Journal of Solid State Chemistry **162,** 1–9 (2001)
doi:10.1006/jssc.2001.9283, available online at http://www.idealibrary.com on IDEAL®

Application of Multivariate Data Analysis Techniques in Modeling Structure–Property Relationships of Some Superconductive Cuprates

K. Lehmus and M. Karppinen[1]

Laboratory of Inorganic and Analytical Chemistry, Helsinki University of Technology, FIN-02015 Espoo, Finland

Received April 4, 2001; in revised form June 14, 2001; accepted June 20, 2001

Multivariate analysis methods are used for the examination of the fine structure and superconductivity properties of the 1212-type superconductive copper oxides with the stoichiometry of $MA_2QCu_2O_{6+z}$ (M = Cu, Hg, Tl/Pb; A = Ba, Sr; Q = rare-earth element, Ca; $z = 0-1$). PCA (principal component analysis) is used for the qualitative evaluation of relations between structural variables such as the cation–oxygen bond lengths, bond angles, and oxygen content z, and the superconductivity transition temperature T_C, adopted from a number of neutron diffraction studies published for $MA_2QCu_2O_{6+z}$ samples with various compositions. The different ways of doping positive holes in the superconductive CuO_2 plane are discussed on the basis of the PCA results. Quantitative modeling of the value of T_C in the $CuA_2QCu_2O_{6+z}$ system is successfully performed by PLS (projections to latent structures by means of partial least squares), resulting in a model with predictive power of ~93%. The present study demonstrates the potential of multivariate analysis methods in studying structure–property relations of inorganic materials with ionic structure. © 2001 Academic Press

Key Words: **multivariate analysis; PCA; PLS; superconductor; neutron diffraction.**

INTRODUCTION

Superconductive copper oxides have layered perovskite-related structures with a varying number of oxide layers and stoichiometries within the layers. Concentration of positive holes, i.e., the valence of copper (or oxygen) in the superconductive CuO_2 plane, is known to be an important parameter in determining the superconductivity transition temperature T_C (1). Upon increasing the hole concentration, the T_C in the so-called underdoped region increases up to a certain optimal level above which the material is overdoped and the T_C starts to drop. Besides hole concentration, distribution of holes in the Cu–O bond and flatness of the CuO_2 plane are believed to be important factors affecting the superconductivity characteristics (2–4). The value of T_C is believed to depend on the crystallographic fine structure, which is described by several structural variables, and consequently the appearance of superconductivity can be considered a multivariate problem.

Utilizing multivariate data analysis methods for the chemical problem solving is called chemometrics. Up to now chemometrics has not found many applications in the field of inorganic chemistry, whereas its use, e.g., in organic chemistry and chemical engineering is more widespread (5). Multivariate data analysis with so-called projection methods enables both qualitative and quantitative examination of correlations among the variables. Quantitative modeling of structure–property relationships has traditionally been applied to environmental chemistry and drug design to model, e.g., the toxicity of an organic compound or the activity of a drug candidate (6–9). The present study demonstrates the applicability of multivariate analysis techniques in evaluating relations between the fine structures and T_C of some superconductive copper-oxide phases, i.e., $MA_2QCu_2O_{6+z}$ phases with a two-CuO_2-plane unit cell. The few studies published so far about the multivariate analysis of copper-oxide superconductors have not dealt with the structure–property relations (10–12).

For the $MA_2QCu_2O_{6+z}$ or M-1212 phases, several stoichiometries and two slightly different crystallographical structures are known, i.e., M-1212:P and M-1212:RS where P stands for a perovskite and RS for a rock-salt-type structure of the MO_z layer (13). The two structures with the nomenclature of the atoms used in the present study are shown in Fig. 1. Note that for M-1212:RS a nomenclature different from the conventionally used one was adopted to make it comparable to that of M-1212:P. In both cases, M-1212:P and M-1212:RS, the hole concentration in the superconductive CuO_2 plane is controlled by changing cation or oxygen stoichiometry. From a structural point of view, creation of positive holes in the CuO_2 plane is seen as shortening of the Cu(2)–O(4) bond or lengthening of the effective A–O(2,3) or Q–O(2,3) distance in the structure.

[1]To whom correspondence should be addressed. Present address: Materials and Structures Laboratory, Tokyo Institute of Technology, Yokohama 226-8503, Japan. Fax: +81-45-924-5365. E-mail: karppinen@rlem.titech.ac.jp.

0022-4596/01 $35.00

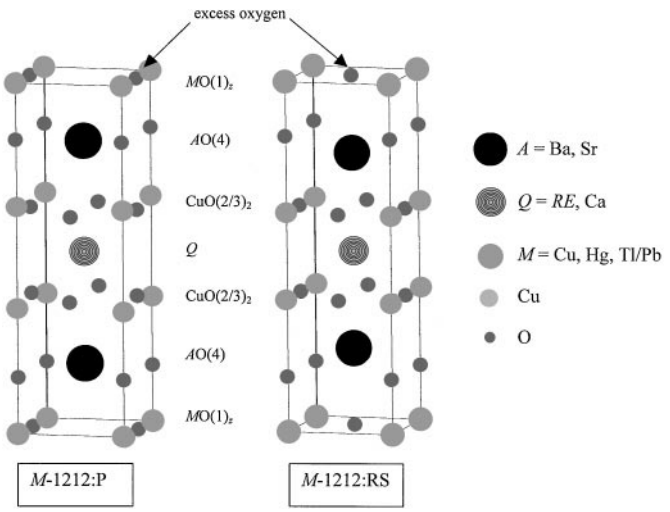

FIG. 1. Structures of $MA_2QCu_2O_{6+z}$ with an MO_z layer of perovskite type (M-1212:P) and rock-salt type (M-1212:RS).

These three different ways, which may work either individually or simultaneously, were defined as hole-doping routes, I, II, and III, respectively (3,4).

The location of the excess oxygen is different in the P- and RS-type structures. In the M-1212:P phases O_z is located close to cation M, while in the M-1212:RS phases O_z enters the rock-salt-type position above the A atom and thus far away from the M cation. Oxygen-stoichiometry variation is the widest in the triple perovskite $CuA_2QCu_2O_{6+z}$, i.e., Cu-1212:P, showing oxygen contents in the range of $0 \leq z \leq 1$ in its CuO_z charge-reservoir block. Moreover, the structure is flexible for various cation substitutions on both Q and A sites. The Cu-1212:P phase is the best known and most thoroughly characterized superconductive copper-oxide phase having a T_C of $\sim$90 K at maximum (14). In terms of the availability of structural data, the M-1212:P system is most suitable for the present multivariate analysis of the fine structure and T_C. The M-1212:RS structure (M = Hg, Tl/Pb) with a rock-salt-type MO_z layer has a higher maximum T_C, i.e., 128 K when M = Hg, Q = Ca, and $z \approx 0.23$ (15, 16). The amount of the excess oxygen in the M-1212:RS phases depends strongly on the M cation. In the case of mercury, the oxygen site in the MO_z layer is less than 50% occupied (17), while for thallium almost full occupancy is achieved (18).

EXPERIMENTAL

Structural Data of the M-1212 Phases

The structural data used in the multivariate analyses were adopted from neutron diffraction studies published for the M-1212:P/RS-type superconductive phases (18–41). The data set consists of 141 observations for $MA_2QCu_2O_{6+z}$ samples with varying M, A, Q, and z (Table 1). Each observation was described by 24 variables containing both structural data and the value of T_C (Table 2). Among the structural variables we find lattice constants, cation–oxygen bond lengths, flatness of cation–oxygen planes, oxidation state of cation Q, amount of excess oxygen z, and some cation–cation distances such as $d(\text{Cu–Cu})_1$ (the distance between two CuO_2 planes over the Q cation layer, i.e., thickness of the superconductive block) and $d(\text{Cu–Cu})_2$ (the distance between two CuO_2 planes over the AO–MO_z–AO layer block, i.e., thickness of the blocking block). The influence of external pressure ($0 \sim 0.6$ GPa) on the structure and T_C was evaluated as well (34, 39).

