Global aspects in the algebraic approach to the genetic code

Michael Forger
Departamento de Matemática Aplicada, Instituto de Matemática e Estatística, Universidade de São Paulo, Caixa Postal 66281, 05315-970 São Paulo, São Paulo, Brazil

Yvone M. M. Hornos
SAPRA, Rua Douter Orlando Damiano 2160, 13560-990 São Carlos, São Paulo, Brazil

José Eduardo Martinho Hornos
Departamento de Física e Ciência dos Materiais, Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, 13560-970 São Carlos, São Paulo, Brazil

(Received 30 May 1997)

The algebraic approach to the genetic code is further developed to incorporate global properties. As a result strong restrictions on the possibilities of assignment of amino acids and codons to representation vectors are found. Extending the search for symmetry breaking schemes to include nonconnected subgroups, a possibility is found based on the exceptional group G_2. [S1063-651X(97)06311-3]

PACS number(s): 87.10.+e

Finding an explanation for the observed degeneracies in the genetic code, or more precisely, in the assignment of codons to amino acids, as shown in Table I, has been a long-standing problem in molecular biology. These degeneracies are an immediate consequence of the fact that genetic information is stored in DNA and RNA in the form of 64 codons, which are triplets built from the four nucleic bases adenine (A), cytosine (C), guanine (G), and thymine (T, in DNA) or uracil (U, in RNA), whereas only 20 amino acids occur in biologically synthesized proteins. The problem of the evolution of the genetic code has been extensively discussed in the literature, for example, by Jukes and coworkers [1], with emphasis on the relation between codons and anticodons and on the conclusions that can be drawn by comparing the standard code with exceptional codes, such as the ones observed in mitochondria.

An alternative approach has recently been suggested [2,3], based on the notion of symmetry breaking [4,5], which over the last few decades has been widely used in particle physics [6], nuclear physics [7], and molecular physics [8]. The phenomenon is also observed in chemical and biological systems; as an example, we may think of the breakdown of chiral invariance. There are various types of symmetry principles that are important in different areas of science, one of the most interesting being the notion of dynamical symmetry: it may not be directly visible in the equations of motion or the Hamiltonian of the system but nevertheless severely restricts its dynamics. As an example, consider the SO(4) symmetry in the Kepler-Coulomb problem, generated by angular momentum together with the Runge-Lenz vector, which is responsible for the fact that the bounded trajectories of the classical Kepler problem are closed orbits and that the quantum mechanical bound states of the hydrogen atom are degenerate not only with the magnetic quantum number m but also with the angular momentum quantum number l. When this symmetry is broken, there appear phenomena such as perihelion rotation or splitting of spectral lines.

The central idea of the algebraic approach to the genetic code is that the presence of synonymous codons expresses invariance of the translation process under the interchange of codons assigned to the same amino acid and reflects the presence of a dynamical symmetry. In order to implement this picture, a systematic search for possible symmetry breaking schemes has been performed, based on the Cartan classification theorem for complex semisimple Lie algebras, by starting out with a 64-dimensional irreducible representation of a given simple Lie algebra (there are just a few of these, which will be referred to as codon representations) and decomposing it along all possible chains of maximal subalgebras. In a first stage, these chains can be pursued until one reaches a direct sum of su(2) subalgebras, the result being that there is *no* chain that would allow one to reproduce exactly the de-

TABLE I. The standard genetic code.

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

TABLE II. Branching rules in the sp(6) chain. The symbol ''hw'' means ''highest weight,'' whereas s_i and m_i denote the spin and the magnetic quantum number with respect to the ith su(2). The shading indicates the multiplets, which in the last step are frozen.

