed (**Figure 14**).[92]

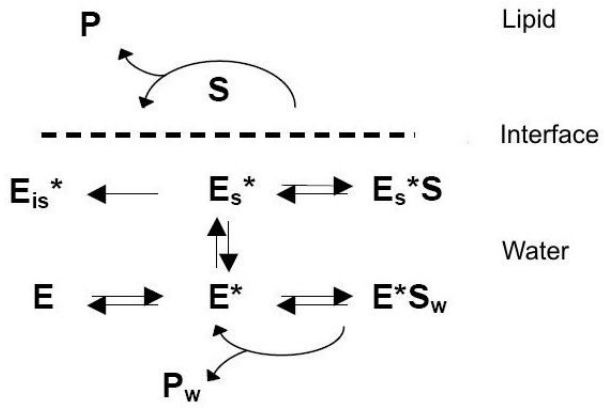


Figure 14. Model of lipase kinetics. The model is an extension of the one proposed by Verger.[93] In aqueous solution, upon activation, the inactive enzyme (E) becomes active (E*) and can either bind to the water-soluble substrate (S_W) and release the product in the water phase (P_W) or be absorbed at the interface (E_S*). Once absorbed, the enzyme can either lose its activity (E_{IS}*) or catalyze the conversion of the water-insoluble substrate (S) to its product (P) (image modified from M. Royter[94]).

The first step comprises the physical binding of the lipase, a step which may include the activation of the enzyme (i.e. lid opening). In the second step, the complex between the lipase and the substrate is formed, giving rise to the product, concurrent with the regeneration of the 'adsorbed' enzyme. **Figure 15** details the mechanism of the lipase reaction.[95]

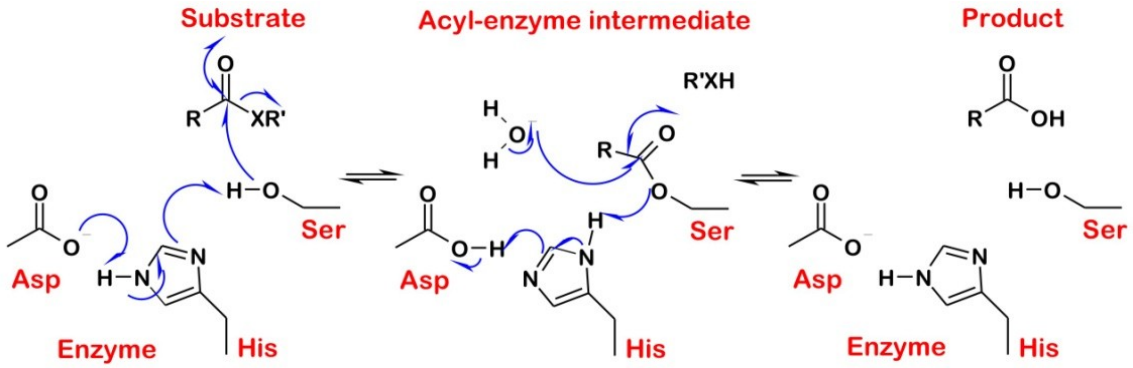


Figure 15. Mechanism of lipase catalysis. Left: Activation of serine and nucleophilic attack of the serine to the carbonyl group of the potential substrate to form an acyl-enzyme intermediate (center). In this step the charges are neutralizing, resulting in the formation of the oxyanion hole. Right: Release of the product and regeneration of the catalytic residues.

First, a serine is activated by proton transfer of the hydroxy proton to aspartate via the mediator histidine residue. As a nucleophile, the hydroxy group of serine attacks the carbonyl group of the potential substrate forming an acyl-enzyme intermediate (**Figure 15**). When a proton from the hydroxy group of the serine is accepted by the nitrogen of the histidine, its positive charge is neutralized with the negative charge of the aspartate, thereby forming the oxyanion hole. The oxyanion hole is important because it helps to stabilize the intermediate state, where the carbonyl oxygen of the substrate bears a partial negative charge. The deacylation step that follows is controlled by the electronegativity of the molecules populating the interface. The mechanism is repeated here, but using water or monoglyceride as a nucleophile instead. These then attack the acylated enzyme to form the tetrahedral intermediate. The intermediate collapses again, releasing the product and regenerating the catalytic residues which can be used in the next catalysis step.

Lipases are widely used to degrade fats and oils in food, flavor, leather, textile, cosmetic, fragrances, and detergent industry. The activity and stability of lipases can be influenced by the addition of surfactants and detergents when added to the reaction medium. However, when it was shown for the first time that they were active in almost 100% of organic solutions[96], lipases began developing as ideal tools for organic chemists.

Lipases can esterify an alcohol in the presence of an acid or transesterify an ester with a second alcohol (alcoholysis), an acid (acidolysis) or another ester. In this way, lipases are used for specific regioselective reactions in organic solvents to produce enantiomers.[97] The hydrolytic/synthetic versatility of lipases makes them the most attractive and widely-used biocatalysts. Commercially, the lipase sector is a million dollar business for which new applications in the areas of wastes management, biodegradation of oil, aliment development, specifically, the reduction of fats or the increase of flavor and fragrance; biofuels, biodegradable polymers, and pesticides production, are being sought for. Of course, there is a strong requirement for identifying new enzymes or optimizing existing ones, pushed forward by a rapidly demand for fine chemicals.[98] Finally, the pharmaceutical field is another important application area, with anticancer, antiviral, antihypertensive, anticholesterol, anti-Alzheimer, and anti-inflammatory drugs being produced in lipase-based procedures.[99]

2 Goal

The basic goal of this thesis is to explore the perspectives of genetic code engineering in applications involving biocatalysts for industrial biotechnology. Here, the lipase from the thermophilic anaerobic bacterium *Thermoanaerobacter thermohydrosulfuricus* (TTL) expressed in *E. coli* was used as a model biocatalyst. The high number of 11 Met residues in TTL indicates that these could play an important role in its catalytic performance. Therefore, the Met side-chain conformational preferences were systematically varied by incorporating the isosteric analog norleucine and the surrogate azidohomoalanine globally. These have similar chain lengths but opposite polarities: Nle is strongly hydrophobic whereas Aha has a strictly hydrophilic side-chain. Accordingly, it is expected that the opposite physicochemical properties directly provided by the synthetic amino acids are fully reflected in the relative lipase congeners.

In order to distinguish between synergistic effects of multiple (all-or-none) replacements and individual effects of key Met residues in the enzyme structure, the first efforts toward the development of a method for the differential replacement of Met residues are also presented. Until now, stop codon suppression-based methods are not applicable to Met analogs. The basic idea behind these systems is to import aminoacyl-tRNA synthetases and tRNAs that show no cross-reactivity with endogenous components. For this, methionyl-tRNA synthetases (MetRSs) from different species should be tested for their cross-reactivity in *E. coli* and for their natural preference toward various Met analogs. This is an important starting point to later evolve the MetRS specificity. Moreover, the natural fact that MetRSs charge two different tRNAs suggests the next point for a differential decoding, i.e. initiator $tRNA^{fMet}$ from *E. coli* is left to initiate translation whereas the imported elongator $tRNA^{Met}$ should be capable to decode internal positions more efficiently. The final aim is to reprogram protein translation *in vivo* whereby the N-terminus is translated with one Met analog, and the internal positions with another one; thus serving as a platform to further engineer biocatalysts of industrial interest.

3 Results and discussion

3.1 Engineering lipase with methionine analogs

The lipase from the thermophilic anaerobic bacterium *Thermoanaerobacter thermohydrosulfuricus* (TTL) was chosen not only because it is efficiently expressed in *E. coli,* but also because it is thermostable, has broad substrate specificity, and exhibits resistance against various organic solvents and detergents.[100] These features make the use of TTL highly advantageous for biotechnological applications since it can be produced at low cost while exhibiting improved stability. In this way, TTL is considered an attractive biocatalyst for studying the effects of Met replacement with both the more hydrophobic analog Nle and the more hydrophilic, azide group-containing Aha (**Figure 16**).

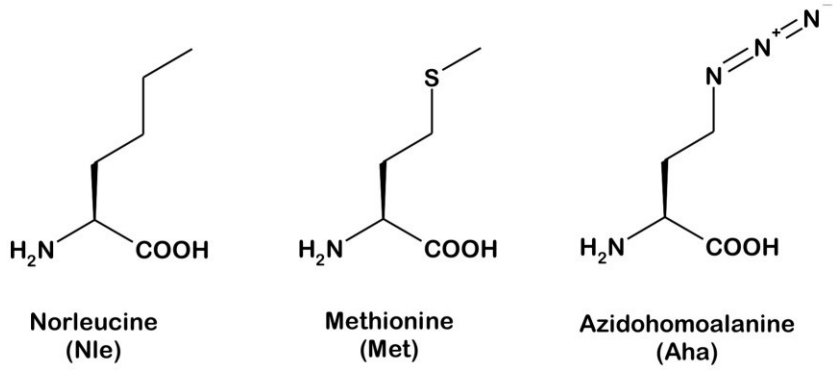


Figure 16. Probing the hydropathy of lipase with two Met analogs. The role of conformational preferences of the Met side-chain in TTL is probed with Nle, which is more hydrophobic than Met, and Aha, that is more hydrophilic than Met.

The high number of Met residues in this particular lipase (11 residues, see Appendix 9.1 for sequence details) might indicate a functional importance. Therefore, it is expected that the TTL[Nle] congener would be more hydrophobic than the native TTL, whereas TTL[Aha] would be a more hydrophilic, ‘clickable’ lipase.

3.1.1 Expression and purification

All lipases were generated by culturing amino acid auxotrophic *E. coli* host strains with Met, Nle, and Aha supplementation during heterologous expression (see 6.2.2.4 for details). The expression of the three lipases was indicated by the presence of bands having an average MW of 30 kDa (**Figure 17A**). Subsequently, all the lipases were purified (close to 95%) by affinity chromatography (**Figure 17B**).

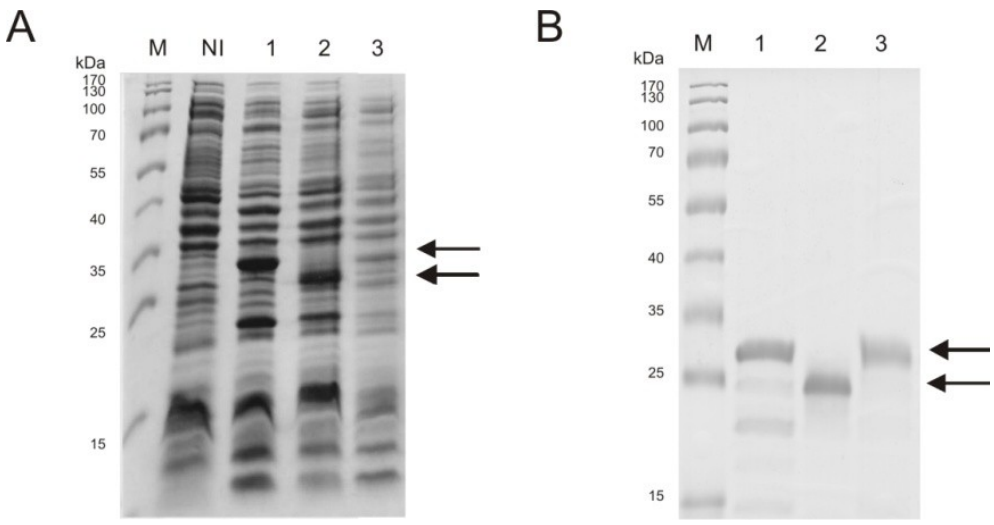


Figure 17. Lipase alloproteins. The TTL congeners were produced by supplementing the medium of auxotrophic *E. coli* strains with 1) Met, 2) Nle, and 3) Aha after expression (A) and purification (B) by using 12% SDS-PAGE and Coomassie staining (indicated with arrows). Note that TTL[Nle] (lane 2) is shifted compared to native TTL (lane 1) and TTL[Aha] (lane 3). M = MW standard marker; NI = Non-induced sample

Notable are the marked differences in the migration patterns observed for TTL[Nle] which moved faster than TTL and TTL[Aha] (**Figure 17**). In electrophoresis, SDS unfolds proteins and gives them a uniform negative charge such that these will migrate through the gel in the electric field based solely upon size. Hydrophobic proteins bind more SDS than hydrophilic ones. Since the lipase congeners have a very similar mass (not more than 100 Da difference; *vida infra*), the 5 kDa difference observed after electrophoresis might indicate that TTL[Nle] has a higher number of negative charges due to its hydrophobic nature compared to native and TTL[Aha].

In general, the SPI method allows the production in the order of milligrams of protein per liter of culture. The simple experimental set-up allows obtaining preparations with a yield of 20 mg/L, indicating the potential of using SPI for industrial applications.

3.1.2 Incorporation efficiency

In all incorporation experiments, comparatively high level of Met substitution is achieved using the SPI method. Here, a high incorporation level of all tested analogs into TTL was readily achieved, as indicated by ESI-MS results (**Figure 18**).

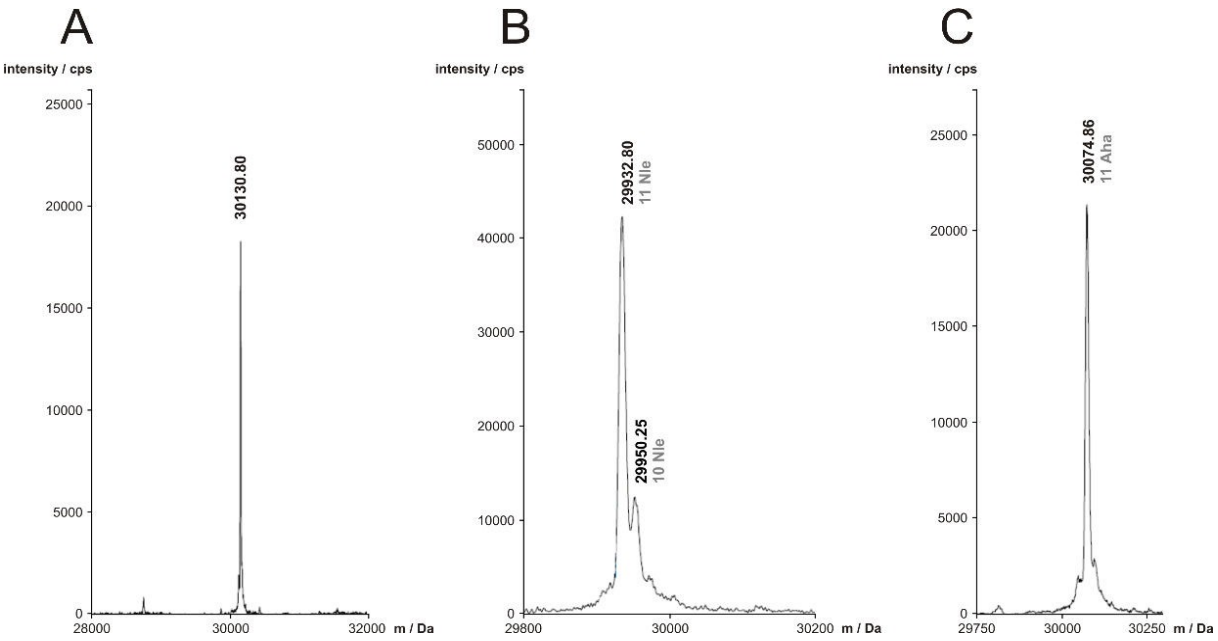


Figure 18. ESI-MS spectra of lipase congeners. Peaks corresponding to calculated masses are labeled accordingly. The theoretical calculated mass for TTL is 30130.2 Da (A), while for fully substituted TTL[Nle], it is 29932.8 Da and the congener with 10 Nle residues, 29950.8 (B); for fully labeled TTL[Aha], the theoretical calculated mass is 30074.6 Da (C).

In the case of TTL, a single peak corresponding to the expected mass was obtained (**Figure 18A**). For TTL[Nle], however, there was an additional mass species detected apart from the fully substituted congener containing only 10 Nle replacements (**Figure 18B**). Nonetheless, the preparation can generally be considered homogenous as the

dominant mass species is the fully Nle-substituted protein. In comparison, Aha was more successfully incorporated into TTL, since the spectra displays just a single peak (**Figure 18C**). In the both incorporations, no traces of native TTL were found.

3.1.3 Basal lipase activity

Lipase activity is normally determined by measuring the hydrolysis of *p*-nitrophenyl palmitate (*p*NPP). After the TTL congeners were characterized, cleavage of *p*NPP was measured at 70 °C in 25 mM Tris·HCl pH 8 according to the protocol of Winkler and Stuckmann (**Table 2**; see 6.2.4.4 for more details).[101]

Table 2. Enzymatic activity of lipase congeners.

Congener	Lipase activity (mU/µg)
TTL	2.3
TTL[Nle]	26.5
TTL[Aha]	0.5

TTL[Nle] displayed more than ten-fold enhanced activity in comparison to native TTL while TTL[Aha] showed almost 5 times lower lipase activity when compared to parent TTL (**Table 2**). The finding that the produced congeners have extremely different basal lipase activities indicates that the substitution of the Met by the more hydrophobic Nle, and the more hydrophilic Aha residues, caused the observed strike differences which can be better understood upon thermal activation (*vida infra*).

3.1.4 Optimal temperature and pH

Parent TTL is a thermoactive lipase that displays its highest enzymatic activity at an optimal temperature (T_{opt}) of ~70 °C and optimal pH (pH_{opt}) of 8.0.[100] In order to determine if the T_{opt} and pH_{opt} of the lipase congeners was affected, lipase activity was determined between 40 and 90 °C and pH 3-11 (**Figure 19**).

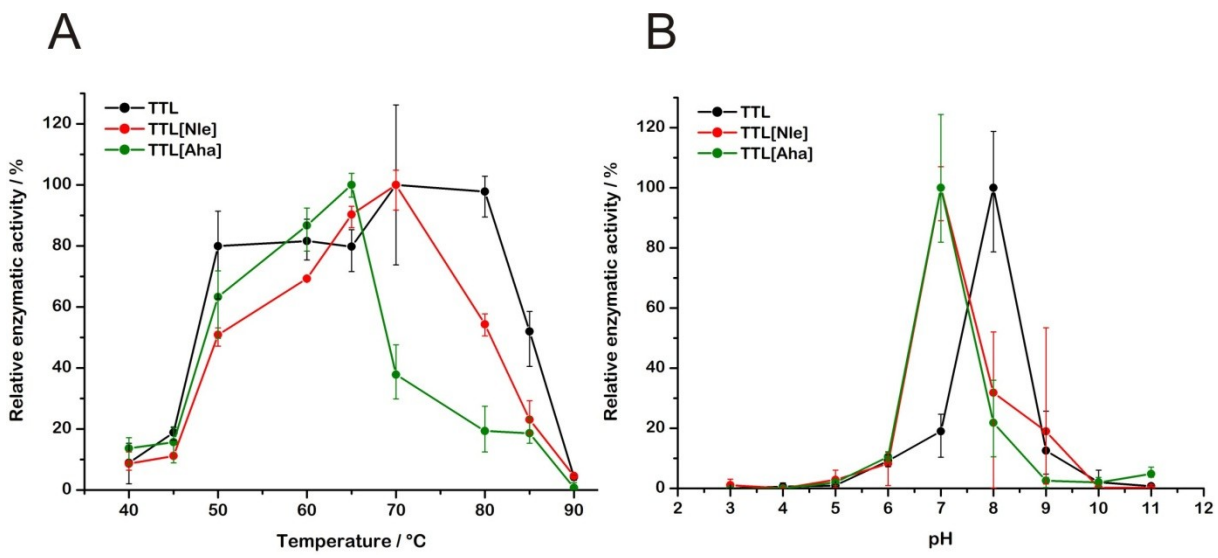


Figure 19. Temperature and pH profiles of lipase congeners. A) The optimal temperature for native TTL and TTL[Nle] is 70 °C, while for TTL[Aha], it is 65 °C. The reactions were performed using *p*NPP as substrate. B) The optimal pH for native TTL is 8, while for TTL[Nle] and TTL[Aha], it is 7. These reactions were carried out at T_{opt} overnight using tricaprylin as substrate. In both cases, enzymatic activity was measured in triplicate.

While the native TTL and TTL[Nle] congeners showed a T_{opt} of 70 °C, the TTL[Aha] exhibited better activity at 65 °C (**Figure 19A**). Note that 60% of TTL[Aha] activity is lost at 70 °C. These differences in the temperature profiles could be due to the acquired physicochemical properties of the lipase congeners. Once the T_{opt} for all lipases was found, these were then incubated at different pH values but using a tricaprylin assay (see 6.2.4.5 for details) given that *p*NPP is autohydrolyzed at extremely basic conditions; the standard lipase assay using *p*NPP as substrate is only reliable in the range of pH 6-9. Optimal lipase activity was observed at pH 8 for native TTL and at pH 7 for TTL[Nle] and TTL[Aha] (**Figure 19B**). These results reveal that all TTL congeners were mostly active between pH 7 – 9; however, the distinct maximum activity values are disproportionately high. Almost 70% of TTL[Nle] and TTL[Aha] activity is lost upon incubation at pH 8 whereas 80% activity of TTL appears to be lost at pH 7. These inconsistencies are intrinsic artifacts of the tricaprylin assay and have been described elsewhere.[102] For this reason, *p*NPP was used in subsequent assays.

3.1.5 Thermostability

To compare thermal stability, the lipases were preincubated at different temperatures and times; thereafter the residual lipase activity was measured (**Figure 20**).

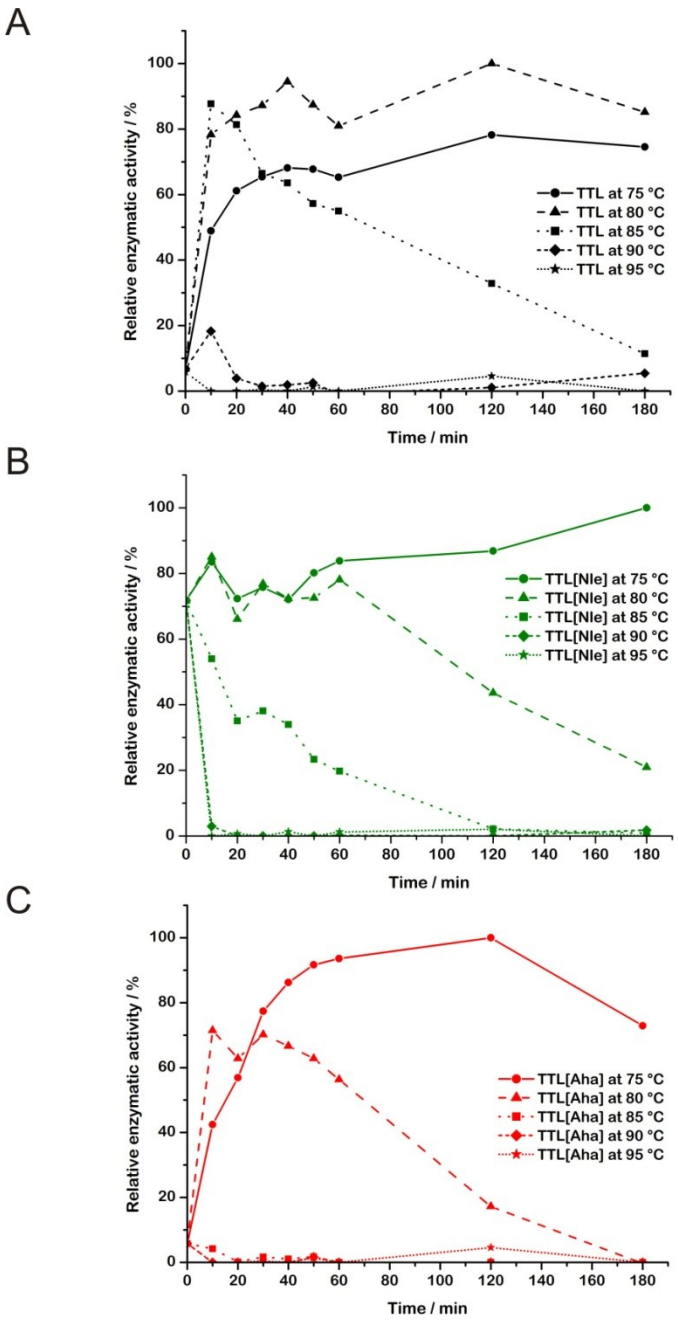


Figure 20. Thermal stability of lipase congeners. A) TTL, B) TTL[Nle], and C) TTL[Aha] were incubated at different temperatures for 10, 20, 30, 40, 50, 60, 120, and 180 minutes. Residual lipase activity of each sample was measured in triplicate for 10 min at T_{opt} and pH 8 using *p*NPP as substrate.

Native TTL is the most thermostable enzyme up to 3 hr at 80 °C (**Figure 20A**), followed by TTL[Nle] (**Figure 20B**) at the same time but incubated at 75 °C, and finally TTL[Aha] for up to 2 hr at 75 °C (**Figure 20C**). At 85 °C, native TTL displayed residual activity for three hours, while TTL[Nle] had residual activity for two hours but TTL[Aha] was almost inactive (**Figure 20**). However, the thermostability of the TTL congeners decreased with time when incubated at 90 °C, with only the native protein showing residual activity after 10 min; none of the enzymes was active at 95 °C (**Figure 20**). The most important finding, however, was that native and TTL[Aha] need thermal activation to be active whereas the TTL[Nle] showed as high as 70% lipase activity at the onset of incubation (**Figure 20**).

3.1.6 Thermal activation

The curves in **Figure 20** indicate that the requirement for thermal activation is restricted to native TTL and TTL[Aha], given that their activities sharply increased within the first 10 to 20 min after incubation at different temperatures. As mentioned previously, however, the most prominent feature among the lipase congeners is a several-fold higher activity of TLL[Nle] in the absence of thermal activation (T = 0 min, **Figure 20B**). Based on these observations, the TTL congeners were re-incubated at their T_{opt} for 10 or 60 min, or at their maximal thermal activation. Maximal thermal activation is defined as the temperature and time period where the lipase exhibited its maximum activity, i.e. native TTL (80 °C for 120 min), TTL[Nle] (75 °C for 10 min) and TTL[Aha] (75 °C for 120 min) (see **Figure 20** for details). In this way, the activity of all lipases was measured after different kinds of thermal activation (**Figure 21**).

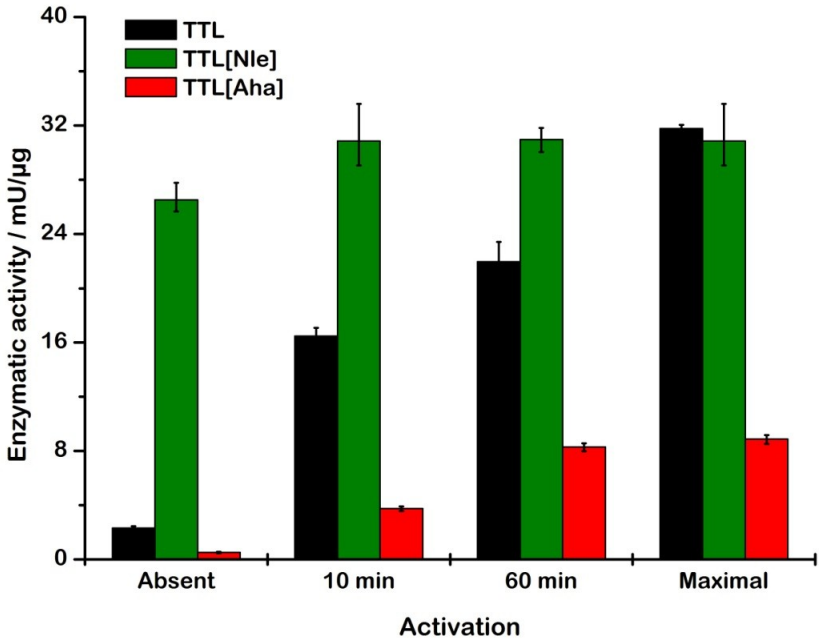


Figure 21. Thermal activation of lipase congeners. The activities of the TTL congeners were measured after activation for 10 and 60 min at their optimal temperature, which is 70 °C for native and TTL[Nle] and 65 °C for TTL[Aha]. For maximal activation, native TTL was incubated at 80 °C for 120 min, TTL[Nle] at 75 °C for 10 min and TTL[Aha] at 75 °C for 120 min. The values are averaged from three independent measurements using *p*NPP as substrate.

The most interesting finding is that the activity of TTL[Nle] did not significantly change after any kind of maximal thermal activation (**Figure 23**). In contrast, native and Aha-containing lipases showed a several-fold increase in activity following different setups of thermal activation (**Figure 23**). This result is rather interesting, since it is expected that the Met residues will enhance the hydrophilic properties of these lipases in contrast to the much more hydrophobic residues of TTL[Nle]. Nevertheless, it is important to note the activity differences between native and TTL[Aha]. Following maximal thermal activation, native TTL reaches an enzymatic activity only little higher than TTL[Nle], while TTL[Aha] loses 75% of the activity of TTL (**Figure 23**). Normally, lipases are not active in aqueous phase without either interfacial or thermal activation. In native TTL and TTL[Aha], the lid domain could be opened upon thermal activation to allow substrate contact; TTL[Nle], however, is always active in the absence of thermal activation, i.e. the enzyme is fully accessible to the substrate in the aqueous

phase. In this congener, the lid-loop must be flipped out to expose the lipid-binding cavity. The structural basis for this aqueous phase activation of TTL[Nle] can be understood in the context of the Met residues distribution across the lipase. To this end, advanced protein modeling techniques (see 6.2.5.1 for details) allowed the generation of a 3D structure homology model for TTL (**Figure 22**).

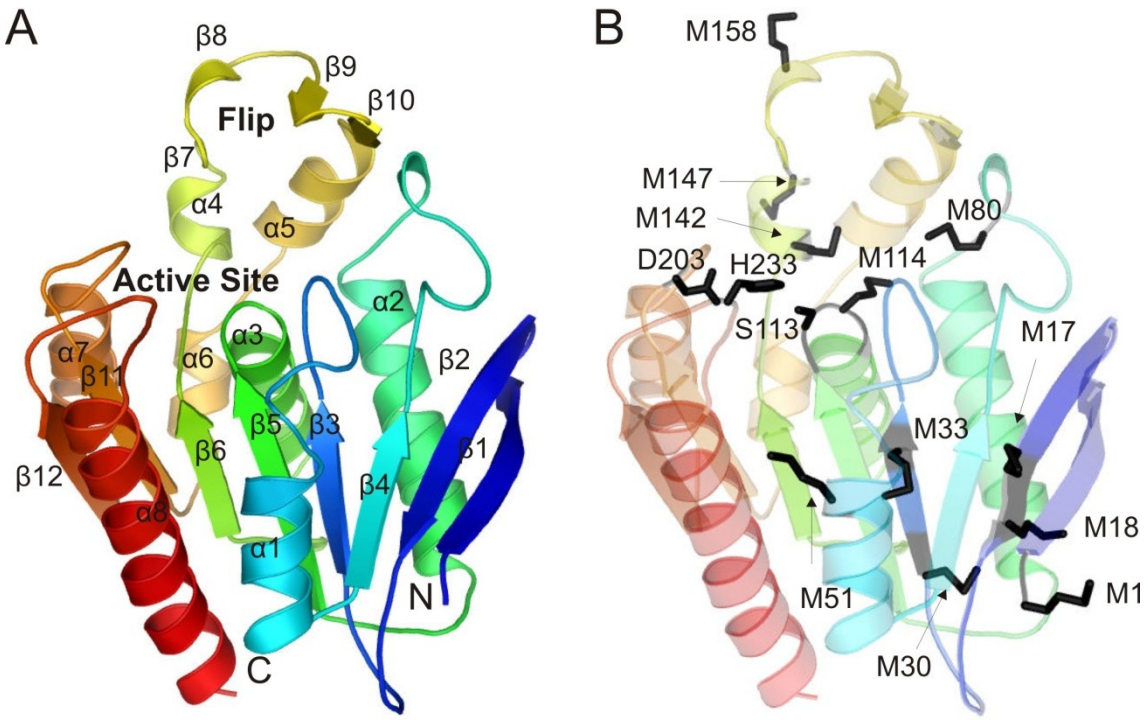


Figure 22. 3D structure homology model for TTL. A) The lipase displays a canonical α/β hydrolase fold (secondary structure elements are shown in black) and the lid. B) The active site of the lipase is composed of the catalytic residues Ser113, Asp203, and His233 (black). The Met residues (labeled black) are located within the hydrolase fold (M1, M17, M18, M30, M33, M51, and M80), close to the active site (M114) or in the lid domain (M142, M147, M158). N = N-terminus; C = C-terminus

Out of a total of 11 Met residues, three are positioned in the lid region; of these, Met147 and Met158 are surface exposed whereas Met142 is buried (**Figure 22B**). Met is a moderately hydrophobic amino acid whereas Nle is extremely hydrophobic (**Figure 6**), and its presence in the lid region should mimic hydrophobicity as delivered by lipid aggregates. As expected, the hydrophobic side of the lid is usually exposed to the lipid phase, enhancing the hydrophobic interactions between the enzyme and the lipid

surface.[103-105] Consequently, the global Met → Nle replacement would result in a hydrophilicity/hydrophobicity balance alteration where the “closed lid” of the native TTL could be converted permanently into a lipase congener having an “open lid”. The open lid congener could be tested against TTL[Aha], which is expected to be more hydrophilic, and in which this “permanently activated” state would not be observable.

Additionally, it is well known that presence or absence of the lipase lid over the active site is not the sole determinant of its activity. Dimerisation is a basic prerequisite for activity in many lipases[106], and the presence of hydrophobic residues in the dimerisation interface could further enhance lipase structural and functional integrity. Of the Met residues, Met1 is part of this interface, while Met17, Met18, and Met30 are in the vicinity (**Figure 22B**). Finally, the replacement of the residues Met80 and M114, which are close to the active center of the enzyme, could have a great impact on the overall lipase activity. In summary, the replacement of Met with more hydrophobic Nle residues across the TTL molecule may act as a coupled network to beneficially promote catalysis.[107]

For all these reasons, the global Met → Nle substitution could be an elegant way to design a stable lipase allowing the substrate to be accessed in the aqueous phase without thermal activation. This feature could be a great advantage in detergent applications where washing-powder lipases are used. In fact, the lid is usually removed in commercial lipases via classical genetic engineering to achieve constant substrate accessibility.[108] The lid region, however, certainly plays an important role in the overall structural integrity of the lipase and its removal might have detrimental effects on the structural and functional integrity of the enzyme.[108] Unfortunately, due to the lack of high-resolution structural data it is difficult to speculate on how the global substitution of Met residues with Nle results in the enormous changes in activation by thermal activation. Nonetheless, all samples were submitted to CD spectroscopy to see whether the secondary structure was affected (**Figure 23**; see 6.2.4.6 for details).

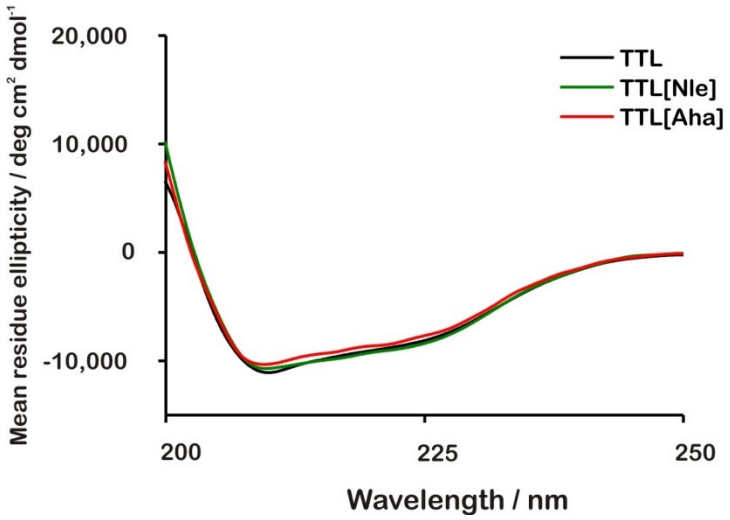


Figure 23. Secondary structure of lipase congeners. The far-UV CD spectra from 200–250 nm was recorded at T_{opt} of 70 °C for native TTL and TTL[Nle], and at 65 °C for TTL[Aha].

The CD spectrum of the two lipase congeners revealed only negligible deviations from that of the native TTL (**Figure 23**). These findings indicate that the lipase secondary structure is not altered, thus suggesting that the lipase activity of the generated congeners may be retained. In support of this, however, NMR or X-ray crystallographic studies are necessary although other proteins have revealed that the incorporation of Nle did not perturb their 3D structures.[48]

3.1.7 Substrate specificity

Lipases hydrolyze triacylglycerols into fatty acids and glycerol, and esterify fatty acids and alcohols. The synthesis of specific compounds depends on the nature of the substrate binding site formed by a hydrophobic tunnel near the catalytic residues. In addition, the fatty acid chain-length of the corresponding substrate depends on the shape of the tunnel and its amino acid composition.[109] It is therefore reasonable to think that the substrate specificity of the lipase congeners could be affected upon incorporation of Nle and Aha in place of Met, since the residues M114 and M142 are located near the active site (**Figure 22**). To determine the substrate specificity of all TTL lipases, two types of substrates were used (**Figure 24**).

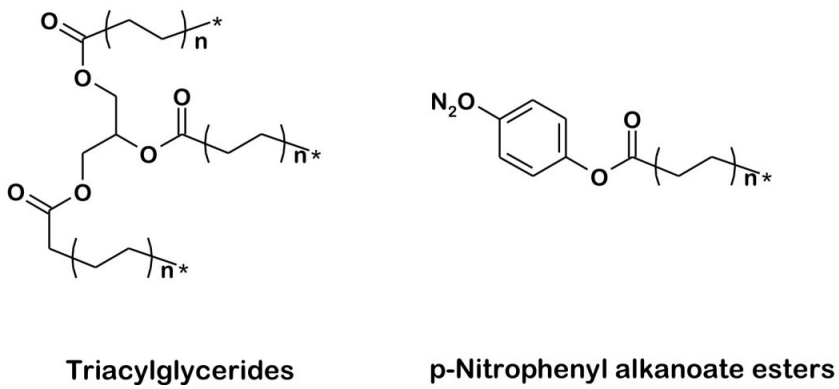


Figure 24. Two types of lipase substrates. Triacylglycerides or p-nitrophenyl alkanoate esters of varying acyl or alkyl side-chain lengths were used to determine substrate preferences of the lipase congeners. n = variable number of substituent; * = CH_3

Lipase activity of all TTL congeners was measured using triacylglycerides of varying acyl side-chain lengths (**Figure 25**; see 6.2.4.5.2 for details).

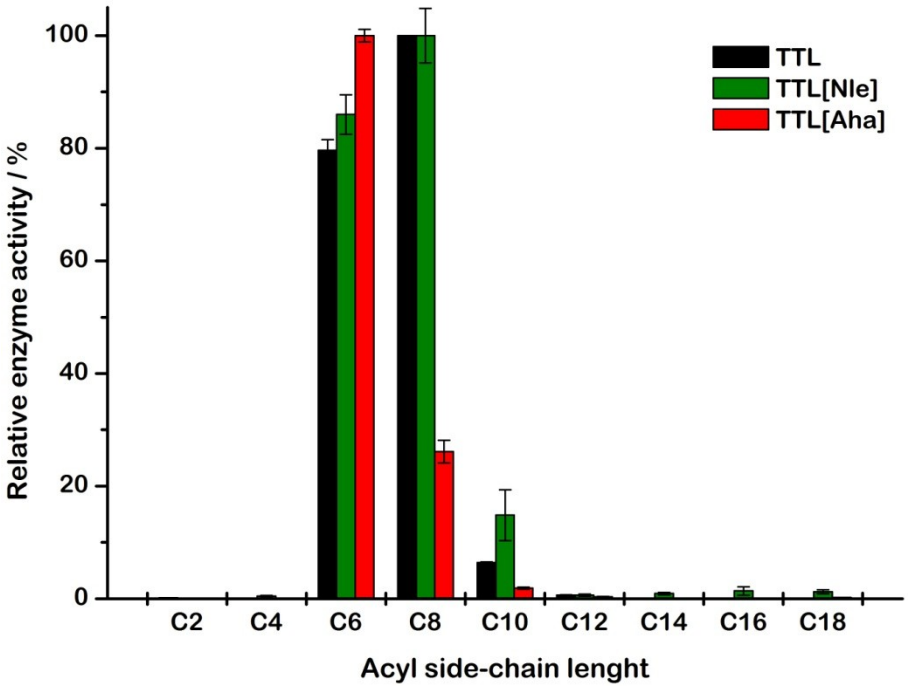


Figure 25. Lipases substrate specificity for tryacylglycerols. The substrates triacetin (C2:0), tributirin (C4:0), tricaproin (C6:0), tricaprylin (C8:0), tricaprin (C10:0), trilaurin (C12:0), trimyristin (C14:0), tripalmitin (C16:0), and tristearin (C18:0) were incubated with all TTLs at T_{opt}, pH 8 for 18 h.

TTL normally shows high activity in the hydrolysis of triacylglycerols: C8:0 > C6:0 > C10:0, but low activity with shorter or longer acyl side-chains.[100] Importantly, the same behavior was detected for TTL but some differences were found for Nle- and TTL[Aha] (**Figure 25**). TTL[Nle] shows a similar pattern as parent TTL, except for a two-fold enhanced activity in tricaprin hydrolysis (C10:0), whereas the more hydrophilic TTL[Aha] accommodates short- over long-chain length acyls better: C6:0 > C8:0 > C10:0 (**Figure 25**).

The enzymatic activity of the other lipases was also determined using varying alkyl side-chain lengths derived from *p*-nitrophenyl alkanoate esters (**Figure 26**; see 6.2.4.4.3 for details).

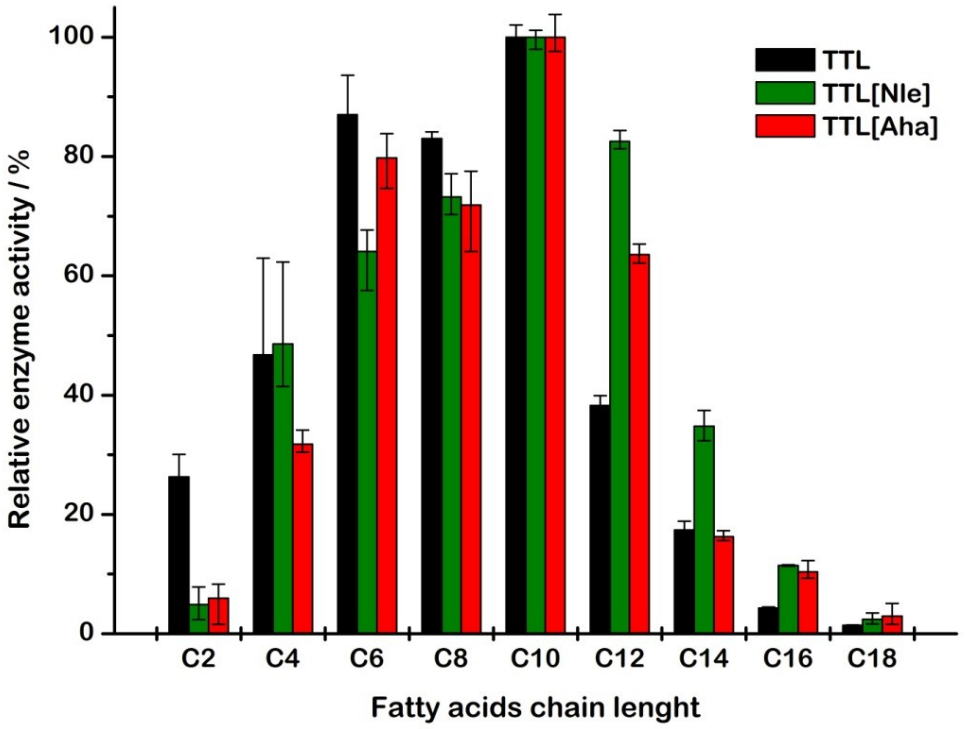


Figure 26. Lipases substrate specificity for *p*-nitrophenyl alkanoate esters. The *p*NP-derived substrates acetate (C2:0), butyrate (C4:0), caproate (C6:0), caprylate (C8:0), caprate (C10:0), laurate (C12:0), myristate (C14:0), palmitate (C16:0), and stearate (C18:0) were incubated with all TTLs at T_{opt}, pH 8 for 10 min.

In this case, the accessibility of the single fatty acid side-chain substrate seems to be less limited than in the side-chains of the triacylglycerides. When hydrolyzing *p*-nitrophenyl alkanoate esters, TTL displays broad substrate specificity: C10:0 > C6:0 > C8:0 > C4:0 > C12:0 > C2:0 (**Figure 26**). TTL[Nle] displayed a different preference when compared to native TTL: C10:0 > C12:0 > C8:0 > C6:0 > C4:0 > C14:0 (**Figure 26**). TTL[Aha] shares more similarities with native than TTL[Nle]: C10:0 > C6:0 > C8:0 > C12:0 > C4:0 > C14:0 (**Figure 26**). It is noteworthy to mention that TTL[Nle] hydrolyzes trilaurin (C:12), trimyristin (C14:0), and tripalmitin (C16:0) twice as better as compared with the native TTL. The substitution of M114 and M142 by Nle could alter the composition of the lipase tunnel, allowing the lipase to achieve more accessibility to substrates with longer side-chains.

3.1.8 Influence of additives on lipase activity

To study the effects of several substances on lipase activity, the purified TTL congeners were pre-incubated with different concentrations of organic solvents, metal ions, surfactants, and inhibitors at RT for 1 hr without substrate (see 6.2.4.4.1 for details); subsequently, the residual lipase activity was measured. It is important to mention that in all the assays for evaluating the effect of different additives on lipase activity, no thermal activation was performed because of practical reasons. Consequently, the comparison of the activities between the different lipases could be considered for both an ‘open’ conformation congener, TTL[Nle], and a ‘closed’ congener, as the case is for TTL and TTL[Aha]. It is noteworthy to mention that absolute effect of the substances is related to the basal lipase activity (BLA) for each congener, i.e. negative values indicate a negative effect on the previously active enzyme.

3.1.8.1 Organic solutions

The nature of the organic solvent is crucial in maintaining the water content required for catalytic activity. More hydrophobic solvents preserve the catalytic activity, leaving

a water layer that adheres to the enzyme surface acting as a protective shell. On the other hand, hydrophilic solvents tend to strip essential water molecules from the enzyme surface, thereby altering its catalytic conformation. To determine the extent of influence of different organic solvents on lipase activity, the TTL congeners were exposed to 90% solutions of the substances of interest. Residual enzyme activities of each lipase were measured thrice and normalized against water (**Figure 27**).

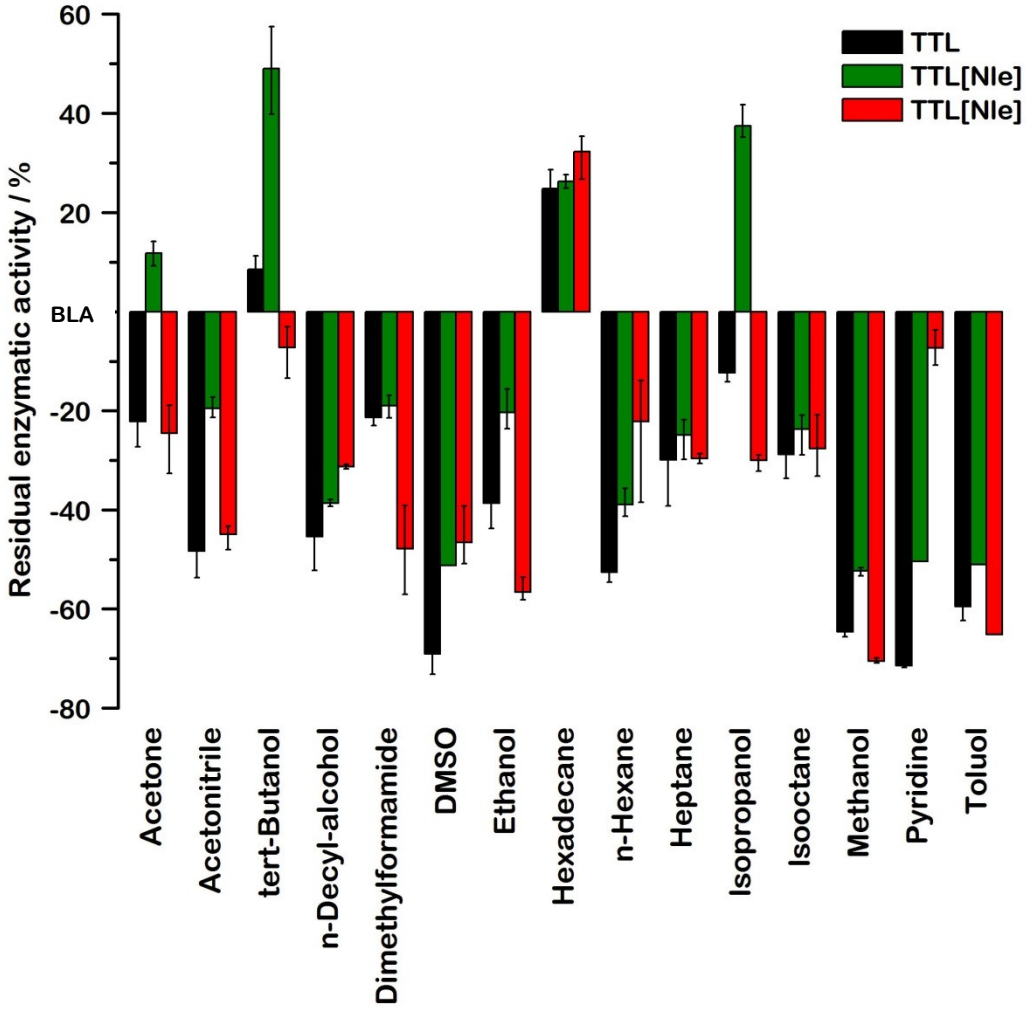


Figure 27. Influence of solvents on lipase activity. The TTL congeners were exposed to 90% of different organic solutions for 60 min and the residual lipase activity was measured in triplicates at T_{opt} and pH 8 with *p*NPP as substrate. All enzymes were active after treatment except TTL[Nle] in pyridine, DMSO, methanol, and toluol; and TTL[Aha] in methanol and toluol. Data is normalized to the basal lipase activity (BLA) of the corresponding congener in the presence of water, and without thermal activation. DMSO = Dimethyl sulfoxide.

Hexadecane was the only solvent that enhanced lipase activity of all congeners, with TTL[Aha] having the highest activity. This finding is very interesting, since this lipase is known to be the least active lipase with or without thermal activation (**Figure 21**). While the incubation with acetone, *tert*-butanol, and isopropanol enhanced the basal lipase activity of TTL[Nle] in 10%, 50%, and 40%, respectively; the activities of native TTL (except for *tert*-butanol) and TTL[Aha] are strongly diminished (**Figure 27**). The rationale behind these differences could be due to the intrinsic hydrophobicity of these substances, allowing the more hydrophobic Nle-containing congener to be more active. This may translate not only into economical savings but also advantages in particular chemical reactions. For instance, it has been suggested that isopropanol favors an open conformation in *Candida rugosa* lipase, allowing a better resolution of 2-(4-chlorophenoxy)-propanoic acid, whose (*R*)-enantiomer lowers cholesterol levels and prevents platelet aggregation.[110] The hydrophobic TTL[Nle] with an 'open' conformation would allow more efficient catalysis to take place in isopropanol.

On exposure to acetonitrile, *N*-decyl-alcohol, dimethylformamide, DMSO, ethanol, methanol, heptane, hexadecane, *n*-hexane, isooctane, pyridine, and toluol the activity of all lipases was strongly affected in different degrees. Nevertheless, after pyridine treatment, TTL[Aha] residual activity was less-detectable when compared to parent and TTL[Nle] (**Figure 27**). Interestingly, there is a great interest in lipase-catalyzed transesterifications between sugars and esters for the synthesis of biodegradable polymers[111]; however, this reaction is relatively inefficient, since the solubility of sugars in inorganic solvents is very low. Pyridine is a solvent widely used for the solubilization of huge amounts of sugar; therefore, the potential of using TTL[Aha] in these reactions should not be underestimated. Overall, these results suggest that in contrast to native TTL, the Nle and Aha congeners could be used in specific esterification reactions that otherwise would not be optimal without incorporating noncanonical amino acids.

3.1.8.2 Metal ions

As previously indicated, lipases are frequently used in the production of enantiomers. However, the lack of enantioselectivity of some lipases still remains a major problem for many organic chemists. To solve this drawback, different strategies have been proposed, including the optimization of the solvent, water content, modifications in the lipase, protein engineering, and additive effects.[112] Of these, the most attractive alternative is the addition of metal ions due to its simplicity and its potential. A lipase-catalyzed enantionselectivity reaction, for example, is increased hundred-fold by adding a small amount of LiCl.[113] Moreover, when the temperature of the reaction decreased, an almost perfect enantioselective reaction took place. This illustrates the importance of using metal ions in lipase-catalyzed reactions. Accordingly, the effect of metal ions on lipase activity was investigated using various salt solutions (**Figure 28**).

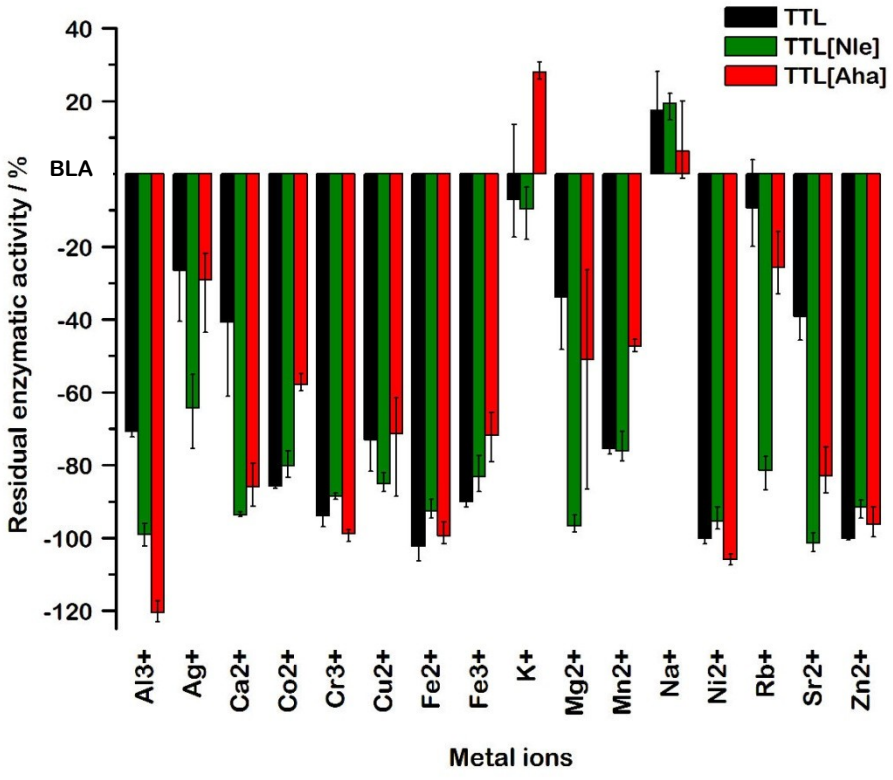


Figure 28. Influence of metal ions on lipase activity. The TTL congeners were exposed to 10 mM of different salt solutions for 60 min and the residual lipase activity was measured at T_{opt} and pH 8 with *p*NPP as substrate. All enzymes were active after treatment except TTL (Fe^{2+}, Ni^{2+}, and Zn^{2+}), TTL[Nle] (Al^{3+} and Sr^{2+}) and TTL[Aha] (Al^{3+}, Cr^{3+}, Fe^{2+}, Ni^{2+} and Zn^{2+}). The data is normalized with respect to the basal lipase activity (BLA) of the corresponding congener in presence of water, and without activation.

The only case where the enzyme activity of all lipases was stimulated over their basal values was observed in the presence of Na^+ (**Figure 28**). The only congener whose activity was enhanced after K^+ treatment was TTL[Aha], while 10% of native and TTL[Nle] activity was lost (**Figure 28**). Except for Ag^+, these results suggest that the monovalent cations did not alter significantly lipase activity. On the other hand, after Rb^+ treatment, TTL was also slightly inhibited (10%), with TTL[Aha] being inhibited to a larger extent (30%), but TTL[Nle] was strongly affected (close to 80%;**Figure 28**). Other metal ions resulted in inhibitory effects on at least 25% of the activities of all lipases (**Figure 28**). The inhibitory effect of calcium ions on lipase activity is well documented.[114] Finally, Cr^{3+}, Fe^{2+}, Ni^{2+}, and Zn^{2+} had the same strong inhibitory effect on all lipases (**Figure 28**). After Mg^{2+} and Rb^2 treatment, both native and TTL[Aha] were much less inhibited than the more hydrophobic TTL[Nle]. The overall inhibition differences could be due to strong interactions between divalent metal ions and the appended C-terminal His·Tag.

3.1.8.3 Surfactants

Lipases are active at the oil-water interface of emulsified substrates. However, emulsions are inherently unstable, with their degree of stability being influenced by the nature and concentration of surfactant used. A surfactant consists of both hydrophobic and hydrophilic groups, allowing it to be marginally soluble in both, polar and non-polar substances. Emulsifiers or surfactants help lowering the interfacial tension between phases or surrounding the emulsion droplet with a thin, charged film. Hence, lipase activity at the oil-water interface could be altered by the presence of these substances. Moreover, there is an overwhelming interest in screening lipases for use in the cosmetic and perfume industry because of their activity in surfactants and their use in aroma production.[115] To further characterize the effect of surfactants, the lipase congeners were incubated with different detergents and other agents (**Figure 29**).

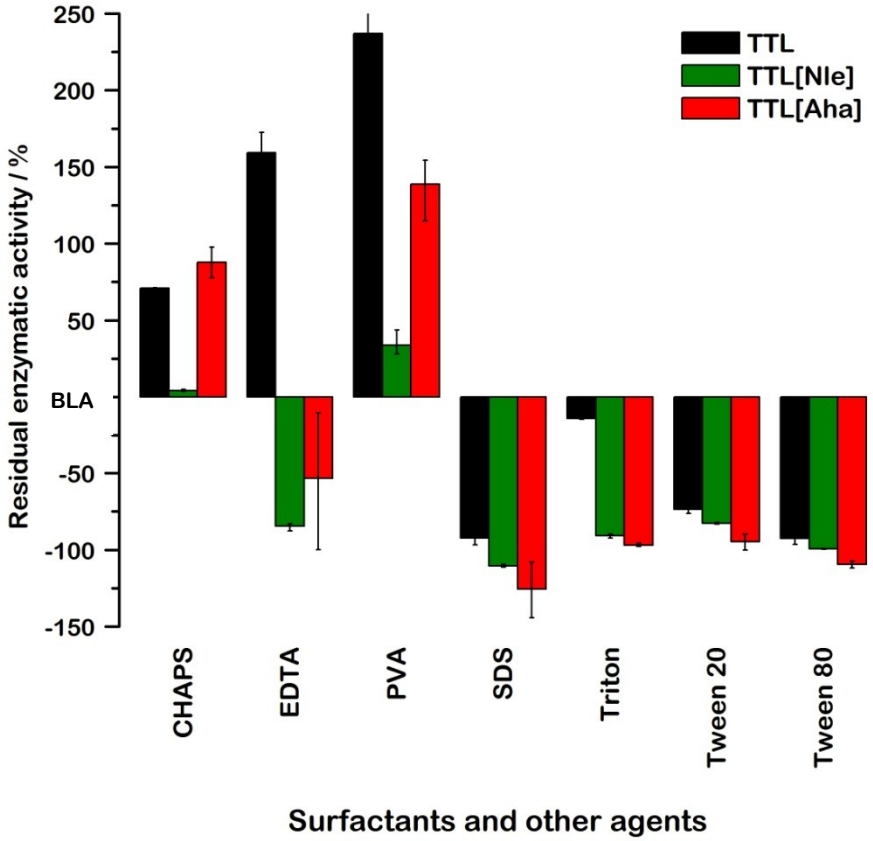


Figure 29. Influence of surfactants in lipase activity. The TTL congeners were exposed to 10% solutions for 60 min and the residual lipase activity was measured thrice at T_{opt} and pH 8 with *p*NPP as substrate. Data is normalized against the basal lipase activity (BLA) of the corresponding congener in water, and without activation. CHAPS = 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate; EDTA = Ethylenediaminetetraacetate; PVA = Polyvinylalcohol; SDS = Sodium dodecyl sulfate; Triton = Triton X-100 or polyethylene glycol p-(1,1,3,3-tetramethylbutyl)-phenyl ether; Tween 20 = Polyoxyethylene (20) sorbitan monolaurate; Tween 80 = Polyoxyethylen (20) sorbitan monooleat

CHAPS had a positive influence on lipase activity of native TTL (70%) and TTL[Aha] (80%), but this improvement is not mirrored in TTL[Nle] (5%; **Figure 29**). A similar but more marked pattern was found in the presence of PVA, where the activity of native lipase increased by more than 230%, TTL[Aha] by 140%, and TTL[Nle] by only 30% when compared to their basal activities (**Figure 29**). The alterations observed are probably attributable to the chemical nature of these substances. CHAPS is a detergent containing hydrophobic regions whereas PVA is a water-soluble synthetic polymer; however, both contain many polar hydroxyl groups. Hence, these

substances could preferentially interact with the exposed hydrophilic Met and Aha residues of native and TTL[Aha], respectively, but not with those hydrophobic Nle residues of the TTL[Nle]. Enhancement of lipase activity upon treatment with CHAPS has been previously reported in the literature.[116-118] Since EDTA can chelate most metal ions, it is used to treat water, and also as an additive in detergents, chemical, paper, food, medical and cosmetic industries. It also has applications in the inactivation of metal-dependent enzymes. While the activity of TTL increased by more than 150% in presence of EDTA, the opposite effect was induced in TTL[Nle] and TTL[Aha] with almost 80% and 50% inhibition, respectively (**Figure 29**). This suggests that the presence of metal ions has more influence on the activity of the generated congeners than parent TTL. These inhibitory effects of EDTA on other lipases have also been previously reported.[114]

A general decrease of all lipase activities was noticed after incubation with SDS, Triton, Tween 20, or Tween 80, where Triton was the only surfactant having a less negative effect on native TTL (10%; **Figure 29**). Triton X-100 is nonionic and contains both hydrophilic and hydrophobic regions, but no net charges. The hydrophilic part is composed of a polyethylene oxide group of 9.5 ethylene oxide units on average. Tween 20 is a polyoxyethylene derivative of sorbitan monolaurate (C12:0). Tween80 is a nonionic surfactant and emulsifier derived from polyethoxylated sorbitan and oleic acid which is very soluble in water and often used in food industry. SDS is one of the most widely used anionic surfactants in cleaning and products for general hygiene including toothpastes, shampoos, and shaving foams. The effects on the lipase activities could only be understood by more in-depth studies of the interactions between the surfactants, the substrate, and the lipases.

3.1.8.4 Inhibitors

The effect of inhibitors on lipases is important because it provides information on its catalytic mechanism. The effect of enzyme can be reversible or irreversible; reversible inhibitors for lipase inhibit its activity by changing the conformation or

interfacial properties, but do not act directly on the active site. Irreversible inhibitors, on the other hand, act directly on the active site. To further characterize the lipase congeners, the irreversible inhibitors Pefabloc and PMSF, as well as denaturing agents were used (**Figure 30**).

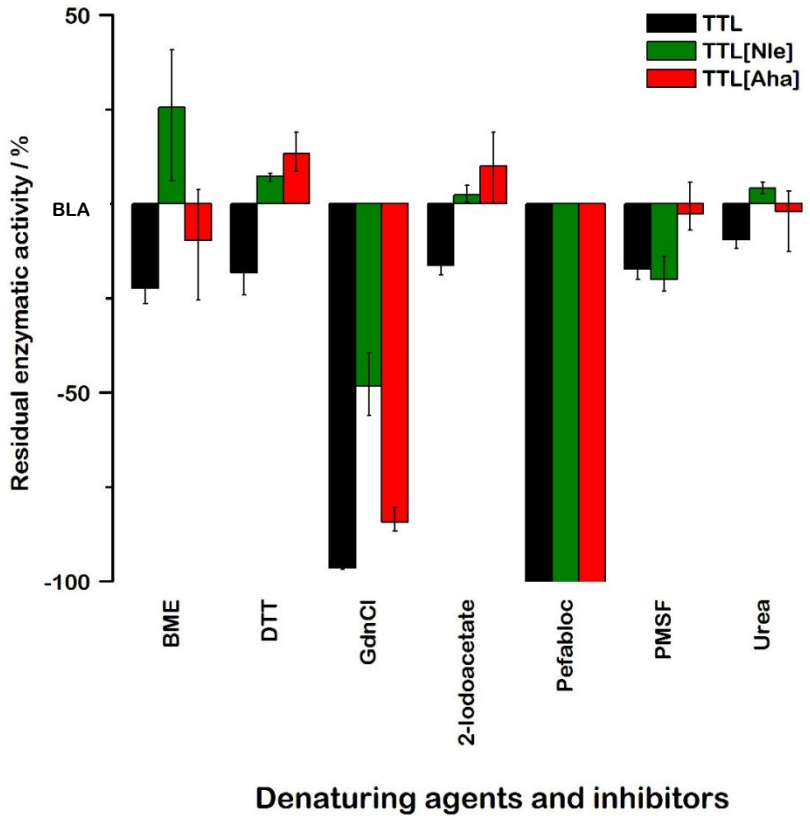


Figure 30. Influence of inhibitors in lipase activity. The TTL congeners were exposed to 1 mM of different inhibitors (500 mM guanidinium chloride and 2M urea) for 60 min and the residual lipase activity was measured three times at T_{opt} and pH 8 with *p*NPP as substrate. Data is normalized against the basal lipase activity (BLA) of the corresponding congener in presence of water, and without activation. BME = β-Mercaptoethanol; DTT = dithiotreitol; GdnCl = Guanidine hydrochloride; Pefabloc = 4-(2-aminoethyl)-benzenesulfonyl fluoride hydrochloride; PMSF = phenylmethylsulfonyl fluoride

The lipase activity of all congeners was strongly affected after treatment with GdnCl, one of the strongest denaturants used in physiochemical studies, and irreversibly inhibited by the water soluble, serine-specific inhibitor Pefabloc (**Figure 30**). The total inactivation of all lipases by Pefabloc is caused by the modification of the essential serine residue that plays a key role in the catalytic mechanism of all known lipases.[119]

Although TTL activity was completely lost when it was preincubated at 30 °C for 90 min with 1 mM PMSF,[100] the same amount of PMSF for 60 min did not affected that much the activity of TTL[Aha] activity, but inhibited as much as 25% the activity of parent TTL and TTL[Nle] (**Figure 30**). These differences could be explained by the length exposure time.

