

Edited by:

Prof. Dr. Gaston Berthier
Université de Paris

Prof. Dr. Hanns Fischer
Universität Zürich

Prof. Dr. Kenichi Fukui
Kyoto University

Prof. Dr. George G. Hall
University of Nottingham

Prof. Dr. Jürgen Hinze
Universität Bielefeld

Prof. Dr. Joshua Jortner
Tel-Aviv University

Prof. Dr. Werner Kutzelnigg
Universität Bochum

Prof. Dr. Klaus Ruedenberg
Iowa State University

Prof. Dr. Jacopo Tomasi
Università di Pisa

Springer-Verlag Berlin Heidelberg GmbH

R. Carbó-Dorca D. Robert Ll. Amat
X. Gironés E. Besalú

Molecular Quantum Similarity in QSAR and Drug Design



Springer

Authors

Ramon Carbó-Dorca
David Robert
Lluís Amat
Xavier Gironés
Emili Besalú

University of Girona
Institute of Computational Chemistry
Campus Montilivi
1071 Girona, Spain

E-mail: quantum@stark.udg.es

Library of Congress Cataloging-in-Publication Data

Molecular quantum similarity in QSAR and drug design / R. Carbo-Dorca ... [et al.].

p. cm. -- (Lecture notes in chemistry, ISSN 0342-4901 ; 73)

Includes bibliographical references and index.

ISBN 978-3-540-67581-5 ISBN 978-3-642-57273-9 (eBook)

DOI 10.1007/978-3-642-57273-9

1. QSAR (Biochemistry) 2. Quantum biochemistry. 3. Quantum pharmacology. 4. Drugs--Design. I. Carbó, Ramón. II. Series.

QP517.S85 M646 2000
572.8'633--dc21

00-041287

ISSN 0342-4901

ISBN 978-3-540-67581-5

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

© Springer-Verlag Berlin Heidelberg 2000

Originally published by Springer-Verlag Berlin Heidelberg New York in 2000

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera ready by author

Printed on acid-free paper SPIN: 10763155 51/3143/du - 543210

Contents

1	Introduction	1
1.1	Origins and evolution of QSAR	2
1.2	Molecular similarity in QSAR	4
1.3	Scope and contents of the book	4
2	Quantum objects, density functions and quantum similarity measures	6
2.1	Tagged sets and molecular description	6
2.1.1	Boolean tagged sets	7
2.1.2	Functional tagged sets	7
2.1.3	Vector semispaces	8
2.2	Density functions	8
2.3	Quantum objects	10
2.4	Expectation values in Quantum Mechanics	10
2.5	Molecular Quantum Similarity	11
2.6	General definition of molecular quantum similarity measures (MQSM)	11
2.6.1	Overlap MQSM	12
2.6.2	Coulomb MQSM	12
2.7	Quantum self-similarity measures	13
2.8	MQSM as discrete matrix representations of the quantum objects .	13
2.9	Molecular quantum similarity indices (MQSI)	14
2.9.1	The Carbó index	15
2.10	The Atomic Shell Approximation (ASA)	15
2.10.1	Promolecular ASA	16
2.10.2	ASA parameters optimization procedure	17
2.10.3	Example of ASA fitting: adjustment to <i>ab initio</i> atomic densities using a 6-311G basis set	17
2.10.4	Descriptive capacity of ASA	18
2.11	The molecular alignment problem	20
2.11.1	Dependence of MQSM with the relative orientation between two molecules	21
2.11.2	Maximal similarity superposition algorithm	21

