Fine Chemical Engineering
http://ojs.wiserpub.com/index.php/FCE/

UNIVERSAL WISER
PUBLISHER

Research Article

Quantitative Structure-Property Relationship of the Rare-Earth Elements-Dibutyl Dithiophosphate Derivative Complexes Using Principal Component Analysis

Nurdeni[1], Heru Agung Saputra[2*], Atje Setiawan Abdullah[3], Budi Nurani Ruchjana[3], Husein Hernandi Bahti[3], Hedi[4]

[1]Indraprasta PGRI University, Jakarta, 12530, Indonesia
[2]Department of Chemistry and Chemistry Institute for Functional Materials, Pusan National University, Busan, 46241, Republic of Korea
[3]Faculty of Mathematics and Natural Sciences, Padjadjaran University, Jatinangor, 45363, Indonesia
[4]Department of Energy Conversion Engineering, Politeknik Negeri Bandung, 40559, Indonesia
Email: heruagungsaputra@pusan.ac.kr

Received: 21 February 2024; **Revised:** 9 June 2024; **Accepted:** 25 June 2024

Abstract: The separation of rare-earth elements (REEs) has increasingly developed, especially using a complexing ligand of dibutyl dithiophosphate (DBDTP) that has numerous advantages as an extractant in the extraction process. Through technology development, this separation utilizes computational chemistry design to scheme the DBDTP ligand and its derivatives. One of the computational chemistry applications is a quantitative structure-property relationship (QSPR), which is useful for designing ligand derivatives by calculating molecular descriptors and connecting molecular structure with its physicochemical properties. In the present study, we aimed to get a dominant factor affecting complex stability formed from the REEs with DBDTP ligands and the REEs with DBDTP derivative ligands using principal component analysis for the QSPR study. The analysis results demonstrated that the stability of gadolinium, terbium, and dysprosium complex compounds was influenced by seven, seven, and six factors with a total variance of 93.93, 93.17, and 91.63%, respectively.

***Keywords*:** complex compound, dibutyl dithiophosphate, DBDTP derivative, principal component analysis, quantitative structure-property relationship, rare-earth elements, separation

Abbreviations

AIC	Akaike's information criterion
DBDTP	dibutyl dithiophosphate
MLR	multiple linear regression
MSE	mean square error
PCA	principal component analysis
PC	principal component
PCR	principal component regression

DOI: https://doi.org/10.37256/fce.5220244491

PLSR	partial least square regression
QSAR	quantitative structure-activity relationship
QSPR	quantitative structure-property relationship
REEs	rare-earth elements
SIC	Schwarz information criterion.

1. Introduction

Along with technological development, the advance in separation of complex compounds has been progressively carried out. Since elements in the complex compounds have similar physicochemical properties, they are difficult to separate individually. An example in this regard is rare-earth elements (REEs). The main issue in obtaining pure REEs is to separate a component from others due to similar characteristics.[1-2] Consequently, several techniques like crystallization, fractionation, ion exchange, and solvent extraction are tough. However, solvent extraction is the most successful technique used so far.[3]

To extract REEs, plentiful studies have been carried out using organic solvents containing complex-forming ligands. Solvent extraction is one of the robust techniques to separate the REEs. The innovation in extraction is the development of the extractant that can form complexes with transition metals and the REEs.[4] In a chromatography study of transition metals, dialkyl dithiophosphate was used as a complexing ligand in the extraction stages of those metals.[5] On the other hand, the separation ability of dibutyl dithiophosphate (DBDTP) as a complexing ligand in the extraction of gadolinium and samarium by the solvent extraction technique is still lacking. Therefore, it is of great importance to develop new derivatives of the DBDTP ligand as a better chelating agent for gadolinium. The fact is that lanthanide elements can form a better chelate compound when combined with phosphate acid, phosphoric acid, or tributyl phosphate. Accordingly, such compounds can be used to extract the lanthanide elements.[6-7]

Quantitative structure-property relationship (QSPR) is one of the applications of computational chemistry. It can be used to help design derivatives of a ligand through the calculation of molecular descriptors, constitutional, topology, electrostatic, geometry, quantum chemistry, and thermodynamics. The relationship between molecular structure and its physicochemical properties can also be predicted.[8] The QSPR is a math equation that connects the response of characteristics with the structure information and physicochemical characteristics in the form of numerical descriptors. It is a statistical approach that connects the physicochemical properties with the structure of a compound. Generally, a relationship studied in the QSPR/quantitative structure-activity relationship (QSAR) analysis is multivariate, yet it also can be done through the relationship among the variables. The statistical approach used to run the QSPR is multiple linear regression (MLR), principal component analysis (PCA), principal component regression (PCR), and partial least square regression (PLSR). These methods have been successfully set to predict a wide range of chemical and physical properties.[9-10]

2. Literature review

2.1 *Quantitative structure-property relationship (QSPR)*

The QSPR is based on the math relationship between the characteristics and one or more descriptors linked to the structure of molecules. The QSPR models are an empirical equation used to predict various physical properties or thermodynamics of a molecule expressed in the following MLR equation (Equation 1).

$$P = a + b \cdot D_1 + c \cdot D_2 + d \cdot D_3 + \ldots \tag{1}$$

where P is the physical property; a, b, c, … is coefficient regression; and D_1, D_2, D_3, ... is the parameters derived from the structure of a molecule (also called descriptors).[11]

