Transfer RNA translocation through the ribosome.

Combining large scale system simulations with experimental data.

Dissertation for the award of the degree
"Doctor rerum naturalium"
Division of Mathematics and Natural Sciences of the
Georg-August-University Göttingen
submitted by
Christian Blau
from
Merseburg

31.12.2013

Thesis Committee

Prof. Dr. Helmut Grubmüller (Reviewer)[1]
Prof. Dr. Holger Stark (Reviewer)[1]
Prof. Dr. Jörg Enderlein[2]

Examination board

Prof. Dr. Kai Tittmann[2]
Dr. Iwan A. T. Schaap[2]
Prof. Dr. Berent L. de Groot[1]

Date of the oral examination

5th of March, 2014

---

[1]Max Planck Institue for Biophysical Chemistry, Göttingen
[2]Georg-August-University Göttingen

Affidavit

Hiermit bestätige ich, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Göttingen, Dezember 2013

Christian Blau

Contents

1. Introduction

Genetic information from DNA is expressed as proteins in a universal two-step mechanism sketched in Fig. 1.1 [1]. During protein expression information on DNA is transcribed to an RNA working copy, then translated to proteins from this RNA working copy.

Replication, transcription, and translation are the general transfers of genetic information. In contrast to the general transfers, special and so-called undetected transfers were omitted from Fig. 1.1. Examples of these undetected and special transfers are small interfering RNAs that inhibit translation of sequence information on RNA [2] and proteins that silence genes on DNA through methylation [3], further proteins that chemically modify other proteins after translation [4] and proteins that post-transcriptionally modify RNA [5].

Each general transfer has its catalyst. DNA polymerases replicate DNA. RNA polymerases transcribe genetic information to the mediating information carrier messenger RNA (mRNA). Ribosomes, the subject of this work, catalyze translation of mRNA to proteins in an elongation cycle. During this cycle amino-acids are sequentially added to a nascent peptide chain.

The ribosome is a huge macromolecular protein-RNA complex of a small and large subunit, comprising more than two-hundred thousand atoms in prokaryotes. Figure 1.2 shows representations of ribosomes, as resolved by cryo electron-microscopy (cryo-EM) and X-ray crystallography.

Translation in ribosomes is performed through a hierarchy of chemical steps, sketched in Fig. 1.3. Peptide elongation is primed by initiation and brought to a controlled halt by elongation termination. Ribosome recycling then provides the two subunits for new rounds of elongation.

Elongation itsself comprises the key kinetic steps decoding, peptide transfer, and tRNA translocation. The small subunit catalyzes translation of genetic

$$\text{DNA} \xrightarrow[\text{RNAPol}]{\text{transcription}} \text{RNA} \xrightarrow[\textit{ribosome}]{\textit{translation}} \text{protein}$$

$$\underset{\text{DNAPol}}{\overset{\text{replication}}{\circlearrowright}}$$
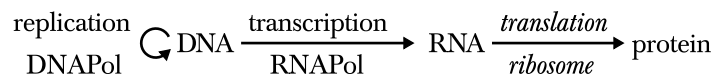
Figure 1.1: The central dogma of molecular biology implies a canonical flow of information (general transfers, indicated by arrows) from DNA to protein. The processes associated with the respective general transfers are shown above the arrows, their catalysts below. Special and undetected information transfers [1] are not shown.
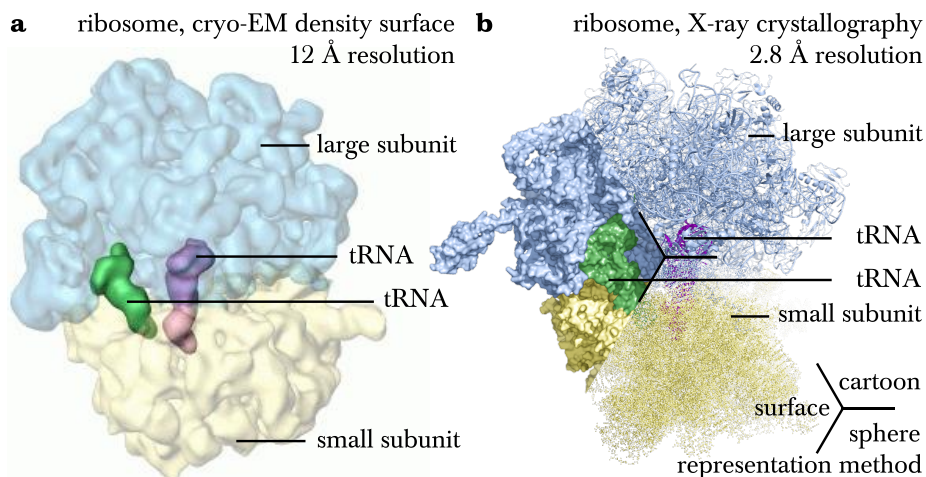
a ribosome, cryo-EM density surface
12 Å resolution

large subunit

tRNA

tRNA

small subunit

b ribosome, X-ray crystallography
2.8 Å resolution

large subunit

tRNA

tRNA

small subunit

cartoon

surface

sphere

representation method

Figure 1.2: The ribosome, as seen by cryo electron microscopy and X-ray crystallography. Shown are the large (light blue) and small subunit (yellow) and two tRNAs (purple, green). (a) Surface representation of a cryo-EM density of the ribosome at 12 Å resolution from [6]. (b) Surface, cartoon and sphere representation of the heavy atoms of the ribosome as resolved by X-ray crystallography at 2.8 Å resolution [7], represented with PYMOL [8]. Note that the methods to define the resolution of cryo-EM densities and X-ray crystallography densities differ.

translation

initiation ⟶ *elongation cycle* ⟶ termination ⟶ recycling
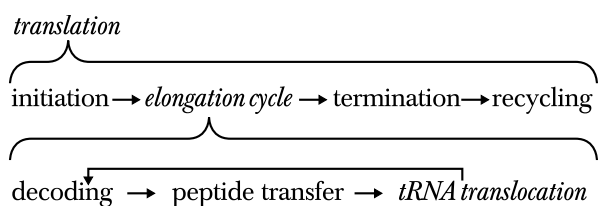
decoding ⟶ peptide transfer ⟶ *tRNA translocation*

Figure 1.3: The hierarchy of processes during translation. *Translation* contains the *elongation cycle*. This work describes tRNA *translocation*, a sub-process of the *elongation cycle*.
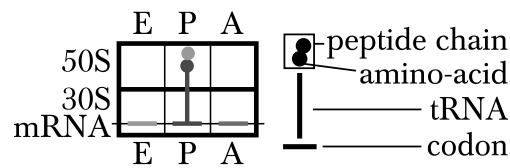
Figure 1.4: The three tRNA binding sites span the small (30S) and large (50S) ribosomal sub-unit. This schematic box-plot representation of the ribosome indicates the three ribosomal binding sites, the amino-acyl site (A site), peptidyl site (P site), and exit site(E site). The tRNA is represented as a vertical line, the amino-acid as disc, the mRNA codons as horizontal bars, and the mRNA as horizontal line. Two discs indicate the nascent peptide chain.

code on the mRNA blueprint to cognate amino-acids by adaptor molecules called transfer RNA (tRNA). Following this decoding step, the large subunit catalyzes elongation of the nascent peptide chain by peptide bond formation [9, 10]. After the ribosome catalyzed codon recognition and peptide transfer, the tRNAs move through the ribosome to prime the ribosome for a new round of elongation.

For efficient catalysis, three specific binding sites position the tRNAs during elongation. Figure 1.4 shows these tRNA binding sites, amino-acyl site (A site), peptidyl site (P site), and exit site (E site). Mechanisms and transition rates for the movement of the tRNAs between their binding sites during elongation were revealed by spectroscopic measurements [11]. The determined reaction mechanism is sketched in Fig. 1.5, and will be laid out in greater detail, with the additional involved factors, further below. During tRNA translocation, the two tRNAs move from the A to the P and from the P to the E site. Fifty intermediate states of spontaneous tRNA movement in the absence of additional factors and GTP hydrolysis have recently been observed using cryo electron microscopy experiments [6].

Spontaneous tRNA translocation is an equilibrium process that consumes no energy. Yet, the tRNAs diffuse more than seven nanometers through the ribosome, despite the many involved interactions and the large size of the ribosome. The energy from thermal fluctuations suffices to induce large-scale, though undirected ribosome motions [6]. The very little energy gained through hydrolysis of guanosine-triphosphate (GTP) by elongation factor G suffices to drive tRNA translocation forward through a "Brownian ratchet" mechanism by using these thermal fluctuations [12, 13].

By observing this Brownian machine through cryo-EM [6], the gap between static structures of the energy minima on the energy landscape and dynamic information has been narrowed in two ways. First, the ensemble of translocation intermediate structures was vastly extended and the free energy of each intermediate determined. A sketch of the resulting free energy landscape is shown in Fig. 1.6. Second, the velocity of the major rate-limiting step that separates so called pre-translocation and post-translocation states was determined, and a rapid equilibrium between the pre-translocation and post-translocation states was found.
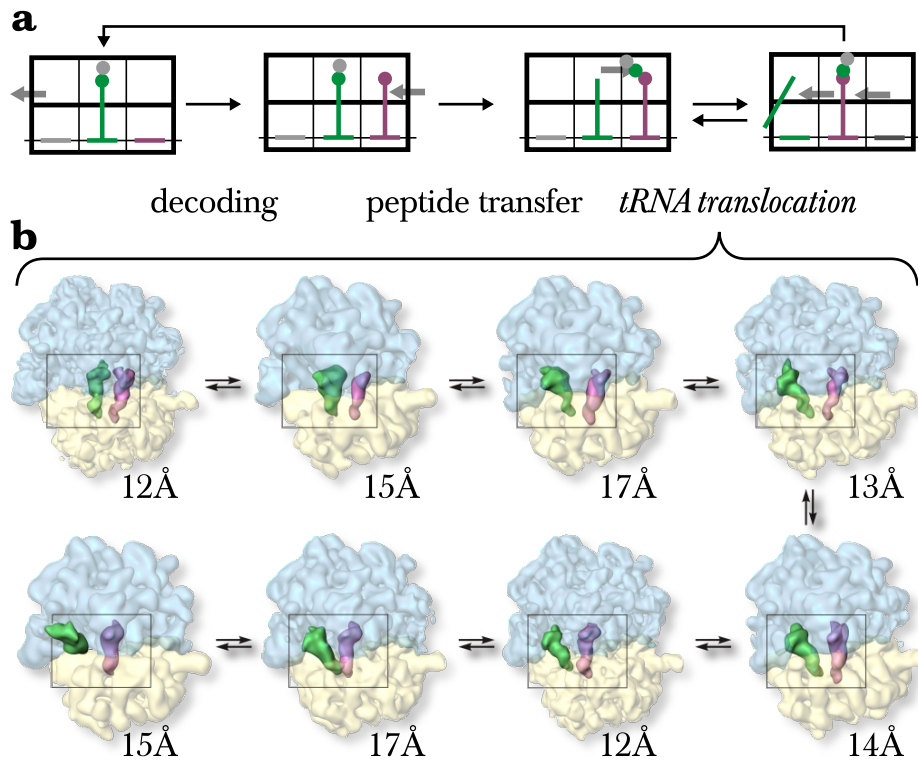
Figure 1.5: The elongation cycle in box-plot representation and tRNA translocation as seen from cryo-EM. Panel (a) sketches the elongation cycle (cf. Fig. 1.4). Panel (b) shows eight representative cyro-EM densities from fifty intermediates of spontaneous tRNA translocation through the ribosome [6]. The gray lines indicates the thick black box shown in (a) for reference. Shown are the large (light blue) and small subunit (yellow) and two tRNAs (purple, green). Resolutions of the respective cryo-EM densities are given below them.
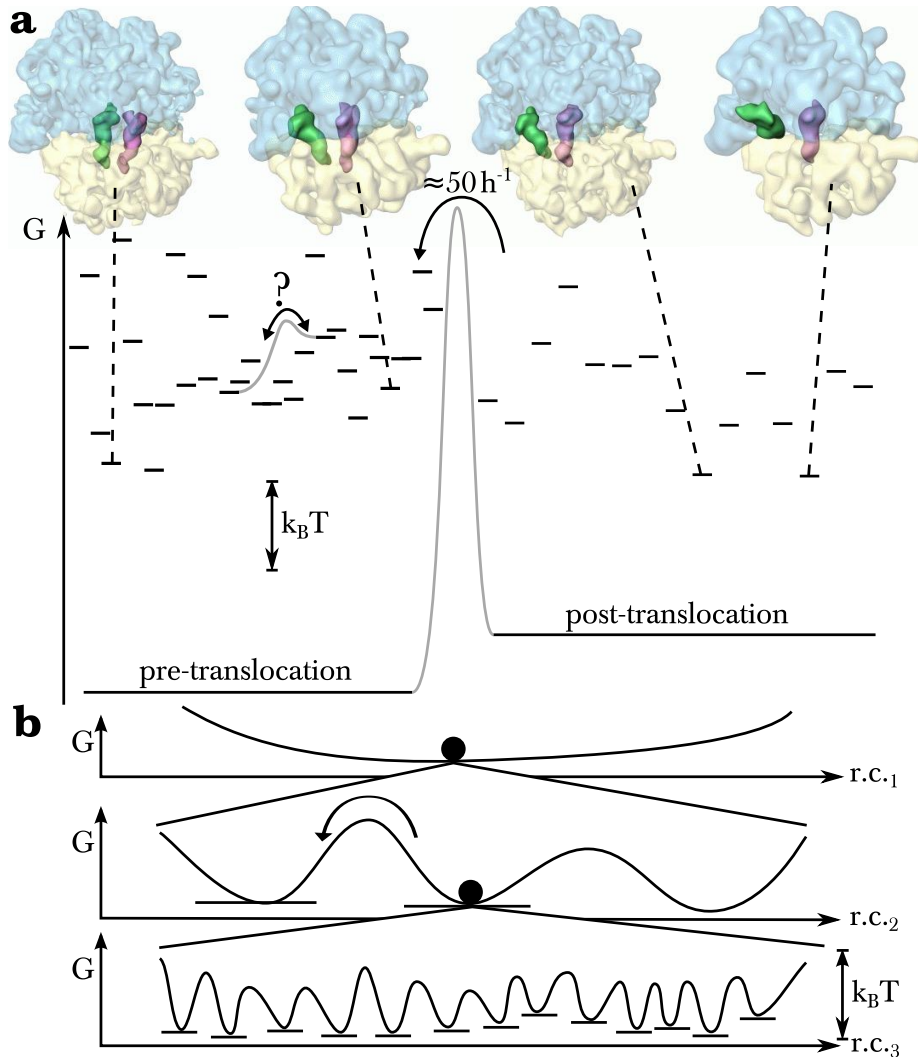
Figure 1.6: the experimental data of tRNA translocation through the ribosome (a) resembles the Frauenfelder picture of a hierarchical energy landscape (b) Panel (a) shows the free energies of translocation intermediates of ribosomal tRNA translocation in equilibrium (to scale). The free energies of the pre- and post-translocation states fluctuate within a few k_BT. They are separated via a large barrier between, and smaller barriers within pre- and post-translocation states, resembling the Frauenfelder picture. Conformational substates are represented as ribosome (blue and yellow) and tRNA (purple and green) surfaces. Data courtesy of Niels Fischer (unpublished and [6]). Panel (b) depicts conformational substates that are separated by a hierarchy of energy barriers that emerges when going from a coarse description (top) to detailed reaction coordinates (bottom), modified from Frauenfelder et al. [14].

The many translocation intermediates differ only by a few $k_B T$ in energy and are seperated by a hierarchy of barriers between the states as shown in Fig. 1.6. Similarly, the study of Frauenfelder proposed a hierarchical energy landscape for proteins using myoglobine as an example [14]. Fig. 1.6 shows the close resemblance of the energy landscape determined by Fischer et al. and the Frauenfelder picture [14, 15].

Yet unresolved are the barriers between the individual pre- and post-translocation states and the dynamics within the resolved energy minima. Further, the cryo-EM data lacks the atomic detail that would allow to name the specific interactions responsible for the barriers.

Figure 1.2 shows an example of one of the high-resolution X-ray structures that provide an atom-by-atom description of residue and nucleotide interactions. The Nobel Prize winning crystal-structures of the ribosome were a break-through in describing translation structurally [16]. They provide the stereochemistry of the binding sites, the interacting ribosome residues and nucleotides for stable, crystallizable states in the energy minima of the ribosome energy landscape.

In this work, we describe the ribosome energy landscape during translocation by refining the cryo electron microscopy data by Fischer et al. [6] to the structural information from X-ray crystallography [17], performing atomistic molecular dynamics (MD) simulations and testing their results against biochemical information from spectroscopic measurements and evolutionary analysis. Though the notion that the ribosome diffuses in a Frauenfelder-type energy landscape has been expressed earlier by Whitford and Sanbonmatsu in the ribosome field [18], we now provide quantitative evidence for that notion. Overall, the statistical mechanics view on ribosomal translation in this work aims to discuss tRNA translocation as proposed by Frauenfelder et al.,

> "The taxonomic attitude is the conventional one of mechanistic chemistry, but the statistical viewpoint requires new concepts. One no longer talks about specific energy levels but about the statistics of the energy landscape." [14]

To estimate the heigth of the free energy barriers that hinder transition of the ribosome from one state to another, we assess the diffusion of the ribosome on its high-dimensional energy landscape. We identify further kinetic steps than the pre- to post-translocation transition and determine the interactions between nucleotides and residues that make transitions between different states so efficient for the ribosome.

Ribosomes evolved to be fast and accurate at little energy cost. They maintain a minute balance between speed of translation, error rate and energy consumption [19]. We employ the working hypothesis that the ribosome is efficient, because its energy landscape has been smoothened evolutionarily; speed and accuracy of translocation are achieved by fine-tuning interactions between ribosome parts, their residues and nucleotides to a near-constant level.

A prerequisite to analyze the fine-tuning of interaction energies on a near-atomic level is a description of all atom contacts on the interface of ribosome parts. Though finding contacting atoms of ribosome parts is conceptually

simple, finding and implementing a fast algorithm for this task is not. For example, between the large and the small subunit of the ribosome there are more then ten billion possible atom-atom contacts. We devised and implemented an efficient algorithm to find contacting atoms on the interface between ribosome parts, which proved to be useful for a variety of biomolecular ensemble data [20].

Before laying out the underlying assumptions in statistical physics that are necessary to bridge energetics and kinetics of translation, we will now detail the biochemistry of protein synthesis in prokaryotes and summarize previous computational approaches on translation.

Our research is published in "g_contacts: Fast contact search in bio-molecular ensemble data" [21] and "Energy barriers and driving forces in tRNA translocation through the ribosome" [22]. An essential part of the assumed research is laid out in the supplementary note complementing the article "Energy barriers and driving forces in tRNA translocation through the ribosome". The data analysis methods that extract interpretable information from the simulation data are presented in this supplement in the context of ribosome simulations, but are valuable in their own right.

1.1   Translation in Procaryotes

Translation

Translation of genetic information from mRNA to proteins occurs in all domains of life in four distinct steps: initiation, elongation, termination, and recycling.

In this work, we investigate translation of the prokaryote *Escherichia coli* (*E. coli*, K-12 strain). This section will place prokaryotic elongation in the context of translation by describing it along with initiation, termination and recycling. In the next section we will give a more detailed description of the elongation cycle, linking experimental and computational studies of elongation.

Initiation prepares ribosomes for the elongation cycle. The Shine-Dalgarno sequence on prokaryotic messenger RNA recruits a small (30S) ribosomal subunit, facilitated by initiation factor 3 (IF3) [23, 24, 25]. The aminoacyl-site (A site) of the small subunit is kept empty by initiation factor 1 (IF1) while initiation factor 2 (IF2) carries an initiator tRNA to the ribosomal peptidyl-site (P site) carrying the amino acid formyl-methionine (fMet, cognate to the AUG start codon) at its CCA-tail. When a large (50S) ribosomal subunit binds, all initiation factors dissociate and the elongation cycle begins with a ribosome, mRNA, and fMet-tRNA$^{\text{fMet}}$ complex as shown in Fig. 1.7I.

During the elongation cycle, sketched in Fig. 1.7, proteins are synthesized by the ribosome one amino acid at a time according to the corresponding mRNA template [26]. After initiation primes the ribosome with an initiator tRNA in the P site (Fig. 1.7I), all decoding of genetic information on the mRNA is performed in the A site through binding of a complex of aminoacyl tRNA (aa-tRNA) and elongation factor thermo unstable (EF-Tu) sketched in Fig. 1.7II and in Fig. 1.8. A site accommodation positions the substrates for the peptide
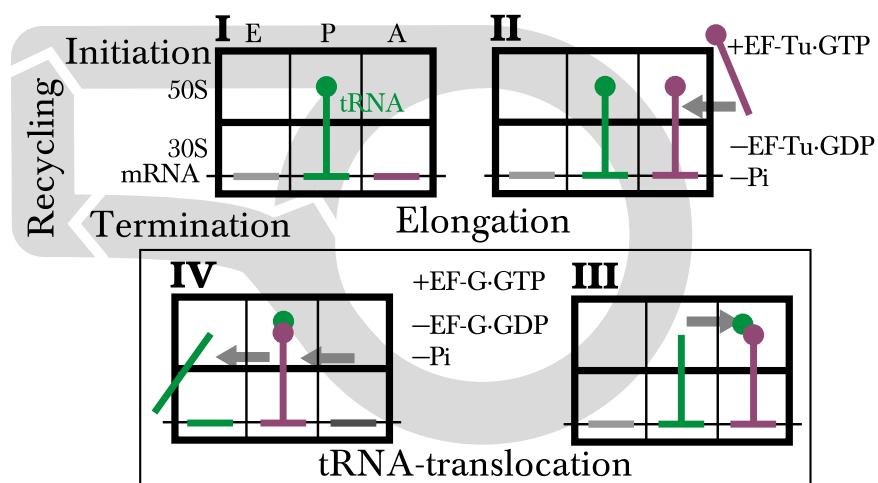
Figure 1.7: The prokaryotic elongation cycle. The ribosome with its small (30S) and large subunit (50S) and its three binding sites, the amino-acyl site (A site), peptidyl site (P site) and exit-site (E site) is sketched in box-representation. **I** fMet-tRNA^{fMet} (green vertical line (tRNA) and ball (amino-acid)) occupies the P site with the cognate start codon (green horizontal line). **II** The A site is occupied with a cognate aminoacylated tRNA (purple line (tRNA) and ball (amino-acid)) after elongation factor thermo unstable (EF-Tu) dissociation.**III** Peptide transfer reaction occurs (indicated by the gray arrow). **IV** Upon EF-G·GTP promoted translocation, tRNA^{fMet} leaves the exit site (E site), preparing the ribosome for the next elongation cycle.

transfer reaction in the peptide transfer center (PTC). After peptide transfer is catalyzed in the PTC (Fig. 1.7III), the nascent peptide chain is bound to the A site tRNA [10, 27]. Subsequent tRNA translocation of the A site tRNA to the P site, and the P site tRNA to the E site, prepares the ribosome for another of elongation cycle (Fig. 1.7IV). Binding of elongation factor GTPase (EF-G) with GTP (EF-G·GTP) drives translocation through GTP hydrolysis [28]. Translocation without EF-G still occurs, but is at least a thousand times slower [29, 30].

An A site stop codon (UAA, UAG or UGA) initializes elongation termination through binding of class-I release factors, release factor 1 (RF1, cognate to UAA and UAG codons) or release factor 2 (RF2, cognate to UAA and UGA codons) [31]. The esther bond that binds the nascent peptide chain to the P site tRNA is cleaved off by hydrolysis caused by the bound class-I release factor. Subsequently class-II release factor 3 (RF3) binds in complex with GTP. GTP hydrolysis at RF3 aids the release of tRNA, mRNA and class-I release factors from the ribosome.

During ribosomal recycling, the small and the large ribosomal subunit disscociate after binding of the ribosomal recycling factor (RRF) and subsequent EF-G·GTP binding, followed by GTP hydrolysis. Both subunits are then re-used in other elongation cycles [32].

Elongation

While giving an overview of the elongation cycle, we will highlight computational studies of different aspects of the process [33]. Computer simulations of ribosomes have been carried out for almost all sub-processes of the elongation cycle [34].

The questions addressed by computer simulations of the ribosome reflect the complexity of the fine-tuned elongation sub-processes and the complexity of the ribosome itself, an asymmetric RNA-protein compound of more than two hundred thousand atoms.

The first layer of complexity of nascent peptide chain elongation, and thus the complexity of the simulation approaches, is reflected in the time-scales its sub-processes span. They reach from pico-seconds for the proton shuttle mechanism of the peptide transfer reaction [35] to minutes required for spontaneous tRNA translocation [29].

A second layer of complexity is encoded in the energetics of translocation. Subtle differences in free energies drastically alter the reaction pathway. For example, a binding affinity difference between cognate and near-cognate tRNA of a single hydrogen-bond regulates rejection or acceptance of an accommodating tRNA at an error rate of one in ten thousand. Other examples are provided by the fact that interaction of antibiotics with single ribosome nucleotides inhibit bacterial growth or that single amino-acid mutations change the translation speed by orders of magnitude.
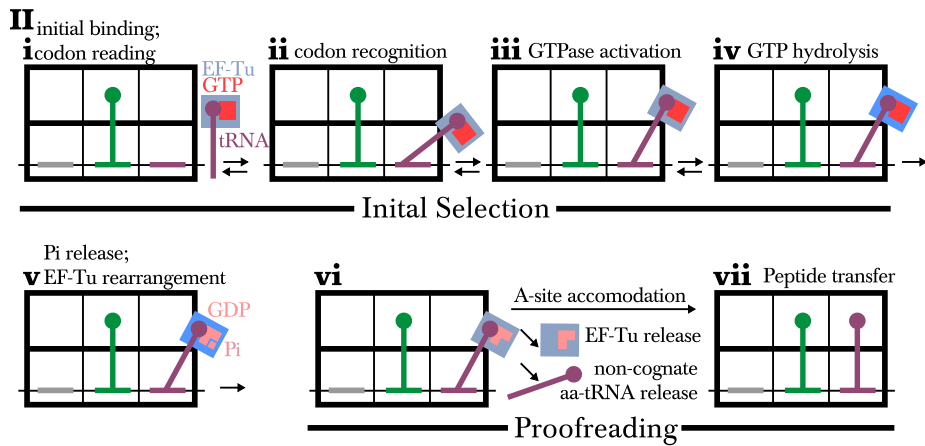
Figure 1.8: **A site decoding, catalyzed by EF-Tu.(cf.Fig 1.7II)** The sub-processes of initial selection and proofreading. Initial binding and subsequent codon reading (**i**) is followed by codon recognition (**ii**). Upon GTPase activation (**iii**), GTP hydrolysis (**iv**), leads to Pi release and EF-Tu rearrangements (**v**) that allow for A site accommodation or, alternatively, tRNA release (**vi**). Once the aa-tRNA is fully accommodated in the A site (**vii**), the nascent peptide chain is transferred to the A site. Modified from Rodnina [36].

A site decoding

The first step of the elongation cycle is A site decoding (Fig. 1.7I). Simulations of A site decoding and the factors involved were one of the first and, to date, computationally most sampled aspects of translation, expedited by the relatively small length-scales involved in A site decoding and its paramount importance to translation. The key question to A site decoding is, how minute free energy differences lead to discrimination between cognate and non-cognate tRNAs, named "Pauling's birthday party paradox" [37, 38].

The omnipresent kinetic and energetic fine-tuning, which we will investigate for the process of tRNA translocation, unravels itself here in the way speed and accuracy of decoding are evolutionarily optimized [19]. Along with tRNA aminoacylation, decoding is the key step in controlling translation fidelity to an error rate of $\approx 10^{-3}$ [39]. A site decoding alone is more accurate than 10^{-2} [40]. This readily implies a binding free energy difference between cognate and near-cognate tRNA larger than $2 \log 10 \, k_B T \approx 4.6 \, k_B T$ in equilibrium. This required energy difference is relatively large compared to the expected free energy difference of the change of a single nucleotide base, e.g. from cytosil to uracil. Two effects, both experimentally confirmed, explain this behavior. First, codon reading is thermodynamically coupled to the whole tRNA, not only its anti-codon [41]. Second, a non-reversible chemical step allows to exploit the chemical difference between cognate and near-cognate tRNA multiple times [19]. Indeed, GTP-hydrolysis separates two phases of decoding, initial selection and proofreading (Fig. 1.8) [42, 43]. During each of these, aa-tRNA can dissociate from the ribosome.

Initial tRNA selection starts when, independent of the codon, the L7/L12-stalk[1] recruites complexes of aa-tRNA, EF-Tu, and GTP (aa-tRNA·EF-Tu·GTP) to the empty A site (Fig. 1.8i).

The connection between residues to their evolutionary conservation was used to classify contacts between tRNA and EF-Tu in the tRNA·EF-Tu complex delivered to the ribosome [44]. We will observe the same connection between tRNA-ribosome contacts and their conservation.

Though the first simulations of ribosomal proteins were carried out on the C-terminal domain of L7/L12 [45, 46], initial tRNA recruitment by the L7/L12-stalk outlines the bounds of simulation studies. Here, the lack of high-resolution models of the complete stalk on the ribosome prevented further investigation by computer experiments [47].

Codon recognition (Fig. 1.8ii) in the ribosomal A site is targeted in antibiotics development and has been studied along with antibiotic resistance in a combination of structural and dynamic measurements [48, 49].

We will find that stochastic processes govern translation, in contrast to the prevailing picture of a sequential progression from one static conformation to another, as was found by in silico models of proposed stochastic gating at the A site [50, 51, 52]. They reaffirm that codon recognition does not solely depend on the local hydrogen bonding pattern between the three coding nucleotides, but involves entropic contributions of the surrounding region. Further, single-molecule Förster resonance energy transfer (sm-FRET) experiments suggested that thermal fluctuations are crucial for cognate tRNA selection [53].

These studies used the region surrounding the A site for simulation. In order to include the expected long-range coupling and the entropic contributions from the ribosome surrounding in our simulations we use the whole ribosome to model tRNA translocation. The significant role of these couplings was confirmed by kinetic [54] and cryo-EM [55] measurements on the decoding center.

A proposed long-range coupling mechanism in the ribosome is the allosteric threE site model of decoding, which proposed that an occupied E site increases the fidelity during decoding. In consequence, we would expect that the ribosome is transiently occupied by three tRNAs at the A, P and E site, resepctively, as seen in crystal structures. This would render the exit of the E site tRNA prior to decoding unlinkely, in contrast to the model of translocation that we assume for our simulations (see Fig. 1.7). However, kinetic measurements largely refuted the allosteric E site model [56, 57]. Further controversies along the lines of structure-dynamics relationship were seen for a new proposed mechanism on ribosomal decoding [58, 59].

Minute conformational changes in the decoding center upon successful codon recognition on the small subunit of the ribosome lead to GTPase activation on the large subunit through a cascade of structural changes (Fig. 1.8iii). The mechanism of GTPase activation in EF-Tu is disputed, explicit quantum mechanical treatment of the chemically active region draws a diverse picture [60, 61, 62]. Simulations show that this GTPase-associated center might pro-

---

[1]Ribosomal proteins of the large and small subunit are numbered sequentially and named L# and S#, respectively.

vide the necessary flexibility for the suggested large-scale L7/L12 movements [63]. Similarily to our work, an all-atom model of the ribosome was refined to cryo-EM data by Trabuco et al. [64]. From this model, a mechanism was proposed on how ribosome induced conformational changes of the GTPase-switch-region control GTP hydrolysis (Fig. 1.8iv) [65]. GTP hydrolysis is then followed by phosphate (Pi) release and EF-Tu unbinding (Fig. 1.8v).

The second discriminatory kinetic step after the irreversible GTP hydrolysis is proofreading (Fig. 1.8v). It ensures that binding of cognate tRNAs is highly favoured over non- and near-cognate binding to the A site [66, 42] (Fig 1.8vi). Only after aminoacyl-tRNA is fully accommodated in the A site, the elongation cycle continues with the peptide transfer reaction (Fig. 1.8vii).

MD studies on A site accommodation upon proofreading showed that entropic contributions together with conformational gating helps ribosomes separate cognate from near-cognate tRNAs during accommodation [67, 68]. The recurring motif we will use also in our work is the relation computational studies establish between conformation, dynamics and the statistics on the underlying energy landscape and the connection of experimental rates with simulation results [69].

After successfull A site decoding and accommodation, the CCA-ends of both tRNAs are precisely positioned in the peptidyl transferase center, PTC. Immediatly afterwards, the peptide transfer reaction transfers the nascent peptide chain from the P site tRNA to the A site tRNA. By this reaction, the nascent peptide chain is elongated by the one amino-acid carried by the A site tRNA.

Peptide transfer

The peptide transfer center is the catalytic site of the large ribosomal subunit. The interpretation of the role of the peptide transfer center experienced two major paradigm shifts. The first paradigm-shifting discovery was that the ribosome is a ribozyme [9]. The fact that its catalytic activity is performed by RNA nucleotides, not protein, provides further evidence for the RNA-world hypothesis [70]. The second shift occured upon the discovery of its general catalytic activity, where it is now believed that the PTC is a versatile catalyst of chemical reactions due to its ability to position substrates through electrostatic and hydrogen-bonding interactions [35]. The ribosomes substrate catalysis by entropy reduction [71] re-emphazises the importance of a combination of structural and dynamical information to understand ribosome function.

Due to the bond-breaking and proton shuttle mechanism of the peptide transfer reaction requires an explicit treatment of the involved quantum mechanics. Hybrid Quantum Mechanics / Molecular Mechanics simulations suggested a variety of proton shuttle mechanisms [72, 73].

tRNA translocation

After the nascent peptide chain has been transferred to the A site tRNA, both tRNAs translocate to their new binding sites as shown in Fig. 1.9. The mRNA moves along with the tRNAs, thus exposing the next codon in the vacated
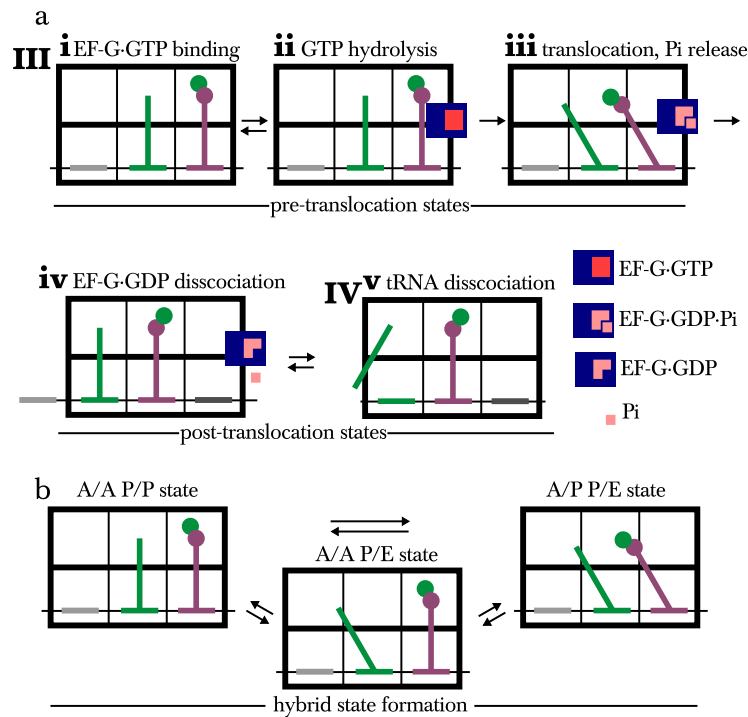
Figure 1.9: Kinetic model of (a) EF-G catalyzed translocation (b) and hybrid state formation during tRNA translocation. (a) Upon peptide bond formation (III, cf. Fig. 1.7), EF-G·GTP binding (i) promotes tRNA translocation through GTP hydrolysis (ii). After tRNA translocation and Phosphate release (iii) the former A site tRNA now occupies the P site and the former P site tRNA the E site (iv). EF-G·GDP (iv) and the E site tRNA (v) dissociate from the ribosome, priming it for a next round of elongation (IV cf. Fig. 1.7). (b) shows the proposed kinetic model of hybrid state formation.

A site. This movement divides translocation in two steps, pre-translocation and post-translocation. The concerted motion of mRNA and tRNA is accelerated by EF-G binding and GTP-hydrolysis [28], but its details remain unclear so far [74]. It is known however, that EF-G stabilizes the so-called hybrid states [75, 76].

