In Silico Modelling and Analysis of Ribosome Kinetics and aa-tRNA Competition

D. Bošnački[1,⋆], T.E. Pronk[2,⋆⋆], and E.P. de Vink[3,⋆⋆⋆]

[1] Dept. of Biomedical Engineering, Eindhoven University of Technology
[2] Swammerdam Institute for Life Sciences, University of Amsterdam
[3] Dept. of Mathematics and Computer Science, Eindhoven University of Technology

Abstract. We present a formal analysis of ribosome kinetics using probabilistic model checking and the tool Prism. We compute different parameters of the model, like probabilities of translation errors and average insertion times per codon. The model predicts strong correlation to the quotient of the concentrations of the so-called cognate and near-cognate tRNAs, in accord with experimental findings and other studies. Using piecewise analysis of the model, we are able to give an analytical explanation of this observation.

1 Introduction

The translation mechanism that synthesizes proteins based on mRNA sequences is a fundamental process of the living cell. Conceptually, an mRNA can be seen as a string of codons, each coding for a specific amino acid. The codons of an mRNA are sequentially read by a ribosome, where each codon is translated using an amino acid specific transfer-RNA (aa-tRNA), building one-by-one a chain of amino acids, i.e. a protein. In this setting, aa-tRNA can be interpreted as molecules containing a so-called anticodon, and carrying a particular amino acid. Dependent on the pairing of the codon under translation with the anticodon of the aa-tRNA, plus the stochastic influences such as the changes in the conformation of the ribosome, an aa-tRNA, arriving by Brownian motion, docks into the ribosome and may succeed in adding its amino acid to the chain under construction. Alternatively, the aa-tRNA dissociates in an early or later stage of the translation.

Since the seventies a vast amount of research has been devoted, unraveling the mRNA translation mechanism and related issues. By now, the overall process of translation is reasonably well understood from a qualitative perspective. The translation process consists of around twenty

⋆ Supported by FP6 LTR ESIGNET.
⋆⋆ Funded by the BSIK project Virtual Laboratory for e-Science VL-e.
⋆⋆⋆ Corresponding author, e-mail evink@win.tue.nl.

small steps, a number of them being reversible. For the model organism *Escherichia coli*, the average frequencies of aa-tRNAs per cell have been collected, but regarding kinetics relatively little is known exactly. Over the past few years, Rodnina and collaborators have made good process in capturing the time rates for various steps in the translation process for a small number of specific codons and anticodons [21, 23, 24, 12]. Using various advanced techniques, they were able to show that the binding of codon and anticodon is crucial at a number of places for the time and probability for success of elongation. Based on these results, Viljoen and co-workers started from the assumption that the rates found by Rodnina et al. can be used in general, for all codon-anticodon pairs as estimates for the reaction dynamics. In [9], a complete detailed model is presented for all 64 codons and all 48 aa-tRNA classes for *E. coli*, on which extensive Monte Carlo experiments are conducted. In particular, using the model, codon insertion times and frequencies of erroneous elongations are established. Given the apparently strong correlation of the ratio of so-called near-cognates vs. cognate and pseudo-cognates, and near-cognates vs. cognates, respectively, it is argued that competition of aa-tRNAs, rather than their availability decides both speed and fidelity of codon translation.

In the present paper, we propose to exploit abstraction and model-checking of continuous-time Markov chains (CTMCs) with Prism [18, 13] for the case of mRNA translation. The abstraction conveniently reduces the number of states and classes of aa-tRNA to consider. The tool provides built-in performance analysis algorithms and path chasing machinery, relieving its user from mathematical calculations. The outcomes are exact, unlike approximations obtained by simulation. More importantly, from a methodological point of view, the incorporated CSL-logic [2] allows to establish quantitative results for parts of the system, e.g. for first-passage time for a specific state. Such piecewise analysis proves useful when explaining the relationships suggested by the data collected from the model. Additionally, in our case, the Prism tool enjoys rather favourably response times compared to simulation.

Related work Measurements in *E. coli* by Sørensen et al. [25] already suggested dependence of the availability of various codons, qualified as 'rare' or 'frequent', and translation rates. Wahab c.s. [27] noted that in *E. coli* strains expressing high levels of $tRNA_1^{Leu}$ isoacceptor, an increase of available *tRNA* led to a decrease of protein production. The present investigation started from the Monte-Carlo experiments of mRNA trans-

lation reported in [9]. A similar stochastic model, but based on ordinary differential equations, was developed in [14]. It treats insertion times, but no translation errors. The model of mRNA translation in [10] assumes insertion rates that are directly proportional to the mRNA concentrations, but assigns the same probability of translation error to all codons.

Currently, there exist various applications of formal methods to biological systems. A selection of recent papers from model checking and process algebra includes [22, 5, 6]. More specifically pertaining to the current paper, [4] applies the Prism model checker to analyze stochastic models of signaling pathways. Their methodology is presented as a more efficient alternative to ordinary differential equations models, including properties that are not of probabilistic nature. Also [13] employs Prism on various types of biological pathways, showing how the advanced features of the tool can be exploited to tackle large models.

In [3], we use the model presented in this paper to perform a formal analysis of amino acid replacement during mRNA translation. Building on the abstract stochastic model of arrival of tRNAs and their processing at the ribosome presented in the sequel, we compute probabilities of the insertion of amino acids into the nascent polypeptide chain. This allows us to construct the substitution matrix containing the probabilities of an amino acid replacing another. Finally in [3], we discuss the analogy of this matrix with standard mutation matrices, and analyze the mutual replacement of biologically similar amino acids. The main contribution of the present paper is the study of the underlying model of mRNA translation itself, exploiting probabilistic model checking and the approach of piecewise analysis.

Organization of the paper Section 2 provides the biological background, discussing the mRNA translation mechanism. Its Prism model is introduced in Section 3. In Section 4, it is explained how error probabilities are obtained from the model and why they correlate with the near-cognate/cognate fraction. This involves adequate estimates of specific stochastic subbehaviour. Insertion times are the subject of Section 5. There too, it is illustrated how the quantitative information of parts of the systems is instrumental in deriving the relationship with the ratio of pseudo-cognate and near-cognates vs. cognates.[1]

[1] An appendix presents supplementary data.

Acknowledgments We are grateful to Timo Breit, Christiaan Henkel, Erik Luit, Jasen Markovski, and Hendrik Viljoen for fruitful discussions and constructive feedback.

2 A kinetic model of mRNA translation

In nature, there is a fixed correspondence of a codon and an amino acid. This is the well-known genetic code, that couples all 61 relevant codons to 20 fundamental amino acids. The three codons not corresponding to an amino acid are so-called stop codons, that guide the termination of the translation process. Thus, an mRNA, as sequence of codons, codes for a unique sequence of amino acids, i.e. protein. However, the match of a codon and the anticodon of a tRNA is different from pair to pair. The binding influences the speed of the actual translation. Here, we give a brief overview of the translation mechanism. Our explanation is based on [23, 17]. The basic idea is that mRNA is transcribed from the cellular DNA. A ribosome, an enzyme catalyzing translation, attaches to an individual mRNA and starts translating the sequence of codons into amino acids. The ribosome processes one codon at the time by recruiting aa-tRNA from the cell. Dependent on the match of the codon at the mRNA and the anticodon of the aa-tRNA, the amino acid carried by the aa-tRNA is transfer to the polypeptide chain under construction, i.e. the nascent protein. Two main phases can be distinguished: peptidyl transfer and translocation.

