Journal of Sciences, Islamic Republic of Iran 13(3): 203-208 (2002)
National Center For Scientific Research, ISSN 1016-1104

AN INDEX REPRESENTING STRUCTURE-CATABOLIC FATE RELATIONSHIPS OF AMINO ACIDS

E. Forouhi*

Department of Biochemistry, Faculty of Medicine, Guilan University of Medical Science, Rasht, Islamic Republic of Iran

Abstract

Based on the Randic suggestion of the resolution of a structure into shape, size, and function an index representing structure-catabolic fate relationships of amino acids is constructed. The index obtained by multiplying three factors; $(n_{b2} + n_r)$, n_c and M; representing shape, size and function respectively, where n_{b2} = number of double bonds, n_r = number of rings, n_c = number of carbon atoms, and M = molecular weight. A systematics of amino acids is obtained and a correlation is observed between the index value and the catabolic fate.

Keywords: Amino acids; Catabolic fate; Structural index

Introduction

From the premise that the chemical properties of a compound depends on its structure [1], generalizations are used to be made in describing the relationship between chemical structure and activity. There are two approaches to a search for an structure-activity relationship [2,3]. The first leads directly to an explanation of the mechanism of actions and responses expressed in chemical structure terms. Examples are: Probing the "active site" of diamine oxidase [4], understanding the substrate specificity of cytochrome p-450 [5], and the mechanism of catalysis of kidney aldehyde reductase [6] *etc*.

The second approach is the possibility of utilizing the structure-activity relationships as a predictive scheme in the design of new and hopefully more effective molecules such as potassium channel activators or openers [7]. Nucleosides [8], activity-dependent neurotrophic factors [9] *etc*.

A useful approach in generalization mentioned above is to seek patterns in the form of relationships between the graph theoretic features or topology of the molecule and its properties such as physico-chemical properties, chemical reactivity and biological activity. An efficient way of coding the topology of a chemical structure is represented by topological indices [10-16]. The application of the molecular topological index has been examined in quantitative structure-activity relationship studies of toxicity of alkyl alcohols on fish [14], of boiling points of alkanes [15], and of taste of various compounds [17].

Graph theory has been applied to quantum chemistry [18], and quantum molecular similarity measures

* *E-mail: medical@gums.ac.ir*

resulted in a theoretical foundation of quantitative structure properties relationships [19]. Molecular orbital theory has been used to derive reactivity indices [29]. Shape similarities of electron density clouds of molecules provide important clue concerning chemical and physical properties, including information about their reactivities in biochemical systems [20]. The concept of topological resolution has been used for quantifying molecular similarities [20]. The calculation of electronic charge from molecular orbital theory has led to a number of useful predictions about structure-activity relationships among drug molecules. From these calculations it was possible to draw certain conclusions about the metabolism of three anesthetic gases [2]. Energetic principles have also been applied to derive structure-activity [21], structure-reactivity [22], and structure-function relationships [23].

In this work, based on graph theoretical features of the molecules, an index representing structure catabolic fate relationship of amino acids is introduced.

Construction of the Index

In application of graph theory to complex molecules, Randic [10] believes that it would help if one could dissect a molecule and focus on individual selected features. So he suggests the following resolution of a structure:

(Shape) (Size) (Function)

"Shape" depends on geometry. "Size" depends on the number of atoms in a compound, and "Function" involves variations in the atomic composition, that is, the presence of heteroatoms. The conditions put forward by Randic [24] for structural ordering of acyclic saturated hydrocarbons may be applied to ordering of other groups of compound such as amino acids, reminding that it is not easy to satisfy simultaneously all the criteria. These conditions are:

(1) The ordering is based on an invariant (index).

(2) The underlying invariant has transparent structural interpretation.

(3) Ordering is complete, *i.e.* no two isomers may have an identical index.

These suggestions of Randic are utilized to construct the index. So, firstly the parameters representing the different structural features of the amino acid molecules are defined.

