REMINISCENCE

CROATICA CHEMICA ACTA

Croat. Chem. Acta **2020**, *93*(4), 247–258
Published online: October 7, 2021
DOI: 10.5562/cca3759

HKD 1926

My Tortuous Pathway Through Mathematical Chemistry and QSAR Research With Memories of Some Personal Interactions and Collaborations With Professors Milan Randić and Mircea Diudea

Subhash C. Basak

Department of Chemistry and Biochemistry, University of Minnesota, Duluth, 1802 Stanford Avenue, Duluth MN 55811, USA
* Author's e-mail address: sbasak@d.umn.edu

RECEIVED: March 16, 2021 * ACCEPTED: March 16, 2021

— THIS PAPER IS DEDICATED TO PROF. MILAN RANDIĆ ON THE OCCASION OF HIS 90TH BIRTHDAY, AND TO THE MEMORY OF PROF. MIRCEA DIUDEA —

Abstract: This article describes my more than four decades of not so straightforward journey through mathematical chemistry and QSAR research with descriptions of some valuable personal interactions and collaborations with Professors Milan Randić and Mircea Diudea.

Keywords: molecular structure, model object, theoretical model, graph theory, quantitative structure-activity relationship (QSAR), topological indices (TIs), three dimensional (3-D) or geometrical descriptors, quantum chemical descriptors, principal component analysis (PCA), congeneric sets, structurally diverse sets, chemodescriptor, biodescriptor, parsimony principle, morphine, naloxone, amphetamine, organophosphate, colligative property, constitutive property, Severe Acute Respiratory Syndrome (SARS) virus, Middle East Respiratory Syndrome (MERS) virus, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2 or COVID-19), connect the dots, peptide vaccine design, mutagenicity, blood-brain barrier (BBB) entry.

At the Beginning: A Dichotomous Academic Existence

"I have become my own version of an optimist. If I can't make it through one door, I'll go through another door - or I'll make a door. Something terrific will come no matter how dark the present."

Rabindranath Tagore

The first spark of my interest in the relationship between structural aspects of molecules and their properties arose out of two factors: a) Teaching physical chemistry to students at a college in Kolkata, India, and b) Carrying out research on the neurochemical effects of psychoactive drugs like morphine, pethidine on rat brain membrane-bound enzymes and toxic effects of various organophosphates on organisms. Colligative properties like vapor pressure lowering, boiling point elevation, freezing point depression, and osmotic pressure of solutions are properties that depend ***only*** upon the concentration of solute molecules or ions, being independent of the constitution or identity of the solute. Constitutive properties, on the other hand, depend on the constitution or structure of the substance. The American Heritage® Dictionary of the English Language, 5th Edition, states the following regarding the word constitutive:[1]

"In physical chemistry, a term introduced by Ostwald to denote those properties of a compound which depend on the constitution of the molecule, or on the mode of union and arrangement of the atoms in the molecule."

The subject matter of my doctoral research on amphetamine[2,3] was a group of psychoactive drugs (amphetamines) where different closely related structures had different stimulant property profiles. My research on biochemical effects of a group of organophosphate compounds also was a case where the major activity of these compounds against the enzyme acetylcholinesterase

(AChE) was guided mainly by the electronic structure of the molecules.[4,5] More subtle effects of minor structural changes on biological property were observable in the action of drugs like morphine and its antagonist naloxone.[6,7]

Biochemical and Pharmacological Observations on the Relationship Between Structure and Biological Activity

For almost a century various researcher in biochemistry and pharmacology generated data on the relation between molecular structure and bioactivity. Probably one of the earliest was the 1928 finding of Quastel and Wooldridge[8] that malonic acid competitively inhibited the activity of the Krebs cycle enzyme succinic dehydrogenase. Although the substrate succinic acid and the inhibitor malonic acid differed by one ($–CH_2$) group the active site of the enzyme still recognized malonic acid. The antibiotic penicillin inhibits cell wall biosynthesis in bacteria by interfering with the transpeptidation reaction responsible for the cross-linking of mucopeptide chains in the cell wall polymer. This is attributed to its putative structural similarity to the D-alanyl-D-alanine portion of the peptide chain.[9,10] The western world is now going through a serious public health crisis of the abuse and overdose death of opioid drugs like morphine, fentanyl[10–12] killing thousands of persons every year in the USA alone.

The effective opioid receptor antagonist naloxone or Narcan is used routinely to treat overdose patients. This action of naloxone originates from its structural similarity to the agonists producing at the same time no mood altering or pain killing effects of the agonists. The antagonistic effects happen by virtue of competitive displacement of the agonist opioid molecules as well as the blockage of opioid access to the receptor sites. We are living now in an age of frequently emerging epidemics and pandemics like Bird Flu, Swine flu as well as coronaviruses, the Severe Acute Respiratory Syndrome (SARS), Middle East Respiratory Syndrome (MERS), and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2 or COVID-19) being some latest instances of the twenty first century. The enzyme neuraminidase is critical for the life cycle of the influenza virus. Analogs or chemicals structurally similar to sialic acid (*N*-acetylneuraminic acid) have been developed as potential treatment for influenza infection.[13]

THE HANSCH QSAR MODEL: Connecting the Dots in the Linear Free Energy Related (LFER) Approach

While wondering about why some physical properties were more dependent on structure as compared to others I accidentally came across the work of Corwin Hansch[14] where he used a combination of electronic, steric, and hydrophobic aspects of structures in correlating physical as well as biological properties of molecules. This approach arose by "connecting the dots" of the different physical organic chemistry methods in bringing together steric, electronic, and hydrophobic factors used separately in relating structural changes to the chemical reactivity of molecules as shown in Figure 1.

This gave me some understanding of the physical/ structural basis of what we call structure-activity relationship.

Chemical graph theory research by our Kolkata group

Around 1974 I started doing research on the development of information-theoretic indices of chemical graphs and their applications in QSAR in collaboration with Dr. A. Roy of Jadavpur University, Kolkata, India. In this research we

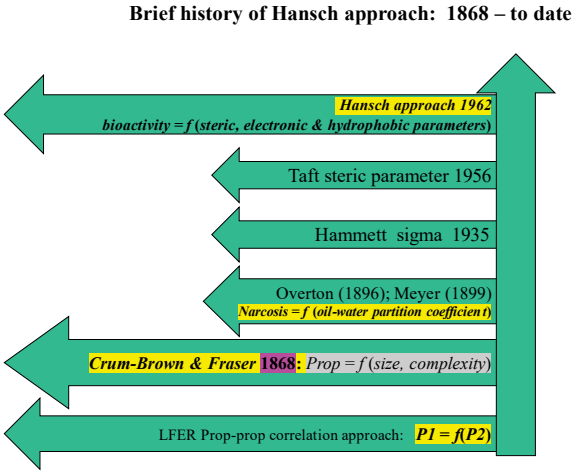

Figure 1. Short history (1868 to date) of the development of linear free energy relationship (LFER) approach to quantitative structure-activity relationship modeling. For more information please see Refs [14,15]. In this approach, a property (P_1) is estimated from another available property (P_2) or a combination of other properties.

