BPEL

Simulation of translation

Olli Niemitalo, Ulf Liebal, André Juffer, Peter Neubauer

Confidential

BPEL

Summary of traineeship

The traineeship took place under the surveillance of triacle biocomputing, André Juffer, and the Bioprocess Engineering Laboratory of the University of Oulu, Peter Neubauer. The core assignment was the validation of a simulation program of the translation apparatus in E. coli, written by Olli Niemitalo.

The objective:

The goal is to build up a model of the translation in E. coli, which allows the estimation of product formation. Because protein degradation in E. coli may governed by first order kinetics, and can be therefore assumed to be linear to protein concentration, the relative protein concentrations[i] might be obtained with use of mRNA half lives and mRNA amount[ii] as seen in Illustration 1. The mRNA degradation is tightly connected with transcription and translation and well balanced. In the case of recombinant gene expression often plasmids are used maintaining, through the T7 RNA polymerase, a huge over transcription of the gene of interest, with exceptional high cellular mRNA concentrations. Under these circumstances, which differ from the genuine case and where transcription, translation and degradation are uncoupled[iii] one might expect that the influence of mRNA concentration to the yield will drop, but the importance of half life will rise. The reason is probably that the differences of mRNA concentrations of different recombinant genes are low compared to the actual concentration, but resides also in the limiting amount of tRNAs, ribosomes and other constituents of the translation system equalizing the translation for the exceptional large amounts of different
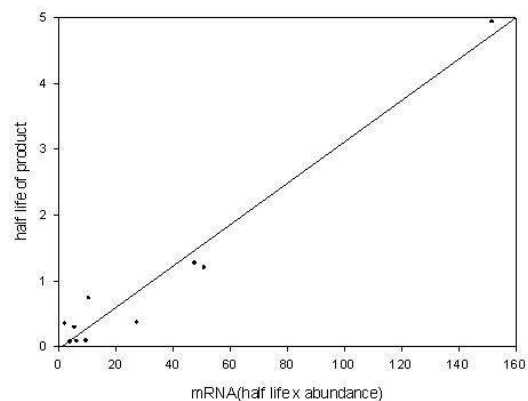

Illustration 1 mRNA half life x abundance vs. relative protein product. Data: half life, abundance: Bernstein et al.[xvii]; relative product: van Bogelen et al.[xvii]
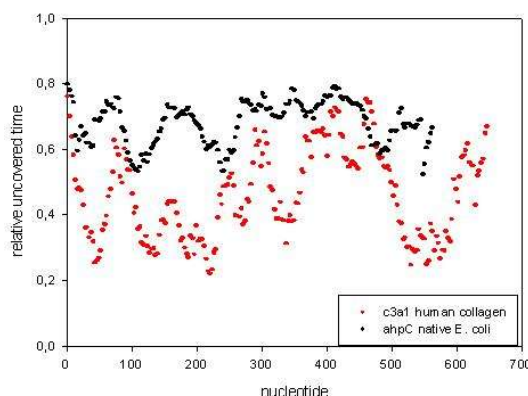

Illustration 2: relative uncovered time for native E. coli and recombinant gene.

recombinant mRNAs. With that assumption a promising issue for further optimization of the product yield is the prolongation of the half life of recombinant mRNAs, but that assumption is also an issue for validation in the fully developed simulation program. The main difficulty in recombinant gene expression is the difference in the tRNA usage of different host organisms. A gene, smoothly expressed in the genuine being, due to the balanced use of tRNAs, may consist of unfavorable tRNAs in a new host, due to different tRNA concentrations and following differences in translation speed[iv,v,vi], and shows therefore often a much more rough translation, what causes ribosome stagging and subsequently RNase cleavage, Illustration 2. As mentioned before ribosome queuing seems to be of central importance to the life time of mRNAs[vii]. Densely packed ribosomes prevent the RNases to cut, as in the case for ribosomes in the queue, but RNase activity is facilitated

BPEL

downstream of the queue. A final and general mechanism for mRNA degradation is still not known, so their significance in the model has to be tested empirically by experiments.

The model:

The model is basing on the stochastic simulation algorithm developed by Gillespie[viii].

A model that can be simulated using the algorithm consists of:

- Molecular species, each containing an integer number of molecules.
- Reaction channels that convert some molecular species into others, the stoichiometries of the reactons.
- Propensity functions associated with the reaction channels.

The meaning of propensity is the probability density over time for a particular reaction channel to fire. For each reaction channel, the propensity is determined by the associated propensity function. A typical second order propensity function is a product of a rate constant and the sizes of two pools, equivalent to a second order rate law.

In the system simulated, there can only be an integer number of each molecular species. Indeed, the Gillespie algorithm generates simulation trajectories consisting of the actual times each elementary reaction took place. These event times are random and they are generated during the simulation from appropriate probability distributions. This is in contrast to deterministic simulations based on ordinary differential equations that deal with real numbers and produce the same trajectory each time.

A chain of reactions can be modeled as a single elementary reaction if one of the reactions is rate-limiting. Otherwise the chain should usually be divided into partial reactions.

The molecular species under investigation are:

- RNA Polymerase
- free mRNA
- ribosome_mRNA
- ribosome_mRNA_tRNA
- translation product

Other considered species are tRNAs, always charged, and ribsosomes, but their concentration is assumed to be buffered and does not change throughout the simulation[ix]. Subsequently the reaction rate does not change and is constrained following values of initiation by Arnold[xvii] and tRNA translation rates calculated by Solomovici et al.[iv]!

These 5 constituents of the system can undergo transformations by the reaction channels:

- mRNA elongation
- ribosome_bind_mRNA
- ribosome_receive_tRNA
- ribosome_translocate
- ribosome_finish_translation
- RNase attack (Not implemented to the expiring of the traineeship)

Each reaction channel is associated with a stochastic reaction rate constant[viii].

The mRNA elongation is the first reaction taking place on the mRNA, and it produces free mRNAs:

RNA_polymerase[x] --> RNA_polymeras[x + 1] + free_mRNA[x]

Where x denotes for nucleotide position. The rate constant is taken from Neidhardt[x], k = 43 nucleotides /s.

BPEL

The initiation of translation, modeled by the ribosome_bind_mRNA reaction channel follows the reaction equation:

free_mRNA[firstcodonpos + ribosomestart .. firstcodonpos + ribosomeend] --> rib_mRNA [firstcodonpos...]

The nucleotides listed in the brackets list the positions which have to be free of ribosomes for that a new ribosome has space to bind. We used a value of k = 0.8 /s which is also used by Arnold[xi]. In the simulations we tried to ensure that initiation is the rate limiting step. In E. coli one assumes, that initiation has to be controlled in order to control the energy expenditure of translation, cf. Yarchuk[xii]. The reaction channel of ribosome_receive_tRNA includes the reactions of initial binding, codon recognition, GTPase activation, EF-Tu conformation change, accommodation and peptidyl transfer[xiii], and is only concerned with the binding of the cognate tRNA. We assume that the tRNAs are in short supply, and therefore rate limiting for the elongation process. We furthermore assume that the binding time of non cognate tRNAs is extremely short and subsequently not interfering or inhibiting binding of cognate tRNAs.

ribosome_mRNA[x] --> ribosome_mRNA_tRNA[x]

The reaction rate is codon depentend and the values are taken from Solomovici et al. pg.: 518 column 6, tsup[iv]!
While the ribosome translocation reaction, the center of the ribosome is positioned 3 nucleotides further downstream releasing 3 nucleotides in the back and covering 3 nucleotides in the front. The space a ribosome covers is always 10 codons[xiv]. The reaction rate is always the same and is much faster than any other reaction and has therefore only minimal effect on the simulation result. A reaction rate of 100000 /s is chosen.
The ribosome_finish_translation follows the reaction scheme:

ribosome_mRNA[terminationpos] --> free_mRNA[terminationpos + ribosomestart .. terminationpos + ribosomeend] + translation_product

In the literature are given nonuniform estimations for termination, and precise values could not be found which would fit in the program. We set emphasize on the issue that the initiation needs to be the rate limiting step, having a value of k_{ini} = 0.8 /s. A probable estimation for termination might be a value of k_{term} = 2 /s, which is faster than initiation, but is still one of the slowest translated codons. The termination value has vital significance for the model, it determines whether global queuing of ribosomes or only local queuing will arouse. A termination value of 2 /s does not induce global queuing, and further comparisons have to be made to estimate a reasonable value for termination. The dependence of the program to the termination value was tested while using the codon demand conditions of Solomovici et al., pg.:518, column7, tdem[iv]!
The RNase attack does not belong to a specific type nor is it sequence dependent. The model will not contain a realistically simulated RNase activity, instead the possibilities of a virtual cut will be counted and processed to obtain the cutting probability over time and nucleotide. This will be the basis for the calculation of the probable first cut position. This modeling does not depend on any proposed degradation mechanism, but can be adopted to any mechanism.