2.2 *Rare-earth elements (REEs)*

The REEs are a scarce element of its existence, yet they have high demand. They consist of lanthanide: La (lanthanum), Ce (Cerium), Pr (Praseodymium), Nd (Neodymium), Pm (Promethium), Sm (Samarium), Eu (Europium), Gd (Gadolinium), Tb (Terbium), Dy (Dysprosium), Ho (Holmium), Er (Erbium), Tm (Thulium), Yb (Ytterbium), and Lu (Lutetium). Besides, Y (Yttrium) and Sc (Scandium) which respectively have the atomic numbers 39 and 21 are also classified into the REEs due to often being found along with lanthanide.[12] In the future, they are strategic elements that are necessary to develop their separation approach continually. The REEs have very similar properties from one element to others causing geologically they are not found in the form of free elements but rather in the form of complex compounds. The main source of the REEs comes from the mineral bastnaesite, monazite, and xenotime.[13-14]

2.3 *Dibutyl dithiophosphate (DBDTP)*

The ligand of DBDTP is used as a complexing agent in the separation and purification of metal ions through the extraction method. Meanwhile, the dialkyl dithiophosphate ligand combines with metal ions to form a neutral complex with a low water solubility due to its hydrophobicity and low polarity. On the other hand, di-n-butyl dithiophosphate acid extracts more metal ions as compared with diethyl dithiophosphate acid. Therefore, di-n-butyl dithiophosphate or derivatives of dithiophosphate with a longer alkyl chain have several advantages when used as an extractant. The distribution ratio in the extraction will be higher when organic solvents with oxygen atoms in the structure like ethyl acetate and isopropyl ether are used.[4,15-16] Figure 1 shows the chemical structure of the DBDTP ligand.

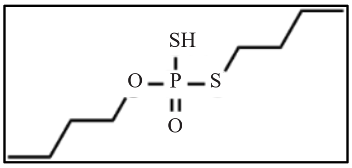


Figure 1. The chemical structure of dibutyl dithiophosphate (DBDTP)

2.4 *Principal component analysis (PCA)*

One of the methods for running the QSPR is PCA. The PCA will be obtained from new free variables that are not correlated, free of each other, but can absorb most of the information contained in the original variables or contribute to the whole of the variant variables. The following are the advantages of using the PCA:[17-18] (1) Being able to eliminate the correlation (correlation = 0) cleanly, so the multicollinearity issue can be completely resolved; (2) Being able to be used for all conditions of the data; (3) Being able to be used without reducing the number of the origin variables; and (4) Despite being highly difficult to use, the PCA provides a more accurate conclusion than that of other methods.

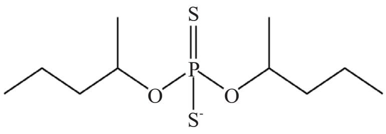
3. Materials and equipment

The hardware consists of a Laptop with a Processor Core i7 CPU 980@3,33 GHz, Memory RAM 8 GB, GPU NVIDIA Ge Force GTX 970 4 GB, Hard drive of 2 TB. The software used in this study was ChemBio 3D Ultra 12.0, BIOVA Discovery Studio 2016, AMBER14, AmberTools 14, MOPAC 2012, Molecular Dynamics (MD), and MATLAB-based algorithm PCA. Meanwhile, the materials used were the structure model of DBDTP ligands and DBDTP derivative ligands obtained from the Statistica package program.

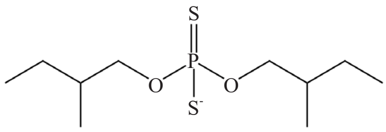
4. Research methods

4.1 *Modeling of DBDTP derivative design*

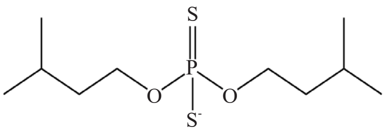
The DBDTP derivatives' structure was drawn manually using ChemBio3D 12.0 with minimum energy. Those twenty different structures of ligand derivatives are shown in Figure 2.



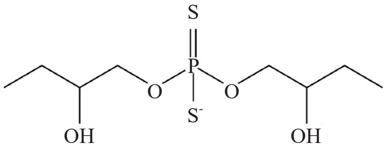
(1)
O,O-di(pentane-2-yl) phosphorodithioate



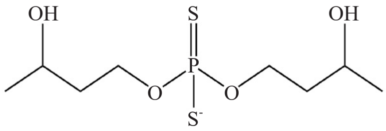
(2)
O,O-bis(2-methyl butyl) phosphorodithioate



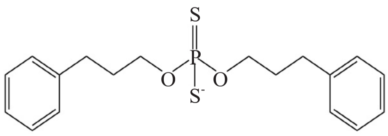
(3)
O,O-di-isopentyl phosphorodithioate



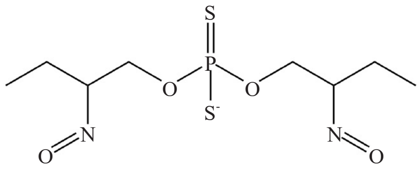
(4)
O,O-bis(2-hydroxy butyl) phosphorodithioate



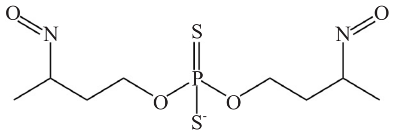
(5)
O,O-bis(1-hydroxybutyl) phosphorodithioate



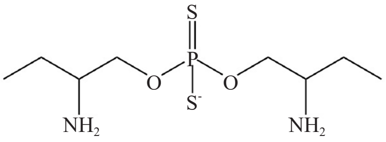
(6)
O,O-bis(3-phenylpropyl) phosphorodithioate



