

Modeling Translation in Protein Synthesis with TASEP: A Tutorial and Recent Developments

R.K.P. Zia · J.J. Dong · B. Schmittmann

Received: 23 December 2010 / Accepted: 15 March 2011 / Published online: 12 April 2011
© Springer Science+Business Media, LLC 2011

Abstract The phenomenon of protein synthesis has been modeled in terms of totally asymmetric simple exclusion processes (TASEP) since 1968. In this article, we provide a tutorial of the biological and mathematical aspects of this approach. We also summarize several new results, concerned with limited resources in the cell and simple estimates for the current (protein production rate) of a TASEP with inhomogeneous hopping rates, reflecting the characteristics of real genes.

Keywords Protein synthesis · TASEP · Nonequilibrium statistical physics

1 Introduction

Nonequilibrium statistical physics remains one of the greatest unsolved challenges of theoretical physics. In recent studies, both the National Academy of Sciences [79] and the Department of Energy [17] have recognized the importance and scientific impact of developing a fundamental and comprehensive understanding of physics far from equilibrium. Unlike string theory or cosmology, this field addresses phenomena in our immediate experience, such as the flocking of birds or fish [6, 41, 92], traffic flow [15, 16, 82], or biological transport [44, 54, 59, 69, 70, 85], to mention just a few examples. Living systems and biological phenomena, in particular, are areas where concepts and methods from nonequilibrium statistical physics find a natural application. On the one hand, nonequilibrium statistical physics is concerned with open many-particle systems, sustaining nontrivial currents of energy or particles; and on the other hand, biological systems are characterized by considerable

R.K.P. Zia · J.J. Dong · B. Schmittmann (✉)
Physics Department, Virginia Tech, Blacksburg, VA, 24061, USA
e-mail: schmittm@vt.edu

R.K.P. Zia
e-mail: rkpzia@vt.edu

J.J. Dong
Physics Department, Hamline University, St. Paul, MN, 55104, USA
e-mail: jdong01@hamline.edu

complexity and depend on energy and matter throughputs for proper functionality. Of specific interest are biological transport phenomena restricted to an effectively one-dimensional track, such as kinesin and dynein on microtubules, RNA polymerase on DNA during transcription, and ribosomes on mRNA since these offer the promise of being accessible via simple model systems. The last process, usually referred to as translation, is a key element of protein synthesis, and will be our prime focus in the following.

One possible approach is to start with a good understanding of the physical and biological processes involved in protein synthesis, so that a detailed model can be developed and tailored to a particular experimental situation. While this promises the possibility of quantitative comparisons between model results and experimental data, any underlying generic characteristics cannot be easily identified, due to the large amount of experimental detail involved. An alternate approach—which has proven immensely powerful in equilibrium statistical mechanics—is to study simple model systems which can provide deep insights into generic behaviors and universal phenomena. Even though such simple models may not allow for immediate comparisons with experimental data, due to the considerable amount of simplification or abstraction involved, they can still guide future experimental work, especially when good data are not yet available. In summary, both approaches have their strengths and drawbacks and offer complementary insights.

In this article, we discuss one particular marriage between statistical physics and biology. We will approach a very complex biological process—protein synthesis—starting from a simple model, the totally asymmetric exclusion process (TASEP). In a twist of history, this model was first suggested in this context in the late 1960s [69, 70] and endowed with as much experimental detail as known at the time. Independently, models of stochastic directed transport were proposed and studied by mathematical physicists [93]. About two decades ago, these were (re-) discovered by the nonequilibrium statistical physics community [63] and became paradigms of this field, much like the Ising model for the study of phase transitions and critical phenomena.

Our goal here is two-fold, namely, to provide a brief tutorial and to discuss some recent developments. The next two sections are devoted to the former and the following one to the latter. Thus, Sect. 2 is designed for those new to biology, providing some basics of how proteins are synthesized in a cell and, more specifically, a part of the process known as “translation.” Complementary, Sect. 3 is designed for readers who may be interested in getting involved with exploiting TASEP as a model for translation. Assuming they are familiar with the basics of statistical mechanics and simulation techniques, this section is meant to provide a self-contained, though by no means comprehensive, review of the simplest TASEP [93] and generalizations relevant for translation. Section 4 is devoted to two recent developments. The first is a brief review of competition between multiple TASEPs. In the second, we are motivated by “silent mutations” of genes to introduce a novel notion in statistical mechanics, namely, a quenched distribution of distributions. We end with a brief summary and outlook in Sect. 5.

2 Rudiments of Protein Synthesis

Essentially every vital process within cells of living organisms involves proteins (a.k.a. polypeptides). These are macromolecules formed from chains of amino acids. For example, hemoglobin is a protein which carries oxygen from our lungs to the rest of our body. The blueprint for how these proteins are synthesized is contained in the DNA of the cell. Over the past century, biochemists gradually discovered the way this information is transformed into a physical molecule. Much of this body of knowledge is quite well established

and can be found in many standard texts of microbiology [3]. For the reader's convenience, this Section provides a brief synopsis of this process.

Protein synthesis involves two stages: *transcription* of genetic information from DNA to messenger RNA (mRNA) by RNA polymerase and *translation* from mRNA to proteins through ribosome translocation. Articulated by F. Crick in [22, 24], this central dogma of molecular biology captures the essence of transferring sequence information to functional macromolecules (e.g. RNAs and proteins) in all life forms. One of the most complex cellular processes, protein synthesis demands concerted actions by hundreds of molecules in sequential steps and typically requires a high level of regulation. Its vast demand for the energy needed to complete the reactions also establishes its crucial role in all metabolic pathways. Therefore, developing a *quantitative understanding* of transcription and translation processes would be most desirable. Indeed, this task has dominated much of recent research in molecular biology, as well as mathematics, physics and emerging cross-disciplinary fields. Of course, any model that encompasses all the biochemical reactions and the structural components in translation will be prohibitively complex. Rather, it is more practical to attempt at gaining some insight into the process of transforming DNA information into a polypeptide chain by restricting our attention to single-cell organisms and identifying the most essential ingredients. To this end, we focus on the bacterium *E. coli*, a well-established model organism in molecular biology of which abundant genetics and kinematics data are available for further analysis. In this article, we will further restrict our attention to the second stage: translation. Our aim here is to condense the relevant biological information into a simple overview of protein synthesis, so that interested readers may actively participate in model-building/analysis.

In most bacteria such as *E. coli*, translation involves three main players: the mRNA (genetic template), the ribosome (assembly machinery), and aminoacyl transfer RNAs (aa-tRNAs), i.e., transfer RNAs "charged" with the corresponding amino acid.

The mRNA carries genetic information, encoded as triplets of nucleotides. Each triplet is known as a "codon". Since there are four nucleotides (A, U, C, G), there are $4^3 = 64$ distinct codons, e.g., AUG, CGG, etc. Except for three "stop codons" (UAA, UGA, and UAG) that signal the termination of translation, each of the remaining 61 "codes" for one of the 20 amino acids. Each of the latter is conveniently denoted by a single letter: A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, and Y. Indeed, we may regard a protein as a word, i.e., a string of letters from the above list, ranging from ~ 10 to ~ 1000 long. With 61 codons and 20 amino acids, there can be, *on the average*, three codons associated with the same amino acid. As a result, a protein with L amino acids can be coded by $\sim 3^L$ distinct sequences (strings of codons or mRNAs). To complicate matters, some amino acids are coded by a single codon (e.g. AUG for methionine/M and UGG for tryptophan/W) while others are associated with as many as 6 (e.g. CGU, CGC, CGA, CGG, AGA, AGG for arginine/R). Therefore, the exact protein-mRNA degeneracy depends on the sequence (though not the precise order of codons). Codons coding for the same amino acid are termed "synonymous." In the example above, CGU and AGG are synonymous codons. Two sequences which differ only by synonymous codons are known as "silent mutations," in the sense that both produce the same protein (polypeptide chain). In addition to the string of codons, there is a long sequence of nucleotides at the beginning of the mRNA, known as the "Shine-Dalgarno sequence" (SD sequence) [90]. This region controls the binding of a ribosome to the mRNA, which is also the start of the process of translation. Known as "initiation" in biology, even this starting event is quite complex [5, 62, 67, 73, 74, 78, 80], requiring the presence of several initiator proteins.

The next major player is the ribosome, a sizable molecule composed of a large and a small subunit. Within the ribosome, there are three sites to which a tRNA can bind and

unbind. Designated as A, P, and E, these are, respectively, the aminoacyl site (for docking of an aa-tRNA), the peptidyl site (for transferring and binding the newly arrived amino acid to the partially formed polypeptide chain) and the exit site (for releasing the tRNA).

The process of translation consists of ribosomes moving along the mRNA without backtracking (from one end to the other, technically known as the 5' end to the 3' end) and is conceptually divided into three major stages: initiation, elongation and termination. Among the three steps, initiation is of the highest complexity and has seen significant developments in unraveling its molecular details [5, 62, 67, 73, 74, 78, 80]. Here, the ribosome interacts with the SD sequence through complementary base-pairing and locates the start codon AUG with the help of several initiator proteins. Translation begins with the assembly of the two subunits of the ribosome, along with a tRNA charged with the M amino acid in the A-site. The next steps, elongation and termination, are also quite intricate [9, 25, 43, 58, 76, 77, 83, 91]. The ribosome moves along the mRNA, "reading" codon by codon, recruiting the appropriate aa-tRNA, "knitting" the latest amino acid into the partially completed chain, and releasing the "discharged" tRNA. This cyclic process consists of the following steps. (i) The last amino acid of a partial chain is attached to its tRNA at the P-site, aligned with a certain codon. (ii) An aa-tRNA, correctly matched with the next codon, docks at the A-site. (iii) The peptide bond between the amino acid and the tRNA at the P-site breaks and reattaches to the new amino acid at the A-site. (iv) The ribosome moves forward so that the two tRNAs are now at the E- and P-sites. (v) The discharged tRNA in the E-site is released, leaving the A-site empty for the next aa-tRNA. Finally, when the ribosome encounters one of the three stop codons, the termination process commences: The ribosome disassociates, while the completed amino acid chain is released (and folds into a functioning protein). The whole process is quite involved and, instead of providing a figure here, we direct interested readers to one of the many helpful animations on the WWW [10].

Obviously, the third major set of players are the tRNAs. One end of this class of molecules consists of one of the many anticodons (e.g., UAC, to match with AUG).¹ The other end is an acceptor stem, to which an appropriate amino acid (one of 20) can be attached, forming a "charged" aa-tRNA. Normally, the mapping from amino acids to anticodons is one-to-many; details may be found in, e.g., [97]. For our purposes, the main concern is the rather dissimilar set of *degeneracies*, i.e., the number of synonymous codons, m_{aa} , for the amino acid aa :

aa	A	C	D	E	F	G	H	I	K	L
m_{aa}	4	2	2	2	2	4	2	3	2	6

aa	M	N	P	Q	R	S	T	V	W	Y
m_{aa}	1	2	4	2	6	6	4	4	1	2

(1)

Note that 3 codons are reserved for termination, so the total here is only 61. Meanwhile, the concentrations of these aa-tRNAs in a typical cell are known to be far from uniform. Indeed, for *E. coli*, the relative abundance can be as much as a factor of 15 [34, 55, 96]. Since the elongation rate is believed to be correlated with the aa-tRNA availability (as a ribosome must "wait" for the appropriate aa-tRNA to arrive before proceeding) [34, 68, 65, 91], the time it takes a ribosome to complete translation can vary widely, depending on the codon sequence. Further, the rate of protein synthesis depends not only on the speed of

¹Simplistically, there would be 61 anticodons to match the 61 codons. However, nature is more complicated. Most cells contain less, due to "wobbling" [23]. For *E. coli*, there are 46 distinct anti-codons. We will ignore this extra complication here and discuss translation as if there were 61 anticodons.

a single ribosome, but also on how many ribosome may be translating simultaneously (on the same mRNA). Therefore, the wide range of degeneracies shown in (1) implies that silent mutations may have serious implications for protein production rates.

