SUPPLEMENTARY INFORMATION FILE

Measurement and numerical modelling of cell-free protein synthesis: combinatorial block-variants of the PURE system

Paolo Carrara [1], Emiliano Altamura [2], Francesca D'Angelo [1], Fabio Mavelli [2] and Pasquale Stano [3,*]

[1] Sciences Department, Roma Tre University; V.le G. Marconi 446, I-00146 Rome, Italy

[2] Chemistry Department, University of Bari "Aldo Moro"; Via Orabona 4, I-70125 Bari, Italy

[3] DiSTeBA, University of Salento; Ecotekne, I-73100 Lecce, Italy

Correspondence: pasquale.stano@unisalento.it; Tel.: +39-0832-298-709

Table of Content

Supplementary Tables S1-S3
Table S1. Composition of PURE system "variants", sample IDs explained
Table S2. Composition of *in silico* PURE system "variants" in terms of modelling species
Table S3. Best-fit parameters referring to data presented in Figure 6a and Figure S3
Table S4. Results from linear and linearized models (Equation 4) and PCA

Supplementary Figures S1
Figure S1. Mechanisms of the *in silico* PURE system model.

Table S1. Correspondence between sample ID and composition of PURE system block-variants.

sample ID	D	E	B	mix "D" = [DNA], nM	mix "E"	mix "B"
333	3	3	3	22.0	3/3 of stock concentration	3/3 of stock concentration
332	3	3	2	22.0	3/3 of stock concentration	2/3 of stock concentration
331	3	3	1	22.0	3/3 of stock concentration	1/3 of stock concentration
323	3	2	3	22.0	2/3 of stock concentration	3/3 of stock concentration
322	3	2	2	22.0	2/3 of stock concentration	2/3 of stock concentration
321	3	2	1	22.0	2/3 of stock concentration	1/3 of stock concentration
313	3	1	3	22.0	1/3 of stock concentration	3/3 of stock concentration
312	3	1	2	22.0	1/3 of stock concentration	2/3 of stock concentration
311	3	1	1	22.0	1/3 of stock concentration	1/3 of stock concentration
233	2	3	3	14.7	3/3 of stock concentration	3/3 of stock concentration
232	2	3	2	14.7	3/3 of stock concentration	2/3 of stock concentration
231	2	3	1	14.7	3/3 of stock concentration	1/3 of stock concentration
223	2	2	3	14.7	2/3 of stock concentration	3/3 of stock concentration
222	2	2	2	14.7	2/3 of stock concentration	2/3 of stock concentration
221	2	2	1	14.7	2/3 of stock concentration	1/3 of stock concentration
213	2	1	3	14.7	1/3 of stock concentration	3/3 of stock concentration
212	2	1	2	14.7	1/3 of stock concentration	2/3 of stock concentration
211	2	1	1	14.7	1/3 of stock concentration	1/3 of stock concentration
133	1	3	3	7.3	3/3 of stock concentration	3/3 of stock concentration
132	1	3	2	7.3	3/3 of stock concentration	2/3 of stock concentration
131	1	3	1	7.3	3/3 of stock concentration	1/3 of stock concentration
123	1	2	3	7.3	2/3 of stock concentration	3/3 of stock concentration
122	1	2	2	7.3	2/3 of stock concentration	2/3 of stock concentration
121	1	2	1	7.3	2/3 of stock concentration	1/3 of stock concentration
113	1	1	3	7.3	1/3 of stock concentration	3/3 of stock concentration
112	1	1	2	7.3	1/3 of stock concentration	2/3 of stock concentration
111	1	1	1	7.3	1/3 of stock concentration	1/3 of stock concentration

Table S2. Results of statistical analyses. Linear and non-linear (linearized Log) multiple regression parameters of 27 PURE system variants. Bold characters indicate statistical significance ($p < 0.05$)

	β_D	β_E	β_B
End-point [eGFP] (μM)			
Linear model	$(-0.1 \pm 0.8) \times 10^{-2}$	**$(9.8 \pm 0.8) \times 10^{-2}$**	**$(4.5 \pm 0.8) \times 10^{-2}$**
Log model	$(-0.4 \pm 5) \times 10^{-2}$	**$(53 \pm 7) \times 10^{-2}$**	**$(21 \pm 6) \times 10^{-2}$**
Inflection time (s)			
Linear model	0 ± 240	**-850 ± 243**	**-1130 ± 243**
Log model	$(0.8 \pm 7) \times 10^{-2}$	**$(-55 \pm 7) \times 10^{-2}$**	**$(-72 \pm 8) \times 10^{-2}$**
Max production rate (μM/s)			
Linear model	$(-3 \pm 5) \times 10^{-6}$	**$(45 \pm 5) \times 10^{-6}$**	**$(22 \pm 5) \times 10^{-6}$**
Log model	$(-7 \pm 6) \times 10^{-2}$	**$(69 \pm 10) \times 10^{-2}$**	**$(26 \pm 7) \times 10^{-2}$**