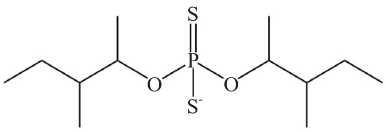
(7)
O,O-bis(2-nitrosobutyl) phosphorodithioate



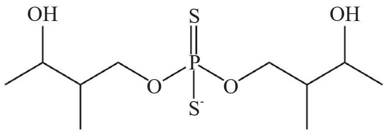
(8)
O,O-bis(3-nitrosobutyl) phosphorodithioate



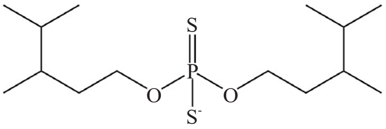
(9)
O,O-bis(2-aminobutyl) phosphorodithioate



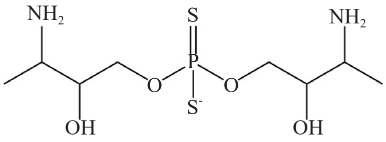
(10)
O,O-bis(3-methylpentan-2-yl)
phosphorodithioate



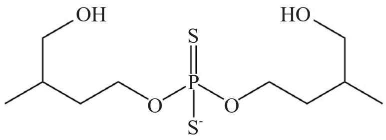
(11)
O,O-bis(3-hydroxy-2-methylbutyl)
phosphorodithioate



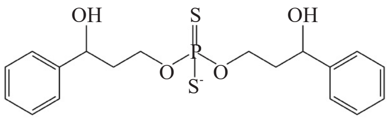
(12)
O,O-bis(3,4-dimethylpentyl)
phosphorodithioate



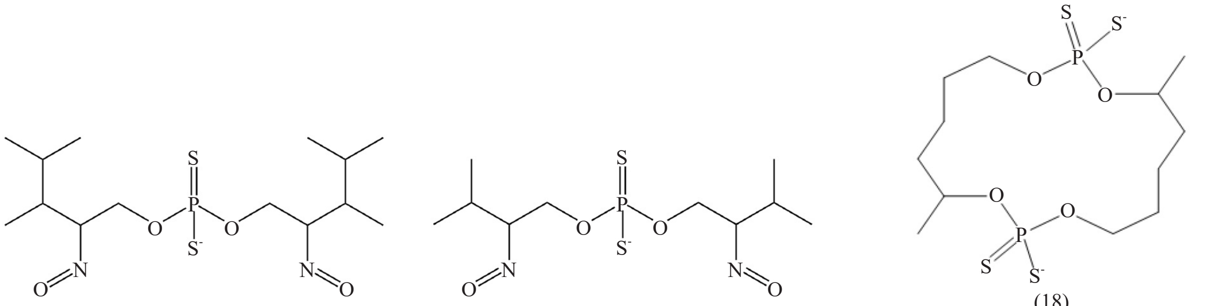
(13)
O,O-bis(3-amino-2-hydroxybutyl)
phosphorodithioate



(14)
O,O-bis(4-hydroxy-3-methylbutyl)
phosphorodithioate



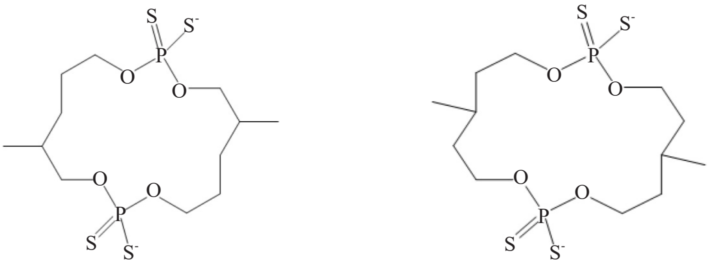
(15)
O,O-bis(3-hydroxy-3-phenylpropyl)
phosphorodithioate



(16)
O,O-bis(3,4-dimethyl-2-nitrosopentyl) phosphorodithioate

(17)
O,O-bis(3-methyl-2-nitrosobutyl) phosphorodithioate

(18)
4,12-dimethyl-1,3,9,11-tetraoxa-2,10-diphosphacyclohexadecane-2,10-bis(thiolate) 2,10-disulfide



(19)
5,13-dimethyl-1,3,9,11-tetraoxa-2,10-diphosphacyclohexadecane-2,10-bis(thiolate) 2,10-disulfide

(20)
6,14-dimethyl-1,3,9,11-tetraoxa-2,10-diphosphacyclohexadecane-2,10-bis(thiolate) 2,10-disulfide

