Computational and Structural Biotechnology Journal 21 (2023) 3768–3795

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research Article

A simple geometrical model of the electrostatic environment around the catalytic center of the ribosome and its significance for the elongation cycle kinetics

Marc Joiret[a,*], Frederic Kerff[b], Francesca Rapino[c], Pierre Close[c], Liesbet Geris[a,d,e]

[a] Biomechanics Research Unit, GIGA in silico medicine, Liège University, CHU-B34(+5) 1 Avenue de l'Hôpital, 4000 Liège, Belgium
[b] UR InBios Centre d'Ingénierie des Protéines, Liège University, Bât B6a, Allèe du 6 Août, 19, B-4000 Liège, Belgium
[c] Cancer Signaling, GIGA Stem Cells, Liège University, CHU-B34(+2) 1 Avenue de l'Hôpital, B-4000 Liège, Belgium
[d] Skeletal Biology & Engineering Research Center, KU Leuven, ON I Herestraat 49 - box 813, 3000 Leuven, Belgium
[e] Biomechanics Section, KU Leuven, Celestijnenlaan 300C box 2419, B-3001 Heverlee, Belgium

ABSTRACT

The central function of the large subunit of the ribosome is to catalyze peptide bond formation. This biochemical reaction is conducted at the peptidyl transferase center (PTC). Experimental evidence shows that the catalytic activity is affected by the electrostatic environment around the peptidyl transferase center. Here, we set up a minimal geometrical model fitting the available x-ray solved structures of the ribonucleic cavity around the catalytic center of the large subunit of the ribosome. The purpose of this phenomenological model is to estimate quantitatively the electrostatic potential and electric field that are experienced during the peptidyl transfer reaction. At least two reasons motivate the need for developing this quantification. First, we inquire whether the electric field in this particular catalytic environment, made only of nucleic acids, is of the same order of magnitude as the one prevailing in catalytic centers of the proteic enzymes counterparts. Second, the protein synthesis rate is dependent on the nature of the amino acid sequentially incorporated in the nascent chain. The activation energy of the catalytic reaction and its detailed kinetics are shown to be dependent on the mechanical work exerted on the amino acids by the electric field, especially when one of the four charged amino acid residues (R, K, E, D) has previously been incorporated at the carboxy-terminal end of the peptidyl-tRNA. Physical values of the electric field provide quantitative knowledge of mechanical work, activation energy and rate of the peptide bond formation catalyzed by the ribosome. We show that our theoretical calculations are consistent with two independent sets of previously published experimental results. Experimental results for *E.coli* in the minimal case of the dipeptide bond formation when puromycin is used as the final amino acid acceptor strongly support our theoretically derived reaction time courses. Experimental `Ribo-Seq` results on *E. coli* and *S. cerevisiae* comparing the residence time distribution of ribosomes upon specific codons are also well accounted for by our theoretical calculations. The statistical queueing time theory was used to model the ribosome residence time per codon during nascent protein elongation and applied for the interpretation of the `Ribo-Seq` data. The hypo-exponential distribution fits the residence time observed distribution of the ribosome on a codon. An educated deconvolution of this distribution is used to estimate the rates of each elongation step in a codon specific manner. Our interpretation of all these results sheds light on the functional role of the electrostatic profile around the PTC and its impact on the ribosome elongation cycle.

Contents

---

* Corresponding author.
E-mail address: marc.joiret@uliege.be (M. Joiret).

https://doi.org/10.1016/j.csbj.2023.07.016
Received 27 April 2023; Received in revised form 17 July 2023; Accepted 19 July 2023

---

1. Introduction

Ribosomes are the cells' manufacturing tools for building up proteins. They decode the 61 sense codons from a primary message encrypted in a messenger RNA (mRNA) single molecule. They translate it with the help of a set of transfer RNAs (tRNAs) into 20 amino acids to be sequentially polymerized in a nascent polypeptide that will eventually pass through the ribosomal exit tunnel and fold into its final structure.

X-ray solved structural representations of ribosomes have been publicly available for different species at atomic resolution for more than 20 years. The peptide bond is formed between the nascent protein chain and a newly incorporated amino acid at the ribosomal large subunit catalytic center, whose salient feature is that the peptidyl transferase center (PTC) is not an enzyme but a ribozyme, composed of ribosomal RNA (rRNA) [1]. There are no ribosomal protein components within a 15 Å radius of the catalytic center [2,3]. The electrostatic environment of the PTC is largely determined by the presence of the negatively charged phosphate moieties belonging to the 23S rRNA backbone in archae or bacteria, and to the 28S rRNA backbone in eukarya. The PTC cavity itself is made of rRNA and the substrates that are processed in the cavity also are RNA molecules: the single amino-acylated tRNA at the A site and the peptidyl-tRNA at the P site. It is thus necessary to quantitatively assess the relative contributions of the 2 tRNAs at the P and A site with respect to the contribution of the 23S rRNA (28S rRNA) to the electrostatic potential profile around the PTC cavity and towards the nascent chain tunnel entry port. Furthermore, in numerous previously published studies, the monovalent K^+ and bivalent Mg^{2+} metal ions have been shown to be key ingredients in the structure of the ribosome and of ribonucleic acids rRNAs and tRNAs [4–6].

Efforts in understanding the kinetics of the protein synthesis in vivo and the factors that affect elongation rate have been conducted for decades by the research community [7,8]. Experimental evidence supports the fact that enzymatic kinetics is affected by the electrostatic environment around the active sites of enzymes. Recently, vibrational Stark effect spectroscopy techniques were used to obtain direct measurements of electric fields at active sites of enzymes [9]. To elucidate the mechanisms of action and the reaction kinetics, there is great interest in knowing quantitatively the potential and the electric field prevailing at the active site of the ribosome, at the peptidyl transferase center and along the ribosome exit tunnel. To this date, there is no experimental protocol that would make the large subunit of the ribosome amenable to such direct measurements, although pioneering efforts were undertaken in the ribosome exit tunnel of rabbit reticulocytes [10]. Given the lack of direct experimental measurements for the potential and the electric field, the only way to obtain the physical quantities of interest is through the use of a mathematical model. Such mathematical models are broadly used to shed light on structure, function and properties of often complex biomolecular systems and are used in numerous studies [11].

In this paper, we start from five publicly available atomic structures of the ribosome across the three domains of life and use the fundamental laws of electrostatics along with the dielectric properties of media inside the ribosome to calculate the potential and the electric field experienced around the center of the peptidyl transferase reaction. We investigate the contribution of the two tRNAs at P and A-sites to the electrostatic potential around the rRNA PTC shell cavity and. The local distribution of the metal ions charges around the PTC and around the A and P site tRNAs was also studied within a distance of the order of the electrostatic screening length away from these RNA molecules. The electric field and the axial forces on test charged amino acids can be quantitatively estimated. The mechanical work exerted by the elec-

Table 1
Universally most conserved 23S/28S rRNA sequences around the PTC in the large subunit of the ribosome across the 3 domains of life, adapted from [17].

Species (domain)	PDB code	P-loop	A-loop	number of nucleotides within 40 Å around PTC
Haloarcula marismortui (archea)	`4V9F`	**A2486** 5’-GGAU**A**AC-3’	**U2620** 5’-GAGCUGGG**U**UUA-3’	487
Escherichia coli (bacteria)	`7K00`	**A2451** 5’-GGAU**A**AC-3’	**U2585** 5’-GAGCUGGG**U**UUA-3’	484
Thermus thermophilus (bacteria)	`4Y4P`	**A2463** 5’-GGAU**A**AC-3’	**U2597** 5’-GAGCUGGG**U**UUA-3’	484
Oryctolagus cuniculus (eukarya)	`7O7Y`	**A4143** 5’-GGAU**A**AC-3’	**U4277** 5’-GAGCUGGG**U**UUA-3’	488
Homo sapiens (eukarya)	`6EK0`	**A4397** 5’-GGAU**A**AC-3’	**U4530** 5’GAGCUGGG**U**UUA-3’	488

trostatic field on the test charged amino acids (E, K, D, E) upon protein elongation displacement can be calculated. The significance of the electrostatic environment around the PTC shell cavity on the peptide bond formation rate is quantitatively studied. We will rely on the statistical queueing time theory to model the ribosome dwelling time per codon during nascent protein elongation and use this theory for the interpretation of `Ribo-Seq` data. The rate calculations will be compared with previously published experimental results from two independent sources.

2. Material and methods

2.1. X-ray crystallographic space position of phosphate moieties and charged amino acid residues around the PTC cavity

We analyzed five publicly available structures of the large ribosomal subunit across the three domains of life: the archeon *Haloarcula marismortui* (PDB code: `4V9F` downloaded from `https://www.rcsb.org/`) obtained from x-ray crystallography at 2.4 Å [12]; two bacteria *Escherichia coli* (PDB code: `7K00`) at 2.0 Å [13], *Thermus thermophilus* (PDB code: `4Y4P`) at 2.5 Å [14]; two eukaryotes, the rabbit *Oryctolagus cuniculus* (PDB code `7O7Y`) at 2.2 Å [15] and the human *Homo sapiens* (PDB code: `6EK0`) at 2.9 Å. To localize the PTC shell cavity for the five species, we relied on two complementary methods: (i) when available in the PDB deposited x-ray solved structures, we used the tips of the acceptor arms of the `tRNAs` at the P-site and/or the A site, i.e. the `O3′` oxygen of the ribose of adenosine 76 at the 3’ end of the tRNAs; (ii) alternatively, we used the 23S or 28S `rRNA` sequence alignment of the nucleotides that are the most universally conserved across the three domains of life as detailed in [16,17]. Table 1 and Fig. 1 show, for each of the five studied species, the nucleotides in the 23S/28S `rRNA` P-loop (between secondary structure helices h90 and h93) and A-loop (helix h89) that are closest to the PTC shell cavity where the peptide bond is catalyzed between the tRNAs at the P and A sites. Once the virtual straight line between the `N3` atoms of the P-loop and A-loop nucleotides most likely involved in the catalytic activity of the PTC has been determined, all nucleotides and phosphate moieties of the 23S/28S `rRNA` molecules that are within a 40 Å distance are selected. These selections are used to show the 3D configurations of the PTC in PyMOL and are used to study the electrostatic environment of the PTC shell cavity. To find the ribosome tunnel entry port and extract the atom coordinates around the PTC, we used a tunnel search algorithm developed by Sehnal et al. [18], implemented in MOLE 2.0 and the web-based MOLEonline 2.0 tool publicly available online [19,20]. We used PyMOL (PyMOL Molecular Graphics System, Version 2.3.2) and exported the relevant selected atom positions’ cartesian coordinates to output files. These files were further processed with algorithms coded in Python to select the charged chemical groups on or near the inner surface of the PTC cavity. To localize the PTC, we used the geometric center of the five known 23S `rRNA` P loop and A loop nucleotides that are known to interact with the 3’-terminal `CCA` end of the tRNA acylated to the carboxy terminal amino acid of the nascent chain at the P site or the 3’-terminal `CCA` end of the aminoacylated tRNA at the A site [3,21–23]. These five nucleotides in the P and A loops of 23S `rRNA` are A2485(2450), A2486(2451) for the P loop and U2620(U2585), U2621(U2586) and U2622(U2585) for the A loop [2]. The numbering of the 23S `rRNA` residues is based on the *H. marismortui* sequence, while the corresponding position in the *E. coli* ribosome is shown in the brackets. For *Haloarcula marismortui*, we translated the crystallographic data model space so that the tunnel entry point would be at the origin and we aligned the direction from the tunnel entry point to the PTC geometric mid-point along the positive z-axis. We selected all charged amino acid residues (`NH2` or `NZ` for arginine or lysine, `OE2` or `OD2` for aspartate or glutamate) belonging to ribosomal proteins and all charged non-bridging oxygen atoms bound to the phosphate moieties in the 23S `rRNA` backbone that are closer than 40 Å from the centerline joining the tunnel entry port to the PTC geometrical midpoint. The cavity around the PTC was approximated by fitting a truncated prolate spheroid (ellipsoid of revolution about the major axis) having its semi-major axis aligned with the direction from the tunnel entry port to the PTC. The half prolate spheroid was also scaled in such a way that its semi-major axis spans the distance between the tunnel entry point and the decoding center P site ($\sim 8.75\,\mathrm{nm}$) and its semi-minor axis spans a half-length of 3 codons ($\sim 4.50/2 = 2.25\,\mathrm{nm}$). We algorithmically set out the 3D equations of the truncated prolate spheroid in this reference frame to calculate the radial distance of the selected atoms to the surface of the prolate spheroid PTC cavity. We confirmed that there are no charged atoms belonging to the `23S rRNA` or to any ribosomal proteins inside the fitted truncated prolate spheroid cavity. Once the PTC cavity and a ribosome exit tunnel is localized, the magnesium, potassium and sodium atoms (and water molecules) around them are counted. The counts are compared to the local charged phosphate moieties count and the positions of all charged groups (phosphate moieties, magnesium, potassium and sodium ions) are extracted to determine the set of the source charges contributing to the electrostatic potential in any given space point around the PTC cavity.

2.2. P-site and A-site tRNAs phosphate moieties and metal ions positions

The x-ray or cryo-electron-microscopy solved structures of the ribosomes of *Thermus thermophilus*(PDB code: `4Y4P`), *Escherichia coli* (PDB code: `7K00`) both include two tRNAs at the P and A site [14]. *Oryctolagus cuniculus* (PDB code: `7O7Y`) includes one tRNA at the P site and *Homo sapiens* (PDB code: `6EK0`) includes one tRNA at the E site. The O3’ oxygen atom of the ribose of the last nucleotide A76 (adenosine 76) in the tRNA at the P-site or the A-site is used to localize the PTC cavity. The space positions of the charged phosphate moieties and of the ions, magnesium, potassium and sodium have been extracted in order to investigate the contribution of both tRNAs and their metal ions to the electrostatic potential around the PTC cavity and along the line towards the tunnel entry port for the ribosome of *Thermus thermophilus*.

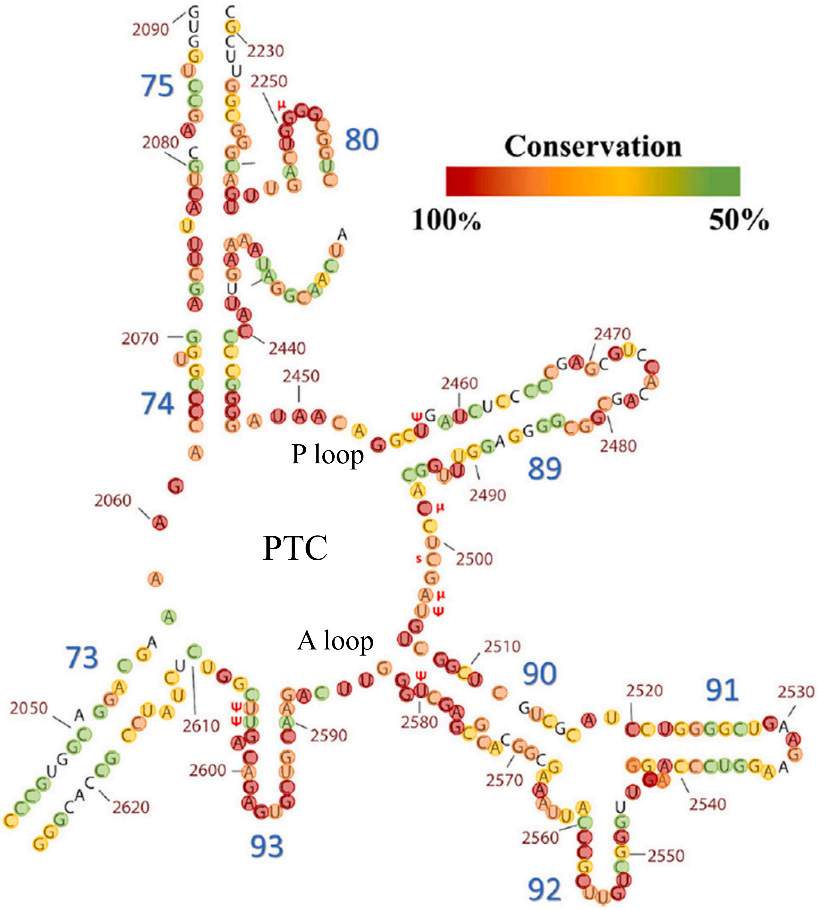


Fig. 1. Fragment of the secondary structure of 23SrRNA belonging to the large subunit of the ribosome of *Thermus thermophilus* which contains the nucleotides within and around the peptidyl transferase center (PTC). Helix numbers are in large blue font. Nucleotides highlighted in dark red are the most universally conserved across the three domains of life. Nucleotides numbering according to *E.coli*. The figure is taken from [16] with permission.

3. Theory and calculations

3.1. Idealized shape model of the ribosomal RNA cavity around the PTC and the Yukawa-Debye-Hückel potential with dielectric screening

After the sphere, the prolate spheroid is the geometrical shape with the smallest surface encompassing the largest volume. An electrostatic model of the cavity around the catalytic center of the large subunit of the ribosome is built using this simplest shape. This fulfills the minimal geometrical constraints which prevail between the ribosome peptide exit tunnel, the mRNA channel and the size of three aminoacylated-tRNAs that are accommodated in the cavity during nascent protein elongation Fig. 3(a). The prolate spheroid was fitted onto the *Haloarcula marismortui* aforementioned publicly available x-ray solved structure of the large ribosomal cavity around the peptidyl transferase center as explained in Material and Methods.

The electrical scalar potential $\Phi(\vec{\mathbf{r}})$ at the observed position $\vec{\mathbf{r}}$ is expressed, in a homogeneous medium of constant permittivity and in the absence of dielectric screening effects, by the Coulomb law:

$$\Phi(\vec{\mathbf{r}}) = \frac{1}{4\pi\epsilon} \int\!\!\int_S \frac{\sigma(\vec{\mathbf{r}}')\, da}{|\vec{\mathbf{r}} - \vec{\mathbf{r}}'|} \tag{1}$$

where $\sigma(\vec{\mathbf{r}}')$ is the surface-charge density (measured in coulombs per square meter) at position $\vec{\mathbf{r}}'$ of the source, da is the two dimensional surface element at $\vec{\mathbf{r}}'$ and ϵ is the permittivity of the dielectric medium (Eq. 1.23 in Jackson [24]) with $\epsilon = \epsilon_r \epsilon_0$, where ϵ_r is the relative permittivity of the medium and ϵ_0 is the permittivity of free space.

It is worth emphasizing here that the position where the physical values of the potential are calculated is $\vec{\mathbf{r}}$, but the variable over which the surface integration is conducted is $\vec{\mathbf{r}}'$ with its elementary surface elements ($d^{(2)}\vec{\mathbf{r}}' = da$) integrated over the surface domain of interest in the 3D space. We actually sum over all the fixed charges located on the surface. There are two ways to describe the charges and their positions. In the first, the charges are considered to be continuously distributed on the surface. A surface charge density σ must be known which can be a function of $\vec{\mathbf{r}}'$ as expressed in equation (1). In the second, the real fixed charges are discrete (with no spatial extensions) but the surface domain on which integration is conducted is still a compact interval in $\mathbb{R}^2$. Implicitly, discrete charges q_i, spatially localized in $\vec{\mathbf{r}}'_i$ are mathematically represented by a generalized function which is the product of the charges q_i by a Dirac distribution [25]: $\sigma(\vec{\mathbf{r}}') = \sum_{i \in \text{all charges}} q_i \delta^{(2D)}(\vec{\mathbf{r}}' - \vec{\mathbf{r}}'_i)$. This second type of description will be used in the next section. In mathematical terms, in both descriptions, the integration domain is the support of the function to be integrated. The support of the function is the interval upon which the function exists and is not null. $\vec{\mathbf{r}}'$ exists in the support of the charges or the support of the Dirac functions one implicitly uses to define the position of the charges that are sources of the field. $\vec{\mathbf{r}}$ is defined in another space, the set of points where one wants to calculate the field, irrespective of the positions of the source charges. In the surface integral calculation conducted below, $\vec{\mathbf{r}}'$ has coordinates u, v in the support of the charges, which is a surface in 3D (a half prolate spheroid) and $\vec{\mathbf{r}}$ has the cartesian coordinates (0, 0, z) of a straight line in 3D, because for the sake of simplicity, the potential and the electric field will only be calculated along the central axis z.

In the presence of polarizable dielectric material, screening effects occur due to induced dipoles or permanent dipoles reorientations in the medium separating the fixed charged sources of the field and the observation point. In this latter case, the electrical scalar potential $\Phi(\vec{\mathbf{r}})$ at the observed position $\vec{\mathbf{r}}$ is expressed by the Yukawa-Debye-Hückel law:

$$\Phi_{\text{Yuk}}(\vec{\mathbf{r}}) = \int\int_{S} \frac{\sigma^*(\vec{\mathbf{r}'})}{4\pi\,\epsilon}\, \frac{e^{-\frac{|\vec{\mathbf{r}}-\vec{\mathbf{r}'}|}{\xi}}}{|\vec{\mathbf{r}}-\vec{\mathbf{r}'}|}\, da \tag{2}$$

where $\sigma^*(\vec{\mathbf{r}'})$ is the actual *formal bare* surface charge density. There is a marked exponential damping of the Coulomb interaction where ξ is a characteristic distance of the screening. We assume the screening effect to be homogeneous around the fixed charged surface, and thus we take the assumption that both ϵ and ξ do not depend on $\vec{\mathbf{r}'}$ or $\vec{\mathbf{r}}$ or at least are piecewise constants in a given space domain [26].

