Artificial Multi-Gene Expression Systems Design Service for Natural Compound Formation and Hetero Protein Complexes

Bernauer, Hubert[1]*Gregor Zipf[1] and Josef Maier[2]

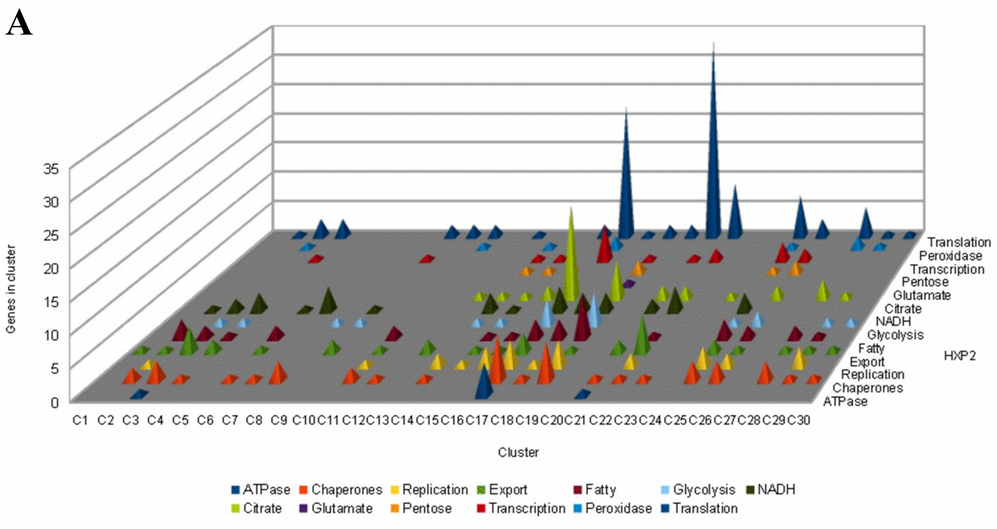
1 **ATG:biosynthetics GmbH**, www.atg-biosynthetics.com, info@atg-biosynthetics.com

2 IStLS Information Services to Life Science, Haerlestr. 24/1, 78727 Oberndorf

Genomic data can be extremely valuable for identifying formal molecular parameters relevant for functionally recoding synthetic genes.

ATG:biosynthetics has developed **analytical** comparative genomics and **constructive *in silico* solutions** for the field of multi-gene construction and multiple-protein expression systems where all orchestral genes are located on one to a few assembled constructs. We achieve this by intended, constructive biological designs based on experience learned during many customer projects. Constructive molecular biology approaches for creating **artificial multi-gene expression systems** in order to express natural compounds - ***in vivo*** and ***in vitro*** - need basic *molecular* d*esign concepts, synthetic bioinformatics services, vector systems and functionally predictive sequence modulations.* ***Multi-parametric*** sequence optimization calculations of regulator regions and coding sequences of pathway genes are performed ***in parallel***. Thorough comparative analyses of related genomes allow the extraction of host-specific design rules, which are applied during sequence optimization. Our setups include iterative design of experiments that makes extensive use of active learning heuristics, in order to speed up the overall pathway optimization processes.

