Symmetry in biology: from genetic code to stochastic gene regulation

A.F. Ramos[1]   G.C.P. Innocentini[1]   F.M. Forger[2]   J.E.M. Hornos[1]

[1]Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, BR-13560-970 São Carlos, S.P., Brazil
[2]Instituto de Matemática e Estatística, Universidade de São Paulo, Caixa Postal 66281, BR-05315-970 São Paulo, S.P., Brazil
E-mail: aramos@ifsc.usp.br

Abstract: Mathematical models, as instruments for understanding the workings of nature, are a traditional tool of physics, but they also play an ever increasing role in biology – in the description of fundamental processes as well as that of complex systems. In this review, the authors discuss two examples of the application of group theoretical methods, which constitute the mathematical discipline for a quantitative description of the idea of symmetry, to genetics. The first one appears, in the form of a pseudo-orthogonal (Lorentz like) symmetry, in the stochastic modelling of what may be regarded as the simplest possible example of a genetic network and, hopefully, a building block for more complicated ones: a single self-interacting or externally regulated gene with only two possible states: 'on' and 'off'. The second is the algebraic approach to the evolution of the genetic code, according to which the current code results from a dynamical symmetry breaking process, starting out from an initial state of complete symmetry and ending in the presently observed final state of low symmetry. In both cases, symmetry plays a decisive role: in the first, it is a characteristic feature of the dynamics of the gene switch and its decay to equilibrium, whereas in the second, it provides the guidelines for the evolution of the coding rules.

1   Introduction

Compared with the huge amount of experimental data on genetics accumulated in the last few decades, which has led some biologists to refer to our present times as the genomic era, it is not unfair to say that true information and knowledge about fundamental biological processes and structures has not increased proportionally. A striking example is gene expression: the sequencing of entire genomes answers the question which protein will be synthesised from a specific gene, but not when and at what rate its synthesis will take place. However, it is becoming increasingly clear how crucial the precise temporal and spatial variation of gene expression is for pattern formation [1] and how enormously complex and strongly species dependent the underlying mechanisms are. Therefore we can think of gene networks as complex dynamical systems whose basic rules are still to be unravelled. From a theoretical point of view, this subject is still in an embryonic stage, since the main theoretical problem is not only the intrinsic complexity of the dynamics of living

matter but even the absence of simple quantitative models that can serve as building blocks and starting points for more elaborate, more detailed and more realistic models.

Our aim in the present review is to demonstrate that the principle of symmetry and its mathematical implementation through group theoretical methods can play a constructive role in the development of such simple quantitative models, and more generally, of models that are relevant for systems biology, even in situations where no symmetry is visible at first sight.

The usefulness of symmetry principles in arts and sciences, including biology, has a long history, with many facets, as can be seen, for example, from the inspiring and entertaining collection of articles in [2, 3] and the books [4, 5]. There can be no doubt that group theoretical methods are able to make concrete statements about the behaviour of complex systems; in fact, it can (and has) been argued that group theory by itself is in some sense a complex system [6].

As a concrete example from biology, think of shape symmetries of flowers, which according to Linnaeus' classification are an essential ingredient for flower taxonomy. The flowers of *Linaria vulgaris*, a normal toadflax, present a low degree of symmetry, with just one mirror plane of invariance. The observation of a rare toadflax whose flowers exhibit a much higher degree of symmetry, with multiple mirror planes of invariance, suggested that one was dealing with a new and different species. However, Linnaeus classified it as a variant of the common toadflax and called it *Peloria* (a greek word for monster), impressed by the fact that the almost asymmetric flower was common and the highly symmetric one was rare. This interpretation has been shown to be correct two hundred years later, with the advent of molecular biology, when it was found that in *Linaria vulgaris*, the gene cycloidea presents a gradient of activation in the young flower bud, which causes an asymmetric development of the flower organs, whereas in the *Peloria* mutants, this gene is completely switched off, thus permitting a symmetric development of the flower. Thus regular flowers can be thought of as ancestors of irregular ones, which reveals the evolutionary aspect of the concept of symmetry breaking in floral symmetry [7–9].

Another example from biology is the symmetry of virus capsids, which is a consequence of energy minimisation at the molecular level [10]. This was realised when it was found that isolated viral particles can spontaneously assemble into rod-shaped, fully functional and infective viruses, indicating that stability of the proteic coat is achieved in a configuration with a certain symmetry. Concretely, an icosahedral symmetry was predicted as a consequence of building up the capsid from identical units [11].

Apart from symmetries of objects, described by point groups and exhaustively used in crystallography, there is also a notion of symmetry in processes, sometimes called dynamical symmetries, which are present in the equations governing a phenomenon. It is to this kind of symmetry that we shall direct our attention in what follows, beginning with the remark that such symmetries play a central role in the development of modern physics. For example, the understanding of atomic spectra from the standpoint of quantum mechanics requires solving the Schrödinger equation in a central force potential, whose rotational symmetry forces the energy states to be organised in multiplets (= irreducible representations) of the rotation group. However, the seminal work of Pauli [12] showed that, in the case of the hydrogen atom, rotational symmetry alone is not sufficient to account for all the degeneracies of the energy spectrum. There is an additional degeneracy with the total angular momentum number l that can only be explained by including the so-called Lenz–Runge vector – a second conserved quantity besides the angular momentum vector – and thus extending the symmetry group of the Coulomb potential from the three-dimensional rotation group SO(3) to the four-dimensional

rotation group SO(4): this is the classical example of a dynamical symmetry. Since then, group theoretical techniques have become useful in a wide variety of physical contexts, including more complex systems than just the hydrogen atom. For example, in atomic physics they can be used to derive a composition rule between hydrogen-like wave functions to furnish an approximate description of the ground state of multi-electronic atoms [13]. At a slightly larger scale, in molecular physics, they are employed in the theoretical calculation of spectra and transition amplitudes [14, 15]. In the opposite direction (i.e. at a much smaller scale), symmetry principles have become a fundamental ingredient in particle physics ever since Gell-Mann and Ne'eman proposed the first classification of particles in terms of multiplets (= irreducible representations) of the group SU(3), the special unitary group in three dimensions, a classification that has become known as the eightfold way [16] and that has opened the way to the so-called standard model of particle physics, which is largely based on group theoretical methods and on the idea of symmetry.

In this review, we consider the application of the idea of a dynamical symmetry to the modelling of gene expression and of genetic code evolution. Our goal is to illustrate with these two examples that group theoretical methods can be used, widely and sometimes perhaps unexpectedly, in biology.

The phenomena of gene transcription and translation are distinguished by an intricate interaction between many elements and by fluctuations that are inherent in chemical reactions [17–64]. Here, the mesoscopic equations of an isolated bimodal gene are shown to possess a symmetry described by the pseudo-orthogonal group SO(2, 1) (similar to the Lorentz group SO(3, 1)), whose representation labels we continue to call 'total angular momentum' and 'azimuthal angular momentum' even though they are now arbitrary real numbers, in contrast to the usual orthogonal group SO(3) for which these quantities are integers (or at best half-integers). This symmetry has clear biological implications, and there is a potential use of group theoretical tools in the elaboration of a composition principle between two or more genes.

On the other hand, there is the question of how the genetic code has evolved [65–89]. It seems to be generally agreed that this evolution has occurred through a process of stepwise incorporation of new amino acids into the machinery of cellular protein synthesis. In the algebraic approach to the question of how precisely this has happened, the existence of synonymous codons is interpreted as a signal of symmetry – a symmetry that was complete at the very beginning but then underwent a dynamical process of symmetry breaking, which finally came to a halt with the onset of 'freezing' [65]. This approach establishes strong restrictions on the routes that could have been followed during code evolution.

The review is organised as follows. In Section 2, we collect some of the necessary concepts from the theory of Lie groups and Lie algebras. Section 3 presents a detailed review of a simple stochastic model for gene expression, which involves only two levels of expression, namely 'on' and (almost) 'off', of its symmetry and of the biological consequences. Section 4 is devoted to a short and non-technical overview of the development of the algebraic model to genetic code evolution. In Section 5, we outline our conclusions.

2 Mathematical theory of symmetries

In this section, we present some concepts of group theory, inasfar as they are essential to the understanding of what follows. Our discussion will be informal and is based on examples; readers interested in a more formal treatment should consult standard mathematical references, such as, for example [90–92].

Abstractly, a group is a set G whose elements g can be multiplied among themselves, subject to certain constraints such as the associativity rule $((g_1 g_2)g_3 = g_1(g_2 g_3))$, the existence of a (necessarily unique) two-sided identity in G denoted by 1 ($g1 = g = 1g$) and the existence of a (necessarily unique) two-sided inverse g^{-1} for any g ($gg^{-1} = 1 = g^{-1}g$). In this review, we shall consider only continuous symmetries that are described by Lie groups. The prototype of such a group is the n-dimensional rotation group SO(n). It is composed of ($n \times n$)-matrices $\boldsymbol{g}$ which are orthogonal, that is, satisfy $\boldsymbol{g}^{\mathrm{T}}\boldsymbol{g} = 1$ (where $\boldsymbol{g}^{\mathrm{T}}$ denotes the transpose of $\boldsymbol{g}$), and have determinant 1. The concrete form of such a matrix in terms of explicit parameters, such as the Euler angles in the case $n = 3$, is somewhat cumbersome and will not be needed here.

The rotation group bears a close relation with the set of skew-symmetric matrices, that is, the set of ($n \times n$)-matrices $\boldsymbol{X}$ satisfying $\boldsymbol{X}^{\mathrm{T}} = -\boldsymbol{X}$: the exponential of any skew-symmetric matrix $\boldsymbol{X}$, defined by the standard series

$$\mathrm{e}^X = \sum_{k=0}^{\infty} \frac{X^k}{k!} = 1 + X + \frac{X^2}{2!} + \cdots \quad (1)$$

is a rotation matrix. This set has a simpler structure than the rotation group itself: it is a vector space, since sums and multiples of skew-symmetric matrices are again skew-symmetric. Moreover, although the product $\boldsymbol{XY}$ of two skew-symmetric matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ is no longer skew-symmetric, their commutator $[\boldsymbol{X}, \boldsymbol{Y}]$, defined by

$$[X, Y] = XY - YX \quad (2)$$

is again skew-symmetric. In mathematical terminology, this is expressed by saying that the set of skew-symmetric matrices forms a Lie algebra: this Lie algebra is denoted by $\mathfrak{so}(n)$.

