ARTICLE

Application of PCA and HCA to the Structure-Activity Relationship Study of Fluoroquinolones

Xiao-hong Li[a], Xian-zhou Zhang[b], Xin-lu Cheng[a*], Xiang-dong Yang[a], Zun-lue Zhu[b]

a. Institute of Atomic and Molecular Physics, Sichuan University, Chengdu 610065, China; b. College of Physics and Information Engineering, He'nan Normal University, Xinxiang 453002, China

(Dated: Received on January 12, 2005; Accepted on April 25, 2005)

Density functional theory (DFT) was used to calculate molecular descriptors (properties) for 12 fluoroquinolone with anti-*S.pneumoniae* activity. Principal component analysis (PCA) and hierarchical cluster analysis (HCA) were employed to reduce dimensionality and investigate in which variables should be more effective for classifying fluoroquinolones according to their degree of an-*S.pneumoniae* activity. The PCA results showed that the variables $E_{\rm LUMO}$, Q_3, Q_5, Q_A, log$P$, MR, VOL and $\triangle E_{\rm HL}$ of these compounds were responsible for the anti-*S.pneumoniae* activity. The HCA results were similar to those obtained with PCA. The methodologies of PCA and HCA provide a reliable rule for classifying new fluoroquinolones with anti-*S.pneumoniae* activity. By using the chemometric results, 6 synthetic compounds were analyzed through the PCA and HCA and two of them are proposed as active molecules with anti-*S.pneumoniae*, which is consistent with the results of clinic experiments.

Key words: Structure-activity relationship, Density functional theory, Principal component analysis, Hierarchical cluster analysis

I. INTRODUCTION

Four decades after the discovery of the first member of the quinolone antibacterial family, nalidixic acid, more than 7000 new analogs have been documented in the literature. Since 1977, this class of synthetic antibacterial agents has been widely used clinically. In recent years, there has been a considerable interest in the development of new fluoroquinolone (FQ) agents [1,2]. Due to the fact that this pathogen reaches millions of people each year, in developed countries and in developing ones, there have been many studies on the main factors responsible for the diseases caused by *S.pneumoniae*. Structure-activity relationships (SAR) and quantitative SAR (QSAR) studies have been extensively used to correlate molecular structures to their biological activities [3-8]. A primary goal of QSAR/SAR methods is to find rules that can lead to reliable classification and prediction of the biological activity for tested, untested or hypothetical compounds. The obtained information can be used for the selection or design of better structures.

The aim of the present work is to investigate the relationship between some molecular properties and their anti-*S.pneumoniae* activities for fluoroquinolones by using quantum chemistry and chemometric methods and afterwards to propose new fluoroquinolones that may be potentially active anti-*S.pneumoniae*. The principal component analysis (PCA) and hierarchical cluster analysis (HCA), which were employed in this work to analyze the data set, are extremely useful to classify the molecules into groups that can be correlated to their anti-*S.pneumoniae* activities.

II. COMPUTATIONAL METHODS

A. Theoretical calculations of the molecular properties

Except the steric and hydrophobic features all the results were obtained using Becke' hybrid three-parameter functional [11-14] B3LYP/3-21G method. The steric and hydrophobic features were calculated using the HyperChem 6.0 program. All the structures have been fully optimized at the B3LYP/3-21G level of theory. The basic structure of fluoroquinolones can be seen in Fig.1.

Structural descriptors of compounds are usually correlated with biological activity. In this work we calculated the following structural descriptors to be correlated with the biological activity: the highest occupied molecular orbital energy ($E_{\rm HOMO}$); the lowest unoccupied molecular orbital energy ($E_{\rm LUMO}$); energy difference between the highest occupied and the lowest unoccupied molecular orbital ($\triangle E_{\rm HL}$); dipole moment (μ); molecular hardness (η); molecular sofeness ($1/\eta$); hydration energy (HE); net atomic charge on atom I (Q_i); net charge on ring (Q_N); Mulliken electronegativity (χ); molecular volume (VOL); molecular polar-

*Author to whom correspondence should be addressed. E-mail: lorna639@163.com, Tel: +86-28-85405526

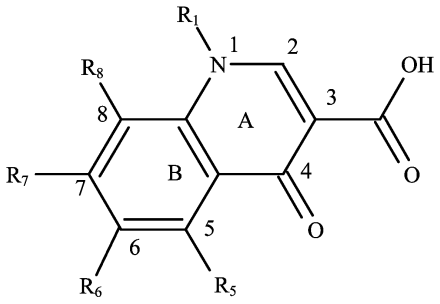


FIG. 1 The basic structure of fluoroquinolones

izability (Pol); partition coefficient (log$P$); molecular refractivity (MR); molecular mass (M); surface area (A).

