PHARMACO CHEMISTRY LIBRARY

10

QSAR IN DRUG DESIGN AND TOXICOLOGY

EDITED BY

DUŠAN HAĎZI AND BORKA JERMAN-BLAŽIČ

ELSEVIER

PHARMACOCHEMISTRY LIBRARY
Editors: W.Th. Nauta†, R.F. Rekker and H. Timmerman

Volume 10

QSAR IN DRUG DESIGN AND TOXICOLOGY

Proceedings of the Sixth European Symposium on Quantitative Stucture–Activity Relationships, Portorož-Portorose (Yugoslavia), 22–26 September, 1986

Edited by

DUŠAN HAĎZI

Lek — Pharmaceutical and Chemical Works and Boris Kidrič Institute of Chemistry, Ljubljana, Yugoslavia

and

BORKA JERMAN-BLAŽIČ

Institute Jožef Stefan, Ljubljana, Yugoslavia

ELSEVIER — Amsterdam — Oxford — New York — Tokyo 1987

ELSEVIER SCIENCE PUBLISHERS B.V.
Sara Burgerhartstraat 25
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

Distributors for the United States and Canada:

ELSEVIER SCIENCE PUBLISHING COMPANY INC.
52, Vanderbilt Avenue
New York, NY 10017, U.S.A.

ISBN 0-444-42767-8 (Vol. 10)
ISBN 0-444-41564-5 (Series)

Printed in The Netherlands

Preface

The perpetual need for new and improved drugs on the one side and the increasing costs of synthesis and testing on the other are pressing for rational approaches to the design of active molecules. The usefulness of QSAR methods is well proven not only in drug design in general but also in the related fields of pharmacokinetics, toxicology and environmental problems. Classical QSAR approaches have in recent years greatly benefited from developments in statistical methods, computer supported molecular modeling, and theoretical calculations of various molecular parameters. Clearly, an understanding of receptor and enzyme mechanisms is of great importance in guiding developments of QSAR approaches. All this was reflected in the profiles of the previous five QSAR Symposia as well as in the conception of the programme of the 6th European Symposium on Quantitative Structure–Activity Relationships. However, every QSAR symposium should have some individual mark, perhaps reflecting the growing interest in some particular type of drug. Peptide drugs are rapidly coming into focus for well known reasons and are attracting increased interest from pharmacologists and medicinal chemists. Although the design of therapeutically promising peptides differs from the design of "classical" drugs and requires a strong background of theoretical methods of energy minimization and molecular dynamics including hydration effects, it is still susceptible to QSAR methods. Thus a special section was devoted to peptides and weight was given to it by inviting several eminent lecturers.

The programme of the 6th European Symposium on Quantitative Structure–Activity Relationships was clearly attractive enough to gather at Portorož-Portorose 156 delegates from 14 European countries, the U.S.A., and Mexico and to keep their attention throughout 11 invited lectures, 27 regular papers and 36 posters. Most of these contributions are collected in this volume. I hope that it will be of interest to a much wider readership than just the delegates present at the Symposium and those who wished to be there but were unable for various reasons to come. The close interest of the pharmaceutical industry was already apparent from the large number of their delegates at the Symposium and it is hoped that the contents of this volume will be of some use in practical applications.

I should like to thank the members of the international and local organizing committees for advice and assistance, respectively, in organizing the Symposium. I am grateful to the UNESCO Scientific Cooperation Bureau for the European and North American Region, the International Committee on Medicinal Chemistry, the Union of Chemical Societies of Yugoslavia and the Lek Works for sponsorhip, and to the pharmaceutical companies and the Research Council of Slovenia for financial assistance. The secretarial assistance of Mrs. Tatjana Krsmanović was invaluable.

November, 1986

D. Hadži

International Committe

E. Darvas
J.C. Dearden
J.L. Fauchère
R. Franke
D. Hadži
R.F. Rekker
J.K. Seydel
M. Tichý

Local Organizing Committee

D. Hadži
B. Jerman-Blažič
D. Kocjan
J. Kidrič
A. Štalc

CONTENTS

[+]invited lectures

SECTION I
CHEMOMETRICS IN DRUG DESIGN

D. Hadži and B. Jerman-Blažič (Editors), *QSAR in Drug Design and Toxicology*

THE DYLOMMS METHOD: INITIAL RESULTS FROM A COMPARATIVE STUDY OF APPROACHES TO 3D QSAR

Richard D. Cramer III and Jeffrey D. Bunce
Tripos Associates, 6548 Clayton Road, St. Louis, Mo. 63317 (USA)

ABSTRACT

When integrated with contemporary molecular modeling software and partial least squares data analysis (PLS) in a preliminary study of four data sets DYLOMMS shows promise as a very simple and powerful method for mapping structure-activity observations into three-dimensional features of an unknown receptor, applicable to the design of new drug molecules.

