match no. 13 pp. 325-332 1982

EDGE- WEIGHTED GRAPHS IN BIOLOGICAL SYSTEMS

R.S.Lall[a] and V.K.Srivastava*

* Biophysical Laboratory, Department of Chemistry Gorakhpur University , Gorakhpur 273001, U.P. India.

[a] Department of Chemistry, St. Andrew's College, Gorakhpur, 273001, U.P., India

(Received: January 14, 1982)

Abstract

Edge - weighted graphs of purine and pyrimidine bases constituting the biological macromolecules (DNA and RNA) are defined. Sum of the topological distances of these bases are calculated defining the distance matrix associated with the edge- weighted graph. Mean topological distances of these bases are correlated with the observed properties.

Introduction

Search of a suitable topological index [1,2] for physico-chemical characterization of the biological macromolecules , is a field of recent interest. This may provide an indirect route for approaching the structural- reactivity relationship[3]. The applicability of topological index in medicine[4] , pharmacology[5] and organic chemistry[6] is very well established.

Molecular topology has recently been used to characterize various structures in unsaturated and saturated hydrocarbons[6-8]. The topological indices defined on the molecular graphs reflect the physical and chemical beha-

viour of the molecules[6-9]. Besides the chemical reactivity of organic compounds with same functional groups has been correlated with structural indices thus representing a topological subgroup additivity model [3, 10].

In the present investigation we wish to construct the representative molecular graphs [11] of purine and pyrimidine bases constituting DNA and RNA macromolecules in order to define a topological index which might be used to characterize their physicochemical properties.

Graphs in Hetero Systems

Molecular graphs[11] of heterosystems are defined by sets of edges and vertices in which the edges represent the C-C bonds and different hetero bonds e.g., C-N, C-O etc. and vertices represent the atoms C,N , O etc. These graphs are thus not simple molecular graphs as we have used in hydrocarbons[7,8]. Thus

$$\text{graph } G = [V(G),\ E(G)]$$

where sets $V(G) = [v_1 , v_2 , v_3 \cdots]$

and $E(G) = [e_1 , e_2 , e_3 \ \ldots]$

In hetero conjugated systems edge and vertex weighted graphs have been used[12] in the framework of Huckel Molecular orbital theory for calculating molecular orbital energies[13] In these graphs different weights are given to the corresponding vertices and edges. The weights correspond to Huckel parameters k and h for hetero atoms[14]. Sachs methods[15] have been employed for different properties of these graphs. However some authors have used chemical formula graphs for

for defining the topological indices for complex molecules such as higher degree connectivity indices in biological systems[16, 17].

Edge-weighted Graphs

For computing the topological characteristics of hetero systems we have to define a representative graph. A hetero system may be represented by a weighted graph.

Edge-weighted graph - An edge-weighted graph G_{Ew} is a graph which has edges associated with appropriate weights corresponding to C-C bonds and hetero bonds C-X.

Weight is identical with the Huckel parameters[14] ks for hetero bonds. Huckel parameters used in hetero systems are defined by

$$\alpha_X = \alpha_C + h\beta_{CC}$$

and $$\beta_{CX} = k\beta_{CC}$$

where α_C and β_{CC} are standard carbon coulomb and C-C resonance integrals respectively. 'h' reflects the difference between the coulomb integral of C atom (α_C) and that of the hetero atom X (α_X). Similarly k ($\neq 1$) signifies the difference between the C-C resonance integral (β_{CC}) and the resonance integral of C-X bond (β_{CX}) respectively.

Examples:

Purine molecule graph G_1

Pyrimidine molecule graph G_2

G_1 and G_2 are the edge-weighted graphs of purine and pyrimidine bases. Vertex o represents a nitrogen atom and • represents a nitrogen atom directly connected with a proton. The values of ks are the weights assigned to edges.

A distance matrix[6,11] may be defined associated with the edge-weighted graph , the elements of which represent the appropriate topological distances.

Weighted distance matrix - A weighted distance matrix is a real NxN matrix $D_w = [d_{ij}]$ where the element d_{ij} is the topological distance between the ith and jth vertices . N is the total number of vertices.

Topological distance - A topological distance is the minimal path between two vertices . Topological paths are treated as graph invariant.

Minimal path - A minimal path is a path which is associated with minimum weight,

$$d_{ij} = \sum_i k_i < \text{weight for any other path}$$

Topological index

Sum of the topological distances or the Weiner number[18] can now be defined on the edge-weighted graph G_{ew} . The Weiner number is half the sum of the off-diagonal elements of weighted distance matrix,

$$W = \frac{1}{2} \sum_{ij} d_{ij}$$

The mean value of the sum of the topological distances is defined as

$$\bar{W} = \frac{2W}{N(N-1)}$$

where N is the total number of vertices.

We have given the appropriate k values to different edges . For C-N= or C=N- bonds k= 0.7 and for C-N-H bond i.e. bond between a C atom and a N atom directly attached to a proton k= 0.9 . For C-O or C=O bonds k= 0.8 and for C-C bonds k= 1.0 as in usual way. These values of Huckel parameters are approximated from the values previously reported by Streitwieser[14] and more recently by Hess and Schaad[19-21]. Some of these values are used by Aihara [13] in resonance energy determination of heteroconjugated systems.