step 1	step 2	step 3		step 4			
sp(6)	sp(4) × su(2)	su(2)×su(2)×su(2)	⊃	su(2)×so(2)×su(2)	⊃	su(2)×so(2)× o(2)	
hw	hw	spin–spin–spin	dim	$spin_1$- m_2-$spin_3$	dim	$spin_1$- m_2-m_3	dim
(1,1,0)	(1,1)–(10)	(1)–(1/2)–(0)	6	(1)-(±1/2)-(0)	6	(1)-(±1/2)-(0)	6
		(1/2)–(1)–(0)	6	(1/2)–(±1)–(0)	4	(1/2)–(±1)–(0)	4
				(1/2)–(0)–(0)	2	(1/2)–(0)–(0)	2
		(1/2)–(0)–(0)	2	(1/2)–(0)–(0)	2	(1/2)–(0)–(0)	2
		(0)-(1/2)-(0)	2	(0)-(1/2)-(0)	2	(0)-(1/2)-(0)	2
	(1,0)–(0)	(1/2)–(0)–(0)	2	(1/2)–(0)–(0)	2	(1/2)–(0)–(0)	2
		(0)–(1/2)–(0)	2	(0)–(1/2)–(0)	2	(0)–(1/2)–(0)	2
	(2,0)-(1/2)	(1/2)-(1/2)-(1/2)	8	(1/2)-(±1/2)-(1/2)	8	(1/2)-(±1/2)-(+1/2)	4
						(1/2)-(±1/2)-(-1/2)	4
		(1)-(0)-(1/2)	6	(1)-(0)-(1/2)	6	(1)-(0)-(+1/2)	3
						(1)-(0)-(-1/2)	3
		(0)–(1)–(1/2)	6	(0)-(±1)-(1/2)	4	(0)-(±1)-(+1/2)	2
						(0)-(±1)-(-1/2)	2
				(0)-(0)-(1/2)	2	(0)-(0)-(+1/2)	1
						(0)-(0)-(-1/2)	1
	(0,1)–(1/2)	(1/2)–(1/2)–(1/2)	8	(1/2)-(±1/2)-(1/2)	8	(1/2)-(±1/2)-(+1/2)	4
						(1/2)-(±1/2)-(-1/2)	4
		(0)–(0)–(1/2)	2	(0)–(0)–(1/2)	2	(0) - (0) - (+1/2)	1
						(0)-(0)-(-1/2)	1
	(0,0)-(1/2)	(0)-(0)-(1/2)	2	(0)-(0)-(1/2)	2	(0) - (0) - (+1/2)	1
						(0)-(0)-(-1/2)	1
	(1,0)–(1)	(1/2)–(0)–(1)	6	(1/2)–(0)–(1)	6	(1/2) - (0) - (+1)	2
						(1/2) - (0) - (-1)	2
						(1/2) - (0) - (0)	2
		(0)–(1/2)–(1)	6	(0)–(±1/2)–(1)	6	(0)- (±1/2) -(+1)	2
						(0)-(±1/2)-(-1)	2
						(0) - (±1/2) -()	2
	6 subspaces	14 subspace		16 sub-spaces		27 subspaces	

generacies of the genetic code. Therefore, a second stage was considered, in which the su(2) subalgebras are allowed to break into Abelian u(1) subalgebras, imposing that the states involved should be labeled in one of two ways: as eigenvectors of L_z or of L_z^2. In the first case, the degeneracy is completely lifted, whereas in the second case, states with ''magnetic quantum number'' $+m$ and $-m$ remain degenerate. With this prescription, it was shown that the degeneracies of the genetic code can be reproduced by starting out from the codon representation of the rank 3 symplectic algebra sp(6) and breaking the symmetry through the following chain of subalgebras:

$$\mathrm{sp}(6)\supset\mathrm{sp}(4)\oplus\mathrm{su}(2)\supset\mathrm{su}(2)\oplus\mathrm{su}(2)\oplus\mathrm{su}(2)$$
$$\supset\mathrm{su}(2)\oplus\mathrm{u}(1)^*\oplus\mathrm{su}(2)\supset\mathrm{su}(2)\oplus\mathrm{u}(1)^*\oplus\mathrm{u}(1)^F. \quad (1)$$

The asterisk in the second u(1) indicates that the states are labeled by $L_{2,z}^2$ rather than by $L_{2,z}$, and the superscript F in the third u(1) means that the some multiplets have during this last step of the symmetry breaking remained unbroken, or frozen. The branching rules for this chain are given in Table II and the weight diagram of the codon representation of sp(6) is shown in Fig. 1.

In the aforementioned analysis, local properties have been taken into account. In mathematical terms, the search was performed in the context of Lie algebras, not of Lie groups. In the present paper, we report on results obtained by incorporating global aspects.

