A Minimum Principle in Codon-anticodon Interaction

A.Sciarrino and P.Sorba

Imposing a minimum principle in the framework of the so called Crystal basis model, we determine the structure of the minimum set of 22 anticodons allowing translation- transcription for animal mitochondrial code. Results in good agreement with the observed anticodons.

Plan:

- *position of the problem*
- *the crystal basis model*
- *a minimum principle*

Mini TAGp 2011

sorba@lapp.in2p3.fr

Sept.16

Position of the problem:

codon: XYZ ---- anti-codon: Z'Y'X'

with nucleotids Z',Y',X' associated to Z, Y, X

- In tRNA process, codon –anticodon pairing does not follow the usual Watson-Crick pattern (i.e. pairing C --- G , U --- A).

- This leads Crick (1966) to propose the wobble hypothesis :

A specified anti-codon can recognize more than one codon differing only in the third nucleotide.

i.e. standard pairing for X--X' and Y—Y' while Z' may pair to different Z.

- Two main hypotheses proposed in this context:
  1- for doublets, the first nucleotide Z' in anticodon should have G (resp. U) to read for codon with Y (resp. R) in third position Z.
  2- the chosen anticodon is the one with first position nucleotide pairing the (third position of the) most abundant codon among synonymous codons.

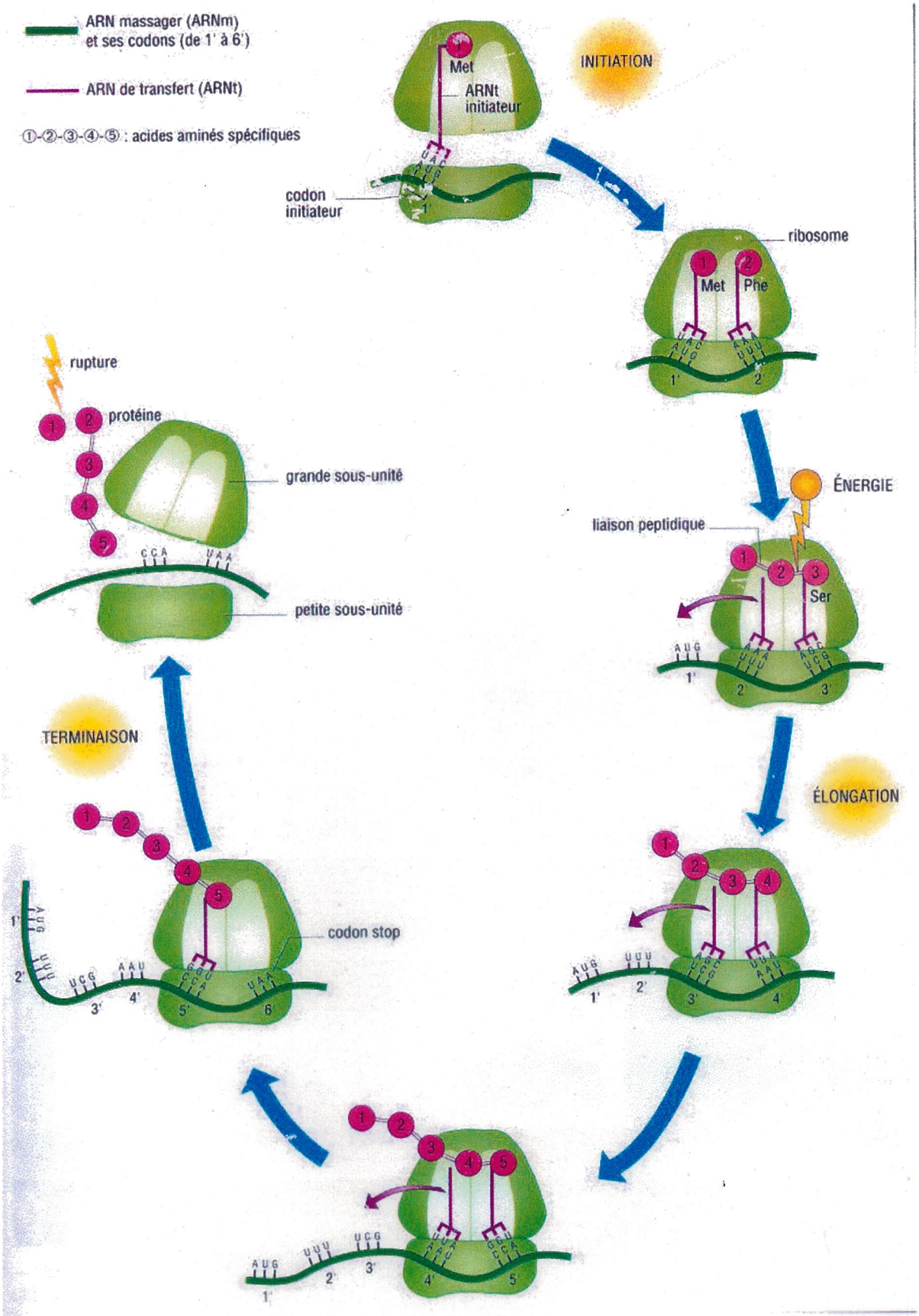
Considering the Mitochondrial Code:

there are: 2 sextets,
6 quadruplets
12 doublets of codons specifying the 20 amino-acids.

So, a minimum number of 22 anticodons is needed.

And this appears to be the case in mitochondria of animals (*Sprinzl et al., 1998*)
Data seem to confirm the empirical rule just above.

It is this set of data that we will consider now in the framework of the Crystal Basis Model.



The different phases of translation in the cytoplasm

codon	a.a.	J_H	J_V	$J_{3,H}$	$J_{3,V}$	anticodon	codon	a.a.	J_H	J_V	$J_{3,H}$	$J_{3,V}$	anticodon
CCC	P	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{3}{2}$		UCC	S	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{1}{2}$	$\frac{3}{2}$	
CCU	P	$(\frac{1}{2}$	$\frac{3}{2})^1$	$\frac{1}{2}$	$\frac{3}{2}$		UCU	S	$(\frac{1}{2}$	$\frac{3}{2})^1$	$-\frac{1}{2}$	$\frac{3}{2}$	
CCG	P	$(\frac{3}{2}$	$\frac{1}{2})^1$	$\frac{3}{2}$	$\frac{1}{2}$	**UGG**	UCG	S	$(\frac{3}{2}$	$\frac{1}{2})^1$	$\frac{1}{2}$	$\frac{1}{2}$	**UGA**
CCA	P	$(\frac{1}{2}$	$\frac{1}{2})^1$	$\frac{1}{2}$	$\frac{1}{2}$		UCA	S	$(\frac{1}{2}$	$\frac{1}{2})^1$	$-\frac{1}{2}$	$\frac{1}{2}$	
CUC	L	$(\frac{1}{2}$	$\frac{3}{2})^2$	$\frac{1}{2}$	$\frac{3}{2}$		UUC	F	$\frac{3}{2}$	$\frac{3}{2}$	$-\frac{1}{2}$	$\frac{3}{2}$	
CUU	L	$(\frac{1}{2}$	$\frac{3}{2})^2$	$-\frac{1}{2}$	$\frac{3}{2}$		UUU	F	$\frac{3}{2}$	$\frac{3}{2}$	$-\frac{3}{2}$	$\frac{3}{2}$	*GAA*
CUG	L	$(\frac{1}{2}$	$\frac{1}{2})^3$	$\frac{1}{2}$	$\frac{1}{2}$	**UAG**	UUG	L	$(\frac{3}{2}$	$\frac{1}{2})^1$	$-\frac{1}{2}$	$\frac{1}{2}$	
CUA	L	$(\frac{1}{2}$	$\frac{1}{2})^3$	$-\frac{1}{2}$	$\frac{1}{2}$		UUA	L	$(\frac{3}{2}$	$\frac{1}{2})^1$	$-\frac{3}{2}$	$\frac{1}{2}$	*UAA*
CGC	R	$(\frac{3}{2}$	$\frac{1}{2})^2$	$\frac{3}{2}$	$\frac{1}{2}$		UGC	C	$(\frac{3}{2}$	$\frac{1}{2})^2$	$\frac{1}{2}$	$\frac{1}{2}$	
CGU	R	$(\frac{1}{2}$	$\frac{1}{2})^2$	$\frac{1}{2}$	$\frac{1}{2}$		UGU	C	$(\frac{1}{2}$	$\frac{1}{2})^2$	$-\frac{1}{2}$	$\frac{1}{2}$	*GCA*
CGG	R	$(\frac{3}{2}$	$\frac{1}{2})^2$	$\frac{3}{2}$	$-\frac{1}{2}$	**UCG**	UGG	W	$(\frac{3}{2}$	$\frac{1}{2})^2$	$\frac{1}{2}$	$-\frac{1}{2}$	
CGA	R	$(\frac{1}{2}$	$\frac{1}{2})^2$	$\frac{1}{2}$	$-\frac{1}{2}$		UGA	W	$(\frac{1}{2}$	$\frac{1}{2})^2$	$-\frac{1}{2}$	$-\frac{1}{2}$	*UCA*
CAC	H	$(\frac{1}{2}$	$\frac{1}{2})^4$	$\frac{1}{2}$	$\frac{1}{2}$		UAC	Y	$(\frac{3}{2}$	$\frac{1}{2})^2$	$-\frac{1}{2}$	$\frac{1}{2}$	
CAU	H	$(\frac{1}{2}$	$\frac{1}{2})^4$	$-\frac{1}{2}$	$\frac{1}{2}$	*GUG*	UAU	Y	$(\frac{3}{2}$	$\frac{1}{2})^2$	$-\frac{3}{2}$	$\frac{1}{2}$	*GUA*