PCA and PLS

The "principal component analysis" (PCA) method is a multivariate projection method that provides us with a tool for modeling the correlation structure of a multivariate data set, usually expressed as a data matrix X containing N observations (here: samples) and K variables (here: structural variables and the value of T_C) (42). PC analysis can be used for searching trends, outliers, dominating variables, groups, and clusters among the data. The "projections to latent structures by means of partial least squares" (PLS) method is an extension of the PC analysis and is used when information on the dependence of two variable blocks X and Y is desired. The number of the y variables may be one or higher (here: T_C). PLS provides information on the importance of each variable on the selected y (42).

In the geometrical presentation of PC analysis, each observation gives a point in a K-dimensional space and thus a group of N points is formed. The average point of the group is centred in the origin of the K-dimensional system of coordinates. PCs are linear combinations of the initial variables and the first PC is the line that best approximates the group of data points. Usually one PC is insufficient to model the variation in a data set, and therefore a second PC must be calculated. The second PC is orthogonal to the first PC and is determined to improve the approximation of the data as much as possible. The distance between an observation and its projection onto the model plane defined by the two PCs is called the residual distance and needs to be examined to find outliers in the third dimension. If there are no deviating observations, the first two PCs are enough to model the data but sometimes more PCs are needed to get a good explanation for the data variation. Two PCs define a plane that can be considered as a window into the K-dimensional space. When all the points are projected onto the plane, the structure of the original data set is visualized in two dimensions. The projection points on the plane are called "scores" (t) or latent variables. To interpret the scores plot, i.e., to explain which original variables affect the location of each point, a "loadings" plot defined by "loading

