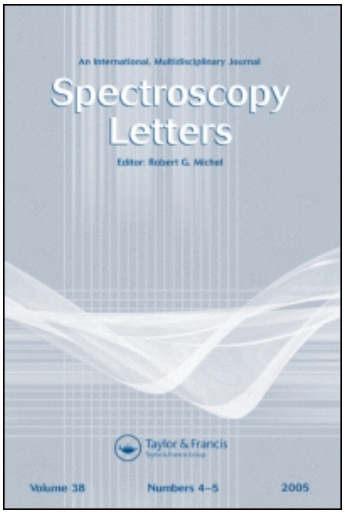

Spectroscopy Letters
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597299

Quantitative Structure-Properties Relationship Study of the ^{29}Si-NMR Chemical Shifts of Some Silicate Species

Nasser Goudarzi[a]; M. H. Fatemi[b]; A. Samadi-Maybodi[b]
[a] Faculty of Chemistry, Shahrood University of Technology, Shahrood, Iran [b] Department of Chemistry, Faculty of Basic Sciences, Mazandaran University, Babolsar, Iran

To cite this Article Goudarzi, Nasser , Fatemi, M. H. and Samadi-Maybodi, A.(2009) 'Quantitative Structure-Properties Relationship Study of the ^{29}Si-NMR Chemical Shifts of Some Silicate Species', Spectroscopy Letters, 42: 4, 186 — 193
To link to this Article: DOI: 10.1080/00387010902809948
URL: http://dx.doi.org/10.1080/00387010902809948

Spectroscopy Letters, 42:186–193, 2009

ISSN: 0038-7010 print/1532-2289 online
DOI: 10.1080/00387010902809948

Taylor & Francis
Taylor & Francis Group

Quantitative Structure–Properties Relationship Study of the ^{29}Si-NMR Chemical Shifts of Some Silicate Species

Nasser Goudarzi[1], M. H. Fatemi[2], and A. Samadi-Maybodi[2]

[1]Faculty of Chemistry, Shahrood University of Technology, Shahrood, Iran
[2]Department of Chemistry, Faculty of Basic Sciences, Mazandaran University, Babolsar, Iran

ABSTRACT In this work, the quantitative structure–properties relationship (QSPR) was applied to modeling and predicting the ^{29}Si-NMR chemical shifts of a series of silicate species (on Q^2 sites). The descriptors that were selected by stepwise multiple linear regression technique were square of alpha polarizability, Moran autocorrelation–lag3/unweighted by atomic Sanderson electronegativities, square of asphericity, and topological path/walk 2-Randic shape index. These descriptors could encode electronic, geometric, and topological characteristics that affect the chemical shifts of the molecules of interest. The results obtained using the multiple linear regression (MLR) model were comparable with the experimental values. The cross-validation test was also performed to evaluate the prediction power of the MLR model obtained. The q^2 and PRESS of this model are 0.976 and 0.44761, respectively, revealing the credibility of the model.

KEYWORDS cross-validation, molecular descriptors, quantitative structure–properties relationship, silicate species, silicon-29 NMR spectroscopy

INTRODUCTION

Aqueous solutions of silicates rapidly equilibrate to a mixture of anionic species, including chains, branched systems, and cyclic units. Evidence for the existence of such polysilicate ions has come from Raman spectroscopy,[1] paper chromatography,[2] trimethylsilylation,[3] and reaction with molybdic acid.[4] Alkali silicates dissolve in water at high pH to form solutions in which a number of anionic species are presented in dynamic equilibrium. Determination of the chemical structures of these species is difficult, but ^{29}Si-NMR technique has achieved some prominence, following the introduction of pulsed methods involving Fourier transformation. A number of papers have already been published on these topics.[5–10] However, because ^{29}Si atoms form a dilute spin system with a natural abundance of 4.7% and the hydroxyl protons exchange rapidly on the NMR timescale, each type of silicon environment normally gives a single resonance, so that assignments are problematic. To a considerable extent, this difficulty is overcome by the use of silica enriched with ^{29}Si.[11]

Received 8 May 2007; accepted 28 October 2008.

Address correspondence to Nasser Goudarzi, Faculty of Chemistry, Shahrood University of Technology, P. O. Box 316, Shahrood, Iran. E-mail: goudarzi@shahroodut.ac.ir

Organic base silicate solutions have long been known to exhibit unique chemical and physical characteristics, including crystallization to distinctive clathrate hydrate structures.[12] As early as 40 years ago, organic base cations, specifically tetraalkyl ammonium (TAA) cations, were used as structure-directing agents in the syntheses of zeolite molecular sieves.[13] Nowadays, tetraalkyl ammonium cations are employed in the production of 4 most commercially important zeolites, and therefore, their interaction with aqueous silicates has a direct and significant impact on a variety of industries ranging from petroleum refining to pollution control.

