ScienSage

Journal of Advanced Scientific Research

Available online through http://www.sciensage.info

ISSN
0976-9595
Research Article

QSAR MODEL DEVELOPMENT FOR DNA BINDING CONSTANT OF COPPER COMPLEXES USING GA-MLR

Sree Kanth Sivan*, Ashwini Khamamkar, Sudeepa Kunche, Gururaj Somadi

Department of Chemistry, Nizam College, Osmania University, Basheerbagh, Hyderabad, Telangana, India

**Corresponding author: sivan.sreekanth@gmail.com*

ABSTRACT

The ability of copper complexes to bind to DNA and exhibit antitumor and antimicrobial activity is well established. The organic ligands of copper complexes do affect and regulate the activity by changing lipophilicity and intercalation to DNA. Quantitative structure activity relationship (QSAR) analysis is widely applied drug design and development method, here we make an attempt to develop a QSAR model for copper complexes, relating their physicochemical and quantum chemical descriptors with DNA binding constant (Kb). A set of 30 copper complexes with reported Kb values were selected from literature, these were built in Chemdraw and optimized using Semi-emprical PM3 method in Hyperchem 7.5. Different physicochemical and quantum chemical descriptors were calculated for the complexes. Double cross validation tool v2.0 was used to develop Multiple Linear Regression (MLR) model using Genetic Algorithm (GA) based variable selection technique. Different QSAR models were developed using 20 complexes as training set and the remaining 10 complexes were used as test set for external validation of the GA-MLR QSAR model. The best QSAR model obtained showed R^2 of 0.9076 and external validation Q^2_{test} of 0.678. The statistically stable model thus obtained gives us an opportunity to understand and develop better QSAR model based on advanced Quantum chemical protocols.

Keywords: DNA Binding, Quantitative structure activity relationship (QSAR), Multiple Linear Regression (MLR), Genetic Algorithm (GA)

1. INTRODUCTION

Copper is one of the essential metal ions required by all living organisms in trace dietary amount as it is involved in key biological process [1, 2]. Copper complexes have shown importance in metal based drug research, as from literature search it is evident that Cu(II) complexes possess antibacterial [3, 4], antifungal and antioxidant activities [5, 6] and DNA-binding and anticancer studies [7-9]. Antitumour activity of various copper(II) complexes have been evaluated for their potential cytotoxic activities on different cell lines [10, 11]. Primary intracellular target in treating a wide range of diseases and for designing of anticancer agents the interaction of metal complexes to DNA is of extensive interest [12,13]. DNA interacts with metal complexes through intercalation, groove and electrostatic interaction, among the three modes of binding intercalative binding is stronger as it is due to insertion of metal complex into base pairs of DNA [14, 15]. QSAR analysis is widely applied drug design and development method, but very few reports of QSAR studies on metal complexes have been reported in literature. One of the earliest reports of QSAR on metal complexes was by Reichert *et al.* [16] who applied QSAR studies on Copper Azamacrocycles and Thiosemicarbazones complexes for predicting the lipophilicity (log Po/w) of several classes of Cu(II)-chelating ligands. Qian *et al.* have reported artificial neural network (ANN) based QSAR studies on flavonoid-metal complexes for predicting their ˙OH and $O2^{-}$˙ free radial scavenging property and anticancer activity [17-19]. Recently Quang *et al*. have reported quantitative structure property relationship (QSPR) model for predicting stability constant of metal thiosemicarbazone complexes by GA-MLR and GA-ANN [20]. QSAR studies on copper complexes relating to DNA binding of copper complexes have not been reported till date, in this article we report a GA-MLR based QSAR model for predicting the DNA binding constant of copper complexes using various physicochemical and quantum chemical descriptors.