A

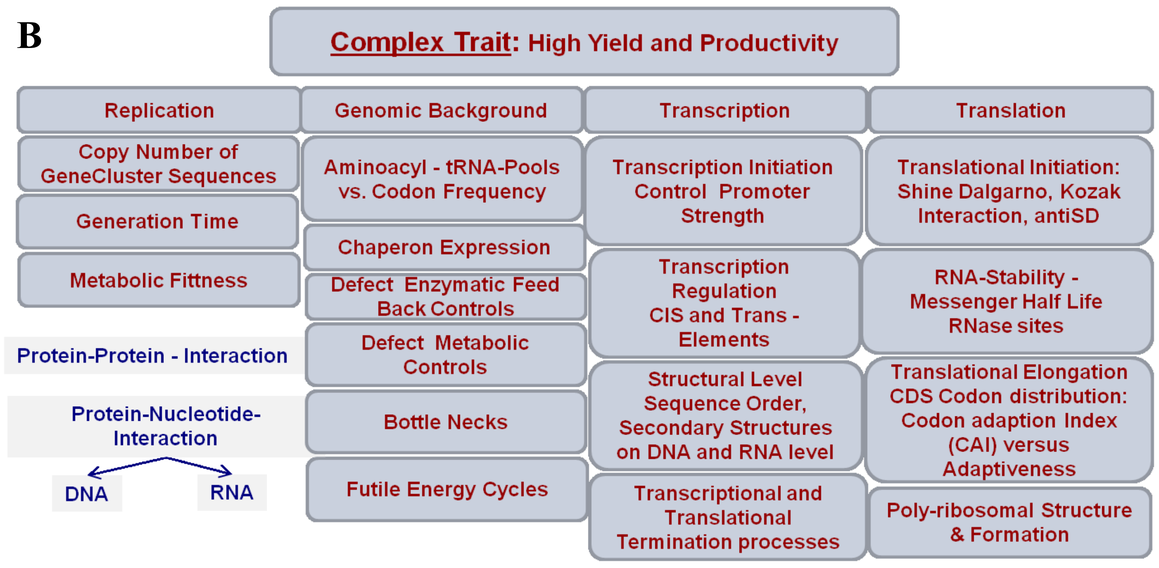


Clustering based on combinatorial (maximum) likelihood optimization reveals patterns of codon usage.

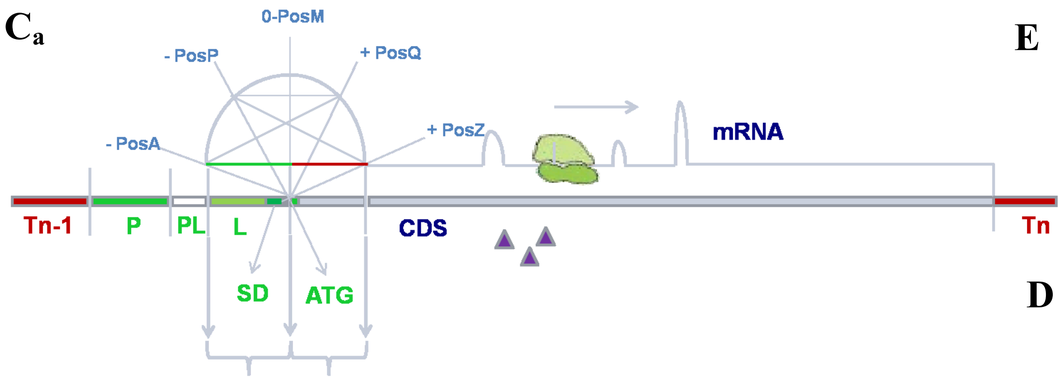
Example of genomic analyses of the whole high expression genes „hxp-member counts“ of the genomic gene set of the microbe ***Myxococcus xanthus***. Different functional sub-gene sets are clustered with significantly different codon usage. This analysis reveals different distinct codon tables for specific functional groups of genes labelled in different colors.