For the prolate spheroid, we can take advantage of the axial symmetry and restrict the observation positions to the spatial points on the z axis, i.e., for $\vec{\mathbf{r}} = (0, 0, z)$. The surface integration is conducted on the support of the fixed source charges. The prolate spheroid's inner wall is geometrically generated by the $\gamma(u)$ curve moving axially along the z-axis from $z = 0$ to $z = -a$ as drawn in Fig. 3(a), where a and b are the semi-major and semi-minor axis of the prolate spheroid:

$$\gamma(u) = (b\sqrt{1-\frac{z^2}{a^2}}\,\cos u,\, b\sqrt{1-\frac{z^2}{a^2}}\,\sin u, z),\; u \in [0,\, 2\pi]. \tag{3}$$

The prolate spheroid's half surface can be written as $S = \phi(K)$ where $K = \{(u, v) \in [0, 2\pi] \times [0, \frac{\pi}{2}]\}$ and where $\phi : \mathbb{R}^2 \to \mathbb{R}^3$:

$$\phi(u, v) = (b\,\sin v\,\cos u,\; b\,\sin v \sin u,\; -a\,\cos v). \tag{4}$$

Another equivalent parametrization of the prolate spheroid's half surface can be written as $S = \phi(K)$ where $K = \{(u, v) \in [0, 2\pi] \times [0, 1]\}$ and where $\phi : \mathbb{R}^2 \to \mathbb{R}^3$:

$$\phi(u, v) = (b\,\sqrt{1-v^2}\,\cos u,\; b\,\sqrt{1-v^2}\,\sin u,\; -a\,v). \tag{5}$$

Upon these parametrizations, the distance between any observation point and any point supporting a source charge can be expressed respectively as:

$$|\vec{\mathbf{r}} - \vec{\mathbf{r}'}| = \sqrt{b^2\,\sin^2 v + (z + a\cos v)^2} \tag{6}$$

$$= \sqrt{b^2(1-v^2) + (z+av)^2}. \tag{7}$$

There is an aperture at the leftmost end of the prolate spheroid where the PTC cavity is connected to the ribosome exit tunnel which is approximated by a cylinder as shown in Fig. 3(a). The radius of the aperture is around 5 Å. Hence, the exact support of the resulting truncated spheroid' surface is parametrized as $S = \phi(K)$ where $K = \{(u, v) \in [0, 2\pi] \times [v_{inf}, v_{sup}]\}$ where $v_{sup} < 1$. $D_u\phi$ is the first partial derivative of the parametric equation of the surface $\phi(u, v)$ with respect to u and $D_v\phi$ is the first partial derivative of the parametric equation of the surface $\phi(u, v)$ with respect to v. In the general Eqs (1) and (2), the surface-charge densities $\sigma(\vec{\mathbf{r}'})$ or $\sigma^*(\vec{\mathbf{r}'})$ are dependent of the position $\vec{\mathbf{r}'}$ on the support of the source charges. Here, we will take the simple approximation that σ or σ^* can be considered a constant parameter over a surface of a given shape, e.g. over the spheroid's surface. This is the surface charge uniform distribution assumption for a given shape. Note however that a space dependence of σ is possible if it is compensated for by a similar space dependence of ϵ ensuring the combined ratio σ/ϵ is constant in a region of interest. This piecewise constant ratio is the strictly necessary assumption for the mathematical surface integration calculations of our models to be analytically tractable. In subsection 4.5, we will discuss how reasonable this assumption is by comparing the potential profiles calculated from the exact x-ray crystallographic structural data of the large subunit of the ribosome and calculated from the idealized spheroidal shape with constant surface charge density and a constant medium permittivity.

The electrostatic scalar Yukawa screened potential (2) results from the surface integral calculation:

$$\Phi(z) = \frac{\sigma^*}{4\pi\epsilon} \iint_{K=\{(u,v)\in[0,2\pi]\times[v_{inf},v_{sup}]\}} \frac{e^{-\frac{\sqrt{b^2(1-v^2)+(z+av)^2}}{\xi}}}{\sqrt{b^2(1-v^2)+(z+av)^2}}\, |D_u\phi \wedge D_v\phi|\, du\, dv \tag{8}$$

$$D_u\phi = (-b\,\sqrt{(1-v^2)}\,\sin u, b\,\sqrt{1-v^2}\,\cos u, 0) \tag{9}$$

$$D_v\phi = (-\frac{b\,v}{\sqrt{1-v^2}}\,\cos u, -\frac{b\,v}{\sqrt{1-v^2}}\,\sin u, -a) \tag{10}$$

$$|D_u\phi \wedge D_v\phi| = \left| det \begin{pmatrix} \vec{e_x} & \vec{e_y} & \vec{e_z} \\ -b\,\sqrt{1-v^2}\,\sin u & b\,\sqrt{1-v^2}\,\cos u & 0 \\ \frac{-b\,v\,\cos u}{\sqrt{1-v^2}} & \frac{-b\,v\,\sin u}{\sqrt{1-v^2}} & -a \end{pmatrix} \right| \tag{11}$$

$$= |(-ab\sqrt{1-v^2}\,\cos u, -ab\sqrt{1-v^2}\,\sin u, b^2\,v)| \tag{12}$$

$$= b\sqrt{a^2(1-v^2)\,+\,b^2\,v^2} \tag{13}$$

$$\Phi(z) = \frac{\sigma^*}{4\pi\epsilon} \int_0^{2\pi} du \int_{v_{inf}}^{v_{sup}} dv \frac{e^{-\frac{\sqrt{b^2(1-v^2)+(z+av)^2}}{\xi}} \cdot b\,\sqrt{a^2(1-v^2)\,+\,b^2 v^2}}{\sqrt{b^2(1-v^2)+(z+av)^2}} \tag{14}$$

$$\Phi(z) = \frac{\sigma^*}{2\epsilon} \int_{v_{inf}}^{v_{sup}} dv\, \frac{e^{-\frac{\sqrt{b^2(1-v^2)+(z+av)^2}}{\xi}} \cdot \sqrt{a^2(1-v^2)\,+\,b^2 v^2}}{\sqrt{(1-v^2)+(\frac{z+av}{b})^2}} \tag{15}$$

$$\Phi(z) = \frac{\sigma^*}{2\epsilon} \int_{v_{inf}}^{v_{sup}} dv \cdot \sqrt{\frac{a^2\,-\,(a^2-b^2)v^2}{(1-v^2)+(\frac{z+av}{b})^2}} \cdot e^{-\frac{\sqrt{b^2(1-v^2)+(z+av)^2}}{\xi}} \tag{16}$$

The integral converges if the denominator in the integrand of Eq. (16) does not vanish, i.e., if $v_{sup} < 1$, meaning that the spheroid (ellipsoid of revolution) is truncated. The ribosome cavity around the PTC further goes towards the ribosome exit tunnel which is approximated by a cylinder surface. By virtue of continuity and physical consistency, it is necessary that the electrostatic potential be a continuous function of space inside the cavity of the catalytic center of the ribosome and inside the ribosome exit tunnel.

According to the x-ray solved structure of the ribosomal large subunit, the length scales for the spheroid are with semi-major axis $a = 8.97$ nm, semi-minor axis $b = 2.25$ nm, giving $v_{sup} = 0.975$ such that the aperture of the spheroid cavity towards the cylindrical ribosome exit tunnel has a radius $R = 5$ Å. The length between the center of the spheroid, at $z = 0$, and the entry point of the ribosome exit tunnel is 8.75 nm. The P loop 23S rRNA nucleotide A2486 (2451) interacting at the P site in the peptide bond formation has its axial position 20.6 Å away from the tunnel entry port. This provides the lower integration limit for the support of the charges on the spheroid's surface $v_{inf} = 0.74536$.

The integral in Eq. (16) converges but cannot be solved analytically as it is an elliptical integral. We resorted to two classical numerical approaches using Newton-Cotes method or Gaussian quadrature where the nodes sampling followed the Gauss-Konrod rule [27]. Both methods gave the same result for the solution of the electrostatic potential $\Phi(z)$ profile along the centerline of the truncated spheroid's cavity.

The axial electric field is obtained from the negative of the first partial derivative with respect to z:

$$\vec{E}_z = -\frac{\partial\Phi(z)}{\partial z}. \tag{17}$$

Finally, the axial force on a test charge can be calculated from:

$$\vec{F}_z = q_e\,\vec{E}_z \tag{18}$$

where q_e is the charge of the test charge.

In Eq. (16), we require the knowledge of at least two phenomenological parameters: the *formal bare* surface charge density σ^* on the truncated spheroid surface in the vicinity of the PTC and the dielectric response (permittivity) ϵ of the medium around the PTC. In fact, only the ratio σ/ϵ is needed. The phenomenological screening length parameter ξ is also required.

3.2. Structural data model of the ribosomal RNA cavity around the PTC and the Yukawa-Debye-Hückel potential with dielectric screening

The Coulomb or Yukawa-Debye-Hückel electrostatic potential can be calculated from the x-ray solved exact distribution of the source charges (phosphate moieties and charged amino acids) for which the positional 3D map is shown in Fig. 3(b) to (e). The method to compute the electrostatic potential based on the real observed atom positions involves a discrete sum and complies with the superposition principle due to the linearity of the electrostatics equations. The Yukawa-Debye-Hückel potential is used and the exact positions $\vec{r}_i^{\,\prime}$ of the sources and their charges q_i are summed over all source charges:

$$\Phi(z) = \sum_{k \in \text{regions}} \sum_{i \in \text{source charges}} \frac{q_{i,k}}{4\pi\epsilon_0\, \epsilon_r(k)} \cdot \frac{e^{-\frac{|\vec{\mathbf{r}}_i^{\,\prime} - (0,0,z)|}{\xi_k}}}{|\vec{\mathbf{r}}_i^{\,\prime} - (0,0,z)|} \tag{19}$$

In this formula, two phenomenological parameters are required which are ϵ_r, the relative permittivity of the medium and the screening length ξ. The screening length ξ is a generic placeholder which can be computed from one of three possible different approaches (Bjerrum, Gouy-chapman or Debye screening length). The Coulomb potential is a particular case of the Yukawa potential when the screening length goes to infinity.

The assumption is made that the two phenomenological parameters are constant (homogeneous) in the media where the potential is computed. The standard or defaulted homogeneous values of these parameters are $\epsilon_r = 78$ (water) and $\xi = 10\,\text{Å}$. Eq. (19) neglects surface charge polarization effects at dielectric media discontinuities.

In the above Coulomb-Yukawa electrostatic potential Eq. (19), different values of $\epsilon_r(k)$ and different ξ_k screening lengths can be used in the different k-indexed regions (or media). The elementary unit charge value of $+|e|$ or $-|e| = -1.602\,10^{-19}\,\text{C}$ is used for each of the charges q_i associated to the positively or negatively charged atoms at their given $\vec{\mathbf{r}}_i^{\,\prime}$ positions.

3.3. Phenomenological parameters: σ^*/ϵ, media permittivities and dielectric screening lengths ξ

Electrostatic interactions in ribonucleoprotein structures are potentially quite strong, but these interactions are mitigated by the screening effects of water, nucleic acids (both rRNA and tRNAs) or nearby protein atoms [28], even in the absence of mobile ions. In addition to the electrostatic screening, surface charge polarization effects also occur due to dielectric response discontinuities at media interfaces [26].

The screening of electrostatic interactions results primarily from electronic polarization, reorientation of dipolar groups in the vicinity of charges and dipoles. These effects are well understood and can be accurately determined for interactions in isotropic, homogeneous media. However, in complex inhomogeneous environments such as those near the surface of ribonucleoproteins, dielectric screening is difficult to predict. Two factors are expected to be especially important in the case of the ribosome cavity around the PTC and the exit tunnel: the confined geometry and composition of the inner wall close to the surface, and whether the interactions involve direct charges or dipoles. The x-ray solved atomic space positions in the immediate 12 Å vicinity of the tunnel wall show that water molecules in addition to tRNAs contribute to the screening of the *formal bare* charges carried by the non-bridging oxygen atoms bound to the phosphorus atoms of 23S/28S rRNA. Due to the dissociation (charge regularization) of surface groups, the rRNA phosphate moieties support surface acquires a net surface charge density that we call σ^*. These *bare* charges do not stay unbalanced due to a screening effect involving water solvent. The water molecules dipole moments re-orient so that a layer of positively charged hydrogens oppose the negatively charged phosphate moieties.

Water as a bulk solvent has a relative electric permittivity of 78 (25°C), while experimental and theoretical evidence suggest that proteins (or the nascent polypeptide) have an average dielectric response that can be approximated with a dielectric constant of about 3-4, see [29] and references therein. The dielectric constant of nucleic acids in bulk solution has been measured to be around 8 [30]. Thus, depending on the abundance of water molecules in the PTC cavity volume, the presence of tRNAs and the presence of the carboxy terminal end of the growing nascent peptide, the cavity micro-environment cannot be viewed as uniform. The dielectric constants ϵ should be used in a range from $\epsilon = 3 - 4$ (polypeptide) to $\epsilon = 78$ (water).

Selecting the most appropriate screening theory reduces to knowing which length scale parameter ξ to use. Three length scales, i.e., the Bjerrum length (λ_B), the Debye length ($\lambda_D = \kappa^{-1}$) and the Gouy-Chapman length (ξ_{GC}) deserve specific attention as highlighted by Van Roij [31]. The expressions of the Bjerrum, Debye and Gouy-Chapman screening lengths are respectively:

$$\lambda_B = \frac{e^2}{4\pi\,\epsilon\,\epsilon_0\,k_B T}. \tag{20}$$

$$\kappa^{-1} = \Big(\frac{2\,e^2\,I\,N_A}{\epsilon\,\epsilon_0\,k_B T}\Big)^{-1/2}. \tag{21}$$

$$\xi_{GC} = \frac{1}{2\pi\,\lambda_B\,\sigma^*}. \tag{22}$$

where e is the elementary charge of the electron, ϵ the relative permittivity of the medium, ϵ_0 the permittivity of the vacuum, I the ionic strength, k_B the Boltzmann constant, N_A the Avogadro number, T the absolute temperature and σ^* is the *bare* surface charge density of the wall.

3.4. Effects of forces on Gibbs free activation energy for the transition state in the peptide bond formation

The peptide bond formation is conducted at the peptidyl transferase center (PTC) in the large subunit of the ribosome. It has been known for more than thirty years that the peptidyl transferase center is not an enzyme but a ribozyme, composed of ribosomal RNA (rRNA) [1]. This ribozyme is a template assisted catalyst using an information rich mRNA single molecule to process two amino acid substrates previously acylated to tRNAs. These two tRNAs were initially accommodated and proofread on the A site for cognate anticodon-codon matching and then transferred to the P site. The detailed mechanisms of the peptide bond formation have largely been studied in the literature and the ribosome catalytic site described both as an entropy water trap and a water trap [32,33]. During the peptidyl transfer reaction, the α-amino group of aminoacyl-tRNA positioned in the A site of the ribosome nucleophilically attacks the carbonyl carbon at the ester bond of the peptidyl-tRNA in the P site, which results in peptidyl-tRNA extended by one amino acid in the A site and deacylated tRNA in the P-site [2]. The peptide bond formation between the two aminoacylated-tRNAs proceeds 10 million times faster when catalyzed by the ribosome than when uncatalyzed in bulk solution [21]. The ribosome catalyzed peptide bond formation kinetics is known to be affected by the particular context of the upstream amino acid sequence and by the amino acid distribution embedded in the nascent chain ribosome exit tunnel [3,26].

The theory of kinetics of catalysis relies on the transition state theory that was introduced by Eyring who linked the rate constant of a chemical reaction to the Gibbs free activation energy of the transition state, TS [34–36]. In this representation, a catalyzer accelerates a (bio)chemical reaction through a significant reduction in the transition Gibbs free energy barrier that the reactants have to overcome as shown in Fig. 2.

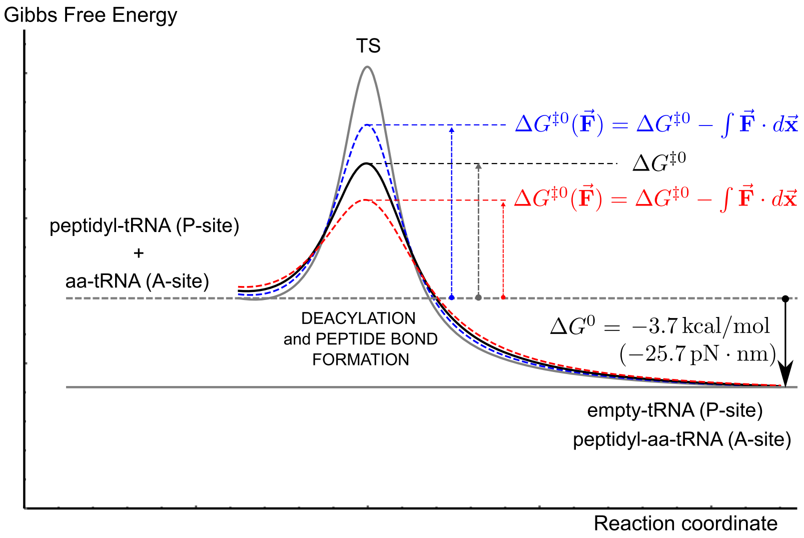


Fig. 2. Gibbs free energy activation barrier to overcome at the transition state (TS) for the peptide bond formation at the peptidyltransferase center of the ribosome as a function of the reaction coordinate. The activation energy for the catalyzed reaction without mechanical forces (thick black line) is $\Delta G^{\ddagger 0}$ as compared to the uncatalyzed transition state (thin black line). The activation energy for the catalyzed reaction can be higher (dashed blue line) or lower (dashed red line) when the mechanical work exerted by any applied force on the peptidyl-tRNA substrate at the P-site during the reaction is negative or positive respectively.

Here, we hypothesize that the physical forces transmitted mainly through the backbone of the peptidyl-tRNA play a role in the reduction of the Gibbs free energy barrier of the transition state. The mechanical work of these physical forces affects the activation Gibbs free energy of the transition state. The modulation of the Gibbs free energy activation barrier changes the reaction rate constant through Eyring's relation [34–36].

The catalyzed reaction scheme for the deacylation and peptide bond formation (transpeptidation) between the two substrates occurs in two steps with the formation of an activation complex (ribozyme complex) in the first step (rate limiting step). This is followed by a fast second step. If the second step is very fast, the overall reaction scheme can be viewed as a nucleophilic substitution (S_N2-like scheme) and there is a single transition state as shown in Fig. 2. The overall reaction can be written:

$$\underbrace{\text{peptidyl tRNA}}_{P-site} + \underbrace{\text{aminoacyl tRNA}}_{A-site} \rightarrow \underbrace{\text{empty tRNA}}_{P-site} + \underbrace{\text{peptidyl}(+1)\text{ tRNA}}_{A-site}. \tag{23}$$

This reaction is exergonic (thermodynamically favourable) as the change in standard Gibbs free energy is $\Delta G^0 \sim -3.7\,\mathrm{kcal/mol}(= -25.7\,\mathrm{pN \cdot nm})$ [37]. The rate constant depends on the Gibbs activation energy of the transition state $\Delta G^{\ddagger 0}$. Applying external forces on molecules involved in catalyzed or uncatalyzed chemical reactions affects the kinetics of the reactions. The mechanical work of these applied mechanical forces can quantitatively be incorporated in the calculation of the activation Gibbs free energy of the transition state as already conceptually introduced by Bell [38], Bustamante [39] and others [40]:

$$\Delta G^{\ddagger 0}(\vec{\mathbf{F}}) = \Delta G^{\ddagger 0}(\mathbf{0}) - \int \vec{\mathbf{F}} \cdot \mathbf{d}\vec{\mathbf{x}} \tag{24}$$

where $\Delta G^{\ddagger 0}(\vec{\mathbf{F}})$ is the activation energy for the transition state in the presence of an applied force acting on the system, $\Delta G^{\ddagger 0}(\mathbf{0}) \sim +14\,\mathrm{kcal/mol} = +97.2\,\mathrm{pN \cdot nm}$) is the activation energy for the transition state without any applied force [39], and $W = \int \vec{\mathbf{F}} \cdot \mathbf{d}\vec{\mathbf{x}}$ is the mechanical work exerted by the force upon a test body along its curvilinear path. The mechanical work W is algebraically positive if the force and the displacement are parallel or negative if they are antiparallel. In the former case, $\Delta G^{\ddagger 0}(\vec{\mathbf{F}})$ is smaller than $\Delta G^{\ddagger 0}(\mathbf{0})$, whereas it is larger in the latter case as can be seen in Fig. 2 (dashed red and dashed blue line respectively). In turn, the modulation of the Gibbs free energy activation barrier changes the reaction rate constant through Eyring's relation [34–36]:

$$k(\vec{\mathbf{F}}) = \kappa \cdot \left(\frac{k_B \cdot T}{h}\right) \cdot e^{-\Delta G^{\ddagger 0}(\vec{\mathbf{F}})/N\, k_B\, T} \tag{25}$$

$$k(\vec{\mathbf{F}}) = \kappa \cdot \left(\frac{k_B \cdot T}{h}\right) \cdot e^{-\left(\frac{\Delta G^{\ddagger 0}(\mathbf{0})}{N\, k_B\, T} - \frac{\int \vec{\mathbf{F}} \cdot \mathbf{d}\vec{\mathbf{x}}}{k_B\, T}\right)} = k(\mathbf{0}) \cdot e^{\frac{\int \vec{\mathbf{F}} \cdot \mathbf{d}\vec{\mathbf{x}}}{k_B\, T}} \tag{26}$$

where $k(\vec{\mathbf{F}})$ is the reaction rate constant of the rate limiting step in the presence of an applied force upon the substrate at the P-site, $k(\mathbf{0})$ is the reaction rate constant in the absence of applied force. N, k_B, h and κ are Avogadro's number, Boltzmann's constant, Planck's constant and transmission coefficient respectively [34–36].

3.5. Modified Michaelis-Menten kinetics of the peptide bond formation

The peptide bond formation kinetics at the catalytic center of the ribosome can be described by the Michaelis-Menten model where the aminoacyl-tRNA (acceptor substrate) is the canonical substrate and where the C-terminal amino acid of the peptidyl-tRNA (donor substrate) behaves as an allosteric substrate.

$$\begin{array}{c} \text{peptidyl tRNA (P-site)} \\ \boxed{\text{substrate 1 (donor)}} \\ \text{aa tRNA (A-site)} \\ \boxed{\text{substrate 2 (acceptor)}} \end{array} + \boxed{\text{ribozyme}} \underset{k_2}{\overset{k_1}{\rightleftarrows}} \boxed{\begin{array}{c}\text{ribozyme}\\ \text{complex}\end{array}} \tag{27}$$

$$\boxed{\begin{array}{c}\text{ribozyme}\\ \text{complex}\end{array}} \xrightarrow{k_3} \boxed{\text{ribozyme}} + \begin{array}{c}\text{empty tRNA}\\ \text{(P-site)}\end{array} + \begin{array}{c}\text{peptidyl}(+1)\text{ tRNA}\\ \text{(A-site)}\end{array} \tag{28}$$

The rate of peptide bond formation, $k_{\text{pep}} = \frac{dP}{dt}$, is written:

$$\frac{dP}{dt} = v'_{max} \cdot \frac{S}{K_{1/2} + S} \tag{29}$$

$$v'_{max} = \frac{v_{max}}{1 + \frac{S_{allo}}{k_{allo}}} \tag{30}$$

where $K_{1/2} = \frac{k_2+k_3}{k_1}$ is the Michaelis constant of substrate 2, S, at the A-site; S_{allo} is the C-terminal amino acid at the peptidyl-tRNA (substrate 1); k_{allo} is the reaction rate constant between substrate 1 and the ribozyme (PTC) at the P-site and v_{max} is the maximum rate in the absence of allosteric effect ($v_{max} = k_3 \cdot [\text{ribozyme}]$). Incorporating the right hand side Maxwell-Boltzmann factor of (26) into (29) to account for the effect of mechanical forces in the kinetics of the ribozyme catalyzed peptide bond formation leads to the final kinetics equation:

$$\frac{dP}{dt} = e^{\frac{\int \vec{\mathbf{F}}\cdot d\vec{\mathbf{x}}}{k_B T}} \cdot v'_{max}(\mathbf{0}) \cdot \frac{S}{K_{1/2} + S} \tag{31}$$

In single molecule experiments, P, when proper normalized, represents the probability of the formation of the peptide bond over time in equation (31).