Finally, for a cell to function properly, the important quantities to be controlled are, presumably, the levels of various proteins. In a steady state, a protein's level depends on the rate of its degradation as well as its production. Assuming the former is the same for all proteins, then the concentration of any particular type follows its production rate closely, and the latter is just the average current associated with translating that mRNA. Thus, our main interest here is the following question: How are protein production rates correlated with specific sequence information? Of course, translation in real biological systems is much more complex, involving subtleties such as prokaryotes vs. eukaryotes, initiation and elongation factors, wobbling, signaling, and regulation in response to external conditions, etc. All are beyond the scope of this article, as we focus on a few key ingredients, build the simplest models, explore their behavior, and make some inroads into the remarkable processes of life.

3 The TASEP and Its Generalizations

In this section, we turn to another topic: the totally asymmetric simple exclusion process (TASEP), which has been exploited to model protein synthesis for over 40 years. After a brief historic introduction, we review recent progress on simple models and generalizations to account for more realistic conditions in biological systems.

3.1 Early History: Studies by Gibbs et al. and Spitzer

Around 1970, TASEP was introduced from two entirely distinct motivations, one from biology and the other from mathematics. Not surprisingly, the two approaches are also quite different. In this subsection, we devote brief paragraphs to each.

Gibbs and collaborators [69, 70] were interested in a quantitative description for translation, the process reviewed briefly above. The mRNA, the codons, and the ribosomes are modeled by, respectively, a one-dimensional open lattice, its sites, and particles (which enter the lattice at one end, hop unidirectionally from site to site, and exit at the other end). Aware of the large size of ribosomes compared to codons, these workers began with particles which "cover" $\ell \geq 1$ sites. Further, their formulation allowed for the possibility that particle hopping rates can be bidirectional, as well as being site- and time-dependent. Setting up the full stochastic problem, with a master equation for the probability to find particles at each site, is relatively easy. However, no one has been able, so far, to solve such a complex problem, even for the steady-state distribution with time-independent rates. Nevertheless, there has been significant progress since 1968, on various fronts. Historically, Gibbs, et. al. focused on a system with uniform and time-independent rates, ignored some correlations between ribosomes yet accounting for the strict exclusion due to $\ell > 1$, and set up recursion relations for ρ_i , the density of particles at site i (i.e., the average occupation of site i by, say, the trailing edge of a particle), with the current J as a to-be-determined unknown. Their analysis of these recursion relations, using both analytic and numerical techniques, led them to several important conclusions. Examples include a non-trivial ℓ -dependent relationship between J and the overall density $\bar{\rho}$, as well as the existence of different phases as functions of the entry/exit rates. Since their model is too far from real biological systems and experimental techniques in biology were too primitive to probe occupations at the molecular level, their results have lain largely dormant until recent years.

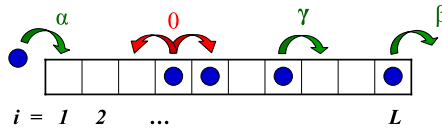


Fig. 1 Schematic summary of the rules of the proto TASEP. Particles cannot hop backwards or onto an occupied site (*red* online). They move forwards to empty sites with rate γ (*green* online). Entry and exit rates are denoted by α and β , respectively (*green* online)

At the other front, Spitzer was interested in Markov processes involving interacting particles [93]. The simplest interaction would be just exclusion, i.e., each site can be occupied by at most one particle—thus the “simple exclusion” part of TASEP. The simplest system would be a one-dimensional periodic lattice, i.e., a ring of L sites. Placing N particles on such a ring and letting them hop randomly to nearest-neighbor sites obviously leads to “simple diffusion.” Even in such a minimal system, highly non-trivial behavior can be found [32], but we will focus on a dynamics more closely resemble that of protein synthesis, namely, hopping is allowed only in one direction. Thus, we turn to the “totally asymmetric” part of TASEP. Though the steady state is still simple, its dynamics displays even more intricate properties [27, 30, 45, 46, 50, 57, 66, 71]. Of course, Spitzer also considered lattices with open boundaries and particles hopping on/off the lattice at the ends. But this problem was sufficiently more complex than the ring that even its steady state distribution remained elusive for another two decades. In the next subsection, we specify Spitzer’s model and briefly summarize some of the progress since the early 90’s.

3.2 The Proto Model and Its Properties

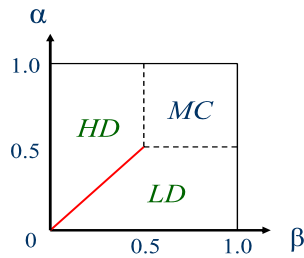
In this model, the sites of a one-dimensional lattice, labeled by $i = 1, \dots, L$, may be occupied by $n_i = 0$ or 1 particle, so that a configuration of the system is specified by $\{n_i\}$. Starting from some initial configuration with $N < L$ particles, the rules of evolution are the following. At each update step, a site is randomly chosen and, provided it is occupied and the next site, located in the selected direction, is empty, an attempt to exchange the particle–hole pair is made. With probability γ , the attempt is termed “successful” and the particle is moved to the next site. For the process on a ring, site L is connected to site 1. Then, a Monte Carlo step (MCS) is defined as L attempts, so that, on the average, each site will be chosen once in a MCS. This way, it makes sense to compare systems with very different L ’s, but run for the same number of MCS. For the open TASEP, two additional rules apply: Site 1, if empty, is filled with probability α and a particle on site L leaves the lattice with probability β . An MCS in this case consists of $L + 1$ attempts. For the rest of this article, we will focus on the open case, since it resembles the process of protein synthesis. Schematically, its rules are summarized in Fig. 1. The central question is: What is $P(\{n_i\}, t)$, the probability for finding the system in configuration $\{n_i\}$ after t attempts? A master equation for P can be easily written (for open TASEP):

$$P(\{n'_i\}, t + 1) - P(\{n'_i\}, t) = \sum_{\{n_i\}} \mathcal{L}(\{n'_i\}, \{n_i\}) P(\{n_i\}, t) \tag{2}$$

where

$$\mathcal{L}(\{n'_i\}, \{n_i\}) \equiv W(\{n'_i\}, \{n_i\}) - \delta(\{n'_i\}, \{n_i\}) \sum_{\{n''_i\}} W(\{n_i\}, \{n''_i\})$$

Fig. 2 Phase diagram of the proto TASEP. Maximal current and high/low density phases are denoted by MC, HD, and LD respectively. The transition between the MC phase and the other two (*dashed lines*) is continuous. The transition across the HD-LD boundary (*solid, red online*) is discontinuous



is known as the Liouvillian (which plays a role similar to the Hamiltonian in quantum mechanics) and δ is the Kronecker delta. Here, $W(\{n'_i\}, \{n_i\})$ is the transition probability from $\{n_i\}$ to $\{n'_i\}$

$$\begin{aligned} & \frac{1}{L+1} \left[\alpha(1-n_1)\delta(n'_1, n_1+1) \prod_{j>1} \delta(n'_j, n_j) \right. \\ & + \sum_{k=1}^{L-1} \gamma n_k(1-n_{k+1})\delta(n'_k, n_k-1)\delta(n'_{k+1}, n_{k+1}+1) \prod_{j\neq k, k+1} \delta(n'_j, n_j) \\ & \left. + \beta n_L \delta(n'_L, n_L-1) \prod_{j<L} \delta(n'_j, n_j) \right] \end{aligned} \tag{3}$$

where the changes $n \rightarrow n \pm 1$ are explicitly displayed. However, finding the solution to (2) is far more difficult. A simpler question is: What is the stationary distribution, $P^*(\{n_i\})$, assuming the system settles into such a t -independent state at large times? Once it is known, other natural questions arise: What are the macroscopic properties of the system in this state? Of particular interest is how α and β control the averages of observables,

$$\langle \mathcal{O} \rangle \equiv \sum_{\{n_i\}} \mathcal{O}(n_i) P^*(\{n_i\}),$$

such as the density profile $\rho_i \equiv \langle n_i \rangle$. Once we have ρ_i , other quantities of interest can be computed. In particular, we will be mostly interested in the overall density, $\bar{\rho} \equiv \sum_i \rho_i / L$, and the average current, $J = \beta \bar{\rho}_L$ (i.e., the average number of particles that enters/exits the lattice in a Monte Carlo step). The answers, some known to Gibbs et al. became more well-established over the last two decades. Setting γ to unity (without loss of generality), the system can be found in three distinct phases in the α - β plane—a half-filled phase with maximal current and high/low density phases [63], denoted by MC, HD, and LD respectively in Fig. 2.

Transitions from the maximal current (MC) phase to the other two phases (HD/LD) are continuous, and display critical behavior similar to second order phase transitions in equilibrium. Indeed, critical properties, such as algebraic decaying correlations, can be found in the entire MC phase. Across the HD-LD boundary, the transition is discontinuous, and, on the line itself, the system displays coexistence of HD and LD. Specifically, the HD-LD regions are macroscopic, separated by a microscopic interface, referred to as a “shock.” As in many equilibrium systems with coexistence, such an interface can be located anywhere. In TASEP, the shock performs a random walk (reflected only from the ends), so that the average density profile is linear in i , interpolating between $\bar{\rho}_{HD}$ and $\bar{\rho}_{LD}$. In the literature,

this line is often referred to as the “shock phase” (SP). Setting up a phenomenological theory for the behavior of this shock, known as domain wall theory, several authors have been successful in predicting many properties of TASEP *outside* the MC region [7, 84]. The exact $P^*(\{n_i\})$ was found [29, 31, 86, 87], from which $J(\alpha, \beta)$ and $\bar{\rho}(\alpha, \beta)$ can be computed analytically for all α, β . In the $L \rightarrow \infty$ limit, these are remarkably simple: $J = \bar{\rho}(1 - \bar{\rho})$ always, while $\bar{\rho} = \{1/2, 1 - \beta, \alpha\}$ in the MC, HD, LD phases, respectively. Thus, $J \leq 0.25$ in general. More recently, considerable progress was made using the powerful Bethe-Ansatz [26, 39, 49, 60, 75, 81, 94], so that the complete spectrum and all the eigenvectors of \mathcal{L} are accessible. Consequently, some of the more complex, dynamic properties of TASEP are also exactly known. Details of this large body of results are beyond the scope of this article. The interested reader may consult several comprehensive reviews such as [11, 28, 47].

