Available online at www.joac.info

ISSN: 2278-1862

Journal of Applicable Chemistry

2015, 4 (4): 1024-1144

(International Peer Reviewed Journal)

Critical Review (Cr)

ChemoInformatics
Part I: Molecular Descriptors in Omnimetrics Research

K RamaKrishna[1], Ch. V. Kameswara Rao[2] and R. Sambasiva Rao[3*]

1. Department of Chemistry, Gitam Institute of Science, Gitam University, Visakhapatnam, AP, **INDIA**
2. Department of Chemistry, Basic sciences and Humanities, GMRIT, Rajam 532 127, AP, **INDIA**
3. School of Chemistry, Andhra University, Visakhapatnam 530 003, **INDIA**

Email: karipeddirk@gmail.com, kamesh.chembolu@gmail.com, rsr.chem@gmail.com

Accepted on 11th July 2015

(Dedicated with profound respects to RSR's father R. Venkata Ramanaiah during his birth centenary year)

Conspectus

Background: In mid eighteenth century, the observation of increase toxicity of alcohols to mammals with decrease in solubility in water was the start of a new discipline - structure activity or specifically structure-toxicity relationships. The seminal contributions of Hammett in relating chemical reactivity (logarithm of rate constant/equilibrium constants) with change in substituent groups in aromatic acids and Hansch modelling of effective/lethal dose of compounds with partition coefficient during 1950 and 1970s initiated 'Linear free energy relationships (LFER)' and 'Quantitative structure activity relation ships (QSAR)'.

Molecular descriptors: The protocol of inspection of chemicals for their reactivity at the start up begins at looking into functional groups, cyclic rings, substituents and 2D-connectivity-structure. Constituent descriptors are simple number counts of atoms/bonds/rings in a molecule throwing light on compositional profile. The connectivity (symmetric) matrix of zeros or ones called adjacency matrix leads to distance matrix (of same size) of path was the foundation stone of topological chemistry.

Graph theory: Wiener index, column or row sum of adjacency matrix, proposed in 1947 for explaining boiling points of acyclic organic molecules was the start of mathematical (graph theoretical) chemistry and emergence of molecular descriptor research, now a coveted discipline. The continual advances of Wiener graph continues even in 2015 and occupying a pivotal position in application arena. Randic index and several distance derived descriptors viz. Hosoya, Rouvray, Detour and Zagreb topological indices are in vogue.

Electrostatic-, quantum-, Eigen- paradigms: Kier and Hall introduced electrostatic descriptors of local category based on non-quantum chemical ground. These charge distribution quantities use partial charge, topo electronic measure and surface area. The next major category involves 3D-optimized geometry available from X-ray/ neutron diffraction/ NMR experiments or computational (ab initio, DFT) quantum chemical approaches. Geometric molecular descriptors like shape/shadow indices speak out molecular size and shape. The core of WHIM and its clones is based on chemically valid 3D-geometry of

a molecule. 3D-MoRSE, RDF_Wtd descriptors found extensive applications. The quantum chemical descriptors are well known CQC derived IP, EA, charges, multi-pole moments, electronic energies and thermodynamic quantities. The chemical hardness/softness, Fuki's parameters shed light on radical / electrophilic/nucleophilic reactions. Eigen matrix analysis has a key role in several molecular descriptor categories primarily to get rid of correlation among descriptors. WHIM descriptors derived from Eigen values of 3D-weighted co-variance matrix encode shape, size and atom distribution. GWHIM takes into account of grid reflecting interaction field dimensions. The minimum and maximum Eigen values of atomic charge, polarizability or H-bonding ability instead of atomic number of diagonal matrix of Burden matrix and Eigen spectra of CQC-generated UV-Vis, IR, NMR profiles are coveted first-order probes per molecule. Many of topological and other types also were subjected to Eigen structure enabling use of classical hard least squares analysis. Information descriptors based on Shannon information theory mirrors mass distribution in the molecules. Bond Environment Descriptor (BED) and Augmented Environment Descriptor (AED) from first and second nearest neighbors and their bonds account for effect of fragments around an active atom.

Hybrid descriptors: The hybridization of descriptors of different origin opened a super-highway embracing the best of both information worlds to predict complex responses in real life materialistic compositions. The components of GETAWAY descriptors are geometric and distance matrix based one. The Topo-Geometric/ Electro-topological state atom (E-state)/ Topo-QC descriptors have high impact in statistical modelling and interpretation perspectives. The classical physico-chemical-biological parameters like logP, refractivity, Lipinski (rule of 5) alert index are also now in the band-wagon of molecular descriptors.

Applications: The applications of molecular descriptors over the last two decades engulfed almost all research disciplines under an umbrella term structure activity relationships (SAR). It comprises of computations of molecular descriptors (simple counting to advanced CQC, graph theory, matrix operations), pruning (selection) to a subset based on statistical, information and process-details basis, modelling strategies (MLR, PCR, NN, SVR etc.), validation procedures (residual analysis), ensemble/forest approaches. The target sciences of focus are chemistry, biology, pharmacy, material synthesis, environment are in the priority list. The typical subclasses of SXR include 'activeSite, AntiOxidant, Blood brain barrier, binding, complexation, drug, genetics, HIV/AntiHIV, inhibition, interaction, $Log\beta_{mlh}$, pharmacokinetic, protein-Ligand-interaction, receptor and toxicity'. This qualifies these cutting edge investigations to be referred as "computational descriptors in Omni_metrics", in equilibrium with Chemometrics, Chemoinformatics, Mathematical chemistry, CQC and nature inspired algorithms.

Molecular descriptors from trustworthy packages for sizable number of compounds from academic research stand point have been used for HIV-1 integrase inhibition, activity against HIV-1 reverse transcriptase, Myorelaxant, Antimalarial activity, Serine protease active site, Ocular toxicity, Opiate/ antinidatory activity, PDE4 inhibition, CYP450 inhibition activity, BBB receptor crossing antagonists, Adenosine receptors, corticosteroids binding, Drug–protein interaction, PharmacoKinetics etc.

In biophysical chemistry, Protein-Ligand interactions, Putative proteins, β-blockers, drugs leading to Phospholipidosis, consequences of amino acid sequence, Chromosomal aberrations were studied with mathematical molecular descriptors. The variation of chromatographic/^{13}C NMR responses of large variety of compounds are interpreted and predicted with a pool of descriptors. The validity and accuracy of prediction of chemically significant mathematically/statistically derived constants (parameters) like logβ of metal-ligand complexes, ΔG, logP, Cell permeability and chemical properties -- solubility Boiling point, melting point, Viscosity refractive index and glass transition temperature-- is significantly increased over classical approaches with properly selected descriptors.

Software: The calculation of molecular descriptors started in manual mode mostly with paper and pencil. Later hand-held calculators played an important role for a molecule of large size. Leaving aside many historical in house programs, ADAPT by Jurs, Package of Kier and Hall are the stand alone tools. CODESSA by Katrizky in late 1990s is a full-fledged software package for generation of over 1000

descriptors and regression procedures for cause-effect model. DRAGON from University of Milan by Todeschini is a mega descriptor generator to over 4500 descriptors along with modelling tools. Some of the typical packages (ADRIANA, MOLGEN-QSPR, PreADMET..)in vogue are compared from the stand point of number and classes of descriptor outputted.

Keywords: Physico- bio-chemical properties, Hammett substituent constants, Hansch parameters, Weiner/Randic index, WHIM, Topological, geometric, electro-topological, quantum-chemical, Eigen, spectra, CODESSA, DRAGON
