

p-ADIC ULTRAMETRIC STRUCTURE OF THE GENETIC CODE

Branko Dragovich

Institute of Physics, Belgrade, Serbia

<http://ipb.ac.rs/~dragovich>

dragovich@ipb.ac.rs

Alexandra Dragovich

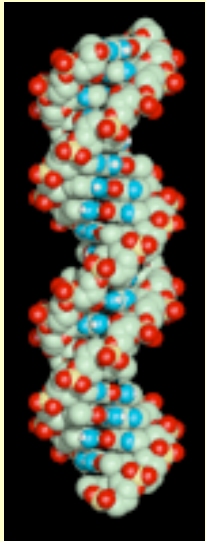
Vavilov Institute of General Genetics

Moscow, Russia

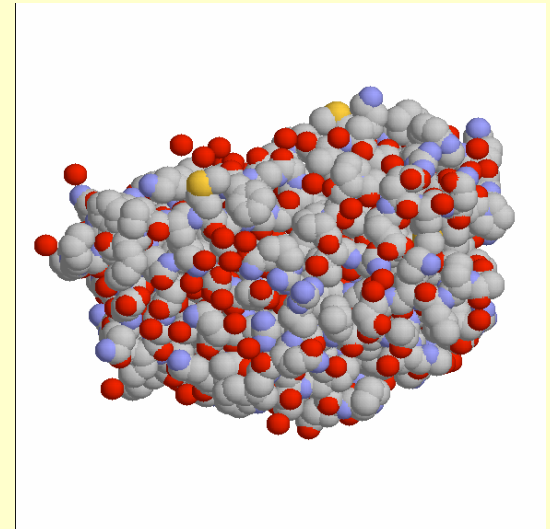
TAGp 2010

Annecy, 20 – 22.10, 2010

From DNA to proteins using the GENETIC CODE



		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

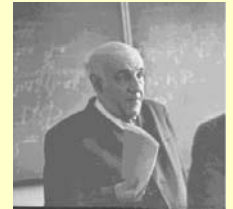


Motivations to Model the Genetic Code

- Genetic code is a map from 64 elements (codons) to the set of 21 elements (20 amino acids and 1 stop signal).
- There is a huge number of possibilities for genetic coding, but exists in living organisms only one genetic code and about 20 its slight modifications.
- What is the rule which established the Genetic Code?
- What is structure of the space of codons?
- What was origin of the genetic code? What was its evolution so far? What is its possible evolution in the future?

