The sound of the DNA language

EILEEN RIEGO[1], ALEJANDRO SILVA[2] and JOSE DE LA FUENTE[1]

[1] Mammalian Cell Genetics Division, Center for Genetic Engineering and Biotechnology, Havana, Cuba.
[2] National Center for Biopreparatives, Bejucal, Havana, Cuba.

A new method is described for the study of DNA language employing the communicative strength of music. An algorithm was created to "translate" codons into musical notes. The analysis of the musical transcriptions of DNA sequences suggests the presence of some structural features of DNA language hidden in human and mouse interferon α1 genes.

Key terms: *DNA, information, language, structure.*

INTRODUCTION

During the evolution of life, DNA molecules of increasing complexity appeared. Evolutionary pressure, through selection, must have preserved from survival many of such DNA molecules, thus avoiding a simple random-base organization of the genome of surviving organisms. This non-random organization of the genome must have left a structural imprinting in the DNA language, preserving some structural features.

Some authors have emphasized the role that palindromes play in the organization of genomic nucleotide sequences and the variety of functions these sequences are known to be involved in (4). Several grammatical rules have been applied to the formulation of DNA language showing that there is at least some degree of organization in the nucleotide sequences found in many organisms (2-3).

Attempting to find some features in the pattern of nucleotide and amino acid sequences, we have created an algorithm to assign musical notes to all different codons. Employing this algorithm we have "translated" into music human and mouse interferon (IFN) α1 genes.

Through this simple method employing the communicative strength of music, we have found that there are some syntactical structures in the DNA language and have revealed some properties of these structures. The method could be employed for the study of the DNA language.

METHODS

An algorithm was created to assign musical notes to all different codons.

Codons were denoted as $X_nY_nZ_n$ and musical notes as N_n.

Every aminoacid

$$(aa)_{n1n2}=X_\beta Y_\beta Z_\beta=N_{n2}N_{n1}=N_{n2}Y_\beta Z_\beta Z_\tau$$

for every $1\leq n1\leq 6$; $1\leq n2\leq 20$; $1\leq \alpha, \beta, \tau\leq 4$

If by definition, aminoacid

$(aa)_{1n2}=X_\alpha Y_\alpha Z_\alpha=N_{n2}$, then

if $n1=1$, $N_{n2}N_{n1}=N_{n2}$

The assignment of the musical note for the first codon for every aminoacid (N_1, N_2, ..., N_{20}) was arbitrary and corresponded to the order given in Table I.

* **Correspondence to:** Eileen Riego, División de Genética de Células de Mamíferos, Centro de Ingeniería Genética y Biotecnología, Casilla Postal 6162, Havana 6, Cuba. Fax: (53-7) 218-070 or 336008

TABLE I

Assignment of musical notes to different aminoacids

Aminoacid	Code	Codons*	Musical notes**
Ala	A	GCA (aa)11	DR (N1)
		GCC (aa)21	DRMR
		GCG (aa)31	DRR
		GCU (aa)41	DRL
Arg	R	CGA (aa)12	R (N2)
		CGC (aa)22	RDR
		CGG (aa)32	RSL
		CGU (aa)42	RTL
		AGA (aa)52	RFS
		AGG (aa)62 RSL	
Asn	N AAC RM		
		AAU	RMTD
Asp	D	GAC D	
		GAU	DTD
Cys	C	UGC	T
		UGU	TTL
Gln	Q	CAA	MF
		CAG	MFR
Glu	E	GAA (aa)17	FS (N7)
		GAG (aa)27	FSR
Gly	G	GGA	SL
		GGC	SLDR
		GGG	SLSL
		GGU	SLTL
His	H	CAC	LT
		CAU	LTM
Ile	I	AUA	TD
		AUC	TDS
		AUU	TDF
Leu	L	CUA	L
		CUC	LS
		CUG	LSF
		CUU	LF
		UUA	LLS
		UUG	LSF
Lys	K	AAA	RD
		AAG	RDR
Met	M	AUG	M
Phe	F	UUC	F
		UUU	FL
Pro	P	CCA	MR
		CCC	MRMR
		CCG	MRR
		CCU	MRL
Ser	S	UCA	S
		UCC	SMR
		UCG	SR
		UCU	SL
		AGC	SDR
		AGU	STL

TABLE I (Continuation)