INTRODUCTION

Three approaches to 3D-QSAR, the "shape difference" methods of Hopfinger and Motoc (refs. 1,2), the "distance geometry" method of Crippen (ref. 3), and the DYLOMMS method of Cramer and Wise (ref. 4), are being comparatively evaluated, using five diverse series of data, steroid binding to carrier globulins (ref. 5), GABA reuptake and inhibition (ref. 6), ACE inhibition (ref. 7), DHFR inhibition (ref. 8), and certain prostaglandin inhibitors (ref. 9). In preliminary work, a new implementation of the DYLOMMS method has demonstrated great data fitting power, good "predictive" power (via cross-validation), good model communicability, and low cost. Because this method has yet to be thoroughly described in the literature, it is reviewed here.

THE DYLOMMS MODEL OF 3D QSAR

The model rests on three axioms: the structure of most receptors is unknown; key aspects of the physical chemistry needed to interpret even interactions involving known receptors are highly uncertain; nevertheless, observed differences in bio-activity can arise only from computable differences among the tested molecules' possible non-covalent interactions. The following hypotheses then form the heart of the DYLOMMS Approach to 3D-QSAR:

A) The "possible non-covalent interactions" may be represented as the steric (van-der-Waals 6-12) and electrostatic (monopole

Coulombic) interactions of a molecule with a "probe atom".

B) The unknown receptor may be replaced by a regular "lattice map" of grid points.(Fig. 1, top). The probe atom is placed successively at each lattice intersection and its steric and electrostatic interactions with a (suitably aligned) drug molecular are calculated. The resulting energies become a row in a data table (Fig. 1, middle). This process is repeated for each tested molecule, yielding additional rows of the data table.

C) Appropriate multivariate analysis will yield an equation which represents those features of the unknown receptor which are implied by the structure/activity data (Fig. 1, middle, 1st column).

CURRENT IMPLEMENTATION OF THE DYLOMMS MODEL

As the result of recent advances in allied fields, these ideas, over a decade old, now form the basis of a practical tool. A new method of multivariate analysis, partial least·squares (PLS), developed by Wold et al. (ref.10), may in this context be described as the critically needed method for extracting a robust "regression equation" from the QSAR table of Figure 1, with its many more columns than rows. PLS "solves" this apparently underdetermined system of equations by a series of rotations of the independent variable matrix, in each rotation maximizing the commonality between the independent and dependent variable matrices. (In contrast, classical least-squares regression rotates the independent variable columns individually and independently, rather than together, thus consuming a degree of freedom for each coefficient estimated.). Confidence in the robustness of the resulting equation, with its extraordinary number of terms, one for every lattice-point-interaction-class/column in the table, arises from "cross-validation". Following this procedure, the equation is rederived omitting a randomly chosen group of compounds and the potencies of the omitted compounds are "predicted" from the rederived equation. This "cmission", "rederivation" and "prediction" process is repeated until the potency of each compound has been predicted. Thus cross-validation yields a set of residuals, much like those from conventional regression, but resulting from a "prediction" rather than a "best fit" procedure. Comparison of the sum of the square of the cross--validated or "predictive residuals", called the "press", with the original variance of the data yields a "cross-validated r^2", analogous to the conventional "r^2", but always lower in value (indeed

negative in value for poorly modelled data sets). A positive cross-validated r^2 is excellent evidence for a robust model, likely to produce biological predictions more accurate than chance alone for as yet untested molecules.

