Chemoinformatics — A quick review

Article · April 2011

DOI: 10.1109/ICECTECH.2011.5942128

CITATIONS
8

READS
2,331

2 authors:

Arulmozhi Varadharajan
Tiruppur Kumaran College for Women.
10 PUBLICATIONS **32** CITATIONS
SEE PROFILE

Reghunadhan Rajesh
Central University of Kerala
102 PUBLICATIONS **1,280** CITATIONS
SEE PROFILE

2011 3rd International Conference on Electronics Computer Technology (ICECT 2011)

Chemoinformatics – A Quick Review

V. Arulmozhi
Department of Computer Science
Tiruppur Kumaran College for Women
Tiruppur, India
arultkc@yahoo.co.in

R. Rajesh
TDC, Network Systems and Technologies
Technopark, Thiruvananthapuram - 695581
Kerala, India
kollamrajeshr@ieee.org

***Abstract*—Chemoinformatics is the application of information technology in design, creation, organization, management, retrieval, analysis, visualization and use of chemical information to solve chemical problems and drug design. This paper presents a quick survey on chemoinformatics, tools and its applications.**

Keywords-chemoinformatics; drug design; chemical visualization; chemical reaction

I. INTRODUCTION

The intrinsic information in huge amounts of data produced in chemistry and pharmaceutical research is often difficult to grasp, since these data contains information about various characteristics of chemical compounds and various methods have to be applied to extract the relevant information. Machine learning, a technique which makes use of several disciplines like cognitive science, computer science, pattern recognitions and statistics, can be used to classify/predict chemical data. The marriage between chemistry and information technology (with machine learning as one of the important component) is recently called as chemoinformatics [1][2][15-29]. Figure 1 shows the picture of Professor Dmitrii Ivanovich Mendeleev.

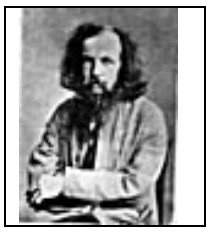

Figure 1: Dmitrii Ivanovich Mendeleev (1834-1907) – An Early Chemoinformatician who discovered Periodic Table

At most basic level what we needed are tools to predict the physical, chemical, and biological properties of small molecules and reactions in order to focus searches and filter search results. Computational methods in chemistry can be organized along a spectrum ranging from computationally intensive Schrödinger equation, Quantum mechanical methods, molecular dynamics methods to fast statistical and machine learning methods.

This paper presents a note on chemoinformatics. This paper is organized as follows. Section 2 presents a view on chemoinformatics. Section 3 presents a few tools based on chemoinformatics and Section 4 concludes the paper.

II. CHEMOINFORMATICS – A SHORT REVEIW

One of the major tasks of chemoinformatics is lead finding and lead optimization within the drug discovery process and its application covers analysis of HTS data, similarity search, design of combinatorial libraries, design of focused libraries, comparison of the similarity/diversity of libraries, virtual screening, docking, prediction of binding affinities, physicochemical properties, pharmacokinetic properties, pharmacological properties, biological activity, modeling, prediction, etc. Drug design can take very long time for preparation and it is expensive too, due to the lack of clinical facilities, animal toxicity, adverse reactions in humans, commercial and formulation issues.

The metabolism of a potential drug has to be considered at an early phase of the drug development process. Biochemical pathways and metabolic reaction networks are another active and rich field for research in chemoinformatics.

Moreover, reaction databases provide a rich source of information for the extraction of knowledge for reaction, prediction and synthesis design. The complex task of chemical reactivity, makes use of this reaction database, is another major task of chemoinformatics.

Reaction prediction and knowledge extraction are time consuming process and develop kinetic models. It creates great interest in the application of chemoinformatics.

The understanding and simulations of chemical reactions is one of the great challenges in chemo informatics. The problems of poor results due to limited understanding of chemical reactions should be solved by deductive and inductive learning methods.

The conformational flexibility of ligand and protein, the prediction of the binding similarity of a substrate to an enzyme, the water desolvation of the ligand, the modeling of protein-protein interaction, the determination of the geometry and the calculation of the strength of the hydrogen bonds, and the prediction of 3D structure of proteins sequence are still challenging works and have to be concentrated in a better way and need to be improved to computer assisted drug design. New drug discovery strategy should be able to extract knowledge from large-scale databases in a shorter time periods and should be able to provide efficient tools to predict properties. More about

978-1-4244-8679-3 /11/$26.00 © 2011 IEEE

chemoinformatics can be had from the Handbook of chemo Informatics [13].

