Computational modeling tool for biological activity assessment of drug candidate

Bramah Hazela[1]
Amity School of Engineering and Technology
Amity University Uttar Pradesh

Vineet Singh[2]
Amity School of Engineering and Technology
Amity University Uttar Pradesh

Abstract: ***Molecular modeling for drug discovery has been recognized as prominent tool in academia and industry. Different molecular properties were successfully related to molecular structure within framework of (QSAR) quantitative structure activity relationship. The efficacy of such modeling largely depends on accurate molecular property calculation. The Quantum mechanical (QM) calculations being used as an accurate molecular modeling route. The QM calculations provided by QM packages (like Gaussian09) provide handful information for the molecular system. The post processing of large amount of data and derive a scientifically meaningful information is time consuming and tedious process while manual handling of data has risk of errors. Here we proposed an automated tool (Q-MAT) for the quantum descriptor calculations. Q-MAT is basically to help Quantum molecular modular and drug designer to deal with huge unprocessed data to bring scientific outcomes in an automated fashion. The Q-MAT output can be use in MATLAB for standard statistical techniques like multiple linear regression analysis (MLR). The Q-MAT was successfully tested for to generate QSAR model of a series of drug related molecules reported in literature.***

IndexTerms - MATLAB, QSAR, DRUG DISCOVERY.

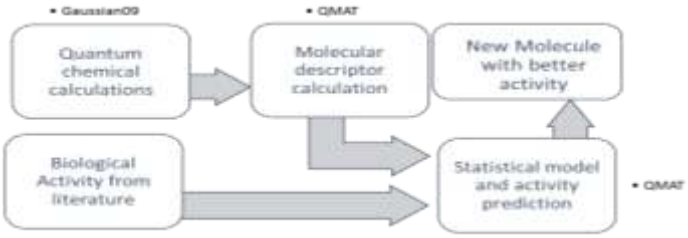
I. INTRODUCTION

The new drug development is very expensive and time-consuming process. On an average a drug discovery takes about 15 years and about 1b USD [1]. Besides a tedious route of drug discovery there is a high risk of Failure. Report says only success rate is less than 10%, further more each developed drug need to get approval from US or European regulators[2]. Thus, there has been continues effort to reduce the cost and time involved in drug discovery. Recent advance in computational modeling got recognition in drug discovery community. The usefulness of molecular modeling has been proven at different step of drug discovery. Mathematical model generation tool and data mining approach will be helpful for drug discovery community. The in-silico quantitative structure activity relationship (QSAR) models mainly used to understand drug action, design new compounds, and screen chemical libraries. In this method the variation of fundamental molecular properties were related as a function of molecular activity. The molecular properties which used as dependent variable, generally referred as descriptor. The descriptor properties may also be correlated with toxicity and property etc within framework of Insilco molecular modeling. The quantitative structure Activity relationships (QSARs), are basically mathematical attempt to relate the structural variation of drug related molecules to its biological activity. The principle of work is an anticipation that structurally similar molecule show similar activity. The quality of model and predictive probability basically depends on two factors: the accuracy of energetics and the structural similarity of molecular series.

The Quantum mechanical (QM) calculations are being used as an accurate molecular modeling route. The QM calculations provided by QM packages like Gaussian09 etc. provide handful information for the molecular system. The post processing of data and derive a scientifically meaningful information is time consuming process and manual handling of data face risk of error. The automated tool (Q-MAT) for the quantum descriptor calculations using post processing analysis of quantum calculations and statistical model generation using MATLAB based standard statistical techniques like multiple linear regression analysis (MLR) is essential requirement.