Comment to PCA. PCA has been carried out on the variables end-point [eGFP], inflection time, max production rate, a_D, a_E, a_B (the latter are the dilution factors of D, E, B sub-groups, corresponding to values as 3/3, 2/3 or 1/3). PCA confirms the role of group E and group B compounds to determine the outcome of protein synthesis in the explored concentration range (whereas DNA does not play a role in the 7-22 nM concentration). Interestingly, a_E and a_B have almost orthogonal projections in the PC1-vs-PC2 plane, and are best distinguished along PC2. The experimental dataset, when represented over the PCs, appears as a regular point array, without cluster formation.

Table S3. Concentration of the species in the *in silico* PURE system model.

		stock conc.	index = 3	index = 2	index = 1
Group	**Species**	**μM**	**μM**	**μM**	**μM**
D	**DNA**	0.022	0.022	0.015	0.007
E	**TXcat**	0.100	0.100	0.067	0.033
E	**TLcat**	1.719	1.719	1.146	0.573
E	**RScat**	0.160	0.160	0.107	0.053
E	**ENcat**	0.080	0.080	0.053	0.027
B	**A**	300.000	300.000	200.000	100.000
B	**T**	1.900	1.900	1.267	0.633
B	**NTP**	1500.000	1500.000	1000.000	500.000
B	**CP**	20000.000	20000.000	13333.333	6666.667

Table S4. Best fit parameters for the simultaneous fitting of 27 eGFP-vs-time profiles (Figure 4b), according to the model of Figure S1.

```
> Kinetic Constants
          k_{TX} =      1.671 1/s
          k_{TL} = 7.176e-02 1/s  Fitting
          k_{RS} =      6.200 1/s
          k_{EN} =    100.000 1/s
      k_{nt,deg} = 7.918e-05 1/s
      k_{TL,deg} = 4.915e-04 1/s  Fitting
       k_{a,fol} = 6.767e-03 1/s  Fitting
       k_{a,mat} = 4.738e-03 1/s  Fitting
      K_{TX,DNA} = 5.000e-03 uM
      K_{TX,NTP} =     80.000 uM
       K_{TL,nt} =     16.634 uM   Fitting
       K_{TL,AT} =     10.000 uM
      K_{TL,NTP} =     10.000 uM
        K_{RS,A} =     23.000 uM
        K_{RS,T} = 7.000e-01 uM
      K_{RS,NTP} =    200.000 uM
       K_{EN,CP} =    200.000 uM
      K_{EN,NXP} =     40.000 uM

    > Hill Coefficients
         hTL(nt) =      1.274    Fitting
        hTL(NTP) =      2.972    Fitting
         hTL(AT) =      0.703    Fitting

    > Initial concentrations
          [NTP]o =   1500.000 uM
            [A]o =    300.000 uM
            [T]o =      1.900 uM
           [CP]o =  20000.000 uM
           [TL]o =      1.721 uM   Fitting

    > Enzyme concentrations
          C(DNA) =  2.000e-02 uM
           C(TX) =  1.000e-01 uM
           C(Rs) =  1.600e-01 uM
           C(En) =  8.000e-02 uM
         length =        238 aa
      n. bases =          4
         n. aa =         20
       n. tRNA =         46
```

