Quantitative Structure−Activity Relationships of the Synthetic Substrates for Elastase Enzyme Using Nonlinear Partial Least Squares Regression

Toshiro Kimura, Yoshikatsu Miyashita,† Kimito Funatsu,* and Shin-ichi Sasaki

Department of Knowledge-Based Information Engineering, Toyohashi University of Technology, Tempaku, Toyohashi 440, Japan

Received September 6, 1995⊗

Eighty-nine synthetic substrates for elastase enzyme and its activities (log $1/K_m$, log k_{cat}, and log k_{cat}/K_m) are treated using *partial least squares* (PLS) and *quadratic partial least squares* (QPLS). Chemical features of synthetic substrates are described using *principal properties*(PPs). By using the QPLS method, we obtain the nonlinear model equations for three properties (log $1/K_m$, log k_{cat}, and log k_{cat}/K_m) with the correlation coefficient 0.736, 0.918, and 0.868, respectively. Also, the predictive correlation coefficients for these model equations are 0.640, 0.865, and 0.793, respectively. By this study, it becomes clear that the z_2 value of the amino acid residue on position A and the size of the side chain for amino acid residues on position B are related to the properties of the synthetic substrates.

1. INTRODUCTION

Many methods are used in chemometrics. The representative method named *quantitative structure−activity relationships* (QSAR) based on mathematics and statistics is often used in the study of chemometrics. Relationships between features and biological activity or chemical property for chemical materials are modeled in QSAR studies, and various methods are used for modeling. In the past, linear modeling methods, such as *multiple linear regression* (MLR), *principal component regression* (PCR), etc., were used widely. In 1984, a new linear modeling method called *partial least squares regression* (PLS) developed by H. Wold and S. Wold was introduced.[1−3] In the PLS algorithm, modeling data is decomposed to orthogonal vectors with the same conception as PCR, and the model equations are built simultaneously. Nowadays, PLS is applied to some QSAR problems, *comparative molecular field analysis* (CoMFA), and other applications.

Recently, the opportunity of treating the nonlinear data was increased following the extension of the QSAR applications. Again, it can be considered that one can get good model equations when one uses the nonlinear modeling method. Therefore, a great variety of nonlinear modeling methods is developed by many researchers. For example, modeling methods with polynomial functions, spline functions, exponential functions, etc., are introduced in many articles.

In 1989, the modeling method within the quadratic polynomial was developed by S. Wold. Since the method is developed by extending the PLS model, it is called the *quadratic PLS* (QPLS) method.[4] In the QPLS calculation, bilinear model equations are used the same as the linear PLS, and bilinear model equations are linked using nonlinear (quadratic) polynomial called nonlinear inner relation equation.

In this study, activity data about synthetic substrate for elastase is treated. It is said that the elastase enzyme is related to some disease. Especially, the relationship between pathogenesis of emphysema and elastase is reported since the 1960s. In 1993, Nomizu et al. reported the result of modeling the activity data about synthetic substrate for elastase.[5] In their paper, log $1/K_m$, log k_{cat}, and log k_{cat}/K_m for synthetic substrate (*Suc-Ala-Ala-Ala-pNa*) and analogous (*Suc-A-B-Ala-pNa*) are modeled using the Free-Wilson/Fujita-Ban method. With this method, the existence of the particular feature in the particular sample is expressed using the indicator variable that takes 0 and 1. In Nomizu et al., kinds of the substituted amino acid residues on A and B are expressed using indicator variables. By using this method, one can express the information about the kind of the amino acid residue contained in particular sample. However, one cannot predict the activity corresponding to the particular sample that includes the amino acid residue that is not contained in the training data.

We thought that we can express the information about the amino acid residues on A and B to make model equations except for the use of an indicator variable, and we can obtain good model equations for the same data set in Nomizu et al. In this paper, the information for the particular sample is expressed using *principal properties* (PPs)[6,7] to avoid the disadvantage mentioned above.