The peptidyl transfer phase runs through the following steps. aa-tRNA arrives at the A-site of the ribosome-mRNA complex by diffusion in a ternary complexation with elongation factor *Tu* (*EF-Tu*) and *GTP* at a rate determined by the interaction of *EF-Tu* and the ribosome. The initial binding is relatively weak. Codon recognition comprises (i) establishing contact between the anticodon of the aa-tRNA and the current codon in the ribosome-mRNA complex, and (ii) subsequent conformational changes of the ribosome. Given a codon, an anticodon can either be a cognate, a near-cognate or a non-cognate. As an aa-tRNA carries precisely one anticodon, we also speak of cognate, near-cognate and non-cognate aa-tRNA. The rates of confirmation of the ribosome are different for cognate and near-cognates. This does not apply to non-cognate anticodons; the aa-tRNA that carries it, dissociates from the ribosome almost immediately. *GTP*ase-activation of the elongation factor *EF-Tu* is largely favoured, because of the conformational changes, in case of a strong complementary matching of the codon and a cognate anticodon.

Otherwise, *GTP*ase-activation is lessened. After *GTP*-hydrolysis, producing inorganic phosphate P_i and *GDP*, the affinity of the ribosome for the *aa-tRNA* reduces. The subsequent accommodation step also depends on the fit of the *aa-tRNA*. Accomodation happens rapidly for cognate *aa-tRNA*, whereas for near-cognate *aa-tRNA* this proceeds slower and the *aa-tRNA* is likely to be rejected.

The subsequent translocation phase will shift the peptide chain in nascent and the *mRNA* including the codon just processed, exposing a new codon for translation and releasing the A-site for the arrival of another *aa-tRNA*. The first step of the translocation phase involves the association of *GTP*. By GTP-hydrolysis of elongation factor *EF-G*, *GDP* and P_i are produced. This results in unlocking and movement of the *aa-tRNA* to the P-site of the ribosome. The latter step is preceded or followed by P_i-release, with *GDP* bound or unbound to P_i, respectively. Reconformation of the ribosome and release of *EF-G* moves the *tRNA*, that has transferred its amino acid to the polypeptide chain, into the E-site of the ribosome. Further rotation eventually leads to dissociation of the used *tRNA*.

Although overall qualitatively well understood, there is, at present, limited quantitative information regarding the translation mechanism and its individual steps. For *E. coli*, a number of specific rates have been collected by Rodnina and co-workers [23, 12]. Some steps are known to be relatively rapid. The fundamental assumption of [9], that we also adopt here, is that experimental data found for the *UUU* and *CUC* codons, regarding their matching to *Phe-tRNA*, extrapolate to other codon-anticodon pairs as well. However, further assumptions are necessary to fill the overall picture. In particular, Viljoen proposes to estimate the delay due to so-called non-cognate *aa-tRNA*, that are blocking the ribosomal A-site, at 0.5ms. Also, accurate rates for the translocation phase are largely missing. Again following [9], we have chosen to assign, if necessary, high rates to steps for which data is lacking. This way these steps will not be rate limiting.

An overview of the reactions involving cognates and near-cognates and the corresponding rates are collected in Table 1. The upper and lower parts of Table 1 correspond respectively to reactions involving cognate and near-cognate tRNAs and they differ only in the reaction rates. The first reaction step represents the arrival of the ternary complex C (N) (*aa-tRNA*, EF-Tu, and GTP) to the ribosome *R1* and the initial binding between those. Subsequent reactions correspond to various conformational changes of the ribosome-mRNA complex. The first selection step happens by means of the inverse reaction that transforms *CR3* (*NR3*)

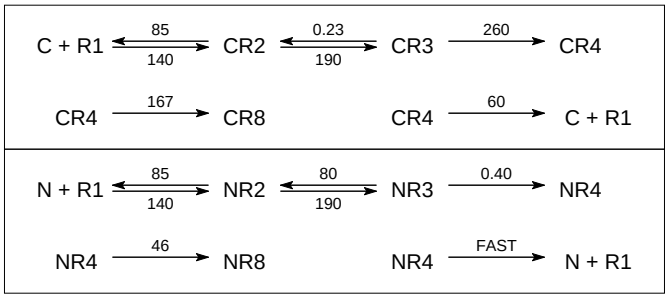

Table 1. Molecular reactions underlying the adapted model. Rates taken from the model in [9] based on experiments reported in [23]. See the main text for an explanation of the individual reactions. Rates for the reactions *CR4* and *NR4* are obtained by merging subsequent reactions, as discussion in Section 3.

back to *CR2* (*NR2*). Because of the higher rate the near- and non-cognate aa-tRNAs have much greater chance to be rejected than the cognate ones. The one way reaction from *CR3* (*NR3*) to *CR4* (*NR4*) includes the GTP hydrolysis step which means that the *aa-tRNA* has passed the first selection test. Reactions from the second row correspond to the proofreading step in which either *aa-tRNA* is definitely accepted and the corresponding amino acid incorporated into the polypeptide chain —represented by conformation state *CR8*— or it is rejected which results in disassociation of the *aa-tRNA*from the ribosome.

We will fit the above model of protein synthesis in the language of the Prism model checker. The experiments confirm the main results of [9], viz. (i) insertion errors are proportional to the quotient of the frequencies of near-cognates and of cognates, (ii) *aa-tRNA* competition better predicts insertion times than *aa-tRNA* availability. In fact, we show in the latter case the stronger result that the ratio of near-cognate and cognate frequencies is an adequate estimate for insertion time. In addition, for the above results, we are able to actually derive the correlation with the quotient of near-cognates and cognates. It is the availability of an explicit model, together with the possibility to obtain, by model checking, quantitative information for parts of the system, that lead to a sharper analysis of the experimental data, that cannot obtained by simulation of a monolithic model.

3 The Prism model

The model employed in the analysis below is an abstraction of the biological model as sketched in the previous section. The abstraction is

twofold: (i) Instead of dealing with 48 individual classes of *aa-tRNA*, that are identified by the their anticodons, we restrict to four types of *aa-tRNA* distinguished by their matching with the codon under translation. (ii) We combine various detailed steps into one transition by accumulation of rates. The first reduction greatly simplifies the model, more clearly eliciting the essentials of the underlying process. The second abstraction is more a matter of convenience, though it helps in compactly presenting the model.

For a specific codon, we distinguish below four types of *aa-tRNA*: cognate, pseudo-cognate, near-cognate, non-cognate. Cognate *aa-tRNA*s have an anticodon that strongly couples with the codon. The amino acid carried by the *aa-tRNA* is always the right one, according to the genetic code. The binding of the anticodon of a pseudo-cognate *aa-tRNA* or a near-cognate *aa-tRNA* is weaker, but sufficiently strong to occasionally result in the addition of the amino acid to the nascent protein. In case the amino acid of the *aa-tRNA* is, accidentally, the right one for the codon, we dubb the *aa-tRNA* of the pseudo-cognate type. If the amino acid does not coincide with the amino acid the codon codes for, we speak in such a case of a near-cognate *aa-tRNA*.[2] The match of the codon and the anticodon can be very poor too. We refer to such *aa-tRNA* as being non-cognate for the codon. This type of *aa-tRNA* does not initiate a translation step at the ribosome.