In contrast to the classic state, where both tRNAs are fully accomodatad in the ribosomal A and P site on the small and large ribosomal subunit, hybrid states as depicted in Fig. 1.9, are pre-translocation complexes, where the tRNAs moved towards their new respective binding sites on the large subunit, but stay put on the small subunit [77, 75, 78]. Hybrid states are named A/P for the tRNA bound to the A site on the small subunit and to the P site on the large subunit and P/E, respectively. These hybrid states are rapidly occupied after peptide bond formation. Peptidylation further increases the affinity of tRNA for the A/P site and promotes EF-G binding [75]. Mutation studies showed that A/P and P/E site movement are seperated events [79], which suggests a kinetic model of hybrid state formation as depicted in Fig. 1.9, where tRNAs fluctuate between classical and hybrid states [80, 78].

Earlier cryo-EM work visualized tRNAs in the A/P and P/E hybrid states without bound factors [81, 82], and further found the P/E tRNA in flexible conformation within the hybrid states [83]. The cryo-EM data we use in our work, shows that pre-translocation classic and hybrid states comprise a continuum of ribosome conformations [6]. A sorting of these configurations according to tRNA conformations lead to the categorization of pre-translocation complexes into states pre1 through pre5, where the pre4 state corresponds to the before-described hybrid state II.

The tRNA movement is coupled to the rotational movement of two small subunit parts, the body and the head with respect to the large subunit [84, 85, 32]. Small subunit body rotation and head swiveling correlate with hybrid state formation. This rotational movement occurs sponateously [86]. In our study we will address whether tRNA movement or the small subunit head and body movement dominate tRNA translocation.

Another movement that was found to correlate with P/E hybrid state formation is the L1-stalk movement into a closed conformation into the ribosome, where it contacts the P/E site tRNA [87, 88]. We will investigate the coupling of the L1-stalk movement to tRNA motion and if, and in what way, the L1-stalk drives the tRNA.

In contrast to the extensive computational studies on the A site, computational studies of translocation are scarce, due to the large scale of movement and the unresolved role of EF-G in the process. The computational cost that goes along with the large scale of the simulation system was partially circumvented using coarse-grained methods, based on shorter, thus less expensive, all-atom simulations [89]. Another approach to reduce the computational cost was to interpret cryo-EM maps using simulations of the movements of tRNA in solution [83].

A ruggedness estimate of the free energy landscape for tRNA has been given by a μs simulation of an intermediate state of translocation, based on the difference between free and effective diffusion [90, 91]. The short time-scale

diffusion in the energy landscape sees barriers of $\approx 1.7 k_B T$. This indicates that the necessary condition for using Kramers' theory is fulfilled that the free energy barriers exceed thermal flucutations for tRNA translocation in the ribosome.

Elongation termination

Elongation is terminated by the recognition of a stop codon. This stop codon recognition does not follow the established static base-pairing picture [92]. The binding free energy estimates from molecular dynamics simulations suggest how the high specificity is achieved for binding of one protein to two different, cognate sequences [93]. This provides another example of a dynamic picture complementing the static-picture, in this case the "tRNA-mimicry" that emerged from previous studies based on X-ray data [94, 95, 96].

2. Concepts and Methods

2.1 Energy Landscapes and Reaction Rates

To describe the rich dynamics of the ribosome and its parts, the tRNA and L1-stalk movements as well as head and body rotations, we use statistical mechanics and thereby describe tRNA translocation in terms of its macrostates and the transitions between them.

During translation, the ribosome visits a series of metastable states, as described by cryo electron microscopy and fluorescence measurements. We connect computer simulation of biomolecules to biochemical measurements, and make predictions of the dynamics between these states, not accessible to experiments. Therefore, we derive relations between underlying statistical mechanics and the measurement of experimental observables. Numerous model assumptions are neccessary to make this connection, which will be presented along with the derivation.

We will conclude with an expression for the transition rates between metastable states in the high friction limit under the assumption of memory-less processes as an extension of Kramers' transition rate theory to multiple dimensions. To arrive at an expression for the transition rates, we describe diffusion on an energy landscape over a set of reaction coordinates. The free energy landscape in turn is derived from the equilibrium properties of a thermodynamical system.

Generalized classical ensembles

An effective potential G, the free energy, for an equilibrium ensemble of ribosomes in configuration x and momenta p is determined by its potential energy $E(x)$, its temperature T and given ensemble properties $\langle A_i \rangle$. The ensemble properties are defined by the microstate occupation probability density $\langle A_i \rangle := \int A_i(r)\rho(r)\,dr$. An explicit expression for the microstate occupation probability density ρ follows from the assumption of maximum entropy

$$\rho(x) = \exp(-\beta E(x) - \sum_i \lambda_i A_i(x))/Z \; [97].\tag{2.1}$$

Here $\beta = (k_B T)^{-1}$ is used with the Boltzmann constant k_B. Z normalizes ρ, i.e. Z is the partition function for the given ensemble and the parameters λ_i are chosen such that $-\partial \ln Z / \partial \lambda_i = k_B \langle A_i \rangle$. The effective potential over a
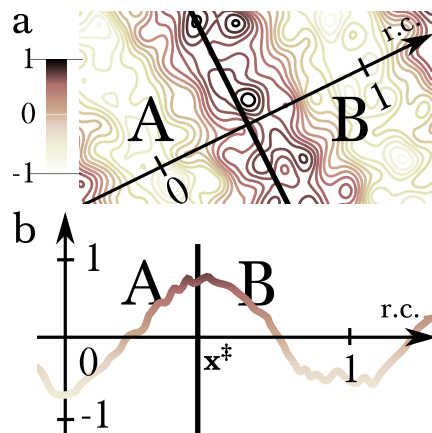
Figure 2.1: Transition between two states described by reaction coordinates.
(**a**) Schematic contour plot of a free energy landscape. An appropriate choice
of a reaction coordinate well sepereates states A and B. The transition seam is
indicated by a gray line. (**b**) Free energy landscape after projection onto one
reaction coordinate.

configuration space subset Ω is then defined as

$$G(\Omega) := -k_B T \ln \int_\Omega \rho = -k_B T \ln \rho_\Omega. \tag{2.2}$$

Reaction coordinates

Figure 2.1 shows how the effective potential G provides a means to reduce the
system description to reaction coordinates $X := f(x)$ by integrating over fast
degrees of freedom that are regarded noise,

$$G(X) = -k_B T \ln \int_{f(x)=X} \rho(x)\, dx. \tag{2.3}$$

To reduce the vast conformational space of ribosome conformations, we de-
scribes the transitions between two states in the relevant subspace Ξ for this
transition.

For systems coupled to a heat-bath this time-scale seperation is valid when
the dynamics on the relevant subspace Ξ is governed by barriers in the ef-
fecitve potential that are larger than thermal fluctuations, $\min(G_{X \in \Xi}) + k_B T \ll
\max(G_{X \in \Xi})$ [98, 99].

To describe transitions by reaction coordinates X, we require that they
seperate the states i, j, i.e. each value of X can be assigned to either state. For
a correct treatment of the dynamics, we further require that the states are
not seperated in any subspace orthogonal to the reaction coordinates, $\Omega_X =
\{x | f(x) = X\}$.

Probability density dynamics in an effective potential

Now that an effective equilibrium potential $G(X)$ over reaction coordinates is defined, a description for the dynamics of an arbitrary probability density distribution $p(X)$ in this potential is obtained, the Smoluchowski equation.

By discretization of the reaction coordainte space into states i, j the treatment changes from probability densities $\rho(X)$ to absolute probabilites P^i, P^j. The time-dependent occupation probability of these states i, j is described by the generalized Master-equation,

$$\dot{P}^i = \sum_j \int_0^t w_{ij}(t - t')P^j(t) - w_{ji}(t - t')P^i(t) \, \mathrm{d}t' \tag{2.4}$$

where $w_{ij}(t)$ is the time-dependent transition probability for a transition from j to i [100]. For future calculations we make the approximation of a memoryless process in reaction coordinate space, which reduces the Master-equation to $\dot{P}^i = \sum_j w_{ij}P^j - w_{ji}P^i$.

The dynamics of the probability density distribution from the Master-equation is obtained from the continuum limit. For our purpose we use the high friction limit, where the momenta are described by a Maxwell velocity distribution, i.e. $p(X, \dot{X}, t) = p(X, t)\exp(-M\dot{X}/2k_BT)$ [101, 102]. We further require that in the long-time limit, the obtained equation describes the equilibrium probability distriubtion $\rho(X)$ for all possible initial probability distributions $p(X)$. With these assumptions, the continuum limit yields the Smoluchowski equation,

$$\dot{p}(X) = \xi^{-1}\nabla \left(\nabla + \nabla\beta G(X)\right) p(X), \tag{2.5}$$

which describes the time dependent behaviour of the probability density with a viscous drag ξ [103, 104].

Transition rates

Now dynamics of our biomolecular system will be described in terms of transition rates k_{AB} between two meta-stable states A and B. For this we assume that a single barrier $X^\ddagger$ seperates the two states which defines a first-order saddle-point on the free-energy landscape $G(X)$. Additionally we assume that initially, all states in A are occupied according to the equilibrium occupation probability and are unoccupied elsewhere.

With these assumptions, the rate of transition from A to B over the barrier $X^\ddagger$ is given by the ratio between the populations in A and barrier $X^\ddagger$. The further assumption of a harmonic well in A with a single minimum at X_0^A and a quasi-harmonic barrier top eventually yields the transition rate approximation used through the rest of this work [105, 106],

$$k \propto \omega_1^\ddagger \frac{\prod_{i=1}^d \omega_i^A}{\prod_{i=2}^d \omega_i^\ddagger} \exp(-\Delta G^\ddagger/k_BT), \tag{2.6}$$

where $\Delta G^\ddagger = G(X^\ddagger) - G(X_0^A)$ and ω_i denote the eigenvalues of the Hessian of G, i.e. the curvature of the free energy surface at X_0^A and $X^\ddagger$, respectively. $\omega_1^\ddagger$ is the curvature in along the transition direction over the barrier top.

Table 2.1: Scope of molecular dynamics simulation. Adapted from van Gunsteren [108].

	Crystalline Solid State	Liquid State Macromolecules	Gas Phase
	Reduction to few degrees of freedom by symmetry	Essential many-particle system	Reduction to few particles by dilution
Quantum Mechanics	possible	hard	possible
Classical Mechanics	easy	**Molecular dynamics**	trivial

Brownian dynamics

We change the treatment of our system from a probability density description to the behaviour of a single system trajectory. From the requirement that the ensemble limit of single system trajectories reproduces the Smoluchowsky equation (2.5), the Brownian dynamics equation is obtained,

$$\dot{X} = -\xi^{-1}\nabla G(X) + \sqrt{2\beta^{-1}\xi}R,\qquad(2.7)$$

where R describes Gaussian white noise with zero mean and unit variance. [1]

2.2   Computer simulation of biomolecules

Our goal is to understand the structure, dynamics and energtics i.e. the physical properties of the ribosome. When relativistic effects are neglected, the most general treatment is to propagate its wave-function $|\psi\rangle$ via the time-dependent Schroedinger equation,

$$H|\psi\rangle = i\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\psi\rangle\,,\qquad(2.8)$$

using its Hamiltonian H. By this equation all physical properties of the ribosome are inherently described. Molecular dynamics simulations [109] is a means to approximate the time-dependent Schroedinger equation to an extent that allows for computationally feasable treatment of biomolecules.
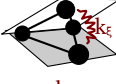
Physical approximations

Three approximations of the Schroedinger equation make a numerical solution tactable for the simulation of biomolecules.

First, the BORN-OPPENHEIMER approximation [110] seperates fast and slow modes and thus reduces the dimensionality of the tensor-product space to

---

[1]Alternatively Brownian dynamics equations are derived as a simplification of the Langevin equation [107]. We made the respective simplifying assumptions in deriving the Somuluchowsky equation above.

Table 2.2: Empirical terms that approximate the electron potential. The k_i denote the force constants, r_{ij} the distance between atoms i and j, P_n the periodicity for the dihedral potential, γ its phase angle and A_{ij}, B_{ij} the van der Waals interaction parameters.

Interaction	Force Field Term	Analogy
Bonded		
bond stretching	$\sum_{\text{bonds}} \frac{1}{2} k_b \left(r_{ij} - r_0 \right)^2$	
bond angle vibrations	$\sum_{\text{angles}} \frac{1}{2} k_\theta \left(\theta - \theta_0 \right)^2$	
extraplanar motions	$\sum_{\text{imp}} \frac{1}{2} k_\xi \left(\xi - \xi_0 \right)^2$	
dihedral torsions	$\sum_{\text{dih}} \frac{k_\phi}{2} \left(1 + \cos(P_n \phi - \gamma) \right)$	
Non-bonded		
van-der-Waals	$\sum_{i,j} - \frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}}$	
Coulomb	$\sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_r\epsilon_0 r_{ij}}$	

that of the direct product of nuclei and electrons, $|\psi\rangle = |\psi_n\rangle|\psi_e\rangle$. Splitting the Hamiltonian in a kinetic part of the nuclei and an electronic part, $H = T_n + H_e$, the approximated Schroedinger equation for the wave function of the nuclei $|\psi_n\rangle$ reads

$$\left(T_n + V\left(r \right) \right) |\psi_n\rangle = i\hbar \frac{d}{dt} |\psi_n\rangle, \tag{2.9}$$

where $V(r)$ describes the dependence of the electronic part of the wave-function on the positions of the nuclei r.

Second, force fields approximate the potential $V(r)$ as a sum of interaction functions as listed in Table 2.2. Additionally to non-bonded interactions that describe interactions between all atoms, interactions between chemically bound atoms are described with extra terms.

Third, instead of the time-development of the wave-function, only the time-development of the expectation value of the nuclei positions is calculated. This approximation is performed through the EHRENFEST theorem. By this approximation, the motion of the nuclei is described by NEWTONS equations of motion,

$$\begin{aligned} \frac{dr}{dt} &= v \\ \frac{dv}{dt} &= -\frac{1}{m}\nabla V\left(r \right) = \frac{1}{m} F\left(r \right). \end{aligned} \tag{2.10}$$
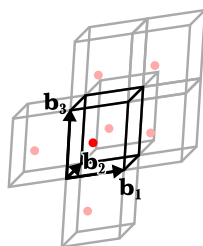
Figure 2.2: Periodic boundary conditions: A particle (red) in the unit cell (black) and its periodic images (light red). When moving into one of the grey cells it is shifted back to the black cell.

Numerical integration of the equations of motion

The leap-frog algorithm numerically integrates eq. (2.10) in a numerically stable, energy-conserving fashion. It uses the following TAYLOR expansion based scheme[2]:

$$
\begin{aligned}
\boldsymbol{r}(t + \Delta t) &= \boldsymbol{r}(t) + \Delta t \boldsymbol{v}(t + \Delta t/2)\,. \\
\boldsymbol{v}(t + \Delta t/2) &= \boldsymbol{v}(t - \Delta t/2) + \frac{\Delta t}{m} \boldsymbol{F}\left(\boldsymbol{r}(t)\right)
\end{aligned}
$$

To ensure numerical stability the integration time step `Δt` is chosen an order of magnitude faster than the time-scale of the fastest motions in our simulations. Since bond-stretching motions of hydrogen atoms are constrained with the LINCS algorithm, and fast bond-angle vibrations with virtual sites, the next fastest motions occur on a timescale of 40 fs. Thus we chose a 4 fs integration time step.

Periodic boundary conditions

Surface effects due to the simulation of small systems are reduced by periodic boundary conditions. As indicated in Fig. 2.2, all simulation particles remain in a unit cell spanned by three linearly independent box vectors $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3$. Periodic boundary conditions are implemented such that any time a particle would leave the unit cell it will be shifted back by a linear combination of the box vectors. Since all particles interact with their periodic images, a sufficient distance between periodic images of the simulated biomolecule needs to be maintained to avoid finite size effects. The number of simulated particles increases with the simulation box size. Thus we chose a rhombic dodecahedron as simulation box that minimizes the number of required solvent molecules due to its near-spherical shape.

Electrostatics and Van-der Waals cut-offs

The number of non-bonded interaction terms scales quadratically with the number of simulated particles, making this the computationally most expen-

---

[2]See [111] for how to cite this method.

sive part of the MD simulation of large systems. Since the VAN-DER-WAALS-interaction drops very fast with the distance, this interaction is cut-off at a defined distance with no significant impact on the simulation results.

In contrast, cut-off approximations of long range electrostatic interactions were found to significantly distort simulations of nucleic acids and proteins [112, 113].

This limitation is overcome by Ewald summation that approximates electrostatic interaction above a cut-off in reciprocal space, accounting for long-range interactions while maintaining $\mathcal{O}$(number of particles) run-time [114].

Pressure and Temperature coupling

To maintain pressure and temperature in our simulations, we use the PARINELLO-RAHMAN [115] barostat and NOSÉ-HOOVER thermostat [116], respectively, which implement a weak coupling scheme. A friction term in pressure, respectively temperature space ensures generation of the grand-canonical ensemble. In contrast to the weak coupling schemes, the more robust BERENDSEN coupling schemes [117] are used to drive systems to equilibium.

Set of starting coordinates

Our simulations rest on experimental data from X-ray diffraction experiments (stored in the pdb-database [118]) and subsequent refinement against three-dimensional cyro electron microscopy data as described below.

Limitations of Molecular Dynamics

Apart from the technical limitation of molecular dynamics to sampling limited phase space, molecular dynamics faces conceptual limitations from the approximations made above.

The effective potentials in Molecular Dynamics simulations of biomolecules are paretrized to reproduce their properties at standard conditions for temperature and pressure. Simulations with parameters far from these conditions will not match observations from other experiments.

By the way the force-field is implemented, chemical bonds cannot break and form. Since quantum effects are subsumed in an effective potential, phenomena as quantum tunneling, excited states, photon absorption, ionization, resonance energy transfer, adiabatic transitions, de-localized electrons cannot be inherently described. Yet, these effect are modelled by MD simulations using multiple effective potentials.

Further, polarization is not explicitly described by Molecular Dynamics.

Numerical description of the diffusion in an effective potential

To numerically estimate transition rates in an energy landscape we solve the Smoluchowsky equation by changing to the single system picture and numerically integrating the Browninan dynamics equation (2.7). The Brownian
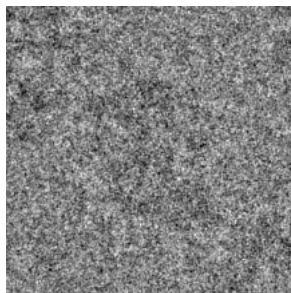
Figure 2.3: Electron microscopy micrograph of a ribosome. The ribosome is visible as the central dark grey region. Note the small signal-to-noise ratio. Data courtesy of Niels Fischer.

dynamics equation is a stochastic differential equation, which we discretize following the Euler-Maruyama scheme [119],

$$X_n = X_{n-1} - \Delta t \xi^{-1} \nabla G(X_{n-1}) + \sqrt{2\beta \xi \Delta t} R. \qquad (2.11)$$

2.3 Cryo-EM fitting

Structure reconstruction form cryo electron microscopy

Cryo electron microsopy (cryo-EM) is an experimental technique that determines the structure of biomolecules through electron scattering on thin layers of cryogenic samples [120]. During the imaging process high energy electrons (100-300 keV) are scattered on the electric field of the nuclei of the sample.

The reconstruction of the three dimensional electron density poses an inverse problem. A direct solution to this inverse problem would yield a probability density distribution for each of the model parameters that are used to describe the observed scattering effects, e.g., voxel-values on a three dimensional grid. The low signal-to-noice ratio of the cryo-EM micrographs as shown in Fig. 2.3 and the large number of these (typically $\gg 10\,000$) renders this computationally prohibitive. Instead, an interative procedure is used that refines a small set of model structures to best match the two-dimensional micrographs.

From each model, two-dimensional reference images are calculated. The micrographs are then classified according to their cross-correlation with the reference images [121, 6]. Following the classification, the procedure is re-iterated with new scattering-density models that are reconstructed from the micrographs in the subclasses.

Cryo-EM works under the assumption that the rapid freezing (10^4K/s) is fast enough to maintain the structural ensemble at the initial temperature. Though radiation damage is reduced through vitrification, it needs to be assumed that the radiation impact on the sample does not alter its conformation.
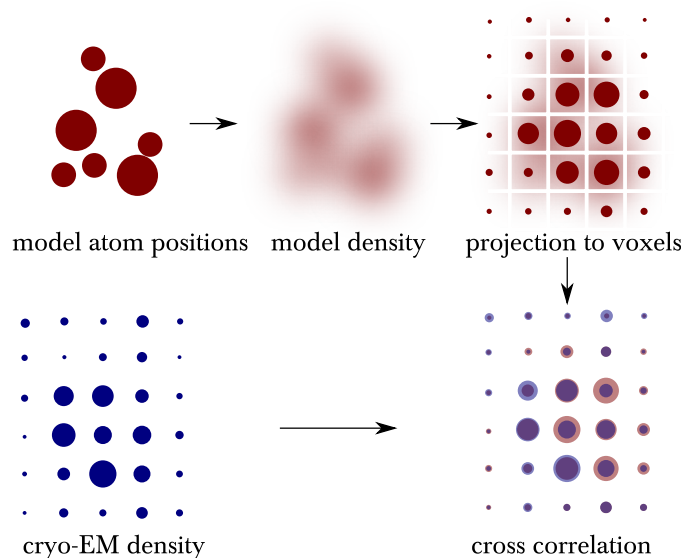
Figure 2.4: Canonical cryo-EM refinement routine. From the model atom positions a density is derived by the convolution of the atom positions with a three dimensional Gaussian function with an amplitude proportional to their atomic number. This density is projected on a voxel grid by integrating of the grid cells. The density values at the voxel is then compared to the values obtained by cryo-EM through cross-correlation.

Refinement of cryo-EM densities

To describe the dynamics of a biomolecule in a state described by cryo-EM we obtained the best matching set of atom coordinates to the cryo-EM data via a method called cryo-EM refinement. Similiarly as above a direct solution comprises probability distributions for the position of each atom obtained from the raw cryo-EM data, i.e. the complete set of two-dimensional micrographs.

The technical limitiations to this approach are overcome by four approximations. First, the reconstructed three-dimensional cryo-EM density, instead of the two-dimensional micrographs, is used for refinement, greatly reducing the amount of data compare to the raw micrographs. Second, instead of an ensemble of structures with occupation probabilities assigned each, the single structure with the highest assigned probability is used. Figure 2.4 shows how in a third approximation, the probability for a structural model to fit to the cryo-EM density is assumed to be proportional to the real space correlation between the structural model convoluted with a Gaussian kernel and the cryo-EM density map. Fourth, a force-field is used to refine the ribosome models. This additional information is needed, because the cryo-EM maps alone do not carry enough information to uniquely determine atom positions in a physically meaningful way.

As force-field in our refinement procedure we used a dynamic elastic network model [122] with additional constraints on the secondary structure of
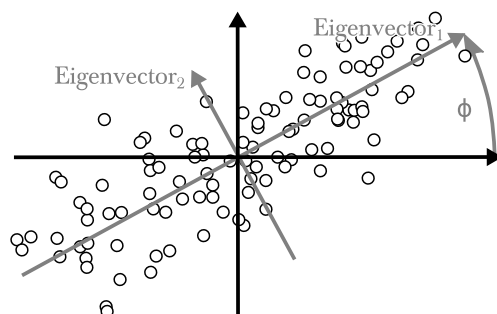
Figure 2.5: PCA in two dimensions. The coordinate system is orthogonally transformed (ϕ), such that the covariance between the projections onto the new vectors eigenvector$_1$ and eigenvector$_2$ is minimized. After Pearson [123].

the ribosomal RNA. This empirical model greatly reduces the computational cost compared to a standard force-field where all long-range interactions are taken into account.

2.4 Ensemble analysis

Principal component analysis

We perform the seperation of time-scales of ribosome motions, i.e., the seperation of fast, noisy modes of ribosome motion from slow, relevant modes of motion, in two steps. First, we perform an orthogonal coordinate transformation that minimizes the covariance between the transformed coordinates as depicted for the 2-dimensional case in Fig. 2.5. Second, we sort these coordinates according to their variance and chose the ones with maximum variance. Technically, the coordinate transformation is obtained from diagonalizing the covariance matrix of the data. The obtained eigenvalues correspond to the variance of the data [123].

Axis and center of rotation for the rotation of a flexible body

The movement of the ribosome small subunit has been approximated as rigid body rotation since the discovery of its ratchet-like movement during translocation [84]. Subsequent studies divided this movement into small subunit body rotation as well as small subunit head rotation, swivel, and tilt. Despite the fact that ribosmal protein and RNA are flexible, an approximate description of these movements as rigid body rotations yields a reaction coordinate that captures most of the flexibility of the small ribosome sub-unit [6].

Determination of pivoting point and axis of rotation has usually been done manually or through an arbitrary definition of "core" residues that have the least flexibility and a subsequent rigid body fit [32, 90].

Figure 2.6 shows our method to determine center and axis of rotation, extended after [124], for a movement of atoms from positions x to x'. We
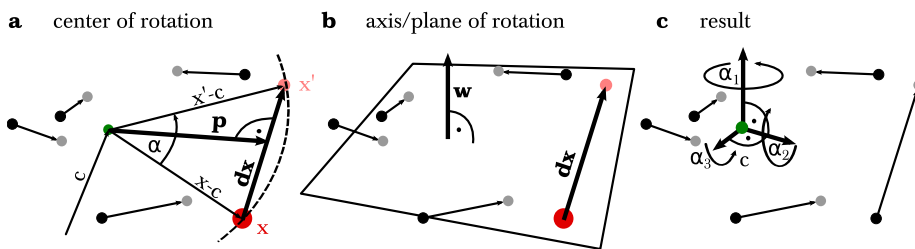
Figure 2.6: Best fitting center and axis of rotation for the the rotation of a flexible body. From the flexible movement of a set of atoms (movement indicated by arrows from black atom positions to grey atom positions), we extract the center and axis of rotation through a non-linear least-squares fit. Panel (a) depicts the constraint used for fitting the center of rotation, where we require that the standard deviation for all atoms from orthogonality of $dx = x' - x$ and $p = dx/2 - c$ and is minimal. Panel (b) shows the fitting constraint to determine the axis of rotation w that requires the standard deviation for all atoms from orthogonality between dx and w is minimal. Panel (c) depicts the obtained parameters from the fitting procedurs, center of rotation and three axis of rotation: the axis of largest rotation, the axis of least rotation and an axis perpendicular to the previous as well as the respective rotation angles.

define the flexible body rotation center with the respective shift vector $dx = x' - x$ as

$$\underset{c}{\mathrm{argmin}} \sum_i \left(dx_i (dx_i - 2c) \right)^2 . \tag{2.12}$$

Now the same diagonalization approach as in the principal components analysis described above yields the center of rotation. Similarly, the axis of rotation is determined from the solution to the following eigenvalue problem,

$$\underset{w}{\mathrm{argmin}} \sum_i (dx_i w)^2 = \underset{w}{\mathrm{argmin}} \, w \left(\sum_i dx_i dx_i^T \right) w^T . \tag{2.13}$$

Here, the other two eigenvectors describe the axis of minimum rotation as well as an axis perpendicular to the axis of maximum and minimum rotation, which are also used to characterize the rotational movement. Overall we obtain a center of rotation as well as three axis of rotation and the respective rotation angles from the movement of atoms from one position to another. To average the determined rotations from the many movements between single snapshots in ensemble data, we use a quaternion-based average [125]. A plain average is used for the center of rotation.

Transition Rate estimates

We will now describe how we estimated the free energy barriers between ribosome translocation intermediates and a global attempt rate A as well as a global barrier calibration factor c. This estimate is also laid out in the Supplementary Information in [22], due to the space limitations in the supplement, in
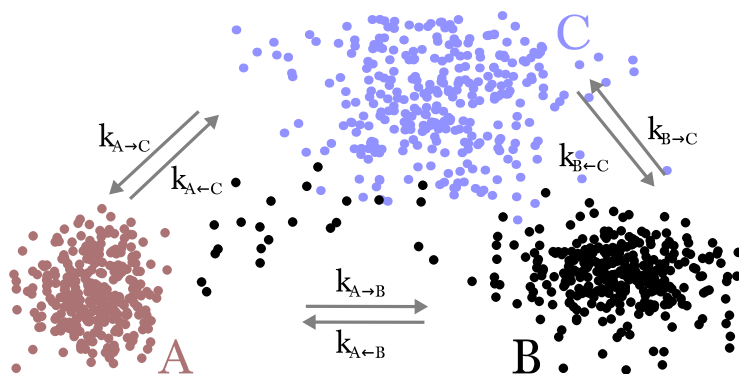
Figure 2.7: From a sparse sampling of an energy landscape in states A **(red),** B **(black), and** C **(blue) we estimate the transition rates** $k_{i \to j}$ **between all pairs of states.**
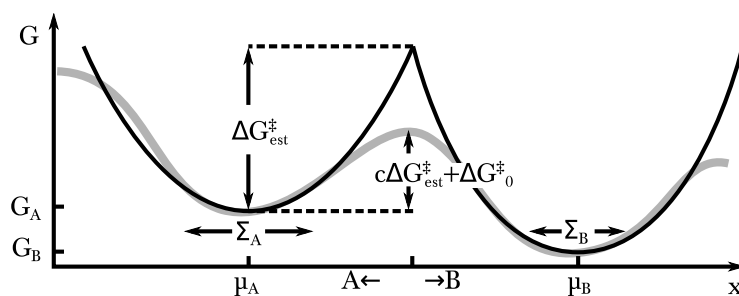
Figure 2.8: Harmonic approximation (black) to the underlying energy landscape (grey) and effect of the global parameters c **and** $\Delta G_0^{\ddagger}$ **to the energy barrier estimate** $\Delta G_{est}^{\ddagger}$ **.** The unknown energy landscape underlying trajectories starting in states A, B is approximated by two harmonic functions with mean $\mu_{A,B}$ and standard deviation $\Sigma_{A,B}$.

a very condensed way that, for example, only implicitely states how the minimum free energy intersection point of two n-dimensional quadratic forms was calculated. Thus, we now expand on the points that were only mentioned in passing in [22].

Fig. 2.7 shows the sparse sampling of an energy landscape states A, B, and C with trajectories $x(t)^A, x(t)^B$, and $x(t)^C$ and the respective tansition rates $k_{A \to B}, k_{A \to C}, k_{B \to C}, \dots$ indicated by arrows. We aim at estimating these transition rates, even though no overlap in sampling these states has been observed (cf. Fig. 2.7, States A and C).

We estimate transition rates from ensembles of trajectories $\{x(t)^i\}$ in three steps. First, we estimate the transition state energy $\Delta G_{\text{est}}^{\ddagger}$ for transitions between all individual pairs of trajectories from a harmonic approximation to the energy landscape using the respective trajectory, as shown in Fig. 2.8.

Second, we combine all local transition energy barrier estimates to estimate two global fitting parameters, a gauge parameter c and global attempt frequency $\tilde{\nu}$ using estimated transition probabilities from transitions observed in the trajectory ensemble.

Third, we estimate the transition rate between all pairs of states, using the gauged transition state energy $\Delta G^{\ddagger} = c \Delta G_{\text{est}}^{\ddagger}$ and global attempt frequency $\tilde{\nu}$ using Kramers' theory.

Harmonic Approximation to the Energy Landscape

The set of coordinates from sampling two states allows us to estimate the energy barrier between them through a harmonic approximation to the energy landscape. First we link the covariance and mean of a trajectory to the Taylor expansion of the local energy landscape, second we use the harmonic approximations to find the isocommiter surface and the saddle point on it.

The Taylor expansion of an energy landscape at a local energy minimum x_0 is

$$G\left(x\right) = G(x_0) + \frac{1}{2}\left(x - x_0\right)^T S \left(x - x_0\right) + \mathcal{O}((x - x_0)^3). \qquad (2.14)$$

The projection of a trajectory to lower dimensions is approximated by a multivariate Gaussian with mean μ and covariance matrix Σ

$$p\left(x\right) = p_0 \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right). \qquad (2.15)$$

From $G(x) = -k_b T \log p(x)$ it follows that

$$G\left(x\right) = G\left(\mu\right) + \frac{1}{2} k_B T \left(x - \mu\right)^T \Sigma^{-1} \left(x - \mu\right). \qquad (2.16)$$

Comparing eqn. (2.14) with eqn. (2.16) shows that the local harmonic approximation to the energy landscape is given by mean and variance of the underlying probability distribution.

Figure 2.9 shows the harmonic apprixmation to two trajectories starting in states A, B with means μ_A and μ_B and covariances Σ_A and Σ_B. Under the assumption that our system shows the same diffusion behaviour on the energy landscape everywhere, $G_A(x) = G_B(x)$ defines the iso-commiter surface,
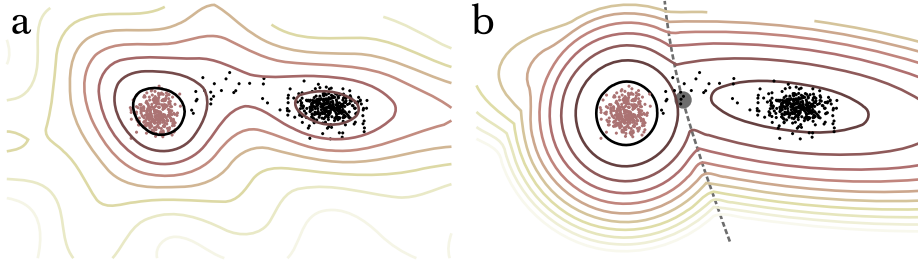
Figure 2.9: Energy barrier estimate from a harmonic approximation to the energy landscape. A randomly generated energy landscape depicted in (**a**) was sampled using a Metropolis-Monte-Carlo algorithmm, starting from two low energy configurations. Each of the sampled point clouds was approximated by their mean and covariance as described in (2.4). The result of this quasi-harmonic approximation to the underlying energy landscape is depicted in (**b**). The grey line depicts the iso-commiter surface, the grey dot the saddle point.

starting from which it is equally likely to end in A and B. The lowest energy value on the iso-commiter surface

$$\min_{G_A(x)=G_B(x)} G_A(x). \tag{2.17}$$

dominates the transition timescales and is used to estimate the transition rate in Kramers' rate theory introduced above. For free energy potentials of form (2.16) for states A and B, a Lagrange multiplier approach,

$$\nabla_{x,\lambda} \left(G_A(x) + \lambda(G_B(x) - G_A(x)) \right) \overset{!}{=} 0,$$

yields the neccessary and sufficient conditions to fulfill 2.17.

$$\left(\Sigma_B^{-1} - \frac{\lambda - 1}{\lambda}\Sigma_A^{-1} \right) x^{\ddagger} - \Sigma_B^{-1}\mu_B = 0$$

$$x^{\ddagger T} \left(\Sigma_B^{-1} - \Sigma_A^{-1} \right) x^{\ddagger T} - 2x^{\ddagger T}\Sigma_B^{-1}\mu_B +$$

$$\mu_B^{T}\Sigma_B^{-1}\mu_B + 2\left(G_B(\mu_B) - G_A(\mu_A) \right) = 0, \tag{2.18}$$

where, without loss of generality, $\mu_A = 0$. Though we found no closed form analytical expression for 2.18, a numerical search on λ is straightforward.

The transition state energy estimate $\Delta G_{\text{est}}^{\ddagger}$ for a transition from state A to B is given by $\Delta G_{\text{est}}^{\ddagger} = G_A(x^{\ddagger}) - G_A(\mu_A)$.

Global Parameters in Trajectory Ensembles

We calibrate the estimated free energy barrier by a linear model

$$\Delta G^{\ddagger} = c\Delta G_{\text{est}}^{\ddagger} + \Delta G_0^{\ddagger} \tag{2.19}$$

with calibration factor c as depicted in Fig. 2.8 and an off-set $\Delta G_0^{\ddagger}$.
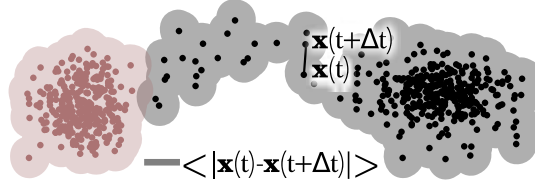
Figure 2.10: A transition occured, when the normalized distance between two states is smaller than unity.

