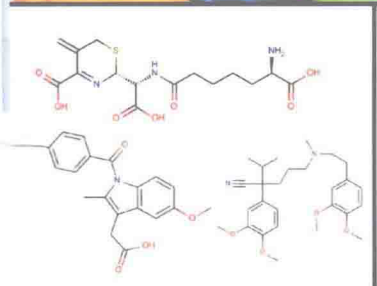
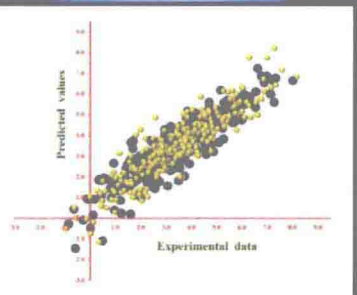
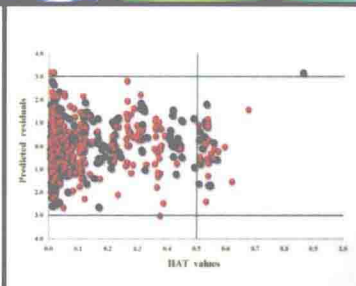
Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment



Response (Y)	X_1	X_2	X_3	...	X_n
1.201	0.522	27.3	1	...	2.399
2.510	0.213	12.8	0	...	5.283
3.583	0.429	50.2	2	...	4.599
6.210	1.201	49.1	3	...	1.963
7.522	1.510	54.3	5	...	3.888
...	...	...	...	...	...
Y_m	...	...	...	...	...





Kunal Roy, Supratik Kar
Rudra Narayan Das

Understanding the BASICS OF QSAR FOR APPLICATIONS IN PHARMACEUTICAL SCIENCES AND RISK ASSESSMENT

KUNAL ROY
Department of Pharmaceutical Technology,
Jadavpur University, Kolkata, West Bengal, India

SUPRATIK KAR
Department of Pharmaceutical Technology,
Jadavpur University, Kolkata, West Bengal, India

RUDRA NARAYAN DAS
Department of Pharmaceutical Technology,
Jadavpur University, Kolkata, West Bengal, India

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
32 Jamestown Road, London NW1 7BY, UK
525 B Street, Suite 1800, San Diego, CA 92101-4495, USA
225 Wyman Street, Waltham, MA 02451, USA
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

ISBN: 978-0-12-801505-6

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress.

Typeset by MPS Limited, Chennai, India
www.adi-mps.com

Printed and bound in the United States of America

Understanding the BASICS OF QSAR FOR APPLICATIONS IN PHARMACEUTICAL SCIENCES AND RISK ASSESSMENT

DEDICATION

In memory of Professor A. U. De

FOREWORD

We are perhaps not far removed from the time when we shall be able to submit the bulk of chemical phenomena to calculation

Joseph-Louis Gay-Lussac, 1808

All of chemistry... would become a branch of mathematical analysis which, taking its constants from observation, would enable us to predict the character of any new compound...

Charles Babbage, 1838

Quantitative structure–activity relationships (QSARs) and quantitative structure–property relationships (QSPRs) have been part of scientific study for many years. As early as 1863, Cros [1] found that the toxicity of alcohols increased with decreasing aqueous solubility. Brown and Fraser [2] recognized in 1868 that "There can be no reasonable doubt but that a relation exists between the physiologic action of a substance and its chemical composition and constitution." Only a few years later, Mills [3] developed mathematical equations (almost certainly the very first QSPRs) to predict the melting and boiling points of several homologous series of chemicals as a function of their chain length.

By far the most widely quoted early work in QSAR is that of Overton [4] and Meyer [5], who independently observed in 1899 that the toxicity of simple organic chemicals to aquatic species increased with their partition coefficient. This discovery was most prescient, for the partitioning process is by far the most important factor controlling the biological activity of chemicals in living organisms.

Despite all this early activity, very little further progress was made in the field until the 1960s, when Corwin Hansch, the acknowledged "father" of QSAR, and his coworkers began to publish research showing that a range of biological activities could be modeled mathematically using simple physicochemical properties. Their first QSAR publication concerned the herbicidal effects of phenoxyacetic acids [6]; this seminal work was the touchstone for most of the thousands of QSAR publications that have followed in the 50+ years since then.