As mentioned earlier, alkaline aqueous solutions of silicates contain a variety of anionic species in dynamic equilibrium. ^{29}Si-NMR spectroscopy has proved to be the most valuable tool for obtaining information about the molecular structures, but isotopic enrichment in ^{29}Si is generally necessary in order to make diagnostic use of splitting patterns caused by (^{29}Si–^{29}Si) scalar coupling.[14,15] To date, 28 anionic silicate structures in solution have been proposed.[14,16] As the Si concentration is increased and the pH of the solution is lowered, many highly condensed species are formed, so that resonance bonds of different silicate species overlap, and the information about individual anions is lost. As a consequence, the oligomers containing Q^1, Q^2, Q^3, and Q^4 silicon sites (where the superscripts indicate the number of silioxane bridges) are well known.[17]

Chemical shifts in NMR spectra are an important parameter for investigating molecular structures. NMR chemical shift prediction algorithms and software have been around for many years, and a number of commercial packages are available. Generally, ^{13}C has been the preferred nucleus for the development of these algorithms; some software is also available that supports the ^{1}H, ^{19}F, ^{29}Si, and ^{31}P chemical shift predictions, all of which are sensitive nuclei.[18] There are several theoretical methods for prediction of chemical shifts of nuclei in different molecules that establish a relationship between the chemical shifts and structure of the molecule. One of the most successful approaches to the prediction of chemical properties starting only from molecular structural information is quantitative structure–activity/structure–property relationship (QSAR/QSPR). In these studies, a correlation is defined between chemical structure and a chosen property. Such studies consist of two main steps. In the first step, the chemical compounds are translated into a computer-readable form, and afterwards the quantitative correlation between chemical structure and its property could be obtained using different statistical and learning procedures, like multiple linear regression (MLR), artificial neural network (ANN), partial least square (PLS), and so forth.[19] There are several reports on prediction of chemical shifts by QSPR approaches.[20–22] Khadikar and Jaiswal studied the effects of ^{13}C-NMR chemical shifts on carbinol carbon atoms by QSPR.[23] Also, Duvenbeck has used topological and geometric descriptors to develop a QSPR model for the prediction of ^{13}C-NMR chemical shifts of some substituted naphthalenes.[24] Alam and Henry developed a partial charge model to investigate hydrolysis in organically modified alkoxy silanes and calculate the ^{29}Si-NMR chemical shielding tensors.[25] In the current work, a quantitative structure–properties relationship was applied to modeling and prediction of the ^{29}Si chemical shifts of some silicate species. We used a linear method (multiple linear regression; MLR) for prediction of ^{29}Si-NMR chemical shifts of silicate species.

The general purpose of multiple regressions is to quantify the relationship between several independent or predictor variables and a dependent variable. A set of coefficients defines the single linear combination of independent variables (molecular descriptors) that best describes molecule chemical shift. The chemical shift value for each molecule would then be calculated as a composite of each molecular descriptor weighted by the respective coefficients. A multilinear model can be represented as:

$$[y = b_1x_1 + b_2x_2 + b_3x_3 \cdots b_mx_m + \varepsilon] \qquad (1)$$

where m is the number of independent variables, $b_1, \ldots, b_m$ are the regression coefficients, and y is the dependent variable. Regression coefficients represent the independent contributions of each calculated molecular descriptor. The algebraic MLR model is defined in Eq. (1) and in matrix notation:

$$[y = Xb + e] \qquad (2)$$

When X is of full rank, the least-squares solution is

$$b = (X^TX)^{-1}X^Ty$$

where b is the estimator for the regression coefficients in b. The MLR model was built using a training set and validated using an external prediction set. Multiple linear regression techniques based on least-squares procedures are very often used for estimating the coefficients involved in the model equation.

MATERIALS AND METHODS

Data Set

The experimental chemical shifts for 18 silicate species in aqueous solutions were extracted from References 14–16. The names of these compounds and their ^{29}Si chemical shifts are tabulated in Table 1. Figure 1 presents the structures of all the silicate species present in solution. The ^{29}Si chemical shifts in the Q^2 sites were selected for modeling. A Q^2 site could be located in different positions in the structure of the molecules. Because each Q^2 site has a different chemical environment, these sites have different chemical shift values (Table 1).