II. MATERIAL AND METHODS

The fundamental quantum chemical properties [3-9] may be using known quantum chemical packages such as Gaussian 09[10]. The molecular structure and energies were calculated by DFT functional b3lyp in conjunction with 6-31g** basis sets as implemented in Gaussian09 program. Several QM descriptors defined in literature namely, HOMO and LUMO energies, Chemical Hardness, softness, electronegativity, chemical potential and electrophilicity index are useful parameters. In order to derive these descriptors values by solving quantum chemical equations we developed a program which intakes fundamental values from Gaussian output files and help us for data mining and descriptor calculation in an automated manner. The outcomes of Q-MAT (given in supporting info) directly can be imported into mat lab for statistical analysis. The descriptor values were correlated with experimental activity using multiple linear regression analysis. The overall workflow is described in following scheme.



Scheme-1: *The workflow for QSAR modeling using Q-MAT*

Multiple Linear Regression Analysis (MLR). MLR analyses performed using MATLAB. The quantum mechanical descriptors used as independent variables and the pIC50 values as the dependent variable. In the statistical analyses, the systematic search performed to determine the significant descriptors. In order to explore the reliability of the proposed model we used the correlation coefficient as a measure.

III. RESULTS

Q-Mat scripts were developed based on fundamental chemical descriptors widely used in literature [11-15]. The short description of molecular descriptors is included here. Recently Yang and Parr has defined[16] the absolute hardness as equation-1

$\eta = \frac{1}{2}(\delta\mu/\delta N)v(r) = \frac{1}{2}(\delta^2 E/\delta N^2)v(r)$ -eq (1)

Upon finite difference approximation of equation-1 we get equation 2,

$\eta = (IP - EA)/2$ -eq (2)

Similarly the chemical potential and electronegativity[17] as equation-3

□□□□□□□□ = (IP +EA)/2. - eq (3)

Similarly Parr et al have defined the electrophilicity index[18] as equation-4

□□□□□□□□□□□□. - eq (4)

The ionization energy and electron affinity of a molecular system are the fundamental requirements to access these properties. According to Koopmans theorem[19], the IP is simply the eigenvalue of the HOMO with change of sign and the EA is the eigenvalue of the LUMO with change of sign.

There are some energy parameters such as electronic energy and Gibbs free energy is useful parameters which may be defined as equation-5 and 6

E= total electronic energy [21] -eq (5)

G = H-TS - eq (6)

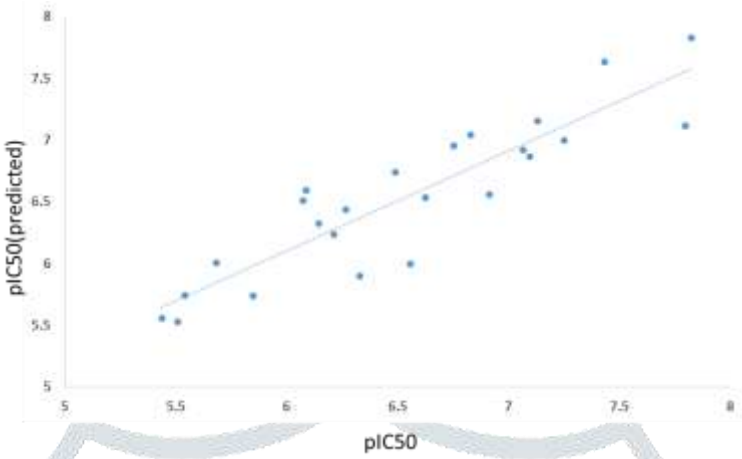
The detail descriptions of these parameters are defined in literature.[12, 20] . The fundamental values to derive these important descriptors may be obtained from quantum chemical calculations using standard program such as Gaussian. The data mining and solution of these complex equations for a series of molecules remains a tedious job. Here Q-MAT is a post processing program developed by us for easy and automated calculation of important descriptors. Q-MAT intakes directly the required input values from Gaussian output files and solve equations 1-6 to provide a tabulated data which may directly imported in MATLAB for the multiple linear regression analysis.