2.11.3	Common skeleton recognition: the topo-geometrical superposition algorithm	24
2.11.4	Other molecular alignment methods	25
3	Application of Quantum Similarity to QSAR	26
3.1	Theoretical connection between QS and QSAR	26
3.1.1	Beyond the expectation value	26
3.2	Construction of the predictive model	27
3.2.1	Multilinear regression	28
3.3	Possible alternatives to the multilinear regression	28
3.3.1	Partial least squares (PLS) regression	29
3.3.2	Neural Network algorithms	29
3.4	Parameters to assess the goodness-of-fit	29
3.4.1	The multiple determination coefficient r^2	30
3.4.2	The standard deviation coefficient σ_N	30
3.5	Robustness of the model	31
3.5.1	Cross-validation by leave-one-out	31
3.5.2	The prediction coefficient q^2	31
3.5.3	Influence on the regression results	32
3.6	Study of chance correlations	33
3.6.1	The randomization test	34
3.7	Comparison between the QSAR models based on MQSM and other 2D and 3D QSAR methods	34
3.7.1	Comparison with 2D methods	36
3.7.2	Comparison with 3D methods built on grids	36
3.8	Limitations of the models based on MQSM	37
3.8.1	Homogeneity of the sets	37
3.8.2	The problem of the bioactive conformation	38
3.8.3	Determination of molecular alignment	38
4	Full molecular quantum similarity matrices as QSAR descriptors	39
4.1	Pretreatment for quantum similarity matrices	39
4.1.1	Dimensionality reduction	39
4.1.2	Variable selection	44
4.2	The MQSM-QSAR protocol	46
4.3	Combination of quantum similarity matrices: the tuned QSAR model	46
4.3.1	Mixture of matrices and coefficient constraints	46
4.3.2	Optimization of the convex coefficients	48

4.4	Examples of QSAR analyses from quantum similarity matrices	49
4.4.1	Activity of indole derivatives	49
4.4.2	Aquatic toxicity of substituted benzenes	55
4.4.3	Single-point mutations in the subtilisin enzyme	62
5	Quantum self-similarity measures as QSAR descriptors	67
5.1	Simple QSPR models based on QS-SM	67
5.2	Characterization of classical 2D QSAR descriptors using QS-SM	68
5.2.1	QS-SM as an alternative to log P values	68
5.2.2	QS-SM as an alternative to Hammett σ constant	70
5.3	Description of biological activities using fragment QS-SM	73
5.3.1	Activity against <i>Bacillus cereus</i> ATCC 11778 (Bc)	77
5.3.2	Activity against <i>Streptococcus faecalis</i> ATCC 10541 (Sf)	79
5.3.3	Activity against <i>Staphylococcus aureus</i> ATCC 25178 (Sa)	81
6	Electron-electron repulsion energy as a QSAR descriptor	84
6.1	Connection between the electron-electron repulsion energy and QS-SM	84
6.2	$\langle V_{ee} \rangle$ as a descriptor for simple linear QSAR models	85
6.3	Evaluation of molecular properties using $\langle V_{ee} \rangle$ as a descriptor	86
6.3.1	Inhibition of spore germination by aliphatic alcohols	86
6.3.2	Inhibition of microbial growth by aliphatic alcohols and amines	88
6.3.3	Aquatic toxicity of benzene-type compounds	90
6.3.4	Activity of alkylimidazoles	92
7	Quantum similarity extensions to non-molecular systems: Nuclear Quantum Similarity	96
7.1	Generality of Quantum Similarity for quantum systems	96
7.2	Nuclear Quantum Similarity	96
7.2.1	Nuclear density functions: the Skyrme-Hartree-Fock model	97

7.3	Structure-property relationships in nuclei	99
7.3.1	The nuclear data set	99
7.3.2	The binding energy per nucleon	99
7.3.3	The mass excess	103
7.4	Limitations of the approach	108
References		109

Foreword

The study of the healing power of chemical compounds, present into the known natural active principles and its subsequent use, shall be seen initially as a pseudoscientific procedure, a continuation of the Chinese and Arabian occultism, which was based in the percentage content of fire, air, earth and water, as well as on the associated qualities: hot-cold, humid-dry..., which the matter was supposed to be formed by. The method, possessing roots in Hippocrates, Dioscorides and Galen was studied, described and polished by Avicenna, Averroës, Bacon and Villanova in the Middle Age. It also was appearing as a study constant during the renaissance and after. On the other hand, the birth of chemistry as a scientific offspring from alchemy propitiated alternative ways of knowledge in order to solve the same problem. In this manner, in the past century, approximately hundred years from now, Sylvester proposed the first molecular description in numerical discrete form, employing ideas which even in present times can be associated within the so called molecular topology. Sylvester's topological model can be considered the seed allowing the origin of this big tree, which is now known as theoretical chemistry.