TABLE 1 Data Set and Corresponding Observed and Predicted Chemical Shifts

Name	Experimental chemical shift of ^{29}Si in Q^2 site	Calculated chemical shift of ^{29}Si in Q^2 site
Linear trimer	−15.51	−15.32
Linear tetramer	−17.27	−17.13
Cyclic trimer	−11.98	−12.03
Monosubstituted cyclic trimer	−11.58	−11.71
Cyclic tetramer	−13.57	−13.55
Monosubstituted cyclic tetramer	−14.91	−14.64
Bridged cyclic tetramer	−15.48	−15.51
Doubly bridged cyclic tetramer	−15.01	−15.12
Bicyclic pentamer	−9.30	−9.04
Tricyclic hexamer a	−14.88	−14.98
Tricyclic hexamer b (transoid)	−16.11	−16.38
Tricyclic hexamer c (sisoid)	−8.30	−5.91
Pentacyclic heptamer	−16.27	−16.80
Tricyclic octamer	−11.12	−10.92
Tricyclic hexamer	−10.60	−10.63
Tetracyclic nonamer	−16.92	−16.09
Hexacyclic nonamer	−17.13	−17.26
Bicyclic octamer	−15.10	−14.91

The ^{29}Si chemical shift values for these species were obtained in the same instrumental conditions at 22°C. The ^{29}Si chemical shifts for the Q^2 sites vary in different silicate species and lie between −73 and −88ppm (with respect to TMS, as reference) or between −7.3 and −18 ppm if Q^0 (monomer) is used as the reference peak.

Descriptor Generations and Screenings

A chemical shift value in NMR is related to the electronic environment of the nucleus of interest in a molecule, as well as to some topological and geometric configurations of the molecule. The values for these molecular features could be encoded quantitatively by numerical values called molecular descriptors. These molecular parameters used to look for the best QSPR model for prediction of a chemical shift value are actually the geometric, electronic, and topological descriptors. The geometric descriptors were calculated using optimized Cartesian coordinates and the van der Waals radius of each atom in the molecule by using the Dragon package.[26,27] The electronic descriptors were calculated using the version 6 of the MOPAC package,[28] and the topological descriptors were calculated using the Dragon package.[29] Because some of the descriptors generated for each compound encoded similar information about the molecule of interest, it was desirable to test each descriptor and eliminate those showing a high correlation ($R = 0.90$). Subsequently, the method of stepwise MLR was used to select the important descriptors and to model construction. The descriptors appearing in the best MLR equation are shown in Table 2. These descriptors are square of alpha polarizability (SAP), Moran autocorrelation–lag3/weighted by atomic Sanderson electronegativities (MATS3e), square of asphericity (SGA), and topological path/walk 2-Randic shape index (PW2).

The first descriptor that appeared in the model is alpha polarizability, which is a kind of electronic descriptor. When a molecule is embedded in a uniform electric field, $\mathbf{E}_o$, in a vacuum, an induced dipole moment, μ_{IND}, arises, defined by the following equation:

$$[\mu_{IND} = \alpha \cdot \mathbf{E}_o] \qquad (3)$$

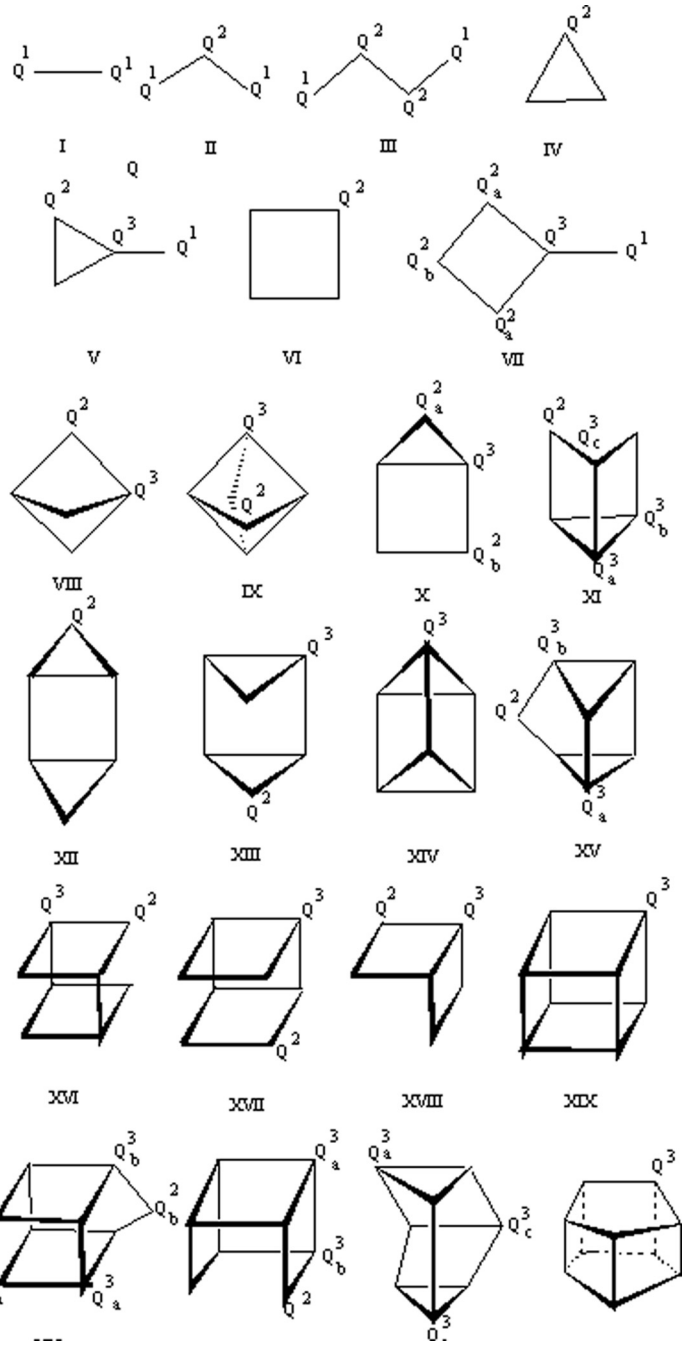

FIGURE 1 **Structures of silicate species. I, dimer; II, linear trimer; III, linear tetramer; IV, cyclic trimer; V, monosubstituted cyclic trimer; VI, cyclic tetramer; VII, monosubstituted cyclic tetramer; VIII, bridged cyclic tetramer; IX, doubly bridged cyclic tetramer; X, bicyclic pentamer; XI, tricyclic hexamer a; XII, tricyclic hexamer b (transoid); XIII, tricyclic hexamer c (sisoid); XIV, prismatic hexamer; XV, pentacyclic heptamer; XVI, bicyclic octamer; XVII, tricyclic octamer; XVIII, bicyclic hexamer; XIX, cubic octamer; XX, tetracyclic nonamer; XXI, hexacyclic nonamer; XXII, hexacyclic octamer; XXIII, prismatic decamer.**

where the scalar constant of proportionality is called the polarizability. This scalar polarizability may be regarded as the sum of the electronic polarizability, α_E, and the atomic polarizability, α_A.

In general, the scalar polarizability, α, is not sufficient to describe the induced polarization, and therefore, a polarizability tensor is used to better encode induced polarization and represent molecular polarizability, which is a tensor not a scalar quantity.

The second descriptor that appeared in the QSPR model was Moran autocorrelation–lag3/weighted (lag by atomic Sanderson electronegativities) (MATS3e). The molecular descriptors based on the autocorrelation function AC_l were defined as:

TABLE 2 Specification of Multiple Linear Regression Model

Descriptor	Notation	Coefficient	Mean effect
Square of alpha polarizability	SAP	−0.037 (±0.001)	−36.933
Moran autocorrelation–lag3/weighted by atomic Sanderson electro-negativities	MATS3e	3.414 (±1.44)	2.371
Square geometric asphericity	SGA	0.682 (±0.358)	1.906
Topological path/walk 2-Randic shape index	PW2	−0.037 (±0.019)	−1.993
Constant	—	−2.928 (±1.469)	—

$R = 0.999$; $R^2 = 0.998$; Standard error of the estimate = 0.1520785; $F = 1417.430$.

$$\left[\mathrm{AC}_l = \int_a^b f(x) \cdot f(x+l) \cdot dx\right] \tag{4}$$

where $f(x)$ is any function of the x, and l is the lag representing an interval of x; a and b define the total interval of the function studied. The function $f(x)$ is usually a time-dependent function such as a time-dependent electrical signal or a spatial-dependent function such as the population density in space. General index of spatial autocorrelation, if applied to a molecular graph, could be defined as

$$\mathbf{I}(d) = \frac{1/\Delta \cdot \sum_{i=1}^{A}\sum_{j=1}^{A} \delta_{ij} \cdot (w_i - w_j) \cdot (w_j - \hat{w})}{1/A \cdot \sum_{i=1}^{A} (w_i - \hat{w})^2} \tag{5}$$

where $\mathbf{I}(d)$ denotes Moran coefficient, w_i is any atomic property, $\hat{w}$ is its average value on the molecule, A is the atomic number, d is the considered topological distance (i.e., the lag in autocorrelation terms), δ_{ij} is a Kronecker delta ($\delta_{ij} = 1$ if d_{ij}=d, zero otherwise), and Δ is the sum of the Kronecker deltas (i.e., the number of vertex pairs at a distance equal to d).

