MEDICINAL CHEMISTRY

A SERIES OF MONOGRAPHS
VOLUME 14

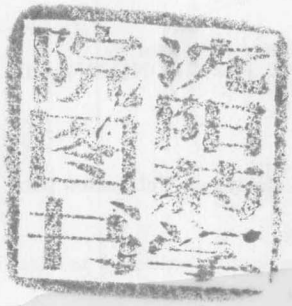
Molecular Connectivity in Chemistry and Drug Research

LEMONT B. KIER

Department of Chemistry
Massachusetts College of Pharmacy
Boston, Massachusetts

LOWELL H. HALL

Department of Chemistry
Eastern Nazarene College
Quincy, Massachusetts



ACADEMIC PRESS
A Subsidiary of Harcourt Brace

ndon 1976

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1

Library of Congress Cataloging in Publication Data

Kier, Lemont Burwell.
Molecular connectivity in chemistry and drug research.

(Medicinal chemistry, a series of monographs)
Includes bibliographical references.
1. Molecular theory. I. Hall, Lowell H., Date joint author. II. Title. III. Series.
QD461.K42 541'.22 76-18696
ISBN 0-12-406560-0

To our fathers

Lemont B. Kier and Lowell H. Hall

for their inspiration, encouragement, and example

PREFACE

The organic or medicinal chemist has, for many years, employed topology in his consideration of molecular structure. A structural formula is, in reality, a topological graph; a skeleton formula is a subgraph; a heterocyclic molecule is depicted with a rooted circuit graph; branched isomers of molecules are distinguished by formulas reflecting different connectivities. There is a well-developed intuition that different molecular structures, described by different topological graphs, have different properties.

Until now this intuition has been qualitative. It is obvious that butane and isobutane structural representations are different. It is not apparent from the structural formula how different they are. Is this difference greater or less than the difference between isopentane and neopentane? In other words, is it possible to assign some numerical value to the graphs of molecules, so that differences in structure could be quantitated? Beyond this, is it possible to differentiate numerically molecular structures sufficiently so that significant correlations are possible with physical, chemical, and biological properties?

These possibilities represent real opportunities to the organic and medicinal chemist in the study of structure–activity relationships. If numerical values could be assigned, or better yet, developed nonempirically, which reflect meaningful aspects of molecular structure, these scientists would have a powerful tool to analyze and predict numerical values of properties of molecules that are of interest to them.

This book describes a new approach to the quantitative evaluation of molecular structure, which we call molecular connectivity. It is a nonempirical derivation of numerical values that encode within them sufficient information to relate to many physicochemical and biological properties. We have discovered these relationships many times.

The method of molecular connectivity is extremely simple, while providing a flexibility to consider important heteroatoms. Furthermore, the method has the inherent ability to describe numerically a molecule at several levels of consideration, each level conveying different information about the connectivity of the molecule. The composite of these extended connectivity values has brought the correlation with some physical properties to a point near the experimental limit of the values.

In this book we will develop the method of molecular connectivity as it has evolved in our laboratory to date. This is followed by a section on the application to physicochemical properties. The next section shows how the method can be applied to structure–activity studies in medicinal chemistry. The final chapter contains some reflections, current challenges, and future areas of investigation of molecular connectivity.

ACKNOWLEDGMENTS

A number of people have contributed significantly to this work and deserve acknowledgment. Our early collaborators, M. Randić and W. J. Murray, made essential contributions during the formative stages of the work. Valuable technical assistance was rendered by J. Fisk, P. Coy, and D. LaLone. Helpful technical discussions were contributed by G. Amidon, A. Cammarata, J. U. Free, R. H. Mann, T. DiPaolo-Chênevert, J. McCloy, W. J. Murray, E. B. Roche, and S. Sickler. Computer assistance was generously supplied by the Eastern Nazarene Computer Center, L. A. Baker, Director, and the Massachusetts College of Pharmacy Computer Center, F. Parmenter, Director. Much of the preliminary draft was typed by D. D. Hall, while the final manuscript was typed by M. L. Kier.

CONTENTS

Chapter Four Molecular Properties and Connectivity

Chapter Five Molar Properties and Molecular Connectivity

Chapter Six Connectivity and Nonspecific Biological Activity

Chapter Seven Substituent Group Structure–Activity Relationships

Chapter Eight Multiple Chi Terms Relating to Biological Activity

Chapter One

STRUCTURE AND PROPERTIES

In chemistry we seek relationships between the fundamental nature of atoms and molecules and their behavior as expressed in experimental quantities. The concept that molecules consist of atoms bound together into stable, identifiable entities has played a vital role in modern chemistry. Physical properties, stability, reactivity, and other characteristics are described and explained in molecular terms. Much creative effort at many levels of theory has been devoted to the development of methods that relate what we know of structure to what we measure as properties.

At the heart of any science is the awareness that changes in composition or structure lead to changes in properties and function. Chemistry is no exception to this rule. Indeed, we are acutely aware of the profound influences that modest structural variation in molecules has upon physical, chemical, and biological properties. As a consequence, a large part of the study of chemistry is devoted to the subject of the definition of structure.