Figure 2. The chemical structure of DBDTP derivatives

4.2 *Modeling of metal complex structure and calculation of descriptors*

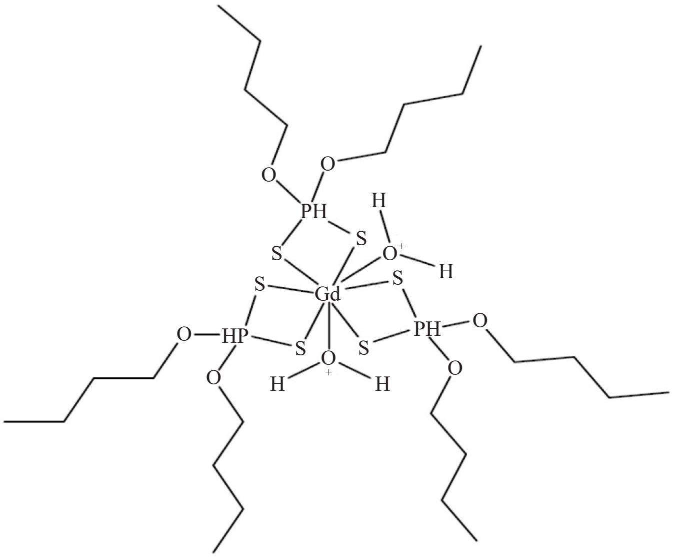


Figure 3. Molecular structure of $[Gd(DBDTP)_3(H_2O)_2]$ complex compound

All of the DBDTP derivatives were linked to each central atom, including Gd, Tb, and Dy on the ChemBio3D 12.0, then they were optimized by MOPAC 2012 with the keyword command "SPARKLE" using the PM7 method. The optimization results were in the form of data descriptors such as the total energy, bond length, bond angle, etc. A representative structure of the metal-ligand complex is shown in Figure 3, where the structure formed in the reaction between the Gd(III) ion and the DBDTP ligand was proposed through molecular modeling.

4.3 *Preparation of data matrix*

All the descriptor data obtained was then arranged to be a data matrix with the dimension of m × n, where m is the number of rows (the result of descriptor calculation, 20 rows) and n is the number of columns (complex compounds of gadolinium, terbium, and dysprosium, 57 columns for each).

4.4 *Data processing using PCA and PCR*

Firstly, the data matrix was processed by PCA using the MATLAB-based algorithm program. After grouping, the PCA data were analyzed by PCR. The error and good models obtained were then used to estimate which ligand is the most suitable for analysis of the Gd complex. At last, the experimental data were used to validate the mathematical model by estimating the correlation and ratio graph.

5. Results and discussion

Based on the modeling of 20 artificial DBDTP derivative ligands and 57 variables of Dy, Gd, and Tb descriptor data, it could be obtained descriptor data for each with the dimension of 20 × 57. Of the corresponding data, the structure descriptors possessed similar data between metals and ligands. Meanwhile, the energy descriptors had a far gap between ligands and metals yet owned a clear characteristic difference between the complexes either energy descriptors or structure descriptors. Since the measurement unit of the descriptor was different, the descriptor data were standardized to avoid the variance between different predictors, so the component determination used a correlation matrix. By applying the correlation matrix, eigenvalues (variance) of descriptor principal components (PCs) of Gd, Tb, and Dy were determined. Afterward, correlation matrix Z, eigenvalues, eigenvector, number of components, and component elements of Gd, Tb, and Dy are presented in Table 1.

For Gd descriptors, the first eigenvalue (variance) of 38.3673 could explain the data variance of 67.50%, which is the first principal component (PC). The second PC with the second largest eigenvalue of 4.5799 could describe a data variance of 8.06% and a total second variance of 75.55%. The third PC with the third largest eigenvalue of 2.9262 could account for a data variance of 5.15% and a total second variance of 80.70%. The fourth PC with the fourth largest eigenvalue of 2.3586 could explain a data variance of 4.15% and a total second variance of 84.85%. The fifth PC with the fifth largest eigenvalue of 2.0351 could define a data variance of 3.58% and a total second variance of 88.43%. The sixth PC with the sixth largest eigenvalue of 1.6848 could explain a data variance of 2.96% and a total second variance of 91.40%. The seventh PC with the seventh largest eigenvalue of 1.4425 could explain a data variance of 2.54% and a total second variance of 93.93%. In this regard, only seven components had eigenvalues larger than 1, where such eigenvalue can retain all factors. The logic dictates that factors worth retaining should have at least a larger variance than any of the original measured variables contained in the factor. Otherwise, an eigenvalue of less than one implies that the score on that component will have negative reliability.[19] For k (clustering) = 7, the scree plot, a line plot of the eigenvalues of factors or PCs in an analysis, owned a sloping shape on the right side as shown in Figure 4 (A). The number of components that explain 57 descriptors was seven components, where the eigenvector (components) suitable with those of seven eigenvalues from the largest to the smallest was C_1, C_2, C_3, C_4, C_5, C_6, and C_7, respectively.

Table 1. Eigenvalues (variance) of complex data correlation matrix of gadolinium, terbium, and dysprosium

Gd			Tb			Dy		
Eigenvalue	Proportion	Cumulative	Eigenvalue	Proportion	Cumulative	Eigenvalue	Proportion	Cumulative
38.3673	67.50%	67.50%	38.872	68.38%	68.38%	39.154	68.82%	68.82%
4.5799	8.06%	75.55%	4.3829	7.71%	76.09%	4.307	7.57%	76.39%
2.9262	5.15%	80.70%	2.859	5.08%	81.11%	3.683	6.47%	82.87%
2.3586	4.15%	84.85%	2.4669	4.34%	85.45%	2.479	4.36%	87.23%
2.0351	3.58%	88.43%	1.7834	3.14%	88.59%	1.397	2.45%	89.68%
1.6848	2.96%	91.40%	1.4682	2.58%	91.17%	1.111	1.95%	91.63%
1.4425	2.54%	93.93%	1.1327	1.99%	93.17%	0.987	1.73%	93.37%
0.9567	1.68%	95.62%	0.9778	1.72%	94.89%	0.947	1.67%	95.03%
0.6508	1.14%	96.76%	0.8951	1.57%	96.46%	0.85	1.49%	96.53%
0.5064	0.89%	97.65%	0.6178	1.09%	97.55%	0.568	1.00%	97.53%
0.4027	0.71%	98.36%	0.4213	0.74%	98.29%	0.387	0.68%	98.21%
0.3434	0.60%	98.97%	0.3548	0.62%	98.91%	0.361	0.64%	98.84%
0.2065	0.36%	99.33%	0.24	0.42%	99.33%	0.247	0.43%	99.28%
0.1524	0.27%	99.60%	0.157	0.28%	99.61%	0.175	0.31%	99.58%
0.1249	0.22%	99.82%	0.1427	0.25%	99.86%	0.128	0.22%	99.81%
0.1045	0.18%	100.00%	0.0791	0.14%	100.00%	0.109	0.19%	100.00%