The Moran coefficient usually takes the interval values −1 and +1. Positive autocorrelation corresponds with positive values of the coefficient, whereas negative autocorrelation produces negative values.

The third descriptor present in the MLR model is asphericity. An anisotropy descriptor that measures the deviation from the spherical shape, and calculated from the eigenvalue λ_i of the inertia matrix, is defined as

$$\Omega_{\mathrm{A}} = 1/2\left[\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_1 - \lambda_3)^2 + (\lambda_2 - \lambda_3)^2}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}\right] \tag{6}$$

and

$$0 \le \Omega_{\mathrm{A}} \le 1$$

where $\Omega_{\mathrm{A}} = 0$ and $\Omega_{\mathrm{A}} = 1$ correspond with the spherical and linear molecules, respectively. For prolate molecules, $\lambda_1 \approx \lambda_2 > \lambda_3$ and $\Omega_{\mathrm{A}} \approx 0.25$, whereas for oblate molecules (disk-shaped), $\lambda_1 > \lambda_2 \approx \lambda_3$ and $\Omega_{\mathrm{A}} \approx 1$.

The fourth descriptor appearing in the model is path/walk 2-Randic shape index (PW2), which is a topological descriptor. Path/walk shape indices are similar to the invariants derived from the distance/distance matrix. Atomic path/walk indices are defined for each atom as the ratio between atomic path count ${}^{m}P_i$ and atomic walk count $awc_i^{(m)}$ of the same length, m, i.e.,

$$(p/w)_i^m = {}^{m}P_i/\mathrm{awc}_i^{(m)} \tag{7}$$

whereas the number of paths in a molecule is bounded. However, being interested only in quotients, the walk count is terminated when it exceeds the maximum allowed length of the corresponding path.

On the other hand, the adjacency matrix A of a simple graph of n vertices, corresponding with a saturated hydrocarbon of n carbon atoms, is a square matrix ($n * n$) with entries $a_{ij} = 1$ if atom i is connected to atom j, and $a_{ij} = 0$ otherwise. The kth power of this matrix, A^k, has $a_{ij}^{(k)}$ entries, which are integers giving the number of walks of length k bonds from atom i to atom j. A walk is any sequence of adjacent graph edges; in walking from i to j, it is allowed to go back and forth or to visit vertices repeatedly. In contrast, a path is a sequence of adjacent graph edge without repetition.

TABLE 3 Values of the Descriptors Used in this Work

Molecule no.	SAP	MATS3e	SGA	PW2
1	240.5601	−1.096	.1482	1.896
2	298.2529	−1.086	.6084	4.265
3	143.5204	−1.076	.0610	4.265
4	134.0964	−1.076	.0538	4.265
5	184.1449	−1.085	.1043	5.086
6	222.3081	−1.090	.4476	7.825
7	239.6304	−1.082	.3014	6.158
8	225.3001	−1.096	.2948	7.986
9	86.4900	−0.9430	.4147	3.028
10	221.4144	−1.097	.2034	6.509
11	259.5321	−1.090	.2256	7.825
12	68.890	−0.988	.3881	1.954
13	264.7129	−1.009	.4045	11.123
14	123.6544	−1.027	.2333	2.924
15	112.3600	−1.035	.2266	4.521
16	286.2864	−1.059	.4733	4.521
17	293.4369	−1.059	.4032	6.161
18	228.0100	−1.086	.3832	4.605

Molecular path/walk indices are defined as the average sum of the atomic path/walk indices of equal length

$$[(p/w)^m = 1/A \sum_{i=1}^{A} (p/w)_i^m] \tag{8}$$

where $(p/w)_i^m$ is the molecular path/walk shape index, and A is the number of atoms in the molecule.

Alternatively, molecular path/walk indices are obtained by separately summing all the paths and walks of the same length and then calculating the ratio between their counts. All of these descriptors were calculated using the Dragon software.[29] Further meanings of these molecular descriptors and their calculation procedure are summarized in the Dragon software and are explained in detail, with related references, in the *Handbook of Molecular Descriptors* by Todeschini and Consonni.[30] The numerical values for all the descriptors are shown in Table 3.

