Physico-chemical properties of amino acids and arrangement of the genetic code.

Robersy Sánchez Rodríguez

Instituto de Investigaciones en Viandas Tropicales (INIVIT). Grupo de Biotecnología, Apdo. 6, Santo Domingo, Villa Clara, CP 53000, Cuba.

E-mail: robersy@cei.uclv.edu.cu

Recibido: 4 de diciembre de 2001 Aceptado: 28 de noviembre de 2002

Palabras clave: Código genético, propiedades de los aminoácidos, hidrofobicidad, energía de transferencia libre, clasificación de aminoácidos.
Key words: genetic code, amino acid properties, hydrophobicity, free energy of transfer, classification of amino acids.

RESUMEN Para investigar la relación existente entre las propiedades físico-químicas de los amino ácidos, la frecuencia de sustitución de unos amino ácidos por otros en las proteínas y la partición natural encontrada en el código genético, se definió un vector de propiedades que refleja, de forma generalizada, los efectos hidrofóbicos y estructurales de los amino ácidos, cuyas coordenadas estuvieron formadas por la energía libre de transferencia de los amino ácidos del agua al octanol, área media superficial accesible al solvente del amino ácido en su estado estándar, área media oculta del amino ácido en la transferencia del estado estándar a la proteína plegada, área media superficial polar accesible del amino ácido en el estado estándar y las relaciones entre la energía y estas variables. El análisis de cluster de los 20 amino ácidos mostró que los amino ácidos con propiedades físico-químicas similares están agrupados por codones con la segunda posición idéntica, excepto aquellos amino ácidos que poseen la guanina como segundo codón, W; R; C; G. Además en el análisis de correlación de la matriz de distancia entre los amino ácidos y las matrices de sustitución de los amino ácidos reportadas por Dayhoff et al. y por Gonnet et al, se obtuvieron correlaciones altamente significativas con valores entre -0.53 y - 0.95 con la matriz de Dayhoff y entre -0.79 y -0.96 con la matriz de Gonnet, para todos los amino ácidos, excepto para la cisteína, mostrando que mientras menor es la distancia entre los amino ácidos mayor es la frecuencia de sustitución de estos en las proteínas. Quedó manifiesta una estrecha conexión entre la organización del código genético, las propiedades físico-químicas de los amino ácidos y la evolución de las proteínas.

ABSTRACT: In order to carry out researches on existing relations between physico-chemical properties of amino acids, the substitution frequency of some amino acids by others in the proteins and natural partitioning found in the genetic code, a property vector was derived showing, in a generalized form, the hydrophobic and structural effects of amino acids. Coordinates were made up of the free energy of transfer of amino acids from water to octanol, the mean of solvent accessible surface area of the residue in standard state, the mean of area buried on transfer from the standard state to the folded protein, the mean of solvent accessible surface polar area of residue in standard state and the rate between this energy and these variables. Cluster analysis for 20 amino acids showed that amino acids with similar physico-chemical properties are grouped by identical second positions in their codon except those amino acids having guanine as a second codon: W; R; C; G. Besides, in the correlation analysis between distance matrix of amino acids and amino acids substitution matrices reported by Dayhoff et al and Gonnet et al, highly significant correlations were obtained from - 0.53 to - 0.95 with Dayhoff matrix and -0.79 to - 0.96 with Gonnet matrix, for all amino acids, except for cysteine, showing that the smaller the distance within amino acids, is the higher substitution frequency in the proteins. A closed connection was also evident between the genetic code organi-zation, physico-chemical properties of amino acids and protein evolution.

INTRODUCTION

Reports on the classification of amino acids are not abundant, particularly in the understanding of existing relationships between genetic codes and physico-chemical properties of amino acids.

Since a research carried out by Sneath in 1966, properties to characterize amino acids have been used as multiple variables; volume,

charge, hydrophobicity, capacity of making hydrogen bridges, etc.[1] He carried out the first efforts to classify amino acids and he used 134 properties to establish existing relations between chemical properties and biological activities of amino acids. However, the same author stated that amino acids were not well grouped. The groups obtained were as follow: L, I, V, A, G, P; Q, N, M, T, S, C; D, E, K, R, Y, F, W; H.