Most lipases do not contain neither free –SH or S–S bridges that play an important role in catalytic activity. Consequently, it is expected that the reducing agents β-mercaptoethanol (BME), DTT, and 2-iodoacetate will have no effect on lipase activity. However, all these agents affected the activity of parent TTL but showed an opposite effect on the activities of TTL[Nle] and TTL[Aha] (except for BME that also diminished lipase activity; **Figure 30**). Interestingly, similar patterns in different instensities were found in lipase activities when BME or urea was used (**Figure 30**). Surprinsingly, the parent TTL is the most vulnerable congener to harsh denaturing and reducing agents, while TTL[Aha] is more more resistant, followed by TTL [Nle]. The rationale behind these results, however, should be carefully done, since the differences in lipase activities are small among the congeners.

In summary, the global substitution of Met residues with the related isosteric analog Nle and the surrogate Aha yielded lipase congeners with both elevated and lowered optimal temperature and pH, thermostability, substrate access and specificity, as well as resistance toward organic solutions, metal ions, surfactants and inhibitors. However, it is difficult to say whether the emergent features of the TTL congeners are due to synergistic effects of multiple (all-or-none) replacements or mostly due to individual effects of key Met residues in the enzyme structure. To this end, the development of a method that allows the site-specific incorporation of Met analogs is necessary.

3.2 Making up the faces of Janus

Currently, no method for the site-specific incorporation of Met analogs into proteins has been developed, probably because the AUG codon has an unambiguous meaning: Methionine. Nevertheless, using AUG sense codons with ambivalent meanings for the site-specific incorporation of Met analogs may allow us to further engineer biocatalysts of industrial interest. This possibility can only be realized by introducing an orthogonal pair, which is composed of an aminoacyl-tRNA synthetase and its cognate tRNA, which does not cross-react with the endogenous components from the host. Several orthogonal pairs from *S. cerevisiae* (Phe), *M. jannaschii* (Tyr), *M. maize* (Pyr), and *M. barkeri* (Pyr) have already been imported into *E. coli*.[8] Most of these, however, rely on stop codon suppression-based methods applicable only to aromatic analogs (Tyr or Phe) or long aliphatic Lys analogs such as Pyr. Nevertheless, the fact that Tirrell and coworkers were able to reassign two degenerate sense codons with two different Phe analogs through a yeast orthogonal pair[120] suggests that in principle, it is feasible to apply the same concept to the AUG sense codon that actually participates both in translation initiation and elongation.

3.2.1 Searching for an orthogonal pair of methionine

The evolution of an orthogonal pair may not be a serendipitous finding[121], considering that an AARS and its cognate tRNA from an organism belonging to a life domain (e.g. bacteria, archaea, or eukarya) usually do not cross-react with counterparts from other evolutionary distinct organisms.[122] For instance, *Ec*MetRS exhibits low affinity for eukaryotic cytoplasmic elongator $tRNA^{Met}$ [123-125], but has a high affinity for initiator tRNAs from other bacteria, as well as mitochondrial and chloroplastic tRNAs from eukaryotes.[124-126] Similarly, MetRSs from mitochondria and chloroplasts efficiently aminoacylate $tRNA^{Met}$ from *E. coli,* whereas eukaryotic cytoplasmic MetRSs aminoacylate mitochondrial, chloroplastic and *E. coli* $tRNAs^{Met}$ with low efficiency.[127-129] Finally, mammalian cytoplasmic MetRSs aminoacylate yeast cytoplasmic $tRNA^{Met}$.[130] Taken together, these cross-aminoacylation experiments suggest that

there are two types of pairs in the Met system (**Figure 31**).[131]

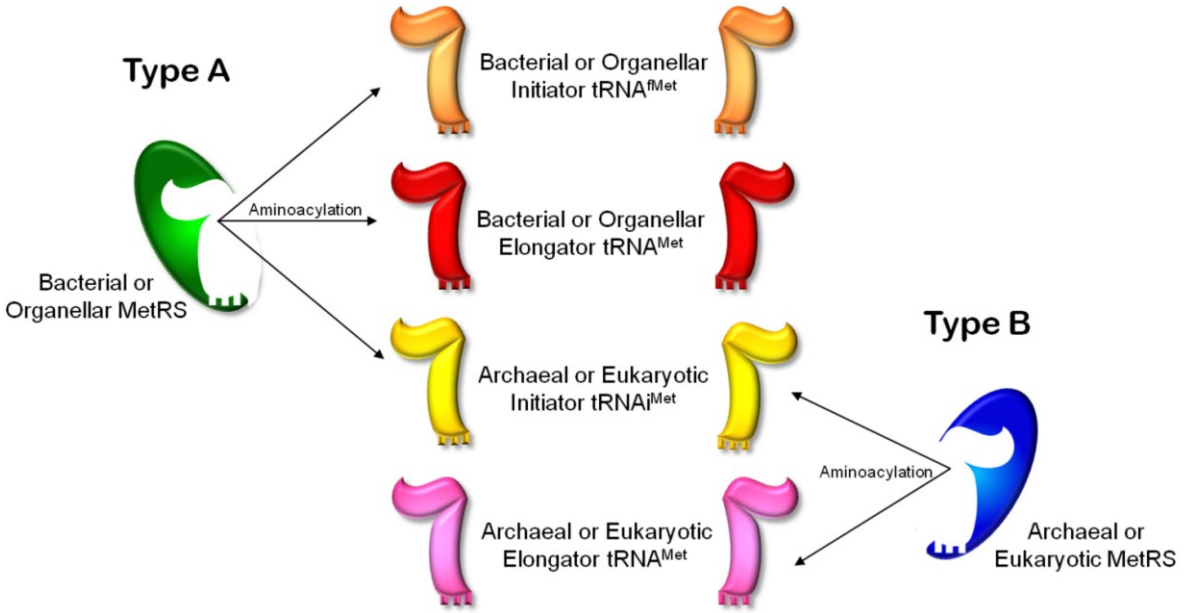


Figure 31. Two types of MetRS:tRNA^{Met} pairs. Type A includes the MetRS from eubacteria and eukaryotic organelles, which are able to preferentially aminoacylate their initiator (tRNA^{fMet}) and elongator (tRNA^{Met}) counterparts, as well as archaeal or eukaryotic cytoplasmic initiator tRNAi^{Met} (tRNAi^{Met}). Type B includes archaeal or eukaryotic cytoplasmic MetRS, which exclusively aminoacylate their cognate elongator tRNAsMet (tRNA^{Met}) and their tRNAi^{Met} (tRNAi^{Met}).

Consequently, it should be feasible to use a type B MetRS:tRNA^{Met} pair that incorporates a Met analog at AUG codons exclusively or at least preferentially during elongation in *E. coli*, since exogenous MetRS would not be able to aminoacylate either the bacterial initiator tRNA^{fMet} or elongator tRNA^{Met}. The overexpression of the orthogonal elongator tRNA^{Met} charged with a Met analog should theoretically be able to efficiently compete with the presumably less abundant, misaminoacylated bacterial elongator tRNA^{Met}. Exogenous tRNA^{Met} should not be aminoacylated by any of the 20 *E. coli* AARSs, including the endogenous MetRS. The type B tRNA^{Met} should also behave as a normal substrate of the bacterial ribosomal machinery. Finally, the Met analog should not be recognized as a substrate of endogenous MetRS nor any other AARS. In summary, the participating molecules have to be orthogonal.[132]

Given this, finding an orthogonal MetRS can be realized by identifying as many enzymes of archaeal or eukaryotic origin as possible. First, the available coding sequences of all MetRSs were obtained from the American National Center for Biotechnology Information (NCBI) databases. Next, the bacterial or archaeal genomic DNA or eukaryotic cDNA of many organisms were acquired from different sources, so that their respective genes amplify by PCR (see 6.2.1.1 for details). Based on these, the MetRS from bacteria (*Aquifex aeolicus, E. coli*), archaea (*Aeropyrum pernix, Methanocaldococcus jannaschii, Natronomonas pharaonis, Pyrobaculum aerophilum, Pyrococcus abyssi, Sulfolobus acidocaldarius,* and *Thermoplasma acidophylum*), and eukarya (*Arabidopsis thaliana, Danio rerio, Drosophila melanogaster, Homo sapiens,* and *Saccharomyces cerevisiae*) were cloned into suitable expression vectors (see 6.1.8.2 for details). Finally, thirteen 'orthogonal candidate' MetRSs, excluding *Ec*MetRS, which is intended as a positive control, were generated. Each MetRS is referred to on the basis of the organism where it comes from, e.g. MetRS from *A. pernix* = *Ap*MetRS, etc.

3.2.1.1 Expression and solubility in *E. coli*

Vectors containing the different MetRS were transformed into suitable *E. coli* strains and protein expression was induced by adding IPTG (see 6.2.2.3 for details). A solubility test was subsequently performed to check if the MetRS was expressed in either in the soluble fraction or as inclusion bodies (see 6.2.2.6 for details). Insoluble exogenous MetRS would be disadvantageous, and are discarded. In all cases, the same amount of non-induced (NI) or ITPG-induced (I) cells, as well as their soluble fraction (SF) and insoluble fraction (IF) was analyzed by SDS-PAGE using a 12% gel (**Figure 32**).

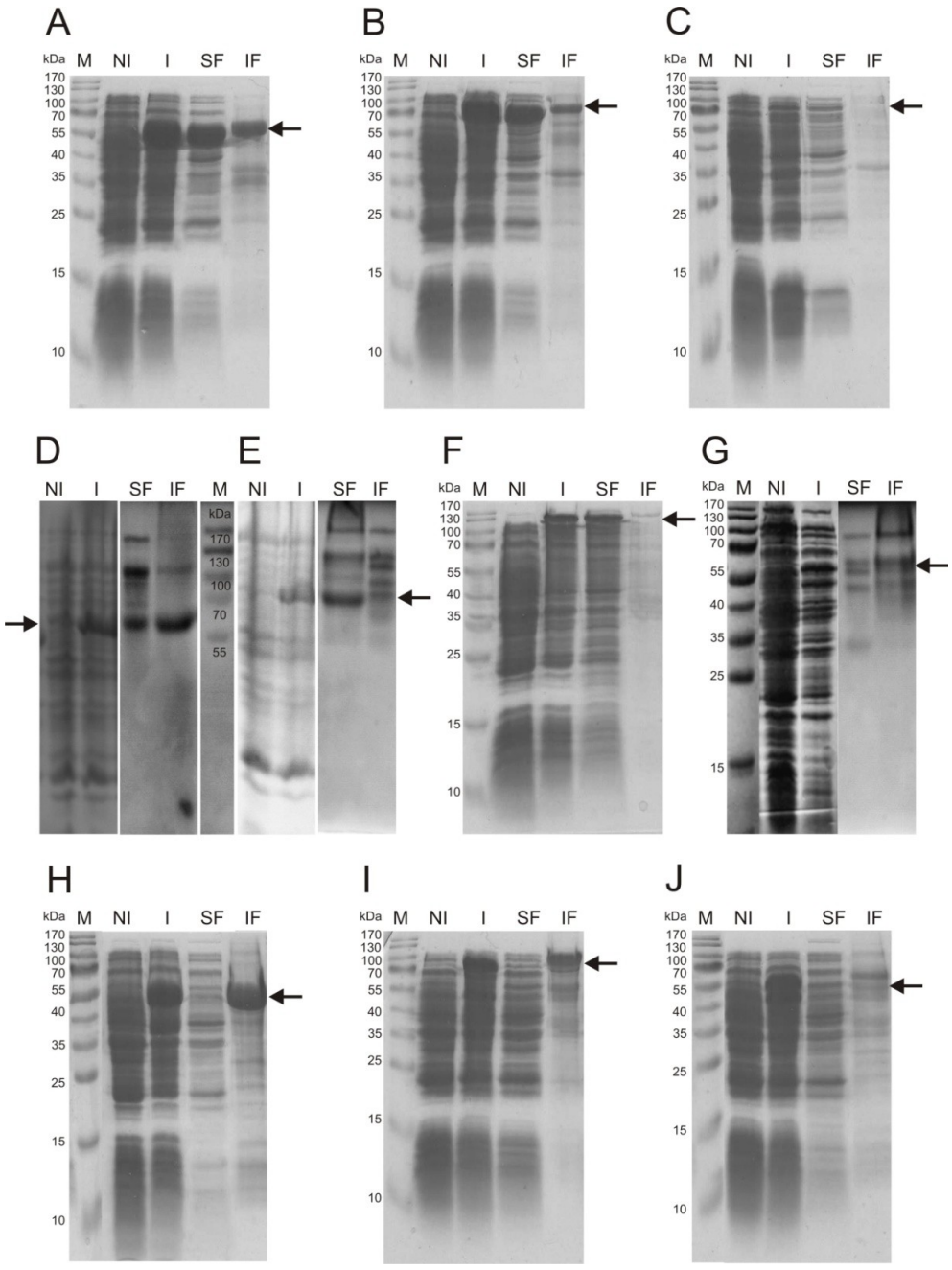


Figure 32. Expression and solubility of different MetRSs in *E. coli*. The MetRSs are indicated with an arrow, with their corresponding theoretical masses in brackets. A) *Aa*MetRS WT (61093.5 Da); B) *Ec*MetRS WT (80398.5 Da); C) *Mj*MetRS WT (77337.9 Da); D) *Ap*MetRS WT (70624.7 Da); E) *Pa*MetRS WT (69896.7 Da); F) *Np*MetRS WT (79924.7 Da); G) *Sa*MetRS WT (70222.9 Da),; H) *Ta*MetRS WT (64681.1 Da); I) *Sc*MetRS WT (87400 Da); and J) *Sc*MetRS TF (6-185Δ; 66825.7 Da). M = MW protein marker; NI = Non-induced cells I = induced cells; SF = Soluble Fraction; IF = Insoluble Fraction

As expected, the bacterial MetRS from *A. aeolicus* and *E. coli* were relatively high concentrations in the soluble form (**Figure 32A/B**). Archaeal MetRS from *A. pernix*, *P. aerophilum*, and *S. acidocaldarius* are similarly well-expressed and soluble (**Figure 32D/E/G**). This is very important, since archaeal MetRSs are type B enzymes. The only question that remains is whether they are active at 37 °C, since all used archaea are hyperthermophylic (optimum grow temperature is from 80 to 100 °C). Archaeal MetRS from *M. jannaschii* was poorly expressed (**Figure 32C**), while MetRS from *N. pharaonis* had an unexpected size (**Figure 32F**). MetRS from *T. acidophylum* was completely insoluble (**Figure 32H**). These patterns might have been caused by the extreme environments in which these organisms normally thrive. For example, *M. jannaschii* is a thermophylic, and has an optimal temperature of 85 °C. *N. pharaonis*, on the other hand, lives optimally in 3.5 M NaCl and at a pH of 8.5; and *T. acidophylum* grows optimally at 56 °C and pH 1.8. Finally, most of the eukaryotic MetRSs could not be expressed in *E. coli* (data not shown, except for the baker´s yeast MetRS or *Sc*MetRS in both WT and truncated forms (6-185Δ), since they were very well expressed, though insoluble; **Figure 32I/J**). The finding that most eukaryotic MetRSs are not even expressed in *E. coli* could be due to the presence of an N-terminal extension. Eukaryotic MetRSs usually contain an N-terminal extension absent in bacterial or archaeal MetRSs (see Appendix 9.4 for more details). This appended region is believed to improve tRNA binding efficiency.[133] Finally, the problem of codon usage can be partly disregarded, since the strains contained plasmids to supplement *E. coli* with rare tRNAs.

3.2.1.2 Purification and analytical characterization

Large-scale LB cultures were used to express soluble MetRSs (see 6.2.2.6 for details). Proteins were subsequently purified to a level approximately between 80 to 95% using Ni-NTA affinity chromatography (see 6.2.2.7 for details). The degree of purity was dependent on the enzyme stability and the yield was in the order of milligrams of protein per liter of culture. As expected, *Ec*MetRS WT was the best-expressed enzyme. The yields for *Ec*MetRS mutants are lower; more mutated

enzymes exhibited both lower stability and a faster precipitation at 4 °C (data not shown). The different *E. coli* mutants where generated because it has been reported that these have a preferential activation toward certain Met analogs. For example, the group of Tirrell demonstrated that mutant *Ec*MetRS L13G is able to activate and efficiently charge azidonorleucine (Anl).[134-136] Similarly, although 6,6,6-trifluoronorleucine (Tfn) does not support significant protein synthesis *in vivo*[137] and *in vitro*[138,139], the same group was able to evolve a triple-mutated *Ec*MetRS that activates Tfn at modest levels.[140] Hence, the single-mutant L13S, double-mutant L13S/Y260L and triple-mutant L13S/Y260L/H301L of *Ec*MetRS were generated and included in the collection to confirm the preferential activities of those Met analogs. The same scheme was followed for the MetRS of *M. jannaschii*, where the mutations from the *E. coli* enzymes were transferred according to primary structure alignment (see Appendix 9.4 for details). Lower yields were obtained from the archaeal MetRSs whose expression was only achieved using *E. coli* strains supplemented with rare tRNAs. Finally, although *Sc*MetRS TF was mainly insoluble, proteins in the order of milligrams could be purified in soluble form. The purification of the enzymes is indicated by the presence of bands with MW of 60 to 80 kDa after 12% SDS-PAGE (**Figure 33**).

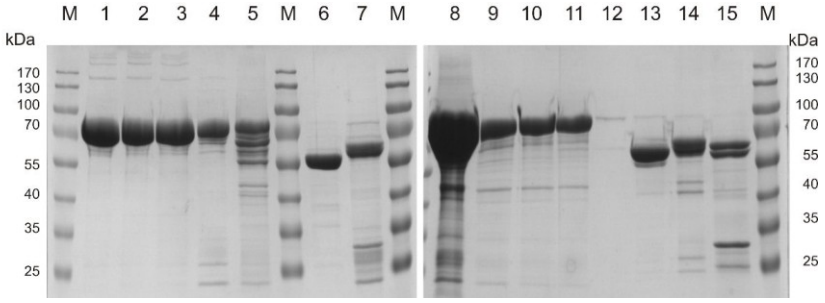


Figure 33. Purification of soluble MetRSs in *E. coli*. Purified MetRSs and their calculated masses: 1. *Ec*MetRS WT (80399.5 Da); 2. *Ec*MetRS L13G (80343.4 Da); 3. *Ec*MetRS L13S (80373.5 Da); 4. *Ec*MetRS L13S/Y260L (80323.5 Da); 5. *Ec*MetRS L13S/Y260L/H301L (80299.5 Da); 6. *Aa*MetRS WT (61094.6 Da); 7. *Sc*MetRS TF (6-185Δ) (66826.7 Da); 8. *Mj*MetRS WT (77338.9 Da); 9. *Mj*MetRS L9G (77282.8 Da); 10. *Mj*MetRS L9S (77312.9 Da); 11. *Mj*MetRS L9S/H254L (77262.8 Da); 12. *Mj*MetRS L9S/H254L/H288L (77238.9 Da); 13. *Ap*MetRS WT (70625.7 Da); 14. *Pa*MetRS WT (69897.7 Da); 15. *Sa*MetRS WT (70223.9 Da). In all cases a defined protein amount was loaded in a 12% SDS-polyacrylamide gel. M = MW protein marker.

After purification, samples were further characterized using ESI-MS (**Figure 34**).

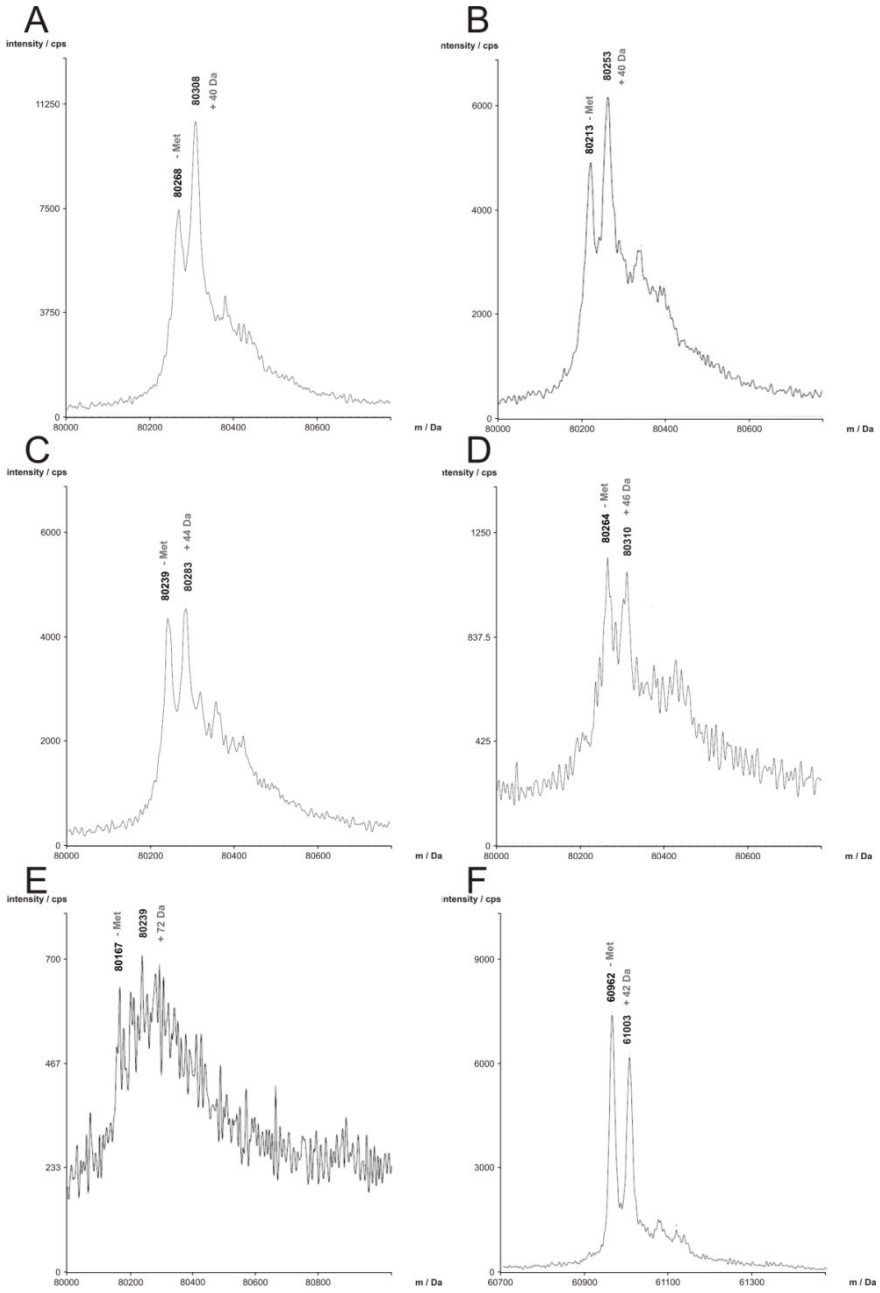


Figure 34. ESI-MS spectra of purified MetRSs. Each MetRS peak is labeled with the determined masses; theoretically calculated masses minus the Met excision are underlined, and indicated parenthetically: A) *Ec*MetRS WT (80399.5 – 131.1 = 80268.3 Da); B) *Ec*MetRS L13G (80343.4 – 131.1 = 80212.2 Da); C). *Ec*MetRS L13S (80373.5 – 131.1 = 80242.3 Da); D) *Ec*MetRS L13S / Y260L (80323.5 – 131.1 = 80192.3 Da); E). *Ec*MetRS L13S / Y260L / H301L (80299.5 – 131.1 = 80168.3 Da); F). *Aa*MetRS WT (61094.6 – 131.1 = 60963.5 Da).

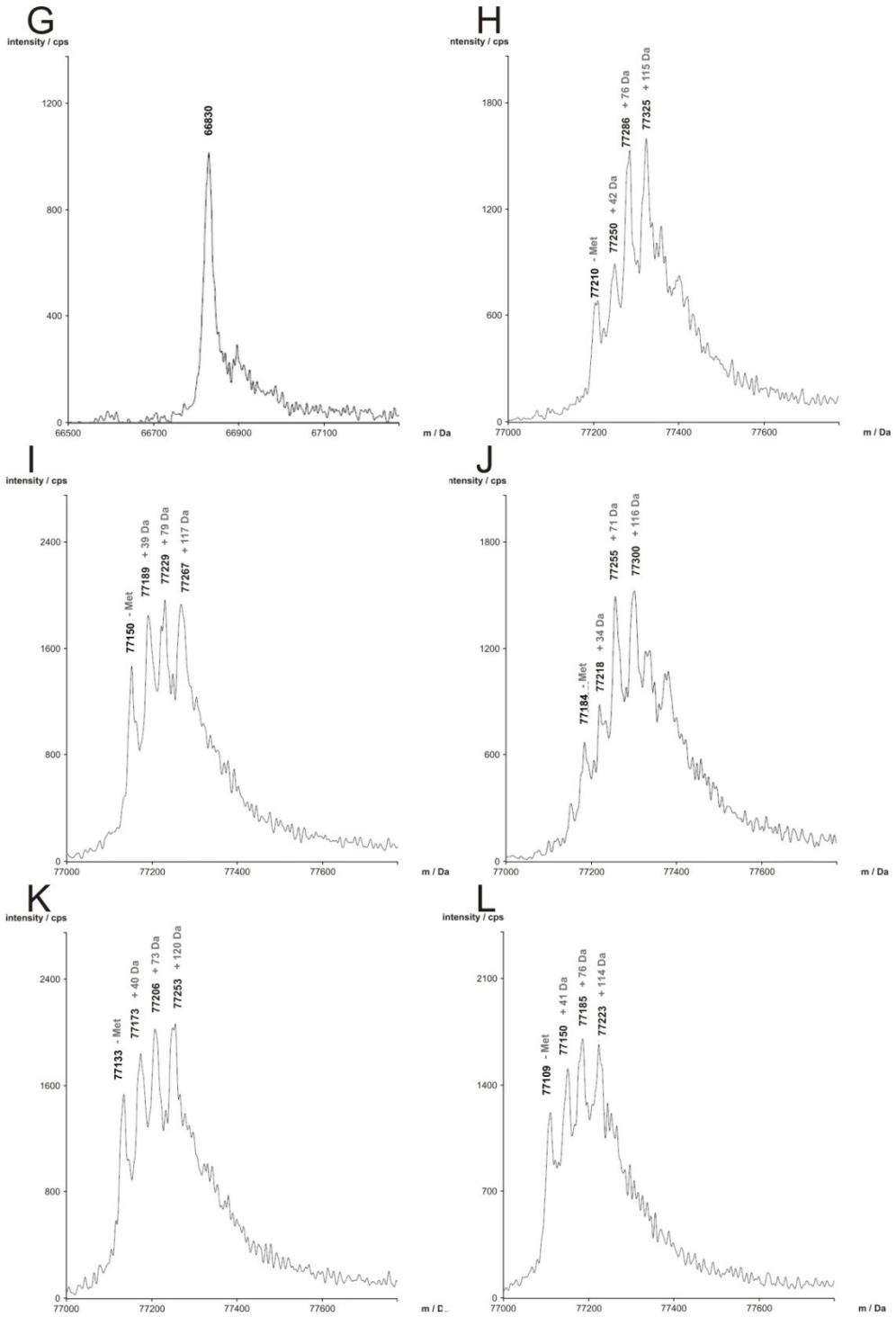


Figure 34. ESI-MS spectra of purified MetRSs (continuation). G) *Sc*MetRS TF (1-185Δ/736-751Δ) (66826.7 Da, no Met excision); H) *Mj*MetRS WT (77338.9 – 131.1 = 77207.7 Da); I) *Mj*MetRS L9G (77282.8 – 131.1 = 77151.6 Da); J) *Mj*MetRS L9S (77312.9 – 131.1 = 77181.7 Da); K) *Mj*MetRS L9S / H254L (77262.8 – 131.1 = 77131.6 Da); L) *Mj*MetRS L9S / H254 L /H288L (77238.9 – 131.1 = 77107.7 Da).

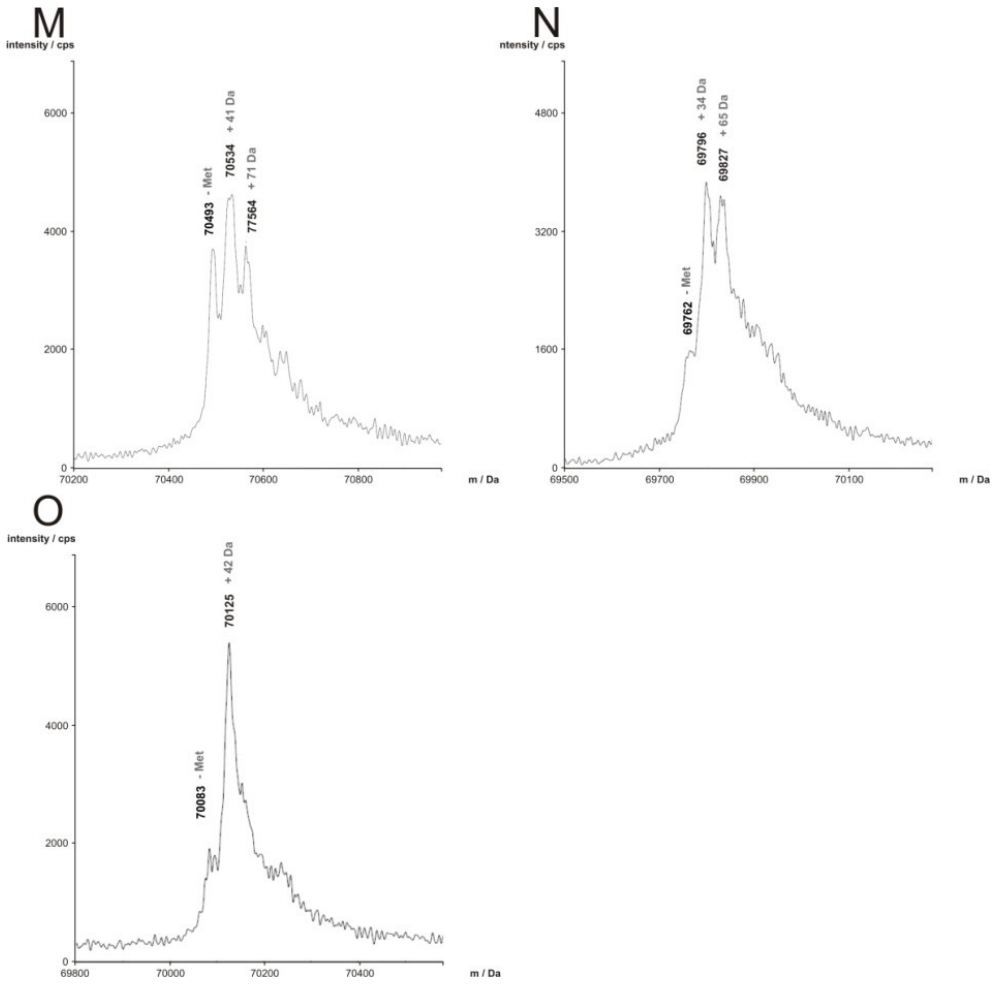


Figure 34. ESI-MS spectra of purified MetRSs (continuation). M) *Ap*MetRS WT (70625.7 – 131.1 = 70494.5 Da); N) *Pa*MetRS WT (69897.7 – 131.1 = 69766.5 Da); O) *Sa*MetRS WT (70223.9 – 131.1 = 70092.7 Da).

In all cases the *N*-terminal Met was excised according to the N-terminal rules[141], except for *Sc*MetRS TF whose second residue is not a bulky amino acid. Thus, 131.1 Daltons (the equivalent of one Met forming a peptide bond) were subtracted to the theoretical calculated mass. The determined masses for all the MetRSs nearly corresponded with the theoretical ones, i.e. in addition to the most prominent expected peak, all preparations revealed another peak of similar size, with an additional 34-46 Da; these are protein adducts of unknown nature (**Figure 34**). More mutated enzymes were less stable, as indicated by the low yield, intensity, and suboptimal spectra display of *Ec*MetRS (**Figure 34D/E**) and *Mj*MetRS (**Figure 34K/L**).

The preparations from *M. jannaschii* had several peaks with *ca.* 71 and 120 Da protein adducts of similarly unknown nature (**Figure 34H/I/J**). In contrast, the rest of the archaeal MetRSs showed spectra with better quality (**Figure 34M/N/O**).

3.2.1.3 *In vitro* cross-aminoacylation experiments

On MetRSs purification and characterization, their activities were measured by using radioactive Met and commercially available *E. coli* initiator $tRNA^{fMet}$ and elongator $tRNA^{Met}$ (present in bulk tRNA; see 6.2.3.1 for details). The methionylation extent for both tRNA preparations by all enzymes is shown in **Figure 35**.

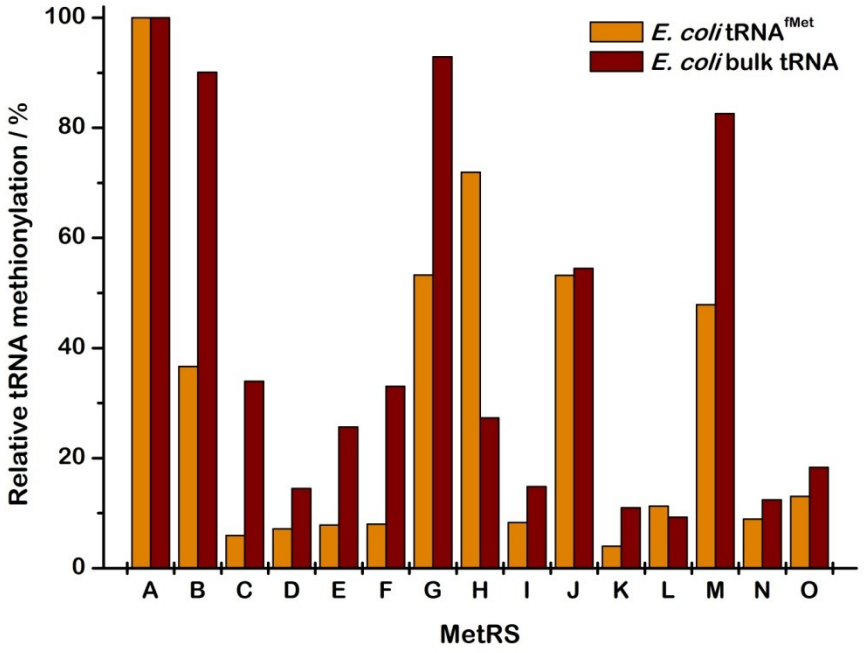


Figure 35. Cross-aminoacylation experiments. The charging efficiency for ^{35}S-Met onto initiator $tRNA^{fMet}$ and bulk tRNA was measured at 37 °C after 20 min using the same concentration of A) *Ec*MetRS WT; B) *Ec*MetRS L13G; C) *Ec*MetRS L13S; D) *Ec*MetRS L13S/Y260L; E) *Ec*MetRS L13S/Y260L/H301L; F). *Aa*MetRS WT; G) *Sc*MetRS TF (6-185Δ); H) *Mj*MetRS WT; I) *Mj*MetRS L9G; J) *Mj*MetRS L9S; K) *Mj*MetRS L9S/H254L; L) *Mj*MetRS L9S/H254L/H288L; M) *Ap*MetRS WT; N) *Pa*MetRS WT; O) *Sa*MetRS WT.

The most active enzyme toward both tRNA preparations was expectedly *Ec*MetRS WT (**Figure 35A**). These values were used as a reference to determine the degree of methionylation for both the different mutants and the different enzymes, which could also yield clues as to its orthogonality in *E. coli.* The *Ec*MetRS mutants displayed less activity toward initiator $tRNA^{fMet}$ than elongator $tRNA^{Met}$ but these activities diminished when the number of mutations increased (**Figure 35B-E**). This findings support the idea that mutations in the MetRS binding pocket may increase the specificity for other Met analogs, but compromise tRNA recognition as a tradeoff. These mutations could be therefore more detrimental than beneficial for MetRS optimal activity. The other bacterial enzyme whose affinity for both tRNA preparations was not significant at 37°C was *Aa*MetRS (**Figure 35F**); however, more than 75% activity for bulk *E.coli* tRNA was detected at 70 °C (data not shown). Thus it could be confirmed that *Aa*MetRS is thermophilic and cross-reacts with *E. coli* tRNAs under suitable conditions. Likewise, other enzymes cross-reacted to some extent with both tRNA preparations, particularly *Sc*MetRS TF (**Figure 35G**), *Mj*MetRS WT (**Figure 35H**), *Mj*MetRS L9S (**Figure 35J**), and *Ap*MetRS WT (**Figure 35M**). Unexpectedly, *Mj*MetRS WT and L3S still showed some activity despite the use of noncognate tRNAs (**Figure 35H/J**), but the double- and triple-mutated enzymes were not as active as the single-mutated or wt (**Figure 35I/K/L**). These results indicate that the eukaryotic MetRS from yeast, as well as the two archaeal MetRS from *M. jannaschii* and *A. pernix* are not suitable for use in *E. coli*. This was not expected since these MetRS are considered as type B. Nonetheless, the less active enzymes with respect to both $tRNA^{Met}$ isoacceptors were the type B MetRS from *P. aerophylum* (**Figure 35N)** and *S. acidocaldarius* (**Figure 35O).** These two archaeal enzymes could be considered as candidate orthogonal MetRSs, since they did not aminoacylate at all *E. coli* tRNAs. The only prerequisite would be that they are as active as *Ec*MetRS at 37 °C.

3.2.1.4 *In vitro* activation of methionine analogs

Since it would obviously be preferred by both the host and all other MetRSs, Met must be depleted from the system to so that Met analogs will be preferentially

incorporated by each MetRS. It is imperative though, that there are differential preferences in terms of charging two different Met analogs, as well as in the subsequent transfer of the aminoacylated parts onto the respective initiator or elongator tRNAs, by both the host and orthogonal MetRS. For this, the different Met analogs were either purchased or synthesized (**Figure 36**; see 6.1.3 for details).

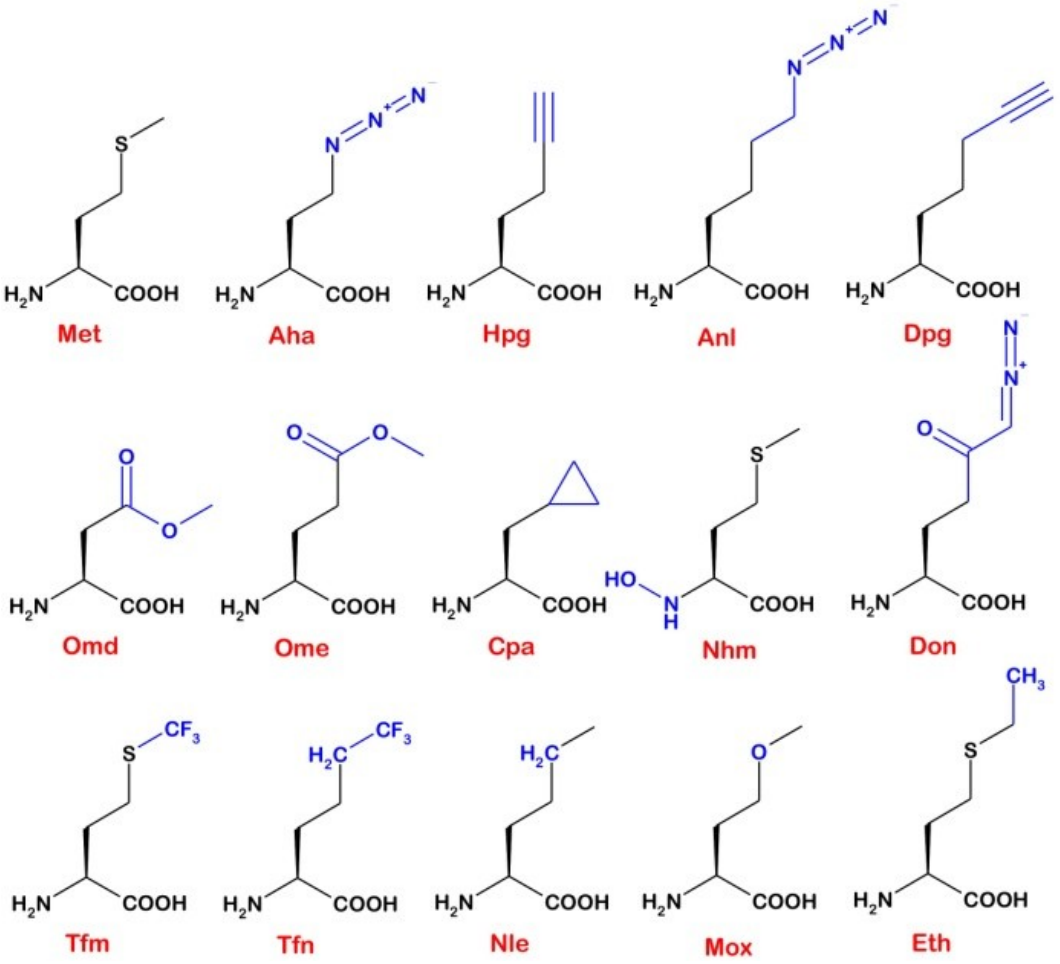


Figure 36. Met analogs. The synthetic amino acids contain chemical groups with single atom substitutions and large side-chains (blue). Met, Methionine; Aha, Azidohomoalanine; Hpg, Homopropargylglycine; Anl, Azidonorleucine; Dpg, Dihomopropargylglycine; Omd, aspartic acid β-methyl esther; Ome, gutamic acid γ-methyl ester; Cpa, β-cyclopropyl-alanine; Nhm, *N*-hydroxy-methionine; Don, 6-diazo-5-oxo-norleucine; Tfm, 5,5,5-trifluoromethionine; Tfn, 6,6,6,-trifluoronorleucine; Nle, norleucine; Mox, methoxinine; and Eth, ethionine.

The substrate specificities of *Ec*MetRS, *Pa*MetRS and *Sa*MetRS for the different Met analogs were subsequently determined using the isotopic ATP:PPi exchange assay (**Figure 37**; see 6.2.3.2 for details).[142]

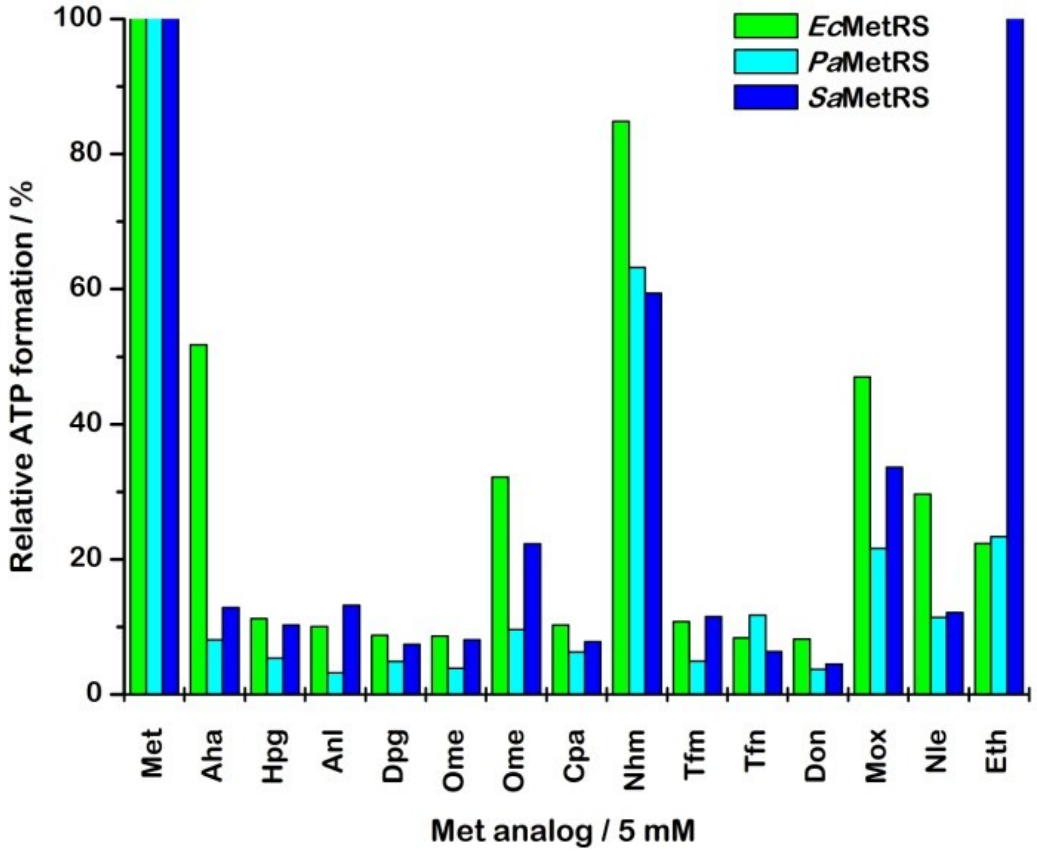


Figure 37. Activation of Met and its analogs by host and 'orthogonal' MetRSs. The activation efficiency of *Ec*MetRS, *Pa*MetRS and *Sa*MetRS for several Met analogs was measured at physiological condition.

Importantly, the only case of differential activation of two different Met analogs by the host *Ec*MetRS and *Sa*MetRS was found for Aha and Eth (**Figure 37**). Here, *Ec*MetRS preferred Aha over Eth, while *Sa*MetRS preferred Eth over Aha. Otherwise, no differential activation for the rest of the Met analogs by both MetRSs was found (**Figure 37**). To rule out any experimental artifact, the activation profiles for Aha and Eth by *Ec*MetRS and *Sa*MetRS were reevaluated at 37 °C (**Figure 38**).

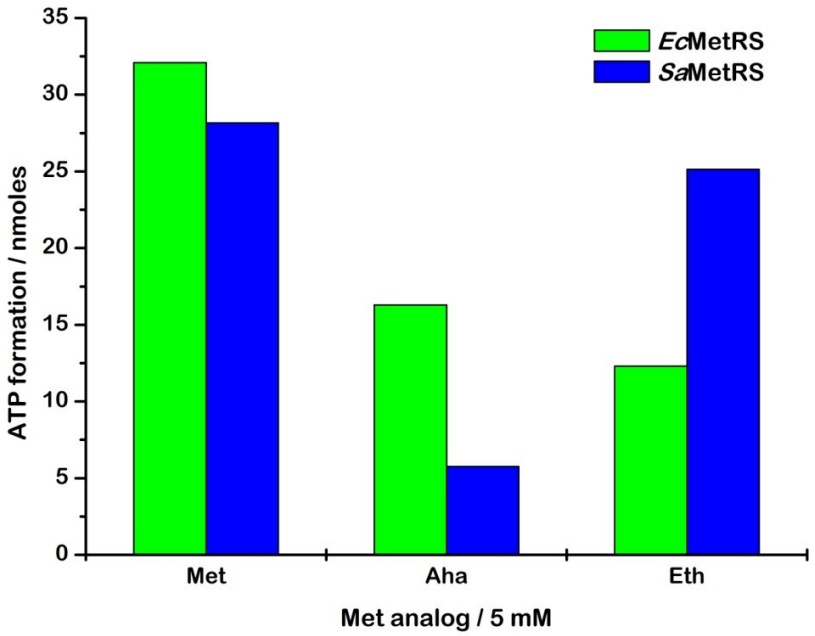


Figure 38. Differential activation of Met analogs by host and 'orthogonal' MetRS. The activation efficiency by *Ec*MetRS and *Sa*MetRS towards Met, Aha, and Eth was measured at 37 °C after 20 min. The bars are depicted without normalization. *Ec*MetRS displayed for Aha 16.3 nmoles and for Eth 12.3 so the ratio of preference for Aha over Eth is 16.3 / 12.3 = 1.3 whereas *Sa*MetRS prefers Eth over Aha for about 25.13 / 5.75 = 4.4.

Strikingly, the *Sa*MetRS showed 4.4 times more enhanced substrate specificity for Eth than Aha, whereas *Ec*MetRS exhibited a 1.3 fold better activation of Aha than Eth (**Figure 38**). Additionally, *Sa*MetRS has very similar activation efficiency for both Met and Eth. This is not surprising for *S. acidocaldarius*, since this extremely thermophilic archaea found in geothermal habitats is a sulfur-metabolizing organism with strains that can even growth in the presence of Eth.[143] On the other hand, Eth is a very toxic amino acid because it inhibits growth in many microorganisms including *E. coli*.[144] In any case, the finding that Aha and Eth are differentially recognized by both the host *Ec*MetRS and the orthogonal *Sa*MetRS is very important since this is a prerequisite for the differential reassignment of AUG codons. These findings are significant but were found *in vitro*. It was thus important to study the *in vivo* preferences of *E. coli* for both Aha and Eth. To this end, a model protein containing both 1 and 2 AUG codons located at appropriate sites was used.

3.2.2 Incorporation of Aha and Eth into barstar 1M

Barstar is an intracellular single-domain inhibitor of the extracellular ribonuclease barnase from *Bacillus amyloliquefaciens, a* protein normally secreted to the environment to degrade foreign RNA as a defense mechanism. The structure of barstar is characterized by a β-α-β motif, typical for a nucleic acid binding proteins. It is widely used in folding studies as a small recombinant protein of 90 amino acids.[145] The barstar mutant P28A/C41A/C83A (ψ-b* from the *Greek* is abbreviated as b*1M for clarity reasons; see Appendix 9.2 for sequence details) was used to estimate the natural preference for Aha and Eth at the N-terminus *in vivo* (**Figure 39**).

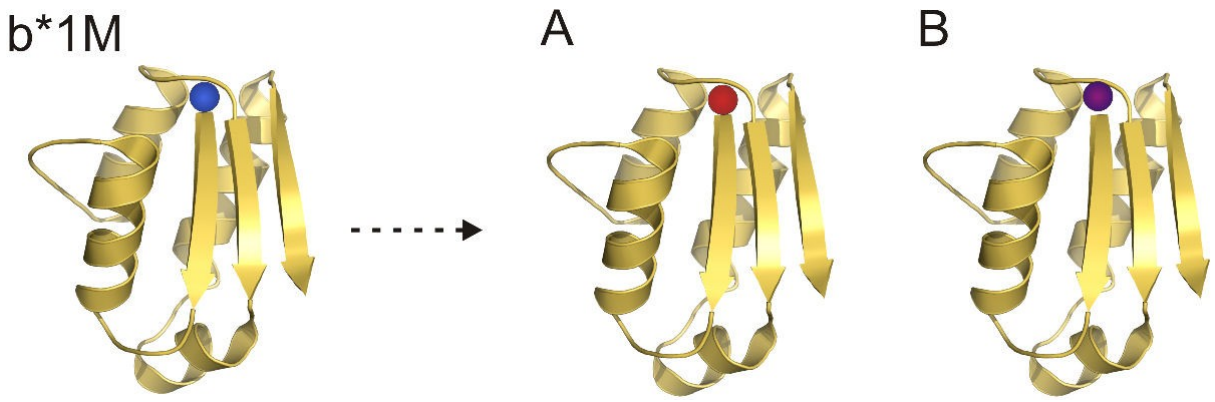


Figure 39. Tandem incorporation of Aha and Eth into barstar 1M. The N-terminal Met (blue) was replaced by Aha (red) or Eth (purple) to generate two congeners: (A) b*1M[Aha] and (B) b*1M[Eth].

This barstar mutant lacks complications arising from oxidation of cysteines or re-folding, which occur in the WT.[146]

E. coli Met auxotrophic strains were transformed with the barstar gene containing only one AUG codon at the starting position and supplemented with a limiting amount of Met. After Met depletion, the cells were transferred to NMM containing different amounts of Aha and Eth (see 6.2.2.8 for details). Before expression, six different conditions were established, i.e. Met, Aha, and Eth (as positive controls), and

different ratios of Aha:Eth (3:1, 1:1, 1:3) to assess if the incorporation of the Met analogs is concentration dependent. Under these conditions, three different b* congeners containing one Met, one Aha, and one Eth were produced in single incorporation experiments, while two congeners in different amounts containing either Aha or Eth at the N-terminus were generated in the tandem incorporations. The expression and purification of b* congeners was indicated by a corresponding band with an average mass of approx. 10 kDa (**Figure 40**).

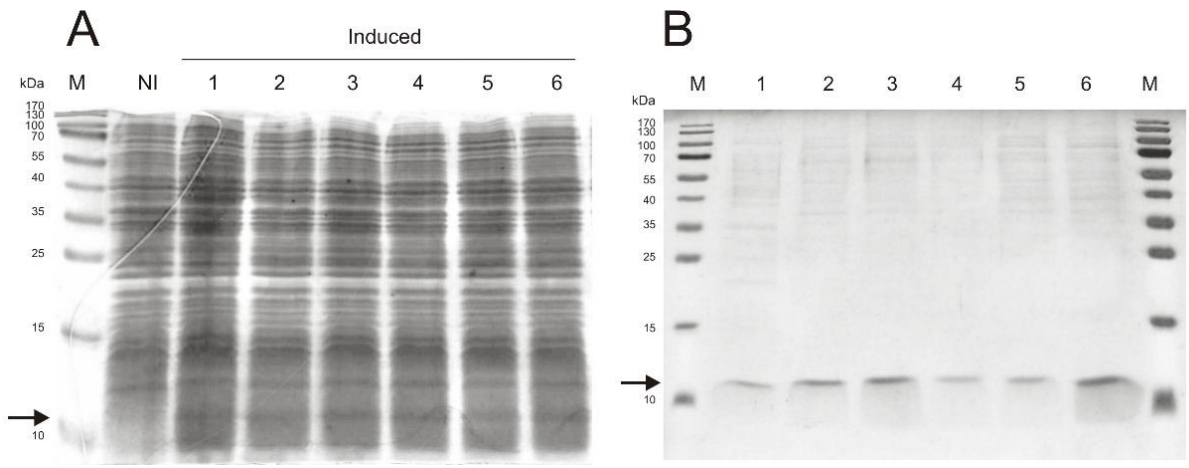


Figure 40. Barstar 1M alloproteins. The b*1M variants were produced through the use of medium supplementats containing 1) Met; 2) Aha; 3) Aha>Eth (3:1); 4) Aha=Eth (1:1); 5) Aha<Eth (1:3); and 6) Eth. These are indicated by arrows in profiles obtained after expression (A) and purification (B) using 17% SDS-polyacrylamide gels and Coomassie staining. M = Standard MW bands; NI = Non-induced culture.

After expression and purification, all b*1M alloproteins were analyzed by ESI-MS and N-terminal sequencing to determine its mass and the amount of Aha or Eth at the N-terminus (**Figure 41**).

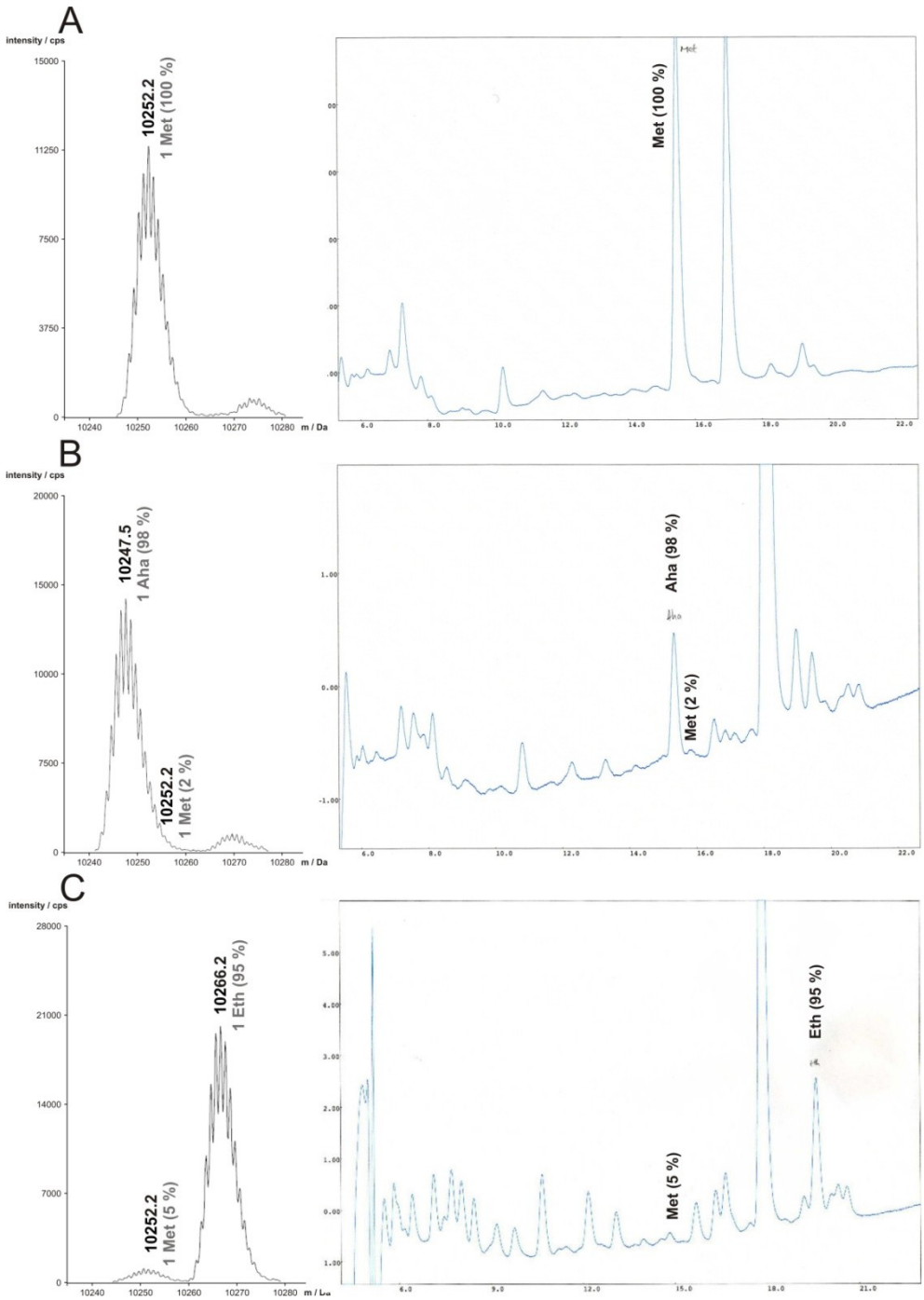


Figure 41. Analytical characterization of barstar 1M congeners. The mass and the N-terminal-sequencing spectra for b*1M supplemented with A) Met; B) Aha; C) Eth are shown in the left and right columns, respectively. In both cases, the combined intensities of the corresponding peaks sum up to 100%, from which each species percentage was calculated and annotated. For ESI-MS, the calculated mass for b* with 1 Met is 10252.6 Da; with 1 Aha is 10247.6 Da; and with 1 Eth is 10266.8 Da. For N-terminal sequencing, Aha and Eth showed retention times at 15 and 19 units, respectively.

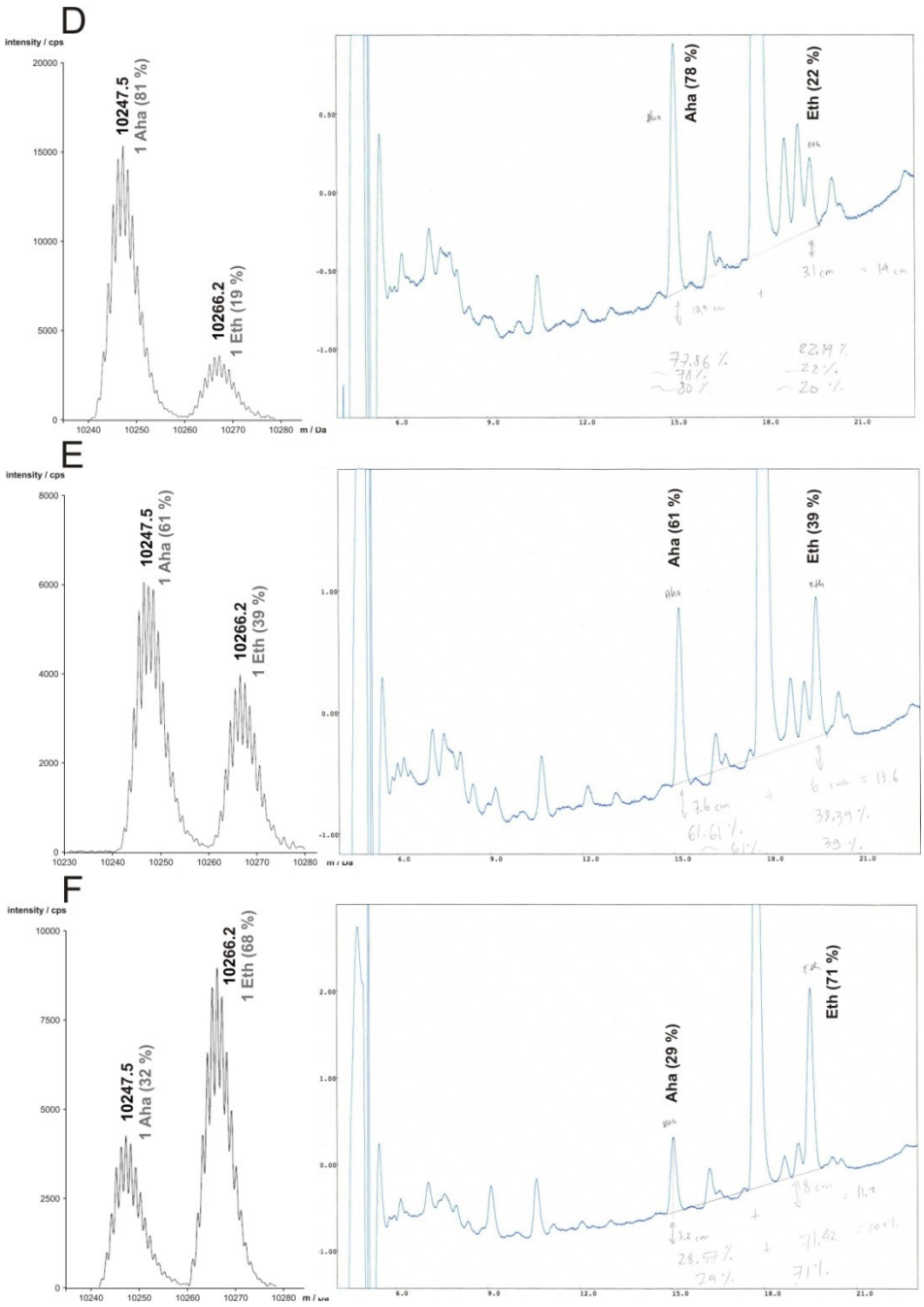


Figure 41. Analytical characterization of barstar 1M congeners (continuation). The mass and the N-terminal-sequencing spectra for b*1M supplemented with D) Aha > Eth (3:1); E) Aha = Eth (1:1); and F) Aha < Eth (1:3) are shown in the left and right columns, respectively. In both cases, the combined intensities of the corresponding peaks sum up to 100%, from which each species percentage was calculated and annotated. For ESI-MS, the calculated mass for barstar with 1 Met is 10252.6 Da; with 1 Aha is 10247.6 Da; and with 1 Eth is 10266.8 Da. For N-terminal sequencing, Aha and Eth show retention times at 15 and 19, respectively.