2. MATERIALS AND METHODS

2.1. Data Set. Elastase is one of the enzymes in the body of a human being. It is said that this enzyme is related to the metabolism of a human being, and it is also considered that elastase enzyme is related to some disease. Especially, relationships between elastase enzyme and emphysema are reported since the 1960s.

In 1993, the previous QSAR study for synthetic substrate is reported by Nomizu et al.[5] In this paper, coefficients related to the reaction rate (log $1/K_m$, log k_{cat}, and log k_{cat}/K_m) are treated. The general formula of the synthetic substrate is expressed as *Suc-A-B-Ala-pNa*. A is substituted with *Gly*, *Ala*, *Val*, *Leu*, *Ile*, *Phe*, *Abu*, *Nva*, and *Nle* residues, and B is substituted with the same residues and *Pro* residue. The modeling data referenced from Nomizu et al.[5] are shown in Table 1.

* Corresponding author.
† Deceased.
⊗ Abstract published in *Advance ACS Abstracts,* March 1, 1996.

0095-2338/96/1636-0185$12.00/0

Table 1. Modeling Data

code	log $1/K_m$	log k_{cat}	log k_{cat}/K_m	code	log $1/K_m$	log k_{cat}	log k_{cat}/K_m
GlyAla	−0.354	0.049	2.695	IlePro	0.071	1.274	4.346
GlyVal	−0.231	−0.105	2.703	IlePhe	−0.188	0.158	2.971
GlyLeu	0.125	−0.796	2.328	IleAbu	−0.086	1.068	3.982
GlyIle	0.036	−0.337	2.704	IleNva	−0.228	1.061	3.833
GlyPro	−0.322	0.068	2.746	IleNle	0.585	0.407	3.992
GlyPhe	−0.260	−0.854	1.881	PheGly	−0.812	0.037	2.225
GlyAbu	0.059	−1.121	2.832	PheAla	−0.425	0.890	3.465
GlyNva	−0.220	0.041	2.882	PheVal	−0.207	0.511	3.303
GlyNle	0.149	−0.155	2.994	PheLeu	−0.344	0.310	2.965
AlaGly	−0.686	0.107	2.422	PheIle	−0.255	0.490	3.236
AlaAla	−0.225	0.936	3.711	PhePro	−0.107	1.000	3.893
AlaVal	−0.049	0.979	3.930	PhePhe	0.013	−0.409	2.605
AlaLeu	−0.193	0.515	3.322	PheAbu	0.053	0.556	3.610
AlaIle	−0.083	0.951	3.869	PheNva	−0.033	0.612	3.579
AlaPro	−0.207	1.436	4.230	PheNle	0.108	0.553	3.660
AlaPhe	−0.072	0.375	3.303	AbuGly	−0.906	0.565	2.658
AlaAbu	−0.114	1.033	3.920	AbuAla	−0.155	1.086	3.931
AlaNva	−0.061	1.045	3.985	AbuVal	0.215	0.619	3.834
AlaNle	0.638	0.622	4.265	AbuLeu	−0.037	0.471	3.435
ValGly	−0.493	0.467	2.974	AbuIle	0.222	0.653	3.876
ValAla	−0.215	1.185	3.970	AbuPro	−0.362	1.581	4.220
ValVal	0.041	0.876	3.920	AbuPhe	0.252	−0.076	3.179
ValLeu	−0.243	0.561	3.318	AbuAbu	0.060	1.045	4.107
ValIle	−0.170	0.880	3.710	AbuNva	−0.244	0.787	4.029
ValPro	0.033	1.476	4.509	AbuNle	0.538	0.589	4.124
ValPhe	0.018	0.210	3.225	NvaGly	−0.377	0.199	2.822
ValAbu	0.276	0.866	4.140	NvaAla	−0.362	1.060	3.699
ValNva	0.060	0.965	4.025	NvaVal	0.051	0.814	3.863
ValNle	0.228	0.864	4.090	NvaLeu	−0.246	0.471	3.226
LeuGly	−0.559	0.188	2.628	NvaIle	−0.228	0.860	3.631
LeuAla	−0.326	1.107	3.781	NvaPro	−0.143	1.380	4.238
LeuVal	−0.185	0.780	3.595	NvaPhe	0.222	−0.102	3.124
LeuLeu	−0.340	0.412	3.072	NvaAbu	0.131	0.925	4.061
LeuIle	−0.452	0.991	3.539	NvaNva	0.316	0.741	4.057
LeuPro	0.142	0.996	4.140	NvaNle	0.569	0.604	4.057
LeuPhe	−0.152	0.033	2.881	NleGly	−0.851	0.581	2.730
LeuAbu	−0.158	1.053	3.895	NleAla	−0.328	1.143	3.814
LeuNva	−0.045	0.925	3.880	NleVal	−0.188	0.840	3.653
LeuNle	0.071	0.956	4.041	NleLeu	−0.371	0.303	2.947
IleGly	−0.708	0.607	2.900	NleIle	−0.068	0.730	3.662
IleAla	−0.371	1.210	3.838	NlePro	−0.196	1.380	4.185
IleVal	−0.252	0.927	3.674	NlePhe	0.244	−0.086	3.158
IleLeu	−0.312	0.358	3.045	NleAbu	0.215	0.812	4.025
IleIle	−0.152	0.738	3.585	NleNva	−0.262	1.176	3.914
				NleNle	0.420	0.736	4.173

