Advanced Application of DIKW Model in Molecular Modelling: An Interdisciplinary Approach

Abhilash Thakur[1], Brij Kishor Tiwari[2], Apurva Trivedi[3] Ashok Barua[4], Mamta Thakur[5*]

[1] Softvision College, Indore, Madhya Pradesh, India

[2] GL Bajaj Institute of Technology and Management, Greater Noida, Uttar Pradesh, India

[3] Shri Jain Diwakar Mahavidyalaya, Indore, Madhya Pradesh, India

[4] Department of Chemistry, Govt. Holkar Science College, Indore, Madhya Pradesh, India

[5] School of Logistics and Supply Chain Management,
Symbiosis University of Applied Sciences, Indore, Madhya Pradesh, India

orcid id : 0000-0002-5662-6403

Email: [5] mamta.thakur@suas.ac.in

Abstract

The computer aided drug design (CADD) methodologies such as quantitative structure activity relationship (QSAR), docking studies, and many combinatorial techniques are now a days employed successfully for designing of new molecules and for prediction of wide variety of molecular activities. These techniques not only saved a lot of animal/human trails, but also amplified the research in many areas such as; medicine, chemical sciences, environmental sciences, physical sciences etc.

It has been observed that theses methodologies also follows the framework of management information system (MIS) i.e., Data, information, knowledge, wisdom (DIKW) model. The study focusing on mapping computer aided drug design procedures on the management information system model or DIKW model.

A case has been selected to map the CADD methods on DIKW pyramid and it has been concluded that CADD methods also follows the DIKW model in exactly the same fashion as it is followed by all organizations in the business management. The study provides a beautiful concoction of management, information technology and sciences.

Keywords: Computational Chemistry, QSAR, Docking, DIKW, Data, Information.

1. Introduction

The basic objective of in-silico methods (also known as computer aided drug design) is to assist researchers, specially chemists and biologist in the preclinical development of the molecules. There are two approaches widely used in Computer aided drug design (CADD) viz., Cheminformatics and Computational Chemistry to develop molecular models. The former approach store and provide data for the study and later approach use this data to model a new molecule(s) with desirable properties.

Cheminformatics is an allied field of information technology that focuses on the collection, storage, analysis and manipulation of the chemical data[1] obtained through previously performed experimental work[2]. It largely includes designing of digital libraries (to store data

obtained from experimental research), storing chemical structure and sub-structures, and retrieving stored data.

On the contrary computational chemistry utilized the data, structures & substructure of the molecules and their experimental properties from digital libraries or a literature. It encompasses the estimation/calculation of different physical, chemical and structural properties of the molecules based on their molecular structures[3]. The method mainly includes; application of data processing tools (e.g. multiple linear regression analysis, factor analysis, nonlinear relationships etc.) on the retrieved and estimated data, generation of mathematical model (causal model, in which experimental properties are dependent variable and calculated parameters are independent parameters), prediction of molecular properties (Physical, Chemical and/or biological) and finally designing of preclinical molecules.

Various simulation methods are used to solve the problems based on the modelling of new molecules. It facilitates user to understand mode of action of molecule(s), before synthesis and clinical trials.

Usually the data used in computer aided drug design (CADD) is a secondary data and obtained from digital libraries, reports, journals, book and periodicals. This data is further processed, using variety of computational tools (Quantitative Structure activity relationship, pharmacophore mapping etc) to acquire the information related to that particular family of molecules. To explore more about the properties and attributes of the molecules, the simulations techniques were used. Based on results of data processing and simulation, the new molecule with comprehensive information is proposed. The cheminformatics is a source of data and computational chemistry is utilizing the data to retrieve the relevant information and designing of preclinical molecule.

In CADD, it has been observed that flow of data and information is executed in the same fashion as it is execute in the business organizations or any other public or private organizations. The flow of data and information within the organizations were studied by various researchers [4-10]. The first computational database and transaction management system was developed by IBM in 1968.[11]

In present study the management information system (MIS model) developed by IBM on the framework of DIKW model has been applied on the CADD. The transformation of data and information at each level of MIS model or DIKW pyramid is discussed in terms of CADD methodologies. The study will provide a concise understanding of drug discovery pipeline and help researchers to follow the systematic protocol to design the molecules. The study also highlights the innovative application of DIKW pyramid in drug discovery.