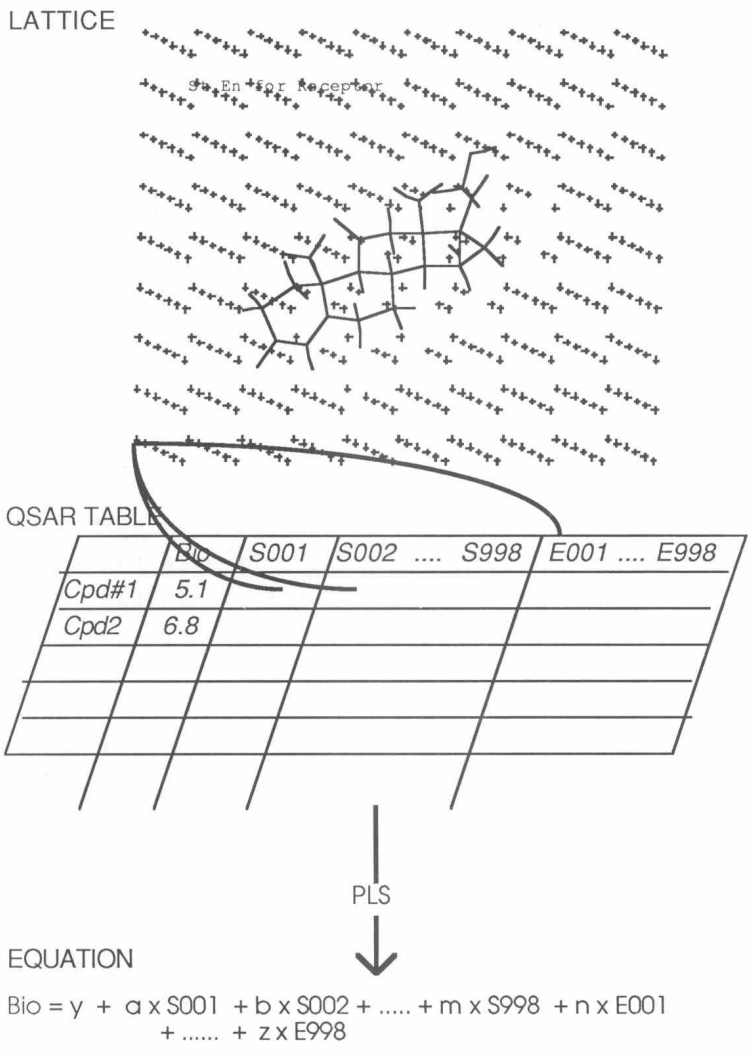

Fig. 1. DYLOMMS method of 3D-QSAR: the internal processes

Advances in the supporting molecular modeling software also make the DYLOMMS approach much more practical. Indeed, a user need have no awareness of the QSAR table or the resulting equation at all. The inputs he must provide, in the form of a "plan", are: at least one molecular model for each compound (multiple conformers can have their individual fields averaged in various ways); an "alignment" rule, specifying how the (rigid) molecules are to be placed in the lattice, and which seems by far the most important variable in deriving a successful model; the "region(s)", that is, a starting point, endpoint, and spacing, in x, y and z, for each lattice; and the force field and probe atom properties to be used for interaction energy calculations.

DYLOMMS outputs are particularly well-suited to graphical presentation. While the equation has many terms, the one-to-one correspondence between a term and a lattice point allows the equation to be presented as an interactive color-coded three-dimensional image, either in the form of a graph visually similar to the top part of Figure 1, the color of a point signifying the magnitude of the corresponding term, or, better, with the term values summarized in contour form, as seen in Figures 2 through 5, to be discussed below.

Using the equation to "predict" the potency of an unsynthesized molecule is straihtforward. The user must provide its molecular model; the software automatically aligns it, computes the field, and plugs the field values into the equation to yield the prediction.

RESULTS

Preliminary results of applying the current implementation of DYLOMMS to four sets of data, involving three series of compounds, are summarized statistically in Table 1.

The generally unimpressive "usual" r^2 values (in parentheses) should actually be reassuring to those concerned about chance correlation with this approach. While one could obtain as high a "usual" r^2 as desired simply by extracting more PLS components, the cross-validation constraint is stringent, and, except for the first instance, at best only one component provides a prediction much better than "the mean of the potencies so far observed". In pursuing this work, one objective is to improve the cross-validated r^2 values for the last three series by refining the associated alignment rules and probably the conformational selection

hypotheses. Indeed, one can envision ways of using the PLS approach itself to optimize the alignment rules for individual compounds (ref. 11). On the other hand, it is possible that each molecule binds uniquely, so that a high cross-validated r^2, from success in predicting its potency from the potencies of others, will not be obtainable. In this situation, one can still obtain as high a conventional r^2 as desired, by including more components in the PLS model. Such a trivial model would not reliably predict the potencies of new molecules, but would provide some information about the regions of receptor space which have not yet been explored synthetically.