Summarising, we have an n-dimensional vector space (the carrier space) and two sets of ($n \times n$)-matrices acting on it, which are related by the exponential map. The former, the Lie group, closes under multiplication whereas the latter, the Lie algebra, closes under addition, multiplication by scalars and commutators, and for this reason is easier to deal with. For example, the Lie algebra $\mathfrak{so}(3)$ can, as a three-dimensional vector space, be thought of as spanned by the three matrices

$$\tau_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \tau_y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}$$

$$\tau_z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (3)$$

which satisfy the commutation relations

$$[\tau_x, \tau_y] = \tau_z, \quad [\tau_y, \tau_z] = \tau_x, \quad [\tau_z, \tau_x] = \tau_y \quad (4)$$

In many applications in science, we have to deal with functions instead of n-component vectors. Working with function spaces, which are also vector spaces but are usually infinite-dimensional, is more difficult from the purely algebraic point of view but becomes possible by making use of analytic techniques, such as derivatives and integrals. In particular, the elements of the pertinent Lie algebra are in this context represented by differential operators, instead of matrices. For example, the matrices in (3) are (up to a factor of i – the imaginary unit) represented by the angular momentum operators

$$K_x = -\mathrm{i}\left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\right), \quad K_y = -\mathrm{i}\left(z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z}\right)$$

$$K_z = -\mathrm{i}\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right) \quad (5)$$

which act in a space of differentiable functions of the variables x, y, z and satisfy the commutation relations

$$[K_x, K_y] = \mathrm{i}K_z, \quad [K_y, K_z] = \mathrm{i}K_x, \quad [K_z, K_x] = \mathrm{i}K_y \quad (6)$$

In this article, we shall deal with another three-dimensional Lie algebra, denoted by $\mathfrak{so}(2, 1)$, which is similar to the rotation algebra $\mathfrak{so}(3)$. It is represented by the differential operators

$$L_x = \eta\frac{\partial^2}{\partial \eta^2} + (2l + 2 - \eta)\frac{\partial}{\partial \eta} + \frac{\eta - 2(1 + l)}{2}$$

$$L_y = -\mathrm{i}\eta\frac{\partial}{\partial \eta} + \mathrm{i}\frac{\eta - 2(1 + l)}{2} \quad (7)$$

$$L_z = \eta\frac{\partial^2}{\partial \eta^2} + (2l + 2 - \eta)\frac{\partial}{\partial \eta} - (1 + l)$$

where η is a complex variable and l is a real constant. These operators satisfy the commutation relations

$$[L_x, L_y] = -iL_z, \quad [L_y, L_z] = iL_x, \quad [L_z, L_x] = iL_y \quad (8)$$

Thus the difference between the two algebras is just a question of sign, but this sign has profound consequences. The relation between the two algebras is the same as that between the (Lie algebras of the) rotation group in four dimensions and the Lorentz group – the symmetry group of special relativity.

One can rewrite the differential operators in (7) in the Cartan basis [90]. The L_x and L_y operators are combined as

$$L_+ = L_x + iL_y \quad \text{and} \quad L_- = L_x - iL_y \quad (9)$$

and are denominated ladder operators with $L_+(L_-)$ as being the uppering (lowering) operator. Those three operators act on a basis composed by eigenfunctions of the L_z operator (the so-called Cartan subalgebra), which respective eigenvalues denominated weights. The action of the uppering (lowering) operator on an element of the basis connects it with another one whose weight is increased (decreased) by one. The commutation relations become

$$[L_z, L_\pm] = \pm L_\pm \quad [L_+, L_-] = -2L_z \quad (10)$$

and are verified by considering the commutation relations from (8).

We also make use of operators that commute with all operators in the algebra: these are called Casimir operators. They play an important role in the theory because in so-called irreducible representations, operators commuting with any operator in the algebra are proportional to identity, in accordance with Schur's lemma [90]. In other words, it defines an invariant of the algebra and its eigenvalue is, consequently, a constant. Here, there is (up to constant multiples) just one of them: in the case of the $\mathfrak{so}(3)$-algebra it is

$$C = L_x^2 + L_y^2 + L_z^2 \quad (11)$$

which is the total angular momentum of a rotating particle in the three-dimensional space.

In the case of the $\mathfrak{so}(2, 1)$-algebra the Casimir operator is

$$C = L_z^2 - L_x^2 - L_y^2 \quad (12)$$

3 Symmetry in stochastic gene expression

One of the motivations for the study of isolated genes is the need for understanding the functioning of the basic building blocks that compose gene networks [17, 18]. Such a bottom-up approach is useful since single genes, even small ones, already perform specific functions in the circuit. However, all such investigations, whether experimental, computational or theoretical, must deal with the issue of fluctuations, that is, the inherent randomness of a low reactant number environment such as the cell [19].

Experimentally, gene networks – both natural and synthetic ones – are tested by applying fluorescence methods to detect the expression level of each gene, that is, the production rate for the protein that it encodes [20–40]. Engineered networks have been used for studying self-regulation [20–22], genetic toggle switches [23], repressilator systems [24], noise propagation [25–27] and so on. They are also useful for classifying sources of noise in the cell as intrinsic or extrinsic [28, 29]. The former type of noise comes from inherent fluctuations of a specific chemical reaction whereas the latter arises from the transmission of fluctuations by components of the cell such as transcription factors or polymerases. Among the natural networks, two systems are of particular interest, the lac operon [38] and the λ-phage [39, 40]. They were useful for the establishment of the basic mechanisms of regulation of gene transcription or decisory gene circuits.

The advances in the quantitative experimental investigation of gene expression have motivated a boom in the modelling of gene networks [41–46, 48–61]. Computational simulations based on Gillespie's algorithm [62] have been applied successfully to the λ-phage decision circuit [41, 42]. The advantage of this approach is that it provides a highly detailed description of the chemical reactions involved in a gene network. However, it requires many hours of processing and involves lots of phenomenological parameters [63]. This is where models with analytic solutions become important, since they furnish rapidly available information about a system. Such solutions can be based on macroscopic or mesoscopic equations [64]. The former are non-linear equations for concentrations whereas the latter provide distribution probabilities for the reactants.

Several deterministic models have been developed to describe experiments with the lac operon [43] and the λ-phage [44], but they do not take into account fluctuations. These can be introduced by the Langevin technique [45] or recent more sophisticated versions thereof [46, 47]. Some studies in terms of master equations are basically committed to understanding the role of fluctuations in protein synthesis by fully calculating the microscopic distribution probabilities [48–51]. Others deal with the noise calculated from the master equations [54, 59–61], which are useful for treating single-peaked distributions but insufficient when multi-peaked distributions are involved. A more accurate description requires explicit analytical solutions for the

probability distributions from the master equations [48, 53, 57, 58].

In this section, we consider the Lie symmetry that appears in the binary model to stochastic gene expression [48]. We solve a pair of coupled master equations analytically by using the generating function technique [53, 58]. The steady-state solutions are given in terms of the celebrated confluent hypergeometric functions, on which very detailed information is available in the literature [93, 94] and which is incorporated in mathematical programs such as Maple, Mathematica and Matlab. It turns out that this system of equations has a Lorentz-like Lie symmetry [the symmetry group is SO(2, 1)], which is the reason underlying its complete integrability [48]. The biological meaning of the symmetry can be summarised as follows: the eigenvalue under the invariant of the algebra (Casimir operator) provides the decaying rate of the system to equilibrium; the eigenvalue under the action of the L_z operator gives the activity level of the gene; the ladder operators act by connecting different stochastic processes.

The symmetries that appear in this elementary gene switch model were identified after its exact solution had already been found. However, it should be kept in mind that the search for symmetries in systems of differential equations is a central method for the calculation of exact solutions. Indeed, the motivation of Sophus Lie that led him to introduce the class of continuous groups known today as Lie groups was precisely the usefulness of this concept in the search for solutions of differential equations.

3.1 Binary stochastic model to a gene

In the simple model adopted here, gene expression is considered as a continuous time Markov process (master equation), where the stochastic variables are (n, m), where n is the number of proteins produced by the gene and m is the state of the gene. For the sake of simplicity, in [52] the gene was supposed to admit only two possible states: an active one and a repressed one. The former prevails when the operator site of the gene is free, whereas the latter is realised when it is occupied by the regulatory protein, leading to the total or at least partial obstruction of protein synthesis. The processes of releasing of the regulatory protein from an occupied operator site and of binding the regulatory protein to an unoccupied operator site lead to transitions of the gene between its active state and its repressed state: this gene switch is governed by a 'releasing rate' f and a 'binding rate' $h_1 n + h_2$, with $h_1 = 0$ for an externally regulated gene, where the regulatory protein is provided by external sources, and $h_2 = 0$ for a self-regulating gene, where the regulatory protein is the one produced by the gene itself. Then denoting by α_n and by β_n the probability for the presence of n protein molecules when the gene is in the active state and in the repressed state, respectively, the master equation for the probability

dynamics reads

$$\frac{d\alpha_n}{dt} = k(\alpha_{n-1} - \alpha_n) + \rho[(n+1)\alpha_{n+1} - n\alpha_n]$$
$$\qquad - (h_1 n + h_2)\alpha_n + f\beta_n$$
$$\frac{d\beta_n}{dt} = \chi k[\beta_{n-1} - \beta_n] + \rho[(n+1)\beta_{n+1} - n\beta_n]$$
$$\qquad + (h_1 n + h_2)\alpha_n - f\beta_n \qquad (13)$$

where k and ρ are protein production and degradation rates, respectively. Here, we admit the possibility that there is still some amount of protein synthesis going on even in the repressed state, that is, repression may not be total: this is represented by the 'damping factor', which should be such that $0 \le \chi < 1$. However, in what follows, we shall mostly consider the case of total suppression, $\chi = 0$.

3.2 Generating function technique

For the sake of definiteness, we shall in this subsection deal only with the case of an externally regulated gene $(h_1 = 0)$, since a self-regulating gene $(h_2 = 0)$ is completely analogous, with the same solution method and the same underlying symmetry; it is discussed in [48]. Exact solutions of (13) are calculated by making use of the generating functions method, in which the individual probabilities are considered as the Taylor coefficients of a function of a complex variable u, say, that is supposed to be analytic near the origin. Within this procedure, the difference-differential equations (13) are transformed into a pair of coupled partial differential equations

$$\frac{\partial \alpha}{\partial t} = (z - 1)\left(k\alpha - \rho\frac{\partial \alpha}{\partial z}\right) - h_2\alpha + f\beta$$
$$\frac{\partial \beta}{\partial t} = (z - 1)\left(\chi k\beta - \rho\frac{\partial \beta}{\partial z}\right) + h_2\alpha - f\beta \qquad (14)$$

which are the master equations in terms of generating functions

$$\alpha(z, t) = \sum_{n=0}^{\infty} \alpha_n(t)z^n, \quad \beta(z, t) = \sum_{n=0}^{\infty} \beta_n(t)z^n \qquad (15)$$

Equations (13) are recovered by direct substitution of (15) into (14).

In terms of the functions (15) the conservation of probability means that

$$\alpha(1, t) + \beta(1, t) = 1 \qquad (16)$$

The nth probabilities (α_n, β_n) and pth order moments $(\langle n^p{}_\alpha\rangle, \langle n^p{}_\beta\rangle)$ of the stochastic process defined by (13) are recovered by differentiating and evaluating the respective generating functions at the points $z = 0$ and $z = 1$,

respectively

$$\alpha_n = \frac{1}{n!}\frac{\partial^n \alpha}{\partial z^n}\bigg|_{z=0}, \qquad \beta_n = \frac{1}{n!}\frac{\partial^n \beta}{\partial z^n}\bigg|_{z=0} \qquad (17)$$

$$\langle n_\alpha^p \rangle = \left(z\frac{\partial}{\partial z}\right)^p \alpha\bigg|_{z=1}, \qquad \langle n_\beta^p \rangle = \left(z\frac{\partial}{\partial z}\right)^p \beta\bigg|_{z=1} \qquad (18)$$

These definitions are verified by direct substitution of (15).