The hydrophobicity of amino acids constitutes one of the most outstanding properties used because it intends to show the magnitude of hydrophobic effects. They describe the thermodynamics of the partitioning of non-polar compounds between water and a non-watery phase. Kauzman stated that hydrophobic effects had the main role in the protein process folding because proteins are surrounded by a watery environment and protein water interactions are considered the leading power in the folding of the polypeptide chain.[2,3] Free energy has been used for transferring amino acids from water to organic solvent, establishing levels in order to estimate the contribution of each amino acid residue into the protein folding, because each residue hides its non-polar surface inside proteins, and exposes the most polar parts to the environment. Epstein used formulas where the molecular weight of amino acids residues, were only considered, as well as the classification in six polarity classes, obtaining identical differences for several amino acid pairs after calculating the Euclidean distance between them.[4] The distance used here was not a metrics; to be exact, it does not satisfy the triangular inequality. Later, Epstein, using distances between amino acids, calculated differences between homologous proteins and concluded that substitutions were not fortuitous.

In 1974, Granthan proposed a formula to find out differences between amino acids, making use of three properties: volume, polarity and composition.[5] Next, the differences between amino acids, were used in the linear regressions against relative substitution frequencies of exchanging amino acids in proteins (RSF) and log (RSF). He found correlation coefficients of -0.66 and 0.72, respectively.[6]

Partition of amino acids into groups taking as similarity criteria a frequency function ,where an amino acid is interchangeable by another, was carried out by Dayhoff et al.[7] Groups obtained were: C; S, T, P, A, G; N, D, E, Q; H, R, K; M, I, L, V; F, Y, W. The cluster method was not used in the obtainment of this partition. Later, Taylor used as many physico-chemical criteria as those of substituting an amino acid by another, starting from groups found by Dayhoff et al., but proceedings used for a further partition were not well defined from the quantitative point of view.[8]

More recently, Stanfel making use of physico-chemical properties of amino acids such as, volume, area, hydrophilicity, polarity, hydrogen bridge, shape and charge, proposed a new procedure. He used, the presentation of amino acids as property vectors, making amino acid-clusters by a new function derived from informational theoretic criterion.[9] The best solution was: A, C, G, I, L, M, P, S, T, V; D, E, N, Q; F, W, Y; H, K, R. Then, the previous procedure is applied to the higher grouping and the following subset is obtained: A, C, G, P; I, L, M, V; S, T.

Here, in order to obtain amino acid groups, similar to natural groups founded in the Genetic Code, we used variables associated with the shape and combined these with the free energy of amino acids transference from water to octanol.

MATERIALS AND METHODS

Data and Data Transformation

Hydrophobicity

The free energy of transfer of amino acids from water to octanol ($\Delta G_{transfer}$) was used to estimate the hydrophobic effect and is the most used value for estimating the contribution of this effect to changes in stability of mutants. These values were also transformed to: $\Delta G_{transfer}$/ A^o; $\Delta G_{transfer}$/ Ab; $\Delta G_{transfer}$/ Ap. The $\Delta G_{transfer}$ values were taken from Fouchere and Pliska[10].

Side chain volume (Scv)

This variable expresses approximately the relationship "Size/complexity" and the values were taken from Klapper.[11, 12].

Areas

1. A^o, The mean of solvent accessible surface area of residue in standard state *.
2. Ab, The mean of area buried on transfer from the standard state to the folded protein*.
3. A^o/Ab, The mean fractional area loss.
4. Ap, The mean of accessible polar area of residue in standard state. $Ap=A^o-A_{non\text{-}polar,}$ where $A_{non\text{-}polar}$ values are the hydrocarbon surface areas of residues and were taken from Karplus.[13]

* The values were taken from Rose et. al.[14]

Shape

This property was selected for the importance of amino acid shape as a determinant of protein shape. This variable represents the amino acid form and values were taken from Stanfel.[9]

Statistical data used in this work, are shown in table 1

Table 1. Statistical data.