In short chemoinformatics has the descriptors, but not limited to the following.

a) Geometrical Descriptors

i) Molecular rsurface area
ii) Solvent-accessible molecular surface area
iii) Molecular volume
iv) Solvent-excluded molecular volume
v) Gravitational indexes
vi) Principal moments of inertia of a molecule
vii) Shadow areas of molecule
viii) Relative shadow areas of a molecule

b) Electrostatic Descriptors

i) Gasteiger-Marsili empirical atomic partial charges
ii) Zefirov's empirical atomic partial charges
iii) Mulliken atomic partial charges
iv) Minimum (most negative) and maximum (most positive) atomic partial charges
v) Polarity parameters
vi) Dipole moment
vii) Molecular polarizability
viii) Molecular hyperpolarizability
ix) Average ionization energy
x) Minimum electrostatic potential at the molecular surface
xi) Maximum electrostatic potential at the molecular surface
xii) Local polarity of molecule
xiii) Total variance of the surface electrostatic potential
xiv) Electrostatic balance parameter

c) Quantum Chemical Descriptors

i) Total energy of the molecule
ii) Total electronic energy of the molecule
iii) Standard heat of formation
iv) Electron-electron repulsion energy for a given atomic species
v) Nuclear-electron attraction energy for a given atomic species
vi) Electron-electron repulsion between two given atoms
vii) Nuclear-electron attraction energy between two given atoms
viii) Nuclear repulsion energy between two given atoms
ix) Electronic exchange energy between two given atoms
x) Resonance energy between given two atomic species
xi) Total electrostatic interaction energy between two given atomic species
xii) Total interaction energy between two given two atomic species
xiii) Total molecular one-center electron-electron repulsion energy
xiv) Total molecular one-center electron-nuclear attraction energy
xv) Total intramolecular electrostatic interaction energy
xvi) Electron kinetic energy density
xvii) Energy of protonation

d) MO related Descriptors

i) Highest Occupied Molecular Orbital (HOMO) energy
ii) Lowest Unoccupied Molecular Orbital (LUMO) energy
iii) Absolute hardness
iv) Activation hardness
v) Fukui atomic nucleophilic reactivity index
vi) Fukui atomic electrophilic reactivity index
vii) Fukui atomic one-electron reactivity index
viii) Mulliken bond orders
ix) Free valence

e) Thermodynamic Descriptors

i) Vibrational enthalpy of the molecule
ii) Translational enthalpy of the molecule
iii) Vibrational entropy of the molecule
iv) Rotational entropy of the molecule
v) Translational entropy of the molecule
vi) Vibrational heat capacity of the molecule
vii) Normal coordinate eigen values (EVA)

f) DFT based Reactivity Descriptors

i) Chemical Potential
ii) Chemical Hardness
iii) Softness
iv) Electrophilicity Index
v) Condensed Fukui Function
vi) Philicity
vii) Group Philicity

The following figure 2 shows the drug discovery and development process.

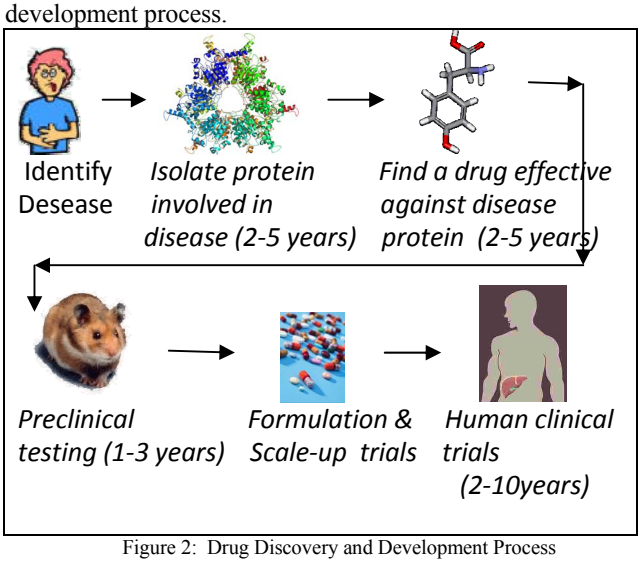

Figure 2: Drug Discovery and Development Process

Chemical information is traditionally stored in many different file types which inhibit reuse of the documents. CML uses XML's portability to help CML developers and chemists design interoperable documents. There are a number of tools that can generate, process and view CML documents. Publishers can distribute chemistry within XML documents by using CML.