TABLE 1
List of Observations Used in the Multivariate Analysis

Observations			
1.[a] $Yb_{0.9}Ca_{0.1}(Ba_{0.8}Sr_{0.2})_2Cu_3O_6$ (19)	41.[a] $Y_{0.9}Ca_{0.1}Ba_2Cu_3O_{6.71}$ (23)	81.[a] $YBa_2Cu_3O_{6.93}$ (27)	115. $YBa_2Cu_3O_{6.93}$ (34) (0.578 GPa)
2.[a] $Yb_{0.8}Ca_{0.2}(Ba_{0.8}Sr_{0.2})_2Cu_3O_6$ (19)	42.[a] $Y_{0.85}Ca_{0.15}Ba_2Cu_3O_{6.59}$ (23)	82. $YBa_2Cu_3O_{6.73}$ (27)	116. $YBa_2Cu_3O_{6.60}$ (34) (0 GPa)
3.[a] $Yb_{0.75}Ca_{0.25}(Ba_{0.8}Sr_{0.2})_2Cu_3O_6$ (19)	43.[a] $NdBa_2Cu_3O_{6.93}$ (24)	83.[a] $YBa_2Cu_3O_{6.64}$ (27)	117. $YBa_2Cu_3O_{6.60}$ (34) (0.103 GPa)
4.[a] $Yb_{0.65}Ca_{0.35}(Ba_{0.8}Sr_{0.2})_2Cu_3O_6$ (19)	44.[a] $Nd_{0.95}Ca_{0.025}Pr_{0.025}Ba_2Cu_3O_{6.95}$ (24)	84. $YBa_2Cu_3O_{6.55}$ (27)	118. $YBa_2Cu_3O_{6.60}$ (34) (0.208 GPa)
5.[a] $Yb(Ba_{0.8}Sr_{0.2})_2Cu_3O_6$ (19)	45. $Nd_{0.9}Ca_{0.05}Pr_{0.05}Ba_2Cu_3O_{6.95}$ (24)	85. $YBa_2Cu_3O_{6.48}$ (27)	119. $YBa_2Cu_3O_{6.60}$ (34) (0.310 GPa)
6.[a] $Yb(Ba_{0.8}Sr_{0.2})_2Cu_3O_{6.06}$ (19)	46.[a] $Nd_{0.8}Ca_{0.1}Pr_{0.1}Ba_2Cu_3O_{7.0}$ (24)	86.[a] $YBa_2Cu_3O_{6.45}$ (27)	120. $YBa_2Cu_3O_{6.60}$ (34) (0.415 GPa)
7.[a] $Yb(Ba_{0.8}Sr_{0.2})_2Cu_3O_{6.38}$ (19)	47. $LaBa_2Cu_3O_{7.06}$ (25)	87. $YBa_2Cu_3O_{6.41}$ (27)	121. $YBa_2Cu_3O_{6.60}$ (34) (0.489 GPa)
8.[a] $Yb(Ba_{0.8}Sr_{0.2})_2Cu_3O_{6.55}$ (19)	48. $PrBa_2Cu_3O_{6.98}$ (25)	88. $YBa_2Cu_3O_{6.38}$ (27)	122. $YBa_2Cu_3O_{6.60}$ (34) (0.563 GPa)
9.[a] $Yb(Ba_{0.8}Sr_{0.2})_2Cu_3O_{6.76}$ (19)	49. $NdBa_2Cu_3O_{6.98}$ (25)	89.[a] $YBa_2Cu_3O_{6.33}$ (27)	123. $HgBa_2CaCu_2O_{6.35}$ (35)
10. $Yb(Ba_{0.8}Sr_{0.2})_2Cu_3O_{6.96}$ (19)	50.[a] $SmBa_2Cu_3O_{6.98}$ (25)	90. $YBa_2Cu_3O_{6.28}$ (27)	124. $HgBa_2CaCu_2O_{6.28}$ (35)
11.[a] $Yb_{0.65}Ca_{0.35}(Ba_{0.8}Sr_{0.2})_2Cu_3O_{6.55}$ (19)	51.[a] $EuBa_2Cu_3O_{6.98}$ (25)	91. $YBa_2Cu_3O_{6.09}$ (27)	125. $HgBa_2CaCu_2O_{6.08}$ (36)
12.[a] $Yb_{0.65}Ca_{0.35}(Ba_{0.8}Sr_{0.2})_2Cu_3O_{6.74}$ (19)	52. $GdBa_2Cu_3O_{6.99}$ (25)	92. $YBa_2Cu_3O_{7.00}$ (28)	126. $HgBa_2CaCu_2O_{6.22}$ (36)
13. $YBa_2Cu_3O_7$ (20)	53. $DyBa_2Cu_3O_{6.98}$ (25)	93. $NdBa_2Cu_3O_7$ (29)	127. $HgBa_2CaCu_2O_{6.22}$ (37)
14.[b] $Y_{0.9}Ca_{0.1}Ba_2Cu_3O_{6.92}$ (20)	54.[a] $YBa_2Cu_3O_7$ (25)	94.[a] $Nd_{0.97}Ca_{0.015}Th_{0.015}Ba_2Cu_3O_{6.98}$ (29)	128. $HgBa_2CaCu_2O_{6.35}$ (37)
15.[b] $Y_{0.8}Ca_{0.2}Ba_2Cu_3O_{6.83}$ (20)	55. $HoBa_2Cu_3O_7$ (25)	95. $Nd_{0.94}Ca_{0.03}Th_{0.03}Ba_2Cu_3O_{6.97}$ (29)	129. $HgBa_2CaCu_2O_{6.22}$ (38)
16.[a] $YBa_2Cu_3O_{6.62}$ (20)	56.[a] $ErBa_2Cu_3O_7$ (25)	96.[a] $Nd_{0.9}Ca_{0.05}Th_{0.05}Ba_2Cu_3O_{6.96}$ (29)	130. $HgBa_2CaCu_2O_{6.23}$ (39) (0 GPa)
17.[a] $Y_{0.9}Ca_{0.1}Ba_2Cu_3O_{6.55}$ (20)	57. $TmBa_2Cu_3O_7$ (25)	97. $Nd_{0.8}Ca_{0.1}Th_{0.1}Ba_2Cu_3O_{6.96}$ (29)	131. $HgBa_2CaCu_2O_{6.23}$ (39) (0.154 GPa)
18.[a] $Y_{0.8}Ca_{0.2}Ba_2Cu_3O_{6.49}$ (20)	58.[a] $YbBa_2Cu_3O_7$ (25)	98.[a] $YBa_2Cu_3O_{6.99}$ (30)	132. $HgBa_2CaCu_2O_{6.23}$ (39) (0.304 GPa)
19. $YBa_2Cu_3O_6$ (20)	59. $LaBa_2Cu_3O_{6.04}$ (25)	99. $Y_{0.9}Ca_{0.05}Th_{0.05}Ba_2Cu_3O_{6.93}$ (30)	133. $HgBa_2CaCu_2O_{6.23}$ (39) (0.461 GPa)
20.[a] $Y_{0.9}Ca_{0.1}Ba_2Cu_3O_6$ (20)	60. $PrBa_2Cu_3O_{6.15}$ (25)	100.[a] $Y_{0.8}Ca_{0.1}Th_{0.1}Ba_2Cu_3O_{6.98}$ (30)	134. $HgBa_2CaCu_2O_{6.23}$ (39) (0.592 GPa)
21.[a] $Y_{0.8}Ca_{0.2}Ba_2Cu_3O_6$ (20)	61.[a] $NdBa_2Cu_3O_{6.12}$ (25)	101. $NdBa_2Cu_3O_7$ (31)	135. $(Tl,Pb)Sr_2CaCu_2O_7$ (40)
22. $YBa_2Cu_3O_{6.94}$ (21)	62. $SmBa_2Cu_3O_{6.11}$ (25)	102. $Y(Ba_{0.75}Sr_{0.25})_2Cu_3O_{6.95}$ (32)	136. $(Tl,Pb)Sr_2CaCu_2O_{6.99}$ (40)
23.[b] $Y_{0.98}Ca_{0.02}Ba_2Cu_3O_{6.91}$ (21)	63.[a] $EuBa_2Cu_3O_{6.13}$ (25)	103. $Y(Ba_{0.5}Sr_{0.5})_2Cu_3O_{6.96}$ (32)	137. $(Tl,Pb)Sr_2CaCu_2O_{6.96}$ (40)
24.[b] $Y_{0.96}Ca_{0.04}Ba_2Cu_3O_{6.93}$ (21)	64. $GdBa_2Cu_3O_{6.06}$ (25)	104. $Y(Ba_{0.375}Sr_{0.625})_2Cu_3O_{6.92}$ (32)	138. $Tl_{0.5}Pb_{0.5}Sr_2CaCu_2O_7$ (41)
25.[b] $Y_{0.95}Ca_{0.05}Ba_2Cu_3O_{6.94}$ (21)	65.[a] $DyBa_2Cu_3O_{6.12}$ (25)	105. $YSr_2Cu_3O_{6.84}$ (33)	139. $TlBa_2Ca_{0.8}Nd_{0.2}Cu_2O_{6.86}$ (18)
26.[b] $Y_{0.93}Ca_{0.07}Ba_2Cu_3O_{6.91}$ (21)	66. $YBa_2Cu_3O_{6.11}$ (25)	106. $Y(Ba_{0.5}Sr_{0.5})_2Cu_3O_{6.98}$ (32)	140. $TlBa_2Ca_{0.5}Nd_{0.5}Cu_2O_{6.86}$ (18)
27.[b] $Y_{0.91}Ca_{0.09}Ba_2Cu_3O_{6.92}$ (21)	67.[a] $HoBa_2Cu_3O_{6.1}$ (25)	107. $Y(Ba_{0.5}Sr_{0.5})_2Cu_3O_{6.8}$ (32)	141. $TlBa_2NdCu_2O_{6.86}$ (18)
28.[b] $Y_{0.9}Ca_{0.1}Ba_2Cu_3O_{6.91}$ (21)	68. $ErBa_2Cu_3O_{6.08}$ (25)	108. $Y(Ba_{0.5}Sr_{0.5})_2Cu_3O_{6.7}$ (32)	
29.[b] $Y_{0.88}Ca_{0.12}Ba_2Cu_3O_{6.9}$ (21)	69.[a] $TmBa_2Cu_3O_{6.1}$ (25)	109. $YBa_2Cu_3O_{6.93}$ (34) (0 GPa)	
30.[b] $Y_{0.86}Ca_{0.14}Ba_2Cu_3O_{6.94}$ (21)	70.[a] $YbBa_2Cu_3O_{6.14}$ (25)		
31.[b] $Y_{0.85}Ca_{0.15}Ba_2Cu_3O_{6.91}$ (21)	71. $YBa_2Cu_3O_{6.95}$ (26)	110. $YBa_2Cu_3O_{6.93}$ (34) (0.102 GPa)	
32.[b] $Y_{0.83}Ca_{0.17}Ba_2Cu_3O_{6.91}$ (21)	72. $YBa_2Cu_3O_{6.84}$ (26)		
33.[b] $Y_{0.8}Ca_{0.2}Ba_2Cu_3O_{6.91}$ (21)	73.[a] $YBa_2Cu_3O_{6.81}$ (26)	111. $YBa_2Cu_3O_{6.93}$ (34) (0.207 GPa)	
34.[b] $Y_{0.8}Ca_{0.2}Ba_2Cu_3O_{6.89}$ (22)	74. $YBa_2Cu_3O_{6.78}$ (26)		
35. $Y_{0.8}Ca_{0.2}Ba_2Cu_3O_{6.52}$ (22)	75.[a] $YBa_2Cu_3O_{6.73}$ (26)	112. $YBa_2Cu_3O_{6.93}$ (34) (0.310 GPa)	
36. $Y_{0.8}Ca_{0.2}Ba_2Cu_3O_{6.5}$ (22)	76. $YBa_2Cu_3O_{6.64}$ (26)		
37. $Y_{0.8}Ca_{0.2}Ba_2Cu_3O_{6.33}$ (22)	77.[a] $YBa_2Cu_3O_{6.58}$ (26)	113. $YBa_2Cu_3O_{6.93}$ (34) (0.416 GPa)	
38. $Y_{0.8}Ca_{0.2}Ba_2Cu_3O_{6.13}$ (22)	78. $YBa_2Cu_3O_{6.45}$ (26)		
39. $YBa_2Cu_3O_{6.76}$ (23)	79. $YBa_2Cu_3O_{6.35}$ (26)	114. $YBa_2Cu_3O_{6.93}$ (34) (0.496 GPa)	
40.[a] $Y_{0.95}Ca_{0.05}Ba_2Cu_3O_{6.75}$ (23)	80. $YBa_2Cu_3O_6$ (26)		

[a] Observations included in the training set of PLS analysis.
[b] Overdoped samples.

vectors" (p) is needed. The loading vectors express the orientation of the plane in the K-dimensional space, i.e., how each variable K contributes to the principal components.

A quantitative measure of the goodness of fit is given by the parameter R^2 describing how much of the variation is explained by the model. The goodness of fit prediction parameter Q^2 describes the predictive power of the model and is to be maximized in the analysis (42). In the present study the first two PCs were found to explain most of the data variation ($R^2 > 75\%$) and thus the results were considered in two dimensions.