3.5.1. Elongation minimal case with dipeptide and puromycin

Studying the catalyzed kinetics of peptide bond formation, as described in the reactional scheme (27) and (28), is possible only when peptidyl transfer is uncoupled from accommodation. The reason is that the accommodation rate of `aa-tRNA` in the A site is in the range $5 - 10\,\text{s}^{-1}$ and peptide bond formation follows instantaneously. Because accommodation precedes peptide bond formation, it limits the rate of product formation because it is much slower than the peptidyl transfer [41]. One way to circumvent the accommodation rate limiting step is to use substrate analogs that bind to the A site rapidly and do not require accommodation. If the full length `aa-tRNA` is replaced by the shorter puromycin substrate as the last acceptor substrate (substrate 2), the kinetics of the catalytic chemical step can be monitored experimentally by the quench flow technique and is amenable to quantitative measurements [21]. These experimental measurements were conducted on prokaryotic ribosomes by Rodnina and coworkers [42] in the elongation minimal case when the donor substrate 1 is the minimal dipeptidyl-tRNA, i.e. $\text{fMet} - \text{X} - \text{tRNA}$ and the acceptor substrate 2 is puromycin, Pmn. X can be any of the 20 natural amino acids (positively charged, neutral or negatively charged). The global reaction scheme (23) becomes:

$$\underbrace{\text{fMet-X-tRNA}}_{P-site} + \underbrace{\text{puromycin(Pmn)}}_{A-site} \rightarrow \underbrace{\text{empty tRNA}}_{P-site} + \underbrace{\text{fMet-X-Pmn}}_{A-site} \tag{32}$$

The product of the catalyzed reaction, dipeptidyl-puromycin $\text{fMet} - \text{X} - \text{Pmn}$, is released from the ribozyme upon completion of the reaction. The experimental initial condition for the puromycin concentration was 20 mM and the reaction rate constant measurements made at $37°C$ were obtained by single exponential fitting [42].

3.5.2. Elongation cycle at codon resolution obtained by ribosome profiling Ribo-seq

The ribosome profiling technology, or Ribo-seq, introduced by Ingolia et al. [43] is based on deep sequencing of ribosome-protected mRNA fragments which produces a detailed account of the ribosome occupancy on specific mRNAs during translation in the sampled cells (bulk Ribo-seq). The density of ribosome occupancy on mRNAs at codon resolution is a proxy for the ribosome residence time (RRT) on each codon under endogenous conditions. A number of ribosome normalized footprint counts (NFC) distributions have been obtained for each codon in a number of previous published studies across multiple species [44–46].

Dana & Tuller analyzed ribosome profiling data for *S.cerevisiae* and *E.coli*, among other species, based on published data in the GEO database, accession number GSE13750 (GSM346111, GSM346114) [44, 45]. They produced normalized footprint counts (NFC) of ribosomes on codons of the same type, originating from different genes, controlling for the transcripts copy number and initiation rates. The NFC enables measuring the relative time a ribosome spends translating each codon in a specific transcript relative to other codons in it, while considering the total number of codons in the transcript. Specifically, these authors denoted by T_j the translation time of codon j in transcript J and denoted the mRNA levels (transcript copy number) of transcript J by m, and its initiation rate by B, and calculated that the ribosome protected read counts RC_j for codon j is [45]:

$$RC_j \sim m \cdot B \cdot T_j \tag{33}$$

and thus

$$NFC_j \sim \frac{m \cdot B \cdot T_j}{m \cdot B \cdot \text{mean}(T)} = \frac{T_j}{\text{mean}(T)} \tag{34}$$

This last relationship, (34), indicates that NFC values represent the time a ribosome spends decoding each codon in a specific transcript relative to other codons in that transcript [45].

Hence, if we divide the NFC by the mean elongation turnover $1/\text{mean}(T)$, i.e., the number of translated codons per unit of time for the species under study, we have:

$$\frac{NFC_j}{\text{turnover}} = \frac{T_j}{\text{mean}(T)} \cdot \frac{1}{1/\text{mean}(T)} \tag{35}$$

$$NFC_j \cdot \text{mean}(T) = T_j \tag{36}$$

We conclude that we can get the distribution of the ribosome residence time for codon j relative to the other codons from the NFC distribution. As the relative comparison is conducted on the translation mean time (elongation mean turnover), we have an averaged residence time of the ribosome for that codon in that species. Equivalently, if we have a probability density function for the NFC_j, multiplying this density by the elongation turnover, while scaling the probability space by the mean elongation time, we have the probability density function of the ribosome residence time for the codon of type j. Explicitly, the probability space scaling by $\text{mean}(T)$ means that if $\text{pdf}_{\text{NFC}_j}(x)$ denotes the probability density function of the normalized footprint count for codon j relative to the other codons, then:

$$f_{\text{RRT}_j}(t) = \frac{1}{\text{mean}(T)}\, \text{pdf}_{NFC_j}(x/\text{mean}(T)) \tag{37}$$

is the probability density function for the ribosome residence time on codon of type j. In equation (37), the turnover, or its inverse, $1/\text{mean}(T)$, just acts as a scaling of the NFC distribution. Given the way the NFCs were generated, one should refrain to interpret the probability space, the x-axis, as absolute times. Relative time ratio can be obtained from the comparison of two RRT distributions. Drawing conclusions on absolute time differences from such distributions might be misleading. In other words, the comparison of RRTs between codons should be made in time fold change, not in absolute time differences.

The time spent by a ribosome on codons provide information on the elongation cycle dynamics at codon resolution but this information is statistically aggregated translatome wide and statistically aggregated for all sequential sub-steps possibly occurring during the elongation cycle. The ribosome footprint count may help infer ribosome residence time on a given codon in a given condition but does not directly provide the temporal breakdown of the sub-steps occurring during one elongation cycle. How long does take the accommodation and proofreading of the tRNA on the codon? How long does take the peptide bond formation at the PTC and how long does take the translocation to the next codon? All these individual sub-step times are aggregated together in a global RRT per codon and translatome wide, i.e. for a large number of transcripts collectively.

Dana and Tuller [44,45] interpreted the distribution of ribosome residence time on codons as the distribution generated by the sum of two independent random variables, i.e., a Gaussian and an exponentially distributed random variable. The convolution of the two distributions leads to the exponentially modified Gaussian distribution. Our interpretation differs from theirs in considering that the empirically observed distribution of the ribosome residence time on a given codon is the sum of (at least) three contributing random variables that are exponentially distributed. The convolution of three such exponential distributions leads to the hypo-exponential distribution, see appendix A.2. Our interpretation derives from the queueing theory of a succession of sequential waiting times. The underlying statistical models in both interpretations result in the same topology for the empirical observed residence time distribution of a ribosome on a given codon. Indeed, the exponentially modified Gaussian distribution, the hypo-exponential distribution, the Gamma distribution (but also the log-normal distribution) convey very similar information as detailed in appendix A.2. In a pioneering paper, Tinoco and Wen used the Gamma distribution to simulate the time spent by a single ribosome on a codon during translation [47]. In a recent paper, estimation of peptide elongation times from ribosome profiling spectra was derived where the copy number of ribosome on any codon on any open reading frame was assumed to be Poisson distributed and where the frequency densities of the peptide elongation times are well fitted to log-normal distributions [48].

The Kullback-Leibler (KL) divergence distance between the exponentially modified Gaussian and its least squared fitted hypo-exponential is very close to zero. From the theory of information, a KL value close to zero indicates that both models are statistically equivalent in practice, when it comes to the characterization of the ribosome residence time on a given codon like `GAA` (Glu-E) or `CGG` (Arg-R), see appendix A.2 and below.

Here, we adopt the point of view that the observed empirical distribution of the residence time of a ribosome on a given codon derives from the sum of three random waiting times that are exponentially distributed. These three time steps are (1) `STEP 1`: tRNA accommodation and proofreading at the A-site decoding center; (2) `STEP 2`: peptide bond formation at the PTC; (3) `STEP 3`: translocation of the ribosome to the next codon downstream the mRNA (after the unloaded-tRNA left the E-site). The ribosome elongation cycle rate limiting step was shown to be `STEP 1` as previously reported by others [3,42,49].

Through the comparisons of the observed empirical ribosome residence time (RRT) on specific codons deciphering positively charged amino acid and negatively charged amino acid, respectively, we can gain access to the Maxwell-Boltzmann's factor of equation (26) modulating the rate of `STEP 2`, the peptide bond formation. We quantitatively compare the peptide bond formation rate when arginine (or lysine) is incorporated to the nascent chain at the PTC with the situation when glutamate (or aspartate) is incorporated. This allows to question whether the *ratios* between these rates are quantitatively consistent with the Maxwell-Boltzmann's factor modulation caused by the change in mechanical work, when a positive or negative `aa` is incorporated at the PTC, while the ribosome is deciphering the corresponding codon on the mRNA.

In summary, the theoretical workflow is to use the observed distribution of RRTs for a positively charged `aa` (resp. negatively charged `aa`), fit the hypo-exponential distribution, estimate the rate of `STEP 2` by educated deconvolution of the hypo-exponential distribution, compute the *ratio* of the two rates (positively and negatively charged `aa`) and compare with the Maxwell-Boltzmann modulation factor. To our knowledge, this workflow shows for the first time how Ribo-seq data at codon resolution, when properly deconvoluted, can be used to indirectly gain information on peptide bond formation relative rates along with the other sub-steps rates in the elongation cycle. The experimental results for the observed NFCs (and the derived RRTs) were extracted from file #007 (excel), publicly available, in the supplementary data of Dana and Tuller [44]. This file provides for each of the 61 sense codons, the exponentially modified Gaussian best fitted parameters to the observed NFC distribution across 4 species.

4. Results

The ribosome exit tunnel was modelled by a cylinder concatenated to a cone frustum in reference [26]. As a result of the scaling exposed in Material and Methods, the peptidyl transferase center (PTC) is approximately surrounded by a minimal surface prolate spheroidal volume accommodating for 3 tRNAs having their anticodon matching the 3 codons at the A, P and E site respectively. The distance between the amino acid residues acylated to their cognate tRNAs and their anticodon loops is 8 nm (the size of the L-shaped tRNA single molecule). The cylinder radius is 5 Å [10,26] and is connected to the PTC cavity by the truncated prolate spheroidal volume as shown in the simplified representation of Fig. 3(a). The angle between the support of the PTC arrow and the direction from the tunnel entry to the tunnel exit port is 38°.

4.1. Overview of the atlas of the 5 maps of the PTC shell cavity across the 3 domains of life and impact on the electrostatic potential profiles

The nucleotides residues of the 23S/28S rRNA which are taking part in the catalytic activity of the peptide bond formation, i.e. at the PTC, are not contiguous. Instead, as shown in Fig. 1, the PTC is formed by a diverse collection of distant residues that, when the rRNA is folded into its tridimensional shape, act together as a catalytic unit [16]. The universally most conserved nucleotides sequences of the 23S/28S rRNA across the three domains of life (better than 90% conservation across archea, prokarya and eukarya) were used to localize the homologous nucleotides that are known to take part in the peptide bond formation [17]. The key nucleotides belonging to the P-loop and A loop respectively are given in Table 1. The key nucleotides that are bold typeset in Table 1 in the P-loop and the A-loop are separated by a distance of 134 nucleotides in the rRNA primary sequence. In the tridimensional 23S/28S rRNA folded structure, these key nucleotides are at a physical distance ranging from 7 to 10 Å only. A number of studies have shown the mutational flexibilities of a number of the nucleotides residues around the PTC. Most mutations do not affect the transpeptidation activity [16]. The PTC is also enriched in modified residues but none of these modifications appear to be essential for peptide bond formation [16]. Here we further inquire if the functional importance of the universal conservation of the rRNA sequence can be visualized in the 3D structures of the PTC shell cavity. Fig. 4 shows the atlas of five 23S/28SrRNA tridimensional maps representing the PTC shell cavity of five species across the 3 domains of life. This atlas compares the tridimensional structures of the PTC shell cavities and shows strong common geometrical features for the 23S/28S rRNA ribose-phosphate backbone across the 5 species that might shed light on important functional roles. There are between 484 and 488 nucleotides of the 23S or 28S rRNA that are within a 40 Å distance of the key nucleotides of the P and A loop of Table 1. We used the 3D space positions of the phosphate moieties to compare the resulting electrostatic profile around the PTC and more specifically, we calculated the electrostatic profile along an arbitrary virtual path in the PTC shell cavity between the A-loop and the P-loop positions and towards the tunnel entry port.

The tridimensional spatial conservation of the 484-488 nucleotides from the 23S or 28S rRNA within a 40 Å distance of the PTC key nucleotides results in the conservation of the electrostatic potential profile along any arbitrary path from the A site to the P site or along the centerline from the PTC towards the ribosome nascent chain tunnel entry point. The common qualitative feature that is shared for the 5 species across the 3 domains of life is the funnel shape of the cavity when moving along an arbitrary path from the A-loop to the P-loop, and then towards the exit tunnel. Any arbitrary path along that direction encounters closer phosphate moieties positions. This causes the electrostatic potential to decrease when moving in the PTC shell cavity from the A-loop to the P-loop and then to the tunnel entry port. A decrease in the electrostatic potential means that the electric field points in the shell cavity from the A-loop to the P-loop and to the entry port of the tunnel. To support the effects of the qualitative observation of the shared common feature across species, Fig. 5 shows the calculated electrostatic potential profiles along an arbitrary straight line path starting around the PTC and heading towards the tunnel entry point for the two bacteria species. The potential also decreases when moving from the A-loop

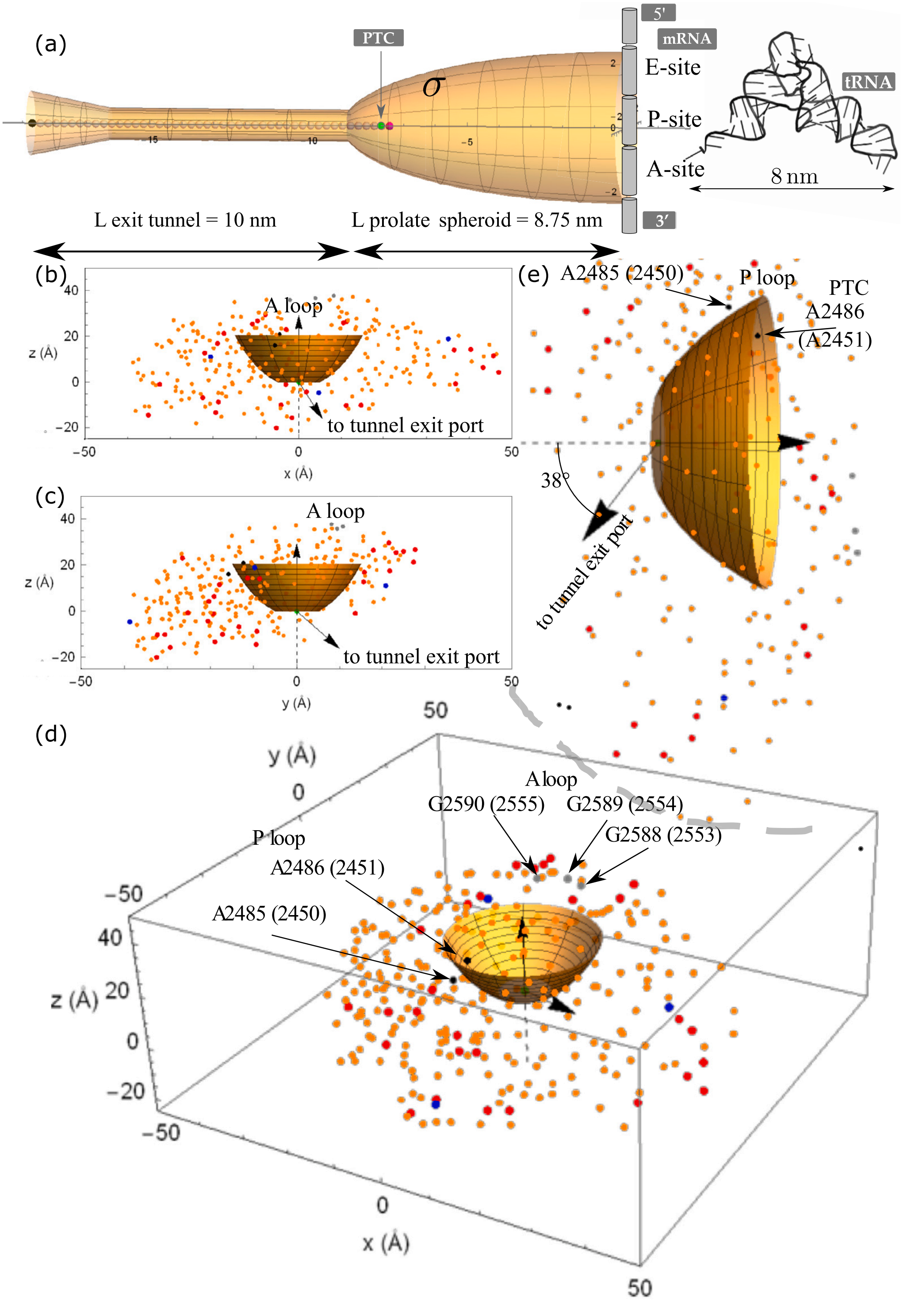


Fig. 3. (a) Ribosome exit tunnel and cavity around the PTC. (b) and (c) Front and right view of the charged groups neighboring the PTC cavity. (d) and (e) 3D scatter plot of the 291 charged atoms around the PTC. Phosphorus atoms bridging the 23S rRNA backbone (orange dots), positively charged moieties of lysine (NH2 atom in PyMol convention) or arginine (NZ atom in PyMol convention) belonging to ribosomal proteins (red dots), negatively charged moieties of aspartate or glutamate (OE2 or OD2 atom in PyMol convention) belonging to ribosomal proteins (blue dots), phosphorus atoms of P loop 23S rRNA nucleotides A2485 and A2486 (black dots), phosphorus atoms of A loop 23S rRNA nucleotides G2588, U2589 and U2590 (gray dots). The atomic positions were retrieved from *H. marismortui* x-ray solved structure of the large subunit of the ribosome.

to the P-loop (not shown here). In the immediate vicinity of the PTC, the electric field points in the direction from the A-loop to the P-loop and in the direction of the tunnel entry port. The calculated values of the electrostatic potential for *E.coli* and *T. thermophilus* show a decrease in the potential from −150 mV to −200 mV and −250 mV. The resulting estimated values for the axial force experienced by a positively charged test particle are in an averaged range of −20 pN to −30 pN. In the chosen z reference frame, the negative sign means the force points towards the tunnel for a positively charged test particle. These results are similar to the ones obtained below for the archeon *H. marismortui*.

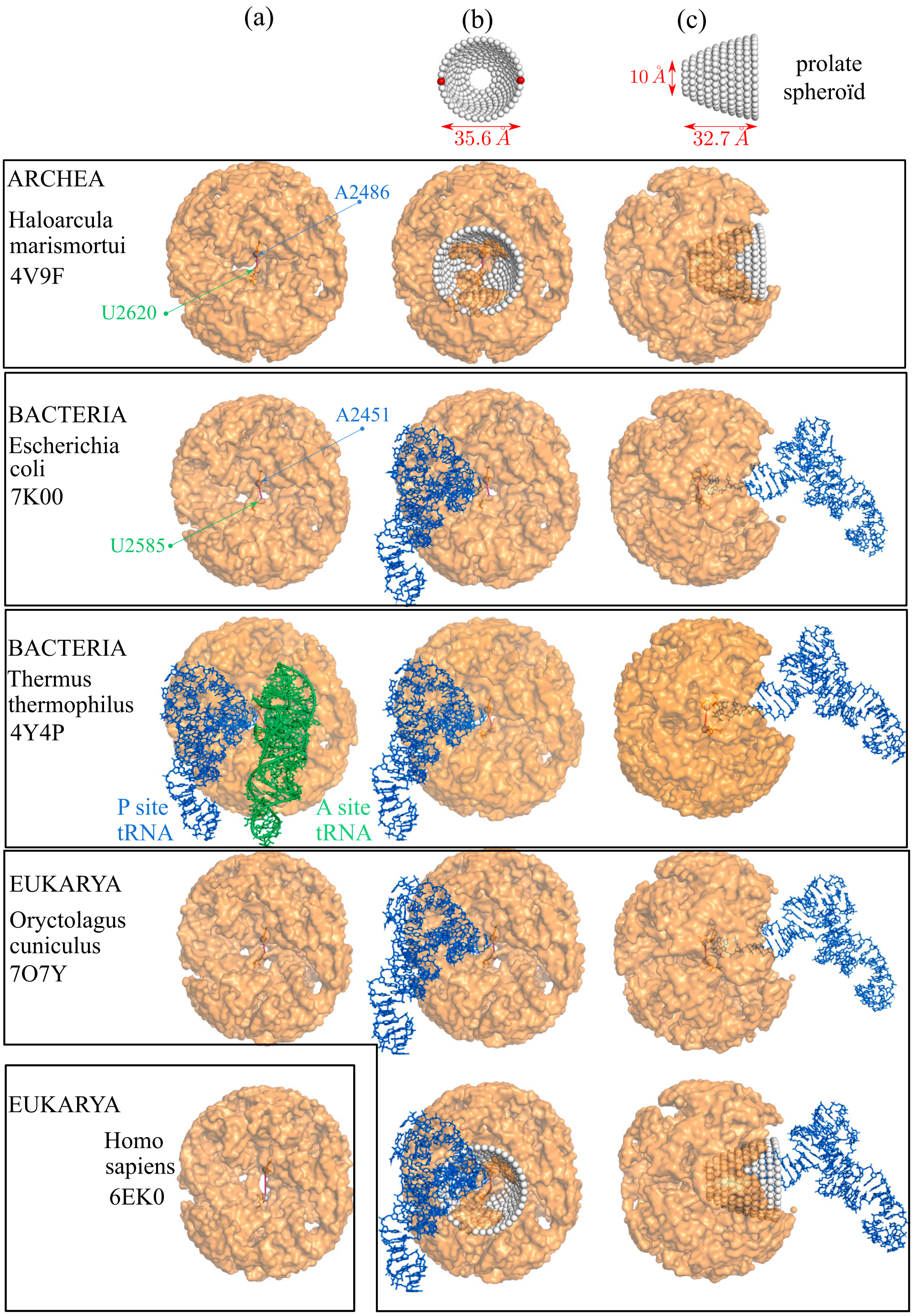


Fig. 4. Atlas of the tri-dimensional PTC shell cavity for 5 species across the 3 domains of life. (a) and (b) Front views of the shell cavity around the PTC. (c) Side views of the shell cavity around the PTC. The atomic positions were retrieved from the 5 atomic solved structures of the large subunit of the ribosome for PDB codes $4V9F$, $7K00$, $4Y4P$, $7O7Y$ and $6EK0$. A synthetic representation of the prolate spheroid was superposed in the front (b) and side views (c) for *Haloarcula marismortui* and *Oryctolagus cuniculus*.