The congeners that only contained Met, Aha, or Eth at the N-terminus displayed a prominent peak at the expected mass (**Figure 41A/B/C**; left panel). Importantly, the N-terminal sequencing results are tightly correlated with the ESI-MS results (**Figure 41A/B/C**; right panel). When the medium was supplemented with different amounts of Aha and Eth, two peaks of different intensities, corresponding to each of the congeners, were found (**Figure 41D/E/F**; left panel). The calculated proportions of the peaks from the ESI-MS spectra were also very similar to those from N-terminal sequencing (**Figure 41D/E/F**; right panel). One can say that both ESI-MS and N-terminal sequencing methods are reliable for accurately determining the incorporation of Aha and Eth into barstar. Furthermore, the summary of the results obtained from both methods in **Table 3** provides an easy way for comparing the natural preferences in terms of ratios of Aha and Eth for the N-terminus of the different b*1M variants.

Table 3. Natural preference of Aha and Eth at the N-terminus of barstar 1M alloproteins.

Amino acids supplemented	*Method*	*N-terminus (%)*		*Ratio*
		Aha	**Eth**	
Aha / Eth (0.75 / 0.25)	ESI-MS	81	19	81 / 19 = **4.0 for Aha**
	N-terminal sequencing	78	22	78 / 22 = **3.5 for Aha**
Aha / Eth (0.50 / 0.50)	ESI-MS	61	39	61 / 39 = **1.5 for Aha**
	N-terminal sequencing	61	39	61 / 39 = **1.5 for Aha**
Aha / Eth (0.25 / 0.75)	ESI-MS	32	68	68 / 32 = **2.0 for Eth**
	N-terminal sequencing	29	71	71 / 29 = **2.5 for Eth**

Interestingly, it was found Aha was found to be incorporated fourfold higher at the N-terminus compared to Eth when it is four times more abundant in the medium than (**Table 3**). On the other hand, when there was four times more Eth in the medium than Aha to the medium, Eth incorporation at the N-terminus was only twice as more than Aha (**Table 3**). This indicates that the extent of incorporation of Aha or Eth at the N-terminus of b*1M could be fine-tuned by changing the concentration of the analogs in the media. Moreover, if one considers the experiment with both analogs present in

equal amounts (**Table 3**), it is plausible to suggest that the *Ec*MetRS exhibited *in vivo* a 1.5 fold Aha-tRNA^{fMet} to Eth-tRNA^{fMet} charging preference *in vivo*. This finding is consistent with previous *in vitro* data where *Ec*MetRS showed a 1.3 fold increase in activation of Aha as compared to Eth (**Figure 38**).

These findings suggest that the activation data of an aminoacyl-tRNA synthetase for a given amino acid not only reflects the charging value with respect to its cognate tRNA, but also its incorporation efficiency into proteins, a very common observation described in the literature.[36] Additionally, it is assumed that the noncanonical amino acids are efficiently up taken by the cells and accumulated several-fold inside them; this factor has also been well documented even for Met analogs in *E. coli*.[147] What would be interesting now would be to assess if there is a preference for Aha or Eth at start or internal AUG codons.

3.2.3 Incorporation of Aha and Eth into barstar 2M

To estimate the natural incorporation preference for Aha and Eth of *E. coli in vivo*, the model protein barstar containing one AUG codon at the start position and at position 47 was used (P28A/C41A/E47M/C83A; b*2M; see Appendix 9.3 for sequence details). Prior to protein expression, the same six different conditions as in b*1M were used (see 6.2.2.8 for details), i.e. Met; Aha; and Eth (as positive controls), and different ratios of Aha:Eth (3:1, 1:1, 1:3) to assess if the incorporation of the Met analogs is concentration-dependent. Under these conditions, three different b*2M congeners containing two Met, two Aha, and two Eth were expressed in single incorporation experiments and four congeners containing either Aha or Eth at the N-terminus or internal positions in different amounts were produced in the following tandem incorporations: 1) b*2M containing two Aha residues (**Figure 42A**); 2) b*2M containing one Aha at the N-terminus and one Eth at the internal position (**Figure 42B**); 3) b*2M containing one Eth at the N-terminus and 1 Aha at the internal position (**Figure 42C**); and 4) b*2M containing two Eth residues (**Figure 42D**).

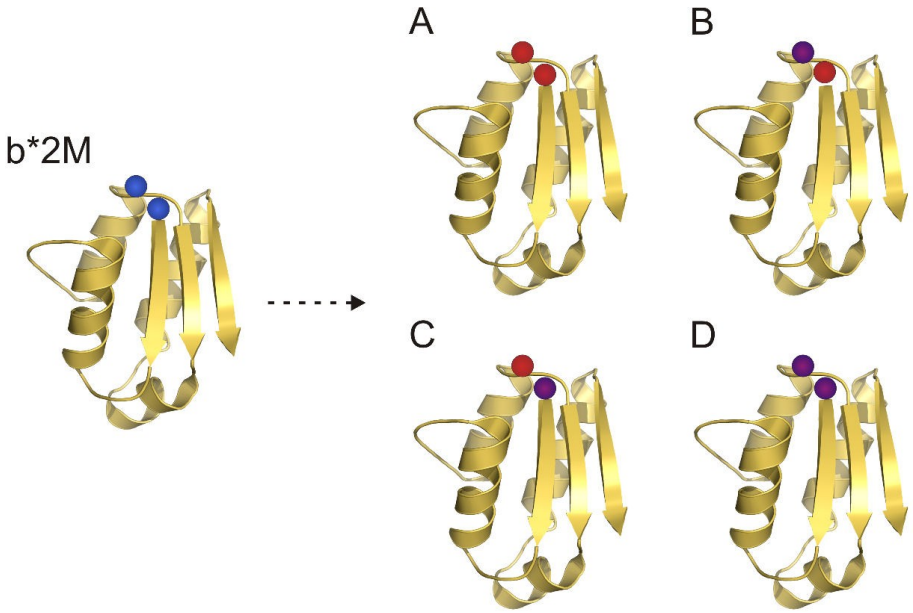


Figure 42. Tandem incorporation of Aha and Eth into barstar 2M. The N-terminal and internal Met (blue) at position 46 from b*2M was replaced by Aha (red) and/or Eth (purple) yielding four different congeners: A) b*2M[Aha1/Aha47]; B) b*2M[Aha1/Eth47]; C) b*2M[Eth1/Aha47]; or D) b*2M[Eth1/Eth47].

The expression and successful purification of b*2M congeners was indicated by a corresponding band with an average mass of 10 kDa (**Figure 43**). Thereafter, all b*2M alloproteins were analyzed by ESI-MS and N-terminal sequencing (**Figure 44**).

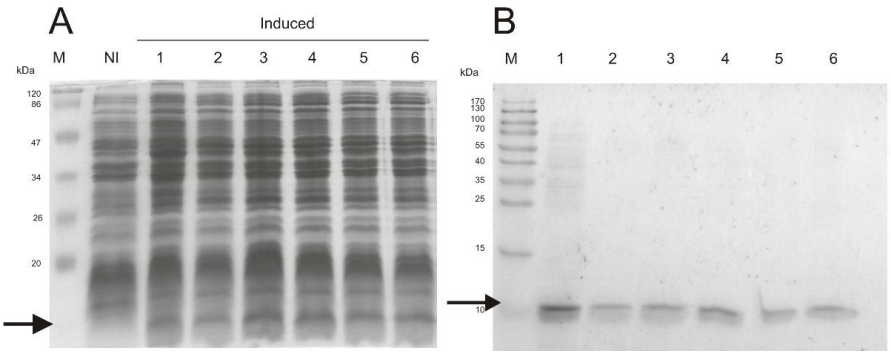


Figure 43. Barstar 2M alloproteins. The b*2M congeners produced in six different conditions (supplemented with: 1) Met; 2) Aha; 3) Aha>Eth; 4) Aha=Eth; 5) Aha<Eth; and 6) Eth) are shown with an arrow after expression (A) and purification (B) using a 17% SDS-polyacrylamide gel and Coomassie staining. M = Standard MW bands; NI = Non-induced culture.

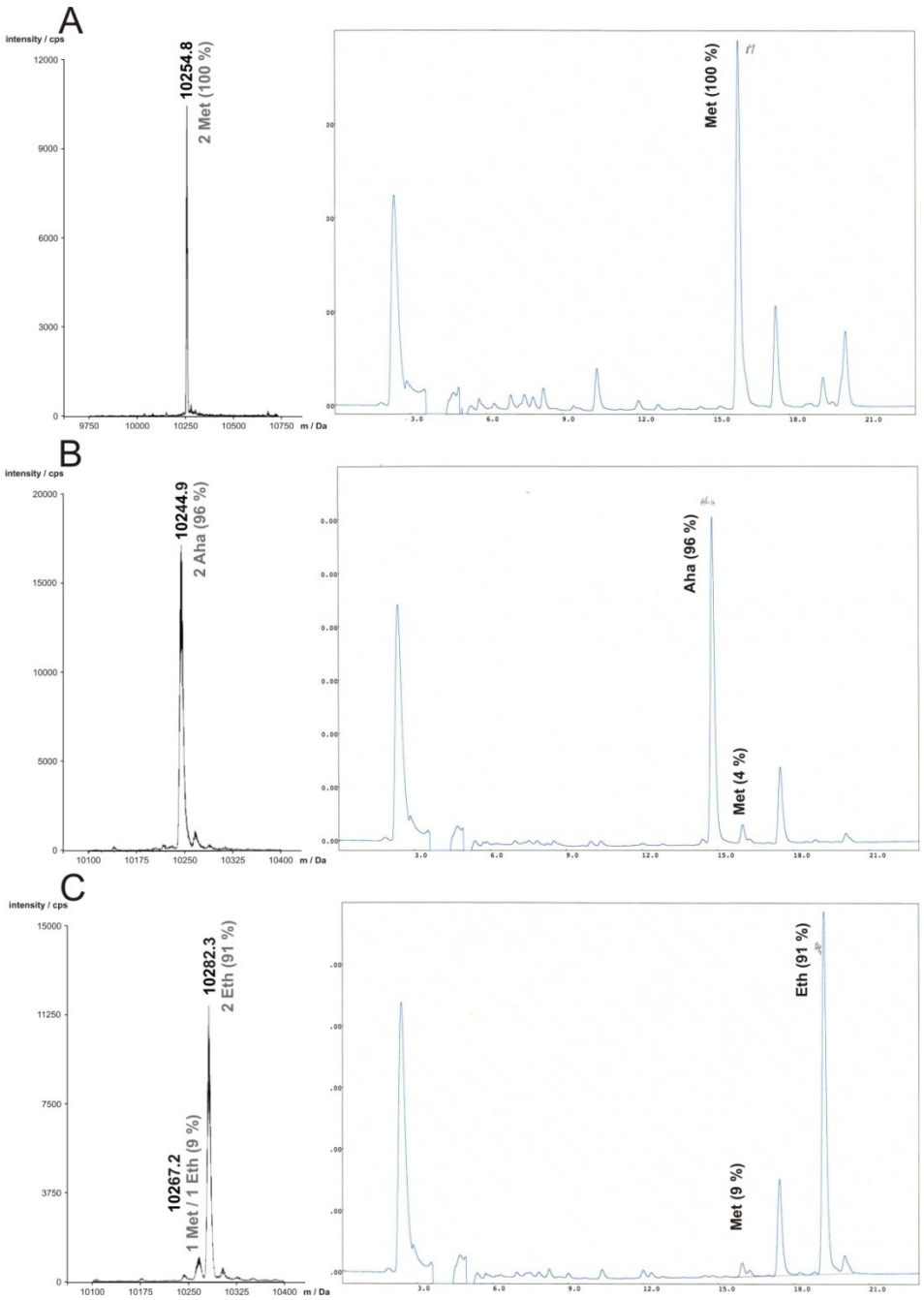


Figure 44. Analytical characterization of barstar 2M congeners. The mass and the N-terminal-sequencing spectra for b*2M supplemented with A) Met; B) Aha; and C) Eth are shown in the left and right columns, respectively. In both cases, the combined intensities of the corresponding peaks sum up to 100%, from which each species percentage was calculated and annotated. For ESI-MS, the theoretical calculated mass for barstar with two Met residues is 10254.5 Da; with 2 Aha residues is 10244.5 Da; and with two Eth residues is 10282.9 Da. For N-terminal sequencing, Aha and Eth show distinctive retention times at 15 and 19, respectively.

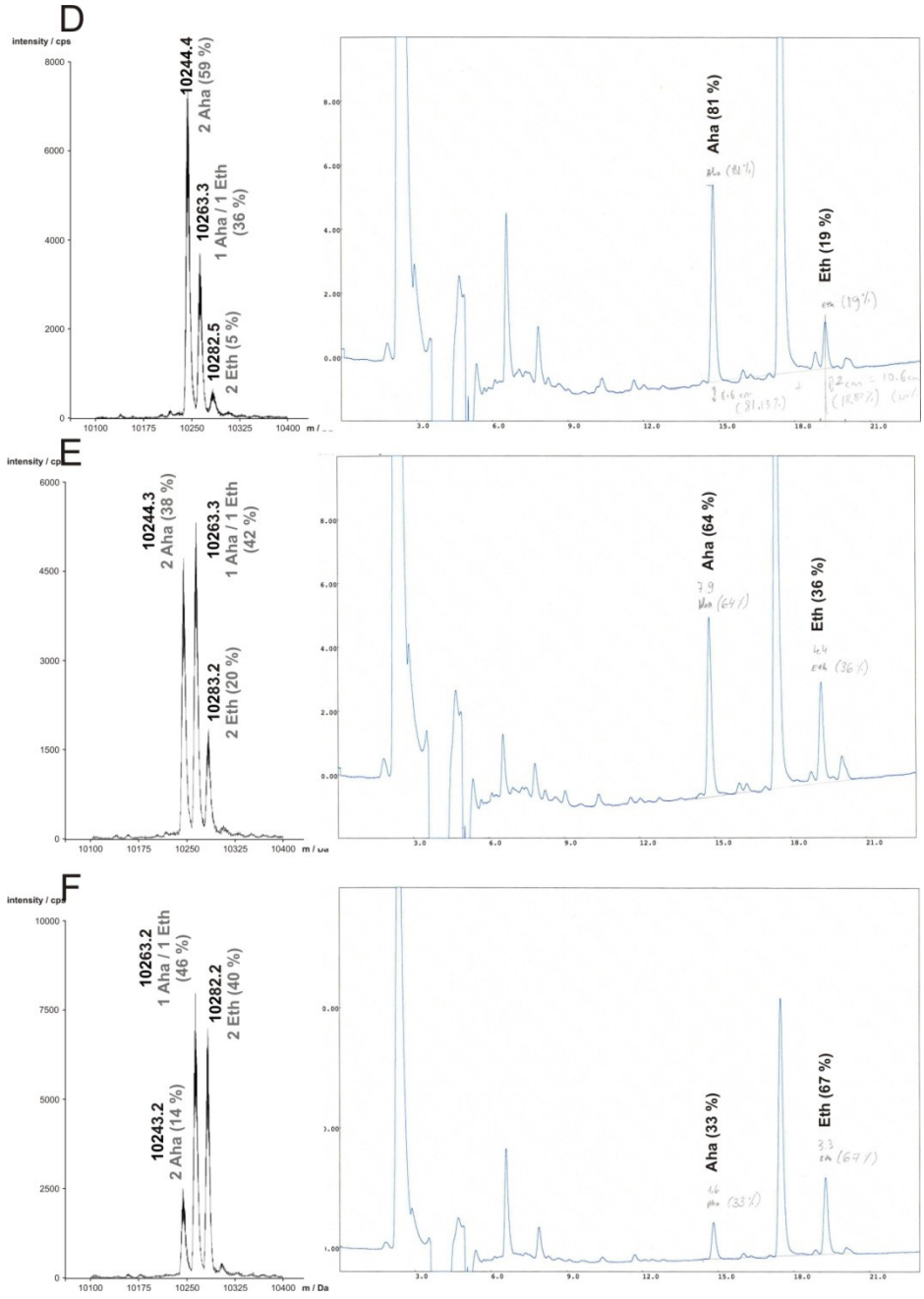


Figure 44. Analytical characterization of barstar 2M congeners (continuation). The mass and the N-terminal-sequencing spectra for b*2M supplemented with D) Aha > Eth (3:1); E) Aha = Eth (1:1); and F) Aha < Eth (1:3) are shown in the left and right columns, respectively. In both cases, the combined intensities of the corresponding peaks sum up to 100%, from which each species percentage was calculated and annotated. For ESI-MS, the theoretical calculated mass for barstar with two Aha residues is 10244.5 Da; with one Aha and one Eth is 10263.7 Da; and with two Eth residues is 10282.9 Da. For N-terminal sequencing, Aha and Eth show distinctive retention times at 15 and 19, respectively.

The congeners with Met, Aha, or Eth at both the N-terminus and internal position displayed a prominent peak where it is expected (**Figure 44A/B/C**; left panel). N-terminal sequencing results are similar to those obtained from ESI-MS in terms of peaks and intensities (**Figure 44A/B/C**; right panel) and could be considered reliable. However, when different concentrations of Aha and Eth were used to supplement the medium, three peaks of different intensities corresponding to the four variants were found (**Figure 44D/E/F**; left panel). The first and third peaks correspond to the Aha and Eth double-substituted b*2M congeners, respectively, while the second peak corresponds to two congeners – one with Aha at the N-terminus and Eth at the internal position and the other exhibiting the opposite distribution. To determine the exact amount of Aha and Eth in each of these variants, N-terminal sequencing was necessary (**Figure 44D/E/F**; right panel). **Table 4** provides an easy way for comparing natural preferences in terms of ratios of both Aha and Eth in the N-terminus and internal position of the different b*2M congeners.

Table 4. Natural preference for Aha and/or Eth at the N-terminus and/or position 47 of barstar 2M.

Amino acids supplemented	***Distribution of Aha and Eth (%)****				***Ratio****	
	A *1Aha* *47Aha*	*B* *1Aha* *47Eth*	*C* *1Eth* *47Aha*	*D* *1Eth* *47Eth*	*N-terminus*	*Internal*
Aha / Eth **(0.75 / 0.25)**	59	22	14	5	**4 for Aha**	**2.7 for Aha**
Aha / Eth **(0.50 / 0.50)**	38	26	16	20	**1.8 for Aha**	**1.2 for Aha**
Aha / Eth **(0.25 / 0.75)**	14	19	27	40	**2 for Eth**	**1.4 for Eth**

*The percentage of b*2M containing two Aha residues (A) or two Eth residues (D) was determined by ESI-MS. To calculate the distribution of Aha at the N-terminus/Eth at the internal position (B) and Eth at the N-terminus/Aha at the internal position (C), the percentage obtained from N-terminal sequencing results(N) was subtracted from the ESI-MS percentage, e.g. when Aha / Eth (0.50 / 0.50), N is 64% for Aha and 36% for Eth, so 64 – 38 = 26 % and 36 – 20 = 16%

**The ratio of preference for the N-terminus was determined with the formula N = A + B / C + D, e.g. 38 + 26 / 16 + 20 = 1.8 when Aha / Eth (0.50 / 0.50); and that for the Internal 47 position with I = A + C / B + D, e.g. 38 + 16 / 26 + 20 = 1.2 when Aha / Eth (0.50 / 0.50).

When there was four times more Aha than Eth in the medium, it was found that Aha occurs at the N-terminus and the internal position 4 and 2.7 times more, respectively (**Table 4**). On the other hand, when there was four times more Eth in the medium than Aha, the preference for Eth at the N-terminus was two times more, while for the internal position, it was 1.3 more (**Table 4**). Since the values for Aha over Eth are almost two-fold different, it is possible to see not only an overall preference for Aha than Eth, but also a particular preference for Aha at the N-terminus over the internal position. Moreover, when both analogs were present in equal amounts, Aha was preferred both at the N-terminus and internal position 47 but differently, i.e. for each Eth, there were 1.8 and 1.2 Aha residues at the N-terminus and internal position, respectively (**Table 4**).

As previously suggested for b*1M, if one assumes that the Met analogs are efficiently taken up by the cells and accumulated several-fold inside them[148], it can be said that the *Ec*MetRS could have a preference to charge both the initiator $tRNA^{fMet}$ and elongator $tRNA^{Met}$ with Aha rather than Eth. Of course, the extent of formylation from both Aha- and Eth-$tRNA^{fMet}$ and their subsequent recognition during translation initiation will certainly play a role, although it is known that the ethionyl-$tRNA^{fMet}$ is recognized and formylated to the same extent as methionyl-$tRNA^{fMet}$ [149], and the same is applicable to azidohomoalinyl-$tRNA^{fMet}$.[150] The elongation factors could likewise differentially accept both Aha- and Eth-$tRNA^{Met}$, but this should be carefully addressed. In the best case, it is desirable that *Ec*MetRS could produce as much as initiator Aha-$tRNA^{fMet}$ as possible, and as less initiator Eth-$tRNA^{fMet}$, elongator Aha-$tRNA^{Met}$ or Eth-$tRNA^{Met}$ as possible, if the goal is to separate AUG-dependent translation. To this end, an efficient *in vivo* expression of the type B MetRS:$tRNA^{Met}$ pair derived from *S. acidocaldarius* that incorporates Eth at AUG codons exclusively, or at least preferentially in translation elongation in *E. coli* is required.

3.2.4 Cloning of orthogonal pair and *Sa*MetRS expression

The construction of an appropriate expression system that allows the expression of the orthogonal MetRS and tRNA^{Met} in *E. coli* was developed (plasmids for Tandem expression in *E. coli* or pTEc system). pTEc plasmids are derived from pRARE, a plasmid containing six rare *E. coli* tRNA genes under their natural promoters, which are incorporated to enhance the expression of exogenous proteins in *E. coli* e.g. the orthogonal *Sa*MetRS. These vectors are resistant to chloramphenicol and contain the *p15A* origin of replication that allows having the orthogonal pair in a mid- to low-copy plasmid number.[151] This system is compatible with the pQE or pET systems because they contain the origins of replication *colE1* or *pBR22*, respectively, and are either resistant to ampicillin or kanamycin. In addition, these plasmids allow an efficient expression of the target gene under the control of either the T5 or T7 inducible promoters. Accordingly, the vector pTEc0 was generated to allow the cloning of the orthogonal pair (see Appendix 9.5 for sequence details). pTEc1 was subsequently developed by cloning the *Sa*MetRS under the control of the constitutive EM7 promoter and by substituting the elongator tRNA^{Met} already present with an elongator tRNA from *S. acidocaldarius* (**Figure 45**).

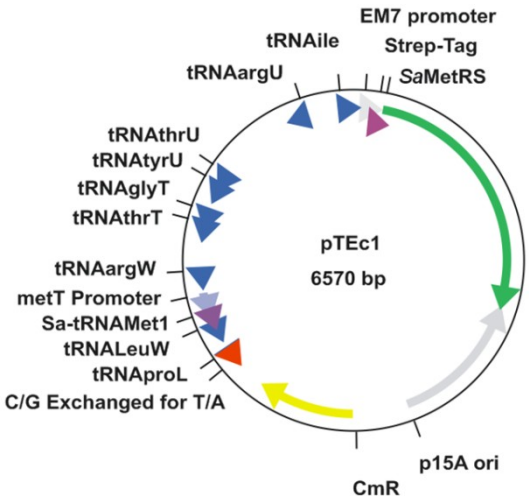


Figure 45. pTEc1. The vector for the tandem expression of the orthogonal pair *Sa*MetRS:*Sa*-tRNA^{Met} in *E. coli* has a bacterial EM7 synthetic constitutive promoter (light gray), followed by the exogenous *Sa*MetRS (green) containing an N-terminal Strep-tag (light purple). The p15A origin of replication is depicted in light gray. The chloramphenicol marker CmR confers resistance (yellow). The rare tRNA genes are indicated in dark blue and the exogenous *Sa*tRNA^{Met} in purple under the natural promoter from the *E. coli* elongator tRNA^{Met} (light blue).

S. acidocaldarius contains three putative Met tRNA isoacceptors coded in the genes *tRNA8*, *tRNA16*, and *tRNA43*.[152] As deduced from its similarities with the initiator tRNA^{fMet} from *E. coli*, the gene *tRNA43* would correspond to the initiator tRNAi^{Met} from *S. acidocaldarius* (**Figure 46**; see **Table 1** for details). Thus, the elongator tRNA^{Met} isoacceptors should be coded by the genes *tRNA8* and *tRNA16* (**Figure 47**).

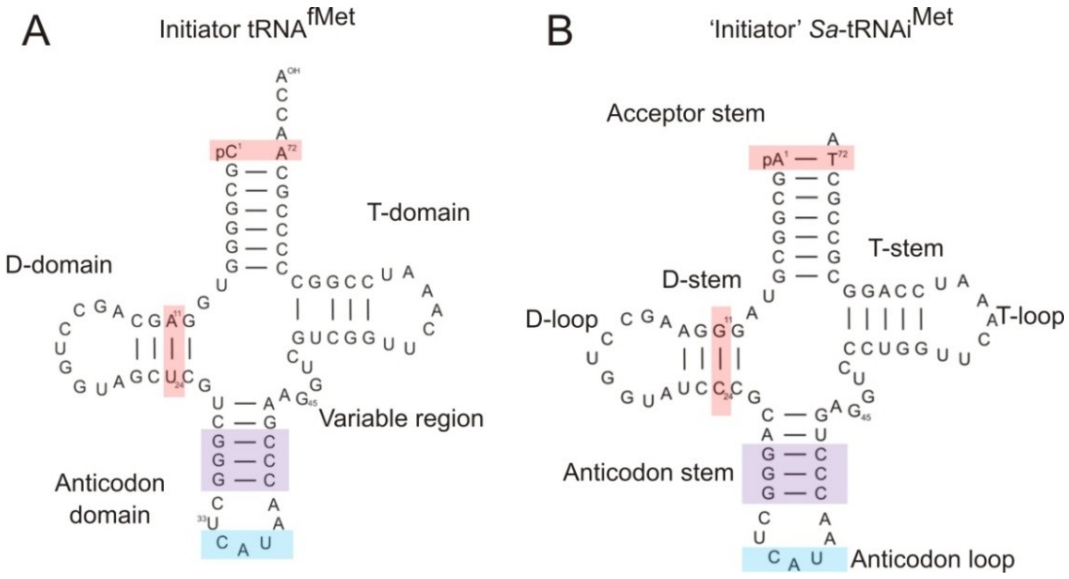


Figure 46. Initiator tRNAs. Secondary structure of the initiator tRNA from A) *E. coli*[72] and B) *S. acidocaldarius* (derived from the gene *tRNA43*).[152] The dihydrouridine (D), thymidine (T), and anticodon domains, as well as the acceptor arm are labeled. Each domain contains a stem and loop region. A variable arm occurs between the anticodon and T domains. The required identity elements for formylation (red), initiation (purple), and aminoacylation (blue) are highlighted for comparison.

Since the rare *E. coli* elongator tRNA^{Met} (**Figure 10B**) was previously present in the pRARE, it was substituted for the *S. acidocaldarius* 'elongator 1' tRNA^{Met} (*Sa*-tRNA^{Met}; **Figure 47A**). The 'elongator 1' tRNA^{Met} was chosen because its D-loop is larger (10 bases; **Figure 47A**) compared to that of the 'elongator 2' tRNA^{Met} (9 bases; **Figure 47B**). This slight difference in the D-loop could resemble the one-base substitution in the anticodon loop that serves as the determinant for specific aminoacylation of type B elongator tRNA^{Met} by type B MetRS but not type A MetRS.[131]

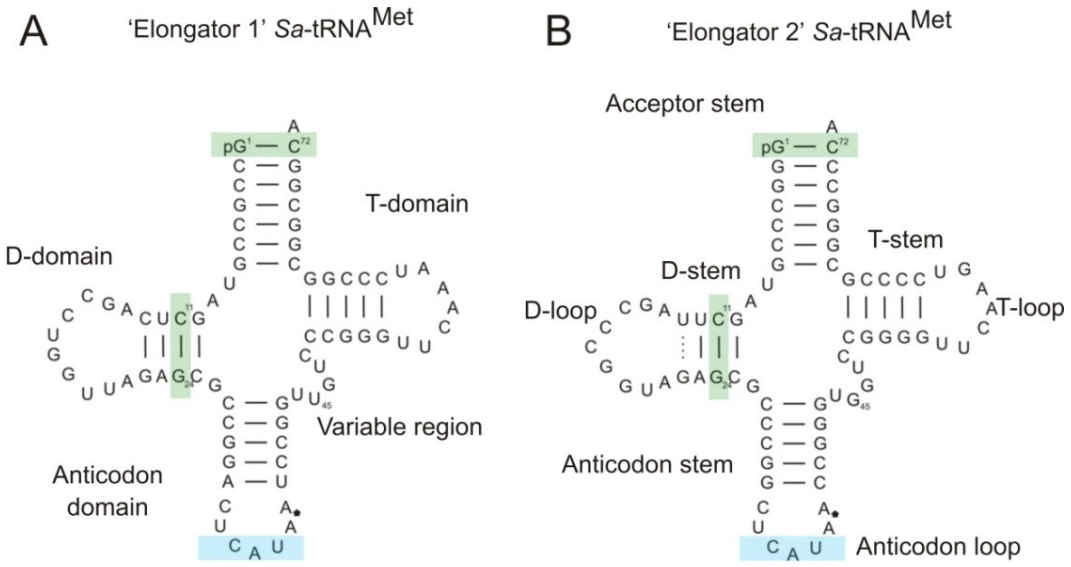


Figure 47. Elongator tRNAs from *S. acidocaldarius*. Secondary structure of the *S. acidocaldarius* elongator tRNA A) No. 1; gene tRNA16 (*Possible intron: 40-59 removed) and B) No. 2; gene tRNA8 (*Possible intron: 39-56 removed).[152] The dihydrouridine (D), thymidine (T), and anticodon domains, as well as the acceptor arm are labeled. The required identity elements for elongation (green) and aminoacylation (blue) are shown. The dotted line indicates that a base modification is required for hydrogen bonding.

A Strep·Tag II was appended to the N-terminus of *Sa*MetRS so that its expression can be assessed by immunoblotting. The Strep·Tag II is an eight-residue minimal peptide sequence (WSHPQFEK) that exhibits intrinsic affinity towards streptavidin. The advantages of using Strep·Tag II are its short size, stability against proteases, and lack of interference with protein folding.[153] To discard the factor of genetic background as a problem for expressing *Sa*MetRS, two different *E. coli* Met auxotrophic strains were transformed with pTEc1. Their protein extracts were subsequently blotted onto nitrocellulose membrane and probed with a mouse anti-Strep·Tag II antibody (see 6.2.1.4 for details). However, no expression of the *Sa*MetRS could be detected after 60 min of exposure time (**Figure 48**).

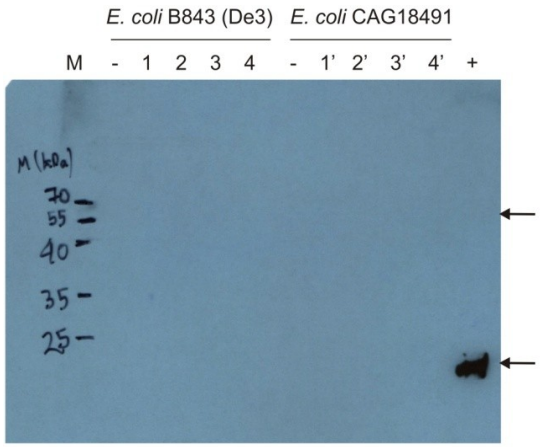


Figure 48. Intracellular expression of *Sa*MetRS in *E. coli* transformed with pTEc1. No expression of *Sa*MetRS (calculated mass of 67.5 kDa, see upper arrow) was detected for any of the four clones from *E. coli* B834 (DE3) (1, 2, 3, 4) or CAG18491 (1', 2', 3', 4'). The negative control (-) corresponds to *E. coli* cells transformed with pTEc0 instead of pTEc1. The positive control (+) is the 17 kDa protein hSOD1 (A2K) containing an N-terminal Strep·Tag II, indicated with an arrow. M = Standard MW.

Since the sequence of pTEc1 was verified (see Appendix 9.6 for details), the lack of expression of *Sa*MetRS could be only attributed to either codon usage or a relatively weak expression of the constitute promoter EM7. To rule out the first possibility, an *in silico* codon optimization analysis for the expression of *Sa*MetRS in *E coli* was performed with the aid of the OptimumGene Codon Optimization Analysis software from GeneScript (**Figure 49**).[154]

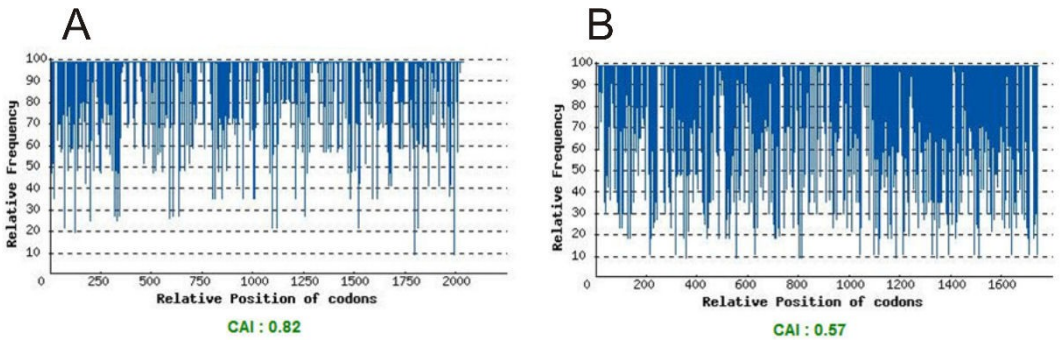


Figure 49. *In silico* expression of *Ec*MetRS and *Sa*MetRS in *E. coli*. The distribution of codon usage frequency for *Ec*MetRS (left panel) and *Sa*MetRS (right panel) confers a codon adaptation index (CAI) of 0.82 and 0.57, respectively. The closer the CAI is to 1.0, the bigger the possibility that the gene can be expressed in *E. coli.*

Codon usage is defined as the specific use of the tRNA population, which is in direct correlation with the codon composition of total mRNA, by each living cell.[155] The fact that the archaeal *Sa*MetRS gene is not expressed in the bacterium *E. coli* could be due to the lack of specific tRNAs to decode it. To approach this "codon usage" problem, a detailed analysis of the rare *E. coli* tRNAs whose codons are present in gene sequence of SaMetRS was performed (**Figure 50**).

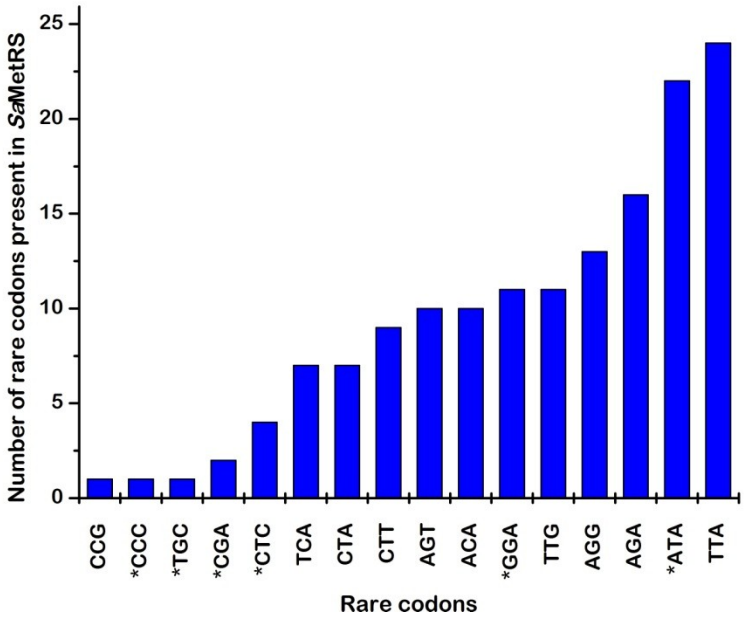


Figure 50. Distribution of codons present in *Sa*MetRS whose tRNAs are the rarest in *E. coli*. The 'rarest' tRNA is $tRNA^{Leu}$, which decodes the TTA codon and is present 24 times along the sequence of the orthogonal MetRS. * are codons that are decoded by rare tRNAs already supplemented by pTEc1.

In order to improve the expression of heterologous proteins in *E. coli*, there are mainly two strategies: 1) supplementation of rare tRNAs or 2) *de novo* synthesis of the target gene.[156] Both methods are considered equally efficient, but supplying tRNA is the most economical approach; therefore, the gene *leuZ* corresponding to the rarest tRNA ($tRNA^{Leu}$ with the anticodon UAA) whose codon is the most frequent along the gene sequence of *Sa*MetRS was cloned into pTEc1. The promoter EM7 was also substituted with the strong constitutive promoter *glnS'* along with the terminator *rho T* (see 6.1.8.4 for details). The *glnS'* promoter is derived from the original glutamine tRNA-synthetase promoter with several mutations is considered one of the strongest

promoters in *E. coli*[157], while *rho T* is a very common naturally occurring release factor that binds nascent RNA molecules in *E. coli*.[158] With these, an optimized vector for *Sa*MetRS expression, pTEc1.1.G-R/L, containing new tRNA, promoter, and terminator sequences, was produced (**Figure 51**; see Appendix 9.7 for sequence details).

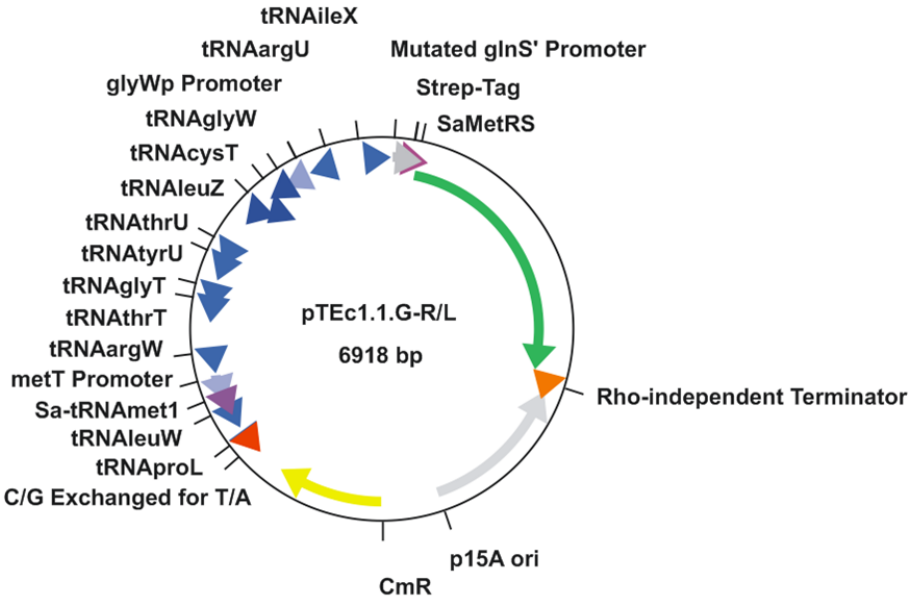


Figure 51. pTEc1.1G-R/L. The vector for the expression of the orthogonal pair *Sa*MetRS:*Sa*-tRNA^{Met} in *E. coli* harbors the strong constitutive promoter *glnS'* (dark gray), followed by the exogenous *Sa*MetRS (green) containing an N-terminal Strep-Tag II (light purple) and flanked by the *rho T* terminator (orange). The p15A origin of replication is depicted in light gray. The chloramphenicol marker CmR confers resistance (yellow). The rare tRNA genes are indicated in dark blue and the exogenous *Sa*tRNA^{Met} in purple under the natural promoter from the *E. coli* elongator tRNA^{Met} (light blue).

To check for the expression of *Sa*MetRS, an immunoblot assay was carried out (see 6.1.5.4 for details). However, even with the cloning of the rarest tRNA and the use of a stronger promoter, no *Sa*MetRS could be detected (**Figure 52**).

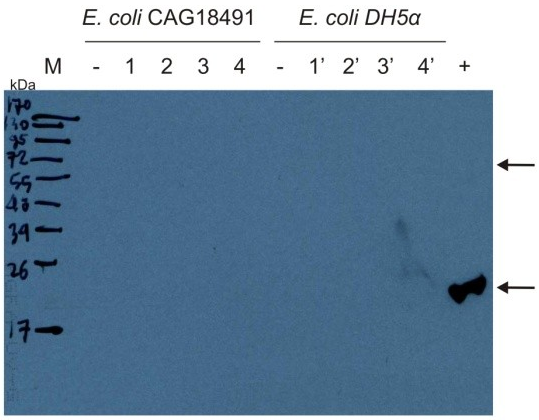


Figure 52. Expression of *Sa*MetRS in *E. coli* transformed with pTEc1.1G-R/L. No expression of *Sa*MetRS (calculated mass of 67 kDa, shown with the upper arrow) is detected for any of the four clones from *E. coli* strains CAG18491 (1, 2, 3, 4) or DH5α (1', 2', 3', 4'). The negative control (-) corresponds to *E. coli* cells transformed with pTEc0 instead of pTEc1.1G-R/L. The positive control (+) corresponds to the 17 kDa protein hSOD1 (A2K) containing a C-terminal-Strep·Tag II, indicated with the lower arrow. M = Standard MW.

To rule out the lack of expression due to low activity, the last experiment was repeated but with an aliquot corresponding to 10 OD_{600} enriched by Strep·Tag II mini-columns. As in the previous cases, no *Sa*MetRS could be detected (**Figure 53**).

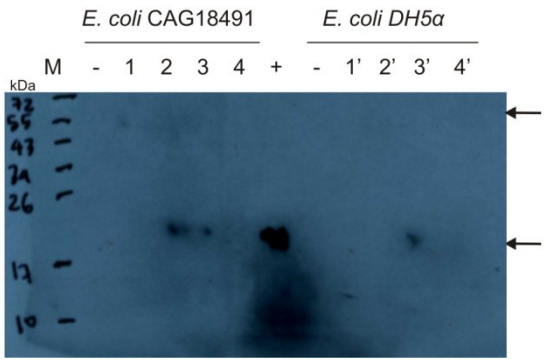


Figure 53. Expression of *Sa*MetRS in *E. coli* transformed with pTEc1.1G-R/L after enrichment. No expression of *Sa*MetRS (calculated mass of 67.5 kDa shown with the upper arrow) is detected in the 1) soluble purified fraction, 2) bound sample after washing, 3) sample after loading, or 4) insoluble fractions for any of the two *E. coli* strains. The negative control (-) corresponds to *E. coli* cells transformed with pTEc0 instead of pTEc1.1G-R/L. The positive control (+) corresponds to the 17 kDa protein hSOD1 (A2K) containing a C-terminal-Strep·Tag II, indicated with the lower arrow. M = Standard MW.

To completely rule out the problem of codon usage, the rare codons present in the gene sequence of *Sa*MetRS were optimized for expressing this gene in *E. coli.* After *de novo* gene synthesis, *Sa*MetRS was cloned into pTEc1.1.G-R/L to yield the plasmid pTEc2, which was thereafter transformed into *E. coli*; however, no expression of the orthogonal *Sa*MetRS could be detected (data not shown), not even after a 10x enrichment of 1 OD_{600} using Strep·Tag II columns, as previously done (**Figure 54**).

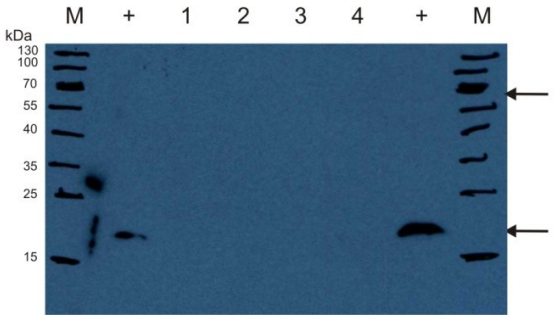


Figure 54. Expression of *Sa*MetRS in *E. coli* transformed with pTEc2 after enrichment. No expression of *Sa*MetRS (calculated mass of 67.5 kDa, shown with the upper arrow) is detected in the 1) sample before loading, 2) sample after loading, 3) bound sample after washing, or the 4) soluble purified fraction from *E. coli* strain CAG18491. The positive control (+) corresponds to the 17 kDa protein hSOD1 (A2K) containing a C-terminal-Strep·Tag II, indicated with the lower arrow. M = Standard MW.

The absence of *Sa*MetRS expression is most likely due to the promoter used. It was recently shown that the *glnS'* promoter is not as efficient as previously described, i.e. the constitutive expression of the AARS alone may not be sufficient to aminoacylate its cognate tRNA.[159] This is very important in the development of a technique for the efficient expression of an orthogonal pair in *E. coli* using an inducible arabinose promoter together with the *glnS'* one for optimal AARS expression. Here, it would make more sense to use an inducible promoter that would allow an efficient expression of *Sa*MetRS and subsequent aminoacylation of the Sa-$tRNA^{Met}$ with Eth. The development of such an expression system is underway. Importantly, a preferential incorporation of the bioorthogonal chemical reporter Aha by *Ec*MetRS and Eth by the *Sa*MetRS at the N-terminal and internal positions, respectively, would allow us to specifically modify the N-termini and retain or even improve the activity of

proteins. In fact, attempts to recode the N-terminus of proteins with noncanonical amino acids for controlled N-terminal protein processing have only been started in started *in vitro*.[160-167] Thus, this system represents the first step towards the *in vivo* control of N-terminal post-translational modifications, as well as the site-specific recoding of a single sense codon by an orthogonal AARS:tRNA pair (**Figure 55**).

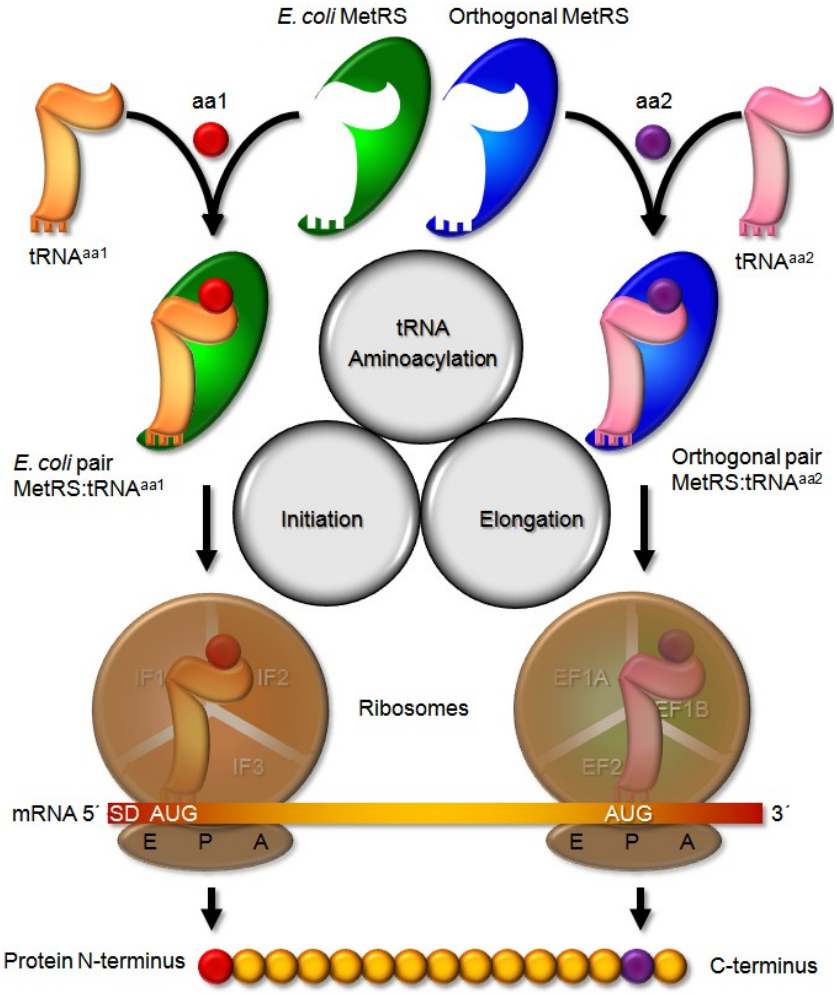


Figure 55. Making up the faces of Janus. Due to its substrate promiscuity, the *E. coli* MetRS can charge a Met analog (aa1; red) onto $tRNA^{fMet}$ (represented as $tRNA^{aa1}$). By combining the SPI and SCS methods, a new functional orthogonal MetRS (blue) would specifically aminoacylate its cognate elongator $tRNA^{aa2}$ (pink) with a Met analog (aa2; purple). Intracellular accumulation of the orthogonal elongator $tRNA^{aa2}$ results in preferrential incorporation of aa2 at internal AUG codons. The exogenous MetRS should not cross-react with endogenous MetRS:$tRNA^{aa1}$ pair. Finally, the tRNAs should be accurately discriminated by the initiation and elongation factors. SD = Shine-Dalgarno sequence

4 Conclusions

This work presents the potentials of genetic code engineering in synthetic biotechnology. Specifically, it demonstrates that it is possible to design lipase congeners with physicochemical properties fully reflected by the global incorporation of the Met analogs Nle and Aha. Instead of causing global perturbations in the protein structure, these noncanonical amino acids were responsible for bringing about striking differences in optimal temperature and pH, thermal stability, substrate access and specificity, as well as resistance toward organic solutions, metal ions, surfactants and inhibitors in the functional congeners. The finding that TTL[Nle] most probably has an open substrate access might pave the way for generating stable lipases with improved characteristics for industrial processes without necessitating the energy and material expenditure required in classical genetic engineering. Furthermore, the generated chemical diversity by globally incorporating noncanonical amino acids can also increase the tolerance of the lipase congeners toward different additives. Doubtless, such chemical diversity is the main cause for the unique and emergent properties observed exclusively in the lipase congeners.

Nevertheless, the SPI method alone is not sufficient to determine if key lipase residues are exclusively responsible for the observed differences. Thus, the development of a new method that allows the simultaneous use of more than one Met analog will definitely enhance the potentials of engineering the genetic code. In this context, the first efforts toward the site-specific incorporation of Met analogs into proteins were presented too. The fact that *E. coli* MetRS prefers activating Aha over Eth and *S. acidocaldarius* MetRS Eth over Aha, could serve as a starting point for further developing their substrate specificity. Additionally, combined together with the fact that the single AUG codon is decoded by two different tRNAs, the importance of using new MetRS:$tRNA^{Met}$ pairs from evolutionary distinct organisms should help developing the required technology for the position-specific AUG recoding. In this thesis, the first steps toward this goal are described.

5 Outlook

Lipases contain one of the most common natural protein folds, the so-called α/β hydrolase fold, which is usually present in most hydrolytic enzymes. In fact, the ESTHER database[168] includes hundreds of proteins that belong to this family, including carboxyl esterases, thioesterases, lipases, peptidases, aminoester hydrolases, deacetylases, or acyl transferases. The fact that the lipase from *T. thermohydrosulfuricus* (TTL) tolerated a relative high number of Met replacements very well indicates that the incorporation of Nle into proteins with the same α/β hydrolase fold could be also beneficial in engineering other biocatalysts of industrial interest. Of course, it would be necessary to obtain 3D structural data of these alloproteins with noncanonical amino acids to be able to correlate its structural and functional properties, and this work is being pursued in this direction. Likewise, the recently successful incorporation of three different noncanonical amino acids using the SPI method with polyauxotrophic *E. coli* strains[169] will permit the generation and characterization of beneficial and emergent properties provided by three or even more chemistries to lipases or other biocatalysts. Currently, we are combining many noncanonical amino acids into TTL (double, triple, quadruple, and quintuple incorporations) to search for lipases with *de novo* properties with interesting applications in white biotechnology.

The nature of scientific research has become increasingly interdisciplinary. In this context, the field of genetic code engineering is not only limited to fundamental academic questions. As a subfield of synthetic biology, it is geared towards new and useful applications in biotechnology. This makes the cross-talk between basic and applied research fundamental: it drives toward innovation. The finding that Nle-TTL's lid is 'open' raises the question: which Met residues play key roles in this conversion? The answer can be obtained by combining different efforts in protein engineering and synthetic biology. Currently, no method that allows the site-specific incorporation of Met analogs exists, but in this work, the first efforts in this direction are presented by combining the SPI and SCS methods. From the results, it is clear that the presented

orthogonal pair has to be optimally cloned and its expression and functionality has to be properly assessed *in vivo*. Such proof-of-principle will bring the next step, namely the insertion of the orthogonal molecules into the host genome. The group of Link recently showed that it is possible to introduce a mutated MetRS with substrate specificity for Anl into the genome of a Met auxotrophic *E. coli* strain without perturbing cell growth while enabling high levels of expression of recombinant proteins containing the Met analog.[170] In our case, the mechanisms governing the translation initiation and AUG-dependent elongation will not only depend on the extent of incorporating an orthogonal MetRS into the host genome, but also on how it would affect the host physiology. The potential pitfalls, however, might be in the form of some difficulties in the regulation of transcription during the expression of the orthogonal pair, given that AARSs complexes have multiple tasks.[171] For this, it would be necessary to engineer metabolic pathwyas. Furthermore, this step might lead towards the next fundamental problem that needs to be solved: would it be possible to evolve heritable changes in the genetic code interpretation?

The major drawback of all existing genetic code engineering techniques is that the coding units, i.e., sense or stop codons are only transiently reassigned. Permanent reassignment of existing coding units inevitably leads to detrimental effects. On one hand, auxotrophic cells cannot grow using just the noncanonical analogs. On the other, a permanent read through of stop codons is lethal for the cells.[172] Thus, for stable and inheritable recoding, new coding units have to be introduced into the genome of the target organism. From the current point of view, the only way to introduce *de novo* coding units is the *de novo* chemical synthesis of a whole genome in which rare codons are exchanged for more frequent ones with the same meaning, which serves as an alternative to the generation of artificial DNA pairs.[173] Thus, rare codons disappear from the genetic code and can subsequently be specifically reintroduced with a new meaning into suitable target genes. However, this approach is still hampered by the rather inefficient and prohibitively expensive chemical DNA synthesis and is currently a perspective for the distant future.

6 Experimental section

6.1 Materials

6.1.1 Equipment

- Autoclave: Varioklav Dampfsterilisator Typ 500 E (H+P Labortechnik GmbH, Oberschleißheim, Germany).
- Balances: TE1502S, BP211D (Sartorius, Göttingen, Germany); GB2002, PC4400 Delta Range (Mettler-Toledo GmbH, Giessen, Germany).
- Blotting apparatus: Trockenblot 250/180, 05/04 (MPI of Biochemistry, Munich, Germany).
- CD spectropolarimeter: Jasco J-715, temperature control by Peltier FDCD attachment PFD-350S/350L (JASCO International Co., Ltd., Tokyo, Japan).
- Centrifuges: Avanti J-25; Avanti J-20 XP (Beckmann, Munich, Germany); Centrifuge 5415 C/D; Zentrifuge 3200 (Eppendorf, Hamburg, Germany); Universal 32R (Hettich Zentrifugen, Tuttlingen, Germany).
- Centrifuge rotors: JA 25.50, JLA 8.1000, JLA 10.500, (Beckmann, Munich, Germany).
- Cuvettes: Hellma 104.002-QS, Hellma 104.002F-QS, Hellma 110-QS, (Hellma, Müllheim, Germany).
- Electroporator: Electroporator 1000 (Stratagene, La Jolla, CA, USA).
- FPLC: Äktaexplorer, Äktabasic (GE Healthcare, Munich, Germany)
- Gel documentation: Eagle Eye II (Stratagene, La Jolla CA, USA), High Performance CCD Camera (COHU, Inc., San Diego CA, USA).
- Incubators: Thermomixer comfort, Thermomixer compact (Eppendorf, Hamburg, Germany), Incubator 3033 (GFL, Burgwedel, Germany).
- Magnetic stirrer: MR 3001 (Heidolph, Kehlheim, Germany).
- Mass spectrometer: MicrOTOF ESI-MS (Bruker Daltonics, Bremen, Germany)
- N-terminal sequencer: Gas-phase sequenator procise cLC (Applied Biosystems, Darmstadt, Germany).

- PCR cycler: Robocycler gradient 96 (Stratagene, La Jolla CA, USA).
- PH meter: MP 220, (Mettler-Toledo GmbH, Giessen, Germany).
- Photometer: UV/VIS Spectrometer Lambda 19 (PerkinElmer Life Sciences, Boston MA, USA); Ultrospec 6300 pro (GE Healthcare, München, Germany); ND 1000 (NanoDrop, Wilmington DE, USA).
- Sonifier: Sonifier 450 Macrotip (Branson, St. Louis, MO, USA).
- Sterile bench: Lamin Air HA244GS (Heraeus, Hanau, Germany).
- Vortex: Vortex Genie 2 (Bender & Hobein AG, Zurich, Switzerland).
- Ultra-Turrax Emulsifier (IKA Werke GmbH & Co. KG, Staufen, Germany)

6.1.2 Extendable materials

- Blotting membrane: Protran 0.1 µm (Whatman International Ltd, Maidstone Kent, UK).
- Centrifugal concentrators: Vivaspin 30 kDa (Sartorius Stedim Biotech, Aubagne, France).
- Columns: HiTrap Chelating HP and HiTrap Q Sepharose (GE Healthcare, Munich, Germany). Waters RP C4 column (Waters GmbH, Eschborn, Germany)
- Dialysis membranes: Spectra/por molecularporous membrane MW cutoff: 3,500 (Spectrum Laboratories, Rancho Dominguez CA, USA).
- Electroporation cuvettes: GenePulser/Micropulser cuvettes, 1 mm gap width (Biorad, Hercules CA, USA).
- Filter paper: Whatman 3MM (Whatman International Ltd, Maidstone Kent, UK).
- X-ray film paper: Curix Ultra UV-G (AGFA, Cologne, Germany).
- Sterile Filters: Millex-HA Filter Unit 0.45 µm/0.22 µm and Steritop GP (Millipore, Billerica MA, USA).

6.1.3 Chemicals

All standard chemicals were purchased from Sigma (Steinheim, Germany) or Merck KGaA (Darmstadt, Germany) unless otherwise specified. L-isomers of all amino acids

were used. Aha and Hpg were synthesized as previously described.[37,174] Anl, Dpg and Nhm were kindly provided by Dr. Shouliang Dong (Lanzhou University, China). Omd, Ome, Cpa, and Don were purchased from Bachem (Weil am Rhein, Germany) and Tfn from Fluorochem (Derbyshire, UK). Tfm, Nle and Eth were obtained from Sigma and Mox from CBL Patras (Patras, Greece). *E. coli* initiator $tRNA^{fMet}$ was purchased from Sigma and total tRNA from *E. coli* MR600 from Roche Diagnostics (Mannheim, Germany).

6.1.4 Commercial reagents

6.1.4.1 Kits and markers

- PCR purification: QIAquick PCR Purification Kit or Gel Extraction Kit (Qiagen GmbH, Hilden, Germany).
- Plasmid preparation: QIAprep Spin Miniprep Kit and Qiagen Plasmid Midi Kit (Qiagen GmbH, Hilden, Germany).
- Chemiluminescence reagent: Supersignal West Pico Chemiluminescence Substrate (Perbio Science Deutschland GmbH, Bonn, Germany).
- Bradford reagent: Biorad Protein Assay (Biorad, Hercules CA, USA).
- DNA MW marker: GeneRuler 1kb Ladder from Fermentas (St. Leon-Rot, Germany).
- Protein MW marker: PageRuler™ Prestained Protein Ladder from Fermentas (St. Leon-Rot, Germany).

6.1.4.2 Enzymes and antibodies

- Protein purification: Lysozyme, DNAse I, and RNAse A were from Sigma.
- PCR: Takara ExTaq polymerase from TakaraBio (Mountain View, CA, USA).
- Cloning: T4 DNA ligase and all restriction enzymes were obtainted from New England Biolabs (Frankfurt am Main, Germany).
- Primary antibody: Monoclonal mouse antibody anti-Streptag II (ZNP, LMU,

Martinsried, Germany).

- Secondary antibody: Goat anti-mouse IgG coupled to alkaline phosphatase (Dianova, Hamburg, Germany).

6.1.5 Buffers and solutions

All aqueous buffers and solutions were prepared using autoclaved, sterile filtered or bi-distilled water when required.

6.1.5.1 DNA electrophoresis

- TAE buffer: Stock solution 50x (2 M Tris·HCl, 1 M glacial acetic acid, 100 mM EDTA, pH 8.5) and working concentration 1x TAE buffer (40 mM Tris·HCl, 20 mM glacial acetic acid, 2 mM EDTA, pH 8.5).
- 6x DNA sample buffer: 0.25% bromophenol blue, 0.25% xylencyanole, 30% glycerol.

6.1.5.2 Protein purification

- Ni-NTA lysis buffer: 50 mM NaH_2PO_4, 300 mM NaCl and 10 mM imidazole, pH 8.
- Ni-NTA high-salt washing buffer: 50 mM NaH_2PO_4, 2 M NaCl and 20 mM imidazole, pH 8.
- Ni-NTA low-salt washing buffer: 50 mM NaH_2PO_4, 300 mM NaCl and 20 mM imidazole, pH 8.
- Ni-NTA elution buffer: 50 mM NaH_2PO_4, 300 mM NaCl/250 mM imidazole, pH 8.
- Ion Exchange (IE) resuspension buffer: 50 mM Tris·HCl, 7.5 M urea, pH 8.
- Ion Exchange washing buffer: 50 mM Tris·HCl, 100 mM NaCl, pH 8.
- Ion Exchange elution buffer: 50 mM Tris·HCl, 1 M NaCl, pH 8.

6.1.5.3 Protein electrophoresis (SDS-PAGE)

- 12% resolving gel: 375 mM Tris·HCl, pH 8, 12% of Acrylamide:Bis-acrylamide [30:0.8%(w/v)], 0.1% SDS, 0.1% APS, 0.1% TEMED.
- 17% resolving gel: 375 mM Tris·HCl, pH 8, 17% of Acrylamide:Bis-acrylamide [30:0.8%(w/v)], 0.1% SDS, 0.1% APS, 0.1% TEMED.
- Stacking gel: 195 mM Tris·HCl, pH 6.8, 5% of Acrylamide:Bis-acrylamide [30:0.8%(w/v)], 0.1% SDS, 0.1% APS, 0.1% TEMED.
- Running buffer: 190 mM glycine, 25 mM Tris·HCl, 3.5 mM SDS.
- 5x SDS-PAGE sample buffer: 450 mM Tris·HCl, pH 6.8; 3.6% SDS, 0.2% bromophenol blue; 30% glycerol; 45% β-mercaptoethanol.
- Coomassie staining solution: 0.1% Coomassie Brilliant Blue R 250; 25% ethanol; 8% acetic acid.
- Coomassie destaining solution: 25% ethanol; 8% acetic acid
- Fixing solution: 5% glycerol; 10% ethanol

6.1.5.4 Western Blot (Protein immunoblotting)

- Transfer buffer: 25 mM Tris·HCl, 192 mM glycine, 20% methanol, 0.1% SDS, pH 8.
- TBS bufffer: 50 mM Tris·HCl, 150 mM NaCl, 0.1% Tween20, pH 8.
- Ponceau S staining solution: 0.5% (w/v) Ponceau S in 1% (v/v) acetic acid.

6.1.5.5 Protein storage

- MetRS buffer: 20 mM Tris·HCl, 150 mM KCl, 15 mM $MgCl_2$ and 5 mM β-Mercaptoethanol, pH 8.
- Barstar buffer: 50 mM Tris·HCl and 100 mM NaCl, pH 8.
- Lipase buffer: 50 mM Tris·HCl pH 8.

6.1.6 Media and supplements

For bacterial growth, fermentation, and protein expression, LB medium and New Minimal Medium (NMM) were used. The components of LB medium were purchased

from BD Biosciences (San José, CA, USA). LB liquid medium was autoclaved for 20 min at 121 °C and 1.5 bars; for agar plates, 1.5% agar was added prior to autoclaving. Antibiotics were filter sterilized before addition to the media.

6.1.6.1 Media

- LB medium: 10 g BactoTM, Tryptone, 5 g BactoTM Yeast Extract, and 10 g NaCl were dissolved in 1 L H_2O and autoclaved before use.
- NMM: 7.5 mM $(NH_4)_2SO_4$, 8.5 mM NaCl, 22.5 mM KH_2PO_4, 50 mM K_2HPO_4, 20 mM Glucose, 50 mg of all amino acids, except the amino acid to be replaced, 1 mM $MgSO_4$, 1 mg Ca^{2+}, 1 mg Fe^{2+}, 10 mg trace elements (Cu^{2+}, Zn^{2+}, Mn^{2+}, MoO_4^{2+}), 10 mg thiamine, 10 mg biotin. Adjust to 1 L with H_2O.

6.1.6.2 Supplements

- Ampicillin: Stock concentration 100 mg/mL in H_2O and final working concentration 100 µg/mL.
- Chloramphenicol: Stock concentration 34 mg/mL in EtOH and final working concentration of 50 µg/mL.
- Kanamycin: Stock concentration 30 mg/mL in H_2O and final working concentration of 15 and 30 µg/ml for cells with a chromosomal and plasmid-containing Kan^RLgene, respectively.
- Streptomycin: Stock concentration 50 mg/mL in H_2O and final working concentration 50 µg/ml.
- Tetracycline: Stock concentration 5 mg/mL in H_2O and final working concentration 12.5 µg/mL.
- IPTG: Stock 1 M IPTG in H_2O and final concentration of 1 mM.