Despite this comprehensive knowledge of \mathcal{L} , there are seemingly simple questions about this system for which simple answers are not available. An example is the power spectrum associated with $N \equiv \sum_i n_i$, the total number of particles on the lattice. Specifically, we record a time series $N(t)$ over a run and construct its Fourier transform, $\tilde{N}(\omega)$. Carrying out many runs and taking the average, $\overline{[\dots]}$, the power spectrum is

$$I(\omega) \equiv \overline{|\tilde{N}(\omega)|^2} = \overline{\left| \frac{1}{T} \sum_{t=0}^T e^{i\omega t} N(t) \right|^2}.$$

Note that this average contains information on the dynamics and is therefore not related to the static average, $\langle \dots \rangle$, above. If the runs are taken when the system is in the steady state, then $I(0)$ is, of course, known: $(\bar{\rho}L)^2$. But, $I(\omega > 0)$ displays more interesting behavior, such as oscillations (in ω) in the HD/LD phases [2, 20]. Although the physics behind these is understandable and approximate theories provide reasonable fits, an exact analytic formula is not known (except formally).² In the remainder of this article, we will look beyond this proto model and focus on generalizations which take into account some other essential ingredients in the process of protein synthesis.

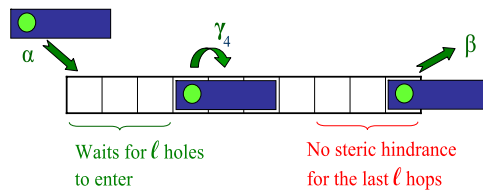
Before continuing, let us point out an equivalent formulation of the open TASEP, but based on a ring. Conceptually simpler and essentially used in simulations, this version will appear to be most natural in the contexts to be presented below. Here, we consider a periodic lattice with $L + 1$ sites filled with a total number, N_{tot} , of particles. Considering the role it plays, we will refer to the extra site, $i = 0$, as the “reservoir” or the “pool.”³ The rules associated with this site are, of course, quite different from those in the bulk: (i) It has unlimited occupation, so that we are guaranteed $n_0 \geq 1$ by imposing $N_{tot} \geq L + 1$. (ii) If it is chosen for updating, one of its particles is moved to $i = 1$ with probability $\alpha(1 - n_1)$. (iii) If site L is chosen and $n_L = 1$, the particle hops into the pool with probability β , regardless of n_0 . By denoting α, β as γ_0, γ_L , we may regard them as part of a full set of site-dependent hopping rates $\{\gamma_i\}$. Incorporating the special rules for site 0, we can replace $[\dots]$ in (3) by a succinct expression

$$\sum_{k=0}^L \gamma_k [n_k + \delta_{k,0}(1 - n_0)] [1 - n_{k+1} + \delta_{k,L}n_L] \prod_{j=0}^L \delta(n'_j, n_j - \delta_{j,k} + \delta_{j,k+1}) \quad (4)$$

²In this case, the difficulties lie mainly in computing the average of nonlocal (in both space and time) operators.

³The notion of particle reservoirs was used in the literature, with one major difference. Unlike here, open TASEPs were coupled to *nwo* unrelated reservoirs, one at each end.

Fig. 3 “Complete entry, incremental exit” rule for an $\ell = 3$ case. The gray dots (green online) denote the “readers.” Since the reader of the second particle is on $i = 4$, it will hop with probability γ_4



(with $n_{k+1} = n_0$, etc.). Note that this $\Pi_j \delta$ includes all the possible changes in $\{n_j\}$. To re-emphasize, $N_{tot} = n_0 + N$ is conserved in this formulation. However, as long as $N_{tot} \geq L + 1$, the properties of the open TASEP above are identical to the $i \in [1, L]$ part of our ring and N can fluctuate in the range $[0, L]$ as before.

3.3 Generalizations of TASEP

As noted above, Gibbs et al. [69, 70] were aware of the size of a ribosome compared to a codon, so that Spitzer’s simple TASEP must be generalized to having “particles” which extend over $\ell \geq 1$ sites. Indeed, from the latest data, $\ell \sim 12$ seems to be the most appropriate [53, 69, 70]. This generalization requires some modifications to the rules. Since the ribosome appears to “read” the codon over its A-site, it is most natural to associate one particular part of the extended particle with the “reader” [89]. After some thought, it is also clear that, as far as TASEP is concerned, which part is labeled the reader is irrelevant. For convenience, we choose the reader to be at the trailing edge of the particle [89]. To “read” the first codon (site 1 on the lattice), the ribosome/particle must enter the lattice and for that to occur, the first ℓ sites must be empty. On the other hand, while the ribosome is “reading” the last ℓ codons, it must be the last particle on the lattice, with no others to impede its progress. Therefore, it can move without hindrance toward the exit end. The new set of entry/exit rules is known as “complete entry, incremental exit” [13]. (See Fig. 3 for an illustration.)

These seemingly modest changes of the rules have profound consequences. At the simplest level, we must now distinguish between particle (ribosome) density and “coverage” density (number of sites “covered” by a particle per unit length). Denoting the former by ρ^r and the latter by ρ , we see that the overall densities differ by a factor of ℓ ($\bar{\rho} = \ell \bar{\rho}^r$), and the two profiles are related by

$$\rho_i = \sum_{k=0}^{\ell-1} \rho_{i-k}^r.$$

If $\rho^h = 1 - \rho$ denotes the hole density, we also have $\rho^r + \rho^h < 1$. Although the phase diagram and the current-density relation $J(\bar{\rho}; \ell)$ are qualitatively unchanged [13, 35, 89], an exact solution (for P^* or J or $\bar{\rho}$) remains elusive. Stationary profiles are much more seriously affected, especially in the HD phase [35]. On the other hand, for a homogeneous TASEP on a ring, P^* is known to be uniform [93], so that an exact $J(\bar{\rho}; \ell)$ can be derived [13, 89].⁴ To be precise, we denote the particle current by J , but write it in terms of $\bar{\rho}$ instead of $\bar{\rho}^r$:

$$J(\bar{\rho}; \ell) = \frac{\bar{\rho}(1 - \bar{\rho})}{\ell + (\ell - 1)\bar{\rho}}. \tag{5}$$

⁴Other exact results (also for a collection of particles with different ℓ ’s) have been found recently. See e.g., [4].

A more elegant version of this formula is

$$J^{-1} = [\bar{\rho}^r]^{-1} + [\bar{\rho}^h]^{-1}, \quad (6)$$

in which the second term accounts for steric hindrance somehow. On the left is the average time between successive particles (with exclusion) exiting the lattice. On the right, we have the sum of such times for *non-interacting* particles and holes. This connection is quite remarkable.

Returning to (5), we see that $J(\bar{\rho})$ still rises from zero, reaches a maximum, and returns to zero. However, its upper bound is lowered to $(1 + \sqrt{\ell})^{-2}$, i.e., by $O(\ell)$ for large ℓ [13, 35, 89]. While this $J(\bar{\rho}; \ell)$ it is not rigorously the same as the one in an open TASEP, it can be argued that, in the $L \rightarrow \infty$ limit, the two should be the same. As noted above, Gibbs, et. al. arrived at the same $J(\bar{\rho}; \ell)$ long ago, by accounting for some effects of the ℓ -exclusion approximately. This is one of the few reasonably well understood aspects of TASEP with extended objects. In passing, we should mention that TASEPs with polydispersed particles on a ring have also been studied [4], though their relevance to protein synthesis seems remote.

A second essential aspect of our problem was also recognized by Gibbs, et. al. [69, 70], namely, site-dependent hopping rates, i.e., inhomogeneous TASEPs. In Sect. 2, we indicated the rationale for considering such a difficult problem: non-uniform aa-tRNA abundance. Since the translation of each codon is mediated by the arrival of the cognate aa-tRNA (up to “wobbling” [23]), it is believed that the concentration (or “abundance”) of these aa-tRNA determines the local translation rate, leading to an inhomogeneous sequence of $\{\gamma_i\}$. These abundances can be measured in experiments on *E. coli* and vary with external conditions, such as access to nutrients [34]. Needless to say, it is prohibitively difficult to determine quantities like J and $\bar{\rho}$ for a TASEP with an arbitrary, fixed set of rates, $\{\gamma_i\}$. Even when restricted to point particles ($\ell = 1$) on a homogeneous ring, the introduction of a single “defective site” (with $\gamma \neq 1$ but no changes to the rules of exclusion) renders the problem insolvable (i.e., no exact P^*) so far. The non-trivial consequences and serious challenges were noted as early as 1992 [56].⁵ With several defects, systematic studies become less manageable, even with approximate or numerical methods. For an open TASEP with a few defect sites, progress was made mainly with Monte Carlo simulations, while some understanding is possible by exploiting mean-field approximations of various levels [37, 42, 61]. Most relevant to modeling translation is the discovery that the current (for $\alpha = \beta = 1$) depends on the location of slow defects ($\gamma < 1$) [14, 35–37, 42]. In particular, if there are two slow sites in the bulk, the distance between them affects J seriously [14, 35, 36, 42]. This implies that protein production rates can be significantly suppressed if codons associated with rare aa-tRNA’s are *clustered* in the gene. At the other extreme, several groups studied TASEPs with a full set of quenched random rates, $\{\gamma_i\}$, each of which is chosen from some distribution (e.g., Gaussian, two-valued, etc.) [8, 48, 52, 64, 95]. These authors considered only point particles and focused on the effect of disorder on the (quenched average) current-density relation. Using simulations and mean field approximations, $J(\bar{\rho})$ is found to develop a plateau in a region around $\bar{\rho} = 1/2$, details of which depend on the variance of the distribution of the inverse rates: $1/\gamma$. The phase diagram remains qualitatively the same, with three phases that resemble MC, HD, and LD. Not surprisingly, the main effect is that the transitions are no longer sharp. Beyond these studies, disordered TASEPs with $\ell > 1$ are yet to be explored.

⁵Remarkably, exact results are available if a single *particle* hops more “slowly” [33, 56, 72, 86].

To model protein synthesis more realistically, we need a combination of at least three ingredients: (a) open boundary conditions, (b) extended objects (say, $\ell = 12$), and (c) inhomogeneous rates, $\{\gamma_i\}$. As we noted from the historic perspective, it took some time to arrive at full solutions—for TASEPs with ingredients (a) or (b). Yet, here we would prefer to include all three aspects and ask for, at the least, the average current $J(\alpha, \beta, \{\gamma_i\}, \ell)$ and the overall density $\bar{\rho}(\alpha, \beta, \{\gamma_i\}, \ell)$. Clearly, this program is extremely ambitious, even if we restrict our investigations to the Monte Carlo approach. In Sect. 4.2, we will present a very simple, yet reasonably reliable, method to arrive at a good estimate for the current.

4 Some Recent Developments

In this section, we present two topics where some recent progress was made. We begin (Sect. 4.1) with an analysis of a particular instance of the cell having limited resources available. In TASEP language, we are exploring how a TASEP is affected by having a *finite* reservoir of particles. The effects of several TASEPs competing for the same pool of particles [1, 19, 21] will be also presented. The rationale behind such pursuits is that a cell has thousands of *copies* of thousands of different *types* of mRNAs, competing for the same pool of ribosomes. Do some “win” while others “lose”? In the language of TASEPs, since we model an mRNA by a sequence $\{\gamma_i\}$, we will be interested in $J(\{\gamma_i\}; N_{tot})$, namely, how the current associated with this sequence depends on N_{tot} , the total number of particles in the pool. Our analysis here will be restricted to homogeneous TASEPs ($\gamma_i = 1$) of differing lengths.