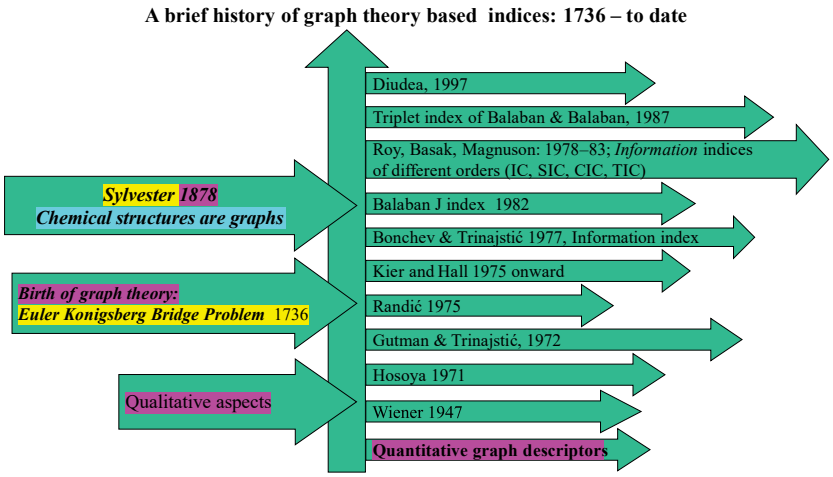

Figure 2. A brief history of graph theory[26] and representation of molecules via chemicographs by Sylvenster.[27] Important topological indices were formulated by Wiener,[28] Hosoya,[29] Randić,[18] Kier & Hall,[17] Bonchev & Trinajstić,[19,20] Balaban et al.,[30,31] Basak, Roy, Harriss and Magnuson,[21,32–35] Diudea[36] and Zagreb group indices of Gutman & Trinajstić.[37]

were highly benefitted by Professor A. T. Balaban's[16] book "Chemical applications of graph theory" and Kier and Hall's[17] book "Molecular connectivity in chemistry and drug research". Through reading the book of Kier and Hall[17] we became aware of the important work of Milan Randić[18] in the formulation of the graph theoretic branching index now popularly known as the connectivity index. Pretty soon our team in Kolkata became familiar with the work of Danail Bonchev and Nenad Trinajstić[19,20] on information theoretic topological indices. We started applying graph theoretic indices in QSAR studies mainly for small and congeneric sets of molecules.[21–25]

A brief chronological account of major developments in graph theory and graph-theoretic numerical descriptors or topological indices is given in Figure 2.

The above factors informed me that chemical structure or molecular graph has a rich content of information and techniques of molecular topology are useful in exploring QSARs using computed graph theoretic descriptors. So, I formed an interdisciplinary team in Kolkata to work on these issues and started publishing results of our research[22–25,33,34] mainly on small and congeneric sets of molecules.

ALL TOPOLOGICAL INDICES ARE NOT CREATED EQUAL: Connecting the Dots in the Topological Index Isle for High Quality QSAR

" शैले शैले न माणिक्यं मौक्तिकं न गजे गजे ।
साधवो नहि सर्वत्र चन्दनं न वने वने ॥"

Chanakya, 375–283 BCE

In Sanskrit

"shaile shaile na maanikyang mauktikang na gaje gaje;
saadhavo nahi sarvatra chandanang na vane vane."

"Not all mountains contain gems in them, nor does every elephant have pearl in it. noble people are not found everywhere, nor is sandalwood found in every forest."

A topological index quantifies different qualitative aspects of molecular structure based on the model object used for the representation of molecules.[38,39] Often an index would be developed to characterize one major qualitative aspect of molecular structure, e.g., branching, complexity, cyclicity, etc. For practical applications in QSARs one specific index or class of indices may not work for large and diverse data sets. Hansch[14] used a combination of different physicochemical factors to derive acceptable QSARs as compared to one kind of property like hydrophobicity (log P) or Hammett's electronic descriptor sigma. This author believes that a similar reasoning is behind the great success of Kier & Hall[17,40] in QSAR using combination of different topological indices. ***When in the 1980s Basak tried to correlate toxicity (LC50 in fathead minnow) of large and diverse sets of aquatic toxicants using one topological index (TI) or a combination of a few TIs at a time or LFER descriptors like log P (octanol-water) everything failed.*** The next pragmatic approach was to use structural information coded in ninety (90) topological indices calculated by the POLLY software[41] to develop quantitative molecular similarity analysis (QMSA)[42] and QSAR models using robust methods like principal components analysis (PCA), ridge regression (RR) and partial least square (PLS) which gave satisfactory results.[43,44]

My collaborations with Professor Milan Randić

I first met Milan Randić in 1983 during the first international mathematical chemistry conference organized by Professors R. B. King and D. H Rouvray on the campus of the University

of Georgia at Athens, USA. Because I live in Minnesota and Milan was in Iowa, two adjoining states, we started talking with each other soon. Subsequently, in the 1990s I invited him to work with my team as a consultant in various research projects which were funded to the University of Minnesota Duluth with myself as the principal investigator (PI), e.g., an NIH grant on the design of anti-epileptic drugs, a United States Air Force grant in predictive toxicology. In the Air Force project, the funding agency gave us the responsibility of developing predictive models for the assessment of toxicity of priority pollutants from the two-dimensional gel electrophoresis (2DE) proteomics patterns obtained by the exposure of cells/organism to the substances. In analogy with topological indices derived from molecular graphs different numerical descriptors of the proteomics maps were called by our University of Minnesota Duluth group "biodescriptors" which found applications in structure-toxicity relationship model development.[45–49]

Another area of biodescriptor development was in the characterization of DNA/RNA sequences using alignment-free mathematical descriptors (AFSDs). I met Ashesh Nandy in 1998 at the First Indo-US Workshop on Mathematical Chemistry organized on the campus of Visva Bharati University, West Bengal, India. Dr. Nandy's group gave a presentation on the comparison of globin genes using his method of plotting nucleic acid sequences. Figure 3 shows the different methods put forward by Gates,[50] Nandy[51] and Leong & Morgenthaler[52] in the graphical representation of nucleic acids. Figure 4 gives the plot of two globin genes of humans and lemur using Nandy's method[51] of plotting described in Figure 3.