4.2. Magnesium, potassium ions counts around the PTC shell cavity

Depending on the presence or not of tRNAs at A and P sites in the x-ray crystallographic solved structure of the ribosomes (depending on the PDB entry codes that are used), it is difficult to discriminate the metal ions (monovalent K^+ or bivalent Mg^{2+}) that are bound to the tRNAs structures from the ones that are bound to the 23S rRNA in the immediate vicinity of the PTC shell cavity or to determine whether these metal ions are free or bound. As shown in Fig. 6 (b) and(c), we localized the `CCA` nucleotides at the 3’ end of the tRNA molecules at the P and A sites of the *Thermus thermophilus* PDB code `4Y4P` solved structure of the ribosome. We counted and mapped the 3D positions of the metal ions (K^+ or Mg^{2+} indistinctively) that are within a $12\,\text{Å}$ distance of the $3'$−oxygen atom at the tip of the ribose of adenosine 76 at the P-site tRNA. Except for a limited number of papers, such as the one by Rozov et al., the bound metal ions counts around the PTC cavity were not of

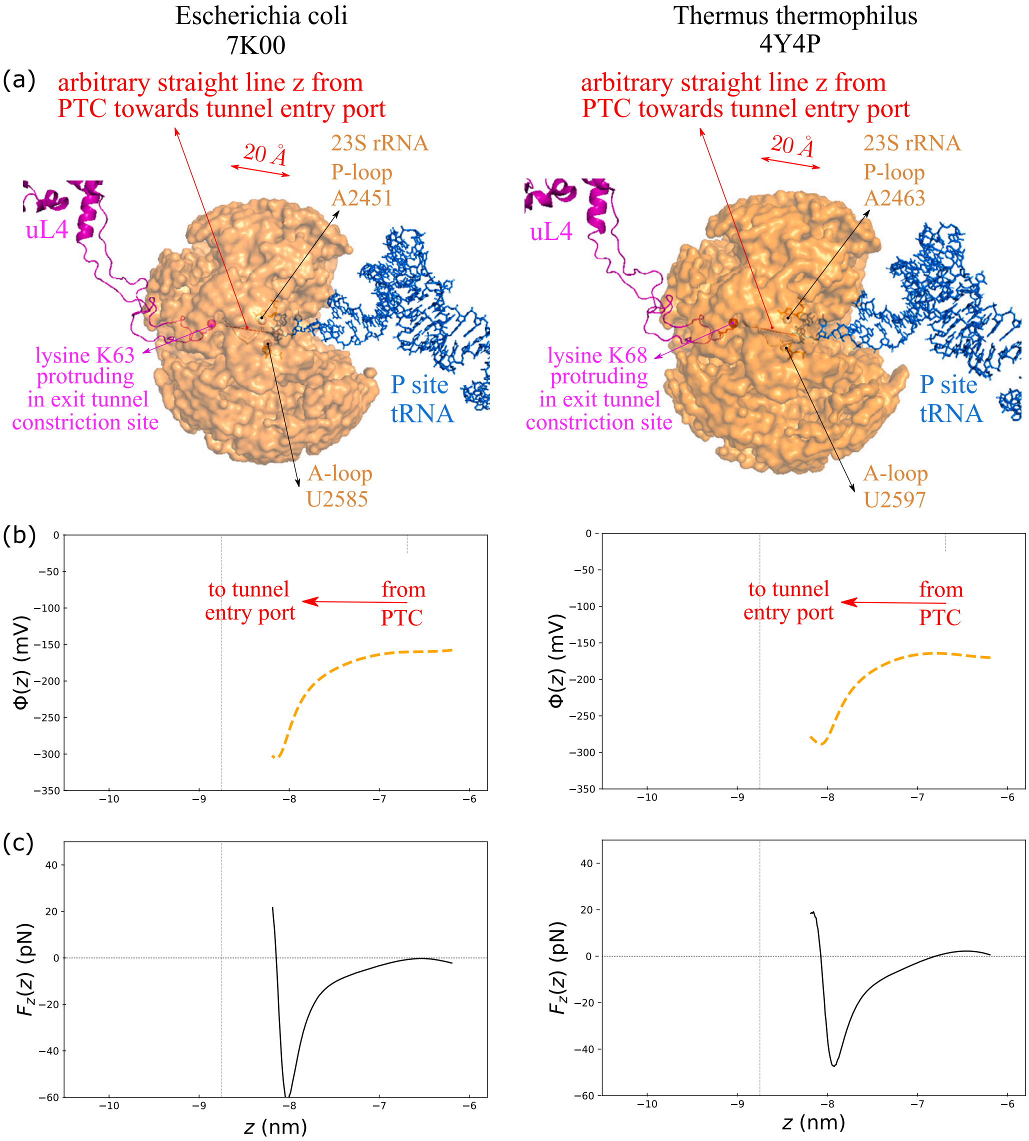


Fig. 5. Electrostatic potential profiles and electric field caused by the funnel shape of the PTC shell cavity. Comparison between species: *E.coli* (left), *T. thermophilus* (right). The arbitrary paths are taken as straight lines from the mid-point between A-loop and P-loop nucleotides to the lysine K amino acid in uL4 protruding in the tunnel constriction site. (a) Comparison of spatial representation of the x-ray solved structures of the 484 nucleotides of 23S rRNA in the immediate vicinity of the PTC. (b) Comparison of electrostatic potential $\Phi(z)$. (c) Comparison of axial force $F_z(z)$ along z path.

specific concerns. In their paper, Rozov *et al.* located seven metal ions next to the inner shell of the PTC for the ribosome with PDB code `4V6F` [4]. In the x-ray solved structure of the ribosome with PDB code `4Y4P` of *Thermus thermophilus*, our count of the bivalent Mg^{2+} ions would be 8 and there would be only one monovalent K^+ ion (and a total of 24 water molecules) within a 12Å distance of the 3′−oxygen atom at the tip of the ribose of adenosine 76 at the P-site tRNA, i.e., very close to the PTC shell cavity as shown in Fig. 6 (a). The metal ions and water molecules that are within a 12Å distance from all the phosphorus atoms belonging to the whole tRNA at the P-site are 34, 1 and 70 for Mg^{2+}, K^+ and water respectively as shown in Fig. 6 (c). Altogether, the metal ions and water molecules count is consistent with their screening role as free mobile ions of the 76 phosphate moieties of a single tRNA molecule at the P-site. For the tRNA molecule at the A site, there are 36, 0 and 68 counts of Mg^{2+}, K^+ and water respectively within a 12Å distance from all the phosphorus atoms belonging to the whole `tRNA` at the A-site (not shown here).

4.3. P-site and A-site tRNAs phosphate moieties and associated screening metal ions positions

The substrate of the peptide bond formation at A-site tRNA that will find itself accommodated in the PTC shell cavity of the ribosome for peptide bond formation is the amino acid residue that was loaded on its cognate tRNA by its specific aminoacyl tRNA synthetase. More precisely, it is the amino group, i.e. $-NH_2$, of the amino acid that will be processed in the PTC shell cavity along with the carboxy terminal part of the peptidyl-tRNA at the P-site. The nitrogen atom of this amino group will be closer to nucleotides of the 23S/28S rRNA (PTC shell cavity) than to any of the 76 nucleotides of the aminoacyl-tRNA. Fig. 6 (a) shows that the first closest phosphate moiety (P76) of the tRNA molecule to the amino group of the amino acid residue is at a distance of 8.7 Å. The second closest phosphate moiety of the tRNA is at a distance of 13 Å. This last distance is already larger than the screening length in the Yukawa-Debye-Hückel model. Most of the 76 phosphorus atoms and their non-bridging oxygen atoms are much more distant than the electrostatic screening length. The electrostatic effect of all the

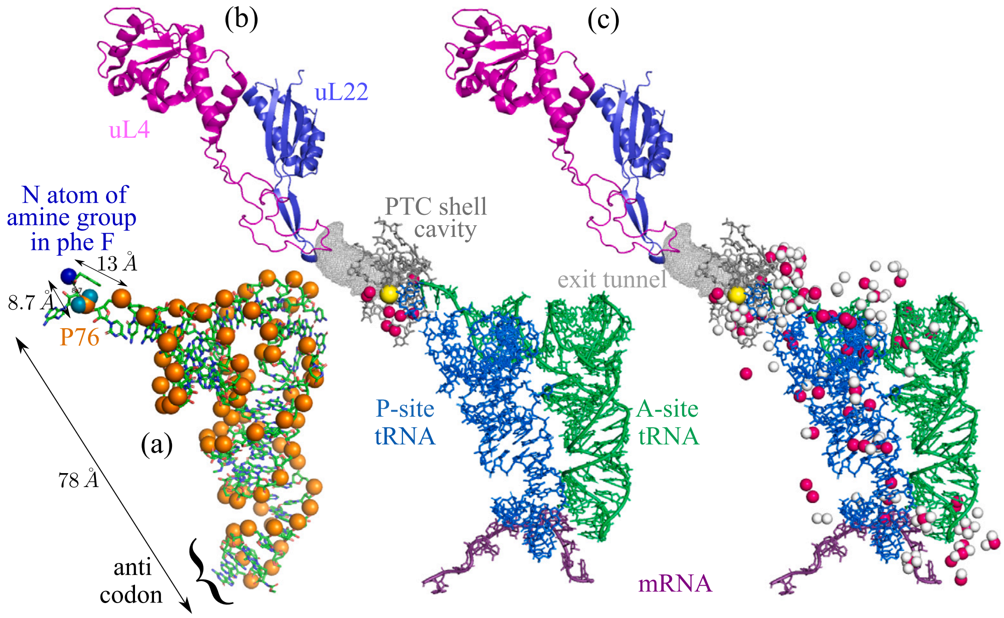


Fig. 6. (a) *Thermus aquaticus* PDB code 1TTT: tRNA structure of aminoacyl-tRNA^{phe} showing distances from amino group of the loaded amino acid to the 2 closest phosphate moieties of the tRNA; 76 phosphorus atoms in tRNA (orange spheres) (b) Mapping of metal ions around PTC shell cavity and tRNAs. *Thermus thermophilus* ribosome X-ray solved structure PDB code 4Y4P: 9 metal ions Mg^{2+} (pink spheres) or K^+ (yellow sphere) within 12Å of the tip of ribose $3'O$ atom of P-site tRNA adenosine 76 in close vicinity of the PTC shell cavity. (c) PDB code 4Y4P: metal ions (pink and yellow spheres) and water molecules (white spheres) within 12Å of all 76 phosphate moieties of the P-site tRNA. Cautionary note: the distinction between the metal ions Mg^{2+} or K^+ is elusive on these standard x-ray crystallographic solved structures of the ribosome and their tRNAs [4].

phosphate moieties belonging to the tRNA molecules is therefore expected to be negligible in the immediate vicinity of the 23S/28S rRNA PTC shell cavity where the peptide bond formation is catalyzed. The number of metal ion equivalents within a 12Å distance to the tRNAs is approximately the same as the number of charged equivalents in the phosphate groups belonging to the tRNAs. These ions, on average, are screening the phosphate groups for the tRNAs. The screening effect of these metal ions along with water molecules is such that the tRNAs taken as a whole with their metal ions do not contribute significantly to the electrostatic profile around the PTC or along the line from the PTC to the tunnel entry port. This is expected since the tRNAs shuttle in and out from the PTC cavity to the cytoplasm where they easily equilibrate with the ionic composition of the cytoplasm. These facts suggest that the tRNAs should not experience significant hindrance of electrostatic origin when they are accommodated at the A site and then at the P site. To substantiate quantitatively the fact that the phosphates belonging to the tRNAs both at P-site and A-site have negligible electrostatic effect around the PTC shell cavity (and taking into account the metal ions within screening distance of these tRNAs), we calculated the electrostatic potential along a virtual observation line close the two tRNAs and in the vicinity of the PTC cavity. We replicated the method used in section 3.2 and extracted 76 + 76 phosphorus atoms' positions of the tRNA molecules, and 35+36 metal ions from PDB file 4Y4P of the solved structure of ribosome for *Thermus thermophilus*, having the tRNAs at both A and P-sites. We constructed the virtual observation line (red straight line in Fig. 7 (c-d)) by taking a parallel straight line at a 5Å distance to the line joining the two oxygen 3-prime atoms of the riboses of adenine 76 at both tRNAs (P site and A site) and in the plane including nucleotide 34 of the P site tRNA (decoding center). This red line is approximately located in the PTC shell cavity where we want to estimate the electrostatic potential and electric field. Fig. 7 (a-b) shows that the contribution to the electrostatic potential and to the electric field caused by these charged moieties of the tRNAs accommodated at P and A sites is negligible when compared to the one caused by the 23S/28S rRNA in the immediate vicinity of the PTC shell cavity (within a distance of 7 – 12Å). The values, calculated with the Yukawa-Debye-Hückel theory, of the electrostatic potential around the PTC contributed by the 2 tRNAs are about −20 mV (Fig. 7(a)) while the values contributed by the 484 nucleotides of the 23S rRNA around the PTC are between −150 mV to −200 mV (Fig. 5 (b)). The potential profile of Fig. 7(a) is flat and the resulting contribution to the electric field or force is negligible: compare Fig. 7(b) (*T. thermophilus*) to Fig. 5(c) (*E.coli* and *T. thermophilus*) or Fig. 8(d) (*H. marismortui*).

At the scale of the tRNA full length (∼ 80 Å), the tRNAs are neutral. For electrostatic charges to be observed as unbalanced, the observation scale must be smaller than the Debye, Bjerrum or Gouy-Chapman screening lengths all of which are smaller than 10 Å.

4.4. Surface charge density and dielectric screening lengths around the PTC

Fig. 3(b-e) shows the 3D scatter plot of 291 charged groups in the region around the PTC and the tunnel entry port. The exact number of phosphate moieties around the PTC was counted from the x-ray solved structural data. The *bare* surface charge density σ^* was calculated by dividing the number of phosphate moieties (= 100) closer than 12 Å from the truncated spheroid by its surface area. The surface area was analytically calculated to be 15.75 nm^2 as detailed in the appendix A. Hence, the *bare* surface charge density σ^* is estimated to be $\sigma^* \sim -\frac{100}{15.75}\frac{|e|}{\mathrm{nm}^2} = -6.35\,\frac{|e|}{\mathrm{nm}^2}$ on the surface of the spheroid cavity around the PTC. This numerical result for the surface charge density is approximately three times higher than the surface charge density prevailing on the inner surface of the ribosome exit tunnel [26].

It is unclear whether Bjerrum or Debye theory should be used for the electrostatic screening length in the medium inside the PTC cavity. The Bjerrum screening length as computed from Eq. (20) is 0.72 nm when water permittivity $\epsilon = 78$ is used at 298 K inside the truncated spheroid medium around the PTC. The Debye screening length as computed from Eq. (21) is 1.1 nm when water permittivity $\epsilon = 78$ and an ionic strength $I = 75\,\mathrm{mM}$ (= $75\,\mathrm{mol/m^3}$) are used at 298 K inside the truncated spheroid medium around the PTC. Note that a Debye screening length equal to the Bjerrum length corresponds to the ionic strength's value of $I = 76\,\mathrm{mM}$ at 25°C and for water permittivity. The mathemat-

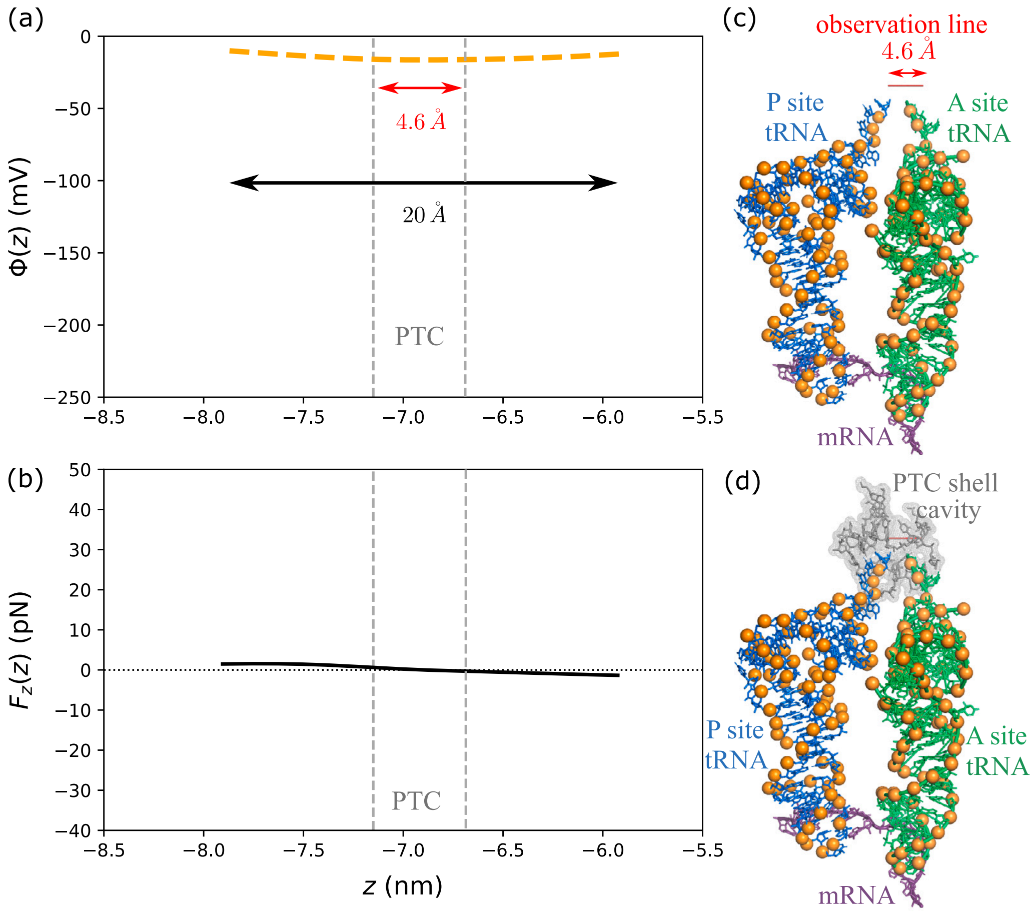


Fig. 7. Electrostatic potential contributed by *Thermus thermophilus* PDB code 4Y4P tRNA structures as accommodated at P and A sites of the ribosome showing phosphate moieties positions of the tRNAs. The potential and axial forces profiles are plotted with the same y-scales as the ones used for the contribution of the 23S/28S rRNA (see Fig. 5 and Fig. 8 for comparison) (a) Electrostatic potential profile (dashed orange) along the red straight line (c-d), in the vicinity of the PTC and contributed by the tRNAs. The potential profile was calculated over a 20 Å extended distance along the red straight line. (b) Axial force profile (black) in the vicinity of the PTC and contributed by the tRNAs. (c) Observation straight line (red line) in the vicinity of the PTC and the 152 phosphorus atoms of the tRNAs (orange spheres). (d) PTC shell cavity near the tRNAs' tips. 71 metal ions (Mg^{2+} or K^{+}) and 138 free water molecules (not shown here) contribute to the ionic strength screening the phosphate moieties.

ical expressions of both theories are then numerically equivalent and knowing which theory prevails is numerically irrelevant.

4.5. Comparison of the potential profiles and fields calculated from the spheroid idealized shape and from the x-ray solved structural data around the PTC

The x-ray solved structure dataset includes the exact 3D coordinate positions of a total of 291 charged atoms that are closest to the ribosome exit tunnel entry point or the PTC. Of these 291 atoms, 113 are at a distance shorter than 12 Å from the truncated spheroid's outer surface. The volume inside the truncated spheroid is empty, i.e., there are no 23S rRNA phosphorus atoms or ribosomal proteins charged amino acid moieties inside the volume. The charges which are at a distance larger than 12 Å from the outer surface of the cavity are so strongly screened by the dielectric medium that they can be neglected as was detailed in [26].

The electrostatic potential profile $\Phi(z)$ for the uniformly charged idealized spheroidal shape, as numerically calculated from Eq. (16), is shown in Fig. 8(c, black solid line). Upon implementing Eq. (19) in Python and using the exact positions of the 113 charged atoms within 12 Å of the spheroid's outer surface (as mapped in Fig. 3), we obtained the electrostatic potential along the spheroid's centerline shown in Fig. 8. The electrostatic potential contributed by the phosphate moieties (Fig. 8(a)) is negative while the potential contributed by the charged amino acid residues (Fig. 8(b)) is positive as arginine or lysine outnumber aspartate or glutamate residues. The net resulting electrostatic potential is negative, orange solid line in Fig. 8(c,) as there are ∼ 100 phosphate groups and only ∼ 13 charged amino acid residues at a distance closer than 12 Å from the PTC surface cavity. The charged amino acid residues are located much further away from the catalytic center than the phosphate moieties. The potential profile is due to the dominant presence of negatively charged phosphate moieties harbored by 23S/28S rRNA on the inner surface cavity in the immediate vicinity of the PTC.