Aminoacid	Code	Codons*	Musical notes**
Thr	T	ACA	FM
		ACC	FMMR
		ACG	FMR
		ACU	FML
Trp	W	UGG	SF
Tyr	Y	UAC	LS
		UAU	LSTD
Val	V	GUA	TL
		GUC	TLS
		GUG	TLSF
		GUU	TLF

Abbreviations: **D** = Do, **R** = Ray, **M** = Me, **F** = Fah, **S** = Solh, **L** = Lah, **T** = Te.
* For some aminoacid codons, the mathematical notation is shown as an example.
** The mathematical notation for the musical note corresponding to the first codon for some aminoacids is shown in parenthesis as an example.

Employing the data on Table I and the proposed algorithm, as an example, let us find the musical notes that correspond to the codons CGC and GAG coding for aminoacids Arg and Glu respectively:

$(aa)_{22} = CGC = N_2GCZ_\tau = R(aa)_{11} = RN_1 = RDR$

$(aa)_{27} = GAG = N_7AGZ_\tau = FS(aa)_{12} = FSN_2 = FSR$

The sequences employed in this report corresponded to the human interferon (HuIFN) -α1 (6) and mouse interferon (MuIFN) -α1 (7) genes.

The random DNA sequence was generated by making the following substitutions in the musical translation of the MuIFN-α1 gene: MexDo; RayxSoh; LahxFah; DoSohxMe; TeTexRay; FahDoxLah.

Musical scores were prepared in a rhythm of waltz.

RESULTS AND DISCUSSION

The problem of studying and understanding the DNA language has focused the attention of many researchers.

Languages of the world belong to two major groups, Indo-Aryan and Ural-Altaic. Yet, the principle of having to have syntactical structures remains the same. Music, as a language, is also governed by syntactical structures. DNA and protein aminoacid sequences are also for communication. Therefore, if considering communication as an evolutionary process, it will not be a surprise if human languages, including music, learned the art of communication from these ancient informational molecules.

In an attempt to try to find some features in the pattern of nucleotide and aminoacid sequences, we have created an algorithm to assign musical notes to all different codons. Employing this algorithm we have "translated" into music human and mouse IFN-α1 genes (Figs 1 and 2). Music, as a highly expressive way of communication, brings the possibility to detect structures otherwise hidden for the human eye.

As a control for randomly originated structures, we created a random DNA sequence by randomly making some arbitrary substitutions in the sequence of the MuIFN-α1 gene (Fig 3). The sequence obtained was compared to the EMBL sequence data bank (release May 20, 1993) employing the program BLASTN 1.2.5 (1) and no significant homologies were found, thus indicating that the sequence obtained was not similar to any known DNA sequence.