After the Indo-US Workshop was over I met Ashesh Nandy in Kolkata in January 1998, had some discussion with him in carrying out collaborative research on the possible creation of multidimensional spaces from a collection of different sequence descriptors in line with what we did with a set of 90 topological indices and a set of 3,692 diverse chemicals[42] and invited him to join my research team at the Natural Resources Research Institute (NRRI, University of Minnesota Duluth, USA). When he came to Minnesota I also asked Milan Randić, Xiaofeng Guo (China), and Marjan Vracko from Slovenia (all three at that time working at NRRI as consultant/visiting scientists) to join in the effort to develop biodescriptors for DNA/RNA

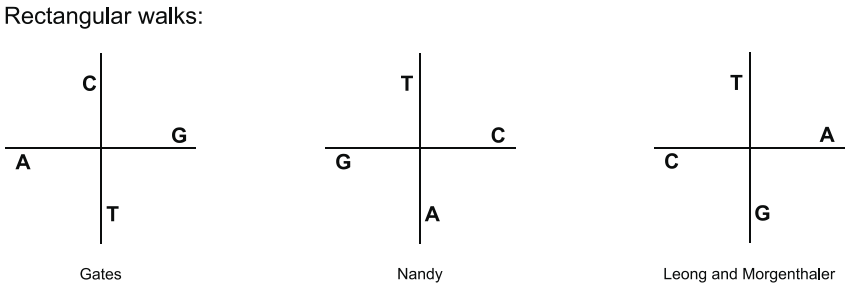

Figure 3. The different modes of 2D graphical representation of the bases of nucleic acids using methods of Gates,[50] Nandy,[51] and Leong & Morgenthaler.[52]

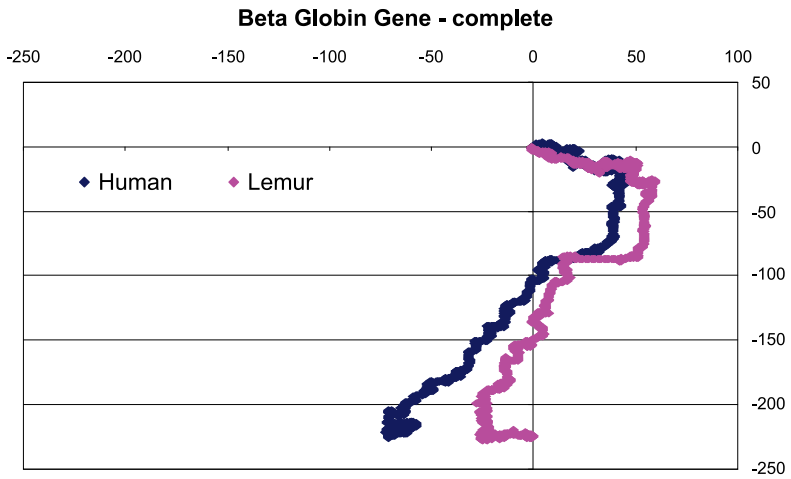

Figure 4. The 2D graphical representations of beta globin genes of humans and lemurs using Nandy's method[51] of plotting described above.

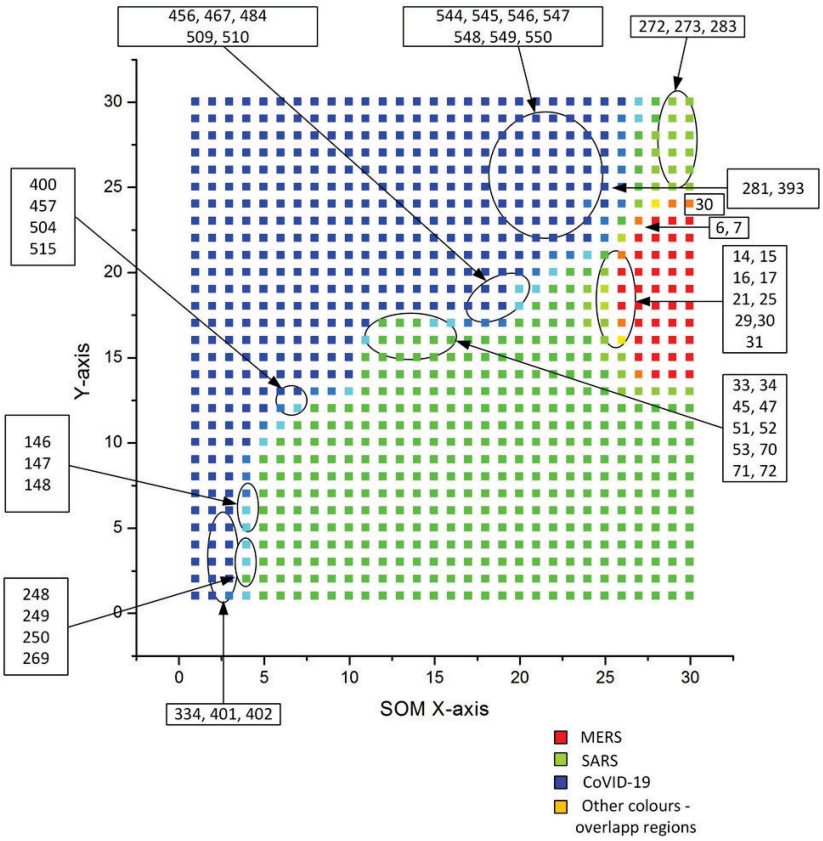

Figure 5. SOM analysis using MERS (32), SARS (252), SARS-CoV-2 (289) using 64 AFSDs.

sequences in a project funded internally by NRRI. This led to a few seminal papers and sparked a worldwide interest in this area as evident by a flurry of published papers.[53–73]

In the realm of practical applications of AFSDs of DNA/RNA sequences there have been two important applications in the surveillance of emerging global pathogens like the Zika, SARS, MERS, and SARS-CoV-2 or COVID-19 viruses and computer-assisted design of novel peptide vaccines.

Surveillance of emerging global pathogens

In recent years we have witnessed the emergence of numerous zoonotic viruses jumping from their usual hosts to cause epidemics and pandemics for the humans.[74] RNA gene containing organisms like the Zika, SARS, MERS, and SARS-CoV-2 are more prone to mutation. As compared to the hosts such viruses possess high mutation rates—up to a million times higher.[75] Such high mutation rates are correlated with augmented viral virulence and evolutionary advantage, traits considered helpful for viruses. Frequent mutations are occurring in SARS-CoV-2 genomes in the genes encoding the Spike protein, RNA polymerase, RNA primase, and nucleoprotein.[76] The new SARS-CoV-2 variant, B.1.1.7, first found in the United Kingdom as well as others emerging in other countries are worrying the public health service agencies worldwide. Therefore, comparison of existing viral sequences with the already known ones can be helpful in the surveillance of global pathogens. Figure 5 shows results of such a study taking a collection of viruses - 32 sequences of MERS, 252 SARS, and 289 SARS-CoV-2 or COVID-19 derived from calculated AFSDs modeled by SOM (self-organizing map) method.[77]

As evident from Figure 5, the SOM model can very strongly discriminate among the three different but closely related classes of beta coronaviruses. Such AFSD-based models can help in the fast surveillance of newly emerging global pathogenic variants.