2. Method

2.1 Study of Molecular Modelling of Aromatic Nitro Compound.

In present study, the dataset of 12 aromatic nitro compounds (**Table 1)** has been taken as a sample for mapping study. A CADD study has already been published for these 12 compound[12]. Thakur et-al have taken the biological activity i.e., rate constant log K of aromatic nitro compounds from the literature[13] as a secondary data.

Other structural parameters were also computed using various computation tool. These computational tools are transforming the molecular structures and coding it into different indices or properties. The complete data set is divided into two types: Secondary data which

is logK (reduction rate constant) and primary data that is structural codes which is computed using computational tools viz., ACD labs, Dragon, Hyperchem etc.

As per the assumption of Quantitative structure activity relationship[14], the biological activity of the molecules depend on its structure. A causal research has been performed and multiple linear regression analysis were performed to establish the relationship between log K and other structural parameters. Fifteen mathematical models has been proposed to describe relationship. These predictive models are represented in **Table 2.** This data processing subsequently unfolds the information regarding structural features influencing the reduction rate of aromatic nitro compounds.

Researcher can utilize this information to gain further insight about the structure activity relationship of aromatic nitro compounds and can design a new molecule within a data set.

2.2 Study of DIKW model of Information System (IS)

A literature review has been performed to explore the application of DIKW model or IS in any scientific methodologies by any researcher. Neither review nor research article has been found in the literature review on the application of IS's DIKW model on drug design methods. However mammoth studies have been performed by various researchers on the application of the MIS or DIKW model on various organizations[4-10,15]. The typical DIKW pyramid **(Figure 1)** is classified into four levels; Transaction system processing, operational, tactical and strategic level. At each level different activities has been performed. In the MIS developed by IBM[11] different modules of the software has been implemented at each level.