The first observation is that the symmetry breaking from su(2) to u(1)* referred to above can be interpreted as a full-fledged symmetry breaking from the group SU(2) to its maximal subgroup $\mathrm{U}(1)^*=\mathrm{Z}_2\times\mathrm{U}(1)\cong\mathrm{O}(2)$, which consists of two connected components and is larger than the maximal connected subgroup $\mathrm{U}(1)\cong\mathrm{SO}(2)$. The distinction between the two can be implemented naturally at the level of Lie groups, but not of Lie algebras. The additional Z_2 factor, which can be represented, e.g., by i times the first Pauli matrix σ_1, has exactly the effect of transforming states of opposite magnetic quantum number, $+m$ and $-m$, into each other. In terms of groups, the symmetry breaking in the model of Ref. [2] can be summarized as follows. Step 1: the primordial symplectic group Sp(6) is broken to its maximal connected subgroup $\mathrm{Sp}(4)\times\mathrm{SU}(2)$; step 2: the Sp(4) factor is broken to its maximal connected subgroup $\mathrm{SU}(2)\times\mathrm{SU}(2)$; step 3: the second SU(2) factor is broken to its maximal subgroup $\mathrm{U}(1)^*=\mathrm{Z}_2\times\mathrm{U}(1)$; step 4: the third SU(2) factor is

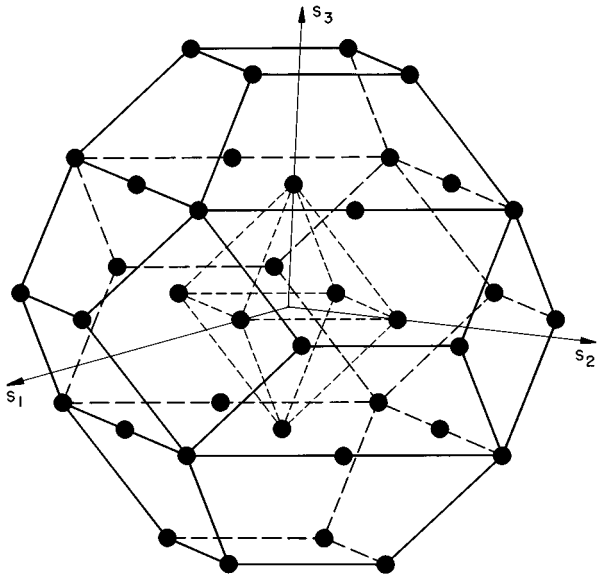

FIG. 1. Weight diagram of the codon representation of sp(6). In the small interior octahedron the weights are fourfold degenerate, in the center of the hexagons the weights are twofold degenerate, and the others are nondegenerate.

TABLE III. The symbol ''hw'' means ''highest weight,'' whereas s_i and m_i denote the spin and the magnetic quantum number with respect to the ith su(2). The shading indicates the multiplets, which in the last step are frozen.

Step 1		Step 2		Step 3			
G_2		$\supset$ su(2) $\oplus$ su(2)		$\supset$ su(2) $\oplus$ u(1)*		$\supset$ su(2) $\oplus$ u(1)	
hw	dim	$s_1 - s_2$	dim	$s_1 - m_2$	dim	$s_1 - m_2$	dim
(1,1)	64	(1)–(2)	15	(1)–(±2)	6	(1)–(+2)	3
						(1)–(−2)	3
				(1)–(±1)	6	(1)–(+1)	3
						(1)–(−1)	3
				(1)–(0)	3	(1)–(0)	3
		(1/2)–(5/2)	12	(1/2)–(±5/2)	4	(1/2)–(+5/2)	2
						(1/2)–(−5/2)	2
				(1/2)–(±3/2)	4	(1/2)–(+3/2)	2
						(1/2)–(−3/2)	2
				(1/2)–(±1/2)	4	(1/2)–(+1/2)	2
						(1/2)–(−1/2)	2
		(1)–(1)	9	(1)–(±1)	6	(1)–(+1)	3
						(1)–(−1)	3
				(1)–(0)	3	(1)–(0)	3
		(3/2)–(1/2)	8	(3/2)–(±1/2)	8	(3/2)–(+1/2)	4
						(3/2)–(−1/2)	4
		(1/2)–(3/2)	8	(1/2)–(±3/2)	4	(1/2)–(+3/2)	2
						(1/2)–(−3/2)	2
				(1/2)–(±1/2)	4	(1/2)–(+1/2)	2
						(1/2)–(−1/2)	2
		(0)–(2)	5	(0)–(±2)	2	(0)–(+2)	1
						(0)–(−2)	1
				(0)–(±1)	2	(0)–(+1)	1
						(0)–(−1)	1
				(0)–(0)	1	(0)–(0)	1
		(1/2)–(1/2)	4	(1/2)–(±1/2)	4	(1/2)–(+1/2)	2
						(1/2)–(−1/2)	2
		(0)–(1)	3	(0)–(±1)	2	(0)–(+1)	1
						(0)–(−1)	1
				(0)–(0)	1	(0)–(0)	1
1 space		8 subspaces		17 subspaces		30 subspaces	

