Ellis Horwood series in chemical science

COMPUTER AIDS TO CHEMISTRY

Editors:

G. VERNIN

and M. CHANON

COMPUTER AIDS TO CHEMISTRY

IMPORTANT NEW & FORTHCOMING BOOKS IN CHEMICAL SCIENCE

JANET E. ASH, et al. CELLULOSE CHEMISTRY AND ITS APPLICATIONS

T. P. NEVELL, UMIST and S. HAIG ZERONION, University of California

COMPUTATIONAL METHODS FOR CHEMISTS A. F. CARLEY and P. H. MORGAN, University College, Cardiff REDUCTIONS IN ORGANIC CHEMISTRY M. HUDLICKY, Virginia Polytechnic Institute and State University

BIOHALOGENATION: Principles, Basic Roles and Applications

S. L. NEIDLEMAN and J. GEIGERT, Cetus Corporation, Emeryville, California ORGANOSILICON AND BIOORGANOSILICON CHEMISTRY: Structure. Bonding. Reactivity and Synthetic Application Editor. H. SAKURAI, Tohoku University, Sendai, Japan COMPUTER AIDS TO CHEMISTRY
Editors: G. VERNIN and M. CHANON, Université de Droit d'Economie et des Sciences, Marseilles, France HANDBOOK OF AQUEOUS ELECTROLYTE SOLUTIONS A. HORVATH, ICI; Mond Division, Runcorn, Cheshare BIOINORGANIC CHEMISTRY R. W. HAY, University of Stirling ORGANIC PHOTOCHEMISTRY J. D. COYLE, Open University PHYSICAL PHOTOCHEMISTRY R. DEVONSHIRE, University of Sheffield INORGANIC PHOTOCHEMISTRY
A. HARRIMAN, The Royal Institution PHOTOCHEMICAL AND PHOTOELECTRICAL CONVERSION OF SOLAR ENERGY A. MILLS, University College, Swansea and J DARWENT, University of London MATHEMATICAL AND COMPUTATIONAL CONCEPTS IN CHEMISTRY N. TRINAJSTIC, Rugjer Boskovic Institute, Zagreb CHEMICAL MONITORING OF OCCUPATIONAL TOXICITY V. FOA, et al., University of Milan HANDBOOK OF LABORATORY WASTE DISPOSAL M. J. PITT, Loughborough University of Technology and E PITT, University of Aston in Birmingham COMMUNICATION, STORAGE AND RETRIEVAL OF CHEMICAL INFORMATION

COMPUTER AIDS TO CHEMISTRY

Editors:

G. VERNIN, M.Sc., Ph.D. Director of Research, National Centre of Scientific Research Marseilles, France

and

M. CHANON
Professor of Chemistry, Université de Droit et des Sciences
Marseilles, France



Publishers Chichester

Press: a division of WILEY & SONS
New York · Chichester · Brisbane · Toronto

First published in 1986 by ELLIS HORWOOD LIMITED

Market Cross House, Cooper Street, Chichester, West Sussex, PO19 1EB, England

The publisher's colophon is reproduced from James Gillison's drawing of the ancient Market Cross, Chichester.

Distributors:

Australia and New Zealand:
Jacaranda-Wiley Ltd., Jacaranda Press,
JOHN WILEY & SONS INC.
GPO Box 859, Brisbane, Queensland 4001, Australia
Canada:
JOHN WILEY & SONS CANADA LIMITED
22 Worcester Road, Rexdale, Ontario, Canada
Europe and Africa:
JOHN WILEY & SONS LIMITED
Baffins Lane, Chichester, West Sussex, England

Baffins Lane, Chichester, West Sussex, England North and South America and the rest of the world: Halsted Press: a division of JOHN WILEY & SONS 605 Third Avenue, New York, NY 10158, USA



© 1986 G. Vernin and M. Chanon/Ellis Horwood Limited

British Library Cataloguing in Publication Data

Computer aids to chemistry —
(Ellis Horwood series in chemical science)
1. Chemistry — Data processing
I. Vernin, G II Chanon, M
542'.8 QD39 3 E46

Library of Congress Card No. 86-10452

ISBN 0-85312-774-3 (Ellis Horwood Limited) ISBN 0-470-20338-2 (Halsted Press)

Phototypeset in Times by Ellis Horwood Limited Printed in Great Britain by The Camelot Press, Southampton

COPYRIGHT NOTICE

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the permission of Ellis Horwood Limited, Market Cross House, Cooper Street, Chichester, West Sussex, England.