Use of alignment-free sequence descriptors in vaccine design

Ever Edward Jenner[78] in 1796 inoculated a child and a few others against smallpox by injecting them with cowpox vaccines of various types are being used worldwide to protect us from serious diseases as well as for the prevention of the spread of those diseases to others. The various forms of classical vaccines are made from dead or inactivated organisms or purified parts of pathogens. It is of mainly four types: (a) Live-attenuated, (b) Inactivated, (c) Subunit, and (d) Toxoid.[79] Because of various difficulties arising out of treatments by classical vaccines researchers are working on the development of an alternative — the peptide vaccines.[80–83]

The major steps in peptide vaccine design project were: a) Find highly conserved amino acid sequences of important viral proteins like the spike glycoprotein of SARS-CoV-2, b) Test the conserved sequences for their potential immunogenicity using online computational tools, c) Ascertain solvent accessibility of the epitope sequence, and d) Test autoimmune threat potential of the sequences selected by steps a–c above. In the first crucial step of finding highly conserved amino acid sequences mathematical sequence descriptors were used for their selection and clustering.[80–83]

My Interaction with Mircea Diudea: Brief but Highly Productive

All that glitters is not gold—
Often have you heard that told.
Many a man his life hath sold
But my outside to behold.

William Shakespeare,
Merchant of Venice, Act II Scene 7

"Theories should be as simple as possible, but no simpler."

Albert Einstein

I met Mircea Diudea many times during the Math/Chem/Comp (MCC) and International Academy of Mathematical Chemistry (IAMC) conferences organized in Dubrovnik, Croatia, both of us being member of IAMC. After I was elected the President of the International Society of Mathematical Chemistry (ISMC) I had some discussions on cooperation between ISMC and the European Society of Mathematical Chemistry which was also launching a journal.[84] Subsequently, in 2016 I visited Cluj to participate at the IAMC conference organized there by Mircea Diudea. After the conference we discussed about collaboration between my University of Minnesota Duluth team and his department of Chemistry and Chemical Engineering group in Babes-Bolyai University, Cluj, Romania, in the use of topological molecular descriptors in predicting bioactivity and toxicity of large and diverse sets of chemicals. We had significant collaborations and publications before he unfortunately passed away in 2019. Our QSAR collaboration involved mutagenicity and BBB (blood-brain barrier) entry of large and structurally diverse sets of chemicals.[85,86] All results of this ongoing collaboration with the Cluj team are not discussed here for brevity. For the BBB entry data of 415 diverse chemicals, 579 descriptors were calculated by Schrodinger[87] and TopoCluj.[88] The second set of 198 descriptors were computed by University of Minnesota software POLLY[41] and Triplet.[89] These latter set of descriptors included both topostructural (TS) and topochemical (TC) subclasses. The former set of indices encodes information strictly on molecular connectivity. The TC indices encode information about chemical features in addition to topological information. These chemical features include atom and bond type. Table 1 provides a list of the TIs from the Basak lab used in this study, along with their brief descriptions.

To summarize the results, both the Cluj set of 579 descriptors and the University of Minnesota set of 198 descriptors gave similar results. Because there were a large number of descriptors in both cases robust or parsimonious methods of statistical model building were used.[42,90–92] The predictive quality of the models was very similar.[84,85] When the combined set of 777 descriptors was used for model building there was no improvement in model quality as compared to the two sets of 579 and 198 descriptors used individually. It is possible that many of the descriptors calculated by either the UMN or the Cluj software may have reached a PLATEAU in the realm of abstraction of relevant STRUCTURAL INFORMATION from molecular structure. It is noteworthy that although Cluj set of indices had some three-dimensional (3-D) and quantum chemical indices they could not make much improvement in model quality over those developed by topological indices only as reported by Basak group[15] using their hierarchical QSAR (HiQSAR) approach.

THE WAY FROWARD: Not only connect the dots, also connect the connectors

"The current state of knowledge is a moment in history, changing just as rapidly as the state of knowledge in the past has ever changed and, in many instances, more rapidly."

Jean Piaget

"To create is to live twice."

Albert Camus - The Myth of Sisyphus

We are currently living in an interesting time when a lot of data are being generated on the effects of chemicals on various levels of biological organization. Property-property correlation using either laboratory test data or experimental data based approach like the LFER method[14] cannot help in assessing the property/bioactivity/toxicity of chemicals now available in the marketplace or those needed for the screening of chemical candidates in the drug discovery pipeline. After the completion of the Human Genome Project[93] the omics technologies like the genomics and proteomics techniques are generating a lot of data that can be used in combination with both available experimental data and algorithmically derived properties that can be computed for any molecule, real or hypothetical, without the input of any other experimental data. ***A unified approach, not only connecting the domain-specific dots, but connecting the connectors are needed***. We envision an overarching integrative CHEMOBIODESCRIPTOR COMBIANTION (Figure 6) where we will take a combined approach through the union of different individual approaches for predicting property/bioactivity of chemicals.[45,94]

Table 1. Symbols, definitions, and classification of structural molecular descriptors.