Most biochemical compounds contain mainly single and double bonds and the main elements in them are carbon, oxygen, hydrogen, nitrogen, phosphorus and sulphur. The following relationships exist between the numbers of atoms, bonds, rings and valences in these compounds.

$$n_H + 2n_O + 2n_S + 3n_N + 4n_C + 5n_P = 2(n_{b1} + 2n_{b2}) \quad (1)$$

$$n_P + N_H + n_N = 2(N + 1) \quad (2)$$

$$n_H + n_O + n_S + n_N + n_C + n_P - (n_{b1} + n_{b2}) + n_r = 1 \quad (3)$$

Where n_H, n_O, n_S, n_N, n_C, and n_P are the numbers of the atoms of hydrogen, oxygen, sulphur, nitrogen, carbon, and phosphorus respectively. n_r is the number of rings; and n_{b1}, and n_{b2} are the numbers of single and double bonds respectively. N is an integer.

Equation (1) is derived from the application of the first theorem of graph theory: In any graph the sum of the degree of vertices is twice the number of edges [25]. Equation (2) is derived from a corollary of this theorem, and Equation (3) is derived from the Euler's theorem [25].

For amino acids $n_p = 0$, and from equations (1)-(3) it is deduced that:

$$n_C - (n_{b2} + n_r) = N \quad (4)$$

N = 1,2,3,4,5 (Table 1)

N depends on n_H and n_N. The number of divalent atoms such as oxygen and sulphur do not appear in equation (4). n_{b2} and n_r are in the same category, that means in writing the isomers of the compounds, a double bond may be replaced by a ring or vice versa. For example there is a double bond (carbonyl group) in open chain form of glucose. In the ringed from this double bond is replaced by a ring. The $(n_{b2} + n_r)$ depends on the geometry of the molecule and therefore may represent its shape. n_C represent the size, having in mind its application to the catabolism of the carbon skeleton of amino acids. For given values of $(n_{b2} + n_r)$ and n_C, the value of N is fixed, so N is a dependent variable. Molecular weight is chosen to represent the function, since so far as the catabolism is concerned, it brings into account the presence of heteroatoms such as nitrogen, oxygen and sulphur. So following form of the index is proposed for ordering of amino acids:

$$\Phi = (n_{b2} + n_r) \cdot \frac{n_C}{(n_C)_{Gly}} \cdot \frac{M}{M_{Gly}} \quad (5)$$

Where $(n_C)_{Gly}$ and M_{Gly} are the number of carbon atoms and the molecular weight of glycine respectively. n_C and M are the number of carbon atoms and the molecular weight of the amino acids respectively.

Table 1. The value of N for different amino acids as predicted from Equation (4)

N				
1	**2**	**3**	**4**	**5**
Gly	Ala	Thr	Val	Met
	Ser	Pro	Arg	Ile
	Cys	Gln	Phe	Leu
	Asn	Glu	Tyr	Lys
	Asp		Trp	
	His			

Results and Discussion

The values of the index representing the structure catabolic fate relationship for the twenty amino acids occurring in proteins calculated from Equation (5) are given in Table 2. The amino acids have been arranged according to their index (Φ) value. As can be seen, dependence is observed between the index value and the catabolic fate of the amino acids (for the catabolic fate of the amino acids see reference [26]). Meanwhile an interesting order is obtained, that is when the amino acids are arranged according to their index value, they seem as if they have been grouped, divided and subdivided according to their values of $n_{b2} + n_r$, n_C, and molecular weight respectively.

Table 2. Systematics of the amino acids and the value of index for different amino acids as predicted from equation (5)

Amino acids	$n_{b2} + n_r$	n_c	M	Φ	The main product of the catabolism	Other products
Gly	1	2	75.1	1	Pyruvate	
Ala	1	3	89.1	1.78		
Ser	1	3	105.1	2.10		
Cys	1	3	121.2	2.42		
Thr	1	4	119.1	3.17		Acetyl-CoA
Val	1	5	117.1	3.90	Succinyl-CoA	Methyl
Met	1	5	149.2	4.07		
Ile	1	6	131.2	5.24		Acetyl-CoA
Leu	1	6	131.2	5.24	Acetoacetate	
Lys	1	6	146.2	5.84		
Asn	2	4	132.1	7.04	Oxaloacetate	
Asp	2	4	133.1	7.09		
Pro	2	5	115.1	7.66	α-ketoglutarate	
Gln	2	5	146.1	9.72		
Glu	2	5	147.1	9.79		
Arg	2	6	174.2	13.92		Urea
His	4	6	155.2	24.80		From-imino
Phe	5	9	165.2	49.49	Acetoacetate	Fumarate
Tyr	5	9	181.2	54.24		
Trp	7	11	204.2	104.68		Pyruvate