partly broken to its maximal connected subgroup U(1), where the term ''partly'' refers to the freezing phenomenon, as explained in Ref. [2].

With this interpretation in mind, we have reexamined the possible chains originating from simple Lie groups of low rank (2 and 3), relaxing the purely algebraic assumptions of Ref. [1] to allow for symmetry breaking through nonconnected subgroups in the second stage of the procedure [breaking of products of SU(2) subgroups]. The result is that besides the chain (1), there appears a single new chain:

$$G_2 \supset su(2)\oplus su(2) \supset su(2)\oplus u(1)^* \supset su(2)\oplus u(1)^F. \qquad (2)$$

The branching rules for this chain are given in Table III, and the weight diagram of the codon representation of G_2 is shown in Fig. 2. In terms of groups, the symmetry breaking in this model can be summarized as follows. Step 1: the primordial group G_2 is broken to its maximal subgroup $SU(2)\times SU(2)$; step 2: the second SU(2) factor is broken to its maximal subgroup U(1)*; step 3: the second SU(2) factor is partly further broken to its maximal connected subgroup U(1), where again the term ''partly'' refers to the freezing phenomenon.

The main advantage of this model is the low rank of the group. However, freezing in the last step is in this case more the rule than the exception: the majority of the $SU(2)\times U(1)^*$ multiplets must be frozen. If no freezing occurred, the scheme would generate 30 amino acids, with rather strange splittings of all sextuplets into triplets and the appearance of many singlets.

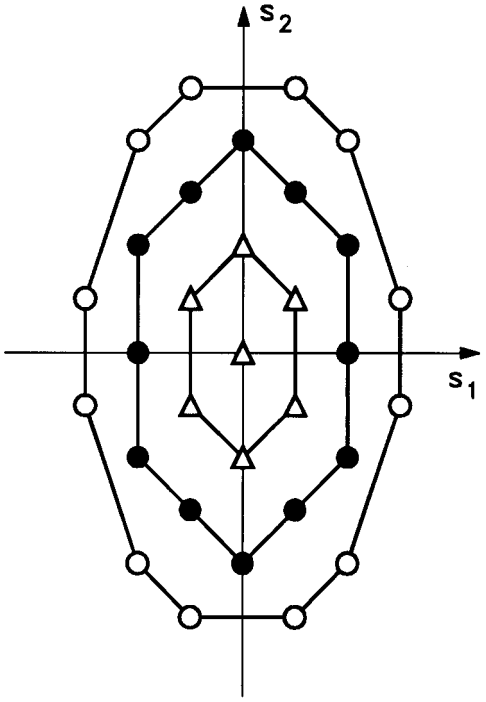

FIG. 2. Weight diagram of the codon representation of G_2. The weights represented by triangles are fourfold degenerate, the black circles are twofold degenerate, and the white circles are nondegenerate.

In both chains, all steps except the last correspond to complete symmetry breaking of groups to subgroups and are therefore consistent with the picture of a spontaneously broken symmetry.