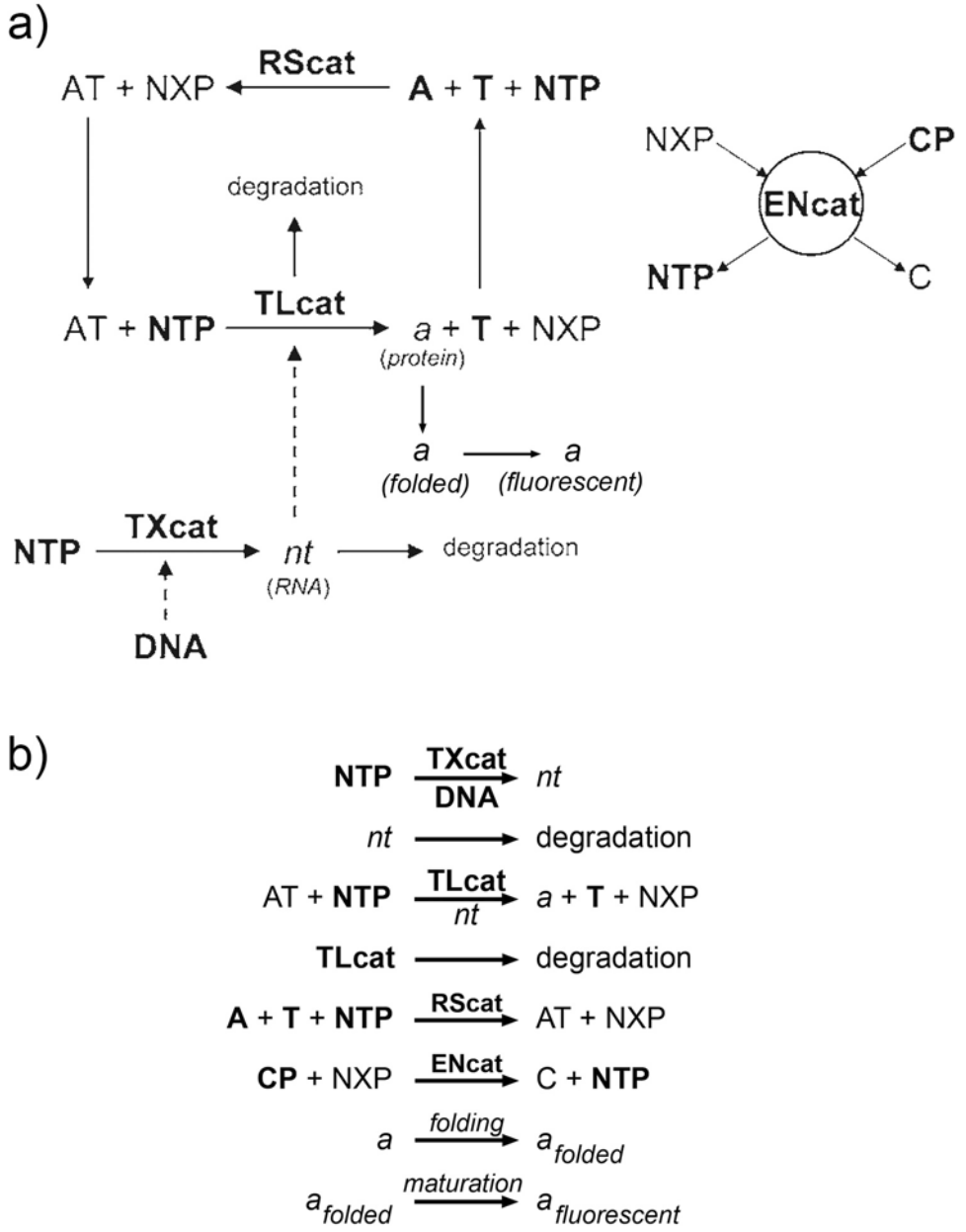

Figure S1. **In silico PURE system model.** General view of the model presented here. (a) The four modules (TX, TL, RS, EN) are interconnected by means of reactants, templates, and energyzing molecules. In bold, the molecules initially present in the PURE system. NTP, nucleoside triphosphate; *nt*, polymerized nucleotides (mRNA); AT, aminoacyl tRNA; A, amino acid; T, tRNA; *a*, polymerized amino acid (protein); NXP, nucleoside mono- or diphosphate; CP, creatine phosphate; C, creatine; TXcat, TLcat, RScat, ENcat represent, respectively, the catalysts (or the set of catalysts) for transcription, translation, aminoacyl-tRNA synthesis, and energy recycling. Dashed lines templating interaction; thick solid lines the reactions. Reproduced from "Mavelli, F.; Marangoni, R.; Stano, P. A Simple Protein Synthesis Model for the PURE System Operation. *Bull. Math. Biol.* **2015**, *77*, 1185–1212" ©2015 Springer, with the permission of Springer. (b) The six reactions that compose the kinetic model. For each, a Michaelis-Menten-like rate law has been used (exceptions: the two degradation reactions, which have been modelled as pseudo-first order reactions).