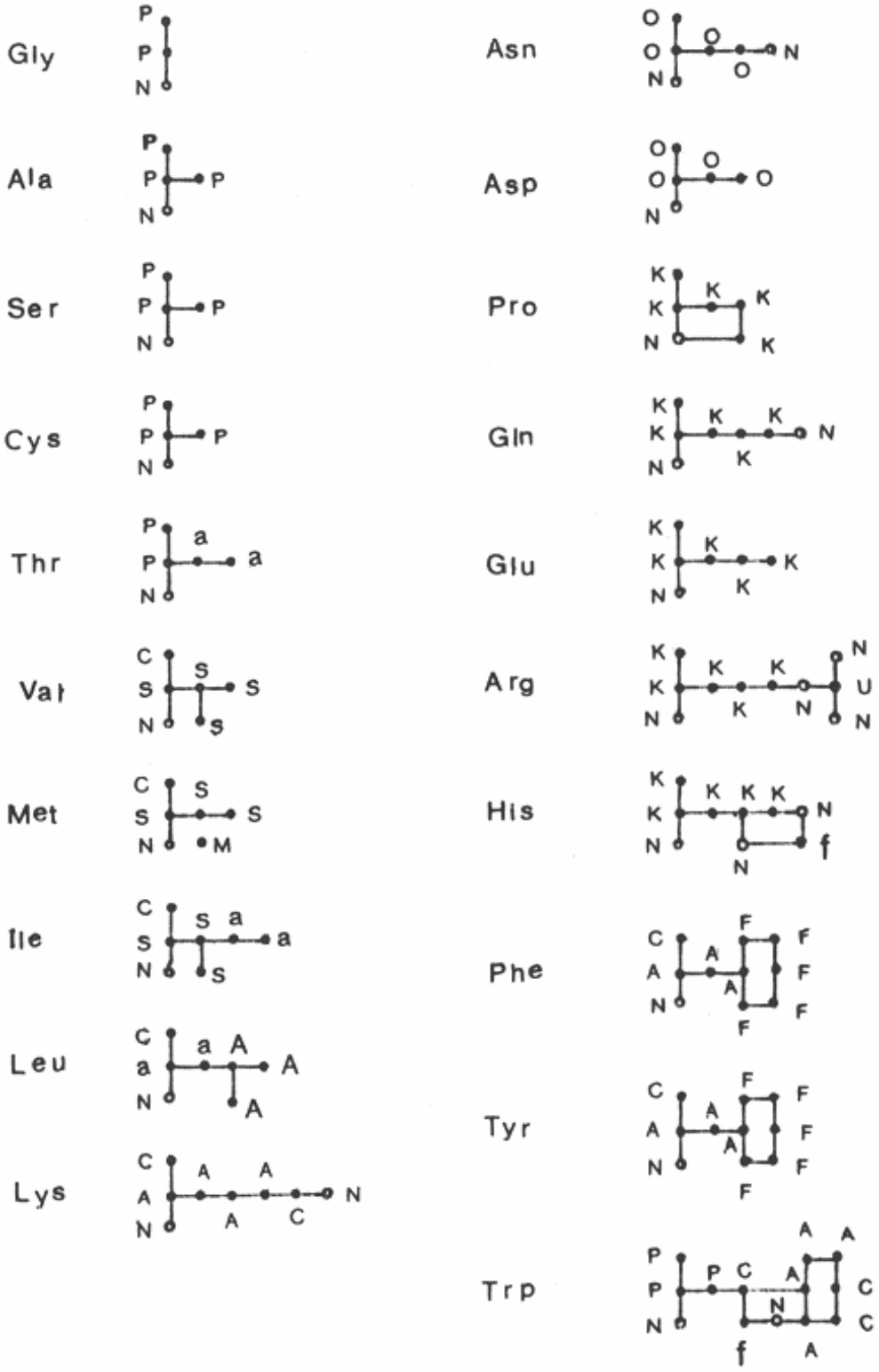

Figure 1. Monovalent and divalent suppressed labeled graph of the amino acids.
P: Pyruvate, a: Acetyl-CoA, S: Succinyl-CoA, C: Co_2, M: Methyl, A: Acetoacetate, O: Oxaloacetate, K: α-Ketoglutarate, f: formic acid or formimino, F: Fumarate, U: Urea, N: Nitrogen(urea), ●:Carbon, ○: Nitrogen.