* Residue	Ao (Å^2)	$A_{non-polar}$ (Å^2)	Ab (Å^2)	$\frac{Ab}{Ao}$	Ap (Å^2)	Shape	Scv	$\Delta G_{transfer}$ $\frac{kcal}{mol}$	$\frac{\Delta G_{transfer}}{Ao}$ $\frac{kcal}{mol\ Å^2}$	$\frac{\Delta G_{tranfer}}{Ab}$ $\frac{kcal}{mol\ Å^2}$	$\frac{\Delta G_{tranfer}}{Ap}$ $\frac{kcal}{mol\ Å^2}$
GLY G	88,1	47	62,9	0,71	41,1	1	5,7	0	0	0	0
ALA A	118,1	86	86,6	0,73	32,1	1,1	22,7	0,42	3,56	4,85	13,08
VAL V	164,5	135	141	0,86	29,5	1,3	56,7	1,66	10,09	11,77	56,27
ILE I	181	155	158	0,87	26	1,45	73,7	2,45	13,54	15,51	94,23
LEU L	193,1	164	164,1	0,85	29,1	1,4	73,7	2,31	11,96	14,08	79,38
PRO P	146,8	124	92,9	0,63	22,8	1,25	45,3	0,98	6,68	10,55	42,98
CYS C	146,1	48	132,3	0,91	98,1	3	34,9	2,09	14,31	15,8	21,3
MET M	203,4	137	172,9	0,85	66,4	3,3	74,6	1,67	8,21	9,66	25,15
PHE F	222,8	194	194,1	0,87	28,8	12	92,1	2,43	10,91	12,52	84,38
TRP W	266,3	236	224,6	0,84	30,3	12,15	120,7	3,06	11,49	13,62	100,99
TYR Y	236,8	154	177,7	0,75	82,8	12,05	100	1,31	5,53	7,37	15,82
HIS H	202,5	129	155,8	0,77	73,5	7	74,7	0,18	0,89	1,16	2,45
THR T	152,5	90	106,5	0,7	62,5	2,1	47,4	0,35	2,3	3,29	5,6
SER S	129,8	56	85,6	0,66	73,8	2	30,4	-0,05	-0,39	-0,58	-0,68
GLN Q	193,2	66	119,2	0,62	127,2	5,3	71	-0,3	-1,55	-2,52	-2,36
ASN N	165,5	42	103,3	0,62	123,5	5,1	54	-0,82	-4,95	-7,94	-6,64
GLU E	186,2	69	113,9	0,61	117,2	5,2	63,5	-0,87	-4,67	-7,64	-7,42
ASP D	158,7	45	97,8	0,62	113,7	5	45,6	-1,05	-6,62	-10,74	-9,23
LYS K	225,8	122	115,5	0,51	103,8	8,5	79,5	-1,4	-6,2	-12,12	-13,49
ARG R	256	89	162,2	0,63	167	8,6	100,4	-1,37	-5,35	-8,45	-8,2

*The three letter codes of the amino acids and then letter code.

A^o: Mean of solvent accessible surface area of residue in standard state.

Ab: Mean of area buried on transfer from the standard state to the folded protein.

A^o/Ab: Mean fractional area loss.

$A_{non-polar}$: Values are the hydrocarbon surface areas of residues.

Ap: Mean of accessible polar area of residue in standard state. $Ap=A^o-A_{non-polar,}$

Shape, the amino acids form.

Scv: Side chain volume.

$\Delta G_{transfer}$: Free energy of transfer of amino acids from water to octanol

$\Delta G_{transfer}$/Ao, $\Delta G_{transfer}$/Ab, $\Delta G_{transfer}$/Ap: Free energy of transfer of amino acids from water to octanol per surface unit.