With $\Delta G^{\ddagger}$ as the lowest saddle point energy on the energy landscape and an attempt rate ν to cross the barrier at a temperature T, Kramers' rate theory gives the transition rate as

$$k = \nu \exp\left(-\frac{\Delta G^{\ddagger}}{k_B T}\right). \tag{2.20}$$

The use of the linear model for the free energy barrier callibration yields our model for the transition rate

$$k = \nu \exp\left(-\frac{\Delta G_0^{\ddagger}}{k_B T}\right) \exp\left(-c\frac{\Delta G_{\text{est}}^{\ddagger}}{k_B T}\right) = \tilde{\nu} \exp\left(-c\frac{\Delta G_{\text{est}}^{\ddagger}}{k_B T}\right) \tag{2.21}$$

where the free energy barrier offset is absorbed into the global attemt rate $\tilde{\nu}$. To avoid overfitting of sparse data, we assume that the diffusiveness and the ruggedness of the energy landscape is similar in all sampled regions and therefore use one global value for the gauge factor c and the attempt rate $\tilde{\nu}$. We now extract the model parameters c and $\tilde{\nu}$ by comparing the probability to observe a transition between two states in simulation time t, using our rate estimates and the two model parameters c and $\tilde{\nu}$ to the actually observed transition frequencies.

Determination of gauge factor c and global attempt rate

To obtain the transition frequencies between states in our simulations, we first define transitions between two states, given our ensemble data. To this aim we introduce a normalized distance between two trajectories shown in Fig. 2.10. It is given as the minimum distance between any $\boldsymbol{a}(t) \in A$ and $\boldsymbol{b}(t) \in B$, normalized by the average distance between subsequent points,

$$d_{AB}(\tau) = \frac{\min_{t<\tau} |\boldsymbol{a}(t) - \boldsymbol{b}(t)|}{\langle |\boldsymbol{r}(t) - \boldsymbol{r}(t+\Delta t)| \rangle_{t,\{A,B\}}}. \tag{2.22}$$

We consider a transition if $d_{AB}(\tau) < 1$. Given our model parameters c, $\tilde{\nu}$ and the estimated free energy barrier, we can now express the probability to observe such a transition within simulation time $p(t)$ from reaction kinetics as

$$p(t) = 1 - \exp\left(-kt\right). \tag{2.23}$$

We determined k above in eq. 2.21, so that we can write the transition probability as

$$p(t, \Delta G_{\text{est}}^{\ddagger}) = 1 - \exp\left[-\tilde{\nu}\exp\left(\frac{-c\Delta G_{\text{est}}^{\ddagger}}{k_{\text{B}}T}\right)t\right]. \tag{2.24}$$

Now, a fit of this model to transition frequencies that were observed in the ensemble data to observed in the simulations, allows to obtain the model parameters $\tilde{\nu}$ and c.

3. Publications

3.1   g_contacts: Fast contact search in bio-molecular ensemble data

Of particular interest to the function of the ribosome are areas where its functional parts, e.g. tRNAs, the small and the large subunit, contact each other. Due to the detailed all-atom description of the ribosome, finding contacts between ribosome parts is computationally very expensive when following a naive approach – checking all pairwise distances of atoms in the small and the large subunit requires more than $1.3 \cdot 10^{10}$ (87 352 times 154 576) distance calculations for one snapshot of the trajectory. In the following paper we implemented an approach that drastically reduces the run-time that is needed to find contacts in large systems.

This article, published in "Computer Physics Communications", was conceived and written by myself under supervision of Helmut Grubmüller.
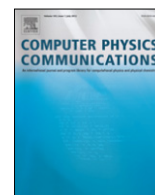
g_contacts: Fast contact search in bio-molecular ensemble data[☆,☆☆]

Christian Blau, Helmut Grubmuller *

Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

ABSTRACT

Short-range interatomic interactions govern many bio-molecular processes. Therefore, identifying close interaction partners in ensemble data is an essential task in structural biology and computational biophysics. A contact search can be cast as a typical range search problem for which efficient algorithms have been developed. However, none of those has yet been adapted to the context of macromolecular ensembles, particularly in a molecular dynamics (MD) framework. Here a set-decomposition algorithm is implemented which detects all contacting atoms or residues in maximum $O(N \log(N))$ run-time, in contrast to the $O(N^2)$ complexity of a brute-force approach.

Program summary

Program title: g_contacts
Catalogue identifier: AEQA_v1_0
Program summary URL: http://cpc.cs.qub.ac.uk/summaries/AEQA_v1_0.html
Program obtainable from: CPC Program Library, Queen's University, Belfast, N. Ireland
Licensing provisions: Standard CPC licence, http://cpc.cs.qub.ac.uk/licence/licence.html
No. of lines in distributed program, including test data, etc.: 8945
No. of bytes in distributed program, including test data, etc.: 981604
Distribution format: tar.gz
Programming language: C99.
Computer: PC.
Operating system: Linux.
RAM: ≈Size of input frame
Classification: 3, 4.14.
External routines: Gromacs 4.6[1]
Nature of problem: Finding atoms or residues that are closer to one another than a given cut-off.
Solution method: Excluding distant atoms from distance calculations by decomposing the given set of atoms into disjoint subsets.
Running time: $\leq O(N \log(N))$
References:
[1] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J.C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and Erik Lindahl, Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics 29 (7) (2013).

1. Introduction

Molecular dynamics (MD) integrators allow simulations of large bio-molecular systems comprising millions of atoms on nanosecond to millisecond time scales [1,2]. These simulations produce a substantial amount of trajectory data, which typically consist of 10^4–10^6 structure "snapshots" (frames). The computational effort to generate the trajectory data scales with $O(N \log(N))$, where N is the number of simulated particles. Efficient analysis tools to extract certain observables from these data are required that exhibit a comparable scaling behavior to the algorithms that generate the trajectory data.

Identifying all atoms of a solute molecule which interact with the solvent, or all close atoms from different subunits of a molecular complex, is a recurring task. From a computational perspective, these tasks require one to identify all pairs of atoms that are closer to one another than a defined minimum contact distance. This task has been described as a spherical range search problem [3].
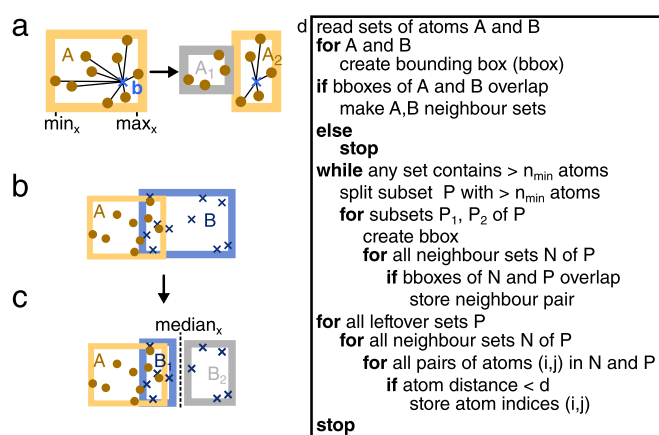
Fig. 1. *Decomposition reduces the number of distance calculations during the contact search between two sets of atoms.* (a) The number of required distance calculations (black lines) is reduced by decomposing the set of atoms A (left) into disjoint subsets A_1, A_2 (right). The distances to $\mathbf{b}$ are only calculated if $\mathbf{b}$ lies within the bounding box of the atoms in A_i or is closer to the bounding box than the minimum contact distance d, indicated by the width of the yellow and gray frames. (b) Generalization of (a) to an arbitrary set B. Another bounding box of the set B is determined and checked for overlap with A. (c) In the next step, the biggest remaining subset is split along the median. The resulting subsets of B are again checked for overlap with A. (d) The implemented set-decomposition scheme given in pseudo-code.

Here, we describe the efficient implementation of an algorithm to obtain contacting atom pairs of two sets of atoms and respective trajectory contact frequencies. The modified k-dimensional tree approach employed has a worst-case run-time of $\propto O(N \log(N))$ for two sets of size N compared with a run-time $\propto O(N^2)$ of a brute-force approach. This high efficiency is achieved by excluding sets of distant atoms from the distance calculation. Combined with the excellent scaling properties of the method on parallel machines, this advantage will be particularly pronounced in future exascale computing applications.

The routine is implemented within GROMACS [4]. Due to the versatile implementation, it can also be applied to other three-dimensional contact searches. Extension to higher dimensions is straightforward.

2. Methods

2.1. Task

Given two sets of labeled atoms, $A = \{\mathbf{a}_i\}$ and $B = \{\mathbf{b}_j\}$, and a minimum contact distance d, the task of the algorithm described here is to identify all contacting atom pairs, i.e., all pairs of atom indices $\{(i, j)\}$ with $\|\mathbf{a}_i - \mathbf{b}_j\| < d$. A brute-force approach would require the calculation of the Euclidean distance between all possible pairs of atoms. The set decomposition scheme implemented here drastically reduces the number of necessary distance calculations and therefore the run-time.

2.2. Algorithm

For simplicity of presentation, we first assume the special case where one of the two sets, B, contains only one atom $\mathbf{b}$ (Fig. 1(a)). This case will subsequently be generalized to arbitrary sets A, B (Fig. 1(b), (c)).

As a first step, the minimum bounding box (bbox, yellow) of set A with sides aligned to the x, y, and z axes is determined. If the distance of $\mathbf{b}$ to the box boundary along the direction of the three coordinates exceeds a given contact distance d, $\mathbf{b}$ is not in contact with A, and the contact search terminates. Otherwise, A is decomposed into two subsets $A = A_1 \cup A_2$ [3]. If the distance of the

two child bboxes (gray, yellow) of subset A_1 to $\mathbf{b}$ or A_2 to $\mathbf{b}$ exceeds d, the respective subset is discarded. Alternatively, the child bbox is further decomposed into two disjoint subsets (not shown), and so on. Decomposition is terminated when all subsets contain less than a given minimum number of atoms. As a final step, the distances of $\mathbf{b}$ to all atoms in the remaining subsets are determined, and the indices i are stored for which $\|\mathbf{a}_i - \mathbf{b}\| < d$.

Fig. 1(b) generalizes the above decomposition procedure to a set B comprising more than one atom. In this case, the bbox is also determined for B (blue), and B is also recursively split into subsets. For each subset B_q, all sets A_p overlapping with B_q are stored. When the decomposition terminates, only the distances for atom pairs i, j in stored pairs of sets A_p, B_q need to be calculated.

2.3. Application to ensemble data

The algorithm is applied to each frame of a given trajectory. Atom pair contacts are counted each frame. From these, the contact frequency is calculated by dividing the contact count by the number of frames analyzed. In addition to atom contact frequencies, residue contact frequencies are determined by defining two residues to be in contact if any of their respective atoms are in contact.

2.4. Efficiency

Four particular properties of the implemented algorithm render it efficient. First, subsets are decomposed along the median atom coordinates, which allows for the application of the median sort algorithm [5] such that the number of atoms in each subset is balanced in minimum run-time. Second, the order of atoms is kept from the previously analyzed frame. Thus the sorting effort is reduced if similar frames are analyzed. Third, after splitting a set A_p into subsets A_{p_1}, A_{p_2}, overlap with subsets $B_q \subset B$ only needs to be checked if B_q overlapped with A_p in the previous step, thus saving a large fraction of overlap checks for newly generated sets. Fourth, decomposition is stopped as soon as the brute-force approach to identify contacts between subsets A_p, B_q becomes on average more efficient than further decomposition at an empirically determined upper boundary for the minimum set size $n_{\min}$.

3. Software structure

The contact search algorithm described here is implemented in C99. It uses the GROMACS application programming interface (API) provided with the MD package GROMACS 4.6 [4].

4. Run description

4.1. Data input

Input arguments are trajectory file names (flagged -f), a GRO-MACS index file name that contains two index groups specifying each set of atoms (-n), a floating-point number that holds the minimum contact distance in nm (-d, by default $d = 0.3$ nm), and the threshold for the largest number of atoms in any node (-bsize). If the option (-resndx) is chosen, a GROMACS structure file (-s) is read.

4.2. Optional switches

The optional switch -nopbc ignores periodic boundary conditions, speeding up the calculation; -resndx calculates the contacts between two groups of residues instead of two groups of atoms.

4.3. Data output

Contact frequencies are written to an output file (name given in -o). If the flag -resndx is set, an additional index file (name

given in `-on`) is written, which contains one index group for each contacting residue and its atom indices.

4.4. Example runs

We performed example runs on a typical test case as well as on a worst-case scenario.

A typical case is provided by a simulation of adenosine triphosphate (ATP) molecules in solution which bind to ribonucleic acid (RNA) [6]. We used an MD simulation of a solvated RNA molecule comprising 1166 atoms and two ATP molecules in solution comprising 86 atoms. The atom pairs and contact frequencies of RNA and ATP that are closer than $d = 0.3$ nm were determined for 20 000 frames of the simulation. The index-file reads:

```
[ RNA ]
1 2 3 4 5 6 7 8 9 10 11 12 13
...
1161 1162 1163 1164 1165 1166
[ ATP ]
1167 1168 1169 1170 1171 1172
...
1248 1249 1250 1251 1252 1253     .
```

The command to analyze the given trajectory `traj.xtc` is:

```
g_contacts -f traj.xtc -n index.ndx   .
```

A worst-case scenario is provided by two highly overlapping sets of atoms, where many set decompositions are required, and only a few subsets can be excluded from the contact search. The example considered here is a contact search in trajectories of a 1 ns simulation of TIP3P water in a periodic cubic water box of 5, 6, 7, 8 and 9 nm length (i.e., 12 426, 21 483, 34 251, 51 393, and 72 768 atoms, respectively), which were screened for contacts between sets of N consecutively labeled atoms. Contacts were searched between $N = 1, 21, 41, \ldots, 981$ atoms. The default distance cut-off of $d = 0.3$ nm was applied.

The respective index-file for $N = 21$ and a 12 426 atom simulation reads:

```
[ group_1 ]
1 2 3 4 5 6 7 8 9 10 11 12 13
14 15 16 17 18 19 20 21
[ group_2 ]
6213 6214  6215 6216 6217 6218
...
6228 6229 6230 6231 6232 6233     .
```

The analysis of the trajectory stored in `traj.xtc` was performed issuing the following command:

```
g_contacts -f traj.xtc -n index.ndx   .
```

The residue-based contact search determines contacts between water molecules, and was performed using

```
g_contacts -s traj.gro -f traj.xtc -resndx   .
```

5. Comparison with other methods

To compare the set-decomposition algorithm and the brute-force approach, the CPU clock cycles were counted that were required for the respective contact search and the storage of the contacts, excluding trajectory-file reading routines.

For the system containing ATP and RNA, the required CPU cycles for contact search and storage were recorded for 101 analyzed frames, which were analyzed every 0.2 ns in a 20 ns trajectory. A speed-up of 9.1-fold was obtained for the implementation of the set decomposition algorithm over the brute-force approach.

In the water box simulation, the required CPU cycles were averaged over the analysis of 45 frames each. Fig. 2 illustrates that our approach exhibits the expected $N \log(N)$ scaling, even in this
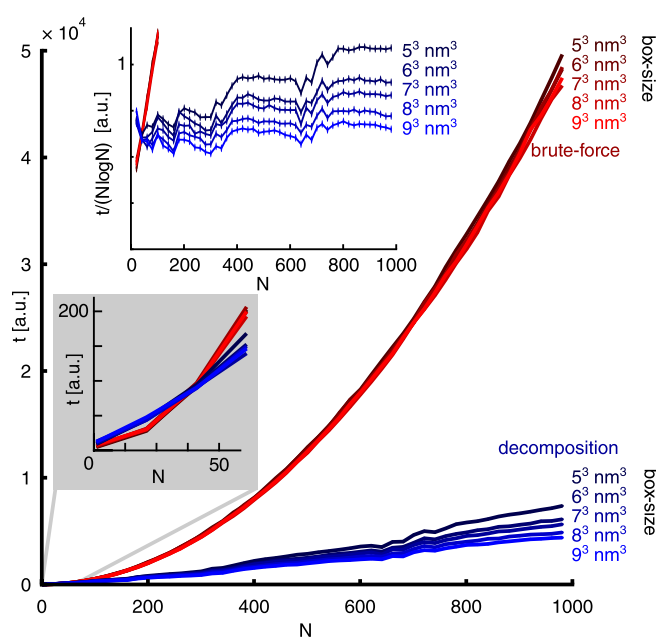
Fig. 2. *Scaling for the set-decomposition approach versus a brute-force contact search between two sets of N atoms for a cubic simulation box of* 5 nm, 6 nm, ..., 9 nm *and* $d = 0.3$ nm *containing water molecules.* The number of CPU cycles (rescaled) used exhibits an $N \log(N)$ scaling for the set-decomposition scheme (blue), while the brute-force algorithm (red) scales with N^2. The gray inset shows the "cross-over" region magnified. The inset above shows the number of CPU cycles used normalized to $N \log(N)$.

worst-case scenario. In contrast, the brute-force approach scales quadratically. For smaller boxes, a deviation from the ideal scaling behavior is observed. We attribute this deviation from $N \log(N)$ scaling to the fact that the number of contacts increases with decreasing box-size. The sorted list of lists approach employed for book-keeping of the contact pairs found results in a slightly worse overall complexity than $N \log(N)$ when very many contacts are found. Further, set decomposition works more efficiently if the subsets are less likely to overlap, which is the case for larger water boxes, explaining the different scaling offsets.

In the current implementation, the "cross-over" in efficiency between the set-decomposition algorithm and the brute-force algorithm (Fig. 2) is seen at set sizes of ≈ 40 atoms each, where the set-decomposition algorithm becomes faster. The set-decomposition algorithm reaches a speed gain of about ten-fold at ≈ 900 atoms per set.

Acknowledgments

References

[1] M. Zink, H. Grubmuller, Mechanical properties of the icosahedral shell of southern bean mosaic virus: a molecular dynamics study, Biophysical Journal 96 (4) (2009) 1350–1363.
[2] J. Klepeis, K.L. Larsen, R. Dror, D. Shaw, Long-timescale molecular dynamics simulations of protein structure and function, Current Opinion in Structural Biology 19 (2) (2009) 120–127.
[3] J. Erickson, P. Agarwal, Geometric range searching and its relatives. advances in discrete and computational geometry, Contemporary Mathematics 223 (1999) 1–56.

[4] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D.V.D. Spoel, B. Hess, E. Lindahl, Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics 29 (7) (2013) 845–854.

[5] D. Knuth, The Art of Computer Programming, Vol. 3: Sorting and Searching, Addison-Wesley, Reading, MA, 1973.

[6] T. Dieckmann, E. Suzuki, J. Feigon, G.K. Nakamura, Solution structure of an atp-binding rna aptamer reveals a novel fold, RNA 2 (7) (1996) 628–640.

3.2 Energy barriers and driving forces in tRNA translocation through the ribosome

In the following publication we describe how tRNA translocate through the ribosome through a combination of MD simulations, cryo-EM, X-ray data, spectroscopic measurements and bioinformatics analysis. In this publication I share first authorship with Lars V Bock. For this publication I predominantly contributed

- Transition rate estimates

- Determination of a kinetic sequence of states

- Check of stereochemical parameters of the models

- Quantification of 30S head and body rotation

I made essential contributions to

- Writing of the manuscript

- Figure preparation

- Refinement of the atomic models against cryo-EM maps

- Comparision to recent crystal structures

- Definition of reaction coordiantes for collective motions

- Check of conservation of contact residues

- Interaction enthalpy estimates

"The proposed manuscript comes across as a summary of a Magnum opus thesis in which gaps are left in the foundations, even with a supplement that is overwhelming in size. [. . .] As it stands, it is impossible to know whether the plausible conclusions are the result of robust ribosome physics shining through unsubstantiated methodology, or a happy accident that might prove irreproducible."

Anonymus

Energy barriers and driving forces in tRNA translocation through the ribosome

Lars V Bock[1,8], Christian Blau[1,8], Gunnar F Schröder[2,3], Iakov I Davydov[4], Niels Fischer[5], Holger Stark[5,6], Marina V Rodnina[7], Andrea C Vaiana[1] & Helmut Grubmüller[1]

During protein synthesis, tRNAs move from the ribosome's aminoacyl to peptidyl to exit sites. Here we investigate conformational motions during spontaneous translocation, using molecular dynamics simulations of 13 intermediate-translocation-state models obtained by combining *Escherichia coli* ribosome crystal structures with cryo-EM data. Resolving fast transitions between states, we find that tRNA motions govern the transition rates within the pre- and post-translocation states. Intersubunit rotations and L1-stalk motion exhibit fast intrinsic submicrosecond dynamics. The L1 stalk drives the tRNA from the peptidyl site and links intersubunit rotation to translocation. Displacement of tRNAs is controlled by 'sliding' and 'stepping' mechanisms involving conserved L16, L5 and L1 residues, thus ensuring binding to the ribosome despite large-scale tRNA movement. Our results complement structural data with a time axis, intrinsic transition rates and molecular forces, revealing correlated functional motions inaccessible by other means.

Ribosomes are molecular machines that synthesize proteins from aminoacyl tRNAs, using mRNA as template. After formation of a peptide bond, the two tRNAs bound to the aminoacyl (A) and peptidyl (P) sites on the small (30S) and large (50S) ribosomal subunits translocate by more than 7 nm to the P and exit (E) sites, respectively, while the next mRNA codon moves into the A site (**Fig. 1a**). During translocation, tRNAs move on the 50S subunit into the hybrid A/P and P/E positions[1] with a concomitant rotation of the 30S subunit relative to the 50S subunit[2–4]. The rate-limiting step of translocation is the displacement of the codon-anticodon complexes on the 30S subunit; this, followed by the reversal of the subunit rotation, yields the post-translocation complex. Translocation is promoted by elongation factor G (EF-G) and is driven by GTP hydrolysis. In the absence of the factor, spontaneous, thermally driven tRNA translocation can occur[5–8], and this seems to involve the same intersubunit interactions that occur in the presence of EF-G[9]. Spontaneous translocation is an equilibrium process, in which the tRNAs make rapid, spontaneous excursions in both forward and backward directions[5,6,10]. Preferential directionality is determined by the affinities of the tRNAs for their respective binding sites[5,6]. The process of translocation entails fluctuations of tRNAs[4,11–14] and of the components of the 50S subunit such as the L1 stalk[3,15–17]. A recent cryo-EM work revealed a large number of different conformational states for spontaneous, thermally driven tRNA movement through the ribosome[10]. However, precisely how the thermal fluctuations of tRNAs and of parts of the ribosome

cooperatively drive the tRNA movement is unclear. It is also unclear whether and how synchronous movements—such as those involving intersubunit rotations, the L1 stalk and tRNA[fMet]—are coupled to one another. Furthermore, it is unknown how efficient tRNA handover from one binding site to another is achieved, despite the considerable structural changes along the translocation path. To address these questions, we combined data from X-ray crystallography and single-particle cryo-EM with molecular dynamics (MD) simulations.

RESULTS

Structural models of translocation intermediates

We refined crystal structures of *E. coli* ribosomes[18] against 13 selected cryo-EM density maps[10] of ribosomes assembled in the post-translocation state, with P-site fMet-Val-tRNA[Val] (P/P state) and deacylated tRNA[fMet] in the E site, and the tRNAs spontaneously moved into their A/A and P/P states, respectively[5] (**Supplementary Fig. 1**, **Supplementary Video 1** and **Supplementary Table 1**, models and refinement described in **Supplementary Note 1**). Several flexible structure-refinement methods have been developed[19–23]. In a recent independent benchmark study[24] comparing four methods[20–23], the DireX refinement method used here was shown to produce on average the most accurate results[23]. All four methods were shown to yield high-quality atomic models from a combination of high-resolution crystal structures and low-resolution data. Indeed, an independent MD-based refinement of the pretranslocation (pre) state pre1b cryo-EM

Figure 1 Structural transitions and interactions along the translocation pathway. (**a**) Left, schematic of tRNA^{Val}, tRNA^{fMet} and L1-stalk motions (indicated by arrows). Right, schematic of 30S head and body rotations. (**b**) Backbone r.m.s. deviation (r.m.s.d.) between the MD structural models of 13 intermediate states of translocation (substates are denoted by a and b) and two crystal structures[26] (solid, whole ribosome; dashed, tRNA-binding region). (**c**) Motions of tRNAs and L1 stalk quantified by reaction coordinates (R. c.) for the 13 states. Arrows denote directionality of motions as in **a**; black bars indicate the fluctuation range covered by each simulation; colored bands denote interaction enthalpies. (**d**) Estimated intrinsic transition rates (denoted by line thickness) between intermediate states of translocation (circles) for motions identified in **a**. (**e**) Changes in the angles of the 30S head tilting, swiveling and 30S body rotation. Black bars indicate the fluctuation range covered by each simulation. (**f**) Molecular driving forces between L1 stalk and tRNA. Two structures representing the range of distances from the fully closed (top left) to open (top right) conformations. Distance dependence of the potential of mean force (PMF, black line, with errors in gray) and interaction enthalpy (circles, colored according to state) between tRNA^{fMet} and L1 stalk. A mechanical analog of the repulsive (sketched gray line, red region) and attractive (green region) regimes is shown below.

map by one of the alternative methods[25] resulted in a structure very similar to that obtained with DireX, with an r.m.s. deviation of 3.2 Å (**Supplementary Notes 2** and **3**).

Validation of the models

We first assessed the quality of our structural models (**Fig. 1b**). The models of the pre1a state closely resemble the crystal structure of the ribosome trapped in the classical state[26], with an r.m.s. deviation of 4.9 Å for the whole ribosome. The structural changes during translocation between the classical state[26] and the other pretranslocation states are captured by r.m.s. deviations of up to 11 Å. After the refinement was completed, the structure of one late-pretranslocation state was reported[26], with the P-site tRNA in the hybrid P/E position and the ribosomal subunits rotated to a high degree; this offered the possibility of testing the quality of one of our structure predictions. Comparison of the pre4 state with this crystal structure gave an r.m.s. deviation of 4.7 Å for the whole ribosome complex and 3.5 Å for the tRNA-binding region. Apparently, our refinement procedure yielded an accuracy of the pre4 model similar to that of the classical pre1 model directly derived from an X-ray structure in the classical state. All other states showed larger deviations, as expected for structurally distinct intermediates. The positions of the tRNAs were accurately predicted, particularly in the functionally relevant anticodon and CCA-end regions (**Supplementary Fig. 2a**). These independent quality checks also suggest that the models of the other pre- and post-translocation states, for which no crystal structures exist, are similarly accurate. The agreement of the models' local stereochemical parameters with those of available crystal structures (**Supplementary Table 2**) confirms their similar stereochemical quality.
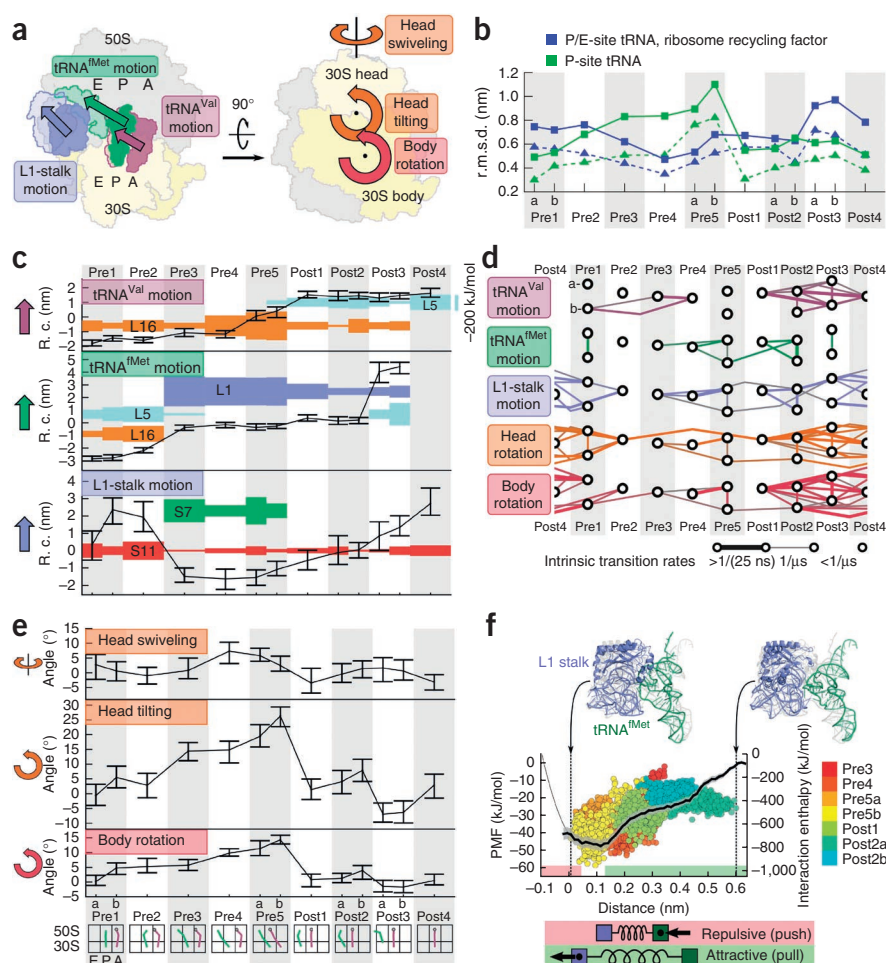
Large-scale conformational motions of the ribosome

The 13 structural models, grouped into nine major pre- and post-translocation (post) states on the basis of tRNA positions, revealed details of tRNA movement as well as large collective motions of the whole ribosome on slow timescales, of seconds to minutes, accessible to the cryo-EM experiments. Structural transitions and interactions along the translocation pathway are shown in **Figure 1**. Six selected motions are indicated (**Fig. 1a**) and characterized (**Fig. 1c,e**). The movement of tRNA^{Val} between the A and P sites and of tRNA^{fMet} between the P and E sites as well as the concomitant movement of the L1 stalk (comprising L1 protein bound to nucleotides 2084–2206 of 23S rRNA) are quantified (**Fig. 1c**).

Understanding of how the ribosome controls tRNA translocation requires (i) resolution of its fast conformational motions, (ii) determination of the rates of these motions and (iii) uncovering of the underlying molecular driving forces. To this end, we carried out all-atom explicit-solvent MD simulations of the entire ribosome for the 13 pre- and post-translocational models (**Supplementary Video 2**, MD setup and simulations described in **Supplementary Note 1**). The system comprised ~2.2 million atoms, and the simulations spanned a total of >1.8 μs. The r.m.s. deviation during the equilibration simulations of each of the 13 structural models served as a third independent control (**Supplementary Fig. 2b**). Large r.m.s. deviations have been shown to point to inaccurate structures[27]. For all simulations, the r.m.s. deviations from the respective starting structure remained small and comparable to the deviations in the simulation started from the crystal structure.

We first analyzed the global swiveling and tilting motions of the 30S head[10,28] and asked how quickly their intermediate translocation

Figure 2 Effect of the tRNAs on the large-angle 30S body rotations. (**a**) Schematic of the tRNA-L5-S13 interaction network. (**b**) Histogram of intersubunit interaction enthalpy for bridge B1b residues, calculated from simulations of the pre5b structural model with (blue) and without tRNA (green). (**c**) Histogram of respective rotation angles derived from MD simulations. The arrow marks the angle of the starting structure as reference. (**d**) Histogram of 30S body-rotation angles derived from cryo-EM analysis of ribosomes in the absence of tRNAs (green) compared to those of states pre1 to pre5 (cyan).

states interconvert. To this end, we determined pivot points, rotation axes and angles for 30S head and body motion (**Fig. 1e**; reaction coordinates described in **Supplementary Note 2**). During the transition from pre1 to pre5, the angles of 30S head tilting and body rotation gradually increase before returning to low values in the pre5-post1 transition. In the post3 state, the rotation angles are even lower than in pre1 or post1, and they finally return to their initial value in post4, thus closing the conformational cycle of the 30S subunit due to tRNA release from the E site. As can be seen from the amplitude of rapid ($>10^7$/s) fluctuations extracted from the simulations, large-scale motions occur not only on the millisecond time scale accessible to single-molecule experiments[29] but also on the submicrosecond time scale. Moreover, the large overlaps between the fluctuation amplitudes of several adjacent states suggest fast transitions between these states, in particular for 30S head swiveling. Conversely, lack of an overlap between adjacent states (for example, between pre4 and pre5 or between pre5 and post1 for tRNA^{Val}) indicates transitions between states that are slower than 100 ns.

Slow tRNA movement governs overall transition rates

To determine which of the observed motions limit the overall interconversion rates between the states, we quantified the overlaps between all states and translated them into free-energy barriers and intrinsic (order of magnitude) transition times of the individual collective motions (**Fig. 1d**, **Supplementary Note 2** and **Supplementary Fig. 3**). Whereas individual motions may be intrinsically fast in isolation, their coupling to slower motions limits the respective overall transition rates. Accordingly, we used the slowest transition between each pair of states in **Figure 1d** to estimate the overall transition rates. For all transitions between states with the exception of post1-post2, the slowest estimated rate is markedly slower than microseconds. This result agrees with rates determined by bulk kinetics[30,31] and single-molecule studies[32–34] for tRNA translocation (milliseconds) as well as with those for L1 stalk–tRNA interactions[12] and intersubunit rotation[29] (seconds). A cluster of high free-energy barriers (i.e., slow transition times) is found at the transition from pre to post states, thus supporting the notion that the pre-to-post transition is the rate-limiting step of spontaneous translocation[5,10]. For the tRNA transitions, low transition barriers for tRNA^{Val} within the pre and post states mostly correlate with high barriers for tRNA^{fMet} and vice versa. Together with the fact that the movements of the two tRNAs are coupled by base-pairing to the mRNA (**Supplementary Fig. 4**), this finding implies that, although intrinsically rapid, the excursions of tRNA^{fMet} between adjacent states (for example, pre5b and post2a) are governed by the slow movement of tRNA^{Val}. Therefore, tRNA motion governs not only the overall rate
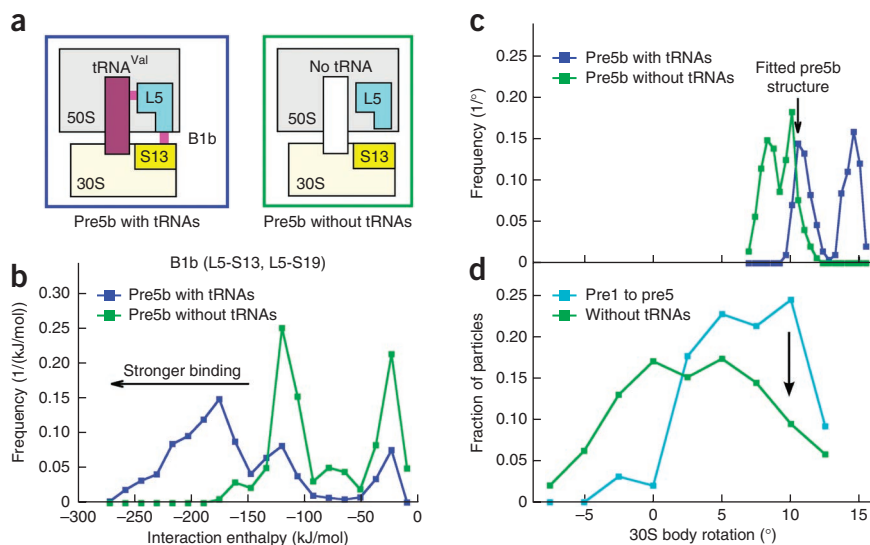
of translocation[35] but also, to a large extent, the dynamics within the pre and post states. In contrast, intrinsic rates for L1-stalk dynamics were present in the submicrosecond range, except for the slower movement from pre to post and from pre2 to pre3. Notably, our simulations revealed a large number of alternative microsecond and submicrosecond transitions of the tRNAs and the L1 stalk (**Fig. 1d**).

Rapid intersubunit rotations couple to slow tRNA movement

All individual transition barriers for 30S motions turned out to be remarkably low, such that the intrinsic motions of the 30S subunit, despite its large size, take place in the submicrosecond range, similarly to the movements of the L1 stalk. The slowest 30S head transitions were found between states pre2 and pre3 as well as between pre5 and post1, consistent with barriers inferred from cryo-EM and biochemical experiments[5,10]. The abundance of low barriers for motions of individual components seems to be a general feature of ribosome dynamics, and this underscores the important role of coupling between collective motions of both tRNAs and the ribosome[3].