In the case of Tb descriptors, the first eigenvalue of 38.872 could describe data variance of around 68.38%, which is the first PC. Then, the second PC with the second largest eigenvalue of 3.829 could explain a data variance of 7.71% and a total second variance of 76.09%. After that, the third PC with the third largest eigenvalue of 2.859 can describe a data variance of 5.03% and a total second variance of 81.11%. Next, the fourth PC with the fourth largest eigenvalue of 2.4669 could elucidate a data variance of 4.34% and a total second variance of 85.45%. The fifth PC with the fifth largest eigenvalue of 1.7834 could describe a data variance of 3.14% and a total second variance of 88.59%. Afterward, the sixth PC with the sixth largest eigenvalue of 1.4682 could explain a data variance of 2.58% and a total second variance of 91.17%. Finally, the last PC with the seventh largest eigenvalue of 1.1327 could clarify a data variance of 1.99% and a total second variance of 93.17%. Here, only seven components also possessed eigenvalues larger than 1. Meanwhile, for k = 7, the scree plot had a sloping shape on the right side as demonstrated in Figure 4 (B). The number of components that describe 57 descriptors was seven components, where the eigenvector (components) suitable with those of seven eigenvalues from the largest to the smallest was C_1, C_2, C_3, C_4, C_5, C_6, and C_7, respectively.

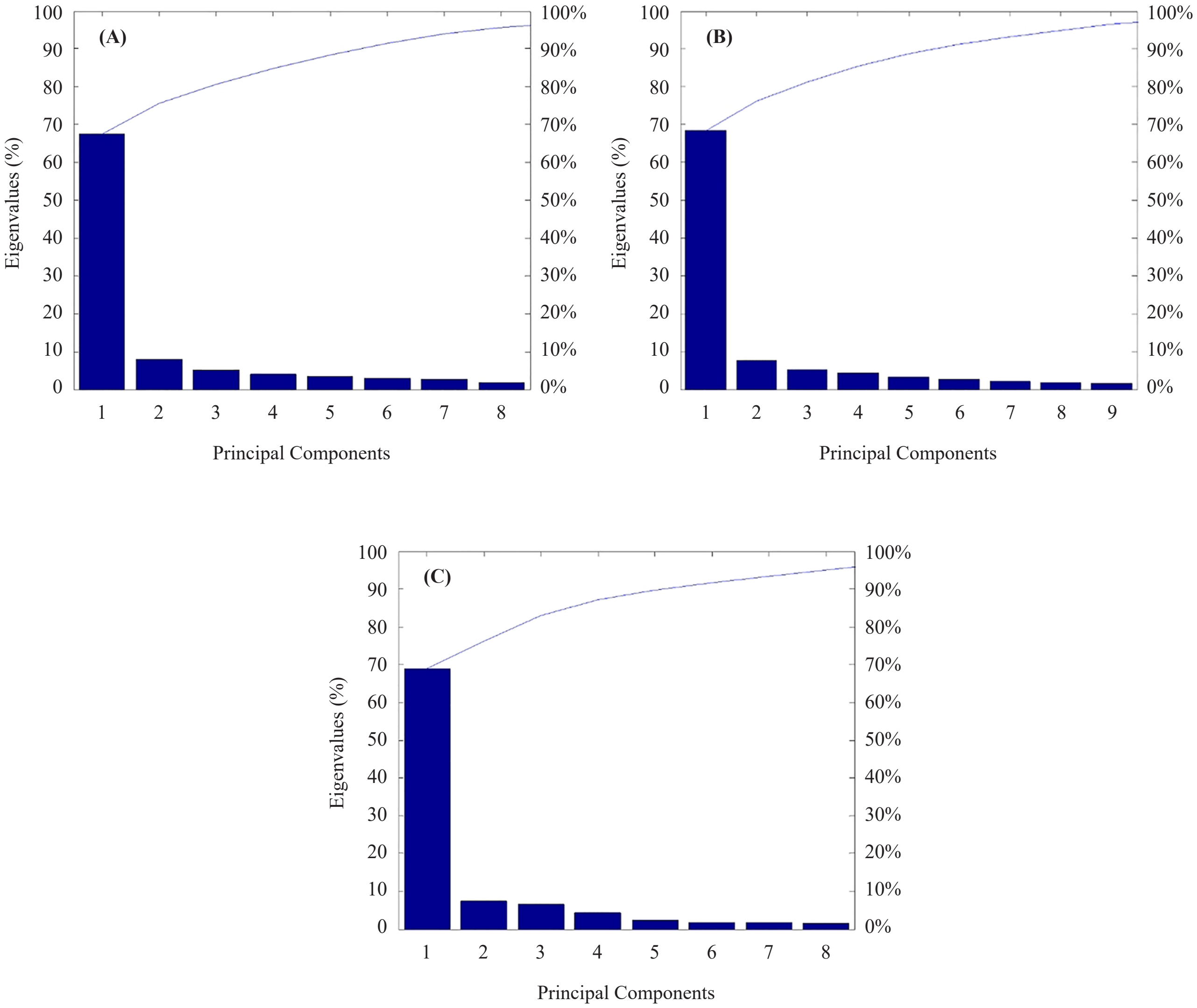


Figure 4. Scree plot of eigenvalues for (A) gadolinium, (B) terbium, and (C) dysprosium