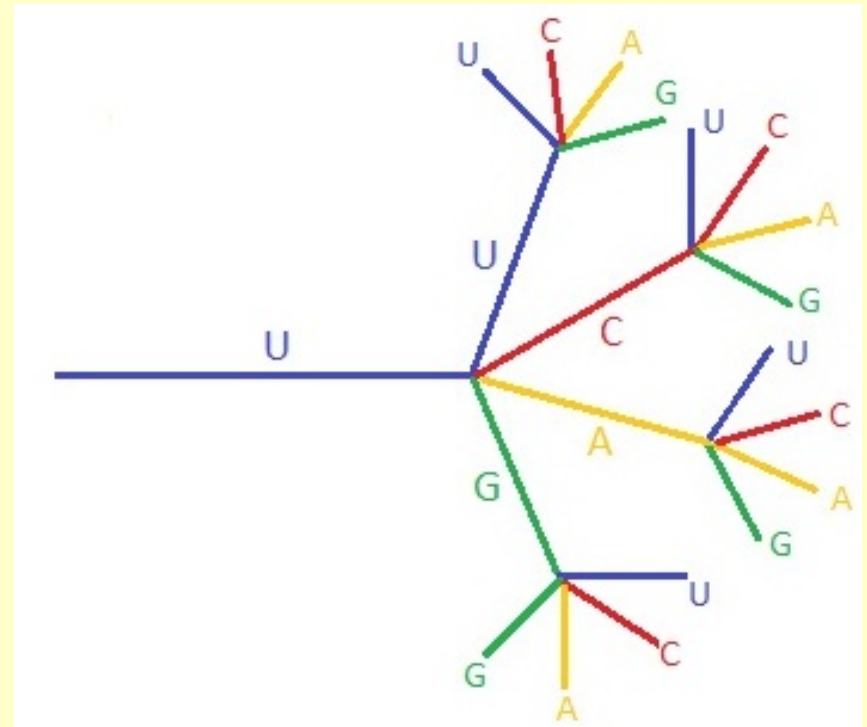
Modeling of the Genetic Code

- G. Gamow (1964-1968): 3-nucleotide codons, diamond code (1954)
- F. Crick (1916-2004): comma-free code (1957)
- Yu. Rumer (1901-1985): first 2 nucleotides emphasized (1966), ...
- Swanson (1984), Rakocevic, ...
- J. Hornos and Y. Hornos (1993), Forger and Sachse (2000)
- Frappat, Sciarrino and Sorba (1998)
- **p-adic approach:** B. Dragovich and A. Dragovich (2006), Khrennikov and Kozyrev (2007), Bradley (2007)

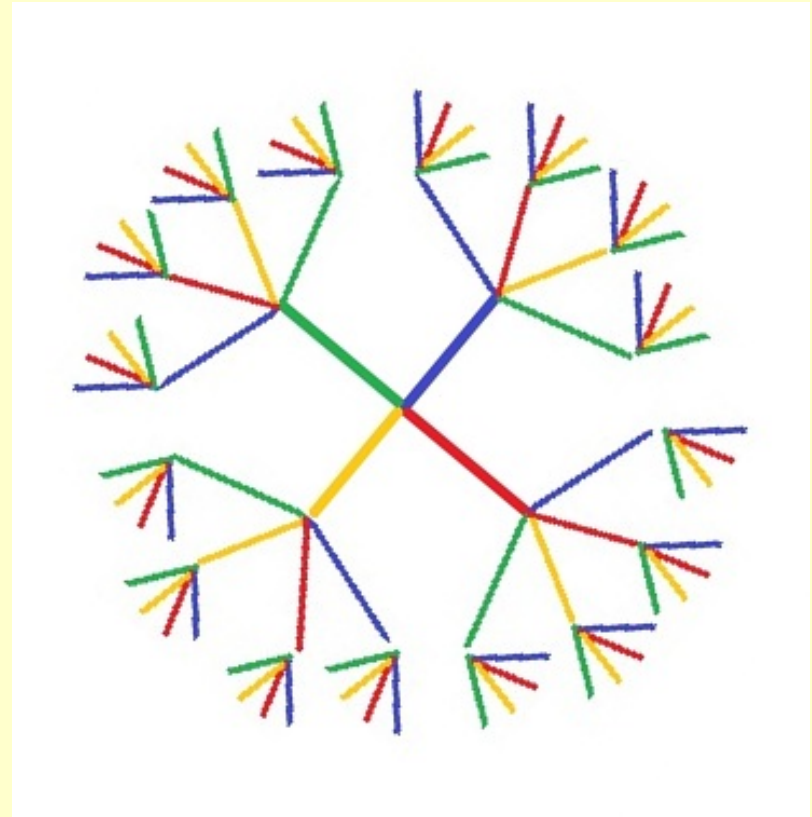
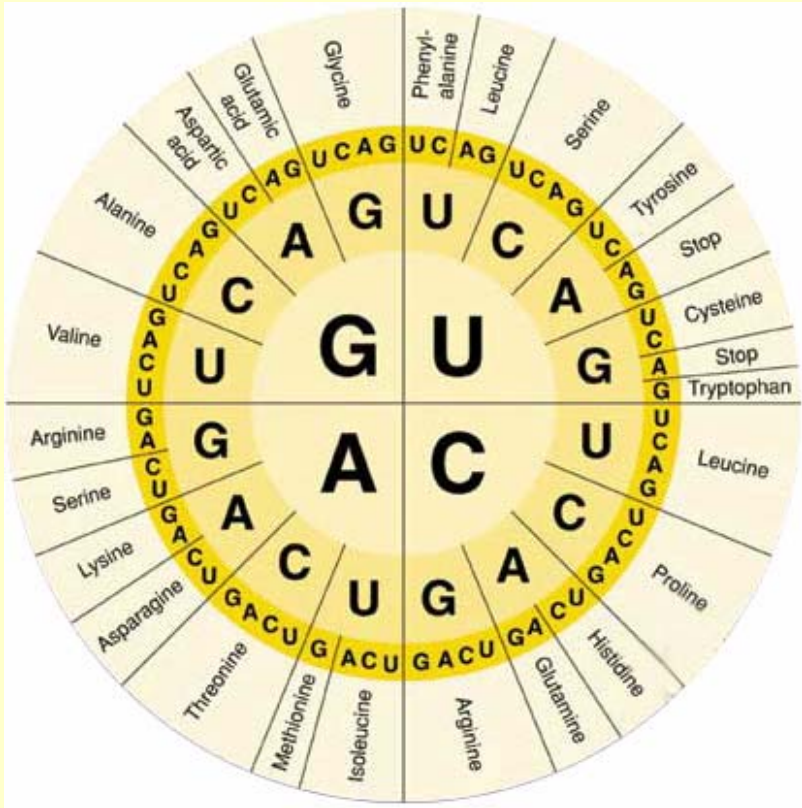