BPEL

For supplementary analysis other additional tools were developed. We suggest the tRNA usage to be a more reliable way of measuring the usage of tRNA in different species as the codon usage is. The tRNA usage is calculated after Equation 1:

$$\mathcal{H}(t) = c_{\text{in } C(t)} * \frac{\mathcal{H}(c) * [t]}{u_{\text{in } T(c)} * [u]}$$

Equation 1

Wherein $\mathcal{H}$ (t) and $\mathcal{H}$ (c) denote the usage frequencies of the tRNA species t and a codon c, respectively, C(t) denotes the set of all codons recognized by the tRNA t and T(c) denotes the set of all tRNAs that recognize the codon c, and [t] and [u] are concentrations of the tRNA species t and u, respectively. But a definite proof for the usability remains, because the computing algorithm does release some irregular outcomes.

The tRNA usage can used for single gene analysis, for that purpose a program was written, calculating the codon usage for a given sequence (freqanalyzer#.perl). For a more effective gene acquisition in E. coli, an algorithm was written, issuing the mRNA sequence of any of the 4288 K12 E. coli genes on the basis of genome analysis of Blattner et al.[xv](gensearch#.perl). For engineering with the codon speed an algorithm was programed, which maps the translation speed to the codon position (posfreq#.perl).

Report overview:

For purpose of validation several tests were performed with the simulation program.

At first the differences between the distinct translation speed sets of Solomovici et al.[iv] should be explained. Solomovici et al. computed translation speeds with the assumption that nature has optimized the translation with respect to tRNA concentrations and translation speeds. In their model the translation time of tRNA(i) - t_i is dependent of the relative tRNA(i) concentration - n_i and the codon anticodon match - f_{ij} with respect to a reference value of k = 198 ms /% tRNA.

$$t_i = \frac{k}{f_{ij} * n_i}$$

Equation 2

The reference value is obtained by applying optimization procedures to the mean translation time of lacZ mRNA. The optimization hypothesis states that for optimal translation conditions first the codon frequency is proportional to the square of the interaction efficiency of codon-anticodon and second that the tRNA specie concentration is proportional to the sum of the square root of cognate codon frequencies. The demand speed arouses due to the assumption, that a hypothetical demand tRNA pool can be calculated (n_i^*), on the basis of an optimization between the physical interaction between codon anticodon (f_{ij}) and the sum of all recognized codon frequencies, $\mu_i = \Sigma \sum \sqrt{\upsilon_{ij}}$ where υ_{ij} symbolizes the codon frequency of codon(i,j):

$$n_i^* = \frac{\mu_i / \sqrt{f_{ij}}}{\mu_i / \sqrt{f_i}}$$

Equation 3

The main difference between offer and demand is such a way, that for codons starting with pyrimidine f_{ij} n_i exceeds f_{ij} n_i^* in most instances, and the reverse holds for codons beginning with a
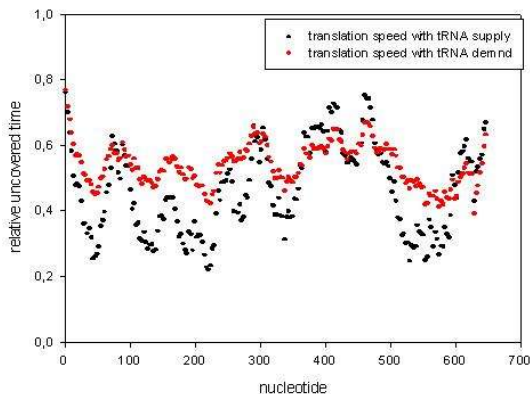
Illustration 3: comparison of demand and supply translation times of Solomovici et al.

purine. Further information about demand and supply conditions and comparisons can be obtained from Solomovici et al.[iv]. Here the effect of the two codon speeds in the simulation program is examined. Table 1 shows the differences of the translation times and Illustration 3 shows the application of these speed on a c3a1 mRNA. As can be seen in Illustration 3 the range of translation time is covering a broader range for the supply state, and shows therefore steeper peaks and hollows. But the general character of the plot remains the same. Owing to slightly irritating expressions in the paper, the first simulations were all performed with the demand conditions, but as Illustration 3 indicate, there are many similarities between the two states. (cf.: over expression report, a_overexpression.doc)

supply case		demand case	
time	**product**	**time**	**product**
26,06	1,000	19,55	1,000
19999,64	8019,000	19999,73	9720,000
time/codon	0,121	time/codon	0,091
product/time	0,401	**product/time**	0,486
statistics of numboundrib		**statistics of numboundrib**	
mean	11,801	mean	10,224
variance	3,481	variance	4,216
rib/codon	0,0549	**rib/cod**	0,0476

<u>Table 1</u>: Comparison several properties between supply and demand case.

The simulation of translation can be run for arbitrary time lengths and therefore arbitrary, and certainly, unrealistic half lives of mRNAs. Long single time runs are taken to gain statistical values and not many small runs, but the effect could be similar, albeit it has to be verified. For statistical analysis of differences the empirical mean squared variance was applied. The relative % variance is obtained by dividing the variance through the mean. For three different time runs, 6000 s, 20,000 s and 50,000 s, three properties of the simulation are compared, the ratio of free mRNA positions, the relative ribosome coverage time per codon and the calculated amount of product. The results indicate that the applied statistical analysis gives consistent results for simulation lengths of 20,000 s which do not alter much for 50,000 s. The ratio of free mRNA position and subsequently the relative ribosome coverage time per codon exhibits a very different pattern in one of the 5 runs for 6000 s as seen in Illustrations 4&5. The wnt1 gene consists of 370 codons. This codon length lies near the average codon length in E. coli. For the
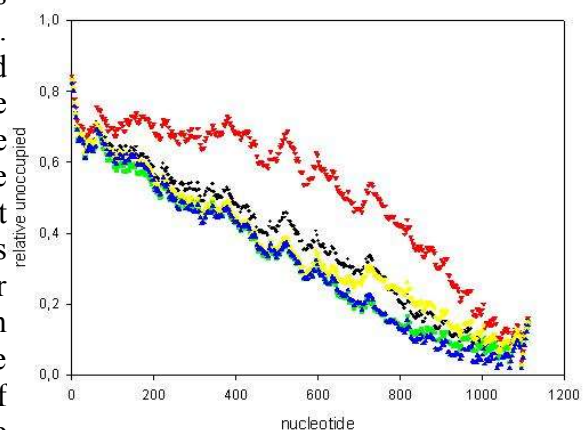

Illustration 4: relative uncovered time for five 6000 s runs.

BPEL

15 genes of the ensemble simulation the mean is 534 codons. Caution is needed for short time runs and large mRNAs.
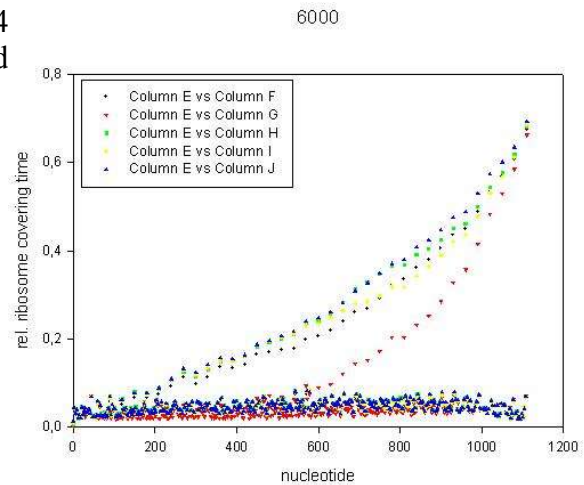Further information in "Statistical quality analysis".