The Prism model can be interpreted as the superposition of four stochastic automata, each encoding the interaction of one of the types of *aa-tRNA*. The automata for the cognates, pseudo-cognates and near-cognates are very similar; the cognate type automaton only differs in its value of the rates from those for pseudo-cognates and near-cognates, while the automata for pseudo-cognates and for near-cognates only differ in their arrival process. The automaton for non-cognates is rather simple. See Figure 1.

Below, we are considering average transition times and probabilities for reachability based on exponential distributions. Therefore, following common practice in performance analysis, there is no obstacle to merge two subsequent sequential transitions with rates λ and μ, say, into a combined transition of rate $\lambda\mu/(\lambda + \mu)$. This way, a smaller model can be obtained that, although differently distributed, is equivalent to the

[2] The notion of a pseudo-cognate comes natural in our modeling. However, the distinction between a pseudo-cognate and a near-cognate is non-standard. Usually, a near-cognate refers to both types of tRNA.

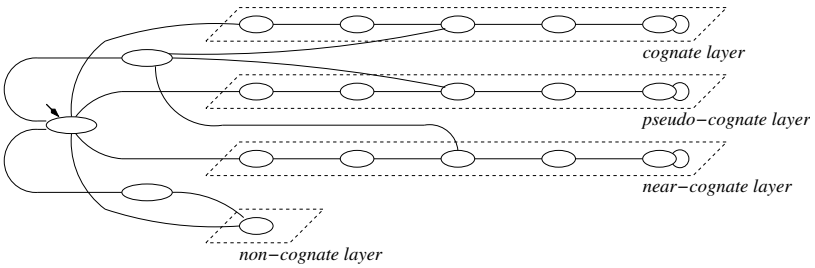

Fig. 1. Overview of Prism model as superposition of four *aa-tRNA* typed automata. Each layer models the processing of a specific type of *aa-tRNA*, viz. cognate, pseudo-cognate, near-cognate and non-cognate *aa-tRNA*.

original regarding expected values. However, it is noted, that in general, such a simplification is not compositional and should be taken with care.

We briefly discuss the Prism code implementing the abstract model. For the modeling of continuous-time Markov chains, Prism commands have the form

$$[\texttt{label}]\ \texttt{guard} \rightarrow \texttt{rate}\colon\ \texttt{update}\ ;$$

In short, from the commands whose guards are fulfilled in the current state, one command is selected proportional to its relative rate. Subsequently, the update is performed on the state variables. So, a probabilistic choice is made among commands. Executing the selected command results in a progress of time according to the exponential distribution for the particular rate. Labels are used to synchronize Prism commands, a feature not used in this paper. We refer to [18, 13] for a proper introduction to the Prism model checker.

Initially, control resides in the common start state `s=1` of the Prism model with four boolean variables `cogn`, `pseu`, `near` and `nonc` set to false.

```
s    : [0..8] init 1 ;
cogn : bool init false ;
pseu : bool init false ;
near : bool init false ;
nonc : bool init false ;
```

Next, an arrival process selects one of the booleans that is to be set to true. This is the initial binding of the ternary *aa-tRNA* complex at the ribosome. The continuation depends on the type of *aa-tRNA*: cognate, pseudo-cognate, near-cognate or non-cognate. In fact, a race is run that

depends on the concentrations c_cogn, c_pseu, c_near and c_nonc of the four types of *aa-tRNA* and a kinetic constant k1f. Concentrations are taken from [7]. For example, following Markovian semantics, the probability in the race for cogn to be set to true (the others remaining false) is the relative concentration c_cogn/(c_cogn + c_pseu + c_near + c_nonc). The rates can therefore also be computed as relative frequencies per cell, as the volume of the cell cancels out. A small C program manipulating Table 4 in the appendix takes care of this. The values of the concentrations are provided to Prism via the command line, since they differ from codon to codon.

```
// initial binding
[ ] (s=1) -> k1f * c_cogn : (s'=2) & (cogn'=true) ;
[ ] (s=1) -> k1f * c_pseu : (s'=2) & (pseu'=true) ;
[ ] (s=1) -> k1f * c_near : (s'=2) & (near'=true) ;
[ ] (s=1) -> k1f * c_nonc : (s'=2) & (nonc'=true) ;
```

As the *aa-tRNA*, that has just arrived, may dissociate too, the reversed reaction is in the model as well. However, control does not return to the initial state directly, but, for model checking purposes, first visits the special state s=0 representing dissociation. At the same time, the boolean that was true is reset. Here, cognates, pseudo-cognates and near-cognates are handled with the shared rate k2b. Non-cognates always dissociate as captured by the separate rate k2bx.

```
// dissociation
[ ] (s=2) &  ( cogn | pseu | near ) -> k2b :
        (s'=0) & (cogn'=false) &
            (pseu'=false) & (near'=false) ;
[ ] (s=2) &  nonc -> k2bx : (s'=0) & (nonc'=false) ;
```

An *aa-tRNA* that is not a non-cognate can continue from state s=2 in the codon recognition phase, leading to state s=3. This is a reversible step in the translation mechanism, so there are transitions from state s=3 back to state s=2 as well. However, the rates for cognates vs. pseudo- and near-cognates, viz. k3bc, k3bp and k3bn, differ significantly (see Table 2), which is essential to the fidelity of the *mRNA*-translation mechanism. Note that the values of the booleans do not change.

```
// codon recognition
[ ] (s=2) & ( cogn | pseu | near ) -> k2f : (s'=3) ;
[ ] (s=3) & cogn -> k3bc : (s'=2) ;
```

```
[ ] (s=3) & pseu -> k3bp : (s'=2) ;
[ ] (s=3) & near -> k3bn : (s'=2) ;
```

The next forward transition, from state `s=3` to state `s=4` in the Prism model, is a combination of several detailed steps of the translation mechanism involving the processing of *GTP*. The transition is one-directional, again with a significant difference in the rate `k3fc` for a cognate *aa-tRNA* compared to the rates `k3fp` and `k3fn` for pseudo-cognate and near-cognate *aa-tRNA*, that are equal.

```
// GTPase activation, GTP hydrolysis
// and  EF-Tu conformation change
[ ] (s=3) & cogn -> k3fc : (s'=4) ;
[ ] (s=3) & pseu -> k3fp : (s'=4) ;
[ ] (s=3) & near -> k3fn : (s'=4) ;
```