6.1.7 Bacterial strains

For plasmid amplification, the DH5α and OmniMax strains were used; for MetRS

protein expression, the BL21(DE3) and strains were used. The Rosetta-gammi 2 strain alleviates codon bias and enhances disulfide bond formation in the cytoplasm when heterologous proteins are expressed in *E. coli*. This strain also carries the chloramphenicol-resistant pRARE2 plasmid, which supplies seven rare tRNAs (Novagen; Merck Chemicals Ltd., Nottingham, UK). For the incorporation of Aha and Nle into lipase and Aha and Eth into barstar, the *E. coli* Met auxotrophic strain CAG18491 (CGSC *E. coli* Genetic Resources at Yale, CGSC# 7464) was used.

6.1.7.1 Plasmid amplification

- DH5α: F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r_K^- m_K^+), λ–.
- OmniMax: F′ [proAB+ lacIq lacZΔM15 Tn10(TetR) Δ(ccdAB)] mcrA Δ(mrr-hsdRMS-mcrBC) φ80(lacZ)ΔM15 Δ(lacZYA-argF) U169 endA1 recA1 supE44 thi-1 gyrA96 relA1 tonA panD.

6.1.7.2 Protein expression

- BL21(DE3): $F^–$ ompT gal dcm lon $hsdS_B$ (r_B^- m_B^-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]).
- Rosetta-gami 2: Δ(ara-leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsL (DE3) F′[lac^+ $lacI^q$ pro] gor522::Tn10 trxB pLysSRARE2 (Cam^R, Str^R, Tet^R).

6.1.7.3 Congener expression

- B834 (DE3): $F^–$ ompT $hsdS_B$ ($r_B^-m_B^-$) gal dcm met.
- CAG18491: LAM- rph-1 metEo-3079::Tn10.

6.1.8 Plasmids

The pET (Novagen) and pQE (Qiagen, Hildesheim, Germany) systems allow the efficient expression of recombinant proteins. In this systems, the *trans*-acting *lac*-repressor blocks protein expression by binding to the *cis*-regulatory *lac*-operator upstream of the promoter sequence. Protein expression is then induced by the addition of the artificial inducer IPTG, which binds to the lac repressor protein and inactivates it. In addition, these systems allow fast, efficient production and purification of N- or C-terminally 6xHis·Tagged proteins. The pET and pQE transcription–translation systems are based on the phage-derived T7 and T5 promoters, respectively. The T7 expression vectors require lysogenic *E. coli* strains for λ-DE3, which can provide the T7 RNA polymerase. The T5 promoter is recognized by any host *E. coli* RNA polymerase. In this work, the lipase was cloned into pQE80L. To generate the MetRS constructs, the pET15b, pET22b, pET28a or pQE80L vectors were used. The site-directed mutagenesis QuikChangeMult Kit (Stratagene, La Jolla, CA, USA) was used for the generation of MetRS mutants. All MetRS and barstar constructs were kindly provided by Petra Birle and Tatjana Krywcun (MPI of Biochemistry, Martinsried, Germany). The pRARE1-derived construct for the orthogonal pair cloning was given by Dr. Yuri Cheburkin (MPI of Biochemistry, Martinsried, Germany). All plasmids sequences were verified (DNA sequencing service at the MPI of Biochemistry, Martinsried, Germany). All primers were synthesized by Metabion International AG (Martinsried, Germany).

6.1.8.1 Lipase construct

pQE80L-TTL: The sequence encoding the lipase from *T. thermohydrosulfuricus* was PCR amplified with the following primers: 5'-tatatctatatcgaattcattaaagaggagaaattaag-catgcaaaaggctgttgaaattacatataac-3' / 5'- cgtgccggcggctgcagttatcagtgatggtgatggtgatg-ggatcctccctttaacaattccttttgaaaaact-3' and cloned into the previously double-digested *EcoRI* and *PstI* cleavage sites (underlined) of pQE80L. A hexahistidine tag was attached to the C-terminus of TTL with an intervening short spacer (Gly-Ser). Additional random nucleotides at the 5'-end were added for efficient restriction.

6.1.8.2 MetRS constructs

The MetRS available coding sequences from different organisms were extracted from NCBI databases, PCR-amplified and cloned into suitable expression vectors. Some *Ec*MetRS mutants have enhanced substrate specificities for Met analogs (see 3.2.1.2 for details).

- pET15b-*Aa*MetRS-WT: The gene encoding *A. aeolicus* MetRS WT (497 residues) was cloned by PCR from genomic DNA using the following primers: 5'-gccatatgacacttatgaagaagttctacg-3 / 5'-gcggatcctcaaccttccctttgggaaatag-3' between the *NdeI* and *BamHI* sites (underlined) of the pET-15b plasmid. Additional random nucleotides at the 5'-end were added for efficient restriction. This plasmid is ampicillin resistant (Amp^R) and contains an N-terminal His·Tag.

- pET28a-*Ec*MetRS-WT: The gene encoding *E. coli* MetRS WT (676 residues) was cloned by PCR from genomic DNA using the following primers: 5'-agctagctgaattcattgaaggccgtactcaagtcgcgaagaaa-3' / 5'-agctagctgaattctcattatttcac-ctgatgacccggttt-3' between the *EcoRI* sites (underlined) of the pET-28a plasmid. This plasmid is kanamycin resistant (Kan^R). Additional random nucleotides at the 5'-end were added for efficient restriction. An Fxa cleavage site (IEGR) was added between the N-terminal His·Tag and the MetRS coding region.

- pET28a-*Ec*MetRS-SM(L13G): pET28a-*Ec*MetRS-WT-derived.

- pET28a-*Ec*MetRS-DM(L13G/Y260L): pET28a-*Ec*MetRS-SM-derived.

- pET28a-*Ec*MetRS-TM(L13G/Y260L/H301L): pET28a-*Ec*MetRS-DM-derived.

- pET28a-*Ap*MetRS-WT: The gene encoding *A. pernix* MetRS WT (572 residues) was cloned by PCR from genomic DNA using the following primers: 5'-ccgccgccgccggaattcatcgagggaagggctaagtatgtagtaacctccgcgtg-3' / 5'- catgcatgcat-

gaagctttcactagtccctgaggagaggggcctct-3' between the *EcoRI* and *HindIII* sites (underlined) of the Kan^R pET-28a plasmid. Additional random nucleotides at the 5'-end were added for efficient restriction. An Fxa cleavage site (IEGR) was added between the N-terminal His·Tag and the MetRS coding region.

- pET15b-*Mj*MetRS-WT: The gene encoding *M. jannaschii* MetRS WT (648 residues) was cloned by PCR from genomic DNA using the following primers: 5'-atgacatatgagatatctaataacaactgcc-3 / 5'-gcggatcctcaactacctgcttttatatccttatc-3' between the *NdeI* and *BamHI* sites (underlined) of the Amp^R pET-15b plasmid. Additional random nucleotides at the 5'-end were added for efficient restriction. The plasmid contains an N-terminal His·Tag.

- pET15b-*Mj*MetRS SM (L9G): pET15b-*Mj*MetRS-WT-derived.

- pET15b-*Mj*MetRS DM (L13G/Y260L): pET15b-*Mj*MetRS-SM-derived.

- pET15b-*Mj*MetRS TM (L13G/Y260L/H301L): pET15b-*Mj*MetRS-DM-derived.

- pET15b-*Np*MetRS-WT: The gene encoding *N. pharaonis* MetRS WT (698 residues) was cloned by PCR from genomic DNA using the following primers: 5'-gccatatgcacgaggagtttccgacc-3' / 5'-gcggatcctactgaaccttcgttcccgg-3' between the *NdeI* and *BamHI* sites (underlined) of the Amp^R pET-15b plasmid. This plasmid contains an N-terminal His·Tag.

- pET28a-*Pa*MetRS-WT: The gene encoding *P. aerophilum* MetRS WT (570 residues) was cloned by PCR from genomic DNA using the following primers: 5'-ccgccgccgccggaattcatcgagggaagggcgaaatacgtaataggctcggcg-3' / 5'-catgcatgcatgaagctttcattatatcactacctgttcccacgggta-3' between the *EcoRI* and *HindIII* sites (underlined) of the Kan^R pET-28a plasmid. Additional random nucleotides at the 5'-end were added for efficient restriction. An Fxa cleavage site (IEGR) was added between the N-terminal His·Tag and the MetRS coding region.

- pET15b-*Pab*MetRS-WT: This Amp^R plasmid containing the MetRS WT sequence from *P. abyssi* was kindly provided by Dr. Mechulam.[133]

- pET28*a*-*Sa*MetRS-WT: The gene encodin*g S. acidocaldarius* MetRS WT (570 residues) was cloned by PCR from genomic DNA using the following primers: 5'-ccgccgccgccg<u>gaattc</u>atcgagggaaggaaggttttagtaacatctgcatggcct-3' / 5'-catgcatgcat-<u>ggcggccgc</u>tcattaccttaaaagatcaggtctttctttttcc-3' between the *EcoRI* and *NotI* sites (underlined) of the Kan^R pET-28a plasmid. Additional random nucleotides at the 5'-end were added for efficient restriction. An Fxa cleavage site (IEGR) was added between the N-terminal His·Tag and the MetRS coding region.

- pET22b-*Ta*MetRS-WT: The gene encoding *T. acidophilum* MetRS WT (698 residues) was cloned by PCR from genomic DNA using the following primers: 5'-tac<u>acatatg</u>gtggttcagatcaagatacttg-3' / 5'-tac<u>agtcgac</u>agataccagggattcaaggtc-3' between the *NdeI* and *SalI* sites (underlined) of the Amp^R pET-22b plasmid. Additional random nucleotides at the 5'-end were added for efficient restriction. The plasmid contains a C-terminal His·Tag.

- pET28a-*At*MetRS-WT: The gene encoding *A. thaliana* MetRS WT (797 residues) was cloned by PCR from cDNA using the following primers: 5'-cgccgccgccgc<u>ggatcc</u>atcgagggaagggaagacgacggcaagagcagccccaag-3' / 5'- catgc-atgcatg<u>ctcgag</u>tcattaccggatcgtgccattgctgatg-3' between the *BamHI* and *XhoI* sites (underlined) of the Kan^R pET-28a plasmid. Additional random nucleotides at the 5'-end were added for efficient restriction. An Fxa cleavage site (IEGR) was added between the N-terminal His·Tag and the MetRS coding region.

- pET28a-*Dr*MetRS-WT: The gene encoding *D. rerio* MetRS WT (922 residues) was cloned by PCR from cDNA using the following primers: 5'-ccgccgccgccg<u>gaattc</u>atcgagggaaggaagctgtttatcggtgagggaaa-3' / 5'- catgcatgcatg-<u>aagctt</u>tcattatttcttcttggttttttgcgcgg-3' between the *EcoRI* and *HindIII* sites (underlined) of the Kan^R pET-28a plasmid. Additional random nucleotides at the 5'-

end were added for efficient restriction. An Fxa cleavage site (IEGR) was added between the N-terminal His·Tag and the MetRS coding region.

- pET28a-*Dm*MetRS-WT: The gene encoding *D. melanogaster* MetRS WT (1022 residues) was cloned by PCR from cDNA using the following primers: 5'-ccgccgccgccggaattcatcgagggaaggataatctacacgaatgatggcaaccc-3' / 5'- catgcatgc-atgctcgagtcactacttcttcttttgcccttgccctgt-3' between the *EcoRI* and *XhoI* sites (underlined) of the Kan^R pET-28a plasmid. Additional random nucleotides at the 5'-end were added for efficient restriction. An Fxa cleavage site (IEGR) was added between the N-terminal His·Tag and the MetRS coding region.

- pET28a-*Hs*MetRS-WT: The gene encoding *H. sapiens* MetRS WT (900 residues) was cloned by PCR from cDNA using the following primers: 5'-ccgccgccgccggaattcatcgagggaaggagactgttcgtgagtgatggcgtc-3' / 5'-catgcatgcatg-aagctttcattacttttttcttcttgcctttagggg-3' between the *EcoRI* and *HindIII* sites (underlined) of the Kan^R pET-28a plasmid. Additional random nucleotides at the 5'-end were added for efficient restriction. An Fxa cleavage site (IEGR) was added between the N-terminal His·Tag and the MetRS coding region.

- pQE80L-*Sc*MetRS-WT: The gene encoding *S. cerevisiae* MetRS WT (751 residues) was cloned by PCR from genomic DNA using the following primers: 5'-agctagctagctggatccattgaaggccgttctttcctcatttcctttgataaatc -3' / 5'- agctagct-agctggatcctcattacacttgttgaccaccatatttg- 3' between the *BamHI* sites (underlined) of the Amp^R pQE80L plasmid. This plasmid contains an N-terminal His·Tag.

- pQE80L-*Sc*MetRS-TF (6-185Δ): The coding region from the residue 1 to 5 and 185 to 751 of *S. cerevisiae* MetRS was cloned by PCR from genomic DNA using the following primers: 5'- cgccgccgccgcggatccatgtctttcctcattgattcagaaattttgcctaagc-caaac-3'; 5'- agctagctagctggatcctcattacacttgttgaccaccatatttg -3' between the *BamHI* sites (underlined) of the Amp^R pQE-80L plasmid. This plasmid contains an N-terminal His·Tag.

6.1.8.3 Barstar constructs

- pQE80L-b*1M: The gene encoding *B. amyloliquefaciens* barstar WT (90 residues) was originally cloned by PCR from genomic DNA between the *EcoRI* and *HindIII* sites of the pKK223-3 (Pharmacia) plasmid. Barstar was subsequently cloned into the same restriction sites of pQE80L, eliminating the N-terminal His·Tag. The barstar WT was then subjected to site-directed mutagenesis to generate barstar P28A (b*) under the control of the inducible T5 promoter. This plasmid is Amp^R.

- pQE80L-b*2M: Derived from pQE80L-b*1M, this plasmid contains barstar P28A/E47M with two AUG codons for Met reassignment.

6.1.8.4 Orthogonal pair constructs

Several strategies were followed to set up a system for the efficient expression of the orthogonal MetRS and tRNA.

- pTEc0: The fragment of plasmid pPICZB (Invitrogen, Carlsbad, CA, USA) containing the EM7 constitutive promoter and the zeocine resistance gene was initially PCR-amplified using the primers: 5'-ggactagtgttgacaattaatcatcggc-3' / 5'-ggaattcgtgaatgtaagcgtgacataac-3' and cloned between the previously double-digested *SpeI* and *EcoRI* sites (underlined) of the pBluescript II SK plasmid (Stratagene, La Jolla, CA, USA). A Multiple Cloning Site (MCS) was then attached next to the EM7 promoter using a three-step PCR procedure: i) Amplification of *SpeI*-EM7 and the first part of the MCS using the 5'-cgacggccagtgagcgcgc-3' / 5'-cctcatatggggccccgatcgtgatcaatgcatggtttagttcctc-accttgtcg-3' primers to generate the first template, ii) Amplification of first template *SpeI*-EM7 and second part of the MCS using the 5'-cgacggccagtgagcgcgc-3' / 5'-ggctcgaggtcgacagatctcacgtggat-caggcctcatatggggccccgatc-3' primers to generate the third template, and iii) the amplification of second template *SpeI*-EM7 and the third part of the MCS-*SacI* using the 5'-cgacggccagtgagcgcgc-3' / 5'-ggtaccgagctcgggcccttaattaacccgggctcg-

aggtcgacagatctc-3' primers. Finally, the EM7-MCS PCR product was subcloned between the previously double-digested *SpeI* and *SacI* sites of the pRARE vector (Novagen Merck Chemicals Ltd., Nottingham, UK). The plasmid pTEc0 contains a chloramphenicol resistance marker, a relatively mid-copy number origin of replication p15A, rare *E. coli* tRNA genes under their natural promoters that enhance the expression of exogenous proteins in *E. coli,* the constitutive promoter EM7 for the expression of orthogonal MetRS, and a MCS for cloning it. See Appendix 9.5 for sequence details.

- pTEc1: Derived from pTEc0, this vector no longer contains the elongator $tRNA^{Met}$ from *E. coli* since it was substituted by the orthogonal tRNA in a two-step PCR procedure: i) Amplification of a region partially containing the *ScaI*-Cam^R and the first part of the orthogonal $tRNA^{Met}$ using the 5'-gcttaatgaattacaacagtactgcg-3' / 5'-agcgccggactcataatccggttgtccggggttcaaatccccgcggcggcaccatctttttttgcgggag-3' primers to generate the first template, and ii) the amplification of first template partially containing the *ScaI*-Cam^R and the second part of the orthogonal $tRNA^{Met}$-*NruI* using the 5'-gcttaatgaattacaacagtactgcg-3' / 5'-tatatatatatatcgcgaaaaaaaa-agatgccgccgtagctcagcctggttagagcgccggactcataatc-3' primers. Additional random nucleotides at the 5'-end were added for efficient restriction. The Cam^R-$tRNA^{Met}$ product was subcloned between the *ScaI* and *NruI* of previously double-digested pTEc0. The orthogonal MetRS was PCR amplified and cloned between previously double-digested *NdeI* and *SalI* sites (underlined) using the 5'-tatatatatatcatatggcaagctggagccacccgcagttcgaaaagggtgcaatgaaggttttagtaacatctgcatgg-3' / 5'-atatatatatatgtcgacttaccttaaaagatcaggtctttctttttc-3' primers. Additional random nucleotides at the 5'-end were added for efficient restriction. A Strep-Tag II sequence was appended to the orthogonal MetRS N-terminus between two linker sequences for immunoblotting detection. See Appendix 9.6 for sequence details.

- pTEc1.1G-R/L: In addition to the rare tRNA genes provided by pTEc1, the extremely rare $tRNA^{Leu}$ coded by the gene *leuZ* is present. This gene was PCR-amplified from genomic *E. coli* DNA using the 5'-atatatatatatgcatgctcaaaagtggtg-

aaaaatatcgttg-3' / 5'-atatatatatatggatcctggtacccggagcgg-3' primers and cloned into the *SphI* and *BamHI* sites (underlined) from the pTEc1plasmid. Additional random nucleotides at the 5'-end were added for efficient restriction. pTEc2 also no longer contains the EM7 promoter, since it was substituted by the stronger constitutive glnS' promoter, as well as the *rho T* independent terminator in a two-step cloning procedure: i) PCR amplification of the *SpeI*-glnS'-BpaRS-rhoT-*SmaI* fragment using the 5'-atatatatatatactagtcatcaatcatccccataatccttg-3' / 5'-atatatatatatcccggg-aaaagcagaaaaaacgccgc-3' primers, followed by cloning between the previously double-digested *SpeI* and *SmaI* sites (underlined) of pTEc1. This introduces the new promoter and terminator via the elimination of the old promoter and the orthogonal MetRS. Additional random nucleotides at the 5'-end were added for efficient restriction. This is followed by ii) PCR amplification of the orthogonal MetRS fragment using the 5'-atatatatatatcatatggcaagctggagcc-3' / 5'-atatatatatatctgcagttaccttaaaagatcaggtctttctttttc-3' primers and its cloning between the previously double-digested *NdeI* and *PstI* sites (underlined) of pTEc1 (additional random nucleotides at the 5'-end were added for efficient restriction.) A Strep-Tag II sequence was appended to the orthogonal MetRS N-terminus between two linker sequences for immunoblotting detection. See Appendix 9.7 for sequence details.

- pTEc2: Derived from pTEc1.1G-R/L, this vector no longer contains the original sequence of the orthogonal MetRS, since it was substituted by a codon bias optimized sequence for expression in *E. coli*. The orthogonal MetRS was designed so that it contains the *NdeI*-linker-Strep·TagII-*XhoI*- sequence at its N-terminus and the -*PstI*-linker-rhoT-*SmaI* sequence at its C-terminus. These should allow its efficient detection by immunoblotting, as well as its efficient substitution by cloning. The entire fragment was synthesized and cloned between the previously double-digested *EcoRV* sites of the pUC57 (Genscript; New Jersey, US) vector. Finally, the whole plasmid was digested with *NdeI* and *SmaI* and the resulting fragment subcloned into pTEc1.1G-R/L. See Appendix 9.8 for sequence details.

6.1.9 Software

- Origin 6.1G (OriginLab Corporation, Northampton, MA, USA) for data analysis.
- ApE for *in silico* cloning.[175]
- OptimumGene Codon Optimization Analysis (GeneScript, New Jersey, US).[154]

6.2 Methods

6.2.1 Molecular biology

6.2.1.1 PCR

The Polymerase Chain Reaction (PCR) allows the amplification of specific nucleic acid sequences. For each reaction, 10 ng/µl target DNA sequence, 1x polymerase buffer, 250 µM of each dNTP, 10 pmol/µl of forward and reverse primer and 0.02 U/µL of ExTaq polymerase in 50 µl were used. The cycle is defined by a nucleic acid denaturation step at 95 °C for 5 min, followed by 25-30 cycles of DNA denaturation at 95 °C for 1 min, primer annealing at 55-60 °C for 1 min, and DNA elongation at 72 °C and 1 kb/min. A 10-min elongation period was applied.

6.2.1.2 Cloning

The DNA generated from PCR was purified with either the QIAquick PCR Purification Kit or Gel Extraction Kit from Qiagen. PCR fragments and vectors in concentrations of 0.5 to 2 µg were double-digested with 10 U of the appropriate restriction enzymes in buffers with or without 10 µg BSA. The reactions were performed in 50-100 µl and incubated at 37 °C for 2-4 hrs, with the exception of reactions involving *SmaI*, for which the optimal temperature is 25 °C. After double-digestion, DNA fragments and vectors were purified with the QIAquick PCR Purification Kit and ligation was performed using 25 ng vector DNA combined with insert DNA having a 5x molar concentration with respect to the vector, 1x T4 ligase buffer, and 200 Units of T4

ligase. The ligation was performed in a 20 µl volume and incubated at 16 °C ON.

6.2.1.3 SDS-PAGE

In Sodium Dodecyl Sulfate Poly-Acrylamide Gel Electrophoresis or SDS-PAGE, proteins are separated according to their molecular mass in a cross-linked polyacrylamide matrix. SDS denatures proteins and surrounds this with a uniform negative charge, making its migration both charge- and conformation-independent. Electrophoresis was performed at 130 V for ca. 1 hr in 12% or 17% poly-acrylamide gels. Protein bands were visualized using the Coomassie staining solution, followed by destaining at RT for 1 h (or ON).

6.2.1.4 Western Blot (Protein immunoblotting)

Western blotting exploits the specific binding of antibodies for the detection of target proteins. After SDS-PAGE, proteins are transferred to a nitrocellulose membrane and the target protein is detected using an antibody that binds specifically to it. This primary antibody is subsequently detected using a secondary antibody linked to an enzyme, typically either horseradish peroxidase (HRP) or alkaline phosphatase (AP). A substrate is then added, and forms a precipitate or a luminescent product at the sites where binding occurred. Here, the HRP-luminal system was used, and the luminescent product is detected in a light sensitive photo paper. Western blotting was performed by incubating SDS-PAGE gels with nitrocellulose membrane covered with two pieces of pre-soaked Whatman paper on both sides for 5 min in transfer buffer. Electrophoresis was then performed at 200 mA for 2 hr. The membrane was then stained with Ponceau S to check the efficiency of protein transfer. After destaining with H_2O, the membrane was incubated in TBS buffer containing 1.5% BSA for 1 h at RT to block non-specific antibody binding sites on the membrane. The membrane was subsequently incubated ON with a murine anti-Strep-Tag II antibody (1:1000 in TBS buffer) at 4 °C. The membrane was then washed with TBS buffer thrice for 10 min and incubated with the secondary anti-mouse antibody (1:4000 in TBS buffer) for

1 hr at RT. This was followed by three more washing steps. All incubation and washing steps were done with low speed shaking. The luminescent reaction was initiated by incubating the membrane with 1:1 luminol:peroxide reagent for 5 min. Detection was performed on the membrane exposed to x-film paper for 5-60 min using the X-OMAT 1000 processor.

6.2.2 Microbiology

6.2.2.1 Production and transformation of electrocompetent cells

Cells were incubated ON at 37 °C in 5 mL of LB medium. Cells were subsequently incubated with 500 mL of LB medium inoculated with a tenth of the previous culture at 37 °C, with shaking at 220 rpm. Incubation is continued until the cells reach an OD_{600} between 0.6 and 0.8. The cells were then harvested, washed twice with ice-cold 10% glycerol and resuspended in 5 mL 10% glycerol. Finally, 100 µL aliquots were frozen in liquid nitrogen and kept at -80 °C until further use.

Thereafter, cells were transformed by electroporation. 50 µL of competent cells aliquot were typically mixed with ~100 ng of the desired plasmid in an electroporation cuvette. A current of 1650 V is then applied on the cell suspension. This is followed by the addition of 1 mL LB medium, and the whole is transferred to a sterile eppendorf tube, which is incubated at 37 °C for 1 h and with shaking at 1000 rpm. Finally, cells are plated on agar supplemented with the appropriate antibiotics. Single colonies that appear the next day are transferred to 5 mL antibiotic-supplemented LB medium and incubated over night at 37 °C and 200 rpm.

6.2.2.2 Limitation test

To successfully incorporate non-canonical amino acids during protein expression, it is absolutely necessary to use strains auxotrophic for the corresponding canonical amino acid. It is important to know the ‘degree of auxotrophy’ by determining the

required amino acid concentration to obtain high biomass production. This can be assessed using a limitation test, where different concentrations of non-canonical amino acid are used and growth is measured. For this, 5 mL NMM supplemented with the appropriate antibiotics and different concentrations of Met were inoculated with 5 µL of pre-culture. Suspensions were incubated ON at 37 °C and 200 rpm. Cell growth was determined the next day by measuring the OD_{600}. The amino acid concentration that allows cells growing up to a 0.6-0.8 OD_{600} was used as limiting concentration for high biomass production and are used as a reference in optimal incorporation experiments.

6.2.2.3 Small-scale expression of lipase, MetRSs, and barstar

To select the best expression clone, a small-scale protein expression test is required. For this, 5 mL of three to five ON cell LB cultures supplemented with the appropriate antibiotics are transferred to a sterile eppendorf tube, where protein expression is induced by adding 1 mM IPTG. All cell cultures were incubated at 30 °C and 200 rpm. Non IPTG-induced samples corresponding to 1 OD_{600} were prepared in parallel. After 4 hours or ON incubation, 1 OD_{600} of cells were harvested, resuspended in 40 µL H_2O containing 10 µL of 5x SDS-sample buffer and heated to 95 °C for 5 min. Protein expression is assessed by SDS-PAGE. Clones exhibiting the highest expression were selected for large-scale protein/congener expression.

6.2.2.4 Large-scale expression of lipase congeners

For the expression of lipase congeners, one tenth of the best expressing clone was inoculated into 1 L of NMM supplemented with a limiting amount of Met and the appropriate antibiotics. After Met depletion in the mid-log phase (OD_{600} 0.6–0.8), which can be observed as a growth arrest, 5 mM of either Met (parent TTL expression) or Aha and Nle (congener TTL expression) were added to the media 15 min prior to induction. Gene expression was induced by the addition of 1 mM IPTG for 4-6 h at 30 °C with vigorous shaking.

6.2.2.5 Purification of lipase congeners

After soluble TTL expression, cells were harvested by low speed centrifugation (3,200 xg, 4 °C, 10 min) and the cell pellet was resuspended in Ni-NTA lysis buffer and 0.1 % Triton X-100. After addition of DNAse, RNAse, and lysozyme, each with a 1 mg/mL concentration, cells were ruptured by sonication and the homogenate cleared from cell debris by high speed centrifugation (30,000 xg) for 30 min at 4 °C. The clear lysate was loaded onto a 5 mL HiTrap Chelating HP column, which was then washed with 10 CV of Ni-NTA low-salt washing buffer, followed by 5 CVs of Ni-NTA high-salt washing buffer, and again 5 CVs of Ni-NTA low-salt washing buffer. Bound proteins were eluted by using an imidazole gradient (20-500 mM) with the Ni-NTA elution buffer. The elution fractions were analyzed by SDS-PAGE and those enriched in the desired congeners were pooled, dialyzed against lipase buffer, and concentrated by ultra-filtration with Vivaspin 20 having a MW cutoff of 10,000 Da. Finally, the concentrated samples were assayed for protein content using the Bradford method, with BSA as a calibration standard.

6.2.2.6 Solubility tests and large-scale expression of MetRSs

Following the small-scale expression test and protein induction and cell harvesting, a volume corresponding to 1 OD_{600} is sonicated and centrifuged at high-speed for 30 min at 4 °C. The soluble phase is separated from the pellet and dissolved in the same volume of 7 M urea. To 40 µl of each fraction, 10 µL of 5x SDS-sample buffer was added. The samples were heated to 95 °C for 5 min. Protein solubility was assessed by SDS-PAGE. For expression of MetRSs, one tenth of the best expressing clone was inoculated in 1 L of LB supplemented with the appropriate antibiotics. In the mid-log phase (OD_{600} 0.6–0.8), gene expression was induced by the addition of 1 mM IPTG. Cultures were maintained at 30 °C with vigorous shaking between 4 h to ON.

6.2.2.7 Purification of MetRSs proteins

After soluble MetRS expression, cells are harvested by low speed centrifugation (3,200g, 4 °C, and 10 min) and the pellet was resuspended in Ni-NTA lysis buffer. After the addition of DNAse, RNAse, and lysozyme, each at a concentration of 1 mg/mL, cells were ruptured by sonication. The homogenate is cleared of cell debris by high speed centrifugation (30,000 xg) for 30 min at 4 °C. The clear lysate was loaded onto a 1 mL HiTrap Chelating HP column, which was then washed with 10 CVs of Ni-NTA low-salt washing buffer. Bound proteins were eluted using an imidazole gradient (20-300 mM) with the Ni-NTA elution buffer. Elution fractions were analyzed by SDS-PAGE and those enriched in the desired congeners were pooled, dialyzed against MetRS buffer, and concentrated by ultra-filtration with Vivaspin 20 with a MW cutoff of 50,000 Da.

6.2.2.8 Large-scale expression of barstar congeners

For expression of barstar congeners, the same protocols used in lipase expression were applied (see section 6.2.2.4), with the modification of adding different concentrations of Aha:Eth (i.e. for a 3:1 ratio, 0.375 mM Aha:0.125 mM Eth; for a 1:1 ratio, 0.25 mM Aha:0.25 mM Eth; and for a 1:3 ratio, 0.125 mM Aha: 0.375 mM Eth) in a final volume of 200 mL of NMM instead.

6.2.2.9 Purification of barstar congeners

After insoluble barstar expression, cells were harvested by low-speed centrifugation (4,000 xg) for 20 min at 4 °C and the cell pellet was resuspended in barstar buffer. After addition of 1 mg/mL lysozyme, cells were ruptured by sonication for 3 min and harvested by high-speed centrifugation (30,000 xg) for 40 min at 4 °C. The supernatant was discarded and the cell pellet, which contains inclusion bodies, was dissolved in IE resuspension buffer. The suspension was separated by high-speed centrifugation (30, 000 xg) for 40 min at 4 °C and the supernatant was transferred into

dialysis tubes (3500 Da cutoff). The first dialysis was performed for 3 h, the second ON and the third for 3 h at 4 °C. All dialyses were performed in IE washing buffer. Solid particles were subsequently separated by high-speed centrifugation (30, 000 xg) for 40 min at 4 °C and the supernatant was passed through a 0.22 µm filter. The clear lysate was loaded onto a 5 mL HiTrap Q Sepharose column, which was then washed with 5 CVs of IE washing buffer. Bound proteins were eluted by using a NaCl gradient (100-1000 mM) with the IE elution buffer. The elution fractions were analyzed by SDS-PAGE and those enriched in the desired congeners were pooled, dialyzed against barstar buffer and concentrated by ultra-filtration with Vivaspin 10 with a MW cutoff of 5,000 Da.

6.2.3 Biochemistry

6.2.3.1 tRNA aminoacylation assay

The tRNA aminoacylation assay is used to determine the rate of amino acid (aa) aminoacylation by an aminoacyl-tRNA synthetase (AARS).[138] In the reaction, the aminoacylated tRNA (aa-tRNA) is formed by transferring the aa, in form of aminoacyladenylate (AMP-aa), to the tRNA. In this assay, radioactively labeled aa (^{35}S-aa) was used to form [^{35}S]-aa-tRNA, which is precipitated onto a filter and separated from the eluate. The activity of the AARS with respect to its cognate aa and tRNA is detected directly by the rate of formation of [^{35}S]-aa-tRNA. The reaction mix (final volume: 30 µl) was incubated at 37 °C in 50 mM HEPES (pH 7.4), 10 mM $MgCl_2$, 50 mM KCl, 3 mM BME, 5 mM ATP, 1 mg/mL BSA, 5 nM [^{35}S]-Met (3 cpm/µmol), 10 -200 µM tRNA, and 1 µM of MetRS. After 15 or 20 min, 20 µl of the reaction mixture was transferred onto a filter and the tRNA was precipitated by adding 1 mL of 10% TCA, followed by washing steps with 1 mL of 5% TCA, 1 mL of absolute EtOH and 1 mL of a 1:1 solution EtOH:Ether. Filters were dried for 5 min and mixed with scintillation solution to determine the amount of radioactivity.

6.2.3.2 ATP:PPi exchange assay

This assay is used to determine the activation rate of an amino acid (aa) by an aminoacyl-tRNA synthetase (AARS).[142] In the reaction, the aminoacyladenylate (AMP-aa) is formed by transferring the AMP group of ATP, accompanied by the release of pyrophosphate (PPi). In the reverse reaction, radioactively labeled PPi with ^{32}P will form [^{32}P]-ATP, the amount of which can be measured. Activated charcoal (aC), which has affinity towards ATP, is used to adsorb [^{32}P]-ATP and separate it from the eluate [^{32}P]-PPi. The activity of the AARS is detected directly by the rate of formation of [^{32}P]-ATP. The reaction mix was incubated at 37 °C in 100 mM Tris·HCl (pH 8.0), 80 mM $MgCl_2$, 5 mM KF, 700 mM BME, 5.5 mM ATP, 0.1 mg/mL BSA, 2.2 mM [^{32}P]-PPi (0.2 cpm/pmol), 5 µM of MetRS, and 5 mM of the L-isomer analogs of Met in a final volume of 200 µl. After 15 or 20 min, 100 µl of the reaction solution was added to 600 µl of 240 mM sodium pyrophosphate solution containing 70% (v/v) perchloric acid. [^{32}P]-ATP formation was followed by specific absorption of 200 µl of 7.5% (w/v) activated charcoal. The suspension was thoroughly mixed and filtered through Whatman GF/F paper. Filters were washed twice with 10 mL of water and mixed with scintillation solution to determine the amount of radioactivity.

6.2.4 Spectroscopy and spectrometry

6.2.4.1 UV/VIS-Spectroscopy

Protein concentrations were determined by the measurement of absorbance at 280 nm, and were calculated according to the Lambert-Beer-Equation $A = \varepsilon \, d \, c$, where A = absorbance; ε = molar extinction coefficient; c = concentration; and d = path length.

6.2.4.2 Bradford protein assay

The Bradford reagent is composed of Coomassie brilliant G250 dye, which forms blue complexes with the positively charged amino groups of proteins in acidic solution. The

dye complex absorbs light at 595 nm; the dye intensity is proportional to protein concentration. All samples are measured at this wavelength and compared against a calibration curve generated using a BSA with known concentrations. For each measurement, 10 µl of sample is mixed with 990 µl of Bradford Reagent and incubated at RT for 5 min, then measured at 595 nm.

6.2.4.3 Electro Spray Ionization Mass Spectrometry (ESI-MS)

For Liquid Chromatography ESI-MS, 20 µL aliquots of the sample were pre-separated on a Waters RP C4 column (300 Å pore size; 3.5 µm particle size; 100 x 2.1 mm) by a 20 minute elution using a gradient from 20 to 90 % 0.05 % (v/v) TFA in acetonitrile in 0.05 % (v/v) TFA in water, with a flow rate of 250 µl/min. The masses of the eluted fractions were analyzed on a MicrOTOF ESI-MS.

6.2.4.4 Spectrophotometric assay with *p*-nitrophenyl palmitate

Lipase activity was determined by measuring the hydrolysis of *p*-nitrophenyl palmitate (*p*NPP). Cleavage of *p*NPP was determined at different temperatures (40-90 °C) in 25 mM Tris·HCl pH 8 according to the protocol of Winkler and Stuckmann.[101] A buffered pNPP suspension containing 25 mM Tris·HCl pH 8.0, 1 mM *p*NPP and 1 mg/mL gum arabic (Acros Organics, Geel, Belgium) was homogenized at 22,000 rpm for 4 min at RT. The reaction was started by mixing 900 µl buffered pNPP suspension with 100 µl of the TTL preparations. Autohydrolysis was assessed by including a blank that contained the same volume of 50 mM Tris·HCl pH 8.0 instead of enzyme. The reaction mixture was incubated at the desired temperature according to the protein for 10 or 15 min with vigorous shaking. Enzymatic hydrolysis of *p*NPP was stopped by the addition of 100 µl of 1 M Na_2CO_3 and chilling on ice for 1 min. Following centrifugation at 20,000g for 10 min at RT, the absorption of the supernatant was measured at 410 nm. All values were determined in triplicates and corrected for autohydrolysis. One unit (1 U) of lipase activity is defined as the amount of enzyme necessary to liberate 1 µmol of *p*NPP (ε at pH 8.0 = 12.75 x 106 M^{-1} cm^{-1}) per minute

under the conditions described above.

6.2.4.4.1 Influence of temperature on lipase activity

The optimal temperature of lipase activity was determined with the standard *p*NPP assay at pH 8. The samples were incubated for 10 min at temperatures between 40 and 90 °C. To determine thermostability, 100 µl lipase solutions were incubated at 75, 80, 85, 90, or 95 °C at different time intervals ranging from 10 and 180 min.

6.2.4.4.2 Influence of additives on lipase activity

To investigate the effects of various substances on lipase activity, the produced congeners were incubated in solutions containing 90% organic solvent, 10% surfactant, or 10 mM inhibitor (500 mM guanidinium chloride and 2M urea) at RT for 60 min. Lipase residual activity was subsequently determined using the *p*NPP assay at the optimal temperature for 10 min at pH 8.

6.2.4.4.3 Substrate specificity for *p*-nitrophenyl esters

The reactions were performed according to the spectrophotometric assay with *p*NPP as described above but using the following *p*-nitrophenyl-derived (*p*NP) substrates (Sigma) at a concentration of 10 mM: acetate (C2:0), butyrate (C4:0), caproate (C6:0), caprylate (C8:0), caprate (C10:0), laurate (C12:0), myristate (C14:0), palmitate (C16:0), and stearate (C18:0). The reactions were incubated at the enzyme's temperature optimum and pH 8 for 10 min.

6.2.4.5 Spectrophotometric assay with tricaprylin

To determine the optimal pH of the lipase enzymes, a modified assay using the formation of copper soaps for detection of free fatty acids was used.[176] The substrate solution was comprised of 10 mM tricaprylin (1,2,3-trioctanoylglycerol) and 5 g/L gum

Arabic in 40 mM universal pH buffer with the desired pH 7 or 8 emulsified at maximum speed for 1 min at RT. An aqueous Copper(II)-acetate-l-hydrate solution (58 mg/mL) with pyridine at pH 6.1, was used as the copper reagent. The dye reagent contained 1 mg/mL diethyldithiocarbamate dissolved in 99.8 % (v/v) ethanol. 200 µl substrate solution was combined with 100 µl enzyme solution with vigorous shaking for 18 h at the optimal temperature for the specific lipase congener. This is followed by the succeeding steps, which are all performed at RT: first, the reaction was stopped by adding 125 µl 3 M HCl. Extraction is then performed by vigorously mixing the sample with 1.5 mL isooctane for 10 min. Phases separation was achieved by spinning in a table-top centrifuge at maximum speed for 10 min. 1.25 mL of the isooctane phase were removed and mixed with 250 µl copper reagent. The emulsion was mixed at 1400 rpm for 5 min and centrifuged again at maximum speed for 5 min. 1 mL of the organic phase was withdrawn and mixed with 200 µl dye reagent. After 5 min, absorption at 430 nm was determined. The contribution of autohydrolysis was assessed by including a blank containing an equal volume of 40 mM universal pH buffer at the appropriate pH. All values were determined in triplicates and corrected for autohydrolysis.

6.2.4.5.1 Influence of pH in lipase activity

The influence of pH on lipase activity was assessed using the tricaprylin assay (see above) between pH 4 and 12 at the optimal temperature for 18 h.

6.2.4.5.2 Substrate specificity for triacylglycerols

The reactions were carried out according to the spectrophotometric assay with tricaprylin as described above but using the following substrates (Sigma) at a concentration of 10 mM: triacetin (C2:0), tributirin (C4:0), tricaproin (C6:0), tricaprylin (C8:0), tricaprin (C10:0), trilaurin (C12:0), trimyristin (C14:0), tripalmitin (C16:0), and tristearin (C18:0). The reactions were incubated at the enzyme's temperature optimum and pH 8 for 18 h.

6.2.4.6 Circular dichroism

Circular Dichroism (CD) is used to estimate the degree of protein folding. Additionally, the relative percentage of secondary structure elements in a sample (α-helix, β-sheet and random coil) can be estimated by comparison to reference spectra. The measurement is based on the differential absorbance of left and right circular polarized light, usually called ellipticity or Θ, by optical active substances at different wavelengths. A CD spectrum is the plot of Θ against the wavelength of circular polarized light. CD spectra of 0.2 mg/mL lipase samples were measured in 10 mM Tris·HCl pH 8.0. Measurements were performed in 110-QS Hellma quartz cells (optical path-length: 0.1 cm) under controlled temperature (Peltier type FDCD attachment, model PFD-350S/350L; JASCO International Co., Ltd., Tokyo, Japan). Ellipticity changes were recorded between 200 nm and 250 nm at an optimal temperature of 65 or 70 °C according to the congener.

6.2.4.7 N-terminal sequencing

The N-terminal sequencing method, or Edman sequencing, consists of the derivatization of N-terminal amino acids with phenylthiohydantoin (PTH), followed by a RP-HPLC step. Samples were subjected to N-terminal sequence analysis using gas-phase sequenator procise cLC (Applied Biosystems GmbH, Darmstadt, Germany) according to manufacturer instructions. Due to their hydrophobicity, the amino acid PTH derivatives display different quantifiable retention times. In this way, the intensities of all peaks were summed up to 100% and compared as in ESI-MS.

6.2.5 Informatics

6.2.5.1 Generation of a 3D structure model for TTL

The primary sequence of TTL (see Appendix 9.1 for details) was queried against the entire PDB database and compared to known 3D structures of structurally similar

proteins, such as lipases, esterases and serine proteases with aid of the software HHpred. This software provides a method for sequence database searching and structure prediction.[177] In addition, the sensitivity of HHpred in finding homologous sequences is comparable to the most powerful profile-profile alignment servers for structure prediction currently available, e.g. x. 100 query-template alignments were obtained from HHpred, of which four were selected according to their high score, including the esterase from the bacterium *Butyrivibrio proteoclasticus* (PDB: 2WTM)[178], the acylaminoacyl serine peptidase from *A. pernix* (PDB: 2HU5)[179], the human monoglyceride lipase (PDB: 3JW8)[180], and the putative serine hydrolase from *Xanthomonas campestris* (PDB: 3KSR). A multiple-sequence alignment of these templates were used to build 100 homology models, out of which the best scoring model was chosen with the aid of the software “Modeller”.[181]

7 List of figures

8 List of tables

9 Appendix

Met residues are indicated in red color, His·Tag residues in blue and mutated residues in green.

9.1 Primary amino acid sequence of TTL

```
  1 MQKAVEITYN GKTLRGMMHL PDDVKGKVPM VIMFHGFTGN KVESHFIFVK 50
 51 MSRALEKVGI GSVRFDFYGS GESDGDFSEM TFSSELEDAR QILKFVKEQP 100
101 TTDPERIGLL GLSMGGAIAG IVAREYKDEI KALVLWAPAF NMPELIMNES 150
151 VKQYGAIMEQ LGFVDIGGHK LSKDFVEDIS KLNIFELSKG YDKKVLIVHG 200
201 TNDEAVEYKV SDRILKEVYG DNATRVTIEN ADHTFKSLEW EKKAIEESVE 250
251 FFKKELLKGG SHHHHHH
```

9.2 Primary sequence of barstar 1M (P28A/C41A/C83A)

```
  1 MKKAVINGEQ IRSISDLHQT LKKELALAEY YGENLDALWD ALTGWVEYPL 50
 51 VLEWRQFEQS KQLTENGAES VLQVFREAKA EGADITIILS
```

9.3 Primary sequence of barstar 2M (P28A/C41A/E47M/C83A)

```
  1 MKKAVINGEQ IRSISDLHQT LKKELALAEY YGENLDALWD ALTGWVEYPL 50
 51 VLMWRQFEQS KQLTENGAES VLQVFREAKA EGADITIILS
```

9.4 MetRS alignment

The MetRS alignment was performed with aid of the software Multalin.[182]

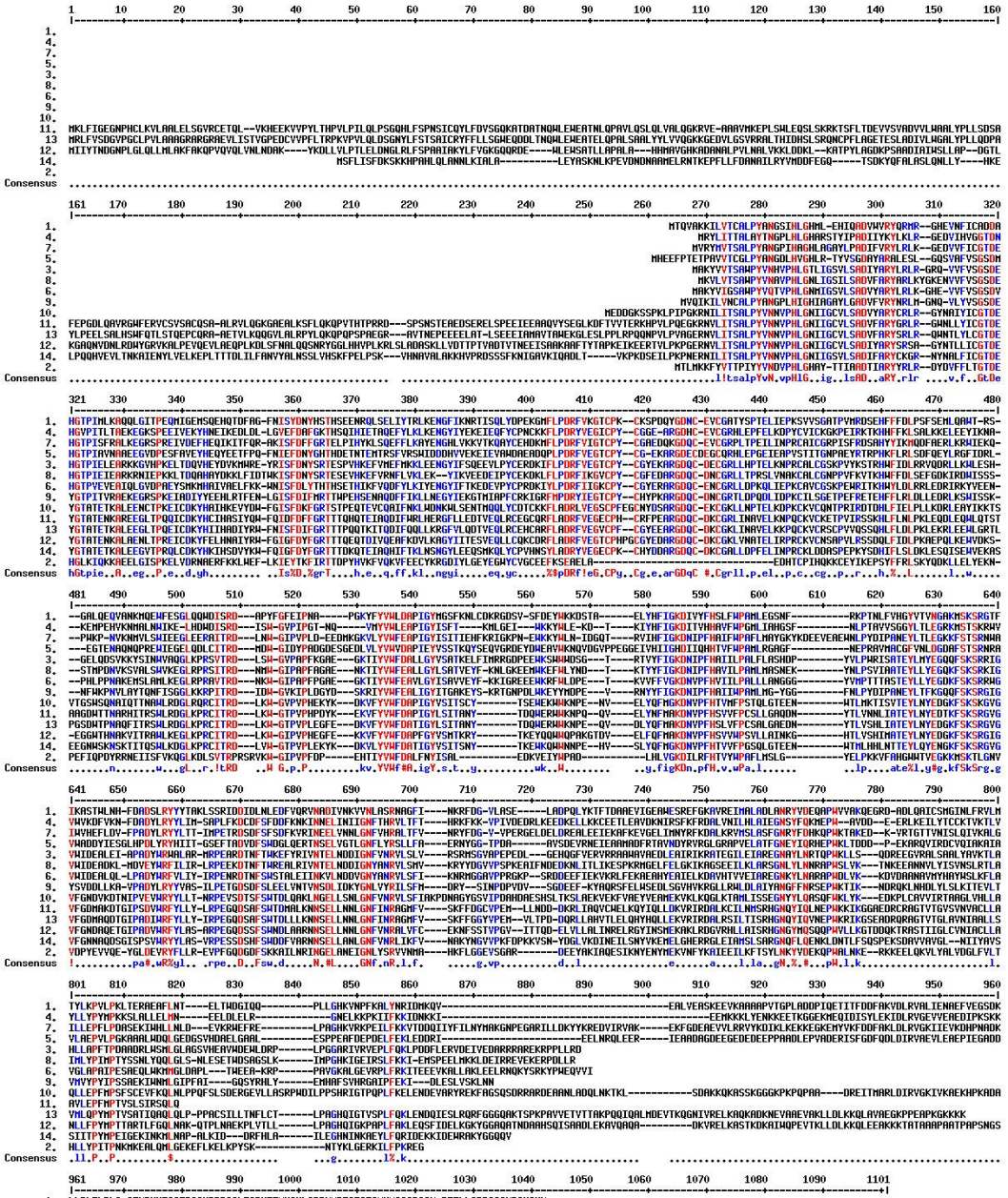
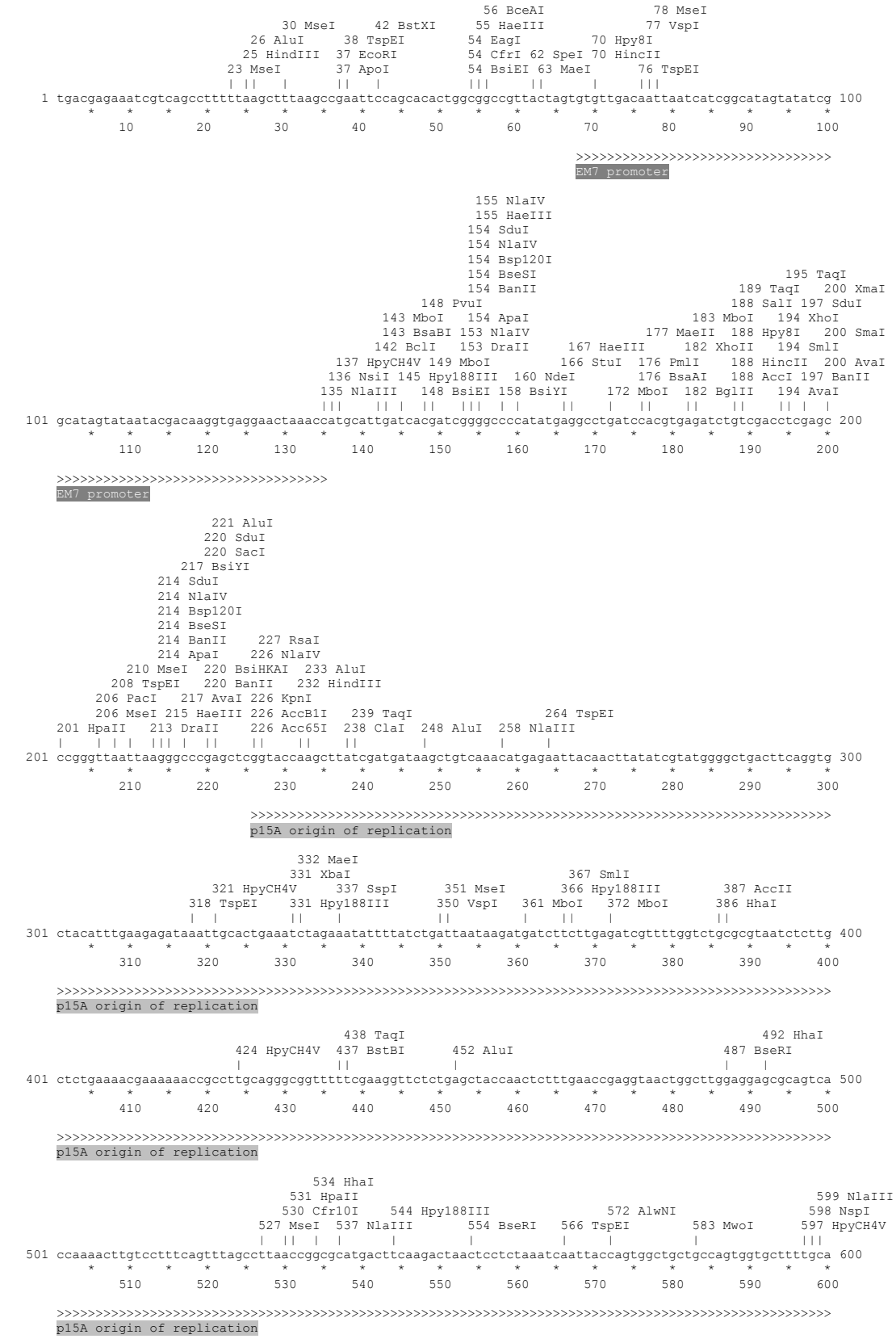
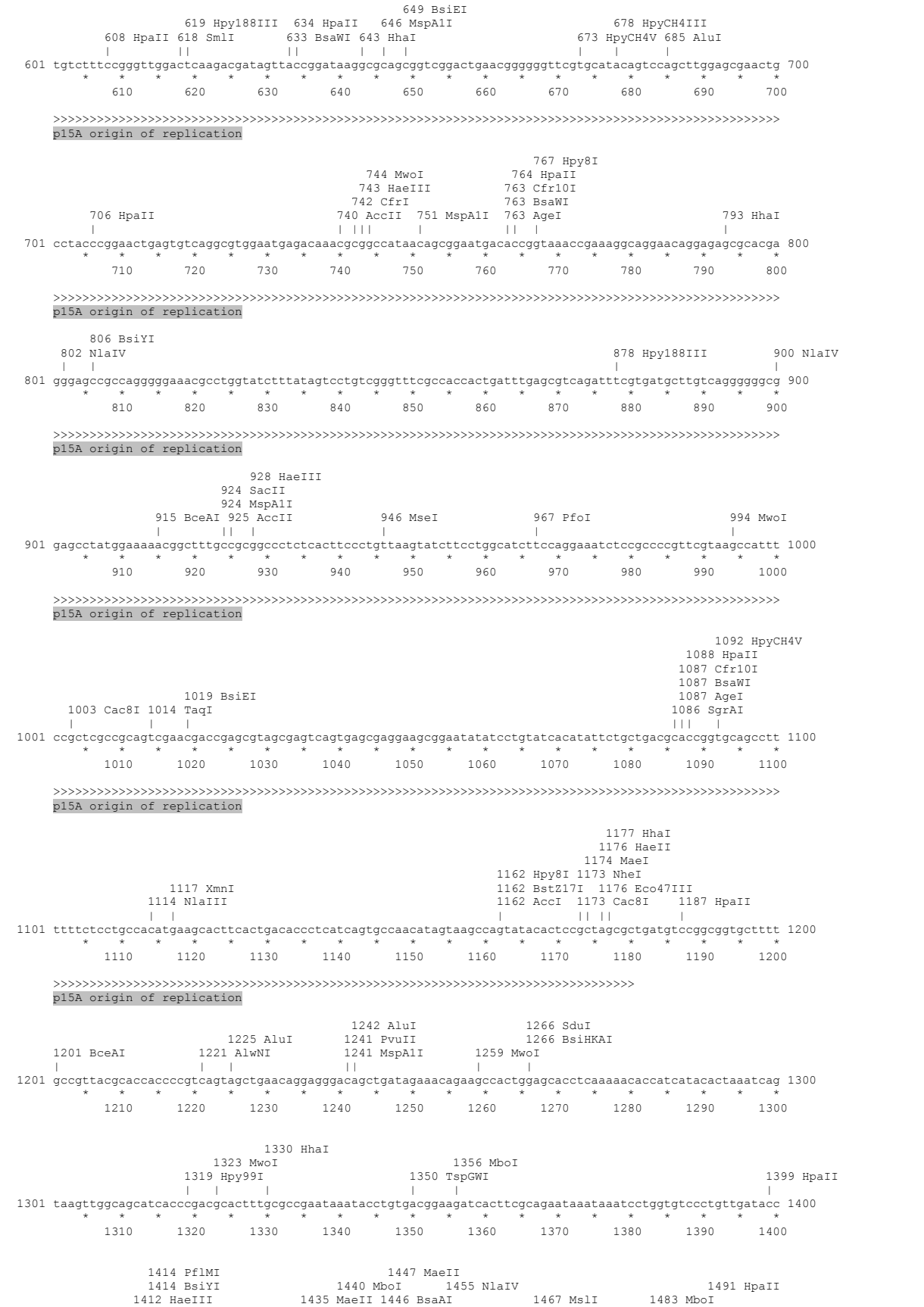
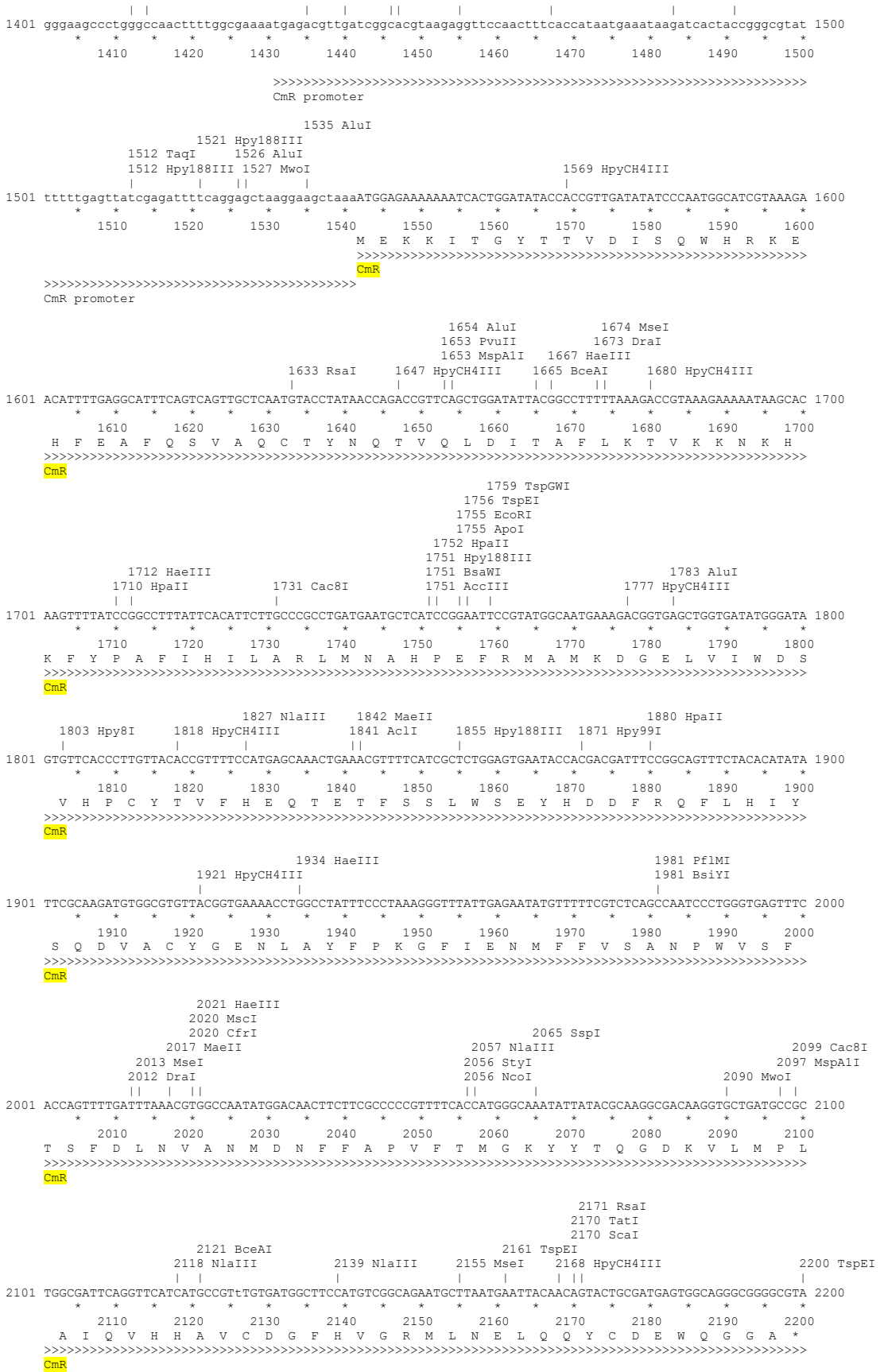


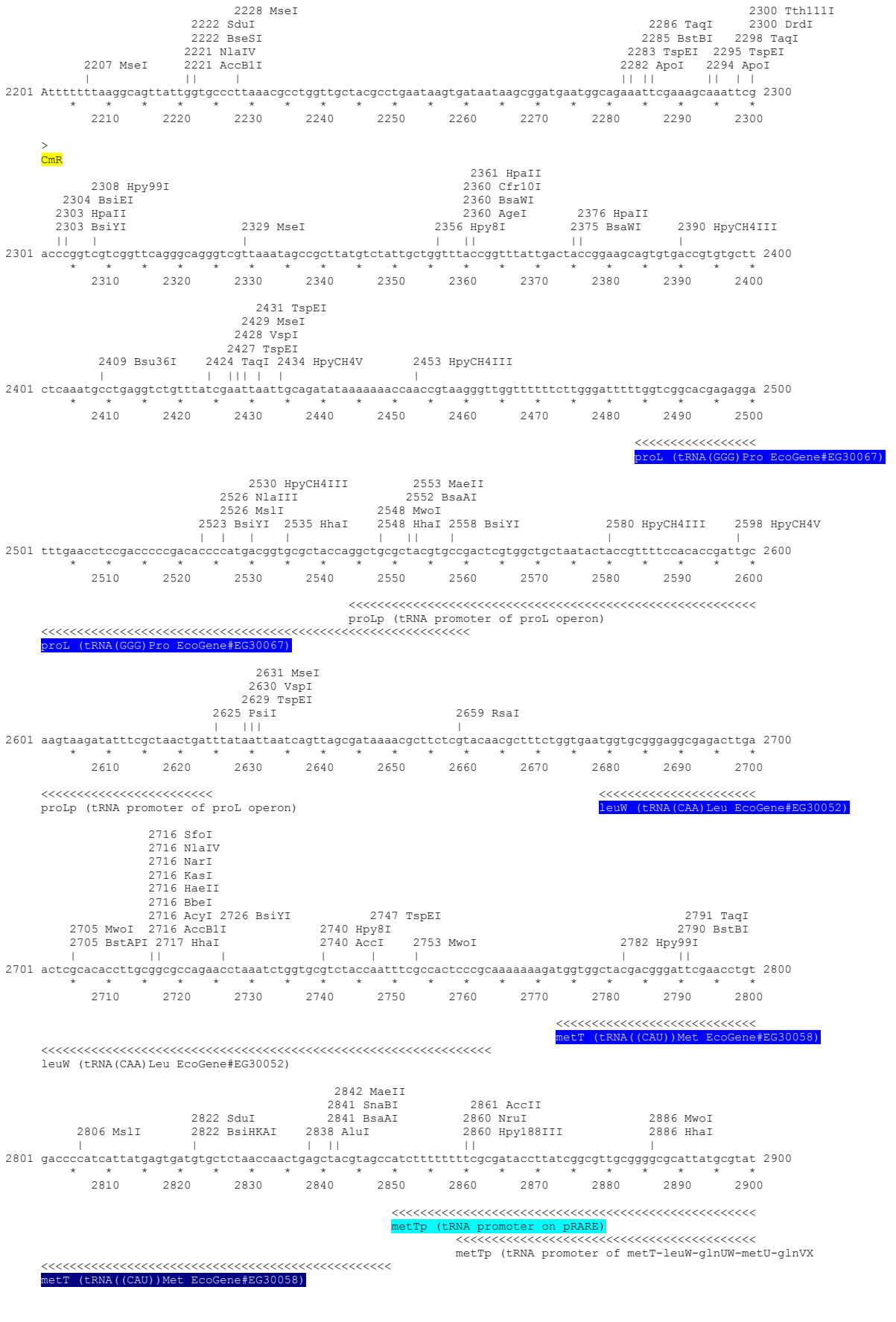
Figure 56. MetRS sequence alignment. Residues with consensus above 90% and between 50-90% are indicated in red and blue respectively. 1. *Ec*MetRS WT (677aa); 2. *Aa* MetRS WT (497aa); 3. *Ap*MetRS WT (572aa); 4. *Mj*MetRS WT (651aa); 5. *Np*MetRS WT (698aa); 6. *Pa*MetRS WT (570aa); 7. *Pab*MetRS WT (722aa); 8. *Sa*MetRS WT (571aa); 9. *Ta*MetRS WT (547aa); 10. *At*MetRS WT (797aa); 11. *Dr*MetRS WT (1022); 12. *Dm*MetRS WT (1022); 13. *Hs*MetRS WT (900aa); 14. *Sc*MetRS WT (751aa)