Similar molecules are expected to have similar properties. The postulate on similarity as mentioned by Randic [10] is formulated in an alternative form as the principles of graduality: "The changes in nature are gradual". The order presented in Table 2 reveals the graduality of changes in the side chain structure of the amino acids, which may be expressed as follows: No double bonds (*e.g.* alanine); more atoms (*e.g.* valine); double bond (*e.g.* asparagine); ring but no double bond (proline); both ring and double bond; aromaticity; one five membered aromatic ring (histidine); one six membered aromatic ring (*e.g.* tyrosine); and finally two aromatic rings (tryptophan). Accumulation of insignificant differences may lead to complete dissimilarity. In Table 2 the set of amino acids begins with the simplest one, glycine, and ends with tryptophan which is more complex than glycine, but the changing process is gradual. The last four amino acids with the highest and most significant index value, absorb ultraviolet light of a wavelength above 240 nm. Figure 1 shows the monovalent and divalent suppressed labeled graphs of the twenty amino acids. The label represent the catabolic fate of the carbons and nitrogens. The graphs of those amino acids with five membered ring (proline, histidine, and tryptophan) are 3-colorable, but the rest are 2-colorable.

The index introduced by Equation (5) reflects in some way the action of the enzymes involved with the catabolism of the amino acids. The structure of a chemical compound is defined by its constitution (the number, kind, and connectivity of the atoms in the molecule), configuration and conformation [27]. These are the determining factors in the specificity of enzyme action. Molecular orbital theory has been utilized to predict the conformational preferences of the amino acids in solution [30]. But in biochemical reactions the conformational preferences of the substrate molecules are determined by the active site of the enzymes catalyzing those reactions. The set of catabolizing enzymes objectify the relationship between the amino acid structure and its catabolic fate. Removal of the α-amino nitrogen by the transamination is most often the initial reaction of amino acid catabolism. Subsequent reactions remove any additional nitrogen and restructure the remaining hydrocarbon skeleton for conversion to amphibolic intermediates [26]. The dependence of enzyme activity on the chain length of the alkyl and acyl groups attached to substrate has been insisted by investigators [28].

The index is some kind of overall reactivity index of the set of metabolites in the amino acids catabolic pathways. It gives an insight into the structural evolution of the amino acids and the development and economic use of the substrates and enzymatic mechanisms in living systems.

The index may help to predict the catabolic fate of compounds similar to amino acids such as drugs to reduce the redundant parts when designing synthetic molecules.

Structure-catabolic fate relationship studies and introduction of the indices such as the one presented in this work help to bridge between the two main topics – structure and metabolism of biomolecules – with which the biochemistry is concerned.

Acknowledgements

Financial support for this project from Guilan University of Medical sciences is gratefully acknowledged.

References

1. Mercier C., Sobel Y. and Dubias J.E. Impact of regularity in structure-property relationships. *J. Math. Chem.* **9**: 351-357 (1992).
2. Kier L.B. and Tute M.S. Theoretical aspects of drug design. In: Principle of Medicinal Chemistry. Foye W.O. (Ed.), Lea & Febiger, Philadelphia, pp. 53-62 (1981).
3. Chu K.C. The quantitative analysis of structure-activity relationships. In: The Basis of Medicinal Chemistry, part 1, 4th Edition, Wolff M.E. (Ed.), John Wiley & Sons, Toronto, pp. 393-418 (1980).
4. High A., Prior T., Bell R.A. and Rangachari P.K. Probing the "active site" of diamine oxidase: structure-activity relations for histamie potentiation by o-alkyl-hydroxylamines on colonic epithelium. *J. Pharmacology and Experimental Therapeutics (JPET)*, **288**: 490-501 (1999).
5. Ekins S., Bravi G., Ring B.J., Gillespie T.A., Gillespie J.S., Vandenbranden M., Wrighton S.A. and Wikel J.H. Three dimensional quantitative structure-activity relationship analyses of substrate for CYP2B6. *Ibid.*, **288**: 21-29 (1999).
6. Bhatnagar A., Si-Qi-Liu and Srivastava S.K. Structure-activity correlations in human kidney aldehyde reductase reduction of para-substituted benzaldelyde by 3-acetyl pyridine adenine dinucleotide phosphate. *Biochim, Biophyn. Acta*, **1077**: 180-186 (1991).
7. Evans J.M. and Taylor S.G. Potassium channel activators; pharmacological methods, models, and structure-activity relationships, progress in Medicinal Chemistry. Ellis G.P. and Luscombe D.K. (Eds.), Elsevier Science, 31, 411-445 (1994).
8. Ohno M. Nucleosides; structure-activity relationships, Medicinal Chemistry. Canady J.M. and Dours J.D. (Eds.), Academic Press, London, 16, 73-130 (1980).
9. Brenneman D.E., Hauser J., Neale E., Rubinraut S., Fridkin M., Davidson A. and Gozes I. Activity-dependent neurotrophic factor: structure-activity relationships of femtomolar-acting peptides. *JPET*, **285**: 619-627 (1998).