2. MATERIAL AND METHODS

2.1. Data set

Copper complexes with experimental DNA binding activity reported as Kb values were searched from literature [21-37]. Total 30 complexes with Kb values determined in similar experimental protocol were selected. The structure of the copper complexes is provided in Fig. 1:

2.2. Calculation of the electronic structure of Copper complexes

All the structures of complexes were built in Chemdraw and pre-optimized using the Molecular Mechanics Force Field (MM+) method. The resulted minimized structures were further refined using the semi empirical PM3 Hamiltonian, which is included in HyperChem release 7.5. A gradient norm limit of 0.001 kcal/Å for geometry optimization was implemented. The QSAR properties module from HyperChem was used to calculate polarizability (Plz), the molar refractivity (MR), partition coefficient octanol/water(log P), molar volume (MV). The Quantum Chemical descriptors HOMO and LUMO energies, total energy (TE), dipole moment (DM), and Hydration energy (HE)were also computed using HyperChem.

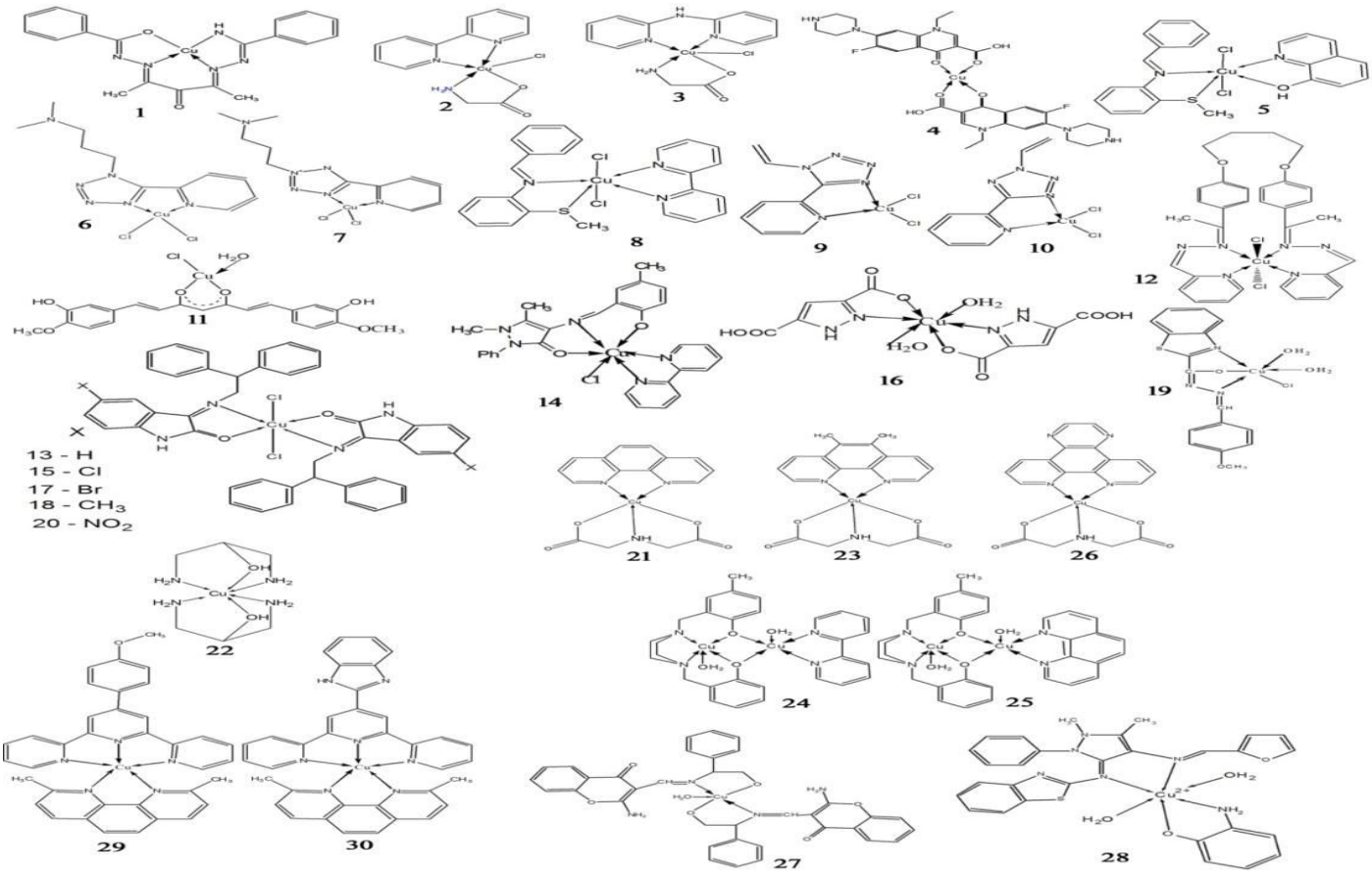

Fig. 1: Structure of Copper complexes

2.3. Development and selection of a MLR model using Double cross-validation technique

DNA binding constant (Kb) values were converted into log Kb and data set was sorted according to their log Kb values. From a range of log Kb value one or two molecules where selected randomly to divided the whole data set into training set consisting of 20 molecules and test set of 10 molecules. Double cross validation tool v2.0 was used to develop Multiple Linear Regression (MLR) model using Genetic Algorithm (GA) based variable selection technique. Descriptor and response values of the training and test set were provided, then k-fold cross validation was set to 10 to develop 'k' models. GA was set with 1000 Iterations, Mutation probability of 0.3 and 50 equation selected from each generation based on FF1(MAE based) fitness

function. MLR models were generated in the first step based on above settings by increasing the number of descriptors in equation incrementally till no further improvement in Q^2 values was observed. In the second step an optimum model was selected with least MAE (validation set) value.

3. RESULTS & DISCUSSION

Copper complexes used in the study were obtained from various literature sources, the complexes has wide variety of ligands like 2-(1H-tetrazol-5-yl)pyridine, hydrazinylidene, pyridinyl, dihydroquinoline-3-carboxylic acid, Azamacrocycles, 1,3-dihydro-2*H*-indol-2-one, 1,3-diaminopropan-2-ol, bipyridyl, benzothioazole, phenanthroline, oxine. DNA binding ability of all these complexes was reported in binding constant (Kb) values determined by UV spectroscopy studies using