The major advantage of the generating function method is the possibility to use well-known mathematical techniques from the theory of differential equations to solve them.

3.3 Stationary regime

Let us consider the stationary or asymptotic solutions where all time derivatives vanish, so that (14) becomes a system of first-order ordinary differential equations that we rewrite in the form

$$0 = (z-1)\left(N\alpha - \frac{d\alpha}{dz}\right) - \epsilon p_\beta \alpha + \epsilon p_\alpha \beta$$
$$0 = (z-1)\left(N\chi\beta - \frac{d\beta}{dz}\right) + \epsilon p_\beta \alpha - \epsilon p_\alpha \beta \qquad (19)$$

where new constants have been introduced as follows

$$N = \frac{k}{\rho}, \quad p_\alpha = \frac{f}{f+h_2}, \quad p_\beta = \frac{h_2}{f+h_2}, \quad \epsilon = \frac{f+h_2}{\rho} \qquad (20)$$

These parameters have been chosen so as to simplify the identification of their biological significance. The first of them, called N, depends on the affinity of the mRNA-polymerase with the promoter site: it is just the protein production rate k in the active state relative to the protein degradation rate ρ. The parameters p_α and p_β represent the asymptotic probabilities of the gene to be 'on' and 'off', respectively, as can be seen by setting $z = 1$ and letting $t \to \infty$ in any one of the equations (19) to obtain

$$-h_2\alpha(1, \infty) + f\beta(1, \infty) = 0 \qquad (21)$$

which together with the relation

$$\alpha(1, \infty) + \beta(1, \infty) = 1 \qquad (22)$$

which follows from (16), implies

$$p_\alpha = \alpha(1, \infty) = \sum_{n=0}^{\infty}\alpha_n(\infty), \quad p_\beta = \beta(1, \infty) = \sum_{n=0}^{\infty}\beta_n(\infty) \quad (23)$$

The parameter p_α can also be referred to as the steady-state activity level of the gene, whereas the parameter $p_\beta = 1 - p_\alpha$ measures the steady-state occupancy of the operator site by the regulatory protein, depending on the affinity between the two. The dimensionless parameter ϵ

measures the relation between the two decay rates to equilibrium, $f + h_2$ for the switch and ρ for protein decay.

To solve the system (19) we use the first equation to express β in terms of α

$$\beta = -\frac{z-1}{\epsilon p_\alpha}\left(N\alpha - \frac{d\alpha}{dz}\right) + \frac{p_\beta}{p_\alpha}\alpha \qquad (24)$$

and inserting this result into the second equation to obtain

$$\frac{z-1}{N}\frac{d^2\alpha}{dz^2} + \left(\frac{1+\epsilon}{N} - (z-1)\right)\frac{d\alpha}{dz} - (1+\epsilon p_\alpha)\alpha = 0 \quad (25)$$

where it has been assumed that $\chi = 0$ (a similar but somewhat more complicated equation holds when $\chi > 0$; it is given in Appendix). This is the celebrated hypergeometric equation, which contains two singularities: a regular one at $z = 1$ and an irregular one at infinity. Its solutions are given in terms of Kummer functions M and U. The Kummer U function is irregular at $z = 1$ and hence must be discarded in order to avoid singularities, so the solutions of interest here can be written exclusively in terms of the Kummer M function, which can be defined by the series [93]

$$M(a, b, \theta) = \sum_{n=0}^{\infty}\frac{(a)_n}{(b)_n}\frac{\theta^n}{n!} \qquad (26)$$

where $(a)_n$ is the Pochhammer symbol

$$(a)_0 = 1, \quad (a)_1 = 1, \ldots,$$
$$(a)_n = a(a+1)(a+2)\ldots(a+n-1) \qquad (27)$$

In these terms, the steady-state solutions are

$$\alpha(z) = p_\alpha M(1 + \epsilon p_\alpha, 1 + \epsilon, N(z-1))$$
$$\beta(z) = (1 - p_\alpha)M(\epsilon p_\alpha, 1 + \epsilon, N(z-1)) \qquad (28)$$

from which the probabilities can be directly calculated using (17) together with the formula for the nth derivative of the Kummer M function (see e.g. (13.4.9) of [93]) and the result is

$$\alpha_n = p_\alpha\frac{N^n}{n!}\frac{(1+\epsilon p_\alpha)_n}{(1+\epsilon)_n}M(1+\epsilon p_\alpha + n, 1+\epsilon + n, -N)$$

$$\beta_n = (1 - p_\alpha)\frac{N^n}{n!}\frac{(\epsilon p_\alpha)_n}{(1+\epsilon)_n}M(\epsilon p_\alpha + n, 1+\epsilon + n, -N)$$

$$(29)$$

when $\chi = 0$ (similar but somewhat more complicated expressions can be obtained when $\chi > 0$; they are given in Appendix).

3.4 Lorentz-like Lie symmetry

The existence of analytical solutions of (25) and the similarity between the differential operator acting on α that appears in

(25) with the L_z of (7) leads us to suspect the existence of a Lie symmetry in this model. This is indeed so [48]. In fact, identifying the variable η in the L_z operator of (7) with $N(z-1)$, the latter becomes

$$L_z = \frac{z-1}{N}\frac{d^2}{dz^2} + \left(2\frac{1+l}{N} - (z-1)\right)\frac{d}{dz} - (1+l) \quad (30)$$

so (25) can be rewritten as an eigenvalue equation for L_z

$$L_z\alpha(z) = m\alpha(z) \quad (31)$$

provided the group theoretical parameters l and m are related to the parameters ϵ and p_α by

$$\epsilon = 2l + 1, \quad p_\alpha = \frac{l+m}{2l+1} \quad (32)$$

The ladder operators defined in (9)

$$L_+ = \frac{z-1}{N}\frac{d^2}{dz^2} + 2\frac{l+1}{N}\frac{d}{dz}$$
$$L_- = \frac{z-1}{N}\frac{d^2}{dz^2} + 2\frac{l+1-N(z-1)}{N}\frac{d}{dz} \quad (33)$$
$$- 2(l+1) + N(z-1)$$

In terms of the parameters m and l, the function $\alpha(z)$ is defined as $(l+m)/(2l+1)M(1+m+l, 2l+2, N(z-1))$. The action of the ladder operators on the Kummer M function defining $\alpha(z)$ or $\beta(z)$ results

$$L_- M(1+m+l, 2l+2, N(z-1))$$
$$= (m-l-1)M(m+l, 2l+2, N(z-1))$$
$$L_+ M(m+l, 2l+2, N(z-1)) \quad (34)$$
$$= (m+l)M(1+m+l, 2l+2, N(z-1))$$

The Casimir operator is given by [90]

$$C = -L_z^2 + L_z + L_+L_- \quad (35)$$

and it has eigenvalue $-l(l+1)$ on α, that is,

$$C\alpha = -l(l+1)\alpha \quad (36)$$

as follows from (31) and (34). In terms of the parameter ϵ,

$$C\alpha = \frac{1-\epsilon^2}{4}\alpha \quad (37)$$

Summarising, the operators L_z, L_+ and L_-, defined in (30) and (33), span the Lie algebra $\mathfrak{so}(2,1)$ whose irreducible representations are labelled by an arbitrary real number l; the index m labelling different states within the representation space is also a real number. The representations are all unbounded, both from above and from below. The relation between these group theoretical labels and the parameters of the model is given by (32).

3.5 Biological meaning of symmetry

We discuss the biological significance of the various operators and parameters. The action of L_+ in α and β produces new amplitudes α_1 and β_1, respectively, for a new stochastic process. By the successive application of this operator, we construct a family of related processes. It is remarkable that β_1 coincides with α and a repressive process ($p_\beta \simeq 1$) is transformed into a non-repressive process ($p_\alpha \simeq 1$).

The invariant of the algebra, the Casimir operator, has a clear biological meaning: its eigenvalue gives the decaying rate of the system to equilibrium, relative to the protein degradation rate ρ. In order to show that, we evaluate the dynamic equations (14) at $z = 1$, writing $P_\alpha(t) \equiv \alpha(1, t)$ and $P_\beta(t) \equiv \beta(1, t)$ for the total probability, at time t, for the gene to be in the active state and in the repressed state, respectively, we have

$$\frac{dP_\alpha}{dt} = -h_2 P_\alpha + f P_\beta$$
$$\frac{dP_\beta}{dt} = h_2 P_\alpha - f P_\beta \quad (38)$$

Noting that according to equation (16), $P_\alpha(t) + P_\beta(t) = 1$, and using (20), we can write down the solution:

$$P_\alpha(t) = p_\alpha + (P_\alpha(0) - p_\alpha)\exp(-\epsilon\rho t) \quad (39)$$

Thus we see that the probability $P_\alpha(t)$ decays exponentially to its equilibrium value p_α, and the rate at which it does so, in terms of the protein degradation rate ρ, is given by the parameter ϵ, or what according to (37) amounts to the same thing, by the eigenvalue of C.

This analysis can also be extended to values of z which are $\neq 1$. In order to show how, we rewrite the dynamic equations (14) in matrix form

$$\frac{1}{\rho}\frac{\partial\Psi}{\partial t} = \mathcal{H}\Psi \quad (40)$$

where

$$\Psi = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad \text{and}$$

$$\mathcal{H} = \begin{pmatrix} N(z-1) - \epsilon p_\beta - (z-1)\frac{\partial}{\partial z} & \epsilon p_\alpha \\ \epsilon p_\beta & -\epsilon p_\alpha - (z-1)\frac{\partial}{\partial z} \end{pmatrix}$$
$$(41)$$

Time-dependent solutions are obtained by expanding Ψ in a basis of eigenfunctions of $\mathcal{H}$

$$\Psi = \sum_\lambda \exp(-\lambda t)\Psi_\lambda \quad (42)$$

The eigenvalue equation is written as

$$\mathcal{H}\Psi_\lambda = -\lambda\Psi_\lambda \qquad (43)$$

or, in components

$$-\lambda\alpha_\lambda = (N(z-1) - \epsilon p_\beta)\alpha_\lambda - (z-1)\frac{d\alpha_\lambda}{dz} + \epsilon p_\alpha\beta_\lambda$$

$$-\lambda\beta_\lambda = \epsilon p_\alpha\beta_\lambda - (z-1)\frac{d\beta_\lambda}{dz} + \epsilon p_\beta\alpha_\lambda \qquad (44)$$

Solving the first of these equations for β_λ and substituting it into the second, we obtain a second-order ordinary differential equation for α_λ. Its solution is obtained by the Frobenius method [95], which starts from a series expansion of $\alpha_\lambda(z)$ around $z = 1$ in the form

$$\alpha_\lambda(z) = (z-1)^s \sum_{n=0}^{\infty} C_n (z-1)^n \qquad (45)$$

where s is a number, called 'indicium', which appears when the exponent of the first term of the series does not vanish, and the $C_n(\lambda)$ are the coefficients of the expansion. The requirement of analyticity forces the indicium to be a non-negative integer, and the aforementioned second-order differential equation gives the result

$$\lambda = \epsilon\rho + s, \quad \text{where } s = 0, 1, 2, \ldots \qquad (46)$$

Therefore we conclude that the fundamental mode ($s = 0$), which dominates the decay to equilibrium, has decay rate equals to $\epsilon\rho$, as before.