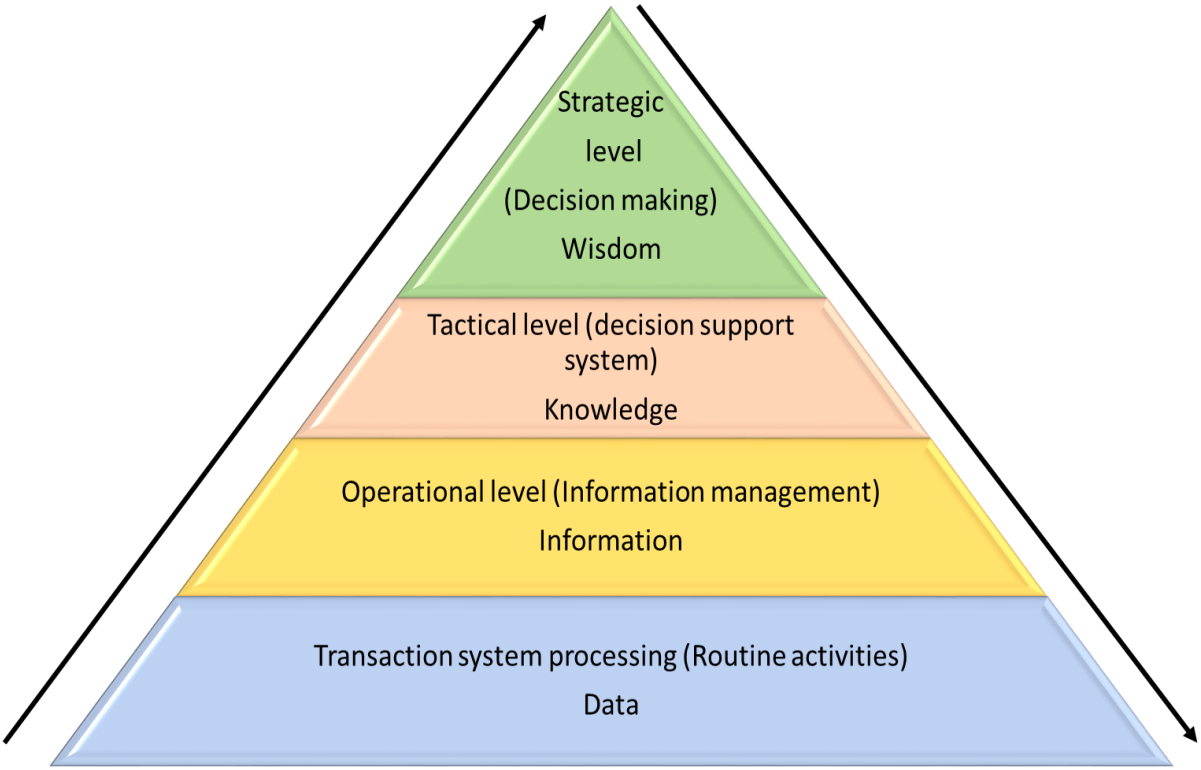

Figure 1: DIKW model of Management Information System in Organizations.

2.3 Mapping of DIKW model and Molecular Model

Firstly all the steps studies in **Section 2.1** in molecular modelling of aromatic nitro compound has been arranged in a sequence. Steps are compared with each level of pyramid in DIKW model of MIS[11].

On the basis of transformation of data used in the study, each step of modelling method is mapped with a different level and activities prescribed in DIKW model of MIS. Nomenclature of each level is mapped with the steps of CADD. Each activities and tools of DIKW model are similarly mapped with procedures and tools used in CADD, the pyramidal representation of mapped model is represented in **Figure 2.**

3. Result and Discussion

3.1 Data Collection:

Table 1 represents dataset of aromatic nitro compounds, its reduction reaction rate constant (logK) and computed variables of all the 12 aromatic nitro compound. The substituents and logK value is collected from secondary sources and remaining variables are computed for the 12 aromatic nitro compounds, therefore considered as primary data. This can be considered as transaction system processing of DIKW model, at this level raw data is collected from different sources.

3.2 Data Processing(Information):

As per the assumption of QSAR, the multiple linear regression (MLR) analysis has been performed to establish linear relationship between dependent parameter (logK) and other computed structural variables (Independent parameters). This data processing subsequently revels the relationship and the information regarding the role of structural parameters on reduction reaction rate of aromatic nitro compounds. All causal models are presented in **Table 2.** As per the DIKW model, required information is extracted from the data by data processing. In the current example, information revel from data processing is shown in **Table 2.**

3.3 Simulation (Knowledge):

On the basis of information obtained after data processing (Multiple linear regression model), it can be easily implicit that model no 13 best describe the relationship between log K and structural variables. Out of all independent variables, net charge at 3^{rd} atom (NC_3) and energy of lowest unoccupied molecular orbitals (LUMO) are the most dominating variable to regulate log K of aromatic nitro compounds. By this knowledge one can interpret that the substituents which increase net charge on third atom and also increase energy of lowest unoccupied molecular orbital will increases log K and vice versa.

3.4 Molecular Design (Wisdom):

On the basis of above knowledge, the decision making can be done, to substitute electron withdrawing group which reduces delocalization of electron due to symmetric pull of electron from opposite end of the benzene ring and hence increasing net charge at 3^{rd} position and makes reduction easier. Also electron withdrawing group breaks conjugation which results in increase in energy of lowest unoccupied molecular orbital and facilitated reduction reaction.

On the basis of above application of DIKW model on the molecular modelling of aromatic nitro compounds, the data/information flow in CADD methodologies is represented in **Figure 2** and comparison of typical DIKW and CADD's DIKW is represented in **Table 3.**

Table 1 : Primary and Secondary data used in CADD study of 12 aromatic nitro compounds

C.No.	Substitutes (X)*	logK	NC_2	NC_3	NC_4	ED_2	ED_3	ED_4	HOMO	LUMO
	Secondary data	Secondary data	Primary data	Primary data	Primary data	Primary data	Primary data	Primary data	Primary data	Primary data
1	NO_2	9.38	3.887	3.721	3.718	0.113	0.279	0.282	-0.311	-0.188
2	CN	9.08	3.940	3.761	3.805	0.059	0.239	0.195	-0.312	-0.158
3	SO_2NH_2	8.90	3.879	3.743	3.759	0.121	0.257	0.241	-0.310	-0.167
4	CHO	9.26	3.933	3.748	3.889	0.066	0.252	0.111	-0.247	-0.163
5	CF_3	8.63	3.975	3.769	3.906	0.025	0.231	0.094	-0.282	-0.153
6	COMe	9.00	3.944	3.753	3.895	0.056	0.247	0.105	-0.242	-0.162
7	CO_2Me	8.93	3.948	3.758	3.880	0.052	0.242	0.120	-0.271	-0.159
8	$CONH_2$	8.63	3.966	3.764	3.881	0.034	0.236	0.119	-0.258	-0.157
9	H	8.52	3.978	3.770	3.897	0.022	0.230	0.103	-0.312	-0.154
10	Me	8.18	4.006	3.778	3.815	0.006	0.222	0.185	-0.312	-0.150
11	OMe	8.04	4.055	3.792	3.632	-0.055	0.208	0.368	-0.261	-0.141
12	OH	7.85	4.055	3.792	3.629	-0.055	0.208	0.371	-0.271	-0.141