Different from the other two, only six components have possession of eigenvalues larger than 1 in the Dy descriptors. The first eigenvalue (variance) of 39.154 could describe a data variance of about 68.82%, which is the first PC. The second PC with the second largest eigenvalue of 4.307 could explicate a data variance of 7.57% and a total second variance of 76.39%. The third PC with the third largest eigenvalue of 3.683 could explain a data variance of 6.47% and a total second variance of 82.87%. The fourth PC with the fourth largest eigenvalue of 2.479 could define a data variance of 4.36% and a total second variance of 87.23%. The fifth PC with the fifth largest eigenvalue of 1.397 could clarify a data variance of 2.45% and a total second variance of 89.68%. The sixth PC with the sixth largest eigenvalue of 1.111 could explain a data variance of 1.95% and a total second variance of 91.63%. Meantime, for k = 7, the scree plot possessed a sloping shape on the right side (depicted in Figure 4 (C)). The number of components that explain 57 descriptors was six components, where the eigenvector (components) suitable with those of six eigenvalues from the largest to the smallest was C_1, C_2, C_3, C_4, C_5, and C_6, respectively.

Table 2. Correlation of C_1 with 57 complex data of gadolinium, terbium, and dysprosium

Descriptor	Gd	Tb	Dy	Descriptor	Gd	Tb	Dy
	C_1	C_1	C_1		C_1	C_1	C_1
Z_1	-0.9783	-0.9823	-0.9805	Z_{30}	-0.9595	-0.9607	-0.9616
Z_2	-0.7649	0.2727	0.3215	Z_{31}	-0.9582	-0.9589	-0.9608
Z_3	-0.8665	-0.8776	-0.8703	Z_{32}	0.8911	0.9013	0.9011
Z_4	-0.8482	-0.8408	-0.8476	Z_{33}	0.8975	0.919	0.9202
Z_5	0.019	-0.1727	-0.1706	Z_{34}	0.9703	0.974	0.9774
Z_6	0.014	0.0036	0.0087	Z_{35}	-0.2478	-0.2961	-0.2790
Z_7	-0.8991	-0.9021	-0.9292	Z_{36}	-0.0827	-0.0515	-0.0700
Z_8	-0.9597	-0.9708	-0.9646	Z_{37}	-0.9594	-0.9567	-0.9574
Z_9	0.7273	0.6849	0.5539	Z_{38}	0.9976	0.9978	0.9977
Z_{10}	-0.0776	0.7274	0.8568	Z_{39}	0.9975	0.9978	0.9977
Z_{11}	-0.0001	0.0016	0.0070	Z_{40}	0.9973	0.9966	0.9970
Z_{12}	0.119	0.1247	0.1190	Z_{41}	0.9896	0.9943	0.9907
Z_{13}	0.9971	0.9971	0.9973	Z_{42}	0.9909	0.9878	0.9917
Z_{14}	0.9971	0.9971	0.9973	Z_{43}	0.9962	0.9974	0.9955
Z_{15}	0.9971	0.9971	0.9973	Z_{44}	0.9941	0.9933	0.9945
Z_{16}	0.9611	0.9964	0.9966	Z_{45}	0.9962	0.9952	0.9954
Z_{17}	0.9969	0.9968	0.9970	Z_{46}	-0.3827	-0.4206	-0.4616
Z_{18}	0.997	0.9971	0.9972	Z_{47}	-0.8154	-0.8078	-0.8112
Z_{19}	-0.9092	-0.9022	-0.8937	Z_{48}	-0.7346	-0.7267	-0.7309
Z_{20}	-0.9971	-0.9971	-0.9972	Z_{49}	0.7296	0.7201	0.7253
Z_{21}	-0.9971	-0.9971	-0.9972	Z_{50}	0.176	0.0964	0.0666
Z_{22}	-0.9534	-0.951	-0.9471	Z_{51}	0.0152	0.2352	0.4371
Z_{23}	-0.9505	-0.9444	-0.9394	Z_{52}	-0.2543	0.0937	0.1898
Z_{24}	-0.9473	-0.947	-0.9484	Z_{53}	0.2542	0.1076	-0.1897
Z_{25}	0.9777	0.9785	0.9793	Z_{54}	0.3759	0.1129	0.0561
Z_{26}	0.6441	0.9843	0.9843	Z_{55}	0.8458	0.8444	0.8442
Z_{27}	0.9369	0.9369	0.9348	Z_{56}	0.9097	0.9167	0.9122
Z_{28}	0.8328	0.8307	0.8300	Z_{57}	0.8774	0.8759	0.8729
Z_{29}	-0.964	-0.9669	-0.9602				