Cluster analysis and Correlations

The partition analysis of amino acids into groups was made by beverage hierarchical cluster analysis using the representation of each amino acid as vector of properties: Ab; A^o/Ab; shape; Scv; $\Delta G_{transfer}$; $\Delta G_{transfer}$/Ao; $\Delta G_{transfer}$/Ab; $\Delta G_{transfer}$/Ap. Values of all variables used for calculating distance were standardized because the numerical scales for properties are different and ever expressed in different unit.The measurement employed here was:

$$m_{ij} = (m_{ij}-\mu_{ij})/\sigma_j$$

where m_{ij}= the raw measurement for amino acid i, property j; μ_{ij} = the mean of values for property j over all amino acids and σ_j = the standard deviation of values for property j over all amino acids. To perform a cluster analysis, the distance function between each pair of mino acids used was:

$d(x,y)= [\Sigma\ (x_i-y_i)^2]^{1/2}$

The Pearsons correlation coefficients were calculated per amino acid between distance matrices obtained from property vectors of amino acids and amino acid substitution matrices reported by Dayhoff et al. and Gonnet et al.[7, 15]

RESULTS AND DISCUSSION

Results of cluster analyses are shown in table 2. Particularly, the set partition of amino acids into 7 subsets: A, E, P, S, T, G; D, E, N, Q; H, Y; L, I, V, M; F, W; R; C. The

amino acids R, C, y W form each one a subset. It is well noticeable that, in the dendogram (Fig.1), histidine and tyrosine form a separate subset. Both amino acids are nearer to polar amino acids than to apolar amino acids (Table 3). The result is in correspondence with Sjostrom and Wold who found that no grouping was detected for amino acids coded by guanine (G) in the second codon position.[16]

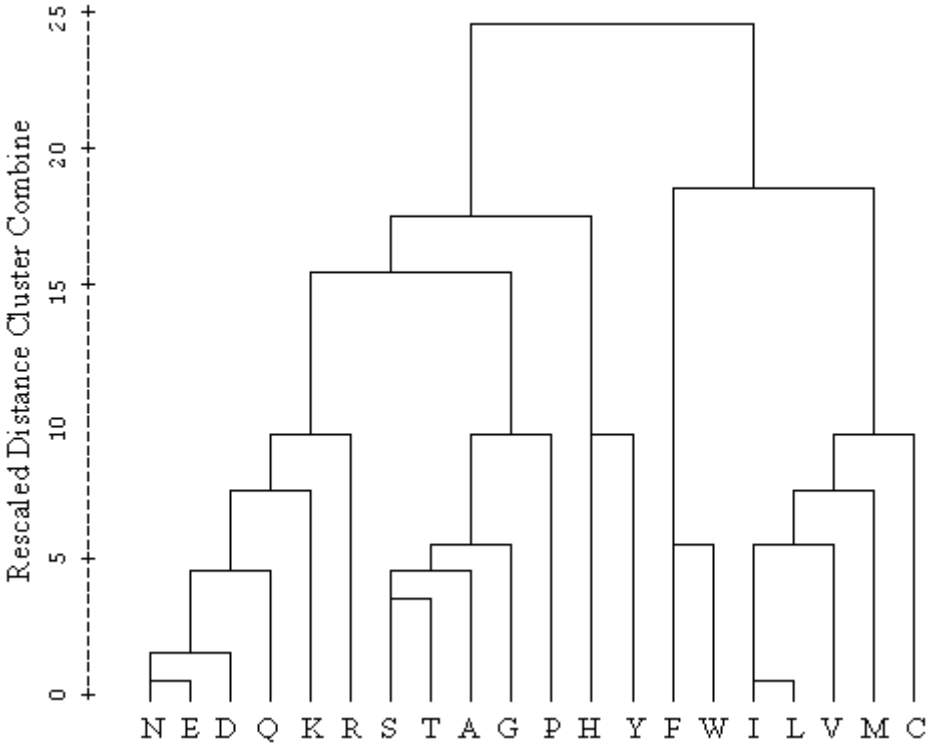

Fig 1. Dendrogram using Average Linkage (Between Groups).