CT-DNA. The log Kb values for these complexes ranged from 2.778 to 5.662, based on the binding constant the molecules were randomly divided into training and test set comprising of 20 and 10 complexes in each respectively, with similar distribution of activity range. Physicochemical descriptors like polarizability (Plzb), the molar refractivity (MR), partition coefficient octanol/water (log P), molar volume (MV), are know to be used in QSAR model development. The Quantum Chemical descriptors HOMO and LUMO, total energy (TE), dipole moment (DM), and Hydration energy(HE)are also used in developing QSAR models to predict the activity of a molecular species [38]. GA-MLR model was developed using Double Cross Validation tool v2.0 [39].

Two nested cross-validation loop of internal and external cross-validation is carried out in this method. Training set is used in internal loop for building models and selecting of a good model where in the set is repeatedly split into calibration and validation sets, model is developed using calibration set and error in the model is analyzed by validation set and the model with least error is selected. This model is assessed by the test set in the external loop where in the predictive power of the model is examined [39]. The splitting of the training set into calibration and validation sets eliminates the bias introduced in variable selection for a single training set of fixed and limited composition [40]. Genetic Algorithm based variable selection was applied for development of MLR model and the Model with the least MAE (Validation set) was selected. The statistical parameters of the developed model are provided in Table 1.

Table 1: Statistical parameters of QSAR models by GA-MLR using Double Cross Validation tool v2.0

Model. No	Descriptors	R^2 (Train)	Q^2_{LOO} (Train)	Q^2 (Test)	MAE(95% data;Train)	MAE(95% data;Test)	$AvgR_m^2$ (Test)	PredQuality (Test)
1	HE LUMO-E	0.6625	0.5296	0.1958	0.3339	0.2255	0.1912	BAD
2	MV HE LUMO-E	0.7186	0.5839	0.4393	0.3153	0.1783	0.413	GOOD
3	MV HE LUMO-E TE	0.847	0.7639	0.5645	0.2469	0.2002	0.5372	MODERATE
4	MV HE Log P LUMO-E TE	0.9076	0.8222	0.6786	0.2297	0.2284	0.6098	GOOD

The value of regression coefficient R^2 of 0.9076 and high value of Q^2_{LOO} of 0.8222 show that models have good internal predictive ability and stability. The true predictive power of the QSAR models is determined by predicting the DNA binding constant of complexes not used in the model development (external prediction). The values of Q^2 (Test) of 0.6786 and average Rm^2(Test) of 0.6098 high and rather similar confirming the good external predictive ability of the developed GA-MLR model. The plot of experimetnal versus predicted log Kb values is shown in Fig. 2.

