WILEY-VCH

Roberto Todeschini, Viviana Consonni

Handbook of Molecular Descriptors

Methods and Principles in Medicinal Chemistry

Volume 11

Edited by
R. Mannhold,
H. Kubinyi,
H. Timmerman

Handbook of Molecular Descriptors

Roberto Todeschini and Viviana Consonni

WILEY-VCH

Methods and Principles in Medicinal Chemistry

Edited by
R. Mannhold
H. Kubinyi
H. Timmerman

Handbook of Molecular Descriptors

Roberto Todeschini and Viviana Consonni

WILEY-VCH

Weinheim · New York · Chichester · Brisbane · Singapore · Toronto

Series Editors:
Prof. Dr. Raimund Mannhold
Biomedical Research Center
Molecular Drug Research Group
Heinrich-Heine-Universität
Universitätsstraße 1
40225 Düsseldorf
Germany

Prof. Dr. Hugo Kubinyi
Combinatorial Chemistry
and Molecular Modelling
ZHF/G, A 30
BASF AG
67056 Ludwigshafen
Germany

Prof. Dr. Hendrik Timmerman
Faculty of Chemistry
Dept. of Pharmacochemistry
Free University of Amsterdam
De Boelelaan 1083
1081 HV Amsterdam
The Netherlands

Authors:
Prof. Dr. R. Todeschini
Dr. V. Consonni
Milano Chemometrics and QSAR Research Group
Dip. di Scienze dell'Ambiente e del Territorio
Universita degli Studi Di Milano-Bicocca
Piazza della Scienza 1
20126 Milano
Italy

Cover: Staatliche Museen zu Berlin – Preußischer Kulturbesitz
Vorderasiatisches Museum

Library of Congress Card No. applied for

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

Die Deutsche Bibliothek – CIP-Cataloguing-in-Publication Data:
A catalogue record for this book is available from Die Deutsche Bibliothek

ISBN 3-52-29913-0

Composition: Kühn & Weyh, D-79111 Freiburg
Printing: Betzdruck GmbH, D-63291 Darmstadt
Bookbinding: Osswald & Co., D-67433 Neustadt (Weinstraße)

Printed in the Federal Republic of Germany.

To our loved ones
R.T. and V.C.

Any alternative viewpoint with a different emphasis leads to an inequivalent description. There is only one reality but there are many viewpoints. It would be very narrow-minded to use only one: we have to learn to be able to imagine several.

Hans Primas

In: Chemistry, Quantum Mechanics and Reductionism
(Springer-Verlag, 1981)

Cover

Dragon
Babylon Ištar Gate (600 – 500 B.C.)
(Pergamon Museum of Berlin)

A molecular descriptor can be thought of as a mythological animal having several different meanings which depend on one's point of view.

Preface

In the late 1930s, the Hammett equation marked a breakthrough in the understanding of organic chemistry. It describes rate and equilibrium constants of the reactions of aromatic acids, phenols and anilines, as well as other compounds, in a quantitative manner, by using reaction and substituent parameters, ρ and σ. In the same manner, the lipophilicity parameter π, derived from *n*-octanol/water partition coefficients by Corwin Hansch, led to a breakthrough in quantitative structure-activity relationships (QSAR) in biology. Like σ, π also is an additive constitutive molecular property. Their combination, later also with molar refractivity values MR, Taft's steric E_s values or the Verloop steric parameters, allowed the derivation of quantitative models for many biological *in vitro* activity values. Nonlinear lipophilicity models describe *in vivo* biological activities, where substance transport through membranes and distribution within the biological systems play an important role.

Practical problems in the estimation of the lipophilicity of araliphatic and aliphatic compounds led to the *f* hydrophobicity scales of Rekker and Leo/Hansch. However, all such descriptor scales depend on experimental determinations. New molecular descriptors were developed from scratch, starting with the work of Randic, Kier and Hall, i.e. the various molecular connectivity parameters χ. Later the electrotopological state parameters and the Todeschini WHIM parameters were added. Whereas topological descriptors are mathematical constructs that have no unique chemical meaning, they are clearly related to some physicochemical properties and are suited to the description of compound similarities in a quantitative manner. Thus, despite several critical comments in the past, they are now relatively widely used in QSAR studies. Only a meaningless and excessive application in quantitative models, as far as the number of tested and included variables is concerned, still deserves criticism.