p-Adic approach: Ultrametric (Tree) Structure of Codons

		Second letter				
		U	C	A	G	
U	UUU } Phe	UCU } Ser	UAU } Tyr	UGU } Cys	U	
	UUC } Phe	UCC } Ser	UAC } Tyr	UGC } Cys	C	
	UUA } Leu	UCA } Ser	UAA Stop	UGA Stop	A	
	UUG } Leu	UCG } Ser	UAG Stop	UGG Trp	G	
C	CUU } Leu	CCU } Pro	CAU } His	CGU } Arg	U	
	CUC } Leu	CCC } Pro	CAC } His	CGC } Arg	C	
	CUA } Leu	CCA } Pro	CAA } Gln	CGA } Arg	A	
	CUG } Leu	CCG } Pro	CAG } Gln	CGG } Arg	G	
A	AUU } Ile	ACU } Thr	AAU } Asn	AGU } Ser	U	
	AUC } Ile	ACC } Thr	AAC } Asn	AGC } Ser	C	
	AUA } Met	ACA } Thr	AAA } Lys	AGA } Arg	A	
	AUG } Met	ACG } Thr	AAG } Lys	AGG } Arg	G	
G	GUU } Val	GCU } Ala	GAU } Asp	GGU } Gly	U	
	GUC } Val	GCC } Ala	GAC } Asp	GGC } Gly	C	
	GUA } Val	GCA } Ala	GAA } Glu	GGA } Gly	A	
	GUG } Val	GCG } Ala	GAG } Glu	GGG } Gly	G	



p-Adic approach: Ultrametric Tree of Codons



p-Adic Distance Between Integers

- Let a and b be two integers, and let $a-b = c = p^k m$, where p is a prime number, k is not negative integer, and m is not divisible by p . Then p -adic distance is:

$$d_p(a, b) = |a - b|_p = |c|_p = |p^k m|_p = p^{-k}$$

which is ultrametric, i.e.

$$d_p(x, y) \leq \max \{d_p(x, z), d_p(z, y)\} \leq d_p(x, z) + d_p(z, y)$$

p-Adic Space of Codons

5-adic natural numbers with three digits different from zero:

$$C[64] = \{n_0 + n_1 5 + n_2 5^2 : n_i = 1, 2, 3, 4\}$$

$$n_0 + n_1 5 + n_2 5^2 \equiv n_0 n_1 n_2$$

C (cytosine) = 1, A (adenine) = 2,

T (thymine) = U (uracil) = 3, G (guanine) = 4

(0 = absence of nucleotide)

5-adic distance between codons a and b

$$d_5(a, b) = |a_0 + a_1 5 + a_2 5^2 - b_0 - b_1 5 - b_2 5^2|_5$$

When codons a and b are different, there are 3 possibilities:

$$d_5(a, b) = 1 \iff a_0 \neq b_0$$

$$d_5(a, b) = \frac{1}{5} \iff a_0 = b_0, a_1 \neq b_1$$

$$d_5(a, b) = \frac{1}{25} \iff a_0 = b_0, a_1 = b_1, a_2 \neq b_2$$

With respect to smallest (1/25) 5-adic distance 64 codons clusterize to 16 quadruplets.