In order to validate the descriptor calculation and adjudge the utility of Q-Mat we have taken a series of drug related molecules especially pyrimidine-urea inhibitors of TNF-α production which is helpful in inflammatory diseases [21, 22] such as Rheumatoid arthritis (RA). The biological activities of these molecules were reported in literature[23]. The quantum chemical calculations performed using Gaussian program. The descriptor calculations were performed using Q-Mat as reported in Table-1 along with the biological activity in terms of pIC50 values. The output of Q-MAT was imported in MATLAB for the multiple linear regression analysis. Based on number of data points 24 we chose maximum 4 descriptors in a forwards selection manner and the best model was found those reported in equation-7. This model gave a high correlation coefficient of 0.8. This model includes energy of highest occupied molecular orbital, Gibbs free energy, electronegativity and electrophilicity index. Based on this model the predicted biological activities are reported in table-1. The accuracy of prediction is revealing in graph-1 plotted between experimental and biological activity. The experimental and biological activities are in a close agreement which proves the utility of QM descriptors. In this way the equation-7 may be use to access the biological activity of further molecule of same series. However the at this stage our interest is to validate Q-MAT based descriptor calculation and statistical model generation. Designing of new molecule is beyond the scope of study however our effort is continue to make advance in Q-MAT capabilities.

Table-1: *The Structure of pyrimidine-urea inhibitors of TNF-α production, QM descriptor values and biological activity in terms of pIC50.*

No.	E_{HOMO}	E_{LUMO}	$E_{Electronic}$	G_{Gibbs}	Hardness	Softness	□	□	□	pIC50	pIC50 (predicted)
1	-0.22091	-0.04974	-1836.51	-1836.14	0.085585	11.68429	-0.13533	0.135325	0.790588	6.914	6.557
2	-0.21046	-0.04486	-1416.23	-1415.83	0.0828	12.07729	-0.12766	0.12766	0.770894	6.754	6.955
3	-0.21543	-0.03576	-1797.17	-1796.83	0.089835	11.13152	-0.1256	0.125595	0.699032	5.509	5.529
4	-0.20858	-0.03569	-1376.9	-1376.52	0.086445	11.56805	-0.12214	0.122135	0.706432	5.438	5.555
5	-0.21229	-0.0352	-1436.81	-1436.47	0.088545	11.29369	-0.12375	0.123745	0.698769	5.541	5.74
6	-0.2176	-0.03273	-1514.04	-1513.68	0.092435	10.81841	-0.12517	0.125165	0.677043	6.074	6.508
7	-0.21709	-0.03302	-1474.92	-1474.57	0.092035	10.86543	-0.12506	0.125055	0.679388	6.268	6.437
8	-0.22163	-0.03585	-2121.59	-2121.14	0.09289	10.76542	-0.12874	0.12874	0.69297	5.848	5.738
9	-0.21781	-0.03238	-1737.28	-1736.9	0.092715	10.78574	-0.1251	0.125095	0.674621	6.144	6.323
10	-0.20869	-0.02129	-1470.85	-1470.38	0.0937	10.67236	-0.11499	0.11499	0.613607	7.432	7.636
11	-0.20598	-0.01824	-1546.06	-1545.59	0.09387	10.65303	-0.11211	0.11211	0.597156	7.824	7.831
12	-0.21692	-0.0301	-1836.5	-1836.14	0.09341	10.70549	-0.12351	0.12351	0.661118	6.627	6.534
13	-0.2125	-0.02471	-1851.79	-1851.39	0.093895	10.65019	-0.11861	0.118605	0.631583	6.83	7.04
14	-0.21215	-0.02441	-1891.11	-1890.69	0.09387	10.65303	-0.11828	0.11828	0.63002	7.252	6.998
15	-0.21511	-0.02753	-2311.39	-2311	0.09379	10.66212	-0.12132	0.12132	0.646764	6.214	6.236
16	-0.21442	-0.02789	-2311.38	-2310.99	0.093265	10.72214	-0.12116	0.121155	0.64952	5.684	6.004
17	-0.21642	-0.02793	-1737.28	-1736.91	0.094245	10.61064	-0.12218	0.122175	0.648178	7.796	7.114
18	-0.20802	-0.02472	-1279.08	-1278.66	0.09165	10.91107	-0.11637	0.11637	0.634861	6.489	6.737
19	-0.20468	-0.02482	-1356.51	-1356.05	0.08993	11.11976	-0.11475	0.11475	0.637996	6.331	5.899
20	-0.20712	-0.02491	-1200.45	-1200.08	0.091105	10.97635	-0.11602	0.116015	0.63671	6.087	6.595