The multiple linear regression equation approach used by Hansch is still employed today for its simplicity and transparency. However, there is now also a plethora of other approaches to QSAR modeling, such as partial least squares, principal component analysis, artificial neural networks, regression trees and random forests, support vector machines, and comparative molecular field analysis (CoMFA), to name but a few. Of course, this is due in part to the huge increases in computing power and speed that we continue to see, as well as to the wide availability of databases and descriptors

(physicochemical, quantum chemical, and structural properties that can be used in a QSAR to model the end point—biological or chemical—under consideration). Since 1999 in particular, there has been an almost exponential rise in the number of QSAR modeling papers published [7], driven in part by the increasing importance of QSAR in environmental and regulatory work.

There are regular conferences dedicated to QSAR, such as the European QSAR Symposia, started by MiloňTichý in Prague in 1973; the International Workshops on QSAR in Environmental and Health Sciences, started by Klaus Kaiser in Hamilton, Ontario, Canada, in 1983; and the Gordon Conferences on QSAR, first organized by Yvonne Martin in Rindge, New Hampshire, in 1975.

Despite all of this activity, very few books have been published specifically on QSAR. Therefore, this book by Roy, Kar, and Das, covering all aspects of QSAR modeling, is greatly to be welcomed. It will be useful to newcomers to QSAR, such as those having to cope with the requirements of the Registration, Evaluation, and Authorization of Chemicals (REACH) legislation, as well as to researchers and seasoned practitioners in pharmaceutical, agricultural, cosmetics, and environmental fields.

John C. Dearden
Emeritus Professor of Medicinal Chemistry,
School of Pharmacy and Biomolecular Sciences,
Liverpool John Moores University, Liverpool, UK

REFERENCES

[1] Cros AFA. Action de l'alcool amylique sur l'organisme (Action of amylalcohol on the organism). Thèse, Faculté de Médecine, Université de Strasbourg, Strasbourg, France; 1863.
[2] Brown AC, Fraser TR. On the connection between chemical constitution and physiological action; with special reference to the physiological action of the salts of the ammonium bases derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia. Trans Roc Soc Edinburgh 1868;25:151–203.
[3] Mills EJ. On melting point and boiling point as related to composition. Phil Mag Ser 1884;5(17): 173–87.
[4] Overton E. Über die allgemeinenosmotischenEigenschaften der Zelle, ihrevermutlichenUrsachen und ihreBedeutungfür die Physiologie (On the general osmotic properties of the cell, their probable origin and their significance for physiology). Vierteljahrsschr Naturforsch Ges Zürich 1899;44: 88–135.
[5] Meyer H. ZurTheorie der Alkoholnarkose. ErsteMitteilung. WelcheEigenschaft der AnastheticabedingtihrenarkotischeWirkung? (On the theory of alcohol narcosis. Part 1. Which property of anaesthetics gives them their narcotic activity?) Naunyn-Schmiedebergs Arch Exp Pathol Pharmakol 1899;42:109–18.
[6] Hansch C, Maloney PP, Fujita T, Muir RM. Correlation of biological activity of phenoxyacetic acids with Hammett constants and partition coefficients. Nature 1962;194:178–80.
[7] Cherkasov A, Muratov E, Fourches D, Varnek A, Baskin I, Cronin MTD, et al. QSAR modeling: Where have you been? Where are you going to? J Med Chem 2014;57:4977–5010.

PREFACE

Quantitative structure–activity relationship (QSAR) modeling, originally evolved from physical organic chemistry, has seen wide application in the fields of medicinal chemistry (for lead optimization) and predictive toxicology (for risk assessment of chemicals). QSAR increases the probability of finding new drug candidates, thus avoiding the synthesis and biological screening of fewer potential molecules and saving time and money. It also helps in screening chemicals for target property or toxicity, thus helping in the prioritization of experimental testing and providing excellent statistical filtering tools. These techniques also have potential applications in other fields, such as materials science, nanoscience, and cosmetic technology for property and activity modeling and predictions. QSAR has now evolved as a well-recognized tool for application in chemistry when a biological activity or property or toxicity is the end point of the study for a series of chemicals of certain degree of structural similarity. The number of QSAR-related publications has grown very rapidly over the last two decades. However, the background knowledge about QSAR has remained confined mainly to researchers in the field and not been disseminated in general to students of and researchers into chemistry and the pharmaceutical sciences. There are many reference texts on advanced topics of QSAR, but there is no good resource covering the basic aspects of QSAR for the beginners in the field. In the present attempt, we have tried to incorporate all the necessary information so that readers may become acquainted with the various areas of the field, at least in a cursory manner.