Illustration 5: relative ribosome covering time for five 6000 s runs.

Nonetheless, the first performed simulation consisted of a poly U chain with 1716 residues to learn more about general behavior of the model. The used translation times are the demand conditions of Solomovici et al.[iv], the simulation length was 15,000 s, initiation rate and termination rate were set to 0.5 s^{-1}. These conditions of slow and balanced rates for initiation and termination assures that a medium ribosome coverage develops so that effects of nucleotide changes in the chain are not covered by a huge global ribosome queue or by the absence of any ribosome interaction. The rate equals the reported[xvi] value for termination with release factor 2. The used U codon has a translation speed in the set of 16.369 s^{-1}, the fast introduced codon is gaa with a speed of 21.739 s^{-1}, and the slow codon is cua translating with 5.051 codons per second. Introduced are always 5 codons to a stated position. Other parameters are listed in "Test of the simulation program with poly U".

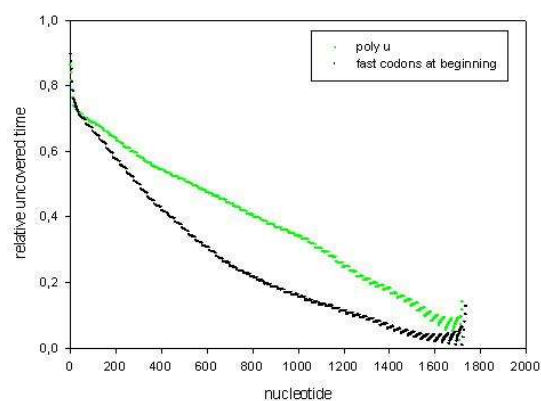*Illustration 6: relative uncovered time for 1716 poly U chain.*

Illustration 6 shows that the relative uncovered time of ribosomes for a nucleotide drops linearly over the chain length. There is a smooth ribosome queue owing to balanced initiation and termination. Under these conditions the introduction of fast codons at the initiation region lowers the relative uncovered time pattern for later codons, Illustration 7, presumably due to a higher ribosome load owing to facilitated clearing of the initiation region. The higher ribosome load results in a higher intensity of ribosome interaction.

Illustration 7: comparison of genuine poly U and poly U with fast codons near initiation.
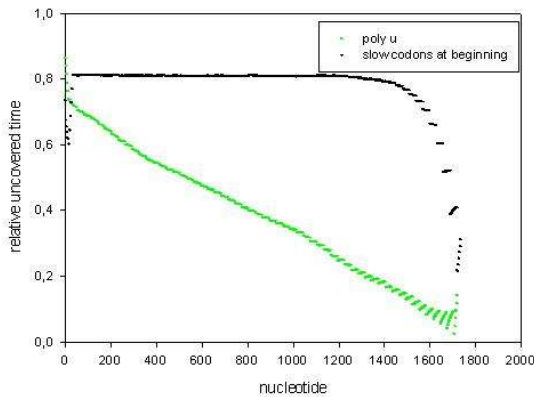
BPEL

Illustration 8: genuine poly U chain and poly U
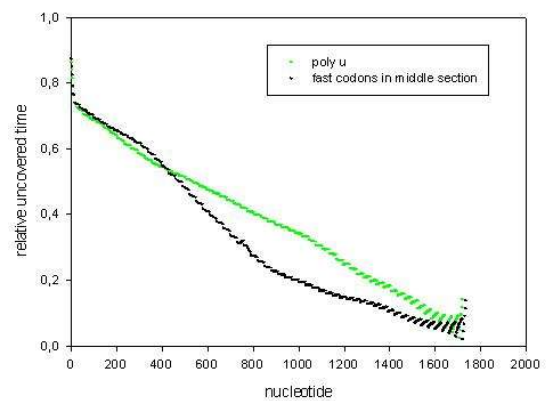with slow codons at initiation.

Illustration 9: genuine poly U chain and poly U
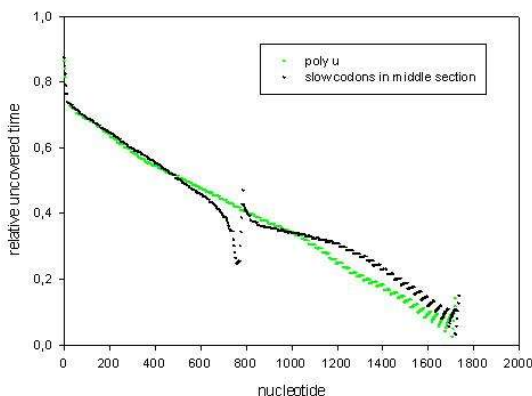with fast codons in the middle.

Illustration 10: genuine poly U and poly U with
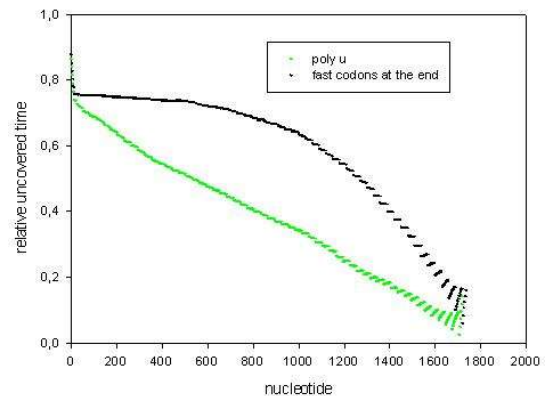slow codons in the middle.

Illustration 11: genuine poly U and poly U with
fast codons at termination.

Are slow codons engineered in the initiation region, as seen in Illustration 8, the effect is reverse. Ribosomes are retarded in the initiation region and hinder new ribosomes to bind causing a much lower ribosome load. The ribosomes are far apart in distance and do not interact with each other, recognizable in the horizontal line at a high relative uncovered time. Interactions occurs only near the slow termination site.

If fast codons are implemented in the middle region of the poly U chain, Illustration 9, after reaching equilibrium for time run of 15,000 s, the relative uncovered time effect precedes the changed position. The reason may be that ribosomes are edged to the end because of the few fast codons in the middle resulting in more ribosome interaction and denser ribosome queues from the end directed to the beginning.

Slow codons in the middle nicely illustrates, Illustration 10, the appearance of a small ribosome queue. There is a remarkable drop in the relative uncovered time followed by a sharp rise of the self. After that interruption the relative uncovered time pattern of the "mutant" slightly exceeds the genuine case indicating lower ribosome concentrations for these nucleotides and therefore lower ribosome interaction.

Although the termination is much slower than any tRNA translation rate, fast codons near the termination do have a significant effect, and are not covered by the slowest rate Illustration 11. The

BPEL

ribosome interaction is weaker compared with the solely poly U chain, due to a lower ribosome load because the ribosomes are getting faster to the termination side than leaving the initiation side.