* Various substituents of 4-X-$C_6H_4NO_2$ reduction by$CH3^{\cdot}CHOH$ in N_2O-saturated aqueous solution.

Table 2 : The Information obtained after data processing using Multiple linear Regression Analysis.

S.No	Dependent variable	Independent variable	MLR model with statistical parameters	Information
1	LogK	**NC2**	**logK = 7.9073(□1.243) NC2 + 8.4074** n = 12, Se = 0.2258, R = 0.8955, R^2 = 0.8019, F = 40.467, Q = 3.97	Net Charge on 2nd atom has a positive linear impact on logK
2	LogK	**ED2**	**logK = -7.8897(□1.1607) ED2 + 39.9735** n = 12, Se = 0.2140, R = -0.9067, R^2= 0.8221, F = 46.203, Q = 4.24	Electron Density at 2nd atom has a negative linear impact on logK
3	LogK	**NC3**	**logK = 22.0128(□2.9421) NC3 + 3.4701** n = 12, Se = 0.1975, R = 0.9211, R^2 = 0.8484, F = 55.982, Q = 4.66	Net Charge on 3rd atom has a positive linear impact on logK
4	LogK	**ED3**	**logK = -22.0128(□2.9421)ED3 + 91.5214**	Electron Density at 3rd atom has a negative linear impact

S.No	Dependent variable	Independent variable	MLR model with statistical parameters	Information
			n = 12, Se = 0.1975, R = -0.9211, R^2= 0.8484, F = 55.982, Q = 4.66	on logK
5	LogK	**LUMO**	**logK= -33.6410(□6.1388)LUMO+ 3.3931** n = 12, Se = 0.2536, R = -0.8661, R^2= 0.7502, F = 30.031, Q = 3.42	Energy of lower unoccupied molecular orbital demonstrate negative impact of rate constant logK
6	LogK	**ED3 & HOMO**	**logK= -22.6015 (± 2.9757) ED3 - 2.3131 (± 2.1787) HOMO . 94.3897** n = 12, Se = 0.1963, R = 0.9302, R2 = 0.8653, F = 28.910, Q = 4.74	Electron density at 3rd atom has negative impact and energy of highest occupied molecular orbital showing positive impact on logK
7	LogK	**NC3 & HOMO**	**logK= 22.6015 (± 2.9757) NC3 + 2.3131 (± 2.1787) HOMO + 3.9835** n = 12, Se = 0.1963, R = 0.9302, R2 = 0.8653, F = 28.910, Q = 4.74	Both Net charge at 3rd atom and energy of highest occupied molecular orbital showing positive impact on logK
8	LogK	**ED3 & NC_4**	**logK= -20.5310 (± 2.7473) ED3 - 1.0151 (± 0.5421) NC_4 + 86.1402** n = 12, Se = 0.1766, R = 0.9439, R2 = 0.8909, F = 36.758, Q = 5.34	Both electron density at 3rd atom and net charge at 4th atom showing negative impact on logK
9	LogK	**NC_3 & NC_4**	**logK= 20.5310 (± 2.7473) NC3 - 1.0151 (± 0.5421) NC4 + 4.0162** n = 12, Se = 0.1766, R = 0.9439, R2 = 0.8909, F = 36.758, Q = 5.34	Net charge at 3rd atom has a positive and net charge at 4th atom shows negative impact on logK
10	LogK	**ED_3 & ED_4**	**logK= -20.5310 (± 2.7473) ED3 + 1.0151 (± 0.5421) ED4 + 82.0798** n = 12, Se = 0.1766, R = 0.9439, R2 = 0.8909, F = 36.758, Q = 5.34	Electron density at 3rd atom shows negative whereas electron density at 4th atom shows positive impact on logK