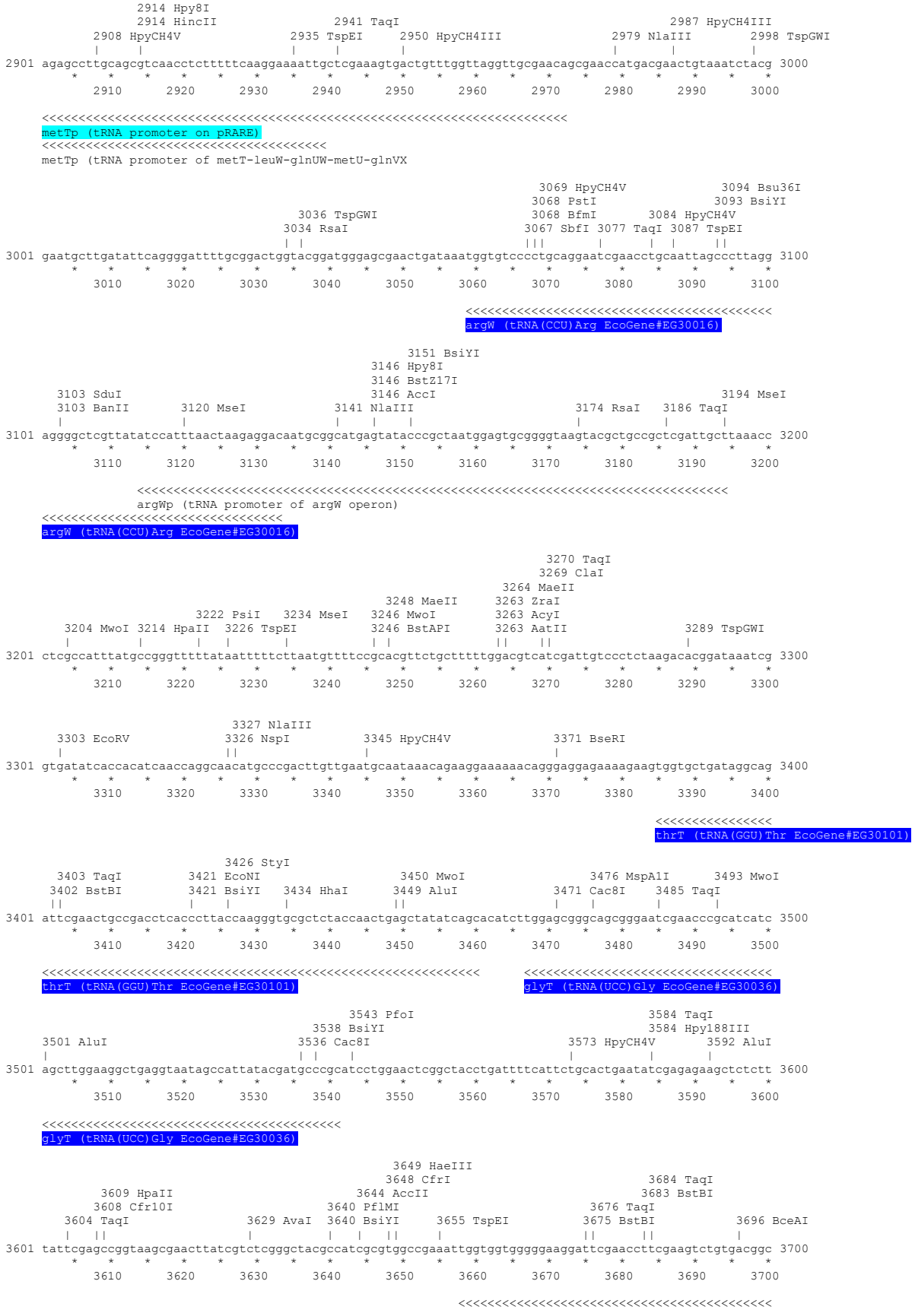
9.5 pTEc0 sequence

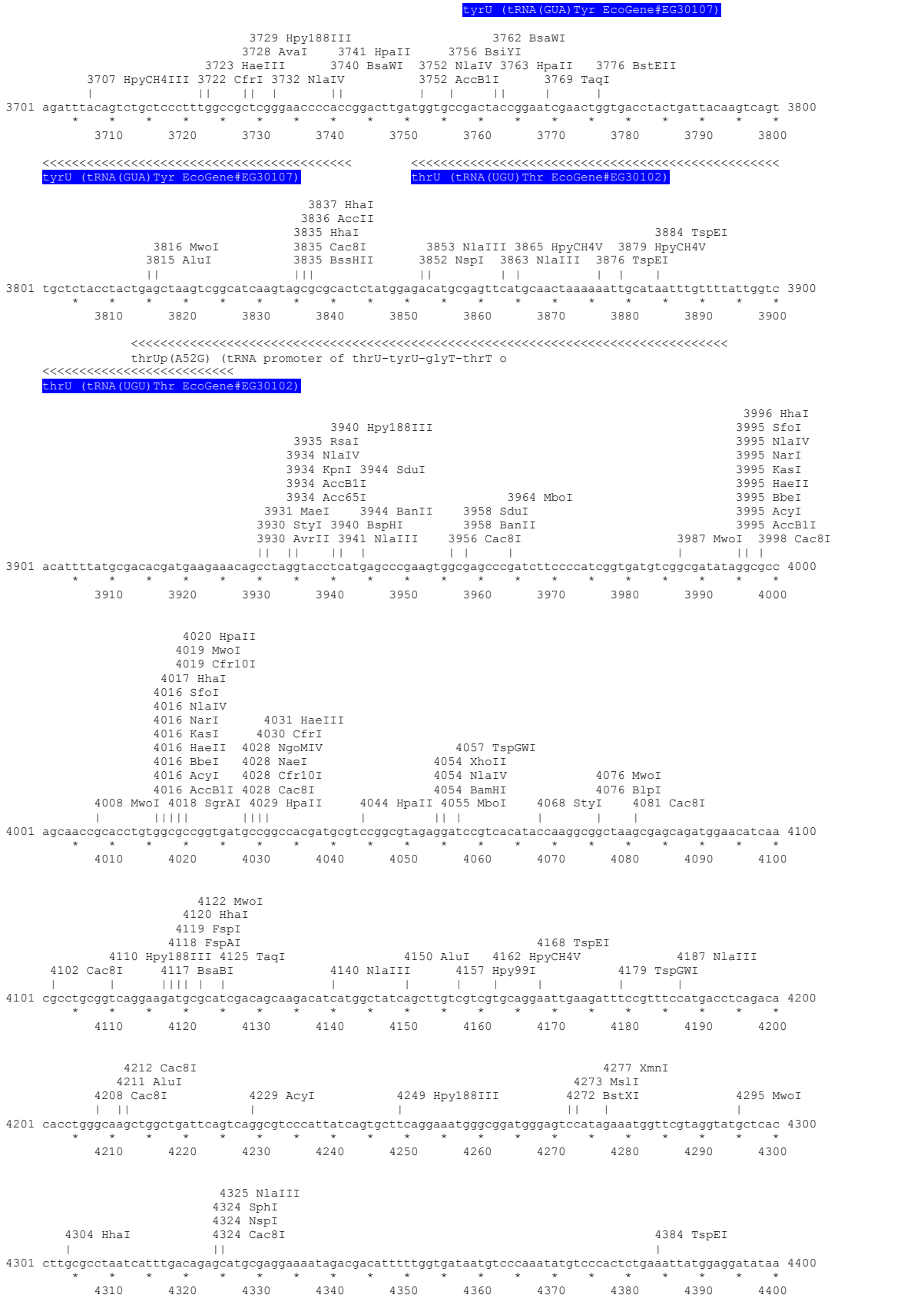



649 BsiEI
619 Hpy188III 634 HpaII 646 MspA1I 678 HpyCH4III
608 HpaII 618 SmlI 633 BsaWI 643 HhaI 673 HpyCH4V 685 AluI
601 tgtctttccgggttggactcaagacgatagttaccggataaggcgcagcggtcggactgaacgggggttcgtgcatacagtccagcttggagcgaactg 700
610 620 630 640 650 660 670 680 690 700
p15A origin of replication
767 Hpy8I
744 MwoI 764 HpaII
743 HaeIII 763 Cfr10I
742 CfrI 763 BsaWI
706 HpaII 740 AccII 751 MspA1I 763 AgeI 793 HhaI
701 cctacccggaactgagtgtcaggcgtggaatgagacaaacgcggccataacagcggaatgacaccggtaaaccgaaaggcaggaacaggagagcgcacga 800
710 720 730 740 750 760 770 780 790 800
p15A origin of replication
806 BsiYI
802 NlaIV 878 Hpy188III 900 NlaIV
801 gggagccgccaggggaaacgcctggtatctttatagtcctgtcgggtttcgccaccactgatttgagcgtcagatttcgtgatgcttgtcaggggggcg 900
810 820 830 840 850 860 870 880 890 900
p15A origin of replication
928 HaeIII
924 SacII
924 MspA1I
915 BceAI 925 AccII 946 MseI 967 PfoI 994 MwoI
901 gagcctatggaaaaacggctttgccgcggccctctcacttccctgttaagtatcttcctggcatcttccaggaaatctccgccccgttcgtaagccattt 1000
910 920 930 940 950 960 970 980 990 1000
p15A origin of replication
1092 HpyCH4V
1088 HpaII
1087 Cfr10I
1087 BsaWI
1019 BsiEI 1087 AgeI
1003 Cac8I 1014 TaqI 1086 SgrAI
1001 ccgctcgccgcagtcgaacgaccgagcgtagcgagtcagtgagcgaggaagcggaatatatcctgtatcacatattctgctgacgcaccggtgcagcctt 1100
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
p15A origin of replication
1177 HhaI
1176 HaeII
1174 MaeI
1162 Hpy8I 1173 NheI
1117 XmnI 1162 BstZ17I 1176 Eco47III
1114 NlaIII 1162 AccI 1173 Cac8I 1187 HpaII
1101 ttttctcctgccacatgaagcacttcactgacaccctcatcagtgccaacatagtaagccagtatacactccgctagcgctgatgtccggcggtgctttt 1200
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
p15A origin of replication
1242 AluI 1266 SduI
1225 AluI 1241 PvuII 1266 BsiHKAI
1201 BceAI 1221 AlwNI 1241 MspA1I 1259 MwoI
1201 gccgttacgcaccaccccgtcagtagctgaacaggagggacagctgatagaaacagaagccactggagcacctcaaaaacaccatcatacactaaatcag 1300
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
1330 HhaI
1323 MwoI 1356 MboI
1319 Hpy99I 1350 TspGWI 1399 HpaII
1301 taagttggcagcatcacccgacgcactttgcgccgaataaatacctgtgacggaagatcacttcgcagaataaataaatcctggtgtccctgttgatacc 1400
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
1414 PflMI 1447 MaeII
1414 BsiYI 1440 MboI 1455 NlaIV 1491 HpaII
1412 HaeIII 1435 MaeII 1446 BsaAI 1467 MslI 1483 MboI


1401 gggaagccctgggccaacttttggcgaaaatgagacgttgatcggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgggcgtat 1500
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
CmR promoter
1535 AluI
1521 Hpy188III
1512 TaqI 1526 AluI
1512 Hpy188III 1527 MwoI
1569 HpyCH4III
1501 tttttgagttatcgagattttcaggagctaaggaagctaaaATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGA 1600
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
M E K K I T G Y T T V D I S Q W H R K E
CmR
CmR promoter
1654 AluI
1653 PvuII
1653 MspA1I
1674 MseI
1673 DraI
1667 HaeIII
1633 RsaI
1647 HpyCH4III
1665 BceAI
1680 HpyCH4III
1601 ACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCAC 1700
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
H F E A F Q S V A Q C T Y N Q T V Q L D I T A F L K T V K K N K H
CmR
1759 TspGWI
1756 TspEI
1755 EcoRI
1755 ApoI
1752 HpaII
1751 Hpy188III
1712 HaeIII
1710 HpaII
1731 Cac8I
1751 BsaWI
1751 AccIII
1783 AluI
1777 HpyCH4III
1701 AAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATA 1800
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
K F Y P A F I H I L A R L M N A H P E F R M A M K D G E L V I W D S
CmR
1827 NlaIII
1842 MaeII
1880 HpaII
1803 Hpy8I
1818 HpyCH4III
1841 AclI
1855 Hpy188III
1871 Hpy99I
1801 GTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATA 1900
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
V H P C Y T V F H E Q T E T F S S L W S E Y H D D F R Q F L H I Y
CmR
1934 HaeIII
1921 HpyCH4III
1981 PflMI
1981 BsiYI
1901 TTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTC 2000
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
S Q D V A C Y G E N L A Y F P K G F I E N M F F V S A N P W V S F
CmR
2021 HaeIII
2020 MscI
2020 CfrI
2065 SspI
2017 MaeII
2057 NlaIII
2099 Cac8I
2013 MseI
2056 StyI
2097 MspA1I
2012 DraI
2056 NcoI
2090 MwoI
2001 ACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGC 2100
2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
T S F D L N V A N M D N F F A P V F T M G K Y Y T Q G D K V L M P L
CmR
2171 RsaI
2170 TatI
2170 ScaI
2121 BceAI
2161 TspEI
2118 NlaIII
2139 NlaIII
2155 MseI
2168 HpyCH4III
2200 TspEI
2101 TGGCGATTCAGGTTCATCATGCCGTtTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTA 2200
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
A I Q V H H A V C D G F H V G R M L N E L Q Q Y C D E W Q G G A *
CmR


2228 MseI
2300 Tth111I
2222 SduI
2300 DrdI
2222 BseSI
2286 TaqI
2285 BstBI
2298 TaqI
2221 NlaIV
2283 TspEI
2295 TspEI
2207 MseI
2221 AccB1I
2282 ApoI
2294 ApoI
2201 Attttttttaaggcagttattggtgcccttaaacgcctggttgctacgcctgaataagtgataataagcggatgaatggcagaaattcgaaagcaaattcg 2300
2210
2220
2230
2240
2250
2260
2270
2280
2290
2300
>
CmR
2361 HpaII
2308 Hpy99I
2360 Cfr10I
2304 BsiEI
2360 BsaWI
2303 HpaII
2360 AgeI
2376 HpaII
2303 BsiYI
2329 MseI
2356 Hpy8I
2375 BsaWI
2390 HpyCH4III
2301 acccggtcgtcggttcagggcagggtcgttaaatagccgcttatgtctattgctggtttaccggtttattgactaccggaagcagtgtgaccgtgtgctt 2400
2310
2320
2330
2340
2350
2360
2370
2380
2390
2400
2431 TspEI
2429 MseI
2428 VspI
2427 TspEI
2409 Bsu36I
2424 TaqI
2434 HpyCH4V
2453 HpyCH4III
2401 ctcaaatgcctgaggtctgtttatcgaattaattgcagatataaaaaaaccaaccgtaagggttggtttttttcttgggatttttggtcggcacgagagga 2500
2410
2420
2430
2440
2450
2460
2470
2480
2490
2500
<<<<<<<<<<<<<<<<<
proL (tRNA(GGG)Pro EcoGene#EG30067)
2530 HpyCH4III
2553 MaeII
2526 NlaIII
2552 BsaAI
2526 MslI
2548 MwoI
2523 BsiYI
2535 HhaI
2548 HhaI
2558 BsiYI
2580 HpyCH4III
2598 HpyCH4V
2501 tttgaacctccgacccccgacaccccatgacggtgcgctaccaggctgcgctacgtgccgactcgtggctgctaatactaccgttttccacaccgattgc 2600
2510
2520
2530
2540
2550
2560
2570
2580
2590
2600
<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
proLp (tRNA promoter of proL operon)
<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
proL (tRNA(GGG)Pro EcoGene#EG30067)
2631 MseI
2630 VspI
2629 TspEI
2625 PsiI
2659 RsaI
2601 aagtaagatatttcgctaactgatttataattaatcagttagcgataaaacgcttctcgtacaacgctttctggtgaatggtgcgggaggcgagacttga 2700
2610
2620
2630
2640
2650
2660
2670
2680
2690
2700
<<<<<<<<<<<<<<<<<<<<<<<<<
proLp (tRNA promoter of proL operon)
<<<<<<<<<<<<<<<<<<<<<<<<<
leuW (tRNA(CAA)Leu EcoGene#EG30052)
2716 SfoI
2716 NlaIV
2716 NarI
2716 KasI
2716 HaeII
2716 BbeI
2716 AcyI
2726 BsiYI
2747 TspEI
2791 TaqI
2790 BstBI
2705 MwoI
2716 AccB1I
2740 Hpy8I
2705 BstAPI
2717 HhaI
2740 AccI
2753 MwoI
2782 Hpy99I
2701 actcgcacaccttgcggcgccagaacctaaatctggtgcgtctaccaatttcgccactcccgcaaaaaaagatggtggctacgacgggattcgaacctgt 2800
2710
2720
2730
2740
2750
2760
2770
2780
2790
2800
<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
metT (tRNA((CAU))Met EcoGene#EG30058)
<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
leuW (tRNA(CAA)Leu EcoGene#EG30052)
2842 MaeII
2841 SnaBI
2861 AccII
2822 SduI
2841 BsaAI
2860 NruI
2886 MwoI
2806 MslI
2822 BsiHKAI
2838 AluI
2860 Hpy188III
2886 HhaI
2801 gaccccatcattatgagtgatgtgctctaaccaactgagctacgtagccatctttttttttcgcgataccttatcggcgttgcggggcgcattatgcgtat 2900
2810
2820
2830
2840
2850
2860
2870
2880
2890
2900
<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
metTp (tRNA promoter on pRARE)
<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
metTp (tRNA promoter of metT-leuW-glnUW-metU-glnVX
<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
metT (tRNA((CAU))Met EcoGene#EG30058)

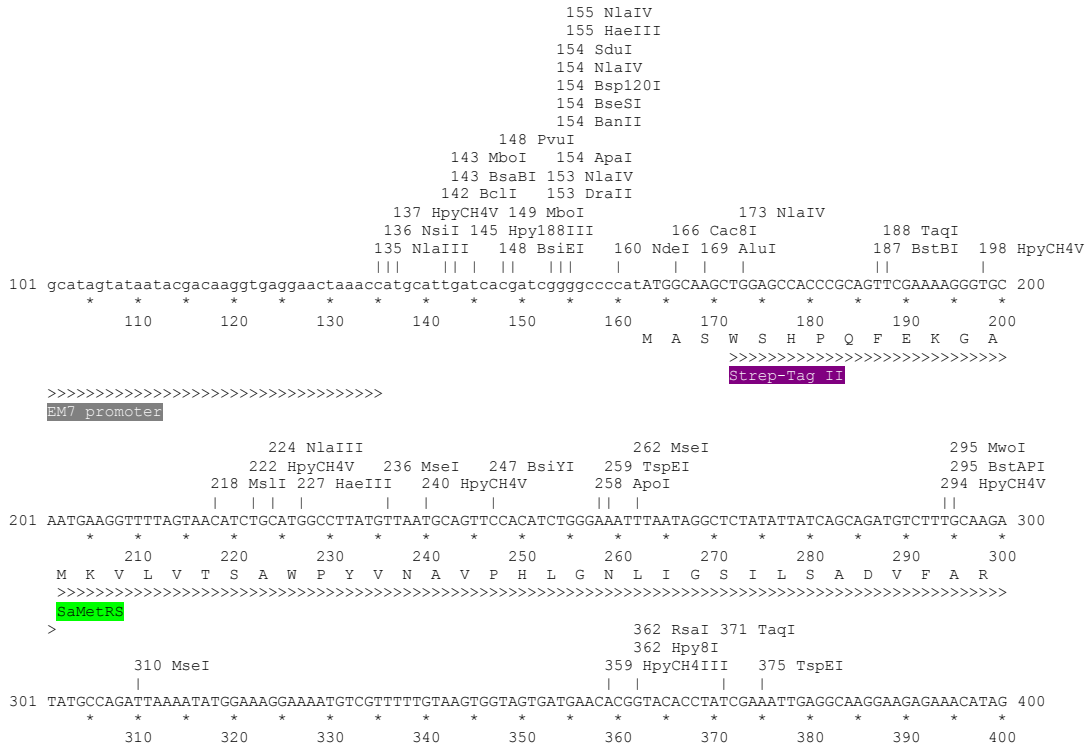

2914 Hpy8I
2914 HincII
2941 TaqI
2987 HpyCH4III
2908 HpyCH4V
2935 TspEI
2950 HpyCH4III
2979 NlaIII
2998 TspGWI
2901 agagccttgcagcgtcaacctctttttcaaggaaaattgctcgaaagtgactgtttggttaggttgcgaacagcgaaccatgacgaactgtaaatctacg 3000
2910
2920
2930
2940
2950
2960
2970
2980
2990
3000
metTp (tRNA promoter on pRARE)
metTp (tRNA promoter of metT-leuW-glnUW-metU-glnVX
3069 HpyCH4V
3094 Bsu36I
3068 PstI
3093 BsiYI
3036 TspGWI
3068 BfmI
3084 HpyCH4V
3034 RsaI
3067 SbfI 3077 TaqI 3087 TspEI
3001 gaatgcttgatattcaggggattttgcggactggtacggatgggagcgaactgataaatggtgtcccctgcaggaatcgaacctgcaattagcccttagg 3100
3010
3020
3030
3040
3050
3060
3070
3080
3090
3100
argW (tRNA(CCU)Arg EcoGene#EG30016)
3151 BsiYI
3146 Hpy8I
3146 BstZ17I
3103 SduI
3146 AccI
3194 MseI
3103 BanII
3120 MseI
3141 NlaIII
3174 RsaI
3186 TaqI
3101 aggggctcgttatatccatttaactaagaggacaatgcggcatgagtatacccgctaatggagtgcggggtaagtacgctgccgctcgattgcttaaacc 3200
3110
3120
3130
3140
3150
3160
3170
3180
3190
3200
argWp (tRNA promoter of argW operon)
argW (tRNA(CCU)Arg EcoGene#EG30016)
3270 TaqI
3269 ClaI
3264 MaeII
3248 MaeII
3263 ZraI
3222 PsiI
3234 MseI
3246 MwoI
3263 AcyI
3204 MwoI 3214 HpaII
3226 TspEI
3246 BstAPI
3263 AatII
3289 TspGWI
3201 ctcgccatttatgccgggtttttataatttttcttaatgttttccgcacgttctgctttttggacgtcatcgattgtccctctaagacacggataaatcg 3300
3210
3220
3230
3240
3250
3260
3270
3280
3290
3300
3327 NlaIII
3303 EcoRV
3326 NspI
3345 HpyCH4V
3371 BseRI
3301 gtgatatcaccacatcaaccaggcaacatgcccgacttgttgaatgcaataaacagaaggaaaaaacagggaggagaaaagaagtggtgctgataggcag 3400
3310
3320
3330
3340
3350
3360
3370
3380
3390
3400
thrT (tRNA(GGU)Thr EcoGene#EG30101)
3426 StyI
3403 TaqI
3421 EcoNI
3450 MwoI
3476 MspA1I
3493 MwoI
3402 BstBI
3421 BsiYI
3434 HhaI
3449 AluI
3471 Cac8I
3485 TaqI
3401 attcgaactgccgacctcacccttaccaagggtgcgctctaccaactgagctatatcagcacatcttggagcgggcagcgggaatcgaaccccgcatcatc 3500
3410
3420
3430
3440
3450
3460
3470
3480
3490
3500
thrT (tRNA(GGU)Thr EcoGene#EG30101)
glyT (tRNA(UCC)Gly EcoGene#EG30036)
3543 PfoI
3584 TaqI
3538 BsiYI
3584 Hpy188III
3501 AluI
3536 Cac8I
3573 HpyCH4V
3592 AluI
3501 agcttggaaggctgaggtaatagccattatacgatgcccgcatcctggaactcggctacctgattttcattctgcactgaatatcgagagaagctctctt 3600
3510
3520
3530
3540
3550
3560
3570
3580
3590
3600
glyT (tRNA(UCC)Gly EcoGene#EG30036)
3649 HaeIII
3648 CfrI
3684 TaqI
3609 HpaII
3644 AccII
3683 BstBI
3608 Cfr10I
3640 PflMI
3676 TaqI
3604 TaqI
3629 AvaI
3640 BsiYI
3655 TspEI
3675 BstBI
3696 BceAI
3601 tattcgagccggtaagcgaacttatcgtctcgggctacgccatcgcgtggccgaaattggtggtgggggaaggattcgaaccttcgaagtctgtgacggc 3700
3610
3620
3630
3640
3650
3660
3670
3680
3690
3700

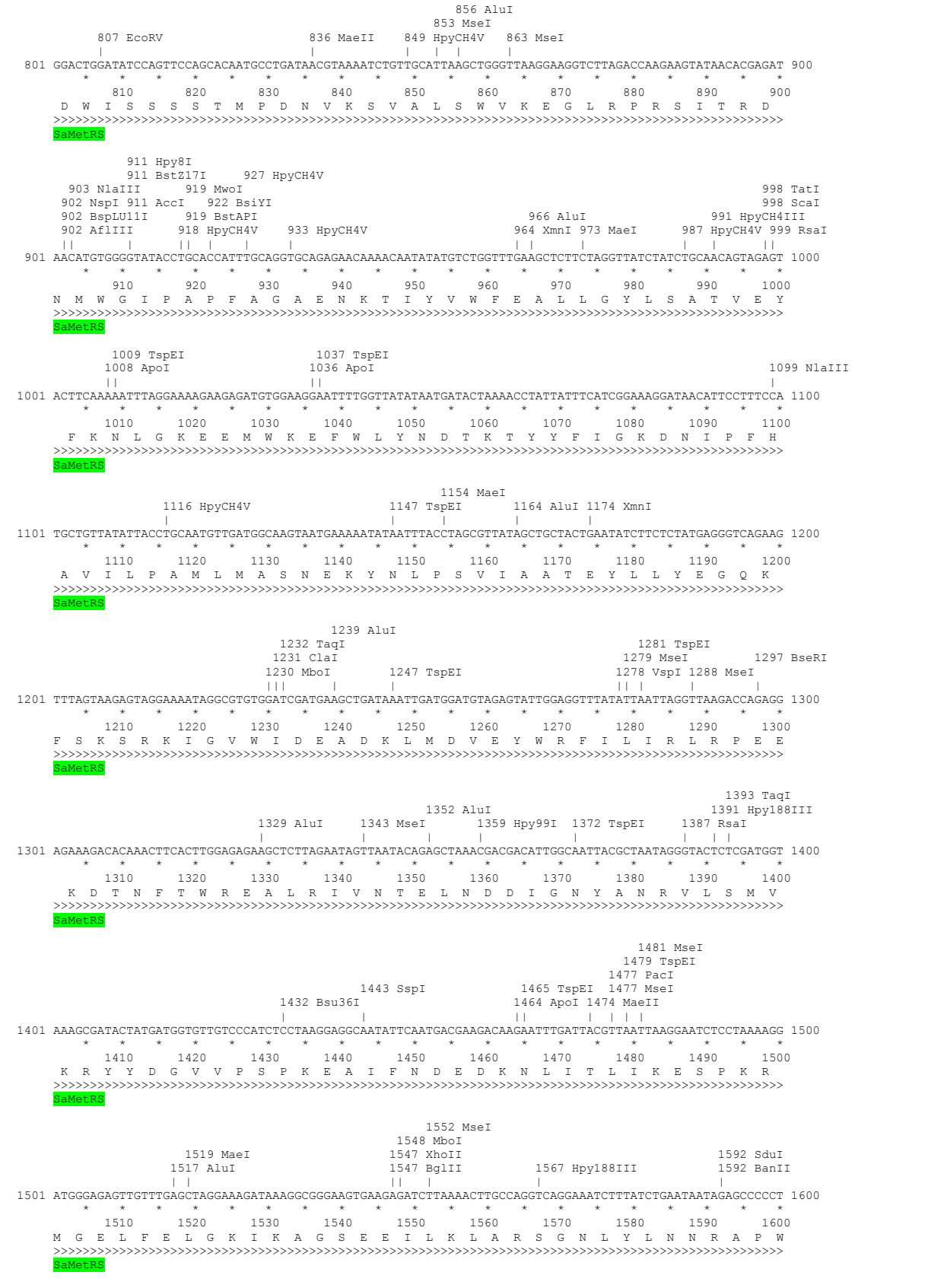

tyrU (tRNA(GUA)Tyr EcoGene#EG30107)
3729 Hpy188III
3762 BsaWI
3728 AvaI
3741 HpaII
3756 BsiYI
3723 HaeIII
3740 BsaWI
3752 NlaIV
3763 HpaII
3776 BstEII
3707 HpyCH4III
3722 CfrI
3732 NlaIV
3752 AccBlI
3769 TaqI
3701 agatttacagtctgctccctttggccgctcgggaaccccaccggacttgatggtgccgactaccggaatcgaactggtgacctactgattacaagtcagt 3800
3710
3720
3730
3740
3750
3760
3770
3780
3790
3800
tyrU (tRNA(GUA)Tyr EcoGene#EG30107)
thrU (tRNA(UGU)Thr EcoGene#EG30102)
3837 HhaI
3836 AccII
3835 HhaI
3884 TspEI
3816 MwoI
3835 Cac8I
3853 NlaIII
3865 HpyCH4V
3879 HpyCH4V
3815 AluI
3835 BssHII
3852 NspI
3863 NlaIII
3876 TspEI
3801 tgctctacctactgagctaagtcggcatcaagtagcgcgcactctatggagacatgcgagttcatgcaactaaaaaattgcataatttgtttttattggtc 3900
3810
3820
3830
3840
3850
3860
3870
3880
3890
3900
thrUp(A52G) (tRNA promoter of thrU-tyrU-glyT-thrT o
thrU (tRNA(UGU)Thr EcoGene#EG30102)
3996 HhaI
3940 Hpy188III
3995 SfoI
3935 RsaI
3995 NlaIV
3934 NlaIV
3995 NarI
3934 KpnI
3944 SduI
3995 KasI
3934 AccB1I
3995 HaeII
3934 Acc65I
3964 MboI
3995 BbeI
3931 MaeI
3944 BanII
3958 SduI
3995 AcyI
3930 StyI
3940 BspHI
3958 BanII
3995 AccB1I
3930 AvrII
3941 NlaIII
3956 Cac8I
3987 MwoI
3998 Cac8I
3901 acatttttatgcgacacgatgaagaaacagcctaggtacctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgcc 4000
3910
3920
3930
3940
3950
3960
3970
3980
3990
4000
4020 HpaII
4019 MwoI
4019 Cfr10I
4017 HhaI
4016 SfoI
4016 NlaIV
4016 NarI
4031 HaeIII
4016 KasI
4030 CfrI
4016 HaeII
4028 NgoMIV
4057 TspGWI
4016 BbeI
4028 NaeI
4054 XhoII
4016 AcyI
4028 Cfr10I
4054 NlaIV
4076 MwoI
4016 AccB1I
4028 Cac8I
4054 BamHI
4076 BlpI
4008 MwoI
4018 SgrAI
4029 HpaII
4044 HpaII
4055 MboI
4068 StyI
4081 Cac8I
4001 agcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatccgtcacataccaaggcggctaagcgagcagatggaacatcaa 4100
4010
4020
4030
4040
4050
4060
4070
4080
4090
4100
4122 MwoI
4120 HhaI
4119 FspI
4118 FspAI
4168 TspEI
4110 Hpy188III
4125 TaqI
4150 AluI
4162 HpyCH4V
4187 NlaIII
4102 Cac8I
4117 BsaBI
4140 NlaIII
4157 Hpy99I
4179 TspGWI
4101 cgcctgcggtcaggaagatgcgcatcgacagcaagacatcatggctatcagcttgtcgtcgtgcaggaattgaagatttccgtttccatgacctcagaca 4200
4110
4120
4130
4140
4150
4160
4170
4180
4190
4200
4212 Cac8I
4277 XmnI
4211 AluI
4273 MslI
4208 Cac8I
4229 AcyI
4249 Hpy188III
4272 BstXI
4295 MwoI
4201 cacctgggcaagctggctgattcagtcaggcgtcccattatcagtgcttcaggaaatgggcggatgggagtccatagaaatggttcgtaggtatgctcac 4300
4210
4220
4230
4240
4250
4260
4270
4280
4290
4300
4325 NlaIII
4324 SphI
4324 NspI
4304 HhaI
4324 Cac8I
4384 TspEI
4301 cttgcgcctaatcatttgacagagcatgcgaggaaaatagacgacatttttggtgataatgtcccaaatatgtcccactctgaaattatggaggatataa 4400
4310
4320
4330
4340
4350
4360
4370
4380
4390
4400

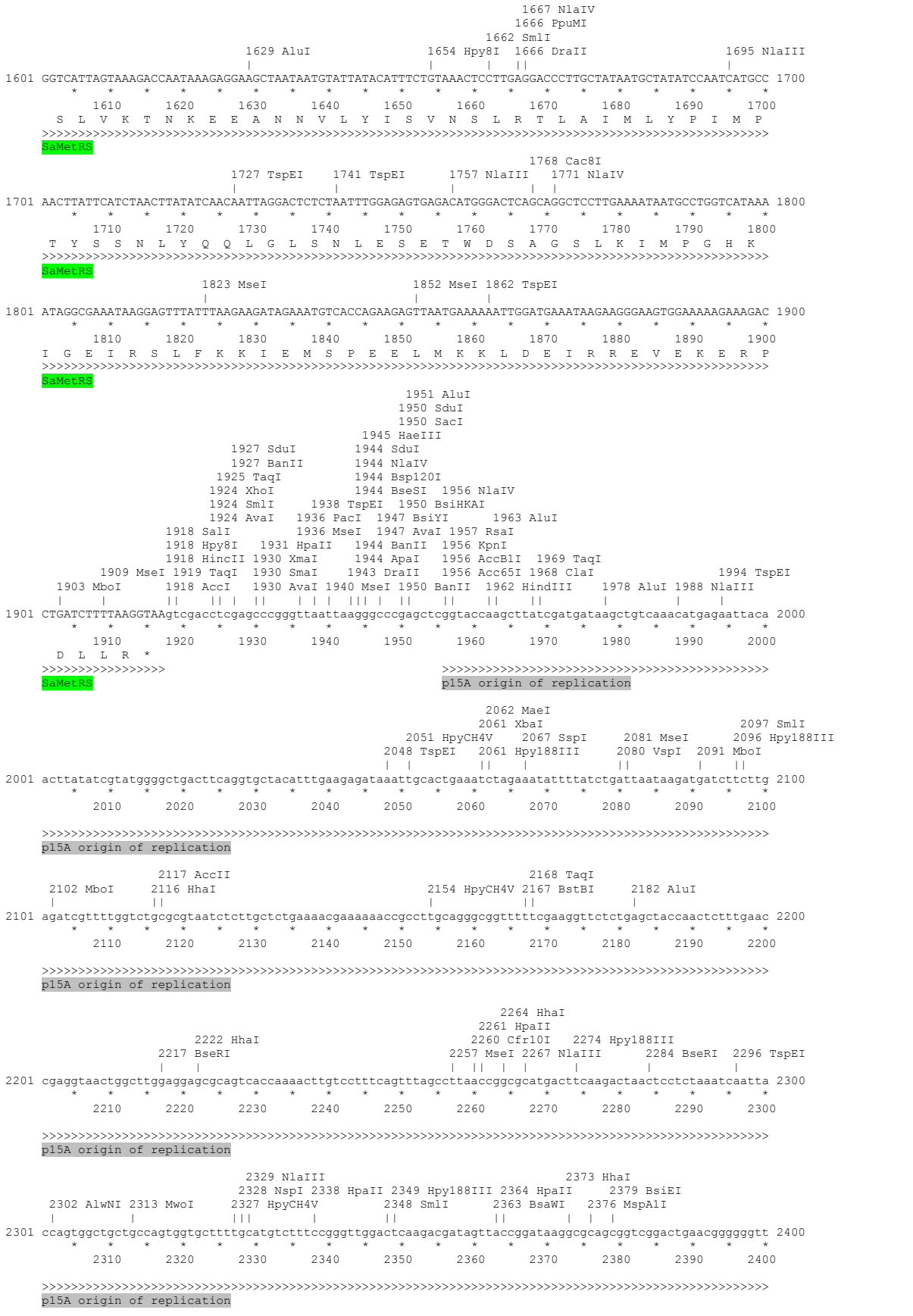
```
                                         4437 HpyCH4V
                                        4436 PstI
                                        4436 BfmI
                                       4435 SbfI
                                   4431 HhaI
                                 4429 MwoI
                                 4429 HhaI
                                 4429 Cac8I
                                 4429 BssHII                                                            4500 HhaI
                                4428 AscI        4445 TaqI                                       4493 MwoI
                       4419 TspEI 4430 AccII    4444 BstBI 4455 HaeIII                          4492 AluI
                       |        ||||   |||      ||         |                                    ||      |
4401 agaaggcgtaactgattgaattgtaatggcgcgccctgcaggattcgaacctgcggcccacgacttagaaggtcgttgctctatccaactgagctaaggg 4500
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
            4410      4420      4430      4440      4450      4460      4470      4480      4490      4500

                               <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
                               argU (tRNA(UCU)Arg EcoGene#EG30014)

                                               4543 Hpy8I                                             4598 AccII
                                               4543 HincII                                         4595 Tth111I
               4511 BsiYI             4534 TspEI                                                  4594 BsiEI
     4501 AccII                 4528 AccII  4540 HpyCH4III                 4571 MboI         4589 Hpy8I
     |         |                |     |     |  |                           |                 |    ||  |
4501 cgcgttgataccgcaatgcggtgtaatcgcgtgaattatacggtcaacccttgctgagtcaatggcttttgatctggttgctgaacaagtgaacgaccgc 4600
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
            4510      4520      4530      4540      4550      4560      4570      4580      4590      4600

       <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
       argUp (tRNA promoter of argU operon)
     <<<
     argU (tRNA(UCU)Arg EcoGene#EG30014)

                                                         4653 AccII                                    4699 HpyCH4V
                            4624 BfmI                   4652 HhaI                                     4698 PstI
                           4623 MwoI               4647 MspA1I     4663 Hpy188III     4682 NlaIII     4698 BfmI
                           ||                      |    ||         |                  |               ||
4601 gtctgatttctgatttatttcgctatagcggcaaacaaacgcacaccgctgcgcgtctgaatcaagaaaacccgtattttcatgtatcaaagtgacctg 4700
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
            4610      4620      4630      4640      4650      4660      4670      4680      4690      4700


                                                      4750 NlaIV
                                                      4750 HaeIII
            4708 MwoI                           4744 TaqI
           4707 AluI                 4733 PsiI 4743 ClaI
          4706 HindIII    4722 TspGWI       4740 BstXI        4758 AluI      4773 Cac8I
          |||             |          |      |  ||     |       |              |
4701 cagccaagcttggattgcgacacggagttactttataatccaatcgattggccccttagctcagtggttagagcaggcgactcataatcgcttggtcgct 4800
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
            4710      4720      4730      4740      4750      4760      4770      4780      4790      4800

                                                      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                                      ileX (tRNA(CAU)Ile EcoGene#EG30046)

                                 4829 MwoI
                                 4829 BglI
                                4828 HaeIII
                               4827 EagI
                               4827 CfrI
                              4826 NotI
                            4824 MspA1I
                      4818 HaeIII
                    4816 NlaIV 4827 BsiEI
                    | |     | ||||
4801 ggttcaagtccagcaggggccaccagcggccgcaaaggc 4839
         *    *    *    *    *    *    *
            4810      4820      4830

     >>>>>>>>>>>>>>>>>>>>>>>>>
     ileX (tRNA(CAU)Ile EcoGene#EG30046)
```

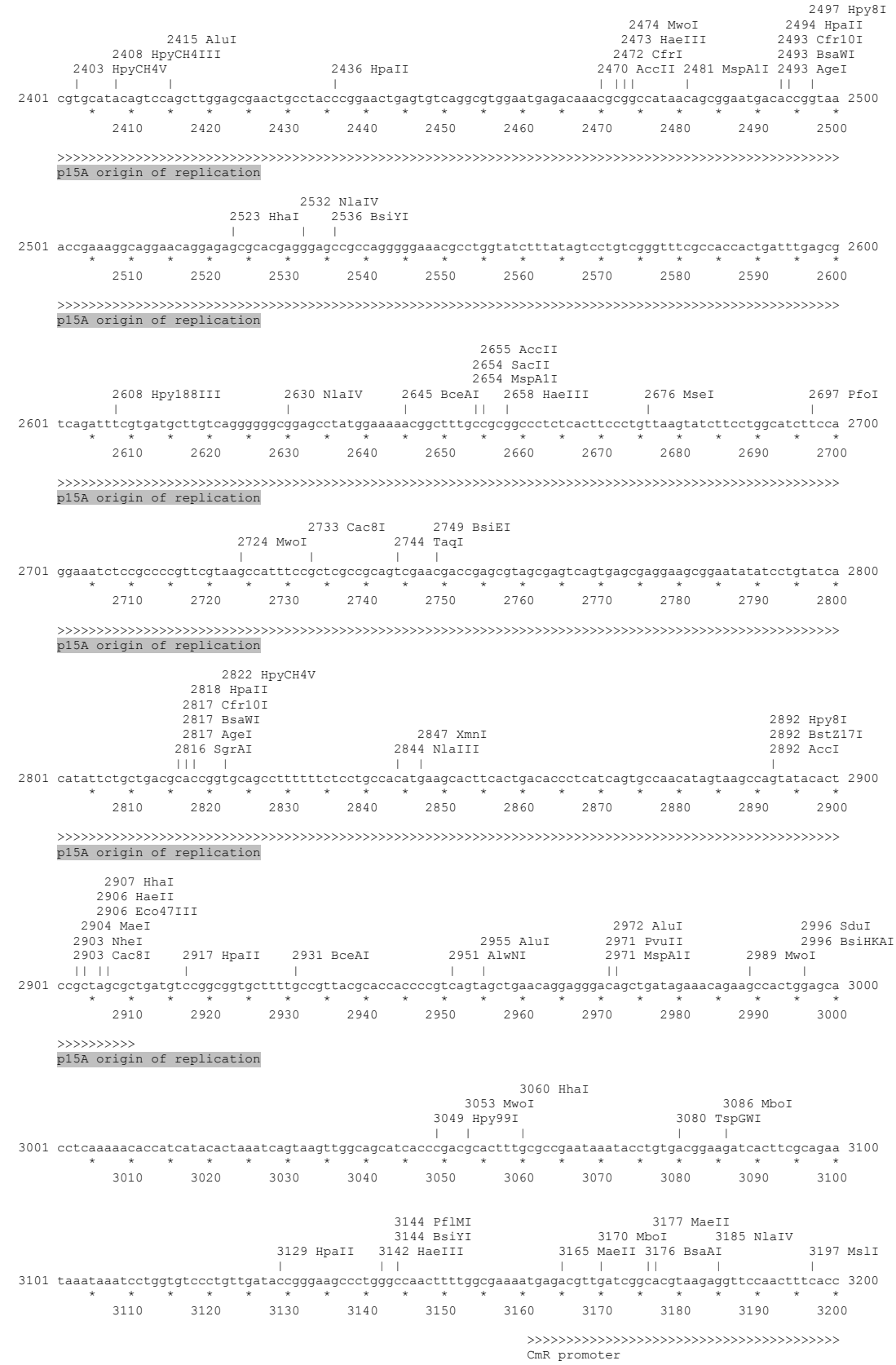
9.6 pTEc1 sequence

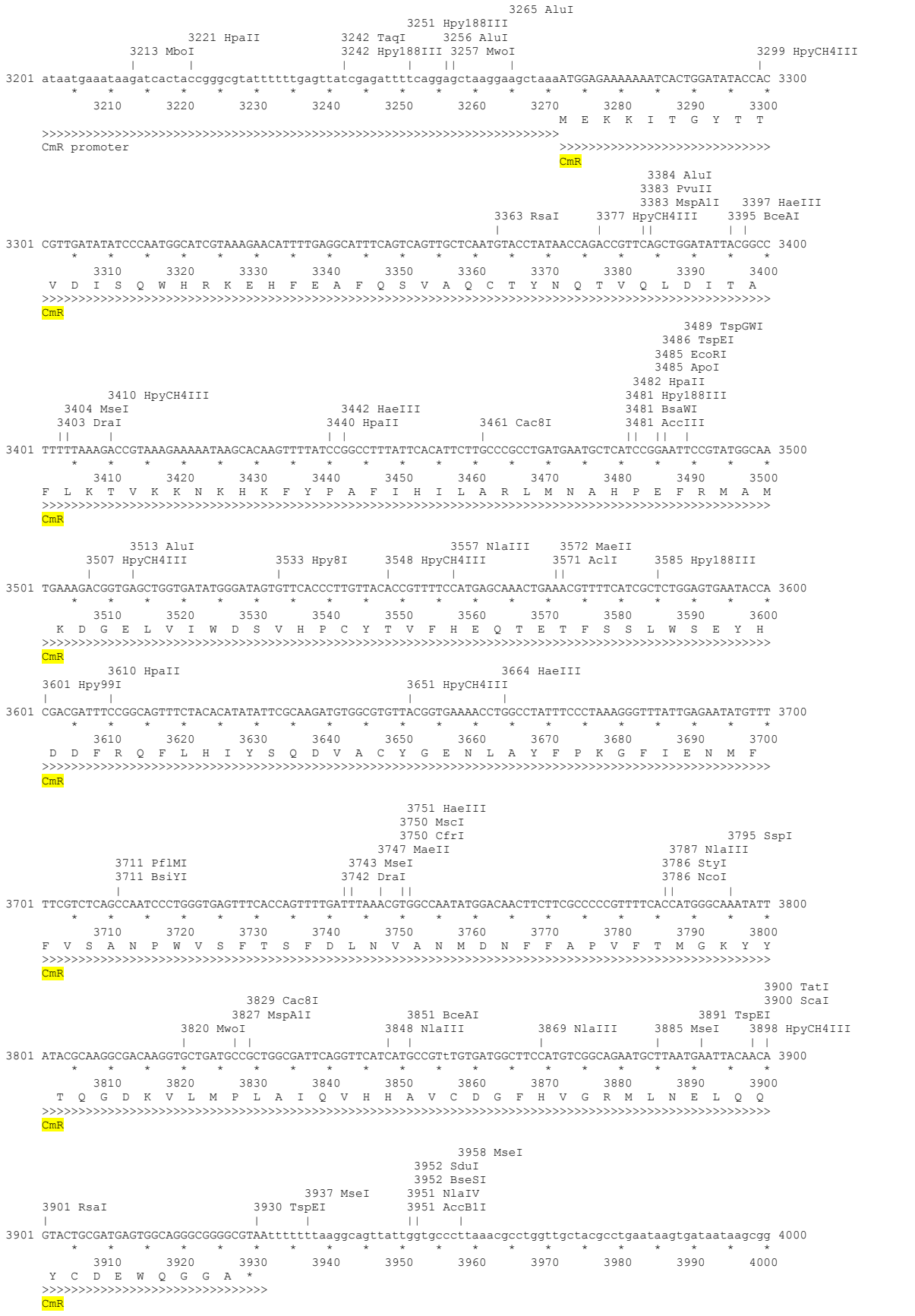
```
                                                            56 BceAI              78 MseI
                                  30 MseI     42 BstXI     55 HaeIII             77 VspI
                              26 AluI     38 TspEI        54 EagI         70 Hpy8I
                             25 HindIII  37 EcoRI         54 CfrI 62 SpeI 70 HincII
                           23 MseI       37 ApoI          54 BsiEI 63 MaeI      76 TspEI
                           | ||   |      ||   |           |||     ||      |     |||
   1 tgacgagaaatcgtcagcctttttaagctttaagccgaattccagcacactggcggccgttactagtgtgttgacaattaatcatcggcatagtatatcg 100
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
             10        20        30        40        50        60        70        80        90       100

                                                                        >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                                                        EM7 promoter
```


155 NlaIV
155 HaeIII
154 SduI
154 NlaIV
154 Bsp120I
154 BseSI
154 BanII
148 PvuI
143 MboI 154 ApaI
143 BsaBI 153 NlaIV
142 BclI 153 DraII
137 HpyCH4V 149 MboI 173 NlaIV
136 NsiI 145 Hpy188III 166 Cac8I 188 TaqI
135 NlaIII 148 BsiEI 160 NdeI 169 AluI 187 BstBI 198 HpyCH4V
101 gcatagtataatacgacaaggtgaggaactaaaccatgcattgatcacgatcggggccccatATGGCAAGCTGGAGCCACCCGCAGTTCGAAAAGGGTGC 200
110 120 130 140 150 160 170 180 190 200
M A S W S H P Q F E K G A
Strep-Tag II
EM7 promoter
224 NlaIII 262 MseI 295 MwoI
222 HpyCH4V 236 MseI 247 BsiYI 259 TspEI 295 BstAPI
218 MslI 227 HaeIII 240 HpyCH4V 258 ApoI 294 HpyCH4V
201 AATGAAGGTTTTAGTAACATCTGCATGGCCTTATGTTAATGCAGTTCCACATCTGGGAAATTTAATAGGCTCTATATTATCAGCAGATGTCTTTGCAAGA 300
210 220 230 240 250 260 270 280 290 300
M K V L V T S A W P Y V N A V P H L G N L I G S I L S A D V F A R
SaMetRS
362 RsaI 371 TaqI
362 Hpy8I
310 MseI 359 HpyCH4III 375 TspEI
301 TATGCCAGATTAAAATATGGAAAGGAAAATGTCGTTTTTGTAAGTGGTAGTGATGAACACGGTACACCTATCGAAATTGAGGCAAGGAAGAGAAACATAG 400
310 320 330 340 350 360 370 380 390 400
Y A R L K Y G K E N V V F V S G S D E H G T P I E I E A R K R N I E
SaMetRS
427 NlaIII
426 SphI
426 NspI
426 Cac8I
418 MboI 427 MslI 468 HindIII
410 TspEI 424 HhaI 443 TspEI 458 NlaIII 469 AluI 478 TspEI 489 Tth111I
401 AACCCAAGAAATTGACCGATCAGGCGCATGCCTATGATAAGAAATTATTTATTGACACATGGAAGATAAGCTTTGACAATTATAGCAGGACTGAGTCTGA 500
410 420 430 440 450 460 470 480 490 500
P K K L T D Q A H A Y D K K L F I D T W K I S F D N Y S R T E S E
SaMetRS
539 TaqI
539 Hpy188III
502 RsaI 529 MaeI 538 XhoI
502 Hpy8I 523 TspEI 538 SmlI
501 TatI 522 ApoI 538 AvaI 578 HpyCH4III
501 AGTACACAAAGAGTTCGTTAGAAATTTTCTAGTAAAACTCGAGAAATACATCAAGGTAGAAGAGGACGAGATTCCATACTGTGAAAAAGACAAGTTATTT 600
510 520 530 540 550 560 570 580 590 600
V H K E F V R N F L V K L E K Y I K V E E D E I P Y C E K D K L F
SaMetRS
604 HpaII
603 Hpy188III 661 MboI
603 BsaWI 635 HpyCH4III 660 BclI 687 Hpy8I
603 AccIII 614 PsiI 631 BsiYI 653 MaeI 673 TspEI 687 HincII 698 MboI
601 CTTCCGGATAGATTTATAAAGGGAGTATGTCCTTACTGTGGCTTTGAAGACGCTAGAGGTGATCAGTGTGATAATTGTGGTAGATTGTTGACTCCAAGAT 700
610 620 630 640 650 660 670 680 690 700
L P D R F I K G V C P Y C G F E D A R G D Q C D N C G R L L T P R S
SaMetRS
720 HpyCH4V
719 SduI
719 Hpy8I
719 BsiHKAI
719 BseSI
719 ApaLI 746 MseI 755 PflMI 782 TspEI
712 MwoI 722 DraIII 745 DraI 755 BsiYI 770 TaqI 781 ApoI
701 CATTAGTAAATGCTAAGTGTGCACTATGTGGTAATCCACCTGTATTTAAAGTTACCAAGCATTGGTTTTTCGATTTATCTGAATTTGGAGATAAAATAAG 800
710 720 730 740 750 760 770 780 790 800
L V N A K C A L C G N P P V F K V T K H W F F D L S E F G D K I R
SaMetRS

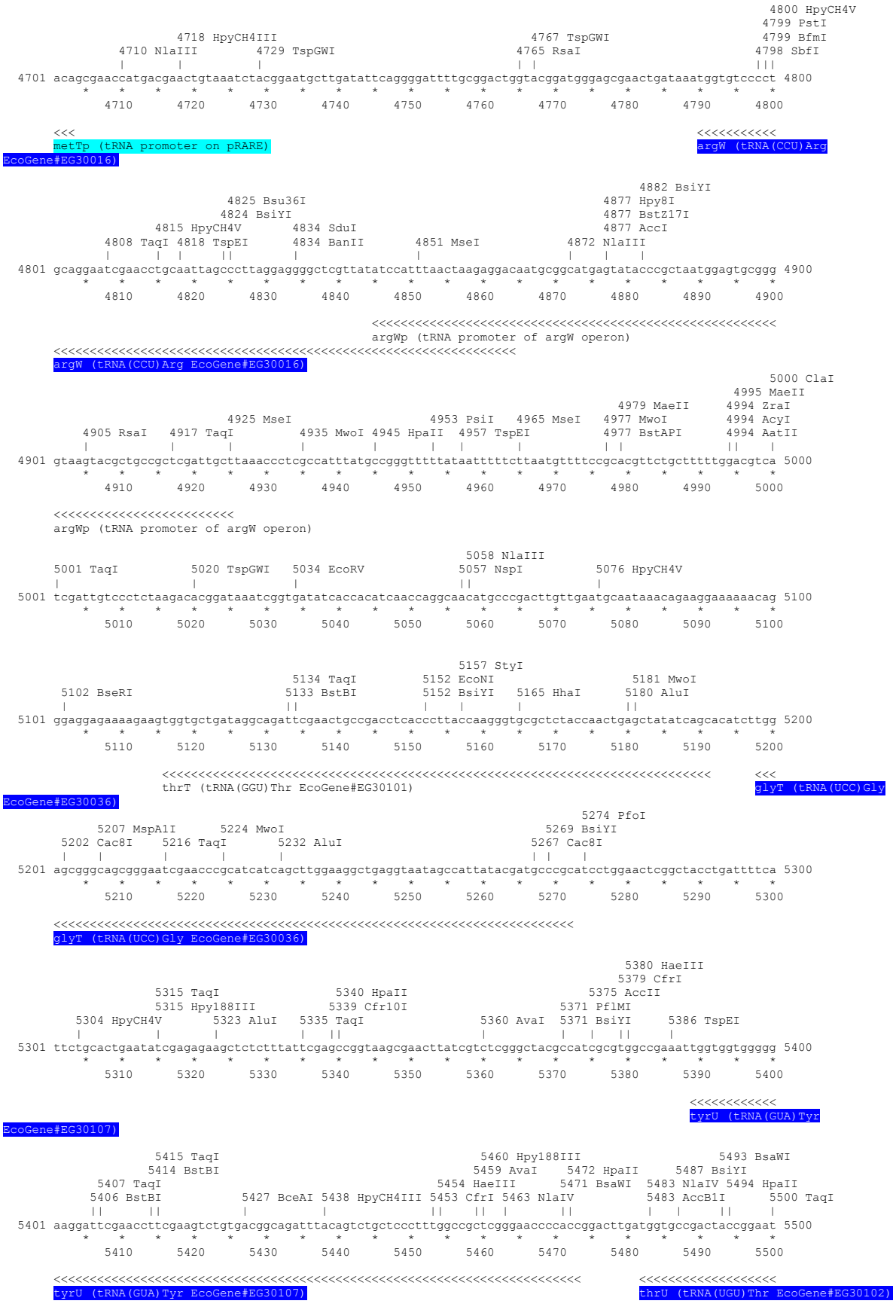

856 AluI
853 MseI
807 EcoRV
836 MaeII
849 HpyCH4V
863 MseI
801 GGACTGGATATCCAGTTCCAGCACAATGCCTGATAACGTAAAATCTGTTGCATTAAGCTGGGTTAAGGAAGGTCTTAGACCAAGAAGTATAACACGAGAT 900
D W I S S S S T M P D N V K S V A L S W V K E G L R P R S I T R D
SaMetRS
911 Hpy8I
911 BstZ17I
927 HpyCH4V
903 NlaIII
919 MwoI
998 TatI
902 NspI 911 AccI
922 BsiYI
998 ScaI
902 BspLU11I
919 BstAPI
966 AluI
991 HpyCH4III
902 AflIII
918 HpyCH4V
933 HpyCH4V
964 XmnI 973 MaeI
987 HpyCH4V 999 RsaI
901 AACATGTGGGGTATACCTGCACCATTTGCAGGTGCAGAGAACAAAACAATATATGTCTGGTTTGAAGCTCTTCTAGGTTATCTATCTGCAACAGTAGAGT 1000
N M W G I P A P F A G A E N K T I Y V W F E A L L G Y L S A T V E Y
SaMetRS
1009 TspEI
1037 TspEI
1008 ApoI
1036 ApoI
1099 NlaIII
1001 ACTTCAAAAATTTAGGAAAAGAAGAGATGTGGAAGGAATTTTGGTTATATAATGATACTAAAACCTATTATTTCATCGGAAAGGATAACATTCCTTTCCA 1100
F K N L G K E E M W K E F W L Y N D T K T Y Y F I G K D N I P F H
SaMetRS
1154 MaeI
1116 HpyCH4V
1147 TspEI
1164 AluI 1174 XmnI
1101 TGCTGTTATATTACCTGCAATGTTGATGGCAAGTAATGAAAAATATAATTTACCTAGCGTTATAGCTGCTACTGAATATCTTCTCTATGAGGGTCAGAAG 1200
A V I L P A M L M A S N E K Y N L P S V I A A T E Y L L Y E G Q K
SaMetRS
1239 AluI
1232 TaqI
1231 ClaI
1230 MboI
1247 TspEI
1281 TspEI
1279 MseI
1297 BseRI
1278 VspI 1288 MseI
1201 TTTAGTAAGAGTAGGAAAATAGGCGTGTGGATCGATGAAGCTGATAAATTGATGGATGTAGAGTATTGGAGGTTTATATTAATTAGGTTAAGACCAGAGG 1300
F S K S R K I G V W I D E A D K L M D V E Y W R F I L I R L R P E E
SaMetRS
1393 TaqI
1352 AluI
1391 Hpy188III
1329 AluI
1343 MseI
1359 Hpy99I
1372 TspEI
1387 RsaI
1301 AGAAAGACACAAACTTCACTTGGAGAGAAGCTCTTAGAATAGTTAATACAGAGCTAAACGACGACATTGGCAATTACGCTAATAGGGTACTCTCGATGGT 1400
K D T N F T W R E A L R I V N T E L N D D I G N Y A N R V L S M V
SaMetRS
1481 MseI
1479 TspEI
1477 PacI
1443 SspI
1465 TspEI
1477 MseI
1432 Bsu36I
1464 ApoI 1474 MaeII
1401 AAAGCGATACTATGATGGTGTTGTCCCATCTCCTAAGGAGGCAATATTCAATGACGAAGACAAGAATTTGATTACGTTAATTAAGGAATCTCCTAAAAGG 1500
K R Y Y D G V V P S P K E A I F N D E D K N L I T L I K E S P K R
SaMetRS
1552 MseI
1548 MboI
1519 MaeI
1547 XhoI
1592 SduI
1517 AluI
1547 BglII
1567 Hpy188III
1592 BanII
1501 ATGGGAGAGTTGTTTGAGCTAGGAAAGATAAAGGCGGGAAGTGAAGAGATCTTAAAACTTGCCAGGTCAGGAAATCTTTATCTGAATAATAGAGCCCCCT 1600
M G E L F E L G K I K A G S E E I L K L A R S G N L Y L N N R A P W
SaMetRS


1667 NlaIV
1666 PpuMI
1662 SmlI
1629 AluI
1654 Hpy8I
1666 DraII
1695 NlaIII
1601 GGTCATTAGTAAAGACCAATAAAGAGGAAGCTAATAATGTATTATACATTTCTGTAAACTCCTTGAGGACCCTTGCTATAATGCTATATCCAATCATGCC 1700
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
S L V K T N K E E A N N V L Y I S V N S L R T L A I M L Y P I M P
SaMetRS
1768 Cac8I
1727 TspEI
1741 TspEI
1757 NlaIII
1771 NlaIV
1701 AACTTATTCATCTAACTTATATCAACAATTAGGACTCTCTAATTTGGAGAGTGAGACATGGGACTCAGCAGGCTCCTTGAAAATAATGCCTGGTCATAAA 1800
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
T Y S S N L Y Q Q L G L S N L E S E T W D S A G S L K I M P G H K
SaMetRS
1823 MseI
1852 MseI
1862 TspEI
1801 ATAGGCGAAATAAGGAGTTTATTTAAGAAGATAGAAATGTCACCAGAAGAGTTAATGAAAAAATTGGATGAAATAAGAAGGGAAGTGGAAAAAGAAAGAC 1900
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
I G E I R S L F K K I E M S P E E L M K K L D E I R R E V E K E R P
SaMetRS
1951 AluI
1950 SduI
1950 SacI
1945 HaeIII
1927 SduI
1927 BanII
1944 SduI
1944 NlaIV
1925 TaqI
1944 Bsp120I
1924 XhoI
1944 BseSI
1956 NlaIV
1924 SmlI
1938 TspEI
1950 BsiHKAI
1924 AvaI
1936 PacI
1947 BsiYI
1963 AluI
1918 SalI
1936 MseI
1947 AvaI
1957 RsaI
1918 Hpy8I
1931 HpaII
1944 BanII
1956 KpnI
1918 HincII
1930 XmaI
1944 ApaI
1956 AccB1I
1969 TaqI
1909 MseI
1919 TaqI
1930 SmaI
1943 DraII
1956 Acc65I
1968 ClaI
1994 TspEI
1903 MboI
1918 AccI
1930 AvaI
1940 MseI
1950 BanII
1962 HindIII
1978 AluI
1988 NlaIII
1901 CTGATCTTTTAAGGTAAgtcgacctcgagcccgggttaattaagggcccgagctcggtaccaagcttatcgatgataagctgtcaaacatgagaattaca 2000
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
D L L R *
SaMetRS
p15A origin of replication
2062 MaeI
2061 XbaI
2097 SmlI
2051 HpyCH4V
2067 SspI
2081 MseI
2096 Hpy188III
2048 TspEI
2061 Hpy188III
2080 VspI
2091 MboI
2001 acttatatcgtatggggctgacttcaggtgctacatttgaagagataaattgcactgaaatctagaaatattttatctgattaataagatgatcttcttg 2100
2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
p15A origin of replication
2117 AccII
2168 TaqI
2102 MboI
2116 HhaI
2154 HpyCH4V
2167 BstBI
2182 AluI
2101 agatcgttttggtctgcgcgtaatctcttgctctgaaaacgaaaaaaccgccttgcagggcggtttttcgaaggttctctgagctaccaactctttgaac 2200
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
p15A origin of replication
2264 HhaI
2261 HpaII
2222 HhaI
2260 Cfr10I
2274 Hpy188III
2217 BseRI
2257 MseI
2267 NlaIII
2284 BseRI
2296 TspEI
2201 cgaggtaactggcttggaggagcgcagtcaccaaaacttgtcctttcagtttagccttaaccggcgcatgacttcaagactaactcctctaaatcaatta 2300
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
p15A origin of replication
2329 NlaIII
2373 HhaI
2328 NspI
2338 HpaII
2349 Hpy188III
2364 HpaII
2379 BsiEI
2302 AlwNI
2313 MwoI
2327 HpyCH4V
2348 SmlI
2363 BsaWI
2376 MspA1I
2301 ccagtggctgctgccagtggtgcttttgcatgtctttccgggttggactcaagacgatagttaccggataaggcgcagcggtcggactgaacgggggttt 2400
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
p15A origin of replication


2497 Hpy8I
2474 MwoI
2494 HpaII
2415 AluI
2473 HaeIII
2493 Cfr10I
2408 HpyCH4III
2472 CfrI
2493 BsaWI
2403 HpyCH4V
2436 HpaII
2470 AccII 2481 MspA1I 2493 AgeI
2401 cgtgcatacagtccagcttggagcgaactgcctacccggaactgagtgtcaggcgtggaatgagacaaacgcggccataacagcggaatgacaccggtaa 2500
2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
p15A origin of replication
2532 NlaIV
2523 HhaI 2536 BsiYI
2501 accgaaaggcaggaacaggagagcgcacgagggagccgccaggggaaacgcctggtatctttatagtcctgtcgggtttcgccaccactgatttgagcg 2600
2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
p15A origin of replication
2655 AccII
2654 SacII
2654 MspA1I
2608 Hpy188III
2630 NlaIV
2645 BceAI
2658 HaeIII
2676 MseI
2697 PfoI
2601 tcagatttcgtgatgcttgtcaggggggcggagcctatggaaaaacggctttgccgcggccctctcacttccctgttaagtatcttcctggcatcttcca 2700
2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
p15A origin of replication
2733 Cac8I
2749 BsiEI
2724 MwoI
2744 TaqI
2701 ggaaatctccgccccgttcgtaagccatttccgctcgccgcagtcgaacgaccgagcgtagcgagtcagtgagcgaggaagcggaatatatcctgtatca 2800
2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
p15A origin of replication
2822 HpyCH4V
2818 HpaII
2817 Cfr10I
2817 BsaWI
2892 Hpy8I
2817 AgeI
2847 XmnI
2892 BstZ17I
2816 SgrAI
2844 NlaIII
2892 AccI
2801 catattctgctgacgcaccggtgcagccttttttctcctgccacatgaagcacttcactgacaccctcatcagtgccaacatagtaagccagtatacact 2900
2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
p15A origin of replication
2907 HhaI
2906 HaeII
2906 Eco47III
2904 MaeI
2972 AluI
2996 SduI
2903 NheI
2955 AluI
2971 PvuII
2996 BsiHKAI
2903 Cac8I
2917 HpaII
2931 BceAI
2951 AlwNI
2971 MspA1I
2989 MwoI
2901 ccgctagcgctgatgtccggcggtgcttttgccgttacgcaccaccccgtcagtagctgaacaggagggacagctgatagaaacagaagccactggagca 3000
2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
p15A origin of replication
3060 HhaI
3053 MwoI
3086 MboI
3049 Hpy99I
3080 TspGWI
3001 cctcaaaaacaccatcatacactaaatcagtaagttggcagcatcacccgacgcactttgcgccgaataaatacctgtgacggaagatcacttcgcagaa 3100
3010 3020 3030 3040 3050 3060 3070 3080 3090 3100
3144 PflMI
3177 MaeII
3144 BsiYI
3170 MboI
3185 NlaIV
3129 HpaII
3142 HaeIII
3165 MaeII 3176 BsaAI
3197 MslI
3101 taaataaatcctggtgtccctgttgataccgggaagccctgggccaacttttggcgaaaatgagacgttgatcggcacgtaagaggttccaactttcacc 3200
3110 3120 3130 3140 3150 3160 3170 3180 3190 3200
CmR promoter