Concerning the range of possible values for the parameters l and m, the mathematics imposes no constraints, but the biological interpretation does. The range of allowed values for m comes from the restriction that a probability must lie in the interval between 0 and 1; this implies, according to (32)

$$-l \leq m \leq l+1 \qquad (47)$$

Similarly, the requirement of positivity of the parameter ϵ, which follows from its definition in (20), implies

$$l \geq -1/2 \qquad (48)$$

3.6 Probability distributions and fluctuations

In Fig. 1 we show the probability distributions for selected values of the algebraic parameters. The protein number is drawn along the the horizontal axis. The graphics were designed by using (55) of Appendix. We have chosen two different values for l: a high ($l = 5$) and a low ($l = -1/4$). For each of these two values, the distributions are plotted for three different values of m: one close to the minimum, one medium and one close to the maximum. The dashed lines represent active state probabilities and the dotted are the repressed. The continuous line shows the sum between the active and repressed states. Note that two-peaked probability distributions appear for $l = -1/4$, corresponding to a slowly switching gene. For high values of l, corresponding to a fast switching gene, the distributions are single-peaked in the sense that the two gene state probabilities are centred around (almost) the same point. They are in accordance with a fast-switching gene. The
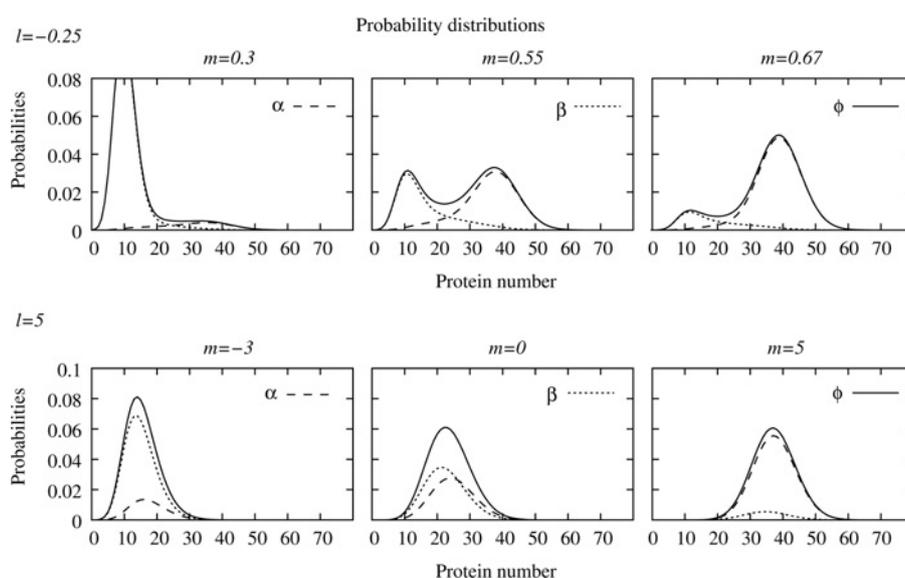

Figure 1 *Probability distributions of the free state and total number of proteins inside the cell*

In all plots $N = 40$ and $\chi = 0.2$. In the top set of plots, $l = -1/4$ (slowly switching gene), whereas in the bottom, $l = 5$ (fast switching gene)

relation between the heights of the peaks depends on the value of m. If m is close to the minimum the repressed peak is higher; if m is close to the maximum, the active peak is the higher. The histograms are displaced to the left or to the right, respectively, when m is lowered or raised. The description of a slowly switching gene requires both active and repressed state probabilities, since the fluctuations in gene state correspond to changing the centre of the distributions, whereas in the case of a fast switching gene, the total probability distribution is sufficient to characterise it, because the two partial distributions that compose it are centred around (almost) the same point.

A bi-peaked probability distribution associated with a slow switch can be observed in protein production of eukaryotes [96]. That is because this DNA is packed in a complex structure, called chromatin, which can be opened and closed to expression at quite slow rates. Otherwise, one-peaked distributions are most frequent in prokaryotic gene expression, since the chemical reactions that regulate the protein synthesis are fast.

In Fig. 2, we plot the Fano factor in terms of the algebraic parameters. It is defined as

$$F = \frac{\langle n^2 \rangle - \langle n \rangle^2}{\langle n \rangle} \quad (49)$$

and characterises the deviation of a distribution from a Poissonian one, whose F is equal to one [97]. Super-Poissonian processes have $F > 1$ and sub-Poissonian processes have $F < 1$. The explicit form of F for the externally regulated gene is

$$F = 1 + \frac{N(1-\chi)^2(l+m)(1-m+l)}{(2l+2)(2l+1)[(m+l)(1-\chi)+(2l+1)\chi]} \quad (50)$$

Note that F is greater or equal to one, so the externally regulated gene has only a super-Fano or a Fano noise regime. Sub-Fano processes only occur for $l + m < 0$, as demonstrated recently for a self-interacting gene [48]. Furthermore, low l valued states are noisier than states with

Fano factor

Figure 2 *Fano factors for two values of N with $\chi = 0.2$*

In order to plot all Fano factors together we substituted m in (50) by the expression $-l + p_\alpha(2l+1)$ that comes from the definition (32). Note that noise tends to one for high values of l

high l. In the limit when l goes to infinity, the Fano factor tends to 1, independent of the value of m, as long as N remains finite. This follows by writing the numerator and denominator of the second part of (50) as polynomials in l. The former has a degree 2 and the latter a degree 3, so in the limit when l goes to infinity, their quotient tends to 0. In fast-switching genes effects of protein creation and degradation phenomena are sufficient for stochastic modelling, whereas in slow systems an accurate modelling also depends on the description of the gene states.

In Fig. 2 we present two plots of the noise as a function of m for a set of values of l and two limit values of N. Each curve presents a maximum at the point

$$\frac{m_{\max}+l}{2l+1} = \frac{\sqrt{\chi}}{1+\sqrt{\chi}}, \quad F_{\max} = 1 + \frac{N}{2}(1-\sqrt{\chi})^2 1 + l \quad (51)$$

Since χ lies in the interval 0 and 1, $m_{\max}$ lies in the interval between $-l$ and 1/2. The value of $F_{\max}$ is linearly proportional to N and tends to one for fast-switching genes. In the limiting values of m the Fano factor tends to one. This means that the gene remains in a Poissonian production degration state. Note also that for a fixed value of the Fano factor, there are two levels of gene activity. This indicates that consideration of variation in protein number only is incomplete in determining gene performance. The probability distributions and Fano factor present an explicit dependence on the algebraic parameters. Slowly switching genes tend to be more fluctuating and present bi-peaked probability distributions: this is in accordance with previous simulation results [56]. Furthermore, it indicates that a stochastic model that accounts for the two states of a gene, even when dealing with conjugated transcription and translation, is enough to indicate a binary behaviour. Otherwise, fast switching genes tend to be Poissonian. The fluctuations in protein number are not affected by gene state variations. Thus, symmetry tells us that high stochasticity is associated with low eigenvalues of the Casimir operator (36). The analogy comes from atomic physics, where quantum mechanical effects are more prominent for energy levels close to ground-states and, low eigenvalues of the Casimir operator. High energised levels tends to present a quasi-classical behaviour, as occurs in Rydberg atoms [98].

4 Symmetry breaking in genetic code evolution

This section is divided into two main parts. In the first, we shall give a brief account of the construction of the algebraic model for the evolution of the genetic code. In the second part, we present the exact Klein symmetry of the standard code discovered in the course of this investigation, as a remnant of a larger symmetry that, according to the algebraic model, has prevailed in earlier
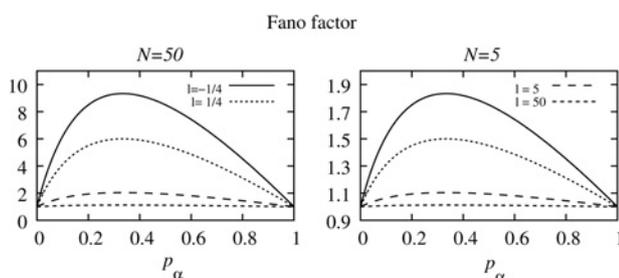
stages of its evolution. Part of this Klein symmetry is the obvious symmetry of the genetic code table (Table 1) under the exchange of U and C in the third base, but the other part is far from obvious. The same Klein symmetry is also found in most of the non-standard codes, in particular the mitochondrial ones, a fact that speaks in favour of its importance in the course of code evolution.

4.1 Construction of an algebraic model to genetic code evolution

Since the expression 'genetic code' is often used erroneously (a quite common abuse consists in confusing it with what should correctly be called the genome, usually of a specific organism or species), it seems in good order to recall that, when interpreted correctly, it refers to the rules of protein synthesis in the ribosome; more precisely, the rules for translating the information contained in a gene, copied to a strand of mRNA, into the sequence of amino acids that constitutes the protein encoded by that gene. It consists of a table specifying the assignment of a specific amino acid to each possible triplet of mRNA nucleotides: such a triplet is called a codon. Note that the four nucleotides U, C, A and G can be combined into 64 different triplets, but all living organisms found in nature use only 20 different amino

Table 1 Standard genetic code − mRNA codons against amino acids

First base	Second base				Third base
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Try	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

acids to build their proteins. Addition of the termination or stop signal implies that there are altogether 21 distinct meanings to be assigned to the 64 codons, so that more than one codon will have the same meaning. This mathematical feature of the table is referred to as the degeneracy of the genetic code, even though biologists or linguists will probably prefer to speak of 'synonymous codons' rather than 'degenerate codons'.

The notion of degeneracy is profoundly related to that of symmetry. Degeneracy means invariance; in the present case, it means that the codon to amino acid assignment is invariant under the replacement of codons by synonymous ones. And invariance means symmetry, in the sense that one can build transformation groups that keep invariant certain properties. This kind of connection between symmetry and invariance can be seen in the spectrum of the hydrogen atom: this is a system with an obvious rotational symmetry, implying that states with the same azimuthal angular momentum quantum number m will have the same energy. But symmetries may be much less obvious than in this case; they may be hidden! And there are many examples where the spectrum of a molecule or atom is a testimony of some hidden symmetry [14]. Thus if we look at the genetic code from this point of view, as if it were some kind of spectrum, we face a straightforward question: is the degeneracy pattern of the code the expression of some hidden symmetry? This promptly suggested performing what we may call 'the search for symmetries in the genetic code'.