In this study, chemical features of synthetic substrates are described using *principal properties* (PPs) developed by S. Wold.[6] PPs is developed for a descriptor of amino acids, and three variables called z_1, z_2, and z_3 are contained. These parameters are determined by using the principal component analysis of 29 parameters for 20 natural amino acids. The first, second, and third principal components are corresponded to z_1, z_2, and z_3, respectively. Afterwards, this parameter is extended for six non-natural amino acids. PPs for natural and non-natural amino acids referenced from S. Wold et al.[6,7] are shown in Table 2.

2.2. Quadratic PLS Method. *Quadratic PLS* (QPLS) is one of the nonlinear modeling methods developed by S. Wold in 1989.[4] Hereafter, assume the matrix **X** and **Y** are as follows: **X** denotes the independent data matrix, and **Y** denotes the dependent data matrix. For example, **X** contains the information of the structure of the particular chemical substances, and **Y** contains the activity or property of the corresponding substances.

The forms of the QPLS model equations are given below

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1' + \mathbf{t}_2\mathbf{p}_2' + ... + \mathbf{t}_A\mathbf{p}_A' + \mathbf{E} \quad (1)$$

$$\mathbf{Y} = \mathbf{u}_1\mathbf{q}_1' + \mathbf{u}_2\mathbf{q}_2' + ... + \mathbf{u}_A\mathbf{q}_A' + \mathbf{F} \quad (2)$$

where $\mathbf{t}_a$ is the latent variable for the independent data matrix denoted by **X**, and $\mathbf{u}_a$ is the latent variable for dependent data matrix denoted by **Y**. $\mathbf{p}_a$ and $\mathbf{q}_a$ are defined to the loading vector corresponding to $\mathbf{t}_a$ and $\mathbf{u}_a$, respectively. **E** and **F** are the model residuals for **X** and **Y**, respectively.

The latent variables corresponding to *k*th component, denoted by $\mathbf{t}_k$ and $\mathbf{u}_k$, are calculated using formulas as follows.

$$\mathbf{t}_k = \{\mathbf{X} - (\mathbf{t}_1\mathbf{p}_1' + \mathbf{t}_2\mathbf{p}_2' + ... + \mathbf{t}_{k-1}\mathbf{p}_{k-1}')\}\mathbf{w}_k \quad (3)$$

$$\mathbf{u}_k = \{\mathbf{Y} - (\mathbf{u}_1\mathbf{q}_1' + \mathbf{u}_2\mathbf{q}_2' + ... + \mathbf{u}_{k-1}\mathbf{q}_{k-1}')\}\mathbf{c}_k \quad (4)$$

The nonlinear inner relation characterizes the QPLS modeling. The form of the inner relation equation is as follows.

$$u_{ja} = b_{0a} + b_{1a}t_{ja} + b_{2a}t_{ja}^2 + h \quad (5)$$

$$u_{ja} = r_{ja} = b_{0a} + b_{1a}t_{ja} + b_{2a}t_{ja}^2 \quad (6)$$

In this study, QPLS parameters, denoted by $\mathbf{w}_a$, $\mathbf{t}_a$, $\mathbf{u}_a$, etc. are determined to minimize the sum of squares for residuals of inner relation.