One example of such coupling is the stabilization of high intersubunit rotation by tRNAs, which has been previously observed[29]. Our simulations showed that, in state pre5b, the large head tilting and body rotation (angles at ~20° and ~15°, **Fig. 1e**) are stabilized by strong interactions involving the tRNAs and a network of contacts between the two subunits, comprising L5 on the 50S subunit and S13 and S19 on the 30S subunit (**Figs. 1c** and **2b**). In the classical state, these interactions form bridge B1b[13] (sketched in **Fig. 2a**). Removal of the tRNAs in state pre5b should thus weaken these intersubunit interactions with respect to the pre1 state. As a consequence, the highly rotated states should also be destabilized, and a smaller population of large rotation angles should thus be observed. To test this hypothesis, we carried out additional MD simulations of ribosomes in the pre5b state from which the tRNAs were removed (**Fig. 2b,c**; rapid angular rearrangement described in **Supplementary Note 3**). These simulations indeed showed weakened interactions between the B1b residues (**Fig. 2a,b**) as well as decreased 30S body-rotation angles (**Fig. 2c** and **Supplementary Fig. 5**). Indeed, additional cryo-EM data of vacant ribosomes (**Supplementary Note 1**) revealed a markedly reduced population of high body-rotation angles (~50% for angles of 10° and above, **Fig. 2d**). Apparently, our structural models and simulations are sufficiently accurate to capture the underlying small energy differences. A similar effect
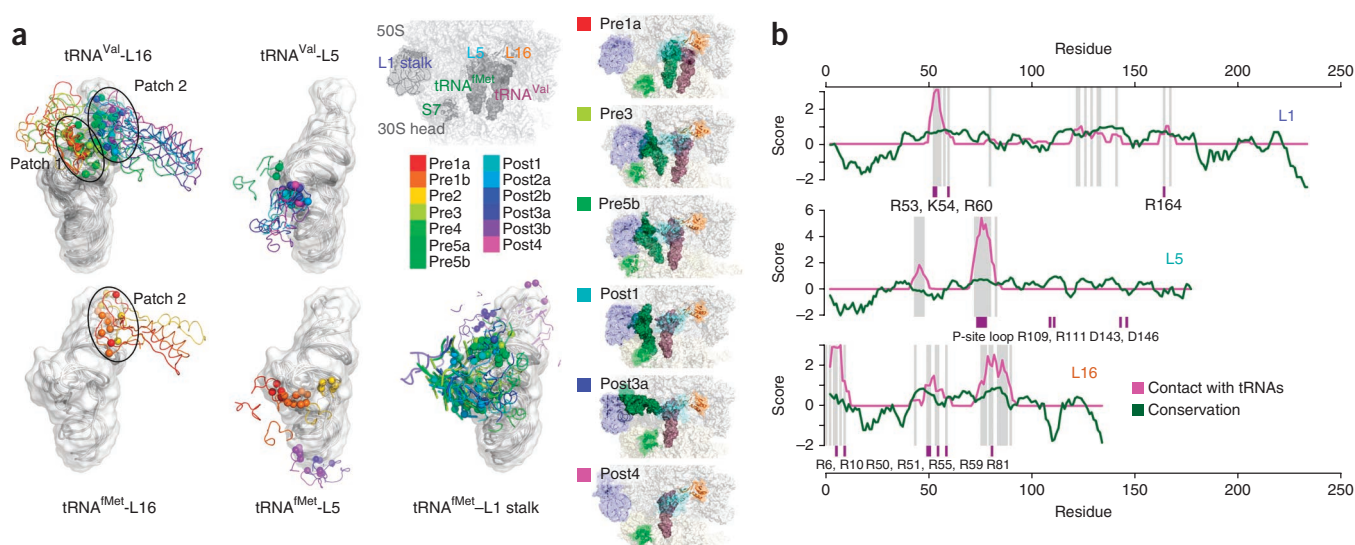
Figure 3 Handover of tRNAs by 50S-subunit proteins during translocation. (**a**) Left, positions and key interacting residues (spheres) of the three main interaction partners, L16, L5 and L1 (ribbons, colored according to the state), relative to tRNA^{Val} (top) and tRNA^{fMet} (bottom). Right, selected conformations relative to the 50S subunit, including the position of S7 on the 30S head for reference. (**b**) Conservation of protein residues contacting tRNAs. Conservation score (higher, more conserved) and contacts score (higher, more frequent contacts) for the L1, L5 and L16 proteins. The scores were smoothed with a 5-aa sliding window. Residues considered contacting are marked with a gray background. Residues highlighted in purple are those specifically mentioned in the text.

on body rotation should be seen for weakened intersubunit interactions, for example, in mutation of the charged L5 residues R109, R111, D143 and D146 to uncharged residues. Indeed, mutation of the corresponding residues in yeast affects translational fidelity[36]. Taken together, these results indicate that previously observed correlated motions[29] are causally connected and explain this coupling in terms of molecular interactions.

Kinetic sequence of translocation intermediate states

The transition-rate estimates determined from the simulations (**Fig. 1d** and **Supplementary Fig. 6a**) enabled us to add time information to the sequence of states that was previously determined by Fischer *et al.*[10] from structural similarity only. To single out the sequence of states that best reflects the motion of the ribosome along the translocation pathway from all (13 − 1)! = 479,001,600 possible linear kinetic sequences that can be formed from permutations of the 12 available conformational states with two tRNAs, we calculated for each of these sequences the overall progression rate from the barrier heights for all five ribosomal components shown in **Figure 1a**. The sequences with the fastest progression rates turned out to be very similar to the one obtained purely from structural similarity of the tRNAs, thus showing that this sequence reflects the kinetics of the system (kinetic sequence of states in **Supplementary Note 2**). Interestingly, removal of any state from this sequence (except pre5b) slows down the progression rate, a result underscoring the kinetic relevance of all states (with the possible exception of pre5b). A systematic scan through all 31 possible subsets of the five conformational motions considered in **Figure 1** (**Supplementary Note 2** and **Supplementary Fig. 6b**) confirms the initial suggestion (based on **Fig. 1d**) that translocation is limited by the motion of the tRNAs. We note that although the preferential overall direction of movement in our experiment is backwards, Fischer *et al.*[10] found that different substates within the ensembles of pre or post states were in rapid equilibrium, and thus at any given time the ribosomes were undergoing transitions in both directions, such that forward and backward directions were equivalent. Similarly, detailed balance also holds for our simulations.

L1 stalk links 30S rotation to translocation by 'pulling' tRNA

Having identified the tRNA motions as the main determinant for translocation efficiency, we expected to find mechanisms that accelerate the rates of these translocation motions. One obvious candidate involves the L1 stalk, which has been suggested to be important for translocation[3,4]. The L1 stalk forms contacts to the tRNA[4,11–13,15,16] as well as to proteins S7 (30S head) and S11 (30S body) (**Figs. 1c**, **3** and **4**). In the pre1 and pre2 states, the L1 stalk is found in an open conformation. In the pre3 state, interactions with S11 are weakened, and the 30S head tilts (**Fig. 1e**), moving S7 into a position that allows it to contact the L1 stalk in the closed conformation. This presumably shifts the L1-stalk equilibrium toward the E site, where it forms strong contacts to tRNA^{fMet} (**Fig. 1c**). Upon decreasing of the tilting angle in post1, contact to S7 is lost, and the L1 stalk moves toward the open
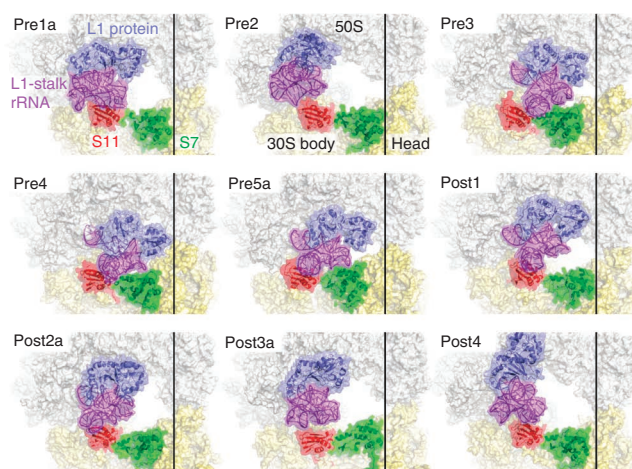
Figure 4 L1-stalk contacts with 30S-subunit proteins S7 and S11. For each state, protein L1 (blue) and L1-stalk rRNA (magenta) as well as proteins S7 (green) and S11 (red) are depicted as ribbons. The 50S and 30S subunits are shown as gray and yellow surfaces. For clarity, the tRNAs are not shown. The vertical black lines serve as reference.

conformation, maintaining its contact to the tRNA[fMet] throughout the post states.

Because structural analyses alone do not provide evidence on whether the L1 stalk accelerates tRNA translocation, we needed to establish that the concerted motion is actually driven by the L1 stalk and not by the tRNA. We calculated the underlying molecular driving forces from our equilibrium simulations of each state (interaction enthalpy) as well as from additional umbrella-sampling simulations (L1 stalk–tRNA[fMet] potential of mean force, **Supplementary Note 2**). The interaction enthalpy and binding free energy changed with the distance between the L1 stalk and tRNA[fMet] (**Fig. 1f**). The positive slopes imply that both the interaction enthalpy and free energy are attractive. These two independent data sets suggest that the L1 stalk actively pulls the tRNA after contact is established in the pre3 state, thereby accelerating tRNA barrier crossing. The opposite repulsive scenario, i.e., the tRNA pushing the L1 stalk, which would slow down barrier crossing, is incompatible with the observed free-energy curve. This result explains both the lower translation rates[37] and the higher stability of the classical tRNA states[14] observed for L1-depleted ribosomes, and it links 30S body and head rotation through the L1 stalk to the motion of the tRNA. That the total interaction enthalpy (**Fig. 1f**) shows the same distance dependency as the binding free energy (albeit with larger values due to partial enthalpy-entropy compensation) suggests that the analysis of interaction enthalpies, to a good approximation, can be used to reveal main interaction sites.

A closer analysis of the interaction enthalpies seen between L1 and the tRNA suggests residues R53, K54, R60 and R164 of L1, which are highly conserved (**Fig. 3b**), as the main interaction sites. L1 residues R53, R60 and R164 contacted residue Ψ55 of the tRNA[fMet] (**Supplementary Table 3**). Mutations of the Ψ55•G18 base pair are known to decrease translocation rates by 80-fold[38], thus supporting the notion that attractive interactions with the L1 stalk actively decrease the barrier for tRNA translocation. A similar effect would be expected from a complementary mutation of the identified L1 residues.

L5 and L16 facilitate tRNA translocation

During translocation, the two tRNAs are handed over from protein L16 to L5 and then to L1. We identified the strong tRNA interactions (identified by enthalpy contributions, **Fig. 1c**) with these large-subunit proteins and analyzed how the position, structure and contact sites of these proteins change from the perspective of the tRNAs in each state along the tRNA pathway from the A to the E site (**Fig. 3a**). Notably, more state-specific contacts to the tRNAs were generally seen for ribosomal proteins than for rRNA (**Supplementary Note 3**). Considerable motions of all involved components are observed; nevertheless, the tRNA remains tightly bound to these proteins, thus enabling accurate adjustment of the tRNA binding free energy. The continuous sequence of interactions might serve to lower free-energy barriers that otherwise would impede tRNA translocation. Indeed, sequence analysis of proteins L1, L5 and L16 revealed significantly higher conservation of the residues identified by contact analysis (overall *P* value = 6.62×10^{-8} obtained from a one-sided permutation test in which the sample comprised $n = 550$ individual amino acids from the proteins), thus corroborating their functional relevance (**Fig. 3b**; conservation of contact residues described in **Supplementary Note 2**).

How are the large structural motions reconciled with the maintained interactions between the tRNA and the ribosome along the translocation path? Closer inspection of interactions revealed two main mechanisms: stepping and sliding. During the A/A-to-P/P
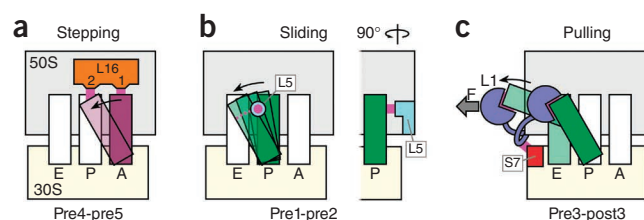
Figure 5 Three mechanisms facilitating tRNA translocation. (**a**) Two distinct contact patches of L16 stabilize A- and P-site positioning of the tRNA (here tRNA[Val], purple) on the 50S subunit. Moving from A to P site, the tRNA steps from patch 1 to patch 2. (**b**) As the tRNA moves into the hybrid P/E conformation (here tRNA[fMet], green), the P loop of L5 maintains a flexible contact (magenta) by sliding along the tRNA (green). (**c**) In pre3, high intersubunit rotations stabilize the closed L1-stalk conformation through S7. After back rotation (post1), this interaction is lost, and the L1 stalk moves into the open conformation, thereby pulling the tRNA into the E site.

transition, tRNA[Val] remained in contact with L16 (**Figs. 3a** and **5**). We used two adjacent interaction patches on L16 (R50, R51, R55 and R59; R6, R10 and R81), each involving positively charged, highly conserved arginines (**Fig. 3b**) that interacted with different parts of the negatively charged tRNA backbone. Upon hybrid-state formation (pre4 to pre5)—when the acceptor stem of tRNA[Val] moved into the P site of the 50S subunit while the L16 conformation remained unchanged—the contact region on tRNA[Val] switched from the first to the second patch. We observed similar interactions with the second patch on L16 for the tRNA[fMet] in its P/P configuration, suggesting that these contacts can stabilize any tRNA in the P site. During the tRNA handover from L16 to L5, the binding of tRNA[Val] to L16 became weaker (**Fig. 1c**), whereas the binding to L5 remained strong in pre5b and throughout the post states. Strong contacts between the tRNA[Val] C56 and the highly conserved P-site loop of L5 (ref. 39) (A74–I78, **Fig. 3b**) were also present in the pre5b state. This finding, together with the large intersubunit rotation angle in the pre5b state, suggests that the tRNA movement is coupled to the intersubunit rotation through L5. In the subsequent post states, the contacts of the L5 P-site loop shifted down the D loop of the tRNA[Val] (**Fig. 3a**), additionally involving nucleotide G19.

In contrast to interactions involved in the stepping motion of the rather rigid protein L16, interactions between tRNA[fMet] and L5 (**Fig. 3a**) were much less localized and were more dynamic. In the pre1 and pre2 states, tRNA[fMet]-L5 contacts shifted from the D loop to the T loop of the tRNA. Upon handover to the L1 stalk in the pre3 state, these contacts to L5 were lost, but new contacts between L5 and the anticodon stem-loop were formed just before the tRNA left the ribosome (in post4). Here, the P-site loop of L5 slides smoothly over the tRNA, thereby flexibly adapting to the changing tRNA position and orientation. Mutations of the P-site loop of L5 in yeast ribosomes impair tRNA binding[39], underscoring the importance of L5 as a guide for the movement of the P-site tRNA. The L1 stalk provides the final contact for tRNA[fMet] after being handed over from L5 and L16. Interactions of L1 with tRNA[fMet], involving the tRNA backbone and the highly conserved, positively charged L1 residues (R53, K54, R60 and R164), were established in the pre3 state and remained unchanged from pre3 to post2.

DISCUSSION

Each of the three different mechanisms facilitating tRNA translocation (summarized in **Fig. 5**) rests on mutual couplings between the tRNAs and parts of their binding region. First, the precise positioning

of the tRNAs in A and P sites is achieved by interactions with L16. To accommodate tRNA motions, L16 'steps' through discontinuous contact patches. Second, the flexible P-site loop of L5 facilitates seamless tRNA sliding across the P site. Third, L1 exerts force to pull the tRNA out of the P site and hence requires internal rigidity. From real-time single-molecule fluorescence resonance energy transfer experiments, the L1 stalk is known to remain associated with the tRNA throughout translocation[12]. Our analysis suggests that L1 forms only one stable, highly conserved contact patch with the tRNA and moves with it as a rather rigid body. The L1 stalk transmits the intersubunit rotation through S7–L1 stalk interactions and promotes tRNA translocation by reducing free-energy barriers between adjacent binding sites. The abundance of interactions between charged, highly conserved residues of ribosomal proteins and the tRNA backbone may provide a general means for the transport of different tRNA species, independently of their sequence, through the ribosome.

Although translocation is a function inherent to the ribosome itself[7,8], the movement, particularly on the 30S subunit, is dramatically facilitated by EF-G. Although the full description of EF-G–dependent translocation is currently not feasible, owing to the lack of structural information, the present work provides insights into the reaction landscape underlying the movement. The simulations provide a view of the energy 'valley' of spontaneous translocation, which allows the tRNAs to move by large distances while maintaining sufficient binding with the ribosome components. Gross deviations from this overall pathway would require a very large energy input and thus seem unlikely even in the presence of EF-G. Rather, EF-G might use the energy of binding and GTP hydrolysis to flatten the energy barriers of the rate-limiting step(s) and to provide the directional bias for forward movement. Understanding the way by which EF-G remodels the free-energy landscape into efficient tRNA translocation is one of the most important future challenges in the field.

Our combined crystallography–cryo-EM–simulation approach reveals fast, large-scale motions of the ribosome, on microsecond time scales, that govern tRNA translocation. Where the small overlap between the fast time scales accessible here and the slower ones observable by bulk kinetics[30,31] and single-molecule studies[12,29,32–34] allows a direct comparison, the obtained transition rates agree within the respective error bounds. Our approach quantifies, from first principles, the picture of a stochastic molecular machine[10,30,33], which fluctuates[12,14,15,29,35] between nearly isoenergetic Frauenfelder-type conformational states through collective and coupled structural transitions[40] (**Supplementary Video 3**). Of all the movements described above, the highest intrinsic barriers are found for tRNA movement throughout the whole translocation pathway. This finding suggests a possible explanation for the strong effect of different tRNA species[5,6] on the rate of spontaneous translocation. Our picture shows how the 30S head and body rotations, through coordinated L1-stalk, S7 and S11 motions, promote and control tRNA translocation. This complements structural[1,3,4,10,11,13,18,26,28,36,39,41–43] and single-molecule data[12,14,15] with intrinsic transition rates of functionally relevant and coupled motions, interaction energies and the underlying molecular driving forces. We begin to uncover the 'gears and wheels' of tRNA translocation through the ribosome from first principles, in terms of a stochastic molecular machine.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Three-dimensional density maps have been deposited at the Electron Microscopy Data Bank, under accession codes EMD-2472 (pre1b), EMD-2473 (pre5a), EMD-2474 (post2b), EMD-2475 (post3a). The atomic coordinates have been deposited at the Protein Data Bank under accession codes 3J4V, 3J52 (pre1a); 3J4W, 3J4X (pre1b); 3J4Z, 3J50 (pre2); 3J4Y, 3J51 (pre3); 3J53, 3J54 (pre4); 3J55, 3J56 (pre5a); 3J57, 3J58 (pre5b); 3J59, 3J5A (post1); 3J5B, 3J5C (post2a); 3J5D, 3J5E (post2b); 3J5F, 3J5G (post3a); 3J5H, 3J5I (post3b); and 3J5J, 3J5K (post4). Details can be found in **Supplementary Table 4**.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

1. Moazed, D. & Noller, H.F. Intermediate states in the movement of transfer RNA in the ribosome. *Nature* **342**, 142–148 (1989).
2. Frank, J. & Agrawal, R.K. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* **406**, 318–322 (2000).
3. Tama, F., Valle, M., Frank, J. & Brooks, C.L. Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. *Proc. Natl. Acad. Sci. USA* **100**, 9319–9323 (2003).
4. Valle, M. *et al.* Locking and unlocking of ribosomal motions. *Cell* **114**, 123–134 (2003).
5. Konevega, A.L. *et al.* Spontaneous reverse movement of mRNA-bound tRNA through the ribosome. *Nat. Struct. Mol. Biol.* **14**, 318–324 (2007).
6. Shoji, S., Walker, S.E. & Fredrick, K. Reverse translocation of tRNA in the ribosome. *Mol. Cell* **24**, 931–942 (2006).
7. Spirin, A.S. Energetics of the ribosome. *Prog. Nucleic Acid Res. Mol. Biol.* **21**, 39–62 (1978).
8. Noller, H.F., Yusupov, M.M., Yusupova, G.Z., Baucom, A. & Cate, J.H. Translocation of tRNA during protein synthesis. *FEBS Lett.* **514**, 11–16 (2002).
9. Liu, Q. & Fredrick, K. Contribution of intersubunit bridges to the energy barrier of ribosomal translocation. *Nucleic Acids Res.* **41**, 565–574 (2013).
10. Fischer, N., Konevega, A.L., Wintermeyer, W., Rodnina, M.V. & Stark, H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* **466**, 329–333 (2010).
11. Agirrezabala, X. *et al.* Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome. *Mol. Cell* **32**, 190–197 (2008).
12. Fei, J., Kosuri, P., MacDougall, D. & Gonzalez, R. Coupling of ribosomal L1 stalk and tRNA dynamics during translation elongation. *Mol. Cell* **30**, 348–359 (2008).
13. Julián, P. *et al.* Structure of ratcheted ribosomes with tRNAs in hybrid states. *Proc. Natl. Acad. Sci. USA* **105**, 16924–16927 (2008).
14. Munro, J.B., Altman, R.B., O'Connor, N. & Blanchard, S.C. Identification of two distinct hybrid state intermediates on the ribosome. *Mol. Cell* **25**, 505–517 (2007).

15. Cornish, P.V. *et al.* Following movement of the L1 stalk between three functional states in single ribosomes. *Proc. Natl. Acad. Sci. USA* **106**, 2571–2576 (2009).
16. Trabuco, L.G. *et al.* The role of L1 stalk-tRNA interaction in the ribosome elongation cycle. *J. Mol. Biol.* **402**, 741–760 (2010).
17. Sanbonmatsu, K.Y., Joseph, S. & Tung, C.-S. Simulating movement of tRNA into the ribosome during decoding. *Proc. Natl. Acad. Sci. USA* **102**, 15854–15859 (2005).
18. Zhang, W., Dunkle, J.A. & Cate, J.H.D. Structures of the ribosome in intermediate states of ratcheting. *Science* **325**, 1014–1017 (2009).
19. Caulfield, T.R., Devkota, B. & Rollins, G.C. Examinations of tRNA range of motion using simulations of cryo-EM microscopy and X-ray data. *J. Biophys.* **2011**, 219515 (2011).
20. Tama, F., Miyashita, O. & Brooks, C.L. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J. Mol. Biol.* **337**, 985–999 (2004).
21. Whitford, P.C. *et al.* An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins* **75**, 430–441 (2009).
22. Tan, R.K.Z., Devkota, B. & Harvey, S.C. YUP.SCX: coaxing atomic models into medium resolution electron density maps. *J. Struct. Biol.* **163**, 163–174 (2008).
23. Schröder, G.F., Brunger, A. & Levitt, M. Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. *Structure* **15**, 1630–1641 (2007).
24. Ahmed, A. & Tama, F. Consensus among multiple approaches as a reliability measure for flexible fitting into cryo-EM data. *J. Struct. Biol.* **182**, 67–77 (2013).
25. Orzechowski, M. & Tama, F. Flexible fitting of high-resolution X-ray structures into cryoelectron microscopy maps using biased molecular dynamics simulations. *Biophys. J.* **95**, 5692–5705 (2008).
26. Dunkle, J.A. *et al.* Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* **332**, 981–984 (2011).
27. de Groot, B.L., Engel, A. & Grubmuller, H. The structure of the aquaporin-1 water channel: a comparison between cryo-electron microscopy and X-ray crystallography. *J. Mol. Biol.* **325**, 485–493 (2003).
28. Schuwirth, B.S. *et al.* Structures of the bacterial ribosome at 3.5 A resolution. *Science* **310**, 827–834 (2005).
29. Cornish, P.V., Ermolenko, D.N., Noller, H.F. & Ha, T. Spontaneous intersubunit rotation in single ribosomes. *Mol. Cell* **30**, 578–588 (2008).
30. Rodnina, M.V., Savelsbergh, A., Katunin, V.I. & Wintermeyer, W. Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. *Nature* **385**, 37–41 (1997).
31. Pan, D., Kirillov, S.V. & Cooperman, B.S. Kinetically competent intermediates in the translocation step of protein synthesis. *Mol. Cell* **25**, 519–529 (2007).
32. Fei, J., Richard, A.C., Bronson, J.E. & Gonzalez, R.L. Transfer RNA–mediated regulation of ribosome dynamics during protein synthesis. *Nat. Struct. Mol. Biol.* **18**, 1043–1051 (2011).
33. Munro, J.B., Sanbonmatsu, K.Y., Spahn, C.M. & Blanchard, S.C. Navigating the ribosome's metastable energy landscape. *Trends Biochem. Sci.* **34**, 390–400 (2009).
34. Chen, C. *et al.* Single-molecule fluorescence measurements of ribosomal translocation dynamics. *Mol. Cell* **42**, 367–377 (2011).
35. Blanchard, S.C., Kim, H.D., Gonzalez, R.L. Jr., Puglisi, J.D. & Chu, S. tRNA dynamics on the ribosome during translation. *Proc. Natl. Acad. Sci. USA* **101**, 12893–12898 (2004).
36. Rhodin, M.H.J. & Dinman, J.D. An extensive network of information flow through the B1b/c intersubunit bridge of the yeast ribosome. *PLoS ONE* **6**, e20048 (2011).
37. Subramanian, A.R. & Dabbs, E.R. Functional studies on ribosomes lacking protein L1 from mutant *Escherichia coli. Eur. J. Biochem.* **112**, 425–430 (1980).
38. Pan, D., Kirillov, S., Zhang, C.M., Hou, Y.M. & Cooperman, B.S. Rapid ribosomal translocation depends on the conserved 18–55 base pair in P-site transfer RNA. *Nat. Struct. Mol. Biol.* **13**, 354–359 (2006).
39. Rhodin, M.H.J. & Dinman, J.D. A flexible loop in yeast ribosomal protein L11 coordinates P-site tRNA binding. *Nucleic Acids Res.* **38**, 8377–8389 (2010).
40. Frauenfelder, H., Sligar, S.G. & Wolynes, P.G. The energy landscapes and motions of proteins. *Science* **254**, 1598–1603 (1991).
41. Yusupov, M.M. *et al.* Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883–896 (2001).
42. Gao, Y.G. *et al.* The structure of the ribosome with elongation factor F trapped in the posttranslocational state. *Science* **326**, 694–699 (2009).
43. Budkevich, T. *et al.* Structure and dynamics of the mammalian ribosomal pretranslocation complex. *Mol. Cell* **44**, 214–224 (2011).

ONLINE METHODS

The ribosome model was based on a crystal structure by Zhang *et al.* (PDB 3I1P and 3I1O)[18]; the crystal structure by Gao *et al.* (PDB 2WRI)[42] was used to model the L1 stalk. To match the cryo-EM setup[10], tRNA structures[42,44] were incorporated into the model. Models were then refined against 13 cryo-EM maps[10] with the real-space refinement program DireX[23]. Explicit-solvent all-atom molecular dynamics simulations were carried out with GROMACS[45] using the amber99sb force field[46]. R.m.s. deviations of refined models relative to crystal structures of the ground-state and rotated structure[26] were calculated after rigid-body fitting, using (i) all 70S Cα and P atoms, except L9 protein atoms, and (ii) all atoms within a 2-nm distance from the two tRNAs. To obtain reaction coordinates for the tRNAs and the L1 stalk, principal component analysis (PCA) was performed on the trajectories. Axes and rotation angles of 30S head and body rotations were obtained from each frame of the trajectories by nonlinear least-squares fitting and then averaged. Initial free energy–barrier heights between different states were estimated with multidimensional transition-state theory, on the basis of atomic fluctuations obtained from the simulations. All barrier heights and transition-rate prefactors were calibrated with barrier crossings that were fast enough to be observed in the simulations. Interaction enthalpies were obtained directly from the force field. The potential of mean force between the L1 stalk and the tRNA was obtained by umbrella-sampling simulations.

Vacant *E. coli* ribosomes were prepared for cryo-EM at 18 °C and imaged with a Titan Krios electron microscope (FEI Company) on a 4,000 × 4,000 CCD camera (FEI company) using two-fold pixel binning (3.2 Å per pixel). The resulting 9,814 ribosome particles and 315,108 pretranslocation-state *E. coli* ribosome particles from an existing cryo-EM data set[10] were analyzed and classified according to 30S body rotation as described. A detailed description of experimental and computational methods is provided in **Supplementary Note 1**.

44. Grishaev, A., Ying, J., Canny, M., Pardi, A. & Bax, A. Solution structure of tRNA Val from refinement of homology model against residual dipolar coupling and SAXS data. *J. Biomol. NMR* **42**, 99–109 (2008).
45. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).
46. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712–725 (2006).

Supplementary Information

Energy barriers and driving forces of tRNA translocation through the ribosome

Lars V. Bock, **Christian Blau**[1], Gunnar F. Schröder, Iakov I. Davydov, Niels Fischer, Holger Stark, Marina V. Rodnina, Andrea C. Vaiana, Helmut Grubmüller

Contents

---

[1]parts where I made a substantial contribution are marked *, parts where I predominantly contributed are marked **

Supplementary Note 3 (Results)

Supplementary Figure 1: All-atom models of pre1a–post4 states obtained from refinement of atomic models against cryo-EM maps[1]. For each state, the refined structure and an isosurface of the cryo-EM map (grey surface) are shown. The ribosomal subunits (50S and 30S) are shown in ribbon representation; tRNA$^{\text{fMet}}$ and tRNA$^{\text{Val}}$ atoms are depicted by magenta and green spheres, respectively.
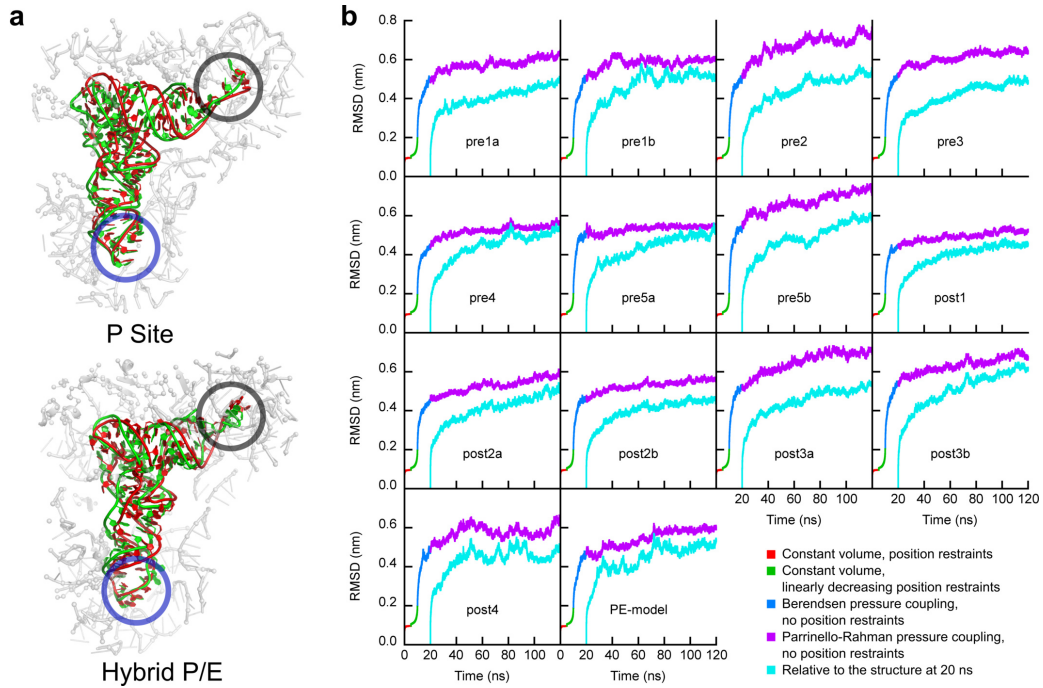
Supplementary Figure 2: Validation of models. (**a**) Comparison of tRNA positions between models and crystal structures in the P-site and P/E hybrid state. The tRNAs from crystal structures[2] and our models (left: pre1a, right: pre4) are shown as red and green ribbons, respectively, after rigid-body fitting of the binding region only (grey ribbons). C^{ff} and P atoms used for fitting are depicted as grey spheres and CCA-tail and acceptor stem regions are indicated by black and blue circles, respectively. (**b**) Structural deviations during the simulations. For each ribosome simulation, started either from the model refined against the cryo-EM map or from the PE-model, the RMSD relative to the starting structure is shown for the different simulation steps (red, green, blue, and magenta curves), and relative to the structure at 20 ns (cyan curve).
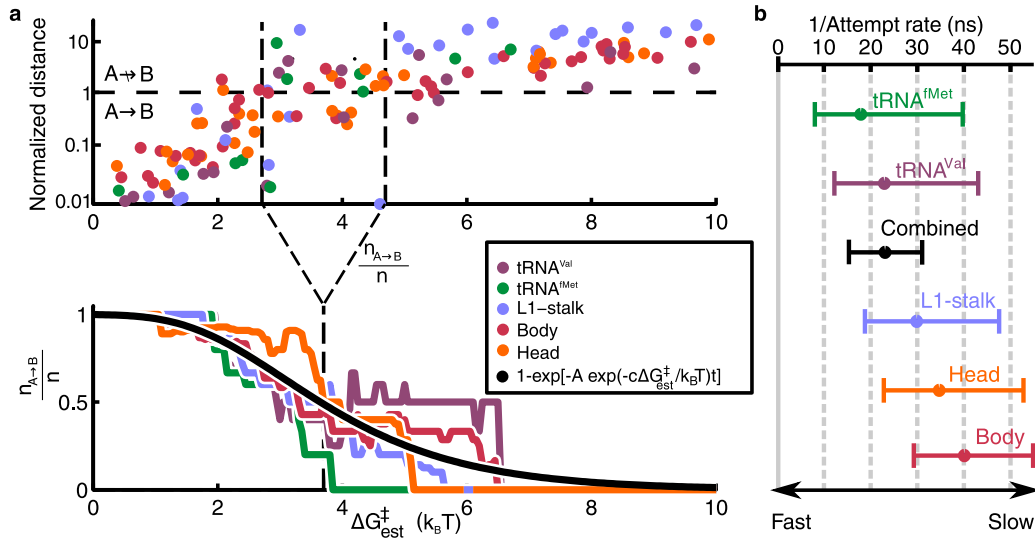
Supplementary Figure 3: Estimation of transition rates. (**a**) Attempt rate and free energy calibration factor. The upper panel shows an excerpt of the normalized distance between the ensembles for each pair of states versus the uncalibrated free energy estimate. This is done for each of the ribosome components (colored circles). A barrier between two states is considered crossed if this distance is smaller than one. The lower panel shows the frequency of barrier crossings $p_{A \to B}^{sim} = (n_{A \to B})/n$ calculated for free energy intervals of $1 \, k_b T$ (colored lines). The probability of barrier crossing $p_{A \to B}$ fitted to $p_{A \to B}^{sim}$ is shown as a black line. (**b**) Statistical uncertainty of the attempt rate of the movement of individual ribosome components. Shown are the medium value of the distribution of the attempt rates A (circles) and standard deviation (bars). The overall attempt rate is shown as reference (black line).