A great effect has the introduction of slow codons in proximity to the termination side as seen in Illustration 12. The graph has the shape of a saturation curve with an always ribosome covered mRNA strand as asymptote. This mRNA is most covered of ribosomes and following the ribosome protection theory, should be most stable against RNase attacks. And indeed, experiments of Gursky et al.[iii] point out that the introduction of slow codons to the 3' end lift the half life and productivity of a mRNA dramatically. Recent theories of mRNA degradation assume a high importance of the RNase E containing degradosome, cf. Carrier et al.[vii], in which RNase E binds to the 5' mRNA end, performs a loop and cuts at specific sites. This contains a 5'-3' cutting preference. If these presumptions are right, results of Gursky et al. are readily explained by the high global ribosome interaction, -queue, induced by
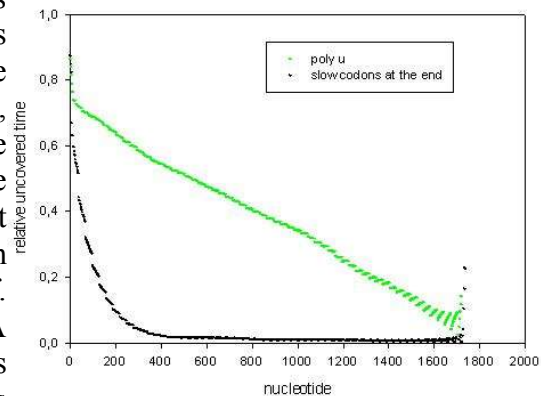

Illustration 12: genuine poly U and poly U with slow codons at termination.

slowed termination. This gives hints about the real termination and initiation parameter and their relativity. The findings of Gursky et al. indicate, that there is no global queue for the examined E. coli cat gene, but can be induced easily by introducing 2 rare and slow codons. This means that the ribosome coverage is in a medium level and sensitive to codon changes at the end. Codons at the end seems to have the highest effect on the global ribosome behavior if initiation and termination are balanced and notedly slower than the slowest tRNA translation rate. A complementary set of experiments, i.e. the influence of slowing silence mutations at the 5' end, were carried out by Deana et al.[xvii]. They observed a drop in the mRNA steady state concentration and a reduced mRNA half life. The shorter half life can be most easily explained in terms of a reduced relative coverage time of the mRNA due to lower ribosome loading originating in a thwarted ribosome clearing of the initiation region. Another decently cause could be an evolving short ribosome queue preceding the slow codons and a relative ribosome free region shortly after, susceptible for RNase attacks. A close examination and weighting of these two findings with the model might give good values for the initiation and termination parameters and are a good proof for the usability of the model.

Further information in "Test of the simulation program with poly U".

From the previous section it becomes essential that tests are performed indicating the dependence of the model to certain parameters in particular values for initiation and termination. The chosen simulation conditions are:

wnt1 mRNA

codon rates: Solomovici et al.[iv], Table 7, column 7(t_{dem})

Ribosome binding and termination rate :

low rates: initiation of v = 0.5 /s,   termination v = 0.5 /s

medium rates: initiation of v = 1.25 /s,   termination v = 2 /s

high rates: initiation of v = 8 /s,   termination v = 5 /s

Ribosome translocation rate (adds to codon rate): 100000 /s

RNA elongation rate per nucleotide: 43 /s (10 /s, 200 /s if stated)

BPEL

Simulation time: 20,000 s

Wherein the low rates represents the presumably lower boundaries and the high rates certainly belongs to a unrealistic high parameter set.
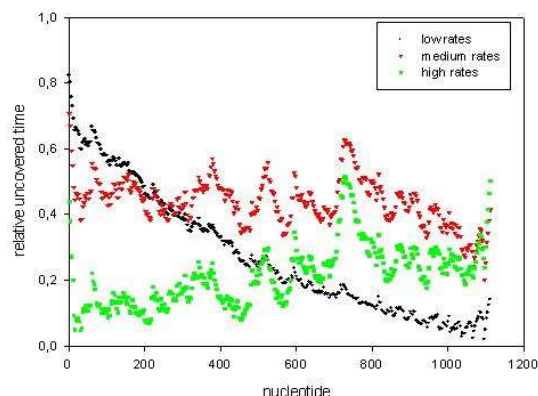

Illustration 13: comparison of different initiation and termination rates, cf. text.

Illustration 13 serves to enlighten the response of the model to the applied parameter sets. As comparable to the poly U case previously discussed the low rate set exhibits a relative uncovered time pattern with ribosome interaction growing linearly to the 3' end. It could may be imagined that this corresponds to a queue of ribosomes, which restrict influences from an averaged codon speed distribution. With that rate conditions changes of rates in the initiation or termination region have global effect on the ribosome interaction.

With medium rates the graph for relative uncovered time remains on one level with high variance basing on the average translation time for 10 codons, equalling the coverage of 1 ribosome. A similar picture is gained for the high rates case, whereas the ribosome interaction gradient rises from 5' to 3' end. This ribosome behaviour will hardly lead to a change of the global ribosome covering structure as will the low case do. Illustration 13 indicates that the effect of different translation speeds for medium and high rates for the ribosome interaction only ranges for limited amount of codons, however, in the middle section of the mRNA the interaction changes for the low rates is weaker. In this set of conditions the results of Gursky et al.[iii] are most easily explained with the low rates, which cause higher ribosome interaction deviations if codon changes are made at the edges of the mRNA. The queue formation is the more restricted the higher rates are applied. The rising rates induces naturally a higher protein yield in the time course of the simulation, cf. table 2. This table also illustrates the statistic of the number of bound ribosomes, and one can recognize first that the equilibrium for the low rates is very noisy, second that the high rates has the highest number of bound ribosomes and third that the medium case has a striking low mean for the number of bound ribosomes, but nonetheless produces more product than the low case with more bound ribosomes.

product release:

	low rates		medium rates		high rates	
time in s	33,8324	**19999,98**	27,846706	**19998,38**	31,94836	**19999,41**
product	1	**6845**	1	**11987**	1	**14657**

number of bound ribosomes:

	low rates	medium rates	high rates
mean	27,0751293	**21,18558506**	**29,4181973**
variance	**50,907706**	7,353733665	2,06160996

Table 2: statistics for the 3 parameter sets.

Differences in the rate of the RNA polymerase do not influence the Ribosome relative covering pattern even if their values differ as much as v=10 s^{-1}, and v=200 s^{-1}. The reason is that the equilibrium forms too fast compared with the 20,000 s simulation run to give the RNA polymerase

BPEL

an impact on the results. Clearly only the release of first protein is concerned by this rate in such long time intervals.
More data in "Sensibility test of different parameter sets".

We will now head for a more comprehensive view of our emulated translation. We have simulated 15 native E. coli genes, while using the supply tRNA reaction rates of Solomovici et al.[iv], a ribosome binding of 0.8 s^{-1} and a termination rate of 2 s^{-1}. The run lasts 20,000 s. The other parameters, which are the same as in other cases, are listed in "Ensemble simulation". We will also focus on correlations between experimental obtained values for mRNA half life and production, obtained by Bernstein et al.[ii], and the relative protein concentration, obtained by van Bogelen et al.[i]. Illustrations 14 - 20 represents the simulated outcome for the relative uncovered time for the 15 genes. It is eye catching that the mean of the genes in the graphs for the applied parameter set are quite stable around a medium value between 0.65 – 0.75 with low variance, also seen in table 3. If a graph is generated reflecting the mRNA half life or the variance of translation speed, basing on the tRNA rates by Solomovici et al., on the abscissa and the variance of the relative uncovered time on the ordinate there is hardly a correlation between them, as seen in Illustration 21. The reason that the half life is not dependent on the used variance is, that this variance is a measurement of the global deviation from the mean, whereas for RNase activity only the local deviation from the mean is of importance. The same accounts for the comparison of the variance of the relative uncovered time and the variance of translation speed, which can have globally different properties than locally. A correlation analyses of the mean translation speed and the product release per time do not exhibit much dependency Illustration 22. From the product release per time an apparent mean translation speed can be obtained, but however as visible from Illustration 22 there is not much correlation with the computed mean tRNA translation rate to be expected. This deviation presumably originates in local queue formations. Illustration 23 presents the logical result that a rise in the amount of ribosome per codon goes along with higher product yield. But the trend curve is admittedly low correlated to the data points, $r^2 = 0.388$.

protein	mean free position	variance in free position	mean translation speed	variance of translation speed
ahpC	0,682	4,396E-03	20,990	60,630
atpD	0,688	2,656E-03	21,084	60,752
carB	0,759	2,608E-03	21,553	60,158
dnaK	0,743	2,502E-03	21,712	52,322
fusA	0,675	2,745E-03	21,185	56,844
grpE	0,729	3,430E-03	21,393	54,351
hns	0,655	6,152E-03	21,088	61,105
nusA	0,673	5,879E-03	21,644	54,502
pnp	0,780	5,174E-03	21,597	57,912
rpoA	0,669	4,569E-03	20,603	65,850
rpoB	0,647	5,927E-03	20,934	60,597
rpsA	0,720	1,921E-03	21,952	50,997
tig	0,703	2,512E-03	22,187	51,094
trxB	0,643	6,245E-03	21,069	68,952
tufA	0,661	5,474E-03	22,019	54,236

Table 3: Comparison of relative uncovered time and translation speed characteristics for the 15 genes.