The calculated parameters were selected so that they can represent the electronic ($\triangle E_{\rm HOMO}$, $\triangle E_{\rm LUMO}$, $\triangle E_{\rm HL}$, χ, μ, $1/\eta$, HE, Q_i, Q_N, MR), steric (A, VOL) and hydrophobic (log$P$) features of the compounds studied. These features are supposed to be important for investigating the SAR of drug [15-17].

B. Analysis method

Principal component analysis is a statistical technique that seeks to find a few mutually orthogonal linear combinations of the original variables to capture most of the variability present within the original system. In other words, a large part of the variance can usually be captured with a small number of components.

In the PCA, the first rows are standardized (unit variance, zero mean) to give a square matrix of moment correlations coefficients between pairs of rows. Computing the principal components of this matrix involves the computation of its eigen values and eigen vectors. The importance of these vectors is that they are orthogonal. In other words, a large proportion of the dispersion through n rows over the m columns may be accounted for by p dimensions. The p-normalized vectors give the directions of a set of p-orthogonal axes in p-dimensional space. The linearly independent principal components are ranked in terms of the amount of total variance that each component explains.

Hierarchical cluster analysis (HCA) technique examines the distances between the samples in a data set and represents this information as a two-dimensional plot called dendrogram. The HCA method is an excellent tool for preliminary data analysis. It is useful for examining data sets for expected or unexpected clusters, including the presence of outliers. It is informative to examine the dendrogram in conjunction with PCA as they give similar information in different forms. In HCA, each point forms an only cluster initially and then the similarity matrix is analyzed. The most similar points are grouped forming one cluster and the process is repeated until all the points belong to an only group [18].

III. RESULTS AND DISCUSSION

A. Pre-processing of the molecular descriptors

Before applying the PCA technique, each of the variables was autoscaled so that they could be compared to each other in the same scale. This method is very important because each variable is weighted equally and this provides a measure of the ability of a descriptor to discriminate classes of compounds [19].

B. Principal component analysis

In the application of PCA to the compounds listed in the Fig.2, several attempts were made to obtain a separation of the active compounds from the inactive compounds. The best separation was obtained when the 19 initial variables were reduced to 11 final variables by using the statistic procedure SPSS12.0. This suggests that the other 8 variables are not so important for classifying these compounds. Moreover, from the 11 variables selected, we obtained the correlation matrix between these variables and the respective calculated values. This correlation matrix is presented in Table I (the selected charges are Mulliken atomic charges in Table I). From Table I, we can see that some variables are correlated to each other (we considered correlated variables that possess correlation coefficients above 0.70) and, according to the results shown in Table I, only 8 variables are important for the separation between active and inactive compounds.

The best separation was obtained by using the following variables: $E_{\rm LUMO}$, Q_3, Q_5, Q_A, log$P$, MR, VOL and $\triangle E_{\rm HL}$, whose values are presented in Table II.

The PCA results show that the first principal component explained 31.241% of the total variance. The first two principal components explained 54.811% of the total variance. The first three principal components explained 72.613% of the total variance and the first four principal components explained 84.442%. The first five principal components explained 93.908% of total variance. Table III shows the loading vectors for PC1, PC2, PC3, PC4, PC5. Figure 3 displays the plot of the loading vectors for these first two principal components (PC1 and PC2). The plot of the score vectors for the first two principal components (PC1×PC2) is shown in Fig.4. From Fig.4, it can be seen that fluoroquinolones studied are separated into two groups, inactive group and active group against anti-*S.pneumoniae*, when the variables $E_{\rm LUMO}$, Q_3, Q_5, Q_A, log$P$, MR, VOL and $\triangle E_{\rm HL}$ were used to calculate the activities.