After the QPLS modeling, we express the model equations using the MLR-like form. The model equations are expressed by nonlinear polynomial containing the quadratic form of the vector of independent variable

$$y_{jl} = d_{0l} + \mathbf{x}^j\mathbf{d}_l + \mathbf{x}^j\mathbf{D}_l\mathbf{x}^{j\prime}$$

$$d_{0l} = b_{01}q_{l1} + b_{02}q_{l2} + ... + b_{0A}q_{lA} \quad (7)$$

$$\mathbf{d}_l = b_{11}\mathbf{w}_1{}^*q_{l1} + b_{12}\mathbf{w}_2{}^*q_{l2} + ... + b_{1A}\mathbf{w}_A{}^*q_{lA} \quad (8)$$

$$\mathbf{D}_l = b_{21}(\mathbf{w}_1{}^*\mathbf{w}_1{}^{*\prime})q_{l1} + b_{22}(\mathbf{w}_2{}^*\mathbf{w}_2{}^{*\prime})q_{l2} + ... + b_{2A}(\mathbf{w}_A{}^*\mathbf{w}_A{}^{*\prime})q_{lA} \quad (9)$$

where $\mathbf{x}^j$ denotes the independent data vector corresponding to the jth sample, and $\mathbf{w}_k{}^*$ is defined as follows.

$$\mathbf{w}_k{}^* = \{\mathbf{I} - (\mathbf{w}_1\mathbf{p}_1{}' + \mathbf{w}_2\mathbf{p}_2{}' + ... + \mathbf{w}_{k-1}\mathbf{p}_{k-1}{}')\}\mathbf{w}_k \quad (10)$$

2.3. Modeling. The model equations are built using the QPLS method, and the number of components in the model equation is determined using a cross-validation technique called leave-one-out. For comparison purposes, linear PLS methods are used for modeling, also.

3. RESULTS AND DISCUSSION

By using the PLS method, linear model equations were obtained. The optimum number of PLS components for the model equations of log $1/K_m$, log k_{cat}, and log k_{cat}/K_m is 1, 5, and 5, respectively. The model equations are shown below

$$\log 1/K_m = -0.1647 + 0.0111a_1{}^A - 0.0083z_2{}^A - 0.0507z_1{}^B + 0.0335z_2{}^B - 0.0076z_3{}^B \quad (11)$$

$$n = 89 \quad A = 1 \quad R^2 = 0.285 \quad Q^2 = 0.248 \quad s = 0.254$$

$$\log k_{cat} = 2.2294 + 0.2152z_1{}^A + 0.3552z_2{}^A - 0.7044z_3{}^A + 0.2041z_1{}^B + 0.2128z_2{}^B - 0.2220z_3{}^B \quad (12)$$

$$n = 89 \quad A = 5 \quad R^2 = 0.465 \quad Q^2 = 0.372 \quad s = 0.392$$

$$\log k_{cat}/K_m = 5.2973 + 0.0573z_1{}^A + 0.1103z_2{}^A - 0.5021z_3{}^A + 0.1237z_1{}^B + 0.2013z_2{}^B - 0.1931z_3{}^B \quad (13)$$

$$n = 89 \quad A = 5 \quad R^2 = 0.330 \quad Q^2 = 0.223 \quad s = 0.483$$

where R^2 denotes the explained variance, and Q^2 denotes predictive explained variance for the model equation. These results show that the poor model equations were obtained when one applies the linear modeling method for this data set. It can be thought that the nonlinear behavior in this data set is responsible for the badness of model equations.