Table of Contents

- 0. 00		METZGER	
Introductio	n		1
Chapter I	CO	MPUTER-AIDED ORGANIC	
. •	SY	NTHESIS (CAOS)	
		Barone and M. Chanon	19
	I.1	Introduction	1
	1.2	Research Groups Involved in Computer-Aided	-
		Synthesis	2
	I.3	*	2
		3.1 Connect chemist's and computer's	_
		capabilities	2
		3.1.1 Representations of molecules.	2
		3.1.2 Reaction representations and storage	3
		3.1.2.1 The 'classical' approaches of	-
		transformations	3
		3.1.2.2 Theoretical approaches to	
		reactions	4
		3.2 Retrosynthetic and synthetic approaches	5
	I.4		
		Synthesis	6
		4.1 Strategy and tactics	6

				and pruning.		62 73
	1.5			Computer-Aided Organic Synthesis		73
it.	1.5					73
				omputers and super computers.		73
						81
				ories	•	82
		5.3 F	iow to	properly 'feed' a CAOS program. rison of CAOS and the abilities of a	•	02
						84
				e breakthroughs as seen by workers	•	04
						85
	1.			OS	•	
	1.6			eviations	•	86 89
	I.7	Kelere	ences		•	67
Chapter II	TE	ACHIN	AG CE	IEMISTRY WITH		
Спарка п			7	JTERS		
	D.	Cabrol	. C. Ca	chet, and R. Cornelius		103
	II.1			on		103
	11.2			outers as Teaching and Learning		
	11.2	Too	le			105
				ducational environment		105
		2.2		ecteristics of microcomputers		107
		4.4		Computational capabilities		108
			222	Data storage and retrieval		108
				Decision-making		109
				Interaction		109
1				Graphics		110
				Electrical input/output		112
	II.3	Ann				113
	11.5	, дрр 3.1	Saftu	are styles	•	113
		3.1	2 1 1	The interaction plane.	•	114
					•	•••
			3.1.2	student.		117
			3.1.3	_		117
		3.2		are for student use	•	118
		3.2		Page turners.	•	118
				Drill and practice	•	119
			2.2.2	Tutorial dialogs	•	121
					•	124
•			3.2.4	Microcomputers in the teaching	•	124
			3.2.5	laboratory		130
			226	Problem-solving approach.		133
					.•	136
			3.2.7	Artificial intelligence approach		138
		3.3				138
			3.3.1	Instructional aids		138
			447	MUCTOCOMPULETS AS DEISOHAI (OOIS		120

FR 1 4	-	Contents	
	-		
1	•		

<u>,</u>	Table of Contents	7
	3.3.3 Computer-assisted test construction	139
į	3.3.4 Computer-managed instruction	139
II.4	Authoring Languages and Available Resources.	140
W.	4.1 Authoring languages	140
. ,	4.2 Available software	142
	4.3 Sources of information	143
11.5	Development of Educational Software	145
	Conclusion	148
II.7	References	149
	IPUTER GRAPHICS:	
	EW TOOL IN CHEMICAL EDUCATION	
J. We	ber	154
£, III'1		154
III.2		155
	2.1 Introduction to graphics	155
	2.2 Graphics hardware	156
	2.2.1 Calligraphic (or vector) system.	156
	2.2.2 Raster-scan system	157
	2.2.3 Direct-view storage tube	157
	2.3 Graphics software	157
III.3	Recent Applications of Computer Graphics .o	
	Chemical Education.	159
	3.1 Low-cost graphics systems.	159
	3.2 Sophisticated graphics systems	159
111.4	Future Developments in Computer Graphics	
	Applied to Chemical Education.	162
111.5	Conclusion	165
III.6		166
Chapter IV KINI	ETICS	
	Larice	169
. IV.1	Introduction	169
IV.2	Stoichiometry and Reaction Schemes	170
	2.1 Introduction	170
	2.2 Matter conservation in chemical	
	reactions.	170
	2.3 The case of non-independent equations .	171
	2.4 Finding the number of degrees of	
	freedom	172
1	2.5 Choice of solution sets: slave and master	
17	compounds	174
	2.6 Significance of the failure of the method.	176
	2.7 Conclusion	177