From Fig.4, it can be seen that PC1 alone is responsible for the separation of compounds with higher and lower anti-*S.pneumoniae*. According to Table III, PC1 can be expressed through the following equation:

$$\begin{aligned} PC1 = {} & 0.629E_{\rm LUMO} + 0.219 \triangle E_{\rm HL} + 0.386Q_3 \\ & + 0.493Q_5 - 0.290Q_A - 0.081{\rm log}P \\ & + 0.939MR + 0.669VOL \end{aligned} \quad (1)$$

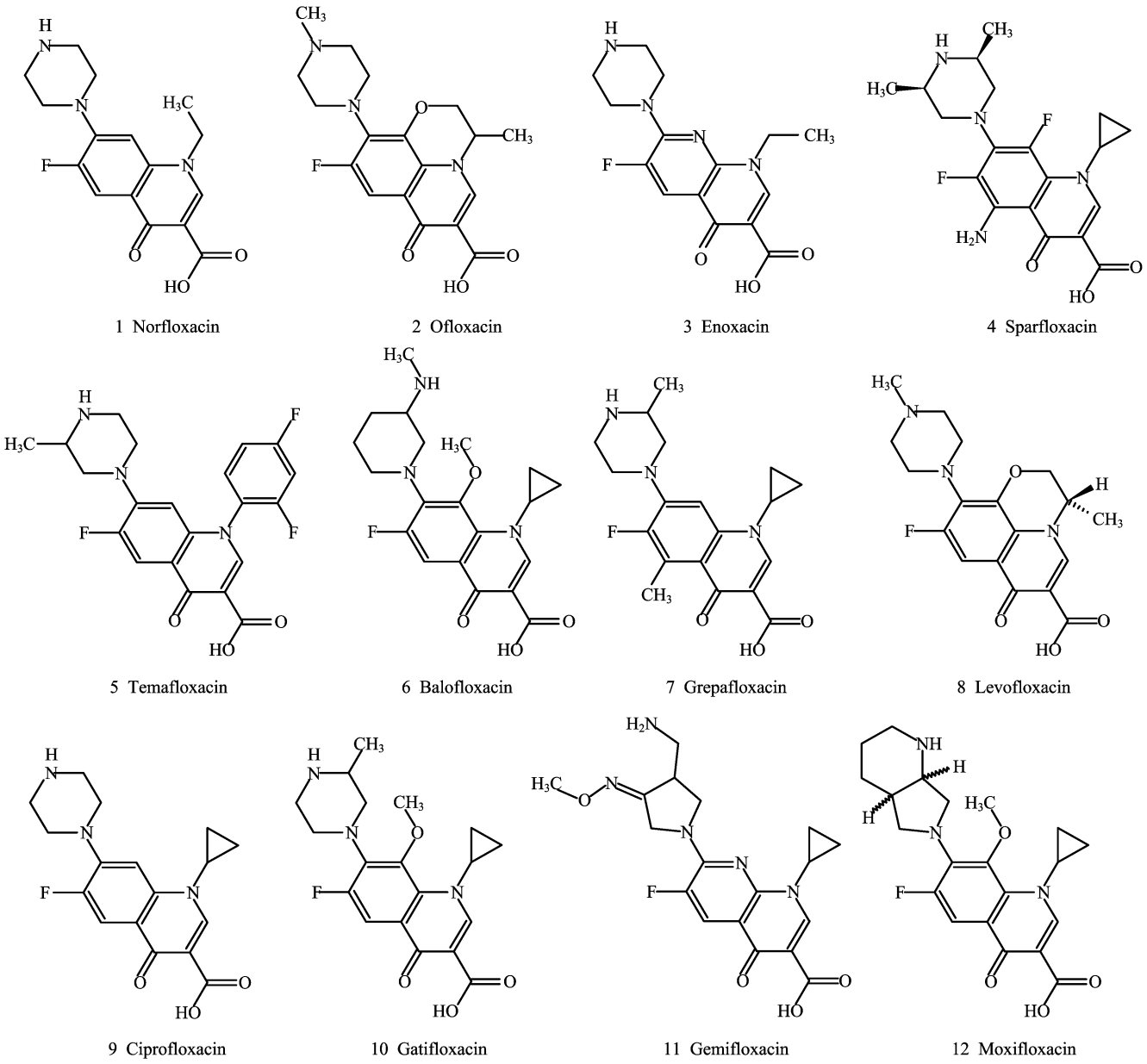


FIG. 2 Molecular structural schematic diagrams of 12 fluoroquinolones

From this equation, we can see that the more active molecules can be obtained when we have bigger values for the variables E_{LUMO}, $\triangle E_{\mathrm{HL}}$, MR, VOL, Q_3, Q_5, combined with smaller values of log$P$ and Q_A. These characteristics can be useful in the design of new compounds with high anti-*S.pneumoniae* activity, as some of these eight variables can be related to properties such as strength of molecular association by electrostatic interaction (Q_3, Q_5, Q_A and MR) and hydrophobicity (log$P$).

C. Hierarchical cluster analysis

In this study, we chose Ward's method to separate the compounds. The method provided an informative dendogram shown in Fig.5. The HCA separated the compounds in a similar way to that of PCA. We can see that the two groups I and II of Fig.5 are the same groups: the active and the inactive of Fig.4 in the PCA. That is to say that HCA and PCA classified the 12 fluoroquinolones under study in exactly the same way.

Starting from the separation of the 12 compounds (training set), we proposed 6 synthetic compounds for analysis (test set). With the application of the PCA and HCA methods of pattern recognition, we obtained a classification of 18 compounds in two groups. According to the method of PCA, the active compounds are expected to be 14 and 18, and the inactive ones to be 13, 15, 16 and 17. Figure 6 gives the separation of the test set into two groups by using PCA.