In analogy to the procedure adopted in nuclear physics [7] or molecular physics [8] when dealing with spectrum generating algebras, each of these symmetry chains can be represented using the Casimir operators of the subalgebras that appear in the chain, which automatically form a family of commuting operators. The operator H associated with the chain is a polynomial in these Casimir operators that to each of the 21 degenerate subspaces appearing at the end of the chain associates a different eigenvalue. For the codon representation of sp(6) it is

$$H_{sp(6)} = h_0 + h_1 C_{sp(4)} + q_1 L_1^2 + q_2 L_2^2 + q_3 L_3^2 + p_2 L_{2,z}^2 + p_3 (L_1^2 + L_2^2)(L_3^2 - 2) L_{3,z}, \qquad (3)$$

and for the codon representation of G_2 it is

$$H_{G_2} = h_0 + q_1 L_1^2 + q_2 L_2^2 + p_2 L_{2,z}^2 + p_3 (L_2^2 - 2)(L_2^2 - 6)(L_2^2 - 35/4) L_{2,z}, \qquad (4)$$

where h_0, q_1, q_2, q_3, p_2, and p_3 are arbitrary constants, $C_{sp(4)}$ is the quadratic Casimir operator of sp(4), L_1^2, L_2^2, L_3^2 are the squared angular momentum operators for the su(2)'s and $L_{2,z}$, $L_{3,z}$ the z components of the angular momentum

operators for the u(1)'s involved; note that L^2 is the Casimir operator of su(2) and L_z is the invariant operator of u(1). The factors that multiply the L_z operators are responsible for the freezing. Note further that in both chains the freezing term is a polynomial in the squared angular momentum operators, of order 6 in the G_2 chain and only of order 4 in the sp(6) chain.

The assignment of amino acids to representation vectors in the codon space must be performed in accordance with the restrictions imposed by the degeneracy of the genetic code. This requirement alone, however, leaves an enormous amount of freedom. Rearrangements of amino acids with the same degeneracy will be possible in 3!5!2!9!2! different ways, reflecting the freedom in assigning the sextuplets, quadruplets, triplets, doublets, and singlets, respectively, totaling more than a billion alternatives. The selection of a particular option must be performed using arguments that go beyond the restrictions imposed by degeneracy alone. A particular assignment has been proposed in Ref. [2], where chemical and biological properties of the amino acids, such as their polarities or their presumed role in the evolution of the first forms of life, have been used as a guide. However, such considerations can be performed from different points of view, depending, for example, on which biological properties are emphasized.

In the following, we show how one can obtain strong restrictions on these degrees of freedom by combining well-established biochemical facts with further global symmetry considerations.

The weight diagram of the codon representation of sp(6) shown in Fig. 1 is formed by a small interior octahedron and a large exterior truncated octahedron, with eight hexagonal faces and six square faces; this geometric figure is well known in solid state physics as the surface of a Wigner-Seitz cell. The weights in the vertices of the truncated octahedron are nondegenerate, the weights in the center of the hexagons are twofold degenerate, and the weights in the vertices of the interior octahedron are fourfold degenerate. It is convenient to divide the weight diagram into five horizontal planes, located at $z=+1$, $+1/2$, 0, $-1/2$, and -1, respectively, and related by a symmetry along the vertical axis, so that the root generators (raising and lowering operators) of the sp(4) subalgebra and in particular of the first two su(2) subalgebras act along the horizontal planes, whereas those of the third su(2) subalgebra act along the vertical axis (third coordinate axis). [In fact, those of the first and/or second su(2) subalgebra act along the first and/or second coordinate axis.]

This leads to a natural distinction between the representations that appear after the first step of the symmetry breaking [from sp(6) to sp(4) $\oplus$ su(2)]; cf. Table II. We shall call representation vectors in the codon space *bosonic* or *tenso-*

TABLE IV. New proposal for codon and amino acid assignments in the sp(6) model. States are labeled as $|k_1,k_2;s_1,s_2,s_3,m_1,m_2,m_3\rangle$, where k_1,k_2 are the components of the highest weight of the representation of the sp(4) subalgebra that the state belongs to, whereas s_i and m_i denote the spin and the magnetic quantum number with respect to the tth su(2) subalgebra. (The assignments marked by an asterisk differ from those of Table II of Ref. [2].)