Amino acids	7 Clusters	6 Clusters	5 Clusters	4 Clusters
ALA	1	1	1	1
ARG	2	2	2	1
ASN	3	3	2	1
ASP	3	3	2	1
CYS	4	4	3	2
GLN	3	3	2	1
GLU	3	3	2	1
GLY	1	1	1	1
HIS	5	5	4	3
ILE	6	4	3	2
LEU	6	4	3	2
LYS	3	3	2	1
MET	6	4	3	2
PHE	7	6	5	4
PRO	1	1	1	1
SER	1	1	1	1
THR	1	1	1	1
TRP	7	6	5	4
TYR	5	5	4	3
VAL	6	4	3	2

Table 2. Cluster Membership

Amino acid	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A		4,54	2,76	3,00	2,99	2,68	2,99	1,19	2,97	4,18	3,85	4,36	3,23	5,39	1,66	1,19	1,11	6,53	4,53	2,78
R	4,54		2,36	2,66	5,86	2,03	1,98	4,82	2,34	6,06	5,63	1,75	4,36	5,40	4,65	3,79	3,56	6,02	3,34	5,17
N	2,76	2,36		0,53	5,07	1,10	0,44	2,66	2,47	5,76	5,34	1,74	4,25	5,92	3,32	1,71	2,08	6,86	4,19	4,53
D	3,00	2,66	0,53		5,44	1,63	0,86	2,68	2,92	6,18	5,78	1,82	4,71	6,36	3,68	1,90	2,46	7,31	4,67	4,94
C	2,99	5,86	5,07	5,44		4,51	5,13	4,02	3,57	2,50	2,31	6,35	2,09	3,85	3,03	3,98	3,19	4,98	4,03	1,57
Q	2,68	2,03	1,10	1,63	4,51		0,87	3,02	1,76	5,02	4,59	1,94	3,47	5,12	2,80	1,90	1,70	5,98	3,31	3,90
E	2,99	1,98	0,44	0,86	5,13	0,87		3,01	2,31	5,71	5,28	1,49	4,16	5,79	3,39	1,99	2,17	6,67	3,98	4,54
G	1,19	4,82	2,66	2,68	4,02	3,02	3,01		3,66	5,32	5,00	4,31	4,32	6,39	2,62	1,16	1,90	7,56	5,37	3,92
H	2,97	2,34	2,47	2,92	3,57	1,76	2,31	3,66		4,10	3,65	3,24	2,22	3,78	3,16	2,80	2,09	4,73	2,10	3,05
I	4,18	6,06	5,76	6,18	2,50	5,02	5,71	5,32	4,10		0,54	6,78	2,23	2,97	3,44	4,99	4,07	3,63	4,07	1,50
L	3,85	5,63	5,34	5,78	2,31	4,59	5,28	5,00	3,65	0,54		6,37	1,74	2,91	3,15	4,61	3,66	3,61	3,72	1,17
K	4,36	1,75	1,74	1,82	6,35	1,94	1,49	4,31	3,24	6,78	6,37		5,25	6,40	4,52	3,33	3,54	7,13	4,38	5,76
M	3,23	4,36	4,25	4,71	2,09	3,47	4,16	4,32	2,22	2,23	1,74	5,25		2,94	3,00	3,76	2,74	3,83	2,63	1,43
F	5,39	5,40	5,92	6,36	3,85	5,12	5,79	6,39	3,78	2,97	2,91	6,40	2,94		4,80	5,82	4,95	1,42	2,47	3,40
P	1,66	4,65	3,32	3,68	3,03	2,80	3,39	2,62	3,16	3,44	3,15	4,52	3,00	4,80		2,15	1,62	5,75	4,15	2,42
S	1,19	3,79	1,71	1,90	3,98	1,90	1,99	1,16	2,80	4,99	4,61	3,33	3,76	5,82	2,15		1,07	6,89	4,54	3,63
T	1,11	3,56	2,08	2,46	3,19	1,70	2,17	1,90	2,09	4,07	3,66	3,54	2,74	4,95	1,62	1,07		5,97	3,73	2,72
W	6,53	6,02	6,86	7,31	4,98	5,98	6,67	7,56	4,73	3,63	3,61	7,13	3,83	1,42	5,75	6,89	5,97		3,19	4,39
Y	4,53	3,34	4,19	4,67	4,03	3,31	3,98	5,37	2,10	4,07	3,72	4,38	2,63	2,47	4,15	4,54	3,73	3,19		3,66
V	2,78	5,17	4,53	4,94	1,57	3,90	4,54	3,92	3,05	1,50	1,17	5,76	1,43	3,40	2,42	3,63	2,72	4,39	3,66	

Table 3. Euclidean distance matrix obtained using the values of standardized variable as coordinates: Ab; Ao/Ab; Shape, Scv, $\Delta G_{transfer}$; $\Delta G_{transfer}$/Ao, $\Delta G_{transfer}$/Ab, $\Delta G_{transfer}$/Ap.