This book is a long-awaited monograph on the various properties and molecular descriptors that are of importance in studies of chemical, physicochemical, and biological properties. It is a must for every research worker who is active in this field because it provides an encyclopedic overview of all known descriptors, whether they are physicochemical or topological in their nature. An exhaustive list of references points the way to the original literature.

The series editors wish this book a wide distribution. It is up to the reader to find out which of the properties and descriptors might be most suitable for describing the data. However, the early warnings by Corwin Hansch, John Topliss, and others should not be forgotten: make your model as simple as possible; test and include only a few parameters; try to achieve an understanding of your model; use a test set to check the external predictivity of your model. Molecular descriptors are powerful tools in QSAR studies – but their abuse may lead nowhere. May this book further contribute to their selective and proper use!

August 2000

Raimund Mannhold, Düsseldorf
Hugo Kubinyi, Ludwigshafen
Hendrik Timmerman, Amsterdam

Contents

Introduction

The effort being made today to organize Knowledge is also a way of participating in the evolution of Knowledge itself. In fact, the significance of attempting such organization can be looked for in its ability not only to give information but also to create know-how: it provides not only a collection of facts – a store of information – but also a contribution to the evolution of Knowledge. The fact is that splitting the organization of Knowledge from its production is completely arbitrary: actually, Knowledge organization is itself one way of doing research.

The true end of an encyclopedic guide is to contribute to the growth of knowledge, but not to knowledge given once and for all, based on some final basic theories, but as a *network of models* in progress. A network primarily consists of knots, i.e. objects, facts, theories, statements, and models, the links between the knots being relationships, comparisons, differences, and analogies: such a network of models is something more than a collection of facts, resulting in a powerful engine for analogical reasoning.

With these purposes in mind, the Authors conceived this *Handbook of Molecular Descriptors* as an encyclopedic guide to molecular descriptors.

First, let us look at the definition of molecular descriptor.

The molecular descriptor is the final result of a logical and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into an useful number or the result of some standardized experiment.

Attention is paid to the term "useful" with its double meaning: it means that the number can give more insight into the *interpretation* of the molecular properties *and/or* is able to take part in a model for the *prediction* of some interesting property of other molecules.

Why must we also accept "or"?

A fundamental task is how to predict and understand experimental facts, i.e. physico-chemical properties, biological activities, and environmental behaviour, from symbolic representations of real objects such as molecules by molecular descriptors.

Because of the huge complexity of this problem, it must be understood that during the development of new descriptors, their interpretation can be weak, provisional, or completely lacking, but their predictive ability or usefulness in application to actual problems can be a strong motive for their use. On the other hand, descriptors with poor predictive power can be usefully retained in models when they are well theoretically founded and interpretable because of their ability to encode structural chemical information.

The incompletely realized comprehension of the chemical information provided by molecular descriptors cannot be systematically ascribed to weakness in the descriptors. Actually, our inability to reduce descriptor meanings to well-established chemical concepts is often because newly emergent concepts need new terms in the language and new hierarchically connected levels for scientific explanation. Thus, what is often considered as scientific failure is sometimes the key to new useful knowledge.

In any case, all the molecular descriptors must contain, to varying extents, chemical information, must satisfy some basic invariance properties and general requirements, and must be derived from well-established procedures which enable molecular descriptors to be calculated for any set of molecules. It is obvious – almost trivial – that a single descriptor or a small number of numbers cannot wholly represent the molecular complexity or model all the physico-chemical responses and biological in-

teractions. As a consequence, although we must get used to living with approximate models (*nothing is perfect*!), we have to keep in mind that "approximate" is not a synonym of "useless".

The field of molecular descriptors is strongly interdisciplinary and involves a mass of different theories. For the definition of molecular descriptors, a knowledge of algebra, graph theory, information theory, computational chemistry, and theories of organic reactivity and physical chemistry is usually required, although at different levels. For the use of the molecular descriptors, a knowledge of statistics, chemometrics, and the principles of the QSAR/QSPR approaches is necessary in addition to the specific knowledge of the problem. Moreover, programming and sophisticated software and hardware are often inseparable fellow-travellers of the researcher in this field.