	Topostructural (TS)
I_D^W	Information index for the magnitudes of distances between all possible pairs of vertices of a graph
$\overline{T}_D^W$	Mean information index for the magnitude of distance
W	Wiener index = half-sum of the off-diagonal elements of the distance matrix of a graph
I^D	Degree complexity
H^V	Graph vertex complexity
H^D	Graph distance complexity
$\overline{IC}$	Information content of the distance matrix partitioned by frequency of occurrences of distance h
M_1	A Zagreb group parameter = sum of square of degree over all vertices
M_2	A Zagreb group parameter = sum of cross-product of degrees over all neighboring (connected) vertices
${}^h\chi$	Path connectivity index of order h = 0–10
${}^h\chi_C$	Cluster connectivity index of order h = 3–6
${}^h\chi_{PC}$	Path-cluster connectivity index of order h = 4–6
${}^h\chi_{Ch}$	Chain connectivity index of order h = 3–10
P_h	Number of paths of length h = 0–10
J	Balaban's J index based on topological distance
$nrings$	Number of rings in a graph
$ncirc$	Number of circuits in a graph
DN^2S_y	Triplet index from distance matrix, square of graph order, and distance sum; operation y = 1–5
DN^21_y	Triplet index from distance matrix, square of graph order, and number 1; operation y = 1–5
$AS1_y$	Triplet index from adjacency matrix, distance sum, and number 1; operation y = 1–5
$DS1_y$	Triplet index from distance matrix, distance sum, and number 1; operation y = 1–5
ASN_y	Triplet index from adjacency matrix, distance sum, and graph order; operation y = 1–5
DSN_y	Triplet index from distance matrix, distance sum, and graph order; operation y = 1–5
DN^2N_y	Triplet index from distance matrix, square of graph order, and graph order; operation y = 1–5
ANS_y	Triplet index from adjacency matrix, graph order, and distance sum; operation y = 1–5
$AN1_y$	Triplet index from adjacency matrix, graph order, and number 1; operation y = 1–5
ANN_y	Triplet index from adjacency matrix, graph order, and graph order again; operation y = 1–5
ASV_y	Triplet index from adjacency matrix, distance sum, and vertex degree; operation y = 1–5
DSV_y	Triplet index from distance matrix, distance sum, and vertex degree; operation y = 1–5
ANV_y	Triplet index from adjacency matrix, graph order, and vertex degree; operation y = 1–5
	Topochemical (TC)
O	Order of neighborhood when ICr reaches its maximum value for the hydrogen– filled graph
O_{orb}	Order of neighborhood when ICr reaches its maximum value for the hydrogen– suppressed graph
I_{ORB}	Information content or complexity of the hydrogen–suppressed graph at its maximum neighborhood of vertices
IC_r	Mean information content or complexity of a graph based on the rth (r = 0–6) order neighborhood of vertices in a hydrogen–filled graph
SIC_r	Structural information content for rth (r = 0–6) order neighborhood of vertices in a hydrogen–filled graph
CIC_r	Complementary information content for rth (r = 0–6) order neighborhood of vertices in a hydrogen–filled graph
${}^h\chi^b$	Bond path connectivity index of order h = 0–6
${}^h\chi_C^b$	Bond cluster connectivity index of order h = 3–6
${}^h\chi_{Ch}^b$	Bond chain connectivity index of order h = 3– 6
${}^h\chi_{PC}^b$	Bond path–cluster connectivity index of order h = 4–6
${}^h\chi^v$	Valence path connectivity index of order h = 0–6
${}^h\chi_C^v$	Valence cluster connectivity index of order h = 3–6
${}^h\chi_{Ch}^v$	Valence chain connectivity index of order h = 3–6
${}^h\chi_{PC}^v$	Valence path–cluster connectivity index of order h = 4–6
J^B	Balaban's J index based on bond types
J^X	Balaban's J index based on relative electronegativities
J^Y	Balaban's J index based on relative covalent radii
AZV_y	Triplet index from adjacency matrix, atomic number, and vertex degree; operation y = 1–5
AZS_y	Triplet index from adjacency matrix, atomic number, and distance sum; operation y = 1–5
ASZ_y	Triplet index from adjacency matrix, distance sum, and atomic number; operation y = 1–5
AZN_y	Triplet index from adjacency matrix, atomic number, and graph order; operation y = 1–5
ANZ_y	Triplet index from adjacency matrix, graph order, and atomic number; operation y = 1–5

Continued on next page

Table 1. Continued from previous page.

DSZ_y	Triplet index from distance matrix, distance sum, and atomic number; operation y = 1–5
DN^2Z_y	Triplet index from distance matrix, square of graph order, and atomic number; operation y = 1–5
nvx	Number of non-hydrogen atoms in a molecule
nelem	Number of elements in a molecule
fw	Molecular weight
${}^h\chi^v$	Valence path connectivity index of order h = 7–10
${}^h\chi^v_{Ch}$	Valence chain connectivity index of order h = 7–10
si	Shannon information index
totop	Total Topological Index *t*
sumI	Sum of the intrinsic state values *I*
sumdelI	Sum of delta-*I* values
tets2	Total topological state index based on electrotopological state indices
phia	Flexibility index ($kp_1 \cdot kp_2/nvx$)
Idcbar	Bonchev-Trinajstić information index
IdC	Bonchev-Trinajstić information index
Wp	Wienerp
Pf	Plattf
Wt	Total Wiener number
knotp	Difference of chi-cluster-3 and path/cluster-4
knotpv	Valence difference of chi-cluster-3 and path/cluster-4
nclass	Number of classes of topologically (symmetry) equivalent graph vertices
NumHBd	Number of hydrogen bond donors
NumHBa	Number of hydrogen bond acceptors
SHCsats	E-State of C sp^3 bonded to other saturated C atoms
SHCsatu	E-State of C sp^3 bonded to unsaturated C atoms
SHvin	E-State of C atoms in the vinyl group, =CH–
SHtvin	E-State of C atoms in the terminal vinyl group, $=CH_2$
SHavin	E-State of C atoms in the vinyl group, =CH–, bonded to an aromatic C
SHarom	E-State of C sp^2 which are part of an aromatic system
SHHBd	Hydrogen bond donor index, sum of Hydrogen E-State values for –OH, =NH, $–NH_2$, –NH–, –SH, and #CH
SHwHBd	Weak hydrogen bond donor index, sum of C–H Hydrogen E-State values for hydrogen atoms on a C to which a F and/or Cl are also bonded
SHHBa	Hydrogen bond acceptor index, sum of the E-State values for –OH, =NH, $–NH_2$, –NH–, >N–, –O–, –S–, along with –F and –Cl
Qv	General Polarity descriptor
NHBinty	Count of potential internal hydrogen bonders (y = 2–10)
SHBinty	E-State descriptors of potential internal hydrogen bond strength (y = 2–10)
	Electrotopological State index values for atoms types: *SHsOH, SHdNH, SHsSH, SHsNH2, SHssNH, SHtCH, SHother, SHCHnX, Hmax Gmax, Hmin, Gmin, Hmaxpos, Hminneg, SsLi, SssBe, Sssss, Bem, SssBH, SsssB, SssssBm, SsCH3, SdCH2, SssCH2, StCH, SdsCH, SaaCH, SsssCH, SddC, StsC, SdssC, SaasC, SaaaC, SssssC, SsNH3p, SsNH2, SssNH2p, SdNH, SssNH, SaaNH, StN, SsssNHp, SdsN, SaaN, SsssN, SddsN, SaasN, SssssNp, SsOH, SdO, SssO, SaaO, SsF, SsSiH3, SssSiH2, SsssSiH, SssssSi, SsPH2, SssPH, SsssP, SdsssP, SssssssP, SsSH, SdS, SssS, SaaS, SdssS, SddssS, SsssssssS, SsCl, SsGeH3, SssGeH2, SsssGeH, SssssGe, SsAsH2, SssAsH, SsssAs, SdsssAs, SssssssAs, SsSeH, SdSe, SsssSe, SaaSe, SdssSe, SddssSe, SsBr, SsSnH3, SssSnH2, SsssSnH, SssssSn, SsI, SsPbH3, SssPbH2, SsssPbH, SssssPb*
	Geometrical (3D) / Shape
kp_0	Kappa zero
kp_1-kp_3	Kappa simple indices
ka_1-ka_3	Kappa alpha indices
V_W	Van der Waals volume
${}^{3D}W$	3D Wiener number based on the hydrogen-suppressed geometric distance matrix
${}^{3D}W_H$	3D Wiener number based on the hydrogen-filled geometric distance matrix
	Quantum Chemical (QC)
E_{HOMO}	Energy of the highest occupied molecular orbital
E_{HOMO-1}	Energy of the second highest occupied molecular
E_{LUMO}	Energy of the lowest unoccupied molecular orbital
E_{LUMO+1}	Energy of the second lowest unoccupied molecular orbital
ΔH_f	Heat of formation
μ	Dipole moment