IV.3	Revi	ew of I	Kinetic Rules	177
	3.1	Introd	luction	177
	3.2	Defini	tion of rate and the laws of	
			os	178
	3.3		ated forms of the kinetic equations	178
	3.4		mination of the order for a simple	
			on	179
		341	Differential method	179
		3.4.2	Measurement of the half-reaction	• • • • • • • • • • • • • • • • • • • •
		J. 1.2	time	180
		3.4.3	Direct Treatment of	100
		3.4.3	time-concentration couples	180
	3.5	Comp	lex reactions that reduce to simple	104
	5.5		ons	180
	3.6	Other	complex reactions	181
IV.4		tment	of Kinetic Data	181
14.7	4.1	Introd	luction	181
	4.2		square estimation problem	182
	4.2		Simple linear regression	182
			Linear regression method.	182
				183
		4.2.3	Estimation of quality of fit Linear transformation	185
			Other mean square applications	185
	4.2		near regression	186
	4.3		Linearization of the residue	186
		4.3.1	Parisher antimination by the	100
			Residue optimization by the	187
			gradient method	10/
		4.3.3		188
		<i>-</i> 1	search method	188
*** *	4.4		usion	189
IV.5			nulation by Numerical Integration.	
		Introd	luction	189
	5.2		t of a numerical integration	189
•	5.3	Integr	ration methods	190 190
			Euler's method	
		5.3.2	Other single step methods	190
			Multistep methods	191
			Extrapolation methods	191
	5.4		racy of the numerical integration	191
	5.5	Stiff s	ystems	192
	5.6		e of the method	193
	5.7		mentation of these methods on a	101
			l computer	
	5.8		f a complete integration package	194
	5.9		erical integration and other	404
			lizations	194
	5 10	Stoch	astic methods	195

	Table of Contents	9
	IV.6 Conclusion	195
	IV.7 References	196
Chapter V	MULTIVARIATE ANALYSIS OF CHEMICAL	
	DATA SET WITH FACTORIAL METHODS	
	J. R. Llinas and J. M. Ruiz,	200
<u> </u>	V.1 Introduction	200
	V.2 Matrix Calculations and Multivariate Data	
	Analysis	202
	2.1 Basic elements for matrix calculations	203
	2.1.1 Data translations	205
	2.1.2 Data rotations	205
	2.2 Information reduction	208
	2.3 Determination of a vector subspace which	
	fits data set	209
	2.3.1 Constraints on axis	212
	2.3.2 Properties of the $(X - \overline{X})(X - \overline{X})$	
	matrix in multidimensional statistics	212
	2.3.3 Properties of the $(X - \overline{X})'(X - \overline{X})$	
	matrix linked to the fact that the	
	matrix is square, symmetric and	
	positive	214
	2.4 Quantification of information	215
	V.3 Practical Factor Analysis (FA)	218
	3.1 The data matrix. Variable transformation.	218
	3.2 Diagonalization of $(x)'(x)$ matrix	219
	3.3 Reproduction of the initial data matrix.	221
	3.3.1 Determination of principal factors .	222
	3.3.2 Significance of the number of	
	principal factors	224
	3.3.3 Example of an application to	
	quantitative analysis	225
	3.4 Physical significance of factors	228
	V.4 Principal Component Analysis (PCA)	229
	4.1 Objectives and methods	230
	4.2 Graphic representation of variables	230
	4.3 Graphic representation of objects	232
	4.4 Application example	234
	4.5 Scope and limitations	239
	V.5 Factor Discriminant Analysis (FDA)	240
	5.1 From canonical analysis to discriminant	1
	analysis	240
	5.2 Theoretical basis for discriminant analysis	241
	5.3 Method of calculation	243
	5.4 Discriminant factors	244
	5.5 Graphical representations	244

		5.6	Assignn	nent of an object to a group	245
				tion example	246
		5.84 5	Scope a	nd limitations	250
	V.6	Gener	al conc	dusion	251
	V.7	Refere	nces .		253
	V.8				255
		11/			
Chapter VI	CON	иРит	ER AI	DS TO CRYSTALLOGRAPHY	
					257
	VI 1	Intr	oductio	on	257
				and Methods in Crystallography.	257
				etric considerations \	257
		2.2		nteraction between X-rays and	
		2.2		ls	259
				The Bragg law	259
		×.		The structure factors and the	237
		0	L. h. h	electron density function	261
		22	Mash	ods for solving the phase problem	262
		2.3			202
			2.3.1	F	263
*			000	molecule	264
	377.3	D:00	2.3.2	How to obtain a phasing model .	267
	VI.3			eters and the Collection of Data	267
		3.1		al description	268
				The goniometer	269
				The computer.	209
		3.2		ol routines and crystallographic	271
			-	ams	2/1
			3.2.1	—	
				orientation and of the unit cell	271
				parameters	271
	4	_		Automatic data collection	272
	VI.4			stems in X-Ray Structure Analysis	213
		4.1		ifferent steps in structure	273
				nination	
		4.2		reduction	273
				er calculation	274
		4.4		methods.	276
				General Considerations	276
				The symbolic addition method.	276
			4.4.3		255
				method: MULTAN program	277
		4.5		nethod of least squares refinement	278
			4.5.1	Formulation	278
			4.5.2	Comments	279
			4.5.3	Choice of weights	280
			4.5.4	Rigid body refinement	281
		4.6		iption of the molecular	
			confor	rmation	281