Summarised briefly, the program runs as follows. First of all, the 64 available codons are considered as forming a basis of a 64-dimensional vector space, called the codon space, on which some (as yet unknown) 'primordial' symmetry group G acts by linear transformations: its elements can be thought of as certain (64×64)-matrices acting on the vectors in the codon space. Initially, the symmetry is complete in the sense that, roughly speaking, it is able to generate any vector in codon space from any other one; the technically correct condition is that the 'primordial' symmetry group G acts irreducibly on codon space. In the language used in physics, this is expressed by saying that all 64 codons form a single multiplet under the 'primordial' symmetry group. Starting out from this situation as an initial condition, the symmetry breaking process consists in selecting a descending chain of subgroups of G, $G \supset G_1 \supset \cdots \supset G_{n-1} \supset G_n$, and successively restricting the symmetry transformations to belong to these subgroups. This picture fits in neatly with the widely accepted idea that the genetic code has evolved from an early stage with few amino acids by gradually incorporating new ones. The process ends when one reaches the 'residual' symmetry group G_n which, ideally, should be such that when considered as acting on the vectors in the codon space, the latter decomposes into the direct sum of 21 disjoint irreducible subspaces: three of dimension 6 to accomodate the three codon sextets for

Arg, Leu and Ser; five of dimension 4 to accomodate the five codon quartets for Ala, Gly, Pro, Thr and Val; two of dimension 3 to accomodate the codon triplet for Ile and the Stop codons; nine of dimension 2 to accomodate the nine codon doublets for Asn, Asp, Cys, Gln, Glu, His, Lys, Phe and Tyr; and finally two of dimension 1 to accomodate the two codon singlets for Met and Try. A more general rule, which formalises the idea of 'freezing' first proposed by Crick [65], allows for the possibility to obtain this final distribution of multiplets from a distribution of multiplets under the penultimate subgroup in the chain, G_{n-1}, by admitting that under the passage to the last subgroup in the chain, G_n, the symmetry breaking may only be partial, that is, some of the G_{n-1}-multiplets break up into the appropriate G_n-multiplets as required by the rules of group theory whereas others that should normally break up as well fail to do so: their breakup is frozen.

The first step to obtain this program up and running is to discover what are the possible ancestor groups G, the basic restriction being that any such group should admit a 64-dimensional irreducible representation. This condition is much more restrictive than may seem and, typically, leads to a finite (and small) list of solutions. In particular, in the world of compact simple Lie groups, the search for such an ancestor group can be performed by combining the Cartan classification theorem, which allows for four series of symmetry groups called 'classical' and another five 'exceptional' groups, with methods from representation theory. The second step of the search consists in analysing, step by step and for each of these possible ancestor groups, all possible chains of subgroups and find out whether symmetry breaking along any one of them is able to produce the distribution of multiplets observed in the genetic code; this procedure relies heavily on the tables of branching rules for irreducible representations of simple Lie algebras of [99].

The first result of such a search has been reported in [72], and the most promising scheme that emerges is based on the symplectic group Sp(6) as the 'primordial' symmetry group, with the following chain of subgroups

$$\text{Sp}(6) \supset \text{Sp}(4) \times \text{SU}(2) \supset \text{SU}(2) \times \text{SU}(2) \times \text{SU}(2)$$
$$\supset \text{SU}(2) \times O(2) \times \text{SU}(2) \supset \text{SU}(2) \times \text{SO}(2) \times \text{SU}(2) \tag{52}$$

(We note in passing that symplectic groups are not totally unfamiliar: their non-compact versions arise naturally in classical mechanics.) The corresponding sequence of branchings, constructed according to the tables of [99], is shown in Fig. 3. The subspaces that appear are represented by boxes with numerical entries indicating the respective dimensions. The symmetry breaking occurs at the branching points where one line may split into two (or more) others, indicating the breakup of the multiplet

written to its left into the two (or more) others written to its right, but the line may just as well continue onward without bifurcating, indicating that the multiplet written to its left does not break up at this point. Thus, for example, at the first step, the 64-dimensional Sp(6)-multiplet breaks into six different (Sp(4) × SU(2))-multiplets of dimension 20, 16, 12, 10, 4 and 2, respectively, corresponding to six primordial amino acids; at the second step, the 20-dimensional (Sp(4) × SU(2))-multiplet breaks into three different (SU(2) × SU(2) × SU(2))-multiplets of dimension 8, 6 and 6, respectively, whereas the two-dimensional (Sp(4) × SU(2))-multiplet does not break up, and so on. The distribution of multiplets with their respective dimensions reached after the last step reproduces exactly the degeneracies found in the standard genetic code, but can be realised only if one allows for the phenomenon of freezing: otherwise, some additional symmetry breakings would occur in the last step, as indicated by the dashed lines. Finally, the amino acid assignments shown in the last step are obtained from principles first established in [73] and further elaborated in [80]. A mathematically complete treatment can be found in [74, 78].

Evidently, this mathematical procedure reflects an evolution of the genetic code in terms of a step-by-step addition of amino acids and corresponding reduction of its degeneracy.

Further insight into this symmetry breaking scheme can be gained by inspecting what is called the weight diagram of the pertinent representation [90], which in this case is the codon representation of the group Sp(6): it provides an intuitive picture of some of its central features that otherwise would be hard to visualise for such a high-dimensional representation. In the case of the group Sp(6), which has rank 3 and dimension 21, this diagram is a three-dimensional array of points, each one associated with a basis vector in the representation. In the present case, we have 64 points that, in Figs. 4 and 5, are depicted as little balls, forming an internal octahedron and an external truncated octahedron. In the external truncated octahedron, the vertices are non-degenerate but the centres of the hexagonal faces are two-fold degenerate, whereas the vertices of the internal octahedron are four-fold degenerate; this is indicated in the figures by drawing several balls at (what is meant to be) the same position but in the picture are slightly displaced in order not to overlap completely. Their coordinates are integers that correspond to the simultaneous eigenvalues of the three commuting generators of the maximal torus of Sp(6). The remaining 18 generators of Sp(6) are divided into nine raising and nine lowering operators that interconnect the points; they can be visualised through the nine directions in which there are segments connecting the balls. The whole figure is invariant under a set of transformations consisting of (i) rotations by 90° around an axis passing through antipodal vertices of the internal octahedron; (ii) rotations by 120°
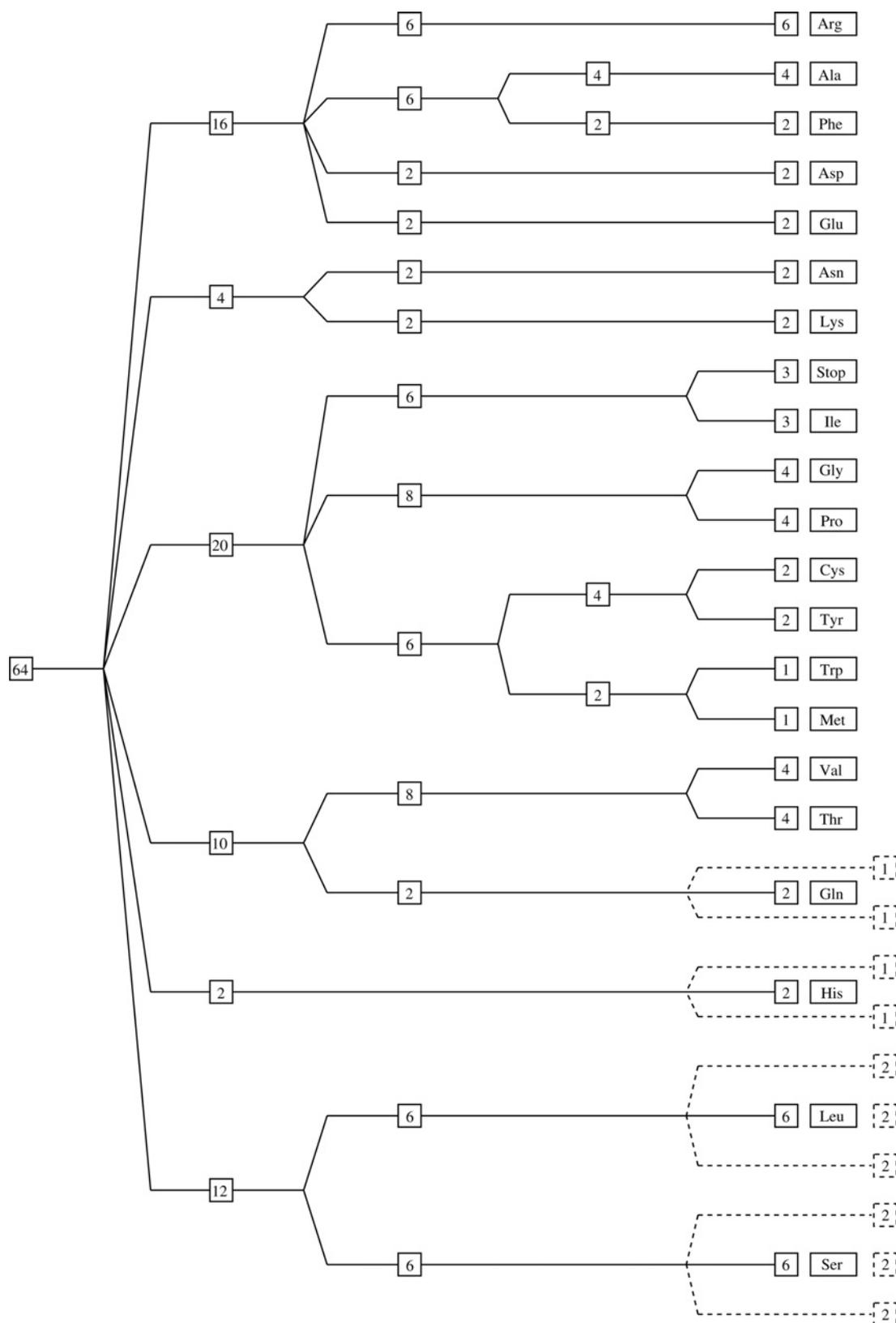
Figure 3 *Tree of evolution of the genetic code with amino acid assignment for the Sp(6)-model*

around an axis passing through antipodal centres of the hexagons of the external truncated octahedron; (iii) rotations by 180° around an axis passing through the centre points of opposite edges of the internal octahedron;

(iv) inversion; and (v) composition of inversion and any of the above operations. These transformations form a finite group of 48 elements known as the 'Weyl group' of the symplectic group Sp(6).

Figure 4 *Weight diagram of the codon representation of Sp(6) before the first symmetry breaking*

Fig. 4 shows the weight diagram before the first symmetry breaking, where all codon vectors are indistinguishable and represented by one and the same colour, whereas Fig. 5 shows the weight diagram after the first symmetry breaking, where codon vectors belonging to different $(Sp(4) \times SU(2))$-multiplets are distinguished by different colours. The operators affiliated with the subgroup $Sp(4) \times SU(2)$ only connect equally painted spheres. At this stage of evolution, a primordial genetic code with six initial amino acids is established.
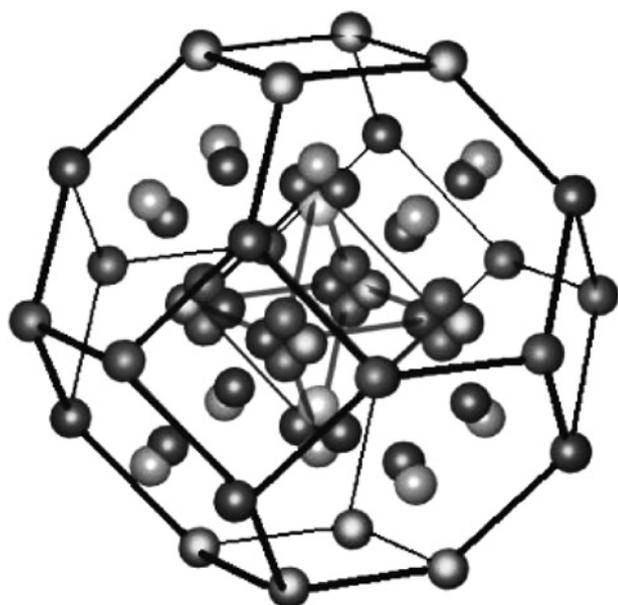

Figure 5 *Weight diagram of the codon representation of Sp(6) before the first symmetry breaking*

Different colours represent the reciprocally orthogonal subspaces

Taking horizontal cross-sections of Figs. 4 and 5 along the five planes at $z = 0$, $z = 1$ and $z = 2$, we arrive at the diagrams of Fig. 6. Each of them appears twice in order to include the tags showing the assignment of amino acids (top) and codons (bottom) to each of the points, according to the rules established in [73]: these have been suppressed in Figs. 4 and 5 to avoid visual congestion.