The potential profiles calculated from the uniformly charged spheroid idealized shape (Eq. (16)) or calculated from the discrete sum of the Yukawa-Debye-Huckel formula over the charged atoms at their exact positions (Eq. (19)) are compared in Fig. 8(c). The two potential profiles look similar and provide an estimation of the order of magnitude of the electrostatic potential along the virtual centerline between the PTC and the entry port of the exit tunnel. The uniformly charged assumption of the surface in the region between the PTC and the tunnel entry port appears to be reasonable. The discrepancy between the two potential profiles (idealized spheroidal shape with uniform surface charge density versus exact atomic positions of the charges, shown in Fig. 8(c)), or axial electrostatic force profiles (Fig. 8(d)), is sensitive to the two phenomenological parameters of the models i.e., permittivity and screening length and their possible local heterogeneity (not shown here). Actual measurements of the electric field through vibra-

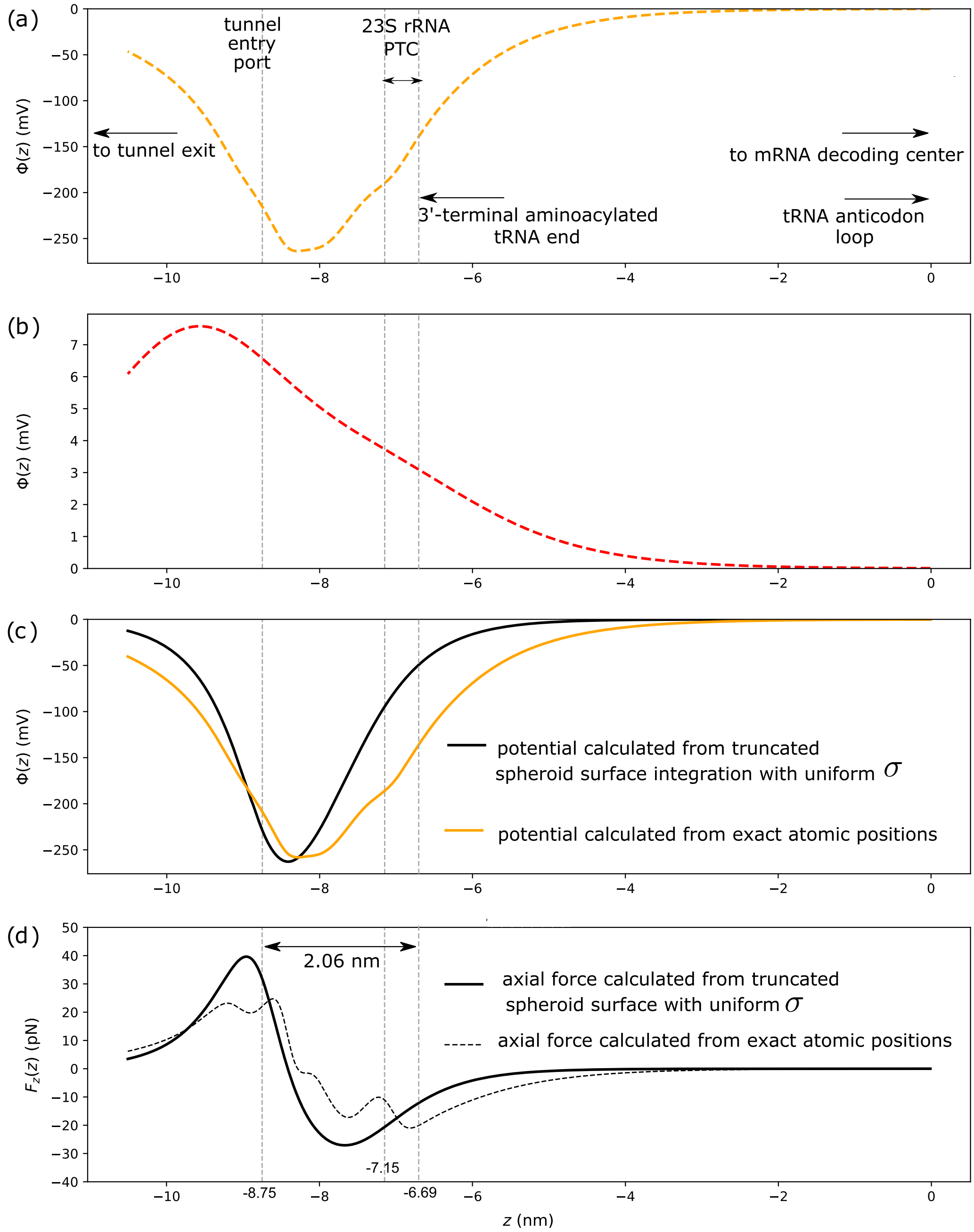


Fig. 8. Electrostatic potential profiles contributed by the ribosomal large subunit cavity around the PTC. (a) Electrostatic potential (mV) computed from the exact atomic positions of the 100 phosphorus within 12 Å of the surface cavity. (b) Electrostatic potential (mV) computed from the exact atomic positions of the 13 charged amino acid within 12 Å of the surface cavity. (c) Electrostatic potential (mV) computed from the exact atomic positions of all 113 charged moieties within 12 Å of the surface cavity (orange solid) compared to the potential resulting from a uniformly charged truncated spheroid surface of $\sigma^* = -6.35|e|/\mathrm{nm}^2$ (black solid). (d) Axial electrostatic force (pN) computed from the exact atomic positions of all 113 charged moieties within 12 Å of the surface cavity (gray dashed) compared to the force resulting from a uniformly charged truncated spheroid surface of $\sigma^* = -6.35|e|/\mathrm{nm}^2$ (black solid). All panels: 3 gray dashed vertical lines, from left to right: tunnel entry port position (radius= 5 Å), P loop A2485 z position, and truncated spheroid position where the cavity radius = 15 Å. A Bjerrum screening length of 0.72 nm was used for the analytical potential calculation and a Debye screening length of 1.1 nm was used in the discrete sum over the exact atomic positions, with ionic strength $I = 75$ mM. The water permittivity was assumed to be $\epsilon = 78$ inside the PTC cavity.

tional Stark effect spectroscopy would provide indirect support and constraints on the ranges of the local values for ϵ and ξ (permittivity and screening length).

The electrostatic potential particular profile results from the funnel shape of the cavity. The impact of the funnel shape on the potential profile is that when moving from the PTC to the tunnel entry port along the $z-$axis, the potential goes from smoothly convex (smaller curvature) to sharply convex (larger curvature) (Fig. 8(c)). The resulting negative inverse bell shaped peak for the electric field (or force) has a larger width in the region from the PTC to the tunnel entry port than the width of the positive bell shaped peak in the region near the tunnel entry port (Fig. 8(d)). The respective electric fields profiles and hence the forces along the $z-$axis centerline are also compared in Fig. 8(d). From these numerically estimated electrostatic force profiles, the maximum force at the center of the cavity neighboring the PTC would be between −21 and −27 pN for a unit positive test charge. The negative sign means that the force would point from the PTC to the tunnel entry port for a positive test charge.

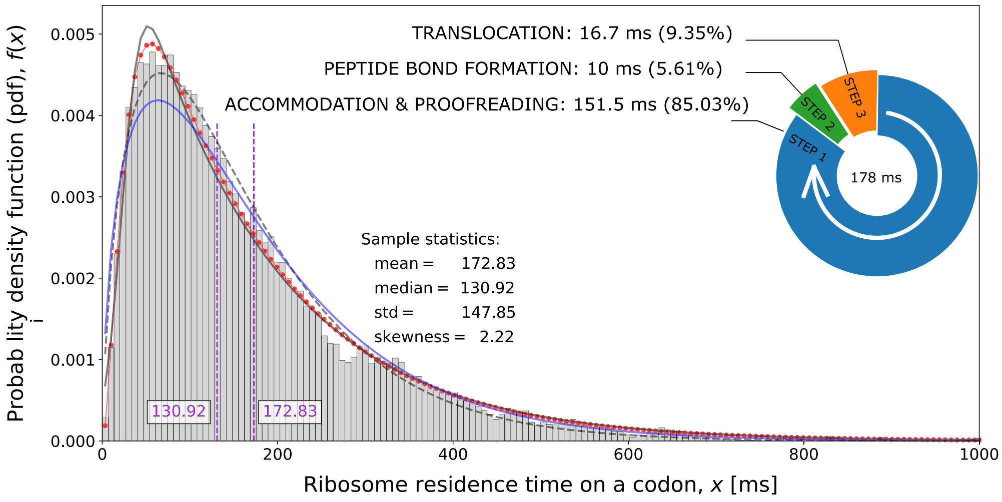


Fig. 9. Ribosome elongation cycle simulation in three sequential time steps. Hypo-exponential distribution with $1/\lambda_i = (151.5, 10.0, 16.7)$: red dots. Grey bars: frequency distribution sampled from the hypo-exponential distribution. Exponentially modified Gaussian distribution least sq. fit to hypo-exponential: grey line. Gamma distribution fit to frequency distribution, least sq.: dashed line; max. likelihood estimation: blue line.

4.6. Electric field estimation in the vicinity of the PTC

The ratio of the *bare* surface charge density over the medium permittivity σ^*/ϵ has the dimension of an electric field. In the absence of a screening layer of water molecules, the order of magnitude of this ratio in the immediate vicinity of the PTC cavity surface is $\sigma^*/\epsilon = -\frac{6.35\,|e|}{78\cdot\epsilon_0} \sim 1.473 \cdot 10^9$ V/m = 14.73 MV/cm, at least if the cavity medium is fully filled with water. If the medium in the PTC is only partially filled with water ($\epsilon = 78$), and the main components are ribonucleic acids (tRNAs) ($\epsilon = 8$) and the carboxy terminal end of the growing nascent protein ($\epsilon = 4$), than the resulting medium coarse-grained permittivity would be $\epsilon \sim 8$ (for a medium composed of 5% water and 95% protein). In this latter case, the estimated electric field, or the ratio $\sigma^*/\epsilon = -\frac{6.35\,|e|}{8\cdot\epsilon_0} \sim 14.36 \cdot 10^9$ V/m = 143.6 MV/cm.

Using vibrational Stark effect spectroscopy, Fried, Boxer and coworkers measured the electric field in a typical enzyme-substrate configuration at the catalytic site of the enzyme ketosteroid isomerase (KSI), with a magnitude of 144 ± 6 MV/cm [9,50]. Our numerical result shows that the order of magnitude of the electric field in the vicinity of the catalytic surface of the peptidyl transferase center of the ribosome is similar to that of the catalytic sites of known protein enzymes.

This intense electric field is possibly and locally electrostatically screened by constitutive water molecules of the PTC cavity inner wall. The reorientation of the permanent electric dipoles of constitutive water molecules on the inner surface of the PTC cavity would strongly damp the electrostatic potential and give rise to an apparent surface charge density $\sigma = \sigma^* \cdot e^{-\frac{\delta}{\xi_{GC}}}$ at a distance δ from the support of the *bare* charges, where $\xi_{GC} = 0.105$ nm is the Gouy-Chapman length. The Gouy-Chapman length is the screening length used when dealing with charges distributed on a surface wall and when the screening is mainly due to water molecules (not mobile ions) [26,31].

The apparent electric field along the z-axis as numerically calculated from the negative of the first derivative of Eq. (16)(Fig. 8(d) black solid line) or of Eq. (19) (Fig. 8(d) black dashed line), is approximately equal to 1.3 MV/cm. The corresponding axial force experienced by a positively charged unit test probe would be 21 – 27 pN around the PTC region (Fig. 8(d)).

The shape of the prolate spheroid with a uniformly charged inner surface results in a particular profile for the electrostatic potential and for the resultant electric field along the centerline of the cavity towards the tunnel entry point.

4.7. Impacts and functional significance of the electrostatic profile around the PTC on the peptide bond formation kinetics and on the complete elongation cycle

In higher eukaryotes, the protein elongation turnover is around 5.6 aa/s, i.e., the average elongation cycle has a time span of ∼ 178 ms. In bacteria and lower eukaryotes (e.g. yeasts) the elongation turnover is around 9.6 aa/s (∼ 105 ms). Three sequential steps are involved in the protein elongation cycle by the ribosome: (1) accommodation and proofreading, (2) peptide bond formation and (3) translocation. The insert of Fig. 9 shows the relative time spans of each step in the complete elongation cycle in the case of higher eukaryotes. The total time spent by a ribosome on a given codon results from the queueing time of these three steps. From a theoretical perspective, the ribosome residence time on an arbitrary codon is a stochastic process that can be viewed as a random variable which is the sum of three elementary independent random variables of the time spent in each step. The mathematical developments of this multi-step queueing theory in the applied fields of probability and stochastic modelling are detailed in appendix A.2. It is important to keep in mind this sequence of three time steps in the elongation cycle and the fact that the electrostatic interaction only affects the rate of `STEP 2` directly. Indeed, the global rate limiting step is `STEP 1`, i.e., the aa-`tRNA` A-site accommodation and proofreading at the `mRNA` decoding center, which occurs at the `16SrRNA-23SrRNA` interface, at a distance ∼ 80 Å away from the PTC. This first step is not dependent on the electrostatic potential around the PTC. The last step, `STEP 3`, i.e., the translocation to the next codon, is not dependent on the electrostatic potential around the PTC either. This explains why the impact of electrostatic interaction on the complete elongation cycle should not be overstated.

In the previous sections, we calculated the electrostatic potential profile around the PTC and from there the electric field and axial force acting on a charged test probe. Specializing on the *Escherichia coli* large ribosomal subunit 50S, the axial force at the PTC is estimated to be 21.2 pN, for a charged probe amino acid residue (Fig. 5 (c) and 8 (d)). This force keeps acting for the short time during the displacement $d\vec{\mathbf{x}}$ along the curvilinear path followed by a charged probe amino acid at transpeptidation. The dot product of the force by the probe small displacement yields the mechanical work upon curvilinear integration, $W = \int \vec{\mathbf{F}} \cdot d\vec{\mathbf{x}}$. This mechanical work is exerted by the surrounding electric field. The Gibbs free activation energy of the peptide bond formation is modulated by this mechanical work according to equations derived in section 3.4. The quantitative values obtained for the axial force are used to estimate the mechanical work and to

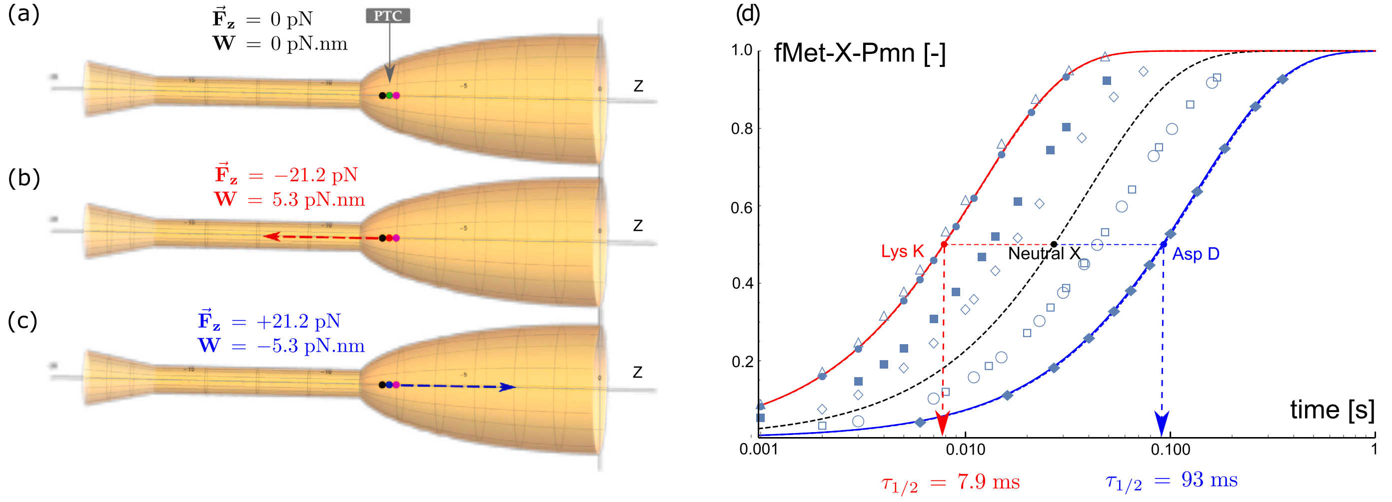


Fig. 10. Elongation minimal case: effect of force on the rate of peptide bond formation when the P-substrate is a dipeptidyl-tRNA (fMet-X-tRNA). Electrostatic force acting on X at the P site: **(a)** X = Neutral amino acid (green sphere). **(b)** X = Positively charged amino acid K or R (red sphere). **(c)** X = Negatively charged amino acid D or E (blue sphere). Methionine (black sphere). Puromycin (magenta sphere). **(d)** Experimentally measured normalized time courses of the Pmn (20mM) reaction with different dipeptidyl-tRNAs: Lys(△), Arg(•), Ala(■), Ser(◇), Phe(○), Val(□), Asp(◆) adapted with permission from reference [42]. Calculated theoretical normalized time courses using Maxwell-Boltzmann factors from equations (26) and (31). Neutral amino acid ($\tau_{1/2} = 27.1\,\text{ms}$), dashed line. Positively charged amino acid C-terminal dipeptidyl-transfer rate: neutral rate $\times \exp(\int \vec{\mathbf{F}} \cdot d\vec{\mathbf{x}}/k_B T)$ = neutral rate $\times \exp(5.3/4.28)$ = neutral rate $\times 3.45$, red line. Negatively charged amino acid dipeptidyl-transfer rate: neutral rate $\times \exp(\int \vec{\mathbf{F}} \cdot d\vec{\mathbf{x}}/k_B T)$ = neutral rate $\times \exp(-5.3/4.28)$ = neutral rate $\times 0.29$, blue line.

Table 2

Experimentally measured dipeptityl transfer rate constants (k_{pep}), apparent affinities of Pmn binding ($K_{\frac{1}{2}}$) and waiting time ($\tau_{1/2}$) to peptide bond formation event with a probability of 0.5; reproduced and adapted with permission from reference [42].

P-site dipeptide substrate	k_{pep} (s^{-1})	$K_{\frac{1}{2}}$ (mM)	$\tau_{\frac{1}{2}}$ (ms)	charge
$fMet - Lys - tRNA^{Lys}$	100 ± 7	14 ± 3	7.2	(+)
$fMet - Arg - tRNA^{Arg}$	90 ± 7	6 ± 3	7.8	(+)
$fMet - Ala - tRNA^{Ala}$	57 ± 4	35 ± 4	13.2	neutral
$fMet - Ser - tRNA^{Ser}$	44 ± 2	30 ± 3	17.0	neutral
$fMet - Phe - tRNA^{Phe}$	16 ± 1	4 ± 1	43.8	neutral
$fMet - Val - tRNA^{Val}$	16 ± 1	6 ± 1	44	neutral
$fMet - Asp - tRNA^{Asp}$	8 ± 1	22 ± 2	91.5	(−)
$fMet - Pro - tRNA^{Pro}$	0.14 ± 0.02	12 ± 4	5,102.2	neutral

calculate theoretically the impact on the peptide bond reaction rate. Any axial force acting on the nascent chain backbone will have similar effects. The electrostatic interaction of the nascent chain with the ribosome exit tunnel can also be estimated in the same way [26]. We compare these theoretical calculations to two independent sets of experimental results published previously by others [42,44]. In the next subsection, we examine the kinetics of the peptide bond formation between a dipeptide-tRNA and puromycin in *E.coli*, [42]. In the subsection after that, we examine the significance of the transpeptidation electrostatic modulation on the dwelling time distribution of the ribosome during the elongation cycle upon specific codons in *Saccharomyces cerevisiae* [44].

4.7.1. Kinetics experimental results for the transpeptidation minimal case with dipeptide-tRNA and puromycin in E.coli

The rate constants of the Pmn reaction with $\text{fMet} - \text{X} - \text{tRNA}^{\text{X}}$ are tabulated by decreasing order in Table 2 reproduced and adapted with permission from reference [42].

In the particular elongation minimal cases studied by Rodnina and coworkers [42], the peptidyl-tRNA is a dipeptidyl-tRNA with a very limited length that does not even reach the ribosome exit tunnel entry point (Fig. 10 (a-c)). When the X amino acid is neutral, no force is exerted at the P-site ($\vec{\mathbf{F}}_z = 0$ pN). On the contrary, if the X amino acid residue is a positively charged arginine R, or lysine K, a pulling force ($\vec{\mathbf{F}}_z = -21.2$ pN) is exerted on the backbone of the dipeptidyl-tRNA at the P site, while if the X amino acid residue is a negatively charged aspartate D, or glutamate E, a pushing force ($\vec{\mathbf{F}}_z = +21.2$ pN) is exerted on the backbone of the dipeptidyl-tRNA during the peptide bond formation when the A-site amino-group of puromycin comes close to the carbonyl carbon of the ester bond in the dipetidyl-tRNA at the P-site. In the former case (X = R or K), the force points toward the exit tunnel (negative $z-$ axis), whereas in the latter case (X = D or E), the force points toward the PTC (positive $z-$ axis). Upon elongation with puromycin as the last acceptor substrate (substrate 2), the displacement upon transpeptidation to the A-site is $d\vec{\mathbf{z}} \sim 0.25$ nm (median distance between two C$-\alpha$ in a peptide bond) and the mechanical work is estimated to be $W = 21.2 \cdot 0.25 = +5.3\,\text{pN} \cdot \text{nm}$ for R or K (force and displacement are parallel) and $W = -5.3\,\text{pN} \cdot$ nm for D or E (force and displacement are antiparallel). Using the final equation for the peptide bond formation kinetics as derived in equation (31) and using the quantitative values for the mechanical work in the two charged cases as compared to the neutral residue case, we predict the Maxwell-Boltzmann factors and the reaction rate constant values tabulated in Table 3. The time courses of the dipeptidyl-tRNA reaction with Pmn are shown in Fig. 10 (d). The red and blue lines are the time course calculated from equation (31) for the positively and negatively charged case respectively as compared to the neutral case (dashed line). These theoretical results are qualitatively and quantitatively consistent with the experimental values obtained by Rodnina and coworkers [42]. This provides indirect evidence of the forces caused by the electrostatic interaction of charged amino residues in the peptidyl-tRNA with the negatively charged phosphate moieties lining the inner surface of the ribosomal RNA around the PTC. These results also support the functional impact of these forces on the kinetics of the peptide bond formation.

4.7.2. Comparison of ribosome residence time empirical distribution results for the complete elongation cycle at specific codons in yeasts and bacteria coding for positively and negatively charged amino acids

In what follows, the elongation turnover for *S. cerevisiae* was taken as 9.5 aa/s [51]. The mean elongation time is $\text{mean}(T) = 105.3\,\text{ms}$ and the probability space for the ribosome residence time on a codon is expressed in millisecond units of time, ms.

Table 3

Maxwell-Boltzmann factors $\exp(\int \vec{\mathbf{F}} \cdot d\vec{\mathbf{x}}/k_B T)$ modulating the dipeptidyl transfer rate constants $k(\vec{\mathbf{F}})$ and the waiting time $\tau_{1/2}$ to peptide bond formation event with a probability of 0.5. Maxwell-Boltzmann factors and reaction rate constants are calculated from equation (26) or (31) at $T = 310.15\,K$, $k_B T = 4.282$ pN.nm. Acceptor substrate at A-site is puromycin.

P-site dipeptide substrate and C-terminal charge at P-site	Mechanical work (pN.nm)	Maxwell Boltzmann factor (-)	Rate constant $k(\vec{\mathbf{F}})$ (s^{-1})	$\tau_{1/2}$ (ms)
$fMet \odot -tRNA(0)$	0.0	1	$k(\mathbf{0})$	27.1
$fMet \oplus -tRNA(+1)$	+5.3	3.45	$3.45 \cdot k(\mathbf{0})$	7.9
$fMet \ominus -tRNA(-1)$	−5.3	0.29	$0.29 \cdot k(\mathbf{0})$	93

The elongation turnover for *E.coli* was taken as 16 aa/s, i.e., the mean elongation time is $\text{mean}(T) = 62.5$ ms at 37°C [52]. Note that the elongation turnover drops to 7.5 aa/s at 28°C in *E.coli* [53].

It is worth recalling that one should refrain from drawing conclusions on absolute times from Ribo-Seq derived data such as the normalized footprint counts (NFC) distributions, considering the definition equation (34) of the NFC. Comparisons of ribosome residence times (RRTs) between codons in terms of time fold changes are meaningful, but not comparisons in terms of absolute times.