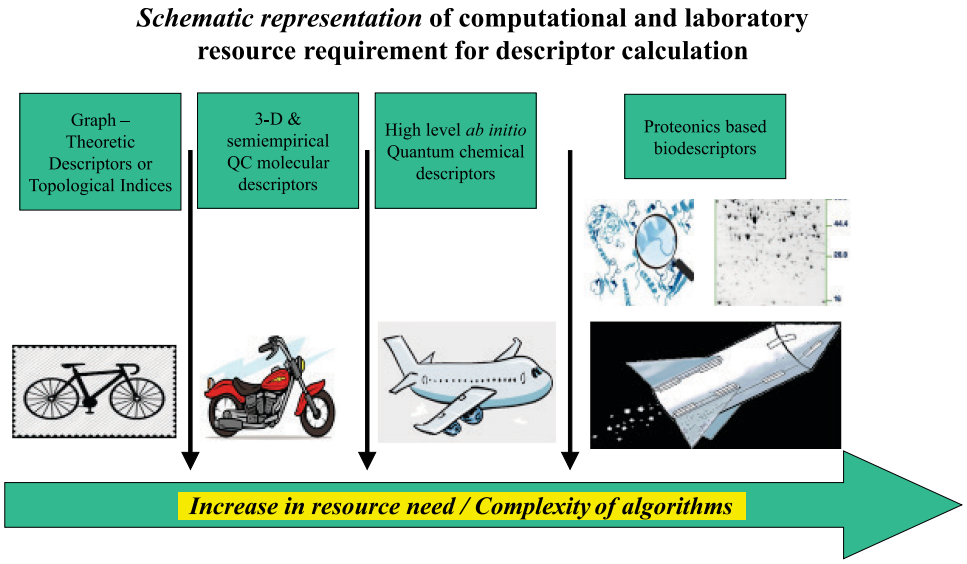

Figure 6. A schematic representation of possible workflow in various approaches in descriptor-based evaluation of chemicals and relative cost/resource needs and complexity of different classes of chemodescripors and biodescriptors.

Dedication. I dedicate this article to Professors Milan Randić and Mircea Diudea for their outstanding contributions to various aspects of chemistry, mathematical chemistry, and chemical graph theory.

Acknowledgment. I have been benefitted by extensive collaborations with many researchers in different areas of experimental biochemistry, mathematical chemistry, predictive toxicology, and computer-assisted drug design all of whom cannot be mentioned here because the list is very long. I would like to specially mention Professor J. J. Ghosh (late Centenary Professor of Biochemistry, Calcutta University, India, Professor A B Roy (Late Professor of the Department of Mathematics, Jadavpur University, Kolkata, India). Fruitful collaboration with Professors A. T. Balaban, M. Randić, M. Diudea, L. B. Kier, N. Trinajstić, M. Vracko, A. Nandy, S. Manna, D. M. Hawkins, G. J. Niemi is acknowledged. I am also thankful to Dr. Subhabrata Majumdar and Dr. C. Lungu, Mr. G. D., Grunwald, Mr. B. D. Gute, and Ms. D. M. Mills for their valuable collaborations.

REFERENCES

[1] The American Heritage® Dictionary of the English Language, 5th Edition. https://www.wordnik.com/words/constitutive#:~:text=Having%20power%20to%20enact%20or,which%20see%2C%20unite%20regulative

[2] E. Costa and S. Garattini (Eds) *Amphetamines and related compounds,* Raven Press, New York, **1970**. http://hdl.handle.net/2027.42/33644

[3] S. C. Basak, Biochemical Action of Amphetamine in Brain (with some biomathematical modelling and interpretations), Ph.D. Thesis, University of Calcutta: Calcutta, India, **1979**.

[4] R. D. O'Brien, *Toxic Phosphorus Esters: Chemistry, Metabolism, and Biological Effects,* Academic Press, **1960**.

[5] R. Sengupta, S. C. Basak, D. Sarkar, J. J. Ghosh, *Pest. Biochem. Physiol*. **1975**, *5*, 52–53. https://doi.org/10.1016/0048-3575(75)90043-7

[6] B. Banerjee, S. C. Basak, J. J. Ghosh, *Brain News* **1980**, *12*, 18–21.

[7] B. Banerjee, S. C. Basak, J. J. Ghosh, *IRCS Med. Sci.* **1981**, *9*, 929–630.

[8] J. H. Quastel, W. R. Wooldridge, *Biochem. J*. **1928**, *22*, 689–702. https://doi.org/10.1042/bj0220689

[9] B. Erlanger, L. Goode, *Nature* **1967**, *213*, 183–184. https://doi.org/10.1038/213183a0

[10] A. Goodman & L. S. Gilman, *The Pharmacological Basis of Therapeutics*, 8th Edition, Pergamon Press, New York, **1990**.

[11] Centers for Disease Control and Prevention (CDC): *Trends and Geographic Patterns in Drug and Synthetic Opioid Overdose Deaths — United States*, **2013–2019**. (accessed 26 February 2021). https://doi.org/10.15585/mmwr.mm7006a4

[12] A. Dahan, L. Aarts, T. W. Smith, *Anesthesiology* **2010**, *112*, 226–238. https://doi.org/10.1097/ALN.0b013e3181c38c25

[13] C. J. Vavricka, C. Muto, T. Hasunuma, et al., *Sci Rep*. **2017**, *7*, 8239. https://doi.org/10.1038/s41598-017-07836-y

[14] C. Hansch, A. Leo, *Exploring QSAR*: Volume 1: *Fundamentals and Applications in Chemistry and Biology*, American Chemical Society; 1st edition, **1995**.

[15] S. C. Basak, *Curr. Comput.-Aided Drug Des.* **2013**, *9*, 449–462. https://doi.org/10.2174/15734099113096660041

[16] A. T. Balaban, (Ed.) *Chemical applications of graph theory*, Academic Press, **1976**.

[17] L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, **1976**.

[18] M. Randic, *J. Am. Chem. Soc.* **1975**, *7*, 6609–6615. https://doi.org/10.1021/ja00856a001

[19] D. Bonchev, N. Trinajstic, *J. Chem. Phys.* **1977**, *67*, 4517–4533. https://doi.org/10.1063/1.434593

[20] D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structures*, Research Studies Press, Chichester, United Kingdom, **1983**.

[21] S. C. Basak, A. B. Roy, and J. J. Ghosh, *Study of the structure-function relationship of pharmacological and toxicological agents using information theory, in Proceedings of the Second International Conference on Mathematical Modelling*, X. J. R. Avula, R. Bellman Y. L. Luke and A. K. Rigler, (Eds.), pp 851–856, University of Missouri-Rolla, Rolla, Missouri, **1979**.