The historical development of molecular descriptors reflects some of the distinctive characteristics of the most creative scientists, i.e. their capability of being at the same time engaged and/or detached, rational and/or quirky, serious and/or not so serious. Science is a game and the best players appreciate not only the beauty of a discovery by precise and logical reasoning, but also the taste of making a guess, of proposing eccentric hypotheses, of being doubtful and uncertain when confronted by new and complex problems. Molecular descriptors constitute a field where the most diverse strategies for scientific discovery can be found.

Molecular descriptors will probably play an increasing role in scientific growth. In fact, the availability of large numbers of theoretical descriptors containing diverse sources of chemical information would be useful to better understand relationships between molecular structure and experimental evidence, also taking advantage of more and more powerful methods, computational algorithms and fast computers. However, as before, deductive reasoning and analogy, theoretical statements and hazardous hypotheses, determination and perplexity still remain fundamental tools.

The *Handbook of Molecular Descriptors* tries to meet the great interest that the scientific community is showing in this topic. In fact, as well as the solid interest in the quantitative modelling of biological activity, physico-chemical properties, and environmental behaviour of compounds, an increasing interest has been shown by the scientific community in recent years in the fields of combinatorial chemistry, high-throughput screening, substructural analysis, and similarity searching, for which several approaches which are particularly suitable for informatic treatment have been proposed. Thus, several disciplines such as chemistry, pharmacology, environmental sciences, drug design, toxicology, and quality control for health and safety derive great advantages from these methodologies in their scientific and technological development.

Although experimental measurements would be the most direct way of obtaining safe and high-quality information, the time, costs and hazards involved in the experimentation are relevant limiting factors if we wish to entrust our knowledge growth to fully experiment-based strategies. Moreover, and most importanly, experiments are always suggested by theories, and theoretical models are the way we try to understand reality. Therefore, both general and local models play a fundamental role in scientific growth, leading to a better understanding of the properties of studied phenomena.

The *Handbook of Molecular Descriptors* collects the definitions, formulas and short comments of molecular descriptors known in chemical literature, our intention being to consider all the known molecular descriptors. The definitions of technical terms, around 1800 in all, are organized in alphabetical order. The importance of a molecular descriptor definition is not related to its length. Only a few old descriptors, abandoned

or demonstrated as wrong, have been intentionally left out to avoid confusion. An effort was also made to collect bibliographic information appropriate for the work proposed in this Handbook. We are sorry if any relevant descriptor and/or work has been missed out; this has not be done deliberately and we take full responsibility for any omissions.

Many molecular descriptors have been grouped into classes using a mixed taxonomy based on different points of view, in keeping with the leading idea of the Handbook to promote learning by comparison. Descriptors have been often distinguished by their *physico-chemical meaning* or the specific *mathematical tool* used for their calculation.

Some basic concepts and definitions of statistics, chemometrics, algebra, graph theory, similarity/diversity, which are fundamental tools in the development and application of molecular descriptors, are also presented in the Handbook in some detail. More attention has been paid to information content, multivariate correlation, model complexity, variable selection, and parameters for model quality estimation, as these are the characteristic components of modern QSAR/QSPR modelling.

The Handbook contains nothing about the combinatorial algorithms for the generation and enumeration of chemical graphs, the basic principles of statistics, algorithms for descriptor calculations, or experimental techniques for measuring physicochemical and biological responses. Moreover, relevant chemometric methods such as Partial Least Squares regression (PLS) and other regression methods, classification methods, cluster analysis, and artificial neural networks, which are also widely applied on molecular descriptors, are simply cited; references are given, but no theoretical aspect is presented. Analogously, computational chemistry methods are only quoted as important tools for calculations, but no claim is made here to their detailed explanation.

Information exchange

The Authors would be grateful to all researchers who would like to send their observations and comments on the Handbook contents, information about new descriptors, and bibliographic references. For e-mail submissions address to: *moldes@disat.unimib. it.*

The Authors are activating a website (*http://disat.unimib.it/chm/*) where recent progress in the molecular descriptor field will be reported, tables of calculated descriptor values for standard data sets collected, and links to other related websites proposed. All information received from interested and collaborative researchers will be useful to develop this Internet tool for sharing ideas and experiences on molecular descriptors and QSAR.

Bibliographic references

The reference list covers a period between 1858 and 2000, and includes 3300 references, about 3000 authors and 250 periodicals. The symbol [R] at the end of a reference denotes a publication with a significant list of references.