10. Randic M. In search of structural invariants. *J. Math. Chem.*, **9**: 97-146 (1992).
11. Plavsic D., Nikolic S., Trinajstic N. and Mihalic Z. On the Harrary index for the characterization of chemical graphs. *J. Math. Chem.*, **12**: 235-250 (1993).
12. Nikolic S., Trinajstic N. and Mihalic Z. Molecular topological index: An extension to hetrosystems. *Ibid.*, **12**: 251-264 (1993).
13. Ivanciuc O., Balaban T.S. and Balaban A.T. Design of the topological index. *Ibid.*, **12**: 309-318 (1993).
14. Juric A., Gagro M., Nikolic S. and Trinajstic N. Molecular topological index: An application in the quantitative structure-activity relationship study of toxicity of alcohols. *Ibid.*, **11**: 179-186 (1992).
15. Morales D.A. and Aranjo O. On the search for the best correlation between graph theoretical invariants, and physicochemical properties. *Ibid.*, **13**: 95-106 (1993).
16. Balaban T.S., Filip P.A. and Ivanciuc O. Computer generation of acyclic graphs based on local vertex invariants and topological indices. *Ibid.*, **11**: 79-105 (1992).
17. Shallenberger R.S. *Taste Chemistry*. Blackie Academic & Professional, London, pp. 424-434 (1993).
18. Gutman I. and Trinajstic N. Graph theory and molecular orbitals, In: Topics in Current Chemistry. Springer-Verlag, Berlin 42, 49-93 (1973).
19. Carbo R., Besalu E., Amat L. and Fredera X. Quantum molecular similarity measures as a natural way leading towards a theoretical foundation of quantitative structure-properties relationships. *J. Math. Chem.*, **18**: 237-246 (1995).
20. Mezey P.G. Shape-similarity relations based on topological resolution. *Ibid.*, **27**: 61-69 (2000).
21. Dubickas A.K. and Petrauskas A.A. Application of linear free energy principles for derivation of nonlinear quantitative structure-activity relationships. *Ibid.*, **13**: 115-132 (1993).
22. Conners K.A. *Chemical Kinetics*, Chap 7: Structure-reactivity relationships. VCH Publishers, Inc, Cambridge, pp. 311-383 (1990).
23. Cera E.D. Site-specific analysis of mutational effects in proteins. *Adv. Protein Chem.*, **51**: 59-119 (1998).
24. Randic M. On structural ordering and branching of acyclic saturated bydrocarbons. *J. Math. Chem.*, **24**: 345-358 (1998).
25. Clark J. and Holton D.A. *A First Look at Graph Theory*. World Scientific, London, pp. 14-16 (1991).
26. Murray R.K., Granner D.K., Moyes P.A. and Rodwell V.W. *Harper's Biochemistry*. Appleton & Lange, London (2000).
27. Eliel E.L., Wilen S.H. and Mander L.N. *Stereochemistry of Organic Compounds*. John Wiley & Sons, Inc., Toronto, pp. 11-49 (1994).
28. Zimmerman J.J. and Feldman S. Physical-chemical properties and biologic activity, in: Principles of Medicinal Chemistry, Foye W.O. (Ed.), Leo & Febiger, Philadelphia, pp. 11-51 (1981).
29. Fukui K. *Theory of Orientation and Steroselection*, Chap 6: Reactivity indices. Springer-Verlag, Berlin, pp. 34-39 (1975).
30. Kier L.B. and George J.M. Molecular orbital consideration of amino acid conformation, In: Molecular Orbital Studies in Chemical Pharmacology. Kier L.B. (Ed.), Springer-Verlag, Berlin, pp. 82-104 (1970).