By using the QPLS method, nonlinear model equations were obtained. The optimum number of QPLS components for model equations of log $1/K_m$, log k_{cat}, and log k_{cat}/K_m is 2, 4, and 3, respectively. The model equations in MLR-like form are shown below.

$$\log 1/K_m = -0.3019 + 0.0074z_1{}^A + 0.0182z_2{}^A - 0.1054z_3{}^A - 0.3260z_1{}^B - 0.0887z_2{}^B - 0.0094z_3{}^B - 0.0135(z_1{}^A)^2 - 0.0002(z_2{}^A)^2 - 0.0062(z_3{}^A)^2 - 0.0573(z_1{}^B)^2 - 0.0056(z_2{}^B)^2 - 0.0267(z_2{}^B)^2 - 0.0020z_1{}^Az_2{}^A - 0.0043z_1{}^Az_3{}^A + 0.0145z_1{}^Az_1{}^B + 0.0097z_1{}^Az_2{}^B - 0.0378z_1{}^Az_3{}^B + 0.0016z_2{}^Az_3{}^A + 0.0069z_2{}^Az_1{}^B + 0.0023z_2{}^Az_2{}^B - 0.0024z_2{}^Az_3{}^B - 0.0331z_3{}^Az_1{}^B - 0.0079z_3{}^Az_2{}^B - 0.0082z_3{}^Az_3{}^B - 0.0338z_1{}^Bz_2{}^B + 0.0137z_1{}^Bz_3{}^B + 0.0119z_2{}^Bz_3{}^B \quad (14)$$

$$n = 89 \quad A = 2 \quad R^2 = 0.542 \quad Q^2 = 0.410 \quad s = 0.205$$

$$\log k_{cat} = 0.3170 - 0.8816z_1{}^A - 0.8370z_2{}^A + 2.2748z_3{}^A + 0.3629z_1{}^B + 0.2006z_2{}^B + 0.3435z_3{}^B - 0.1632(z_1{}^A)^2 - 0.2350(z_2{}^A)^2 - 1.2638(z_3{}^A)^2 + 0.0504(z_1{}^B)^2 - 0.0034(z_2{}^B)^2 - 0.1609(z_3{}^B)^2 - 0.3777z_1{}^Az_2{}^A + 0.8703z_1{}^Az_3{}^A - 0.0117z_1{}^Az_1{}^B + 0.0284z_1{}^Az_2{}^B - 0.0131z_1{}^Az_3{}^B + 1.0793z_2{}^Az_3{}^A + 0.0202z_2{}^Az_1{}^B + 0.1018z_2{}^Az_2{}^B - 0.1058z_2{}^Az_3{}^B - 0.0143z_3{}^Az_1{}^B - 0.1017z_3{}^Az_2{}^B + 0.1163z_3{}^Az_3{}^B + 0.1225z_1{}^Bz_2{}^B - 0.0188z_1{}^Bz_3{}^B + 0.1801z_2{}^Bz_3{}^B \quad (15)$$

$$n = 89 \quad A = 4 \quad R^2 = 0.842 \quad Q^2 = 0.748 \quad s = 0.212$$

$$\log k_{cat}/K_m = 1.4645 - 0.3331z_1{}^A - 1.3024z_2{}^A + 1.3469z_3{}^A - 0.4608z_1{}^B - 1.0500z_2{}^B + 1.4025z_3{}^B - 0.0111(z_1{}^A)^2 - 0.1698(z_2{}^A)^2 - 0.1929(z_3{}^A)^2 - 0.0286(z_1{}^B)^2 - 0.1342(z_2{}^B)^2 - 0.1638(z_3{}^B)^2 - 0.0862z_1{}^Az_2{}^A + 0.0903z_1{}^Az_3{}^A - 0.0293z_1{}^Az_1{}^B - 0.0726z_1{}^Az_2{}^B + 0.0982z_1{}^Az_3{}^B + 0.3425z_2{}^Az_3{}^A - 0.1271z_2{}^Az_1{}^B - 0.2549z_2{}^Az_2{}^B + 0.3353z_2{}^Az_3{}^B + 0.1007z_3{}^Az_1{}^B + 0.3324z_3{}^Az_2{}^B - 0.4515z_3{}^Az_3{}^B - 0.0682z_1{}^Bz_2{}^B + 0.1025z_1{}^Bz_3{}^B + 0.3202z_2{}^Bz_3{}^B \quad (16)$$