With the application of the HCA method, the result is the same as that of PCA. Two compounds were classified as active by the two methods. The results of

TABLE I Correlation matrix between the selected variables

	E_{LUMO}	Q_2	Q_3	Q_4	Q_5	Q_6	Q_A	log$P$	MR	VOL	$\triangle E_{\mathrm{HL}}$
E_{LUMO}	1.00	0.38	−0.09	0.55	0.45	0.42	−0.28	−0.37	0.62	0.03	0.12
Q_2		1.00	0.22	−0.02	0.74	−0.18	0.21	−0.10	0.24	0.15	0.21
Q_3			1.00	−0.25	−0.01	−0.29	0.20	−0.23	0.42	0.31	0.10
Q_4				1.00	0.44	0.62	−0.76	0.26	0.28	0.09	−0.17
Q_5					1.00	−0.13	0.01	−0.01	0.24	0.17	0.10
Q_6						1.00	−0.92	0.49	0.27	0.11	−0.30
Q_A							1.00	−0.63	−0.25	−0.27	0.37
log$P$								1.00	−0.15	0.38	−0.26
MR									1.00	0.64	0.15
VOL										1.00	0.11
$\triangle E_{\mathrm{HL}}$											1.00

TABLE II Values obtained for the eight most important properties (descriptors) that classified the 12 compounds of the training set as active and inactive molecules and the values of MIC_{90} for the compounds

Compound	E_{LUMO}[a]	Q_3[a]	Q_5[a]	Q_A[a]	log$P$	MR	VOL/Å^3	ΔE_{HL}[a]	MIC_{90}/(mg/L) [20-27]
1	−0.2000	−0.2062	−0.1966	−0.1248	4.76	47.98	875.69	0.1583	12.50
2	−0.1755	−0.2045	-0.2051	−0.1433	4.77	57.40	947.01	0.1354	3.13
3	−0.1972	−0.2041	−0.1450	0.1795	4.43	47.08	862.84	0.1494	25.0
4	−0.1857	−0.2045	0.3471	−0.1430	5.02	62.51	1024.06	0.1489	1.50
5	−0.1909	−0.2092	−0.1981	−0.2471	7.08	46.23	1045.95	0.1403	1.00
6	−0.2023	−0.2009	−0.1967	−0.1543	5.15	66.04	1049.57	0.1621	0.39
7	−0.1898	−0.2163	−0.0350	−0.1474	5.54	59.47	981.69	0.1525	0.50
8	−0.1761	−0.2044	−0.2050	−0.1449	4.77	57.40	945.68	0.1361	0.78
9	−0.1904	−0.2187	−0.1983	−0.1553	4.81	50.44	895.69	0.1498	0.78
10	−0.1808	−0.2077	−0.2055	−0.1785	4.99	61.28	1030.12	0.1386	0.50
11	−0.2046	−0.2052	−0.1501	0.1448	4.25	56.64	1055.95	0.1608	0.03
12	−0.1945	−0.1998	−0.1949	−0.1584	5.11	68.68	1055.62	0.1530	1.25

[a] The unitsof energy and charge are all atomic unit.