A deeper insight into this analysis can be gained from "Ensemble simulation".
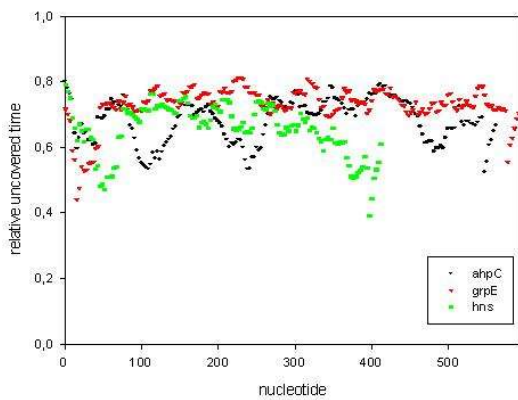
BPEL

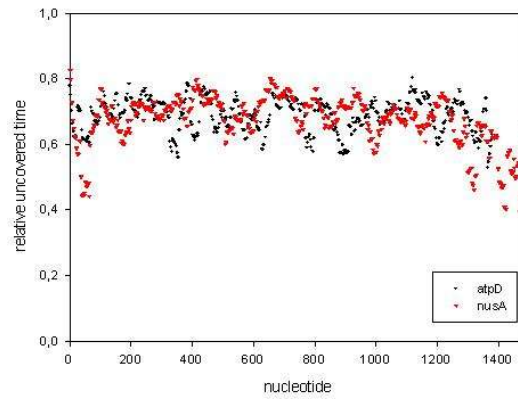*Illustration 14: relative uncovered time for E. coli genes ahpC, grpE and hns.*

Illustration 15: relative uncovered time for E. coli genes atpD and nusA.

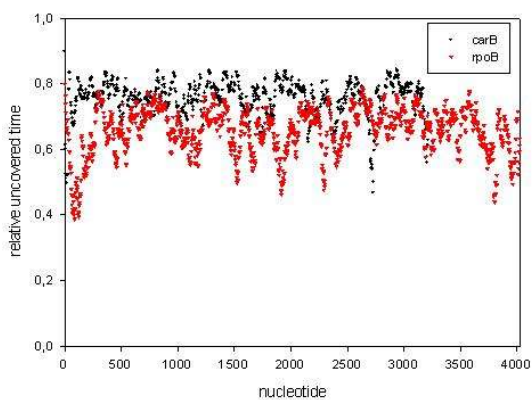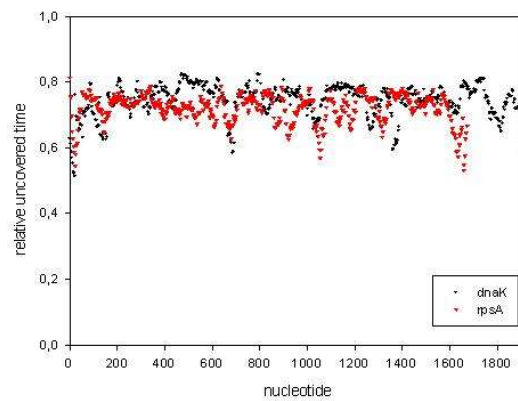*Illustration 16: relative uncovered time for E. coli genes carB and rpoB.*

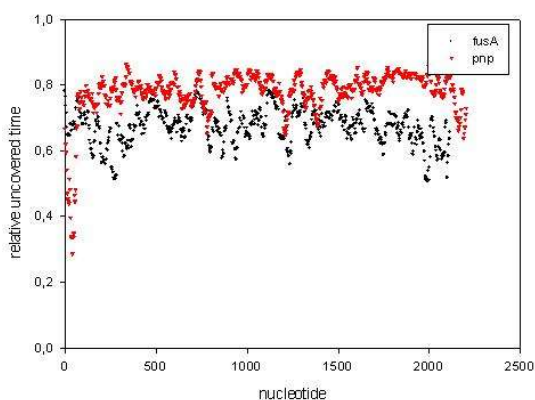*Illustration 17: relative uncovered time for E. coli genes dnaK and rpsA.*

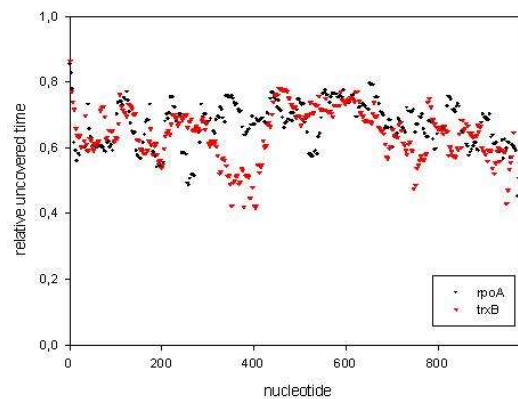*Illustration 18: relative uncovered time for E. coli genes fusA and pnp.*

Illustration 19: relative uncovered time for E. coli genes rpoA and trxB.
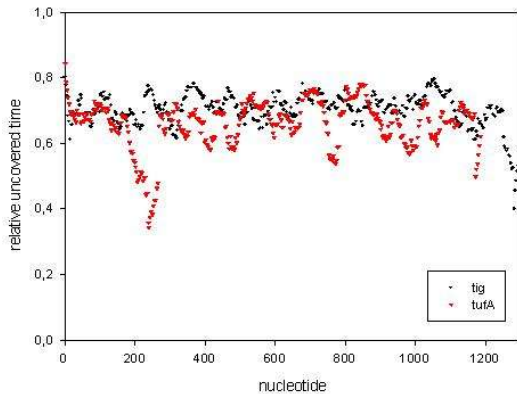
triacle
biocomputing

BPEL

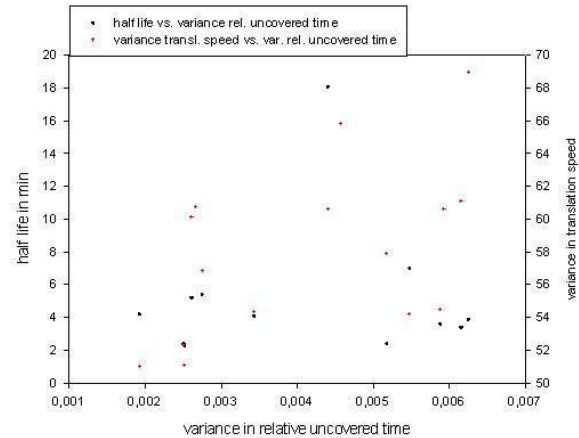Illustration 20: relative uncovered time for E. coli genes tig and tufA.