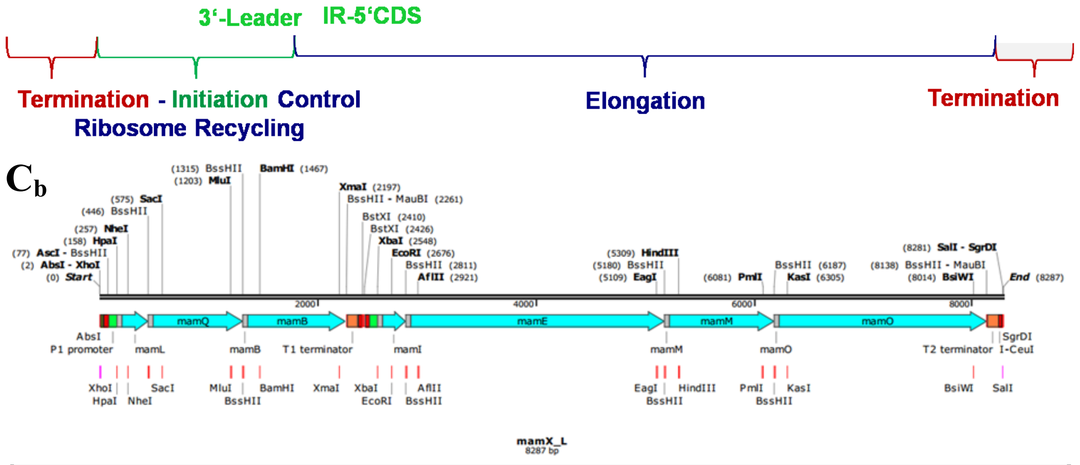
In addition the methodology allows for the identification of clusters of evolutionary closely related genes like secondary metabolite pathways but also molecular differentiation processes where physiological bias reflects higher expression of specific functionally related gene sets. The method can be applied for virtually all genomes of interest.

B



C*a*



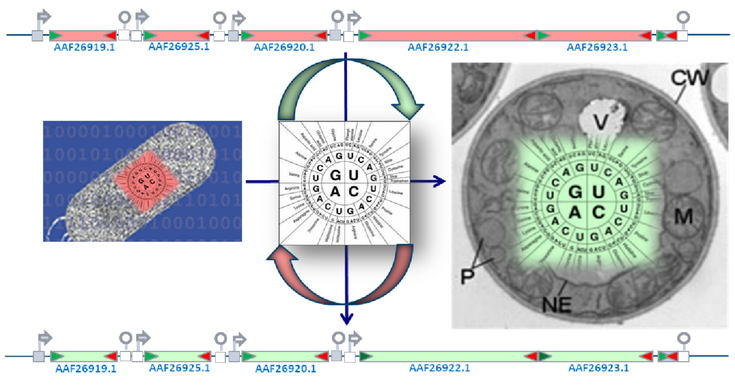


C*b*

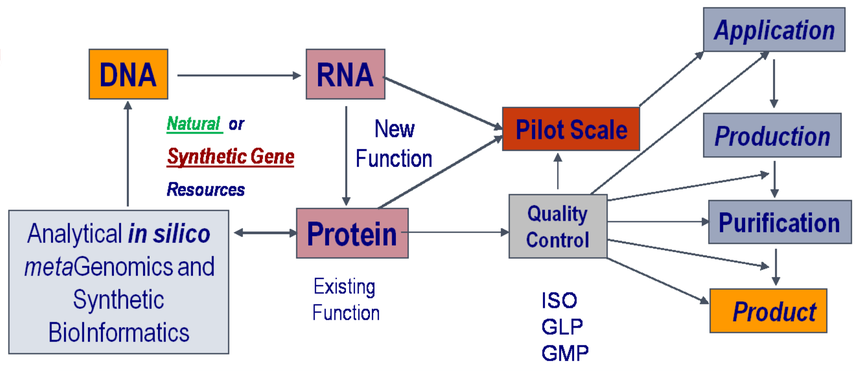
E

ATG’s primary interests are supporting customers in molecular genetics and genomic analyses around their projects on a very high level of knowledge. ATG can solve questions where genomic analyses and formal-functional molecular designs of genes play an important role. An comprehensive experience over the years is that most, if not all, functional features are encoded at least implicitly in the DNA and that intended functional behavior can be formally treated in case the formal molecular prerequisites are known.

Professional re-design of individual synthetic genes and the retro-design of specific gene clusters e.g. in heterologous expression of hetero-protein complexes and artificial biochemical pathway constructions require detailed genomic knowledge in order to predict the functional behavior of genes and artificially composed gene clusters. In addition to formal-functional molecular design flexible modular structural designs are required for implementing functional features into sequence contexts optimized for molecular handling properties.