The equation obtained for best model 4 includes five descriptors that includes two physicochemical descriptor Molar volume (MV), and partition coefficient

(log P) and three quantum chemical descriptors total energy, Hydration Energy (HE) and energy of lowest unoccupied molecular orbital (LUMO-E). The possibility of co-linearity was avoided by applying a inter-correlation coefficient cut-off of 0.95 with variance 0.0001, the correlation matrix is shown in Table-2 clearly indicate that the descriptors used in GA-MLR model are poorly correlated with each other.

The DNA binding constant (log Kb) and descriptors used in model generation of the copper complex is provided in table 3 along with predicted log Kb value for model equation and the residual value (difference between experimental and predicted log Kb). Complex 1 in training set showed the highest deviation of 0.479 and complex 14 in test set showed a deviation of 0.371, this clearly indicates the predicting power of the GA-MLR model obtained. The residuals to the experimental log Kb values by GA- MLR model for training and test set is shown in Fig.3:

Positive signs in the equation for coefficients of TE and MV show that these descriptors are in direct correlation with the DNA binding constant while negative coefficient of HE, log P and LUMO indicates its indirect correlation. Hydration energy and LUMO-E parameters have more significant impact on the model equation relating to DNA binding constant. The complexes with lower Hydration energy (more negative value) and LUMO-E value should have better DNA binding ability.

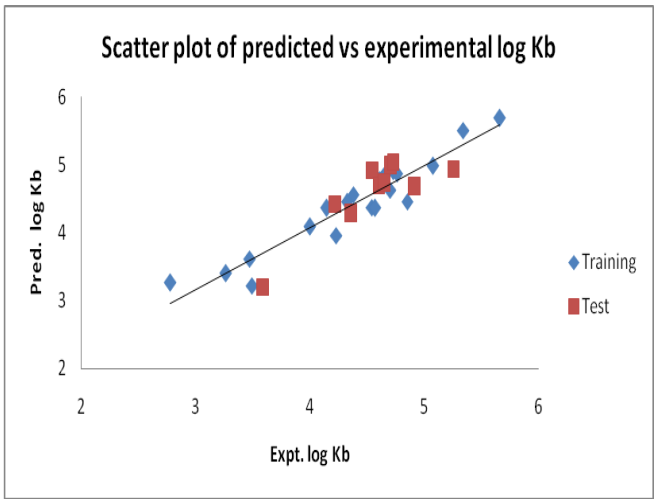

Fig. 2: Scatter plot experimental vs predicted DNA binding constant (log Kb) of Best GA-MLR model

MLR equation for Model 4:

log Kb = [0.0046(±0.0008)**MV**]-[0.067(±0.0 091) **HE**]-[0.0804(±0.0265)**logP**]-[0.2462(±0.0357) **LUMO-E**] + [0.0369(±0.0069) **TE**] + [2.2916(±0.2273)]

Table 2: Correlation matrix showing correlation between descriptors used in MLR

	Log Kb	TE	MV	HE	log P	LUMO-E
Log Kb	1	0.438	0.43	0.568	0.063	0.661
TE	0.438	1	0.91	0.231	0.132	0.337
MV	0.43	0.91	1	0.045	0.33	0.31
HE	0.568	0.231	0.045	1	0.344	0.149
log P	0.063	0.132	0.33	0.344	1	0.25
LUMO-E	0.661	0.337	0.31	0.149	0.25	1

Table 3: Calculated descriptor values, experimental and predicted log Kb values of complexes