3265 AluI
3251 Hpy188III
3221 HpaII
3242 TaqI
3256 AluI
3213 MboI
3242 Hpy188III 3257 MwoI
3299 HpyCH4III
3201 ataatgaaataagatcactaccgggcgtatttttttgagttatcgagattttcaggagctaaggaagctaaaATGGAGAAAAAAATCACTGGATATACCAC 3300
3210 3220 3230 3240 3250 3260 3270 3280 3290 3300
M E K K I T G Y T T
CmR promoter
CmR
3384 AluI
3383 PvuII
3383 MspA1I
3397 HaeIII
3363 RsaI
3377 HpyCH4III
3395 BceAI
3301 CGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCC 3400
3310 3320 3330 3340 3350 3360 3370 3380 3390 3400
V D I S Q W H R K E H F E A F Q S V A Q C T Y N Q T V Q L D I T A
CmR
3489 TspGWI
3486 TspEI
3485 EcoRI
3485 ApoI
3482 HpaII
3410 HpyCH4III
3481 Hpy188III
3404 MseI
3442 HaeIII
3481 BsaWI
3403 DraI
3440 HpaII
3461 Cac8I
3481 AccIII
3401 TTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTCCGTATGGCAA 3500
3410 3420 3430 3440 3450 3460 3470 3480 3490 3500
F L K T V K K N K H K F Y P A F I H I L A R L M N A H P E F R M A M
CmR
3513 AluI
3557 NlaIII
3572 MaeII
3507 HpyCH4III
3533 Hpy8I
3548 HpyCH4III
3571 AclI
3585 Hpy188III
3501 TGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCA 3600
3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
K D G E L V I W D S V H P C Y T V F H E Q T E T F S S L W S E Y H
CmR
3610 HpaII
3664 HaeIII
3601 Hpy99I
3651 HpyCH4III
3601 CGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTT 3700
3610 3620 3630 3640 3650 3660 3670 3680 3690 3700
D D F R Q F L H I Y S Q D V A C Y G E N L A Y F P K G F I E N M F
CmR
3751 HaeIII
3750 MscI
3750 CfrI
3795 SspI
3747 MaeII
3787 NlaIII
3711 PflMI
3743 MseI
3786 StyI
3711 BsiYI
3742 DraI
3786 NcoI
3701 TTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATT 3800
3710 3720 3730 3740 3750 3760 3770 3780 3790 3800
F V S A N P W V S F T S F D L N V A N M D N F F A P V F T M G K Y Y
CmR
3900 TatI
3900 ScaI
3829 Cac8I
3827 MspA1I
3851 BceAI
3891 TspEI
3820 MwoI
3848 NlaIII
3869 NlaIII
3885 MseI
3898 HpyCH4III
3801 ATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTtTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACA 3900
3810 3820 3830 3840 3850 3860 3870 3880 3890 3900
T Q G D K V L M P L A I Q V H H A V C D G F H V G R M L N E L Q Q
CmR
3958 MseI
3952 SduI
3952 BseSI
3937 MseI
3951 NlaIV
3901 RsaI
3930 TspEI
3951 AccB1I
3901 GTACTGCGATGAGTGGCAGGGCGGGGCGTAAttttttttaaggcagttattggtgcccttaaacgcctggttgctacgcctgaataagtgataataagcgg 4000
3910 3920 3930 3940 3950 3960 3970 3980 3990 4000
Y C D E W Q G G A *
CmR

```
                                       4033 HpaII
                                       4033 BsiYI
                                    4030 Tth111I                                                 4091 HpaII
                      4016 TaqI     4030 DrdI                                                   4090 Cfr10I
                     4015 BstBI   4028 TaqI                                                     4090 BsaWI
                   4013 TspEI  4025 TspEI   4038 Hpy99I                                         4090 AgeI
                  4012 ApoI   4024 ApoI 4034 BsiEI               4059 MseI                  4086 Hpy8I
                  || ||       ||  | |  ||   |                    |                          |   ||
  4001 atgaatggcagaaattcgaaagcaaattcgacccggtcgtcggttcagggcagggtcgttaaatagccgcttatgtctattgctggtttaccggtttatt 4100
           *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
              4010      4020      4030      4040      4050      4060      4070      4080      4090      4100

                                                                   4161 TspEI
                                                                 4159 MseI
                                                                4158 VspI
            4106 HpaII                                         4157 TspEI
           4105 BsaWI     4120 HpyCH4III     4139 Bsu36I    4154 TaqI 4164 HpyCH4V       4183 HpyCH4III
           ||             |                  |              |  ||| |  |                  |
  4101 gactaccggaagcagtgtgaccgtgtgcttctcaaatgcctgaggtctgtttatcgaattaattgcagatataaaaaaccaaccgtaagggttggtttt 4200
           *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
              4110      4120      4130      4140      4150      4160      4170      4180      4190      4200

                                                                  4260 HpyCH4III         4283 MaeII
                                                              4256 NlaIII               4282 BsaAI
         4203 SmlI                                            4256 MslI            4278 MwoI
        4202 Hpy188III                                     4253 BsiYI  4265 HhaI    4278 HhaI 4288 BsiYI
        ||                                                 |  |   |    |            |   ||    |
  4201 ttcttgagattttggtcggcacgagaggatttgaacctccgacccccgacaccccatgacggtgcgctaccaggctgcgctacgtgccgactcgtggct 4300
           *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
              4210      4220      4230      4240      4250      4260      4270      4280      4290      4300

                                                                                <<<<<<<<<<<<<<<<<<<<<<<<<<<<
                                                                                proLp (tRNA promoter of proL operon)
                    <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
                    proL (tRNA(GGG)Pro EcoGene#EG30067)
             <
             G Exchanged for A
                                                                   4361 MseI
                                                                  4360 VspI
                                                                 4359 TspEI
                4310 HpyCH4III    4328 HpyCH4V               4355 PsiI                         4389 RsaI
                |                 |                          |   |||                           |
  4301 gctaatactaccgttttccacaccgattgcaagtaagatatttcgctaactgatttataattaatcagttagcgataaaacgcttctcgtacaacgcttt 4400
           *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
              4310      4320      4330      4340      4350      4360      4370      4380      4390      4400

       <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
       proLp (tRNA promoter of proL operon)
                                                    4446 SfoI
                                                    4446 NlaIV
                                                    4446 NarI
                                                    4446 KasI
                                                    4446 HaeII
                                                    4446 BbeI
                                                    4446 AcyI 4456 BsiYI           4477 TspEI
                                         4435 MwoI  4446 AccB1I             4470 Hpy8I
                                         4435 BstAPI 4447 HhaI              4470 AccI    4483 MwoI
                                         |          ||        |             |      |     |
  4401 ctggtgaatggtgcgggaggcgagacttgaactcgcacaccttgcggcgccagaacctaaatctggtgcgtctaccaatttcgccactcccgcaaaaaaa 4500
           *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
              4410      4420      4430      4440      4450      4460      4470      4480      4490      4500

               <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
               leuW (tRNA(CAA)Leu EcoGene#EG30052)

                  4512 AccII
                 4511 SacII
                 4511 MspA1I                                                   4573 BceAI
              4508 BsiYI                  4536 HpaII      4552 HhaI         4570 MwoI             4592 AccII
          4504 NlaIV                     4535 BsaWI      4551 HaeII        4569 AluI             4591 NruI
          4504 AccB1I             4528 HpaII           4549 HpaII      4565 BlpI                 4591 Hpy188III
          |   |  ||               |      ||            | ||            |   ||  |                 ||
  4501 gatggtgccgccgcggggatttgaaccccggacaaccggattatgagtccggcgctctaaccaggctgagctacggcggcatcttttttttcgcgatacc 4600
           *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
              4510      4520      4530      4540      4550      4560      4570      4580      4590      4600

                                                                                                    <<<<<<<
                                                                                                    metTp (tRNA promoter on
pRARE)

                                                                                                    <<<<<<<
                                                                                                    metTp (tRNA promoter of
metT-leuW-glnUW-metU-glnVX
         <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
         Sa-tRNAMet1
                                                   4645 Hpy8I
                       4617 MwoI                   4645 HincII                4672 TaqI
                       4617 HhaI             4639 HpyCH4V               4666 TspEI     4681 HpyCH4III
                       |                     |     |                    |     |        |
  4601 ttatcggcgttgcggggcgcattatgcgtatagagccttgcagcgtcaacctctttttcaaggaaaattgctcgaaagtgactgtttggttaggttgcga 4700
           *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
              4610      4620      4630      4640      4650      4660      4670      4680      4690      4700

       <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
       metTp (tRNA promoter on pRARE)
       <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
       metTp (tRNA promoter of metT-leuW-glnUW-metU-glnVX
```

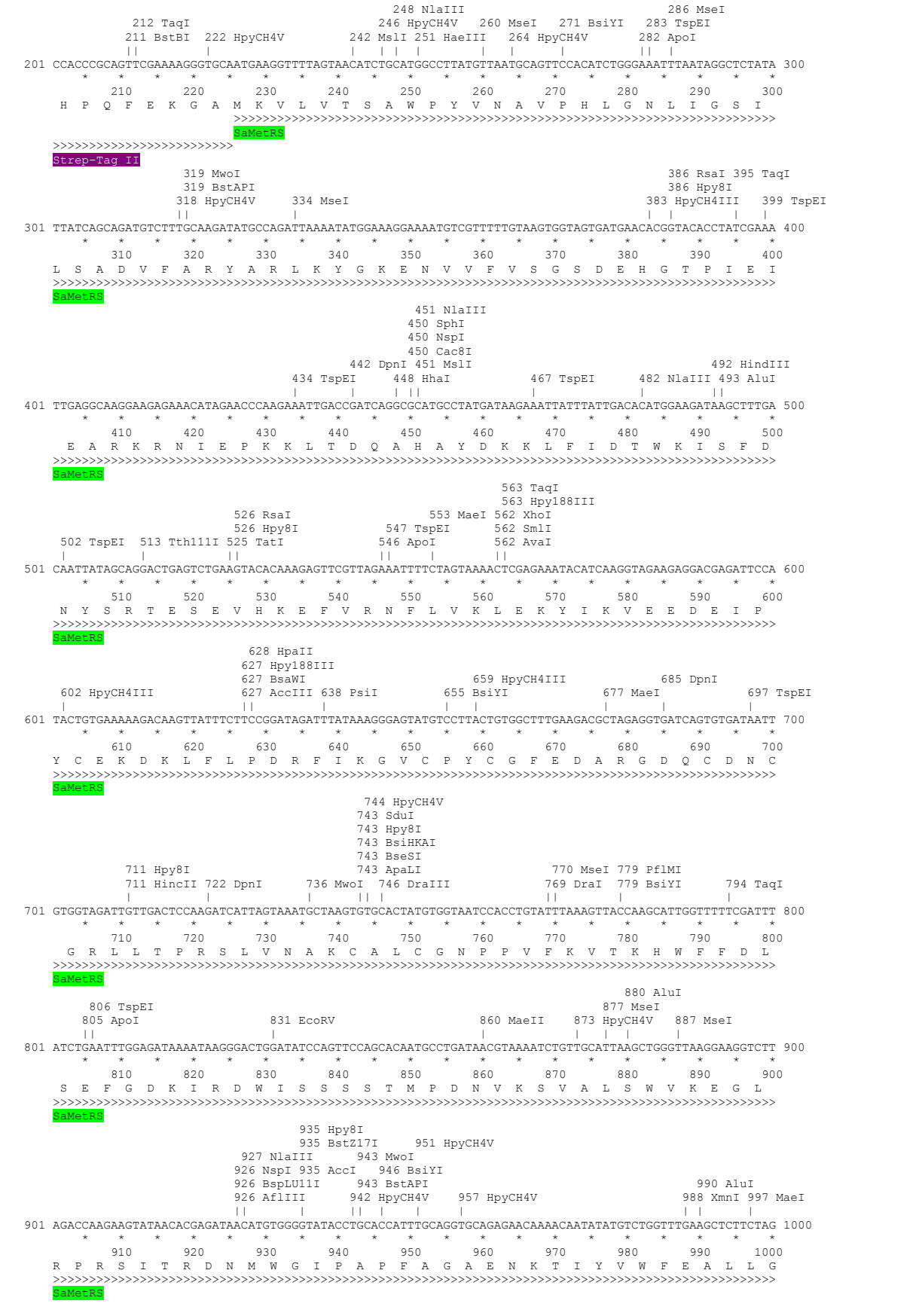

4800 HpyCH4V
4799 PstI
4718 HpyCH4III
4767 TspGWI
4799 BfmI
4710 NlaIII
4729 TspGWI
4765 RsaI
4798 SbfI
4701 acagcgaaccatgacgaactgtaaatctacggaatgcttgatattcaggggattttgcggactggtacggatgggagcgaactgataaatggtgtcccct 4800
4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
metTp (tRNA promoter on pRARE)
argW (tRNA(CCU)Arg EcoGene#EG30016)
4882 BsiYI
4825 Bsu36I
4877 Hpy8I
4824 BsiYI
4877 BstZ17I
4815 HpyCH4V
4834 SduI
4877 AccI
4808 TaqI 4818 TspEI
4834 BanII
4851 MseI
4872 NlaIII
4801 gcaggaatcgaacctgcaattagcccttaggaggggctcgttatatccatttaactaagaggacaatgcggcatgagtatacccgctaatggagtgcggg 4900
4810 4820 4830 4840 4850 4860 4870 4880 4890 4900
argWp (tRNA promoter of argW operon)
argW (tRNA(CCU)Arg EcoGene#EG30016)
5000 ClaI
4995 MaeII
4979 MaeII
4994 ZraI
4925 MseI
4953 PsiI
4965 MseI
4977 MwoI
4994 AcyI
4905 RsaI
4917 TaqI
4935 MwoI 4945 HpaII
4957 TspEI
4977 BstAPI
4994 AatII
4901 gtaagtacgctgccgctcgattgcttaaaccctcgccatttatgccgggtttttataatttttcttaatgttttccgcacgttctgctttttggacgtca 5000
4910 4920 4930 4940 4950 4960 4970 4980 4990 5000
argWp (tRNA promoter of argW operon)
5058 NlaIII
5001 TaqI
5020 TspGWI
5034 EcoRV
5057 NspI
5076 HpyCH4V
5001 tcgattgtccctctaagacacggatgaaatcggtgatatcaccacatcaaccaggcaacatgcccgacttgttgaatgcaataaacagaaggaaaaaacag 5100
5010 5020 5030 5040 5050 5060 5070 5080 5090 5100
5157 StyI
5134 TaqI
5152 EcoNI
5181 MwoI
5102 BseRI
5133 BstBI
5152 BsiYI
5165 HhaI
5180 AluI
5101 ggaggagaaaagaagtggtgctgataggcagattcgaactgccgacctcacccttaccaagggtgcgctctaccaactgagctatatcagcacatcttgg 5200
5110 5120 5130 5140 5150 5160 5170 5180 5190 5200
thrT (tRNA(GGU)Thr EcoGene#EG30101)
glyT (tRNA(UCC)Gly EcoGene#EG30036)
5274 PfoI
5207 MspA1I
5224 MwoI
5269 BsiYI
5202 Cac8I
5216 TaqI
5232 AluI
5267 Cac8I
5201 agcgggcagcgggaatcgaacccgcatcatcagcttggaaggctgaggtaatagccattatacgatgcccgcatcctggaactcggctacctgattttca 5300
5210 5220 5230 5240 5250 5260 5270 5280 5290 5300
glyT (tRNA(UCC)Gly EcoGene#EG30036)
5380 HaeIII
5379 CfrI
5315 TaqI
5340 HpaII
5375 AccII
5315 Hpy188III
5339 Cfr10I
5371 PflMI
5304 HpyCH4V
5323 AluI
5335 TaqI
5360 AvaI
5371 BsiYI
5386 TspEI
5301 ttctgcactgaatatcgagagaagctctctttattcgagccggtaagcgaacttatcgtctcgggctacgccatcgcgtggccgaaattggtggtgggggg 5400
5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
tyrU (tRNA(GUA)Tyr EcoGene#EG30107)
5415 TaqI
5460 Hpy188III
5493 BsaWI
5414 BstBI
5459 AvaI
5472 HpaII
5487 BsiYI
5407 TaqI
5454 HaeIII
5471 BsaWI
5483 NlaIV 5494 HpaII
5406 BstBI
5427 BceAI 5438 HpyCH4III 5453 CfrI 5463 NlaIV
5483 AccB1I
5500 TaqI
5401 aaggattcgaaccttcgaagtctgtgacggcagatttacagtctgctccctttggccgctcgggaaccccaccggacttgatggtgccgactaccggaat 5500
5410 5420 5430 5440 5450 5460 5470 5480 5490 5500
tyrU (tRNA(GUA)Tyr EcoGene#EG30107)
thrU (tRNA(UGU)Thr EcoGene#EG30102)

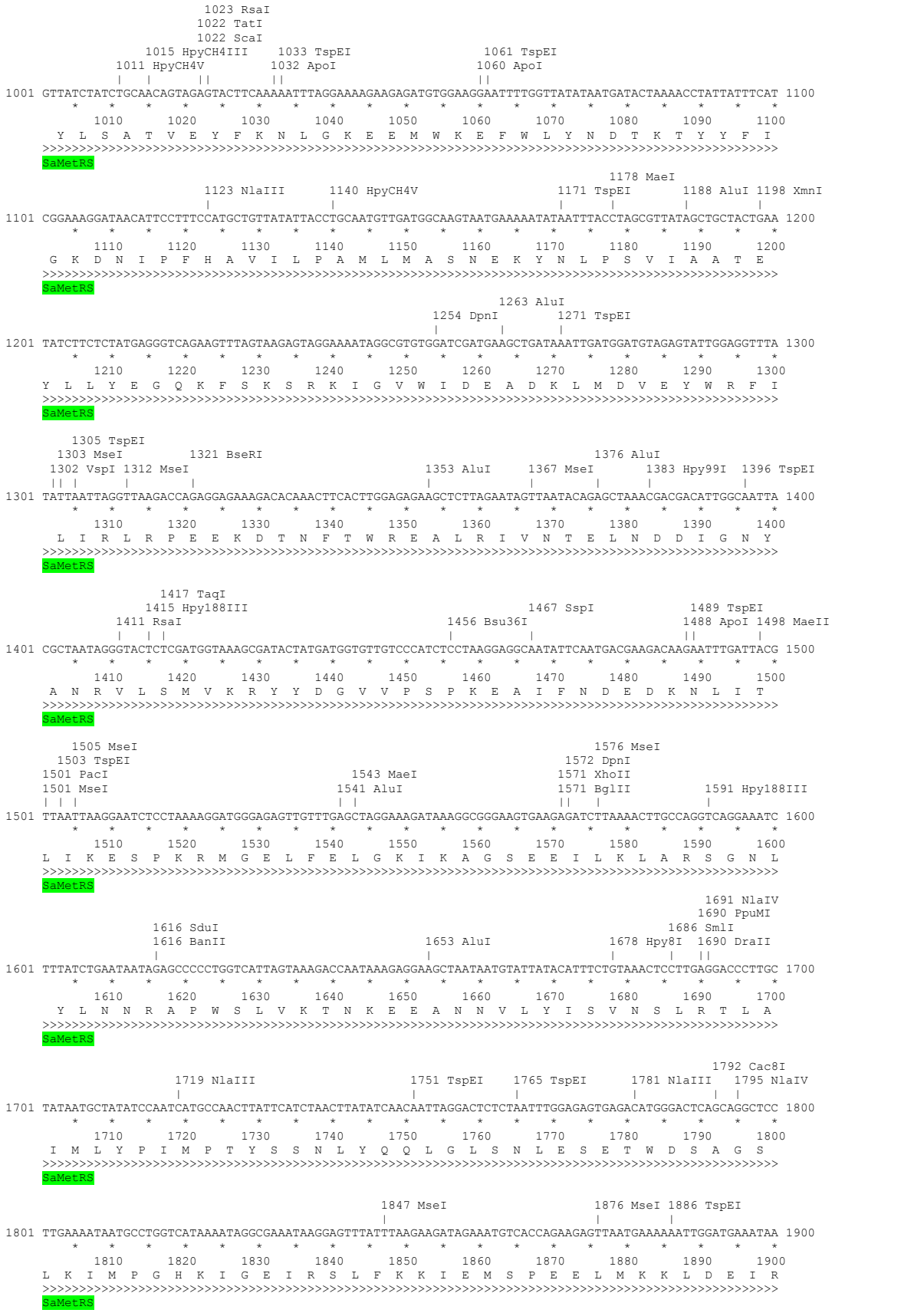
```
                                                                       5568 HhaI
                                                                      5567 AccII
                                                                     5566 HhaI
                                                   5547 MwoI         5566 Cac8I        5584 NlaIII 5596 HpyCH4V
          5507 BstEII                             5546 AluI          5566 BssHII      5583 NspI  5594 NlaIII
          |                                        ||                |||              ||          | |
5501 cgaactggtgacctactgattacaagtcagttgctctacctactgagctaagtcggcatcaagtagcgcgcactctatggagacatgcgagttcatgcaa 5600
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
          5510      5520      5530      5540      5550      5560      5570      5580      5590      5600

                                                  <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
                                                  thrUp(A52G) (tRNA promoter of thrU-tyrU-glyT-thrT o
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     thrU (tRNA(UGU)Thr EcoGene#EG30102)
                                                                          5671 Hpy188III
                                                                     5666 RsaI
                                                                    5665 NlaIV
                                                                    5665 KpnI 5675 SduI
                                                                    5665 AccB1I
                                                                    5665 Acc65I                    5695 MboI
              5615 TspEI                                         5662 MaeI    5675 BanII    5689 SduI
         5610 HpyCH4V                                           5661 StyI 5671 BspHI        5689 BanII
      5607 TspEI                                                5661 AvrII 5672 NlaIII    5687 Cac8I
      |  |    |                                                 ||  ||    ||  |            | |     |
5601 ctaaaaaattgcataatttgttttattggtcacattttatgcgacacgatgaagaaacagcctaggtacctcatgagcccgaagtggcgagcccgatctt 5700
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
          5610      5620      5630      5640      5650      5660      5670      5680      5690      5700

     <<<<<<<<<<<<<<<<<<<<<<<<
     thrUp(A52G) (tRNA promoter of thrU-tyrU-glyT-thrT o

                                                    5751 HpaII
                                                   5750 MwoI
                                                   5750 Cfr10I
                              5727 HhaI          5748 HhaI
                             5726 SfoI          5747 SfoI
                             5726 NlaIV         5747 NlaIV
                             5726 NarI          5747 NarI      5762 HaeIII
                             5726 KasI          5747 KasI     5761 CfrI
                             5726 HaeII         5747 HaeII  5759 NgoMIV              5788 TspGWI
                             5726 BbeI          5747 BbeI   5759 NaeI             5785 XhoII
                             5726 AcyI          5747 AcyI   5759 Cfr10I           5785 NlaIV
                             5726 AccB1I  5739 MwoI 5749 SgrAI 5760 HpaII         5785 BamHI
                     5718 MwoI  5729 Cac8I        5747 AccB1I 5759 Cac8I      5775 HpaII 5786 MboI    5799 StyI
                     |       || |          |       |||||       ||||            |         || |           |
5701 ccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatccgtcacatacc 5800
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
          5710      5720      5730      5740      5750      5760      5770      5780      5790      5800

                                                      5853 MwoI
                                                    5851 HhaI
                                                   5850 FspI
           5812 Cac8I                             5849 FspAI                                    5899 TspEI
      5807 MwoI                           5841 Hpy188III 5856 TaqI             5881 AluI   5893 HpyCH4V
      5807 BlpI                   5833 Cac8I     5848 BsaBI           5871 NlaIII      5888 Hpy99I
      |    |                      |       |      |||| |  |            |         |      |     |     |
5801 aaggcggctaagcgagcagatggaacatcaacgcctgcggtcaggaagatgcgcatcgacagcaagacatcatggctatcagcttgtcgtcgtgcaggaa 5900
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
          5810      5820      5830      5840      5850      5860      5870      5880      5890      5900

                                              5943 Cac8I
                 5918 NlaIII                5942 AluI
         5910 TspGWI                      5939 Cac8I          5960 AcyI          5980 Hpy188III
         |       |                       |  ||                |                   |
5901 ttgaagatttccgtttccatgacctcagacacacctgggcaagctggctgattcagtcaggcgtcccattatcagtgcttcaggaaatgggcggatggga 6000
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
          5910      5920      5930      5940      5950      5960      5970      5980      5990      6000

                                                          6056 NlaIII
       6008 XmnI                                         6055 SphI
   6004 MslI                       6035 HhaI             6055 NspI
  6003 BstXI             6026 MwoI                       6055 Cac8I
  ||   |                 |        |                      ||
6001 gtccatagaaatggttcgtaggtatgctcaccttgcgcctaatcatttgacagagcatgcgaggaaaatagacgacatttttggtgataatgtcccaaat 6100
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
          6010      6020      6030      6040      6050      6060      6070      6080      6090      6100

                                                                       6168 HpyCH4V
                                                                      6167 PstI
                                                                      6167 BfmI
                                                                     6166 SbfI
                                                                6162 HhaI
                                                              6160 MwoI
                                                              6160 HhaI
                                                              6160 Cac8I
                                                              6160 BssHII
                                                             6159 AscI       6176 TaqI
              6115 TspEI                          6150 TspEI 6161 AccII    6175 BstBI 6186 HaeIII
              |                                   |        ||||   |||      ||         |
6101 atgtcccactctgaaattatggaggatataaagaaggcgtaactgattgaattgtaatggcgcgccctgcaggattcgaacctgcggcccacgacttaga 6200
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
          6110      6120      6130      6140      6150      6160      6170      6180      6190      6200

                                                                <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
                                                                argU (tRNA(UCU)Arg EcoGene#EG30014)
```

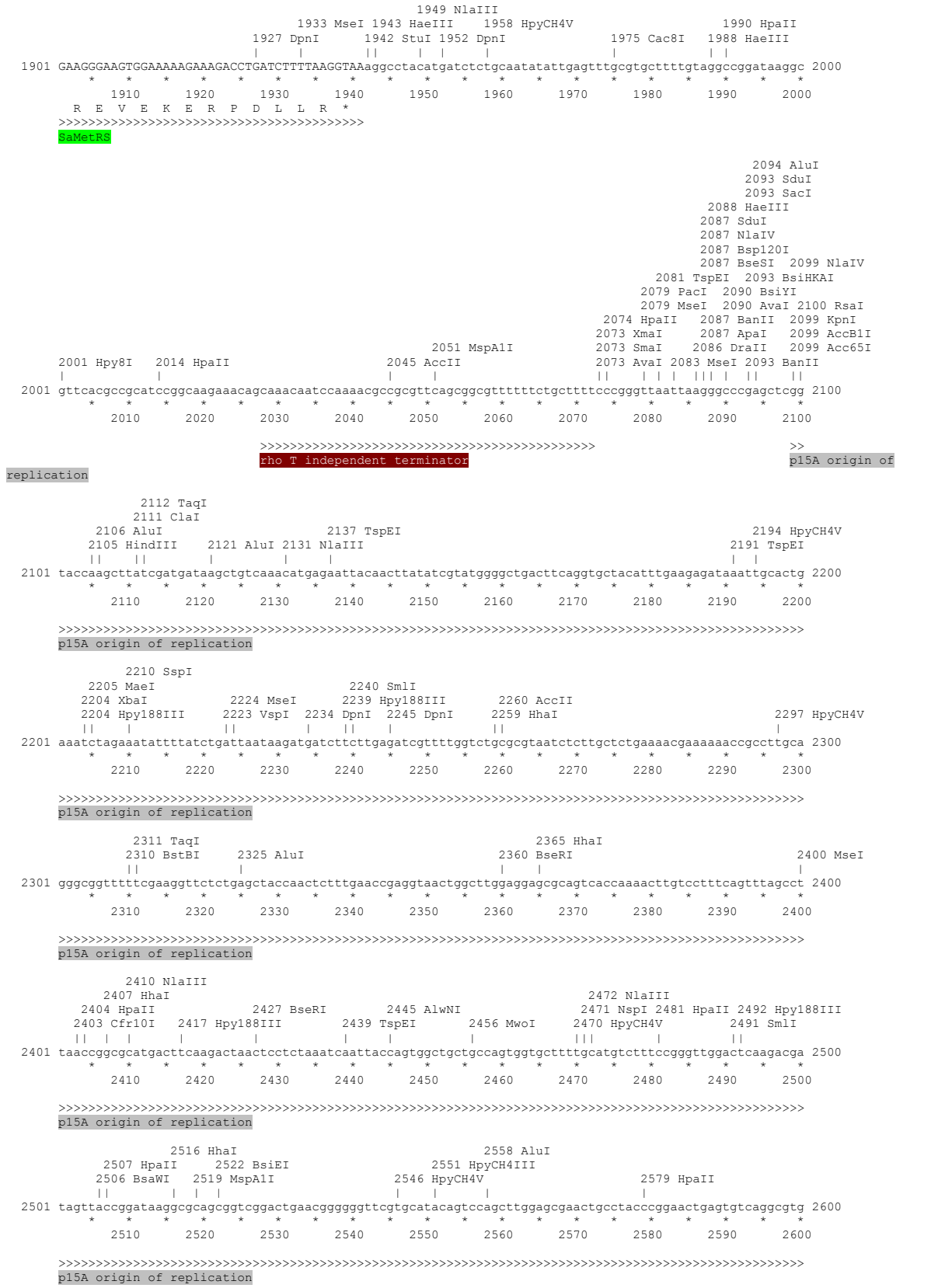
```
                                    6232 AccII                                6274 Hpy8I
                                   6231 HhaI                                  6274 HincII
                            6224 MwoI                               6265 TspEI
                           6223 AluI          6242 BsiYI       6259 AccII  6271 HpyCH4III
                           ||      ||         |                |     |     |  |
6201 aggtcgttgctctatccaactgagctaagggcgcgttgataccgcaatgcggtgtaatcgcgtgaattatacggtcaacccttgctgagtcaatggcttt 6300
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
            6210      6220      6230      6240      6250      6260      6270      6280      6290      6300

                                      <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
                                      argUp (tRNA promoter of argU operon)
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     argU (tRNA(UCU)Arg EcoGene#EG30014)

                                 6329 AccII
                              6326 Tth111I                                      6384 AccII
                             6325 BsiEI                    6355 BfmI                   6383 HhaI
      6302 MboI         6320 Hpy8I                        6354 MwoI               6378 MspA1I     6394 Hpy188III
      |                 |    ||  |                        ||                      |    ||         |
6301 tgatctggttgctgaacaagtgaacgaccgcgtctgattttctgatttatttcgctatagcggcaaacaaacgcacaccgctgcgcgtctgaatcaagaa 6400
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
            6310      6320      6330      6340      6350      6360      6370      6380      6390      6400

     <<<<<<<<<<<<<<<
     argUp (tRNA promoter of argU operon)

                                          6438 AluI                                  6481 NlaIV
                                         6437 HindIII                                6481 HaeIII
                                  6430 HpyCH4V                                 6475 TaqI
                                 6429 PstI                          6464 PsiI 6474 ClaI
                 6413 NlaIII     6429 BfmI 6439 MwoI     6453 TspGWI       6471 BstXI        6489 AluI
                 |               ||      |||             |          |      |  ||     |       |
6401 aacccgtattttcatgtatcaaagtgacctgcagccaagcttggattgcgacacggagttactttataatccaatcgattggccccttagctcagtggtt 6500
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
            6410      6420      6430      6440      6450      6460      6470      6480      6490      6500

                                                                                     >>>>>>>>>>>>>>>>>>>>
                                                                                     ileX (tRNA(CAU)Ile EcoGene#EG30046)

                                                                6560 MwoI
                                                                6560 BglI
                                                               6559 HaeIII
                                                              6558 EagI
                                                              6558 CfrI
                                                             6557 NotI
                                                           6555 MspA1I
                                                     6549 HaeIII
        6504 Cac8I                                 6547 NlaIV 6558 BsiEI
        |                                          | |     | ||||
6501 agagcaggcgactcataatcgcttggtcgctggttcaagtccagcaggggccaccagcggccgcaaaggc 6570
         *    *    *    *    *    *    *    *    *    *    *    *    *    *
            6510      6520      6530      6540      6550      6560      6570

     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     ileX (tRNA(CAU)Ile EcoGene#EG30046)
```

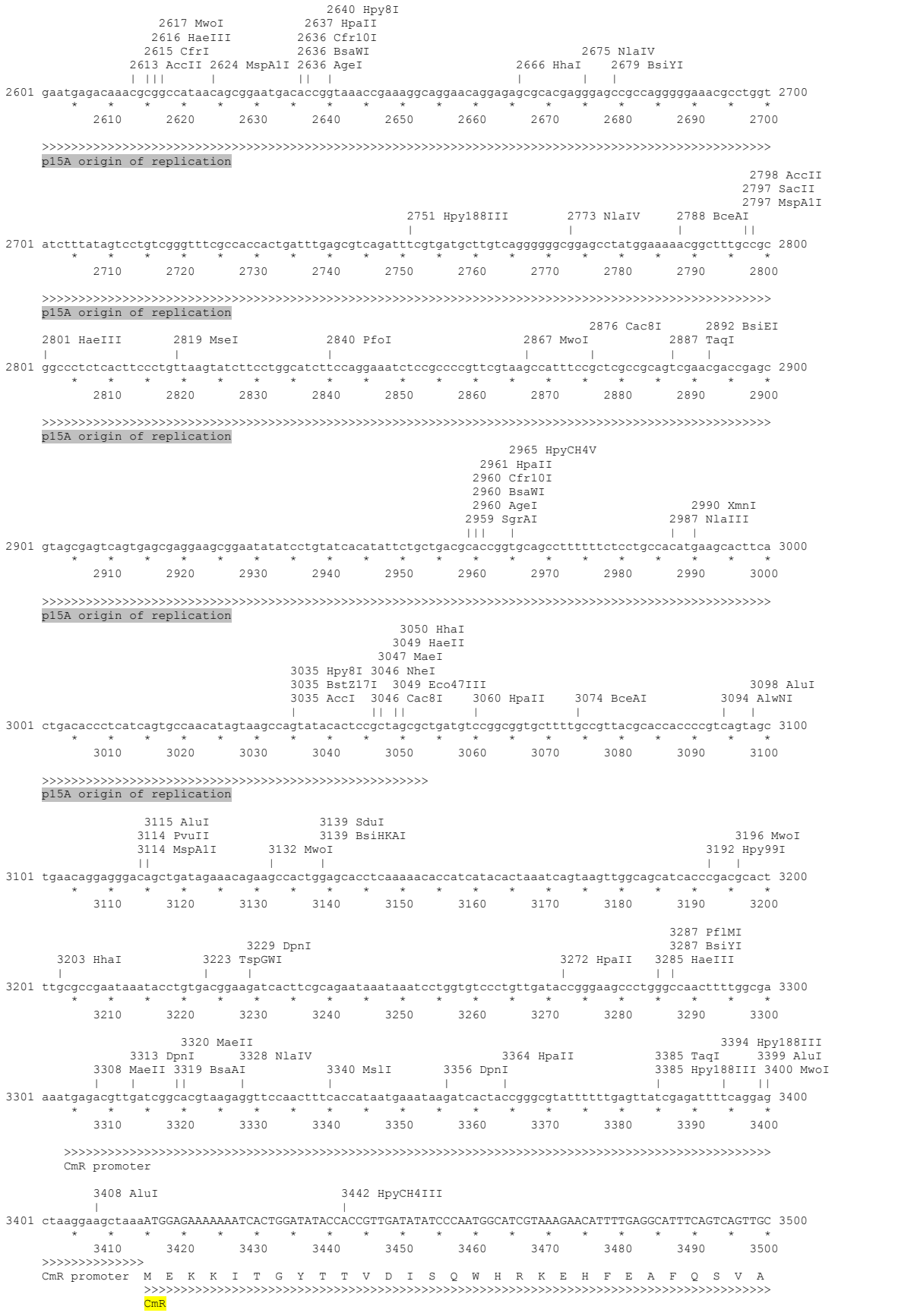
9.7 pTEc1.1G-R/L sequence

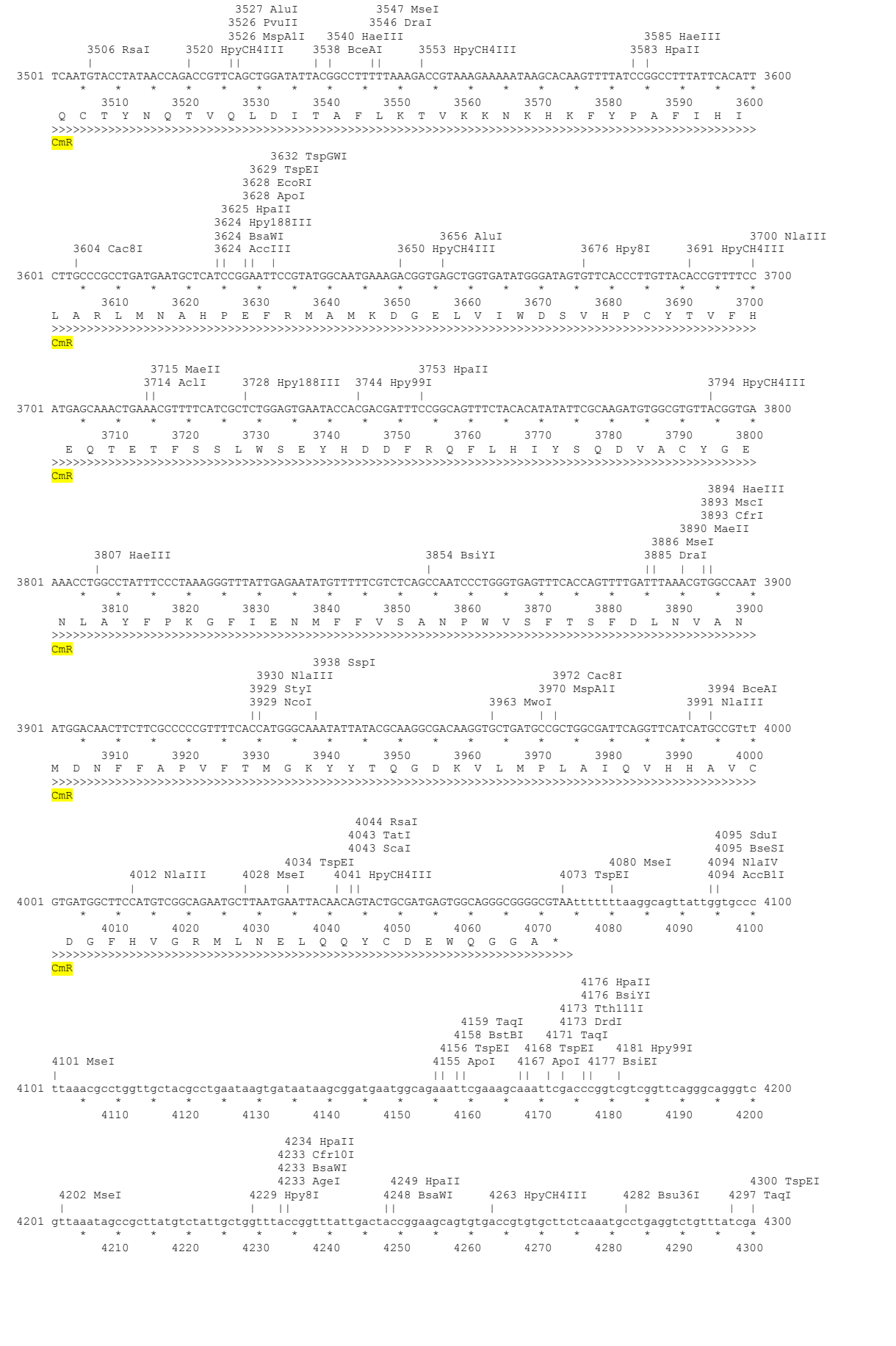
```
                                                            56 BceAI
                                  30 MseI     42 BstXI     55 HaeIII
                              26 AluI     38 TspEI        54 EagI
                             25 HindIII  37 EcoRI         54 CfrI 62 SpeI
                           23 MseI       37 ApoI          54 BsiEI 63 MaeI
                           | ||   |      ||   |           |||     ||
   1 tgacgagaaatcgtcagcctttttaagctttaagccgaattccagcacactggcggccgttactagtcatcaatcatccccataatccttgttagattat 100
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
             10        20        30        40        50        60        70        80        90        100

                                                                        >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                                                        glnS' promoter
                                                                        >>>>>>>>>>>>>>>>>>>>>>>>>>>
                                                                        Extra sequence upstream

          106 MseI                                                                               197 NlaIV
         105 DraI             127 MwoI                        159 MaeII                     190 Cac8I
      102 TspEI      117 HpyCH4III         139 MseI    151 BceAI    164 Hpy8I           184 NdeI 193 AluI
      |  ||          |         |           |           |       |    |                   |     |  |   |
 101 caatttaaaaaactaacagttgtcagcctgtcccgctttaatatcatacgccgttatacgttgtttacgctttgaggaatcccatATGGCAAGCTGGAG 200
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
             110       120       130       140       150       160       170       180       190       200
                                                                                           M  A  S  W  S
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     glnS' promoter
                                           >                                                       >>>>>
                                           glnS Promoter deletion of A between GC                  Strep-Tag II
                                               >                                       >>>
                                               glnS promoter mutation G -> T           glnS promoter mutation ACG -> CAT
```

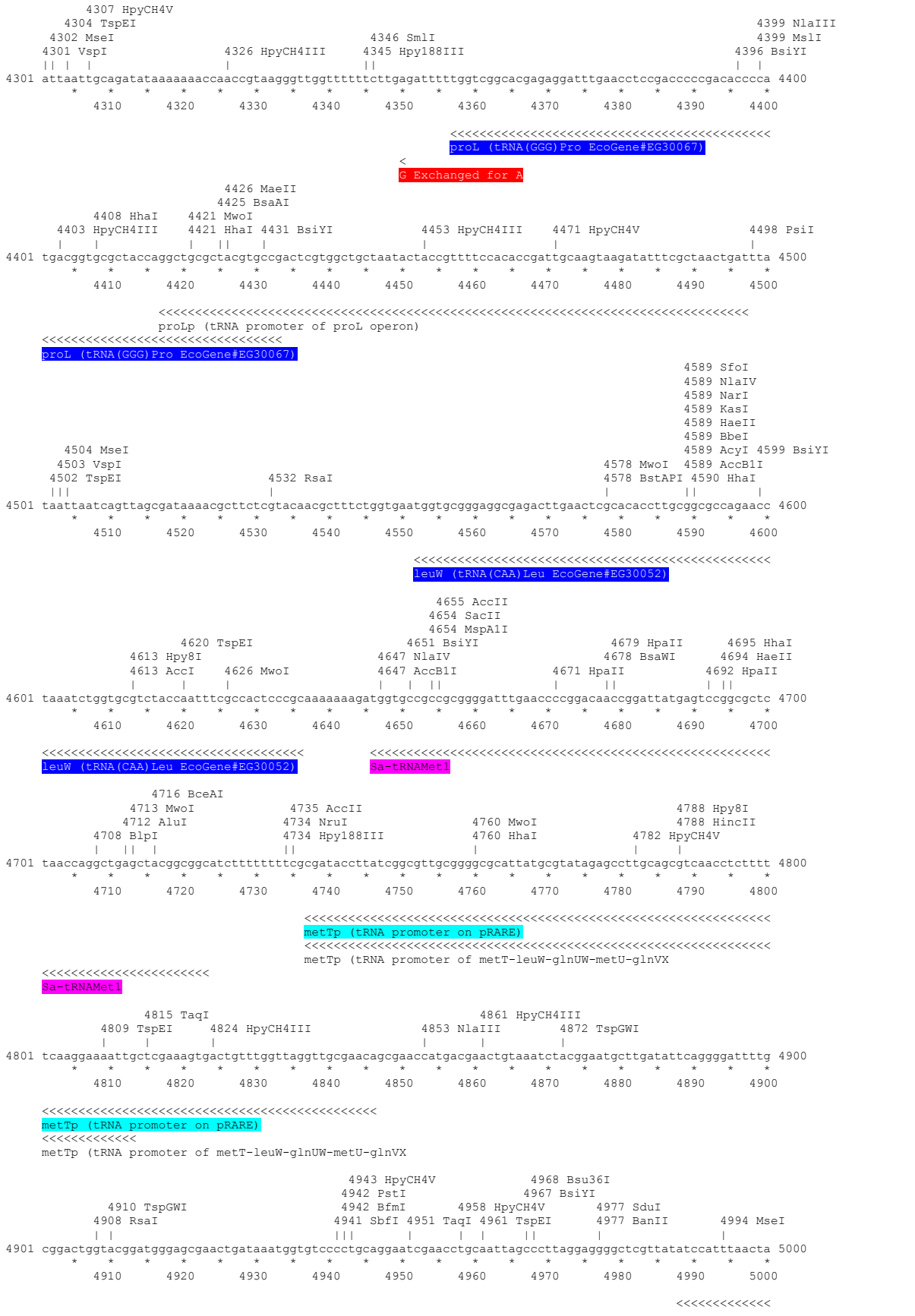

248 NlaIII
286 MseI
212 TaqI
246 HpyCH4V 260 MseI 271 BsiYI 283 TspEI
211 BstBI 222 HpyCH4V
242 MslI 251 HaeIII 264 HpyCH4V 282 ApoI
201 CCACCCGCAGTTCGAAAAGGGTGCAATGAAGGTTTTAGTAACATCTGCATGGCCTTATGTTAATGCAGTTCCACATCTGGGAAATTTAATAGGCTCTATA 300
210 220 230 240 250 260 270 280 290 300
H P Q F E K G A M K V L V T S A W P Y V N A V P H L G N L I G S I
SaMetRS
Strep-Tag II
319 MwoI
319 BstAPI
318 HpyCH4V 334 MseI
386 RsaI 395 TaqI
386 Hpy8I
383 HpyCH4III 399 TspEI
301 TTATCAGCAGATGTCTTTGCAAGATATGCCAGATTAAAATATGGAAAGGAAAATGTCGTTTTTGTAAGTGGTAGTGATGAACACGGTACACCTATCGAAA 400
310 320 330 340 350 360 370 380 390 400
L S A D V F A R Y A R L K Y G K E N V V F V S G S D E H G T P I E I
SaMetRS
451 NlaIII
450 SphI
450 NspI
450 Cac8I
442 DpnI 451 MslI
492 HindIII
434 TspEI 448 HhaI 467 TspEI 482 NlaIII 493 AluI
401 TTGAGGCAAGGAAGAGAAACATAGAACCCAAGAAATTGACCGATCAGGCGCATGCCTATGATAAGAAATTATTTATTGACACATGGAAGATAAGCTTTGA 500
410 420 430 440 450 460 470 480 490 500
E A R K R N I E P K K L T D Q A H A Y D K K L F I D T W K I S F D
SaMetRS
563 TaqI
563 Hpy188III
526 RsaI 553 MaeI 562 XhoI
526 Hpy8I 547 TspEI 562 SmlI
502 TspEI 513 Tth111I 525 TatI 546 ApoI 562 AvaI
501 CAATTATAGCAGGACTGAGTCTGAAGTACACAAAGAGTTCGTTAGAAATTTTCTAGTAAAACTCGAGAAATACATCAAGGTAGAAGAGGACGAGATTCCA 600
510 520 530 540 550 560 570 580 590 600
N Y S R T E S E V H K E F V R N F L V K L E K Y I K V E E D E I P
SaMetRS
628 HpaII
627 Hpy188III
627 BsaWI
659 HpyCH4III
685 DpnI
602 HpyCH4III
627 AccIII 638 PsiI
655 BsiYI
677 MaeI
697 TspEI
601 TACTGTGAAAAAGACAAGTTATTTCTTCCGGATAGATTTATAAAGGGAGTATGTCCTTACTGTGGCTTTGAAGACGCTAGAGGTGATCAGTGTGATAATT 700
610 620 630 640 650 660 670 680 690 700
Y C E K D K L F L P D R F I K G V C P Y C G F E D A R G D Q C D N C
SaMetRS
744 HpyCH4V
743 SduI
743 Hpy8I
743 BsiHKAI
743 BseSI
711 Hpy8I
743 ApaLI
770 MseI 779 PflMI
711 HincII 722 DpnI 736 MwoI 746 DraIII
769 DraI 779 BsiYI
794 TaqI
701 GTGGTAGATTGTTGACTCCAAGATCATTAGTAAATGCTAAGTGTGCACTATGTGGTAATCCACCTGTATTTAAAGTTACCAAGCATTGGTTTTTCGATTT 800
710 720 730 740 750 760 770 780 790 800
G R L L T P R S L V N A K C A L C G N P P V F K V T K H W F F D L
SaMetRS
880 AluI
806 TspEI
877 MseI
805 ApoI
831 EcoRV
860 MaeII
873 HpyCH4V
887 MseI
801 ATCTGAATTTGGAGATAAAATAAGGGACTGGATATCCAGTTCCAGCACAATGCCTGATAACGTAAAATCTGTTGCATTAAGCTGGGTTAAGGAAGGTCTT 900
810 820 830 840 850 860 870 880 890 900
S E F G D K I R D W I S S S S T M P D N V K S V A L S W V K E G L
SaMetRS
935 Hpy8I
935 BstZ17I
951 HpyCH4V
927 NlaIII
943 MwoI
926 NspI 935 AccI 946 BsiYI
926 BspLU11I
943 BstAPI
990 AluI
926 AflIII
942 HpyCH4V
957 HpyCH4V
988 XmnI 997 MaeI
901 AGACCAAGAAGTATAACACGAGATAACATGTGGGGTATACCTGCACCATTTGCAGGTGCAGAGAACAAAACAATATATGTCTGGTTTGAAGCTCTTCTAG 1000
910 920 930 940 950 960 970 980 990 1000
R P R S I T R D N M W G I P A P F A G A E N K T I Y V W F E A L L G
SaMetRS

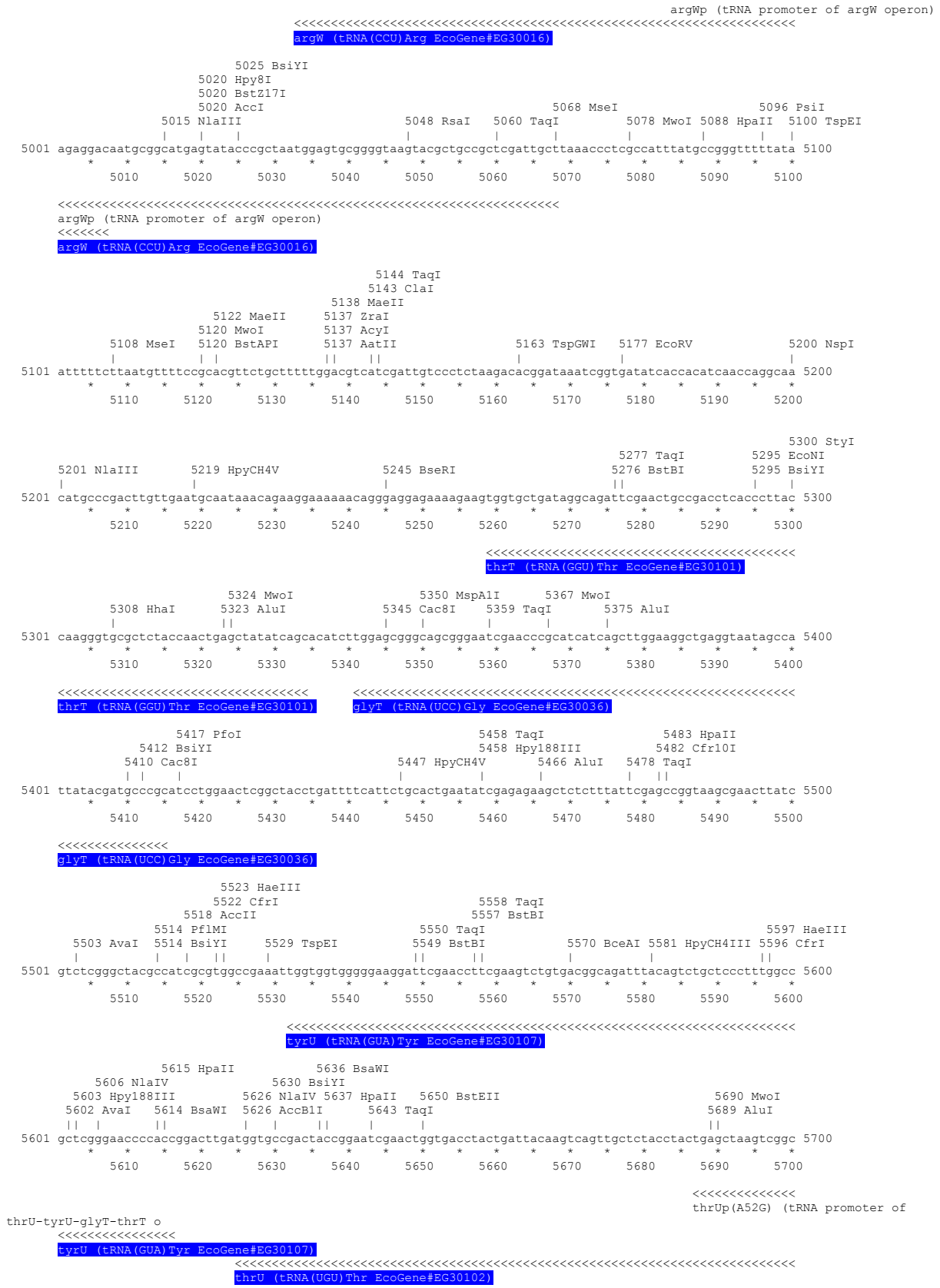

1023 RsaI
1022 TatI
1022 ScaI
1015 HpyCH4III
1033 TspEI
1061 TspEI
1011 HpyCH4V
1032 ApoI
1060 ApoI
1001 GTTATCTATCTGCAACAGTAGAGTACTTCAAAAATTTAGGAAAAGAAGAGATGTGGAAGGAATTTTGGTTATATAATGATACTAAAACCTATTATTTCAT 1100
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
Y L S A T V E Y F K N L G K E E M W K E F W L Y N D T K T Y Y F I
SaMetRS
1178 MaeI
1123 NlaIII
1140 HpyCH4V
1171 TspEI
1188 AluI
1198 XmnI
1101 CGGAAAGGATAACATTCCTTTCCATGCTGTTATATTACCTGCAATGTTGATGGCAAGTAATGAAAAATATAATTTACCTAGCGTTATAGCTGCTACTGAA 1200
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
G K D N I P F H A V I L P A M L M A S N E K Y N L P S V I A A T E
SaMetRS
1263 AluI
1254 DpnI
1271 TspEI
1201 TATCTTCTCTATGAGGGTCAGAAGTTTAGTAAGAGTAGGAAAATAGGCGTGTGGATCGATGAAGCTGATAAATTGATGGATGTAGAGTATTGGAGGTTTA 1300
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
Y L L Y E G Q K F S K S R K I G V W I D E A D K L M D V E Y W R F I
SaMetRS
1305 TspEI
1303 MseI
1321 BseRI
1376 AluI
1302 VspI
1312 MseI
1353 AluI
1367 MseI
1383 Hpy99I
1396 TspEI
1301 TATTAATTAGGTTAAGACCAGAGGAGAAAGACACAAACTTCACTTGGAGAGAAGCTCTTAGAATAGTTAATACAGAGCTAAACGACGACATTGGCAATTA 1400
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
L I R L R P E E K D T N F T W R E A L R I V N T E L N D D I G N Y
SaMetRS
1417 TaqI
1415 Hpy188III
1467 SspI
1489 TspEI
1411 RsaI
1456 Bsu36I
1488 ApoI
1498 MaeII
1401 CGCTAATAGGGTACTCTCGATGGTAAAGCGATACTATGATGGTGTTGTCCCATCTCCTAAGGAGGCAATATTCAATGACGAAGACAAGAATTTGATTACG 1500
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
A N R V L S M V K R Y Y D G V V P S P K E A I F N D E D K N L I T
SaMetRS
1505 MseI
1503 TspEI
1576 MseI
1572 DpnI
1501 PacI
1543 MaeI
1571 XhoII
1501 MseI
1541 AluI
1571 BglII
1591 Hpy188III
1501 TTAATTAAGGAATCTCCTAAAAGGATGGGAGAGTTGTTTGAGCTAGGAAAGATAAAGGCGGGAAGTGAAGAGATCTTAAAACTTGCCAGGTCAGGAAATC 1600
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
L I K E S P K R M G E L F E L G K I K A G S E E I L K L A R S G N L
SaMetRS
1691 NlaIV
1690 PpuMI
1686 SmlI
1616 SduI
1616 BanII
1653 AluI
1678 Hpy8I
1690 DraII
1601 TTTATCTGAATAATAGAGCCCCCTGGTCATTAGTAAAGACCAATAAAGAGGAAGCTAATAATGTATTATACATTTCTGTAAACTCCTTGAGGACCCTTGC 1700
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
Y L N N R A P W S L V K T N K E E A N N V L Y I S V N S L R T L A
SaMetRS
1792 Cac8I
1719 NlaIII
1751 TspEI
1765 TspEI
1781 NlaIII
1795 NlaIV
1701 TATAATGCTATATCCAATCATGCCAACTTATTCATCTAACTTATATCAACAATTAGGACTCTCTAATTTGGAGAGTGAGACATGGGACTCAGCAGGCTCC 1800
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
I M L Y P I M P T Y S S N L Y Q Q L G L S N L E S E T W D S A G S
SaMetRS
1847 MseI
1876 MseI
1886 TspEI
1801 TTGAAAATAATGCCTGGTCATAAAATAGGCGAAATAAGGAGTTTATTTAAGAAGATAGAAATGTCACCAGAAGAGTTAATGAAAAAATTGGATGAAATAA 1900
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
L K I M P G H K I G E I R S L F K K K I E M S P E E L M K K L D E I R
SaMetRS

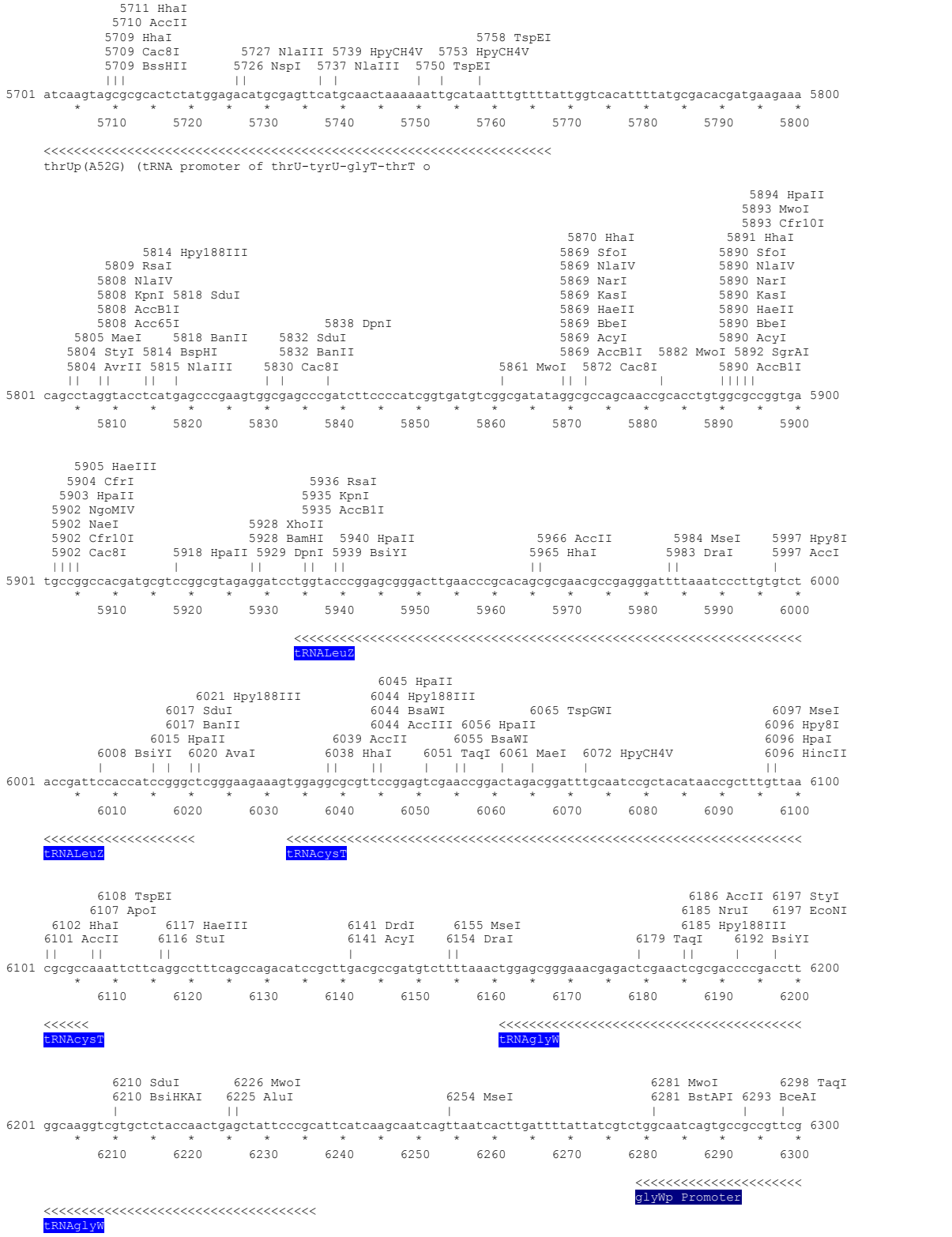

1949 NlaIII
1933 MseI 1943 HaeIII 1958 HpyCH4V 1990 HpaII
1927 DpnI 1942 StuI 1952 DpnI 1975 Cac8I 1988 HaeIII
1901 GAAGGGAAGTGGAAAAAGAAAGACCTGATCTTTTAAGGTAAaggcctacatgatctctgcaatatattgagtttgcgtgcttttgtaggccggataaggc 2000
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
R E V E K E R P D L L R *
SaMetRS
2094 AluI
2093 SduI
2093 SacI
2088 HaeIII
2087 SduI
2087 NlaIV
2087 Bsp120I
2087 BseSI 2099 NlaIV
2081 TspEI 2093 BsiHKAI
2079 PacI 2090 BsiYI
2079 MseI 2090 AvaI 2100 RsaI
2074 HpaII 2087 BanII 2099 KpnI
2073 XmaI 2087 ApaI 2099 AccB1I
2051 MspA1I 2073 SmaI 2086 DraII 2099 Acc65I
2001 Hpy8I 2014 HpaII 2045 AccII 2073 AvaI 2083 MseI 2093 BanII
2001 gttcacgccgcatccggcaagaaacagcaaacaatccaaaacgccgcgttcagcggcgtttttctgctttcccgggttaattaagggcccgagctcgg 2100
2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
rho T independent terminator
p15A origin of replication
2112 TaqI
2111 ClaI
2106 AluI 2137 TspEI 2194 HpyCH4V
2105 HindIII 2121 AluI 2131 NlaIII 2191 TspEI
2101 taccaagcttatcgatgataagctgtcaaacatgagaattacaacttatatcgtatggggctgacttcaggtgctacatttgaagagataaattgcactg 2200
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
p15A origin of replication
2210 SspI
2205 MaeI 2240 SmlI
2204 XbaI 2224 MseI 2239 Hpy188III 2260 AccII
2204 Hpy188III 2223 VspI 2234 DpnI 2245 DpnI 2259 HhaI 2297 HpyCH4V
2201 aaatctagaaatattttatctgattaataagatgatcttcttgagatcgttttggtctgcgcgtaatctcttgctctgaaaacgaaaaaaccgccttgca 2300
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
p15A origin of replication
2311 TaqI 2365 HhaI
2310 BstBI 2325 AluI 2360 BseRI 2400 MseI
2301 gggcggtttttcgaaggttctctgagctaccaactctttgaaccgaggtaactggcttggaggagcgcagtcaccaaaacttgtcctttcagtttagcct 2400
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
p15A origin of replication
2410 NlaIII
2407 HhaI 2472 NlaIII
2404 HpaII 2427 BseRI 2445 AlwNI 2471 NspI 2481 HpaII 2492 Hpy188III
2403 Cfr10I 2417 Hpy188III 2439 TspEI 2456 MwoI 2470 HpyCH4V 2491 SmlI
2401 taaccggcgcatgacttcaagactaactcctctaaatcaattaccagtggctgctgccagtggtgcttttgcatgtctttccgggttggactcaagacga 2500
2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
p15A origin of replication
2516 HhaI 2558 AluI
2507 HpaII 2522 BsiEI 2551 HpyCH4III
2506 BsaWI 2519 MspA1I 2546 HpyCH4V 2579 HpaII
2501 tagttaccggataaggcgcagcggtcggactgaacgggggttcgtgcatacagtccagcttggagcgaactgcctacccggaactgagtgtcaggcgtg 2600
2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
p15A origin of replication