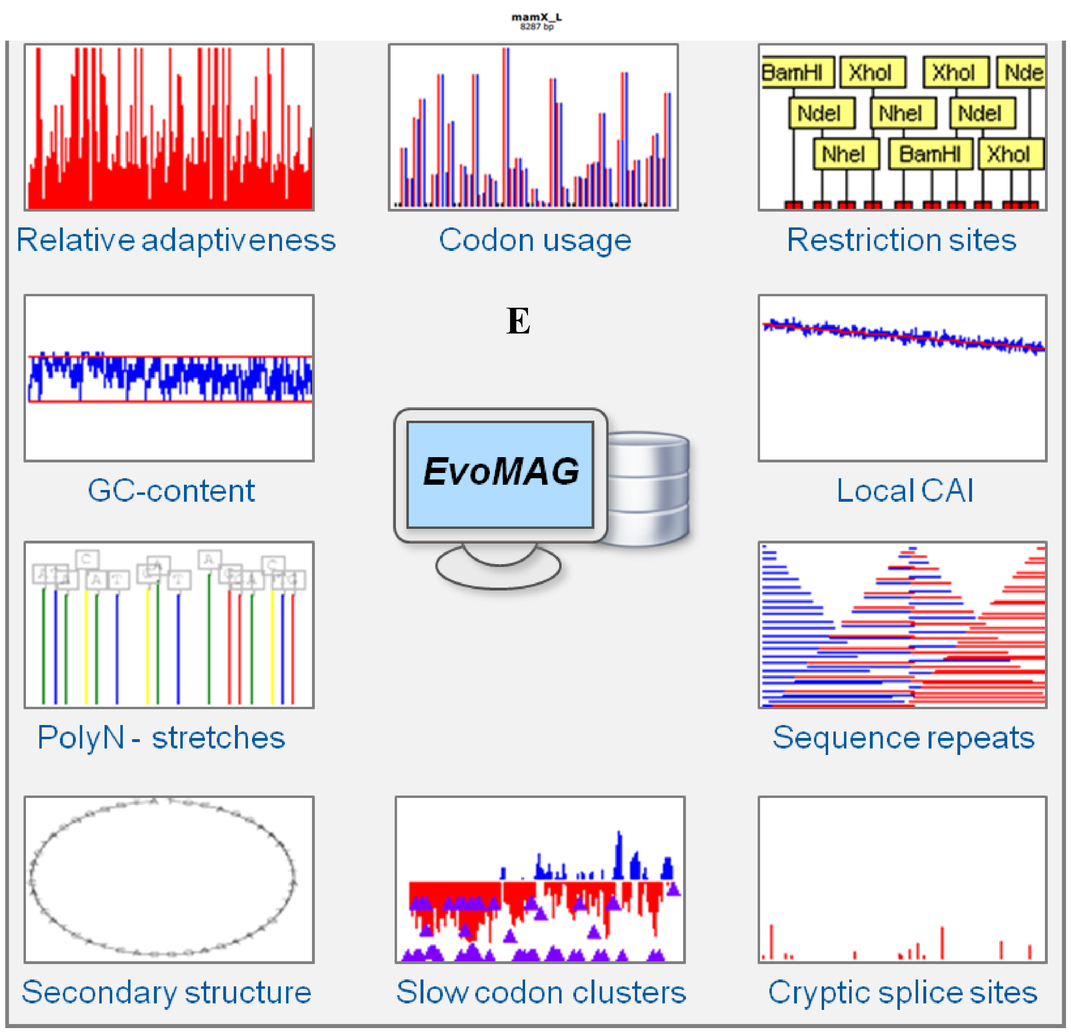
D



Detailed comparative genomic knowledge assessing all genes and its qualitative composition has a deep impact in a formal way on understanding the meaning of gene function qualitatively as well as quantitatively. Moreover besides the quantitative expression yield qualitative traits like specific activity of proteins and its native solubility are mirrored in the genomic information pool.