The workflow detailed in section 3.5.2 provided the following results for the comparison of the deconvoluted elongation cycle for specific pairs of amino acid and their codons. In the four vertical panels of Fig. 11 and Fig. 12, pairwise comparison of the RRT distributions for positive versus negative amino acids deciphering codons are conducted. The first row plot in each panel shows the red and blue curve for the positive and negative amino acid respectively: Fig. 11(a(i), b(i)) and Fig. 12(a(i), b(i)). The second and third row plots shows, for the positively (Fig. 11 and 12 (ii)) respectively the negatively (Fig. 11 and 12 (iii)) charged amino acid, the Levenberg-Marquardt least squares fitting of the observed empirical exponentially modified Gaussian (full line) with the hypo-exponential distribution (dotted line). The inserts of the second and third row plots in each panel display the inferred individual rates $\lambda_1, \lambda_2, \lambda_3$ of the three exponentials as deconvoluted components of the hypo-exponentials, plus the time shift Δ. The quality of the fit, or similarity of the two statistical models, is assessed by the Kullback-Leibler (KL) divergence distance (relative entropy). A KL divergence distance close to zero indicates that the two models are very similar and convey the same information. For all eight best least squares fitting, the Kullback-Leibler divergence measures were smaller than 0.013, i.e., $KL(\text{EMG} \,||\, \text{HYPO}) < 0.013$.

A first broad inspection of Fig. 11 and Fig. 12 immediately shows that the ribosome on average spends significantly more time on average on triplets coding for glutamate or aspartate, i.e., the negatively charged amino acids, than for triplets coding for arginine or lysine *in S.cerevisiae*, i.e., the positively charged amino acids.

The educated deconvolution of the hypo-exponential distribution least square fitting into its three independent exponential components, plus time shift, provides an estimated breakdown of the three elongation sub-steps as shown on Fig. 11 a(ii), a(iii), b(ii), b(iii) and Fig. 12 a(ii), a(iii), b(ii), b(iii) and summed up in Table 4.

The smallest value of the λ_i, or largest of $1/\lambda_i$, is interpreted as the rate limiting step in the elongation cycle, i.e. the rate of accommodation and proofreading, `STEP 1`, with the other two λ values corresponding to `STEP 2` and `STEP 3`. Note that the time shift Δ can be interpreted as a time lag being part of.

The ratios of λ_2 are calculated for each pair of positively charged versus negatively charged amino acids in *Saccharomyces cerevisiae*.

Overall, the educated deconvolution results show that the ratio of the fast rate (for + amino acid) over the slow rate (for - amino acid) of STEP 2 is ~ 11.90. The inferred fold change as compared to the median rate for a neutral amino acid is $\sqrt{11.90} = 3.45$. The Maxwell-Boltzmann factor is quantitatively close to 3.45 (at 37 Celsius) and is consistent with our derivation of section 4.7.1.

5. Discussion

We studied the electrostatic environment around the catalytic center of the ribosome by using the Yukawa-Debye-Hückel theory applied to structural data from 5 publicly available x-ray solved structures of the ribosome across the three domains of life. The salient feature of the catalytic center of the ribosomal large subunit is that it is made only of nucleic acids (ribozymes). This is uncommon as the catalytic activity in biochemical processes are carried out mostly by proteins (enzymes). By contrast to proteins, nucleic acids have a molecular backbone harboring a very large number of phosphate moieties. Our study results suggest that the exact tridimensional distribution of these phosphate moieties has a functional role in the peptide bond formation and can affect its rate. The comparison of 5 species across the 3 domains of life of the ribosome catalytic center shell cavity shows that a common feature lies in the 3D shape of the distribution of non-adjacent 484 phosphate moieties within the polynucleotide backbone of 23S/28S rRNA at a 40 Å distance from the peptidyl transfer center (PTC). The tridimensional phosphate moieties distribution in the immediate vicinity of the PTC appears to be the main contributing structure that determines the electrostatic profile around the catalytic center. A simple truncated prolate spheroid shape fits the catalytic center of the ribosome. This funnel shape feature and its surface charge density generated by the fixed phosphate moieties explains the decrease in the electrostatic potential when a test charged probe moves from the A-loop to the P-loop and towards the tunnel entry port. To quantitatively determine the electrostatic potential profile and the electric field, the exact path followed by a newly incorporated amino acid should be known. To this date, we do not have this information. For simplicity, the charged amino acid residues participating in the reaction were approximated by immobile point charges, apparently equivalent to the positions of the α-atoms of the corresponding residues. However, in reality, these charges sit on the tip of rather large and highly mobile aliphatic chains. Lysine (K) and arginine (R) are notoriously large entities (~ 11Å), and even aspartate (D) and glutamate (E) are by no means small (6.5 and 8 Å, respectively, in a fully extended conformation). Their aliphatic segments are largely unconstrained, and rotation around the C-C bonds occurs with sub-nanosecond rates, which means that during the relatively slow transpeptidation, the charges carried on the tip of these amino acids can in principle sample a large volume around them. This volume is actually comparable to that of the entire PTC cavity (the distance between the PTC and the tunnel entry port is ~ 2 nm, and the amino acid displacement upon transpeptidation is only 0.25 nm). If these charges are indeed so mobile (i.e. their exact position is subject to high uncertainty), then, at this spatiotemporal scale, it might be difficult, by using the proposed framework, to measure the exact mechanical forces they experience. This mobility may even dampen the effect of the electrostatic field on the reaction energy barrier. In the best case scenario, since the conformational changes of the side chains must be quite rapid, one could envisage that arginine and lysine, directed by the field, would be relatively immobile and stretched in the direction of the tunnel entry port, bringing their positive charges 1 nm farther than expected from the α-carbon approximation. Based on Fig. 8(d), this would mean that they would experience an even stronger axial electrostatic force (closer to 30 pN). By contrast, aspartate and glutamate, stretched in the opposite direction, would suffer a lesser drag than expected by the theory. This side chain mobility is difficult to incorporate in the current model and is a limitation of the proposed coarse-grained framework.

The relative permittivity prevailing in the medium around the PTC is not known with accuracy and is not homogeneous. The medium inside the PTC cavity is more aqueous than the medium in the more confined micro-environment of the ribosome exit tunnel. It was hypothesized in this study that a coarse-grained permittivity should be taken in the

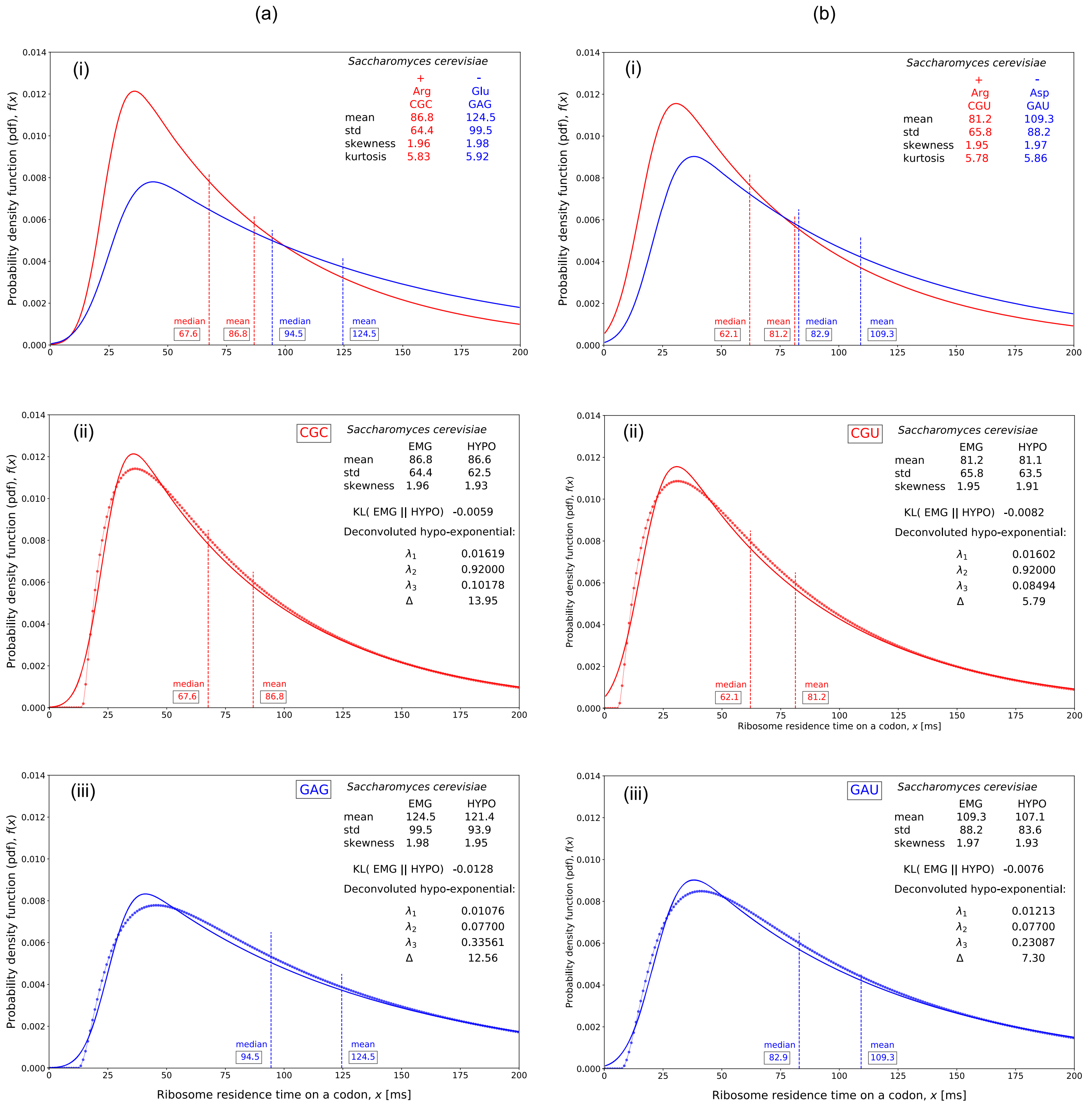


Fig. 11. Ribosome residence time distribution on specific codons in *Saccharomyces cerevisiae*. a,b-(i) Exponentially modified Gaussian empirical distributions as experimental reference comparing (+) red and (-) blue lines aa encoding codons. Data from [44] with permission. a,b-(ii) Least squares fitting with shifted hypo-exponential distribution for the (+) aa, red dotted line. a,b-(iii) Least squares fitting with shifted hypo-exponential distribution for the (-) aa, blue dotted line.

range corresponding to a mixture of protein, nucleic acid and water, i.e. at least between $\epsilon_r = 8$ and $\epsilon_r = 78$. The uncertainty on the numerical value of the permittivity on an extended spatial domain mainly affects the y-axis scale of the electrostatic potential. This is also modulated by the screening length. Some authors only provide a so-called arbitrary scale for the electrostatic potential to avoid these uncertainties on the empirical parameters [6,54].

Metal ions, Mg^{2+} and K^+ specifically, have long been described in the literature to be key players in stabilizing nucleic acid secondary and tertiary structures and also in stabilizing the ribosomal subunits [5,16]. For this reason, it is relevant to incorporate the positions of the metal ions when available in the structural data of the ribosome and in the vicinity of the PTC. Unfortunately, methodological shortcomings in the metal ions assignments and counts in the large ribosome deposited x-ray solved structures put the research community in a quandary [55]. Due to the elusive distinction between bivalent Mg^{2+}, monovalent K^+ or even oxygen of water molecules, the charge equivalents (valence) attributed to the metal ions is uncertain if the exact positions of these metal ions are used in the Yukawa-Debye-Hückel theory as source charges of the electrostatic potential and field [4]. The convenient way to deal with this metal ions assignment uncertainty in the field is to keep the metal ions, Mg^{2+} and K^+ free mobile ions indistinctively, as

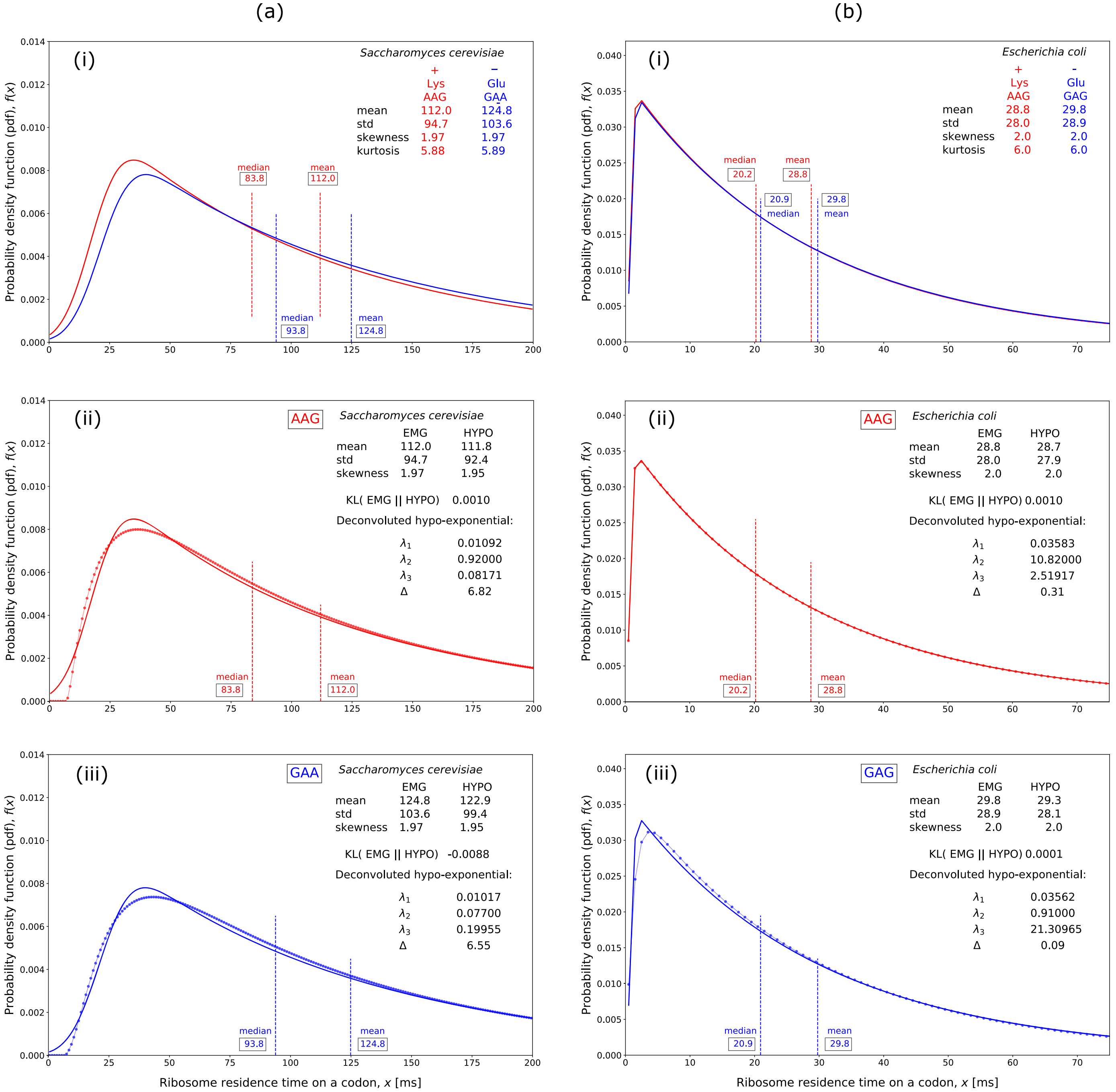


Fig. 12. Ribosome residence time distribution on specific codons in *Saccharomyces cerevisiae* and *Esherichia coli*. a,b-(i) Exponentially modified Gaussian empirical distributions as experimental reference comparing (+) red and (-) blue lines aa encoding codons. Data from [44] with permission. a,b-(ii) Least squares fitting with shifted hypo-exponential distribution for the (+) aa, red dotted line. a,b-(iii) Least squares fitting with shifted hypo-exponential distribution for the (-) aa, blue dotted line.

part of the empirical screening length parameter ξ generally adopted in the literature for the Yukawa-Debye-Hückel formula to calculate the electrostatic potential, without using the exact positions of these metal ions. The comparison of solved structures of ribosomes, around the PTC cavity for different species, shows that there is heterogeneity in the distribution of the metal ions across different x-ray or cryo-EM structural data from different species or PDB entries [4,6]. Given that, the assumption that the metal ions can be considered as free mobile ions screening the phosphate moieties fixed charges is reliable, at least around the PTC or around the tRNA molecules at the P and A sites.

Our study is based on x-ray or cryo-EM solved structures of the ribosome. These high resolution solved structures inherently are static data. In this static framework, the conformational rearrangements and rotation motions that occur during elongation, such as tRNAs transitioning to hybrid states and ribosome ratcheting were not taken into account. Dynamic behavior of rRNA nucleotides was also ignored. How these dynamic rearrangements could alter the electrostatic environment around the PTC is left for later molecular dynamics studies.

The present work does not explain the peptide bond formation rate dependence on the nature of the substrate at the A-site. Both pair of amino acids at the P and A-sites in the PTC shell cavity must play a role in the peptide bond formation rate. This has been partially covered by other studies. Organic chemistry steric hindrance, pH-sensitivity, water trap and entropy trap explain why different A site substrates incorporate

Table 4

Ribosome residence time hypo-exponential distribution deconvolution into its exponential rate components and time steps breakdown of the elongation cycle for codons deciphering positive and negative amino acids (λ, ms^{-1}; $1/\lambda$, Δ and RRT, ms).

				STEP 1		STEP 2		STEP 3			
				ACCOMMODATION & PROOFREADING		PEPTIDE BOND FORMATION		E-SITE EVICTION & TRANSLOCATION			ELONGATION CYCLE
species	charge	amino acid	codon	λ_1	$1/\lambda_1$	λ_2	$1/\lambda_2$	λ_3	$1/\lambda_3$	Δ	RRT
S. cerevisiae	(+)	`Arg R`	`CGC`	0.01619	61.76	0.92	1.09	0.10178	9.82	13.95	86.6
S. cerevisiae	(−)	`Glu E`	`GAG`	0.01076	92.9	0.077	12.90	0.33561	3.0	12.6	121.4
		λ_2 rate fold change for + over − =			$\frac{\lambda_2^+}{\lambda_2^-}$ =	11.90					
S. cerevisiae	(+)	`Arg R`	`CGU`	0.01602	62.41	0.92	1.09	0.08494	11.77	5.79	81.1
S. cerevisiae	(−)	`Asp D`	`GAU`	0.01213	82.5	0.077	12.90	0.23087	4.30	7.30	107.1
		λ_2 rate fold change for + over − =			$\frac{\lambda_2^+}{\lambda_2^-}$ =	11.90					
S. cerevisiae	(+)	`Lys K`	`AAG`	0.01092	91.62	0.92	1.09	0.08171	12.24	6.82	111.8
S. cerevisiae	(−)	`Glu E`	`GAA`	0.01017	98.4	0.077	12.90	0.19955	5.0	6.5	122.9
		λ_2 rate fold change for + over − =			$\frac{\lambda_2^+}{\lambda_2^-}$ =	11.90					
E. coli	(+)	`Lys K`	`AAG`	0.03583	27.91	10.82	0.09	2.51917	0.40	0.31	28.7
E. coli	(−)	`Glu E`	`GAG`	0.03562	28.1	0.91	1.1	21.30965	0.0	0.1	29.3
		λ_2 rate fold change for + over − =			$\frac{\lambda_2^+}{\lambda_2^-}$ =	11.90					

Cautionary note: comparisons between codons in terms of rate fold changes or time fold changes are meaningful but not comparisons in terms of absolute times.

at different rates in the PTC [32,33,56]. The A-site substrates context was also investigated in references [48,57,58]. Some amino acid like proline even require the help of specific elongation factors, e.g., EF-P or eIF5A [59–62]. Our model implication focuses on how the electrostatic environment around the immediate vicinity of the PTC (within a distance less than $5 - 10$Å) has its most significant impact in the energy barrier to overcome the deacylation of the peptidyl-tRNA at the P-site and this particularly appears when charged amino acids are at the carboxy-terminal end of the peptidyl-tRNA at the P-site.

The contribution of the tRNA molecules themselves at the P and A sites to the electrostatic environment in the catalytic shell cavity of the PTC was also considered. The analysis of structural data of the x-ray solved large ribosomal subunit including tRNA molecules showed that except for the last nucleotide at the $3'$−end of the tRNAs, all 76 phosphate moieties belonging to each of the tRNA at P and A sites are much further away from the PTC cavity than the Debye screening length. The electrostatic potential calculated contribution from the two tRNAs is much smaller than the calculated contribution from the 484-488 nucleotides of the 23S/28S rRNA in the vicinity of the PTC. The electrostatic field contributed by the two tRNAs fixed phosphate moieties is negligible in the vicinity of the PTC.

The significance and functional consequences of the electrostatic potential profile around the catalytic center are affecting the kinetics of the protein elongation rate. Following the literature in mechano-biochemistry, we have applied the classical Eyring theory of catalysis and incorporated a modulation of the Gibbs free energy activation energy barrier by the mechanical work of physical forces acting on at least one substrate. We hypothesized that the physical forces transmitted mainly through the backbone of the peptidyl-tRNA play a role in the reduction of the Gibbs free energy barrier of the transition state. During the peptidyl transfer reaction, the α-amino group of aminoacyl-tRNA positioned in the A site of the ribosome nucleophilically attacks the carbonyl carbon at the ester bond of the peptidyl-tRNA in the P site, which results in peptidyl-tRNA extended by one amino acid in the A site and deacylated tRNA in the P-site. Our analysis advocates that the biochemical deacylation of the peptidyl-tRNA at the P site appears to be facilitated (or hampered) by a pulling (pushing) force exerted on the backbone of oligopeptide attached at the $3'$−tip of tRNA at the P-site. If the carboxy-terminal end of the peptidyl-tRNA is a positively charged amino acid, a pulling force of electrostatic origin reduces the activation energy while if the carboxy-terminal end of the peptidyt-RNA is a negatively charged amino acid, a pushing force increases the activation energy. Overall, the 23S/28S rRNA catalytic shell cavity of the PTC functionally works like some kind of electrostatic bottle cap opener where the bottle part is the tRNA at the P-site and the cap is the carboxy-terminal end of the nascent peptide chain. The opening itself would be the deacylation of the peptidyl-tRNA at the P-site during the peptide bond formation. The development of the Michaelis-Menten kinetics incorporating the mechanochemical effect of the forces acting at the peptidyl transfer center led to theoretical quantitative predictions in the relative rates of peptide bond formation when comparing lysine or arginine to glutamate or aspartate as substrates at the carboxy-terminal end of the P-site tRNA. The consistency of these predictions with previously published experimental data sheds light on the significance and functional consequence of the electrostatic profile around the PTC, on the kinetics of the peptide bond formation and on its dependence on the nature of the second last incorporated amino acid residues. The peptide bond formation is the second step of the protein elongation cycle in the ribosome and the only one affected by the electrostatic interaction around the PTC shell cavity. We claimed that the global residence time of the ribosome on a given codon is a stochastic process resulting from the sum of three independent elementary queueing times, each of which being exponentially distributed. We relied on the queueing time theory in probability and statistics and used the convolution product to show that the ribosome residence time on a given codon should be hypo-exponentially distributed. The hypo-exponential distribution is the distribution best describing the sequential sub-steps involved in the elongation process. In the previous literature in the field, the exponentially modified Gaussian distribution and the Gamma distribution have been used as stochastic models for the RRT. In practice, we showed that the information conveyed by these distributions, HYPO, EMG or GAMMA, is very similar. These three distributions are statistically equivalent from the perspective of the information theory. However, the HYPO, when deconvoluted into its three exponential components has the advantage to provide the temporal breakdown of the sub-steps occurring during the elongation cycle. The educated deconvolution of the hypo-exponential distribution allows to use Ribo-Seq data or normalized footprint counts to infer kinetics information on the three sub-steps of elongation. To the knowledge of the authors, this is the first time that Ribo-Seq data were interpreted using a deconvolution of the hypo-exponential distributions that were previously fit to the ribosome normalized footprint count on codons.