21	-0.20483	-0.02525	-1239.77	-1239.38	0.08979	11.1371	-0.11504	0.11504	0.640606	6.559	5.995
22	-0.21913	-0.03367	-1223.26	-1222.92	0.09273	10.784	-0.1264	0.1264	0.681549	7.066	6.918
23	-0.20771	-0.0249	-1085.93	-1085.6	0.091405	10.94032	-0.11631	0.116305	0.636207	7.097	6.868
24	-0.21557	-0.03035	-1109.96	-1109.64	0.09261	10.79797	-0.12296	0.12296	0.663859	7.131	7.157



$pIC50_{(predicted)}$ = -2929.911851* E_{HOMO} + 1.252891763·10^{-3}* G_{Gibbs} + 4893.444041*□□ + 356.4420112*□□- 257.9825846 -eq(7)
Correlation coefficient r2= .80, standard error= 0.33, n=24

IV. CONCLUSION

The QM based QSAR studies are in practice since last decades. In QSAR analysis large numbers of molecules are considered for extensive calculations. QM calculation generate significant amount of data for a single molecule. The useful information for the QSAR analysis may be derived using Q-Mat in an automated manner. The output of Q-Mat can be imported into statistical package such as MATLAB to perform MLR analysis for QSAR modeling. Though the Q-MAT is capable of handling only Gaussian output as this stage with a limited number of descriptor calculation but effort is continue to add more descriptor calculations and to make it generic to be useful for other popular QM codes

REFERENCES

[1] Kraljevic, S., P.J. Stambrook, and K. Pavelic, 2004, *Accelerating drug discovery - Although the evolution of '-omics' methodologies is still in its infancy, both the pharmaceutical industry and patients could benefit from their implementation in the drug development process.* Embo Reports, **5**(9): p. 837-842.

[2] Kola, I. and J. Landis, 2004, *Can the pharmaceutical industry reduce attrition rates?* Nature Reviews Drug Discovery, **3**(8): p. 711-715.

[3] Karelson, M., Lobanov V.S., and Katritzky A.R., 1996, *Quantum-chemical descriptors in QSAR/QSPR studies.* Chemical Reviews, **96**(3): p. 1027-1043.

[4] Katritzky, A.R., et al., 2004, *Quantitative measures of solvent polarity.* Chemical Reviews, **104**(1): p. 175-198.

[5] Katritzky, A.R., et al., 2010, *Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction.* Chemical Reviews, **110** (10): p. 5714-5789.

[6] Katritzky, A.R., Lobanov V.S., and Karelson M., 1995, *Qspr - the Correlation and Quantitative Prediction of Chemical and Physical-Properties from Structure.* Chemical Society Reviews, **24**(4): p. 279-&.

[7] Colombo, A., et al., 2008, *The proposal of architecture for chemical splitting to optimize QSAR models for aquatic toxicity.* Chemosphere, **72**(5): p. 772-780.

[8] Murugan, R., et al., 1994, *Predicting Physical-Properties from Molecular-Structure.* Chemtech, **24**(6): p. 17-23.

[9] Karelson, M., et al., 1999, QSPR and QSAR models derived using large molecular descriptor spaces. A review of CODESSA applications. Collection of Czechoslovak Chemical Communications, **64**(10): p. 1551-1571.