Acknowledgements

The idea of producing a *Handbook of Molecular Descriptors* was welcomed by several colleagues whom we warmly thank for their suggestions, revisions, bibliographic information, and moral support. We particularly thank Alexander Balaban,

Subhash Basak, Laura Belvisi, Pierre-Alain Carrupt, Claudio Chiorboli, Johann Gasteiger, Paola Gramatica, Peter Jurs, Lemont Kier, Douglas Klein, Hugo Kubinyi, Silvia Lanteri, Alessandro Maiocchi, Marjana Novic, Demetrio Pitea, Lionello Pogliani, Milan Randic, Gianfranco Tantardini, Bernard Testa, Giuseppe Triacchini, and Jure Zupan.

We are also grateful to Bracco SPA of Milan, EniRicerche of Milan, and the National Institute of Chemistry of Ljiubliana for support in the bibliographic search.

Roberto Todeschini and Viviana Consonni

Milano, June, 2000

User's Guide

This handbook consists of definitions of technical terms in alphabetical order, each technical term being an *entry* of the Handbook.

Each topic is organized in a hierarchical fashion. By following cross-references (→ and typeset in italics) one can easily find all the entries pertaining to a topic even if they are not located together. Starting from the topic name itself, one is referred to more and more specific entries in a top-down manner.

Each entry begins with an entry line.

There are three different kinds of entries: *regular*, *referenced*, and *synonym*.

A **regular entry** has its definition immediately after the entry line. A regular entry is typeset in bold face; it is followed by its (ACRONYM and/or SYMBOL), if any, and by its (*: synonyms*), if any. For example:

Wiener index *(W) (: Wiener number)*

A **referenced entry** has its definition in the text of another entry indicated by the symbol → and typeset in bold face. For example:

Wiener orthogonal operator → **algebraic operators**

A **synonym entry** is followed by the symbol " *:* " and its synonym typeset in italics. To find the definition of a synonym entry, if the synonym is a regular entry, one goes directly to the text under the entry line of the synonym word; otherwise, if the synonym is a referenced entry, one goes to the text of the entry indicated by → , typeset in bold face letters. For example:

Wiener number *: Wiener index*

walk number *: molecular walk count* → **walk counts**

The text of a regular entry may include the definition of one or more referenced entries highlighted in bold face. When there are many referenced entries collected under one regular entry, called a "mega" entry, they are often organized in hierarchical fashion, denoting them by the symbol •. The sub-entries can be in either alphabetic or logical order. For example, in the mega entry "steric descriptors", the first sub-entries, each followed by the corresponding text, are:

- **gravitational indices**
- **Kier steric descriptor**
- **Austel branching index**

Finally, a referenced entry within a sub-entry has its definition in the text of the sub-entry denoted by the symbol (⊙ ...). For example:

WHIM shape → **WHIM descriptors** (⊙ global WHIM descriptors)

indicates that the index "WHIM shape" is defined in the sub-entry "global WHIM descriptors" of the main entry "WHIM descriptors".

In the text of a regular entry one is referred to other relevant terms by words in italics indicated by → . In order to reach a complete view of the studied topic, we highly recommend reading also the definitions of these words in conjunction with the original entry. For example:

count descriptors

These are simple molecular descriptors based on counting the defined elements of a compound. The most common chemical count descriptors are → *atom number A*, → *bond number B*, → *cyclomatic number C*, → *H-bond acceptor index* and → *H-bond donor index* counts, → *distance-counting descriptors*, → *path counts*, → *walk counts*. ...

Finally, words in italics not indicated by → in the text of a main entry (or sub-entry) denote relevant terms for the topic which are not further explained or whose definition is reported in a successive part of the same entry.

The symbol 🕮 at the end of each entry denotes a list of suggested bibliographic references.

We have made a special effort to keep mathematical notation simple and uniform. A collection of the most often appearing symbols are in the next paragraph Notations and Symbols. Moreover, a list of acronyms helps to decipher and locate the full terminologies given in the book.

Notations and Symbols

The notations and symbols used in the Handbook are listed below. In some cases, notations slightly different from those proposed by the Authors are used to avoid confusion with other descriptors and quantities.