At this time, quantum mechanics is the ultimate approach to the quantification of molecular structure. Given the coordinates and atomic numbers for a collection of atoms, the Schroedinger equation, in principle, can be solved for the eigenvalues and eigenvectors that describe the energy and electron distribution. The stable arrangement of these atoms in molecular form corresponds to the lowest energy arrangement. Other properties of the molecular aggregate are derived from the wave function and energy.

For example, if we were to conduct a sophisticated quantum mechanical calculation of the eigenvalues and eigenvectors of a molecule containing four carbon and ten hydrogen atoms, we must introduce the

numbers and kinds of atoms. Solution would yield the structures of the two most stable isomers, butane and isobutane. Our chemical experience tells us that these are the only stable combinations of these atoms under reasonable conditions.

Our intuition, based on classical notions of valence, would lead us to the same prediction. Normally the chemist could not accurately describe the electronic structure or energies of the two isomers without a quantum mechanical calculation.

There are two levels of structural information concerning a molecule. The complete structure, both electronic and geometrical, is obtained through quantum mechanics. At an intermediate level it is possible to write the structural formulas of isomers based on intuitive notions of chemical bonding. This intermediate level of structural information is the bonding or branching pattern in the molecule. Structural information concerning branching, atom connections, shape, and size can be classified under the general term *topology*.

I. Structural Influences on Physicochemical Properties

Numerous examples are available illustrating the influence of structure on experimental properties of molecules. Some of these are presented in Table I. It is apparent that each property bears a relationship to the molecular structure, although the nature of this relationship is variable. The molecular weight is strictly additive in terms of the numbers and kinds of atoms in the molecule. Additivity is fundamental to the concept of an homologous series in organic chemistry. The heat of atomization in the hydrocarbon series is additive within experimental error, the increment per methylene unit being 280.03 kcal/mole. Molar volume and molar refraction are also perceived to be additive in this series.

In contrast, pure additivity is not found in this series for the properties of boiling point and specific gravity. In each of these cases, the increment between successive members in the homologous series slowly decreases.

In this hydrocarbon homologous series, as well as in others, such properties as heat of atomization and molar refraction show excellent linear correlation with the number of carbon atoms. In contrast, the properties of boiling point and density show a nonlinear correlation.

The relationship between molecular structure and properties is less direct when we consider molecules that are branched. Table II illustrates some of these properties for the isomeric hexanes. None of the

TABLE I

Structural Influences on Selected Properties of Alkane Homologous Series

Compound	Heat of atomization[a]	Molar refraction[b]	Molar volume[c]	Molecular weight	Refractive index[d]	Boiling point[d]	Specific gravity[d]
Butane	1234.96	—	—	58.13	—	−5.0	—
Pentane	1514.80	25.27	115.22	72.15	1.3575	36.07	0.6262
Hexane	1794.72	29.91	130.68	86.17	1.3749	68.74	0.6594
Heptane	2074.75	34.54	146.52	100.19	1.3876	98.43	0.6838
Octane	2354.86	39.19	162.58	114.21	1.3974	125.67	0.7025
Nonane	2634.76	43.83	178.69	128.23	1.4054	150.81	0.7176
Decane	2914.84	48.47	194.84	142.25	1.4119	174.12	0.7301

[a] Heat of atomization in kcal mol^{-1} taken from Cox and Pilcher [19, Table 34].

[b] Molar refraction R_m calculated as $[(n^2 - 1)/(n^2 + 1)](M/d)$, in $cm^3\ mol^{-1}$, where n is refractive index and d density.

[c] Molar volume V_m calculated as M/d, where M is molecular weight and d density.

[d] Data taken from Handbook of Tables for Organic Compound Identification, CRC Press, Cleveland, Ohio.

TABLE II

Structural Influences on Selected Properties of an Alkane Isomeric Series[a]

Compound	Heat of atomization	Molar refraction	Molar volume	Molecular weight	Refractive index	Boiling point	Specific gravity
n-Hexane	1794.72	29.91	130.68	86.17	1.3749	68.74	0.6594
3-Methylpentane	1795.93	29.80	129.72	86.17	1.3765	63.28	0.6643
2-Methylpentane	1796.57	29.95	131.92	86.17	1.3715	60.27	0.6532
2,3-Dimethylbutane	1797.41	29.82	130.25	86.17	1.3750	57.99	0.6616
2,2-Dimethylbutane	1799.28	29.93	132.73	86.17	1.3687	49.74	0.6492

[a] All definitions and data sources are the same as in Table I.

properties is the same for any two isomers, hence number of atoms in the molecule is insufficient to describe all of the salient features of the structure that govern the magnitude of the property.

In a superficial analysis, we have listed the hexanes in increasing order of branching based on our perceived intuition of this structural characteristic. The boiling points are seen to decrease in this order, while the heats of atomization increase. In contrast, the molar volumes, molar refractions, and specific gravities are all apparently poorly correlated with respect to this intuitive ordering of the molecules. In each case, however, the properties have a different value for each isomer. Properties in this series, therefore, depend on structure, but that structural quantitation is not always predictable from simple intuitive notions of degree of branching.