CAG	Q	$(\frac{1}{2}$	$\frac{1}{2})^4$	$\frac{1}{2}$	$-\frac{1}{2}$		UAG	**Ter**	$(\frac{3}{2}$	$\frac{1}{2})^2$	$-\frac{1}{2}$	$-\frac{1}{2}$	—
CAA	Q	$(\frac{1}{2}$	$\frac{1}{2})^4$	$-\frac{1}{2}$	$-\frac{1}{2}$	*UUG*	UAA	**Ter**	$(\frac{3}{2}$	$\frac{1}{2})^2$	$-\frac{3}{2}$	$-\frac{1}{2}$	—
GCC	A	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{1}{2}$		ACC	T	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	
GCU	A	$(\frac{1}{2}$	$\frac{3}{2})^1$	$\frac{1}{2}$	$\frac{1}{2}$		ACU	T	$(\frac{1}{2}$	$\frac{3}{2})^1$	$-\frac{1}{2}$	$\frac{1}{2}$	
GCG	A	$(\frac{3}{2}$	$\frac{1}{2})^1$	$\frac{3}{2}$	$-\frac{1}{2}$	**UGC**	ACG	T	$(\frac{3}{2}$	$\frac{1}{2})^1$	$\frac{1}{2}$	$-\frac{1}{2}$	**UGU**
GCA	A	$(\frac{1}{2}$	$\frac{1}{2})^1$	$\frac{1}{2}$	$-\frac{1}{2}$		ACA	T	$(\frac{1}{2}$	$\frac{1}{2})^1$	$-\frac{1}{2}$	$-\frac{1}{2}$	
GUC	V	$(\frac{1}{2}$	$\frac{3}{2})^2$	$\frac{1}{2}$	$\frac{1}{2}$		AUC	I	$\frac{3}{2}$	$\frac{3}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	
GUU	V	$(\frac{1}{2}$	$\frac{3}{2})^2$	$-\frac{1}{2}$	$\frac{1}{2}$		AUU	I	$\frac{3}{2}$	$\frac{3}{2}$	$-\frac{3}{2}$	$\frac{1}{2}$	*GAU*
GUG	V	$(\frac{1}{2}$	$\frac{1}{2})^3$	$\frac{1}{2}$	$-\frac{1}{2}$	**UAC**	AUG	M	$(\frac{3}{2}$	$\frac{1}{2})^1$	$-\frac{1}{2}$	$-\frac{1}{2}$	
GUA	V	$(\frac{1}{2}$	$\frac{1}{2})^3$	$-\frac{1}{2}$	$-\frac{1}{2}$		AUA	M	$(\frac{3}{2}$	$\frac{1}{2})^1$	$-\frac{3}{2}$	$-\frac{1}{2}$	*CAU*
GGC	G	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{3}{2}$	$-\frac{1}{2}$		AGC	S	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	
GGU	G	$(\frac{1}{2}$	$\frac{3}{2})^1$	$\frac{1}{2}$	$-\frac{1}{2}$		AGU	S	$(\frac{1}{2}$	$\frac{3}{2})^1$	$-\frac{1}{2}$	$-\frac{1}{2}$	*GCU*
GGG	G	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{3}{2}$	$-\frac{3}{2}$	**UCC**	AGG	**Ter**	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{1}{2}$	$-\frac{3}{2}$	—
GGA	G	$(\frac{1}{2}$	$\frac{3}{2})^1$	$\frac{1}{2}$	$-\frac{3}{2}$		AGA	**Ter**	$(\frac{1}{2}$	$\frac{3}{2})^1$	$-\frac{1}{2}$	$-\frac{3}{2}$	—
GAC	D	$(\frac{1}{2}$	$\frac{3}{2})^2$	$\frac{1}{2}$	$-\frac{1}{2}$		AAC	N	$\frac{3}{2}$	$\frac{3}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	
GAU	D	$(\frac{1}{2}$	$\frac{3}{2})^2$	$-\frac{1}{2}$	$-\frac{1}{2}$	*GUC*	AAU	N	$\frac{3}{2}$	$\frac{3}{2}$	$-\frac{3}{2}$	$-\frac{1}{2}$	*GUU*
GAG	E	$(\frac{1}{2}$	$\frac{3}{2})^2$	$\frac{1}{2}$	$-\frac{3}{2}$		AAG	K	$\frac{3}{2}$	$\frac{3}{2}$	$-\frac{1}{2}$	$-\frac{3}{2}$	
GAA	E	$(\frac{1}{2}$	$\frac{3}{2})^2$	$-\frac{1}{2}$	$-\frac{3}{2}$	*UUC*	AAA	K	$\frac{3}{2}$	$\frac{3}{2}$	$-\frac{3}{2}$	$-\frac{3}{2}$	*UUU*