Illustration 21: dependence of mRNA half life and variance in translation speed to the variance of relative uncovered time. Half lifes from Bernstein et al., translation speeds from Solomovici et al.
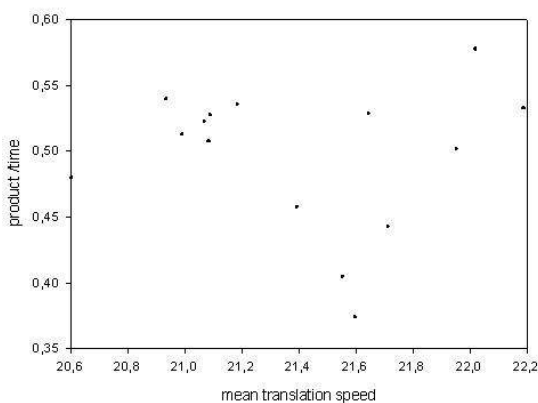

Illustration 22: dependence of product per time release to mean translation speed. Translation speeds from Solomovici et al.
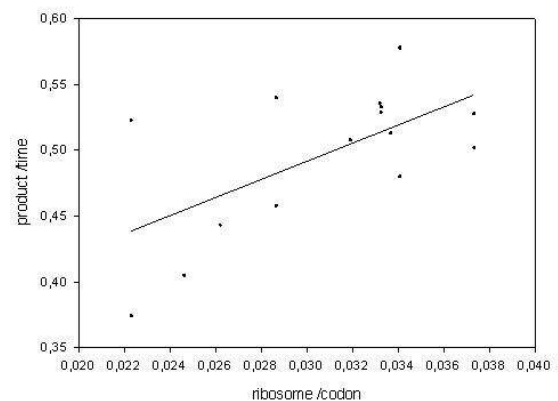

Illustration 23: correlation between product release per time and ribosomes per codon. r² = 0.388

A number of studies were performed with recombinant mRNAs in which codon changes were introduced and the effect on the model is observed. The first case to be presented handles with wnt4, a gene in the developmental cascade of eukaryotes, which is under investigation for effective recombinant production in E. coli. The goal here is to interpret the relative uncovered time pattern and to alter it in a way that a RNase activity to the mRNA might be reduced. A reduced RNase activity on the mRNA is assumed to arouse if the coverage of the mRNA with ribosomes rises, therefore the pattern for the relative uncovered time approximates the shape of the inverse saturation curve seen for the poly U case with slow codons near the termination, or fast at the beginning, Illustrations 7&12. The simulation was performed with the demand tRNA translation times of Solomovici et al., an initiation and binding rate of 0.5 s⁻¹, a simulation time of 6000 s and the other parameters which remain mainly the same like RNA polymerase with $v = 43$ s⁻¹, a Ribosome footprint of 30 nucleotides and a very fast translocation rate of $v = 100000$ s⁻¹. To obtain a more favorable ribosome interaction behavior than in the genuine case, a slow codon at position 3 was changed into a fast one, and at the end at position 345 a fast codon was changed into a slow one.

triacle
biocomputing

BPEL

program validation

Ulf Liebal
19.07.2004 – 22.10.2004

page: 14

The results can be seen in Illustration 24, and they elucidate that the best drop in the curve can be achieved while installing a slow end.
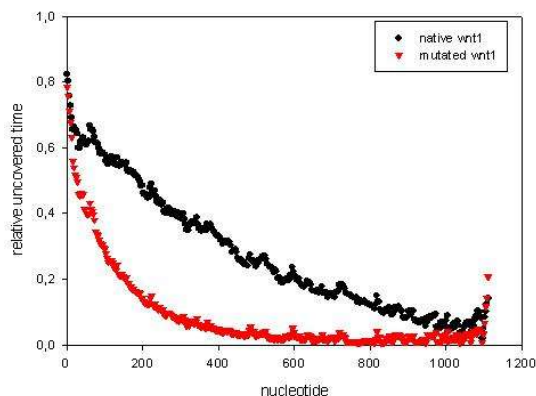Details in "Case study wnt4".

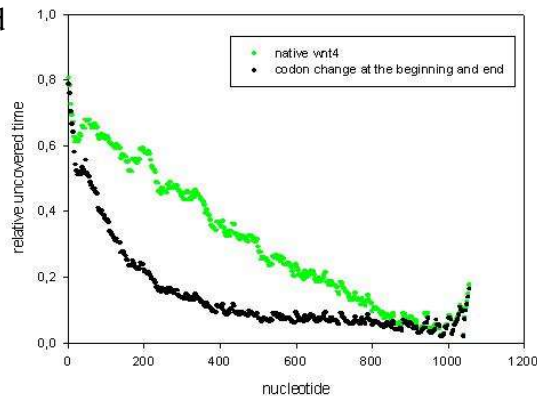Illustration 24: comparison of native wnt4 and optimized wnt4.

Illustration 25: comparison of native and optimized wnt1.

The same procedure, with the very much parameters as in the wnt4 case, was applied for wnt1. More Mutations are suggested and the effect is reproduced in Illustration 25 The relative uncovered time is lowered and might therefore hobble RNase attacks.

A little bit more can be read in "Case study wnt1".

The report "Case study c1a1" nicely illustrates how easily expectations and, at the first glance, positive results can lead to mistaken derivations. The codon amount of the mRNA is very high, nearly 1500 codons. Even if the number of bound ribosomes has seemingly reached equilibrium this is not necessarily the case for the relative uncovered time pattern, due to the large amount of codons. In the Statistical analysis previously described and more detailed in report "Statistical analysis" the wnt1 gene is used as basis for analysis, which has significantly fewer codons. Apparently the high codon amount and the relative short simulation time has lead to the huge alteration in the program outcome. This can be deduced with the graph of the relative uncovered time where the splitting of the curve first occurs at nucleotide 1000 what is definitely far after the changed position of 24. The impressive feature in that report is only the tempt-ability of mind.
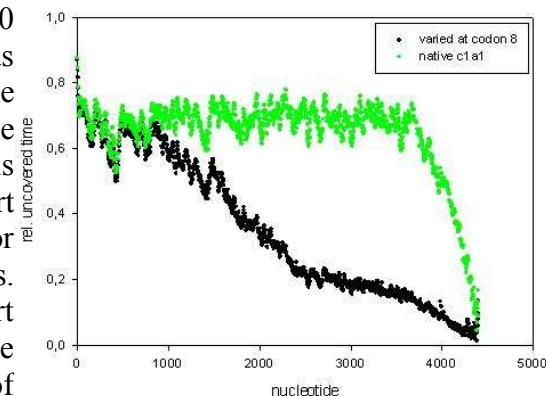

Illustration 26: probable statistical deviations of two simulation runs.

The case study with c3a1, known as 309, seems to be statistically more sure. The codon length is with 215 rather small, and the simulation was run for a time scale of 20,000 s. Apart from that, the other parameters remained the same, i.e. demand translation times, balanced initiation and termination with rate 0.5 s^{-1}, other parameters in "Case study c3a1". It points out that even the dramatic change of 11 codons does not alter the ribosome coverage in dramatic way, Illustration 27.
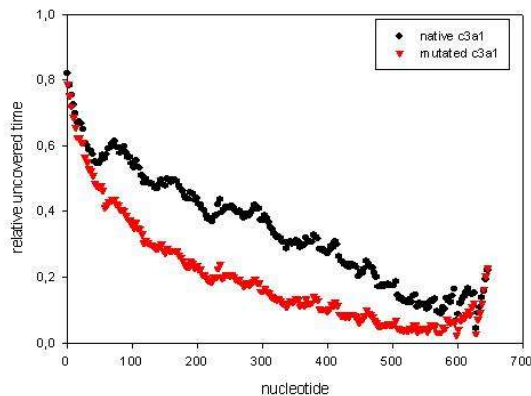
BPEL

Illustration 27: native and optimized c3a1 gene.

The His-codon row in the beginning is quite slow and rather restricting for attempts of codon speed engineering, since all alternative His-codons are slow. Caused by the introduction of slow codons the yield of protein drops.

A more extensive view in "Case study c3a1".

Parameters used so far are more or less restricting for the influence of all codons positioned in the middle of the mRNA, and only codons near initiation or, perhaps more important, termination effects the ribosome coverage. If one assumes the RNase E activity with loop formation as valid, Carrier et al.[vii], it seems also very likely that local ribosome queues determine the functional half life of a mRNA, Yarchuk et al.[xii]. But local jamming of ribosomes is very weak in a parameter set with balanced initiation and termination at a low rate of 0.5 s[-1], since they are covered mainly by global jamming. To account for a situation with emphasized local ribosome interaction the parameters are changed similar to the reported ones in "Sensibility test of different parameter sets" to 0.8 s[-1] for initiation as in Arnold[xi] and 2 s[-1] for termination. Still the demand translation times of Solomovici et al.[iv] are used, the simulation lasts 20,000 s, the RNA polymerase works with speed of 43 nucleotides per second, the ribosome covers 10 codons and the translocation rate is very fast with v = 100000 s[-1]. In "Case study 309/II" the approach for a possibly optimized sequence is derived. However, the goal in optimizing with that medium rates is to smoothen the appearance of the relative uncovered time function to minimize clefts, which might be susceptible for RNase attacks. Illustration 28 shows the relative uncovered time pattern for the chosen conditions. It is visible that the ribosome interaction, which might be analogous to the uncovered time behavior, oscillates around a certain value, with many peaks and hollows.