Geometric presentation described for PC analysis is possible also for PLS. The difference between PC analysis and PLS is that PLS calculates scores for the data matrix X so that they will both approximate well X and correlate well with Y (42). One of the major application areas of PLS is "quantitative structure–property relationships modeling" (QSPR), which is a tool for estimating properties of a group of compounds based on their structure. In the present study QSPR was used for modeling the structure–property relationships in the M-1212:P system with M = Cu. PC analysis was used in defining the training set, i.e., the set of observations that is used for training the model. Selection of

TABLE 2
List of Original Variables and Their Explanation

Variable	Explanation
T_C	Superconductivity transition temperature
z	Amount of excess oxygen in MO_z layer
a, b, c	Lattice constants
V(Q)	(Average) oxidation state of Q cation
Cu–O(2), Cu–O(3), Cu–O(4)	Cation–oxygen bond lengths
M–O(1), M–O(4)	
Q–O(2), Q–O(3)	
A–O(1), A–O(2), A–O(3), A–O(4)	
Cu–A	Distance between Cu and A cation
Cu–Q	Distance between Cu and Q cation
d(Cu–Cu)$_1$	Thickness of the superconductive block
d(Cu–Cu)$_2$	Thickness of the blocking block
f(Cu–O(2))	Flatness of O(2)–Cu–O(2) plane
f(Cu–O(3))	Flatness of O(3)–Cu–O(3) plane
f(A–O(4))	Flatness of O(4)–A–O(4) plane

a representative and informative training set is essential to obtain a model with good predictive power, whereas validation of the model is of high importance in order to test the predictive power of the model in practice. As QSPR is a semiempirical method the models obtained are only locally valid, i.e., they embrace only compounds that are chemically and structurally alike. The model was made for underdoped samples only, and therefore the data of the overdoped samples, i.e., the Ca(II)-for-*RE*(III) substituted (*RE* = rare-earth element) oxygenated $CuBa_2(RE_{1-x}Ca_x)Cu_2O_{6+z}$ samples (Table 1, nos. 12, 14, 15, and 23–34), were excluded from the training set. Forty-three observations of underdoped samples out of altogether 122 observations of *M*-1212:P samples were chosen for the training set whose composition is indicated in Table 1. The remaining 47 underdoped samples were used for the external validation of the model.

As a pretreatment of the data, the values of different variables were scaled to unit variance in order to give them the same weight in the analyses. Since the values of the oxidation state of Q cation deviate strongly from normal distribution they were transformed by negative logarithm transformation (42). The PC analysis and PLS modelings were carried out with Simca-P 8.0 software provided by Umetrics. A detailed description of PC analysis and PLS algorithms used by the software can be found in the literature (5, 42). For both PC analysis and PLS the number of significant dimensions was determined by cross-validation (42).

RESULTS AND DISCUSSION

PC Analysis

Results of the PC analysis are interpreted visually by examining the scores and the loadings plots in which the first two PCs are described by two orthogonal lines. In the scores plot the scales of the axes are relative to the degree of explanation for data variation of the PC while in the loadings plot the scale tells the relative importance of each original variable on the PCs. Observations (their numbers) in the scores plot and variables in the loadings plot appear in the same place in respect to the PCs, and therefore the effect of a certain variable on the observations is obtained just by comparing the locations of observations and variables in the two plots. Furthermore, information on the correlation between variables is obtained from the loadings plot since variables grouped together in the plot correlate positively and those located on opposite sides of the origin correlate negatively with each other. In the scores plot the 95% tolerance region defined by Hotelling's T^2 is described as an ellipse, which provides a means for detecting deviating observations in the data set (42).

In the PC analysis based on the data of both *M*-1212:P and *M*-1212:RS samples (Table 1) only the structural variables related to the superconductive block were included since otherwise the differences resulting from the different structure of the MO_z layer in the two systems would dominate the analysis. However, as seen from the scores plot in Fig. 2a, observations corresponding to the *M*-1212:P- and *M*-1212:RS-type samples separated into two groups. Grouping of the observations is mainly due to the longer Cu–O(4) bonds and the flatter CuO_2 planes in the *M*-1212:RS samples than in the *M*-1212:P samples as explained by the loadings plot in Fig. 2b. Owing to the clear grouping of the *M*-1212:P and *M*-1212:RS samples, it was decided to analyze the two systems separately.

The M-1212:P system ($M = Cu$). In the PC analysis of the *M*-1212:P system (Fig. 3) the first principal component is determined mainly by the amount of excess oxygen z and variables closely related to z, such as the distances *M*-O(4), *A*-O(2), and *A*-O(3), flatness of the *A*O plane, and the T_C, which all correlate positively with each other (Fig. 3b). A negative correlation in the direction of the first PC is observed between z and the distances Cu–O(4) and d(Cu–Cu)$_2$. The high impact of z on the deviation of the observations ($R^2_{PC1} \approx 46\%$, Fig. 3a) was an expected result since many of the samples in the data set have z values of either 0 or 1, and thus significant differences in the fine structures and the T_C values. The second PC is strongly loaded by the variables depending on the size of the cations Q and A, i.e., the lattice constants a and c, and the distances Cu–O(2), *A*–O(4), and *Q*–O(2,3) ($R^2_{PC2} \approx 31\%$).

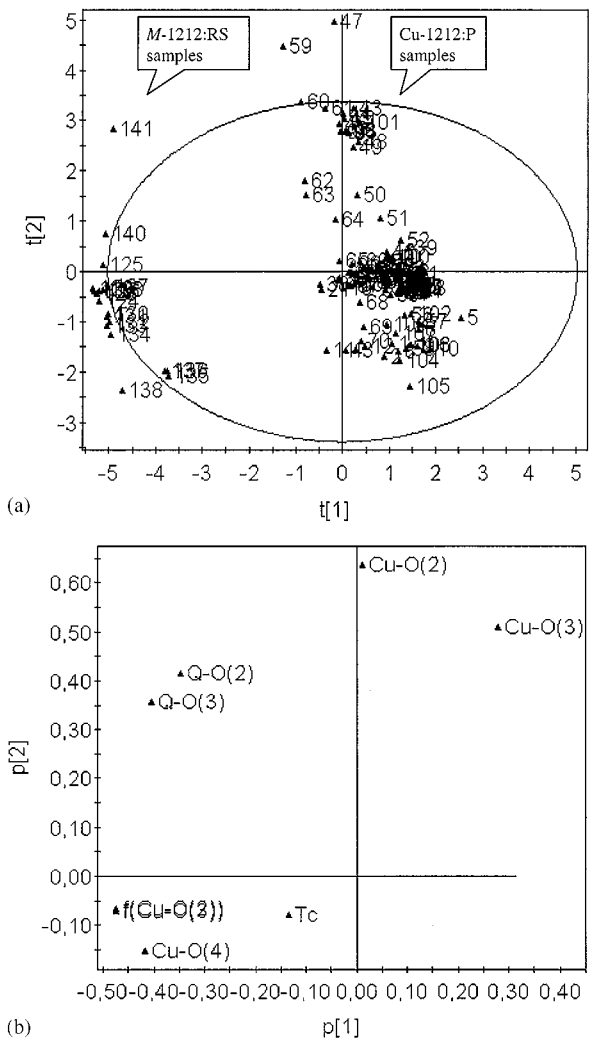

FIG. 2. PCA (a) scores and (b) loadings plots for the structural data of all M-1212-type samples. The first PC explains 52% (R^2_{PC1}) and the second PC 23% (R^2_{PC2}) of the data variation.