[22] S. K. Ray, S. C. Basak, C. Raychaudhury, A. B. Roy, J. J. Ghosh, *Indian J. Chem.* **1981**, *20B*, 894–897.

[23] S. C. Basak, V. R. Magnuson, *Arzneim.-Forsch./Drug Res.* **1983**, *33*, 501–503.

[24] S. C. Basak, D. P. Gieschen, V. R. Magnuson, D. K. Harriss, *IRCS Med. Sci.* **1982**, *10*, 619–620.

[25] S. C. Basak, D. P. Gieschen, D. K. Harriss, V. R. Magnuson, *J. Pharm. Sci.* **1983**, *72*, 934–937. https://doi.org/10.1002/jps.2600720823

[26] F. Harary, *Graph theory*, 2nd ed., Addison-Wesley: Reading, MA, **1969**. https://doi.org/10.21236/AD0705364

[27] J. J. Sylvester, *Amer. J. Math.* **1878**, *1*, 64–125.

[28] H. Wiener, *J. Am. Chem. Soc.* **1947**, *69*, 17–20. https://doi.org/10.1021/ja01193a005

[29] H. Hosoya, *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339. https://doi.org/10.1246/bcsj.44.2332

[30] P. A. Filip, A. T. Balaban, T. S. Balaban, *J. Math. Chem.* **1987**, *1*, 61–83. https://doi.org/10.1007/BF01205338

[31] A. T. Balaban, *Chem. Phys. Lett.* **1982**, *89*, 399–404. https://doi.org/10.1016/0009-2614(82)80009-2

[32] S. C. Basak, *Information theoretic indices of neighborhood complexity and their applications*, S. C. Basak, in *Topological Indices and Related Descriptors in QSAR and QSPR*, J. Devillers and A. T. Balaban, Eds., Gordon and Breach Science Publishers, The Netherlands, **1999**, pp. 563–593.

[33] S. C. Basak, S. K. Ray, C. Raychaudhury, A. B. Roy, J. J. Ghosh, *IRCS Med. Sci.* **1982**, *10*, 145–146.

[34] S. C. Basak, *Med. Sci. Res.* **1987**, *15*, 605–609. https://doi.org/10.1007/BF03306651

[35] A. B. Roy, S. C. Basak, D. K. Harriss, and V. R. Magnuson, Neighborhood complexities and symmetry of chemical graphs and their biological applications, In: Mathematical Modelling in Science and Technology, X. J. R. Avula, R. E. Kaman, A. I. Lapis, and E. Y. Rodin, Eds., Pergamum Press, **1984**, pp. 745–750. https://doi.org/10.1016/B978-0-08-030156-3.50138-7

[36] M. V. Diudea, *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 292–299.

[37] I. Gutman, N. Trinajstic, *Chem. Phys. Lett.* **1972**, *17*, 535–538. https://doi.org/10.1016/0009-2614(72)85099-1

[38] M. Bunge, *Method, model and matter*, Boston, Reidel, **1973**. https://doi.org/10.1007/978-94-010-2519-5

[39] S. C. Basak, HYLE **2013**, *19*, 3–17.

[40] L. B. Kier, L. H. Hall, *Molecular Structure Description: The Electrotopological State*, Academic Press, San Diego, CA, **1999**.

[41] S. C. Basak, D. K. Harriss, V. R. Magnuson, POLLY v.2.3: **1988**; Copyright of the University of Minnesota, USA

[42] S. C. Basak, V. R. Magnuson, G. J. Niemi, R. R. Regal, *Discrete Appl. Math.* **1988**, *19*, 17–44. https://doi.org/10.1016/0166-218X(88)90004-2

[43] D. M. Hawkins, S. C. Basak, X. Shi, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 663–670. https://doi.org/10.1021/ci0001177

[44] D. M. Hawkins, S. C. Basak, D. Mills, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 579–586. https://doi.org/10.1021/ci025626i

[45] S. C. Basak, *Drug Dev. Res.* **2010**, *72*, 1–9. https://doi.org/10.1002/ddr.20428

[46] S. C. Basak, B. D. Gute, *Curr. Opin. Drug Discov. Devel.* **2008**, *11*, 320–326.

[47] S. C. Basak, B. D. Gute and F. Witzmann, *WSEAS Transactions on Information Science and Applications* **2006**, *7*, 996–1001.

[48] M. Vracko, S. C. Basak, K. Geiss, F. Witzmann, *J. Chem. Inf. Model.* **2006**, *46*, 130–136. https://doi.org/10.1021/ci0502597

[49] M. Randic, F. Witzmann, M. Vracko, S. C. Basak, *Med. Chem. Res.* **2001**, *10*, 456–479.

[50] M. A. Gates, *J. Theor. Biol.* **1986**, *119*, 319–328. https://doi.org/10.1016/S0022-5193(86)80144-8

[51] A. Nandy, *Curr. Sci.* **1994**, *66*, 309–314. https://doi.org/10.1524/ract.1994.6667.special-issue.309

[52] P. M. Leong, S. Morgenthaler, *Comput. Appl. Biosci.* **1995**, *11*, 503–507. https://doi.org/10.1093/bioinformatics/11.5.503

[53] M. Randic, M. Vracko, A. Nandy, S. C. Basak, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1235–1244. https://doi.org/10.1021/ci000034q

[54] A. Nandy, S. C. Basak, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 915–919. https://doi.org/10.1021/ci990117a

[55] A. Nandy, P. Nandy, S. C. Basak, *Internet Electron. J. Mol. Des.* **2002**, *1*, 367–373. http://www.biochempress.com

[56] A. Nandy, S. C. Basak, B. D. Gute, *J. Chem. Inf. Model.* **2007**, *47*, 945–951. https://doi.org/10.1021/ci600558w

[57] *Similarity and dissimilarity of DNA/RNA sequences*. D. Bielinska-Waz, P. Waz, W. Nowak, A. Nandy, S. C. Basak, *Computation in Modern Science and Engineering, Proceedings of the International Conference on Computational Methods in Science and Engineering 2007* (ICCMSE 2007) (Eds.: T. E. Simos, G. Maroulis), American Institute of Physics: Melville, New York, **2007**; 28–30.