RESULTS AND DISCUSSION

In the QSPR model obtained, four descriptors correlate linearly with a ^{29}Si chemical shift in the molecules of interest. The descriptors numerically encode the electronic, geometric, and structural features of the selected molecules. Table 1 shows the data set and corresponding observed and MLR predicted values of the chemical shifts for the selected Q^2 sites for the molecules studied in this work. The selected MLR model is presented in Table 2. The numerical values for the descriptors are shown in Table 3. As it can be observed in this table, four descriptors appeared in the MLR model: namely, (1) square of polarizability, (2) Moran autocorrelation–lag3/weighted by atomic Sanderson electronegativities, (3) square of asphericity, and (4) topological path/walk 2-Randic shape index.

These variables encode different topological, geometric, and electronic aspects of a molecular structure. Because the chemical shift of a nucleus depends on the chemical (electronic) environment of the nucleus, these parameters could encode the different aspects of electronic, topological, and geometric features of molecules. The first and the most important parameter that appeared in the model is square of polarizability. This parameter is influenced by the chemical shifts of the sites of interest (Q^2 site in silicate species) for the molecules in various positions. This parameter has the largest (absolute) mean effect (−36.933). The negative sign for the mean value of polarizability reveals that when polarizability (or to be more exact, square of polarizability) is increased (nucleus is shielded), the chemical shift of a Q^2 site in a silicate species is shifted to a lower frequency (higher chemical shift). It is known that when electron density around a nucleus is increased, its polarizability increases, which causes the nucleus to be shielded against an external applied magnetic field. As an example, in linear trimer (II) silicate species, δ_{Si} is −15.51 ppm with a polarizability value of 240.5601, but in bicyclic pentamer (X), δ_{Si} is −9.30 ppm with a polarizability value of 86.4900. These values indicate that when polarizability in a molecule (on the Q^2 site) is increased, electron density around the ^{29}Si atom is increased significantly and its chemical shift is displaced to a lower frequency (higher chemical shift value). As another example, it could be seen that the calculated polarizability value for tricyclic hexamer (XIII) is 112.3600, but for tetracyclic nonnumeric (XX) it is 286.2864 (Table 3). These values would justify the difference between the chemical shifts of XIII and XX, −10.60 and −16.92 ppm, respectively. It could therefore be concluded that this

descriptor has an important role in the chemical shift value and has the highest weight in the MLR model in this work (see Table 2).

The second descriptor in this model that affects a chemical shift value is MATS3e. This descriptor depends on some atomic properties such as electronegativity. This descriptor is also related to the atomic distances, bond angles, and implied strain in the molecule. Because silicate species have different shapes and molecular structures (see Fig. 1), the Q^2 sites of the silicate species have different electron densities, causing differentiation in shielding of the Q^2 sites. This effect is encoded in the model by the MATS3e descriptor. The silicate groups in linear silicate species consist of tetrahedral structures. However, in branched and cyclic silicate species, the bond angles vary significantly from the optimal value, i.e., 109.5°, which depends on the molecular structure of the silicate and could be deviated up to 10°. This structural deviation applies to a torsional tension in the molecule, which is known as steric hindrance. Thus when the O–Si–O bond angle in a silicate species is varied, the electron density around Q^2 site would be changed. Therefore, by varying the electron density around Q^2 site, the chemical shift of the nucleus is shifted. As an example, in linear trimer silicate species, δ_{Si} is -15.51 but in cyclic trimer, δ_{Si} is -11.98 ppm. These values indicate that when a molecule has a cyclic structure (meaning increase in steric effect), electron density around ^{29}Si atom (Q^2 sites here) is decreased and its chemical shift is displaced to a higher frequency (low negative value). The same effect was likewise shown clearly in linear tetramer ($\delta_{Si} = -17.27$ ppm) and cyclic tetramer ($\delta_{Si} = -13.57$ ppm).

One of the effective parameters affecting the chemical shift value is direction of the molecule in the applied magnetic field. The sensitive (or observed) magnetic field for each nucleus (Q^2 sites in this respect) could also be varied by the symmetry or spherical shape of the molecule. Thus asphericity, which shows the deviation from the spherical shape, could be presented in this model for prediction of the chemical shift value. Because the silicate species have different geometric shapes (linear trimer, cyclic trimer, tetramer, cyclic tetramer, and other molecules), the asphericity term is expected to be different for the corresponding species. The Ω_A value depends on the shape of a molecule. The silicate species in solution that could be encoded by SGA descriptor can appear as prolate, oblate, or spherical. The inclusion of SGA in the MLR model reveals that the chemical shifts of the Q^2 sites are related to the shapes of the silicate species. For instance, we considered the molecules XII [tricyclic hexamer (transoid)] and XIII [tricyclic hexamer (sisoid)], the former having nearly a prolate shape ($\Omega_A = 0.25$), whereas the shape of the latter is nearly oblate ($\Omega_A \approx 1$). The experimental ^{29}Si-NMR data show that the chemical shifts of the Q^2 sites for the above molecules (XII and XIII) are considerably different ($\delta_{XII} = -16.11$ ppm and $\delta_{XIII} = -8.3$ ppm).