The strong correlation observed among the oxygen content z, the T_C, and the distances A–O(2), A–O(3), and Cu–O(4) (Fig. 3b) suggests that adding oxygen to the structure increases the hole concentration in the superconductive CuO_2 plane through the hole-doping routes I and II. This is in accordance with the conclusions made for the $Cu(Ba,Sr)_2(Yb,Ca)Cu_2O_{6+z}$ system based on neutron diffraction results (19), showing that incorporation of oxygen into the CuO_z chain results in a shift of O(4) toward the CuO_2 plane (route I) and Ba/Sr away from the in-plane O(2, 3) atoms (route II) as the oxidation proceeds. Substitution of trivalent Yb by divalent Ca in the $Cu(Ba,Sr)_2(Yb,Ca)Cu_2O_{6+z}$ system was shown to increase the concentration of holes in the CuO_2 plane by lengthening the effective (Yb,Ca)–O(2,3) bond length according to route III (19). In the PC analysis the effects of Ca(II)-for-RE(III) substitution at the Q site are explained mainly by the parameter V(Q), which is the average valence of the Q cation. V(Q) correlates negatively with the T_C, while the Q–O(2,3) bond length, which increases not only with Ca substitution but also with the size of the trivalent Q cation, does not directly correlate with the T_C (Fig. 3b).

The outlying observations in Fig. 3a correspond to the samples with the largest Q constituents, i.e., Pr and La (nos. 47, 59, 60), or the smallest A constituent, i.e., Sr (no. 105). In contrast to the other oxygenated $CuBa_2QCu_2O_{6+z}$ samples, the sample with Q = Pr is not superconductive and thus deviates from the data set. The phase with Q = La is known to be somewhat different with oxygen content higher than unity in the CuO_z layer ($z > 1$) and the excess oxygen atoms being located on both the a and b axes (25). The difference of

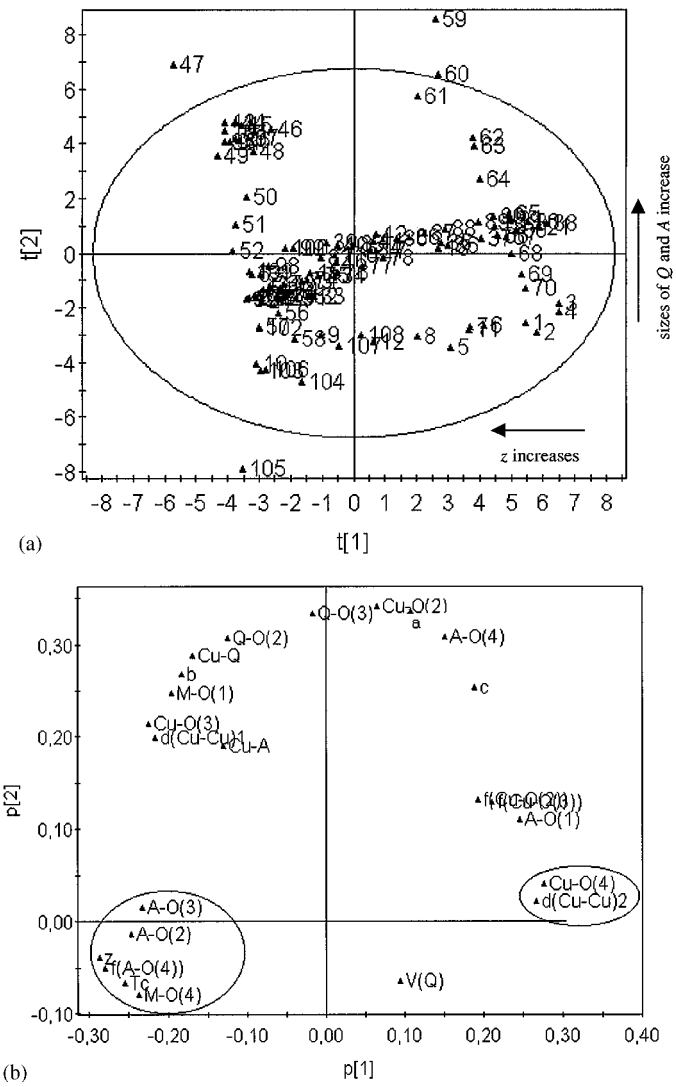

FIG. 3. PCA (a) scores and (b) loadings plots for the M-1212:P samples ($R^2_{PC1} \approx 46\%$ and $R^2_{PC2} \approx 31\%$). The variables correlating most strongly with z are circled.

the Sr-containing sample no. 105 may be explained by the fact that it was synthesized through high-pressure synthesis.

The effect of an external pressure on the structure T_C was examined utilizing the data reported for the $CuBa_2YCu_2O_{6+z}$ phase ($z = 0.60$ and 0.93) in a pressure range of $0 \leq p \leq 0.578$ GPa (34). The first PC ($R^2_{PC1} \approx 77\%$) is determined by the oxygen content and thus separates the samples into two groups due to the two oxygen content levels (Fig. 4). The pressure effect is included in the second PC ($R^2_{PC2} \approx 17\%$) and consequently the observations are distributed in the direction of the second PC according to the pressure applied. The variables strongly affected by the external pressure are M–O(1), Cu–O(3), and the lattice parameter b. As the pressure is increased, the b axis (and to some extent also c) is compressed due to the shortening of the M–O(1) and Cu–O(3) bonds. These variables dominate so strongly that the effects of the pressure on the other variables are not readily discernable. Upon increasing the pressure, charge transfer from M to Cu is supposed to occur (34). This seems to happen through lengthening of the M–O(4) bond and shortening of the Cu–O(4) bond (route I). The minor deviation of the z-dependent variables in the direction of the second PC shows (Fig. 4b) that the values of M–O(4) and T_C increase slightly with pressure. However, the dependence of these variables on the external pressure is quite weak as compared to their dependence on the oxygen content.

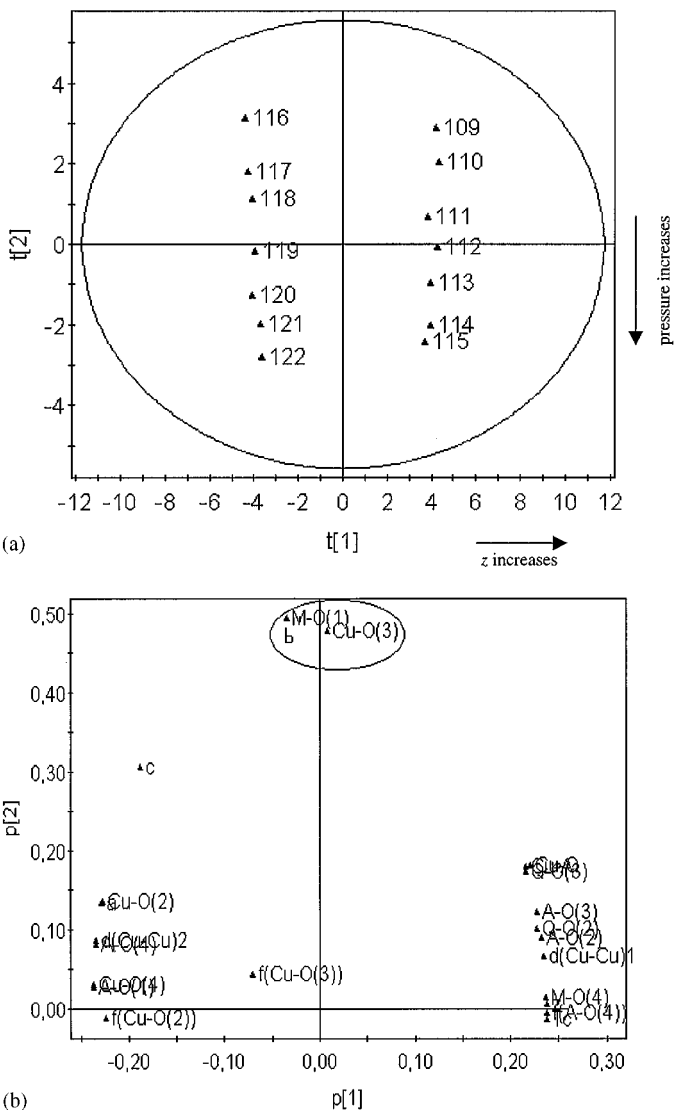

FIG. 4. PCA (a) scores and (b) loadings plots for the data of $CuBa_2YCu_2O_{6+z}$ samples ($z = 0.6$ and $z = 0.93$) measured under pressures of 0–0.578 GPa ($R^2_{PC1} \approx 77\%$ and $R^2_{PC2} \approx 17\%$). Structural variables most strongly affected by the external pressure are circled.