2-adic distance between 5-adic quadruplet codons

- Denote codons inside 5-adic quadruplet by **a, b, c, d** . Then 2-adic distance is:

$$d_2(a, c) = |(3 - 1)5^2|_2 = \frac{1}{2}$$

$$d_2(b, d) = |(4 - 2)5^2|_2 = \frac{1}{2}$$

Now every quadruplet decays into two 2-adic doublets. **These 32 doublets make basic structure of space of 64 codons.**

Vertebral Mitochondrial Code

111 CCC	Pro	211 ACC	Thr	311 UCC	Ser	411 GCC	Ala
112 CCA	Pro	212 ACA	Thr	312 UCA	Ser	412 GCA	Ala
113 CCU	Pro	213 ACU	Thr	313 UCU	Ser	413 GCU	Ala
114 CCG	Pro	214 ACG	Thr	314 UCG	Ser	414 GCG	Ala
121 CAC	His	221 AAC	Asn	321 UAC	Tyr	421 GAC	Asp
122 CAA	Gln	222 AAA	Lys	322 UAA	Ter	422 GAA	Glu
123 CAU	His	223 AAU	Asn	323 UAU	Tyr	423 GAU	Asp
124 CAG	Gln	224 AAG	Lys	324 UAG	Ter	424 GAG	Glu
131 CUC	Leu	231 AUC	Ile	331 UUC	Phe	431 GUC	Val
132 CUA	Leu	232 AUA	Met	332 UUA	Leu	432 GUA	Val
133 CUU	Leu	233 AUU	Ile	333 UUU	Phe	433 GUU	Val
134 CUG	Leu	234 AUG	Met	334 UUG	Leu	434 GUG	Val
141 CGC	Arg	241 AGC	Ser	341 UGC	Cys	441 GGC	Gly
142 CGA	Arg	242 AGA	Ter	342 UGA	Trp	442 GGA	Gly
143 CGU	Arg	243 AGU	Ser	343 UGU	Cys	443 GGU	Gly
144 CGG	Arg	244 AGG	Ter	344 UGG	Trp	444 GGG	Gly

p-Adic Properties of the Vertebral Mitochondrial Code

- **T-symmetry: doublets-doublets and quadruplets-quadruplets invariance**
- **5-Adic distance gives quadruplets**
- **2-Adic distance inside quadruplets gives doublets**
- **Degeneration of the genetic code has p-adic structure**
- **p-Adic degeneracy principle: *Codons code amino acids and stop signals by doublets which are result of combined 5-adic and 2-adic distances***
- **Modern assignment of codon doublets to particular amino acids is a result of coevolution of the genetic code and amino acids: single nucleotide code – 4 amino acids, dinucleotide code – 16 amino acids, trinucleotide code 20 amino acids.**
- **Other (19) codes may be regarded as slight modifications of the Vertebral Mitochondrial Code**

Examples of p-Adic Codon Spaces

$$\Gamma_p [(p-1)^m]$$

1-nucleotide codon space: $p=5$, $m=1$

2-nucleotide codon space: $p=5$, $m=2$

3-nucleotide codon space: $p=5$, $m=3$

Possible evolution of codon space:

$$\Gamma_5(4) \rightarrow \Gamma_5(4^2) \rightarrow \Gamma_5(4^3) = \Gamma_5(64)$$

CONCLUSIONS

- **Codon space has ultrametric structure.**
- **p-Adic distance plays important role in the genetic code.**
- **4 nucleotides are structural elements of 5-adic space of 64 codons.**
- **Codons clusterize according 5-adic and 2-adic distances.**
- **Degeneration of the Genetic Code has p-adic structure.**
- **This p-adic approach can be extended to some other aspects of genomics and proteomics.**

Mathematical side of Life contains p-Adics !

REFERENCES

- B. Dragovich and A. Dragovich, “A p-Adic Model of DNA Sequence and Genetic Code,” *p-Adic Numbers, Ultrametric Analysis and Applications* **1**, 34-41 (2009); [arXiv:q-bio.GN/0607018](#).
- B. Dragovich and A. Dragovich, “p-Adic Modelling of the Genome and the Genetic Code,” *The Computer Journal* **53**, 432-442 (2010); [arXiv:0707.3043v1](#) [q-bio.OT]