4.2 Klein symmetry preservation in evolution of genetic code

The weight diagrams of Figs. 4 and 5 possess an obvious octahedral symmetry that is not preserved in the final stages of the symmetry breaking process, but it is remarkable that there is a subgroup of the octahedral group, known in mathematics as the Klein group, which remains untouched. This symmetry group is composed of four operators that act on three-dimensional vectors with coordinates x, y, z as follows: (i) the identity; (ii) the reflection in the xz-plane, taking (x, y, z) to $(x, -y, z)$; (iii) the reflection in the yz-plane, taking (x, y, z) to $(-x, y, z)$; and (iv) the $180°$ rotation around the z-axis, taking (x, y, z) to $(-x, -y, z)$. Indeed, all the planar diagrams of Fig. 6 are manifestly invariant under these transformations [84]. Even more remarkable is the fact that this non-trivial Klein symmetry of the standard genetic code is shared by almost all of the presently known non-standard codes, in particular the mitochondrial codes [84].

In Fig. 7 we show the phylogenetic tree for the evolution of the non-standard mitochondrial codes. The numbers inside the brackets represent different codon reassignments (as compared to the standard code) that have been found. The ones that have occurred in the main line of evolution are labelled by (1), (2), (3) and (7). The remaining numbers represent changes that have occurred in the side branches of the tree. For example, in the transitions (1) and (3), the standard codons UGA and AUA for Stop and Ile are reassigned so as to code for Trp and Met, respectively, whereas in the transitions (2) and (7), the standard codons AGA and AGG for Arg are reassigned so as to code for Ser in (2) and Stop in (7).

The four transitions occurred in the main line of evolution preserve the Klein symmetry, as can be seen by inspecting Fig. 6. Moreover, the changes (1) and (3) even increase the symmetry of the internal octahedron by including the reflection in the horizontal plane (xy-plane). Even though the reflection in the horizontal plane is not a symmetry of the external truncated octahedron, this still means that the evolution of the mitochondrial codes has a tendency towards increasing their symmetry.

The reassignments in the side branches are of three kinds. The first group consists of those that preserve the Klein symmetry completely, namely (6), (9) and (10), which occur in yeast (*Saccharomyces cerevisiae*, *Torulopsis glabrata*), echinoderms and tunicates, respectively. Transition (6)
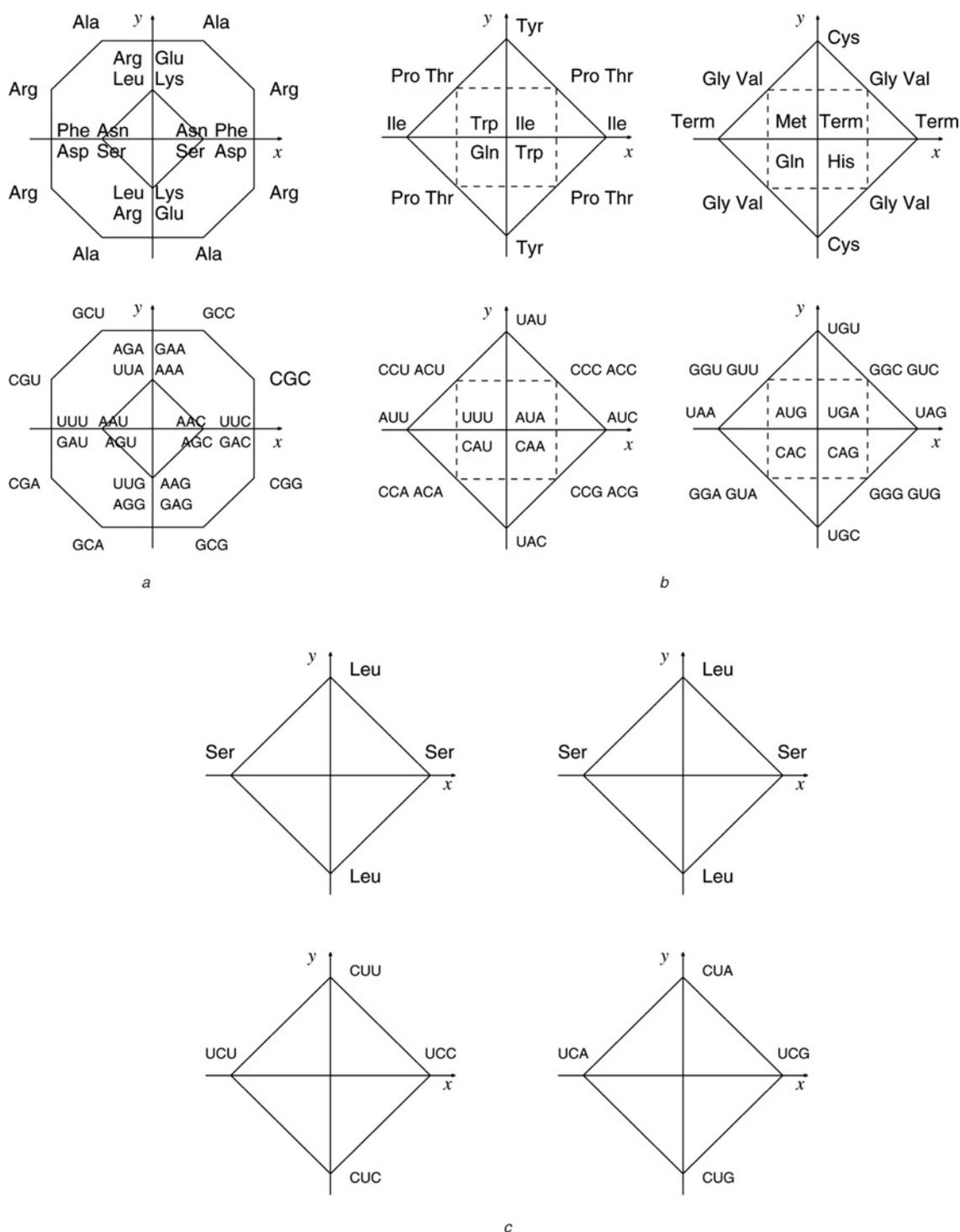
Figure 6 *Codon assignments in terms of the weights of the 64-dimensional representation of the Sp(6) in the bidimensional diagram*

a Lies in the plane $z = 0$
b Horizontal sections in planes $z = 1$ and $z = -1$
c Elements of the planes $z = 2$ and $z = -2$

represents the reassignment of the standard Leu codons CUN to Thr, (9) reverts (3), reassigning the codon AUA back to its standard meaning, from Met to Ile, and (10) reassigns the pair of standard Arg codons AGA and AGG to Gly.

The second group is formed by those that preserve the Klein symmetry partially, namely they preserve the symmetry under reflection in one of the two planes, xz or xy. These are (4) and (5), indicating the reassignment of the standard Lys and Stop codons to Asn and Tyr, as
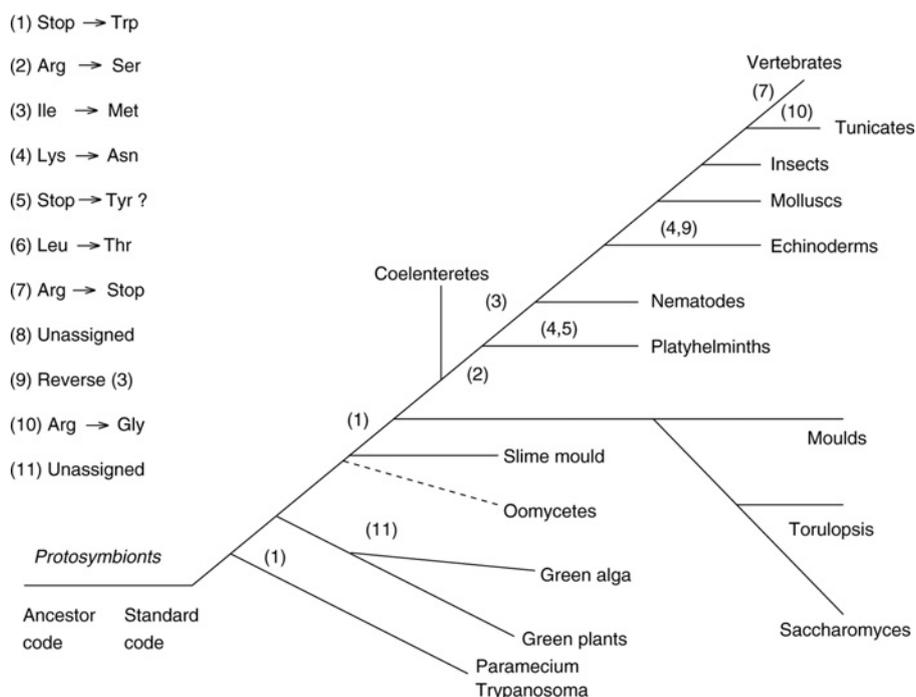
Figure 7 *Phylogenetic tree for the evolution of mitochondrial codes [66]*

observed in the mitochondrial codes of platyhelmints and echinoderms, respectively.

The remaining transitions are (8) and (11), in which unassigned and nonsense codons arise. The first is found in yeast mitochondrial codes (*Torulopsis glabrata*) and affects the standard Arg codons CGN, whereas the second is observed in green algae, affecting the codons CGG, UGA and UAG. According to the codon capture theory, these codons are in a transient state for future reassignment. Therefore even though in (8) the Klein symmetry is preserved and in (11) it is entirely broken, these exceptional cases are probably not important.

5 Conclusions

Lie group theory provides a well-developed mathematical machinery for modelling symmetry in biological systems. It provides not only a quantitative framework but also leads to biological insights about the processes that are modelled, as shown by the examples presented in this review. In the stochastic model for a two-state gene, symmetry has practical implications: the eigenvalue of the diagonal operator characterises the dynamics of the gene switch and the affinity between the regulatory protein and the gene operator site, whereas the non-diagonal operators connect the probability distributions of the two states. In addition, noise analysis leads to the conclusion that fast switching genes give rise to Poissonian distributions whereas slowly switching genes have broader or bi-peaked distributions. In the algebraic model for the evolution of the genetic code, possible pathways for this evolution arise naturally, but are

strongly restricted. The picture of evolution by a stepwise incorporation of new amino acids fits perfectly with that of dynamical symmetry breaking. The Klein symmetry that has remained preserved can serve as an underlying principle that has conducted the evolution of the standard code as well as that of non-standard codes.