Section 4.2 will be devoted to our search for a simple (“quick and dirty”) way to estimate $J(\alpha, \beta, \{\gamma_i\}, \ell)$ for a single TASEP but with a fully inhomogeneous sequence of hopping rates, $\{\gamma_i\}$. This search leads us to a novel form of quenched randomness, which we named⁶ “distribution of distributions.” Recall that a protein is a fixed sequence of L amino acids and can be coded by $O(e^L)$ different mRNAs. As an illustration, suppose we wish to synthesize an artificial protein consisting of only one amino acid, R, repeated L times. According to the table provided in (1), R can be represented by 6 different codons, namely, CGU, CGC, CGA, CCG, AGA, or AGG. Hence, there are 6^L possible mRNA sequences which code for this codon. Each of these corresponds to a particular realization of a quenched random sequence of codons, chosen from a single distribution of 6 values (i.e., codons). This procedure is standard for problems involving quenched disorder⁷. Turning to the more typical case of a naturally occurring protein—the “wild type”—the L amino acids will of course be different, each with its own characteristic degeneracy. Thus, the sequence of degeneracies will be non-trivial and fixed (e.g., 4266224 for the amino acid string PQLRFEV). Thus, instead of choosing codons from a single distribution to construct all possible mRNAs (as in the artificial RRRRRR case or in all previous studies of quenched disorder), they must be chosen from a *fixed sequence of different* distributions. Pursuing these ideas further, we discovered a remarkable fact about *E. coli*. Simulating 5000 randomly chosen sequences for each of 10 specific genes, we find that the average currents lie in a narrow range (within 25% of each other). However, the currents associated with the wild types typically lie very far above the average. These are intriguing findings, from the perspectives of both, the statistical physics of quenched disorder and the specific realization “chosen” by the living organism.

⁶This notion was presented in [99].

⁷See, e.g., [38, 40, 98], etc.

4.1 Competition for Ribosomes: TASEP with Finite Particle Reservoirs

In a living cell, ribosomes are constantly synthesized and degraded. On the other hand, it is believed that some are also “recycled,” i.e., after termination in translating one gene, the subunits reassemble to translate another gene. Of course, there are multiple facets to “ribosome recycling.” Chou considered the enhancement of initiation rates on a gene due to the proximity of a ribosome which unbinds from the same mRNA [12]. We consider a different aspect. Ignoring synthesis/degradation, let us model the number of ribosomes in a cell by a constant, N_{tot} , to be shared by all the genes. Then we ask: What is the effect of multiple TASEPs competing for a single pool with a finite number of particles? As a base-line study, we first focus on the effects of finite N_{tot} ’s on just one homogeneous TASEP [1, 19, 20]. For example, we seek $\bar{\rho}(N_{tot})$, the dependence of the overall density on the total particle number. This study is then extended to include multiple (homogeneous) TASEPs [21] with possibly different L ’s. Will the overall densities and currents be the same or different? If the latter, how are they controlled by N_{tot} ? So far, all studies are based on point particles and uniform entry/exit rates.

For these investigations, it is clear that the alternative representation of an open TASEP in Sect. 3.2 is most natural, with n_0 being the number of particles in the pool. In the single TASEP case, novel behavior already arises when we introduce only one modification: allowing N_{tot} to be lowered below $L + 1$. Ha and den Nijs coined this the “parking garage problem” and provided many interesting results [51]. To model how translation might be affected by the scarcity of ribosomes, we let the binding rate of a ribosome to the mRNA, γ_0 , depend on the ribosome concentration. In particular, when the ribosome concentration is very low, we let γ_0 be proportional to it. At the opposite extreme, it should have no effect on γ_0 , which should take on some intrinsic value—denoted by α —associated purely with the binding kinetics. In our model, n_0 is proportional to this concentration, so that we simply choose a convenient $\gamma_0(n_0)$ which interpolates between 0 and α . In all the simulation studies [1, 19, 20], we have

$$\gamma_0(n_0) = \alpha \tanh(n_0/N^*) \tag{7}$$

where N^* is some crossover parameter (chosen to be $O(L)$ for convenience). By contrast, the exit rate should not be affected by n_0 , so that we simply have $\gamma_L = \beta$. Since $N_{tot} = n_0 + N$, both n_0 and N will be small as N_{tot} is increased from 0, and the system first finds itself in the LD phase. At the other extreme, n_0 is necessarily large as well (since $N \leq L$), so that we will arrive at an ordinary open TASEP associated with (α, β) . A crossover occurs when N_{tot} reaches $O(N^*) = O(L)$. As may be expected, the LD-LD and LD-MC crossovers are uneventful, since no discontinuities are encountered. The response of the TASEP can be well approximated by a self-consistent equation for $\bar{\rho}(N_{tot})$:

$$\bar{\rho} = \gamma_0 = \alpha \tanh((N_{tot} - \bar{\rho}L)/N^*) \tag{8}$$

More interesting is the LD-HD crossover, since it spans a discontinuous boundary. The response is well described by the following. Raising N_{tot} from 0, the average density is given by the above equation, until a critical value, $N_{tot}^- \equiv \beta L + N^* \tanh^{-1}(\beta/\alpha)$, is reached. Lowering N_{tot} from ∞ , $\bar{\rho}$ remains at the HD value of $\bar{\rho}_{HD} \equiv 1 - \beta$, until N_{tot} reaches another critical value: $N_{tot}^+ \equiv (1 - \beta)L + N^* \tanh^{-1}(\beta/\alpha)$. Between N_{tot}^- and N_{tot}^+ , all increases in N_{tot} are absorbed by the TASEP (while n_0 and γ_0 stay constant). Thus, $\bar{\rho}$ rises linearly: $\bar{\rho}(N_{tot}) = \beta + (N_{tot} - N_{tot}^-)L$. Such a response has an analog in equilibrium first order transitions, corresponding to, e.g., the linear section in an isotherm in the P - V diagram of a

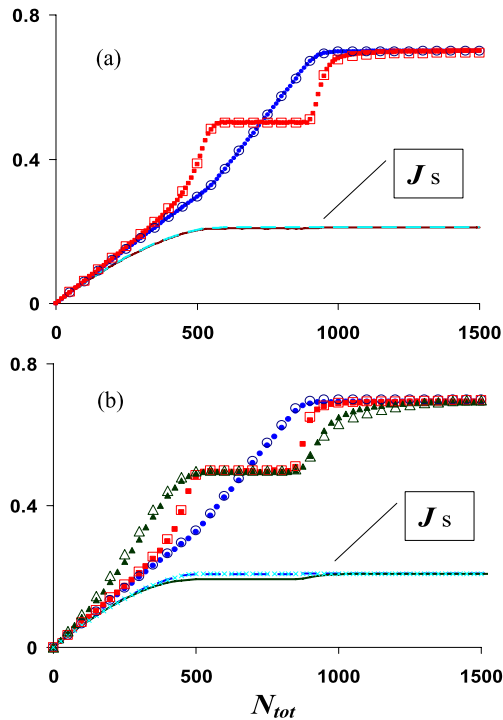
binary mixture. Furthermore, the average profile (ρ_i) in this regime is also noteworthy. Instead of being linear in i (as in an unconstrained TASEP), it resembles a stationary shock. The underlying physics is understandable: The feedback from the pool prevents the shock from wandering throughout the lattice. Instead the shock is localized to a position controlled by N_{tot} while its fluctuations are controlled by another detail of the feedback: $\partial\gamma_0/\partial n_0$. Domain wall theory, so successful in providing good approximations for an ordinary TASEP, can be generalized to account for the feedback to give excellent “zero-parameter fits” to simulation data [19]. In passing, let us mention that even more remarkable structures appear in the special case of LD-SP (i.e., setting $\alpha = \beta$ and varying N_{tot}). In all cases, the current displays no major surprises, mainly following the $J(\bar{\rho})$ curve of an unconstrained TASEP. The interested reader is referred to [19] for details.

Next, we turn to multiple TASEPs and their competition for a finite pool of particles [21]. To model different genes and the many possibilities of regulation, we need (at the least) three parameters for each type (μ) of TASEP: $L_\mu, \alpha_\mu, \beta_\mu$. A systematic study in the full parameter space of M TASEPs becomes quickly unmanageable, so ours is restricted to $M = 2, 3$. On the other hand, as an attempt at more realistic models, one unpublished study [18] simulated the competition of 10 genes from *E. coli* (with L 's ranging from 109 to 558; details in the next subsection). With $\ell = 12$ and the appropriate sets of γ_i 's, the only unrealistic part of this study is setting $\alpha_\mu = 1$ for all genes. Not surprisingly, for $N_{tot} \sim O(1)$, the currents are all the same, being controlled by the same small entry rate: γ_0 . For $N_{tot} \gtrsim 300$, each TASEP is saturated in their MC-like phases, so that the currents differ by a factor of ~ 2 . The approximate value of 300 can be expected from $10(\text{genes}) \times 400(\text{typical length})/12(\ell)$. Meanwhile, crossovers occur at N_{tot} 's in the range of 100–250. The conclusion of this limited study is that, while the first attempt has been made at the question of mRNA competition for ribosome in a real cell, much more remains to be explored before meaningful insights can be developed.

Focusing on a more systematic (though less “realistic”) study of competition, we consider two TASEPs. The model here consists of two lattices with $L_{1,2}$ sites, joined at one site (site 0, the pool), so that it has the topology of two rings joined at one point. When site 0 is chosen, with equal probability a particle attempts to move onto one of the two lattices. Once a lattice is chosen, it enters with the rates, $\alpha_{1,2}$. As usual, there is no exclusion at site 0, so that particles simply hop from sites $L_{1,2}$ into the pool with rates $\beta_{1,2}$. For simplicity, we let $\alpha_{1,2} = \alpha$ and $\beta_{1,2} = \beta$ in this initial study [21]. Perhaps to be expected, when the two TASEPs are identical (i.e., $L_1 = L_2$), the symmetry is not spontaneously broken. The two response curves are the same, within statistical fluctuations. However, when the lengths are very different, a new pattern emerges. In particular, for $L_1 = 1000$ and $L_2 = 100$, we find roughly *five* regimes in the LD-HD crossover. While the longer TASEP displays essentially the same behavior as in the single TASEP case (three regimes), the shorter one experiences more variety (Fig. 4).