HuIFN-alpha1

ATG	GCC	TCG	CCC	TTT	GCT	TTA	CTG	ATG	GTC	CTG	GTG	GTG	CTC	AGC
M	A	S	P	F	A	L	L	M	V	L	V	V	L	S
M	DRMR	SR	MRMR	FL	DRL	LLS	LSF	M	TLS	LSF	TLSF	TLSF	LS	SDR
TGC	AAG	TCA	AGC	TGC	TCT	CTG	GGC	TGT	GAT	CTC	CCT	GAG	ACC	CAC
C	K	S	S	C	S	L	G	C	D	L	P	E	T	H
T	RDR	S	SDR	T	SL	LSF	SLDR	TTL	DTD	LS	MLR	FSR	FMMR	LT
AGC	CTG	GAT	AAC	AGG	AGG	ACC	TTG	ATG	CTC	CTG	GCA	CAA	ATG	AGC
S	L	D	N	R	R	T	L	M	L	L	A	Q	M	S
SDR	LSF	DTD	RM	RLS	RLS	FMMR	LSF	M	LS	LSF	DR	MF	M	SDR
AGA	ATC	TCT	CCT	TCC	TCC	TGT	CTG	ATG	GAC	AGA	CAT	GAC	TTT	GGA
R	I	S	P	S	S	T	L	M	D	R	H	D	F	G
RSF	TDS	SL	MRL	SMR	SMR	TTL	LSF	M	D	RSF	LTM	D	FL	SLFL
TTT	CCC	CAG	GAG	GAG	TTT	GAT	GGC	AAC	CAG	TTC	CAG	AAG	GCT	CCA
F	P	Q	E	E	F	D	G	N	Q	F	Q	K	A	P
FL	MRMR	MFR	FSR	FSR	FL	DTD	SLDR	RM	MFR	F	MFR	RDR	DRL	MR
GCC	ATC	TCT	GTC	CTC	CAT	GAG	CTG	ATC	CAG	CAG	ATC	TTC	AAC	AAA
A	I	S	V	L	H	E	L	I	Q	Q	I	F	N	K
DRMR	TDS	SL	TLS	LS	LTM	FSR	LSF	TDS	MFR	MFR	TDS	F	RM	RD
GAT	TCA	TCT	GCT	GCT	TGG	GAT	GAG	GAC	CTC	CTC	TTT	ACC	ACA	CTA
D	S	S	A	A	W	D	E	D	L	L	F	T	T	H
DTD	S	SL	DRL	DRL	SF	DTD	FSR	D	LS	LS	FL	FMMR	FM	LTM
GAC	AAA	TTC	TGC	ACC	GAA	CTC	TAC	CAG	CAG	CTG	AAT	GAC	TTG	GAA
D	K	F	C	T	E	L	Y	Q	Q	L	N	D	L	E
D	RD	F	T	FMMR	FS	LS	LS	MFR	MFR	LS	RMTD	D	LSF	FS
GCC	TGT	GTG	ATG	CAG	GAG	GAG	AGG	GTG	GGA	GAA	ACT	CCC	CTG	ATG
G	C	V	M	Q	E	E	R	V	G	E	T	P	L	M
SLDR	TTL	TLSF	M	MFR	FSR	FSR	RSL	TLSF	SL	FS	FML	MRMR	LS	M
AAT	GCG	GAC	TCC	ATC	TTG	GCT	GTG	AAG	AAA	TAC	TTC	CGA	AGA	ATC
N	A	D	S	I	L	A	V	K	K	Y	F	Q	R	I
RMTD	DRR	D	SMR	TDS	LSF	DRL	TLSF	RDR	RD	LS	F	MFR	RFS	TDS
ACT	CTC	TAT	CTG	ACA	GAG	AAG	AAA	TAC	AGC	CCT	TGT	GCC	TGG	GAG
T	L	Y	L	T	E	K	K	Y	S	P	C	A	W	E
FML	LS	LSTD	LS	FMF	SR	RDR	RD	LS	SDR	MRL	TTL	DRMR	SF	FSR
GTT	GTC	AGA	GCA	GAA	ATC	ATG	AGA	TCC	CTC	TCT	TTA	TCA	ACA	AAC
V	V	R	A	E	I	M	R	S	L	S	L	S	T	N
TLF	TLS	RFS	DR	FS	TDS	M	RSF	SMR	LS	SL	LLS	S	FM	RM
TTG	CAA	GAA	AGA	TTA	AGG	AGG	AAG	GAA						
L	Q	E	R	L	R	R	K	E						
LSF	MF	FS	RSF	LLS	RSL	RSL	RDR	FS						

Fig 1. Nucleotide and aminoacid sequences of HuIFN-α1, with corresponding musical notes for each codon. Abbreviations described in Table I.

After the application of the algorithm for translating into music the DNA sequences, resulting arrays of musical notes were organized in scores (Figs 4-6). Melodies that corresponded to the sequences of Hu and Mu IFN-α1 genes resembled a Re-dorical mode and revealed the existence of some syntactical structures in the DNA language, such as palindromic and repetitive segments, that carried numerous overlapping patterns, used for communicating to other molecules and structures the details of specific interactions and entire processes. For the randomly generated sequence, no structured melody could be found.