In the modelling of gene networks, group theoretical tools can be useful for the search for a composition rule between two or more genes. Another feature is the possibility to model single genes that present more than two levels of regulation. The construction of a dynamical system for the evolution of the genetic code is also a possible future application of group theoretical methods in biology.

6 Acknowledgments

7 References

[1] JAEGER J., BLAGOV M., KOSMAN D., *ET AL.*: 'Dynamical analysis of regulatory interactions in the gap gene system of *Drosophila melanogaster*', *Genetics*, 2004, **167**, (4), pp. 1721–1737

[2] HARGITTAI I. (ED.): 'Symmetry: unifying human understanding', *Comp. Math. Appl.*, 1986, **12B**, (1–4), pp. 1–1045

[3] HARGITTAI I. (ED.): 'Symmetry 2: unifying human understanding', *Comp. Math. Appl.*, 1988, **16**, (5–8), pp. 351–669

[4] GOLUBTSKY M., STEWART I.: 'The symmetry perspective', Progress in Mathematics' (Birkhäuser Verlag, Basel, 2002)

[5] HARGITTAI M., HARGITTAI I.: 'Visual symmetry' (World Scientific, Singapore, 2009)

[6] FOOTE R.: 'Mathematics and complex systems', *Science*, 2007, **318**, pp. 410–412

[7] COEN E.S., NUGENT J.M., LUO D., ET AL.: 'Evolution of floral symmetry', *Philos. Trans. R. Soc. Lond. B*, 1995, **350**, pp. 35–38

[8] ENDRESS P.K.: 'Evolution of floral symmetry', *Curr. Opin. Plant Biol.*, 2001, **4**, pp. 86–91

[9] NEAL P.R., DAFNI A., GIURFA M.: 'Floral symmetry and its role in plant-pollinator systems: terminology, distribution and hypotheses', *Annu. Rev. Ecol. Syst.*, 1998, **29**, pp. 345–373

[10] FRAENKEL-CONRAT H., SINGER B., WILLIAMS R.C.: 'Infectivity of viral nucleic acid', *Biochem. Biophys. Acta*, 1957, **25**, pp. 87–96

[11] CRICK F.H.C., WATSON J.D.: 'Structure of small viruses', *Nature*, 1956, **177**, pp. 473–475

[12] PAULI W.: 'On the hydrogen spectrum from the standpoint of the new quantum mechanics', *Z. Phys.*, 1926, **36**, pp. 336–363

[13] SLATER J.C.: 'The theory of complex spectra', *Phys. Rev.*, 1929, **34**, pp. 1293–1322

[14] IACHELLO F., LEVINE R.D.: 'Algebraic theory of molecules' (Oxford University Press, Oxford, 1995)

[15] HORNOS J.E.M., IACHELLO F.: 'The overtone spectrum of acetylene in the vibron model', *J. Chem. Phys.*, 1989, **90**, pp. 5284–5291

[16] GELL-MANN M., NE'EMAN Y.: 'The eightfold way' (Benjamin, New York, 1964)

[17] HARTWELL L.H., HOPFIELD J.J., LEIBLER S., MURRAY A.W.: 'From molecular to modular cell biology', *Nature*, 1999, **402**, (6761 Suppl), pp. C47–52

[18] SHEN-ORR S.S., MILO R., MANGAN S., ALON U.: 'Network motifs in the transcriptional regulation network of *Escherichia coli*', *Nat. Genet.*, 2002, **31**, (1), pp. 64–68

[19] DELBRÜCK M.: 'Statistical fluctuations in autocatalytic reactions', *J. Chem. Phys.*, 1940, **8**, pp. 120–124

[20] BECSKEI A., SERAPHIN B., SERRANO L.: 'Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion', *EMBO J.*, 2001, **20**, (10), pp. 2528–2535

[21] BECSKEI A., SERRANO L.: 'Engineering stability in gene networks by autoregulation', *Nature*, 2000, **405**, (6786), pp. 590–593

[22] SIMPSON M.L., COX C.D., SAYLER G.S.: 'Frequency domain analysis of noise in autoregulated gene circuits', *Proc. Natl. Acad. Sci. USA*, 2003, **100**, (8), pp. 4551–4556

[23] GARDNER T.S., CANTOR C.R., COLLINS J.J.: 'Construction of a genetic toggle switch in *Escherichia coli*', *Nature*, 2000, **403**, (6767), pp. 339–342

[24] ELOWITZ M.B., LEIBLER S.: 'A synthetic oscillatory network of transcriptional regulators', *Nature*, 2000, **403**, (6767), pp. 335–338

[25] HOOSHANGI S., THIBERGE S., WEISS R.: 'Ultrasensitivity and noise propagation in a synthetic transcriptional cascade', *Proc. Natl. Acad. Sci. USA*, 2005, **102**, (10), pp. 3581–3586

[26] HOOSHANGI S., WEISS R.: 'The effect of negative feedback on noise propagation in transcriptional gene networks', *Chaos*, 2006, **16**, (2), p. 026108

[27] PEDRAZA J.M., VAN OUDENAARDEN A.: 'Noise propagation in gene networks', *Science*, 2005, **307**, (5717), pp. 1965–1969

[28] ELOWITZ M.B., LEVINE A.J., SIGGIA E.D., SWAIN P.S.: 'Stochastic gene expression in a single cell', *Science*, 2002, **297**, (5584), pp. 1183–1186

[29] RASER J.M., O'SHEA E.K.: 'Control of stochasticity in eukaryotic gene expression', *Science*, 2004, **304**, (5678), pp. 1811–1814

[30] OZBUDAK E.M., THATTAI M., KURTSER I., GROSSMAN A.D., VAN OUDENAARDEN A.: 'Regulation of noise in the expression of a single gene', *Nat. Genet.*, 2002, **31**, (1), pp. 69–73

[31] KAERN M., BLAKE W.J., COLLINS J.J.: 'The engineering of gene regulatory networks', *Annu. Rev. Biomed. Eng.*, 2003, **5**, pp. 179–206

[32] BLAKE W.J., KAERN M., CANTOR C.R., COLLINS J.J.: 'Noise in eukaryotic gene expression', *Nature*, 2003, **422**, (6932), pp. 633–637

[33] ISAACS F.J., HASTY J., CANTOR C.R., COLLINS J.J.: 'Prediction and measurement of an autoregulatory genetic module', *Proc. Natl. Acad. Sci. USA*, 2003, **100**, (13), pp. 7714–7719

[34] ROSENFELD N., PERKINS T.J., ALON U., ELOWITZ M.B., SWAIN P.S.: 'A fluctuation method to quantify *in vivo* fluorescence data', *Biophys. J.*, 2006, **91**, (2), pp. 759–766

[35] RASER J.M., O'SHEA E.K.: 'Noise in gene expression: origins, consequences, and control', *Science*, 2005, **309**, pp. 2010–2013

[36] BHALLA U.S., IYENGAR R.: 'Emergent properties of networks of biological signaling pathways', *Science*, 1999, **283**, (5400), pp. 381–387

[37] CAI L., FRIEDMAN N., XIE X.S.: 'Stochastic protein expression in individual cells at the single molecule level', *Nature*, 2006, **440**, (7082), pp. 358–362

[38] METTETAL J.T., MUZZEY D., PEDRAZA J.M., OZBUDAK E.M., VAN OUDENAARDEN A.: 'Predicting stochastic gene expression dynamics in single cells', *Proc. Natl. Acad. Sci. USA*, 2006, **103**, (19), pp. 7304–7309

[39] LITTLE J.W., SHEPLEY D.P., WERT D.W.: 'Robustness of a gene regulatory circuit', *EMBO J.*, 1999, **18**, (15), pp. 4299–4307

[40] PTASHNE M.: 'A genetic switch: phage lambda and higher organisms' (Cell Press and Blackwell Science, Cambridge, 1992, 2nd edn.)

[41] MCADAMS H.H., ARKIN A.: 'Stochastic mechanisms in gene expression', *Proc. Natl. Acad. Sci. USA*, 1997, **94**, (3), pp. 814–819

[42] ARKIN A., ROSS J., MCADAMS H.H.: 'Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells', *Genetics*, 1998, **149**, (4), pp. 1633–1648

[43] SANTILLAN M., MACKEY M.C.: 'Influence of catabolite repression and inducer exclusion on the bistable behavior of the lac operon', *Biophys. J.*, 2004, **86**, (3), pp. 1282–1292

[44] REINITZ J., VAISNYS J.R.: 'Theoretical and experimental analysis of the phage lambda genetic switch implies missing levels of co-operativity', *J. Theor. Biol.*, 1990, **145**, (3), pp. 295–318

[45] AURELL E., SNEPPEN K.: 'Epigenetics as a first exit problem', *Phys. Rev. Lett.*, 2002, **88**, (4), p. 048101

[46] ZHU X.M., YIN L., HOOD L., AO P.: 'Robustness, stability and efficiency of phage lambda genetic switch: dynamical structure analysis', *J. Bioinform. Comput. Biol.*, 2004, **2**, (4), pp. 785–817

[47] KWON C., AO P., THOULESS D.J.: 'Structure of stochastic dynamics near fixed points', *Proc. Natl. Acad. Sci. USA*, 2005, **102**, (37), pp. 13029–13033

[48] RAMOS A.F., HORNOS J.E.M.: 'Symmetry and stochastic gene regulation', *Phys. Rev. Lett.*, 2007, **99**, (10), p. 108103

[49] PECCOUD J., YCART B.: 'Markovian modelling of gene product synthesis', *Theor. Popul. Biol.*, 1995, **48**, pp. 222–234

[50] KEPLER T.B., ELSTON T.C.: 'Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations', *Biophys. J.*, 2001, **81**, (6), pp. 3116–3136

[51] THATTAI M., VAN OUDENAARDEN A.: 'Intrinsic noise in gene regulatory networks', *Proc. Natl. Acad. Sci. USA*, 2001, **98**, (15), pp. 8614–8619

[52] SASAI M., WOLYNES P.G.: 'Stochastic gene expression as a many-body problem', *Proc. Natl. Acad. Sci. USA*, 2003, **100**, (5), pp. 2374–2379

[53] HORNOS J.E.M., SCHULTZ D., INNOCENTINI G.C.F., ET AL.: 'Self-regulating gene: an exact solution', *Phys. Rev. E*, 2005, **72**, (5 Part 1), p. 051907

[54] PAULSSON J.: 'Models of stochastic gene expression', *Phys. Life. Rev.*, 2005, **2**, pp. 157–175

[55] TIAN T., BURRAGE K.: 'Stochastic models for regulatory networks of the genetic toggle switch', *Proc. Natl. Acad. Sci. USA*, 2006, **103**, (22), pp. 8372–8377

[56] KAERN M., ELSTON T.C., BLAKE W.J., COLLINS J.J.: 'Stochasticity in gene expression: from theories to phenotypes', *Nat. Rev. Genet.*, 2005, **6**, (6), pp. 451–464

[57] FRIEDMAN N., CAI L., XIE X.S.: 'Linking stochastic dynamics to population distribution: an analytical framework of gene expression', *Phys. Rev. Lett.*, 2006, **97**, (16), p. 168302