3
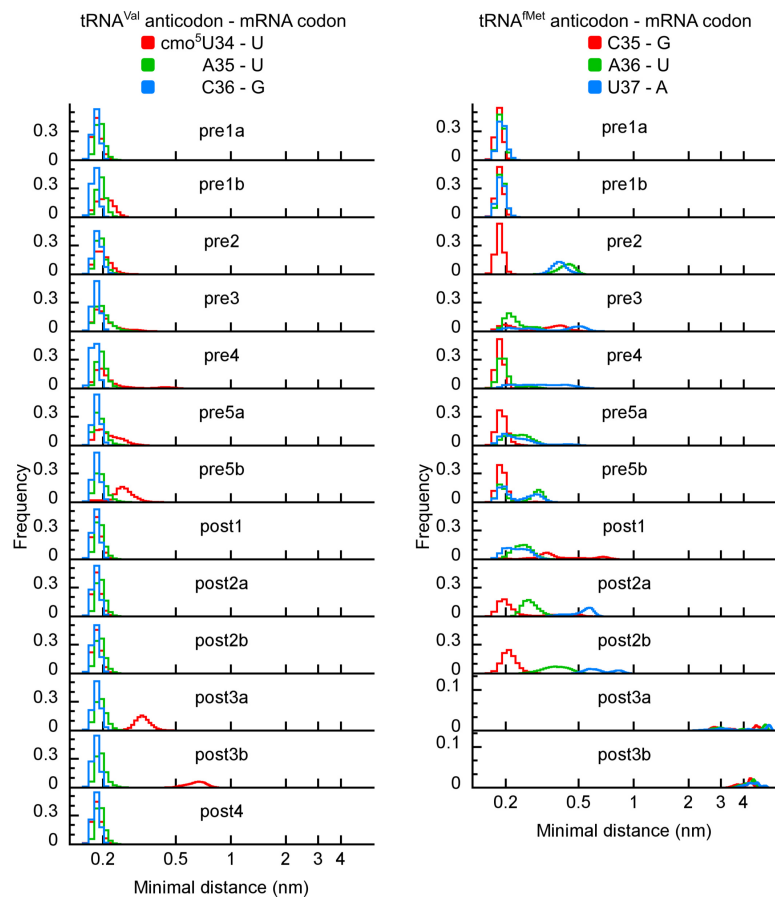
Supplementary Figure 4: Quality of tRNA-mRNA base-paring. For each state, histograms of the distances between codon residues of the mRNA and the corresponding anticodon residues of the two tRNAs are shown.
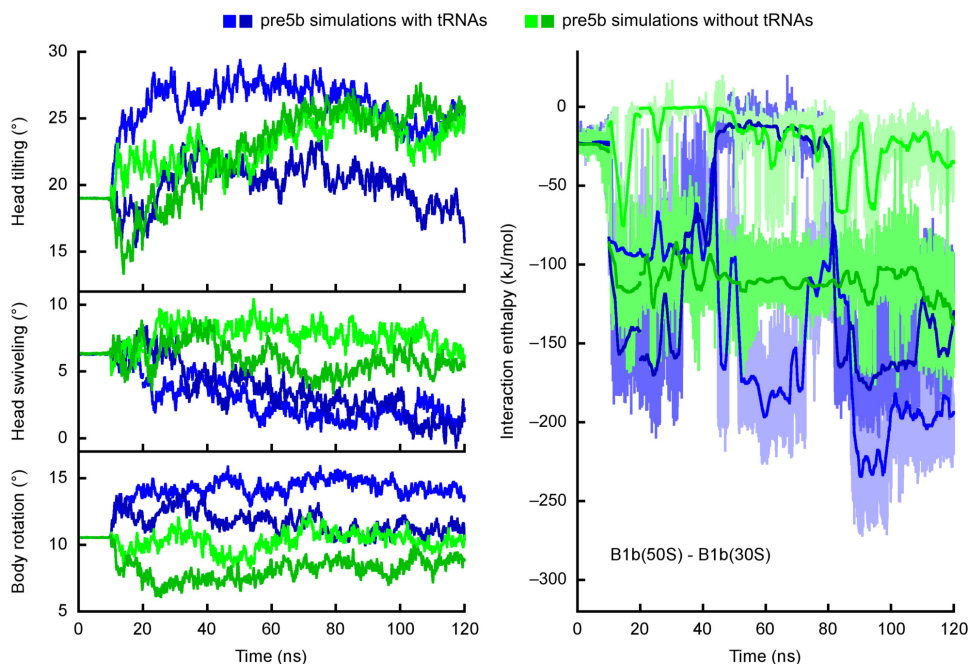
Supplementary Figure 5: Fast relaxation motions of the ribosome after tRNA removal during the simulations. Shown are time-traces of 30S head tilting, head swiveling, and body rotation angles (left panel), as well as of interaction enthalpies (right panel) for intersubunit bridge B1b, derived from four independent simulations. Blue curves refer to the two simulations started from the refined structure of the pre5b state with bound tRNAs, the green ones refer to simulations started from the same structure after removal of the tRNAs.
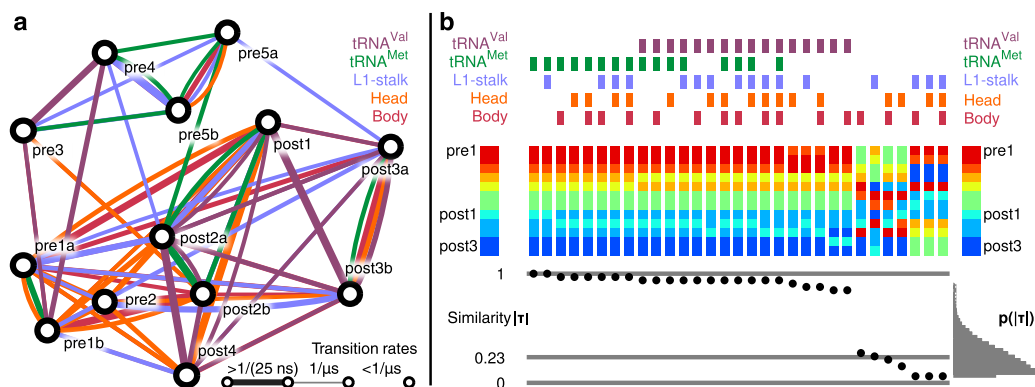
Supplementary Figure 6: Transiton Rates. (**a**) Schematic representation of the translocation intermediate states as a Markov model. Circles denote states, connecting lines encode the transition time estimates for L1-stalk, tRNA$^{\text{fMet}}$, tRNA$^{\text{Val}}$ motion as well as body and head rotation. We thank Benoit Roux for providing the idea. (**b**) Fastest progression sequences of translocation intermediate states ranked according to similarity to the sequence proposed by Fischer et al.[1] For all 31 possible combinations of ribosome components (top, color scheme as in Fig. 1a,d), the fastest progression sequence was determined as in 2.10. The similarity of each of the identified sequences (mid, columns) to the sequence given by Fischer et al.[1] was described using the absolute Kendall rank correlation coefficient τ (bottom). As a reference the mean τ value for random sequences (0.23) and their probability distribution $p(\tau)$ is shown.

Supplementary Table 1: Sequence of structure refinements against cryo-EM maps. For each state the starting model which was used for refinement against the corresponding map is shown. The pre5c structure was not used for simulations.

AP-states		PE-states		P-state	
state of refined structure	starting model	state of refined structure	starting model	state of refined structure	starting model
pre1b	AP-model	post1	PE-model	post4	P-model
pre1a	pre1b	post2a	post1		
pre2	pre1b	post2b	post2a		
pre4	pre1b	post3b	post2a		
pre3	pre4	post3a	post3b		
pre5a	pre1b				
pre5c	pre5a				
pre5b	pre5c				

5

Supplementary Table 2: Stereochemical parameters of our models compared to those of crystal structures. Shown are deviation of the distributions of key stereochemical parameters from distributions found in the protein data bank (PDB) for the 50S and 30S subunits. The upper part of each table shows deviations from the mean values in standard deviations σ. The relative width of the distribution compared to the standard distribution is shown in the lower part.

(a) 30S

PDB id	Models(pre1a-post4) 100ns	refined	Dunkle et al.[2] 3R8N	3R8O	Zhang et al.[3] 3I1Q	3I1Z
Ramachandran	-3.12	-7.62	-7.44	-7.52	-6.73	-6.96
χ_1-χ_2-rotamers	-2.04	-5.97	-6.97	-7.10	-6.26	-6.25
Backbone	-2.23	-4.66	-3.27	-3.47	-2.41	-2.39
Bond length	1.40	1.57	0.85	0.85	0.59	0.57
Bond angles	1.91	2.24	1.20	1.22	1.20	1.15
Omega angles	1.15	2.17	1.62	1.66	0.93	0.87
Side chain planarity	0.96	2.30	0.39	0.38	0.23	0.21
Improper dihedrals	1.04	1.90	0.75	0.72	0.53	0.46
Inside-Outside dist.	0.99	1.03	1.01	1.02	1.02	1.03

(b) 50S

PDB id	Models(pre1a-post4) 100ns	refined	Dunkle et al.[2] 3R8S	3R8T	Zhang et al.[3] 3I1R	3I20
Ramachandran	-3.05	-7.67	-5.75	-6.52	-6.01	-7.15
χ_1-χ_2-rotamers	-2.00	-5.95	-5.89	-6.23	-6.05	-5.78
Backbone	-1.98	-5.13	-1.92	-2.04	-2.78	-3.64
Bond length	1.33	1.48	1.49	1.04	0.77	0.61
Bond angles	1.98	2.34	1.53	1.38	1.40	1.21
Omega angles	1.14	2.22	1.52	1.47	1.20	0.87
Side chain planarity	1.00	2.32	0.50	0.41	0.31	0.20
Improper dihedrals	1.05	1.95	0.92	0.79	0.72	0.44
Inside-Outside dist.	0.99	1.02	0.99	0.99	1.00	1.03

Supplementary Table 3: Accession codes for models of spontaneous tRNA translocation. Shown are the EM-DataBank accession codes for the cryo-EM densities used for refinement and the resolution of the respective cryo-EM maps. PDB-id codes are given for the 30S and 50S ribosomal subunit.

state	EMDB id	resolution [Å]	PDB id 30S	PDB id 50S
pre1a	1716	12	3J4V	3J52
pre1b	2472	12	3J4W	3J4X
pre2	1717	20	3J4Z	3J50
pre3	1718	17	3J4Y	3J51
pre4	1719	13	3J53	3J54
pre5a	2473	15	3J55	3J56
pre5b	1720	17	3J57	3J58
post1	1721	12	3J59	3J5A
post2a	1722	17	3J5B	3J5C
post2b	2474	17	3J5D	3J5E
post3a	2475	20	3J5F	3J5G
post3b	1723	15	3J5H	3J5J
post4	1724	9	3J5J	3J5K

Supplementary Table 4a: Contacting residues between tRNA^Val and 23S.

tRNA^Val residue	23S residue	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4
		pre							post					
C56	H38 A896													
G57	H38 A896													
Ψ55	H38 A896													
U17	H38 G882													
G19	H38 A896													
U17	H38 G881													
G18	H38 G882													
G19	H38 G882													
G19	H38 G883													
C56	H38 U895													
C56	H38 U894													
G20	H38 G883													
C56	H38 G882													
G18	H38 C897													
U17	H38 C898													
G18	H38 A896													
U17	H38 C897													
U17	H38 G880													
C56	H38 C897													
C56	H38 G881													
G20	H38 G882													
G19	H38 G881													
C56	H38 G880													
G57	H38 G880													
G57	H38 G881													
C25	H69 C1914													
U12	H69 U1915													
A38	H69 A1913													
G24	H69 C1914													
C11	H69 U1915													
U12	H69 A1916													
h⁶m¹A37	H69 A1913													
C25	H69 A1913													
G39	H69 A1913													
C11	H69 C1914													
A26	H69 C1914													
G24	H69 U1915													
G10	H69 C1914													
G10	H69 U1915													
C27	H69 A1913													
C36	H69 A1913													
C27	H69 C1914													
U12	H69 G1910													
C13	H69 G1910													
C13	H69 C1909													
U12	H69 C1924													
U12	H69 U1923													
C11	H69 U1923													
C13	H69 C1924													
A69	H69 C1908													
A69	H69 G1907													
C70	H69 G1907													
C11	H69 C1909													
C72	H71 C1942													
A73	H71 U1943													
C71	H71 C1942													
C74	H71 U1943													
C74	H71 U1944													
C71	H71 C1941													
C72	H71 A1966													
C71	H71 A1966													
C71	H71 C1965													
V77	H74 A2451													
fMet78	H74 C2063													
fMet78	H74 A2439													
fMet78	H74 A2062													
fMet78	H74 A2451													
V77	H74 C2064													
V77	H74 A2450													
fMet78	H74 A2450													
fMet78	H74 C2064													
fMet78	H74 G2061													
C75	H74 A2451													
A76	H74 A2450													
A76	H74 C2064													
A76	H74 C2065													
C75	H74 A2450													
A76	H74 A2451													
V77	H74 C2063													
C75	H74 C2065													
A76	H74 C2063													
V77	H74 A2439													
fMet78	H74 U2441													
C75	H74 C2064													
V77	H74 A2062													
A76	H80 G2252													
A76	H80 G2253													
C75	H80 G2252													
A76	H80 C2254													
C75	H80 G2251													
A76	H80 G2251													
C74	H80 G2253													
C75	H80 G2253													
G3	H80 C2254													
G3	H80 G2255													
G2	H80 C2254													
C74	H80 G2252													
U4	H80 G2255													

tRNA^Val residue	23S residue	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4
		pre							post					
A73	H80 G2251													
C72	H80 G2253													
A73	H80 G2252													
A73	H80 G2253													
G2	H80 G2255													
C74	H80 C2254													
G63	H81 G2279													
C56	H84 A2309													
C56	H84 G2308													
G19	H84 G2308													
G19	H84 A2309													
C56	H84 C2310													
G57	H84 A2309													
G57	H84 C2310													
A58	H84 A2309													
m⁵U54	H85 A2327													
Ψ55	H85 C2326													
G52	H89 A2469													
U64	H89 C2483													
G52	H89 G2470													
U64	H89 A2482													
C51	H89 G2470													
G63	H89 C2483													
A76	H89 C2452													
G63	H89 A2482													
G52	H89 A2482													
G1	H89 C2462													
G63	H89 G2484													
fMet78	H89 C2452													
C51	H89 A2469													
V77	H89 C2452													
G53	H89 A2469													
V77	H89 U2506													
fMet78	H89 U2506													
A76	H89 U2506													
fMet78	H89 G2505													
C75	H89 C2452													
C74	H89 G2494													
C75	H90 C2507													
C74	H90 C2573													
A76	H90 C2507													
A76	H90 C2573													
C75	H92 G2553													
C74	H92 U2555													
C75	H92 U2554													
A73	H92 C2556													
C75	H92 U2555													
C74	H92 C2556													
A76	H92 G2553													
A76	H93 U2584													
A76	H93 A2602													
fMet78	H93 U2585													
A76	H93 G2583													
A76	H93 U2585													
C75	H93 G2583													
C75	H93 U2584													
V77	H93 U2585													
fMet78	H93 U2586													
C74	H93 A2602													
A73	H93 A2602													
fMet78	H93 U2584													
C75	H93 A2602													
A76	H93 U2604													
V77	H93 A2602													
V77	H93 U2584													
A76	H93 C2601													
A76	H93 A2600													
fMet78	H93 G2583													
A76	H93 U2586													
C74	H93 A2600													
C74	H93 C2601													
C72	H93 C2594													

Supplementary Table 4b: Contacting residues between tRNA^Val and L5.

tRNA^Val residue	L5 residue	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4
		pre							post					
C56	A74													
C56	V73													
C56	G75													
Ψ55	A74													
C56	S72													
G20	K77													
C56	R79													
G19	A74													
G57	A74													
G19	R79													
G19	K77													
G57	V73													
G57	K77													
A58	K77													
C56	I78													
C56	K77													
C56	Q80													
G19	G75													
G19	V73													
G20	V73													
G19	F76													
G57	G75													
C56	K71													
G57	S72													
G20	F76													

Supplementary Table 4c: Contacting residues between tRNA^Val and L16.

tRNA^{Val} residue	L16 residue	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4
				pre							post			
G53	R51													
m^5U54	R51													
G52	R55													
G53	R55													
G63	M1													
C62	M1													
m^5U54	R50													
C62	R44													
G63	R51													
m^5U54	T54													
G53	T54													
C62	R51													
G63	R44													
m^5U54	R59													
G53	R59													
U64	R44													
G53	R50													
A73	R81													
G1	N88													
G1	P77													
G1	R81													
G52	M1													
G63	Q3													
C62	R6													
m^5U54	K5													
G63	R10													
C72	R81													
A73	V80													
C74	V80													
U64	R10													
U64	Q3													
C75	R81													
G53	K5													
G63	R6													
G1	K76													
G2	N88													
G1	G87													
C51	M1													
C74	R81													
G1	E90													
G1	T74													
C65	Q3													
C65	R10													
G1	E75													
G2	G87													
A73	P77													
G52	R6													
G52	K5													
G53	R6													
C65	M1													
C51	K5													
A66	M1													
A66	Q3													
G2	K84													
G3	K84													
U64	R6													
U64	N88													
G53	K8													
G63	K8													
C65	R6													
G1	L78													
G52	K8													
m^5U54	K8													
C62	K8													
G1	G85													
G1	K86													
G2	K86													
G1	K84													
G1	V80													
G2	R81													
C65	K86													
tRNA^{Val}	L16	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4

Supplementary Table 4d: Contacting residues between tRNA^{Val} and L27.

tRNA^{Val} residue	L27 residue	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4
G3	ACE5													
tRNA^{Val}	L27	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4

Supplementary Table 4e: Contacting residues between tRNA^{Val} and L33.

tRNA^{Val} residue	L33 residue	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4
U17	ACE2													
U17	I4													
U17	G3													
U17	R27													
G19	R27													
C56	ACE2													
C56	R27													
tRNA^{Val}	L33	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4

Supplementary Table 4f: Contacting residues between tRNA^{Val} and 16S.

tRNA^{Val} residue	16S residue	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4
A35	h18 G530													
cmo^5U34	h18 G530													
C36	h18 G530													
C36	h18 C518													
A35	h18 C518													
A38	h24 A790													
G39	h24 A790													
h^6m^1A37	h24 A790													
tRNA^{Val}	16S	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4

---

continued from last page

tRNA^{Val} residue	16S residue	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4
C31	h29 A1339													
G40	h29 A1339													
A41	h29 G1338													
C30	h29 A1339													
A41	h29 A1339													
C31	h29 A1340													
C32	h29 A1340													
C32	h29 U1341													
G39	h29 A1339													
G42	h29 G1338													
G40	h29 G1338													
C30	h29 G1338													
U29	h30 A1229													
C30	h30 C1230													
C30	h30 A1229													
C31	h30 C1230													
U29	h30 C1230													
C30	h30 G1231													
U29	h30 C1228													
cmo^5U34	h31 G966													
cmo^5U34	h34 C1054													
cmo^5U34	h34 A1196													
cmo^5U34	h34 A1197													
h^6m^1A37	h44 A1493													
A38	h44 A1493													
A35	h44 A1493													
C36	h44 A1493													
A38	h44 G1494													
cmo^5U34	h44 C1400													
tRNA^{Val}	16S	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4

Supplementary Table 4g: Contacting residues between tRNA^{Val} and S9.

tRNA^{Val} residue	S9 residue	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4
cmo^5U34	R129													
C32	K128													
C31	K128													
C31	R129													
U33	K128													
cmo^5U34	K128													
A35	K128													
tRNA^{Val}	S9	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4

Supplementary Table 4h: Contacting residues between tRNA^{Val} and S13.

tRNA^{Val} residue	S13 residue	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4
G44	K113													
G44	NH$_2$114													
G45	K113													
G43	K113													
G42	K113													
G44	P111													
G42	NH$_2$114													
G43	R112													
C28	K113													
G43	P111													
G43	NH$_2$114													
tRNA^{Val}	S13	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4

Supplementary Table 4i: Contacting residues between tRNA^{fMet} and 23S.

tRNA^{fMet} residue	23S residue	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4
A76	H13 C249													
G52	H38 G880													
G52	H38 G881													
G4	H68 G1850													
G4	H68 U1851													
C71	H68 C1893													
G70	H68 C1893													
C3	H68 U1851													
C71	H68 C1892													
C71	H68 U1851													
G70	H68 U1851													
C71	H68 G1852													
C71	H68 G1850													
A72	H68 G1891													
G70	H68 G1850													
G70	H68 C1892													
G4	H68 U1852													
G5	H68 U1851													
A76	H68 C1870													
A76	H68 A1871													
C75	H68 A1871													
C13	H69 C1924													
G12	H69 U1923													
G12	H69 C1924													
G12	H69 C1909													
C13	H69 C1909													
G12	H69 G1910													
U24	H69 U1923													
A11	H69 G1910													
U24	H69 G1922													
C25	H69 G1922													
A14	H69 C1925													
C13	H69 C1925													
C13	H69 G1907													
tRNA^{fMet}	23S	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4

tRNA^fMet residue	23S residue	pre 1a	1b	2	3	4	5a	5b	post 1	2a	2b	3a	3b	4
U36	H69 G1921													
U36	H69 G1922													
A37	H69 G1922													
A37	H69 U1923													
A76	H74 A2450													
C75	H74 C2064													
A76	H74 A2451													
A76	H74 C2064													
C75	H74 A2451													
A76	H74 C2065													
C75	H74 A2432													
A76	H74 A2432													
C75	H74 A2433													
C74	H74 A2433													
C74	H74 A2432													
U17A	H76 G2112													
G19	H76 G2112													
G19	H76 U2113													
G18	H76 G2112													
U17A	H76 U2111													
D20	H76 U2113													
D20	H76 G2112													
U17A	H76 U2180													
A21	H76 G2112													
U17A	H76 U2110													
A59	H76 G2112													
C17	H76 U2180													
C17	H76 U2181													
C56	H77 A2169													
G19	H77 A2169													
C56	H77 A2170													
C56	H77 U2122													
Ψ55	H77 A2169													
U17A	H78 C2145													
C3	H80 G2255													
C75	H80 G2251													
G2	H80 G2255													
G2	H80 C2254													
C1	H80 C2254													
C1	H80 G2253													
A76	H80 G2251													
C75	H80 G2252													
C3	H80 C2254													
C74	H80 G2252													
C74	H80 G2253													
C1	H80 G2255													
A76	H80 G2252													
C75	H80 G2253													
A76	H80 G2253													
C3	H80 G2256													
C74	H80 C2254													
C51	H84 A2309													
G52	H84 C2310													
G52	H84 A2309													
A76	H88 C2395													
A76	H88 C2394													
A76	H88 C2422													
C74	H88 C2422													
A76	H88 G2421													
C75	H88 C2422													
C75	H88 G2421													
A76	H88 G2396													
A76	H88 G2397													
C75	H88 G2397													
C75	H88 G2396													
C1	H88 G2397													
A76	H88 C2424													
A76	H88 U2423													
C74	H93 C2594													
A76	H93 A2602													
C74	H93 C2600													
C74	H93 A2602													
C75	H93 A2602													
A76	H93 C2601													
tRNA^fMet	23S	pre							post					

Supplementary Table 4j: Contacting residues between tRNA^fMet and L1.

tRNA^fMet residue	L1 residue	pre 1a	1b	2	3	4	5a	5b	post 1	2a	2b	3a	3b	4
Ψ55	R60													
m5U54	R60													
C62	K54													
G63	K54													
C62	R53													
C62	S55													
G63	R53													
C56	Q129													
Ψ56	R164													
C56	P133													
C56	G132													
G53	S55													
G63	D56													
G52	S55													
G63	S55													
G64	K54													
m5U54	R164													
m5U54	K141													
C56	G128													
G64	Q203													
m5U54	N58													
G4	R53													
C56	R164													
G53	K141													
m5U54	S55													
m5U54	N139													
C62	D56													
C61	R53													
tRNA^fMet	L1	pre							post					

tRNA^fMet residue	L1 residue	pre 1a	1b	2	3	4	5a	5b	post 1	2a	2b	3a	3b	4
G18	K167													
A58	R53													
G18	R53													
G53	R53													
A57	K167													
C56	R134													
Ψ55	P133													
C62	D51													
G53	D56													
m5U54	D56													
Ψ55	R53													
m5U54	R53													
A72	R122													
A73	M121													
A72	G125													
A72	Q126													
A73	R122													
A76	K141													
C75	V123													
A76	E98													
C74	Q80													
C74	G81													
C75	Q80													
C75	R122													
C75	Q126													
A76	Q80													
A76	M97													
A76	K105													
A76	Q126													
tRNA^fMet	L1	pre							post					

Supplementary Table 4k: Contacting residues between tRNA^fMet and L5.

tRNA^fMet residue	L5 residue	pre 1a	1b	2	3	4	5a	5b	post 1	2a	2b	3a	3b	4
C56	V73													
G19	G75													
C56	S72													
G19	K77													
C56	A74													
G18	K77													
C56	G75													
C56	R79													
A57	V73													
A57	A74													
Ψ55	V73													
C56	K77													
A57	G75													
A57	K77													
C56	I78													
G53	S72													
G53	R79													
m5U54	R79													
Ψ55	R79													
G52	V73													
G52	A74													
G52	S72													
A59	K77													
G30	K46													
G42	K47													
G29	K47													
G31	K46													
C34	R79													
A35	R79													
A43	K47													
G30	A44													
G31	I43													
G31	A44													
G31	Y82													
Cm32	K77													
C41	D45													
C41	K47													
G30	Y82													
U33	K77													
G30	I43													
G30	R79													
G31	K77													
G31	R79													
tRNA^fMet	L5	pre							post					

Supplementary Table 4l: Contacting residues between tRNA^fMet and L16.

tRNA^fMet residue	L16 residue	pre 1a	1b	2	3	4	5a	5b	post 1	2a	2b	3a	3b	4
C65	R6													
C75	R81													
G64	R6													
A76	R81													
C51	R6													
G64	R10													
C1	K76													
C1	V80													
C1	G85													
G52	R6													
C1	N88													
C1	P77													
C1	K84													
C1	K86													
C1	G87													
C74	K84													
C1	R6													
C66	R6													
G2	R6													
tRNA^fMet	L16	pre							post					

tRNA^fMet residue	L16 residue	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4
				pre							post			

Supplementary Table 4m: Contacting residues between tRNA^fMet and L27.

tRNA^fMet residue	L27 residue	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4
				pre							post			
G2	G6													
G2	ACE5													
C1	ACE5													
tRNA^fMet	L27	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4
				pre							post			

Supplementary Table 4n: Contacting residues between tRNA^fMet and L28.

tRNA^fMet residue	L28 residue	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4
				pre							post			
C75	H19													
tRNA^fMet	L28	1a	1b	2	3	4	5a	5b	1	2a	2b	3a	3b	4
				pre							post			

Supplementary Table 4o: Contacting residues between tRNA^fMet and L33.

tRNA^fMet residue	L33 residue
Ψ55	R27
Ψ55	T28
C56	R27
C56	T28
C65	R27
C65	K26
C66	K26
C51	K29
G52	K29
C66	R27
G52	P30
G64	R27
C65	I4
C66	I4
C1	I4
C1	E6
C74	G3
A73	ACE2
A73	G3
C74	ACE2
C74	I4
C67	R27
G64	K52
C65	K52
A37	P30
A38	P30
U36	P30
G29	K29
U36	T28
U36	K29
C28	K29
C34	R27
A35	R27
A35	T28

Column headers: 1a 1b 2 3 4 5a 5b (pre) | 1 2a 2b 3a 3b 4 (post)

Supplementary Table 4p: Contacting residues between tRNA^fMet and 16S.

tRNA^fMet residue	16S residue
A37	h23 G693
A38	h23 A694
A37	h23 A694
A38	h23 A695
U36	h23 G693
C39	h23 A694
A38	h23 G693
C39	h23 G693
C39	h24 A790
A38	h24 A790
A37	h24 A790
U36	h24 A790
U36	h24 U789
A37	h24 U789
U33	h28 C1383
C34	h28 C1383
C41	h29 A1339
G30	h29 A1339
G31	h29 A1340
C40	h29 A1339
Cm32	h29 U1341
G31	h29 A1339
C41	h29 G1338
Cm32	h29 A1340
G29	h29 G1338
G42	h29 G1338
G31	h29 U1341
G42	h29 A1339
A43	h29 G1338
G29	h29 A1339
G30	h29 A1340
C41	h29 A1340
G30	h29 G1338
U36	h29 A1339
G29	h30 A1229
G30	h30 C1230
G42	h30 A1229
A44	h30 C1228
tRNA^fMet	16S

tRNA^fMet residue	16S residue
	(continued from last page)
G30	h30 A1229
G29	h30 C1230
A43	h30 C1228
A43	h30 A1229
G31	h30 C1230
G29	h30 C1228
C28	h30 C1228
C34	h31 G966
A35	h31 G966
C34	h44 C1400
tRNA^fMet	16S

Column headers: 1a 1b 2 3 4 5a 5b (pre) | 1 2a 2b 3a 3b 4 (post)

Supplementary Table 4q: Contacting residues between tRNA^fMet and S7.

tRNA^fMet residue	S7 residue
Cm32	Q85
U33	S82
U33	T83
A38	T83
C40	A146
C41	R142
G42	R142
Cm32	T83
Cm32	Y84
C40	M143
C41	K135
G29	K135
U33	Q85
C34	R78
C40	N147
C41	A146
Cm32	R78
U33	R78
tRNA^fMet	S7

Column headers: 1a 1b 2 3 4 5a 5b (pre) | 1 2a 2b 3a 3b 4 (post)

Supplementary Table 4r: Contacting residues between tRNA^fMet and S9.

tRNA^fMet residue	S9 residue
U33	R129
C34	R129
Cm32	R129
A35	R129
Cm32	K128
G31	R129
G30	R129
G31	K128
tRNA^fMet	S9

Column headers: 1a 1b 2 3 4 5a 5b (pre) | 1 2a 2b 3a 3b 4 (post)

Supplementary Table 4s: Contacting residues between tRNA^fMet and S11.

tRNA^fMet residue	S11 residue
A38	R52
C39	R55
C39	R52
tRNA^fMet	S11

Column headers: 1a 1b 2 3 4 5a 5b (pre) | 1 2a 2b 3a 3b 4 (post)

Supplementary Table 4t: Contacting residues between tRNA^fMet and S13.

tRNA^fMet residue	S13 residue
A44	K109
A43	K109
A44	R112
G45	R112
A44	P111
G29	R112
C28	K113
A44	K113
C28	R112
C28	NH2 114
A43	K113
A43	NH2 114
U27	K113
G42	NH2 114
A43	R112
G42	K113
U27	NH2 114
A44	NH2 114
G45	K113
G29	NH2 114
tRNA^fMet	S13

Column headers: 1a 1b 2 3 4 5a 5b (pre) | 1 2a 2b 3a 3b 4 (post)

Supplementary Note 1 (Methods)

1.1   General molecular dynamics setup

All molecular dynamics (MD) simulations were carried out with GROMACS 4[4] using the amber99sb force field[5], and the SPC/E water model[6]. Each simulated model was first solvated in a dodecahedron box keeping a minimum distance of 1.5 nm between the model atoms and the box boundaries. K^+ and Cl^- forcefield parameters were taken from Joung and Cheatham[7]. Long-range electrostatic interactions, beyond 0.9 nm were calculated by particle-mesh Ewald summation[8] with a grid spacing of 0.12 nm. Lennard-Jones qinteractions were calculated within a distance of 1.4 nm. Coordinates were recorded for analysis every 2 ps. Unless stated otherwise, an integration time step of 4 fs was used, applying virtual site constraints[9]. All bond lengths were constrained with the LINCS algorithm[10]. The system temperature was kept constant at $T = 300$ K using velocity rescaling[11] with a coupling time constant of $\tau_T = 0.1$ ps. Protonation states of amino acids were determined with WHATIF[12].

1.2   Models of the ribosome including tRNAs

Three initial atomic models of the *E.coli* ribosome were built, which were subsequently refined against the cryo-EM maps provided by Fischer et. al[1]: First, a model of the ribosome with a P-site fMetVal-tRNA^{Val} (P-model), second, a model with a P-site fMetVal-tRNA^{Val} and an E-site tRNA^{fMet} (PE-model), and third, a model with an A-site fMetVal-tRNA^{Val} and P-site tRNA^{fMet} (AP-model).

   All models were constructed from the crystal structure by Zhang et al.[3]. This was the best resolved (resolution: 3.19 Å) and most complete *E.coli* ribosome structure at the time of modeling. All structural information (pdb ids: 3I1P, 3I1O), including the crystallographic water molecules and ions, was used.

   For the L1 protein and the parts of the L1-stalk rRNA which are not resolved in the structure of Zhang et al., the *T.thermophilus* ribosome structure of Gao et al.[13] (pdb id: 2WRI) was used. A homology model of the L1 protein was built using the swissmodel server (swissmodel.expasy.org) with the *E.coli* sequence and the *T.thermophilus* structure as a template (44% sequence identity).

   In the *E.coli* structure, 68 nucleotides (2111–2179) are not resolved in the L1-stalk rRNA. These were also modeled using the *T.thermophilus* structure. The corresponding structurally aligned nucleotides[13] as well as ten nucleotides upstream and downstream, which form ten base pairs at the stem of the L1-stalk, were extracted from the *T.thermophilus* structure. The extracted nucleotides were mutated to match the *E.coli* sequence using WHATIF[12] (58% sequence identity).

   The whole L1-stalk, comprising the mutated rRNA and the homology model of the L1 protein, was energy minimized in vacuum with position restraints on the P and C^{ff} atoms (position restraints force constant: 1000 kJ mol^{-1} nm^{-2}). Next, water and K^+Cl^- ions at a concentration of 0.154 mol was added. This system was energy minimized and subsequently equilibrated for 1 ns, maintaining position restraints. After that, the stem base pairs were superimposed to the matching nucleotides in the *E.coli* structure. Finally, an MD simulation of the L1-stalk was carried out, restraining the positions of heavy atoms of the

stem to those of the target *E.coli* structure. The force constant of the restraining potential was linearly increased from 0 to 500 kJ mol^{-1} nm^{-2} within 5 ns. The resulting L1-stalk structure was finally incorporated into the model.

Helix 38 (H38) of the large ribosomal subunit forms contacts to the small subunit via intersubunit bridge B1a[14]. Nucleotides G879–C897 of this helix are not resolved in the *E.coli* structure. Secondary structure prediction software S2S[15] predicted a pentaloop for the tip of the helix. The nucleotides of a pentaloop from an NMR structure[16] (pdb id: 1NA2) were mutated to match H38 sequence using WHATIF. To fit this rRNA structural motif into the model of the ribosome, the same protocol as for the L1-stalk rRNA was used.

The ribosomes used for the cryo-EM experiments contained an fMetVal-tRNA^{Val}. Since no high resolution structure was available, a solution structure of tRNA^{Val} from a refinement of a homology model against residual dipolar coupling and SAXS data (pdb id: 2K4C)[17] was used for the models.

The *T.thermophilus* structure by Yusupova et al.[18] (pdb id: 2HGP) contains a P-site tRNA^{Phe} and a 50 nucleotide long mRNA. This structure was rigid-body fitted to our model using structurally aligned nucleotides[18] from 16S rRNA of the small subunit. The fitted coordinates of the tRNA phosphates and of the mRNA were stored for later use. From this mRNA structure, the A-, P-, and E-site codons as well as three upstream and downstream nucleotides were extracted. These nucleotides were mutated with WHATIF to match the sequence of the mRNA used in the cryo-EM experiments. Appropriate tRNA modifications and the dipeptide were added to match experimental conditions[1]. Atom types for fMet were obtained with ANTECHAMBER[19], partial charges were determined using DFT-b3lyp with a 6-31/G* basis set. Side chain charges are the same as in Met. On the backbone, only the charges of the formylamino cap changed more than 5 %. The modified fMetVal-tRNA^{Val} structure was then fitted as a rigid body to the P atoms obtained from the fitted Yusupova structure. Next, a 1 ns simulation of the tRNA and the mRNA in solvent with position restraints on the P and C1′ atoms was carried out. In the subsequent 5 ns simulation, positions of P and C1′ atoms were restrained to those of the fitted Yusupova structure, thereby linearly increasing the force constant from 0 to 1000 kJ mol^{-1} nm^{-2} during the simulation. The resulting tRNA structure was then included into the ribosome model, yielding the P-model.

The *T.thermophilus* structure by Gao et al.[13] (pdb id: 2WRI) contains an E-site tRNA^{fMet}. The 30S subunit structure was rigid body fitted to our model using structurally aligned nucleotides[13] from 16S rRNA. Nucleotide modifications were added to the tRNA to match the modifications of the tRNAs in the cryo-EM experiments and a 1 ns simulation of the tRNA in solvent with position restraints on the P and C1′ atoms was performed. The resulting tRNA structure was then included into the P-model, yielding the PE-model.

For the AP-model, the same tRNA structures were used as for the PE-model, but the tRNA^{Val} was fitted into the A- and the tRNA^{fMet} into the P-site. Almost the same protocol as for the addition of the P-site tRNA^{Val} was used, except that the tRNA^{Val} has 77 nucleotides and the tRNA^{Phe} from the Yusupova structure which was used for fitting has 76 nucleotides. All the nucleotides, except for 5 nucleotides upstream and 5 nucleotides downstream of the insertion, were used for the rigid body fitting and the position restraints in the simulation.