Objects

X	molecular descriptor
M	molecule, compound
$\mathcal{M}$	experimental measure
$\mathcal{P}$	experimental property
G	graph, molecular graph
MG	multigraph, molecular multigraph

Sets

V	set of vertices of a graph
E	set of edges of a graph
F	set of fragments of a molecule partition
G	set of points in a 3D grid
${}^{m}P_{ij}$	set of atoms of the path of order m from the ith to the jth atoms
${}^{m}\mathcal{P}$	set of paths of order m

Counts

A	number of atoms of a molecule
B	number of bonds of a molecule
C	number of cycles of a molecule
C^{+}	number of cycles with overlapping of a molecule
G	number of equivalence classes
h_a	number of hydrogens bonded to the atom a
L	principal quantum numbers
n	number of objects, data, molecules
n_x	number of elements with an x-value
N	generic number of elements
N_X	number of atoms, groups, fragments of X-type
M	number of significant principal components or latent variables
p	number of variables
${}^{m}P$	number of paths of length m
P	total number of paths of a graph
Z_a	atomic number of the atom a

Matrix operators

C	column sum operator
D	diagonal operator
R	row sum operator
S	total sum operator
W	Wiener operator

Indices and characteristic symbols

a	index on the atoms of a molecule
b	index on the bonds of a molecule
g	index on the equivalence classes
i,j,k,l,f,m	generic indices
x,y,z	geometric coordinates
d	data distances
d	topological distances
r	geometric distances
δ	vertex degree
δ^b	bond vertex degree
δ^v	valence vertex degree
ε	edge degree
η	atom eccentricity
π	bond order
σ	vertex distance degree
m	order of a descriptor, exponent
w	weights, atom properties
p	probability
q	atomic charge
ℓ_{jm}	PCA loadings of the mth component for the jth variable
λ	eigenvalue
$\lambda_j(\mathbf{M})$	jth eigenvalue from the matrix **M**
$[\mathbf{M}]_{ij}$, m_{ij}	i-j element of the matrix **M**
t_{im}	ith score of the mth component from PCA or PLS
$\mathbf{t}_m$	mth vector score
v	generic column vector
$\mathbf{p}_i$	ith grid point of coordinates (x,y,z)
a	vector of atoms in a path
D	dimension (0,1,2,3)
D	diameter
R	radius
I	binary or indicator variable
I	information content

Acronyms

The most well-known acronyms used to define research fields, methods, and molecular descriptors are listed below, in alphabetical order.

AAA	Active Analog Approach
AAC	Augmented Atom Codes
AID	Atomic ID number
ANN	Artificial Neural Networks
ATS	Autocorrelation of a Topological Structure
AWC	Atomic Walk Count
BP-ANN	Back-Propagation Artificial Neural Networks
BIC	Bonding Information Content
BID	Balaban ID number
BLOGP	Bodor LOG*P*
CADD	Computer-Aided Drug Design
CAMD	Computer-Aided Molecular Design
CAMM	Computer-Aided Molecular Modelling
CASE	Computer-Automated Structure Evaluation
CHEMICALC	Combined Handling of Estimation Methods Intended for Completely Automated Log*P* Calculation
CIC	Complementary Information Content
CID	Connectivity ID number
CLOGP	Calculated LOG*P*
CoMFA	Comparative Molecular Field Analysis
CoMMA	Comparative Molecular Moment Analysis
CoMSIA	Comparative Molecular Similarity Indices Analysis
COSV	Common Overlap Steric Volume
CP-ANN	Counter-Propagation Kohonen Artificial Neural Networks
CPK	Corey-Pauling-Koltun volume
CPSA	Charged Partial Surface Areas
CR	Continuum Regression
CSA	Cluster Significance Analysis
DARC	Description, Acquisition, Retrieval Computer system
DD	Drug Design
DFT	Density Functional Theory
DG	Distance Geometry
EAID	Extended Adjacency ID number
EC	Extended Connectivity
ECA	Extended Connectivity Algorithm
ECI	Electronic Charge Index
EEVA	Electronic EigenVAlue descriptors
EVA	EigenVAlue descriptors
FEVA	First EigenValue Algorithm
FW	Free-Wilson analysis
GA	Genetic Algorithms
GAI	General a_N-Index
GA-VSS	Genetic Algorithms – Variable Subset Selection
GCSA	Generalized Cluster Significance Analysis
GERM	Genetically Evolved Receptor Models