[10] Frisch M. J., G.W.T., Schlegel H. B., Scuseria G. E., Robb M. A., Cheeseman J. R., Scalmani G., Barone V., Mennucci B., Petersson G. A., Nakatsuji H., Caricato M., X. Li, Hratchian H. P., Izmaylov A. F., Bloino J., Zheng G., Sonnenberg J. L., M. Hada, Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima T., Honda Y., Kitao O., Nakai H., Vreven T., Montgomery J. A., Peralta Jr., J. E., Ogliaro F., Bearpark M., Heyd J. J., Brothers E., Kudin K. N., Staroverov V. N., Kobayashi R., Normand J., Raghavachari K., Rendell A., Burant J. C., Iyengar S. S., Tomasi J., Cossi M., Rega N., Millam J. M., Klene M., Knox J. E., Cross J. B., Bakken V., Adamo C., Jaramillo J., Gomperts R., Stratmann R. E., Yazyev O., Austin A. J., Cammi R., Pomelli C., Ochterski J. W., Martin R. L., Morokuma K., Zakrzewski V. G., Voth G. A., Salvador P., Dannenberg J. J., Dapprich S., Daniels A. D., Farkas Ö., Foresman J. B., Ortiz J. V., Cioslowski J., and Fox D. J., 2009 *Gaussian-09*. 2009, Gaussian, Inc.: Wallingford CT,.

[11] Pasha, F.A., Srivastava H.K., and Singh P.P., 2005, *Comparative QSAR study of phenol derivatives with the help of density functional theory.* Bioorganic & Medicinal Chemistry, **13**(24): p. 6823-6829.

[12] Singh, P.P., Pasha F.A., and Srivastava H.K., 2003, *DFT based atomic softness and its application in site selectivity.* Qsar & Combinatorial Science, **22**(8): p. 843-851.

[13] Singh, P.P., Pasha F.A., and Srivastava H.K., 2004, *Novel application of softness parameter for regioselectivity and reaction mechanism.* Indian Journal of Chemistry Section B-Organic Chemistry Including Medicinal Chemistry, **43**(5): p. 983-991.

[14] Singh, P.P., Srivastava H.K., and Pasha F.A., 2004, *DFT-based QSAR study of testosterone and its derivatives.* Bioorganic & Medicinal Chemistry, **12**(1): p. 171-177.

[15] Srivastava, H.K., Pasha F.A., and Singh P.P., 2005, *Atomic softness-based QSAR study of testosterone.* International Journal of Quantum Chemistry, **103**(3): p. 237-245.

[16] Yang, W.T. and Parr R.G., 1985, *Hardness, Softness, and the Fukui Function in the Electronic Theory of Metals and Catalysis.* Proceedings of the National Academy of Sciences of the United States of America, **82**(20): p. 6723-6726.

[17] Parr, R.G., et al., 1978, *Electronegativity - Density Functional Viewpoint.* Journal of Chemical Physics, **68**(8): p. 3801-3807.

[18] Chattaraj, P.K., Sarkar U., and Roy D.R., 2006, *Electrophilicity index.* Chemical Reviews, **106**(6): p. 2065-2091.

[19] Koopmans, T.C., Physica 1934. **1**: p. 104-113.

[20] Pasha, F.A., et al., 2010, *In silico QSAR studies of anilinoquinolines as EGFR inhibitors.* Journal of Molecular Modeling, **16**(2): p. 263-277.

[21] Dinarello, C.A., 1991, *Inflammatory Cytokines - Interleukin-1 and Tumor-Necrosis-Factor as Effector Molecules in Autoimmune-Diseases.* Current Opinion in Immunology, **3**(6): p. 941-948.

[22] Kontoyiannis, D., et al., 2001, *Interleukin-10 targets p38 MAPK to modulate ARE-dependent TNF mRNA translation and limit intestinal pathology.* Embo Journal, **20**(14): p. 3760-3770.

[23] Mouchlis, V.D., et al., 2012, *Molecular Modeling on Pyrimidine-Urea Inhibitors of TNF-alpha Production: An Integrated Approach Using a Combination of Molecular Docking, Classification Techniques, and 3D-QSAR CoMSIA.* Journal of Chemical Information and Modeling, **52**(3): p. 711-723.