It is remarkable that, in the central section, $\bar{\rho}(N_{tot})$ for both are linear, with the shorter one being a constant! It turns out that the shock in this TASEP is delocalized, but acts as a control for the shock in the longer one. The motion of the two shocks is completely *anti*-correlated, so that n_0 , the pool particle number, remains essentially fixed. As a result, the average density profiles are quite different, being strictly linear for the short one. For the longer TASEP, the profile can be readily described by the profile of a single constrained TASEP, but “smeared out” over a distance of L_2 (length of the shorter lattice). As can be seen from Fig. 4a, the generalized domain wall theory is quite successful at capturing all this novel behavior. Finally, these insights can be exploited to understand the behavior of three TASEPs in competition. Though not dramatically different, new features do appear, especially in cases where the lengths are widely separated. For example, Fig. 4b is an illustration

Fig. 4 Average overall densities and currents as a function of N_{tot} when two/three TASEP are competing for a single pool of particles. In both cases, $\alpha = 0.7$ and $\beta = 0.3$. The currents in all cases follow approximately the same curve of the single TASEP case. Denoted by essentially indistinguishable *solid* and *dashed lines* (color online), they are marked by the call-out “ J_s ”. Simulations (*solid symbols*) and predictions from generalized domain wall theory (*open symbols*) of the densities are as follows. (a) $L_1 = 1000$ (circles, blue online) and $L_2 = 100$ (squares, red online). (b) $L_1 = 1000$, $L_2 = 100$, and $L_3 = 10$ (triangles, green online)



of $\bar{\rho}(N_{tot})$ for $L_1 = 10$, $L_2 = 100$, and $L_3 = 1000$. Many other results, such as remarkable properties of the stationary $P^*({k_\mu})$, the probability to find the domain wall located at site k_μ in lattice L_μ , are available. Beyond the scope of this article, these may be found in [21]. Of course, we have taken only a minuscule step towards modeling competition in real cells. In addition of containing thousands of different genes (e.g., 5416 in one strain of *E. coli*), there can be thousands of copies of each type, not to mention that we should include the three essential ingredients pointed out at the end of Sect. 3. Finally, looking far ahead, we can consider the genes competing for finite pools of the 46 varieties of aa-tRNA, a problem involving feedback from the details of the average occupation at each site. Clearly, this is a gargantuan task and much remains to be investigated before we can claim to understand competition for finite resources in a real cell.

4.2 A Simple Estimate for Currents in the Inhomogeneous TASEP

In this subsection, we return to a single open TASEP with extended particles ($\ell > 1$) hopping along with a fully inhomogeneous set of rates $\{\gamma_i\}$. This set can be chosen on the basis of aa-tRNA abundances, if we want to model translation, or from different quenched random distributions, if we are interested in the effects of randomness on the physics of TASEP. We may even choose an arbitrary, specific sequence of positive, real numbers between 0 and 1, just to illustrate an argument. In any case, we will focus only on the average current $J(\alpha, \beta, \{\gamma_i\}; \ell)$. The task of predicting this J is clearly beyond our present analytic abilities. Faced with this impasse, one reasonable question is: Is there a simple way to estimate it? Of course, in the limit of $\alpha \ll \gamma_{i>0}$, the particle density will be exceedingly low so that the particles are non-interacting, to a good approximation. Then, we simply have $J \simeq \alpha$, since

the total time it takes a particle to traverse the lattice (mRNA) becomes irrelevant.⁸ Note that the exclusion plays no role, so that ℓ is also irrelevant. This consideration, based on the idea of the “worst bottleneck,” can be used to provide the most naive estimate of J :

$$J_{worst\ b'neck}(\alpha, \beta, \{\gamma_i\}) \sim \gamma_{\min} \tag{9}$$

where γ_{\min} is the minimum in the set $\{\gamma_i\}$. Another possible estimate is to use the averages and variances of the entire set $\{\gamma_i^{-1}\}$, which we denote by $\bar{\eta}$ and σ_{η}^2 , respectively (following the notation in [52]). These quantities proved quite successful in the analysis of currents resulting from quenched random averages over $\{\gamma_i\}$ chosen from different distributions [52, 64, 95]. However, there are limitations for both estimates when addressing issues of interest here, namely, finding a reasonably good value of the current for a *specific* sequence $\{\gamma_i\}$. As in the estimate $J \simeq \alpha$, (9) is useful only when the bottleneck is very severe. But realistic rates are typically not so extreme that γ_{\min} is drastically smaller than the rest of the γ 's. More crucially, if we have two or more sites with the same γ_{\min} , then J will depend on their locations. For example, studies of just two slow sites [14, 35, 42] showed that, having them as neighbors (as opposed to being far apart) lowers J by as much as a factor of 2. Indeed, if the sequence contains a consecutive string of k such sites with $k \gg \ell$, then we can regard this stretch as an open, homogeneous TASEP in its MC phase. The considerations around (5) then provide us with $J \sim \gamma_{\min}(1 + \sqrt{\ell})^{-2}$. Thus, the clustering of many slow sites indeed suppresses J , by as much as a factor of 20 for $\ell = 12$. On the other hand, if these k slow sites are very far apart, then the estimate for J due to a *single* slow site [14, 35, 36, 61, 88] (which reduces to $J_{worst\ b'neck}$, to lowest order in γ_{\min}) should suffice. Similar limitations for the other estimate exist. Given a particular set $\{\gamma_i\}$ (e.g., a real gene found in nature), we may compute $\bar{\eta}$ and σ_{η}^2 by *assuming* that this set is a good representative of the underlying distribution of γ 's. Yet, neither of these quantities contains any information on the *location* of slow sites. Thus, we face quite a range of uncertainties when attempting to provide a good estimate. In the remainder of this article, we propose a rough and simple, yet tolerably reliable, estimate for $J(\alpha, \beta, \{\gamma_i\})$.

Since the clustering of slow sites appears to play an important role, our attempt relies on considering a “coarse-grained” set of rates. In particular, we follow the notion introduced in [88] and define

$$(K_s)_i \equiv \left[\frac{1}{s} \sum_{k=0}^{s-1} \frac{1}{\gamma_{i-k}} \right]^{-1}. \tag{10}$$

The sum in this expression is recognizable as the typical time for a (free) particle to traverse a stretch of s sites before site i . Thus, $(K_s)_i$ can be regarded as a “coarse-grained” rate associated with hopping from site i . Obviously, by setting $s = i = L$ we recover a quantity that resembles $\bar{\eta}$, but our interest is more mesoscopic, e.g., $s \sim \ell$, since that would account for some of the effects of clustering of slow sites. Combining this notion with the idea of the bottleneck being the limiting factor, we propose that $K_{\ell, \min}$, the smallest rate in $\{(K_{\ell})_i\}$, can be exploited to give a good estimate for J . Note that we are *not* proposing $J \cong K_{\ell, \min}$, since the (maximal) current for a homogeneous TASEP would be $K_{\ell, \min}(1 + \sqrt{\ell})^{-2}$, a value 20 times lower than $K_{\ell, \min}$ in the case of $\ell = 12$! Instead, our hope is that a linear relationship $J \propto K_{\ell, \min}$ would be adequate. To be specific, let us focus only on $\ell = 12$ TASEPs

⁸Actually, many exact results exist for the full stochastic problem of just a single particle hopping on such a ring. But we will not pursue this line further, since our main interest will be the many-body problem on open lattices.

I	W	A	M	S
3	1	4	1	6
AUA [12.37] AUC [12.37] AUU [12.37]	UGG [5.02]	GCC [3.57] GCA [20.97] GCG [20.97] GCU [20.97]	AUG [13.99]	AGC [5.67] AGU [5.67] UCA [7.36] UCC [4.03] UCG [8.81] UCU [11.39]
0.421	0.171	{0.122, 0.695}	0.477	{0.193, 0.251, 0.137, 0.300, 0.388}

Fig. 5 A 5-amino acid “designer gene” IWAMS with its associated degeneracies (m) in the *second* row. The *third* row shows explicitly the m synonymous codons and aa-tRNA cellular concentrations from [34]. The last row are the corresponding hopping rates used in our simulations, defined in (12)

with large entry/exit rates, i.e., $\alpha = \beta = \gamma_{\max}$, and use Monte Carlo techniques to find $J(\gamma_{\max}, \gamma_{\max}, \{\gamma_i\}; 12)$. To simplify the notation, this average current will be denoted simply by $J(\{\gamma_i\})$. Then, introducing a phenomenological slope, A , we will test how well

$$J(\{\gamma_i\}) \simeq AK_{12,\min} \tag{11}$$

is obeyed.

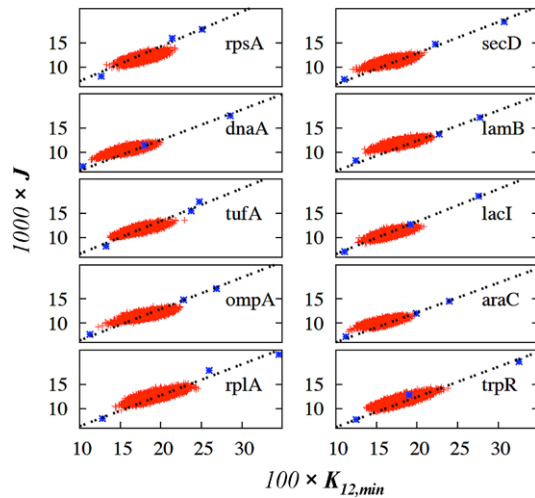
Before describing the results of such a test, let us provide some details on the ensemble of genes we will use, as well as the concept of a quenched “distribution of distributions.” As pointed out above, if we wish to synthesize a particular protein (a specific sequence of L amino acids: $\{aa_i\}$), we can use the codes from $\prod_i m(aa_i)$ different mRNAs, using the appropriate degeneracies given in (1). To help the readers, let us provide a simple example: a fictitious $L = 5$ “protein chain,” IWAMS, shown in the first row of Fig. 5. From (1), we see that the sequence of m ’s is 31416, shown in the second row. So, there are $72 (= 3 \cdot 1 \cdot 4 \cdot 1 \cdot 6)$ possible “genes” which can code for this “protein.” All the possible codons are shown in the third row, so all 72 can be read off, e.g., AUCUGGGCCAUGUCC.

One natural question is: If these 72 possibilities are generated with equal probability, what is the distribution of the currents? To answer this, we must deal with another complication. Each codon is translated with the help of one or a few (due to “wobbling”) aa-tRNAs. Their relative abundances can be measured for *E. coli*, under a specified growth condition [34], and are shown within [...] next to each codon in the third row of Fig. 5. As one can easily see, some abundances are the same, and others differ. In our simulations, we normalized the hopping rate associated with the highest abundance—29.35—to unity, and so, in the fourth row, we list the corresponding γ ’s. As a consequence, for the 72 possible “genes,” there are only 10 ($= 1 \cdot 1 \cdot 2 \cdot 1 \cdot 5$) distinct sequences of $\{\gamma_i\}$. Therefore, we should alert the reader to another complication when considering our ensemble of 72 (equally probable) “genes,” namely, the multiplicity associated with each of the 10 *distinct* sequences $\{\gamma_i\}$. As an illustration, the sequence $\{0.421, 0.171, 0.695, 0.477, 0.388\}$ is three times more likely to occur than $\{0.421, 0.171, 0.122, 0.477, 0.388\}$, since 3 codons out of 4 coding for A have the same abundance, 20.97. Hence, the $\{\gamma_i\}$ ’s are no longer equally probable, due to different multiplicities. And finally, since the $\{\gamma_i\}$ ’s determine the average production rates, the distribution of the J ’s will be non-uniform.

With this illustration in mind, let us define our notations for an explicit formulation.