In state `s=4`, the *aa-tRNA* can either be rejected, after which control moves to intermediate state `s=5`, or accommodates, i.e. the ribosome reconforms such that the *aa-tRNA* can hand over the amino acid it carries, so-called peptidyl transfer. In the latter case, control changes to state `s=6`. As before, rates for cognates and those for pseudo-cognates and near-cognates are of different magnitudes. From the intermediate rejection state `s=5`, with all booleans set to false again, control returns to the start state `s=1`.

```
// rejection
[ ] (s=4) & cogn -> k4rc : (s'=5) & (cogn'=false) ;
[ ] (s=4) & pseu -> k4rp : (s'=5) & (pseu'=false) ;
[ ] (s=4) & near -> k4rn : (s'=5) & (near'=false) ;

// accommodation, peptidyl transfer
[ ] (s=4) & cogn -> k4fc : (s'=6) ;
[ ] (s=4) & pseu -> k4fp : (s'=6) ;
[ ] (s=4) & near -> k4fn : (s'=6) ;
```

After some movement back-and-forth between state `s=6` and state `s=7`, the binding of the EF-G complex becomes permanent. In the detailed translation mechanism a number of (mainly sequential) steps follows, that are summarized in the Prism model by a single transition to a final state `s=8`, that represents elongation of the protein in nascent with the amino acid carried by the *aa-tRNA*. The synthesis is successful if the *aa-tRNA* was either a cognate or pseudo-cognate for the codon under

translation, reflected by either `cogn` or `pseu` being true. In case the *aa-tRNA* was a near-cognate (non-cognates never pass beyond state `s=2`), an amino acid that does not correspond to the codon in the genetic code has been inserted. Thus, in this case, an insertion error has occurred.

```
// EF-G binding
[ ] (s=6) -> k6f : (s'=7) ;
[ ] (s=7) -> k7b : (s'=6) ;

// GTP hydrolysis, unlocking, tRNA movement
// Pi release, rearrangements of ribosome and EF-G
// dissociation of GDP
[ ] (s=7) -> k7f : (s'=8) ;
```

A number of transitions, linking the dissociation state `s=0` and the rejection state `s=5` back to the start state `s=1`, where a race of *aa-tRNA*s of the four types commences anew, and looping at the final state `s=8`, complete the Prism model. The transitions are deterministically taking, as no other transitions leave these states. Having no biological counterpart the transitions are assigned a high-rate making the time they take negligible.

```
// no entrance, re-entrance at state 1
[ ] (s=0) -> FAST : (s'=1) ;
// rejection, re-entrance at state 1
[ ] (s=5) -> FAST : (s'=1) ;
// elongation
[ ] (s=8) -> FAST : (s'=8) ;
```

Table 2 collects the rates as compiled from the biological literature and used in the Prism model above.

`k1f`	140	`k3fc`	260	`k4rc`	60	`k6f`	150
`k2f`	190	`k3fp, k3fn`	0.40	`k4rp, k4rn`	`FAST`	`k7f`	145.8
`k2b`	85	`k3bc`	0.23	`k4fc`	166.7	`k7b`	140
`k2bx`	2000	`k3bp, k3bn`	80	`k4fp, k4fn`	46.1		

Table 2. Rates of the Prism model, adapted from [9, 26]. Rate `k2bx` is based on the estimate of the average delay of non-cognate arrivals of 0.5ms. Rates `k4fc`, `k4fp`, `k4fn` and `k7f` are accumulative rates of sequentially composed transitions.

In the next two sections, we will study the Prism model described above for the analysis of the probability for insertion errors, i.e. extension of the

peptidyl chain with a different amino acid than the codon codes for, and of the average insertion times, i.e. the average time it takes to process a codon up to elongation.

4 Insertion errors

In this section we discuss how the model checking features of Prism can be exploited to predict the misreading frequencies for individual codons. The translation of *mRNA* into a polypeptide chain is performed by the ribosome machinery with high precision. Experimental measurements show that on average, only one in 1,000 to 10,000 amino acids is added wrongly (cf. [12]). [3]

For a codon under translation, a pseudo-cognate anticodon carries precisely the amino acid that the codon codes for. Therefore, although different in codon-anticodon bound, successful matching of a pseudo-cognate does not lead to an insertion error, as –accidentally– the right amino acid has been used for elongation. In our model, the main difference of cognates vs. pseudo-cognates and near-cognates is in the kinetics. At various stages of the peptidyl transfer the rates for true cognates differ from those for pseudo-cognates and near-cognates up to three orders of magnitude.

Figure 2 depicts the relevant abstract automaton, derived from the Prism model discussed above. See also Table 1. In case a transition is labeled with two rates, e.g. 0.23/80, the leftmost number, viz. 0.23, concerns the processing of a cognate *aa-tRNA*, while the rightmost number, viz. 80, that of a pseudo-cognate or near-cognate. In three states a probabilistic choice has to be made: in state 2 leading to state 0 or 3, in state 3 leading back to state 3 or forward to state 4, and in state 4 leading to rejection in state 5 or eventually to success via state 6. The probabilistic choice in state 2 is the same for cognates, pseudo-cognates and near-cognates alike, the ones in state 3 and in state 4 depend on the type of *aa-tRNA*, cognates and pseudo-cognates vs. near-cognates.
A cognate *aa-tRNA* starting in state 1 will move to state 2 with probability 1. From here, it will dissociate with probability $85/(85+190) \approx 0.309$, moving to state 0, or will be recognized with the complementary probability $190/(85+190) \approx 0.691$, moving to state 3. The same holds for pseudo-cognate and near-cognate *aa-tRNA*. However, after recognition in state 3, a cognate *aa-tRNA* will go through the hydrolysis phase leading to state 4 for a fraction 0.999 of the cases (computed as $260/(0.23+260)$), a

[3] Our findings, see Table 5, based on the kinetic rates available and the assumptions made, are well within these boundaries.

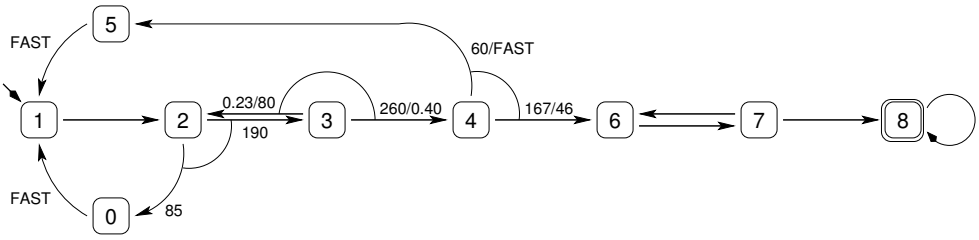

Fig. 2. Abstract automaton summarizing the Prism code. See also Table 1.

fraction being close to 1. In contrast, for a pseudo-cognate or near-cognate *aa-tRNA* this is $0.40/(0.40 + 80) \approx 0.005$ only. A similar difference can be noted in state 4 itself. Cognates will accommodate and continue to state 6 with probability 0.736, while pseudo-cognates and near-cognates will do so with the small probability 0.044, the constant FAST being set to 1000 in our experiments as in [9]. Since the transition from state 4 to state 6 is irreversible, the rates of the remaining transitions are not of importance here.