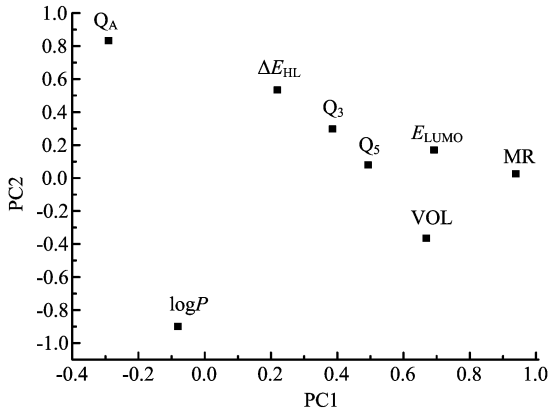


FIG. 3 PCA loading vectors (PC1 and PC2) for the eight variables responsible for the anti-*S.pneumoniae* activity for the 12 fluoroquinolones.

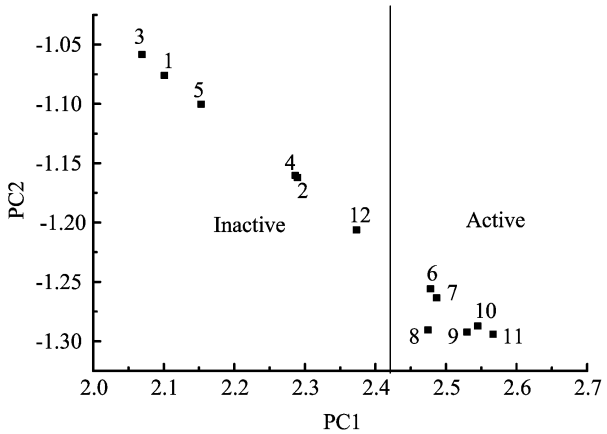


FIG. 4 PCA score (PC1 and PC2) for the 12 fluoroquinolones with anti-*S.pneumoniae* activity. The PC analysis leads to a separation into two groups: inactive group and active group.

this classification are summarized in Table IV. According to the result of clinic experiment, we know that the activities [28-31] in anti-*S.pneumoniae* for the six compounds are 125 L/g (Lomefloxacin), 2.564×10^3 L/g (Tosufloxacin), 80 L/g (Fleroxacin), 125 L/g (Pefloxacin), 62.5 L/g (Rufloxacin), 4 L/g (Trovafloxacin), respectively, which is consistent with our classification result.

TABLE III The loading vectors for PC1, PC2, PC3, PC4 and PC5

Variable	PC1	PC2	PC3	PC4	PC5
E_{LUMO}	0.692	0.170	−0.651	−0.167	−0.150
E_{HL}	0.219	0.534	0.154	0.618	−0.458
Q_3	0.386	0.297	0.677	−0.320	0.284
Q_5	0.493	0.079	−0.403	0.448	0.618
Q_A	−0.290	0.833	0.240	0.175	0.201
Log$P$	−0.081	−0.900	0.177	0.331	0.037
MR	0.939	0.026	0.107	−0.205	−0.145
VOL	0.669	−0.365	0.505	0.228	−0.020

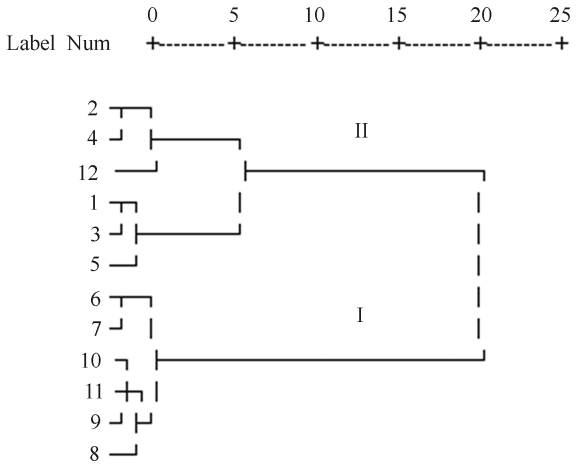


FIG. 5 Dendogram of the training set, separated into two groups: I (active) and II (inactive)