The fourth subset of the genetic code has guanine as the second base in the codon and is integrated by amino acids without common physico-chemical properties. If these amino acids can't be classified within other subsets, then, a great connection would be present between this partition and the established in the genetic code (Table 4).

Partition	Subset *						
	I	**II**	**III**	**IV**	**V**	**VI**	**VII**
Five groups	**A P S T** G	**D N E Q K** R	**I L M V** C	Y H	F W		
Six groups	**A P S T** G	**D N E Q K**	**I L M V** C	Y H	F W		R
Seven groups	**A P S T** G	**D N E Q K**	**I L M V**	**C**	H Y	F W	R
Dayhoff et al	**A P S T** G	**N D E Q**	**I L M V**	**C**	H R K	F W Y	
Genetic Code	**A P S T**	**D N E Q K H Y**	**I L M V F**	**R S W CG**			

*The amino acids that coincide in a subset with the genetic code are in black.

Table 4. Comparison of the partitions with the genetic code and Dayhoff partition.

This connection involves genetic code arrangement with protein structures. Chisuano et al, after investigating the relationships between base composition of coding sequences and secondary structures of encoded proteins, found that the physical-chemical properties of protein structures, strongly dependent on amino acids composition are indeed correlated with well defined choices in second base position of corresponding coding sequences.[17]

From mentioned partitions, the one of highest correspondence with the natural partition is carried out by Dayhoff et al.[7] Table 3 shows the correspondence between the genetic code, the new obtained partitions and the one obtained by Dayhoff.
The partition in seven subsets is able to classify amino acids more correctly than Dayhoff's partition, in relation to the Genetic code.

However, it is important to keep in mind that partitions start from different criterions and while the given classification takes as base physico-chemical criteria, the Dayhoff's partition is deduced from substitution surveys of amino acids within a family of isofunctional proteins.

The high correspondence between the two classifications makes us think that amino acids substitutions (mutations) within isofunctional protein families don't take place at random, but on similarity base of physico-chemical properties of exchanged amino acids. Particularly, the role played by hydrophobic interactions may be valued, because most of the variables used are closely bound together with this factor. The correlation analysis between the distance matrix used in the cluster analysis and the amino acid substitution matrices reported by Dayhoff et al. and Gonnet et al. support this latter idea (Table 5). [7, 15] In general, there is a high matrix correlation with both matrices. It is important to stand out that correlation coefficients for Gonnet matrix are greater than Dayhoff matrix, which is in agreement with the generalized criterion that Gonnet matrix shows with greater accuracy the substitution frequencies of an amino acid by another. The high value of coefficients means that, there are not discrepancies between isofunctional amino acid groupings based on comparisons of physicochemical properties of amino acids and those based on their mutational substitutions. This implies that the above distance matrix could be used as an amino acid index matrix like that reported by Kawashima et al.[19]

	ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR
Dayhoff	-0,91	-0,65	-0,95	-0,94	-0,48*	-0,87	-0,91	-0,95	-0,83	-0,70	-0,73	-0,82	-0,53	-0,88	-0,82	-0,88	-0,86	-0,61	-0,58
Gonnet et al	-0,92	-0,90	-0,95	-0,98	-0,66	-0,93	-0,96	-0,85	-0,81	-0,86	-0,85	-0,95	-0,80	-0,96	-0,79	-0,95	-0,88	-0,84	-0,87

- Significant for $p < 0.05$. The other coefficients are significant for $p < 0.01$.

Table 5. Correlation coefficients between the Euclidean distance matrix used in the cluster analysis and amino acids substitution matrices reported by Dayhoff and by Gonnet et al.