The meaning of these messages is, in most of the cases, unknown. In fact, the method employed by us is not intended for decod-

MuIFN-alpha1

ATG	GCT	AGG	CTC	TGT	GCT	TTC	CTG	ATG	GTC	CTG	GCG	GTG	ATG	AGC
M	A	R	L	C	A	F	L	M	V	L	A	V	M	S
M	DRL	RLS	LS	TTL	DRL	F	LSF	M	TLS	LSF	DRR	TLSF	M	SDR
TAC	TGG	CCA	ACC	TGC	TCT	CTA	GGA	TGT	GAC	CTG	CCT	CAG	ACT	CAT
Y	W	P	T	C	S	L	G	C	D	L	P	Q	T	H
LS	SF	MR	FMMR	T	SL	L	SL	TTL	D	LSF	MRL	MFR	FML	LTM
AAC	CTC	AGG	AAC	AAG	AGA	GCC	TTG	ACA	CTC	CTG	GTA	CAA	ATG	AGG
N	L	R	N	K	R	A	L	T	L	L	V	Q	M	R
RM	LS	RSL	RM	RDR	RFS	DRMR	LSF	FM	LS	LSF	TL	MF	M	RSL
AGA	CTC	TCC	CCT	CTC	TCC	TGC	CTG	AAG	GAC	AGG	AAG	GAC	TTT	GGA
R	L	S	P	L	S	C	L	K	D	R	K	D	F	G
RFS	LS	SMR	MRL	LS	SMR	T	LSF	RDR	D	RSL	RDR	D	FL	SL
TTC	CCG	CAG	GAG	AAG	GTG	GAT	GCC	CAG	CAG	ATC	AAG	AAG	GCT	CAA
F	P	Q	E	K	V	D	A	Q	Q	I	K	K	A	Q
F	MRR	MFR	FSR	RDR	TLSF	DTD	DRMR	MFR	MFR	TDS	RDR	RDR	DRL	MF
GCC	ATC	CCT	GTC	CTG	AGT	GAG	CTG	ACC	CAG	CAG	ATC	CTG	AAC	ATC
A	I	P	V	L	S	E	L	T	Q	Q	I	L	N	I
DRMR	TDS	MRL	TLS	LSF	STL	FSR	LSF	FMMR	MFR	MFR	TDS	LSF	RM	TDS
TTC	ACA	TCA	AAG	GAC	TCA	TCT	GCT	GCT	TGG	AAT	GCA	ACC	CTC	CTA
F	T	S	K	D	S	S	A	A	W	N	A	T	L	L
F	FM	S	RDR	D	S	SL	DRL	DRL	SF	RMTD	DR	FMMR	LS	L
GAC	TCA	TTC	TGC	AAT	GAC	CTC	CAC	CAG	CAG	CTC	AAT	GAC	CTG	CAA
D	S	F	C	N	D	L	H	Q	Q	L	N	D	L	Q
D	S	F	T	RMTD	D	LS	LT	MFR	MFR	LS	RMTD	D	LSF	MF
GGT	TGT	CTG	ATG	CAG	CAG	GTG	GGG	GTG	CAG	GAA	TTT	CCC	CTG	ACC
G	C	L	M	Q	Q	V	G	V	Q	E	F	P	L	T
SLTL	TTL	LSF	M	MFR	MFR	TLSF	SLSL	TLSF	MFR	FS	FL	MRMR	LSF	FMMR
CAG	GAA	GAT	GCC	CTG	CTG	GCT	GTG	AGG	AAA	TAC	TTC	CAC	AGG	ATC
Q	E	D	A	L	L	A	V	R	K	Y	F	H	R	I
MFR	FS	DTD	DRMR	LSF	LSF	DRL	TLSF	RSL	RD	LS	RD	LT	RSL	TDS
ACT	GTG	TAC	CTG	AGA	GAG	AAG	AAA	CAC	AGC	CCC	TGT	GCC	TGG	GAG
T	V	Y	L	R	E	K	K	H	S	P	C	A	W	E
FML	TLSF	LS	LSF	RFS	FSR	RDR	RD	LT	SDR	MRMR	TTL	DRMR	SF	FSR
GTG	GTC	AGA	GCA	GAA	GTC	TGG	AGA	GCC	CTG	TCT	TCC	TCT	GCC	AAT
V	V	R	A	E	V	W	R	A	L	S	S	S	A	N
TLSF	TLS	RFS	DR	FS	TLS	SF	RFS	DRMR	LSF	SL	SMR	SL	DRMR	RMTD
GTG	CTG	GGA	AGA	CTG	AGA	GAA	GAG	AAA						
V	L	G	R	L	R	E	E	K						
TLSF	LSF	SL	RFS	LSF	RFS	FS	FSR	RD						

Fig 2. Nucleotide and aminoacid sequences of MuIFN-α1, with corresponding musical notes for each codon. Abbreviations described in Table I.

ifying such messages. As U Eco mentioned in his "Treatise of General Semiotic" (5), music appears as a semiotic system in which any expressive situation can be interpreted in different ways. Through the method described here, we can only search for syntactical structures and some features of such structures by analyzing the scores and the melodies obtained from these scores, but we can not help in trying to understand them.