Type of state	Quantum numbers $\lvert k_1,k_2;s_1,s_2,s_3,m_1,m_2,m_3\rangle$	Amino acid	Codons
Vector bosons	$\lvert 1,0;0,1/2,1;0,\pm 1/2,(\pm 1,0)\rangle$	Leu	*CUU*, *CUC*, *CUA* *CUG*, *UUA*, *UUG*
	$\lvert 1,0;1/2,0,1;\pm 1/2,0,(\pm 1,0)\rangle$	Ser	*UCU*, *UCC*, *UCA* *UCG*, *AGU*, *AGC*
Scalar bosons	$\lvert 1,1;1,1/2,0;(\pm 1,0),\pm 1/2,0\rangle$	Arg	*CGC*, *CGG*, *CGU* *CGA*, *AGA*, *AGG*
	$\lvert 1,1;1/2,1,0;\pm 1/2,\pm 1,0\rangle$	Ala	*GCC*, *GCG*, *GCU*, *GCA*
	$\lvert 1,1;1/2,1,0;\pm 1/2,0,0\rangle$	Phe*	*UUU*, *UUC*
	$\lvert 1,1;0,1/2,0;0,\pm 1/2,0\rangle$	Asp	*GAU*, *GAC*
	$\lvert 1,1;1/2,0,0;\pm 1/2,0,0\rangle$	Glu	*GAA*, *GAG*
	$\lvert 1,0;0,1/2,0;0,\pm 1/2,0\rangle$	Asn*	*AAU*, *AAC*
	$\lvert 1,0;1/2,0,0;\pm 1/2,0,0\rangle$	Lys*	*AAA*, *AAG*
Fermions	$\lvert 0,1;1/2,1/2,1/2;\pm 1/2,\pm 1/2,+1/2\rangle$	Thr	*ACU*, *ACC*, *ACA*, *ACG*
	$\lvert 0,1;1/2,1/2,1/2;\pm 1/2,\pm 1/2,-1/2\rangle$	Val	*GUU*, *GUC*, *GUA*, *GUG*
	$\lvert 2,0;1/2,1/2,/1/2;\pm 1/2,\pm 1/2,+1/2\rangle$	Pro	*CCU*, *CCC*, *CCA*, *CCG*
	$\lvert 2,0;1/2,1/2,1/2;\pm 1/2,\pm 1/2,-1/2\rangle$	Gly	*GGU*, *GGC*, *GGA*, *GGG*
	$\lvert 0,1;0,0,1/2;0,0,\pm 1/2\rangle$	Gln*	*CAA*, *CAG*
	$\lvert 0,0;0,0,1/2;0,0,\pm 1/2\rangle$	His*	*CAU*, *CAC*
	$\lvert 2,0;0,1,1/2;0,\pm 1,+1/2\rangle$	Tyr*	*UAU*, *UAC*
	$\lvert 2,0;0,1,1/2;0,\pm 1,-1/2\rangle$	Cys*	*UGU*, *UGC*
	$\lvert 2,0;0,1,1/2;0,0,+1/2\rangle$	Trp	*UGG*
	$\lvert 2,0;0,1,1/2;0,0,-1/2\rangle$	Met	*AUG*
	$\lvert 2,0;1,0,1/2;(\pm 1,0),0,+1/2\rangle$	Ile	*AUU*, *AUC*, *AUA*
	$\lvert 2,0;1,0,1/2;(\pm 1,0),0,-1/2\rangle$	Term	*UAA*, *UAG*, *UGA*

rial if they belong to representations of the third su(2) subalgebra with integer spin and *fermionic* or *spinorial* if they belong to representations of the third su(2) subalgebra with half-integer spin. The tensorial representations are associated to the rotation group SO(3) while the spinorial ones require its universal covering group SU(2): both have the same Lie algebra but are globally different. In the sp(6) weight diagram, the bosonic states are the ones located in the planes $z=+1,\ 0,\ -1$ and the fermionic states are the ones located in the planes at $z=+1/2,\ -1/2$; note that there are exactly 32 bosonic states and 32 fermionic states.

We now present an amino acid and codon assignment that is slightly different from the tentative assignment presented in Ref. [2], based on combining two important biological facts with certain discrete symmetries in the codon space. It is well known that the two helices of DNA are complementary under the Watson-Crick pairing rule, which states that *C* pairs with *G* and *U* pairs with *A*; this is also true for the first two bases in the codon-anticodon recognition. Mathematically, this rule can be expressed as a principle of duality: every nucleic base X has a canonical dual nucleic base $X^\dagger$:

$$A^\dagger=U,\quad C^\dagger=G,\quad G^\dagger=C,\quad U^\dagger=A.$$

This duality will be referred to as *Watson-Crick* (or WC) *duality*; it reminds us of the symplectic symmetry in mechanics or thermodynamics where dynamical variables always come in canonically conjugate pairs. Similarly, every codon XYZ has a canonical WC dual codon $(XYZ)^\dagger=Z^\dagger Y^\dagger X^\dagger$; note the inversion of order, which is mathematically compelling and corresponds to the biological fact that the two helices in a DNA molecule run in antiparallel directions. Another obvious and useful fact is the weak dependence of the meaning of a codon on the third base, which has led molecular biologists to organize the rules of the genetic code in the (by now standard) form of Table I, where codons are assembled in family boxes: all codons starting with the same two bases form a family box, and in 8 of the 16 family boxes, they all code for the same amino acid. This leads us to introduce a notion of *partial Watson-Crick duality* for codons, which refers only to the first two bases: thus the partial WC dual of a codon XYN is, by definition, the codon $Y^\dagger X^\dagger N$.

The new assignment is obtained by imposing the following two invariance principles: (1) *Principle of family box completeness:* Codons in the same family box (XYN; $N=U,C,A,G$) are either all bosonic or all fermionic. (2) *Principle of Watson-Crick (or WC) dual completeness:* For any codon that is bosonic or fermionic, the corresponding partial WC dual codon must also be bosonic or fermionic, respectively.

A simple inspection of the dimensions of the multiplets in the sp(6) model (Table II) shows that the triplets and singlets are fermionic and the sextuplets are bosonic. Family box completeness then requires the phenylalanine codons to lie in the bosonic sector and the cysteine and tyrosine codons to lie in the fermionic sector, in order to complete the family boxes CGN (arginine), AGN (serine, arginine), UCN (serine), CUN (leucine), UUN (phenylalanine, leucine) in the bosonic sector and UAN (tyrosine, termination), UGN (cysteine, termination, tryptophan), AUN (isoleucine, methionine) in the fermionic sector. Next, WC dual completeness forces the codons in the family boxes GAN (the dual of UCN), for aspartic and glutamic acid, and AAN (the dual of UUN), for aspartine and lysine, to belong to the bosonic sector and the codons in the family box CAN (the dual of UGN), for histamine and glutamine, to belong to the fermionic sector. All that remains to be done is the allocation of the five quadruplets: one of them will be bosonic and the other four will be fermionic. Family box completeness will be fulfilled by any assignment, but WC dual completeness requires the bosonic quadruplet to be self-dual. The quadruplets are proline (CCN), glycine (GGN), valine (GUN), threonine (ACN), and alanine (GCN), and since only the alanine codons are self-dual, we arrive at the following list of family boxes:

bosonic: UUN, AAN, CGN, AGN, CUN, UCN, GAN, GCN,

fermionic: UAN, AUN, UGN, CAN, CCN, GGN, GUN, ACN.

The resulting codon and amino acid assignment is shown in Table IV. Interestingly, one arrives at exactly the same assignment if, maintaining the requirement of family box completeness, one replaces the condition of WC dual completeness by the condition of *invariance under permutation of the first two bases*.

Concluding, we reemphasize the statement already made in Ref. [2] that symmetry considerations alone cannot replace a microscopic model but just establish a general background.

Beyond the description of the evolution of the genetic code as a symmetry breaking process, the symmetry principles can also serve as a guiding line for the formulation of a dynamical model.

This work was partially supported by Conselho Nacional de Pesquisa (CNPq) and Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP).

[1] S. Ozawa, T. H. Jukes, K. Watanabe, and A. Muto, Microbiol. Rev. **56**, 229 (1992).

[2] J. E. M. Hornos, and Y. M. M. Hornos, Phys. Rev. Lett. **71**, 4401 (1993).

[3] J. E. M. Hornos and Y. M. M. Hornos, J. Biol. Phys. **20**, 289 (1994).

[4] J. Maddox, Nature (London) **367**, 111 (1994).

[5] I. Stewart, New Sci. **141** (1915), 16 (1994).

[6] M. Gell-Mann and Y. Ne'eman, *The Eightfold Way* (Benjamin, New York, 1964).

[7] A. Arima and F. Iachello, *The Interacting Boson Model* (Cambridge University Press, Cambridge, 1987).

[8] F. Iachello and R. D. Levine, *Algebraic Theory of Molecules* (Oxford University Press, Oxford, 1995).