[58] INNOCENTINI G.C.F., HORNOS J.E.M.: 'Modeling stochastic gene expression under repression', *J. Math. Biol.*, 2007, **55**, (3), pp. 413–431

[59] KIM K.Y., LEPZELTER D., WANG J.: 'Dynamics and intrinsic statistical fluctuations of a gene switch', *J. Phys. Chem. B*, 2007, **111**, (10239), pp. 10239–10247

[60] KIM K.Y., LEPZELTER D., WANG J.: 'Single molecule dynamics and statistical fluctuations of gene regulatory networks: a repressilator', *J. Chem. Phys.*, 2007, **126**, (3), p. 034702

[61] LU T., SHEN T., BENNETT M.R., WOLYNES P.G., HASTY J.: 'Phenotypic variability of growing cellular populations', *Proc. Natl. Acad. Sci. USA*, 2007, **104**, (48), pp. 18982–18987

[62] GILLESPIE D.T.: 'Exact stochastic simulation of coupled chemical reactions', *J. Phys. Chem.*, 1977, **81**, (25), pp. 2340–2351

[63] HASTY J., MCMILLEN D., ISAACS F., COLLINS J.J.: 'Computational studies of gene regulatory networks: in numero molecular biology', *Nat. Rev. Genet.*, 2001, **2**, (4), pp. 268–279

[64] ANDRIEUX D., GASPARD P.: 'Temporal disorder and fluctuation theorem in chemical reactions', *Phys. Rev. E*, 2008, **77**, p. 031137

[65] CRICK F.H.C.: 'The origin of the genetic code', *J. Mol. Biol.*, 1968, **38**, pp. 367–379

[66] OSAWA S., JUKES T.H., WATANABE K., MUTO A.: 'Recent evidence for evolution of the genetic code', *Microbiol. Rev.*, 1992, **56**, (1), pp. 229–264

[67] TRIFONOV E.N.: 'Consensus temporal order of amino acids and evolution of the triplet code', *Gene*, 2000, **261**, (1), pp. 139–151

[68] WOESE C.R.: 'Translation: in retrospect and prospect', *RNA*, 2001, **7**, (8), pp. 1055–1067

[69] ARBER W.: 'The evolutionary strategy of DNA acquisition as a possible reason for a universal genetic code', *Hist. Philos. Life Sci.*, 2006, **28**, (4), pp. 525–532

[70] LUO L., LI X.: 'Construction of genetic code from evolutionary stability', *Biosystems*, 2002, **65**, pp. 83–97

[71] FREELAND S.J., KNIGHT R.D., LANDWEBER L.F., HURST L.D.: 'Early fixation of an optimal genetic code', *Mol. Biol. Evol.*, 2000, **17**, (4), pp. 511–518

[72] HORNOS J.E.M., HORNOS Y.M.M.: 'Algebraic model for the evolution of the genetic code', *Phys. Rev. Lett.*, 1993, **71**, (26), pp. 4401–4404

[73] FORGER M., HORNOS Y.M.M., HORNOS J.E.M.: 'Global aspects in the algebraic approach to the genetic code', *Phys. Rev. E*, 1997, **56**, pp. 7078–7082

[74] HORNOS J.E.M., HORNOS Y.M.M., FORGER M.: 'Symmetry and symmetry breaking: an algebraic approach to the genetic code', *Int. J. Mod. Phys. B*, 1999, **13**, pp. 2795–2885

[75] BASHFORD J.D., TSOHANTJIS I., JARVIS P.D.: 'A supersymmetric model for the evolution of the genetic code', *Proc. Natl. Acad. Sci. USA*, 1998, **95**, (3), pp. 987–992

[76] FORGER M., SACHSE S.: 'Lie superalgebras and the multiplet structure of the genetic code I. Codon representations', *J. Math. Phys.*, 2000, **41**, pp. 5407–5422

[77] FORGER M., SACHSE S.: 'Lie superalgebras and the multiplet structure of the genetic code II. Branching schemes', *J. Math. Phys.*, 2000, **41**, pp. 5423–5444

[78] ANTONELI F., BRAGGION L., FORGER M., HORNOS J.E.M.: 'Extending the search for symmetries in the genetic code: finite groups', *Int. J. Mod. Phys. B*, 2003, **17**, pp. 3135–3204

[79] ANTONELI F., FORGER M., HORNOS J.E.M.: 'The search for symmetries in the genetic code: finite groups', *Mod. Phys. Lett. B*, 2004, **18**, pp. 971–978

[80] ANTONELI F., FORGER M., GAVIRIA P.A., HORNOS J.E.M.: 'On amino acid and codon assignment in algebraic models for the genetic code', *Int. J. Mod. Phys. B*, 2010, **24**, pp. 435–463

[81] ALBERTI S.: 'The origin of the genetic code and protein synthesis', *J. Mol. Evol.*, 1997, **45**, (4), pp. 352–358

[82] DELARUE M.: 'An asymmetric underlying rule in the assignment of codons: possible clue to a quick early evolution of the genetic code via successive binary choices', *RNA*, 2007, **13**, (2), pp. 161–169

[83] FINDLEY G.L., FINDLEY A.M., MCGLYNN S.P.: 'Symmetry characteristics of the genetic code', *Proc. Natl. Acad. Sci. USA*, 1982, **79**, (22), pp. 7061–7065

[84] HORNOS J.E.M., BRAGGION L., MAGINI M., FORGER M.: 'Symmetry preservation in the evolution of the genetic code', *IUBMB Life*, 2004, **56**, (3), pp. 125–130

[85] SANCHEZ R., GRAU R., MORGADO E.: 'A novel Lie algebra of the genetic code over the Galois field of four DNA bases', *Math. Biosci.*, 2006, **202**, (1), pp. 156–174

[86] BALAKRISHNAN J.: 'Symmetry scheme for amino acid codons', *Phys. Rev. E*, 2002, **65**, (2 part 1), p. 021912

[87] KNIGHT R.D., LANDWEBER L.F.: 'The early evolution of the genetic code', *Cell*, 2000, **101**, (6), pp. 569–572

[88] TAYLOR F.J., COATES D.: 'The code within the codons', *Biosystems*, 1989, **22**, (3), pp. 177–187

[89] SCHULTZ D.W., YARUS M.: 'On malleability in the genetic code', *J. Mol. Evol.*, 1996, **42**, (5), pp. 597–601

[90] WYBOURNE B.G.: 'Classical groups for physicists' (John Wiley, New York, 1974)

[91] BARUT A.O., RACZKA R.: 'Theory of group representations and applications' (World Scientific, Singapore, 1986, 2nd edn.)

[92] GILMORE R.: 'Lie groups, Lie algebras, and some of their representations' (Dover, New York, 2006, 3rd edn.)

[93] ABRAMOWITZ M., STEGUN I.A. (ED.): 'Handbook of mathematical functions' (US GPO, Washington, 1972)

[94] BATEMAN H., ERDELYI A.: 'Higher transcendental functions' (McGraw-Hill, New York, 1953, 1st edn.)

[95] BUTKOV E.: 'Mathematical physics' (Addison-Wesley, 1968)

[96] PALDI A.: 'Stochastic gene expression during cell differentiation: order from disorder?', *Cell. Mol. Life Sci.*, 2003, **60**, (9), pp. 1775–1778

[97] FANO U.: 'Ionization yield of radiations II. The fluctuations of the number of ions', *Phys. Rev.*, 1947, **72**, pp. 26–29

[98] GALLAGHER T.F.: 'Rydberg atoms' (Cambridge University Press, Cambridge, 1994)

[99] MCKAY W., PATERA J.: 'Table of dimensions, indices and branching rules for representations of simple Lie algebras' (Marcel Dekker, New York, 1981)

8   Appendix

Most of the explicit formulas stated in Section 3 have been given for the special case $\chi = 0$ only. In order to deal with the case where $\chi \neq 0$, the following substitutions must be made:

The differential equation 25 is replaced by

$$\frac{z-1}{N(1-\chi)}\frac{\mathrm{d}^2\alpha}{\mathrm{d}z^2} + \frac{1+\epsilon - N(1+\chi)(z-1)}{N(1-\chi)}\frac{\mathrm{d}\alpha}{\mathrm{d}z}$$
$$- \frac{1 + \epsilon p_\alpha + \chi\epsilon p_\beta - N\chi(z-1)}{1-\chi}\alpha = 0 \qquad (53)$$

The expression for its solution given by (28) is replaced by

$$\alpha(z) = p_\alpha \exp[N\chi(z-1)]M(1+\epsilon p_\alpha, 1+\epsilon, N(1-\chi)(z-1))$$
$$\beta(z) = (1 - p_\alpha)\exp[N\chi(z-1)]M(\epsilon p_\alpha, 1+\epsilon, N(1-\chi)(z-1)) \qquad (54)$$

The expressions for the probability distributions given by (29)

are replaced by

$$\alpha_n = p_\alpha e^{-N\chi}\frac{N^n(1-\chi)^n}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\left(\frac{\chi}{1-\chi}\right)^{n-k}$$
$$\times \frac{(1+\epsilon p_\alpha)_k}{(1+\epsilon)_k}M(1+\epsilon p_\alpha + k, 1+\epsilon + k, N(\chi-1))$$
$$\beta_n = (1 - p_\alpha)e^{-N\chi}\frac{N^n(1-\chi)^n}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\left(\frac{\chi}{1-\chi}\right)^{n-k}$$
$$\times \frac{(\epsilon p_\alpha)_k}{(1+\epsilon)_k}M(\epsilon p_\alpha + k, 1+\epsilon + k, N(\chi-1)) \qquad (55)$$

and the operators of (30) and (33) are replaced by

$$L_z = \frac{z-1}{N(1-\chi)}\frac{\mathrm{d}^2}{\mathrm{d}z^2} + \frac{2(l+1) - N(1+\chi)(z-1)}{N(1-\chi)}\frac{\mathrm{d}}{\mathrm{d}z}$$
$$+ \frac{N\chi(z-1) - (1+\chi)(1+l)}{1-\chi}$$
$$L_+ = \frac{z-1}{N(1-\chi)}\frac{\mathrm{d}^2}{\mathrm{d}z^2} + 2\frac{(l+1) - N\chi(z-1)}{N(1-\chi)}\frac{\mathrm{d}}{\mathrm{d}z}$$
$$+ \chi\frac{N\chi(z-1) - 2(1+l)}{1-\chi}$$
$$L_- = \frac{z-1}{N(1-\chi)}\frac{\mathrm{d}^2}{\mathrm{d}z^2} + 2\frac{(l+1) - N(z-1)}{N(1-\chi)}\frac{\mathrm{d}}{\mathrm{d}z}$$
$$+ \frac{N(z-1) - 2(1+l)}{1-\chi} \qquad (56)$$

Finally, (50) has been obtained from the formulae

$$\phi(z) = \alpha(z) + \beta(z) = \exp[N\chi(z-1)]$$
$$\times M(m+l, 2l+1, N(1-\chi)(z-1)),$$
$$\langle n^p \rangle = \left[\left(z\frac{\mathrm{d}}{\mathrm{d}z}\right)^p \phi\right]_{z=1} \qquad (57)$$

where we employed (13.4.3) from [93].