At this point it is possible to presume that a complete quantum mechanical treatment of a series of molecules may not be necessary to develop enough information about structure to correlate with some physicochemical properties. If a quantification of the topology of molecules, which we call *molecular connectivity*, could carry with it

TABLE III

Physical Properties with Limited Dependence on Topology

Compound	Ionization potential (eV)	Base ionization constant
1-Chlorobutane	10.67	
2-Chlorobutane	10.65	
1-Chloro-2-methylpropane	10.66	
2-Chloro-2-methylpropane	10.66	
1-Aminobutane	8.71	10.61
2-Aminobutane	8.70	—
1-Amino-2-methylpropane	8.70	10.72
2-Methyl-2-aminopropane	8.64	10.68
Ethane	11.5	
Propane	11.1	
Butane	8.64	
Isobutane	9.23	
Pentane	10.35	
Isopentane	10.32	
Neopentane	10.35	

sufficient structural information, a close correlation may be possible. This is the objective of the approach described in this book and termed molecular connectivity [1–4].

Certain properties do not have a strong dependence on molecular topology. Ionization potential I_p, arising primarily from a single Schroedinger equation eigenvalue, may be strongly dependent on the presence of a single structural feature. As shown in Table III, chlorobutanes all have essentially the same value for I_p, whereas saturated noncyclic alkanes show a relation to structure similar to those properties in Table I and II. Base dissociation constants for aminobutanes also reveal weak dependence on topology. The information derived from molecular topology may be insufficient to establish a basis for good correlation to these properties. The more complete quantum mechanical approach is required.

II. Applications of Structure Definition

The principal value of structural information, whether it is derived from quantum mechanics or from an intermediate topological level, is the explanation and prediction of physical and chemical properties. This approach, generally termed structure–activity relationship (SAR) studies, has found wide application in chemistry in the prediction of both properties and the course of reactions. A classic example is the use of a numerical value, assigned to an atom or a chemical group to predict its electronic influence on another portion of the molecule. This is exemplified by the Hammett linear free energy relationship. The value, designated σ, is derived from relative values of the pK_a's of aromatic acids, substituted on the ring. The ratio of the K_a values is considered to be a measure of varying electronic influences of the ring substituents.

It should be stated here for the sake of rigorous definition that the σ values of Hammett are ratios derived from one property, used to relate influences on other properties. As Norrington has pointed out, this is an example of property–activity relation (PAR) [5]. The σ value is not a structural characterization but a manifestation of the structure.

Biological properties of interest to medicinal chemists, such as relative potency of drugs, also depend on molecular structure. The topological influence is illustrated by some selected data in Table IV. Nonlinear dependence on the number of atoms is frequently observed. Various patterns of dependence on the degree of branching is also typical. The combination of these two factors has rendered difficult the development of relationships between molecular structure and biological activity.

TABLE IV

Structural Influence on Drug Activity of Selected Alcohols

Compound	log MBC (mM)[a]	pC[b]	log(1/c)[c]	pC[d]
Methanol	3.09			
Ethanol	2.75			
Propanol	2.40			
n-Butanol	1.78			
n-Pentanol	1.20			
n-Hexanol	0.56			
n-Heptanol	0.20			
n-Butanol		1.46	1.42	0.87
Isobutanol		1.54	1.35	—
sec-Butanol		1.16	—	0.60
tert-Butanol		0.98	0.89	0.46

[a] The logarithm of the minimum blocking concentration for nonspecific local anesthesia from D. Agin, L. Hersch, and D. Holtzmann, *Proc. Nat. Sci.* **53,** 952 (1965).

[b] The negative logarithm of concentration from C. Hansch and W. J. Dunn, *J. Pharmacol. Sci.* **61,** 1 (1972).

[c] Relative activity for tadpole narcosis from Overton, *Studies on Narcosis,* Fischer, Jena, Germany (1901).

[d] Relative activity on Madison Fungus from R. H. Baechler, *Proc. Am. Wood Preserv. Assoc.* **43,** 94 (1947).

As an approach to this problem, investigators have sought relationships between experimentally observable properties and biological activity. For example, the partition coefficients between oil and water of series of drug molecules have been used to analyze biological activity [6]. Other properties include the Hammett σ term, molar refractivity, and empirical terms depicting steric influences. These efforts have resulted in some good correlations with biological activity but have not truly achieved a structure definition.

Other studies on drug molecules have considered the electronic structures and reactivity indices derived from approximate quantum mechanical calculations [7]. This is an example of SAR. Unfortunately, the approximate nature of these methods, necessary for large molecules, results in a substantial loss in the information content regarding the structure. As a result this approach is still in its promising infancy.

We have seen from our introductory considerations that quantum mechanics, in principle, gives a complete structural description of a molecule. At an intermediate level, we can consider a molecule as an