Table 2: The vertebral mitochondrial code. The upper label denotes different irreducible representations. We list the most used anticodons for mitochondria of animals, see (Sprinzl et al. , 1998). In bold-red (italic-blue) the anticodons reading quadruplets (resp. doublets).

THE MODEL:

4 bases : purines : (A, G) *and* bases
pyrimidines : $(C, T/U)$ complementarity

$\Rightarrow (\frac{1}{2}, \frac{1}{2})$ representation of $SU(2) \times SU(2)$.

$$
\begin{array}{ccc}
 & \overset{su(2)_H}{\longleftrightarrow} & \\
C \equiv (+,+) & & U \equiv (-,+) \\
su(2)_V \updownarrow & & \updownarrow su(2)_V \\
G \equiv (+,-) & \underset{su(2)_H}{\longleftrightarrow} & A \equiv (-,-)
\end{array}
$$

Analogy between quark (q) and baryon $(3q)$
and base (b) and codon $(3b)$.

But :

$|p> \sim |uud> + |udu> + |duu>$ (implicit spin structure)

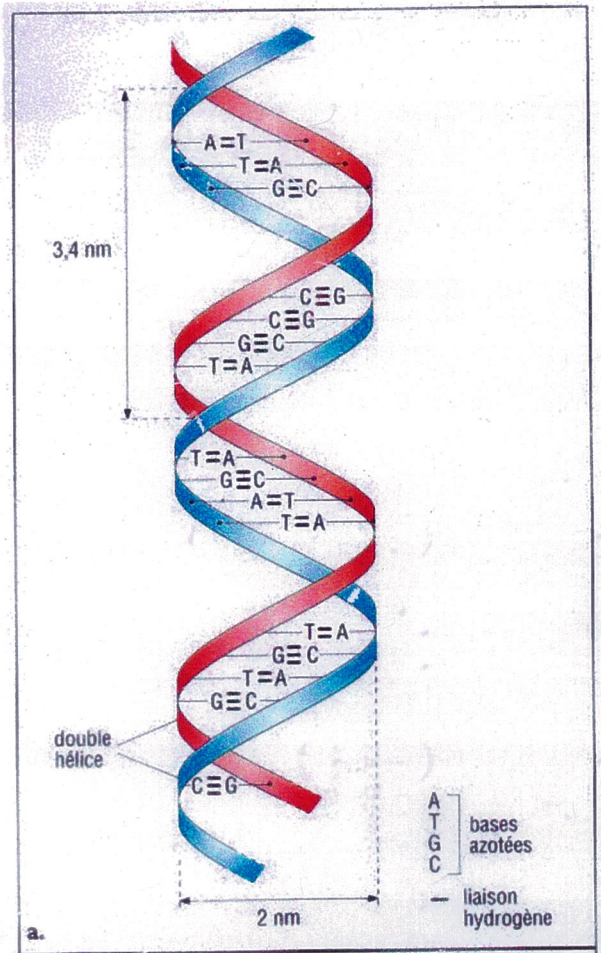
while:

$UAG \neq AUG$ in codons. *no mixing*

$\Longrightarrow$ Limit of the quantum (deformed) algebra
$\mathcal{U}_q[sl(2) \oplus sl(2)]$ when $q \to 0$

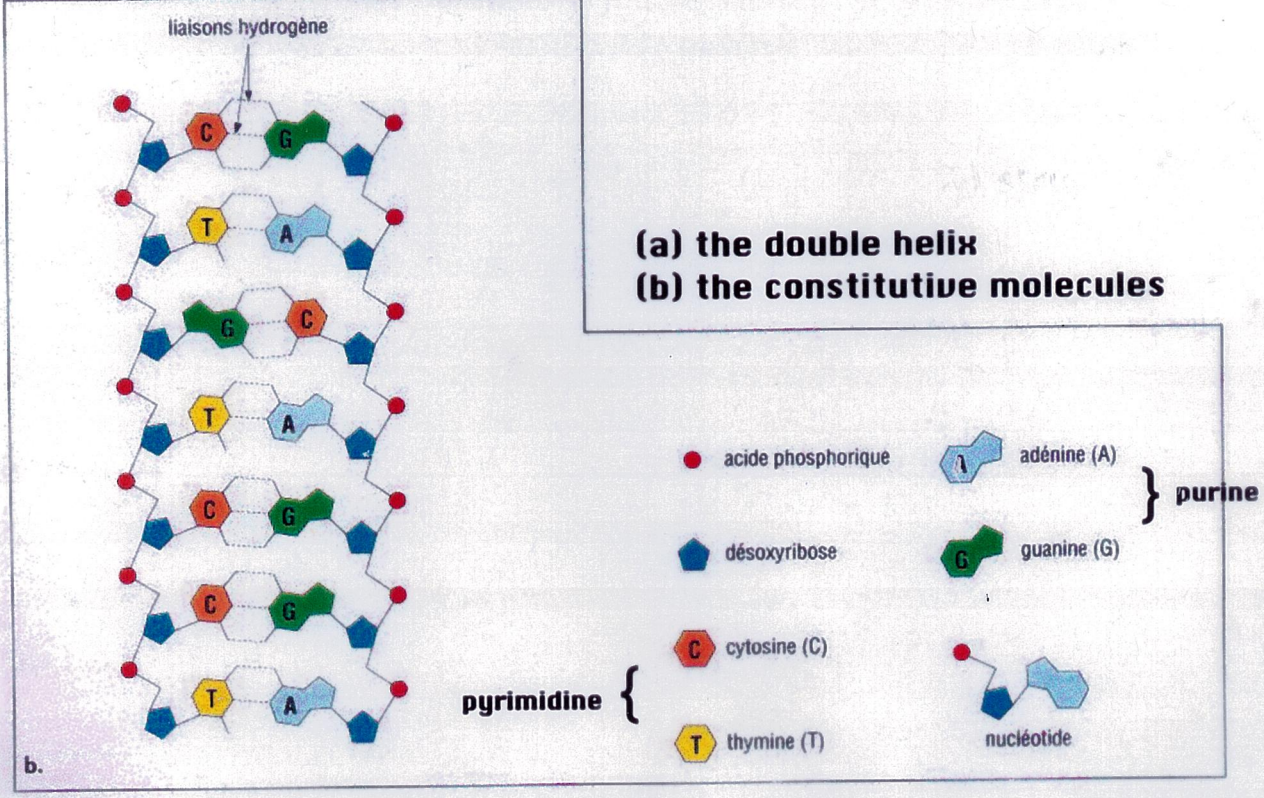
(remind: $q \to 1$ usual $\mathcal{U}[sl(2) \oplus sl(2)]$).
Then:

tensorial product of representations
=
"pure" states of constituent states
in **crystal bases**.



The DNA Structure

(DesoxyriboNucleic Acid)



(a) the double helix
(b) the constitutive molecules

DI-NUCLEOTIDS TABLE

in Crystal basis Model

(XY) C U A G

(J_H, J_V) :

$(0,0) = (CA)$

$(1,0) = (CG \text{ Arg} \quad UG \quad UA)$

$(0,1) = \begin{pmatrix} CU \text{ Leu} \\ GU \text{ Val} \\ GA \end{pmatrix}$

$(1,1) = \begin{pmatrix} CC \text{ Pro} & UC \text{ Ser} & UU \\ GC \text{ Ala} & AC \text{ Thr} & AU \\ GG \text{ Gly} & AG & AA \end{pmatrix}$

quadruplets $Q > 0$

doublets $Q < 0$

$Q = J_{3H} + \frac{1}{4} C_V (J_{3V} + 1) - \frac{1}{4}$

In our model: a codon $\in (\frac{1}{2},\frac{1}{2}) \otimes (\frac{1}{2},\frac{1}{2}) \otimes (\frac{1}{2},\frac{1}{2})$.

product of two representations

$$(\tfrac{1}{2},\tfrac{1}{2}) \otimes (\tfrac{1}{2},\tfrac{1}{2}) = (1,1) \oplus (1,0) \oplus (0,1) \oplus (0,0)$$

$\rightarrow$ $su(2)_H$ $(J_H, J_V) = (0,0)$ (CA) $(1,0)$ $(CG \;\; UG \;\; UA)$

$\downarrow$

$su(2)_V$ $(0,1)$ $\begin{pmatrix} CU \\ GU \\ GA \end{pmatrix}$ $(1,1)$ $\begin{pmatrix} CC & UC & UU \\ GC & AC & AU \\ GG & AG & AA \end{pmatrix}$

Property:

$Q = J_{3H} + \frac{1}{4} C_V (J_{3V} + 1) - \frac{1}{4}$

quadruplets $Q > 0$
(as well as those in sextets) s.t. $J_{3,H} > 0$
or $J_{3,H} = 0$ and $J_{3,V} \geq 0$, $J_V \neq 0$

doublets $Q < 0$
(and others: triplet, singlets) s.t. $J_{3,H} < 0$
or $J_{3,H} = 0$ and $J_{3,V} < 0$ or $J_V = 0$