6. Concluding remarks and future perspectives

One of the expected future developments in the field of protein synthesis and translational control will be in biochemical kinetics and dynamics. In this study, the structural data from x-ray crystallography was used to provide an inherently static picture of the peptide bond catalytic center. Molecular dynamics studies are expected to provide a better understanding of the dynamical interactions between key nucleotides of the P-loop and A-loop and the amino acid substrates loaded on the tRNAs at the P and A sites. Experimental studies using Stark effect spectroscopy, Förster resonance energy transfer and optical tweezers have been used and will continue to be used to probe the details of the elongation cycle and especially of the peptide bond formation dynamics in the ribosome in different contexts. To date, Ribo-Seq experimental results are difficult to interpret. The research community will greatly benefit from the development of computational biology models that will be able to generate simulated synthetic data under controlled settings. The comparison of data patterns between the simulated ribosome footprints on arbitrary transcripts and their real experimentally observed ribosome footprints (Ribo-Seq) will eventually help to disentangle the complex factors that are modulating the elongation cycle rates. The statistical queueing theory and the convolution product of the probability distribution for the queueing times of the sequential steps involved in the elongation cycle developed in this study provide fundamental insights in the stochastic behavior of the ribosome when a mRNA is translated *in singulo*. Our proposed interpretation of the normalized ribosome footprint profiles at codon resolution as a hypo-exponential distribution will improve data mining and parameter learning of existing Ribo-Seq big repositories like the Sequence Read Archive (SRA) and the European Read Archive (ENA). Future bioinformatics and machine learning studies trained on these repositories will help to gain better quantitative knowledge on the elongation cycle sub-steps and on the time spent by ribosomes on individual codons in different experimental settings and for species with different codon usage.

CRediT authorship contribution statement

Marc Joiret: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft. **Frederic Kerff:** Formal analysis, Methodology, Validation. **Francesca Rapino:** Conceptualization. **Pierre Close:** Conceptualization. **Liesbet Geris:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

All authors declare the absence of any conflict of interest.

Acknowledgements

This work was supported by the FNRS-FWO EOS grant n^o30480119 (Join-t-against-Osteoarthritis), the FNRS-WELBIO grant n^o*WELBIO-CR-2017S-02* (THERAtRAME) in Belgium and the European Research Council under the European Union's Horizon 2020 Framework Program (H2020/2014-2020) /ERC grant agreement n^o772418 (INSITE). We gratefully acknowledge the anonymous referees for their valuable contribution to the revised version of the manuscript.

Appendix A

A.1. Analytical solution for the area of the truncated prolate spheroid as a surface of revolution of a truncated ellipse

A prolate spheroid is a surface $S \in \mathbb{R}^3$ generated by the revolution of an ellipse about its major axis. A parametric representation of a simple ellipse in the plane Oxy in $\mathbb{R}^3$ having semi-major axis a and semi-minor axis b is

$$\gamma(u) = (a\cos u, b\sin u, 0). \tag{A.1}$$

The parametric representation of the surface of revolution is

$$\phi(u,v) = \big(|\gamma(u) \wedge \vec{e}|\cos(v), |\gamma(u) \wedge \vec{e}|\sin(v), <\gamma(u), \vec{e}>\big), \tag{A.2}$$

for all $(u,v) \in K$, with $K = [u_{lower}, u_{upper}] \times [0, 2\pi]$ and $\vec{e}$ the unit vector about which the revolution takes place. The area of the surface is determined by the formula known in elementary mathematical analysis

$$A_s = 2\pi \int_{u_{lower}}^{u_{upper}} |\gamma(u) \wedge \vec{e}| \sqrt{(D_u|\gamma \wedge \vec{e}|)^2 + (<D_u\gamma, \vec{e}>)^2}\, du. \tag{A.3}$$

Taking the revolution about the semi-major axis, we have $\vec{e} = (1,0,0)$ and

$$|\gamma(u) \wedge \vec{e}| = \left| det \begin{pmatrix} a\cos u & b\sin u & 0 \\ 1 & 0 & 0 \\ \vec{e}_x & \vec{e}_y & \vec{e}_z \end{pmatrix} \right| \tag{A.4}$$

$$= |-(b\sin u)\vec{e}_z| = b\sin u. \tag{A.5}$$

$$D_u|\gamma(u) \wedge \vec{e}| = b\cos u \tag{A.6}$$

$$<\gamma(u), \vec{e}> = a\cos u \tag{A.7}$$

$$D_u <\gamma(u), \vec{e}> = -a\sin u \tag{A.8}$$

$$A_s = 2\pi \int_{u_{lower}}^{u_{upper}} b\sin u \sqrt{b^2\cos^2 u + a^2\sin^2 u}\, du \tag{A.9}$$

$$= 2\pi \int_{u_{lower}}^{u_{upper}} b\sin u \sqrt{a^2(1-\cos^2 u) + b^2\cos^2 u}\, du \tag{A.10}$$

Substituting $t = \cos u$, $dt = -\sin u\, du$, we have

$$A_s = 2\pi\, b \int_{\cos u_{lower}}^{\cos u_{upper}} -\sqrt{a^2 - t^2(a^2-b^2)}\, dt \tag{A.11}$$

$$= 2\pi\, ab \int_{\cos u_{upper}}^{\cos u_{lower}} \sqrt{1-e^2t^2}\, dt \tag{A.12}$$

where, in the previous line, the eccentricity of the ellipse $e = \sqrt{\frac{a^2-b^2}{a^2}}$ was used. Substituting again $et = \sin w$, $e\,dt = \cos w\, dw$, the last integral turns into

$$A_s = 2\pi\, \frac{ab}{e} \int_*^* \sqrt{1-\sin^2 w} \cos w\, dw \tag{A.13}$$

$$= 2\pi\, \frac{ab}{e} \int_*^* \cos^2 w\, dw \tag{A.14}$$

$$= 2\pi\, \frac{ab}{e} \int_*^* (\frac{1}{2} + \frac{1}{2}\cos 2w)\, dw \tag{A.15}$$

$$= 2\pi\, \frac{ab}{e} [\frac{w}{2} + \frac{1}{4}\sin 2w]_*^* \tag{A.16}$$

$$= \frac{\pi ab}{e} [w + \frac{1}{2} 2\sin w\cos w]_*^* \tag{A.17}$$

$$= \frac{\pi ab}{e} \left[\arcsin(et) + et\sqrt{1-e^2t^2} \right]_{t_{lower}=\cos u_{upper}}^{t_{upper}=\cos u_{lower}}. \tag{A.18}$$

The area of the truncated prolate spheroid of semi-major axis $a = 8.9744\,\text{nm}$ and semi-minor axis $b = 2.25\,\text{nm}$ (eccentricity $e = 0.9681$), for t in the range $t_{lower} = \cos u_{upper} = 0.7454$ and $t_{upper} = \cos u_{lower} = 0.9750$, is calculated from this last formula to be $15.75\,\text{nm}^2$. Note that, when $t \in [0,1]$ or $u \in [\frac{\pi}{2}, \pi]$, the surface of the half prolate spheroid is $\frac{\pi ab}{e}\big(\arcsin(e) + e\sqrt{1-e^2}\big) = 102.23\,\text{nm}^2$ with the adopted values of a and b.

A.2. Queueing time theory: sum of exponentially distributed random variables with arbitrary parameters, hypo-exponential, gamma and exponentially modified Gaussian density distributions

This appendix reviews important mathematical results of the probability and statistics theory related to the density of sums of independent random variables having each a specific density distribution.

In all probability and statistics introductory courses, the Gaussian distribution plays a key role because of the so-called Central Limit Theorem (CLT). The strong version of the Central Limit Theorem (Laplace-Lyapounov) asserts that if a random variable X is the sum of a large number of independent random variables X_i, each with unknown arbitrary distinct and standardized distributions (re-centered and re-scaled), then $X = 1/n \sum_{i=1}^{n} X_i$ is approximately normally distributed and $X \sim N(x; \mu = 0, \sigma^2 = 1)$ when $n \to \infty$. The CLT theorem does not tell how fast the convergence is to the normal distribution. In fact, for a limited number of independent contributing random variable terms, the normal distribution may not be a good approximation at all, especially if the contributing variables have asymmetric distributions (one-sided tails or skewed to one direction).

Here, we want to focus on cases where the independent random variables taken as the terms of the sum are present in a limited finite number, e.g. 2, 3 or 5 contributing independent distributions (not a number close to infinity); may have different pairwise means or variances; and may be highly skewed. We are not looking for an asymptotic approximation but for exact or quasi-exact results. More specifically, we re-derive below the hypo-exponential probability density function as being that for the sum of independent exponentials having pairwise distinct arbitrary parameters. We further assess whether highly skewed classical distributions with 2 (or 3) parameters such as the (shifted) Gamma distribution or the exponentially modified Gaussian distribution or the (shifted) hypo-exponential distribution can be suitably fitted to one another. The quality of the fit will be assessed by three criteria, i.e., the Kullback-Leibler (KL) divergence criterion, the Akaike Infomartion Criterion (AIC) and the Bayesian Information Criterion (BIC).

Exponentially distributed random variables are prevalent in the applied fields of probability and in stochastic modelling. One of the most used of such models, the Poisson process, the interarrival times of events are independent exponential random variables as a consequence of its postulates [63,64]. Combining such processes in the development of his method of stages (or steps or phases) in queueing models, the Danish statistician Agner Erlang was led to introduce what is now the familiar Erlang distribution in the context of assessing the reliability of communication call centers between 1909 and 1920 [65]. The Erlang distribution is the distribution of an integer sum of independent and identically distributed (*iid*) exponentials with a common rate parameter λ ($\lambda_i = \lambda$, $\forall i$). The Erlang distribution is a particular case of a Gamma distribution for which the shape parameter α is an integer number and the scale parameter $\beta = 1/\lambda$, the inverse of the rate parameter λ, is such that $\beta = 1/\lambda = \sum_{i=1}^{n} 1/\lambda_i$ of the contributing *iid* exponential terms.

A.2.1. Probability density function for the sum of random variables as a convolution product of the probability density functions of the terms in the sum

Let X_f and Y_g be two independent random variables having, as probability distribution functions, respectively $f(x)$ and $g(x)$; then the random variable Z being the sum of X_f and Y_g has a probability density function which is the convolution product of $f(x)$ and $g(x)$:

$$f_Z(x) = (f \star g)(x) = \int_{\mathbb{R}} f(y) \cdot g(x-y)dy \tag{A.19}$$

Note that X_f and Y_g need not to be identically distributed. Formula (A.19) can be generalized to more than two random variables and also can be generalized in $\mathbb{R}^n$. Both commutativity and associativity hold for the convolution product operation.

If the family of the distribution for the convolution product belongs to the same family as the contributing distributions in the sum of the random variables, this family is said to be stable. By definition, a distribution is called stable if a linear combination of two random variables drawn from it also has the same distribution (up to location and scale parameters). For instance, a sum of Poisson distributed random variables is still a Poisson random variable. This also holds for the sum of Gaussian random variables, hypo-exponentially distributed random variables (if all the contributing rates are pairwise distinct), chi-squared distributed random variables, Erlang distributed random variables, Gamma distributed random variables. All these families are stable. But the exponentially modified Gaussian family is not stable.

A.2.2. Poisson process, exponential distribution and the memoryless property

Poisson experiments are experiments where the number of outcomes occurring during a given time interval, or in a specified region, are counted. The given time interval may be of any length. The number of outcomes is a random variable called a Poisson random variable and its probability distribution is a Poisson density function. The Poisson distribution is discrete as the outcome count takes discrete integer values. A Poisson experiment is derived from the Poisson process which has the three following properties:

1) The number of outcomes occurring in one time interval (or specified region of space) is independent of the number that occur in any other disjoint time interval (or region). It is said that Poisson process has no memory.
2) The probability that a single outcome will occur during a very short time interval (or in a very small region) is proportional to the length of the time interval and does not depend on the number of outcomes occurring outside this time interval.
3) The probability that more than one outcome will occur in such a short time interval is negligible.

The mean number of outcomes is computed from $\mu = \lambda t$ where t is the specific unit time interval. So, λ is the mean number of outcomes per unit of time and is called the rate (of occurrence of outcomes). From the three above properties, it follows that the Poisson probability density function is given by [64]:

$$f_{X_{POISSON}}(x; \lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} \tag{A.20}$$

$$f_{X_{POISSON}}(x; \mu) = \frac{e^{-\mu}\mu^x}{x!} \tag{A.21}$$

The salient feature of the Poisson distribution is that it depends on a single parameter μ or the rate λ and that the mean, μ, is equal to the variance, $\sigma^2 = \mu$ (for the Poisson distribution).

The exponential probability density function is related to the Poisson process. Consider a random variable X_{EXP} describing the time required for the first Poisson event to occur in a Poisson process. Using the Poisson distribution, we find that the probability of no events occurring in the span up to time t is given by:

$$P(X_{POISSON} = 0) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} \tag{A.22}$$

$$= e^{-\lambda t} \tag{A.23}$$

The probability that X_{EXP} will exceed x is the same as the probability that no Poisson event occurs within the time span x, the latter was just given by $e^{-\lambda x}$. Hence,

$$P(X_{EXP} > x) = e^{-\lambda x}. \tag{A.24}$$

Thus the cumulative distribution function (CDF) for X_{EXP} is given by:

$$P(0 \le X_{EXP} \le x) = 1 - P(X_{EXP} > x) \tag{A.25}$$

$$CDF(x) = 1 - e^{-\lambda x}. \tag{A.26}$$

The probability density function for X_{EXP} results from the derivative of its CDF:

$$f_{X_{EXP}}(x) = \lambda e^{-\lambda x}. \tag{A.27}$$

Let X_i be a random variable having the exponential distribution with rate parameter $\lambda_i > 0$. Then its probability density function, $f_{X_i}(t)$, is given by:

$$f_{X_i}(t) = \begin{cases} \lambda_i e^{-\lambda_i t} & t \geq 0 \\ 0 & t < 0. \end{cases} \tag{A.28}$$

The main descriptive statistics parameters of the exponential density are:

$$\mathbb{E}(X_{EXP}) = \beta = 1/\lambda \tag{A.29}$$

$$\mathrm{VAR}(X_{EXP}) = \beta^2 = 1/\lambda^2 \tag{A.30}$$

$$\text{skewness} = 2 \tag{A.31}$$

$$\text{Fisher kurtosis} = 6. \tag{A.32}$$

The median is $\ln 2/\lambda$, which is always smaller than the mean. The exponential distribution is asymmetric with a heavy one-sided right tail (skewed to the right).

The exponential distribution has a memoryless property (lack of memory). This lack of memory means that the conditional probability for X being larger than $t + t_0$ given that X was larger than t_0 is equal to the probability that X was larger than t. So if the waiting time or queueing time 'makes it' to t_0, the probability of waiting an additional t time is the same as the probability of waiting t time in the first place (as if you would have reset the queueing time from the start). There is no 'benefit' from patience or no 'punishment' through wear that may have ensued for lasting the first t_0. Thus the exponential distribution is appropriate when the memoryless property is justified. When patience pays off or ageing effects occur, other distributions than the exponential distribution are more appropriate as will be seen in the next subsections.

A.2.3. Erlang and Gamma distributions and the loss of the memoryless property

Before we proceed to the Gamma distribution and its particular case, the Erlang distribution, the Euler gamma function must be recalled. The Euler's gamma function is defined by:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \quad \text{for } \alpha > 0. \tag{A.33}$$

The following properties hold for the gamma function:

(a) $\Gamma(n) = (n-1)!$ for positive integer n.
(b) $\Gamma(1) = 1$.
(c) $\Gamma(1/2) = \sqrt{\pi}$

The Gamma probability density function is defined by:

$$f_{X_{GAMMA}}(x; \alpha, \beta) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x \geq 0 \\ \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{A.34}$$

where the shape parameter is a positive real $\alpha > 0$ and the scale parameter is a positive real $\beta = 1/\lambda > 0$. The main descriptive statistics parameters of the Gamma density are:

$$\mathbb{E}(X_{GAMMA}) = \alpha/\beta \tag{A.35}$$

$$\mathrm{VAR}(X_{GAMMA}) = \alpha/\beta^2 \tag{A.36}$$

$$\text{skewness} = 2/\sqrt{\alpha} \tag{A.37}$$

$$\text{Fisher kurtosis} = 6/\alpha \tag{A.38}$$

The median has no analytically closed form but can be calculated by numerical quadrature.

The maximum likelihood estimators for the shape and rate parameters of the Gamma distribution when a data sample of size n is available are:

$$\hat{\beta} = \frac{n^2}{n \sum_{i=1}^n x_i \ln x_i - \sum_{i=1}^n \ln x_i \sum_{i=1}^n x_i}, \quad \text{biased} \tag{A.39}$$

$$\tilde{\beta} = \frac{n-1}{n} \hat{\beta}, \quad \text{unbiased} \tag{A.40}$$

$$\hat{\alpha} = \frac{\hat{\beta}}{n} \sum_{i=1}^n x_i, \quad \text{biased} \tag{A.41}$$

$$\tilde{\alpha} = \hat{\alpha} - \frac{1}{n} \left(3\hat{\alpha} - \frac{2}{3} \frac{\hat{\alpha}}{\hat{\alpha}+1} - \frac{4}{5} \frac{\hat{\alpha}}{(\hat{\alpha}+1)^2} \right), \quad \text{unbiased.} \tag{A.42}$$

Assumption: the data are supposed to have been sampled from a Gamma distribution. Note the differences in the above formula with respect to the classical formula for data that would have been sampled from a Gaussian distribution.

The exponential distribution is a particular case of a Gamma distribution when $\alpha = 1$. The Erlang distribution is a particular case of the Gamma distribution when α is a positive integer. The Erlang distribution results from the sum of α *iid* exponentials having all a common rate $\lambda = 1/\beta$. Erlang distributions are the distributions of the waiting time for a specific number (more than one) of Poisson events to occur with the assumption that all these Poisson events have the same rate λ. Erlang distributions are stable.

The Gamma distribution can be viewed as a generalization of the Erlang when α, the shape parameter is not restricted to be an integer but is extended to real positive values.

The memoryless property does not hold for the Erlang or the Gamma distributions. If a queueing time or an event waiting time is a result of a gradual or slow wear or results from multiple time steps that need to proceed sequentially, then the memoryless property does not hold and Gamma distributions or hypo-exponential distributions can be appropriate stochastic models in these situations.

A.2.4. Hypo-exponential density as a sum of independent exponentials having arbitrary pairwise distinct parameters

The sum of n mutually independent exponential random variables, X_i, with pairwise distinct parameters, $\lambda_i, i = 1, \ldots, n$, respectively, has the hypo-exponential probability density function, given by [64]:

$$f_{HYPO}(t) = \Big(\prod_{i=1}^n \lambda_i \Big) \sum_{j=1}^n \frac{e^{-\lambda_j t}}{\prod_{k=1,\neq j}^n (\lambda_k - \lambda_j)}, \quad t \geq 0. \tag{A.43}$$

The condition that the λ_i's be distinct is essential as the formula (A.43) is undefined for any instance where $\lambda_i = \lambda_j$ for $i \neq j$. In the case where $n = 3$, we explicitly have:

$$f_{HYPO}(t) = \lambda_1 \lambda_2 \lambda_3 \left(\frac{e^{-\lambda_1 t}}{(\lambda_3 - \lambda_1)(\lambda_2 - \lambda_1)} + \frac{e^{-\lambda_2 t}}{(\lambda_3 - \lambda_2)(\lambda_1 - \lambda_2)} + \frac{e^{-\lambda_3 t}}{(\lambda_2 - \lambda_3)(\lambda_1 - \lambda_3)} \right). \tag{A.44}$$

The main descriptive statistics parameters of the hypo-exponential density are:

$$\mathbb{E}(X_{HYPO}) = \sum_{i=1}^n 1/\lambda_i \tag{A.45}$$

$$\mathrm{VAR}(X_{HYPO}) = \sum_{i=1}^n 1/\lambda_i^2 \tag{A.46}$$

$$\text{skewness} = 2 \Big(\sum_{i=1}^n 1/\lambda_i^3 \Big) / \Big(\sum_{i=1}^n 1/\lambda_i^2 \Big)^{3/2} \tag{A.47}$$

$$\text{Fisher kurtosis} = \text{no simple closed form.} \tag{A.48}$$

The median has no analytically closed form but can be calculated by numerical quadrature.

Hypo-exponential distributions are stable if all the contributing exponentials have pairwise distinct rates.

It can easily be shown that the memoryless property does not hold for the distribution of the sum of two independent exponential distributions [66].

A supplementary material `.mp4` animation shows the geometric interpretation of the convolution product to calculate the probability distribution function resulting from the sum of three exponentially distributed random variables.