S.No	Dependent variable	Independent variable	MLR model with statistical parameters	Information
11	LogK	**NC_3 & ED_4**	**logK= 20.5310 (± 2.7473) NC3 + 1.0151 (± 0.5421) ED4 - 0.0442** n = 12, Se = 0.1766, R = 0.9439, R2 = 0.8909, F = 36.758, Q = 5.34	Net charge at 3rd atom and electron density at 4th atom shows positive impact on logK
12	LogK	**ED_3 & LUMO**	**logK= -50.5310 (± 14.1313) ED3 + 46.4115 (± 22.9668) LUMO + 204.4902** n = 12, Se = 0.1727, R = 0.9464, R2 = 0.8957, F = 38.664, Q = 5.48	Electron density at 3rd atom shows negative whereas energy of lowest unoccupied molecular orbital shows positive impact on logK
13	LogK	**NC_3 & LUMO**	**logK= 50.0925 (± 14.1313) NC3 + 46.4115 (± 22.9668) LUMO + 4.1203** n = 12, Se = 0.1727, R = 0.9464, R2 = 0.8957, F = 38.664, Q = 5.48	Both Net charge at 3rd atom and lowest unoccupied molecular orbital shows positive impact on logK

Table 3 : Mapping of typical DIKW model and CADD DIKW model

S.No	Typical DIKW	CADD DIKW	Source of reference in present study
1	Source of data is usually primary in nature and collected from market or sometimes secondary (internal data of the company) or external data recorded in government records or periodicals.	Source of data is usually secondary taken from the literature, magazines, and cheminformatics data bases. Simulation based software are employed to obtain primary data of the molecules.	**Table 1**
2	Excel applications, SPSS etc. tools are used for data processing to obtained information	The Multiple linear regression analysis has been performed in the given case to obtain required information	**Table 2**
3.	Data analysis and recommendations are prepared after gaining knowledge from the information	The recommendation on structural aspects has been proposed after gaining insights from data processing	**Section 3.3**
4.	Decision Making on the basis of recommendations	Molecular design on the basis of recommendations and requirements.	**Section 3.4**