		Table of Contents	11
		4.6.1 The conformational parameters .	281
		4.6.2 Graphic illustrations	282
	4.	-	
		4.7.1 Mini or large computers?	282
		4.7.2 System development in	مثمام
		crystallography	283
		onclusion	286
	VI.6 R	eferences	287
Chapter VII	MASS	SPECTRA AND KOVATS' INDICES	
	DATA	BANK OF VOLATILE AROMA	
	COM	POUNDS	
	G. Ve	nin, M. Petitjean, J. C. Poite and J. Metzger	
		uisse and K-N. Suon	. 294
	VII.1	Introduction	. 294
	VII.2	Various approaches for interpreting mass	
		spectrometry data	. 295
		2.1 Theoretical spectrum interpretation.	. 295
		2.2 Pattern recognition	. 297
		2.3 Library searches	299
	VII.3		
		3.1 Various search algorithms	
		3.2 Mass spectra abbreviation	305
		3.3 Quality criterion for mass spectra .	. 307
		3.4 Mass spectral acquisition and validation	. 308
	VII.4	Kovats' Indices Databank Construction .	310
	A 11:4	4.1 Choice of chromatographic data	. 310
		4.2 Properties of Kovats' indices	. 311
		4.3 Choice of a comparison method	. 314
	VII.5	SPECMA Databank Realization	. 314
	V 11.0	5.1 File structures	. 315
		5.2 Program structure	. 317
•		5.3 Databank utilization	319
	VII.6	Applications	. 323
	VII.7	=	323
	VII.8	Addendum	323
		References	. 324
			총
Chapter VIII	ONL	NE ACCESS TO CHEMICAL	420.4
		RMATION	
		ssanaly and H. J. M. Dou	334
	VIII.1		. 334
		1.1 Importance of chemical information	
		1.2 Growth of chemical information	. 335
	VIII 2	From hard conv to online access	335

/•.

j.

Table of Contents

VIII.3	How to Make Information Available in an	
	Interactive Form	336
	3.1 Reference	336
	3.2 Batch access	337
	3.3 Interactive searches	337
	3.4 Boolean operators	338
	3.5 Example	340
VIII.4	Chemical Abstracts Text Files	341
	4.1 General presentation	342
	4.2 Structure of references	343
	4.3 Survey of the reference fields	343
	4.4 Strategies and languages	344
	4.5 Hosts and formats	345
•	4.6 Various examples	345
	4.7 Access to the full abstracts	347
VIII.5		349
, 111.0	5.1 Chemical dictionaries	349
	5.2 Telesystemes DARC	350
	5.3 CAS online	351
VIII.6	The Patent Files: Description and	
V 111.U	Examples	353
	6.1 Patents in Chemical Abstracts	353
	6.2 Patents in WPI, WPIL (Derwent)	355
	6.3 Patents in USP	355
	6.4 Patents in INPI	355
VIII.7	Other Files Dealing with Chemistry or	333
V 111./	Related Topics	359
	7.1 The dissemination of scientific	339
		359
	information.	337
	7.2 How to deal with large information	360
	sources	360
	7.3 Specialized databases available	
	7.3.1 Environline	361
	7.3.2 NTIS	362 363
	7.3.3 CRDS	
		365
	7.3.5 Citation index	368
VIII.8	Downloading	368
VIII.9	Cost of Online Access to Chemical	260
	Information	369
	9.1 Invoicing structure	369 369
	9.1.1 Communication network	309
	9.1.2 Connect time with the	250
	computer	370
	9.1.3 References printed	370
	9.2 Textual files	370
	9.3 Structural files	370

Table of Contents		13
9.4 Equipment		370
9.5 The SDI		371
VIII.10 Conclusion		371
VIII.11 References		371
Recent Literature on the Use of Computers in		
Chemistry	•	373
Subject Index		375
Author Index		393

ij

.

•

Foreword

The development over the past twenty years of big computers has accelerated calculations in many fields—quantum chemistry, molecular mechanics, computer-assisted chemical synthesis are only some typical examples. A new surge in the contribution made by computers to progress in chemistry is the result of more recent commercialization of the high-performance personal computers (PCs). Thanks to its relatively low price and easy availability, such equipment has become commonplace in most chemistry laboratories throughout the world. The purpose of this book is to show chemists, with the aid of judiciously selected examples, how such recently developed and easily available techniques can be of invaluable assistance.