CML is capable of supporting a wide range of chemical concepts including:

a) molecules
b) reactions
c) spectra and analytical data
d) computational chemistry
e) chemical crystallography and materials

III. TOOLS

Cory and Wipke in 1969 first showed how the computers were used for chemical synthesis by using the program OCCS [14]. Since then, development of software and tools for computer-assisted organic synthesis have still under development. There are vast number of tools/ specification/ notations are available for the study of chemoinformatics.

SMILES (Simplified Molecular Input Line Entry Specification) [4][5] is one such general purpose chemical nomenclature and data exchange format, originally developed by Arthur Weininger and David Weininger in 1980 and further extended by open source chemistry community (called Open SMILES). In graph-based computational procedure, SMILES is a string obtained by printing the symbols nodes using techniques of depth-first tree traversal of chemical graph. It removes hydrogen atoms and cycle are turned it to spanning tree.

Some of the chemoinformatics software's are Chemwindow, Alchemy, Chemdraw [9][10][11], CLIFF[12], etc. Other popular softwares are, not limited to, the following.

a) Drawing the chemical structure - ISIS/Draw [6][7][8]
b) Representation of Organic structure Description Arranged Linearly – ROSDAL
c) To specify 3D image of the molecular structure with additional lighting effects to enhance the impression of the depth in spaces - Rasmol, Swiss PDB Viewer, PyMOL, Accelry
d) Tools used for drawing chemical structure - Molecular editor and Viewers
e) Generating 3D structure - CORINA and CONCORD
f) Biomolecular codes – Amber
g) Sequence Analysis codes for finding structural motifs in nucleic acid sequences – RNAMotif
h) Provides a programming environment for geometric and force-field manipulations of nucleic acids (and proteins as well), now part of Amber - NAB (Nucleic Acid Builder)
i) Predicts nitrogen, carbon and proton chemical shifts in proteins and proton shifts in nucleic acids – SHIFTS
j) Finds good binding positions for small molecules in protein or nucleic acid receptors; include a variety of scoring protocols, including those using the Amber/GB solvation model - UCSF DOCK
k) A complete system of *ab initio* programs for molecular electronic structure calculations – MOIL
l) Develop Common Component Architecture (see http://cca-forum.org/) interfaces for quantum chemistry – MPQC
m) Large-scale Atomic/Molecular Massively Parallel Simulator - Classical molecular dynamics code LAMMPS

Some other software are ALOGPS, CLOGP, KOWWIN ,XLOGP, IA_LOGP, AMSOL, APBS, ArgusLab, Babel, CLOGP, KOWWIN,XLOGP, IA_LOGP, BioSolveIT, ChemTK, Chimera, CLIFF, COSMOlogic, Dragon, gOpenMol, Grace, JOELib, Jmol, Lammps, MIPSIM, Mol2Mol, MOLCAS, Molexel, ICM-Pro, ORTEP, Packmol, Polar, PREMIER Biosoft, Q-chem, Qmol, SageMD, TINKER, Transient, TURBOMOLE, UNIVIS, VMD, WHATIF, etc. and more information about these and other softwares can be had from [12].

IV. CONCLUSION AND FUTURE WORKS

Chemo informatics has developed over the last 40 years to an established discipline that has applications in many areas of chemistry. This paper has given an introductory note on Chemoinformatics.

There are still many problems that await a solution, namely, drug design, multivariate optimization, combination of bioinformatics & chemoinformatics, etc. Therefore still will need new developments and research in the development of hardware and software to tackle the various problems in chemoinformatics.

ACKNOWLEDGMENT

The first authors is thankful to Mother Teresa Women's University, Kodaikannal and the Chairman, Principal, and all the staff members of the Tiruppur Kumaran College for Women for their valuable support. The second author is thankful to all the team members of TDC group, NeST for their valuable support. The second author is an Assistant Professor (on leave) in the Department of Computer Applications, Bharathiar University and he is thankful to Bharathiar University authorities for their valuable support.

REFERENCES

[1] R.V. Williams and M.E. Bowden, Chronology of Chemical Information Science(at http:/www.libsci.sc.edu)

[2] W.L. Chen, Chemoinformatics: past, present and future, Journal of Chemical Information and Modelling 46 (2006)2230-2255

[3] F.K.Brown, Chemoinformatics: What is it a d how does it impact drug discovery?, Annual Reports in Medicinal Chemistry 33,1998, 375-384

[4] Weininger, D. (1988), SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28, 31-36.