During all the past time from the first topological model of Sylvester up to now, the proliferation of numerical parameters to describe molecular structures has not ceased to grow larger. Some of these parameters have played a very important role for the understanding of the organic molecules behavior and, by extension, for the comprehension and evaluation of their physical as well as biological properties. In the mind of every specialist are the Hammett's σ , the Taft constants or the octanol-water partition coefficient. Other numerical parameters, such as those derived from the modern topological molecular representation are in a process of constant revision and growing. Thus, the Hosoya and Randic indices, or the Kier's connectivities, among several not so well known numerical data are usual reference descriptors. They are put at the researchers' disposition, and are easily deducible from any molecular representation in form of ordered sets of numerical figures. All of them are profusely studied and employed in present times. The main idea consists into the use of these numerical data in order to obtain information on the molecular trends to possess or acquire certain properties and, even better than this, to determine in which degree or intensity molecules present everything.

At the turn of the almost past century, in a broadly manner first, in a refined way after the twenties, just at the dawn of the thirties, quantum mechanics was precluding the possibility to mathematically describe the molecular structures in

the best of complete and precise ways. Quantum theory, based in a physico-mathematical formalism as well as in experimental evidences about the composition and behavior of matter, aroused the hope to find out the definitive procedure for the a priori theoretical evaluation of the molecular properties. In both, the quantum mechanical Schrödinger's description of the hydrogen atom as well as in the relativistic Dirac formalism, there is the seed of the algorithms to find out the definitive manner to study new molecular entities, which even did not to have to be synthesized, in order to analyze theoretically their structure and properties.

Suddenly, the application of quantum mechanics to study molecular structure got an unprecedented help from the first born electronic computers. Their use, thirty years after the Schrödinger works, became a road nevermore abandoned. It is obvious that, after the first computer models offered by UNIVAC and IBM, the future of the solution of Schrödinger's equation was inexorably tied up to numerical procedures, in a great majority already described in the nineteenth century, others waiting to be developed in parallel along the theoretical development of applied quantum mechanics. Everything was connected to the computational needs related to the search for wavefunctions, constituting the solutions of Schrödinger's equation, also called wave equation in remembrance of its origins. According to the quantum mechanical postulates, all the information about a given molecular system in particular, about a quantum object in general, is contained within the so-called system's density (probability) function. The density function is nothing but the squared module of the system's wavefunction, followed by a manipulation to obtain a drastic reduction of the number of implicit function variables. This means that, in order to obtain any information about the molecular properties, even if the molecule is simply constituted by a virtual formula, that information can be obtained from the associated density function. From here the molecular characteristics could be obtained with ease, applying anew quantum mechanical ideas based in turn in well-established theoretical statistical definitions.

In this way, the medieval dream can be considered fulfilled, several centuries after, by means of the conjunction of a physico-mathematical theory and the technology connected with electronic computing. However, to determine the molecular properties, physico-chemical or biological, even in our days presents several problems not exempt of difficulties, whose roots are nearby located into the zone of the unsolvable. In present times, the quantum mechanical parameters are often employed in the same fashion as those coming from topological origins, which were previously referred, without any deeper connection with the quantum mechanical theory than to have its source on Schrödinger equation and wavefunctions. This way to act has an easy explanation in the fact that the wave equation solution and the density function computation were seen, along the computational development of quantum chemistry, as dissociated processes, independent one from another. The appearance of alternative points of view in the nearby time period, culminating with the 1998 Nobel prize award, allow to think

into the leading role the density function has in connection with the Schrödinger equation itself, may be throughout the so called Hohenberg-Kohn theorem or by some extension or variant of it.