Chapter 1 deals with the background knowledge and historical evolution of QSAR. Chapter 2 discusses chemical information and introduces the concept of descriptors. In the next chapter, classical QSAR, encompassing the Free–Wilson model, Fujita–Ban modification, the Hansch model, and the mixed approach, all of which have created the foundation of modern QSAR, are covered in detail. Chapter 4 deals with graph theory-based QSAR and topological descriptors. The use of computers in medicinal chemistry, computational chemistry, and related aspects are discussed in Chapter 5. The selected statistical methods used to develop linear and nonlinear QSARs are discussed in Chapter 6. The validation tools used for QSAR models are covered in Chapter 7. Chapter 8 discusses important 3D-QSAR models. Chapter 9 introduces some newer QSAR techniques. Other related (non-QSAR) techniques including docking and pharmacophore mapping used in drug design are dealt with in Chapter 10. Some examples of successful application of QSAR modeling in drug discovery are cited in Chapter 11. The last chapter discusses future avenues

for further application of QSAR. We hope that the readers will get at least a bird's-eye view on different aspects of QSAR and related tools. Readers may consult the original references cited in the book for a more detailed discussion.

In this opportunity, we thank Professor C. Sengupta and Professor J. K. Gupta, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India, for giving us inspiration for writing this book. We must acknowledge the help and inspiration received from our respective family members during the development of the book. We are thankful to the publisher for releasing this book, which we hope will be of use to QSAR learners. We also thank Dr. Kristine Jones and Molly M. McLaughlin of Elsevier for their cooperation.

Kunal Roy
Supratik Kar
Rudra Narayan Das

CONTENTS

CHAPTER 1

Background of QSAR and Historical Developments

Contents

Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment.
ISBN: 978-0-12-801505-6, DOI: http://dx.doi.org/10.1016/B978-0-12-801505-6.00001-6

1.1 INTRODUCTION

Chemicals are essential components of human civilization. The applications of chemicals in the modern era span a wide range, from industrial to household environments. The esteemed goal of a chemist lies in the development of chemicals with the desired profile of their better behavioral manifestation; that is, activity/property. In cases of chemicals with pronounced biological activity or drugs, the chemist aims in enhancing the efficacy of the molecule while reducing any toxic effects exerted. Now, following the general axiom, development of a new chemical can be achieved either by designing a complete new entity or by modifying an existing one. In both cases, chemists need to possess sufficient knowledge regarding the nature of the chemical and its potential for interaction with the biological system. It is obvious that the biological activity (including toxicity) and property of any chemical (e.g., drugs, pharmaceuticals, and carcinogens) depend upon its interaction with the biological system concerned. Hence, the primary goal of a chemical designer lies in establishing a rational explanation of the mechanism of action of the chemical, which can lead to the derivation of a suitable theoretical basis and thus enabling the tailoring of its structure. Sometimes the lack of suitable explanation for a drug action limits the derivation of a hypothetical basis. In order to establish a proper theoretical basis for the action of chemicals, one would need to have a deep insight into the fundamental chemistry of the system that controls its physicochemical and structural behavior [1].

The biological activity elicited by chemicals is attributed by various interactions of the molecule at the critical reaction site in the biological system. The ligand molecule is recognized by a particular receptor followed by formation of a ligand–receptor complex involving different physicochemical forces. The complex then undergoes some conformational changes that lead to a series of events giving rise to the activity. The features possessed by a chemical entity get modified during its encounter with the biological system. Hence, elicitation of biological response is controlled by the way in which the chemical reacts at the active site of the biological system. In other words, the biological system plays a crucial role in determining the structural features of a chemical needed to elicit a desired response. Owing to the limited available space for this chapter, we shall not elaborate on this issue; instead, we would like to present a brief overview of the important steps and interactions involved at the biological reaction site when a chemical comes in contact with a biological system.

Before eliciting a desired response, drugs and other bioactive molecules are subjected to a complex path inside the biological system governed by their pharmacokinetic and pharmacodynamic behavior. Following administration, the drug molecules are subjected to the uncertainties arising from absorption, metabolism, excretion, and the *random walk* toward the critical reaction site, where it binds with a suitable receptor molecule. Figure 1.1 depicts different stages of pharmacokinetic movement of a drug