[5] Weininger, D.; Weininger, A.; Weininger, L. (1989) SMILES. 2. Algorithm for generation of unique SMILES notation J. Chem. Inf. Comput. Sci. 29, 97-101.

[6] Li, Z.; Wan, H.; Shi, Y.; Ouyang, P. (2004). "Personal Experience with Four Kinds of Chemical Structure Drawing Software: Review on ChemDraw, ChemWindow, ISIS/Draw, and ChemSketch". J. Chem. Inf. Comput. Sci. 44 (5): 1886–1890. doi:10.1021/ci049794h. PMID 15446849.

[7] Johann Gasteiger, Thomas Engel (2003). Chemoinformatics: A Textbook. Wiley-VCH. p. 143. ISBN 3527306811.

[8] Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L. et al. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited, Journal of Chemical Information and Computer Sciences, 1992, 32, 244-255.

[9] Mills, N. (2006). "ChemDraw Ultra 10.0". J. Am. Chem. Soc. 128 (41): 13649–13650. doi:10.1021/ja0697875.

[10] Li, Z.; Wan, H.; Shi, Y.; Ouyang, P. (2004). "Personal Experience with Four Kinds of Chemical Structure Drawing Software: Review on ChemDraw, ChemWindow, ISIS/Draw, and ChemSketch". J. Chem. Inf. Comput. Sci. 44 (5): 1886–1890. doi:10.1021/ci049794h. PMID 15446849.

[11] Dieter Strack (2001). "ChemOffice Ultra 2000". Phytochemistry 57 (1): 144. doi:10.1016/S0031-9422(00)00503-3.

[12] http://www.ccl.net/chemistry/links/software/index.shtml

[13] Johann Gasteiger "Handbook of Chemoinformatics: From Data to Knowledge" Wiley- VCH | 2003.

[14] E.J. Corey and W.T. Wipke, Computer-Assisted Design of Complex Organic Synthesis. Science 166 (1969), pp. 178–192

[15] Leach, Andrew R.; Gillet, Valerie J. An Introduction to Chemoinformatics. Kluwer, 2003. ISBN 1-4020-1347-7

[16] Engel, Thomas. Chemoinformatics: A Textbook. Wiley-VCH, expected date of publication: August 2003. ISBN 3-527-30681-1

[17] Encyclopedia of Computational Chemistry, Schleyer, P. von R.; Allinger, N.L.; Clark, T.; Gasteiger, J.; Kollman, P.A.; Schaefer, H.F.; Shreiner, P.R. (Eds.). 5 v. Wiley, Chichester, 1998.

[18] Gasteiger, Johann J., ed. Handbook of Chemoinformatics: From Data to Knowledge. 4 v. Wiley-VCH, expected date of publication August 2003. ISBN 3-527-30680-3

[19] Reviews in Computational Chemistry. Wiley-VCH, 1990-

[20] Paris, Greg. Bibliography: Chemical Information Retrieval and 3D Searching. http://panizzi.shef.ac.uk/cisrg/links/grep/chemDB.4.html

[21] SIRCh: Chemical Informatics Home Page at Indiana University http://www.indiana.edu/~cheminfo/informatics/cinformhome.html

[22] Brown, F.K. "Chemoinformatics, what it is and how does it impact drug discovery." Annual Reports in Medicinal Chemistry, 1998, 33, 375-384.

[23] Glen, Robert. "Developing tools and standards in molecular informatics." Chemical Communications, 2002, (23), 2745-2747.

[24] Hann, Mike; Green, Richard. "Chemoinformatics—a new name for an old problem?" Current Opinion in Chemical Biology, 1979, 3, 379-383.

[25] Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. "Experimental and computational approaches to estimate the solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews, 1997, 23, 3-15.

[26] Rosso, Eugene. "Chemistry plans a structural overhaul." Nature (Naturejobs) 12 September 2002, 419(6903).

[27] http://www.nature.com/naturejobs/careersandrecruitment/2002.html

[28] Rothman, L. David. "Information management for research in the chemical industry." Abstracts of Papers, 223rd ACS National Meeting, Orlando, FL, United States, April 7-11, 2002 (2002), CINF-044.

[29] Smith, Chris. "Cheminformatics: Redefining the crucible." The Scientist, 2002, 16(8), 40.