It will be certainly early to affirm that the relationship between the molecular structure, their properties and quantum mechanics are completely established and that all details and problems are solved. For the moment it seems out of question that the computational process, associated to the theory, which can be called the Quantum Quantitative Structure-Properties Relationships, it is not related to a unique molecule, but at least in the biological activity environment, has to agglutinate the information simultaneously connected to several molecular structures. From here, the quantum mechanical connection between structure and properties can be unequivocally established by means of the quantum similarity theory, an extension of the quantum theory, developed from a first paper published in 1980. The molecular similarity of quantum origin description is founded in the mathematical structure, which contains the density functions as objects of study, and in extracting the information content from the density function, which it is related in turn with the similarities, which are present within the structures made by microscopic particles. This constitutes an important question, which has been already mentioned as a fundamental element of quantum theory. In practice, the so-called quantum molecular similarity consists in the simple comparison of various density functions, associated to two or more molecules, with as much precision as possible. Such comparison is realized by means of numerical calculations, related with the mathematical concept of measure. The best approximate definition of this comparison process result can be based in the fact that a measure can be considered a sort of generalized volume.

From this starting point can be easily deduced that the equations related with structure-properties relationships can be directly connected with the quantum postulates, in a pure fashion, without additional suppositions. The fashionable empirical structure-properties equations more than a century alive, acquire a legitimization right in this manner and they can be considered from now on as more or less coarse approximations of the theoretical quantum procedure. More than this, from the quantum theory of structure-properties relationships it can be deduced that some empirical parameters of experimental origin, like the octanol-water partition coefficient, shall be strongly correlated with the quantum self-similarity measures: those comparisons where only a density function of a unique molecule intervene. For example, due that it is possible to consider as a self-similarity the quantum Coulomb repulsion potential between the electrons of a molecule, at the same time this molecular property can be considered a substitute of the usual empirical parameters, within molecular series with some degree of homogeneity. Another interesting example, for instance, may be related with the substitution of the classical topological matrices connected to some molecular structure. Such matrices can be refined by means of intramolecular quantum similarity matrices constructed by comparison the atomic electronic densities or these belonging to the polyatomic fragments, which form the molecular structure. Search for pharmacophores can be performed in this last case by means of processes connected to quantum mechanics.

The quantum similarity measures involving one or several density functions cannot be considered uniquely as another step to legitimate previous empirical processes. They constitute a manner, not influenced by the research operator, to find out molecular parameters acting as descriptors, unbiased ordered numerical arrays, obtained without any previous supposition but quantum theoretical considerations. From this initial step, a quantum structure-property relationship can be obtained for any known molecular collection. The quantum similarity measures constitute a basis so general for this specific purpose that they can be used to any microscopic system, from atoms and nuclei up to proteins. In fact the well known atomic periodic table can be deduced as a consequence of the application on the atomic electronic structure of the quantum similarity theory. Of course, quantum similarity permits to obtain a relationship between quantum objects and any associated property, according to the spirit linked to the subjacent quantum theory. It is possible, as a consequence, if one wishes to, to establish molecular periodic tables where, in the same way as in the homologous atomic one, the relationships between the implied molecules permit to obtain information about the behavior of new structures. It is important to repeat now, once more, that also it is feasible to deduce, by using the basic postulates of the quantum mechanical theory only, the form, which have to take the quantum structure-property relationships, and it can be said that this form coincides with the linear appearance employed up to present time within the empirical environment. But this is not the unique similarity between both methodologies.

It must be said, finally, that the evolution of the quantum similarity theory and the numerical tests performed along the past twenty years on a large number of varied molecular systems confirm unequivocally that the theory is correct and completely general. The models deduced from the new theory of the quantum structure-property relationships, by means of the quantum molecular similarity, are comparable in precision, if not better than the empirical results. However, the quantum models differ from the empirical ones in the fact that there is no need to obtain them, any other manner to operate based on any other suppositions, than those most usual among all the possible ones, related with the logical structure of applied quantum mechanics.

All these topics constitute the basis of the present book, and will be developed in detail along the different chapters, illustrating each discussion with real-life application examples, most of them unpublished. Finally, we want to gratefully acknowledge professors Robert Ponec and Paul G. Mezey for their enlightening comments and lively discussions. This work has been partially supported by the CICYT grant SAF-96-0158 and the European Commission contract ENV4-CT97-0508. Thanks also to the *Fundació Maria Francisca de Roviralta* for financial support.

Girona, January 2000