The method can be employed also to generate musical scores from the available sequence data bank. These scores can be employed to study gene families and the evolution of conserved genes by searching for homology in the respective melodies'

Random sequence

```
GAC CAA  TCA GAA  GAA  CGA TTC  CAA TTC  TTC TGG  GAC  TGC GAA   TTC TGG
 D   Q    S   E    E    R   F    Q   F    F   W    D    C   E     F   W
 D   MF    SFS     FS     RF     MF  F     FSF     D     TFS       FSF

ATG TCA   TGC TTC TGG  GAC  TCA ATG  GAA  TGG  ATG  CTA ATG  TGC TGG TTC
 M   S     C   F   W   D    S   M    E    W    M    L   M    C   W   F
  MS         TFSF      D     SM      FS   SF   M     LM      T   SF  F

TCA TTC   CGA TTC  GAC  TTC TGG  CAA  GAC GAA  CTT  TTC ATA   TCA GAC  GAA
 S   F     R   F    D    F   W   Q    D   E    L    F   I    S   D    E
  SF       RF       D     FSF     MF    DFS     LF    FTD      SD      FS

TCA TGG   TCA GAC   TCA ATG   TCA GAA   ATG ATG   TTC TGG  CTA GAA  TTC TGG
 S   W     S   D     S   M     S   E     M   M     F   W    L   E    F   W
  SSF       SD        SM        SFS       MM       FSF     L   FS    FSF

TGC TTC   GAC TTC  GAC  TCA TGG   TCA GAA  GAA  TCA ATG  CAA GAA  TCA ATG
 C   F     D   F    D    S   W    S   E    E    S   M    Q   E    S   M
  TF        DF      D     SSF      SFS     FS    SM      MF  FS    SM

TGC  TTC TGG   TCA ATG  GAC  TCA TGG   TCA ATG  GAC TTC TTC  TGG TTC  ATG TCA
 C    F   W     S   M    D    S   W    S   M    D   F   F   W   F    M   S
 T     FSF       SM      D     SSF      SM      D    FF     SF  F     MS

GAC GAA   GAA TCA   TCA ATG   TGC TTC TGG  GAT  ATG ATG   GAC GAA   GAC GAA
 D   E     E   S     S   M     C   F   W   D    M   M     D   E     D   E
  DFS       FSS       SM        TFSF      DTD   MM       DFS       DFS

TGC ATG   TCA ATG   TCA ATG  CAA  GAC TTC   ATG ATG   TGC ATG  CAA  TGC GAA
 C   M     S   M     S   M    Q    D   F    M   M     C   M    Q    C   E
  TM        SM        SM      MF    DF       MM       TM      MF    TFS

TTC TGG   TCA TGC TTC   GAA TCA   TTC TGG   CTA ATG   GAC GAA   GAC GAA
 F   W     S   C   F     E   S     F   W     L   M     D   E     D   E
  FSF       STF       FSS       FSF       LM        DFS       DFS

TGC ATG   TTC TGG   TCA GAC   TGC ATG  TTC CTA TCA  TCA ATG  GAC TCA TGG CAA
 C   M     F   W     S   D     C   M    F   L   S    S   M    D   S   W   Q
 TM        FSF       SD        TM       F   L   S    SM       D   S   SF  MF

ATG TTC TGG  ATC GAT  ATG  CTA ATG  GAA TTC GAC TCA TTC TGC  ATC GAT  GAC GAA
 M   F   W    I   D    M    L   M    E   F   D   S   F   C   I   D    D   E
  MF     SF     SDTD    M      LM      FS  F   D   S   F   T     SDTD    D   FS

TTC TGC  GAC GAA  GAC GAA GAA  ATC GAT  GAC  TTC TGG   GAC TTC   TGG TGC TTC
 F   C    D   E    D   E   E    I   D    D    F   W     D   F     W   C   F
  FT       DFS      DFS     FS    SDTD    D     FSF       DF        SFTF

CGA TTC  TTC TGG  GAC  GAC GAA  GAC GAA   TGC TTC TGG   TGG TGG  TGC TTC TGG
 R   F    F   W    D    D   E    D   E    C   F   W     W   W    C   F   W
  RF       FSF     D      DFS      DFS       TFSF        SFSF      TFSF

GAC GAA  GAA  TTC TTC  ATG ATG  TTC TGG   CTA ATG  GAC GAA  GAA GAT  ATG ATG
 D   E    E    F   F    M   M    F   W    L   M    D   E    E   D    M   M
  DFS      FS    FF      MM       FSF      LM      DFS      FS DTD    MM

TTC TGG  GAA TTC  CAA  TGC TTC TGG  TCA TGG   TCA GAC  GAA  TCA GAC  TTC TGC
 F   W    E   F    Q    C   F   W    S   W    S   D    E    S   D    F   C
  FSF      FSF     MF     TFSF        SSF      SD      FS    SD       FT

TCA TGG   TGC ATG  CTT  TGC TTC TGG  GAA  TTC TGG   TCA GAA   GAA TCA
 S   W     C   M    L    C   F   W   E    F   W    S   E     E   S
  SSF       TM      LF     TFSF      FS     FSF      SFS       FSS

TCA ATG   TCA GAC   TTC TGC   TCA ATG   ATG ATG   CGA TTC   ATG ATG  TGG
 S   M     S   D     F   C     S   M     M   M     R   F     M   M    W
  SM        SD        FT        SM        MM        RF        MM      SF

GAA TCA   TGC TTC TGG   TGC GAA   TCA GAA  ATG GAA  TGC GAA  TGG  TCA GAA
 E   S     C   F   W     C   E     S   E    M   E    C   E    W    S   E
  FSS       TFSF        TFS       SFS      M   FS     TFS     SF     SFS

ATG ATG   TTC TGG   TCA TTC   TCA ATG  TGG  ATG ATG   ATC GAT   TGC TTC TGG
 M   M     F   W     S   F     S   M    W    M   M     I   D     C   F   W
  MM        FSF       SF        SM       SF    MM       SDTD      TFSF

TTC TGG  TGG  TCA GAA   TTC TGG   TCA GAA  GAA  GAA TCA   TCA GAC
 F   W    W    S   E     F   W     S   E    E    E   S     S   D
  FSF      SF     SFS       FSF       SFS      FS    FSS      SD
```