On this basis the values of W and $\overline{W}$ are calculated for a number of purine and pyrimidine bases and their derivatives. The values for parent compounds ,purine and pyrimidine are ,

for purine W = 63.2 and $\overline{W}$ = 1.755

and for pyrimidine W = 21.0 and $\overline{W}$ = 1.400

The purines and pyrimidines in nucleic acids are derivatives of the parent compounds . Adenine and guanine are the derivatives of parent compound purine. The pyrimidine derivatives commonly found in nucleic acids are , cytosine , uracil and thymine. In addition to these there are some unusual bases that occur in specific cases e.g., the methylated cytosines in viral nucleic acids and a variety of methylated bases in transfer RNA . The values of W and $\overline{W}$ for bases in DNA and RNA are reported in table I ,

TABLE I

Topological distances of bases in DNA and RNA

Base	W	$\overline{W}$
Adenine	85.5	1.900
Guanine	116.1	2.110
Cytosine	50.7	1.810
Uracil	54.6	1.913
Thymine	76.6	2.127

For a number of purines and pyrimidines we have calculated W and $\overline{W}$ values and correlated with HOMO energies LUMO - HOMO energy gaps and resonance energies[22]. On the bases of results of these correlations following conclusions can be made,

a) Mean topological distances varies almost linearly with HOMO energy values (in β units) . This indicates a structural depandance of pi-energies and hence of ionization potentials and chemical reactivity.

b) Purines and pyrimidine bases have almost same order of structural regularity in terms of topological distances.

c) For pairs of adenine -thymine or cytosine-guanine sum of the W values is of the same order showing the structural complimentarity of these pairs in DNA molecules,

$$W_A + W_T \simeq W_C + W_G$$

This forms a topology model of DNA structure.

d) For DNA bases the difference between W values i.e. Δ W is proportional to the strength between the bonding of AT or CG pairs. Thus Δ W value for complimentary base pairs shows the number of hydrogen bonds between these pairs.

e) In all the bases lactam states have higher values of W and $\bar{W}$ than those of lactim states,

$$W_{lactam} > W_{lactim}$$

The topology model of relative stability in lactam -lactim tautomerism could thus be presented.

f) The order of W and $\bar{W}$ for purines and pyrimidines of DNA and RNA is

$$W_C < W_U < W_T < W_A < W_G$$

If we assume that topological distances vary with the stability this is in agreement with the stability order with respect to resonance energy per pi-electron (REPE) for these bases. The order of REPE is[22],

$$C > U > T \quad \text{and} \quad A > G$$

Concluding Remarks

The results of the topological model proposed in this investigation may provide the basis for suitable topology models of DNA molecule under different experimental situations leading to studies of thermal stabilisation or destabilisation in presence of alkali, alkaline, and transition metal cations. Further work is in progress in this direction[23].

Acknowledgements- Prof. E.V.Krishnamurthy , Department of applied Mathematics , Indian Institute of Science , Bangalore, brought our attention to his recent works for which we are thankful to him.

References

1. H.Hosoya, J.Chem.Documen. 12 ,181(1972).
2. D.H.Rouvray,American Scientist 61, 729 (1973).
3. R.S.Lall, (Communicated) .
4. L.B.Kier and L.H.Hall, ' Molecular Connectivity in Chemistry and Drug Research', Academic ,New York, 1976.
5. W.Conley, Inter.J.Math.Educ.Sci.and Tech. 12, 425(1981).
6. D.Bonchev, Ov.Mekenyan and N.Trinajstić, Inter.J.Quant.Chem. 17, 845 (1980); Ov.Mekenyan, D.Bonchev and N.Trinajstić. Inter.J.Quant.Chem. 19 , 929 (1981).
7. R.S.Lall, Curr.Sci. 50 , 668(1981).
8. R.S.Lall, Curr.Sci. 50, 846 (1981).
9. M.Randić, J.Am.Chem.Soc. 97, 6609 (1975).
10. J.A.MacPhee, A.Ponaye and J.E.Dubois, J.Org.Chem. 45, 1164 (1980).
11. D.H.Rouvray in ' Chemical Applications of Graph Theory', ed. A.T.Balaban, Academic , New York, 1976.
12. A.Graovac ,O.E.Polansky , N.Trinajstić and N.Tyutyulkov. Z.Naturforsch, 30a , 1969(1975).
13. J.Aihara, J.Am.Chem.Soc. 98, 2750, 6840 (1976).
14. A.Steitwieser, Jr. ' Molecular Orbital Theory for Organic Chemists', Wiley , New York , 1961.
15. N.Trinajstić, Croat.Chem.Acta 49, 593(1977).
16. L.B.Kier and L.G.Hall , European J.Medicinal Chem. 12, 307 (1977).
17. I.Motoc and R.Vălceanu, Math.Chem. 8 , 333(1980).
18. H.Wiener, J.Am.Chem.Soc. 69 , 2636 (1947).
19. B.A.Hess,Jr., L.J.Schaad and C.W.Holyoke Jr., Tetrahedron. 28, 3657 (1972).
20. B.A.Hess Jr. and L.J.Schaad, J.Am.Chem.Soc. 95, 3907(1973).
21. B.A.Hess Jr., L.J.Schaad and C.W.Holyoke Jr. , Tetrahedran, 31, 295(1975).
22. B.Pullman and A.Pullman, ' Quantum Biochemistry',Interscience, New York, 1963, Chap.V p. 185.
23. V.K.Srivastava and R.S.Lall (in preparation).