1.3   Refinement of the atomic models against cryo-EM maps*

The three ribosome models obtained from crystal structures as described in the previous paragraph were used as starting structures for subsequent refinement against the 13 different cryo-EM density maps, to provide an all-atom interpretation of each individual conformational state. Initial placement of a starting model into a density map was done using the rigid-body fit feature of the program Chimera[20]. The real-space refinement program DireX[21] was then used for all refinements. DireX computes a density map from an atomic model and refines the atomic coordinates to maximize the overlap between this model map and the cryo-EM density map. The quality of the refinement by DireX has been found to be similar or better than that of other established methods in a comparative study[22]. The model density maps were generated using a Gaussian kernel with a width adapted to the resolution of the corresponding cryo-EM density map.

For each refinement, 2000 steps were performed which took 36 hours on average on one core of an Intel Core 2 Quad Processor Q9300 (2.5 GHz).

The initial AP-, PE-, and P-models were refined against the map which most closely resembled the state of the model (respectively: pre1b, post1, and post4). Subsequently, fitted structures were used as starting models for refinement against the remaining 10 maps in the sequence described in Supplementary Table 1.

Due to a program bug during the refinement process with DireX, in the 30S subunit, the tRNAs and the mRNA, several amino acids and nucleotides had wrong chiralities. In the structures extracted after 20 ns pre-equilibration, the errors were corrected by placing the chiral center atom on the opposite side of the plane defined by the three chiral neighbor heavy atoms. To that goal, the bond vector between the chiral center atom and the respective bound hydrogen atom was used for shifting the chiral center and the bound hydrogen atom. Subsequently, the bound hydrogen atom was flipped to the other side of the chiral center atom using the same bond vector. Alternatively, for the C2' in nucleotide sugars, chiral errors were corrected by swapping the positions of the O2' and H2' atoms. All corrected structures were again energy-minimized. Since all the simulations were started from the structures containing these errors, we performed a 60 ns simulation from the energy-minimized corrected pre5b structure at 20 ns to make sure that these errors do not influence our results and conclusions. The pre5b state was chosen, because the refined structure contained the highest number of errors of all the structures. The root mean square deviation (RMSD) relative to the structure at 20 ns was calculated, a Principal Component Analysis (PCA) of tRNA$^{\text{Val}}$, tRNA$^{\text{fMet}}$, and L1-stalk motions was carried out and the intersubunit rotation angles were calculated as described in 2.3. The results were compared to the two independent pre5b simulations which started from different structures at 20 ns in order to see the effect of different starting structures (representing the same state) compared to the effect of changed chiralities in two simulations starting from the same structure. The PCAs of the tRNAs and the L1-stalk motions as well as the intersubunit rotation angles showed larger differences between the simulations starting from different structures than the difference due to the changed chiralities. The RMSD of the simulation started from the corrected structure was not markedly different from that of the other two simulations of the pre5b state. Hence, possible inaccuracies due to changed chiralities are found to be smaller than the statistical uncertainty due to limited sampling and, therefore, not significant. Next, we investigated the effect the changed chiralities have on our identification of residues involved in contacts

between the tRNAs and proteins L1, L5 and L16. Each of these residues was characterized by the maximum of all its contact frequencies. To measure the similarity of contacting patterns in two simulations, we calculated the ratio of the number residues with the same level of contact frequency (12.5–25%, 25–50%, 50%–100%) in both simulations to the number of residues with different levels. The ratio extracted from the simulations with wrong chiralities and different starting structures was 0.56 and the ratio for the two simulations with the same starting structure but different chiralities was 1.16. Again, the possible inaccuracies due to changed chiralities are smaller than the estimated error range due to limited sampling. The energy-minimized corrected structures were submitted to the pdb-database under pdb-ids shown in Supplementary Table 3. Amino acids and nucleotides which were corrected for chirality are listed in the header of the pdb-files.

1.4   Choice of models for simulation

For each of the major states of spontaneous retro-translocation (pre1 to post4), the fit to the one or two cryo-EM maps with the highest resolution were used as starting structures for MD simulations: pre1a (12 Å), pre1b (12 Å), pre2 (15 Å), pre3 (17 Å), pre4 (13 Å), pre5a (15 Å), post1 (12 Å), post2a (17 Å), post2b (17 Å), post3b (15 Å), and post4 (9 Å). Two additional structures were chosen for simulations, because of extreme intersubunit rotation angles (pre5b) and an extreme tRNA$^{\text{fMet}}$ conformation (post3a), to capture a large range of conformations accessible by the ribosome and the tRNAs. To estimate the effect of refinement accuracy on our conclusions, the refined structure of the pre3 state was perturbed such that the conformation of tRNA$^{\text{fMet}}$ was closer to the pre2 state, but within the resolution limits set by the cryo-EM density.

1.5   MD simulations of the refined models

Atomic models, including the crystallographic resolved ions, obtained from the flexible fitting to 13 cryo-EM maps were solvated, and the system was neutralized with K^+ ions before adding additional explicit salt (7 mM $MgCl_2$ and 150 mM KCl) using the GENION program from the GROMACS suite[4] to mimic the conditions used for the cryo-EM experiments[1]. The system was then equilibrated in four steps:

- 0–5 ns: Constant volume and position restraints on all ribosomal heavy atoms with a force constant of 1000 kJ mol^{-1} nm^{-2}, 2 fs time step.

- 5–10 ns: Constant volume and linearly decreasing the position re straints force constant to zero.

- 10–20 ns: The pressure was coupled to a Berendsen barostat[23] with a coupling constant $\tau_p = 1$ ps and an isotropic compressibility of $4.5 \cdot 10^{-5}$ bar^{-1}.

- 20–120 ns: The pressure was coupled to a Parrinello-Rahman barostat[24] with a coupling constant $\tau_p = 1$ ps and an isotropic compressibility of $4.5 \cdot 10^{-5}$ bar^{-1}.

For each state as well for the initial PE-model, one simulation was carried out except for state pre5b where two independent simulations were carried out. Additionally, two simulations of the pre5b state without tRNAs were

completed. Here, the tRNAs were removed from the structure refined against the pre5b state and the ribosome structure was subsequently solvated and equilibrated in the same way as the other structures.

1.6   Cryo-EM analysis of global ribosome conformation

Cryo-EM grids of vacant *E.coli* ribosomes were prepared at 18°C under controlled environmental conditions[25] and imaged under cryo conditions with a Titan Krios electron microscope (FEI Company, the Netherlands) at 300 kV and defoci ranging from 1 μm to 2.5 μm on a Eagle 4k x 4k CCD camera (FEI Company, the Netherlands) using 2fold pixel binning, resulting in a final pixel size of 3.2 Å. Ribosome particles were selected semi-atomically with Boxer[26] and corrected locally for the CTF[27]. The resulting 9814 ribosome particles were coarsened twofold to a pixel size of 6.4 Å and classified according to 30S body rotation in 2.5° steps as described[1]. Pre-translocation state ribosome particles (315108 in total) were obtained from an existing dataset of *E.coli* ribosome complexes prepared for cryo-EM at different time-points of the retro-translocation reaction, using the same buffer and temperature of 18°C as for the vacant ribosomes[28;1]. Hierarchical classification resulted in 34 groups of ribosome particles representing structurally distinct pre-translocation states[1]. For each population of pre-translocation ribosomes, the 30S body rotation of the corresponding cryo-EM reconstruction was determined. In Fig. 2, the fraction of particles as a function of 30S body rotation was plotted in 2.5° steps for all pre-translocation state ribosomes and vacant ribosomes, respectively. Image processing was generally performed using IMAGIC-5[29] and exhaustive alignment[30].

Supplementary Note 2 (Analysis)

2.1   Comparison to recent crystal structures*

To compare our models to existing crystal structures for each state, an average structure was calculated from the last 10 ns of each trajectory. The root mean square deviations (RMSD) of these structures relative to two *E.coli* ribosome crystal structures (pdb ids: 3R8S, 3R8T[2]) were calculated after rigid-body fitting using all resolved 70S C^{ff} and P atoms, except L9 protein atoms (Fig. 1b).

Of particular importance for our analysis is the quality of our models in the tRNA binding region. To assess the accuracy of the models in this region, the RMSD of the C^{ff} and P atoms which are within a 2-nm distance to the two tRNAs in any of the models was calculated after rigid-body fitting (Fig. 1b).

2.2   Independent MD based refinement

In order to provide an independent test of our refinement procedure, we performed an additional refinement of the AP-model against the pre1b cryo-EM map using an all-atom explicit solvent MD simulation with an additional biasing potential and no further restraints. This biasing potential maximizes correlation between the atomic model, using an adapted Gaussian kernel as described above, and the cryo-EM map[31;32]. The model density maps were calculated for each simulation time step, an effective potential constant[31] of $k = 10^6$ kJ/mol and a total simulation time of 2.7 ns were used. The AP-model was solvated in a cubic simulation box which matches the geometry of the cryo-EM map.

2.3   Definition of reaction coordinates for collective motions*

Principal component analysis*

Principal component analysis (PCA)[33] was carried out for tRNA^{Val}, tRNA^{fMet}, and the L1-stalk using all the simulations of the whole ribosome. To define common sub-spaces, all trajectories were first superimposed by least square fit, using C^{ff} and P atoms of the 50S subunit excluding the L1-stalk. Next, the C^{ff} and P atoms of the tRNAs and the L1-stalk were extracted from the trajectories. For each of the three ribosomal components the extracted trajectories of all states were concatenated, and the atomic displacement covariance matrix was calculated. The trajectories of each state were then projected on the first eigenvector of this matrix. The projections divided by $\sqrt{N}$, where N is the number of atoms used to construct the covariance matrix, yielded the reaction coordinates (r.c.). The minimum and maximum of this reaction coordinate for each state are shown in Fig. 1c. The projections on the first three eigenvectors were used to estimate transition rates (see below).

In order to calculate the distance dependence of the interaction enthalpies, i.e. the sum of electrostatic and Lennard-Jones interactions, between tRNA^{fMet} and the L1 protein, a distance coordinate was obtained from a PCA using C^{ff} and P atoms of the tRNA and the L1-stalk. The atomic displacement covariance matrix was constructed from the trajectories of those states in which the tRNA and the L1 were in contact (pre3–pre2b). The projection on the first eigenvector of this matrix, divided by $\sqrt{N}$, where N is the number of atoms used to construct the covariance matrix, was used as the distance coordinate, where the smallest value observed in the simulations was set to zero.

30S head and body rotation**

Head and body rotations were quantified by comparing structures of each state, extracted from the respective trajectories at 200 ps intervals, to the post1a structure at 20 ns. The post1a structure was used to define zero degree body and head rotations. To define the axes of rotation and pivoting points, we extended a non-linear least squares fitting method[34] to also include the axis of minimal and median rotation. Final mean pivoting point and axes of rotation for head and body movement were determined by quaternion-based averaging[35] over all rotations obtained from all structures of all states. Rotation angles were then calculated relative to the mean axes of rotation and pivoting points.

2.4   Transition rate estimates**

Transition rates were estimated in two steps. Firstly, initial estimates for the free energy barrier heights $\Delta G^{\ddagger}_{est}$ were obtained from a fluctuation analysis of the trajectories of all states. Secondly, these barrier estimates were calibrated by comparing passage frequencies obtained from the initial barrier estimates to passage frequencies actually observed in the simulations, to yield the free energy barriers $\Delta G^{\ddagger}$. The calibrated free energy barrier heights $\Delta G^{\ddagger}$ were then used to calculate the transition rates shown in Fig. 1d.

Free energy barrier estimates**

The initial free energy barrier estimates $\Delta G^{\ddagger}_{est}$ for the transitions between all states for the motions of the L1-stalk, both tRNAs, and 30S head and body

rotations were obtained as follows. Each trajectory (state) was projected onto 3-dimensions spanned by the dominant PCA eigenvectors for the L1-stalk and tRNA motions and by the three rotation angles for the 30S head and body rotation. The mean μ and the 3×3 covariance matrix Σ of each projected trajectory define a three-dimensional multivariate Gaussian probability distribution function,

$$p(x) = \frac{w_{exp}}{\sqrt{2\pi^3 \|\Sigma\|}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}, \tag{1}$$

for the corresponding state, where the weighting factor w_{exp} accounts for the experimentally measured population of the state[1]. The free energy landscape

$$G(x) = -k_B T \ln p(x) \tag{2}$$

arising from such a distribution $p(x)$ describes a three-dimensional quasi-harmonic approximation to the underlying free energy landscape. For each pair of states, the intersection of the two quasi-harmonic approximations obtained for a given motion defines a hypersurface, the free energy minimum of which was used as an estimate for the barrier height $\Delta G_{est}^{\ddagger}$ between the two states. To test whether barrier estimates change with the number of dimensions used for the analysis in PCA space, up to 20 dimensions were used for the L1-stalk and tRNA trajectory projections. Even though the absolute values for the free energy barrier estimates increased, their ratios to the respective mean value did not change markedly. Therefore, we assume $\Delta G^{\ddagger} = c\Delta G_{est}^{\ddagger}$ with a constant calibration factor c.

To determine the statistical uncertainty of the free energy barrier estimate we randomly drew N points from distributions with given mean μ and covariance matrix Σ and recalculated mean μ' and covariance matrix Σ' from the drawn samples. The number of stochastically independent data points that define the distributions was set to $N = N_{total}\frac{t_{ac}}{t}$, where t_{ac} denotes the autocorrelation time of the data points and t the total simulation time. This was repeated until the standard deviation of the free energy estimates for each transition with these newly defined harmonic potentials converged. The statistical uncertainty of the free energy barriers for each transition is the converged standard deviation of the free energy estimates.

Calibration of Arrhenius transition rates**

Several conformational transitions between the 13 conformation states defined in Fig. 1c were actually observed during our 100 ns simulations. Comparison of the statistics of the observed transitions with transition rates obtained from the above barrier height estimates using Arrhenius' law therefore allowed to refine all barrier heights and transition rates by a common factor. To this aim, the barrier height estimates were grouped into height intervals of 2.494 kJ/mol width. For each interval, the fraction $p_{A \to B}^{sim}$ of trajectories for which conformational transitions were observed was determined (cf. Supplementary Fig. 3a) as follows. The trajectories were projected onto the reaction coordinates for the six collective motions defined above and analyzed in 200 ps intervals for the rotational movements and 10 ps intervals for the projections in PCA space. The distance $d_{A,B}$ between two state ensembles A and B was defined as the minimum distance of all projections onto the reaction coordinates. A barrier between A and B states was considered to be crossed if the average distance within an ensemble was found to be larger than the

distance $d_{A,B}$. For each motion, the frequency of barrier crossing $p^{sim}_{A \rightarrow B}$ in an energy interval $I = [\Delta G^{\ddagger}_{est} - \Delta G, \Delta G^{\ddagger}_{est} + \Delta G]$ was defined as

$$p^{sim}_{A \rightarrow B}\left(\Delta G^{\ddagger}_{est}\right) = \frac{n_{A \rightarrow B}}{n}, \tag{3}$$

where $n_{A \rightarrow B}$ is the number of observed barrier crossing from ensemble A to B with estimated free energy barriers in I, and n is the total number of barriers in the same energy interval. The probability of observing a transition from a state A to a state B in a time interval $[0, t]$ is also known from reaction kinetics to be

$$p_{A \rightarrow B} = 1 - \exp\left(-kt\right), \tag{4}$$

where k is the transition rate from A to B. Transition rates k are estimated by Arrhenius' law

$$k = A \exp\left(-\Delta G^{\ddagger}/k_B T\right), \tag{5}$$

where k_B is the Boltzmann constant, and T is the temperature. Assuming the same linear calibration $\Delta G^{\ddagger} = a + c\Delta G^{\ddagger}_{est}$, of all barrier heights and attempt rates, respectively, the calibration factors A and c were determined from a least square fit of the Arrhenius transition probability

$$p_{A \rightarrow B} = 1 - \exp\left[-A \exp\left(\frac{-c\Delta G^{\ddagger}_{est}}{k_B T}\right)t\right] \tag{6}$$

to the respective fraction $p^{sim}_{A \rightarrow B}$ observed in the simulations. From the refined energy barrier heights, transition time estimates $\tau = 1/k$ were obtained from Arrhenius' law, Eq. (5).

The statistical uncertainty of attempt rate A and calibration factor c were determined by reconstructing $p^{sim}_{A \rightarrow B}$ from the free energy barrier estimate for a given transition. We randomly determined whether this transition would occur within a 100 ns simulation time using Eq. (6) with a random shift in $\Delta G^{\ddagger}_{est}$, that accounts for the error estimated for $\Delta G^{\ddagger}_{est}$. Fitting the reconstructed $p^{sim}_{A \rightarrow B}$ to Eq. (6) yields a new calibration factor and attempt rate. When this procedure is repeated, the standard deviation of the redetermined calibration factors and attempt rates converges to the statistical uncertainty of the calibration factor and attempt rate.

2.5   tRNA contacts with the ribosome and mRNA*

To assess the residue-residue contacts and the interaction enthalpy between the tRNAs and the ribosome, for each simulation, all pairs of atoms, respectively from the tRNAs and the ribosome, whose distances were below 3 Å were identified using g_contacts[36]. A residue pair was considered to be in contact if the distance between any two atoms (one from each residue) was found to be below 3 Å in at least one frame of the 100 ns trajectory. Interaction enthalpies between tRNA residues and contacting residues of L1, L5, or L16 proteins were calculated from the MD force field as the sum of electrostatic and Lennard-Jones interactions and averaged over the trajectory of each state (Fig. 1c).

In order to show the distance dependence of the L1-tRNA^{fMet} interaction enthalpy, the interaction enthalpy and the L1-tRNA^{fMet} distance (see section 2.3) were extracted from each simulation at intervals of 1 ns. For this, only simulations of the pre3–post2b states were taken into account (Fig. 1f).

To monitor tRNA-mRNA base-pairing, the minimal distance between the atoms from each codon nucleotide on the mRNA and the corresponding anticodon residue on the tRNA was calculated from each simulation.

Contacts between tRNA nucleosides and residues of the ribosomal proteins as well as the rRNA were found to have different levels of state-specifity, e.g some contacts are only present in a single state and others are present in several states. To asses the level of state-specifity, a contact entropy was calculated as follows: $\sum_{i=1}^{13} -f_i \ln(f_i)$, where f_i is the normalized contact frequency in state i. With this definition, contacts that are more state-specific have a smaller contact entropy than less specific ones.

2.6   Conservation of contact residues in L1, L5, and L16*

Protein sequences of L1, L5 and L16 proteins were retrieved from the UniProt database[37]. 6,029 individual sequences of L1; 6,125 sequences of L5; and 6,031 sequences of L16 protein were used. To reduce computational complexity, sequences that had more than 90% identity were combined to a single cluster, each cluster represented by a single characteristic sequence. 1,174 characteristic sequences for L1; 1,106 sequences for L5; and 859 sequences for L16 protein were analyzed. Individual sequences were manually curated in order to exclude incomplete sequences leaving 1,153 sequences of L1, 1,079 sequences of L5 and 852 sequences of L16 protein.

Multiple sequence alignments were performed using Muscle software[38]. To construct a phylogenetic tree, incomplete positions of multiple sequence alignments were eliminated using Gblocks[39] software. Phylogenetic trees were constructed based on maximum-likelihood with JTT[40] model using PhyML[41]. Rate4Site software with JTT model was used to calculate conservation[42]. Calculation was performed using empirical Bayesian approach, which was shown to be superior to the maximum-likelihood method for site-specific conservation scores[42]. Conservation score was calculated for each individual position of the complete multiple sequence alignment. *E. coli* sequences were used as a reference. The conservation scores calculated by Rate4Site were inverted such that values higher than zero indicate conservation degree which is higher than for the protein in general, whereas values lower than zero indicate less than average conservation.

In order to calculate the contact score, frequencies of contacts between tRNA and protein in every substate were used. For each pair of tRNA-protein contacts, the maximum contact frequency over all substates was determined. For every protein residue, the contact score was calculated as a sum of frequencies from all of its contact frequencies of different interactions. Residues that had a contact score of > 0.8 were considered contacting.

2.7   L1-stalk interaction with the 30S subunit

To monitor the interaction between the L1-stalk and the 30S subunit, the interaction enthalpy between the L1-stalk rRNA and proteins S7 and S11 was calculated. To that aim, the sum of electrostatic and Lennard-Jones interactions were averaged over the trajectory of each state (Fig. 1c).

2.8   L1-tRNA$^{\text{fMet}}$ potential of mean force

The potential of mean force (PMF) between the L1-stalk and the tRNA$^{\text{fMet}}$ was calculated using the extended umbrella sampling simulations[43]. The motion of the system was restricted at selected positions along the vector

describing the distance between L1-stalk and tRNA (see section 2.3). First, 20 equally spaced positions x_i ($i = 1, \cdots, 20$) were chosen between the minimum and maximum value of the projection onto the distance vector observed in the simulations (pre3–post2b). For each i, the structure of the L1-stalk and the tRNA whose projection onto the vector was closest to x_i was then extracted from the trajectories. Next, the obtained structures were solvated (as described in section 1.1) and subsequently energy minimized. An additional harmonic umbrella potential with a force constant of $k_U = 100$ kJ/mol/nm^2 and centered at x_i was applied to the C$^{\text{ff}}$ and P atoms of the tRNA and the L1-stalk to restrain the movement along the distance vector.

Next, the solvent was equilibrated for 5 ns using position restraints on tRNA and the L1-stalk heavy atoms with a force constant of $k = 1000$ kJ/mol/nm^2. Subsequently, the system was simulated for 20 ns with the umbrella potential, but without position restraints.

The distances extracted from the 20 20-ns simulations were used to construct the free energy landscape (Fig. 1f) using the weighted histogram analysis method (WHAM)[44]. The WHAM implementation g_wham[45] was used and the statistical errors were calculated by bootstrapping new trajectories based on the umbrella histograms.

2.9 Bridge B1b interaction enthalpy*

The 50S part of intersubunit bridge B1b[14] was defined as the set of residues of protein L5 that are in contact (see section 2.5) with the 30S residues in at least one of the states. The 30S part was defined as the set of all the 30S residues that are in contact with L5 in at least one of the states. The interaction enthalpy between 50S and 30S part of the intersubunit bridge was extracted as in section 2.5 from the pre5b simulation at intervals of 2 ps.

2.10 Kinetic sequence of states**

From all reaction sequences (i.e. one permutation of all states, {pre1a, pre1b, pre2a, ..., post3b}), we determined the one which best matches the observed set of transition rates as follows. We assume the best matching reaction sequence to be the one that yields the shortest overall half-time τ. This overall half-time is proportional to the sum of the half-times of the state transitions $\tau_{A \to B} \propto \exp\left(\Delta G^{\ddagger}_{A \to B}\right)$. The highest transition barrier estimates dominate the overall half-time, so only the barrier for the ribosomal component governing the transition for a given pair of states was taken into account. The sequence with the shortest overall half-time of states minimizes $\sum_i^{\text{transitions}} \exp\left(\Delta G^{\ddagger\text{max}}_i\right)$, where $\Delta G^{\ddagger\text{max}}_i$ is the highest barrier of all the barriers estimated for the individual ribosome components for transition i. To avoid that only very high barrier estimates with large errors dominate the calculation, barrier estimates higher than 50 kT were set to 50 kT.

To check whether the omission of possible "off-track" states allows for a faster reaction sequence, we performed the above analysis for the fastest progression sequences with single states omitted. Two cases have been excluded here, which trivially enhance the progression rate. First, if end-states are omitted, the reaction sequence trivially becomes faster (e.g. choosing the simple sequence "pre3 pre4"). Second, if two barriers that include at least one capped barrier are replaced by just one capped barrier (e.g. replacing "pre2 pre3 pre4" by "pre2 pre4"). Excluding the two above trivial accelerations of the overall rate, we calculated the fastest progression rate for all

left non-trivial combinations of omitting states from the fastest progression sequences.

2.11 Kinetic sequence of states for subsets of ribosome component movements**

To assess which movements of individual ribosome components dominate the kinetic sequence of tRNA translocation, we calculated the fastest progression sequence using the movements of all ribosome components, of individual ribosome components, or of a combination of them. To that aim, only the maximum free energy barriers from the respective subsets of ribosome components were used for calculating the fastest progression sequences. To quantify the similarity of these sequences to the sequence introduced by Fischer et al.[1], the Kendall rank correlation coefficient τ was used, which reflects the minimum number of swaps of neighbouring states that are required to yield the desired sequence. Because a fully reversed sequence with negative τ leaves the progression rate unchanged, the absolute value of τ was used. Further, to account for the cases where the sequence presented by Fischer et al.[1] contains multiple substates for single states that have no specific order assigned (pre1a, pre1b, etc.) the maximum absolute τ for comparison to any permutation of substates was used.

After translocation, the tRNAs occupy other positions than before translocation, while body, head and L1-stalk return to their initial positions. To quantify how the derived sequences are affected by this fact, we determined a second set of reaction sequences with the first position fixed to a pre1 state.

Supplementary Note 3 (Results)

3.1 Refinement of atomic models

The AP-, PE-, and E-model were refined against 13 cryo-EM maps using DireX[21]. The complete set of all-atom ribosome models for the different states is depicted in Supplementary Fig. 1.

3.2 Comparison to recent crystal structures*

Supplementary Figure 2a compares tRNA positions in our models with tRNA positions in two crystal structures[2] after aligning the binding site regions. For the comparison the pre1a and pre4 models were used which have the lowest RMSDs to these crystal structures (Fig. 1b). Even though the tRNAs used in our simulations and the tRNAs from the crystal structure represent different tRNA-species, they adopt almost the same position, especially in the functionally important anticodon and CCA-tail regions.

3.3 Structural deviation during the simulations

For each simulation, the root mean square deviations (RMSD) relative to the respective starting structure and relative to the structure at 20 ns are shown in Supplementary Fig. 2b. In all simulations an RMSD of less than 8 Å was obtained, which is very low for a system of this size. Typical values obtained by other authors are $\sim$10 Å[46]. Note that the PE-model simulation started from a crystal structure and thus can serve as a benchmark for the quality of the EM-fitted structures[47]. Notably, the RMSD values for the fitted structures are similar or only slightly larger, thus underscoring the quality of these structures.

3.4 Independent MD based refinement

The independent MD based refinement against the pre1b cryo-EM map resulted in an RMSD of 3.26 Å relative to the corresponding structure obtained using DireX. Further, an RMSD of 4.54 Å was observed between the MD refined structure and the above mentioned crystal structure with a ribosome including a P-site tRNA[2]. A comparable RMSD of 5.31 Å was found for the DireX refined pre1b structure (cf. Fig. 1b, solid green line).

The fact that the two structures obtained by the two refinement methods are more similar to each other than to the reference crystal structure suggests that the refinement quality is independent of the choice of an elastic network based or an MD based refinement method.

3.5 Stereochemical parameters of the models**

To further assess the quality of our models, we have carried out a statistical analysis of various stereochemical quantities. Supplementary Table 2 shows the deviation of the distributions of our models from those obtained from the protein data bank (PDB). To fully incorporate the additional effect of the simulations, the energy-minimzed refined structures and the energy-minimized structures after 100 ns of free MD simulation were used. Stereochemical parameters were calculated using WHATCHECK[48] and then averaged over all models (pre1a–post4). The ribosomal environment provides different conditions for protein and RNA folds than for most of the protein and RNA

structures in the PDB. For this reason, we also checked ribosome crystal structures[2,3]. Supplementary Table 2 shows that upon initial refinement and energy-minimization through the force-field, the stereochemical parameters of our models are comparable to these of recent crystal strucutures. The stereochemical parameters of the models after 100 ns of free MD simulation, and subsequent energy-minimization further approached the parameters expected from an extensive analysis of the pdb-database[48]. Overall our results suggest that the stereochemical quality of our models is similar to that of crystal structures of comparable complexes.

3.6   30S head and body rotation**

The 30S body rotation pivot point was found to be close to the 16S RNA residues G242 and U562, and the 30S head rotation pivot point is close to the 16S RNA residues A923, U1194, and G1386.

3.7   Transition rates**

The fit of the probability of barrier crossing $p_{A \to B}$ to the frequency of barrier crossings $p_{A \to B}^{sim}$ obtained from the simulations (see Supplementary Fig. 3a), yielded an attempt rate of $A = (22.4 ns)^{-1}$ with an error interval from $(15.95 ns)^{-1}$ to $(30.72 ns)^{-1}$ and a calibration factor $c = 0.601 \pm 0.069$ at $t = 100$ ns. The relative statistical uncertainty of the free energy barrier estimates is 57%. Together these errors affect the statistical uncertainty of the transition rates by

$$\frac{\Delta k}{k} = c \sqrt{\left(\frac{\Delta A}{A} \frac{1}{c\Delta G^{\ddagger}} \right)^2 + \left(\frac{\Delta \Delta G^{\ddagger}}{\Delta G^{\ddagger}} \right)^2 + \left(\frac{\Delta c}{c} \right)^2} \tag{7}$$

$$\frac{\Delta k}{k} = 0.601 \sqrt{0.429 \left(\frac{1}{c\Delta G^{\ddagger}} \right)^2 + 0.338}. \tag{8}$$

The relative statistical uncertainty of the transition rate estimates ranges from 52% for fast transitions to 35% for slow transitions. The attempt rates and error estimates for the individual ribosome components are shown in Supplementary Fig. 3b. Notably, the attempt rates for the individual ribosome components do not differ markedly from the overall attempt rate, which was determined by combining the data from all transitions of all ribosome components.

3.8   tRNA-mRNA base-pairing

For tRNA^{Val} the base-pairing to the anticodon is maintained in all the simulations with minimal distances around 0.19 nm (Supplementary Fig. 4), except for the wobble base pair cmo^{5}U35-U which shows larger minimal distances in states pre5b, post3a and post3b. For tRNA^{fMet} the base-pairing is partially disrupted in the simulations of the pre2 and pre3 states and, as expected, when the tRNA is moving out of the E site (post1–post4).

3.9   Rapid angular rearrangement after tRNA removal

The rotation angle between the 50S and 30S subunits changes quickly after tRNA removal (Supplementary Fig. 5). Compared to the simulations with bound tRNA (red curves), consistently lower body rotation angles are observed for the two simulations of the tRNA depleted ribosome structure

(green). The interaction enthalpies of intersubunit bridge B1b between residues of the 50S protein L5 (109–111, 114, 135, 141, 143, 145–146, 177) and 30S proteins S13 (2, 6, 8, 56-57, 60, 63–65, 69–70) and S19 (26) were calculated. The interaction enthalpies were found to be weaker for simulations without tRNAs, indicating that the presence of tRNAs leads to a stabilization of large rotation angles by bridge B1b in the pre5b state. Figure 2b,c show histograms of B1b interaction enthalpies and 30S body rotation angles extracted from the last 50 ns of the simulations.

3.10   Influence of refinement accuracy on our conclusions*

To estimate if and to which extent our conclusions might be affected by possible structural inaccuracies resulting from our refinement of an X-ray structure against the 13 cryo-EM maps has on our conclusions, we repeated a simulation and our analyses using the pre3 structure with the tRNA^{fMet} in an artificially perturbed conformation. This perturbed structure also obeys all stereochemical and energetic constraints set by the force field and is within the resolution limits set by the cryo-EM map of the pre3 state, with a tRNA conformation closer to that of the pre2 state. From a simulation starting from this perturbed structure, for the reaction coordinates (r.c.) shown Figs. 1c,e, one would expect the changes to mostly affect tRNA^{fMet} motion, such that it is similar to the motion in the pre2 state. Indeed, the r.c. values for this motion are lower in the perturbed simulation, close to those observed in the pre2 state, and the tRNA^{fMet}-L1 interaction enthalpy is weakened (data not shown). The tRNA^{Val}-L16 interaction enthalpies are weaker in the pre3 state than in the neighbouring states. Any change in the tRNA^{Val} conformation in this state is expected to result in a stronger interaction enthalpy, as is the case for the perturbed simulation. The changes observed, however, would not have changed the qualitative picture of the interaction enthalpies offered by Fig. 1c. Interestingly, the changes in body rotation angle observed in the perturbed pre3 state reflect the coupling between tRNA conformation and body rotation reported in the main text (compare Fig. 2). The influence of the perturbation on the other motions is small.

Upon replacement of the pre3 state by the perturbed pre3 state in the the transition rate estimation, the barrier heights for head rotation and tRNA^{Val} motion did not change significantly. In particular, only transition barriers larger than ∼4 kT are affected corresponding to rates slower than $1/\mu$s, thus leaving Fig. 1d unchanged.

3.11   Markov-state like representation of states and transition barriers**

Supplementary Fig. 6a shows a schematic representation of the translocation intermediate states as a Markov model.

3.12   Kinetic sequence of states**

The sequences of translocation intermediate states with the shortest overall half-time within the range of the stochastical uncertainties of the correction factor and barrier heights are {pre1a, pre1b, pre2, pre4, pre3, pre5b, pre5a, post2a, post2b, post1, post3a, post3b} and {pre1a, pre1b, pre2, pre3, pre4, pre5b, pre5a, post2a, post1, post2b, post3a, post3b}. They closely resemble the sequence derived by Fischer et al.[1] based on structural similarity.

Omitting states in the proposed sequences did not yield a faster overall half-time, with the exception of omitting the pre5b state. According to this analysis only pre5b is a possible "off-track" state, suggesting that all other states are kinetically relevant.

3.13   Kinetic sequence of states for subsets of ribosome component movements**

As shown in 3.12, sequences based on structural similarity reflect kinetic sequences, i.e. the ones with the fastest progression rate for the whole ribosome. Since sequence of translocation intermediate states presented by Fischer et al.[1] is based on structural similarity of the tRNAs, the fastest progression sequence based only on the barriers heights impeding tRNA movements should also match this structure based sequence. To test this idea, we determined the sequences, including all 12 states with two tRNAs, with the fastest progression rate for all 31 possible subsets of ribosome components. Indeed, Supplementary Fig. 6b shows that all fastest progression sequences including only tRNA movements match the structure based sequence. Further, the inclusion of movements of other ribosome components leaves sequence of states essentially unchanged, suggesting that the movement of the tRNAs dominates the movements of the other ribosome components. As a negative control, we looked at the fastest progression sequences from body, head and L1-stalk movements without tRNA movements. Indeed, completely different sequences were obtained, whose τ values are comparable to those of randomly drawn sequences (Supplementary Fig. 6b). Some of these sequence contain multiple pre to post transtions.

Since the fastest progression sequences with the first position fixed to a pre1 state are almost as similar to the structure based sequence as the ones with an unfixed first state (data not shown), we conclude that the fact that the tRNAs occupy different positions before and after translocation, while body, head and L1-stalk return to their initial positions does not markedly influence the determined sequences.

3.14   Contacts between the tRNAs and the ribosome and conservation of involved protein residues*

The frequencies of contacts between the two tRNAs and the ribosome were determined for all states from the respective trajectories as described above. The contacting residues of L1, L5, and L16 are generally found to be clustered, e.g., D51–R60, R122–K141 and R164–K167 in L1, I43–K47, S72–Y82 in L5, M1–R10, R44–R59 and K76–E90 in L16 protein (Fig. 3b). All these contacting fragments have a high conservation degree and appear more conserved than the protein on average.

To test whether contacting residues had in fact a higher degree of conservation than the rest of the protein, we used a one-sided permutation test[49]. Differences in means were calculated for the contacting and non-contacting surface residues. Then, the same difference was calculated for every possible permutation of contacting and non-contacting groups. The p-values were calculated as a proportion of sampled mean differences larger than the observed value. The significance level was set at 0.05 (or 5%). Indeed, tRNA-contacting residues have a higher conservation degree for L1 (p-value=0.0019), L5 (p-value=0.027) and L16 (p-value=1.05×10^{-5}) proteins, with an overall p-value=6.62×10^{-8}. The high degree of evolutionary conservation of the protein residues which were identified to contact tRNA during

translocation provides an independent evidence for their potential functional importance.

Overall, the contact entropy of tRNA-rRNA contacts is larger than that of tRNA-protein contacts: 64 % of the tRNA-rRNA contacts, but only 38 % of the tRNA-protein contacts have a contact entropy above 0.5. This means that there are more state-specific contacts between the tRNAs and the proteins, which is an interesting finding in itself.