$$n = 89 \quad A = 3 \quad R^2 = 0.754 \quad Q^2 = 0.629 \quad s = 0.290$$

These equations show that the QPLS model equations are reliable compared with the PLS model equations. The Q^2 for log $1/K_m$, log k_{cat}, and log k_{cat}/K_m are improved compared with the PLS model equations, also. Figure 1−3 show the results of calculation using eqs 14−16, respectively.

In Figure 1, it is found that there are some samples with large errors between observed and calculated values of log $1/K_m$. These samples contain the Nle residue on position B. It can be considered that these samples behave as outliers in the QPLS modeling. The two-component QPLS model equation was obtained when we built the model equation for the data set without samples containing the Nle residue on position B. The R^2 and Q^2 for this model equation are 0.6367 and 0.4916, respectively.

In this study, we investigate the behavior of the linear and nonlinear terms in model eqs 14−16.

Table 2. *Principal Properties* (PPs) for Natural and Non-Natural Amino Acids

no.	code	z_1	z_2	z_3
1	Ala	0.07	−1.73	0.09
2	Val	−2.69	−2.53	−1.29
3	Leu	−4.19	−1.03	−0.98
4	Ile	−4.44	−1.68	−1.03
5	Pro	−1.22	0.88	2.23
6	Phe	−4.92	1.30	0.45
7	Trp	−4.75	3.65	0.85
8	Met	−2.49	−0.27	−0.41
9	Lys	2.84	1.41	−3.14
10	Arg	2.88	2.52	−3.44
11	His	2.41	1.74	1.11
12	Gly	2.23	−5.36	0.30
13	Ser	1.96	−1.63	0.57
14	Thr	0.92	−2.09	−1.40
15	Cys	0.71	−0.97	4.13
16	Tyr	−1.39	2.32	0.01
17	Asn	3.22	1.45	0.84
18	Gln	2.18	0.53	−1.14
19	Asp	3.64	1.13	2.36
20	Glu	3.08	0.39	−0.07
21	Orn	4.11	0.51	−1.61
22	Cit	0.77	1.28	−1.58
23	Abu	−0.61	−2.51	−0.01
24	Nle	−3.62	−1.36	−0.88
25	Nva	−2.50	−1.92	−0.35
26	Omt	−1.43	−1.80	−1.78

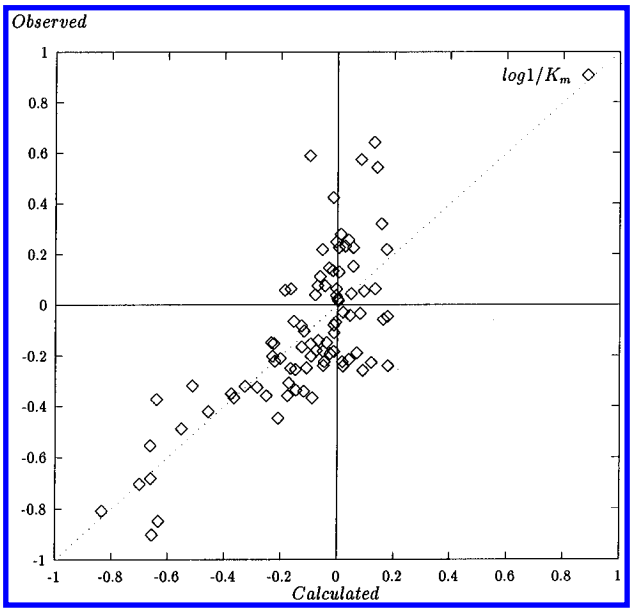

Figure 1. Results of QPLS modeling, plot of observed against calculated values of log $1/K_m$.