Fig 3. Nucleotide and aminoacid sequences of randomly originated sequence, with corresponding musical notes for each codon. Abbreviations described in Table I.

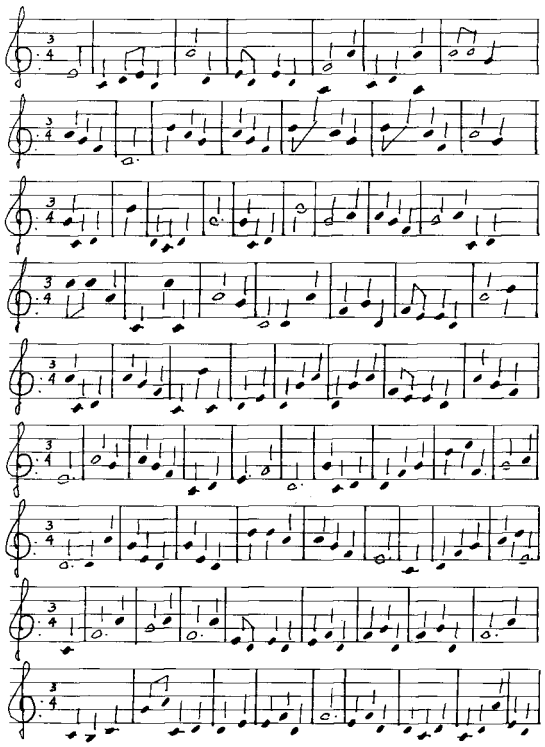

Fig 4. Scores derived from the arrays of musical notes described in Figure 1 for HuIFN-α1.