For cognates, pseudo-cognates and near-cognates, the probability of reaching state 8 in one attempt can be easily computed, solving a small system of equations by hand or by using Prism. In the latter case, we have Prism evaluate the CSL-formula

```
P=? [ (s!=0 & s!=5) U (s=8) {(s=2) & cogn} ]
```

against our model. The formula asks to establish the probability for all paths where `s` is not set to `0` nor `5`, until `s` have been set to `8`, starting from the (unique) state satisfying `s=2 & cogn`. The expression $\{(s = 2)\&cogn\}$ is a so-called filter construction as supported by Prism. We obtain $p_s^c = 0.508$, $p_s^p = 0.484 \cdot 10^{-4}$ and $p_s^n = 0.484 \cdot 10^{-4}$, with p_s^c the probability for a cognate to end up in state 8 —and elongate the peptidyl chain— without going through state 0 nor state 5; p_s^p and p_s^n the analogues for success of pseudo- and near-cognates, respectively. Note that these values are the same for every codon.

Different among codons in *E. coli* are the concentrations of cognates, pseudo-cognates and near-cognates.[4] Ultimately, the frequencies f_c, f_p and f_n of the types of *aa-tRNA* in the cell, i.e. the actual number of molecules of the kind, determine the concentration of the *aa-tRNA*.

[4] See Table 4 in the appendix.

Hence, under the usual assumption of homogeneous distribution, the frequencies determine the total rates for the arrival process of an anticodon. The probability for an anticodon arriving to be a cognate, pseudo-cognate or near-cognate can then be calculated from this.

As concluded in [9] based on simulation results, the probability for an erroneous insertion, is strongly correlated with the quotient of the number of near-cognate anticodons and the number of cognate anticodons. See Figure 3. As an advantage of the present setting, this correlation actually

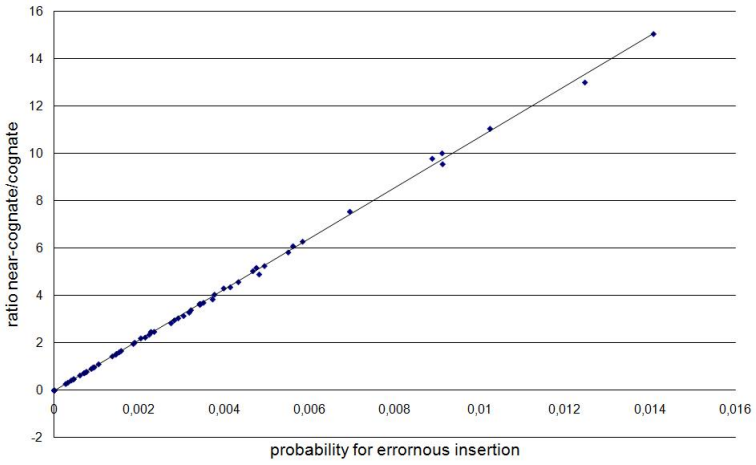

Fig. 3. Correlation of the ratio f_n/f_c of the frequency of near-cognates over the frequency of cognates vs. the probability of an insertion error. See also Table 5 in the appendix.

can be formally derived. This is as follows. We have that an insertion error occurs if a near-cognate succeeds to attach its amino acid. Note that we already have established $p_s^p, p_s^n \ll p_s^c$. Therefore,

$$\begin{aligned} P(\text{error}) &= P(\text{near \& elongation} \mid \text{elongation}) \\ &= \frac{p_s^n \cdot (f_n/tot)}{p_s^c \cdot (f_c/tot) + p_s^p \cdot (f_p/tot) + p_s^n \cdot (f_n/tot)} \approx \frac{p_s^n \cdot f_n}{p_s^c \cdot f_c} \sim \frac{f_n}{f_c} \end{aligned}$$

with $tot = f_c + f_p + f_n$, and where we have used that

$$P(\text{elongation}) = (f_c/tot) \cdot p_s^c + (f_p/tot) \cdot p_s^p + (f_n/tot) \cdot p_s^n \,.$$

Note, the ability to precalculate the probabilities p_s^c, p_s^p and p_s^n is instrumental in obtaining the above result. As such, it illustrates the approach

of piecewise analysis, first establishing quantities for part of the system to obtain a quantity for the system as a whole.

5 Competition and insertion times

In this section, we continue the analysis of the Prism model for translation and discuss the correlation of the average insertion time for the amino acid specified by a codon, on the one hand, and and the *aa-tRNA* competition, i.e. the relative abundance of pseudo-cognate and near-cognate *aa-tRNA*s, on the other hand. The insertion time of a codon is the average time it takes to elongate the protein in nascent with an amino acid.

The average insertion time can be computed in Prism using the concept of *rewards*, also known as *costs* in Markov theory. Each state is assigned a value as its reward. Further, the reward of each state is weighted per unit of time. Hence, it is computed by multiplication with the average time spent in the state. The cumulative reward of a path in the chain is defined as a sum over all states in the path of such weighted rewards per state. Thus, by assigning to each state the value 1 as reward, we obtain the total average time for a given path. For example, in Prism the cumulative reward formula `R=? [ F (s=8) ]` which asks to compute the expected time to reach state `s=8`. Recall, in state `s=8` the amino acid is added to the polypeptide chain. The formula returns the average reward of all the paths that lead from the initial state 1 to state 8. As explained above, in order to obtain the average time for insertion, we assign each state the value 1 as a reward, which in Prism can be done using the following code

```
rewards true: 1 endrewards
```

The construct expresses the fact that `1` is assigned to any state that satisfy the condition `true` (which is trivially satisfied by all states).

So, a script calling Prism for model checking the above formula then yields the expected insertion time per codon. Table 6 in the appendix lists the results. Although the correlation of cognate frequency and insertion times is limited, the qualitative claim of [25] of 'rare' codons being translated slow and 'frequent' codons being translated fast is roughly confirmed by the model. E.g., the codons AGC and CCA have amongst the lowest frequencies, 420 and 617, the lowest and two but lowest frequency, respectively, and translates indeed the slowest, 1.4924 and 1.5622 seconds, respectively. However, the codon CCA with an availability of 581, of one but lowest frequency, is translated at a moderate rate of 0.5525 seconds on the average. Thus, in line with our considerations, cognate availability per

se does not sufficiently predict translation time. Comparably, the fastest insertion times, 52.7 and 64.5 milliseconds, are realized by the codons CUG and CGU, of the codons corresponding to amino acids the one and two but most abundant. The codon CUG of the highest frequency 5136, excluding stop codons, though has an average insertion time of 102.8 milliseconds.

p_s^c 0.5079	p_f^c 0.4921	T_s^c 0.03182	T_f^c $9.342 \cdot 10^{-3}$
p_s^p $4.847 \cdot 10^{-4}$	p_f^p 0.9995	T_s^p 3.251	T_f^p 0.3914
p_s^n $4.847 \cdot 10^{-4}$	p_f^n 0.9995	T_s^n 3.251	T_f^n 0.3914

Table 3. Exit probabilities and exit times (in seconds) for three types of *aa-tRNA*, superscripts c, p and n for cognate, pseudo-cognate and near-cognate *aa-tRNA*, respectively. Failure for exit to states `s=0` or `s=5`, subscript f; success for exit to state `s=8`, subscript s.