First, linear and squared terms of the independent variable in model equations are treated. For this purpose, we define the variable C_A and C_a as follows.

$$C_A = b_{1A}z_1^A + b_{2A}z_2^A + b_{3A}z_3^A + b_{11A}(z_1^A)^2 + b_{22A}(z_2^A)^2 + b_{33A}(z_3^A)^2$$

$$C_B = b_{1B}z_1^B + b_{2B}z_2^B + b_{3B}z_3^B + b_{11B}(z_1^B)^2 + b_{22B}(z_2^B)^2 + b_{33B}(z_3^B)^2$$

Variable C_A represents the total amount of the linear and squared term which is related to the amino acid residue on position A. Similarly, variable C_B represents the total amount of the linear and squared term which is related to the amino acid residue on position B. Behavior of the variables C_A and C_B corresponding to the model eqs 14−16 for each amino acid residue are regarded.

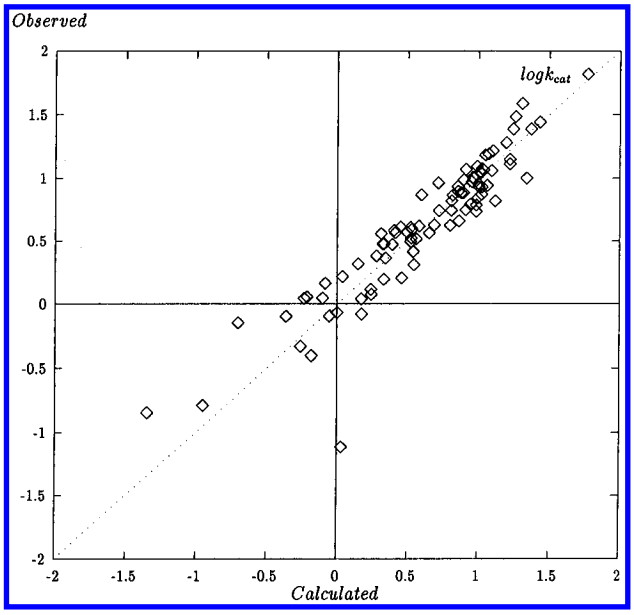

Figure 2. Results of QPLS modeling, plot of observed against calculated values of log k_{cat}.

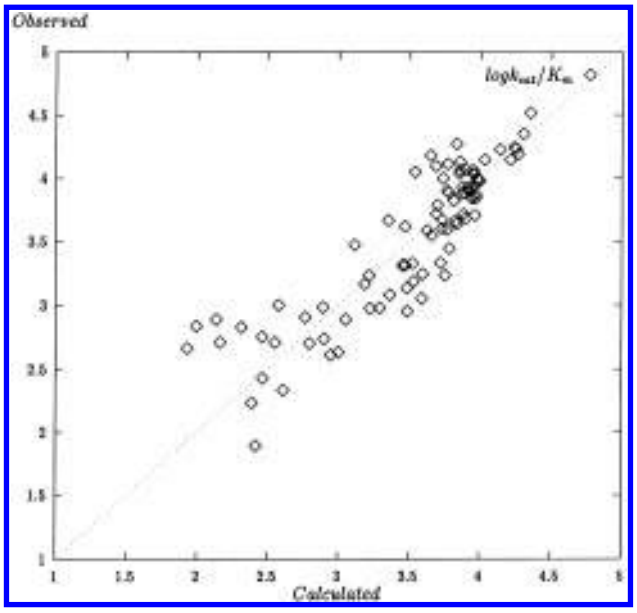

Figure 3. Results of QPLS modeling, plot of observed against calculated values of log k_{cat}/K_m.

Next, relationships between the variety of amino acid residues on A, B and the tendency of the calculated value of cross-terms in model equations are investigated. For this purpose, the total amount of cross-terms in the model equation denoted by C_{AB} is defined as follows.

$$C_{AB} = b_{1A1B}z_1^A z_1^B + b_{1A2B}z_1^A z_2^B + ... + b_{3A3B}z_3^A z_3^B$$

According to the calculation of C_{AB}, it becomes clear that the amino acid residue on position B is mainly affected by cross-terms in the model equation of log $1/K_m$, and position A is mainly affected by cross-terms in the model equations of log k_{cat} and log k_{cat}/K_m. The results of C_{AB} calculation for log $1/K_m$, log k_{cat}, and log k_{cat}/K_m are regarded.