A.2.5. Exponentially modified Gaussian density

The sum of a normally distributed random variable having the two parameters (mean μ and variance σ^2), with an independent exponentially distributed random variable having one parameter (rate λ), is a random variable that has an exponentially modified Gaussian probability distribution (EMG) having 3 parameters μ, σ, λ. The probability density function resulting from this sum of random variables is expressed by:

$$f_{EMG}(x) = \frac{\lambda}{2} e^{\frac{\lambda^2\sigma^2}{2}} e^{-\lambda(x-\mu)} \,\mathrm{Erfc}\left[\frac{1}{\sqrt{2}}\left(\lambda\sigma - \frac{x-\mu}{\sigma}\right)\right] \tag{A.49}$$

where $\mathrm{Erfc}(x)$ is the complement of the error function:

$$\mathrm{Erfc}(x) = 1 - \mathrm{Erf}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,dt \tag{A.50}$$

$$= \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2}\,dt. \tag{A.51}$$

The main descriptive statistics parameters of the exponentially modified Gaussian density are:

$$\mathbb{E}(X_{EMG}) = \mu + 1/\lambda \tag{A.52}$$

$$\mathrm{VAR}(X_{EMG}) = \sigma^2 + 1/\lambda^2 \tag{A.53}$$

$$\text{skewness} = \frac{2}{(\sigma\lambda)^3}\left(1 + \frac{1}{\sigma^2\lambda^2}\right)^{-3/2} \tag{A.54}$$

$$\text{Fisher kurtosis} = \frac{3\left(1 + \frac{2}{\sigma^2\lambda^2} + \frac{3}{\sigma^4\lambda^4}\right)}{\left(1 + \frac{1}{\sigma^2\lambda^2}\right)^2} - 3. \tag{A.55}$$

The median has no analytically closed form but can be calculated by numerical quadrature.

Exponentially modified Gaussian random variables are not stable.

The exponentially modified Gaussian distribution is used in chromatography as a theoretical model for the shape of the peak of a given chemical solute during elution on a chromatographic column [67]. It was also applied as a statistical model of the intermitotic time in dividing cells and of the ribosome residence time on specific codons during `mRNA` translation [44].

A.2.6. A note on the number of the required parameters to determine the density of a distribution, their relation to the mean, variance, skewness and definition domains

A noticeable shortcoming of the exponentially modified Gaussian distribution is that the support of this density is $\mathbb{R}$, i.e. negative or positive real values. The negative values do not make sense in the framework of a queueing time theory as a queueing time is always positive. The exponential distribution, the hypo-exponential distribution and the Gamma distribution have their support on $\mathbb{R}^+$: the probability space is the set of all positive or null real values. This is fully consistent with a queueing time theory. The Gamma distribution requires 2 (3 for shifted gamma) parameters to be unequivocally defined. The hypo-exponential distribution requires 3 parameters (4 if shifted). The exponentially modified Gaussian requires 3 parameters. In all these distributions, the mean and variance are not independent, in contrast to the Gaussian distribution. The mean and the variance are not sufficient to determine these distributions, except for the Gamma distribution (2 parameters). Hence, the measure of asymmetry, i.e., the skewness, of these distributions is the third parameter that can help specifying them completely. In this respect, it is informative to know that the skewness of the exponential distribution is always equal to 2, while the skewness of the Gamma distribution is $2/\sqrt{\alpha}$, and the skewness of the hypo-exponential and the exponentially modified Gaussian are dependent on their other parameters in complex expressions. The skewness of a symmetric distribution is equal to zero. It is also extremely important to be aware that the classical sample mean and variance are not necessarily equal to the maximum likelihood estimators of the true distribution parameters as would be the case for normally (Gaussian) distributed data. Refer for instance to equations (A.39)-(A.42) for the expressions of the maximum likelihood estimators of the Gamma distribution.

A.2.7. Brute force mutual comparison of distributions and quality of the fit of a (shifted) hypo-exponential or a (shifted) Gamma to the exponentially modified Gaussian density

We will conduct the comparison using two comparison criteria exposed in (1) and (2) below. The comparison using the two methods is carried out after that.

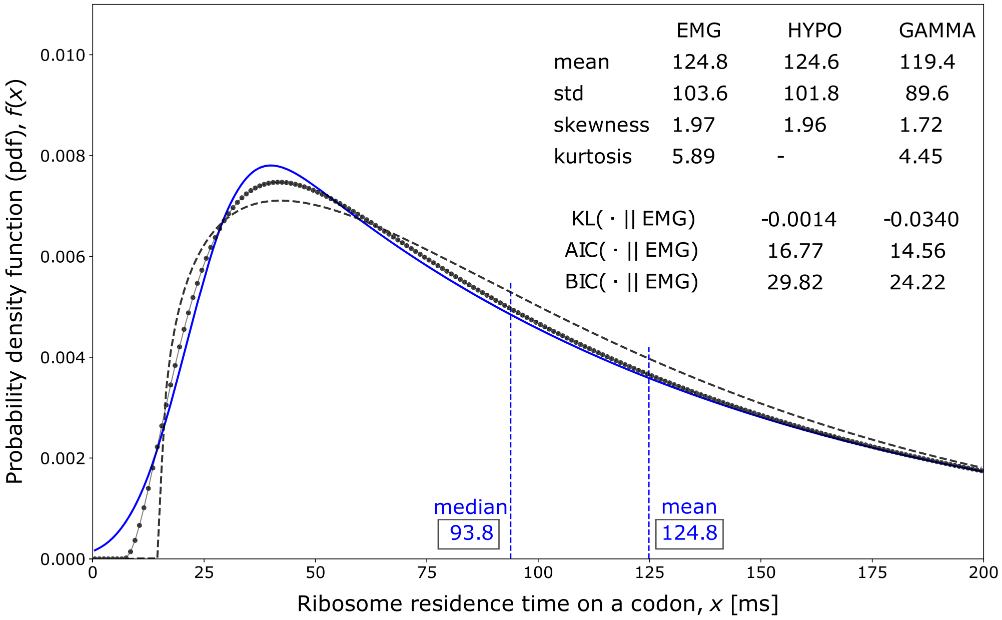
1. Kullback-Leibler (KL) divergence criterion.
   In mathematical statistics and information theory, the Kullback-Leibler divergence, also called relative entropy, noted $KL(P||Q)$ is a statistical distance measuring how one reference probability distribution P is different from a second, probability distribution Q [68]. A simple interpretation of the divergence of P from Q is the expected excess surprise from using Q as a model when the actual distribution is P. This distance is not a metric, is asymmetric and the triangle inequality does not hold. In the simple case, a relative entropy of 0 indicates that the two distributions in question convey the exact same information.
   Consider two probability distributions P and Q. Usually, P represents the data, the observations, or a measured probability distribution. Distribution Q represents instead a theory, a model, a description or an approximation of P. The Kullback–Leibler divergence is then interpreted as the average difference of the number of bits required for encoding samples of P using a code optimized for Q rather than one optimized for P. Note that the roles of P and Q can be reversed in some situations where it is easier to compute.
   In the discrete case, the Kullback-Leibler divergence is:

   $$\mathrm{KL}(P||Q) = \sum_{x_i \in \mathcal{X}} P(x) \ln \frac{P(x)}{Q(x)} \tag{A.56}$$

   for P and Q defined on the same probability space $\mathcal{X}$. In other words, it is the expectation of the logarithmic difference between the probabilities P and Q, where the expectation is taken using the probabilities P.
   For the continuous case, the KL divergence definition is:

   $$\mathrm{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) \ln\left(\frac{p(x)}{q(x)}\right) dx. \tag{A.57}$$

2. Akaike information criterion (AIC) and Bayesian information criterion (BIC).
   The Akaike information criterion (AIC) is an estimator of prediction error and thereby relative quality of statistical models for a given set of data [69]. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Thus, AIC provides a means for model selection.
   AIC is founded on information theory. When a statistical model is used to represent the process that generated the data, the representation will almost never be exact; so some information will be



	EMG	HYPO	GAMMA
mean	124.8	124.6	119.4
std	103.6	101.8	89.6
skewness	1.97	1.96	1.72
kurtosis	5.89	-	4.45
KL(· || EMG)		-0.0014	-0.0340
AIC(· || EMG)		16.77	14.56
BIC(· || EMG)		29.82	24.22

Fig. 13. Best least squared fit of shifted hypo-exponential (black dotted line) or shifted Gamma distribution (dashed line) to the exponentially modified Gaussian distribution (blue line) as empirical model for the ribosome residence time on the GAA codon (in yeast). EMG parameters were taken from the *Saccharomyces cerevisiae* GAA codon reported in Dana and Tuller [44].

lost by using the model to represent the process. AIC estimates the relative amount of information lost by a given model: the less information a model loses, the higher the quality of that model.

Suppose that we have a statistical model of some data. Let k be the number of estimated parameters in the model. Let $\hat{L}$ be the maximum value of the likelihood function for the model. Then the AIC value of the model is the following:

$$\mathrm{AIC} = 2k - 2\ln(\hat{L}). \tag{A.58}$$

Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. Thus, AIC rewards goodness of fit (as assessed by the likelihood function), but it also includes a penalty that is an increasing function of the number of estimated parameters. The penalty discourages overfitting, which is desired because increasing the number of parameters in the model almost always improves the goodness of the fit.

AIC is founded in information theory. Suppose that the data is generated by some unknown process f. We consider two candidate models to represent f: g_1 and g_2. If we knew f, then we could find the information lost from using g_1 to represent f by calculating the Kullback–Leibler divergence directly, $\mathrm{KL}(f||g_1)$; similarly, the information lost from using g_2 to represent f could be found by calculating $\mathrm{KL}(f||g_2)$. We would then, generally, choose the candidate model that minimized the information loss.

We cannot choose with certainty, because we do not know f. Akaike showed, however, that we can estimate, via AIC, how much more (or less) information is lost by g_1 than by g_2. The estimate, though, is only valid asymptotically; if the number of data points is small, then some correction is often necessary.

The BIC is defined as [70]

$$\mathrm{BIC} = k\ln n - 2\ln(\hat{L}) \tag{A.59}$$

where

- $\hat{L}$, the maximized value of the likelihood function of the model M, i.e. $\hat{L} = p(x \mid \widehat{\theta}, M)$, where $\widehat{\theta}$ are the parameter values that maximize the likelihood function;
- x, the observed data;
- n, the sample size;
- k, the number of parameters estimated by the model.

How does the exponentially modified Gaussian distribution compare to the hypo-exponential or the Gamma distributions?

We conducted two comparisons:

(a) How similar is the 4-parameter (shifted) hypo-exponential distribution, called Q_1, to the 3-parameter exponentially modified Gaussian, called P?

(b) How similar is the 3-parameter (shifted) Gamma distribution, called Q_2, to the 3-parameter exponentially modified Gaussian, called P?

In both cases, we took the same reference EMG density as adapted from the normalized footprint counts of Dana & Tuller [44] to estimate the ribosome residence time on the specific codon `GAA` decoding for Glutamate `Glu-E`, in yeast. This EMG density, i.e., P, is arbitrary taken as the 'true' data generating model. The Q distributions used in the calculation of the goodness of the fit criteria are the ones for which the parameters were determined either by the maximum likelihood approach or by the non-linear least squared fit obtained with the Levenberg-Marquardt algorithm. The three probability density function plots are represented in Fig. 13 for comparison. The calculated values of the three information criteria (Kullback-Leibler, AIC and BIC) for the similarity and goodness of the fit are given in the table at the insert of Fig. 13.

The shifted hypo-exponential distribution essentially conveys the same information as the exponentially modified Gaussian with a Kullback-Leibler divergence $KL(\mathrm{EMG}||\mathrm{shifted}-\mathrm{HYPO}) = -0.0014$. This is slightly better than the Gamma distribution for which $KL(\mathrm{EMG}||\mathrm{shifted}-\mathrm{GAMMA}) = -0.0340$. If the number of parameters of the distributions are taken as a penalty to assess the goodness of the fit, as is the case with AIC and BIC criteria, then the shifted Gamma (3 parameters) ranks better than the shifted hypo-exponential (4 parameters) with respect to the exponentially modified Gaussian distribution (3 parameters). The shifted hypo-exponential extra parameter may cause some overfitting in the goodness of the fit. In any case, any of these three distributions could be used to represent the ribosome residence time data on any specific codon without affecting significantly the conveyed information when moving from one to the other. Other authors have shown that the log-normal distribution can be a good fit as well for the exponentially modified Gaussian distribution [45].

A supplementary material `.mp4` animation shows the geometric interpretation of the convolution product to calculate the probability distribution function resulting from the sum of three exponentially distributed random variables. The animation also compares the quality of the fit between the hypo-exponential distribution, the Gamma distribution and the exponentially modified Gaussian distribution.

Appendix B. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csbj.2023.07.016.

References

[1] Noller HF, Hoffarth V, Zimniak L. Unusual resistance of peptidyl transferase to protein extraction procedures. Science 1992;256:1416.

[2] Simonovic M, Steitz T. A structural view on the mechanism of the ribosome-catalyzed peptide bond formation. Biochim Biophys Acta - Gene Regul Mech 2009;1789:612.

[3] Rodnina M, Beringer M, Wintermeyer W. Mechanism of peptide bond formation on the ribosome. Q Rev Biophys 2006;39:203.

[4] Rozov A, Khusainov I, t. El Omari K. Importance of potassium ions for ribosome structure and function revealed by long-wavelength x-ray diffraction. Nat Commun 2019;10.

[5] Nierhaus KH. Mg2+, K+, and the ribosome. J Bacteriol 2014;196:3817–9.

[6] Wang J, Karki C, Xiao Y, Li L. Electrostatics of prokaryotic ribosome and its biological implication. Biophys J 2020;118:1205.

[7] Rodnina M. The ribosome in action: tuning of translational efficiency and protein folding. Protein Sci 2016;25.

[8] Simpson LJ, Tzima E, Reader JS. Mechanical forces and their effect on the ribosome and protein translation machinery. Cells 2020;9.

[9] Fried SD, Boxer SG. Electric fields and enzyme catalysis. Annu Rev Biochem 2017;86:387.

[10] Lu J, Kobertz W, Deutsch C. Mapping the electrostatic potential within the ribosomal exit tunnel. J Mol Biol 2007;371:1378.

[11] Brooks B, et al. CHARMM: the biomolecular simulation program. J Comput Chem 2009;30:1545.

[12] Gabdulkhakov A, Nikonov S, Garber M. Revisiting the *Haloarcula marismortui* 50S ribosomal subunit model. Acta Crystallogr, D Biol Crystallogr 2013;69:997.

[13] Watson ZL, Ward FR, Méheust R, Ad O, Schepartz A, Banfield JF, et al. Structure of the bacterial ribosome at 2 Å resolution. eLife 2020;9:e60482.

[14] Polikanov Y, Melnikov S, Soll D, Steitz T. Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. Nat Struct Mol Biol 2015;22.

[15] Bhatt PR, Scaiola A, Loughran G, Leibundgut M, Kratzel A, Meurs R, et al. Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome. Science 2021;372:1306.

[16] Tirumalai M, Rivas M, Tran Q, Fox G. The peptidyl transferase center: a window to the past. Microbiol Mol Biol Rev 2021;85.

[17] Doris S, Smith D, Beamesderfer J, Raphael B, Nathanson J, Gerbi S. Universal and domain-specific sequences in 23S–28S ribosomal RNA identified by computational phylogenetics. RNA 2015;21.

[18] Sehnal D, Svobodová Vařeková R, Berka K, Pravda L, Navrátilová V, Banáš P, et al. MOLE 2.0: advanced approach for analysis of biomacromolecular channels. J Cheminform 2013;5:39.

[19] Berka K, Hanak O, Sehnal D, Banás P, Navrátilová V, Jaiswal D, et al. MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels. Nucleic Acids Res 2012;40:W222.

[20] Pravda L, Sehnal D, Toušek D, Navrátilová V, Bazgier V, Berka K, et al. MOLEonline: a web-based tool for analyzing channels, tunnels and pores (2018 update). Nucleic Acids Res 2018;46.

[21] Beringer M, Rodnina M. The ribosomal peptidyl transferase. Mol Cell 2007;26:311.

[22] Lang K, Erlacher M, Wilson DN, Micura R, Polacek N. The role of 23S ribosomal RNA residue A 2451 in peptide bond synthesis revealed by atomic mutagenesis. Chem Biol 2008;15:485.

[23] Polikanov Y, Steitz T, Innis C. A proton wire to couple aminoacyl-tRNA accommodation and peptide-bond formation on the ribosome. Nat Struct Mol Biol 2014;21(9):787.

[24] Jackson JD. Classical electrodynamics. third edition. Wiley & Sons; 1998. p. 32.

[25] Barton G. Elements of Green's functions and propagation. Oxford University Press; 1989. p. 7–38.

[26] Joiret M, Kerff F, Rapino F, Close P, Geris L. Ribosome exit tunnel electrostatics. Phys Rev E 2022;105:1.

[27] Johansson R. Numerical python. second edition. Apress; 2019. p. 267–93.

[28] Lockhart D, Kim P. Electrostatic screening of charge and dipole interactions with the helix backbone. Science 1993;260:198.

[29] Sharp K, Honig B. Electrostatic interactions in macromolecules: theory and applications. Annu Rev Biophys Biophys Chem 1990;19:301.

[30] Cuervo A, Dans P, Carrascosa J, Orozco M, Gomila G, Fumagalli L. Direct measurement of the dielectric polarization properties of DNA. Proc Natl Acad Sci USA 2014;111.

[31] van Roij R. Electrostatics in liquids: electrolytes, suspension, and emulsions, in: Lecture Notes, Institute of Theoretical Physics, Utrecht, the Netherlands, 16 July 2009, unpublished.

[32] Sievers A, Beringer M, Rodnina M, Wolfenden R. The ribosome as an entropy trap. Proc Natl Acad Sci USA 2004;101:7897.

[33] Wallin G, Aqvist J. The transition state for peptide bond formation reveals the ribosome as a water trap. Proc Natl Acad Sci USA 2010;107:1888.

[34] Eyring H. The activated complex in chemical reactions. J Chem Phys 1935;3:107.

[35] Laidler KJ, King MC. Development of transition-state theory. J Phys Chem 1983;87:2657.

[36] Anslyn E, Dougherty D. Energy surfaces and kinetic analysis. In modern physical organic chemistry. Saucalito, CA, USA: University Science Books; 2005. p. 355–419.

[37] Kaiser C, Tinoco I. Probing the mechanisms of translation with force. Chem Rev 2014;114:3266.

[38] Bell G. Models for the specific adhesion of cells to cells. Science 1978;200:618.

[39] Bustamante C, Chemla YR, Forde NR, Izhaky D. Mechanical processes in biochemistry. Annu Rev Biochem 2004;73:705.

[40] Ribas-Arino J, Marx D. Covalent mechanochemistry: theoretical concepts and computational tools with applications to molecular nanomechanics. Chem Rev 2012;112:5412.

[41] Pape T, Wintermeyer W, Rodnina M. Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the E. coli ribosome. EMBO J 1999;17:7490.

[42] Wohlgemuth I, Brenner S, Beringer M, Rodnina MV. Modulation of the rate of peptidyl transfer on the ribosome by the nature of substrates. J Biol Chem 2008;283:32229.

[43] Ingolia N, Ghaemmaghami S, Newman J, Weissman J. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 2009;324:218.

[44] Dana A, Tuller T. The effect of tRNA levels on decoding times of mRNA codons. Nucleic Acids Res 2014;42(14):9171.

[45] Dana A, Tuller T. Properties and determinants of codon decoding time distributions. BMC Genomics 2014;15:S13.

[46] Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 2011;147:789.

[47] Tinoco I, Wen J-D. Simulation and analysis of single-ribosome translation. Phys Biol 2009;6:025006.

[48] Pavlov MY, Ullman G, Ignatova Z, Ehrenberg M. Estimation of peptide elongation times from ribosome profiling spectra. Nucleic Acids Res 2021;49:5124.

[49] Rodnina M, Wintermeyer W. Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanism. Annu Rev Biochem 2001;70:415.

[50] Fried S, Bagchi S, Boxer S. Extreme electric fields power catalysis in the active site of ketosteroid isomerase. Science 2014;346:1510.

[51] Shah P, Ding Y, Niemczyk M, Kudla G, Plotkin J. Rate-limiting steps in yeast protein translation. Cell 2013;153:1589.

[52] Nieß A, Siemann-Herzberg M, Takors R. Protein production in *Escherichia coli* is guided by the trade-off between intracellular substrate availability and energy cost. Microb Cell Fact 2019;18(1):8.

[53] Farewell A, Neidhardt FC. Effect of temperature on in vivo protein synthetic capacity in escherichia coli. J Bacteriol 1998;180:4704.

[54] Dao Duc K, Batra S, Bhattacharya N, Cate J, Song Y. Differences in the path to exit the ribosome across the three domains of life. Nucleic Acids Res 2019;47.

[55] Auffinger P, Ennifar E, D'Ascenso L. Deflating the Mg2+ bubble. Stereochemistry to the rescue! RNA 2021;27:243.

[56] Johansson M, Ieong K-W, Trobro S, Strazewski P, Åqvist J, Pavlov MY, et al. pH-sensitivity of the ribosomal peptidyl transfer reaction dependent on the identity of the a-site aminoacyl-trna. Proc Natl Acad Sci 2011;108:79.

[57] Melnikov S, Mailliot J, Rigger L, Neuner S, Shin B-S, Yusupova G, et al. Molecular insights into protein synthesis with proline residues. EMBO Rep 2016;17:1776.

[58] Pavlov M, Watts E, Tan Z, Cornish V, Ehrenberg M, Forster A. Slow peptide bond formation by proline and other N-alkylamino acids in translation. Proc Natl Acad Sci USA 2009;106:50.

[59] Doerfel L, Wohlgemuth I, Kothe C, Peske F, Urlaub H, Rodnina M. EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. Science 2012;339.

[60] Peil L, Starosta A, Lassak J, Atkinson G, Virumäe K, Spitzer M, et al. Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor ef-p. Proc Natl Acad Sci 2013;110.

[61] Starosta AL, Lassak J, Peil L, Atkinson GC, Virumäe K, Tenson T, et al. Translational stalling at polyproline stretches is modulated by the sequence context upstream of the stall site. Nucleic Acids Res 2014;42:10711.

[62] Doerfel LK, Wohlgemuth I, Kubyshkin V, Starosta AL, Wilson DN, Budisa N, et al. Entropic contribution of elongation factor p to proline positioning at the catalytic center of the ribosome. J Am Chem Soc 2015;137:12997.
[63] Levy E. On the density for sums of independent exponential, Erlang and gamma variates. arXiv:2006.12428v4, 2020.
[64] Ross S. Introduction to probability models. 11th ed. San Diego, CA, USA: Academic Press; 2014. p. 282–361.
[65] Erlang AK. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. In: Brockmeyer E, Halstrøm HL, Jensen Arne, editors. The life and works of A.K. Erlang, Transactions of the Danish Academy of Technical Sciences, 2, Akademiet for de Tekniske Videnskaber; 1948. p. 138–55. archived from the original (PDF) on July 19, 2011.
[66] Oguntunde P, Odetunmibi O, Adejumo AO. On the sum of exponentially distributed random variables: a convolution approach. Eur J Stat Probab 2014;2:1.
[67] Grushka E. Characterization of exponentially modified Gaussian peaks in chromatography. Anal Chem 1972;44:1733.
[68] Kullback S, Leibler RA. On information and sufficiency. Ann Math Stat 1951;22:79.
[69] Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control 1974;19:716.
[70] Schwarz G. Estimating the dimension of a model. Ann Stat 1978;6:461.