The eight chapters deal with such varied topics as organic synthesis, chemical education, chemical kinetics, analysis of chemical data, crystallography, structural analysis through data banks, and chemical information. Although far from exhaustive, this list is quite representative of the types of aid that microcomputers can offer to chemists. The presentation of the mathematical basis of these computational methods is reasonably condensed, allowing the average chemist to understand their development and applications. The contributors are experimental chemists applying computational methods to their own fields of research or teaching which is at least one reason why this text is so naturally assimilable.

This excellent book fills a gap in chemical education and will without doubt be of valuable help to a large population of chemists in research, teaching and industry. I am quite confident it will do just this.

Introduction

Chemistry is an evergreen subject. This is so because chemistry has roots in conceptual fields which range from quantum mechanics to the molecular organization of life, and because chemistry deals with such a wide variety of problems in diverse fields, such as analytical, synthetic, physical, and industrial, living as well as inanimate matter, solid state as well as solution and gas phase systems, and from the simplest molecules to the most complex structures. Consequently, advances in nearly any area of scientific endeavor has an effect on chemistry, and vice versa.

One of the more recent revolutions to affect the field of chemistry is the advancement of spectroscopic techniques such that they have become ubiquitous in the everyday life of the chemist. Although far from over, this spectroscopic revolution is already undergoing its con revolution as a consequence of a new advance affecting the lives of all chemists. The new advance is, of course, the effect of the computer sciences on chemistry, made possible by the commercial availability of computers with ever-increasing capabilities and steadily-decreasing prices. Thirty years ago, computers were already known and used, but perhaps 90% or more of the chemists then working functioned quite well without them. However, this is no longer the case; within the next twenty years, quite the opposite will no doubt hold true.

The goal of this book is to show why the swing to computer aids to chemistry is occurring as well as examples of how. Since we accept the multifaceted nature of the field of chemistry, we have selected some representative, yet quite different, facets which exemplify the changes computers have made in the activity of chemists.

Of the eight chapters, three (Chapters IV, V, and VI) treat typical quantitative data. They have been written so as to be understandable to any chemist willing to understand how a computer makes such fundamental tools as X-ray Crystallography, Kinetics, and Data Analysis/Experimental Design easier to use. Chemists often complain of being inundated by information from all directions and thus being unable to keep abreast of current topics either in general or specific areas of chemistry. Three chapters, one devoted to information access (Chapter VIII) and two treating the specific areas of spectrometry databases (Chapter VII) and synthetic organic chemistry (Chapter I) bring good news in these areas. The computer can ease the burden of keeping abreast of the wave of chemical information not only by efficient information management, but also in the 'pretreatment' of information for a given chemical speciality.

Finally, each of us is familiar with our own process of learning chemistry, both as a student and throughout our careers. Two chapters describe how computers are affecting this aspect of our lives as chemists. Chapter II summarizes some of the exciting uses of computers as complements to the teaching and learning of chemistry. Chapter III deals more specifically with computer graphics and its effects far beyond activities concerned only with teaching. Chemistry has often been the battlefield and proving ground for mathematical models of all sorts. It was often fashionable to denigrate the chemist's intuition (which is somewhat impressionistic) in favor of the precision of mathematicians. The cross-fertilization of these approaches was strongly inhibited, since even though the chemists had an idea of the phenomena they had extracted from their laboratory experiments, they often felt unable to understand the mathematics involved in the elaboration of various models. With modern computer graphics this situation changes, since every mathematical model can be transformed into something tangible and observable. Thus, the problem of communication is solved, and the chemist can say to the mathematician, 'Although I do not fully understand the mathematics involved, your model does (or does not) fit my knowledge and understanding of my experiments, in such and such respects'.

Early in his history (actually in pre-history), Man realized that tools, adapted to specific tasks, extended the usefulness and efficiency of his own hands and fingers. Now, at the close of the twentieth century and on the verge of the twenty-first, it becomes apparent that a tool is now being introduced to extend the usefulness and efficiency of Man's own brain. This is a fantastic advance, and chemists should not attempt to avoid it. However, a good tool in awkward and inexperienced hands does a poor job. Thus, although each chapter attempts to show what computers have brought to a subject, we also try to illustrate in which ways the human brain must direct and decide, drawing comparisons between the abilities of man and the computer as often as possible.

This book is illustrative rather than exhaustive, as the extent of the topic precludes even an attempt at exhaustivity. One still may ask, 'Why no chapters dealing with quantum chemistry, infrared and n.m.r. databases, drug design, instrumentation interfacing, robotics, etc.?' Our only answer is the