TABLE 1

Summary statistics for DYLOMMS analysis of four sets of data

Activity (compounds)	# cpds	# PLS compnts	r^2 pred (usual)	std.err of press(resid)
corticosterone binding globulin (steroids)	20	5	.860 (.992)	.434 (.045)
testosterone binding globulin (steroids)	20	1	.116 (.515)	1.117 (.827)
GABA reuptake inhibition (amino acids)	15	1	-.315 (.536)	1.084 (.577)
ACE inhibitors (dipeptides & analogs)	28	1	.207 (.457)	.753 (.623)

In each of these series, differences in steric fields seem to be more than ten times as important as differences in electrostatic fields in rationalizing the observed potencies. However, this may yet be some artifact of the way the fields are calculated, specifically a "cutoff" of 30.0 kcal/mole which replaces the unrealistically high steric energy calculated when the probe is "inside" an atom of a molecule.

Nevertheless, it is only the steric aspects of the "receptor feature maps" which are shown in Figures 2 through 5. Each Figure contains two panels, each panel in turn showing two orthogonal views of the same object, a "receptor feature map" surrounding a molecule. These "receptor feature maps" are actually contours surrounding those regions whose equation coefficients, multiplied by the variance of the corresponding column in the QSAR table, are higher or lower than a selected cutoff value (for attraction, +1.0; for repulsion, -1.5). In the upper panel, the contours surround

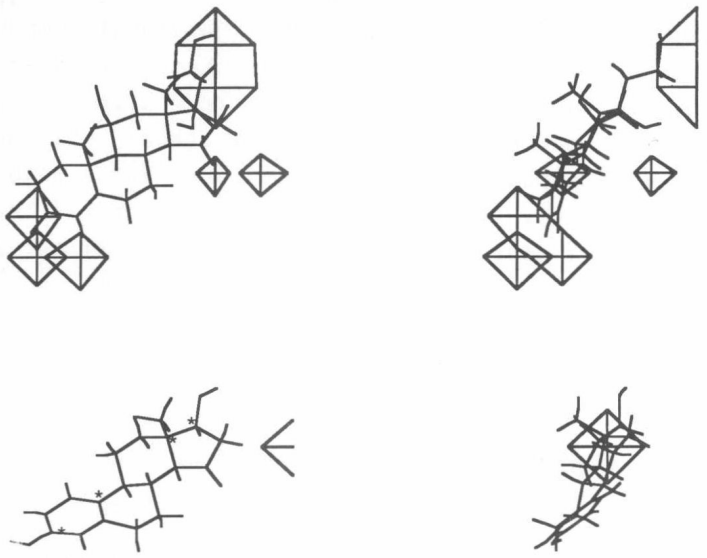

Fig. 2. The most important features of the binding site of corticosteroid binding globulin. See text for explanation.

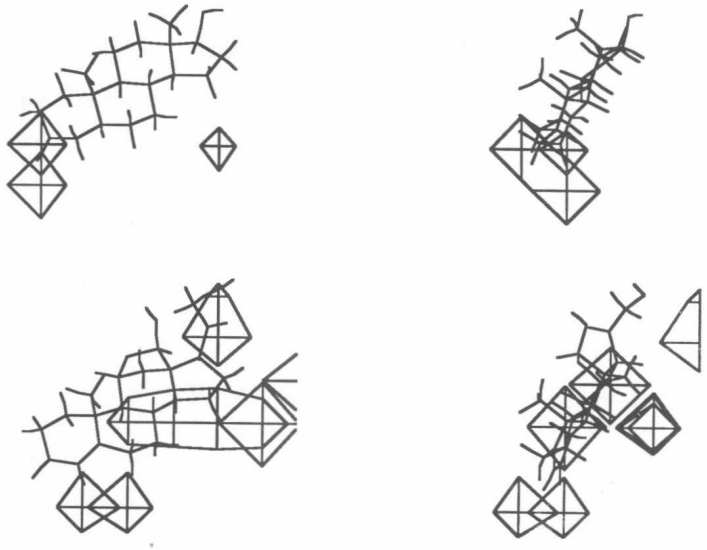

Fig. 3. The most important features of the binding site of testosterone binding globulin. See text for explanation.