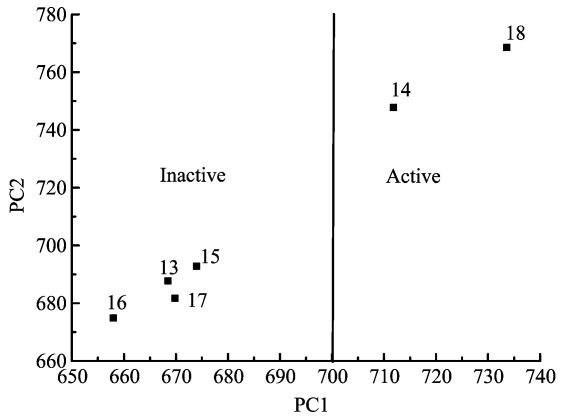
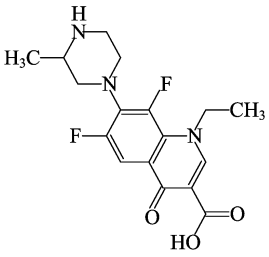
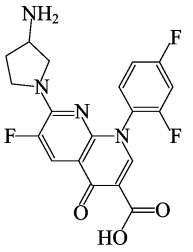
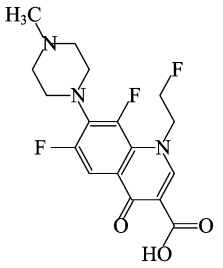
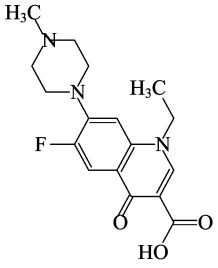
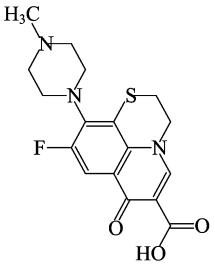
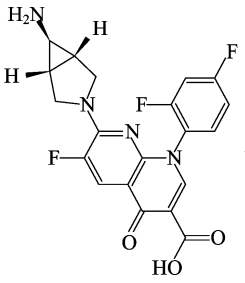


FIG. 6 PCA score (PC1 and PC2) for the 6 fluoroquinolones (test set) with anti-*S.pneumoniae* activity.

TABLE IV The results of two pattern recognitions for six new synthetic compounds: active compound (+) and inactive compound (−)

Compound	PCA	HCA	Compound	PCA	HCA
 13 Lomefloxacin	−	−	 14 Tosufloxacin	+	+
 15 Fleroxacin	−	−	 16 Pefloxacin	+	+
 17 Rufloxacin	−	−	 18 Trovafloxacin	+	+

IV. CONCLUSIONS

The application of the pattern recognition methods (PCA) in fluoroquinolones with activity in anti-*S.pneumoniae* showed that the compounds studied in this work can be correctly classified into two groups: active and inactive molecules. The PCA results showed that the variables E_{LUMO}, $\triangle E_{\mathrm{HL}}$, MR, VOL, log$P$, Q_3, Q_5, Q_A are responsible for the separation of active and inactive compounds. The HCA results were similar to those obtained with PCA, i.e. both methods classified the 12 fluoroquinolones under study in exactly the same way. With the application of the PCA and HCA methods of pattern recognition to other 6 synthetic compounds, the results of PCA and HCA are the same, 14 and 18 are active and 13, 15, 16, 17 are inactive, which is consistent with the result of clinic experiment. So the methodologies of PCA and HCA provide a reliable rule for classifying new fluoroquinolones with anti-*S.pneumoniae* activity. This also provides a theoretical basis for reasonable pharmic molecular design.

V. ACKNOWLEDGMENT

This work was supported by National Nature Science Foundation of China and China Academy of Engineering Physics (No. 10376021) and Provincial National Science Foundation of He'nan (No. 2004601107).