A little bit more ingenuity is needed to establish average exit times, for example for a cognate to pass from state `s=2` to state `s=8`. The point is that conditional probabilities are involved. However, since dealing with exponential distributions, elimination of transitions in favour of adding their rates to that of the remaining ones, does the trick. Various results, some of them used below, are collected in Table 3. (The probabilities of failure and success for the non-cognates are trivial, $p_f^x = 1$ and $p_s^x = 0$, with a time per failed attempt $T_f^x = 0.5 \cdot 10^{-3}$ seconds.)

There is a visible correlation between the quotient of the number of near-cognate *aa-tRNA* over the number of cognate *aa-tRNA* and the average insertion time. See Figure 4. In fact, the average insertion time for a codon is approximately proportional to the near-cognate/cognate ratio. This can be seen as follows. The insertion of the amino acid is completed if state `s=8` is reached, either for a cognate, pseudo-cognate or near-cognate. As we have seen, the probability for either of the latter two is negligible, $p_s^p, p_s^n = 4.847 \cdot 10^{-4}$. Therefore, the number of cognate arrivals is decisive. With p_f^c and p_s^c being the probability for a cognate to fail, i.e. exit at state `s=0` or `s=5`, or to succeed, i.e. reach of state `s=8`, the insertion time T_{ins} can be regarded as a geometric series. (Note the exponent i below.) Important are the numbers of arrivals of the other *aa-tRNA* types per single cognate arrival, expressed in terms of frequencies. An arrival occurring for the $(i+1)$st arrival of a cognate has spent $(i \times T_f^c) + T_s^c$ processing cognate *aa-tRNA*. The number of pseudo-cognate,

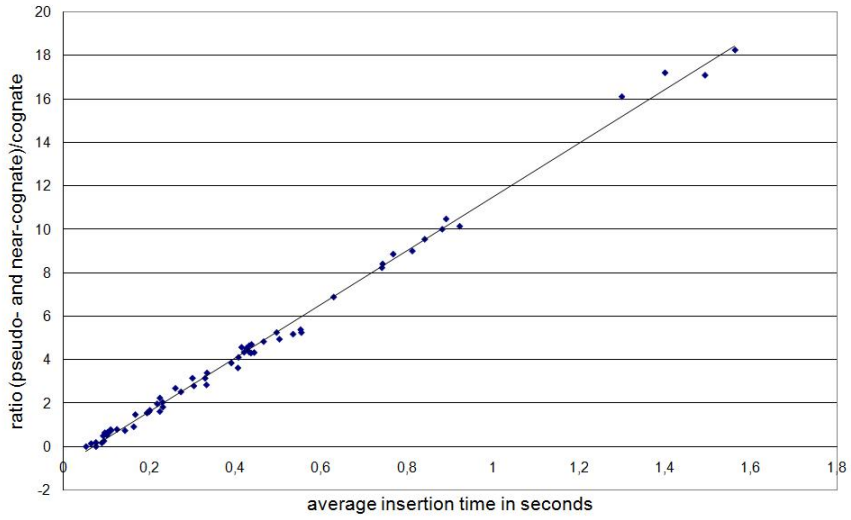

Fig. 4. Correlation of the ratio $(f_p + f_n)/f_c$ of total frequency of pseudo-cognates and near-cognates over the frequency of cognates vs. average insertion times. See also Table 6 in the appendix.

near-cognate and non-cognate arrivals per individual cognate arrival are, on the average, the relative fractions $\frac{f_p}{f_c}$, $\frac{f_n}{f_c}$, and $\frac{f_x}{f_c}$, respectively (with f_p, f_n, and f_c as before in Section 4, and f_x the frequency of non-cognate *aa-tRNA*). See Figure 5. Summing over i, the number of failing cognate *aa-tRNA* for a successful cognate insertion, yields

$$\begin{aligned} T_{ins} &= \textstyle\sum_{i=0}^{\infty} (p_f^c)^i p_s^c \cdot (\text{delay for } i \text{ failing and 1 successful cognate arrivals}) \\ &= \textstyle\sum_{i=0}^{\infty} (p_f^c)^i p_s^c \cdot \big((i+1) \cdot (\frac{f_p}{f_c} T_f^p + \frac{f_n}{f_c} T_f^n + \frac{f_x}{f_c} T_f^x) + i \cdot T_f^c + T_s^c \big) \\ &\approx \frac{f_p+f_n}{f_c}\, p_s^c\, T_f^n \, \textstyle\sum_{i=0}^{\infty} (i+1) \cdot (p_f^c)^i \\ &\sim \frac{f_p+f_n}{f_c}\,. \end{aligned}$$

Fig. 5. Accumulated delay after three cognate arrivals: (p) delay $(f_p/f_c)\cdot T_f^p$ for failing pseudo-cognates, (n) delay $(f_n/f_c)\cdot T_f^n$ for failing near-cognates, (x) delay $(f_x/f_c)\cdot T_f^x$ for non-cognates, (cf) exit time T_f^c for a failing cognate, (cs) exit time T_s^c for a successful cognate.

Here, we have used that T_f^c and T_s^c are negligible, T_f^p equals T_f^n, and $\frac{f_x}{f_c} T_f^x$ is relatively small, from which it follows that $\frac{f_p+f_n}{f_c} T_f^n$ is the dominant summand. Note that the estimate is not accurate for small values of $f_p + f_n$. Nevertheless, closer inspection shows that for these values the approximation remains order-preserving. Again, the results obtained for parts of the systems are pivotal in the derivation.

6 Concluding remarks

In this paper, we presented a stochastic model of the translation process based on presently available data of ribosome kinetics [12, 9]. We used the model checking facilities of the Prism tool for continuous-time Markov chains. Compared to [9] that uses simulation, our approach is computationally more reliable (independent on the number of simulations) and has faster response times (taking seconds rather then minutes or hours). More importantly, model checking allowed us to perform piecewise analysis of the system, yielding better insight in the model compared to just observing the end-to-end results with a monolithic model. Based on this, we improved on earlier observations, regarding error probabilities and insertion times, by actually deriving the correlation suggested by the data.

In [7] a correlation was reported between the number of copies (concentrations) of cognate *tRNA*s and the frequency of usage of particular codons in the most abundant proteins in *E. coli*. It is suggested that this optimization is favorable for the cell growth: when they are urgently needed the most used proteins are translated with maximum speed and accuracy. On the other hand, we observed that there is a high correlation (0.86) between the cognate *tRNA* concentrations and the ration near-cognates vs. cognates which, according to our model, determines the error probabilities. Consequently, it would be interesting to check if there exists even better correlation between the near-cognates/cognates ratios and the codon usage frequencies than between the latter and the concentrations.

In conclusion, we have experienced *aa-tRNA* competition as a very interesting biological case study of intrinsic stochastic nature, falling in the category of the well known lambda-phage example [1]. Our model opens a new avenue for future work on biological systems that possess intrinsically probabilistic properties. It would be interesting to apply our method to processes which, similarly to translation, involve small numbers

of molecules, like *DNA* replication [16, 19], *DNA* repair [11, 20], charging of the *tRNA*s with amino acids [8, 15], etc., thus rendering approaches based on ordinary differential equations less attractive.