- Let $\nu = 1, 2, \dots$ label the various proteins in a cell. Typically, there would be thousands. Below, we will study just 10 in *E. coli*: five highly expressed ones (dnaA, ompA, rspA,

Fig. 6 Relation between J and $K_{12,min}$ for synthesizing 10 proteins in *E. coli*. Stars (*, blue online) are from the “abysmal”, wild type and “optimal” sequences. The dash line (blue online) is the best linear fit through them and the origin. The elliptical cluster (red online) is from 5000 randomly generated sequences which code for the 10 indicated proteins



- rplA, tufA) and five poorly expressed ones (araC, lamB, lacI, secD, trpR). Each is a specific sequence of L_v amino acids, which we denote by $\{aa_i\}_v; i = 1, \dots, L_v$. Of course, $1 \leq aa \leq 20$, associated with the 20 alphabets in the first row of table (1).
- To synthesize each v , there are M_v distinct mRNAs (sequences of codons), which we denote as $\{c_i\}_v$. Let us label these sequences by $\mu = 1, \dots, M_v$. Obviously, all sequences $\{c_i\}_v$ have the same length as $\{aa_i\}_v$ so that $L_\mu = L_v$. Here, the variable c lies between 1 and 61, but c_i (its value at site i) is linked to the value of aa_i (via the aa-codon mapping). Recall that this mapping is one-to-many ($1-m_{aa}$), as the first three rows in table (1) illustrate. Thus, $M_v = \prod_{i=1}^{L_v} m_{aa_i}$ is large number, typically $O(\exp L_v)$.
 - Depending on the conditions in which a cell finds itself, different aa-tRNAs are found with varying concentrations (e.g., ref. [34] for *E. coli*). Following typical notation, we write $[c]$ for the concentration of the aa-tRNA associated with codon c . As we see in the third row of table (1), the c - $[c]$ mapping is also often one-many. For simplicity, we assume the ribosome’s hopping rate, from site i to $i + 1$, to be proportional to $[c_i]$. Normalizing these rates so that unity is associated with the largest concentration, $[\max]$, we use

$$\gamma_i \equiv \frac{[c_i]}{[\max]} \tag{12}$$

for our simulations.

With this framework in place, let us discuss our findings from performing the following simulations. For each of the 10 proteins shown above, we generated 5000 sequences $\{c_i\}_v$ with no bias, and compiled the associated $\{\gamma_i\}_v$ accordingly. For each member in this ensemble, we computed $K_{12,min}$ and simulated the associated TASEP (with $\ell = 12$ particles) to obtain its current J ($\{\gamma_i\}_v$). These pairs of values are plotted in the J - K plane. They generally form an elliptical cluster (5K indiscernible points, red online), as shown in Fig. 6 for each of the 10 proteins.

Two other features appear in these plots: a dashed line and three points (stars), all being blue online. The lowest point corresponds to an “abysmal” sequence, formed by having the lowest allowed γ at each site. Thus, it produces the lowest possible (J, K) . Similarly, the highest point is associated with an “optimal” sequence, with the highest possible (J, K)

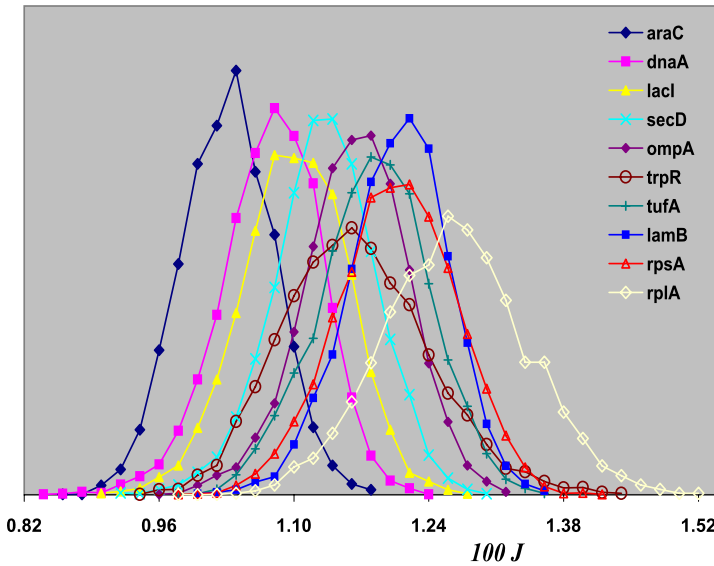


Fig. 7 Current distributions for 5000 TASEP sequences (modeling the silent mutations) which code for the 10 indicated proteins

for this protein. The point in the middle is derived from the wild type (naturally occurring) sequence. Finally, the dashed line is the best linear fit through the three points, *constrained* to pass through the origin ($J = K = 0$). Remarkably, the 5K points lie reasonably close to the dashed line, giving us hope that the expression (11) might be quite good.

Before detailing quantitative aspects of the analysis, let us comment on a remarkable aspect of this data. From the 5K simulated J 's, we compiled histograms to form a current distribution for each v . Shown in Fig. 7, no major surprises are apparent: All distributions seem normal, with means in the approximate range of 1.00–1.25 (for $100 \times J$) and standard deviations of 0.05–0.10. Their skewness and kurtosis (both unitless measures of deviations from pure Gaussians) fall in the ranges of, respectively, $[-0.3, 0.3]$ and $[-0.1, 0.4]$. However, with even a casual glance at the panels in Fig. 6, the reader may notice that all but two of the wild types lie well *above* the cluster of 5K points. Indeed, five of them are more than 6.5 standard deviations above the mean. We can only speculate that natural evolution optimized the production rates! Work is in progress to study the rest of the 5416 proteins in *E. coli* and, if this systematic deviation persists, to consider possible deeper underpinnings of this phenomenon.

Returning to the more practical issue at hand, we seek a quantitative description in an attempt to test expression (11). For a particular protein v , we consider an ensemble in which all M_v sequences $\{c_i\}_v$ are equally probable. However, as illustrated by the last two rows of table (1), each distinct $\{\gamma_i\}_v$ sequence can result from several $\{c_i\}_v$ sequences (depending on conditions on the cell, and other complications which we ignore here). Due to this complicated $c_i-\gamma_i$ connection, there are far fewer $\{\gamma_i\}_v$ sequences, so that the distribution of γ 's, which we denote by $\mathcal{P}_v[\{\gamma_i\}]$, will not be trivial. Since J depends only on $\{\gamma_i\}$ and not $\{c_i\}$, this \mathcal{P}_v will control the average J over our ensemble. Of course, $\mathcal{P}_v[\{\gamma_i\}]$ is still a product distribution, $\prod_{i=1}^{L_v} p_i(\gamma_i)$, since no correlations between sites are assumed. But, unlike previous studies, the γ 's here must be chosen from *site-dependent* distributions—thus the subscript i on p_i . To clarify, let us return to our illustration, in which

$\mathcal{P}_{\text{IWAMS}}[\{\gamma_i\}] = \prod_{i=1}^5 p_i(\gamma_i)$ with, explicitly,

$$\begin{aligned} p_1(\gamma) &= \delta(\gamma - 0.421) \\ p_2(\gamma) &= \delta(\gamma - 0.171) \\ p_3(\gamma) &= \{\delta(\gamma - 0.122) + 3\delta(\gamma - 0.695)\}/4 \\ p_4(\gamma) &= \delta(\gamma - 0.477) \\ p_5(\gamma) &= \{\delta(\gamma - 0.137) + 2\delta(\gamma - 0.193) + \delta(\gamma - 0.251) \\ &\quad + \delta(\gamma - 0.300) + \delta(\gamma - 0.388)\}/6. \end{aligned}$$

Since the sequence of p_i 's are fixed by the amino acid sequence, $\{aa_i\}$, we are constrained by a quenched distribution of different p 's. Thus, we arrive at the notion of a quenched "distribution of distributions."

With this framework in mind, we can define another average of J , associated with all possible ways of producing protein ν in our ensemble of M_ν (equally probable) silent mutations:

$$\langle J \rangle_\nu = \int \mathcal{D}\gamma J(\{\gamma_i\}) \mathcal{P}_\nu[\{\gamma_i\}]$$

where $\mathcal{D}\gamma$ denotes $\prod_{i=1}^{L_\nu} d\gamma_i$. In a similar vein, $K_{12,\min}$ also depends only on $\{\gamma_i\}$ (and not $\{c_i\}$), so that

$$\langle K_{12,\min} \rangle_\nu = \int \mathcal{D}\gamma K_{12,\min}(\{\gamma_i\}) \mathcal{P}_\nu[\{\gamma_i\}]$$

The simulations for each ν in Fig. 6 are a 5K-point sampling of this \mathcal{P}_ν . Thus, the coordinates of the "center of mass" of the (roughly elliptical) cluster are just $\langle J \rangle_\nu$ and $\langle K_{12,\min} \rangle_\nu$. Of course, we can consider other quantities of interest, such as $\langle \delta(J - J(\{\gamma_i\})) \rangle_\nu$, corresponding to the histograms in Fig. 7. Other obvious possibilities are the second moments, which will provide us with the two axes and the orientation of each cluster, as well as a measure of the J - K correlation. Here, we are content to focus only the averages and their ratios, $\langle J \rangle_\nu / \langle K_{12,\min} \rangle_\nu$, for these 10 proteins. Remarkably, though both $\langle J \rangle_\nu$ and $\langle K \rangle_\nu$ range by 25% (over the 10 ν 's), this ratio is essentially constant! This observation motivates us to define A in (11) by a further average:

$$A \equiv \frac{1}{10} \sum_\nu \frac{\langle J \rangle_\nu}{\langle K_{12,\min} \rangle_\nu}.$$

From our data, we find

$$A \cong 0.0656$$

which is, interestingly, comparable to $(1 + \sqrt{12})^{-2} \cong 0.0502$. As a test of its "predictive power," we computed $AK_{12,\min}$ for all 5000×10 $\{\gamma\}$'s and compared them to the values of the currents obtained from simulations. Specifically, the average of (these 50K values of) $AK_{12,\min}/J$ is within 0.4% of unity, while the standard deviation is about 5%. Rarely does this ratio range more than 15% from unity. In this sense, we are hopeful that, when we extend this study to the other 5406 genes in *E. coli*, we will confirm $AK_{12,\min}(\{\gamma_i\})$ as a simple and reliable estimate for $J(\{\gamma_i\})$. It is truly remarkable that the value of A appears to be universal for different genes, and possibly even across the whole genome! It remains to be

seen whether this holds true for *E. coli* under different growth conditions or other bacteria, leading to tests of this seeming universality.

Despite being quite involved and extensive, this study has answered few questions in biology, since it remains quite remote from the goal of understanding protein synthesis in real cells. It does, however, open up rich new ground for exploring nonequilibrium statistical mechanics. From the biological perspective, the main insight can be summarized as follows: Should we wish to design a silent mutation of a gene that will vastly outperform the wild type, either by enhancing or suppressing the production rate of this protein, then we have discovered a reliable and simple method to search for the “perfect” mutation.