Supplementary Tables 4a–4t list the frequencies of contacts between tRNA$^{\text{Val}}$ and the ribosome for each state. Supplementary Tables **??**–**??** list the contacts between tRNA$^{\text{fMet}}$ and the ribosome. The gray-scale level of the cells indicates the frequency of atom-atom contacts corresponding to the residue pairs, white (0–12.5%), light gray (12.5–25%), dark gray (25–50%), and black (50–100%).

Supplementary References

[1] Fischer, N., Konevega, A., Wintermeyer, W., Rodnina, M. & Stark, H. Ribosome dynamics and trna movement by time-resolved electron cryomicroscopy. *Nature* **466**, 329–333 (2010).

[2] Dunkle, J. A. *et al.* Structures of the bacterial ribosome in classical and hybrid states of trna binding. *Science* **332**, 981 (2011).

[3] Zhang, W., Dunkle, J. A. & Cate, J. H. D. Structures of the ribosome in intermediate states of ratcheting. *Science* **325**, 1014 (2009).

[4] Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory. Comput.* **4**, 435–447 (2008).

[5] Hornak, V. *et al.* Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712–725 (2006).

[6] Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).

[7] Joung, I. S. & Cheatham, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **112**, 9020–9041 (2008).

[8] Essmann, U. *et al.* A smooth particle mesh ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).

[9] Feenstra, K. A., Hess, B. & Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* **20**, 786–798 (1999).

[10] Hess, B. P-lincs: a parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**, 116–122 (2008).

[11] Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).

[12] Vriend, G. What if: a molecular modeling and drug design program. *J. Mol. Graphics* **8**, 52–56 (1990).

[13] Gao, Y. *et al.* The structure of the ribosome with elongation factor g trapped in the post-translocational state. *Science* **326**, 694–699 (2009).

[14] Yusupov, M. *et al.* Crystal structure of the ribosome at 5.5 angstrom resolution. *Science* **292**, 883–896 (2001).

[15] Jossinet, F. & Westhof, E. Sequence to structure (s2s): display, manipulate and interconnect rna data from sequence to structure. *Bioinformatics* **21**, 3320 (2005).

[16] Theimer, C. A., Finger, L. D., Trantirek, L. & Feigon, J. Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase rna. *P. Natl. Acad. Sci. USA* **100**, 449–454 (2003).

[17] Grishaev, A., Ying, J., Canny, M. D., Pardi, A. & Bax, A. Solution structure of trna val from refinement of homology model against residual dipolar coupling and saxs data. *J. Biomol. NMR* **42**, 99–109 (2008).

[18] Yusupova, G., Jenner, L., Rees, B., Moras, D. & Yusupov, M. Structural basis for messenger rna movement on the ribosome. *Nature* **444**, 391–394 (2006).

[19] Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**, 247–260 (2006).

[20] Pettersen, E. F. *et al.* Ucsf chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

[21] Schröder, G., Brunger, A. & Levitt, M. Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. *Structure* **15**, 1630–1641 (2007).

[22] Ahmed, A. & Tama, F. Consensus among multiple approaches as a reliability measure for flexible fitting into cryo-em data. *Journal of Structural Biology* **182**, 67–77 (2013).

[23] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).

[24] Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamicsmethod. *J. Appl. Phys.* **52**, 7182 (1981).

[25] Bellare, J. R., Davis, H. T., Scriven, L. E. & Talmon, Y. Controlled environment vitrification system: an improved sample preparation technique. *J. Electron Micr. Tech.* **10**, 87–111 (1988).

[26] Ludtke, S. J., Baldwin, P. R. & Chiu, W. Eman: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).

[27] Sander, B., Golas, M. M. & Stark, H. Automatic ctf correction for single particles based upon multivariate statistical analysis of individual power spectra. *J. Struct. Biol.* **142**, 392–401 (2003).

[28] Konevega, A. L. *et al.* Spontaneous reverse movement of mrna-bound trna through the ribosome. *Nat. Struct. Mol. Biol.* **14**, 318–324 (2007).

[29] van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the imagic image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).

[30] Sander, B., Golas, M. M. & Stark, H. Corrim-based alignment for improved speed in single-particle image processing. *J. Struct. Biol.* **143**, 219–228 (2003).

[31] Orzechowski, M. & Tama, F. Flexible fitting of high-resolution x-ray structures into cryo-electron microscopy maps using biased molecular dynamics simulations. *Biophysical Journal* **95**, 5692–5705 (2008).

[32] Ratje, A. H. *et al.* Head swivel on the ribosome facilitates translocation by means of intra-subunit trna hybrid sites. *Nature* **468**, 713–716 (2010).

[33] Amadei, A., Linssen, A. B. & Berendsen, H. J. C. Essential dynamics of proteins. *Proteins* **17**, 412–425 (1993).

[34] Halvorsen, K., Lesser, M. & Lundberg, A. A new method for estimating the axis of rotation and the center of rotation. *J. Biomech.* **32**, 1221 – 1227 (1999).

[35] Gramkow, C. On averaging rotations. *Int. J. Comput. Vision* **42**, 7–16 (2001).

[36] Blau, C. & Grubmuller, H. g_contacts: Fast contact search in bio-molecular ensemble data. *Computer Physics Communications* – (2013). Doi:10.1016/j.cpc.2013.07.018.

[37] Apweiler, R. *et al.* Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res.* **39**, D214–219 (2011).

[38] Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).

[39] Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).

[40] Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282 (1992).

[41] Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

[42] Mayrose, I., Graur, D., Ben-Tal, N. & Pupko, T. Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior. *Mol. Biol. Evol.* **21**, 1781–1791 (2004).

[43] Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **91**, 275–282 (1995).

[44] Kumar, S., Rosenberg, J., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *J. Comput. Chem.* **13**, 1011–1021 (1992).

[45] Hub, J. S., de Groot, B. L. & van der Spoel, D. g_wham - a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *J. Chem. Theory Comput.* **6**, 3713–3720 (2010).

[46] Brandman, R., Brandman, Y. & Pande, V. S. A-site residues move independently from p-site residues in all-atom molecular dynamics simulations of the 70s bacterial ribosome. *PLoS One* **7**, e29377 (2012).

[47] de Groot, B. L., Engel, A. & Grubmüller, H. The structure of the aquaporin-1 water channel: a comparison between cryo-electron microscopy and x-ray crystallography. *J. Mol. Biol.* **325**, 485–493 (2003).

[48] Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. Errors in protein structures. *Nature* **381**, 272–272 (1996).

[49] Röhmel, J. Precision intervals for estimates of the difference in success rates for binary random variables based on the permutation principle. *Biometrical journal* **38**, 977–993 (2007).

4. Discussion

We will discuss the findings that were presented in the publications "Energy barriers and driving forces in tRNA translocation through the ribosome"[22] and the respective supplementary information with a focus on my personal contributions. Then we will, in retrospect, discuss the assumptions, limitations and potential improvements to the methods used to obtain these results in the following sections. We will show, where different approaches might have been used and highlight new possible applications of the developed methods.

We built an all-atom model of the ribosome using X-ray crystallography data and refined it to thirteen cryo-EM densities from Fischer et. al [6] (Supplementary Fig. 1 in [22]). By that refinement we obtained an all-atom description of intermediate states of tRNA translocation.

Validity and predictive power of our refined ribosome models has been confirmed by the agreement between our model and ribosome conformations later found in X-ray structures [32] (Fig.1b in [22]). Bond length and angle distribution quality as well as the overall stereochemical quality of our ribosome models matches that of X-ray structures (Supplementary Table 2 in [22]). We conclude from our findings that models from cryo-EM refinement can parallel the accuracy of X-ray structures, if the refinement is based on high-resolution cryo-EM data and a good starting model. This high model quality further allows to reliably perform MD simulations with the refined models as starting structures.

Quantifying ribosome movements from cryo-EM densities requires manual guidance in segmenting cryo-EM densities (Fig. 1.2a shows a segmented ribosome cryo-EM density). Instead, our refined models directly show the movement of the tRNAs, L1-stalk and small subunit body and head. We generated ensembles from the refined models of the thirteen substates by 100 ns MD simulations. To identify the interaction patterns on ribosome parts that are crucial for tRNA translocation, we first identified the movements that govern translocation by exploring the structure of the ribosome free energy landscape.

We initially proposed that the structure of the free energy landscape of tRNA translocation is governed by a hierarchy of barriers, as suggested by Frauenfelder (Fig. 1.6). Indeed, we found that this postulated hierarchy of energy barriers impedes the movement of the tRNAs, the L1-stalk as well as body and head rotations (Fig. 1d in [22]). The Frauenfelder picture, in which a hierarchy of energy barriers governs transitions between states with similar

free energy (Fig. 2.1), was initially proposed for myoglobin [14]. We find it confirmed for the much larger and more complex ribosome.

Whereas the cryo-EM experiment determined the occupation probabilities of the energy minima in the energy landscape, our free energy barrier estimates through a harmonic approximation to the ensemble data estimated the probability to find ribosome parts in transition states between two translocation intermediates. Our simulations further allowed us not only to determine the free energies of the transition states, but also how fast transitions occur from one state to another. To that aim, we linked the energy barriers to transition rates using the fact that our molecular dynamics method (eq. (2.11)) provides the correct thermodynamical ensemble and, additionally, dynamic information. We exploited this dynamic information by determining an overall attempt rate and geometry factor by comparing observed transitions in our simulation time to the expected transition probabilities from the free energy barrier estimate (section 2.4 and Supplementary Information in [22]).

The uncovered hierarchy of barriers (Fig. 1d in [22]) suggests a classification of ribosome substates into early pre states (pre1 and pre2), late pre-states (pre3-pre5), early post (post1-post2) and late post states (post3). The free energy barrier between the pre2 and the pre3 state has been identified before from the population analysis of small subunit body and head conformations [6]. We now find that not only body and head rotation impede the transition from pre2 to pre3, but also the movement of both tRNAs. In fact, in the majority of other states it is the tRNA movement, not body and head rotations, that impedes transitions.

Though translocation is presented as a sequence of intermediates, the actual transition dynamic is a stochastic process in an energy landscape [6, 126]. So far, it was unclear whether the sequence of translocation intermediates in [6] is kinetically relevant. We tested the kinetic relevance of this linear sequence based on structural similarity by comparing it to the sequence of states that would show the fastest overall progression. To perform the metadynamics analysis that describes translocation as this stochastic process, we discretized the continuum of translocation substates by a Markov model (Supplementary Fig. 5), where we assume that transitions in our Markov model are given by the slowest of our estimated intrinsic transition rates. Using our Markov state model, we obtained the fastest progression sequence of substates that almost coincides with the sequence derived from purely geometric sorting, except for a swap between the post1 and post2 states. Together with the barriers seen between these states, our data would now suggest to group the post1 and post2 states into one.

We have also addressed the question whether "dead-end" states exist in tRNA translocation using the same framework (Supplementary Information section 3.12). On the one hand this seems unlikely in the context of a ribosome that evolved to be highly efficient, on the other hand these states might exist as kinetic controls. Our results suggested pre5b is a "dead end", though removal of this state from the fastest progression sequence did not significantly alter the overall sequence progression rate. More precise kinetic data on the transition rates from the pre5 states would be required to increase the significance

of this finding. The current rate estimates, especially for transitions from pre5 states to the early post states are coarse due to the large barriers that slow the pre- to post-translocation state transition.

After describing the ribosome dynamics as a concerted motion of tRNA, L1-stalk, body, and head movements, we assessed the influence of the individual parts on the meta-dynamics described by the Markov model above by excluding the rate limiting effect of individual movements, as well as combinations of ribosome movements (Supplementary Information Fig. 6b). The analysis allowed us to conclude that tRNA movements determine the kinetic sequence of states.

We extracted how ribosome parts individually contribute to the barriers between states and showed a possible way to connect them to the overall ribosome transitions with our Markov model. These rates for the movement of single ribosome parts were obtained using complete ribosome simulations, whereas, to estimate the transition rates, we treated them as independent from the ribosome. This procedure allowed us to extract "intrinsic" transition rates, and to disentangle the mutual influence of the ribosome parts.

We further asked, what interactions create the free energy barriers and how they are overcome. Detailed protein residue and RNA nucleotide interactions are not accessible to cryo-EM, but through our thirteen all-atom descriptions of translocation intermediates. The stability of these interactions is an ensemble property whereas our refined models represent single ribosome conformations. This ensemble property of cryo-EM densities, is reflected in the smeared out cryo-EM densities of the flexible ribosome parts (compare the green and purple tRNA densities of the two 12Å resolution cryo-EM densities in Fig. 1.5b), but lacks the resolution to identify interactions on a per-residue level. From our simulation ensemble information we extracted the interaction stability, i.e. the contact frequency between the tRNAs and ribosomal proteins (Fig. 1c in [22]). This extraction of contact frequencies from 50 000 conformational snapshots was only possible through the efficient contact finding algorithm we developed [21].

We identified three main tRNAs interaction partners on the large ribosomal subunit, L1, L5 and L16 (Fig. 1c in [22]). The tRNAs interact in distinct patches with L16 and L5, stepping along L16 and sliding along L5 (Fig. 5 in [22]), and interact in one strong contact patch with the L1-stalk that serves to pull the tRNA out of the ribosome.

To address whether the identified mechanisms are anecdotal or universal to prokaryotic translocation, we further analysed the obtained contact frequencies. We assumed that the contacting residues should be more than average conserved if they are universally relevant. Through a comparison of more than 6 000 protein sequences we found that contact patches are indeed conserved (Fig. 3b in [22]).

The underlying experimental data as well as each of the above described steps, the model building, the cryo-EM refinement, MD simulations, contacts analysis, free energy barrier estimates, attempt rate and geometry factor determination as well as the Markov model of translocation are based on numerous assumptions, which we will discuss now. Further, we will show where the de-

veloped methods are not limited to the ribosome, but applicable to related systems.

4.1   Biochemical background

The biochemical setup used to gather the cryo-EM data for our ribosome model refinement is special in many ways. First, it uses an in vitro model assay of *E. coli* ribosomes. Second, it does not use EF-G and GTP to drive translocation, but thermal fluctuations in equilibrium. Third, the used buffer conditions and the specific tRNA pair $tRNA^{Val}$/$tRNA^{fMet}$ favor retro-translocation, not the canonical forward translocation [127].

In vitro assays neglect cellular effects. In vivo, translocation rates are dominated by the availability of cognate tRNA [128] and influenced, e.g., by molecular crowding in the cell and the number of ribosomes reading mRNA in close proximity to each other [129, 130]. Though these effects do affect the overall protein expression rate, we assume that they do not alter the fundamental properties of ribosome function we want to explore. Their influence on ribosome function is only understood, if we understand translation in its most basic setting.

Escherichia coli is an established model organism in spectroscopic, X-ray, and cryo-EM measurements. Mechanisms that are specific for *E. coli* might be mistaken as general features of prokaryotic translocation. However, the extremely high evolutionary conservation of ribosomes and the large structural similarity to extromphile prokaryotes from completely different environments, such as *Haloarcula marismortui*, *Thermus thermophilus*, and *Deinococcus radiodurans* renders it unlike that the features we describe are specific to *E. coli*. In contrast, we find many key features of prokaryotic ribosomes even in eukaryotes [131, 132] — to an extent that a common naming convention for ribosomal proteins in eukaryotes and prokaryotes is being developed.

The biological relevance of the cryo-EM data we used has been challenged on the grounds that the tRNAs translocate "backwards" through the ribosome. Due to initial excess of $tRNA^{fMet}$ in solution, the preferred $tRNA^{fMet}$ pathway is the seemingly paradox $tRNA^{fMet}$ entry through the ribosomal exit site, and subsequent movement to the peptidyl site. However, after 20 minutes, the excess concentration of $tRNA^{fMet}$ is depleted [6] and the system is equilibrated[1]. In an equilibrium system, all processes are time-independent, thus the retro-translocation setup cannot have any influence on the translocation pathway. Indeed, different translocation pathways for forward and retro-translocation in equilibrium would generate directed motion without energy cost, a perpetuum mobile. Further, Fischer et al. showed that during equilibration only the ratio between pre-and post-translocation state population changes, not the populations within pre- and post-translocation states.

Yet, the special buffer conditions and the combination tRNA species required to induce retro-translocation in an excess of $tRNA^{fMet}$ might affect the

---

[1]More precisely it approached equilibrium beyond measurable deviations, true equilibrium is only reached in the infinite limit.

translocation pathway. Though the tRNA species does affect the overall rate of translocation [133], we expect little influence on the tRNA translocation pathway. First, the tRNAs are very similar in structure and ribosome binding sites very narrowly fit the tRNAs, sterically prohibiting large deviations. Second, the observed order-of-magnitude difference in rates translates to a free energy difference of $(\log 10)k_BT \approx 2.3k_BT$, which is below the error margin of our rate estimates.

Though we do expect an influence of the buffer conditions on our estimated transition rates, a rigidifying effect of Mg^{2+} ions on RNA [134], for example, we do not expect a large effect on our conclusions, because ion rearrangement on RNA is very slow compared to our simulation time-scales [134].

Through the use of GTP, EF-G drives translocation forward, i.e. in presence of EF-G and GTP, the pre-translocation states are rapidly depleted and the occupation of post-translocation states is favoured. However, it is unknown to what extent EF-G alters the reaction pathway during tRNA translocation. Yet, we assume that the fundamental principles of movement of tRNA through the ribosome remain unchanged, because the available structural data [135, 136, 137, 138] suggests that EF-G mainly acts to prevent backward translocation. Further, our estimated interaction enthalpies largely exceed the energy provided by GTP hydrolysis, rendering a complete change in the free-energy valley that defines tRNA translocation unlikely.

Still, EF-G catalyzed GTP hydrolysis might act as a "gate-opener" to a distinct free energy "valley" in which tRNAs move during EF-G catalyzed translocation. Here, by the very long (20 min) equilibration time that the used cryo-EM setup allowed, we would very likely already observe such a second "valley". Overall, though the translocation kinetics are definitely affected by the presence of EF-G, we assumed that translocation paths and ribosome-tRNA interactions are very similar with and without EF-G present.

4.2 Cryo-EM reconstruction

For our refinement, we used a subset of the sharpened cryo-EM density maps instead of the raw data, ca. 2 million cryo-EM micrographs. Though the cryo-EM reconstruction method carfully tested against overfitting of the data through crossover refinement, a small inherent bias towards the ensemble of obtained cryo-EM densities and to the assigned free energies remains. This is due to the fact that for a reconstruction of cryo-EM maps configurational space is binned to yield a sufficient number of micrographs to reconstruct a cryo-EM density. The hierarchical sorting routine that was employed, for example, bins body rotations in $5°$ steps. Fig. 4.1 shows that any partitioning of state space changes the free energy landscape. Uneven partitioning might even introduce artificial high energy transition state intermediates, an effect that was avoided in the hierarchical sorting routine employed by Fischer et al.

The fact that a different sorting routine, e.g. a shift of the 30S body rotation bins by one degree, would result in a slightly different reconstructed ensemble of structures with different assigned free energies shows that the cryo-EM raw
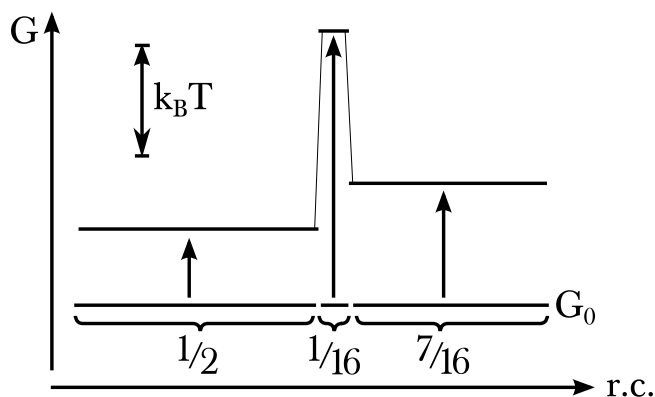
Figure 4.1: Partioning in state space changes a flat free energy landscape G_0 and introduces a free energy barrier.

data ensemble information is not fully exploited.

A potential way to further exploit the cryo-EM raw data would be the assignment of free energies to ribosome configurations that were generated in our MD simulations using the 2d projection images from cryo-EM. To perform this task, the ribosome configurations would be projected into two dimensions and subsequently assigned to cryo-EM micrographs. A subsequent counting of the assigned projections would yield an occupation probability for a ribosome configuration directly from the cryo-EM raw data. Sifting through the most likely configurations would in turn yield a reconstruction method similar to electron density reconstruction methods from free-electron laser diffraction images.

4.3 Cryo-EM refinement

From the ensemble of fifty available cryo-EM density maps we choose a subset of thirteen translocation intermediates. First, we selected the best resolved cryo-EM densities of major tRNA configurations (pre1-post4 states). Where cryo-EM maps with significantly differing features for these major tRNA configurations had been resolved, we complemented our initial set with these maps (in the pre1, pre5, post2 and post3 state) to best represent the complete set of fifty resolved cryo-EM densities.

Few of the many possible atom configurations that fit cryo-EM data are biologically and physically meaningful. Cryo-EM refinement poses an incomplete problem where additional constraints, a force-field, need to be introduced to obtain the desired, physically and biologically relevant, atom configurations. With these additional constraints however, the refined structures are an overinterpretation of cryo-EM data if not cross-validated independently.

Of the variety of cryo-EM refinement procedures we chose to maximize the cross-correlation of the atom structure to the cryo-EM density over using the inverted cryo-EM density as an additional attractive potential energy term in the force-field [139]. This eliminated the need for strong constraints on

secondary structure that were necessary in previous approaches to refine cryo-EM densities of the ribosome [64].

To test the influence of the refinement procedure choice on the resulting conformation, we used DireX [122] and implemented a refinement based on the algorithm of Orzechowski et al. [140], using a physical potential as a force-field instead of the dynamic elastic network model used in DireX. The improved results with the physical potential at the expense of increased computational cost suggests that initial configurations obtained by refinement with DireX should be further refined using a physical potential.

Though all atom positions were available from the refinement procedures and our refinement did predict a crystal structure conformation (Fig.1b in [22]) within the range of thermal fluctuations, we made sure not to over-interpret our refined structures by ensemble and spacial averages. The ensemble averages were performed over all frames of MD simulations of the refined structures and thus minimzed the effect of artifacts in the single refinement structure. The spacial averages were performed over protein residues and RNA nucleotides, when describing local interactions.

Two key questions are to be addressed in future refinement approaches. First, present refinement procedures yield a single conformation, whereas the cryo-EM map represents a conformation ensemble. This issue is currently being addressed by Vaiana et al. through implementing a target function that maximizes the cross-correlation to an average map from an ensemble of structures, as proposed in [141].

The second question is, how strongly a force-field may influence the cryo-EM refinement without overfitting the data. A force-field free refinement, will yield a very high correlation at the cost of an unphysical, chemically meaningless structure, whereas a refinement guided predominantly by the force-field would result in poor correlation to the input density. The weighting λ between force-field $V(x)$ and experimental input $C(x)$ might be addressed using a Hamiltonian replica-exchange method [142], where multiple refinement Hamiltonians $H(\lambda) = \lambda V(x) + (1 - \lambda)C(x)$ are used simultaneously. The ensemble reweighting technique would further yield the required ensemble of refined structures as described above and might prevent the common refinement problem where refinement algorithms are trapped in local minima.

4.4 Force-field and sampling

The conclusions we draw for mechanisms of tRNA translocation from our simulations rely on a sufficiently accurate description by the MD force-field and sampling of ribosome conformations.

Apart from the conventional limitations sketched in section 2.2, due to the high flexibility and charge of RNA compared to proteins, parametrization of RNA force-fields for simulations is especially challenging [134, 143]. Correctly accounting for long-range electrostatic interactions is crucial for RNA simulations [143]. To accurately model these long-range electrostatic interactions, we used the particle-mesh-ewald method [144]. The accurate description of RNA-ion interactions poses an additional challenge to the force-field

[143]. Though the ion parameters by Joung et al. [145] largely improved the simulation properties of monovalent ions and especially prevents their unphysical aggregation in solution as observed in previous ribosome simulations [146], the interactions of Mg^{2+} with RNA are not fully described by current models. However, multivalent ions in our simulations were placed in their binding sites as resolved in crystal structures. Because ion rearrangement on RNA is much slower than the time-scales of our simulations, we assume that these short-comings in force-field parametrization do not affect our conclusions [134].

Influence of the force-field parameters on the simulation results is usually tested by performing the same set of simulations with another force-field. We did not perform this canonical test against another force-fields, because no other force-field for RNA and protein simulations with modified RNA nucleotide parameters was available that was parametrized independently from the amber99sb-ildn force-field [147] we used.

Despite the vast computational time ($\approx 2 \cdot 10^6$ CPU hours) that we used for sampling ribosome conformations with highly efficient algorithms [148], we cannot rule out that single simulations are "stuck" in improbable excursions in the free energy landscape. To test how robust our conclusions are with respect to excursions in the energy landscape, we performed one of our simulations with an improbable, distorted, tRNA conformation in the pre3 state. We found that our conclusions were not affected and that the overall conclusions remain valid. Multiple parallel simulations from the same refined conformations have now been performed to address this concern further. Yet, essential movements on the meso-scale between the 100ns dynamics from our simulations and the much larger time-scales reflected in the cryo-EM data might not be covered due to our limited sampling.

To address our limited sampling of ribosome conformations, we extended our sampling capacities over the past three years. Accessible ribosome simulation time-scales, and thus the available sampling, increased by a factor of ten through an increase in computer power and more efficient simulation algorithms, such as domain-decomposition, electro-static calculations and the use of graphical processing units.

4.5   Transition rate estimates

The limited sampling time will also have a direct effect on the estimated free energy barriers of the ribosome. Figure 4.2 shows a one-dimensional example system, where limited sampling of two intial configurations creates an apparent free energy barrier in a flat free energy landscape. The estimated free energy barrier in this system only vanishes after infinitely long sampling.

Also in our simulations, barriers might be overestimated due to limited sampling. First, the large differences in free energy barrier estimates from conformationally close states speak against a locally flat ribosome energy landscape. Second, the observed range of attempt rates for ribosome parts between $(20\,\text{ns})^{-1}$ and $(40\,\text{ns})^{-1}$ (Supplementary Fig.3b in [22]) indicates that in the same simulation different ribosome parts experience different dynamics, also
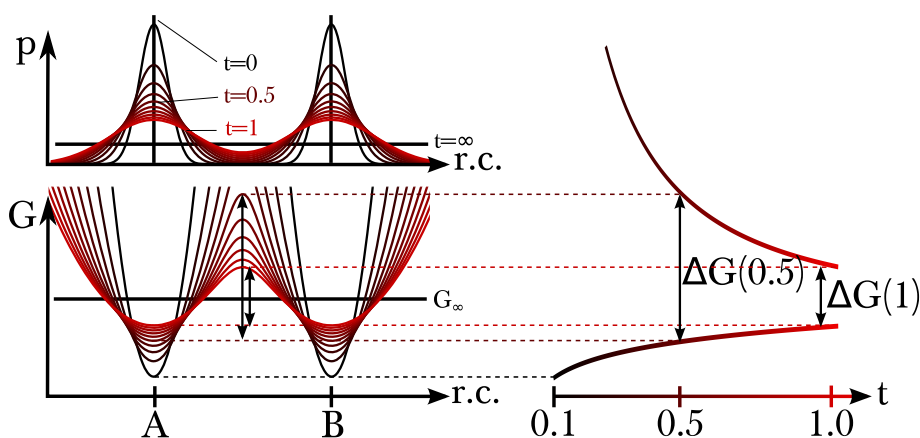
Figure 4.2: Sampling on a flat energy landscape from two distinct conformations within limited time induces a barrier between states. Sampling in a flat energy landscape (G_∞ in the chart on the bottom left) started simultaneously in states A and B at $t = 0$ (left charts). Propagation of the probability densities through the Smoluchowsky equation up to t=1 shows that the barrier between states A and B only vanishes for $t \to \infty$, even assuming perfect sampling at all time points.

contradicting a locally flat free energy landscape. Overestimation of barriers due to limited sampling is further accounted for by the global geometry factor *c*.

The different approximations that were made to obtain the free energy barrier estimates and then, via global attempt rate and geometry factor, the transition rates, introduce systematic errors. For example, the harmonic approximation yields poor results, if the underlying energy-landscape is very anharmonic. Note however, that for the relatively small structural deviations from the starting structures during our simulations, the second order Taylor-expansion of the energy landscape captures the essential features of the local free energy landscape. The quality of the estimates will be increased in future applications by taking into account the local geometry at the barrier top and in the minima via eq. (2.20).

Fig. 4.1 shows that the way state space is partitioned strongly influences the obtained free energy profile and the estimated barriers. A method we used to partition ribosome conformational space is the choice of a few reaction coordinates (by principal component analysis) as shown in Fig. 2.1. If, for example, the reaction coordinate shown in Fig. 2.1 is oriented at another angle, the separation of states A and B is less pronounced. Fig. 4.3 shows an example, where the seperation of states vanishes all together. A rotation of the reaction coordinates in Fig 4.3 by $\pi/2$, however, will separate the two states in one reaction coordinate. A method that finds these rotations is the full correlation analysis [149]. This method empirically minimizes the mutual information, thus maximizes the separation, between reaction coordinates by a series of orthogonal transformations of the reaction coordinates obtained by
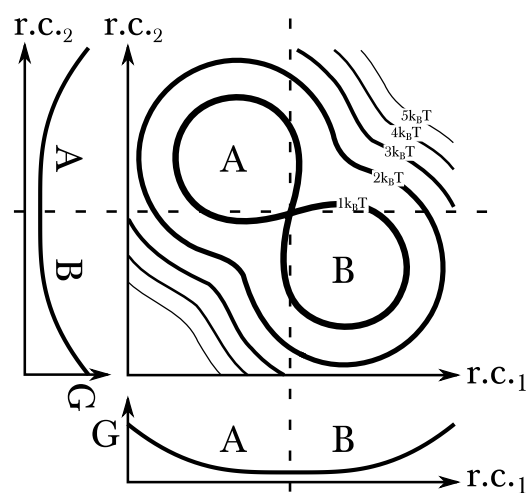
Figure 4.3: A free energy barrier in a two dimensional energy landscape disappears upon projection on the individual reaction coordinates. Two states A and B are well separated by an energy barrier in the two-dimensional product space (top right) of reaction coordinate one and two (r.c. 1 and r.c.2), but have no barrier in the individual reaction coordinates. In consequence, the inclusion of multiple dimensions is necessary for good free energy barrier estimates. If r.c.1 and r.c.2 describe different ribosome parts, this shows that coupling between ribosome parts might induce free energy barriers that are invisible if the parts are treated independently.

PCA. Applied to our simulations, it will yield reaction coordinates that better pronounce free energy barriers. However, we expect that we well capture the free energy barriers impeding the movement of ribosome parts with three reaction coordinates, since the inclusion of additional reaction coordinates did not change the relative height of the estimated free energy barriers (data not shown). On the other hand, the relative free energy barrier heights did change significantly with the inclusion of a second and third reaction coordinate, indicating that a naive one- or two-dimensional harmonic approximation to the free energy landscape is insufficient.

We extracted a single attempt rate for all transitions from our ribosome simulations by comparing the frequency of succesfull attempts to overcome energy barriers in our simulations to the estimated height of these barriers. Longer simulations would also see attempts to overcome larger barriers, and in turn lead to a lower attempt rate estimate, thus the single attempt rate we observe is coupled to the time-scale of the simulation and to the height of the free energy barriers we analyze. Dropping the assumption of one single attempt rate and instead using subsets of translocation intermediates to an attempt rate for each subset of states might provide a tool to uncover special features of the free energy landscape of a limited subset of translocation intermediates, though more extended sampling of ribosome configurations will be needed for such an approach.

In modeling the transition rates we did not consider the following three effects. First, we did not analyze other 50S subunit movements than the L1-stalk. Due to the very similar conformation of the 50S subunit, L1-stalk excluded, throughout all tRNA translocation states, we did not expect a significant contribution to the overal ribosome movement. Second, for the ribosome to transition from from one translocation intermediate state to another, the concerted motion of all its parts will be required. We assumed, that the slowest of these motions dominates the transition. Especially if the time-scales for the individual transition of the slowest ribosome parts is similar, this will affect the overall ribosome transition rate even if these movements are energetically uncoupled. We did not consider this second effect in our analysis, because we expect the error in the estimates of the individual rates to be larger than this "concerted motion" effect. Third, as sketched in Fig. 4.3, the transitions might be hindered by larger energy barriers than estimated due to energetic coupling between different ribosome parts, which we did not yet cover in our analysis.

To overcome high barriers hindering transitions and reveal the pathways that connect the minima on the free energy landscape, sampling might be enhanced at the cost of restricting the system to chosen reaction coordinates. One of these methods is umbrella sampling, which we already employed to yield the free energy profile of the tRNA L1-stalk interaction (Fig.1f in [22]). Possible other approaches are temperature replica exchange simulations or simulations where the cryo-EM density of another state is used as a biasing potential. Our harmonic approximation in the free energy landscape also estimates the location of the free energy barrier (eq. 2.18) and thus yields a transition state approximation we did not employ so far.

The transition rate estimates might help to perform meta-dynamics on other sparsely sampled energy landscapes, e.g. the energy landscape of protein folding and also to extract local diffusion properties in energy landscapes [150] from local attempt rate estimates.

4.6 Interaction enthalpies and contact frequencies

Interaction enthalpies have been described on a per-residue basis between tRNA and the large subunit proteins L1, L5, and L16 after filtering relevant residue-nucleotide interactions according to contact frequencies. Other interactions will also drive tRNA translocation, e.g. the interactions between the large and the small ribosomal subunit. The analysis of these interactions went beyond the scope of the publication [22]. Preliminary analysis by Bock et al. showed that, similar to tRNA translocation, 30S body and head rotation is made efficient through levelling the interaction enthalpy [151]. Though umbrella sampling and interaction enthalpy analysis revealed that the L1-stalk pulls the tRNA, it remains unclear what drives the L1-stalk.

The contact frequencies between tRNA and ribosomal proteins L1, L5, and L16 from our simulations (Supplementary Tables 4a-4t in [22]) reflect the strength of residue-nucleotide interactions per translocation intermediate state. Thus they directly suggest residues for mutagenesis that might stabilize or destabilize specific translocation intermediates.

4.7 Fast contacts search algorithm and implementation

Though initially developed for finding contacts between ribosome parts, the implementation of the fast contacts search algorithm addresses the issue of the scaling of data analysis algorithms in the simulation field. Extracting meaningful information from simulation data is as important as the data generation itself, but most efforts so far have been spent on increasing the efficiency of sampling algorithms, whereas data analysis tools scale poorly. We implemented an algorithm that scales approximately on the order of $n \log n$, where n is this particle number. This scaling now is similar to the MD algorithms [148].

Yet, algorithms for finding atoms in near vicinity to each other were previously employed in molecular dynamics simulations. One example is the "neighbor-list" approach used in MD simulations, where a contact search is made efficient through keeping a list of neighbours from previous simulation frames. We did not use this approach explicitly in our algorithm, because the frames we analyzes are too uncorrelated to efficiently employ the neighbor-information from previously analyzed frames. However, by initially reusing the atom sorting from previous frames, we implicitly use this neighbor information in our implementation.

Another frequently used approach is a "grid-search" that assignes atoms to grid-cells, or, even more efficiently, to nodes of a balanced octree and reduces

computational cost by calculating distances only for atoms that were assigned to related grid cells or nodes, respectively. Such an algorithm is very efficient for multiple distance queries, but not used in our approach, because of the large overhead of initially creating the grid or octree structure. Our algorithm implicitly uses the advantages of the octree approach. It is implicit in the so-called bboxes and their repeated splitting (cf. Fig.1 in [21]). In contrast to a standard octree approach, only the nodes of the octree structure (called bboxes in [21]) are created that contribute to find contacts within the given distance.

A theoretical proof for the lower- and upper bounds of scaling and data storage needs for the fast contact search went beyond the scope of our work, but exhaustive empirical testing did not show significant deviation from the assumed $n \log n$ scaling (Fig. 2 in [21] and scaling of intersubunit contacts search between differently sized parts of ribosome subunits, data not shown). The scaling behaviour of our fast contact search appears worse than the reporting time for other range searches, $\log n$. However, the pre-processing time is the crucial factor for our applications, because we query each data set only once. Thus we match the expected $n \log n$ scaling behaviour of other reported efficient algorithms if the necessary pre-processing is considered [152].