From these results, the following becomes clear. (1) For log $1/K_m$, amino acid residues on position B are influenced

greatly compared to the residues on position A. So C_B is greatly influenced compared to the coefficient C_A, and C_{AB} is influenced by the kind of amino acid residue on position B. Furthermore, Gly, Pro, and Phe residues on position B are significant compared to the other residues on position B. It can be considered that the size of side chain of the amino acid residue at position B is reflected in log $1/K_m$. (2) For log k_{cat}, the kind of amino acid residue on position A is influenced greatly compared to the residue on position B. Ala, Abu, and Nva residues on position A are not significant compared with the other residues on position A. (3) The amino acid residues on positions A and B influences similar effects to the model equation for log k_{cat}/K_m, but Gly, Ala, Abu, and Nva residues on position A are more effective than the other residues on position A. It can be considered that log k_{cat}/K_m is related to the z_2 value of the amino acid residue on position A.

4. CONCLUSION

In this study, data set of log $1/K_m$, log k_{cat}, and log k_{cat}/K_m for synthetic substrates of elastase enzyme are treated using the QPLS modeling method. By this modeling, we obtain the nonlinear model equations for log $1/K_m$, log k_{cat}, and log k_{cat}/K_m with the correlation coefficients 0.736, 0.918, and 0.868, respectively. Also, the predictive correlation coefficients for these model equations are 0.640, 0.865, and 0.793, respectively. These results calculated by QPLS modeling method are favorable compared with the result of calculation by PLS method.

We investigate the QPLS model equations after the transformation of the QPLS model equations into the MLR-like forms. From this investigation, it becomes clear that the z_2 value of the amino acid residues on position A and the size of the side chain for amino acid residues on position B are related to the properties of the synthetic substrates.

ACKNOWLEDGMENT

The sudden loss of Professor Miyashita grives all of us, and it is to his memory that we dedicate this dissertation. We wish to express our sincerest gratitude for his utmost kindness while he was still living. To his bereaved family we also present our deepest sympathies and condolences while reminiscing their past unwavering hospitality. May Professor Miyashita rest in peace.

REFERENCES AND NOTES

(1) Geladi, P.; Kowalski, B. R. Partial Least-Squares Regression: A Tutorial. *Anal. Chim. Acta* **1986**, *185*, 1−17.
(2) Höskuldsson, A. PLS Regression Methods. *J. Chemometrics* **1988**, *2*, 211−228.
(3) Helland, I. S. On the Structure of Partial Least Squares Regression. *Commun. Statist.-Simula.* **1988**, *17(2)*, 581−607.
(4) Wold, S.; Wold, N. K.; Skagerberg, B. Nonlinear PLS Modelling. *Chemometrics Intell. Lab. Syst.* **1989**, *7*, 53−65.
(5) Nomizu, M.; Iwaki, T.; Yamashita, T.; Inagaki, Y.; Asano, K.; Akamatsu, M.; Fujita, T. Quantitative structure-activity relationship-(QSAR) study of elastase substrates and inhibitors. *Int. J. Peptide Protein Res.* **1993**, *42*, 216−226.
(6) Hellberg, S.; Sjöström, M.; Skagerberg, B.; Wold, S. Peptide Quantitative Structure−Activity Relationships, a Multivariate Approach. *J. Med. Chem.* **1987**, *30*, 1126−1135.
(7) Wold, S.; Eriksson, L.; Hellberg, S.; Jonsson, J.; Sjöström, M.; Skagerberg, B.; Wikström, C. Principal property values for six non-natural amino acids and their application to a structure−activity relationships for oxytocin peptide analogues. *Can. J. Chem.* **1987**, *65*, 1814−1820.

CI9501103