Complex	MV	HE	Log P	HOMO-E	LUMO-E	TE *10^3	Expt. log Kb	Predicted log Kb	Residual
1	846.79	-8.50	-2.32	-8.599	-2.069	-113.688	2.778	3.257	-0.479
2	748.98	-9.09	-0.79	-8.146	-2.171	-96.154	3.265	3.396	-0.131
3	1150.37	-6.56	-0.46	-7.698	-1.550	-130.763	3.470	3.616	-0.146
4	782.44	-10.98	0.87	-8.647	-1.459	-100.248	3.491	3.217	0.275
5*	925.49	-6.70	-2.01	-8.553	-2.008	-120.591	3.591	3.204	0.387
6	1464.12	-8.43	2.24	-11.302	-4.822	-176.483	4.000	4.086	-0.086
7	1172.97	-4.48	0.45	-10.648	-4.945	-130.313	4.145	4.360	-0.216
8*	569.10	-16.41	-1.88	-12.856	-4.620	-77.834	4.223	4.426	-0.203
9	943.07	-10.35	-4.76	-6.540	-3.747	-126.571	4.230	3.958	0.272
10	706.50	-9.89	2.93	-11.436	-6.518	-84.736	4.322	4.446	-0.124
11*	692.98	-12.80	3.36	-12.556	-5.434	-84.737	4.362	4.278	0.084

12	1026.50	-12.73	4.79	-12.541	-5.717	-117.368	4.378	4.558	-0.179
13	913.60	-5.41	2.84	-11.342	-5.786	-99.832	4.544	4.369	0.175
14*	1676.50	-23.40	-0.49	-11.325	-2.804	-199.993	4.550	4.921	-0.371
15	1384.78	-13.73	-1.87	-13.971	-6.317	-187.794	4.568	4.357	0.211
16*	1389.60	-8.07	0.36	-9.831	-4.991	-155.184	4.602	4.698	-0.096
17	1668.92	-24.41	-1.31	-10.815	-3.466	-210.879	4.611	4.781	-0.171
18	1428.51	-13.04	-2.21	-10.622	-8.336	-193.192	4.643	4.838	-0.194
19*	898.08	-9.62	3.26	-11.336	-6.509	-99.907	4.653	4.721	-0.068
20	1895.99	-7.79	6.97	-10.645	-5.733	-210.157	4.702	4.631	0.071
21*	1234.51	-32.64	-0.76	-8.726	-0.898	-147.944	4.708	4.980	-0.273
22	1845.73	-12.29	6.66	-11.439	-5.402	-203.243	4.725	4.900	-0.175
23*	1920.75	-11.78	6.22	-10.945	-6.643	-217.102	4.725	5.041	-0.316
24	1927.65	-18.26	1.17	-11.908	-5.388	-236.994	4.762	4.869	-0.108
25	1477.37	-7.06	1.10	-10.026	-3.806	-161.366	4.851	4.455	0.397
26*	1951.01	-11.78	6.76	-11.242	-5.006	-218.852	4.916	4.669	0.248
27	892.05	-36.94	-2.86	-12.098	-4.481	-141.492	5.079	4.982	0.097
28*	1488.20	-8.55	0.53	-8.259	-4.979	-161.318	5.255	4.941	0.315
29	1488.07	-18.19	-3.43	-10.337	-4.697	-170.145	5.342	5.509	-0.167
30	1011.72	-27.27	-0.45	-11.147	-6.283	-126.435	5.663	5.690	-0.027