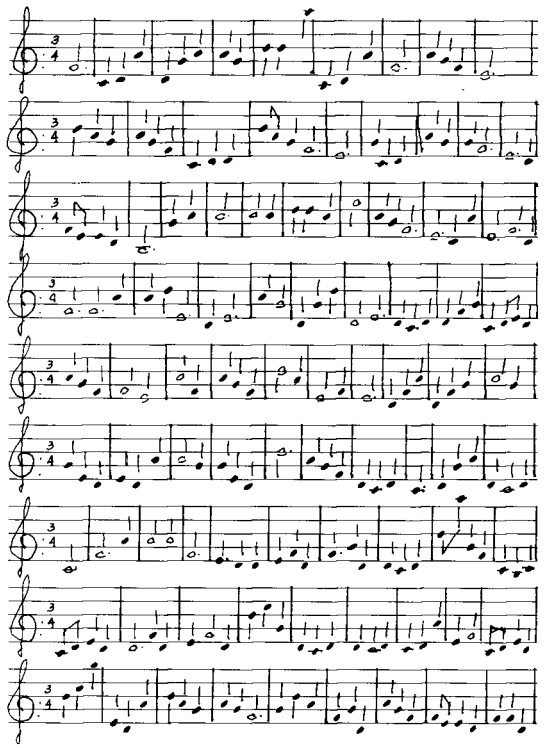

Fig 5. Scores derived from the arrays of musical notes described in Figure 2 for MuIFN-α1.

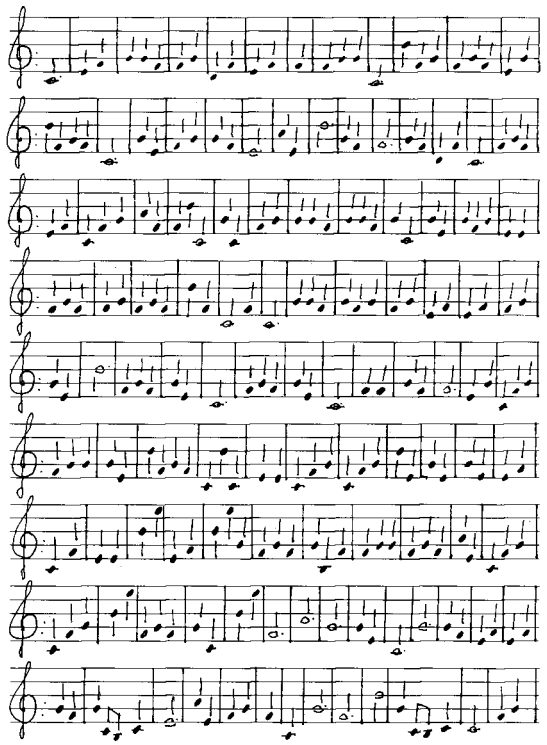

Fig 6. Scores derived from the arrays of musical notes described in Figure 3 for the randomly originated sequence.

modes and calculating the similarity between sequences by correlating their scores as a mean to detect both closely and distantly related sequences.

We have found that the genome appears organized as a score of necessary messages, although we do understand only few of them.

ACKNOWLEDGEMENTS

The authors will like to thank N García and S García for preparing musical scores. N García was also of invaluable help for playing and recording scores. Scores were kindly revised by Mrs C Rodríguez and Mr G Silva.

REFERENCES

1. ALTSCHUL S F, GISH W, MILLER W, MYERS EW, LIPMAN DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410
2. BECKMANN JS, TRIFONOV EN (1991) Splice junctions follow a 205-base lader. Proc Natl Acad Sci USA 88: 2380-2383
3. BOLSHOY A, McNAMARA P, HARRINGTON RE, TRIFONOV EN (1991) Curved DNA without A-A:

Experimental estimation of all 16 DNA wedge angles. Proc Natl Acad Sci USA 88: 2312-2316
4. DE ALMEIDA DF, VASCONCELOS AT, COIMBRA CA, CORREA G, PFEFFER AV, MARTINS LC (1992) Probing the extragenic DNA language. Braz J Genet 15 (suppl 1): 175-176
5. ECO U (1988) Tratado de Semiótica General. España: Editorial Lumen. pp 145-150
6. NAGATA S, MANTEI N, WEISSMANN C (1980) The structure of one of the eight or more distinct chromosomal genes for human interferon-α. Nature 287: 401-408
7. SHAW GD, BOLL W, TAIRA H, MANTEI N, LENGYEL P, WEISSMANN C (1983) Structure and expression of cloned murine IFN-α genes. Nucleic Acids Res 11: 555-573