product of three representations = codons

$$(\tfrac{1}{2},\tfrac{1}{2}) \otimes (\tfrac{1}{2},\tfrac{1}{2}) \otimes (\tfrac{1}{2},\tfrac{1}{2}) = (\tfrac{3}{2},\tfrac{3}{2}) \oplus 2(\tfrac{3}{2},\tfrac{1}{2}) \oplus 2(\tfrac{1}{2},\tfrac{3}{2}) \oplus 4(\tfrac{1}{2},\tfrac{1}{2})$$

$$(\tfrac{3}{2},\tfrac{3}{2}) \equiv \begin{pmatrix} CCC & UCC & UUC & UUU \\ GCC & ACC & AUC & AUU \\ GGC & AGC & AAC & AAU \\ GGG & AGG & AAG & AAA \end{pmatrix}$$

$$(\tfrac{3}{2},\tfrac{1}{2}) \equiv \begin{pmatrix} CCG & UCG & UUG & UUA \\ GCG & ACG & AUG & AUA \end{pmatrix}$$

$$(\tfrac{3}{2},\tfrac{1}{2})' \equiv \begin{pmatrix} CGC & UGC & UAC & UAU \\ CGG & UGG & UAG & UAA \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{3}{2}) \equiv \begin{pmatrix} CCU & UCU \\ GCU & ACU \\ GGU & AGU \\ GGA & AGA \end{pmatrix} \qquad (\tfrac{1}{2},\tfrac{3}{2})' \equiv \begin{pmatrix} CUC & CUU \\ GUC & GUU \\ GAC & GAU \\ GAG & GAA \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2}) \equiv \begin{pmatrix} CCA & UCA \\ GCA & ACA \end{pmatrix} \qquad (\tfrac{1}{2},\tfrac{1}{2})' \equiv \begin{pmatrix} CGU & UGU \\ CGA & UGA \end{pmatrix}$$

$$(\tfrac{1}{2},\tfrac{1}{2})'' \equiv \begin{pmatrix} CUG & CUA \\ GUG & GUA \end{pmatrix} \qquad (\tfrac{1}{2},\tfrac{1}{2})''' \equiv \begin{pmatrix} CAC & CAU \\ CAG & CAA \end{pmatrix}$$

Applications of this model have been provided in a series of papers in collaboration with *L.Frappat and A.Sciarrino* in the years 1998-2005.

Among them:

- study of codon usage probabilities for vertebrates, invertebrates and plants.

- elaboration and verification of sum rules for codon usage probabilities in the case of vertebrates.

- relations between physico-chemical properties of amino-acids and predictions for not yet measured quantities.

- investigations on aspects of mRNA editing.

- more "mathematical" aspects : construction of an operator giving the correspondence between amino-acids and codons for any known genetic code; attempt to describe mutations, etc.

The Minimum Principle

Consider the operator:

$$T\,(\text{anticodon, codon}) = c_H \vec{J}^{c}_{H} \cdot \vec{J}^{a}_{H} + c_V \vec{J}^{c}_{V} \cdot \vec{J}^{a}_{V}$$

where : $\vec{J}^{c} \cdot \vec{J}^{a} = \frac{1}{2} \{ (\vec{J}^{c} + \vec{J}^{a})^2 - (\vec{J}^{c})^2 - (\vec{J}^{a})^2 \}$

and: $\vec{J} = (J_1, J_2, J_3)$ generators of $su(2)_{\substack{H \\ V}}$ group.

Then define:

- for quadruplets: taking as an example Val (GUN; N= C,U,G,A) and as a possible anticodon CAC:

$$T\text{aver.}\,(\text{CAC, Val}) = \sum_N P^{val}_N \cdot T\,(\text{CAC, GUN})$$

with : $P^{q}_{C} + P^{q}_{U} + P^{q}_{G} + P^{q}_{A} = 1$

- for doublets : taking as an ex. Asp (GAC, GAU) and as a possible anticodon CUC:

$$T\text{aver.}\,(\text{CUC, Asp}) = \sum_Y P^{Asp}_Y \cdot T\,(\text{CUC, GUY})$$

with : $P^{d}_{Y} = P^{d}_{C} + P^{d}_{U} = 1$ (and $P^{d}_{R} = P^{d}_{G} + P^{d}_{A} = 1$)

Question:
Can we determine c_H and c_V such that for each given quadruplet (or doublet) of codons, the anticodon minimizing Taver. is the one given by the data ?