4.8 Applications of fast contacts search

On the ribosome, during A-site decoding, small configurational changes are transmitted to trigger a response in distant sites (cf. section 1.1). Further intersubunit interactions might serve as a "dial control" of tRNA translocation. An analysis of contact patch changes that is now made possible with our implementation of efficient contacts search, can serve as a tool to explain how these configurational changes are transmitted through the ribosome.

Our implementation of a fast contact search algorithm in g_contacts, allowed to identify the contacts that are involved in allosteric transitions in hemoglobin [20]. Further it has been applied to study the impact of point mutations to EF-Tu, and as a tool to identify flipped lipid head groups in coarse-grain simulations (Leonov and Bubnis, personal communication).

Another application of our fast contact search implementation lies in the characterization of intrinsically disordered proteins. These proteins do not have a defined fold, but explore a large variety of conformations. Standard methods to characterize ensemble data fail in this case, e.g. the root mean square deviation measure with respect to a reference structure, because a well defined reference structure is missing. Also contact maps that describe the pair-wise interactions of protein residues are not applicable due to the undefined fold and thus wide variety of residue-residue interactions. This problem is solved by the extension of this contact map to a contact frequency map that reflects the averaged contact map over a large ensemble of conformations. Our implementation of fast contacts search now allows calculation of this contact frequency map, which was previously prohibitively expensive.

Further the conatct frequency map provides a similarity measure comparable to the widely used RMSD, but which does not rely on a reference structure.

Here the deviation of the contact frequency map averaged over a simulation window from the contact frequency map of the whole simulation is used.

4.9   Follow-up projects

With our ribosome simulations and the respective analysis tools we laid the foundations for a series of follow up studies that we will present now.

Translation comprises a hierarchy of chemical processes (Fig. 1.3). The methods developed in this work, the ribosome model, cryo-EM refinement routines, the fast contact search, and our method to characterize the ribosome free energy landscape, will facilitate the study of the processes adjacent to elongation: A-site decoding, peptide bond formation and termination, and, further up the tier, ribosomal initiation, termination, and recycling.

The prerequisite for these further studies is accompanying structural data to refine and validate our all-atom ribosome model. With this data on hand, the same methods to characterize the ribosome free energy landscape and the interactions of its parts, will indeed aid to advance to the "post-structural" era of ribosome research also in other processes than translocation [126].

This structural data is already on hand for a biochemical complex that occurs during selenocysteine (Sec) incoorporation in peptide elongation. Selenocystein is a non-canonical amino-acid that is essential for a variety of prokaryotes, archea and eukaryotes. It is part of the catalytic center of selenoproteins that catalyses redox-reactions. The mechanism of selenocystein incorporation into proteins is unusual, because the tRNA that delivers Sec to the ribosome reads a stop codon. This reading is only performed if the downstream mRNA exhibits a special fold, which is triggered through the mRNA selenocystein insertion (SECIS) sequence. This special fold triggers selB biding, which in turn aids Sec-tRNA^{Sec} accommodation to the ribosme similar to EF-Tu. Due to the low yield of the ribosomal elongation factor selB promoted accommodation of Sec-tRNA^{Sec}, structure determination has been difficult. Now, a high-resolution cryo-EM density will allow to model the ribosome in atomic detail in complex with SelB (Fischer et al., personal communication). From that starting point, we aim at understanding how the SECIS fold is recognised by selB and how selB aids selenocystein incorporation through the ensemble information that is provided by MD simulations of this complex. In contrast to the approach of this thesis, where we extracted general features of translocation from a special setup, we aim here at the selenocystein specific ribosome interactions. The current problems in refining high-resolution cryo-EM densities of the selB-ribosome complex might be overcome by the Hamiltonian replica exchange protocol described above.

Translocation in eukaryotes is yet less understood than in prokaryotes. The novel high-resolution crystal structures of eukaryotic ribosomes [131] allow MD simulations on the resolved states, as performed by Sanbonmatsu et al. (conference proceedings), though the structural detail of translocation intermediate states and detailed kinetic information of tRNA translocation in eukaryotes is still lacking and renders a similar approach to ours challenging at the time [132].

An outstanding challenge that is closest related to the presented work is to understand the role of EF-G in translocation. The transition rate estimates in our work already provide a hypothetical mechanism for EF-G (Fig.1d in [22]). By binding to the pre3 state EF-G binding might block the excursions of the tRNA^{Val} as well as the body from the pre3 and pre4 to the pre1 state conformations (cf. purple and red lines connecting pre3, pre4 and pre1b in Fig. 1d [22]). Until recently, the lack of structural data prohibited a computational study of EF-G dynamics on the ribosome. Now, new crystal structures of EF-G bound to the ribosome have been solved and provide atomic models for EF-G binding to the ribosome [136, 137, 138]. Here, cryo-EM structures of further translocation intermediates with EF-G would provide the necessary additional information to follow the same approach as in this thesis.

5. Acknowledgements

I would like to thank Prof. Helmut Grubmüller for his honest support and a fair share of his time. Lars Bock and Andrea Vaiana, for their patience with me in a truly colloborative project. Niels Fischer for his time to explain and discuss his cryo-EM data. Holger Stark and Marina Rodnina for their aid in conceiving the manuscript for [22] and insight into the current state-of-the art in cryo-EM and the ribosome field. Wolfgang Wintermeyer for his advice on the manuscript for [22]. Martin Fechner and Ansgar Eszterman for system administration beyond expected. Eveline Heineman for efficient and pragmatic organisation of the department. Petra Kellers for thorough manuscript reading as well as teaching of style and writing. Benjamin von Ardenne, Sarah Rauscher and Timo Graen for reading the thesis manuscript. Vytautas Gapšys, Bela Voss, Greg Bubnis and Martin Hoefling for scientific discussions. Martin Stumpe and Martin Vesper for a good start in the department and practical advise. I thank Antje Erdmann, Frauke Bergmann and Michaela Böttcher for their support through the GGNB program and patience in answering my questions.

Eventually, I shall acknowledge funding from the Excellence Stipends of the International Max-Planck Research School as well as the DFG Excellence Cluster 171 (Nanoscale Microscopy and Molecular Physiology of the Brain).

6. Curriculum Vitae

Personal Data

Address	Lange-Geismar-Strasse 2
	37073 Göttingen
Email	cblau@gwdg.de
Place of Birth	Merseburg
Nationality	German

Education

2010-date	PhD student at MPI-BPC Göttingen
Jan 2011 – Jun 2013	Excellence Stipend of the International Max-Planck Research School
2009	Diploma in Physics at Georg-August-University Göttingen Effect of posttranscriptional modifications and Mg^{2+} ions on tRNA structure and flexibility
Sep 2003	Internship at Institute for Low Temperature Plasma Physics Greifswald
2003	Abitur Georg Cantor Gymnasium Halle/Saale

Work Experience

Nov 2006 – Feb 2007	Logistics Admin InBev Praha
Sep 2006 – Nov 2006	Communication Specialist (Quality Assistance) Icon Communication Centres Praha

Bibliography

[1] F Crick. Central Dogma Of Molecular Biology, *Nature*, **227**(5258), 561–, 1970.

[2] A. J Hamilton and D. C Baulcombe. A species of small antisense RNA in posttranscriptional gene silencing in plants., *Science*, **286**(5441), 950–2, 1999.

[3] A Bird. Perceptions of epigenetics., *Nature*, **447**(7143), 396–8, 2007.

[4] S Prabakaran, G Lippens, H Steen, and J Gunawardena. Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding., *Wiley interdisciplinary reviews. Systems biology and medicine*, **4**(6), 565–83, 2012.

[5] H Grosjean. DNA and RNA modification enzymes: structure, mechanism, function and evolution. Landes Bioscience Austin, TX, 2009.

[6] N Fischer, A Konevega, W Wintermeyer, M Rodnina, and H Stark. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy, *Nature*, **466**(7304), 329–333, 2010.

[7] M Selmer, C. M Dunham, F. V Murphy, A Weixlbaumer, S Petry, A. C Kelley, J. R Weir, and V Ramakrishnan. Structure of the 70S ribosome complexed with mRNA and tRNA, *Science*, **313**(5795), 1935–1942, 2006.

[8] W DeLano. *The PyMOL Molecular Graphics System*, 2002.

[9] T. R Cech. Structural biology - The ribosome is a ribozyme, *Science*, **289**(5481), 878–879, 2000.

[10] M. V Rodnina, M Beringer, and W Wintermeyer. How ribosomes make peptide bonds, *Trends in Biochemical Sciences*, **32**, 20–26, 2006.

[11] J Chen, A Tsai, S O'Leary, A Petrov, and J Puglisi. Unraveling the dynamics of ribosome translocation., *Current Opinion in Structural Biology*, **22**(6), 804–14, 2012.

[12] A Spirin. The ribosome as a conveying thermal ratchet machine., *Journal of biological chemistry*, **284**(32), 21103–19, 2009.

[13] J Frank and R Gonzalez. Structure and Dynamics of a Processive Brownian Motor: The Translating Ribosome, *Annual Review of Biochemistry*, **79**(1), 381–412, 2010.

[14] H Frauenfelder, S. G Sligar, and P. G Wolynes. The energy landscapes and motions of proteins., *Science*, **254**(5038), 1598–603, 1991.

[15] J Munro, K Sanbonmatsu, , and S Blanchard. Navigating the ribosome's metastable energy landscape, *Trends in Biochemical Sciences*, **34**(8), 390–400, 2009.

[16] J Noeske and J. H. D Cate. Structural basis for protein synthesis: snapshots of the ribosome in motion., *Current Opinion in Structural Biology*, **22**(6), 743–9, 2012.

[17] W Zhang, J. A Dunkle, and J. H. D Cate. Structures of the Ribosome in Intermediate States of Ratcheting, *Science*, **325**(5943), 1014–1017, 2009.

[18] P. C Whitford, R. B Altman, P Geggier, D. S Terry, and J. B Munro. in *Ribosomes : Structure, Function, and Dynamics*, (Springer Vienna, VIENNA). pages 303–319. 2011.

[19] I Wohlgemuth, C Pohl, J Mittelstaet, A Konevega, and M Rodnina. Evolutionary optimization of speed and accuracy of decoding on the ribosome., *Philosophical transactions - Royal Society. Biological sciences*, **366**(1580), 2979–86, 2011.

[20] M Vesper and B de Groot. Collective dynamics underlying allosteric transitions in hemoglobin., *PLoS Computational Biology*, **9**(9), e1003232, 2013.

[21] C Blau and H Grubmuller. g_contacts: Fast contact search in bio-molecular ensemble data, *Computer Physics Communications*, **184**(12), 2856 – 2859, 2013.

[22] L Bock, C Blau, G Schroeder, I Davydov, N Fischer, G Schroder, H Stark, M Rodnina, A Vaiana, and H Grubmuller. Energy barriers and driving forces in tRNA translocation through the ribosome., *Nature structural & molecular biology*, **20**(12), 1390–1396, 2013.

[23] B Laursen, H Sorensen, H Sorensen, K Mortensen, and H. S Petersen. Initiation of protein synthesis in bacteria., *Microbiology and molecular biology reviews*, **69**(1), 101–23, 2005.

[24] A Myasnikov, A Simonetti, S Marzi, and B Klaholz. Structure-function insights into prokaryotic and eukaryotic translation initiation., *Current Opinion in Structural Biology*, **19**(3), 300–9, 2009.

[25] P Milon and M Rodnina. Kinetic control of translation initiation in bacteria., *Critical reviews in biochemistry and molecular biology*, **47**(4), 334–48, 2012.

[26] B Cooperman, Y Goldman, C Chen, I Farrell, and J Kaur. in *Ribosomes : Structure, Function, and Dynamics*, (Springer Vienna, VIENNA). pages 339–348. 2011.

[27] E. K. Y Leung, N Suslov, N Tuttle, R Sengupta, and J Piccirilli. The mechanism of peptidyl transfer catalysis by the ribosome, volume 80 of *Annual Review of Biochemistry*. Stanford University Press, Palo Alto, Calif., 2011.

[28] M Rodnina, A Rodnina, V Savelsbergh, W Katunin, and Wintermeyer. Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome, *Nature*, **385**(6611), 37–41, 1997.

[29] L. P Gavrilova, O. E Kostiashkina, V. E Koteliansky, N. M Rutkevitch, and A. S Spirin. Factor-free (non-enzymic) and factor-dependent systems of translation of polyuridylic acid by Escherichia coli ribosomes., *Journal of Molecular Biology*, **101**(4), 537–52, 1976.

[30] S Shoji, S Walker, and K Fredrick. Reverse translocation of tRNA in the ribosome., *Molecular Cell*, **24**(6), 931–942, 2006.

[31] B Klaholz. Molecular recognition and catalysis in translation termination complexes., *Trends in Biochemical Sciences*, **36**(5), 282–92, 2011.

[32] J. A Dunkle, L Wang, M. B Feldman, A Pulk, V. B Chen, G. J Kapral, J Noeske, J. S Richardson, S. C Blanchard, and J. H. D Cate. Structures of the Bacterial Ribosome in Classical and Hybrid States of tRNA Binding, *Science*, **332**(6032), 981–984, 2011.

[33] J Panecka and J Trylska. in *Theroy and applications in computational chemistry: the first decade of the second millenium*, (American Institute of Physics, New York). volume 1456 of *AIP Conference Proceedings*. pages 207–214. 2012.

[34] J Aqvist, C Lind, J Sund, and G Wallin. Bridging the gap between ribosome structure and biochemistry by mechanistic computations., *Current Opinion in Structural Biology*, **22**(6), 815–23, 2012.

[35] M Rodnina. The ribosome as a versatile catalyst: reactions at the peptidyl transferase center., *Current Opinion in Structural Biology*, **23**(4), 595–602, 2013.

[36] M Rodnina. in *Ribosomes : Structure, Function, and Dynamics*, (Springer Vienna, VIENNA). pages 199–212. 2011.

[37] H Noller. How does the ribosome sense a cognate tRNA?, *Journal of Molecular Biology*, **425**(20), 3776–7, 2013.

[38] P Moore. Ribosomal ambiguity made less ambiguous., *Proceedings of the National Academy of Sciences of the United States of America*, **110**(24), 9627–8, 2013.

[39] N Reynolds, B Lazazzera, and M Ibba. Cellular mechanisms that control mistranslation., *Nature Reviews Microbiology*, **8**(12), 849–56, 2010.

[40] I Wohlgemuth, C Pohl, and M Rodnina. Optimization of speed and accuracy of decoding in translation., *EMBO Journal*, **29**(21), 3701–9, 2010.

[41] L Cochella and R Green. An active role for tRNA in decoding beyond codon:anticodon pairing., *Science*, **308**(5725), 1178–80, 2005.

[42] M Rodnina, K Gromadski, U Kothe, and H.-J Wieden. Recognition and selection of tRNA in translation., *FEBS Letters*, **579**(4), 938–42, 2005.

[43] R. A Marshall, C Aitken, M Dorywalska, and J Puglisi. Translation at the Single-Molecule Level, *Annual Review of Biochemistry*, **77**, 177–203, 2008.

[44] J Eargle, A Black, A Sethi, L Trabuco, and Z. L Schulten. Dynamics of Recognition between tRNA and elongation factor Tu., *Journal of Molecular Biology*, **377**(5), 1382–405, 2008.

[45] J Aqvist, M Leijonmarck, O Tapia, and W. F van Gunsteren. A molecular dynamics study of the C-terminal fragment of the L7/L12 ribosomal protein. Secondary structure motion in a 150 picosecond trajectory., *Journal of Molecular Biology*, **183**(3), 461–77, 1985.

[46] Dagget and M Levitt. A molecular dynamics simulation of the c-terminal fragment of the L7/L12 ribosomal-protein in solution, *Chemical Physics*, **158**(2-3), 501–512, 1991.

[47] M Diaconu, F Schlunzen, U Kothe, N Fischer, J Harms, A Tonevitsky, H Stark, M Rodnina, and M Wahl. Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation., *Cell*, **121**(7), 991–1004, 2005.

[48] H Demirci, F Murphy, E Murphy, S Gregory, A Dahlberg, and G Jogl. A structural basis for streptomycin-induced misreading of the genetic code., *Nature communications*, **4**, 1355, 2013.

[49] A Tsai, S Uemura, M Johansson, E Puglisi, R. A Marshall, C Aitken, J Korlach, M Ehrenberg, and J Puglisi. The impact of aminoglycosides on the dynamics of translation elongation., *Cell Reports*, **3**(2), 497–508, 2013.

[50] A. C Vaiana, E Westhof, and P Auffinger. A molecular dynamics simulation study of an aminoglycoside/A-site RNA complex: conformational and hydration patterns., *Biochimie*, **88**(8), 1061–73, 2006.

[51] K. Y Sanbonmatsu. Energy landscape of the ribosomal decoding center., *Biochimie*, **88**(8), 1053–9, 2006.

[52] A Vaiana and K Sanbonmatsu. Stochastic gating and drug-ribosome interactions., *Journal of Molecular Biology*, **386**(3), 648–61, 2009.

[53] T.-H Lee, S Blanchard, H Kim, J Puglisi, and S Chu. The role of fluctuations in tRNA selection by the ribosome., *Proceedings of the National Academy of Sciences of the United States of America*, **104**(34), 13661–5, 2007.

[54] P Khade, X Shi, and S Joseph. Steric complementarity in the decoding center is important for tRNA selection by the ribosome., *Journal of Molecular Biology*, **425**(20), 3778–89, 2013.

[55] X Agirrezabala, E Schreiner, L Trabuco, J Lei, R. O Meoz, K Schulten, R Green, and J Frank. Structural insights into cognate versus near-cognate discrimination during decoding., *EMBO Journal*, **30**(8), 1497–507, 2011.

[56] Y. P Semenkov, W Wintermeyer, and M. V Rodnina. The ”allosteric three-site model” of elongation cannot be confirmed in a well-defined ribosome system from Escherichia coli., *Proceedings of the National Academy of Sciences of the United States of America*, **93**(22), 12183–8, 1996.

[57] A Petropoulos and R Green. Further in vitro exploration fails to support the allosteric three-site model., *Journal of Biological Chemistry*, **287**(15), 11642–8, 2012.

[58] N Demeshkina, L Jenner, E Westhof, M Yusupov, and G Yusupova. A new understanding of the decoding principle on the ribosome., *Nature*, **484**(7393), 256–9, 2012.

[59] N Demeshkina, L Jenner, E Westhof, M Yusupov, and G Yusupova. New structural insights into the decoding mechanism: translation infidelity via a GůU pair with Watson-Crick geometry., *FEBS Letters*, **587**(13), 1848–57, 2013.

[60] B. L Grigorenko, M. S Shadrina, I. A Topol, J. R Collins, and A. V Nemukhin. Mechanism of the chemical step for the guanosine triphosphate (GTP) hydrolysis catalyzed by elongation factor Tu., *Biochimica et Biophysica Acta*, **1784**(12), 1908–17, 2008.

[61] A Adamczyk and A Warshel. Converting structural information into an allosteric-energy-based picture for elongation factor Tu activation by the ribosome., *Proceedings of the National Academy of Sciences of the United States of America*, **108**(24), 9827–32, 2011.

[62] A Aleksandrov and M Field. Mechanism of activation of elongation factor Tu by ribosome: catalytic histidine activates GTP by protonation., *RNA*, **19**(9), 1218–25, 2013.

[63] I Besseova, K Reblova, N Leontis, and J Sponer. Molecular dynamics simulations suggest that RNA three-way junctions can act as flexible RNA structural elements in the ribosome., *Nucleic Acids Research*, **38**(18), 6247–64, 2010.

[64] L. G Trabuco, E Villa, K Mitra, J Frank, and K Schulten. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics, *Structure*, **16**(5), 673–683, 2008.

[65] E Villa, J Sengupta, L. G Trabuco, J LeBarronc, W. T Baxterc, T. R Shaikhc, R. A Grassuccid, P Nissene, M Ehrenberg, K Schulten, and J Frank. Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis, *Proceedings of the National Academy of Sciences of the United States of America*, **106**(4), 1063–1068, 2009.

[66] R. C Thompson and P. J Stone. Proofreading of the codon-anticodon interaction on ribosomes., *Proceedings of the National Academy of Sciences of the United States of America*, **74**(1), 198–202, 1977.

[67] P Whitford, P Geggier, R Altman, S Blanchard, J Onuchic, and K Sanbonmatsu. Accommodation of aminoacyl-tRNA into the ribosome involves reversible excursions along multiple pathways, *RNA*, **16**(6), 1196–1204, 2010.

[68] T Caulfield and B Devkota. Motion of transfer RNA from the A/T state into the A-site using docking and simulations., *Proteins*, **80**(11), 2489–500, 2012.

[69] P Whitford, J Onuchic, and K Sanbonmatsu. Connecting energy landscapes with experimental rates for aminoacyl-tRNA accommodation in the ribosome., *Journal of the American Chemical Society*, **132**(38), 13170–1, 2010.

[70] H White. Coenzymes as fossils of an earlier metabolic state, *Journal of Molecular Evolution*, **7**(2), 101–104, 1976.

[71] A Sievers, M Beringer, M Rodnina, and R Wolfenden. The ribosome as an entropy trap., *Proceedings of the National Academy of Sciences of the United States of America*, **101**(21), 7897–901, 2004.

[72] S Trobro and J Aqvist. Mechanism of peptide bond synthesis on the ribosome, *PNAS*, **102**, 12395–12400, 2005.

[73] P Sharma, Y Xiang, M Kato, and A Warshel. What are the roles of substrate-assisted catalysis and proximity effects in peptide bond formation by the ribosome?, *Biochemistry*, **44**(34), 11307–14, 2005.

[74] M Rodnina and W Wintermeyer. The ribosome as a molecular machine: the mechanism of tRNA-mRNA movement in translocation., *Biochemical Society transactions*, **39**(2), 658–62, 2011.

[75] Y. P Semenkov, M. V Rodnina, and W Wintermeyer. Energetic contribution of tRNA hybrid state formation to translocation catalysis on the ribosome, *Nature structural biology*, **7**(11), 1027–1031, 2000.

[76] P. C Spiegel, D. N Ermolenko, and H. F Noller. Elongation factor G stabilizes the hybrid-state conformation of the 70S ribosome, *RNA*, **13**(9), 1473–1482, 2007.

[77] D Moazed and H. F Noller. Intermediate states in the movement of transfer RNA in the ribosome., *Nature*, **342**(6246), 142–8, 1989.

[78] J. B Munro, R. B Altman, N O'Connor, and S. C Blanchard. Identification of two distinct hybrid state intermediates on the ribosome, *Molecular Cell*, **25**(4), 505–517, 2007.

[79] S. E Walker, S Shoji, D Pan, B. S Cooperman, and K Fredrick. Role of hybrid tRNA-binding states in ribosomal translocation, *Proceedings of the National Academy of Sciences*, **105**(27), 9192–9197, 2008.

[80] H Kim, J Yeom, H Ha, J Kim, and K Lee. Functional analysis of the residues C770 and G771 of E coli 16S rRNA implicated in forming the intersubunit bridge B2c of the ribosome, *Journal of microbiology and biotechnology*, **17**(7), 1204–1207, 2007.

[81] X Agirrezabala, J Lei, J Brunelle, R Ortiz-Meoz, R Green, and J Frank. Visualization of the Hybrid State of tRNA Binding Promoted by Spontaneous Ratcheting of the Ribosome, *Molecular Cell*, **32**(2), 190–197, 2008.

[82] P Julian, A Konevega, S Scheres, M Lazaro, D Gil, and W Wintermeyer. Structure of ratcheted ribosomes with tRNAs in hybrid states, *Proceedings of the National Academy of Sciences of the United States of America*, **105**(44), 16924–16927, 2008.

[83] W Li and J Frank. Transfer RNA in the hybrid P/E state: Correlating molecular dynamics simulations with cryo-EM data, *Proceedings of the National Academy of Sciences of the United States of America*, **104**(42), 16540–16545, 2007.

[84] J Frank and R. K Agrawal. A ratchet-like inter-subunit reorganization of the ribosome during translocation., *Nature*, **406**(6793), 318–22, 2000.

[85] A Ratje, J Loerke, A Mikolajka, M Brunner, P Hildebrand, and A Starosta. Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites, *Nature*, **468**(7324), 713–U143, 2010.

[86] P Cornish, D Ermolenko, H Noller, T Ha, and D Ermolenko. Spontaneous intersubunit rotation in single ribosomes., *Molecular Cell*, **30**(5), 578–588, 2008.

[87] J Fei, A Richard, J Bronson, and R Gonzalez. Transfer RNA-mediated regulation of ribosome dynamics during protein synthesis, *Nature structural & molecular biology*, **18**(9), 1043–U106, 2011.

[88] L Trabuco, E Schreiner, J Eargle, P Cornish, and T Ha. The Role of L1 Stalk-tRNA Interaction in the Ribosome Elongation Cycle, *Journal of Molecular Biology*, **402**(4), 741–760, 2010.

[89] Z Zhang, K Sanbonmatsu, and G Voth. Key intermolecular interactions in the E. coli 70S ribosome revealed by coarse-grained analysis., *Journal of the American Chemical Society*, **133**(42), 16828–38, 2011.

[90] P. C Whitford, S. C Blanchard, J. H. D Cate, and K. Y Sanbonmatsu. Connecting the Kinetics and Energy Landscape of tRNA Translocation on the Ribosome, *PLoS Computational Biology*, **9**(3), e1003003, 03 2013.

[91] R Zwanzig. Diffusion in a rough potential., *Proceedings of the National Academy of Sciences of the United States of America*, **85**(7), 2029–30, 1988.

[92] I Fernandez, C Ng, A Kelley, G Wu, Y.-T Yu, and V Ramakrishnan. Unusual base pairing during the decoding of a stop codon by the ribosome., *Nature*, **500**(7460), 107–10, 2013.

[93] J Sund, M Ander, and J Aqvist. Principles of stop-codon reading on the ribosome., *Nature*, **465**(7300), 947–50, 2010.

[94] M Laurberg, H Asahara, A Korostelev, J Zhu, S Trakhanov, and H Noller. Structural basis for translation termination on the 70S ribosome., *Nature*, **454**(7206), 852–7, 2008.

[95] A Weixlbaumer, H Jin, C Neubauer, R Voorhees, S Petry, A Kelley, and V Ramakrishnan. Insights into translational termination from the structure of RF2 bound to the ribosome., *Science*, **322**(5903), 953–6, 2008.

[96] A Korostelev, H Asahara, L Lancaster, M Laurberg, A Hirschi, J Zhu, S Trakhanov, W Scott, and H Noller. Crystal structure of a translation termination complex formed with release factor RF2., *Proceedings of the National Academy of Sciences of the United States of America*, **105**(50), 19684–9, 2008.

[97] E Jaynes. Information Theory and Statistical Mechanics, *Physical review*, **106**(4), 620–630, 1957.

[98] J Bryngelson and P Wolynes. Intermediates and Barrier Crossing in a Random Energy-Model (with Applications to Protein Folding), *Journal of physical chemistry*, **93**(19), 6902–6915, 1989.

[99] P Hanggi, P Talkner, and M Borkovec. Reaction-Rate Theory - 50 Years after Kramers, *Reviews of modern physics*, **62**(2), 251–341, 1990.

[100] V Kenkre, E Montroll, and M Shlesinger. Generalized master equations for continuous-time random walks, *Journal of Statistical Physics*, **9**(1), 45–50, 1973.

[101] H Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions, *Physica*, **7**, 284–304, 1940.

[102] H Brinkman. Brownian Motion in a Field of Force and the Diffusion Theory of Chemical Reactions, *Physica*, **22**(U), 29–34, 1956.

[103] M Von Smoluchowski. Uber Brownsche Molekularbewegung unter Einwirkung ausserer Krafte und deren Zusammenhang mit der verallgemeinerten diffusionsgleichung, *Ann. Phys*, **48**, 1103, 1915.

[104] S Chandrasekhar. Stochastic problems in physics and astronomy, *Revs Modern Physics*, **15**(1), 2–89, 1943.

[105] R Landauer and J Swanson. Frequency Factors in Thermally Activated Process, *physical review*, **121**(6), 1668–&, 1961.

[106] J Langer. Theory of Nucleation Rates, *Physical Review Letters*, **21**(14), 973–&, 1968.

[107] P Langevin. Sur la théorie du mouvement brownien, *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, **146**, 530–533, 1908.

[108] W. E van Gunsteren and H. J. C Berendsen. Computer Simulation Of Molecular Dynamics Methodology Applications and Perspectives In Chemistry, *Angewandte Chemie. International edition in English*, **29**(9), 992–1023, 1990.

[109] B. J Alder and T. E Wainwright. Studies in Molecular Dynamics. I. General Method, *The Journal of Chemical Physics*, **31**(2), 459–466, 1959.

[110] M Born and R Oppenheimer. Zur Quantentheorie der Molekeln, *Annalen der Physik*, **389**(20), 457–484, 1927.

[111] E Hairer, C Lubich, and G Wanner. Geometric numerical integration illustrated by the Stormer-Verlet method, *Acta numerica*, **12**, 399–450, 2003.

[112] P Auffinger and E Westhof. RNA solvation: A molecular dynamics simulation perspective, *Biopolymers*, **56**(4), 266–274, 2000.

[113] J Norberg and L Nilsson. Molecular dynamics applied to nucleic acids, *Accounts of Chemical Research*, **35**(6), 465–472, 2002.

[114] T Darden, D York, and L Pedersen. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems, *Journal of Chemical Physics*, **98**, 10089–10092, 1993.

[115] S Nose and M. L Klein. Constant Pressure Molecular-dynamics for Molecular-systems, *Molecular Physics*, **50**(5), 1055–1076, 1983.

[116] W Hoover. Canonical Dynamics - Equilibrium Phase-Space Distributions, *Physical review. A, Atomic, molecular, and optical physics*, **31**(3), 1695–1697, 1985.

[117] H. J. C Berendsen, J. P. M Postma, W. F van Gunsteren, A DiNola, and J. R Haak. Molecular dynamics with coupling to an external bath, *Journal of Chemical Physics*, **81**(8), 3684–3690, October 1984.

[118] F. C Bernstein, T. F Koetzle, G. J. B Williams, E. F Meyer, M. D Brice, J. R Rodgers, O Kennard, T Shimanouchi, and M Tasumi. Protein Data Bank - Computer-Based Archival File For Macromolecular Structures, *Journal of Molecular Biology*, **112**(3), 535–542, 1977.

[119] D Ermak and J McCammon. Brownian Dynamics with Hydrodynamic Interactions, *the Journal of chemical physics*, **69**(4), 1352–1360, 1978.

[120] J Dubochet, M Adrian, J Lepault, J. J Chang, J. C Homo, A. W McDowall, and P Schultz. Cryo-electron microscopy of vitrified specimens., *Quarterly reviews of biophysics*, **21**(2), 129–228, 1988.

[121] B Sander, H Stark, and M. M Golas. Corrim-based alignment for improved speed in single-particle image processing., *Journal of structural biology*, **143**(3), 219–28, 2003.

[122] G. F Schroder, A. T Brunger, and M Levitt. Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution, *Structure*, **15**(12), 1630–1641, 2007.

[123] K Pearson. On lines and planes of closest fit to systems of points in space., *Philosophical Magazine*, **2**(7-12), 559–572, 1901.

[124] K Halvorsen, M Lesser, and A Lundberg. A new method for estimating the axis of rotation and the center of rotation, *Journal of Biomechanics*, **32**(11), 1221 – 1227, 1999.

[125] C Gramkow. On Averaging Rotations, *International Journal of Computer Vision*, **42**, 7–16, 2001.

[126] P Moore. How Should We Think About the Ribosome?, *Annual review of biophysics*, **41**(1), 1–19, 2012.

[127] A. L Konevega, N Fischer, Y. P Semenkov, H Stark, W Wintermeyer, and M. V Rodnina. Spontaneous reverse movement of mRNA-bound tRNA through the ribosome, *Nature structural & molecular biology*, **14**(4), 318–324, 2007.

[128] W. F Anderson. The effect of tRNA concentration on the rate of protein synthesis., *Proceedings of the National Academy of Sciences of the United States of America*, **62**(2), 566–73, 1969.

[129] J. R Warner, A Rich, and P. M Knopf. A multiple ribosomal structure in protein synthesis., *Proceedings of the National Academy of Sciences of the United States of America*, **49**(1), 122–9, 1963.

[130] S Proshkin, A. R Rahmouni, A Mironov, and E Nudler. Cooperation between translating ribosomes and RNA polymerase in transcription elongation., *Science*, **328**(5977), 504–8, 2010.

[131] A Ben-Shem, N de Loubresse, S Melnikov, L Jenner, G Yusupova, and M Yusupov. The structure of the eukaryotic ribosome at 3.0 Å resolution., *Science*, **334**(6062), 1524–9, 2011.

[132] D Wilson. The Structure and Function of the Eukaryotic Ribosome, *Cold Spring Harbor perspectives in biology*, **4**(5), 2012.

[133] H Gingold and Y Pilpel. Determinants of translation efficiency and accuracy., *Molecular Systems Biology*, **7**, 481, 2011.

[134] D. E Draper. A guide to ions and RNA structure, *RNA*, **10**(3), 335–343, 2004.

[135] Y Gao, M Selmer, C Dunham, A Weixlbaumer, A Kelley, and V Ramakrishnan. The Structure of the Ribosome with Elongation Factor G Trapped in the Posttranslocational State, *Science*, **326**(5953), 694–699, 2009.

[136] A Pulk and J. H. D Cate. Control of ribosomal subunit rotation by elongation factor G., *Science*, **340**(6140), 1235970–U77, 2013.

[137] D Tourigny, I Fernandez, A Kelley, V Ramakrishnan, and I FernÃąndez. Elongation factor G bound to the ribosome in an intermediate state of translocation., *Science*, **340**(6140), 1235490–U148, 2013.

[138] J Zhou, L Lancaster, J Donohue, and H Noller. Crystal structures of EF-G-ribosome complexes trapped in intermediate states of translocation., *Science*, **340**(6140), 1236086–U69, 2013.

[139] A Ahmed and F Tama. Consensus among multiple approaches as a reliability measure for flexible fitting into cryo-EM data., *Journal of structural biology*, **182**(2), 67–77, 2013.

[140] M Orzechowski and F Tama. Flexible Fitting of High-Resolution X-Ray Structures into Cryoelectron Microscopy Maps Using Biased Molecular Dynamics Simulations, *Biophysical Journal*, **95**(12), 5692–5705, 2008.

[141] J Pitera and J Chodera. On the Use of Experimental Observations to Bias Simulated Ensembles, *Journal of Chemical Theory and Computation*, **8**(10), 3445–3451, 2012.

[142] K Hukushima and K Nemoto. Exchange Monte Carlo method and application to spin glass simulations, *Journal Of The Physical Society Of Japan*, **65**(6), 1604–1608, 1996.

[143] Y Hashem and P Auffinger. A short guide for molecular dynamics simulations of RNA systems, *Methods*, **47**(3), 187–197, 2009.

[144] U Essmann, L Perera, M. L Berkowitz, T Darden, H Lee, and L. G Pedersen. A smooth particle mesh Ewald method, *Journal of Chemical Physics*, **103**(19), 8577–8593, 1995.

[145] I Joung. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations, *The journal of physical chemistry. B*, **112**(30), 9020–9041, 2008.

[146] K Sanbonmatsu, S Joseph, and C Tung. Simulating movement of tRNA into the ribosome during decoding, *Proceedings of The National Academy of Sciences of the United States*, **102**(44), 15854–15859, Nov 1 2005.

[147] V Hornak, R Abel, A Okur, B Strockbine, A Roitberg, and C Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins: Structure, Function, and Bioinformatics*, **65**(3), 712–725, 2006.

[148] B Hess, C Kutzner, D van der Spoel, and E Lindahl. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, *Journal of Chemical Theory and Computation*, **4**(3), 435–447, 2008.

[149] O. F Lange and H Grubmueller. Full correlation analysis of conformational protein dynamics, *Proteins*, **70**(4), 1294–1312, 2008.

[150] B Wang, J Kuo, S Bae, and S Granick. When Brownian diffusion is not Gaussian., *Nature Materials*, **11**(6), 481–5, 2012.

[151] C Blau, L Bock, G Schroeder, N Fischer, H Stark, A Vaiana, and H Grubmuller. Contacts Between Ribosome Parts Refined by Molecular Dynamics Simulations, *Biophysical journal*, **102**(3), 69A–69A, 2012.

[152] J Erickson and P Agarwal. Geometric range searching and its relatives. Advances in Discrete and Computational Geometry, *Contemporary Mathematics*, **223**, 1–56, 1999.