The last descriptor in the MLR model obtained that affects the chemical shift is path/walk 2-Randic shape index (PW2), which is a topological descriptor. Typically, this descriptor describes the degree of branching of the molecule. By increasing the PW2 parameter, the electronic movement through the molecule is increased, and this causes an increase in the electron density of the Si atom, and a Q^2 site in the molecule is shielded (moved to a higher chemical shift). As an example, linear trimer and linear tetramer have PW2 values of 1.896 and 4.265 and chemical shift values of -15.51 and -17.27 ppm, respectively. In conclusion, different structural descriptors appearing in the MLR model could encode different electronic, topological, and geometric characteristics of the molecular site affecting the chemical shift.

It is worth mentioning that contrary to the traditional regression model, cross-validation evaluates the validity of the model by how well it predicts the data rather than by how well it fits the data. The analysis uses a "leave-one-out" scheme; a model was built with n−1 compounds and the *n*th compound was predicted. Each compound was left out of the model deviation and predicted in turn. Then the values for q^2 (cross-validated correlation coefficient) and $PRESS_{val}$ (prediction error sum of squares) of the model obtained were calculated from the following equations:

$$PRESS_{val} = \sum_{i=1}^{n} (yi - \hat{y}_i) \tag{9}$$

$$\left[q^2 = 1 - \left[PRESS_{val} \Big/ \sum_{i=1}^{n} ((yi - \hat{y}_i)^2) \right] \right] \tag{10}$$

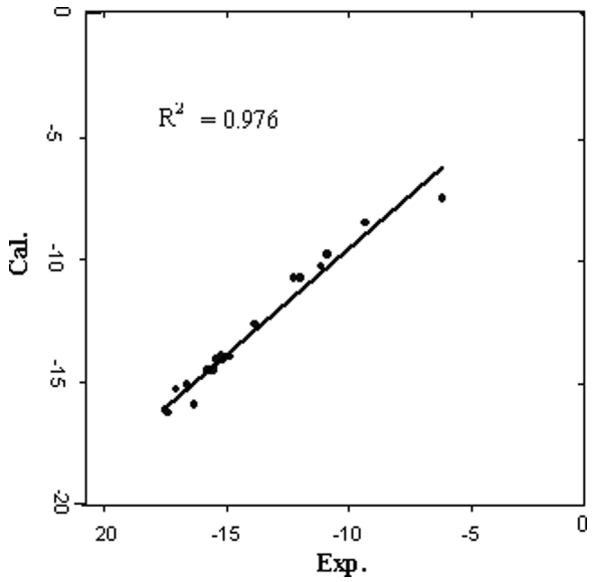

FIGURE 2 **Plot of the calculated chemical shift against the experimental values.**

where n is the number of compounds in the training set, y_i is the experimental chemical shift for the ith compound, and $\hat{y}_i$ is the value predicted by the model built without compound i. $\text{PRESS}_{\text{val}}$ is the PRESS calculated when all compounds from the training set are included in the model. The calculated values for q^2 and PRESS were 0.976 and 0.44671, respectively, which reveal the credibility of the model obtained. Figure 2 shows the plot of the predicted MLR against the experimental chemical shift values for the molecules included in the data set.

CONCLUSION

In the current study, MLR was used as a feature mapping technique for prediction of the chemical shifts of some silicate species. The results obtained indicated that MLR could establish a good relationship between a chemical shift value and molecular properties. Thus, descriptors appearing in this QSAR model provide information related to different molecular properties such as electronic, geometric, and topological properties, which could participate in the physiochemical processes that affect the chemical shifts of the silicate species in NMR.