Illustration 28: relative uncovered time for native and optimized c3a1 with initiation of 0.8 /s and termination of 2 /s.

The red line indicates an optimization attempt with 9 mutations. The product statistic is listed in table 4 and hints that more product is released for the mutated mRNA. This is probably due to the introduction of 6 faster codon whereas only 3 codons were engineered to be slower. The gain in translation speed rises the product per time release and therefore the productivity. Yet, the validity of the approach has to be justified by experiments.
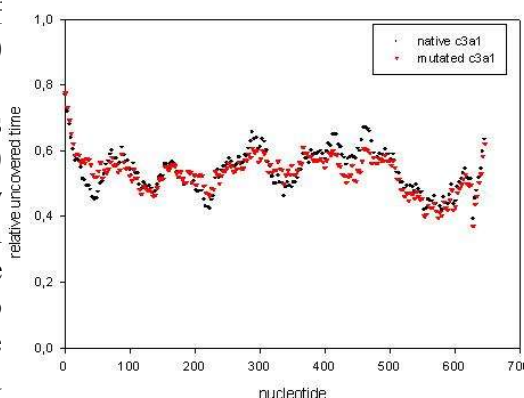
	time	product
native	20,672	1
	19998	9710
mutant	17,833	1
	19999	9938

Table 4: appearance of first product, total yield for native and mutant.

Read more in "Case study 309/II".

BPEL

Another interesting feature is the over expression of rare tRNAs with the intuition that a higher concentration of these tRNAs will lead to a higher translation speed. This is incorporated of the translation speed derivation of Sol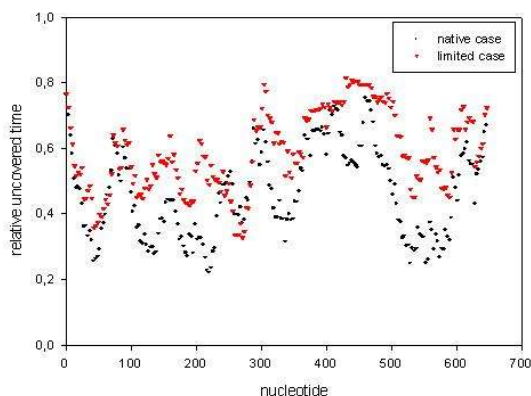omovici et al.[iv], where one of the basic assumption is the correlation between speed and abundance, cf. the objective part. This newly adopted tRNA translation speeds are then thought to provide a more smooth translation. In the model we assume that the tRNAs are rate limiting in the translation process, Solomovici et al. [iv], and that there are only very few available for a translating ribosome. Furthermore, the time in which a mismatching tRNA is bound to the ribosome is very short and does not interfere with the binding of other tRNAs. As a consequence the only parameter, determining the translation speed of a tRNA is solely the concentration of the matching tRNA. This consideration is important when analyzing the effect of over expressed tRNAs, whether this manipulation which changes the global tRNA relations also alters the

Illustration 29: comparison of the limited case of the over expression of tRNAs and a native case.

translation speeds of all other tRNAs. Investigations were also made with global changing translation rates but we emphasize the limited speed change. In "Analysis of over expression" the simulation result is analyzed for the global situation, the limited situation, the mutated situation and the situation for a mixed mutation and global case. The parameters given to the program are the supply translation times of Solomovici et al., a ribosome binding rate of 0.8 s^{-1}, a termination rate of 2 s^{-1}, a simulated time of 20,000 s, the other usual parameters are listed in "Analysis of over expression". A comparison of the limited case, where only the speeds of the tRNA with enriched concentration rise, with the native case shows that the mRNA is globally covered with fewer ribosomes as seen in Illustration 29, the relative uncovered time is mostly above the nature case, and the value for ribosomes per nucleotide is significantly lower than in the native case, table 5. Additionally the program reports a higher product per time release for the limited case, table 5, due to the risen speed of 6 codons, the sum of the translation speeds of all codons of the mRNA is risen and therefore the ribosome translating time drops. The reported success of tRNA over expression can be explained in the limited case with a higher productivity.
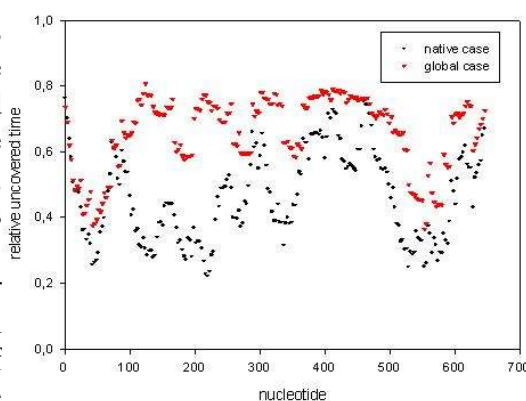

Illustration 30: comparison of the global case of the overexpression of tRNAs and a native case.

The global case, where the speed of over expressed tRNAs rises, but for the other tRNAs goes down, because of lower tRNA relative concentrations, shows another picture of the scheme. The relative uncovered time pattern in Illustration 30 seems to be less rough, albeit on a high uncovered level. The product release per time is hardly changed, indicating that the sum of translating speeds for the codons is similar. The main difference is the ribosome coverage which is below the ribosome per codon value as for the native case. Altogether the lower ribosome coverage and a constrained product per time release indicates a smoother translation as in the native case, this

triacle
biocomputing

BPEL

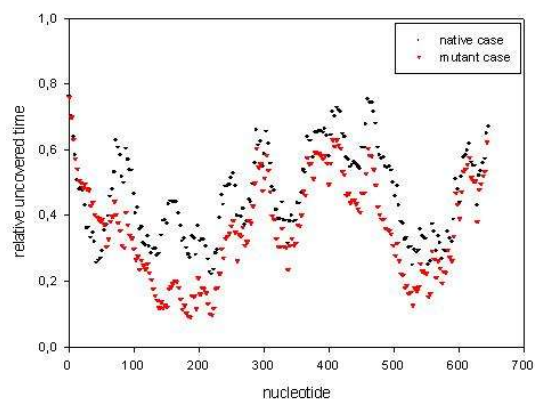would be the only explanation for higher yield in the global case.

Illustration 31: comparison of a mutation optimization and a native case.

The mutations were now engineered with the rationality to erase first peaks which might be the cutting points for a looping RNase E. This was established by introduction of 9 silent mutations, where 8 codons becomes faster and one slower. The effect for the relative uncovered time is represented in Illustration 31. The product per time release is in the magnitude of the limited case, presumably owing to the very same reason, i.e. the up speeding by the mRNA translation speed sum. The aimed mutations lead to a higher ribosome coverage, reproduced by the ribosome per codon measurement in table 5, and consequently to more ribosome interaction and hindered RNase attacks.

A more complete discussion can be found in "Analysis of over expression".

	native case	limited case	global case	mutation case	global + mutation case
	product	product	product	product	product
20,000 s	8019	8740	8186	8692	8952
time/codon	0,121	0,077	0,084	0,101	0,087
product/time	0,401	0,437	0,409	0,435	0,448
statistics of number of bound ribosomes					
mean	11,801	9,080	7,630	13,966	8,285
variance	3,481	3,104	2,514	4,262	3,527
rib/codon	0,0549	0,0422	0,0355	0,0650	0,0385

Table 5: Influence of the different improvement procedures.