[1] A. Bryskier and J. F. Chantot, Drugs **49**, 16 (1995).
[2] Y. Asahina, T. Ishizaki and S. Suzue, Prog. Drug. Res. **38**, 57 (1992).
[3] D. S. Lawrence, J. T. Zilfou and C. D. Smith, J. Med. Chem. **42**, 4932 (1999).
[4] I. Abi-Zeid, Q. Yang and L. Lamontagne, Lect. Notes Artif. Int. **358**, 1650 (1999).
[5] A. M. Khawaja, Y. C. Lui and D. F. Rogers, Eur. J. Pharmacol. **384**, 173 (1999).
[6] A. Ruhmann, I. Bonk and A. K. Kopke, Peptides **20**, 1311 (1999).
[7] J. Hines, S. J. Fluharty and R. R. Sakai, Eur. J. Pharmacol. **384**, 81 (1999).
[8] P. Madsen, J. M. Lundbeck, P. Jakobsen, *et al.*, Bioorg. Med. Chem. **8**, 2277 (2000).
[9] M. J. Frisch, G. W. Truck, H. B. Schlegel, *et al.*, *GUASSIAN 03*, Gaussian, Inc., Pittsburgh PA.USA (2003).
[10] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford, New York (1989).
[11] C. Lee, W. Yang and R. G. Parr, Phys. Rev. B **37**, 785 (1980).
[12] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[13] A. D. J. Becke, Chem. Phys. **98**, 5648 (1992).
[14] W. Khon, A. D. Becke and R. G. J. Parr, J. Phys. Chem. **100**, 12974 (1996).
[15] C. N. Alves, J. C. Pinheiro, A. J. Camargo, *et al.* J. Mol. Struct. (Theochem) **491**, 123 (1999).
[16] G. Cruciani, M. Baroni, D. Bonelli, *et al.* Quant, Struct. Act. Relat. **9**, 101 (1990).
[17] M. G. Ford and D. J. Livingstone, Quant. Struct. Act. Relat. **9**, 107 (1990).
[18] B. R. Kowalski and C. F. Bender, J. Am. Chem. Soc. **94**, 5632 (1972).
[19] J. C. Lindon, E. Holmes and J. K. Nicholson, Prog. Nucl. Magn. Reson. Spectroscopy **39**, 01 (2001).
[20] C. J. Alexandeer, E. J. C. Goldstein, D. M. Citron and A. H. Decherney, Clinical medicine **6**, 313 (2000).
[21] G. G. Zgabel and A. M. Notrffin, Current Opinion in Pharmacology **1**, 459 (2001).
[22] B. D. Hoellman, G. Lin, M. R. Jacobs and P. C. Appelbaum, J. Antimicrob. Chemother. **43**, 645 (1999).
[23] M. A. Visalli, M. R. Jacobs and P. C. Appelbaum, Antimicrob. Agents. Chemother. **41**, 2786 (1997).
[24] L. M. Ednie, M. R. Jacobs and P. C. Appelbaum, Antimicrob. Agents. Chemother. **42**, 1269 (1997).
[25] T. Nakane, S. Iyobe, K. Sato and S. Mitsuhashi, Antimicrob. Agents. Chemother **39**, 2822 (1995).
[26] M. G. Cormican and R. N. Jones, Antimicrob. Agents. Chemother. **41**, 204 (1997).
[27] J. M. Woodcock, J. M. Andrews, J. F. Boswell, N. P. Brenwald and R. Wise, Antimicrob. Agents. Chemother. **41**, 101 (1997).
[28] D. Felmingham, M. D. O'Hare, M. J. Robbins, *et al.* Drugs Exp. Clin. Res. **11**, 317 (1985).
[29] R. N. Grüneberg, D. Felmingham and M. D. O'Hare, J. Antimicrob. Chemother. **22**(Suppl. C), 9 (1988).
[30] G. L. Ridgway, M. D. O'Hare, D. Felmingham, *et al.* Drugs. Exp.Clin. Res. **11**, 259 (1985).
[31] A. Bauernfeind and J. Antimicrob. Chemother. **40**, 639 (1997).