REFERENCES

1. Freund, E. *Bull. Soc. Chim. France* **1973**, 2244.
2. Wieker, W.; Hoebbel, D. Chemische untersuchungen von silicaten. XXXVI. Die papierchromatographische untersuchung von kondensierten silicaten und kieselsauren. *Z. Anorg. Chem.* **1969**, *366*, 139–151.
3. Lentz, C. W. Silicate Minerals as Sources of Trimethylsilyl Silicates and Silicate Structure Analysis of Sodium Silicate Solutions. *Inorg. Chem.* **1964**, *366*, 139.
4. Thilo, E.; Wieker, W.; Stade, H. Chemische untersuchungen von silicaten, XXXI. Über beziehungen zwischen dem polymerisationsgrad silicatischer anionen und ihrem reaktionsvermögen mit molybdänsäure. *Z. Anorg. Chem.* **1965**, *340*, 261.
5. Marsmann, H. C. *Chem. Ztg.* **1973**, *97*, 128.
6. Engelhardt, G.; Jancke, H.; Hoebbel, D.; Wieker, W. *Z. Chem.* **1974**, *14*, 109.
7. Marsmann, H. C. Derivatives of silicic acids as model compounds for silicates. *Z. Natureforsch. B.* **1974**, *29*, 495.
8. Gould, R. O.; Lowe, B. M.; MacGlip, N. A. *J. Chem. Soc. Chem. Comm.* **1974**, 720.
9. Engelhardt, G.; Zeigan, D.; Jancke, H.; Hoebble, D.; Wieker, W. ^{29}Si-NMR-Spektroskopie an silicatlösungen. V. über die ^{29}Si-NMR-Spektren niedermolekularer kieselsäuren. *Z. Anorg. Chem.* **1975**, *418*, 17.
10. Engelhardt, G. *Z. Chem.* **1975**, *15*, 495.
11. Engelhardt, G.; Michel, D. *High-resolution Solid-State NMR of Silicates and Zeolites*; Wiley: New York, 1987.
12. Iler, R. K. *The Chemistry of Silica*; Wiley: New York, 1979.
13. Barrer, R. M.; Denny, P. J. *J. Chem. Soc.* **1961**, 71.
14. Kinrade, S. D.; Knight, C. T. G.; Pole, D. L.; Syvitski, R. T. Silicon-29 NMR studies of tetraalkylammonium silicate solutions. 2. Polymerization kinetics. *Inorg. Chem.* **1998**, *37*, 272.
15. Harris, R. K.; Knight, C. T. G. Silicon-29 nuclear magnetic resonance studies of aqueous silicate solutions. Part 6. Second-order patterns in potassium silicate solutions enriched with silicon-29. *J. Chem. Soc. Faraday Trans.* **1983**, *79*, 1539.
16. Knight, C. T. G. A two-dimensional silicon-29 nuclear magnetic resonance spectroscopic study of the structure of the silicate anions present in an aqueous potassium silicate solution. *J. Chem. Soc. Dalton Trans.* **1988**, 1457.
17. Harris, R. K.; O'Connor, M. J.; Curzon, E. H.; Howarth, O. W. Two-dimensional silicon-29 NMR studies of aqueous silicate solutions. *J. Magn. Reson.* **1984**, *57*, 115.
18. Martin, G. E.; Hadden, C. E. Long-range (1)H-(15)N heteronuclear shift correlation at natural abundance. *J. Nat. Prod.* **2000**, *63*, 543.
19. Fatemi, M. H. Prediction of ozone tropospheric degradation rate constant of organic compounds by using artificial neural networks. *Anal. Chim Acta.* **2006**, *556*, 355.
20. Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Dobrynin, A.; Gutman, I.; Domotor, G. The Szeged Index and an Analogy with the Wiener Index. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547.
21. Khadikar, P. V.; Karmarkar, S.; Agrawal, K. V. A novel PI index and its applications to QSPR/QSAR studies. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 934.
22. Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Karmarkar, S.; Agrawal, K. V. *J. Math. Chem.* **2001**, *29*, 134.
23. Khadikar, P. V.; Jaiswal, M. QSAR Study on ^{13}C NMR chemical shifts on carbinol carbon atoms. *Bioorg. & Med. Chem.* **2004**, *12*, 1793.
24. Duvenbeck, C. *Topological and Geometrical Approach to Develop Models for prediction of ^{13}C NMR Chemical Shifts*; Bochum: Germany, 1995.
25. Alam, T. M.; Henry, M. Empirical calculations of ^{29}Si NMR chemical shielding tensors: A partial charge model investigation of hydrolysis in organically modified alkoxy silanes. *Phys. Chem. Chem. Phys.* **2000**, *2*, 23.
26. Whalden, E. K.; Jurs, P. C. A simple method for the representation, quantification, and comparison of the volumes and shapes of chemical compounds. *Anal. Chem.* **1981**, *53*, 2184.
27. Stouch, T. R.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 4.
28. Stewart, J. J. P. MOPAC, semi empirical molecular orbital program. *QCPE* **1983**, *6*, 455.
29. Available at: www.disat.unimib.it/chm/
30. Consonni, V.; Todeschini, R. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, 2000; pp. 392–430.