In "Codon translation speed analysis" we try to find evolutionary clues which might have directed the codon usage in the initiator region. This might be the case as the translation speed in the initiation region partly determines the initiation efficiency and the ribosome load of the mRNA. Another hint comes from Lesnik et al.[xviii]. If there is ribosome jamming it might defined as a situation in which the ribosome active side has exactly a distance to the next side of the ribosome footprint. If so, the waiting ribosomes can not translate the codons with optimal translation speed, but they can only advance if the slowest ribosome at the queue top advances. The thought is now that in such a case codons with the exact ribosome
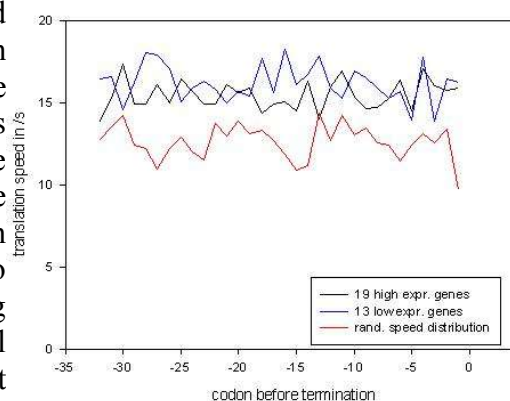

Illustration 32: codon speed statistics for the termination preceding codons. Termination =0. Translation speeds demand after Solomovici et al.

covering distance to another are not codon speed optimized. This should be visible if the codon speeds of gene ensembles are plotted against their position. Because the termination is rather slow also there one might expect tracks. The original goal was to get a value for the ribosome coverage, which is differing in the literature. As Illustrations 32 - 34 indicate there are no slower codons than the random distribution and the variance of the speed is lower for high expressed genes and slightly higher for low expressed genes than a random speed distribution, table 6. This is not only true for the demand translation times of Solomovici et al. but as well as for the supply times. The distinction of high and low expressed genes shows with the selected genes that low expressed genes starts with faster codons and high expressed genes have somehow lower termination codons.

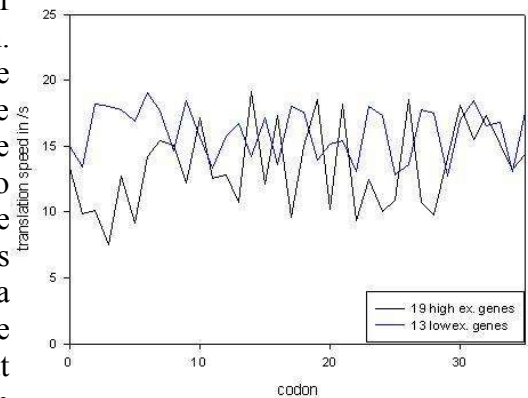The analyze report is made in "Codon translation speed analysis" and "Ensemble Simulation".

Illustration 33: codon speed statistics for the first 30 codons of high and low expressed genes. translation speeds demand after Solomovici et al.

statistic for the last 30 codons	high expr. genes	low expr. genes	rand. speed distr.
mean transl. speed demand	15.4477 s	16.184 s	12.6106 s
variance	0.7658	1.2629	1.1287

Table 6: Solomovici et al. demand translation values applied for the last 30 codons of high and low expressed genes.
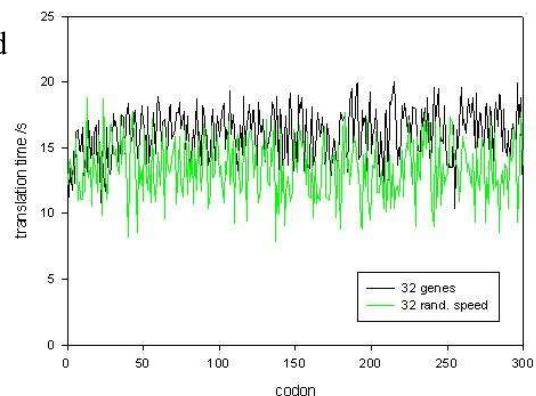

Illustration 34: codon speed statistics for a gene selection and a randomly distributed speed. Translation speeds demand from Solomovici et al.

An interesting insight into the working procedure of the program provides Illustration 35. If certain rates for initiation and termination are taken, that is an initiation of 0.8 s^{-1} and a termination of 2 s^{-1}, then the simulation outcome resembles obviously an averaged translation speed over the ribosome footprint of 30 nucleotide on the mRNA.
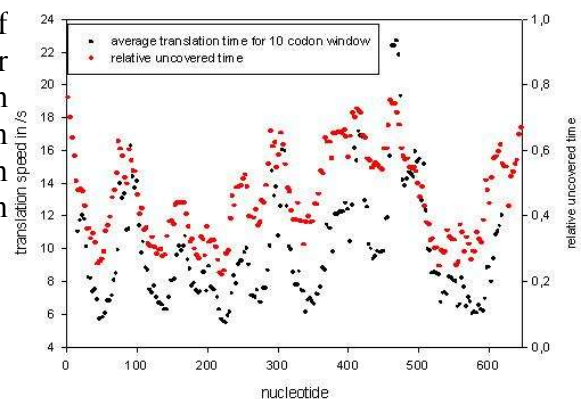

Illustration 35: similarities between the computed relative uncovered time of the simulation program and a 30 nucleotide averaged translation speed. Translation speed values from Solomovici et al.

[i] Neidhardt, Escherichia coli and Salmonella, 1996-2nd ed,, vanBogelen et al., pg: 2101

[ii] Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. Proc Natl Acad Sci U S A. 2002 Jul 23;99(15):9697-702. Epub 2002 Jul 15

[iii] Gursky YG, Beabealashvilli RSh., The increase in gene expression induced by introduction of rare codons into the C terminus of the template., Gene. 1994 Oct 11;148(1):15-21

[iv] Solomovici J, Lesnik T, and Reiss, C. Does Escherichia coli Optimize the Economics of the Translation Process?. J. theor. Biol. 1997; 185, 511-521

[v] Dong H, Nilsson L, Kurland CG. Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. J Mol Biol. 1996 Aug 2;260(5):649-63

[vi] Ikemura. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes. J Mol Biol. 1981 Feb 15;146(1):1-21

[vii] Carrier TA, Keasling JD. Mechanistic modeling of prokaryotic mRNA decay. J Theor Biol. 1997 Nov 21;189(2): 195-209

[viii] Gillespie D.T. Exact Stochastic Simulation of Coupled Chemical Reactions. J. Phys. Chem., 1977; 81:2340-2361

[ix] Kierzek AM, Zaim J, Zielenkiewicz P. The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression. J Biol Chem. 2001 Mar 16;276(11):8165-72. Epub 2000 Nov 02

[x] Neidhardt, Escherichia coli and Salmonella, 1996-2nd ed., Richardson, Greenblatt, control of chain elongation and termination, pg.: 822

[xi] Arnold, Sabine. Kinetic Modelling of Gene Expression. Dissertation, Fakultät Maschinenbau, Institut für Bioverfahrenstechnik, Stuttgart, 30.10.2003

[xii] Yarchuk O, Jacques N, Guillerez J, Dreyfus M., Interdependence of translation, transcription and mRNA degradation in the lacZ gene. J Mol Biol. 1992 Aug 5;226(3):581-96

[xiii] Gromadski KB, Rodnina MV. Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. Mol Cell. 2004 Jan 30;13(2):191-200

[xiv] Stryer L., Biochemistry, 1995, 4th ed., pg.: 895

[xv] Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y, The complete genome sequence of Escherichia coli K-12, Science. 1997 Sep 5;277(5331):1453-74

[xvi] Freistroffer DV, Pavlov MY, MacDougall J, Buckingham RH, Ehrenberg M., Release factor RF3 in E.coli accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner, EMBO J. 1997 Jul 1;16(13):4126-33

[xvii] Deana A., Ehrlich R., Reiss C., Silent mutations in the Escherichia coli ompA leader peptide region strongly affects transcription and translation in vivo, Nucl. Ac. Res., 1998, Vol. 26, No. 20, 4778-4782

[xviii] Lesnik T., Solomovici J., Deana A., Ehrlich R., Reiss C., Ribosome Traffic in E. coli and Regulation of Gene Expression, J. theor. Biol. 2000; 202, 175 -185