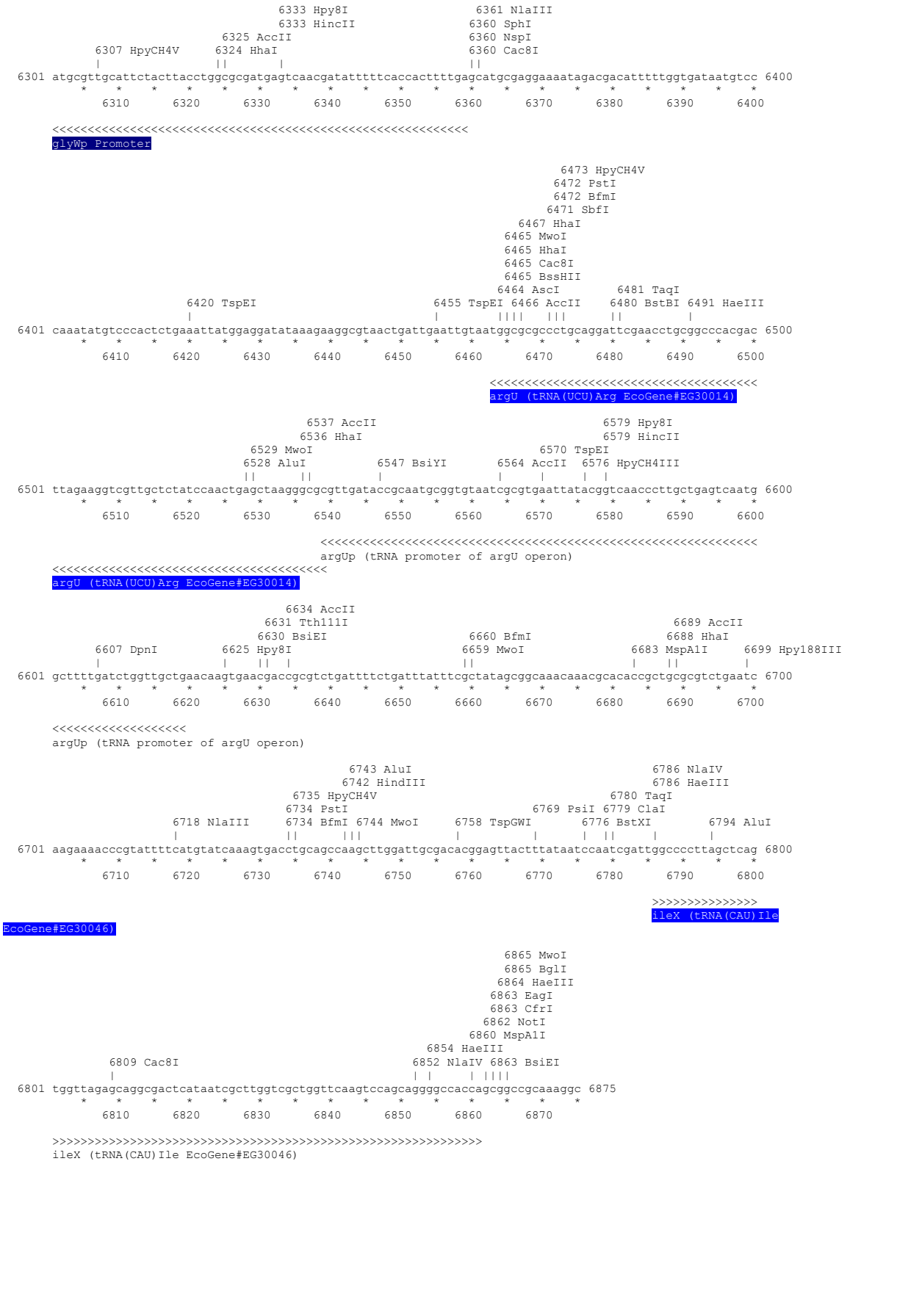

2640 Hpy8I
2617 MwoI
2637 HpaII
2616 HaeIII
2636 Cfr10I
2615 CfrI
2636 BsaWI
2675 NlaIV
2613 AccII 2624 MspA1I 2636 AgeI
2666 HhaI
2679 BsiYI
2601 gaatgagacaaacgcggccataacagcggaatgacaccggtaaaccgaaaggcaggaacaggagagcgcacgagggagccgccaggggaaacgcctggt 2700
2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
p15A origin of replication
2798 AccII
2797 SacII
2797 MspA1I
2751 Hpy188III
2773 NlaIV
2788 BceAI
2701 atctttatagtcctgtcgggtttcgccaccactgatttgagcgtcagatttcgtgatgcttgtcaggggggcggagcctatggaaaaacggctttgccgc 2800
2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
p15A origin of replication
2876 Cac8I
2892 BsiEI
2801 HaeIII
2819 MseI
2840 PfoI
2867 MwoI
2887 TaqI
2801 ggccctctcacttccctgttaagtatcttcctggcatcttccaggaaatctccgccccgttcgtaagccatttccgctcgccgcagtcgaacgaccgagc 2900
2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
p15A origin of replication
2965 HpyCH4V
2961 HpaII
2960 Cfr10I
2960 BsaWI
2960 AgeI
2990 XmnI
2959 SgrAI
2987 NlaIII
2901 gtagcgagtcagtgagcgaggaagcggaatatatcctgtatcacatattctgctgacgcaccggtgcagccttttttctcctgccacatgaagcacttca 3000
2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
p15A origin of replication
3050 HhaI
3049 HaeII
3047 MaeI
3035 Hpy8I 3046 NheI
3035 BstZ17I 3049 Eco47III
3098 AluI
3035 AccI 3046 Cac8I
3060 HpaII
3074 BceAI
3094 AlwNI
3001 ctgacaccctcatcagtgccaacatagtaagccagtatacactccgctagcgctgatgtccggcggtgcttttgccgttacgcaccaccccgtcagtagc 3100
3010 3020 3030 3040 3050 3060 3070 3080 3090 3100
p15A origin of replication
3115 AluI
3139 SduI
3114 PvuII
3139 BsiHKAI
3196 MwoI
3114 MspA1I
3132 MwoI
3192 Hpy99I
3101 tgaacaggagggacagctgatagaaacagaagccactggagcacctcaaaaacaccatcatacactaaatcagtaagttggcagcatcacccgacgcact 3200
3110 3120 3130 3140 3150 3160 3170 3180 3190 3200
3287 PflMI
3287 BsiYI
3229 DpnI
3203 HhaI
3223 TspGWI
3272 HpaII
3285 HaeIII
3201 ttgcgccgaataaatacctgtgacggaagatcacttcgcagaataaataaatcctggtgtccctgttgataccgggaagccctgggccaacttttggcga 3300
3210 3220 3230 3240 3250 3260 3270 3280 3290 3300
3320 MaeII
3394 Hpy188III
3313 DpnI
3328 NlaIV
3364 HpaII
3385 TaqI
3399 AluI
3308 MaeII 3319 BsaAI
3340 MslI
3356 DpnI
3385 Hpy188III 3400 MwoI
3301 aaatgagacgttgatcggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgggcgtatttttttgagttatcgagattttcaggag 3400
3310 3320 3330 3340 3350 3360 3370 3380 3390 3400
CmR promoter
3408 AluI
3442 HpyCH4III
3401 ctaaggaagctaaaATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGC 3500
3410 3420 3430 3440 3450 3460 3470 3480 3490 3500
CmR promoter M E K K I T G Y T T V D I S Q W H R K E H F E A F Q S V A
CmR


3527 AluI
3547 MseI
3526 PvuII
3546 DraI
3526 MspA1I
3540 HaeIII
3585 HaeIII
3506 RsaI
3520 HpyCH4III
3538 BceAI
3553 HpyCH4III
3583 HpaII
3501 TCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATT 3600
3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
Q C T Y N Q T V Q L D I T A F L K T V K K N K H K F Y P A F I H I
CmR
3632 TspGWI
3629 TspEI
3628 EcoRI
3628 ApoI
3625 HpaII
3624 Hpy188III
3624 BsaWI
3656 AluI
3700 NlaIII
3604 Cac8I
3624 AccIII
3650 HpyCH4III
3676 Hpy8I
3691 HpyCH4III
3601 CTTGCCCGCCTGATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCC 3700
3610 3620 3630 3640 3650 3660 3670 3680 3690 3700
L A R L M N A H P E F R M A M K D G E L V I W D S V H P C Y T V F H
CmR
3715 MaeII
3753 HpaII
3714 AclI
3728 Hpy188III
3744 Hpy99I
3794 HpyCH4III
3701 ATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGA 3800
3710 3720 3730 3740 3750 3760 3770 3780 3790 3800
E Q T E T F S S L W S E Y H D D F R Q F L H I Y S Q D V A C Y G E
CmR
3894 HaeIII
3893 MscI
3893 CfrI
3890 MaeII
3886 MseI
3807 HaeIII
3854 BsiYI
3885 DraI
3801 AAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAAT 3900
3810 3820 3830 3840 3850 3860 3870 3880 3890 3900
N L A Y F P K G F I E N M F F V S A N P W V S F T S F D L N V A N
CmR
3938 SspI
3930 NlaIII
3972 Cac8I
3929 StyI
3970 MspA1I
3994 BceAI
3929 NcoI
3963 MwoI
3991 NlaIII
3901 ATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTtT 4000
3910 3920 3930 3940 3950 3960 3970 3980 3990 4000
M D N F F A P V F T M G K Y Y T Q G D K V L M P L A I Q V H H A V C
CmR
4044 RsaI
4043 TatI
4095 SduI
4043 ScaI
4095 BseSI
4034 TspEI
4080 MseI
4094 NlaIV
4012 NlaIII
4028 MseI
4041 HpyCH4III
4073 TspEI
4094 AccB1I
4001 GTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAAttttttttaaggcagttattggtgccc 4100
4010 4020 4030 4040 4050 4060 4070 4080 4090 4100
D G F H V G R M L N E L Q Q Y C D E W Q G G A *
CmR
4176 HpaII
4176 BsiYI
4173 Tth111I
4159 TaqI
4173 DrdI
4158 BstBI
4171 TaqI
4156 TspEI
4168 TspEI
4181 Hpy99I
4101 MseI
4155 ApoI
4167 ApoI
4177 BsiEI
4101 ttaaacgcctggttgctacgcctgaataagtgataataagcggatgaatggcagaaattcgaaagcaaattcgacccggtcgtcggttcagggcagggtc 4200
4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
4234 HpaII
4233 Cfr10I
4233 BsaWI
4233 AgeI
4249 HpaII
4300 TspEI
4202 MseI
4229 Hpy8I
4248 BsaWI
4263 HpyCH4III
4282 Bsu36I
4297 TaqI
4201 gttaaatagccgcttatgtctattgctggtttaccggtttattgactaccggaagcagtgtgaccgtgtgcttctcaaatgcctgaggtctgtttatcga 4300
4210 4220 4230 4240 4250 4260 4270 4280 4290 4300

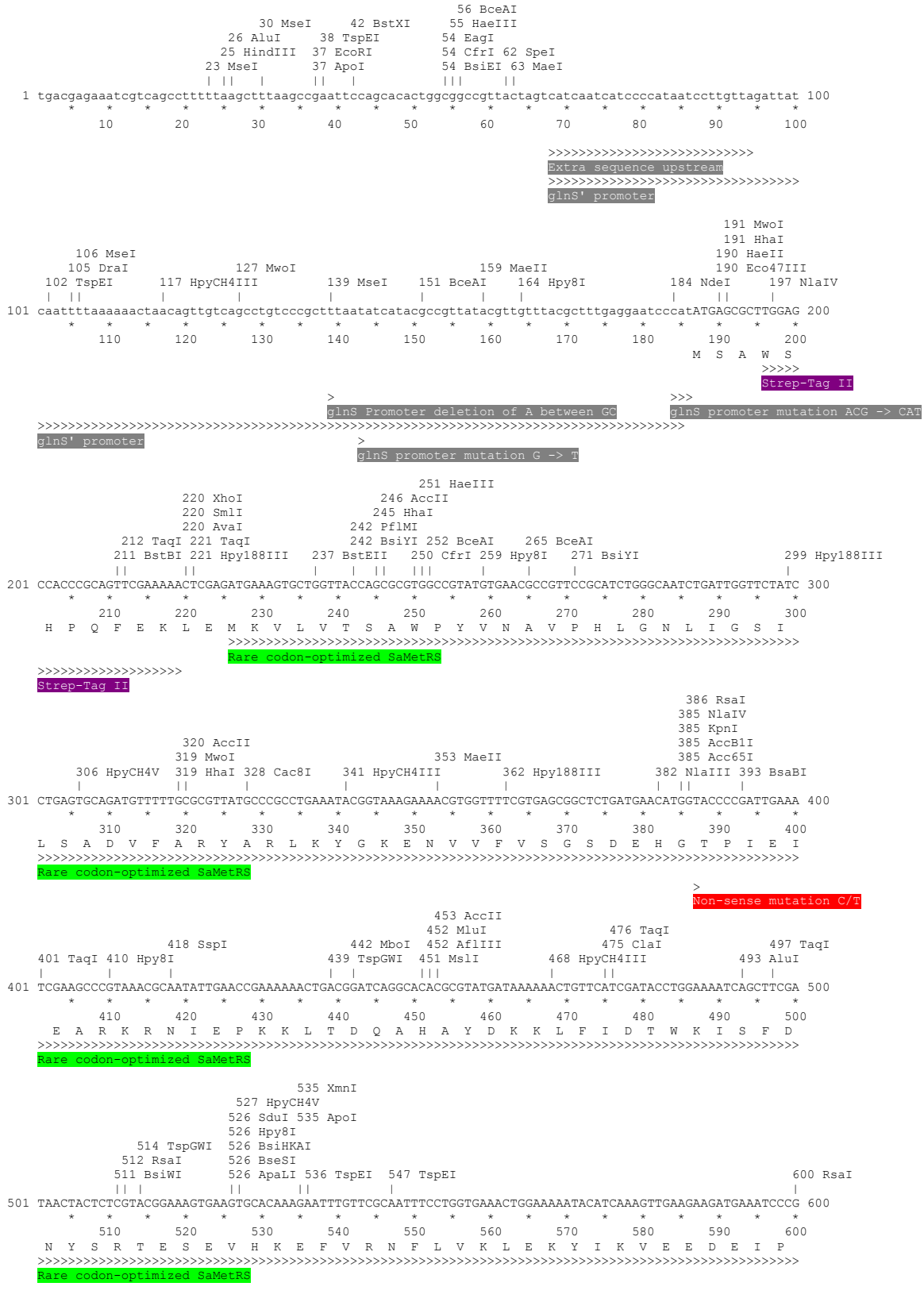

4307 HpyCH4V
4304 TspEI
4399 NlaIII
4302 MseI
4346 SmlI
4399 MslI
4301 VspI
4326 HpyCH4III
4345 Hpy188III
4396 BsiYI
4301 attaattgcagatataaaaaaaccaaccgtaagggttggttttttcttgagatttttggtcggcacgagaggatttgaaccttccgaccccccgacacccca 4400
4310 4320 4330 4340 4350 4360 4370 4380 4390 4400
proL (tRNA(GGG)Pro EcoGene#EG30067)
G Exchanged for A
4426 MaeII
4425 BsaAI
4408 HhaI
4421 MwoI
4403 HpyCH4III
4421 HhaI 4431 BsiYI
4453 HpyCH4III
4471 HpyCH4V
4498 PsiI
4401 tgacggtgcgctaccaggctgcgctacgtgccgactcgtggctgctaatactaccgttttccacaccgattgcaagtaagatatttcgctaactgattta 4500
4410 4420 4430 4440 4450 4460 4470 4480 4490 4500
proLp (tRNA promoter of proL operon)
proL (tRNA(GGG)Pro EcoGene#EG30067)
4589 SfoI
4589 NlaIV
4589 NarI
4589 KasI
4589 HaeII
4589 BbeI
4589 AcyI 4599 BsiYI
4504 MseI
4503 VspI
4578 MwoI 4589 AccB1I
4502 TspEI
4532 RsaI
4578 BstAPI 4590 HhaI
4501 taattaatcagttagcgataaaacgcttctcgtacaacgctttctggtgaatggtgcgggaggcgagacttgaactcgcacaccttgcggcgccagaacc 4600
4510 4520 4530 4540 4550 4560 4570 4580 4590 4600
leuW (tRNA(CAA)Leu EcoGene#EG30052)
4655 AccII
4654 SacII
4654 MspA1I
4620 TspEI
4651 BsiYI
4679 HpaII
4695 HhaI
4613 Hpy8I
4647 NlaIV
4678 BsaWI
4694 HaeII
4613 AccI
4626 MwoI
4647 AccB1I
4671 HpaII
4692 HpaII
4601 taaatctggtgcgtctaccaatttcgccactcccgcaaaaaaagatggtgccgccgcgggggatttgaaccccggacaaccggattatgagtccggcgctc 4700
4610 4620 4630 4640 4650 4660 4670 4680 4690 4700
leuW (tRNA(CAA)Leu EcoGene#EG30052)
Sa-tRNAMet1
4716 BceAI
4713 MwoI
4735 AccII
4788 Hpy8I
4712 AluI
4734 NruI
4760 MwoI
4788 HincII
4708 BlpI
4734 Hpy188III
4760 HhaI
4782 HpyCH4V
4701 taaccaggctgagctacggcggcatcttttttttcgcgataccttatcggcgttgcggggcgcattatgcgtatagagccttgcagcgtcaaccttcttt 4800
4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
metTp (tRNA promoter on pRARE)
metTp (tRNA promoter of metT-leuW-glnUW-metU-glnVX
Sa-tRNAMet1
4815 TaqI
4861 HpyCH4III
4809 TspEI
4824 HpyCH4III
4853 NlaIII
4872 TspGWI
4801 tcaaggaaaattgctcgaaagtgactgtttggttaggttgcgaacagcgaaccatgacgaactgtaaatctacggaatgcttgatattcaggggatttttg 4900
4810 4820 4830 4840 4850 4860 4870 4880 4890 4900
metTp (tRNA promoter on pRARE)
metTp (tRNA promoter of metT-leuW-glnUW-metU-glnVX
4943 HpyCH4V
4968 Bsu36I
4942 PstI
4967 BsiYI
4910 TspGWI
4942 BfmI
4958 HpyCH4V
4977 SduI
4908 RsaI
4941 SbfI 4951 TaqI 4961 TspEI
4977 BanII
4994 MseI
4901 cggactggtacggatgggagcgaactgataaatggtgtcccctgcaggaatcgaacctgcaattagcccttaggaggggctcgttatatccatttaacta 5000
4910 4920 4930 4940 4950 4960 4970 4980 4990 5000

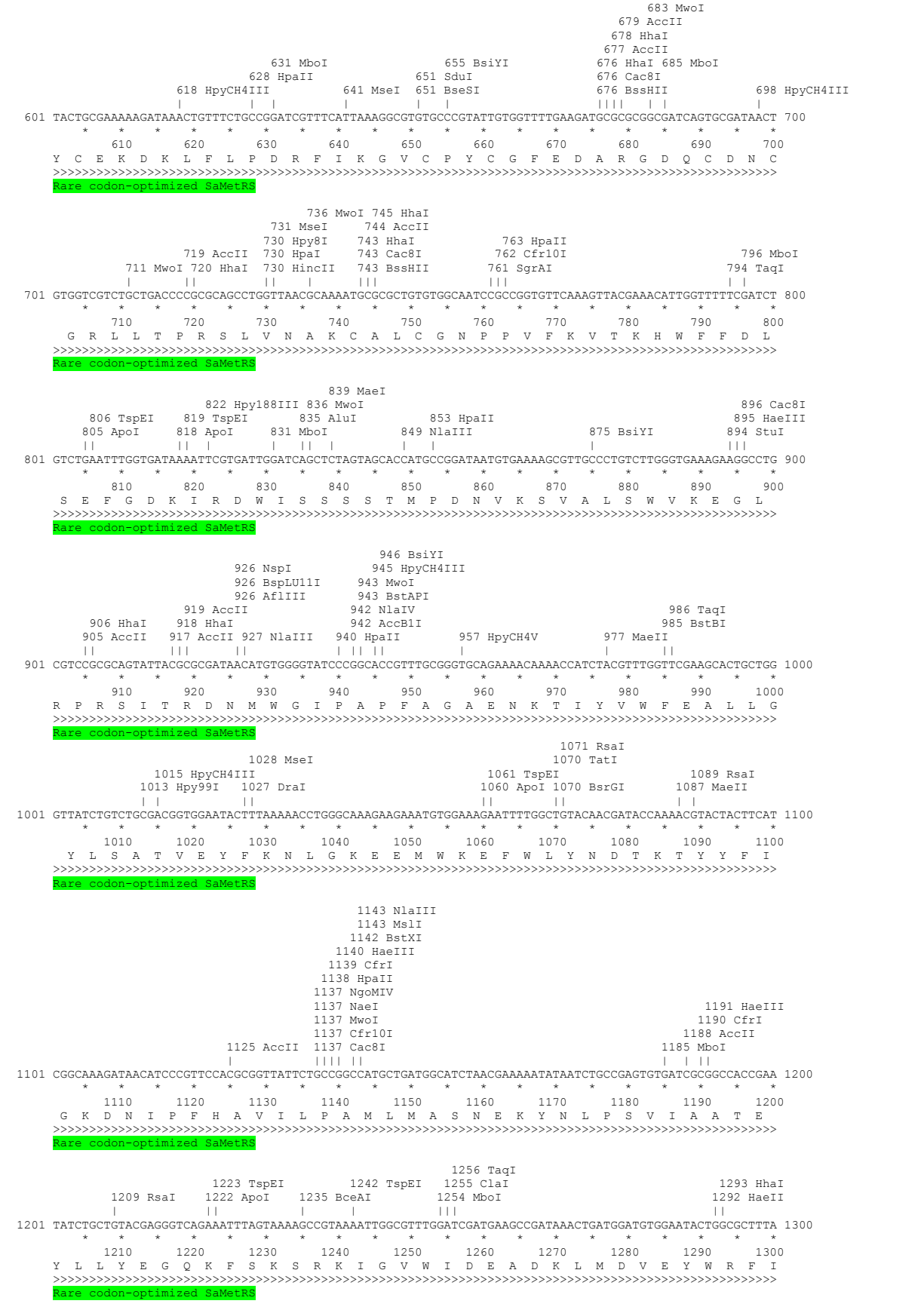

argWp (tRNA promoter of argW operon)
argW (tRNA(CCU)Arg EcoGene#EG30016)
5025 BsiYI
5020 Hpy8I
5020 BstZ17I
5020 AccI
5068 MseI
5096 PsiI
5015 NlaIII
5048 RsaI
5060 TaqI
5078 MwoI
5088 HpaII
5100 TspEI
5001 agaggacaatgcggcatgagtatacccgctaatggagtgcggggtaagtacgctgccgctcgattgcttaaaccctcgccatttatgccgggtttttata 5100
5010
5020
5030
5040
5050
5060
5070
5080
5090
5100
argWp (tRNA promoter of argW operon)
argW (tRNA(CCU)Arg EcoGene#EG30016)
5144 TaqI
5143 ClaI
5138 MaeII
5122 MaeII
5137 ZraI
5120 MwoI
5137 AcyI
5108 MseI
5120 BstAPI
5137 AatII
5163 TspGWI
5177 EcoRV
5200 NspI
5101 atttttcttaatgttttccgcacgttctgctttttggacgtcatcgattgtccctctaagacacggataaatcggtgatatcaccacatcaaccaggcaa 5200
5110
5120
5130
5140
5150
5160
5170
5180
5190
5200
5300 StyI
5277 TaqI
5295 EcoNI
5201 NlaIII
5219 HpyCH4V
5245 BseRI
5276 BstBI
5295 BsiYI
5201 catgcccgacttgttgaatgcaataaacagaaggaaaaaacagggaggagaaaagaagtggtgctgataggcagattcgaactgccgacctcacccttac 5300
5210
5220
5230
5240
5250
5260
5270
5280
5290
5300
thrT (tRNA(GGU)Thr EcoGene#EG30101)
5324 MwoI
5350 MspA1I
5367 MwoI
5308 HhaI
5323 AluI
5345 Cac8I
5359 TaqI
5375 AluI
5301 caagggtgcgctctaccaactgagctatatcagcacatcttggagcgggcagcgggaatcgaacccgcatcatcagcttggaaggctgaggtaatagcca 5400
5310
5320
5330
5340
5350
5360
5370
5380
5390
5400
thrT (tRNA(GGU)Thr EcoGene#EG30101)
glyT (tRNA(UCC)Gly EcoGene#EG30036)
5417 PfoI
5458 TaqI
5483 HpaII
5412 BsiYI
5458 Hpy188III
5482 Cfr10I
5410 Cac8I
5447 HpyCH4V
5466 AluI
5478 TaqI
5401 ttatacgatgcccgcatcctggaactcggctacctgattttcattctgcactgaatatcgagagaagctctctttattcgagccggtaagcgaacttatc 5500
5410
5420
5430
5440
5450
5460
5470
5480
5490
5500
glyT (tRNA(UCC)Gly EcoGene#EG30036)
5523 HaeIII
5522 CfrI
5558 TaqI
5518 AccII
5557 BstBI
5514 PflMI
5550 TaqI
5597 HaeIII
5503 AvaI
5514 BsiYI
5529 TspEI
5549 BstBI
5570 BceAI
5581 HpyCH4III
5596 CfrI
5501 gtctcgggctacgccatcgcgtggccgaaattggtggtgggggaaggattcgaaccttcgaagtctgtgacggcagatttacagtctgctccctttggcc 5600
5510
5520
5530
5540
5550
5560
5570
5580
5590
5600
tyrU (tRNA(GUA)Tyr EcoGene#EG30107)
5615 HpaII
5636 BsaWI
5606 NlaIV
5630 BsiYI
5603 Hpy188III
5626 NlaIV
5637 HpaII
5650 BstEII
5690 MwoI
5602 AvaI
5614 BsaWI
5626 AccB1I
5643 TaqI
5689 AluI
5601 gctcgggaaccccaccggacttgatggtgccgactaccggaatcgaactggtgacctactgattacaagtcagttgctctacctactgagctaagtcggc 5700
5610
5620
5630
5640
5650
5660
5670
5680
5690
5700
thrUp(A52G) (tRNA promoter of
thrU-tyrU-glyT-thrT o
tyrU (tRNA(GUA)Tyr EcoGene#EG30107)
thrU (tRNA(UGU)Thr EcoGene#EG30102)

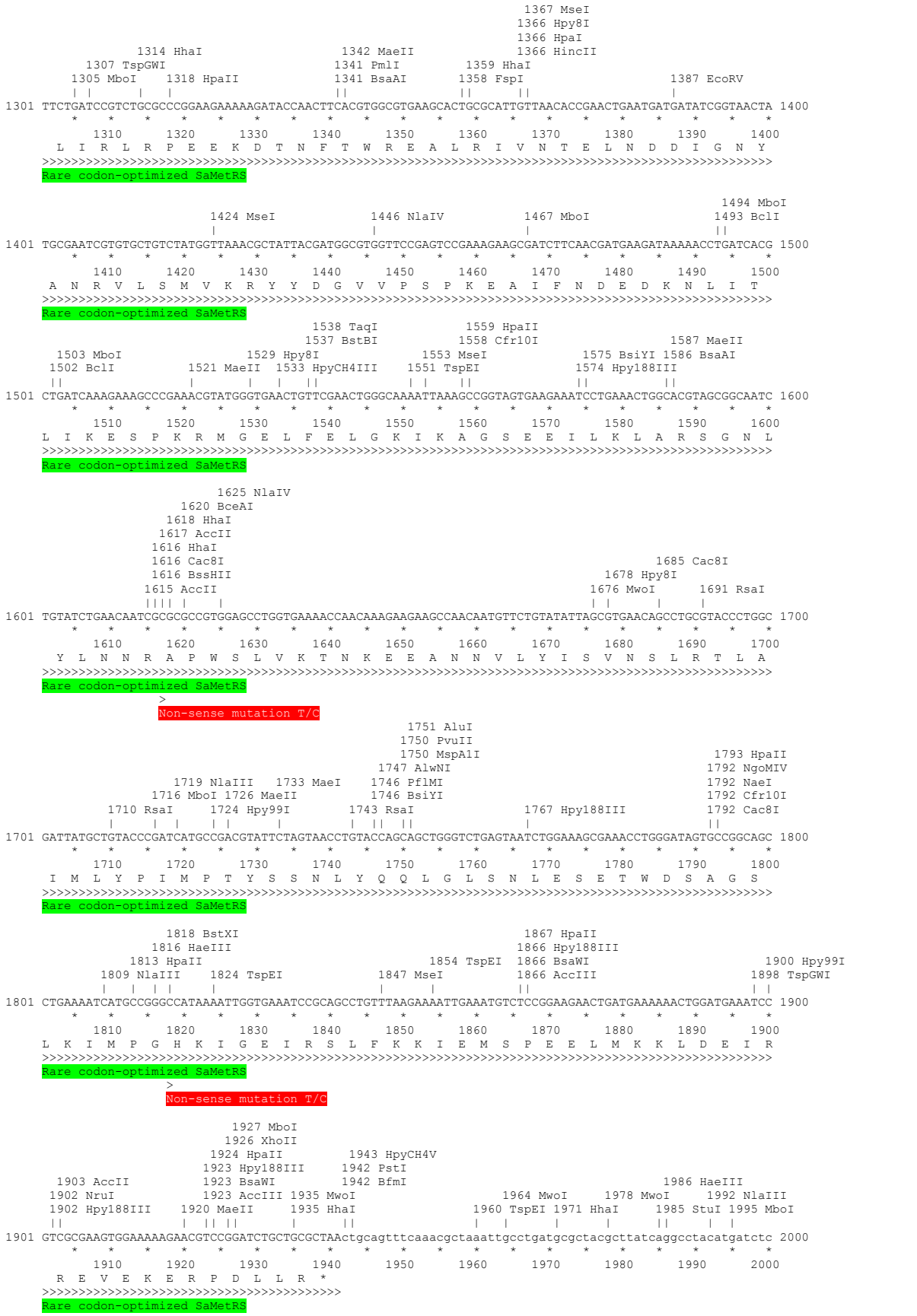

5711 HhaI
5710 AccII
5709 HhaI
5709 Cac8I
5709 BssHII
5758 TspEI
5727 NlaIII 5739 HpyCH4V 5753 HpyCH4V
5726 NspI 5737 NlaIII 5750 TspEI
5701 atcaagtagcgcgcactctatggagacatgcgagttcatgcaactaaaaaattgcataatttgttttattggtcacattttatgcgacacgatgaagaaa 5800
5710 5720 5730 5740 5750 5760 5770 5780 5790 5800
thrUp(A52G) (tRNA promoter of thrU-tyrU-glyT-thrT o
5894 HpaII
5893 MwoI
5893 Cfr10I
5870 HhaI
5891 HhaI
5814 Hpy188III
5809 RsaI
5808 NlaIV
5808 KpnI 5818 SduI
5808 AccB1I
5808 Acc65I
5838 DpnI
5805 MaeI 5818 BanII 5832 SduI
5804 StyI 5814 BspHI 5832 BanII
5804 AvrII 5815 NlaIII 5830 Cac8I
5869 SfoI
5869 NlaIV
5869 NarI
5869 KasI
5869 HaeII
5869 BbeI
5869 AcyI
5869 AccB1I 5882 MwoI 5892 SgrAI
5861 MwoI 5872 Cac8I
5890 SfoI
5890 NlaIV
5890 NarI
5890 KasI
5890 HaeII
5890 BbeI
5890 AcyI
5890 AccB1I
5801 cagcctaggtacctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtga 5900
5810 5820 5830 5840 5850 5860 5870 5880 5890 5900
5905 HaeIII
5904 CfrI
5903 HpaII
5902 NgoMIV
5902 NaeI
5902 Cfr10I
5902 Cac8I
5936 RsaI
5935 KpnI
5935 AccB1I
5928 XhoII
5928 BamHI 5940 HpaII
5918 HpaII 5929 DpnI 5939 BsiYI
5966 AccII
5965 HhaI
5984 MseI
5983 DraI
5997 Hpy8I
5997 AccI
5901 tgccggccacgatgcgtccggcgtagaggatcctggtacccggagcgggacttgaacccgcacagcgcgaacgccgagggatttttaaatccccttgtgtct 6000
5910 5920 5930 5940 5950 5960 5970 5980 5990 6000
tRNALeuZ
6045 HpaII
6021 Hpy188III
6044 Hpy188III
6017 SduI
6017 BanII
6015 HpaII
6039 AccII
6044 BsaWI
6044 AccIII 6056 HpaII
6055 BsaWI
6065 TspGWI
6097 MseI
6096 Hpy8I
6096 HpaI
6008 BsiYI 6020 AvaI
6038 HhaI
6051 TaqI 6061 MaeI 6072 HpyCH4V
6096 HincII
6001 accgattccaccatccgggctcgggaagaaagtggaggcgcgttccggagtcgaaccggactagacggatttgcaatccgctacataaccgctttgttaa 6100
6010 6020 6030 6040 6050 6060 6070 6080 6090 6100
tRNALeuZ
tRNAcysT
6108 TspEI
6107 ApoI
6102 HhaI
6117 HaeIII
6101 AccII
6116 StuI
6141 DrdI
6141 AcyI
6155 MseI
6154 DraI
6186 AccII 6197 StyI
6185 NruI 6197 EcoNI
6185 Hpy188III
6179 TaqI
6192 BsiYI
6101 cgcgccaaattcttcaggcctttcagccagacatccgcttgacgccgatgtcttttaaactggagcgggaaacgagactcgaactcgcgaccccgacctt 6200
6110 6120 6130 6140 6150 6160 6170 6180 6190 6200
tRNAcysT
tRNAglyW
6210 SduI
6210 BsiHKAI
6226 MwoI
6225 AluI
6254 MseI
6281 MwoI
6281 BstAPI 6293 BceAI
6298 TaqI
6201 ggcaaggtcgtgctctaccaactgagctattcccgcattcatcaagcaatcagttaatcacttgattttattatcgtctggcaatcagtgccgccgttcg 6300
6210 6220 6230 6240 6250 6260 6270 6280 6290 6300
glyWp Promoter
tRNAglyW

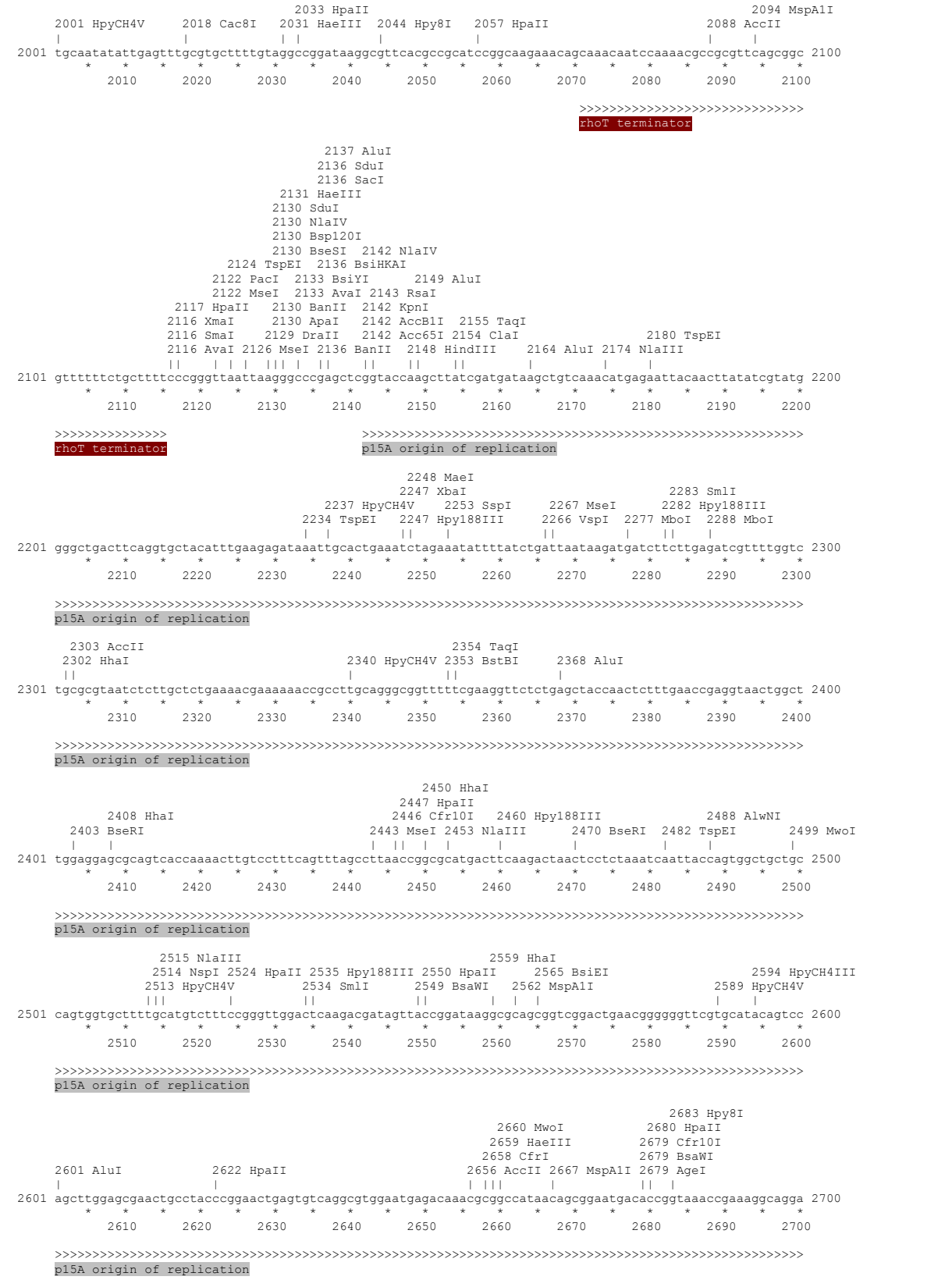

6333 Hpy8I
6333 HincII
6325 AccII
6307 HpyCH4V
6324 HhaI
6361 NlaIII
6360 SphI
6360 NspI
6360 Cac8I
6301 atgcgttgcattctacttacctggcgcgatgagtcaacgatatttttcaccacttttgagcatgcgaggaaaatagacgacatttttggtgataatgtcc 6400
6310 6320 6330 6340 6350 6360 6370 6380 6390 6400
glyWp Promoter
6473 HpyCH4V
6472 PstI
6472 BfmI
6471 SbfI
6467 HhaI
6465 MwoI
6465 HhaI
6465 Cac8I
6465 BssHII
6464 AscI
6481 TaqI
6420 TspEI
6455 TspEI 6466 AccII
6480 BstBI 6491 HaeIII
6401 caaatatgtcccactctgaaattatggaggatataaagaaggcgtaactgattgaattgtaatggcgcgccctgcaggattcgaacctgcggcccacgac 6500
6410 6420 6430 6440 6450 6460 6470 6480 6490 6500
argU (tRNA(UCU)Arg EcoGene#EG30014)
6537 AccII
6536 HhaI
6529 MwoI
6528 AluI
6547 BsiYI
6579 Hpy8I
6579 HincII
6570 TspEI
6564 AccII
6576 HpyCH4III
6501 ttagaaggtcgttgctctatccaactgagctaagggcgcgttgataccgcaatgcggtgtaatcgcgtgaattatacggtcaacccttgctgagtcaatg 6600
6510 6520 6530 6540 6550 6560 6570 6580 6590 6600
argUp (tRNA promoter of argU operon)
argU (tRNA(UCU)Arg EcoGene#EG30014)
6634 AccII
6631 Tth111I
6630 BsiEI
6689 AccII
6688 HhaI
6607 DpnI
6625 Hpy8I
6660 BfmI
6659 MwoI
6683 MspA1I
6699 Hpy188III
6601 gcttttgatctggttgctgaacaagtgaacgaccgcgtctgattttctgatttatttcgctatagcggcaaacaaacgcacaccgctgcgcgtctgaatc 6700
6610 6620 6630 6640 6650 6660 6670 6680 6690 6700
argUp (tRNA promoter of argU operon)
6743 AluI
6742 HindIII
6735 HpyCH4V
6734 PstI
6786 NlaIV
6786 HaeIII
6780 TaqI
6769 PsiI 6779 ClaI
6718 NlaIII
6734 BfmI 6744 MwoI
6758 TspGWI
6776 BstXI
6794 AluI
6701 aagaaaacccgtattttcatgtatcaaagtgacctgcagccaagcttggattgcgacacggagttactttataatccaatcgattggccccttagctcag 6800
6710 6720 6730 6740 6750 6760 6770 6780 6790 6800
ileX (tRNA(CAU)Ile
EcoGene#EG30046)
6865 MwoI
6865 BglI
6864 HaeIII
6863 EagI
6863 CfrI
6862 NotI
6860 MspA1I
6854 HaeIII
6809 Cac8I
6852 NlaIV 6863 BsiEI
6801 tggttagagcaggcgactcataatcgcttggtcgctggttcaagtccagcaggggccaccagcggccgcaaaggc 6875
6810 6820 6830 6840 6850 6860 6870
ileX (tRNA(CAU)Ile EcoGene#EG30046)

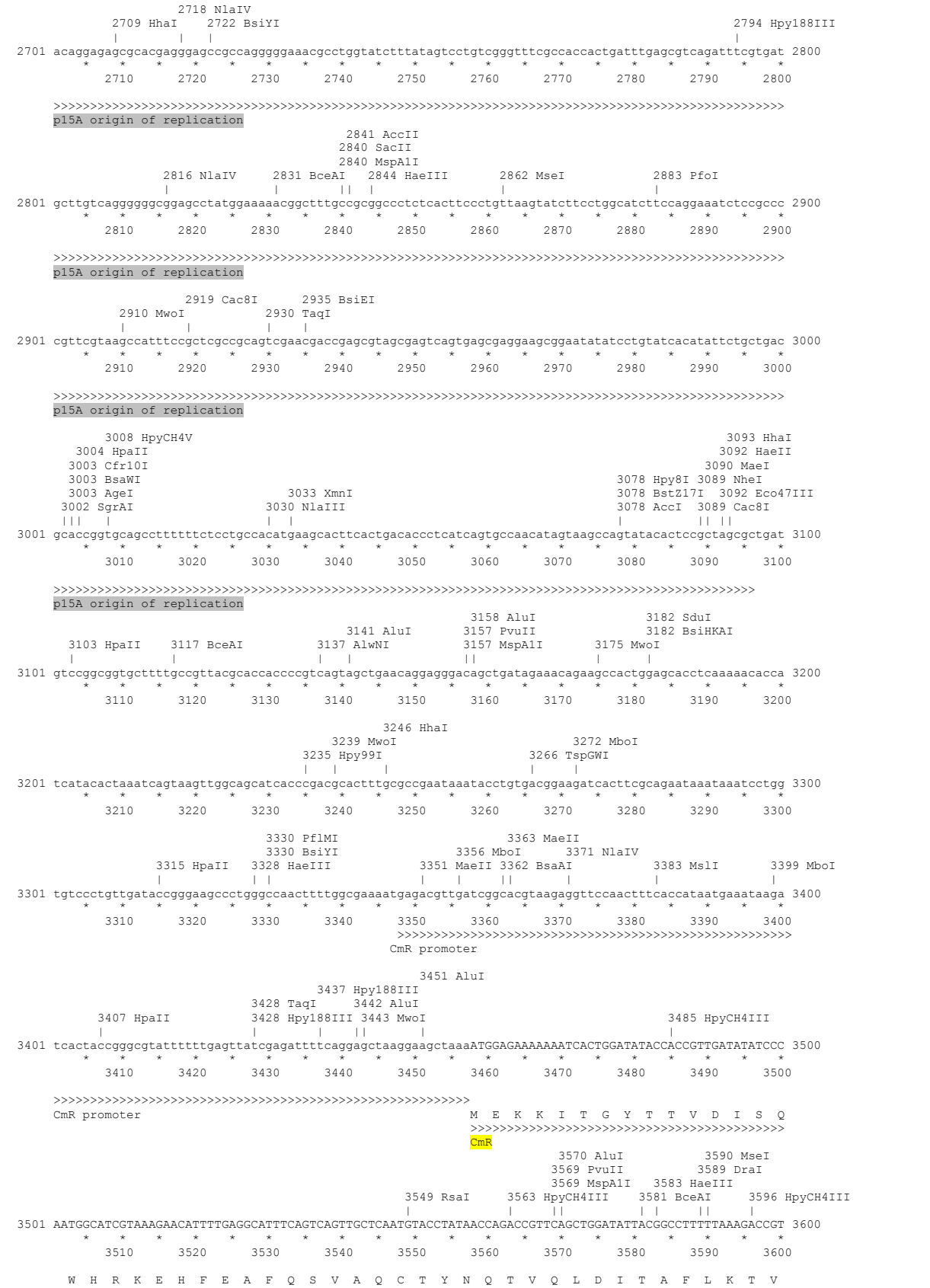
9.8 pTEc2 sequence

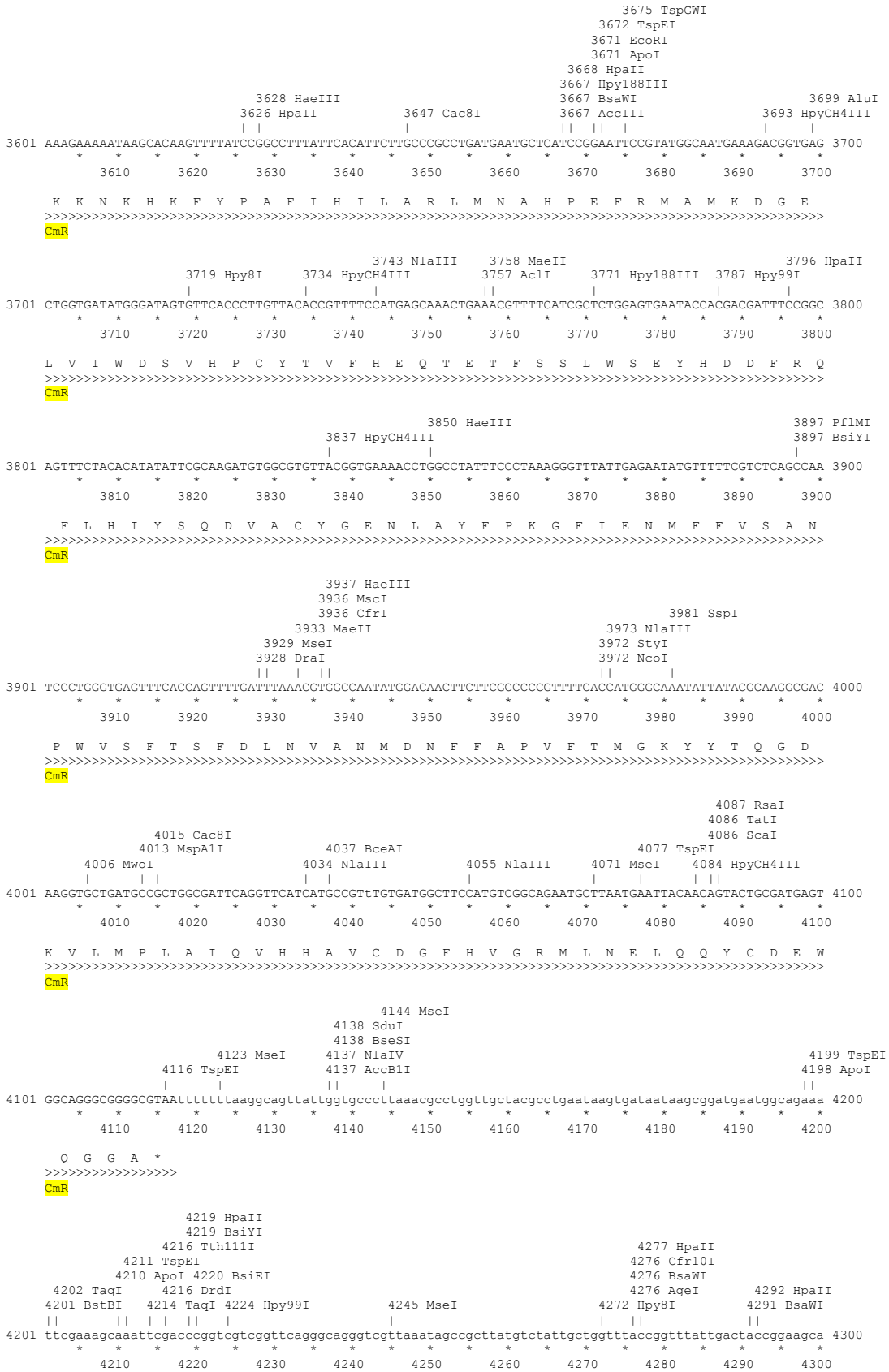


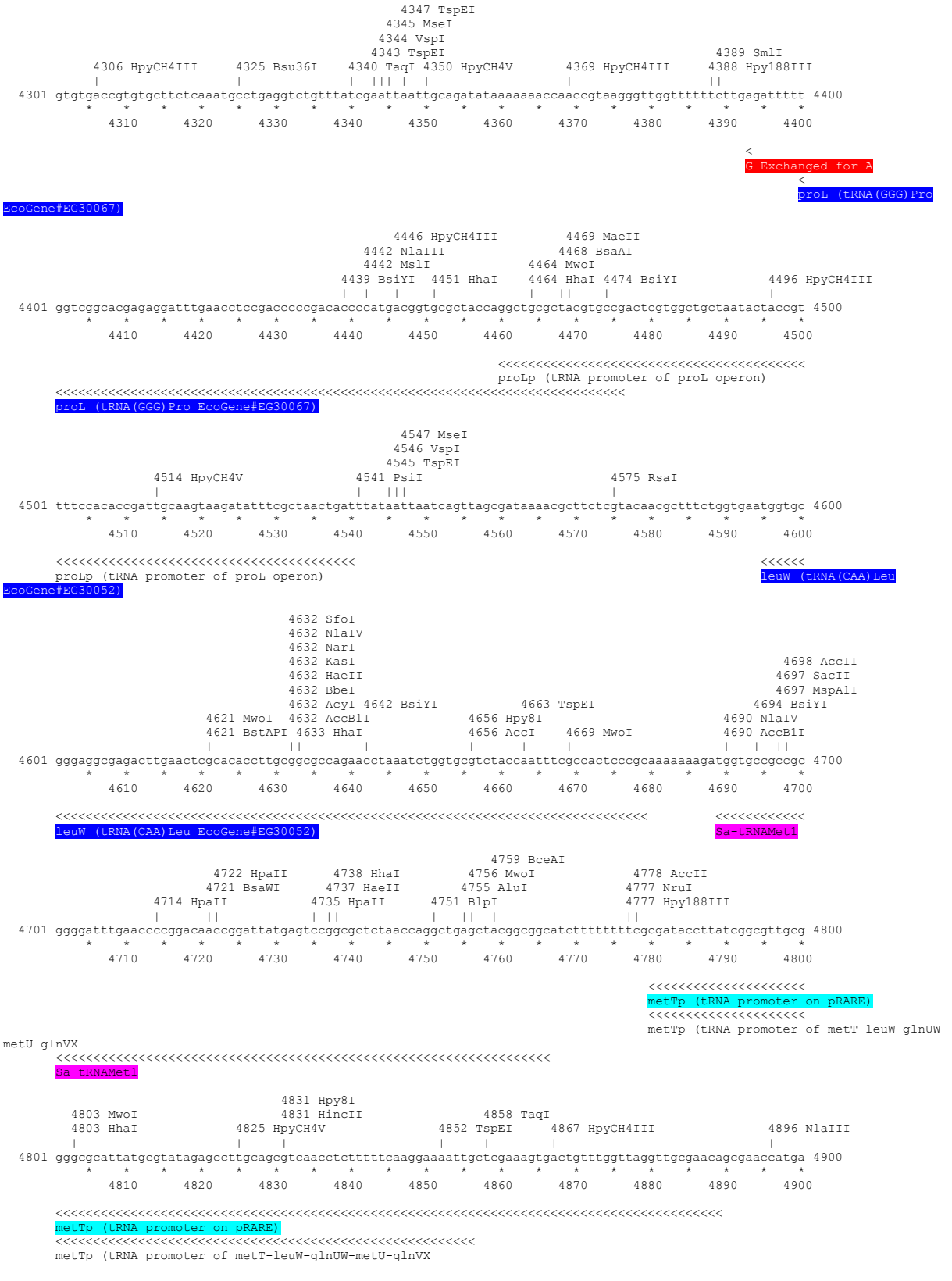

683 MwoI
679 AccII
678 HhaI
677 AccII
631 MboI
655 BsiYI
676 HhaI 685 MboI
628 HpaII
651 SduI
676 Cac8I
618 HpyCH4III
641 MseI
651 BseSI
676 BssHII
698 HpyCH4III
601 TACTGCGAAAAAGATAAACTGTTTCTGCCGGATCGTTTCATTAAAGGCGTGTGCCCGTATTGTGGTTTTGAAGATGCGCGCGGCGATCAGTGCGATAACT 700
610 620 630 640 650 660 670 680 690 700
Y C E K D K L F L P D R F I K G V C P Y C G F E D A R G D Q C D N C
Rare codon-optimized SaMetRS
736 MwoI 745 HhaI
731 MseI 744 AccII
730 Hpy8I 743 HhaI 763 HpaII
719 AccII 730 HpaI 743 Cac8I 762 Cfr10I 796 MboI
711 MwoI 720 HhaI 730 HincII 743 BssHII 761 SgrAI 794 TaqI
701 GTGGTCGTCTGCTGACCCCGCGCAGCCTGGTTAACGCAAAATGCGCGCTGTGTGGCAATCCGCCGGTGTTCAAAGTTACGAAACATTGGTTTTTCGATCT 800
710 720 730 740 750 760 770 780 790 800
G R L L T P R S L V N A K C A L C G N P P V F K V T K H W F F D L
Rare codon-optimized SaMetRS
839 MaeI
822 Hpy188III 836 MwoI
806 TspEI 819 TspEI 835 AluI 853 HpaII 896 Cac8I
805 ApoI 818 ApoI 831 MboI 849 NlaIII 875 BsiYI 894 StuI 895 HaeIII
801 GTCTGAATTTGGTGATAAAATTCGTGATTGGATCAGCTCTAGTAGCACCATGCCGGATAATGTGAAAAGCGTTGCCCTGTCTTGGGTGAAAGAAGGCCTG 900
810 820 830 840 850 860 870 880 890 900
S E F G D K I R D W I S S S S T M P D N V K S V A L S W V K E G L
Rare codon-optimized SaMetRS
946 BsiYI
926 NspI 945 HpyCH4III
926 BspLU11I 943 MwoI
926 AflIII 943 BstAPI
919 AccII 942 NlaIV 986 TaqI
906 HhaI 918 HhaI 942 AccB1I 985 BstBI
905 AccII 917 AccII 927 NlaIII 940 HpaII 957 HpyCH4V 977 MaeII
901 CGTCCGCGCAGTATTACGCGCGATAACATGTGGGGTATCCCGGCACCGTTTGCGGGTGCAGAAAACAAAACCATCTACGTTTGGTTCGAAGCACTGCTGG 1000
910 920 930 940 950 960 970 980 990 1000
R P R S I T R D N M W G I P A P F A G A E N K T I Y V W F E A L L G
Rare codon-optimized SaMetRS
1071 RsaI
1028 MseI 1070 TatI
1015 HpyCH4III 1061 TspEI 1089 RsaI
1013 Hpy99I 1027 DraI 1060 ApoI 1070 BsrGI 1087 MaeII
1001 GTTATCTGTCTGCGACGGTGGAATACTTTAAAAACCTGGGCAAAGAAGAAATGTGGAAAGAATTTTGGCTGTACAACGATACCAAAACGTACTACTTCAT 1100
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
Y L S A T V E Y F K N L G K E E M W K E F W L Y N D T K T Y Y F I
Rare codon-optimized SaMetRS
1143 NlaIII
1143 MslI
1142 BstXI
1140 HaeIII
1139 CfrI
1138 HpaII
1137 NgoMIV
1137 NaeI 1191 HaeIII
1137 MwoI 1190 CfrI
1137 Cfr10I 1188 AccII
1125 AccII 1137 Cac8I 1185 MboI
1101 CGGCAAAGATAACATCCCGTTCCACGCGGTTATTCTGCCGGCCATGCTGATGGCATCTAACGAAAAATATAATCTGCCGAGTGTGATCGCGGCCACCGAA 1200
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
G K D N I P F H A V I L P A M L M A S N E K Y N L P S V I A A T E
Rare codon-optimized SaMetRS
1256 TaqI
1223 TspEI 1242 TspEI 1255 ClaI 1293 HhaI
1209 RsaI 1222 ApoI 1235 BceAI 1254 MboI 1292 HaeII
1201 TATCTGCTGTACGAGGGTCAGAAATTTAGTAAAAGCCGTAAAATTGGCGTTTGGATCGATGAAGCCGATAAACTGATGGATGTGGAATACTGGCGCTTTA 1300
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
Y L L Y E G Q K F S K S R K I G V W I D E A D K L M D V E Y W R F I
Rare codon-optimized SaMetRS

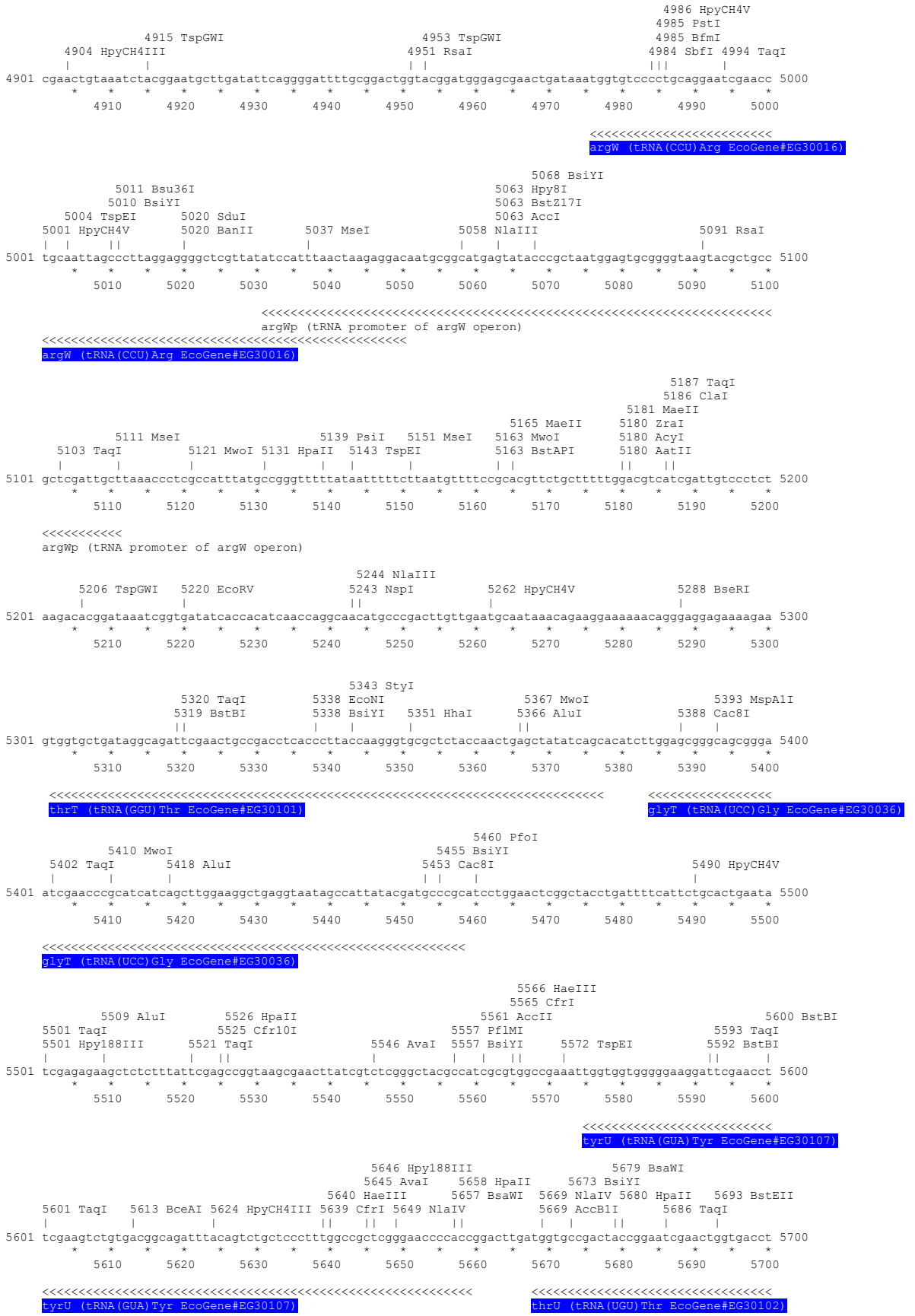

1367 MseI
1366 Hpy8I
1366 HpaI
1366 HincII
1314 HhaI
1342 MaeII
1307 TspGWI
1341 PmlI
1359 HhaI
1305 MboI
1318 HpaII
1341 BsaAI
1358 FspI
1387 EcoRV
1301 TTCTGATCCGTCTGCGCCCGGAAGAAAAAGATACCAACTTCACGTGGCGTGAAGCACTGCGCATTGTTAACACCGAACTGAATGATGATATCGGTAACTA 1400
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
L I R L R P E E K D T N F T W R E A L R I V N T E L N D D I G N Y
Rare codon-optimized SaMetRS
1494 MboI
1424 MseI
1446 NlaIV
1467 MboI
1493 BclI
1401 TGCGAATCGTGTGCTGTCTATGGTTAAACGCTATTACGATGGCGTGGTTCCGAGTCCGAAAGAAGCGATCTTCAACGATGAAGATAAAAACCTGATCACG 1500
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
A N R V L S M V K R Y Y D G V V P S P K E A I F N D E D K N L I T
Rare codon-optimized SaMetRS
1538 TaqI
1559 HpaII
1537 BstBI
1558 Cfr10I
1587 MaeII
1503 MboI
1529 Hpy8I
1553 MseI
1575 BsiYI 1586 BsaAI
1502 BclI
1521 MaeII 1533 HpyCH4III
1551 TspEI
1574 Hpy188III
1501 CTGATCAAAGAAAGCCCGAAACGTATGGGTGAACTGTTCGAACTGGGCAAAATTAAAGCCGGTAGTGAAGAAATCCTGAAACTGGCACGTAGCGGCAATC 1600
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
L I K E S P K R M G E L F E L G K I K A G S E E I L K L A R S G N L
Rare codon-optimized SaMetRS
1625 NlaIV
1620 BceAI
1618 HhaI
1617 AccII
1616 HhaI
1616 Cac8I
1685 Cac8I
1616 BssHII
1678 Hpy8I
1615 AccII
1676 MwoI
1691 RsaI
1601 TGTATCTGAACAATCGCGCGCCGTGGAGCCTGGTGAAAACCAACAAAGAAGAAGCCAACAATGTTCTGTATATTAGCGTGAACAGCCTGCGTACCCTGGC 1700
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
Y L N N R A P W S L V K T N K E E A N N V L Y I S V N S L R T L A
Rare codon-optimized SaMetRS
Non-sense mutation T/C
1751 AluI
1750 PvuII
1750 MspA1I
1793 HpaII
1747 AlwNI
1792 NgoMIV
1719 NlaIII 1733 MaeI
1746 PflMI
1792 NaeI
1716 MboI 1726 MaeII
1746 BsiYI
1792 Cfr10I
1710 RsaI
1724 Hpy99I
1743 RsaI
1767 Hpy188III
1792 Cac8I
1701 GATTATGCTGTACCCGATCATGCCGACGTATTCTAGTAACCTGTACCAGCAGCTGGGTCTGAGTAATCTGGAAAGCGAAACCTGGGATAGTGCCGGCAGC 1800
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
I M L Y P I M P T Y S S N L Y Q Q L G L S N L E S E T W D S A G S
Rare codon-optimized SaMetRS
1818 BstXI
1867 HpaII
1816 HaeIII
1866 Hpy188III
1813 HpaII
1854 TspEI
1866 BsaWI
1900 Hpy99I
1809 NlaIII
1824 TspEI
1847 MseI
1866 AccIII
1898 TspGWI
1801 CTGAAAATCATGCCGGGCCATAAAATTGGTGAAATCCGCAGCCTGTTTAAGAAAATTGAAATGTCTCCGGAAGAACTGATGAAAAAACTGGATGAAATCC 1900
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
L K I M P G H K I G E I R S L F K K I E M S P E E L M K K L D E I R
Rare codon-optimized SaMetRS
Non-sense mutation T/C
1927 MboI
1926 XhoII
1924 HpaII
1943 HpyCH4V
1923 Hpy188III
1942 PstI
1903 AccII
1923 BsaWI
1942 BfmI
1986 HaeIII
1902 NruI
1923 AccIII 1935 MwoI
1964 MwoI
1978 MwoI
1992 NlaIII
1902 Hpy188III
1920 MaeII
1935 HhaI
1960 TspEI 1971 HhaI
1985 StuI 1995 MboI
1901 GTCGCGAAGTGGAAAAAGAACGTCCGGATCTGCTGCGCTAActgcagtttcaaacgctaaattgcctgatgcgctacgcttatcaggcctacatgatctc 2000
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
R E V E K E R P D L L R *
Rare codon-optimized SaMetRS