The M-1212:RS system. The number of observations in the PC analysis of the M-1212:RS system was significantly smaller ($N = 19$) than in the analysis of the M-1212:P system and furthermore the data were more inhomogeneous since they involved samples with different M cations (18, 35–41). Consequently, the variables depending on the MO_z layer were excluded from the analysis. The resulting PC analysis plots show, e.g., that the M = Hg samples possess the highest T_C values and the longest Cu–O(4) and A–O(4) bonds among the M-1212:RS samples (Fig. 5).

As for the Cu-1212:P system, the effect of an external pressure on the Hg-1212:RS structure was considered in a set of observations corresponding to the samples with the same cation composition. The positive effect of pressure on the T_C is clearly seen in the plots of Fig. 6. Upon increasing the pressure the unit cell shrinks as the bond lengths A–O(4), Cu–O(4), Cu–O(2,3), and Q–O(2,3) are shortened. The calculated first two principal components explain 76% of the data variation.

In the Hg-1212:RS system the oxygen content z and the bond length A–O(2,3) have a strong positive correlation with each other (Fig. 6b), which indicates that increasing the oxygen content produces holes into the CuO_2 plane through route II. The difference in the hole-doping routes being active when loading the M-1212:P and M-1212:RS phases with oxygen was explained based on bond-valence-sum considerations by the different crystallographical positions of the excess oxygen in the two structures (3,4). The oxygen atom in the RS-type MO_z charge-reservoir block is only loosely bonded to M since it is located above the A cation quite far away from M. Consequently, an increase in the oxygen content mainly shortens the bond from the excess oxygen atom to A and lengthens the bond from A to the in-plane oxygen atom O(2,3), resulting in hole doping into the CuO_2 plane by route II. According to the present analysis the flatness of CuO_2 plane increases with the oxygen content z but does not correlate with the value of T_C in the Hg-1212:RS system.

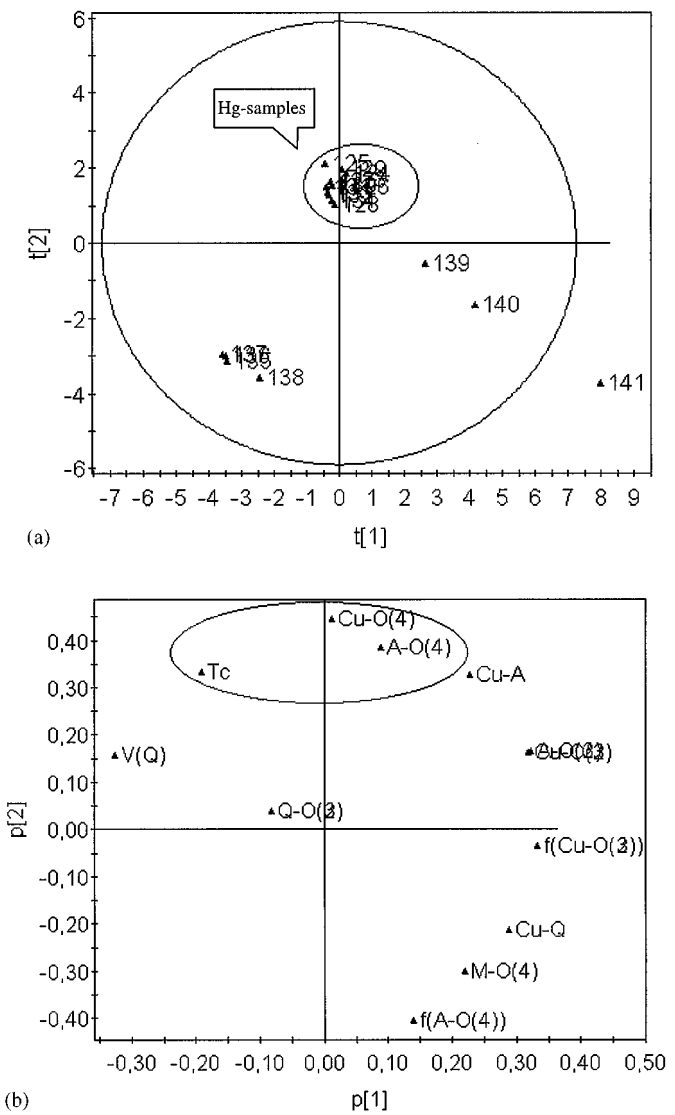

FIG. 5. PCA (a) scores and (b) loadings plots for the M-1212:RS samples ($R^2_{PC1} \approx 51\%$ and $R^2_{PC2} \approx 31\%$). The samples with M = Hg and the variables most strongly characterizing them are circled.

QSPR Modeling of T_C in the M-1212:P System

The PLS analysis based on the training set of underdoped M-1212:P samples (M = Cu) resulted in a model with five principal components and the goodness-of-fit parameters of $R^2X \approx 96\%$ and $R^2Y \approx 96\%$. The predictive power of the model, $Q^2 \approx 93\%$, can be considered excellent for a model based on data adopted from different studies. The plot of variable importance in Fig. 7 shows that the oxidation state of the Q cation, the bond length A–O(1), and the oxygen content z are the most significant x variables in determining the T_C in the model. As expected, the bond lengths Cu–O(4) (hole-doping route I) and A–O(2) (route II) are also of great importance in determining the T_C. Moreover, the flatness of the A–O(4) plane and the A–O(3) bond length are essential variables, whereas variables a and Cu–O(2) are less meaningful for the model. Quite interestingly, according to the present PLS analysis, flatness of the CuO_2 plane is a less significant factor in determining the T_C. Omitting the seven least meaningful variables from the modeling did not markedly weaken the predictive power of the model but resulted in a model of three PCs with $Q^2 \approx 92\%$.

The T_C values for the underdoped samples are well predicted by the PLS model as illustrated in the plot of the observed and predicted T_C values (Fig. 8a). As being quite understandable, the model is not capable of predicting the T_C values for the overdoped samples correctly but the predicted values are higher than the observed ones.

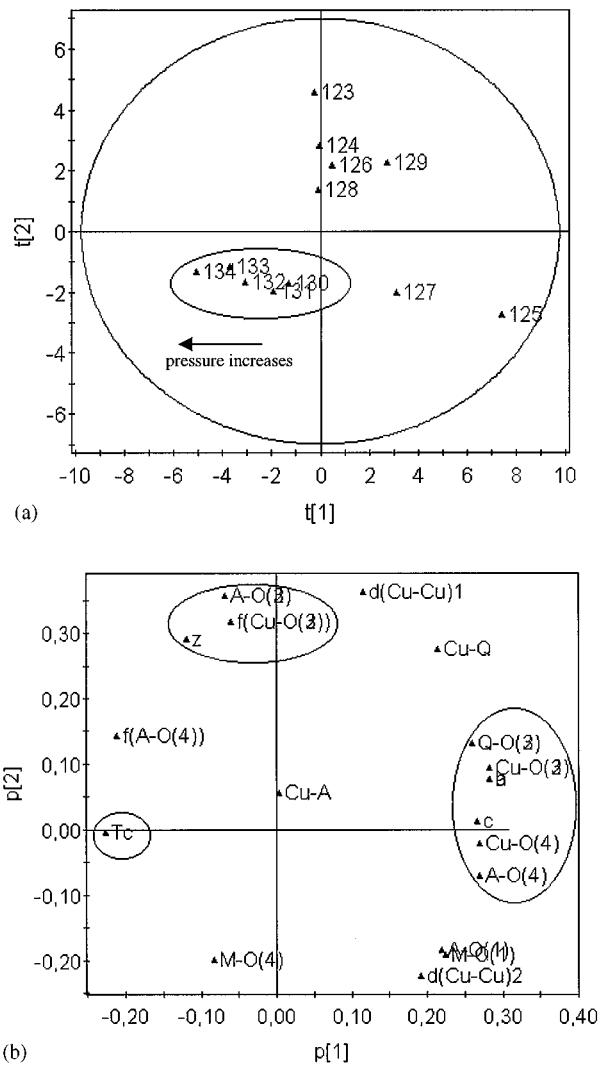

FIG. 6. PCA (a) scores and (b) loadings plots for the data of $HgBa_2CaCu_2O_{6+z}$ samples measured under pressures of 0–0.592 GPa ($R^2_{PC1} \approx 50\%$ and $R^2_{PC2} \approx 26\%$). The variables affected by the external pressure and those correlating positively with z are circled.