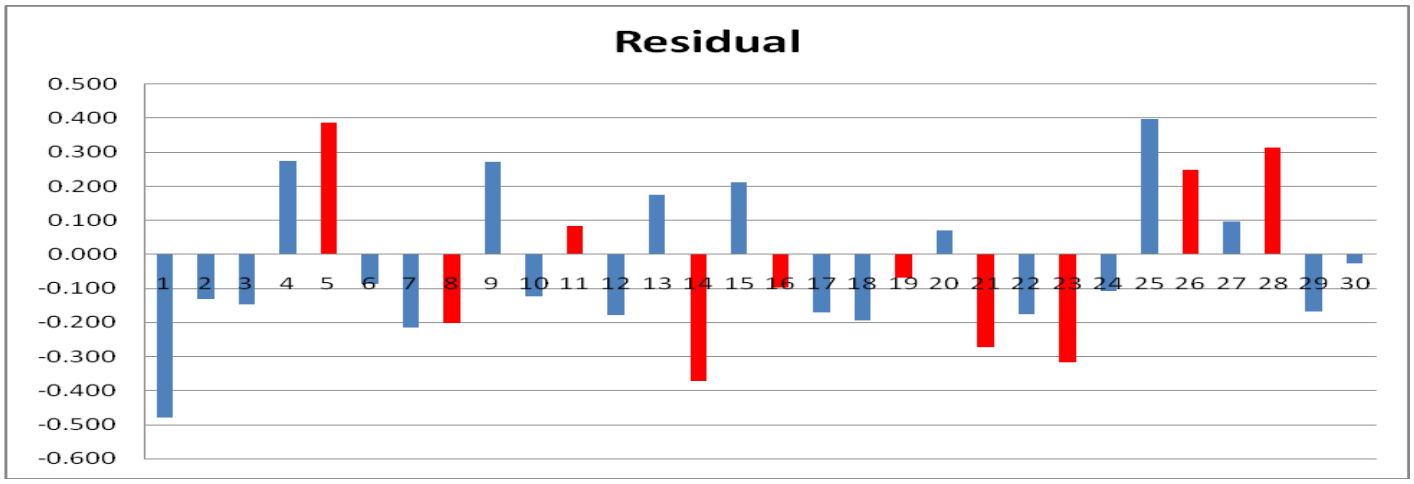

Fig. 3: Plot of residuals to the experimental log Kb values by GA- MLR model (Blue – Training and Red – Test)

4. CONCLUSION

GA-MLR based QSAR analysis was carried out on 30 copper complexes having diverse structure. A stastically stable and realiable QSAR model was developed for understanding the DNA binding constatn Kb of copper complexes based on physicochemical and quantum chemical descriptors. Semi-empirical PM3 calculations was carried out on the copper complexes, to calcualte HOMO-LUMO energies, total energy, hydration energy which served as quantum chemical descriptor, and the physicochemcial descriptor consisted of log P and molar volume. LUMO energy was considered to be important parameters that was found to be relateto DNA binding constant of the complexes. GA-MLR based model with very good statistical fit as evident from R^2 of 0.9076 and high value of Q^2_{LOO} of 0.8222. The model was validated using external test set that provide good Q^2 (Test) of 0.6786 and average R_m^2(Test) of 0.6098 confirming predictive ability of the developed GA-MLR model.

5. ACKNOWLEDGEMENT

The authors are grateful to Kunal Roy, developers of Double Cross Validation tool v2.0. for providing the software (http://teqip.jdvu.ac.in/QSAR_Tools/), the author also grateful to Department of Chemistry, Nizam

College and Osmania University for Computer Lab and software facilities.

Conflict of interest

Authors have no conflict of interest for present study.