Table 2 presents the calculation results of the correlation between PCs and 57 descriptors. For Gd descriptors, a strong correlation with the first PCs was the descriptors $Z_1 \sim Z_4$, $Z_7 \sim Z_{26}$, Z_{22}, Z_{23}, $Z_{27} \sim Z_{34}$, $Z_{37} \sim Z_{45}$, $Z_{47} \sim Z_{49}$, $Z_{55} \sim Z_{57}$, so corresponding descriptors were as component C_1. Meanwhile, the strong correlation of Tb descriptors with the PCs was the descriptors Z_1, Z_3, Z_4, $Z_7 \sim Z_{10}$, $Z_{13} \sim Z_{34}$, $Z_{37} \sim Z_{49}$, $Z_{55} \sim Z_{57}$. These descriptors were as component C_1. In Dy descriptors, the strong correlation with the first PCs was the descriptors Z_1, Z_3, Z_4, $Z_7 \sim Z_{26}$, Z_{22}, Z_{23}, $Z_{27} \sim Z_{34}$, $Z_{37} \sim Z_{49}$, $Z \sim Z_{57}$. Similarly, these descriptors were also as component C_1. According to the descriptors of each complex compound of Gd, Tb, and Dy, the same descriptors affect the REE complex stability, namely thermodynamic energy and dipole moment. It is in good agreement with the reported work that there is a significant influence of thermodynamic and dipole moment toward the stability of Gd with the DBDTP derivative ligands.[20]

Regression analysis using seven PCs and gadolinium, terbium, and dysprosium response is presented in Equation 2-4, respectively. Meanwhile, a scree plot of response against prediction for gadolinium, terbium, and dysprosium is depicted in Figure 5 (A-C), respectively.

$$Y = 77.1599 + 6.8368C_1 - 4.4153C_2 + 5.4449C_3 - 2.0248\,C_4 + 2.2152C_5 - 4.7531C_6 + 1.3410C_7 \quad (2)$$

$$Y = 77.267 + 6.7696C_1 + 3.7722C_2 - 0.3020C_3 - 7.5792C_4 + 0.3486C_5 - 5.4903\,C_6 + 4.9115C_7 \quad (3)$$

$$Y = 77.2917 + 6.7446C_1 + 4.1182C_2 - 1.7449C_3 - 7.7165C_4 - 1.0080C_5 - 2.8312C_6 \quad (4)$$

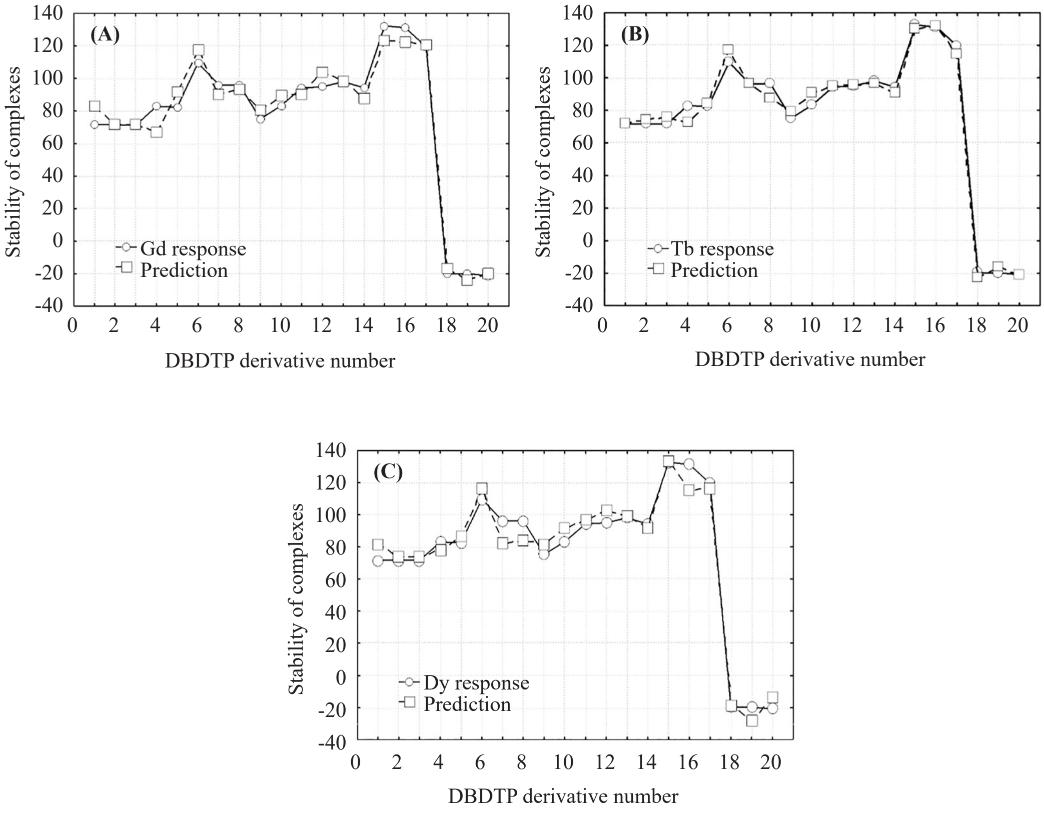


Figure 5. Scree plot of response vs prediction for (A) gadolinium, (B) terbium, and (C) dysprosium

According to the plot of three complex compounds, the model prediction graph fits with the observed graph. To select the best time series model was required to use several model selection standards. Several methods that could be used to gain the best regression model included the Akaike Information Criterion (AIC) and Schwarz Information Criterion (SIC) methods. The AIC is a mathematical method utilized to evaluate how well a model fits the data it was created from.[21] On the other hand, the SIC is among the most widely known statistical model selection to serve as an asymptotic approximation to a transformation of the Bayesian posterior probability of a candidate model.[22] In this study, the model goodness ratio was determined based on the mean square error (MSE), Akaike information criterion (AIC), and Schwarz information criterion (SIC) values, as presented in Table 3. Referring to AIC and SIC methods, the best regression model was the one that had the smallest AIC and SIC values.[23]