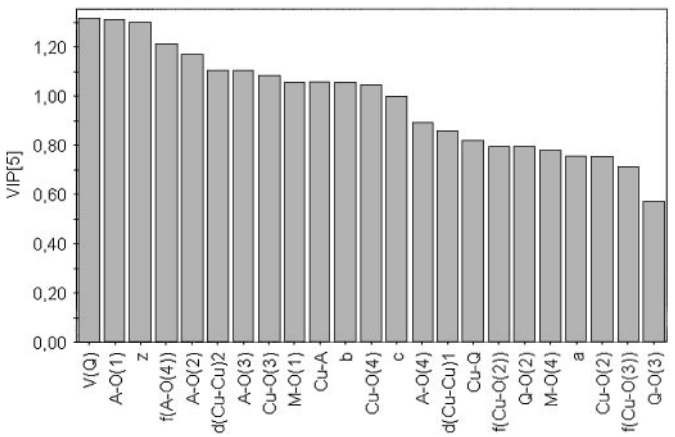

FIG. 7. Relative importance of variables in the PLS analysis of the M-1212:P system.

Consequently, the fully oxygenated and Ca(II)-for-Q(III) substituted samples (nos. 12, 14, 15, 23–34) diverge from the general trend in Fig. 8b. On the other hand, the Sr(II)-for-Ba(II) substituted samples (nos. 10, 102–108) do not deviate from the model, even though Sr-for-Ba substitution has been proposed to increase the hole concentration within the CuO_2 plane (43). In the modeling, the effects of Sr-for-Ba substitution are probably taken into account due to the $Cu(Ba_{0.8}Sr_{0.2})_2(Yb,Ca)Cu_2O_{6+z}$ samples included in the training set. As expected, the T_C values for the oxygenated samples with Q = La or Pr (nos. 47, 48) are not properly predicted. The effect of pressure on the structure and T_C in the $CuBa_2YCu_2O_{6+z}$ system is too small ($dp/dT_C \approx$ 5 K/GPa (34)) to be detected from the plot.

In order to improve the predictive power of the model the error sources in the sample characterization should be minimized. One of the most important parameters in determining the T_C in the M-1212:P phases is the oxygen content z. Inaccuracy in the oxygen content determination would cause an observation to deviate from the model. The oxygen contents of the samples have been determined only in few studies by means of chemical analysis. The chemical analysis methods are known to be highly accurate (1), while refinement of neutron diffraction data gives less precise results. The T_C values reported may also contain a variation of a few degrees in different studies depending on the way the measurement has been carried out. Furthermore, in the samples with intermediate oxygen contents the transition is not sharp, which may cause inaccuracy in the reported T_C values. Modeling of the samples with low T_C values is less precise than those with high T_C values partially due to the large number of different nonsuperconductive samples in the data set for all of which the T_C value was set at 0 K.

If the cation composition of the M, A, and Q sites and the oxygen content of the structure were varied in a systematic way and the sample characterization were performed in a consistent way, multivariate modeling of T_C within the M-1212 system would probably result in a model with even higher predictive power. Moreover, extending the multivariate analysis to cover all superconductive copper-oxide structures with different numbers of layers in both the superconductive block and the blocking block would be an interesting challenge.

CONCLUSIONS

Suitability of multivariate data analysis methods in studying the structure–property relationships of ionic compounds was demonstrated by applying PC analysis and PLS techniques to the structural data of $MA_2QCu_2O_{6+z}$

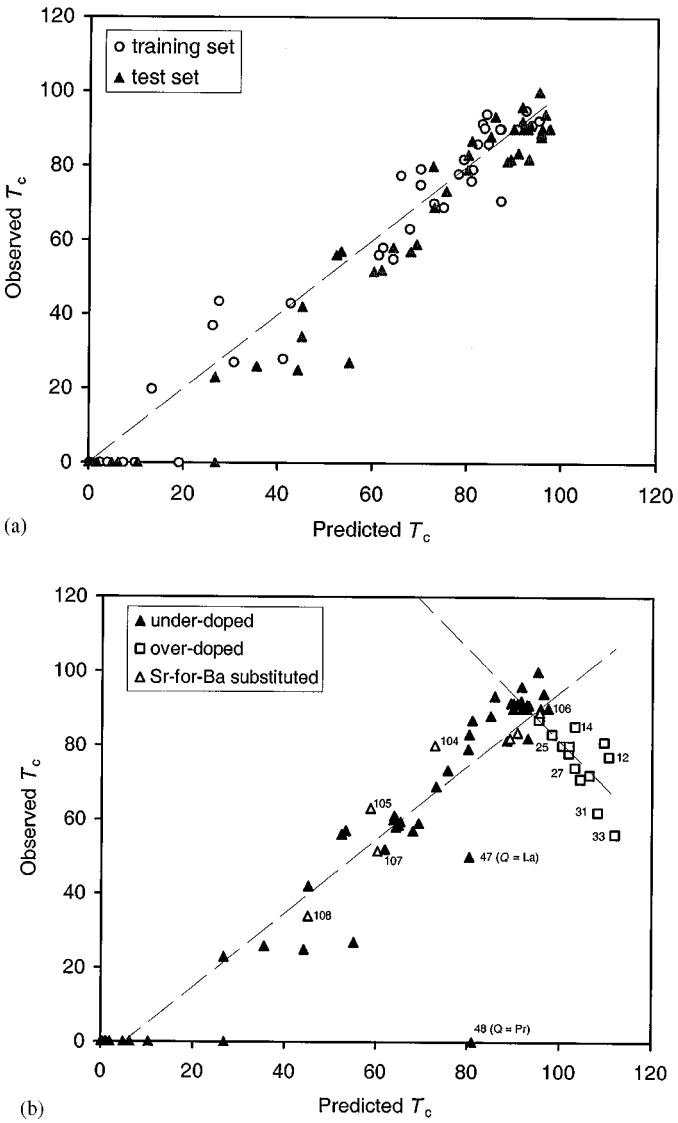

FIG. 8. Observed *versus* predicted T_C (a) for the underdoped and (b) for both the under- and overdoped M-1212:P samples. The overdoped and the Sr-for-Ba substituted samples are indicated.

superconductors. PC analysis indicated groupings of the samples, revealed the deviating samples, and provided qualitative information on the relationships between different variables, i.e., the structural variables and the T_C values obtained from neutron diffraction studies of the M-1212:P (M = Cu) and M-1212:RS (M = Hg, Tl/Pb) systems. PLS modeling was performed for the M-1212:P system, leading to a model with a predictive power of $\sim$93% for T_C, which is considered very good for a model based on data collected from a number of different studies. Careful experimental design and characterization of the sample series would enable multivariate modeling of T_C with even higher accuracy. The results obtained provide a positive example of the application of PLS for the modeling of structure–property relationships of ionic compounds in general.