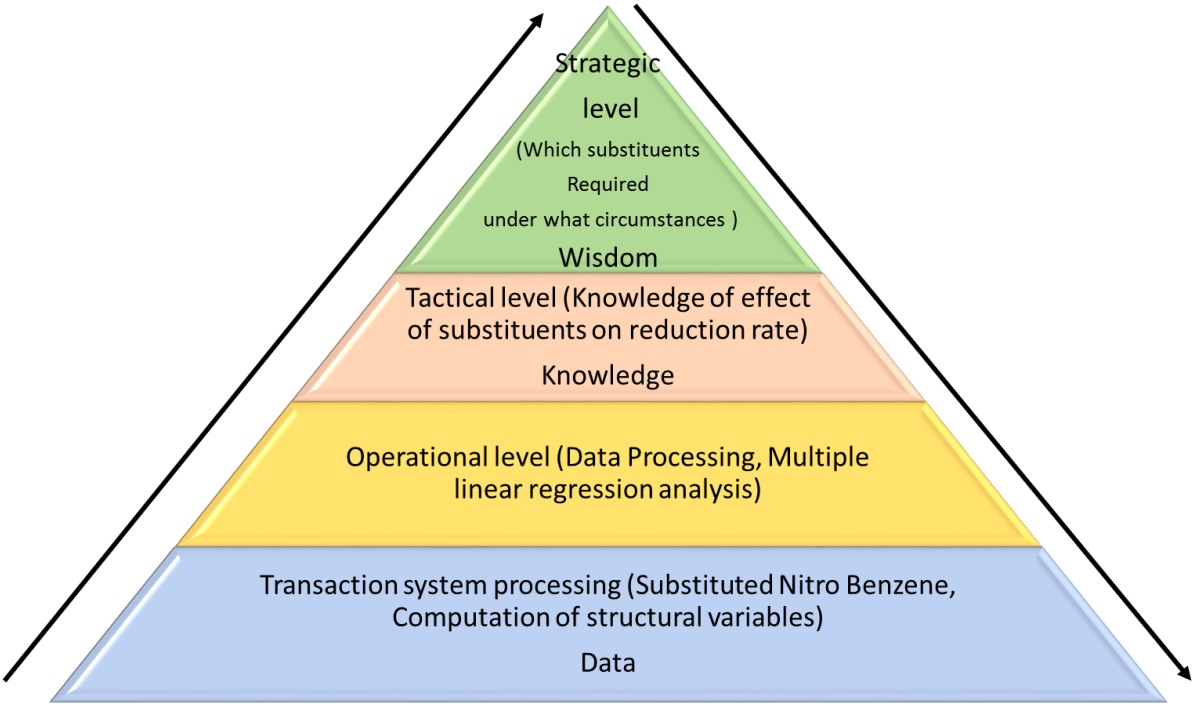

Figure 2: DIKW model for the molecular modelling of aromatic nitro compounds.

4. Conclusion:

The molecular modelling approach used for the drug design and drug discovery follows the same conceptual framework as data information knowledge wisdom (DIKW) model of management information system. In present study, the molecular modelling of aromatic nitro compounds has been taken as a sample to map with the DIKW model.

However being linear in nature DIKW model sometimes deviates from its typical pattern in some management related problem, but it is best fitted for the computer aided drug design methods. In management sometimes information and knowledge can be used as an interchangeable terms, but in CADD it can be clearly differentiated. In management one or the other level of pyramid can be bypassed in reality, but CADD studies cannot proceed to another level without passing lower levels. The study reveals that DIKW model is more applicable to scientific studies than management studies.

Reference

[1] David S. Wishart Introduction to Cheminformatics, current protocols in Bioinformatics 2007, Curr Protoc Bioinformatics. 2016 Mar 24;53:14.1.1-14.1.21.

[2] Cheminformatics: a tool for decision-makers in drug discovery. Olsson T, Oprea TI. Curr Opin Drug Discov Devel. 2001 May;4(3):308-13.

[3] https://arxiv.org/ftp/arxiv/papers/1706/1706.05413.pdf

[4] James O.Hicks,Jr - Management Information System(Third edition , Virginia Polytechnic Institute and State University), 2003

[5] Falkenberg, E. D., Hesse, W., Lindgreen, P., Nilsson, B. E., Oei, J. L. H., Rolland, C., Stamper, R. K., Assche, F. J. M. V., Verrijn-Stuart, A. A. & Voss, K. 1998. A Framework of Information System Concepts: The FRISCO Report, The Netherlands, Department of Computer Science, University of Leiden.

[6] Fayyad, U. M., Piatetsky-Shapiro, G. & Smyth, P. From data mining to knowledge discovery: an overview. AI Magazine, 1996, 17, 37-54.
[7] Frické, M. The knowledge pyramid: a critique of the DIKW hierarchy. Journal of Information Science, 2009, 35, 131–142.
[8] Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. J Inf Sci. 2007, 33(2), 163-80.
[9] Zins, Chaim, Conceptual approaches for defining data, information, and knowledge. Journal of the American Society for Information Science and Technology, 2007, 58: 479-493.
[10] R.L. Ackoff, From data to wisdom, Journal of Applied Systems Analysis, 1989, 16, 3–9.
[11] https://www.ibm.com/ibm/history/ibm100/us/en/icons/ibmims
[12] A.Thakur, M.Thakur, S.Baghel, T.R.Thapak, G.L.Badole, A.Bharadwaj, Reduction of aromatic nitro compounds by CH3ÿCHOH: 3D SAR approach; Organic Chemistry : An Indian Journal, 2010, 6,1,1-7
[13] C.Hansch, A.Kurup, R.Garg, H.Gao; Chem.Rev., Chem-bioinformatics and QSAR: a review of QSAR lacking positive hydrophobic terms, 2001, 101, 619–672.
[14] Verma J, Khedkar VM, Coutinho EC. 3D-QSAR in drug design-A review. Curr Top Med Chem. 2010, 10(1), 95–115.
[15] AfërdÖta Berisha – Shaqir; Management Information System and Decision-Making; Academic Journal of Interdisciplinary Studies, 2014, 3, 2, 19-23.