References

1. A. Arkin, J. Ross, and H.H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics*, 149:1633–1648, 1998.
2. C. Baier, J.-P. Katoen, and H. Hermanns. Approximate symbolic model checking of continuous-time Markov chains. In J.C.M. Baeten and S. Mauw, editors, *Proc. CONCUR'99*, pages 146–161. LNCS 1664, 1999.
3. D. Bošnački, H.M.M. ten Eikelder, M.N. Steijaert, and E.P. de Vink. Stochastic analysis of amino acid substitution in protein synthesis. In M. Heiner and A.M. Uhrmacher, editors, *Proc. CMSB 2008*, pages 367–386. LNCS 5307, 2008.
4. M. Calder, V. Vyshemirsky, D. Gilbert, and R. Orton. Analysis of signalling pathways using continuous time Markov chains. In *Transactions on Computational Systems Biology VI*, pages 44–67. LNBI 4220, 2006.
5. N. Chabrier and F. Fages. Symbolic model checking of biochemical networks. In C. Priami, editor, *Proc. CMSB 2003*, pages 149–162. LNCS 2602, 2003.
6. V. Danos, J. Feret, W. Fontana, R. Harmer, and J. Krivine. Rule-based modelling of cellular signalling. In L. Caires and V. Thudichum Vasconcelos, editors, *Proc. CONCUR*, pages 17–41. LNCS 4703, 2007.
7. H. Dong, L. Nilsson, and C.G. Kurland. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *Journal of Molecular Biology*, 260:649–663, 1996.
8. O. Nureki et al. Enzyme structure with two catalytic sites for double-sieve selection of substrate. *Science*, 280:578–582, 1998.
9. A. Fluitt, E. Pienaar, and H. Viljoen. Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis. *Computational Biology and Chemistry*, 31:335–346, 2007.
10. M.A. Gilchrist and A. Wagner. A model of protein translation including codon bias, nonsense errors, and ribosome recycling. *Journal of Theoretical Biology*, 239:417–434, 2006.
11. M.F. Goodman. Coping with replication 'train wrecks' in *Escherichia coli* using Pol V, Pol II and RecA proteins. *Trends in Biochemical Sciences*, 25:189–195, 2000.
12. K.B. Gromadski and M.V. Rodnina. Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. *Molecular Cell*, 13(2):191–200, 2004.
13. J. Heath, M. Kwiatkowska, G. Norman, D. Parker, and O. Tymchyshyn. Probabilistic model checking of complex biological pathways. In C. Priami, editor, *Proc. CMSB 2006*, pages 32–47. LNBI 4210, 2006.
14. A.W. Heyd and D.A. Drew. A mathematical model for elongation of a peptide chain. *Bulletin of Mathematical Biology*, 65:1095–1109, 2003.
15. M. Ibba and D. Söll. Aminoacyl-tRNAs: setting the limits of the genetic code. *Genes & Development*, 18:731–738, 2004.
16. K.A. Johnson. Conformational coupling in DNA polymerase fidelity. *Annual Reviews in Biochemistry*, 62:685–713, 1993.
17. G. Karp. *Cell and Molecular Biology, 5th ed.* Wiley, 2008.

18. M. Kwiatkowska, G. Norman, and D. Parker. Probabilistic symbolic model cheking with Prism: a hybrid approach. *Journal on Software Tools for Technology Transfer*, 6:128–142, 2004. See also http://www.prismmodelchecker.org/.
19. S.A. Martomo and C.K.Mathews. Effects of biological DNA precursor pool asymmetry upon accuracy of DNA replication in vitro. *Mutation Research*, 499:197–211, 2002.
20. Ming Ni, Si-Yan Wang, Ji-Kun Li, and Qi Ouyang. Simulating the temporal modulation of inducible DNA damage response in *Escherichia coli*. *Biophysical Journal*, 93:62–73, 2007.
21. T. Pape, W. Wintermeyer, and M. Rodnina. Complete kinetic mechanism of elongation factor Tu-dependent binding of aa-tRNA to the A-site of *E. coli*. *EMBO Journal*, 17:7490–7497, 1998.
22. C. Priami, A. Regev, E. Shapiro, and W. Silverman. Application of a stochastic name-passing calculus to represent ation and simulation of molecular processes. *Information Processing Letters*, 80:25–31, 2001.
23. M.V. Rodnina and W. Wintermeyer. Ribosome fidelity: tRNA discrimination, proofreading and induced fit. *TRENDS in Biochemical Sciences*, 26(2):124–130, 2001.
24. A. Savelsbergh et al. An elongation factor G-induced ribosome rearrangement precedes tRNA–mRNA translocation. *Molecular Cell*, 11:1517–1523, 2003.
25. M.A. Sørensen, C.G. Kurland, and S. Pedersen. Codon usage determines translation rate in *Escherichia coli*. *Journal of Molecular Biology*, 207:365–377, 1989.
26. H. Viljoen. private communication, 2008.
27. S.Z. Wahab, K.O. Rowley, and W.M. Holmes. Effects of $tRNA_1^{Leu}$ overproduction in *Escherichia coli*. *Molecular Microbiology*, 7:253–263, 1993.

A Appendix: suplementary figures and data

```
// translation model


stochastic


// constants
const double ONE=1;
const double FAST=1000;

// tRNA rates
const double c_cogn ;
const double c_pseu ;
const double c_near ;
const double c_nonc ;

const double k1f = 140;
const double k2b =  85;
const double k2bx=2000;
const double k2f = 190;
const double k3bc=   0.23;
const double k3bp=  80;
const double k3bn=  80;
const double k3fc= 260;
const double k3fp=   0.40;
const double k3fn=   0.40;
const double k4rc=  60;
const double k4rp=FAST;
const double k4rn=FAST;
const double k4fc= 166.7;
const double k4fp=  46.1;
const double k4fn=  46.1;
const double k6f = 150;
const double k7b = 140;
const double k7f = 145.8;
```

```
module ribosome


s : [0..8] init 1 ;
cogn : bool init false ;
pseu : bool init false ;
near : bool init false ;
nonc : bool init false ;


// initial binding
[ ] (s=1) -> k1f * c_cogn : (s'=2) & (cogn'=true) ;
[ ] (s=1) -> k1f * c_pseu : (s'=2) & (pseu'=true) ;
[ ] (s=1) -> k1f * c_near : (s'=2) & (near'=true) ;
[ ] (s=1) -> k1f * c_nonc : (s'=2) & (nonc'=true) ;
[ ] (s=2) &  ( cogn | pseu | near ) -> k2b : (s'=0) &
      (cogn'=false) & (pseu'=false) & (near'=false) ;
[ ] (s=2) &  nonc -> k2bx : (s'=0) & (nonc'=false) ;

// codon recognition
[ ] (s=2) & ( cogn | pseu | near ) -> k2f : (s'=3) ;
[ ] (s=3) & cogn -> k3bc : (s'=2) ;
[ ] (s=3) & pseu -> k3bp : (s'=2) ;
[ ] (s=3) & near -> k3bn : (s'=2) ;

// GTPase activation, GTP hydrolysis, reconformation
[ ] (s=3) & cogn -> k3fc : (s'=4) ;
[ ] (s=3) & pseu -> k3fp : (s'=4) ;
[ ] (s=3) & near -> k3fn : (s'=4) ;

// rejection
[ ] (s=4) & cogn -> k4rc : (s'=5) & (cogn'=false) ;
[ ] (s=4) & pseu -> k4rp : (s'=5) & (pseu'=false) ;
[ ] (s=4) & near -> k4rn : (s'=5) & (near'=false) ;

// accommodation, peptidyl transfer
[ ] (s=4) & cogn -> k4fc : (s'=6) ;
[ ] (s=4) & pseu -> k4fp : (s'=6) ;
[ ] (s=4) & near -> k4fn : (s'=6) ;
```

```
// EF-G binding
[ ] (s=6) -> k6f : (s'=7) ;
[ ] (s=7) -> k7b : (s'=6) ;

// GTP hydrolysis, unlocking,
// tRNA movement and Pi release,
// rearrangements of ribosome and EF-G,
// dissociation of GDP
[ ] (s=7) -> k7f : (s'=8) ;

// no entrance, re-entrance at state 1
[ ] (s=0) -> FAST*FAST : (s'=1) ;
// rejection, re-entrance at state 1
[ ] (s=5) -> FAST*FAST : (s'=1) ;
// elongation
[ ] (s=8) -> FAST*FAST : (s'=8) ;

endmodule

rewards
  true : 1;
endrewards
```