Results are supported by the fact that hydrophobicity and volume variables have a significant contribution to Dayhoff matrix.[17] They are two important factors affecting the amino acid substitution during evolution.[5] Besides, Rose and Wolfenden[3] found that $\Delta G_{transfer}$ is a good indication of changes in stability for most amino acids substitutions in the protein core. Mentioned variables in the previous works are well represented in the property vector of amino acids used in the cluster analysis. These variables may be grouped like those which are the reflection of a generalized hydrophobic effect: Ab; Ao/Ab, $\Delta G_{transfer}$; $\Delta G_{transfer}$/ Ao, $\Delta G_{transfer}$/Ab, $\Delta G_{transfer}$/Ap, and generalized structural effect: Shape and side chain volume.

CONCLUSIONS

The property vector of amino acids, which has been used in this work, reflects the hydrophobic and structural generalized effects of amino acids on protein folding. This vector shows a closed connection with the genetic code arrangement and the substitution frequencies of amino acids in proteins. The amino acids that have a small euclidian distance within respective vectors are grouped, in general, by identical second positions in their codons. Moreover, within these amino acids, while the substitution frequency of one by another increases, the distance within them decreases and so on.

These results show that the genetic code arrangement into four sets is not at random. That is, this partition reflects the amino acids with similar physico-chemical properties grouped by identical second positions in their codon, except those amino acids having guanine as second codon, W; R; C; G.

BIBLIOGRAPHY

1. Sneath P. Relation between Chemical and Biological Activity in Peptides. **J Theor Biol**, **12**, 157, 1966.
2. Kauzmann W. Some factors in the interpretation of protein denaturation. **Advances in Protein Chemistry**, **14**, 1, 1959.

3. Rose G. D. and Wolfenden R. Hydrogen bonding hydrophobicity, packing, and protein folding. **Ann Rev Biophys Biomol Struct**, **22**, 381, 1993.
4. Epstein C. Non randomnes of amino-acid changes in the evolution of homologous proteins. **Nature**, **215**, 355, 1967.
5. Granthan R. Amino Acid Difference Formula to Help Explain Protein Evolution. **Sciences**, **185**, 862, 1974.
6. McLachlan A D. Repeating Sequences and Gene Duplication in Proteins. **J Mol Biol**, **64**, 417, 1972.
7. Dayhoff, M., Schwartz, R. & Orcutt B. A model of evolutionary change in proteins. In: Atlas of Protein Sequence and Structure. Silver Spring, MD: National Biomedical Research Foundation, 345-352, 1978.
8. Taylor W. The classification of amino acid Conservation. **J Theor Biol**, **119**, 205, 1986.
9. Stanfel, L E. A new approach to clustering the amino acids. **J Theor Biol**, **183**, 195, 1996.
10. Fauchere J.L and Pliska V. Hydrophobic parameters pi of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. **Eur J Med Chem - Chim Ther**, **18**, 369, 1983.
11. Klapper, M.H. Nature of the protein interior. **Biochim Biophys Acta**, **229**, 557, 1971.
12. Dufton, M. J. Genetic code synonym quotas and amino acid complexity: Cutting the Cost of Proteins? **J Theor Biol**, **187**, 165, 1997.
13. Karplus P. A. Hydrophobicity regained. **Protein Science**, **6**, 1302, 1997.
14. Rose, G. D, Hydrophobicity of amino acid residues in globular proteins. **Sciences**, **229**, 834, 1985.
15. Gonnet, G. H., Cohen, M. A., and Benner, S. A. Exhaustive matching of the entire protein sequence database. **Science**, **256**, 1443, 1992.
16. Sjostrom M and Wold S. A multivariate study of relationship between the genetic code and the physical-chemical properties of amino acids. **J Mol Evol**, **22**, 272, 1985.
17. Chisuano, M. L. Second codon position of gene and secondary structure of proteins. Relationship and implications for the origen of the genetic code. **Gene, 261**, 63, 2000.
18. Kawashima S, Ogata H, Kanehisa M. AAindex: amino acid index database. **Nucleic Acids Res**, **27**, 368, 1999.