Improving product yields and specificity in bio-productions is a multifactorial process which includes the optimization of culturing conditions like culturing media and all upstream processing measures one can take in order to adjust the rigt conditions for yield optimization. In addition to classical strain improvements by selection processes advanced technical genetics (ATG) serves with front edge tools for improving forward designs, But also reverse engineering approaches can be increasingly improved with applying rational to semi– rational approaches in genetic engineering.

E



F

Bioproduction Product	Pathway in kb	Application	Constructional Design	Expression Optimization	Leader-Library	Specific Vector: Design & Syntheses
Epothilone[1]	65	Cytostatic	yes	yes	yes/no	yes
ω-3-PUFA	31,4	Essent. Fatty Acid	yes	yes	yes	yes
Argyrin	37,2	Antibiotic	yes	yes	yes	yes
Methyl-MalonylCoA	18,8	Metabolic	yes	yes	no	yes
MyxoChromid	32	Antibiotic	yes	yes	no	yes
Coronatin	36	Phytotoxin	yes	no	no	yes
Bottromyicin	20	Antibiotic	yes	no	no	yes
Myxopyronin	53,4	Antibiotic	yes	no	no	yes
Glumicin	35,2	Antibiotic	yes	no	no	yes
Magnetit	8,9	Biomagnetism	yes	no	no	yes
Hetero-Protein Complexes	10,4	Regulation on Protein Level	yes	yes	no	yes
Multiple Glycosyl-transferase genes	9,5	Glycosyl-ation	yes	yes	no	yes

F: Table 1. Track record of successful customer projects.

A very high expression of artificial pathways in heterologous expression need integrative modulations and improvements on different technological levels of the desired system in total (A). Starting from the genetic engineering level to the process optimization level upstream even downstream processing technology is usually required. **B** Complex, composite production traits with required features, like high production yield, robustness etc. are encoded in complex sequence pattern. For the improvement of artificial biosyntheses advantageous features of living cells like being self replicating systems, information processing, compartimentation, systems behavior, modularity, orthogonality and cellular differentiation are deployed. Strain development is by ***forward*** selection.

ATG:biosynthetics services organizes the value chain from financing efforts to bioinformatics, from lab scale optimization to pilot scale and production.

On all molecular levels sequence amendments with advantageous features are optimized by use of proprietary ***in silico*** technology. **C** Reverse engineering of functional multi-gene constructions like artificial metabolic pathways or multiple -hetero-protein complexes require reprogramming of individual genes (**C*a*,, D**) and its integration into a perfectly orchestrated molecular system (**C*b***). Despite high production yields this can mean features like increasing robustness and reproducibility of gene cluster expression (**E**). The ATG:***Evo*MAG** computer aided molecular gene design and multi-target parameter synthetic bioinformatic approaches are the backbone of all planning phases of Synthetic Biology projects which are performed at ATG.

Future planning and realization of Synthetic Biology Projects need analyses of natural systems, high level structuring of information and its systematization. The planning phase of projects in Synthetic Biology is most important. The integration of all proprietary information and published data are leading to a model which heuristically accumulates all experimental experiences in a "**proof of concept**" and finally in a **pilot scale (A)**. Functional traits like high production yields are composed of individual features of lower complexity from different molecular levels and can be resolved down towards individual features of molecular interaction **(B)**. Based on the knowledge gained molecular systems ***in vivo*** can be subjected to simplification, standardization and modularization. In general the development of Synthetic Biology applications with desired artificial production schemes **(A)** ***in vivo*** are focused on the generation of defined production strains with either an output of highest specific production yields possible or highest regulated dosage of compound production.

Osswald et al. (2012) DOI: 10.1021/sb300080t— www.atg-biosynthetics.de