5 Conclusions and Outlook

In this article, we have touched upon two fundamental issues in two very different fields: understanding nonequilibrium steady states and developing quantitative models for protein production. These two seemingly disparate problems converge in a simple one-dimensional transport model, the totally asymmetric simple exclusion process and its modifications. The TASEP is a paradigmatic far-from-equilibrium model, characterized by open boundaries and a systematic particle current through the system. Due to the exclusion, the particles are interacting, and so it is highly non-trivial to find steady-state and dynamic properties. Still, a considerable body of exact results is available for the standard TASEP. In particular, despite the one-dimensional nature of the model, it displays three distinct phases, separated by first order and continuous transitions. If the model is modified to include extended particles and inhomogeneous hopping rates, it is generally accessible only via simulations or approximate (mean-field) methods.

With these modifications, the model becomes a more realistic—but still highly simplified—description of protein synthesis. The one-dimensional lattice models the mRNA template, with sites and extended particles representing codons and ribosomes, respectively. Further, we allow non-uniform hopping rates, to reflect the variability of the aa-tRNA concentrations associated with different codons. The particle current through the TASEP is simply the protein production rate. An interesting feature of translation is the sophisticated degeneracy: 61 codons code for 20 amino acids (mediated by 46 tRNAs in *E. coli*). In other words, there are many distinct sequences (“silent mutations”) which code for the same protein but are characterized by different production rates.

In this article, we presented a brief introduction to the main findings for TASEP and the basics of protein synthesis, designed with non-experts in mind. We also described the modeling of translation in terms of a generalized TASEP, summarizing both well-established and more recent results. Amongst the latter we discussed two specific topics: first, the effects of limited availability of particles, and second, simple but remarkably good estimates for currents in the fully inhomogeneous case. The first project is motivated by the observation that ribosomes are large molecules so that their synthesis is costly for the cell. Hence, it is reasonable to expect them to be in limited supply. Considering only the simplest case—fully uniform rates and particles covering only one site—we asked: How are currents and density profiles affected if a single, or several, TASEPs compete for particles from a finite reservoir? Remarkable results, such as multiple, distinct regimes in density profiles and shock localization were discovered. The second discussion centered on two questions: Is it possible to arrive at simple yet reliable estimates for currents associated with fully inhomogeneous sequences? And how do the currents associated with the “ensemble” of silent mutations compare to that of the wild type? The answer to the first question relies on computing the

typical time any particular codon is covered by a ribosome. In the language of TASEP, this is the time it takes a particle to traverse a stretch of 12 sites around a given site. This quantity can be determined from sequence information with minimal effort (provided the aa-tRNA concentrations are known). Its inverse is effectively a coarse-grained rate associated with hopping from that site. It turns out that the lowest of these rates (in a given sequence), denoted by $K_{12,\min}$, provides a good estimate for the average current. Specifically, Monte Carlo simulations for 5000 randomly selected silent mutations of 10 different proteins show a reliable linear relation between currents and $K_{12,\min}$'s, with a proportionality constant that appears to be the *same* for all the proteins studied. Moreover, the current and $K_{12,\min}$ of the wild type also obey this linear relation even though both fall well above the typical values for randomly chosen sequences.

Clearly, the explorations reported here leave many questions unanswered, both on the statistical physics and the biology side. We just cite a few which will hopefully spark future research. The central fundamental question concerns the “stability” of steady-state properties with respect to model modifications. Which changes of microscopic model details (e.g., hopping rates) will lead to changes of microscopic or macroscopic behaviors? While notions of universality and independence from certain dynamic details are well understood for equilibrium systems, we have taken only initial steps towards extending them to nonequilibrium steady states [100, 101]. Further, little if anything is known about how these general concepts apply to specific models. On the quantitative biology side, even relatively simple questions remain open: Are aa-tRNA concentrations really the limiting factor for protein production rates? Are there other intrinsic rates, or is initiation the critical bottle neck? Secondary structures are known to be important [65], but how exactly do they affect production rates? Why are the currents of wild type genes so optimized? Clearly, fundamental insights and close collaborations between physicists and biologists are needed before we will begin to understand biological processes—which are generically far from equilibrium—at a quantitative level.

Acknowledgements We are grateful for numerous illuminating discussions with many colleagues, especially T. Chou, L.J. Cook, C.V. Finkelstein, R.J. Harris, K. Mallick, and L.B. Shaw. We thank J.L. Lebowitz for the invitation to participate in this special issue of the Journal of Statistical Physics. This work is supported in part by grants from the NSF: DMR-0705152 and DMR-1005417.

References

1. Adams, D.A., Schmittmann, B., Zia, R.K.P.: Far-from-equilibrium transport with constrained resources. *J. Stat. Mech. Theory Exp.* **2009**, P06009 (2008)
2. Adams, D.A., Zia, R.K.P., Schmittmann, B.: Power spectra of the total occupancy in the totally asymmetric simple exclusion process. *Phys. Rev. Lett.* **99**, 020601 (2007)
3. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P.: *How Cells Read the Genome: From DNA to Protein*. Garland Science, New York (2007)
4. Alcaraz, F.C., Lazo, M.J.: The exact solution of the asymmetric exclusion problem with particles of arbitrary size: Matrix product ansatz. *Braz. J. Phys.* **33**, 533 (2003)
5. Antoun, A., Pavlov, M.Y., Lovmar, M., Ehrenberg, M.: How initiation factors tune the rate of initiation of protein synthesis in bacteria. *EMBO J.* **25**(11), 2539–2550 (2006)
6. Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Orlandi, A., Parisi, P., Procaccini, A., Viale, M., Zdravkovic, V.: Empirical investigation of starling flocks: a benchmark study in collective animal behaviour. *Anim. Behav.* **76**, 201 (2008)
7. Belitzky, V., Schütz, G.M.: Diffusion and scattering of shocks in the partially asymmetric exclusion process. *Electron. J. Probab.* **7**(11), 1 (2002)
8. Bengrine, M., Benyoussef, A., Ez-Zahraoui, H., Mhirech, F.: Traffic model with quenched disorder. *Phys. Lett. A* **253**(3–4), 135 (1999)

9. Beyer, D., Skripkin, E., Wadzack, J., Nierhaus, K.H.: How the ribosome moves along the mrna during protein synthesis. *J. Biol. Chem.* **269**(48), 30713–30717 (1994)
10. BioStudio. www.biostudio.com/demo_freeman_protein_synthesis.htm
11. Blythe, R.A., Evans, M.R.: Nonequilibrium steady states of matrix-product form: a solver's guide. *J. Phys. A, Math. Gen.* **40**(46), R333 (2007)
12. Chou, T.: Ribosome recycling, diffusion, and mrna loop formation in translational regulation. *Biophys. J.* **85**, 755 (2010)
13. Chou, T., Lakatos, G.: Totally asymmetric exclusion processes with particles of arbitrary size. *J. Phys. A, Math. Gen.* **36**, 2027 (2003)
14. Chou, T., Lakatos, G.: Clustered bottlenecks in mrna translation and protein synthesis. *Phys. Rev. Lett.* **93**, 198101 (2004)
15. Chowdhury, D., Santen, L., Schadschneider, A.: Statistical physics of vehicular traffic and some related systems. *Phys. Rep.* **329**, 199 (2000)
16. Chowdhury, D., Santen, L., Schadschneider, A.: Vehicular traffic: A system of interacting particles driven far from equilibrium. *Curr. Sci.* **77**, 411 (2000)
17. Basic Energy Sciences Advisory Committee: Directing Matter and Energy: Five Challenges for Science and the Imagination. Department of Energy Publications, Washington (2007)
18. Cook, L.J.: private communication (2010)
19. Cook, L.J., Zia, R.K.P.: Feedback and fluctuations in a totally asymmetric simple exclusion process with finite resources. *J. Stat. Mech. Theory Exp.* **2009**, P02012 (2009)
20. Cook, L.J., Zia, R.K.P.: Power spectra of a constrained totally asymmetric simple exclusion process. *J. Stat. Mech. Theory Exp.* **2010**, P07014 (2010)
21. Cook, L.J., Zia, R.K.P., Schmittmann, B.: Competition between many totally asymmetric simple exclusion processes for a finite pool of resources. *Phys. Rev. E* **80**(3), 031142 (2009)
22. Crick, F.H.C.: On protein synthesis. *Symp. Soc. Exp. Biol.* **XII**, 139–163 (1958)
23. Crick, F.H.C.: Codon–anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* **19**(2), 548–555 (1966)
24. Crick, F.H.C.: Central dogma of molecular biology. *Nature* **227**, 561–563 (1970)
25. Czworkowski, J., Moore, P.B.: The elongation phase of protein synthesis. *Prog. Nucleic Acid Res. Mol. Biol.* **54**, 293–332 (1996)
26. de Gier, J., Essler, F.H.L.: Exact spectral gaps of the asymmetric exclusion process with open boundaries. *J. Stat. Mech. Theory Exp.* **2006**, P12011 (2006)
27. De Masi, A., Ferrari, P.A.: Self diffusion in one dimensional lattice gases in the presence of an external field. *J. Stat. Phys.* **38**, 603 (1985)
28. Derrida, B.: Non-equilibrium steady states: fluctuations and large deviations of the density and of the current. *J. Stat. Mech.* **2007**, P07023 (2007)
29. Derrida, B., Domany, E., Mukamel, D.: An exact solution of a one-dimensional asymmetric exclusion model with open boundaries. *J. Stat. Phys.* **69**, 667 (1992)
30. Derrida, B., Evans, M.R., Pasquier, V.: Exact solution of a 1d asymmetric exclusion model using a matrix formulation. *J. Phys. A, Math. Gen.* **26**, 1493 (1993)
31. Derrida, B., Evans, M.R., Hakim, V., Pasquier, V.: Exact solution of a 1d asymmetric exclusion model using a matrix formulation. *J. Phys. A, Math. Gen.* **26**, 1493 (1993)
32. Derrida, B., Gerschenfeld, A.: Current fluctuations of the one dimensional symmetric simple exclusion process with step initial condition. *J. Stat. Phys.* **136**(1), 1 (2009)
33. Derrida, B., Janowsky, S.A., Lebowitz, J.L., Speer, E.R.: Exact solution of the totally asymmetric simple exclusion process: shock profiles. *J. Stat. Phys.* **73**, 813 (1993)
34. Dong, H., Nilsson, L., Kurland, C.G.: Co-variation of trna abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.* **260**(5), 649–63 (1996)
35. Dong, J.J., Schmittmann, B., Zia, R.K.P.: Inhomogeneous exclusion processes and protein synthesis. *Phys. Rev. E* **76**, 051113 (2007)
36. Dong, J.J., Schmittmann, B., Zia, R.K.P.: Towards a model for protein production rates. *J. Stat. Phys.* **128**(1), 21 (2007)
37. Dong, J.J., Zia, R.K.P., Schmittmann, B.: Understanding the edge effect in tasep with mean-field theoretic approaches. *J. Phys. A, Math. Theor.* **42**(1), 015002 (2009)
38. Dotsenko, V.: Introduction to the Replica Theory of Disordered Statistical Systems. Cambridge University Press, Cambridge (2001)
39. Dudzinski, M., Schütz, G.M.: Relaxation spectrum of the asymmetric exclusion process with open boundaries. *J. Phys. A, Math. Gen.* **33**(47), 8351 (2000)
40. Edwards, S.F., Anderson, P.W.: Theory of spin glasses. *J. Phys. F, Met. Phys.* **5**, 965 (1975)
41. Feder, T.: Statistical physics is for the birds. *Phys. Today* **60**(10), 28 (2007)
42. Ebrahim Foulaadvand, M., Kolomeisky, A.B., Teymouri, H.: Asymmetric exclusion processes with disorder: Effect of correlations. *Phys. Rev. E* **78**(6), 061116 (2008)