Let us remind that the possible anticodons to the codon XY N is N'Y'X' with X -- X' and Y --Y' related by the "usual pairing" (i.e. C -- G, U -- A) and N' is any nucleotide C,G,U, A).

Results:

- for quadruplets: choose simply $c_H > 0$ and $c_V < 0$ to be in accordance with data.

-for doublets: choose $c_V > 0$
and the sign of c_H such that:

c_H is $>$ for the doublets : UUY, UAY, AUY, AAY
CAR, UGR, AGR, GAR

c_H is $<$ for the other doublets: UUR, UAR, AUR, AAR
CAY, UGY, AGY, GAY

(c_H of opposite sign for two doublets with same dinucleotide but ending with a purine or a pyrimidine).

For doublets, we remark that, with the choice of sign of c_H above specified and $c_V > 0$ for all a.a., the anticodons minimizing the average value of $\mathcal{T}$ are in agreement with the observed anticodon, see (Sprinzl et al. , 1998) and Table 2. We summarize in Table 1 the results for the doublets.

a.a	sign c_H	anticodon	note
His	-	GUG	$P_C^d > 0,25$
Gln	+	UUG	$P_G^d > 0,25$
Phe	-	GAA	
Leu	+	UAA	
Cys	+	GCA	
Trp	-	UCA	
Tyr	-	GUA	
Ser	+	GCU	
Asp	+	GUC	$P_C^d > 0,25$
Glu	-	UUC	$P_G^d > 0,25$
Ile	+	GAU	
Met	-	CAU	
Asn	-	GUU	
Lys	+	UUU	

Table 1: Anticodon minimizing the operator $\mathcal{T}$, averaged over the two codons, for any amino acid encoded by a doublet, specifying the sign of c_H.

Let us remark that we find that for Met the anticodon is not UAU, as it should be expected from the empirical rule above quoted, but CAU which seems in agrement with the data, see (Sprinzl et al. , 1998).

4. Conclusions

We have found that the anticodons minimizing the conjectured operator $\mathcal{T}$ given in eq.(2), averaged over the concerned multiplets, are in very good agreement, the results depending only on the signs of the two coupling constants, with the observed ones, even if we have made comparison with a limited database.

The fact that the crystal basis model is able to explain, in a relatively simple way, the observed anticodon-codon pairing which has its roots on the stereochemical properties of nucleotides (Lim and Curran , 2001) strongly suggests that our modelisation is able to incorporate some crucial features of the complex physico-chemical structure of the genetic code. Incidentally let us remark that the model explains the symmetry codon anticodon remarked in (Wilhelm and Nikolajewa , 2004). Let us stress that our modelisation has a very peculiar feature which makes it very different from the standard 4-letter alphabet, used to identify the nucleotides, as well as with the usual modelisation of nucleotide chain as spin chain. Indeed the identification of the nucleotides with the fundamental

Conclusion:

- Anticodons minimizing the conjectured operator T aver. in very good agreement with the observed ones for mitochondria of animals
- Results depending only of the sign of two coupling constants.

One may expect a more complicated pattern in the general case, following the biological species. It might happen that the "universal" feature of c and c should be released, and that the expression of the T operator modified. The crystal basis model allows a lot of possibilities, for example by adjunction of a term of "spin-spin" interaction of the type:

$$g_H \overset{c}{J}_{H,3} \overset{a}{J}_{H,3} + g_V \overset{c}{J}_{V,3} \cdot \overset{a}{J}_{V,3}$$

Remark:

The above obtained result, joined to previous ones, encourages us to imagine a kind of double " BIO-SPIN ", leading to the description of any ordered sequence of n nucleotides as a state of an (irreducible) representation of $\mathcal{U}_q\big(su(2) + su(2)\big)$ and allowing to describe interactions using the standard powerful mathematical language already well adapted in physical spin models.