Table 3. The MSE, AIC, and SIC values for PCR of Dy, Gd, and Tb complex compounds

Compound	MSE	AIC	SIC
Gd	48,5623	108,0775	160,9576
Tb	19,8986	44,2852	65,953
Dy	53,1805	107,0925	151,7446

6. Conclusion

Three DBDTP complex compounds had the same variables unaffecting the stability, which included bond length variable of P number one with O number one, bond length variable of P number one with O number two, bond length variable of P number two with O number three, bond length variable of P number two with O number four, bond angle variable of P number two with S number for and O number three, bond angle variable of P number two with S number four and O number two, gradient norm, dipole, potential ionization, homo, and lumo. The response prediction of Gd, Tb, and Dy resulted from response regression with PCs was selected through the plot of three complex compounds according to the observation graph. Based on the MSE, AIC, and SIC calculation, the Tb complex compounds possessed the smallest value among others, and therefore this complex was the best PCR model.

Acknowledgment

This work was supported by the Higher Education Services Institutes and Institute of Research and Community Service funded by Indraprasta PGRI University (No. 032/KM/PNT/2018 and No. 0293/SKP.LT/LPPM/UNINDRA/III/2018).

Author contributions

Nurdeni: Methodology, Writing of Manuscript Draft. Heru Agung Saputra: Writing of Manuscript Draft, Final Manuscript Writing. Atje Setiawan Abdullah: Data Collection and Analysis. Budi Nurani Ruchjana: Data Collection and Analysis. Husein Hernandi Bahti: Final Manuscript Writing. Hedi: Writing-review & editing.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Grupta, C. K.; Krishnamurthy, N. *Extractive Metallurgy of Rare Earths*; CRC Press: Boca Raton, 2005.
[2] Balaram, V. *Minerals* **2023,** *13*, 425.
[3] Opare, E. O.; Struhs, E.; Mirkouei, A. *Renew. Sust. Energ. Rev.* **2021,** *143*, 110917.
[4] Pratomo, U.; Anggraeni, A.; Muthalib, A.; Soedjanaatmadja, U. M. S.; Pinarti, I.; Hidayat, A. T.; Bahti, H. H. *Procedia Chem.* **2015,** *17*, 207-215.
[5] Bahti, H. H. *Chromatographic Studies of Metal Dialkyldithiophosphate*; The University of New South Wales: Sydney, 1990.
[6] Sofyatin, T.; Nurlina, N.; Anggraeni, A.; Bahti, H. H. *J. Chem. Nat. Acta. (Engl. Transl.)* **2016,** *4*, 47-51.
[7] Ozturk, T.; Ertas, E.; Mert, O. *Chem. Rev.* **2010,** *110*, 3419-3478.
[8] Liu, F.; Cao, C.; Cheng, B. *Int. J. Mol. Sci.* **2011,** *12*, 2448-2462.
[9] Roy, K.; Kar, S.; Das, R. N. *A Primer on QSAR/QSPR Modeling Fundamental Concepts*; Springer Champ: Kolkata, 2015.
[10] Toropov, A. A.; Toropova, A. P. *Molecules* **2020,** *25*, 1292.
[11] Yousefinejad, S.; Hemmateenejad, B. *Chemom. Intell. Lab. Syst.* **2015,** *149*, 177-204.
[12] Saputra, H. A.; Anggraeni, A.; Mutalib, A.; Bahti, H. H. *J. Kim. Sains Apl.* **2021,** *5*, 177-184.
[13] Balaram, V. *Geosci. Front.* **2019,** *10*, 1285-1303.
[14] Balaram, V. *J. Geol. Soc. India* **2022,** *98*, 1210-1216.
[15] Dewi, R.; Sastyarina, Y.; Janitra, R.; Hardianto, A.; Yusuf, M.; Mutalib, A.; Anggraeni, A. *Asian J. Chem.* **2022,** *34*, 3105-3107.
[16] Budiman, S.; Hardian, A.; Virdasari, L.; Nurdeni, N.; Herman, H.; Mutalib, A.; Anggraeni, A.; Bahti, H. H. *J. Trop. Pharm. Chem.* **2022,** *6*, 64-70.
[17] Jolliffe, I. T.; Cadima, J. *Philos. Trans. Royal Soc. A* **2016,** *374*, 20150202.
[18] Greenacre, M.; Groenen, P. J., Hastie, T.; d'Enza, A. I.; Markos, A.; Tuzhilina, E. *Nat. Rev. Methods Primers*. **2022,** *2*, 100.
[19] Cliff, N. *Psychol. Bull.* **1988,** *103*, 276-279.
[20] Nurdeni. The Influence of Thermodynamic Energy and Dipole Moment on the Stability of Gadolinium Complexes with DBDTP Ligands Using Multiple Regression (Engl. Transl.). In *Proceedings of the National Mathematics Seminar, Depok (Engl. Transl.)*; Nurdeni, 2017.
[21] Kim, Y. M.; Han, G.; Kim, H.; Oh, T. M.; Kim, J. S.; Kwon, T. H. *Appl. Sci.* **2020,** *10*, 8727.
[22] Cavanaugh, J. E.; Neath, A. A. *Commun. Stat.-Theory Methods* **1999,** *28*, 49-66.
[23] Agus, W. *Econometrics Theory and Applications (Engl. Transl.)*; Ekonisia FE UII: Yogyakarta, 2007.