6. REFERENCES

1. Zoroddu MA, Aaseth J, Crisponi G, Medici S, et.al. *J. Inorg. Biochem.,* 2019; **195:**120-129
2. Sorenson JR. *Prog. Med. Chem.,* 1989; **26:**437-568.
3. Patel MN, Dosi PA, Bhatt BS. *Polyhedron,* 2010; **29:**3238-3245.
4. Saha DK, Sandbhor U, Shirisha K, Padhye S, et. al. *J. Inorg. Biochem.,* 1997; **65:**183-189.
5. Dimiza F, Fountoulaki S, Papadopoulos AN, Kontogiorgis CA, et.al, *Dalton Trans.,* 2011, **40:**8555-8568.
6. Dimiza F, Perdih F, Tangoulis V, Turel I, et.al., *J. Inorg. Biochem.,* 2011; **105**:476-489.
7. Powell AK. *Bioorg. Med. Chem. Lett.,* 2004; **14**:3027-3032.
8. Barve A, Kumbhar A, Bhat M, Joshi B, et.al., *Inorg. Chem.,* 2009; **48:**9120-9132.
9. Weder JE, Dillon CT, Hambley TW, et al. *Coord. Chem. Rev.,* 2002; **232(1-2):**95-126
10. Farrell N. *Coord. Chem. Rev.,* 2002; **232:**1-4.
11. Li MX, Zhang LZ, Chen CL, Niu JY, Ji BS. *J. Inorg. Biochem.,* 2012; **106**:117-125.
12. Zhang QL, Liu JG, Chao H, Xue GQ, Ji LN. *J. Inorg. Biochem*, 2001; **83(1):**49-55.
13. Kramer R. *Coord. Chem. Rev*., 1999; **182**:243-261.
14. Delaney S, Pascaly M, Bhattacharya PK, Han K, Barton JK. *Inorg. Chem.,* 2002; **41(7):**1966-1974.
15. Bergeron KL, Murphy EL, Olulade M, Muñoz LD, et.al., *Mutat. Res.,* 2009; **673(2):**141-148.
16. Peter W, Yoo J, Welch MJ, Reichert DE. *J. Med. Chem.,* 2005; **48**:5561-5569.
17. Wang BC, Qian JZ, Fan Y, Tan J. *J. Mol. Struct*., 2017; **1075:**204-212.
18. Qian JZ, Wang BC, Fan Y, Tan J, Huange J. *J. Coord. Chem.,* 2014; **67(17):**2867-2884.
19. Qian JZ, Wang BC, Fan Y, Tan J, Yang X. *J. Struct. Chem.,* 2015; **56(2):**338-345.
20. Quang NM, Mau V, Nhung NTA, An TNM, Tat PV. *J. Mol. Struct.*, 2019; **1195:**95-109.
21. Vafazadeh R, Willis AC, Heidari MM, Hasanzade N. *Acta. Chim. Slov.,* 2015; **62:**122-129.
22. Mohemeda MS, Shoukry A, Ali AG. *Spectrochimica Acta (Part A),* 2012; **86:**562-570.
23. Zivec P, Perdih F, Turel I, Giester G, Psomas G. *J. Inorg. Biochem.,* 2012; **117:**35-47.
24. Amaravathi C, Mustafa S, Babu MSS. *Ind. J. Adv. Chem. Sci.,* 2016; **4(4):**394-400.
25. Kulaksizoglu S, Gokce C, Gup R. *Turk. J. Chem.,* 2012; **36:**717-733.
26. Gomathi R, Ramu A. *Inter. J. Innov. Res. Sci. Eng. Tech.,* 2013; **2(9):**4852-4865.
27. Arjmand F, Sayeed F. *J. Mol. Struct.,* 2010; **965:**14-22.
28. Anbu S, Swamy M. *Polyhedron,* 2011; **30:**123-131.
29. Arjmand F, Sayeed F, Muddassir M. *J. Photochem. Photobio. B:Bio.,* 2011; **103:**166-179.
30. Rajalakshmi S, Weyhermüller T, Dinesh M, Nair BU. *J. Inorg. Biochem.,* 2012; **117:**48-59.
31. Neelaveni R, Vasantha S, Keerthana R, Sivakolunthu R, Angeline T. *Asian J. Pharma. Clin. Res.,* 2016; **9(3):**277-281.
32. Rajesh J, Rajasekaran M, Rajagopal G, Athappan P. *Spectrochimica. Acta Part A: Mol. Biomol. Spec.,* 2012; **97:**223-230
33. Anupama B, Aruna A, Manga V, Sivan SK, et.al; *J. Fluoresc.,* 2017**; 27:**953-965.
34. Srivastava AK, Kumari N, Khan RA, Rai R, et.al. *Ind. J. Chem.,* 2013; **52A:**835-844.
35. Sudeepa K, Aparna B, Ravi M, Ranjithreddy P, et.al., *IJRPC,* 2015; **5(4):**668-682.
36. Selvakumar B, Rajendiran V, Maheswari PU, Evans HS, Palaniandavar M. *J. Inorg. Biochem*, 2006; **100(3):**316-330.
37. Leelavathy C, Arulantony S. *Asian. J. Chem.,* 2015; **27(7):**2527-2532.
38. Musa AY, Jalgham RTT, Mohamad AB. *Corros Sci.,* 2012; **56:**176-183.
39. Roy K, Ambure P. *Chemom. Intell. Lab. Sys.,* 2016; **159:**108-126.
40. Baumann D, Baumann K. *J. Cheminformatics*, 2014; **6(1):**47.