GFA	Genetic Function Approximation
GIPF	General Interaction Properties Function
GOLPE	Generating Optimal Linear PLS Estimations
G-WHIM	Grid-Weighted Holistic Invariant Molecular descriptors
HASL	Hypothetical Active Site Lattice
HBA	Hydrogen Bond Acceptor
HBD	Hydrogen Bond Donor
HQSAR	Hologram QSAR
HFED	Hydration Free Energy Density
HINT	Hydrophatic INTeractions
HOC	Hierarchically Ordered extended Connectivity
HOMO	Highest Occupied Molecular Orbital
HSA	Hydrated Surface Area
HXID	Hu-Xu ID number
IC	neighbourhood Information Content
ILGS	Iterated Line Graph Sequence
ILS	Intermediate Least Squares regression
ISA	Isotropic Surface Area
IVEC	Iterative Vertex and Edge Centricity algorithm
IVS-PLS	Interactive Variable Selection – Partial Least Squares
K-ANN	Kohonen Artificial Neural Networks
KLOGP	Klopman LOG*P*
LASRR	Linear Aromatic Substituent Reactivity Relationships
LFER	Linear Free Energy Relationships
LSER	Linear Solvation Energy Relationships
LOMO	Lowest Occupied Molecular Orbital
LOVI	LOcal Vertex Invariant
LSER	Linear Solvation Energy Relationships
LUMO	Lowest Unoccupied Molecular Orbital
MCD	MonteCarlo version of MTD
MCIs	Molecular Connectivity Indices
MCS	Maximum Common Substructure
MEP	Molecular Electrostatic Potential
MFTA	Molecular Field Topology Analysis
MID	Molecular ID number
MLP	Molecular Lipophilicity Potential
MQSI	Molecular Quantum Similarity Indices
MQSM	Molecular Quantum Similarity Measures
MSA	Molecular Shape Analysis
MSD	Minimal Steric Difference
MSG	Molecular SuperGraph
MTD	Minimal Topological Difference
MTI	Molecular Topological Index
MUSEUM	MUtation and SElection Uncover Models
MWC	Molecular Walk Count
NN	Neural Networks
OASIS	Optimized Approach based on Structural Indices Set
OLS	Ordinary Least Squares regression
PAR	Property-Activity Relationships
PCA	Principal Component Analysis

PCR	Principal Component Regression
PELCO	Pérturbation d'un Environnement Limité Concentrique Ordonné
PID	Prime ID number
PLS	Partial Least Squares regression
PSA	Polar Surface Area
QSAR	Quantitative Structure-Activity Relationships
QShAR	Quantitative Shape-Activity Relationships
QSiAR	Quantitative Similarity-Activity Relationships
QSPR	Quantitative Structure-Property Relationships
QSRC	Quantitative Structure/Response Correlations
QSRR	Quantitative Structure-Reactivity Relationships
RID	Ring ID number
RR	Ridge Regression
RBSM	Receptor Binding Site Model
RSM	Receptor Surface Model
SA	Surface Area
SAR	Structure-Activity Relationships
SASA	Solvent-Accessible Surface Area
SAVOL	Solvent-Accessible VOLume
SBL	Smallest Binary Label
SIBIS	Steric Interactions in BIological Systems
SIC	Structural Information Content
SID	Self-returning ID number
SPP	Submolecular Polarity Parameter
SPR	Structure-Property Relationships
SRC	Structure/Response Correlations
SRW	Self-Returning Walk
SWIM	Spectral Weighted Invariant Molecular descriptors
SWM	Spectral Weighted Molecular signals
SWR	StepWise Regression
TI	Topological Index
TIC	neighbourhood Total Information Content
TLP	Topological Lipophilicity Potential
TLSER	Theoretical Linear Solvation Energy Relationships
TMSA	Total Molecular Surface Area
TOSS-MODE	TOpological SubStructure MOlecular DEsign
VFA	Voronoi Field Analysis
VR	Variable Reduction
VS	Variable Selection
VSS	Variable Subset Selection
WHIM	Weighted Holistic Invariant Molecular descriptors
WID	Weighted ID number
WLN	Wiswesser Line-formula Notation
3D-MoRSE	3D-Molecule Representation of Structures based on Electron diffraction descriptors