codon	cognate	pseudo-cognate	near-cognate	non-cognate	codon	cognate	pseudo-cognate	near-cognate	non-cognate
UUU	1037	0	2944	67493	GUU	5105	0	0	66369
UUC	1037	0	9904	60533	GUC	1265	3840	7372	58997
UUG	2944	0	2324	66206	GUG	3840	1265	1068	65301
UUA	1031	1913	2552	65978	GUA	3840	1265	9036	57333
UCU	2060	344	0	69070	GCU	3250	617	0	67607
UCC	764	1640	4654	64416	GCC	617	3250	8020	59587
UCG	1296	764	2856	66558	GCG	3250	617	1068	66539
UCA	1296	1108	1250	67820	GCA	3250	617	9626	57981
UGU	1587	0	1162	68725	GGU	4359	2137	0	64978
UGC	1587	0	4993	64894	GGC	4359	2137	4278	60700
UGG	943	0	4063	66468	GGG	2137	4359	0	64978
UGA*	6219	0	4857	60398	GGA	1069	5427	11807	53171
UAU	2030	0	0	69444	GAU	2396	0	4717	64361
UAC	2030	0	3388	66056	GAC	2396	0	10958	58120
UAG*	1200	0	5230	65044	GAG	4717	0	3464	63293
UAA*	7200	0	4576	59698	GAA	4717	0	10555	56202
CUU	943	5136	4752	60643	AUU	1737	1737	2632	65368
CUC	943	5136	1359	64036	AUC	1737	1737	6432	61568
CUG	5136	943	2420	62975	AUG	706	1926	4435	64407
CUA	666	5413	1345	64050	AUA	1737	1737	6339	61661
CCU	1301	900	4752	64521	ACU	2115	541	0	68818
CCC	1913	943	2120	66498	ACC	1199	1457	4338	64480
CCG	1481	720	5990	63283	ACG	1457	1199	4789	64029
CCA	581	1620	1430	67843	ACA	916	1740	2791	66027
CGU	4752	639	0	66083	AGU	1408	0	1287	68779
CGC	4752	639	2302	63781	AGC	1408	0	5416	64650
CGG	639	4752	6251	59832	AGG	420	867	6318	63869
CGA	4752	639	2011	64072	AGA	867	420	4248	65939
CAU	639	0	6397	64438	AAU	1193	0	1924	68357
CAC	639	0	3308	67527	AAC	1193	0	6268	64013
CAG	881	764	6648	63181	AAG	1924	0	6523	63027
CAA	764	881	1886	67943	AAA	1924	0	2976	66574

Table 4. Frequencies of cognate, pseudo-cognate, near-cognate and non-cognates for *E. coli* as molecules per cell [7]. Stop codons UGA, UAG and UAA.

UUU 27.4e-4	CUU 46.7e-4	GUU 1.12e-10	AUU 14.4e-4
UUC 91.2e-4	CUC 13.6e-4	GUC 55.0e-4	AUC 35.0e-4
UUG 7.59e-4	CUG 4.49e-4	GUG 2.68e-4	AUG 58.3e-4
UUA 23.5e-4	CUA 18.9e-4	GUA 22.3e-4	AUA 34.4e-4
UCU 2.81e-10	CCU 34.1e-4	GCU 1.77e-10	ACU 2.73e-10
UCC 56.1e-4	CCC 10.4e-4	GCC 12.5e-4	ACC 34.2e-4
UCG 20.3e-4	CCG 37.6e-4	GCG 3.187e-4	ACG 31.7e-4
UCA 9.09e-4	CCA 22.8e-4	GCA 28.2e-4	ACA 29.1e-4
UGU 6.97e-4	CGU 1.21e-10	GGU 1.32e-10	AGU 8.70e-4
UGC 30.4e-4	CGC 4.59e-4	GGC 9.40e-4	AGC 37.2e-4
UGG 39.8e-4	CGG 88.7e-4	GGG 2.72e-10	AGG 140.7e-4
UGA 7.50e-4	CGA 3.98e-4	GGA 100.3e-4	AGA 48.1e-4
UAU 2.81e-10	CAU 91.1e-4	GAU 18.6e-4	AAU 15.2e-4
UAC 15.7e-4	CAC 47.5e-4	GAC 43.2e-4	AAC 49.3e-4
UAG 41.3e-4	CAG 69.4e-4	GAG 7.09e-4	AAG 32.1e-4
UAA 6.04e-4	CAA 22.7e-4	GAA 21.4e-4	AAA 14.6e-4

Table 5. Probabilities per codon for erroneous elongation.

UUU 0.3327	CUU 0.8901	GUU 0.0527	AUU 0.2733
UUC 0.8404	CUC 0.6286	GUC 0.7670	AUC 0.4373
UUG 0.1245	CUG 0.1028	GUG 0.1041	AUG 0.8115
UUA 0.4436	CUA 0.9217	GUA 0.2604	AUA 0.4321
UCU 0.0893	CCU 0.4202	GCU 0.0756	ACU 0.0943
UCC 0.7409	CCC 0.1992	GCC 1.5622	ACC 0.4658
UCG 0.3035	CCG 0.4257	GCG 0.1010	ACG 0.4073
UCA 0.2313	CCA 0.5535	GCA 0.3002	ACA 0.5025
UGU 0.1432	CGU 0.0645	GGU 0.0924	AGU 0.1636
UGC 0.3296	CGC 0.1010	GGC 0.1673	AGC 0.3905
UGG 0.4360	CGG 1.3993	GGG 0.2308	AGG 1.4924
UGA 0.1098	CGA 0.0962	GGA 1.2989	AGA 0.5517
UAU 0.0758	CAU 0.8811	GAU 0.2180	AAU 0.2242
UAC 0.2008	CAC 0.5341	GAC 0.4144	AAC 0.4959
UAG 0.4319	CAG 0.7425	GAG 0.1106	AAG 0.3339
UAA 0.0963	CAA 0.4058	GAA 0.2243	AAA 0.1945

Table 6. Estimated average insertion time per codon in seconds.