ACKNOWLEDGMENTS

Professor H. Yamauchi is thanked for continuous discussions concerning the superconductive materials and their structure–property relations. Dr. T. Siimes and M.Sc. T. Rajalahti are acknowledged for fruitful advice concerning the multivariate analysis. Financial support from the Academy of Finland (Decision 46039) and the Jenny and Antti Wihuri Foundation is gratefully acknowledged.

REFERENCES

1. M. Karppinen and H. Yamauchi, *Mater. Sci. Eng. R* **26,** 51 (2000).
2. O. Chmaissem, J. D. Jorgensen, S. Short, A. Knizhnik, and Y. Eckstein, *Nature* **397,** 45 (1999).
3. M. Karppinen and H. Yamauchi, *Philos. Mag. B* **79,** 343 (1999).
4. M. Karppinen and H. Yamauchi, *Int. J. Inorg. Mater.* **2,** 589 (2000).
5. S. Wold, M. Sjöström, and L. Eriksson, *in* "Encyclopedia of Computational Chemistry" (P. von Ragué Schleyer, Ed.), pp. 2006–2022, Wiley, New York, 1999.
6. H. J. M. Verhaar, E. U. Ramos, and J. L. M. Hermens, *J. Chemom.* **10,** 149 (1996).
7. L. Eriksson, J. Jonsson, and R. Berglind, *Environ. Toxicol. Chem.* **12,** 1185 (1993).
8. J. T. McCloskey, M. C. Newman, and S. B. Clark, *Environ. Toxicol. Chem.* **15,** 1730 (1996).
9. G. Schüürmann, K. Rayasamuda, and U. Kristen, *Environ. Toxicol. Chem.* **15,** 1702 (1996).
10. I. I. Bondarenko, B. A. Treiger, V. Rezvitskii, and L. N. Mazalov, *Analyst* **117,** 795 (1992).
11. O. Grossmann and A. N. Turanov, *Anal. Chim. Acta* **257,** 195 (1992).
12. X. Wang, H. Song, G. Qiu, and D. Wang, *J. Mater. Sci. Technol.* **16,** 327 (2000).
13. H. Yamauchi, M. Karppinen, and S. Tanaka, *Physica C* **263,** 146 (1996).
14. M. K. Wu, J. R. Ashburn, C. J. Torng, P. H. Hor, R. L. Meng, L. Gao, Z. J. Huang, Y. Q. Wang, and C. W. Chu, *Phys. Rev. Lett.* **58,** 908 (1987).
15. S. N. Putilin, E. V. Antipov, O. Chmaissem, and M. Marezio, *Nature* **362,** 226 (1993).
16. A. Schilling, M. Antoni, J. D. Guo, and H. R. Ott, *Nature* **363,** 56 (1993).
17. M. Marezio, J.-J. Capponi, P.-G. Radaelli, P. P. Edwards, A. R. Armstrong, and W. I. F. David, *Eur. J. Solid State Inorg. Chem.* **31,** 843 (1994).
18. C. Michel, E. Suard, V. Caignaert, C. Martin, A. Maignan, M. Hervieu, and B. Raveau, *Physica C* **178,** 29 (1991).
19. M. Karppinen, H. Yamauchi, K. Fujinami, T. Nakane, K. Peitola, H. Rundlöf, and R. Tellgren, *Phys. Rev. B.* **60,** 4378 (1999).
20. P. Berastequi, S.-G. Eriksson, L. G. Johansson, M. Kakihana, M. Osada, H. Mazaki, and S. Tochihara, *J. Solid State Chem.* **127,** 56 (1996).
21. G. Böttger, I. Mangelschots, E. Kaldis, P. Fischer, Ch. Krüger, and F. Fauth, *J. Phys.: Condens. Matter* **8,** 8889 (1996).
22. J. Hejtmánek, Z. Jirák, K. Knízek, M. Dlouhá, and S. Vratislav, *Phys. Rev. B* **54,** 16,226 (1996).
23. V. P. S. Awana, S. K. Malik, and W. B. Yelon, *Physica C* **262,** 272 (1996).
24. P. Lundqvist, Ö. Rapp, R. Tellgren, and I. Bryntse, *Phys. Rev. B* **56,** 2824 (1997).
25. M. Guillaume, P. Allenspach, W. Henggeler, J. Mesot, B. Roessli, B. Staub, P. Fischer, A. Furrer, and V. Trounov, *J. Phys.: Condens. Matter* **6,** 7963 (1994).
26. R. J. Cava, A. W. Hewat, E. A. Hewat, B. Batlogg, M. Marezio, K. M. Rabe, J. J. Krajewski, W. F. Peck, Jr., and L. W. Rupp, Jr., *Physica C* **165,** 419 (1990).
27. J. D. Jorgensen, B. W. Veal, A. P. Paulikas, L. J. Nowicki, G. W. Crabtree, H. Claus, and W. K. Kwok, *Phys. Rev. B* **41,** 1863 (1990).
28. J. J. Capponi, C. Chaillout, A. W. Hewat, P. Lejay, M. Marezio, N. Nguyen, B. Raveau, J. L. Soubeyroux, J. L. Tholence, and R. Tournier, *Europhys. Lett.* **3,** 1301 (1987).
29. P. Lundqvist, C. Tengroth, Ö. Rapp, R. Tellgren, and Z. Hegedüs, *Physica C* **269,** 231 (1996).
30. M. Andersson, Ö. Rapp, and R. Tellgren, *Solid State Commun.* **81,** 425 (1992).
31. M. J. Kramer, S. I. Yoo, R. W. McCullum, W. B. Yelon, H. Xie, and P. Allenspach, *Physica C* **219,** 145 (1994).
32. F. Licci, A. Gauzzi, M. Marezio, G. P. Radaelli, R. Masini, and C. Chaillout-Bougerol, *Phys. Rev. B* **58,** 15208 (1998).
33. A. Gauzzi, E. Gilioli, F. Licci, M. Marezio, S. Massidda, F. Bernardini, A. Continenza, and P. G. Radaelli, *Proc. SPIE* **4058,** 12 (2000).
34. J. D. Jorgensen, S. Pei, P. Lightfoot, D. G. Hinks, B. W. Veal, B. Dabrowski, A. P. Paulikas, and R. Kleb, *Physica C* **171,** 93 (1990).
35. E. V. Antipov, J. J. Capponi, C. Chaillout, O. Chmaissem, S. M. Loureiro, M. Marezio, S. N. Putilin, A. Santoro, and J. L. Tholence, *Physica C* **218,** 348 (1993).
36. P. G. Radaelli, J. L. Wagner, B. A. Hunter, M. A. Beno, G. S. Knapp, J. D. Jorgensen, and D. G. Hinks, *Physica C* **216,** 29 (1993).
37. Q. Huang, J. W. Lynn, R. L. Meng, and C. W. Chu, *Physica C* **218,** 356 (1993).
38. S. M. Loureiro, E. V. Antipov, J. L. Tholence, J. J. Capponi, O. Chmaissem, Q. Huang, and M. Marezio, *Physica C* **217,** 253 (1993).
39. B. A. Hunter, J. D. Jorgensen, J. L. Wagner, P. G. Radaelli, D. G. Hinks, H. Shaked, R. L. Hitterman, and R. B. Von Dreele, *Physica C* **221,** 1 (1994).
40. D. M. Ogborne and M. T. Weller, *Physica C* **220,** 389 (1994).
41. J. B. Parise, P. L. Gai, M. A. Subramanian, J. Gopalakrishnan, and A. W. Sleight, *Physica C* **159,** 245 (1989).
42. L. Eriksson, E. Johansson, N. Kettaneh-Wold, and S. Wold, "Introduction to Multi- and Megavariate Data Analysis using Projection Methods (PCA & PLS)." Umetrics Ab, Umeå, 1999.
43. R. S. Liu, C. Y. Chang, and J. M. Chen, *Inorg. Chem.* **37,** 5527 (1998).