[58] A. Ghosh, A. Nandy, P. Nandy, B. Gute, S. C. Basak, *J. Chem. Inf. Model.* **2009**, *47*, 2627–2638. https://doi.org/10.1021/ci9001662

[59] A. Nandy, S. C. Basak, *Curr. Comp.-Aided Drug Des.* **2010**, *6*, 283–289. https://doi.org/10.2174/1573409911006040283

[60] A. Nandy, T. Sarkar, S. C. Basak, P. Nandy, S. Das, *Curr. Comp.-Aided Drug Des.* **2014**, *10*, 285–302. https://doi.org/10.2174/1573409911666150318203621

[61] S. C. Basak, A. Nandy, *Curr. Comp.-Aided Drug Des*. **2016**, *12*, 1–3. https://doi.org/10.2174/1573409912999160315115502

[62] A. Nandy, S. Dey, S. C. Basak, D. Bielinska-Waz, P. Waz, *Curr. Comp.-Aided Drug Des.* **2016**, *12*, 87–97. https://doi.org/10.2174/1573409912666160401115812

[63] D. Sen, S. Dasgupta, I. Pal, S. Manna, S. C. Basak, A. Nandy, G. D. Grunwald G. *Curr. Comp.-Aided Drug Des.* **2016**, *12*, 216–228. https://doi.org/10.2174/1573409912666160525111918

[64] D. Panas, P. Waz, D. Bielinska-Waz, A. Nandy, S. C. Basak, *MATCH Commun. Math. Comput. Chem.* **2017**, *77*, 321–332.

[65] S. Dey, A. Nandy, S. C. Basak, P. Nandy, S. Das, *Comput. Biol. Chem.* **2017**, *68*, 143–152. https://doi.org/10.1016/j.compbiolchem.2017.03.002

[66] D. Sen, P. Roy, A. Nandy, S. C. Basak, S. Das, *Chem. Phys.* **2018**, *513*, 156–164. https://doi.org/10.1016/j.chemphys.2018.07.031

[67] A. Nandy, M. Harle, S. C. Basak, *ARKIVOC* **2006**, *9*, 211–238. https://doi.org/10.3998/ark.5550190.0007.907

[68] X. Guo, M. Randic, S. C. Basak, *Chem. Phys. Lett.* **2001**, *350*, 106–112. https://doi.org/10.1016/S0009-2614(01)01246-5

[69] M. Randic, X. Guo, S. C. Basak, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 619–626. https://doi.org/10.1021/ci000120q

[70] M. Randic, S. C. Basak, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 561–568. https://doi.org/10.1021/ci0000981

[71] M. Randic, J. Zupan, A. T. Balaban, D. Vikić-Topić, D. Plavsic, *Chem. Rev.* **2011**, *111*, 790–862 https://doi.org/10.1021/cr800198j

[72] A. Perez-Bello, C. R. Munteanu, F. M. Ubeira, A. L. DeMagalhaes, E. Uriarte, H. Gonzalez-Diaz, *J. Theor. Biol.* **2009**, *256*, 458–466. https://doi.org/10.1016/j.jtbi.2008.09.035

[73] C. Li, W. Fei ,Y. Zhao, X. Yu, *Appl. Sci.* **2016**, *6*, 63 https://doi.org/10.3390/app6030063

[74] World Health Organization (WHO) report on Coronavirus disease (COVID-19) pandemic https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (Accessed 1 November 2021)

[75] S. Duffy, *PLoS Biol. 16*(*8*), e3000003. https://doi.org/10.1371/journal.pbio.3000003

[76] C. Yin, Genotyping coronavirus SARS-CoV-2: methods and Implications. https://arxiv.org/pdf/2003.10965.pdf

[77] M. Vračko, S. C. Basak, T. Dey, A. Nandy, *Curr Comput Aided Drug Des.* **2021** ahead of print. https://doi.org/10.2174/1573409917666210202092646

[78] T. W. Mak, M. E. Saunders, *Vaccines and clinical immunization, the immune response*, Academic Press, 2006, pp 695–749. https://doi.org/10.1016/B978-012088451-3.50025-9

[79] https://www.vaccines.gov/ (Accessed, 1 November 2021).

[80] S. Chatterjee, S. Dey, A. Nandy, S. C. Basak, *A Computational Search for Peptide Vaccines Using Novel Mathematical Descriptors of Sequences of Emerging Pathogens* In: *Topics in Medicinal Chemistry*, **2020**, Springer, Berlin, Heidelberg. https://doi.org/10.1007/7355_2020_108

[81] P. K. Panda et al., Sci. Adv. **2020**, *6*(*28*), eabb8097. https://doi.org/10.1126/sciadv.abb8097

[82] S. C. Basak, A. Nandy, S. Manna, S. Dey, T. Dutta, S. Biswas, T. Dey, S. Chatterjee, M. Vracko, *Computer-aided new approach for the surveillance and peptide vaccine design for emerging global pathogens like*

Zika and Covid-19, Presented at the Vaccines Summit, Ohio (VSOHIO21), March 1–3, 2021, https://biogatesc.com/events/va-cc-in-e/index.html

[83] *In Silico Approach for Peptide Vaccine Design for CoVID 19*, S. Biswas, S. Chatterjee, T. Dey, S. Dey, S. Manna, A. Nandy, S. C. Basak, *MOL2NET*, **2020**, *6*, ISSN: 2624–5078 1. https://sciforum.net/manuscripts/6787/manuscript.pdf

[84] Journal of the European Society of Mathematical Chemistry, http://chem.ubbcluj.ro/jesmc/

[85] S. Majumdar, S. C. Basak, C. N. Lungu, M. V. Diudea, G. D. Grunwald, SAR QSAR *Environ. Res.* **2018**, *29*, 579–590. https://doi.org/10.1080/1062936X.2018.1496475

[86] S. Majumdar, S. C. Basak, Claudiu N. Lungu, Mircea V. Diudea and G. D. Grunwald, Mol. Inf. 38 (2019) 1800164. https://doi.org/10.1002/minf.201800164

[87] *Small-Molecule Drug Discovery Suite*, **2009**, Schrödinger, LLC, New York, NY.

[88] O. Ursu and M. V. Diudea, TOPOCLUJ software program, Cluj, Romania: Babes-Bolyai University, **2005**.

[89] S. C. Basak, G. Grunwald, A. T. Balaban, TRIPLET, Copyright of the Regents of the University of Minnesota, **1993**.

[90] S. C. Basak, M. Vracko, *Current Computer-Aided Drug Design*, **2020**, *16*, 1–5. https://doi.org/10.2174/1573409916012001061228 54

[91] Simplicity: Stanford encyclopedia of philosophy, https://plato.stanford.edu/entries/simplicity/

[92] J. Vandekerckhove, D. Matzke, and E. J. Wagenmakers, Model comparison and the principle of parsimony. In Busemeyer J. R., Wang Z., Townsend J. T., & Eidels A., (Eds.), Oxford handbook of computational and mathematical psychology, **2015** pp. 300–319. https://doi.org/10.1093/oxfordhb/9780199957996.013.14

[93] The Human Genome Project, https://www.genome.gov/human-genome-project

[94] D. M. Hawkins, S. C. Basak, J. J. Kraker, K. T. Geiss and F. A. Witzmann, J. Chem. Inf. Model., 46 (2006) 9–16. https://doi.org/10.1021/ci050252p