43. Fraser, C.S., Hershey, J.W.B.: Movement in ribosome translocation. *J. Biol.* **4**(2), 8 (2005)
44. Friedman, M.H.: Principles and Models of Biological Transport. Springer, Berlin (2008)
45. Golinelli, O., Mallick, K.: Bethe ansatz calculation of the spectral gap of the asymmetric exclusion process. *J. Phys. A, Math. Gen.* **37**(10), 3321 (2004)
46. Golinelli, O., Mallick, K.: Spectral gap of the totally asymmetric exclusion process at arbitrary filling. *J. Phys. A, Math. Gen.* **38**(7), 1419 (2005)
47. Golinelli, O., Mallick, K.: The asymmetric simple exclusion process: an integrable model for non-equilibrium statistical mechanics. *J. Phys. A, Math. Gen.* **39**(41), 12679 (2006)
48. Greulich, P., Schadschneider, A.: Single-bottleneck approximation for driven lattice gases with disorder and open boundary conditions. *J. Stat. Mech. Theory Exp.* **2008**, P04009 (2008)
49. Gupta, S., Majumdar, S.N., Godrèche, C., Barma, M.: Tagged particle correlations in the asymmetric simple exclusion process: Finite-size effects. *Phys. Rev. E* **76**, 021112 (2007)
50. Gwa, L.-H., Spohn, H.: Bethe solution for the dynamical scaling exponent of the noisy burgers equation. *Phys. Rev. A* **46**, 844 (1992)
51. Ha, M., den Nijs, M.: Macroscopic car condensation in a parking garage. *Phys. Rev. E* **66**, 036118 (2002)
52. Harris, R.J., Stinchcombe, R.B.: Disordered asymmetric simple exclusion process: Mean-field treatment. *Phys. Rev. E* **70**, 016108 (2004)
53. Heinrich, R., Rapoport, T.A.: Mathematical modelling of translation of mrna in eucaryotes; steady states, time-dependent processes and application to reticulocytes. *J. Theor. Biol.* **86**(2), 279–313 (1980)
54. Howard, J.: Mechanics of Motor Proteins and the Cytoskeleton. Sinauer, Sunderland (2001)
55. Ikemura, T.: Correlation between the abundance of Escherichia coli transfer mnas and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.* **146**(1), 1–21 (1981)
56. Janowsky, S.A., Lebowitz, J.L.: Finite-size effects and shock fluctuations in the asymmetric simple-exclusion process. *Phys. Rev. A* **45**, 618 (1992)
57. Kim, D.: Bethe ansatz solution for crossover scaling functions of the asymmetric xxz chain and the Kardar-Parisi-Zhang-type growth model. *Phys. Rev. E* **52**, 3512 (1995)
58. Kisselev, L.L., Frolova, L.Y.: Termination of translation in eukaryotes: new results and new hypotheses. *Biochemistry (Moscow)* **64**, 8–16 (1999)
59. Kolomeisky, A.B., Fisher, M.E.: Molecular motors: a theorist's perspective. *Annu. Rev. Phys. Chem.* **58**, 675 (2007)
60. Kolomeisky, A.B., Schütz, G.M., Kolomeisky, E.B., Straley, J.P.: Phase diagram of one-dimensional driven lattice gases with open boundaries. *J. Phys. A, Math. Gen.* **31**(33), 6911 (1998)
61. Kolomeisky, A.B.: Asymmetric simple exclusion model with local inhomogeneity. *J. Phys. A, Math. Gen.* **31**(4), 1153 (1998)
62. Kozak, M.: Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiol. Rev.* **47**(1), 1–45 (1983)
63. Krug, J.: Boundary-induced phase transitions in driven diffusive systems. *Phys. Rev. Lett.* **67**, 1882 (1991)
64. Krug, J.: Phase separation in disordered exclusion models. *Braz. J. Phys.* **30**, 97–104 (2000)
65. Kudla, G., Murray, A.W., Tollervey, D., Plotkin, J.B.: Coding-sequence determinants of gene expression in Escherichia coli. *Science* **324**(5924), 255–258 (2009)
66. Kutner, R., van Beijeren, H.: Influence of an external force on tracer diffusion in a one-dimensional lattice gas. *J. Stat. Phys.* **39**, 317 (1985)
67. Laursen, B.S., Sørensen, H.P., Mortensen, K.K., Sperling-Petersen, H.U.: Initiation of protein synthesis in bacteria. *Microbiol. Mol. Biol. Rev.* **69**(1), 101 (2005)
68. Maaloe, O.: Regulation of the protein-synthesizing machinery—ribosomes, trna, factors and so on. In: Goldberger, R.F. (ed.) *Biological Regulation and Development*. Plenum, New York (1979)
69. MacDonald, C.T., Gibbs, J.H.: Concerning the kinetics of polypeptide synthesis on polyribosomes. *Biopolymers* **7**, 707 (1969)
70. MacDonald, C.T., Gibbs, J.H., Pipkin, A.C.: Kinetics of biopolymerization on nucleic acid templates. *Biopolymers* **6**(1) (1968)
71. Majumdar, S.N., Barma, M.: Tag diffusion in driven systems, growing interfaces, and anomalous fluctuations. *Phys. Rev. B* **44**, 5306 (1991)
72. Mallick, K.: Shocks in the asymmetry exclusion model with an impurity. *J. Phys. A, Math. Gen.* **29**, 5375 (1996)
73. Merrick, W.: Mechanism and regulation of eukaryotic protein synthesis. *Microbiol. Mol. Biol. Rev.* **56**(2), 291 (1992)
74. Moldave, K.: Eukaryotic protein synthesis. *Annu. Rev. Biochem.* **54**(1), 1109–1149 (1985)
75. Nagy, Z., Appert, C., Santen, L.: Relaxation times in the asep model using a dmrg method. *J. Stat. Phys.* **109**, 623 (2002)

76. Nakamura, Y., Ito, K.: How protein reads the stop codon and terminates translation. *Genes Cells* **3**, 265–278 (1998)
77. Negrutskii, B., El'Skaya, A.: Eukaryotic translation elongation factor 1 [alpha]: Structure, expression, functions, and possible role in aminoacyl-trna channeling. *Prog. Nucleic Acid Res.* **60**, 47–78 (1998)
78. Nygård, O., Nilsson, L.: Translational dynamics: Interactions between the translational factors, trna and ribosomes during eukaryotic protein synthesis. *Eur. J. Biochem.* **191**(1), 1–17 (1990)
79. Board on Physics and Astronomy: Condensed-Matter and Materials Physics: The Science of the World Around Us. Academies Press, Washington (2007)
80. Pestova, T., Hellen, C.: The structure and function of initiation factors in eukaryotic protein synthesis. *Cell. Mol. Life Sci.* **57**(4), 651 (2000)
81. Pierobon, P., Parmeggiani, A., von Oppen, F., Frey, E.: Dynamic correlation functions and Boltzmann Langevin approach for driven one dimensional lattice gas. *Phys. Rev. E* **72**, 036123 (2005)
82. Popkov, V., Santen, L., Schadschneider, A., Schütz, G.M.: Boundary-induced phase transitions in traffic flow. *J. Phys. A, Math. Gen.* **34**, L45 (2001)
83. Riis, B., Rattan, S., Clark, B., Merrick, W.: Eukaryotic protein elongation factors. *Trends Biochem. Sci.* **15**, 420–424 (1990)
84. Santen, L., Appert, C.: The asymmetric exclusion process revisited: Fluctuations and dynamics in the domain wall picture. *J. Stat. Phys.* **106**, 187 (2002)
85. Schliwa, M., Woehlke, G.: Molecular motors. *Nature* **422**, 759 (2003)
86. Schütz, G.M.: Time-dependent correlation functions in a one-dimensional asymmetric exclusion process. *Phys. Rev. E* **47**, 4265 (1993)
87. Schütz, G.M., Domany, E.: Phase transitions in an exactly solvable one-dimensional exclusion process. *J. Stat. Phys.* **72**, 277 (1993)
88. Shaw, L.B., Kolomeisky, A.B., Lee, K.H.: Local inhomogeneity in asymmetric simple exclusion processes with extended objects. *J. Phys. A, Math. Gen.* **37**, 2105 (2004)
89. Shaw, L.B., Zia, R.K.P., Lee, K.H.: Modeling, simulations, and analyses of protein synthesis: Driven lattice gas with extended objects. *Phys. Rev. E* **68**, 021910 (2003)
90. Shine, J., Dalgarno, L.: Determinant of cistron specificity in bacterial ribosomes. *Nature* **254**(5495), 34 (1975)
91. Sørensen, M.A., Pedersen, S.: Absolute in vivo translation rates of individual codons in *Escherichia coli*: The two glutamic acid codons gaa and gag are translated with a threefold difference in rate. *J. Mol. Biol.* **222**(2), 265–280 (1991)
92. Spector, L., Klein, J., Perry, C., Feinstein, M.: Emergence of collective behavior in evolving populations of flying agents. In: Cantu-Paz, E., Foster, J.A., Deb, K., Davis, L.D., Roy, R., O'Reilly, U.-M., Beyer, H.-G., Standish, R., Kendall, G., Wilson, S., Harman, M., Wegener, J., Dasgupta, D., Potter, M.A., Schultz, A.C., Dowsland, K.A., Jonoska, N., Miller, J. (eds.) *Proceedings of the Genetic and Evolutionary Computation Conference*. Springer, Berlin (2003)
93. Spitzer, F.: Interaction of Markov processes. *Adv. Math.* **5**, 246 (1970)
94. Takesue, S., Mitsudo, T., Hayakawa, H.: Power-law behavior in the power spectrum induced by Brownian motion of a domain wall. *Phys. Rev. E* **68**, 015103(R) (2003)
95. Tripathy, G., Barma, M.: Driven lattice gases with quenched disorder: Exact results and different macroscopic regimes. *Phys. Rev. E* **58**, 1998 (1998)
96. Varenne, S., Buc, J., Lloubes, R., Lazdunski, C.: Translation is a non-uniform process: Effect of trna availability on the rate of elongation of nascent polypeptide chains. *J. Mol. Biol.* **180**(3), 549–576 (1984)
97. Wikipedia. Genetic code. en.wikipedia.org/wiki/Genetic_code
98. Young, A.P. (ed.): *Spin Glasses and Random Fields*. World Scientific, Singapore (1998)
99. Zia, R.K.P., Dong, J.J., Schmittmann, B.: Estimating currents in totally asymmetric simple exclusion process with extended particles and inhomogeneous hopping rates. meetings.aps.org/Meeting/MAR09/Event/93724
100. Zia, R.K.P., Schmittmann, B.: A possible classification of nonequilibrium steady states. *J. Phys. A, Math. Gen.* **39**(41), L407–L413 (2006)
101. Zia, R.K.P., Schmittmann, B.: Probability currents as principal characteristics in the statistical mechanics of nonequilibrium steady states. *J. Stat. Mech. Theory Exp.* **2007**, P07012 (2007)