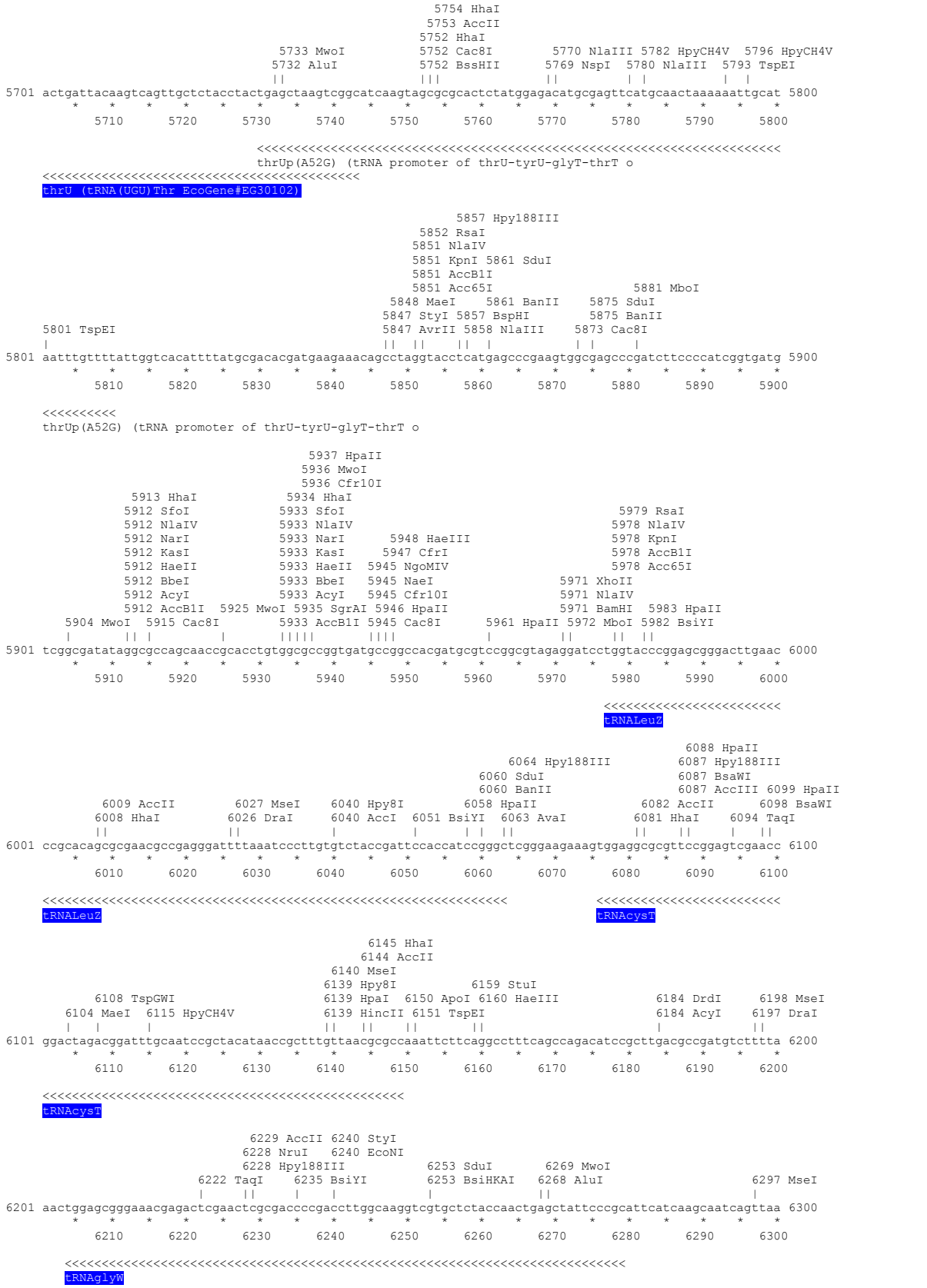

2033 HpaII
2094 MspA1I
2001 HpyCH4V
2018 Cac8I
2031 HaeIII
2044 Hpy8I
2057 HpaII
2088 AccII
2001 tgcaatatattgagtttgcgtgcttttgtaggccggataaggcgttcacgccgcatccggcaagaaacagcaaacaatccaaaacgccgcgttcagcggc 2100
2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
rhoT terminator
2137 AluI
2136 SduI
2136 SacI
2131 HaeIII
2130 SduI
2130 NlaIV
2130 Bsp120I
2130 BseSI 2142 NlaIV
2124 TspEI 2136 BsiHKAI
2122 PacI 2133 BsiYI 2149 AluI
2122 MseI 2133 AvaI 2143 RsaI
2117 HpaII 2130 BanII 2142 KpnI
2116 XmaI 2130 ApaI 2142 AccB1I 2155 TaqI
2116 SmaI 2129 DraII 2142 Acc65I 2154 ClaI 2180 TspEI
2116 AvaI 2126 MseI 2136 BanII 2148 HindIII 2164 AluI 2174 NlaIII
2101 gttttttctgcttttcccgggttaattaagggcccgagctcggtaccaagcttatcgatgataagctgtcaaacatgagaattacaacttatatcgtatg 2200
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
rhoT terminator
p15A origin of replication
2248 MaeI
2247 XbaI
2283 SmlI
2237 HpyCH4V 2253 SspI 2267 MseI 2282 Hpy188III
2234 TspEI 2247 Hpy188III 2266 VspI 2277 MboI 2288 MboI
2201 gggctgacttcaggtgctacatttgaagagataaattgcactgaaatctagaaatattttatctgattaataagatgatcttcttgagatcgttttggtc 2300
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
p15A origin of replication
2303 AccII
2302 HhaI
2354 TaqI
2340 HpyCH4V 2353 BstBI 2368 AluI
2301 tgcgcgtaatctcttgctctgaaaacgaaaaaaccgccttgcagggcggtttttcgaaggttctctgagctaccaactctttgaaccgaggtaactggct 2400
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
p15A origin of replication
2450 HhaI
2447 HpaII
2408 HhaI
2446 Cfr10I 2460 Hpy188III 2488 AlwNI
2403 BseRI
2443 MseI 2453 NlaIII 2470 BseRI 2482 TspEI 2499 MwoI
2401 tggaggagcgcagtcaccaaaacttgtcctttcagtttagccttaaccggcgcatgacttcaagactaactcctctaaatcaattaccagtggctgctgc 2500
2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
p15A origin of replication
2515 NlaIII
2559 HhaI
2514 NspI 2524 HpaII 2535 Hpy188III 2550 HpaII 2565 BsiEI 2594 HpyCH4III
2513 HpyCH4V 2534 SmlI 2549 BsaWI 2562 MspA1I 2589 HpyCH4V
2501 cagtggtgcttttgcatgtctttccgggttggactcaagacgatagttaccggataaggcgcagcggtcggactgaacgggggggttcgtgcatacagtcc 2600
2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
p15A origin of replication
2683 Hpy8I
2660 MwoI
2680 HpaII
2659 HaeIII
2679 Cfr10I
2658 CfrI
2679 BsaWI
2601 AluI
2622 HpaII
2656 AccII 2667 MspA1I 2679 AgeI
2601 agcttggagcgaactgcctacccggaactgagtgtcaggcgtggaatgagacaaacgcggccataacagcggaatgacaccggtaaaccgaaaggcagga 2700
2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
p15A origin of replication

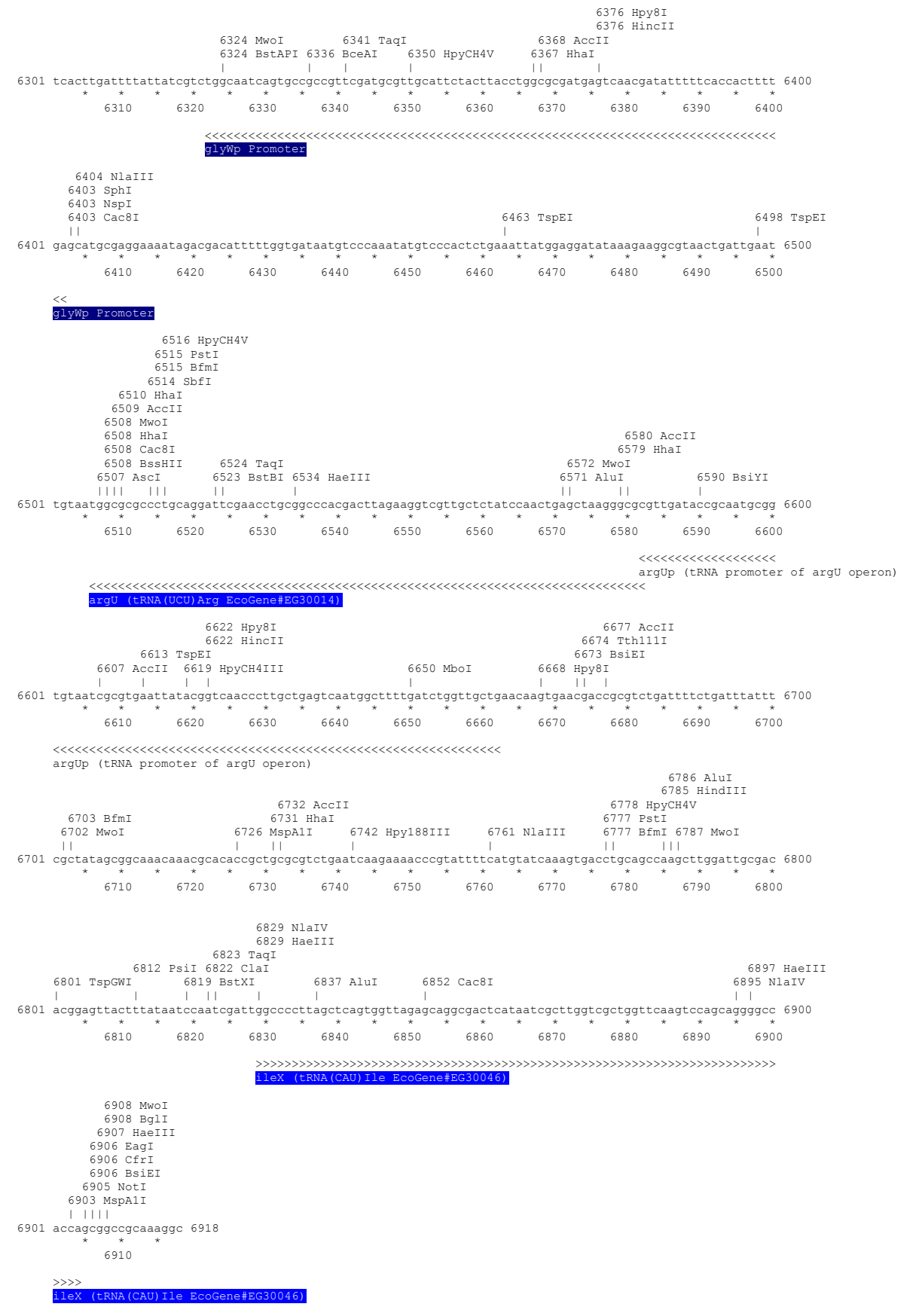

2718 NlaIV
2709 HhaI 2722 BsiYI 2794 Hpy188III
2701 acaggagagcgcacgagggagccgccaggggggaaacgcctggtatctttatagtcctgtcgggtttcgccaccactgatttgagcgtcagatttcgtgat 2800
2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
p15A origin of replication
2841 AccII
2840 SacII
2840 MspA1I
2816 NlaIV 2831 BceAI 2844 HaeIII 2862 MseI 2883 PfoI
2801 gcttgtcaggggggcggagcctatggaaaaacggctttgccgcggccctctcacttccctgttaagtatcttcctggcatcttccaggaaatctccgccc 2900
2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
p15A origin of replication
2919 Cac8I 2935 BsiEI
2910 MwoI 2930 TaqI
2901 cgttcgtaagccatttccgctcgccgcagtcgaacgaccgagcgtagcgagtcagtgagcgaggaagcggaatatatcctgtatcacatattctgctgac 3000
2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
p15A origin of replication
3008 HpyCH4V 3093 HhaI
3004 HpaII 3092 HaeII
3003 Cfr10I 3090 MaeI
3003 BsaWI 3078 Hpy8I 3089 NheI
3003 AgeI 3033 XmnI 3078 BstZ17I 3092 Eco47III
3002 SgrAI 3030 NlaIII 3078 AccI 3089 Cac8I
3001 gcaccggtgcagccttttttctcctgccacatgaagcacttcactgacaccctcatcagtgccaacatagtaagccagtatacactccgctagcgctgat 3100
3010 3020 3030 3040 3050 3060 3070 3080 3090 3100
p15A origin of replication
3158 AluI 3182 SduI
3141 AluI 3157 PvuII 3182 BsiHKAI
3103 HpaII 3117 BceAI 3137 AlwNI 3157 MspA1I 3175 MwoI
3101 gtccggcggtgcttttgccgttacgcaccaccccgtcagtagctgaacaggagggacagctgatagaaacagaagccactggagcacctcaaaaacacca 3200
3110 3120 3130 3140 3150 3160 3170 3180 3190 3200
3246 HhaI
3239 MwoI 3272 MboI
3235 Hpy99I 3266 TspGWI
3201 tcatacactaaatcagtaagttggcagcatcacccgacgcactttgcgccgaataaatacctgtgacggaagatcacttcgcagaataaataaatcctgg 3300
3210 3220 3230 3240 3250 3260 3270 3280 3290 3300
3330 PflMI 3363 MaeII
3330 BsiYI 3356 MboI 3371 NlaIV
3315 HpaII 3328 HaeIII 3351 MaeII 3362 BsaAI 3383 MslI 3399 MboI
3301 tgtccctgttgataccgggaagccctgggccaacttttggcgaaaatgagacgttgatcggcacgtaagaggttccaactttcaccataatgaaataaga 3400
3310 3320 3330 3340 3350 3360 3370 3380 3390 3400
CmR promoter
3451 AluI
3437 Hpy188III
3428 TaqI 3442 AluI
3407 HpaII 3428 Hpy188III 3443 MwoI 3485 HpyCH4III
3401 tcactaccgggcgtatttttttgagttatcgagattttcaggagctaaggaagctaaaATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCC 3500
3410 3420 3430 3440 3450 3460 3470 3480 3490 3500
CmR promoter
M E K K I T G Y T T V D I S Q
CmR
3570 AluI 3590 MseI
3569 PvuII 3589 DraI
3569 MspA1I 3583 HaeIII
3549 RsaI 3563 HpyCH4III 3581 BceAI 3596 HpyCH4III
3501 AATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGT 3600
3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
W H R K E H F E A F Q S V A Q C T Y N Q T V Q L D I T A F L K T V
CmR


3675 TspGWI
3672 TspEI
3671 EcoRI
3671 ApoI
3668 HpaII
3667 Hpy188III
3628 HaeIII
3626 HpaII
3647 Cac8I
3667 BsaWI
3667 AccIII
3699 AluI
3693 HpyCH4III
3601 AAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGGTGAG 3700
3610 3620 3630 3640 3650 3660 3670 3680 3690 3700
K K N K H K F Y P A F I H I L A R L M N A H P E F R M A M K D G E
CmR
3743 NlaIII
3758 MaeII
3796 HpaII
3719 Hpy8I
3734 HpyCH4III
3757 AclI
3771 Hpy188III
3787 Hpy99I
3701 CTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGC 3800
3710 3720 3730 3740 3750 3760 3770 3780 3790 3800
L V I W D S V H P C Y T V F H E Q T E T F S S L W S E Y H D D F R Q
CmR
3850 HaeIII
3897 PflMI
3897 BsiYI
3837 HpyCH4III
3801 AGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAA 3900
3810 3820 3830 3840 3850 3860 3870 3880 3890 3900
F L H I Y S Q D V A C Y G E N L A Y F P K G F I E N M F F V S A N
CmR
3937 HaeIII
3936 MscI
3936 CfrI
3981 SspI
3933 MaeII
3973 NlaIII
3929 MseI
3928 DraI
3972 StyI
3972 NcoI
3901 TCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGAC 4000
3910 3920 3930 3940 3950 3960 3970 3980 3990 4000
P W V S F T S F D L N V A N M D N F F A P V F T M G K Y Y T Q G D
CmR
4087 RsaI
4086 TatI
4086 ScaI
4015 Cac8I
4013 MspA1I
4037 BceAI
4077 TspEI
4006 MwoI
4034 NlaIII
4055 NlaIII
4071 MseI
4084 HpyCH4III
4001 AAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTtTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGT 4100
4010 4020 4030 4040 4050 4060 4070 4080 4090 4100
K V L M P L A I Q V H H A V C D G F H V G R M L N E L Q Q Y C D E W
CmR
4144 MseI
4138 SduI
4138 BseSI
4123 MseI
4137 NlaIV
4199 TspEI
4116 TspEI
4137 AccB1I
4198 ApoI
4101 GGCAGGGCGGGGCGTAAttttttaaggcagttattggtgcccttaaacgcctggttgctacgcctgaataagtgataataagcggatgaatggcagaaa 4200
4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
Q G G A *
CmR
4219 HpaII
4219 BsiYI
4216 Tth111I
4277 HpaII
4211 TspEI
4276 Cfr10I
4210 ApoI
4220 BsiEI
4276 BsaWI
4202 TaqI
4216 DrdI
4276 AgeI
4292 HpaII
4201 BstBI
4214 TaqI
4224 Hpy99I
4245 MseI
4272 Hpy8I
4291 BsaWI
4201 ttcgaaagcaaattcgacccggtcgtcggttcagggcagggtcgttaaatagccgcttatgtctattgctggtttaccggtttattgactaccggaagca 4300
4210 4220 4230 4240 4250 4260 4270 4280 4290 4300


4347 TspEI
4345 MseI
4344 VspI
4343 TspEI
4389 SmlI
4306 HpyCH4III
4325 Bsu36I
4340 TaqI
4350 HpyCH4V
4369 HpyCH4III
4388 Hpy188III
4301 gtgtgaccgtgtgcttctcaaatgcctgaggtctgtttatcgaattaattgcagatataaaaaaaccaaccgtaagggttggttttttcttgagattttt 4400
4310 4320 4330 4340 4350 4360 4370 4380 4390 4400
G Exchanged for A
proL (tRNA(GGG)Pro EcoGene#EG30067)
4446 HpyCH4III
4469 MaeII
4442 NlaIII
4468 BsaAI
4442 MslI
4464 MwoI
4439 BsiYI
4451 HhaI
4464 HhaI
4474 BsiYI
4496 HpyCH4III
4401 ggtcggcacgagaggatttgaacctccgacccccgacaccccatgacggtgcgctaccaggctgcgctacgtgccgactcgtggctgctaatactaccgt 4500
4410 4420 4430 4440 4450 4460 4470 4480 4490 4500
proLp (tRNA promoter of proL operon)
proL (tRNA(GGG)Pro EcoGene#EG30067)
4547 MseI
4546 VspI
4545 TspEI
4514 HpyCH4V
4541 PsiI
4575 RsaI
4501 tttccacaccgattgcaagtaagatatttcgctaactgatttataattaatcagttagcgataaaacgcttctcgtacaacgctttctggtgaatggtgc 4600
4510 4520 4530 4540 4550 4560 4570 4580 4590 4600
proLp (tRNA promoter of proL operon)
leuW (tRNA(CAA)Leu EcoGene#EG30052)
4632 SfoI
4632 NlaIV
4632 NarI
4632 KasI
4632 HaeII
4632 BbeI
4632 AcyI 4642 BsiYI
4663 TspEI
4698 AccII
4697 SacII
4697 MspA1I
4694 BsiYI
4621 MwoI
4632 AccB1I
4656 Hpy8I
4690 NlaIV
4621 BstAPI 4633 HhaI
4656 AccI
4669 MwoI
4690 AccB1I
4601 gggaggcgagacttgaactcgcacaccttgcggcgccagaacctaaatctggtgcgtctaccaatttcgccactcccgcaaaaaaagatggtgccgccgc 4700
4610 4620 4630 4640 4650 4660 4670 4680 4690 4700
leuW (tRNA(CAA)Leu EcoGene#EG30052)
Sa-tRNAMet1
4759 BceAI
4722 HpaII
4738 HhaI
4756 MwoI
4778 AccII
4721 BsaWI
4737 HaeII
4755 AluI
4777 NruI
4714 HpaII
4735 HpaII
4751 BlpI
4777 Hpy188III
4701 ggggatttgaaccccggacaaccggattatgagtccggcgctctaaccaggctgagctacggcggcatctttttttttcgcgataccttatcggcgttgcg 4800
4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
metTp (tRNA promoter on pRARE)
metTp (tRNA promoter of metT-leuW-glnUW-metU-glnVX
Sa-tRNAMet1
4831 Hpy8I
4803 MwoI
4831 HincII
4858 TaqI
4803 HhaI
4825 HpyCH4V
4852 TspEI
4867 HpyCH4III
4896 NlaIII
4801 gggcgcattatgcgtatagagccttgcagcgtcaacctctttttcaaggaaaattgctcgaaagtgactgtttggttaggttgcgaacagcgaaccatga 4900
4810 4820 4830 4840 4850 4860 4870 4880 4890 4900
metTp (tRNA promoter on pRARE)
metTp (tRNA promoter of metT-leuW-glnUW-metU-glnVX


4986 HpyCH4V
4985 PstI
4915 TspGWI
4953 TspGWI
4985 BfmI
4904 HpyCH4III
4951 RsaI
4984 SbfI 4994 TaqI
4901 cgaactgtaaatctacggaatgcttgatattcaggggatttgcggactggtacggatgggagcgaactgataaatggtgtcccctgcaggaatcgaacc 5000
4910 4920 4930 4940 4950 4960 4970 4980 4990 5000
argW (tRNA(CCU)Arg EcoGene#EG30016)
5068 BsiYI
5011 Bsu36I
5063 Hpy8I
5010 BsiYI
5063 BstZ17I
5004 TspEI
5020 SduI
5063 AccI
5001 HpyCH4V
5020 BanII
5037 MseI
5058 NlaIII
5091 RsaI
5001 tgcaattagcccttaggaggggctcgttatatccatttaactaagaggacaatgcggcatgagtatacccgctaatggagtgcggggtaagtacgctgcc 5100
5010 5020 5030 5040 5050 5060 5070 5080 5090 5100
argWp (tRNA promoter of argW operon)
argW (tRNA(CCU)Arg EcoGene#EG30016)
5187 TaqI
5186 ClaI
5181 MaeII
5165 MaeII
5180 ZraI
5111 MseI
5139 PsiI
5151 MseI
5163 MwoI
5180 AcyI
5103 TaqI
5121 MwoI 5131 HpaII
5143 TspEI
5163 BstAPI
5180 AatII
5101 gctcgattgcttaaaccctcgccatttatgccgggtttttataatttttcttaatgttttccgcacgttctgctttttggacgtcatcgattgtccctct 5200
5110 5120 5130 5140 5150 5160 5170 5180 5190 5200
argWp (tRNA promoter of argW operon)
5244 NlaIII
5206 TspGWI
5220 EcoRV
5243 NspI
5262 HpyCH4V
5288 BseRI
5201 aagacacggataaatcggtgatatcaccacatcaaccaggcaacatgcccgacttgttgaatgcaataaacagaaggaaaaaacagggaggagaaaagaa 5300
5210 5220 5230 5240 5250 5260 5270 5280 5290 5300
5343 StyI
5320 TaqI
5338 EcoNI
5367 MwoI
5393 MspA1I
5319 BstBI
5338 BsiYI
5351 HhaI
5366 AluI
5388 Cac8I
5301 gtggtgctgataggcagattcgaactgccgacctcacccttaccaagggtgcgctctaccaactgagctatatcagcacatcttggagcgggcagcggga 5400
5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
thrT (tRNA(GGU)Thr EcoGene#EG30101)
glyT (tRNA(UCC)Gly EcoGene#EG30036)
5460 PfoI
5410 MwoI
5455 BsiYI
5402 TaqI
5418 AluI
5453 Cac8I
5490 HpyCH4V
5401 atcgaacccgcatcatcagcttggaaggctgaggtaatagccattatacgatgcccgcatcctggaactcggctacctgattttcattctgcactgaata 5500
5410 5420 5430 5440 5450 5460 5470 5480 5490 5500
glyT (tRNA(UCC)Gly EcoGene#EG30036)
5566 HaeIII
5565 CfrI
5509 AluI
5526 HpaII
5561 AccII
5600 BstBI
5501 TaqI
5525 Cfr10I
5557 PflMI
5593 TaqI
5501 Hpy188III
5521 TaqI
5546 AvaI
5557 BsiYI
5572 TspEI
5592 BstBI
5501 tcgagagaagctctctttattcgagccggtaagcgaacttatcgtctcgggctacgccatcgcgtggccgaaattggtggtgggggaaggattcgaacct 5600
5510 5520 5530 5540 5550 5560 5570 5580 5590 5600
tyrU (tRNA(GUA)Tyr EcoGene#EG30107)
5646 Hpy188III
5679 BsaWI
5645 AvaI
5658 HpaII
5673 BsiYI
5640 HaeIII
5657 BsaWI
5669 NlaIV 5680 HpaII
5693 BstEII
5601 TaqI
5613 BceAI 5624 HpyCH4III 5639 CfrI 5649 NlaIV
5669 AccB1I
5686 TaqI
5601 tcgaagtctgtgacggcagatttacagtctgctccctttggccgctcgggaaccccaccggacttgatggtgccgactaccggaatcgaactggtgacct 5700
5610 5620 5630 5640 5650 5660 5670 5680 5690 5700
tyrU (tRNA(GUA)Tyr EcoGene#EG30107)
thrU (tRNA(UGU)Thr EcoGene#EG30102)


5754 HhaI
5753 AccII
5752 HhaI
5733 MwoI
5732 AluI
5752 Cac8I
5752 BssHII
5770 NlaIII 5782 HpyCH4V 5796 HpyCH4V
5769 NspI 5780 NlaIII 5793 TspEI
5701 actgattacaagtcagttgctctacctactgagctaagtcggcatcaagtagcgcgcactctatggagacatgcgagttcatgcaactaaaaaattgcat 5800
5710 5720 5730 5740 5750 5760 5770 5780 5790 5800
thrUp(A52G) (tRNA promoter of thrU-tyrU-glyT-thrT o
thrU (tRNA(UGU)Thr EcoGene#EG30102)
5857 Hpy188III
5852 RsaI
5851 NlaIV
5851 KpnI 5861 SduI
5851 AccB1I
5851 Acc65I
5881 MboI
5848 MaeI 5861 BanII 5875 SduI
5847 StyI 5857 BspHI 5875 BanII
5801 TspEI
5847 AvrII 5858 NlaIII 5873 Cac8I
5801 aatttgttttattggtcacattttatgcgacacgatgaagaaacagcctaggtacctcatgagcccgaagtggcgagcccgatcttccccatcggtgatg 5900
5810 5820 5830 5840 5850 5860 5870 5880 5890 5900
thrUp(A52G) (tRNA promoter of thrU-tyrU-glyT-thrT o
5937 HpaII
5936 MwoI
5936 Cfr10I
5913 HhaI
5934 HhaI
5912 SfoI
5933 SfoI
5979 RsaI
5912 NlaIV
5933 NlaIV
5978 NlaIV
5912 NarI
5933 NarI
5948 HaeIII
5978 KpnI
5912 KasI
5933 KasI
5947 CfrI
5978 AccB1I
5912 HaeII
5933 HaeII
5945 NgoMIV
5978 Acc65I
5912 BbeI
5933 BbeI
5945 NaeI
5971 XhoII
5912 AcyI
5933 AcyI
5945 Cfr10I
5971 NlaIV
5912 AccB1I
5925 MwoI 5935 SgrAI 5946 HpaII
5971 BamHI 5983 HpaII
5904 MwoI 5915 Cac8I
5933 AccB1I 5945 Cac8I
5961 HpaII 5972 MboI 5982 BsiYI
5901 tcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcctggtacccggagcgggacttgaac 6000
5910 5920 5930 5940 5950 5960 5970 5980 5990 6000
tRNALeuZ
6088 HpaII
6064 Hpy188III
6087 Hpy188III
6060 SduI
6087 BsaWI
6060 BanII
6087 AccIII 6099 HpaII
6009 AccII
6027 MseI
6040 Hpy8I
6058 HpaII
6082 AccII
6098 BsaWI
6008 HhaI
6026 DraI
6040 AccI
6051 BsiYI
6063 AvaI
6081 HhaI
6094 TaqI
6001 ccgcacagcgcgaacgccgagggatttaaatcccttgtgtctaccgattccaccatccgggctcgggaagaaagtggaggcgcgttccggagtcgaacc 6100
6010 6020 6030 6040 6050 6060 6070 6080 6090 6100
tRNALeuZ
tRNAcysT
6145 HhaI
6144 AccII
6140 MseI
6139 Hpy8I
6159 StuI
6108 TspGWI
6139 HpaI
6150 ApoI 6160 HaeIII
6184 DrdI
6198 MseI
6104 MaeI 6115 HpyCH4V
6139 HincII 6151 TspEI
6184 AcyI
6197 DraI
6101 ggactagacggatttgcaatccgctacataaccgctttgttaacgcgccaaattcttcaggcctttcagccagacatccgcttgacgccgatgtctttta 6200
6110 6120 6130 6140 6150 6160 6170 6180 6190 6200
tRNAcysT
6229 AccII 6240 StyI
6228 NruI 6240 EcoNI
6228 Hpy188III
6253 SduI
6269 MwoI
6222 TaqI 6235 BsiYI
6253 BsiHKAI
6268 AluI
6297 MseI
6201 aactggagcgggaaacgagactcgaactcgcgacccccgaccttggcaaggtcgtgctctaccaactgagctattcccgcattcatcaagcaatcagttaa 6300
6210 6220 6230 6240 6250 6260 6270 6280 6290 6300
tRNAglyW


6376 Hpy8I
6376 HincII
6324 MwoI
6341 TaqI
6368 AccII
6324 BstAPI 6336 BceAI
6350 HpyCH4V
6367 HhaI
6301 tcacttgattttattatcgtctggcaatcagtgccgccgttcgatgcgttgcattctacttacctggcgcgatgagtcaacgatatttttcaccactttt 6400
6310 6320 6330 6340 6350 6360 6370 6380 6390 6400
glyWp Promoter
6404 NlaIII
6403 SphI
6403 NspI
6403 Cac8I
6463 TspEI
6498 TspEI
6401 gagcatgcgaggaaaatagacgacatttttggtgataatgtcccaaatatgtcccactctgaaattatggaggatataaagaaggcgtaactgattgaat 6500
6410 6420 6430 6440 6450 6460 6470 6480 6490 6500
glyWp Promoter
6516 HpyCH4V
6515 PstI
6515 BfmI
6514 SbfI
6510 HhaI
6509 AccII
6508 MwoI
6508 HhaI
6508 Cac8I
6508 BssHII
6524 TaqI
6580 AccII
6579 HhaI
6572 MwoI
6507 AscI
6523 BstBI 6534 HaeIII
6571 AluI
6590 BsiYI
6501 tgtaatggcgcgccctgcaggattcgaacctgcggcccacgacttagaaggtcgttgctctatccaactgagctaagggcgcgttgataccgcaatgcgg 6600
6510 6520 6530 6540 6550 6560 6570 6580 6590 6600
argUp (tRNA promoter of argU operon)
argU (tRNA(UCU)Arg EcoGene#EG30014)
6622 Hpy8I
6622 HincII
6677 AccII
6674 Tth111I
6613 TspEI
6673 BsiEI
6607 AccII
6619 HpyCH4III
6650 MboI
6668 Hpy8I
6601 tgtaatcgcgtgaattatacggtcaacccttgctgagtcaatggcttttgatctggttgctgaacaagtgaacgaccgcgtctgattttctgatttattt 6700
6610 6620 6630 6640 6650 6660 6670 6680 6690 6700
argUp (tRNA promoter of argU operon)
6786 AluI
6785 HindIII
6732 AccII
6778 HpyCH4V
6703 BfmI
6731 HhaI
6777 PstI
6702 MwoI
6726 MspA1I
6742 Hpy188III
6761 NlaIII
6777 BfmI 6787 MwoI
6701 cgctatagcggcaaacaaacgcacaccgctgcgcgtctgaatcaagaaaacccgtattttcatgtatcaaagtgacctgcagccaagcttggattgcgac 6800
6710 6720 6730 6740 6750 6760 6770 6780 6790 6800
6829 NlaIV
6829 HaeIII
6823 TaqI
6812 PsiI 6822 ClaI
6897 HaeIII
6801 TspGWI
6819 BstXI
6837 AluI
6852 Cac8I
6895 NlaIV
6801 acggagttactttataatccaatcgattggccccttagctcagtggttagagcaggcgactcataatcgcttggtcgctggttcaagtccagcaggggcc 6900
6810 6820 6830 6840 6850 6860 6870 6880 6890 6900
ileX (tRNA(CAU)Ile EcoGene#EG30046)
6908 MwoI
6908 BglI
6907 HaeIII
6906 EagI
6906 CfrI
6906 BsiEI
6905 NotI
6903 MspA1I
6901 accagcggccgcaaaggc 6918
6910
ileX (tRNA(CAU)Ile EcoGene#EG30046)

10 References

1. Gibson, D. et al. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **319**, 1215-20 (2008).
2. Knight, T. Idempotent Vector Design for Standard Assembly of Biobricks. in *IT Synthetic Biology Working Group* (2003).
3. Basu, S., Gerchman, Y., Collins, C., Arnold, F. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130-4 (2005).
4. Lu, T. & Collins, J. Dispersing biofilms with engineered enzymatic bacteriophage. *Proc Natl Acad Sci U S A* **104**, 11197-202 (2007).
5. Ro, D. et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940-3 (2006).
6. Anderson, J., Clarke, E., Arkin, A. & Voigt, C. Environmentally controlled invasion of cancer cells by engineered bacteria. *J Mol Biol* **355**, 619-27 (2006).
7. Martin, C., Nielsen, D., Solomon, K. & Prather, K. Synthetic metabolism: engineering biology at the protein and pathway scales. *Chem Biol* **16**, 277-86 (2009).
8. Voloshchuk, N. & Montclare, J. Incorporation of unnatural amino acids for synthetic biology. *Mol Biosyst* **6**, 65-80 (2010).
9. Dong, H., Nilsson, L. & Kurland, C. Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J Mol Biol* **260**, 649-63 (1996).
10. Giegé, R., Sissler, M. & Florentz, C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res* **26**, 5017-35 (1998).
11. Walsh, C., Garneau-Tsodikova, S. & Gatto, G.J. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **44**, 7342-72 (2005).
12. Budisa, N. *Engineering the Genetic Code*, 312 (Wiley-VCH Verlag GmbH & Co. KGaA, 2005).

13. Dougherty, D. Unnatural amino acids as probes of protein structure and function. *Curr Opin Chem Biol* **4**, 645-52 (2000).
14. Hohsaka, T. & Sisido, M. Incorporation of non-natural amino acids into proteins. *Curr Opin Chem Biol* **6**, 809-15 (2002).
15. Link, A., Mock, M. & Tirrell, D. Non-canonical amino acids in protein engineering. *Curr Opin Biotechnol* **14**, 603-9 (2003).
16. Budisa, N. Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire. *Angewandte Chemie International Edition in English* **43**, 6426-6463 (2004).
17. Hendrickson, T., de Crécy-Lagard, V. & Schimmel, P. Incorporation of nonnatural amino acids into proteins. *Annu Rev Biochem* **73**, 147-76 (2004).
18. Wang, L. & Schultz, P. Expanding the genetic code. *Angew Chem Int Ed Engl* **44**, 34-66 (2004).
19. Budisa, N. et al. Toward the experimental codon reassignment in vivo: protein building with an expanded amino acid repertoire. *FASEB J* **13**, 41-51 (1999).
20. Wang, L., Brock, A., Herberich, B. & Schultz, P. Expanding the genetic code of Escherichia coli. *Science* **292**, 498-500 (2001).
21. Young, T. & Schultz, P. Beyond the canonical 20 amino acids: expanding the genetic lexicon. *J Biol Chem* **285**, 11039-44 (2010).
22. Liu, C. & Schultz, P. Adding New Chemistries to the Genetic Code. *Annu Rev Biochem* (2010).
23. Neumann, H., Peak-Chew, S. & Chin, J. Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat Chem Biol* **4**, 232-4 (2008).
24. Yanagisawa, T. et al. Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(epsilon)-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem Biol* **15**, 1187-97 (2008).
25. Nguyen, D. et al. Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA Synthetase/tRNA(CUA) pair and click chemistry. *J Am Chem Soc* **131**, 8720-1 (2009).

26. Rodriguez, E., Lester, H. & Dougherty, D. In vivo incorporation of multiple unnatural amino acids through nonsense and frameshift suppression. *Proc Natl Acad Sci U S A* **103**, 8650-5 (2006).
27. LEVINE, M. & TARVER, H. Studies on ethionine. III. Incorporation of ethionine into rat proteins. *J Biol Chem* **192**, 835-50 (1951).
28. COWIE, D. & COHEN, G. Biosynthesis by Escherichia coli of active altered proteins containing selenium instead of sulfur. *Biochim Biophys Acta* **26**, 252-61 (1957).
29. van Hest, J. & Tirrell, D. Efficient introduction of alkene functionality into proteins in vivo. *FEBS Lett* **428**, 68-70 (1998).
30. Budisa, N. et al. 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**, 788-96 (1995).
31. Hendrickson, W., Horton, J. & LeMaster, D. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J* **9**, 1665-72 (1990).
32. Budisa, N. et al. High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in Escherichia coli. *European Journal of Biochemistry* **230**, 788-96 (1995).
33. Garner, D. et al. Reduction potential tuning of the blue copper center in Pseudomonas aeruginosa azurin by the axial methionine as probed by unnatural amino acids. *J Am Chem Soc* **128**, 15608-17 (2006).
34. Boles, J. et al. Bio-incorporation of telluromethionine into buried residues of dihydrofolate reductase. *Nat Struct Biol* **1**, 283-4 (1994).
35. van Hest, J., Kiick, K. & Tirrell, D. Efficient incorporation of unsaturated methionine analogues into proteins in vivo. *Journal of the American Chemical Society* **122**, 1282-8 (2000).

36. Kiick, K., Saxon, E., Tirrell, D. & Bertozzi, C. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc Natl Acad Sci U S A* **99**, 19-24 (2002).
37. Merkel, L., Cheburkin, Y., Wiltschi, B. & Budisa, N. In vivo chemoenzymatic control of N-terminal processing in recombinant human epidermal growth factor. *Chembiochem* **8**, 2227-32 (2007).
38. Tornøe, C., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J Org Chem* **67**, 3057-64 (2002).
39. Kolb, H., Finn, M. & Sharpless, K. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Ed Engl* **40**, 2004-2021 (2001).
40. Jewett, J.C. & Bertozzi, C.R. Cu-free click cycloaddition reactions in chemical biology. *Chemical Society Reviews* **39**, 1272-1279 (2010).
41. Merkel, L., Beckmann, H.S.G., Wittmann, V. & Budisa, N. Efficient N-terminal glycoconjugation of proteins by the N-end rule. *Chembiochem* **9**, 1220-1224 (2008).
42. Link, A. & Tirrell, D. Cell surface labeling of Escherichia coli via copper(I)-catalyzed [3+2] cycloaddition. *J Am Chem Soc* **125**, 11164-5 (2003).
43. Beatty, K. et al. Fluorescence visualization of newly synthesized proteins in mammalian cells. *Angew Chem Int Ed Engl* **45**, 7364-7 (2006).
44. Dieterich, D., Link, A., Graumann, J., Tirrell, D. & Schuman, E. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc Natl Acad Sci U S A* **103**, 9482-7 (2006).
45. Strable, E. et al. Unnatural Amino Acid Incorporation into Virus-Like Particles. *Bioconjug Chem* (2008).
46. Moroder, L. & Budisa, N. Synthetic Biology of Protein Folding. *Chemphyschem* (2010).
47. Wolschner, C. et al. Design of anti- and pro-aggregation variants to assess the effects of methionine oxidation in human prion protein. *Proc Natl Acad Sci U S A* **106**, 7756-61 (2009).

48. Budisa, N. et al. Atomic mutations in annexin V--thermodynamic studies of isomorphous protein variants. *Eur J Biochem* **253**, 1-9 (1998).
49. Yuan, T. & Vogel, H. Substitution of the methionine residues of calmodulin with the unnatural amino acid analogs ethionine and norleucine: biochemical and spectroscopic studies. *Protein Sci* **8**, 113-21 (1999).
50. Anfinsen, C. & Corley, L. An active variant of staphylococcal nuclease containing norleucine in place of methionine. *J Biol Chem* **244**, 5149-52 (1969).
51. Duewel, H., Daub, E., Robinson, V. & Honek, J. Incorporation of trifluoromethionine into a phage lysozyme: implications and a new marker for use in protein 19F NMR. *Biochemistry* **36**, 3404-16 (1997).
52. Holzberger, B., Rubini, M., Möller, H. & Marx, A. A highly active DNA polymerase with a fluorous core. *Angew Chem Int Ed Engl* **49**, 1324-7 (2010).
53. Budisa, N. et al. Efforts towards the design of 'teflon' proteins: in vivo translation with trifluorinated leucine and methionine analogues. *Chem Biodivers* **1**, 1465-75 (2004).
54. Cirino, P., Tang, Y., Takahashi, K., Tirrell, D. & Arnold, F. Global incorporation of norleucine in place of methionine in cytochrome P450 BM-3 heme domain increases peroxygenase activity. *Biotechnol Bioeng* **83**, 729-34 (2003).
55. Gilles, A. et al. Conservative replacement of methionine by norleucine in Escherichia coli adenylate kinase. *J Biol Chem* **263**, 8204-9 (1988).
56. Ring, M. & Huber, R. The properties of beta-galactosidases (Escherichia coli) with halogenated tyrosines. *Biochem Cell Biol* **71**, 127-32 (1993).
57. Brooks, B., Phillips, R. & Benisek, W. High-efficiency incorporation in vivo of tyrosine analogues with altered hydroxyl acidity in place of the catalytic tyrosine-14 of Delta 5-3-ketosteroid isomerase of Comamonas (Pseudomonas) testosteroni: effects of the modifications on isomerase kinetics. *Biochemistry* **37**, 9738-42 (1998).
58. Budisa, N., Wenger, W. & Wiltschi, B. Residue-specific global fluorination of Candida antarctica lipase B in Pichia pastoris. *Mol Biosyst* (2010).

59. YOSHIDA, A. & YAMASAKI, M. Studies on the mechanism of protein synthesis; incorporation of ethionine into alpha-amylase of Bacillus subtilis. *Biochim Biophys Acta* **34**, 158-65 (1959).
60. Cirino, P.C., Tang, Y., Takahashi, K., Tirrell, D.A. & Arnold, F.H. Global incorporation of norleucine in place of methionine in cytochrome P450 BM-3 heme domain increases peroxygenase activity. *Biotechnology and Bioengineering* **83**, 729-734 (2003).
61. Walasek, P. & Honek, J. Nonnatural amino acid incorporation into the methionine 214 position of the metzincin Pseudomonas aeruginosa alkaline protease. *BMC Biochem* **6**, 21 (2005).
62. Schoffelen, S., Lambermon, M., van Eldijk, M. & van Hest, J. Site-specific modification of Candida antarctica lipase B via residue-specific incorporation of a non-canonical amino acid. *Bioconjug Chem* **19**, 1127-31 (2008).
63. Budisa, N. Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire. *Angew Chem Int Ed Engl* **43**, 6426-63 (2004).
64. Gellman, S. On the role of methionine residues in the sequence-independent recognition of nonpolar protein surfaces. *Biochemistry* **30**, 6633-6 (1991).
65. Shacter, E. Quantification and significance of protein oxidation in biological samples. *Drug Metab Rev* **32**, 307-26 (2000).
66. Berlett, B. & Stadtman, E. Protein oxidation in aging, disease, and oxidative stress. *J Biol Chem* **272**, 20313-6 (1997).
67. Vogt, W. Oxidation of methionyl residues in proteins: tools, targets, and reversal. *Free Radic Biol Med* **18**, 93-105 (1995).
68. Mechulam, Y. et al. Crystal structure of Escherichia coli methionyl-tRNA synthetase highlights species-specific features. *J Mol Biol* **294**, 1287-97 (1999).
69. Serre, L. et al. How methionyl-tRNA synthetase creates its amino acid recognition pocket upon L-methionine binding. *J Mol Biol* **306**, 863-76 (2001).

70. Brunie, S., Zelwer, C. & Risler, J. Crystallographic study at 2.5 A resolution of the interaction of methionyl-tRNA synthetase from Escherichia coli with ATP. *J Mol Biol* **216**, 411-24 (1990).
71. Cory, S., Marcker, K., Dube, S. & Clark, B. Primary structure of a methionine transfer RNA from Escherichia coli. *Nature* **220**, 1039-40 (1968).
72. Dube, S., Marcker, K., Clark, B. & Cory, S. Nucleotide sequence of N-formyl-methionyl-transfer RNA. *Nature* **218**, 232-3 (1968).
73. Schmitt, E. et al. Molecular recognition governing the initiation of translation in Escherichia coli. A review. *Biochimie* **78**, 543-54 (1996).
74. Laursen, B., Sørensen, H., Mortensen, K. & Sperling-Petersen, H. Initiation of protein synthesis in bacteria. *Microbiol Mol Biol Rev* **69**, 101-23 (2005).
75. Old, I., Phillips, S., Stockley, P. & Saint Girons, I. Regulation of methionine biosynthesis in the Enterobacteriaceae. *Prog Biophys Mol Biol* **56**, 145-85 (1991).
76. Waller, J. The NH2-terminal residues of the proteins from cell-free extracts of E. coli. *J Mol Biol* **7**, 483-96 (1963).
77. Kennell, D. & Riezman, H. Transcription and translation initiation frequencies of the Escherichia coli lac operon. *J Mol Biol* **114**, 1-21 (1977).
78. Tang, W. & Zhao, H. Industrial biotechnology: tools and applications. *Biotechnol J* **4**, 1725-39 (2009).
79. Benkovic, S. & Hammes-Schiffer, S. A perspective on enzyme catalysis. *Science* **301**, 1196-202 (2003).
80. Farinas, E., Bulter, T. & Arnold, F. Directed enzyme evolution. *Curr Opin Biotechnol* **12**, 545-51 (2001).
81. Arnold, F. Combinatorial and computational challenges for biocatalyst design. *Nature* **409**, 253-7 (2001).
82. Burton, S., Cowan, D. & Woodley, J. The search for the ideal biocatalyst. *Nat Biotechnol* **20**, 37-45 (2002).
83. Davis, B. Chemical modification of biocatalysts. *Curr Opin Biotechnol* **14**, 379-86 (2003).
84. Ollis, D. et al. The alpha/beta hydrolase fold. *Protein Eng* **5**, 197-211 (1992).

85. Nardini, M. & Dijkstra, B. Alpha/beta hydrolase fold enzymes: the family keeps growing. *Curr Opin Struct Biol* **9**, 732-7 (1999).
86. SARDA, L. & DESNUELLE, P. [Actions of pancreatic lipase on esters in emulsions.]. *Biochim Biophys Acta* **30**, 513-21 (1958).
87. Rathi, P., Bradoo, S., SAxena, R. & Gupta, R. A hyper-thermostable, alkaline lipase from Pseudomonas sp. with the property of thermal activation *Biotechnology Letters* **22**, 495-498 (2000).
88. Jaeger, K., Ransac, S., Koch, H., Ferrato, F. & Dijkstra, B. Topological characterization and modeling of the 3D structure of lipase from Pseudomonas aeruginosa. *FEBS Lett* **332**, 143-9 (1993).
89. Jaeger, K. et al. Bacterial lipases. *FEMS Microbiol Rev* **15**, 29-63 (1994).
90. Noble, M., Cleasby, A., Johnson, L., Egmond, M. & Frenken, L. The crystal structure of triacylglycerol lipase from Pseudomonas glumae reveals a partially redundant catalytic aspartate. *FEBS Lett* **331**, 123-8 (1993).
91. Uppenberg, J., Hansen, M., Patkar, S. & Jones, T. The sequence, crystal structure determination and refinement of two crystal forms of lipase B from Candida antarctica. *Structure* **2**, 293-308 (1994).
92. Verger, R. Interfacial enzyme kinetics of lipolysis. *Annu Rev Biophys Bioeng* **5**, 77-117 (1976).
93. Verger, R. & De Haas, G. Enzyme reactions in a membrane model. 1. A new technique to study enzyme reactions in monolayers. *Chem Phys Lipids* **10**, 127-36 (1973).
94. Royter, M. Technical University Hamburg-Harburg (2006).
95. Reis, P., Holmberg, K., Watzke, H., Leser, M. & Miller, R. Lipases at interfaces: a review. *Adv Colloid Interface Sci* **147-148**, 237-50 (2009).
96. Zaks, A. & Klibanov, A. Enzymatic catalysis in organic media at 100 degrees C. *Science* **224**, 1249-51 (1984).
97. Reetz, M. Lipases as practical biocatalysts. *Curr Opin Chem Biol* **6**, 145-50 (2002).
98. Jaeger, K. & Eggert, T. Lipases for biotechnology. *Curr Opin Biotechnol* **13**, 390-7 (2002).

99. Houde, A., Kademi, A. & Leblanc, D. Lipases and their industrial applications: an overview. *Appl Biochem Biotechnol* **118**, 155-70 (2004).
100. Royter, M. et al. Thermostable lipases from the extreme thermophilic anaerobic bacteria Thermoanaerobacter thermohydrosulfuricus SOL1 and Caldanaerobacter subterraneus subsp. tengcongensis. *Extremophiles* **13**, 769-83 (2009).
101. Winkler, U. & Stuckmann, M. Glycogen, hyaluronate, and some other polysaccharides greatly enhance the formation of exolipase by Serratia marcescens. *J Bacteriol* **138**, 663-70 (1979).
102. Verger, R. 'Interfacial activation' of lipases: facts and artifacts. *Trends in Biotechnology* **15**, 32-38 (1997).
103. Brzozowski, A. et al. A model for interfacial activation in lipases from the structure of a fungal lipase-inhibitor complex. *Nature* **351**, 491-4 (1991).
104. Blow, D. Enzymology. Lipases reach the surface. *Nature* **351**, 444-5 (1991).
105. van Tilbeurgh, H. et al. Interfacial activation of the lipase-procolipase complex by mixed micelles revealed by X-ray crystallography. *Nature* **362**, 814-20 (1993).
106. Turner, N., Needs, E., Khan, J. & Vulfson, E. Analysis of conformational states of Candida rugosa lipase in solution: implications for mechanism of interfacial activation and separation of open and closed forms. *Biotechnol Bioeng* **72**, 108-18 (2001).
107. Agarwal, P., Billeter, S., Rajagopalan, P., Benkovic, S. & Hammes-Schiffer, S. Network of coupled promoting motions in enzyme catalysis. *Proc Natl Acad Sci U S A* **99**, 2794-9 (2002).
108. Svendsen, A. Lipase protein engineering. *Biochim Biophys Acta* **1543**, 223-238 (2000).
109. Yang, J., Koga, Y., Nakano, H. & Yamane, T. Modifying the chain-length selectivity of the lipase from Burkholderia cepacia KWI-56 through in vitro combinatorial mutagenesis in the substrate-binding site. *Protein Eng* **15**, 147-52 (2002).

110. Colton, I., Ahmed, S. & Kazlaukskas, R. Isopropanol Treatment Increases Enantioselectivity of Candida rugosa Lipase toward Carboxylic Acid Esters. *J. Org. Chem* **60**, 212-217 (1995).
111. Kitagawa, M. & Tokiwa, Y. Synthesis of polymerizable sugar ester possessing long spacer catalyzed by lipase from Alcaligenes sp. and its chemical polymerization. *Biotechnology Letters* **20**, *627–630* (1998).
112. Fritz. Enhancement of Selectivity and Reactivity of Lipases by Additives. *Tetrahedron* **56**, 2905-2919 (2000).
113. Okamoto, T., Yasuhito, E. & Ueji, S. Metal ions dramatically enhance the enantioselectivity for lipase-catalysed reactions in organic solvents. *Org Biomol Chem* **4**, 1147-53 (2006).
114. Finkelstein, A., Strawich, E. & Sonnino, S. Characterization and partial purification of a lipase from Pseudomonas aeruginosa. *Biochim Biophys Acta* **206**, 380-91 (1970).
115. Pandey, A. et al. The realm of microbial lipases in biotechnology. *Biotechnol Appl Biochem* **29 (Pt 2)**, 119-31 (1999).
116. Mozaffar, Z., Weete, J.D. & Dute, R. Influence of surfactants on an extracellular lipase from Pythium ultimum. *Journal of the American Oil Chemists' Society* **71**, 75-79 (1994).
117. Freeman, K.S., Tang, T.T., Shah, R.D.E., Kiserow, D.J. & McGown, L.B. Activity and Stability of Lipase in AOT Reversed Micelles with Bile Salt Cosurfactant. *The Journal of Physical Chemistry* **104**, 9312–9316 (2000).
118. Zhang, J., Lin, S. & Zeng, R. Cloning, expression, and characterization of a cold-adapted lipase gene from an antarctic deep-sea psychrotrophic bacterium, Psychrobacter sp 7195. *J Microbiol Biotechnol* **17**, 604-10 (2007).
119. Nam, K., Kim, S., Priyadarshi, A., Kim, H. & Hwang, K. The crystal structure of an HSL-homolog EstE5 complex with PMSF reveals a unique configuration that inhibits the nucleophile Ser144 in catalytic triads. *Biochem Biophys Res Commun* **389**, 247-50 (2009).
120. Kwon, I., Kirshenbaum, K. & Tirrell, D. Breaking the degeneracy of the genetic code. *J Am Chem Soc* **125**, 7512-3 (2003).

121. Schimmel, P. & Söll, D. When protein engineering confronts the tRNA world. *Proc Natl Acad Sci U S A* **94**, 10007-9 (1997).
122. Kwok, Y. & Wong, J. Evolutionary relationship between Halobacterium cutirubrum and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes. *Can J Biochem* **58**, 213-8 (1980).
123. Petrissant, G., Boisnard, M. & Puissant, C. [Purification of a methionine tRNA from rabbit liver]. *Biochim Biophys Acta* **213**, 223-5 (1970).
124. RajBhandary, U. & Ghosh, H. Studies on polynucleotides. XCI. Yeast methionine transfer ribonucleic acid: purification, properties, and terminal nucleotide sequences. *J Biol Chem* **244**, 1104-13 (1969).
125. Gillemaut, P. & Weil, J. Aminoacylation of Phaseolus vulgaris cytoplasmic, chloroplastic and mitochondrial tRNAsMet and of Escherichia coli tRNAsMet by homologous and heterologous enzymes. *Biochim Biophys Acta* **407**, 240-8 (1975).
126. Blanquet, S., Petrissant, G. & Waller, J. The mechanism of action of methionyl-tRNA synthetase. 2. Interaction of the enzyme with specific and unspecific tRNAs. *Eur J Biochem* **36**, 227-33 (1973).
127. Schneller, J., Schneider, C. & Stahl, A. Distinct nuclear genes for yeast mitochondrial and cytoplasmic methionyl-tRNA synthetases. *Biochem Biophys Res Commun* **85**, 1392-9 (1978).
128. Carias, J., Mouricout, M. & Julien, R. Chloroplastic methionyl-tRNA synthetase from wheat. *Biochem Biophys Res Commun* **98**, 735-42 (1981).
129. Schwob, E., Sanni, A., Fasiolo, F. & Martin, R. Purification of the yeast mitochondrial methionyl-tRNA synthetase. Common and distinctive features of the cytoplasmic and mitochondrial isoenzymes. *Eur J Biochem* **178**, 235-42 (1988).
130. Mirande, M., Cirakoğlu, B. & Waller, J. Seven mammalian aminoacyl-tRNA synthetases associated within the same complex are functionally independent. *Eur J Biochem* **131**, 163-70 (1983).

131. Meinnel, T., Mechulam, Y., Fayat, G. & Blanquet, S. Involvement of the size and sequence of the anticodon loop in tRNA recognition by mammalian and E. coli methionyl-tRNA synthetases. *Nucleic Acids Res* **20**, 4741-6 (1992).
132. Wang, L. & Schultz, P. Expanding the genetic code. *Chem Commun (Camb)*, 1-11 (2002).
133. Crepin, T., Schmitt, E., Blanquet, S. & Mechulam, Y. Structure and function of the C-terminal domain of methionyl-tRNA synthetase. *Biochemistry* **41**, 13003-11 (2002).
134. Link, A. et al. Discovery of aminoacyl-tRNA synthetase activity through cell-surface display of noncanonical amino acids. *Proc Natl Acad Sci U S A* **103**, 10180-5 (2006).
135. Tanrikulu, I., Schmitt, E., Mechulam, Y., Goddard, W.r. & Tirrell, D. Discovery of Escherichia coli methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. *Proc Natl Acad Sci U S A* **106**, 15285-90 (2009).
136. Schmitt, E. et al. Switching from an induced-fit to a lock-and-key mechanism in an aminoacyl-tRNA synthetase with modified specificity. *J Mol Biol* **394**, 843-51 (2009).
137. Kiick, K., CM, J., Hest, v. & Tirrell, D. Expanding the Scope of Protein Biosynthesis by Altering the Methionyl-tRNA Synthetase Activity of a Bacterial Expression Host. *Angew Chem Int Ed Engl* **39**, 2148-2152 (2000).
138. Hartman, M., Josephson, K. & Szostak, J. Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proc Natl Acad Sci U S A* **103**, 4356-61 (2006).
139. Hartman, M., Josephson, K., Lin, C. & Szostak, J. An expanded set of amino Acid analogs for the ribosomal translation of unnatural peptides. *PLoS ONE* **2**, e972 (2007).
140. Yoo, T. & Tirrell, D. High-throughput screening for methionyl-tRNA synthetases that enable residue-specific incorporation of noncanonical amino acids into recombinant proteins in bacterial cells. *Angew Chem Int Ed Engl* **46**, 5340-3 (2007).

141. Hirel, P., Schmitter, M., Dessen, P., Fayat, G. & Blanquet, S. Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. *Proc Natl Acad Sci U S A* **86**, 8247-51 (1989).
142. Lawrence, F., Blanquet, S., Poiret, M., Robert-Gero, M. & Waller, J. The mechanism of action of methionyl-tRNA synthetase. 3. Ion requirements and kinetic parameters of the ATP-PPi exchange and methionine-transfer reactions catalyzed by the native and trypsin-modified enzymes. *Eur J Biochem* **36**, 234-43 (1973).
143. Grogan, D. Exchange of genetic markers at extremely high temperatures in the archaeon Sulfolobus acidocaldarius. *J Bacteriol* **178**, 3207-11 (1996).
144. Richmond, M. The effect of amino acid analogues on growth and protein synthesis in microorganisms. *Bacteriol Rev* **26**, 398-420 (1962).
145. Hartley, R. Barnase and barstar: two small proteins to fold and fit together. *Trends Biochem Sci* **14**, 450-4 (1989).
146. Golbik, R., Fischer, G. & Fersht, A. Folding of barstar C40A/C82A/P27A and catalysis of the peptidyl-prolyl cis/trans isomerization by human cytosolic cyclophilin (Cyp18). *Protein Sci* **8**, 1505-14 (1999).
147. Kiick, K. University of Massachusetts Amherst (2001).
148. Kadner, R. Regulation of methionine transport activity in Escherichia coli. *J Bacteriol* **122**, 110-9 (1975).
149. Old, J. & Jones, D. A comparison of ethionine with methionine in Escherichia coli in vitro polypeptide chain initiation and synthesis. *FEBS Lett* **66**, 264-8 (1976).
150. Wang, A., Winblade Nairn, N., Johnson, R., Tirrell, D. & Grabstein, K. Processing of N-terminal unnatural amino acids in recombinant human interferon-beta in Escherichia coli. *Chembiochem* **9**, 324-30 (2008).
151. Chang, A. & Cohen, S. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol* **134**, 1141-56 (1978).

152. Juhling, F. et al. tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Research* **37**, D159-D162 (2009).
153. Schmidt, T. & Skerra, A. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat Protoc* **2**, 1528-35 (2007).
154. http://www.genscript.com/cgi-bin/tools/rare_codon_analysis. GenScript Rare Codon Analysis Tool.
155. Osawa, S., Jukes, T., Watanabe, K. & Muto, A. Recent evidence for evolution of the genetic code. *Microbiol Rev* **56**, 229-64 (1992).
156. Burgess-Brown, N. et al. Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study. *Protein Expr Purif* **59**, 94-102 (2008).
157. Ryu, Y. & Schultz, P. Efficient incorporation of unnatural amino acids into proteins in Escherichia coli. *Nat Methods* **3**, 263-5 (2006).
158. Richardson, J. Rho-dependent transcription termination. *Biochim Biophys Acta* **1048**, 127-38 (1990).
159. Young, T., Ahmad, I., Yin, J. & Schultz, P. An enhanced system for unnatural amino acid mutagenesis in E. coli. *J Mol Biol* **395**, 361-74 (2010).
160. Taki, M. et al. Leucyl/Phenylalanyl-tRNA-protein transferase-mediated chemoenzymatic coupling of N-terminal Arg/Lys units in post-translationally processed proteins with non-natural amino acids. *Chembiochem* **7**, 1676-9 (2006).
161. Watanabe, T., Miyata, Y., Abe, R., Muranaka, N. & Hohsaka, T. N-terminal specific fluorescence labeling of proteins through fourbase codon-mediated incorporation of fluorescent hydroxy acid. *Nucleic Acids Symp Ser (Oxf)*, 363-4 (2007).
162. Muranaka, N., Miura, M., Taira, H. & Hohsaka, T. Incorporation of unnatural non-alpha-amino acids into the N terminus of proteins in a cell-free translation system. *Chembiochem* **8**, 1650-3 (2007).
163. Goto, Y. et al. Reprogramming the Translation Initiation for the Synthesis of Physiologically Stable Cyclic Peptides. *ACS Chem Biol* (2008).

164. Connor, R., Piatkov, K., Varshavsky, A. & Tirrell, D. Enzymatic N-terminal Addition of Noncanonical Amino Acids to Peptides and Proteins. *Chembiochem* (2008).
165. Heal, W. et al. Site-specific N-terminal labelling of proteins in vitro and in vivo using N-myristoyl transferase and bioorthogonal ligation chemistry. *Chem Commun (Camb)*, 480-2 (2008).
166. Watanabe, T., Miyata, Y., Abe, R., Muranaka, N. & Hohsaka, T. N-terminal specific fluorescence labeling of proteins through incorporation of fluorescent hydroxy acid and subsequent ester cleavage. *Chembiochem* **9**, 1235-1242 (2008).
167. Goto, Y., Murakami, H. & Suga, H. Initiating translation with D-amino acids. *RNA* **14**, 1390-8 (2008).
168. Hotelier, T. et al. ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins. *Nucleic Acids Res* **32**, D145-7 (2004).
169. Lepthien, S., Merkel, L. & Budisa, N. *In vivo* double and triple labeling of proteins using synthetic amino acids. *Angewandte Chemie International Edition in English* (in press).
170. Abdeljabbar, D., Klein, T., Zhang, S. & Link, A. A single genomic copy of an engineered methionyl-tRNA synthetase enables robust incorporation of azidonorleucine into recombinant proteins in E. coli. *J Am Chem Soc* **131**, 17078-9 (2009).
171. Hausmann, C.D. & Ibba, M. Aminoacyl-tRNA synthetase complexes: molecular multitasking revealed. *Fems Microbiology Reviews* **32**, 705-721 (2008).
172. Conti, E. & Izaurralde, E. Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Curr Opin Cell Biol* **17**, 316-25 (2005).
173. Hirao, I. et al. An unnatural base pair for incorporating amino acid analogs into proteins. *Nat Biotechnol* **20**, 177-82 (2002).
174. Dong, S., Merkel, L., Moroder, L. & Budisa, N. Convenient syntheses of homopropargylglycine. *J Pept Sci* **14**, 1148-50 (2008).
175. http://www.biology.utah.edu/jorgensen/wayned/ape/.

176. Schmidt-Dannert, C., Sztajer, H., Stöcklein, W., Menge, U. & Schmid, R. Screening, purification and properties of a thermophilic lipase from Bacillus thermocatenulatus. *Biochim Biophys Acta* **1214**, 43-53 (1994).
177. Söding, J., Biegert, A. & Lupas, A. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**, W244-8 (2005).
178. Goldstone, D. et al. Structural and functional characterization of a promiscuous feruloyl esterase (Est1E) from the rumen bacterium Butyrivibrio proteoclasticus. *Proteins* **78**, 1457-69 (2010).
179. Kiss, A. et al. The acylaminoacyl peptidase from Aeropyrum pernix K1 thought to be an exopeptidase displays endopeptidase activity. *J Mol Biol* **368**, 509-20 (2007).
180. Bertrand, T. et al. Structural basis for human monoglyceride lipase inhibition. *J Mol Biol* **396**, 663-73 (2010).
181. Eswar, N. et al. Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* **Chapter 5**, Unit 5.6 (2006).
182. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**, 10881-90 (1988).