Electronic Supplementary Material (ESI) for Analytical Methods

Supporting Information

Rapid Determination of Enantiomeric Excess of Protected Amino Acids by Catalytic Amounts of Chiral Reagent

Xiangou Zhu [a], Jun Jiang[b], Xinxiang Lei[b], Xiaojing Chen*[a]

[a] College of Physics and Electronic Engineering Information,

[b] College of Chemistry and Materials Engineering and Wenzhou University, Chashan University Town, Wenzhou, Zhejiang Province 325035, People's Republic of China.

* Corresponding author: chenxj10@yahoo.cn (Xiaojing Chen), Fax: +86 577 86689027 Tel: +86 577 86689027

Supplementary Figure

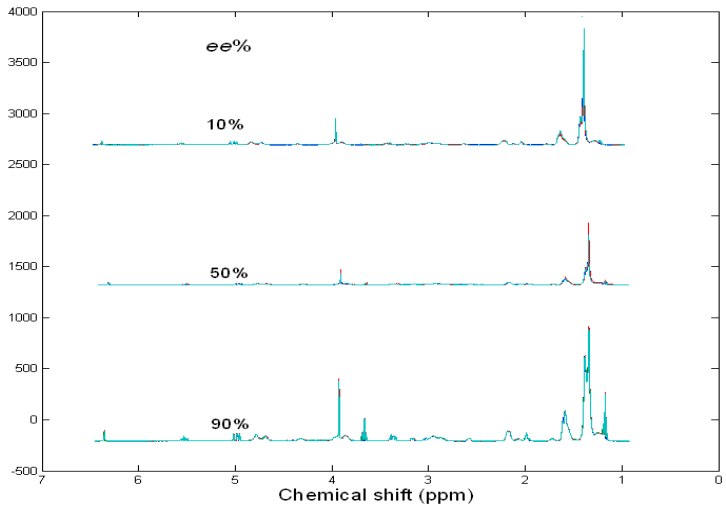

Figure S1 Original NMR spectra of Boc-protected piperidine-2-carboxylic at different ee values (10%, 50%, and 90%) with quinine.

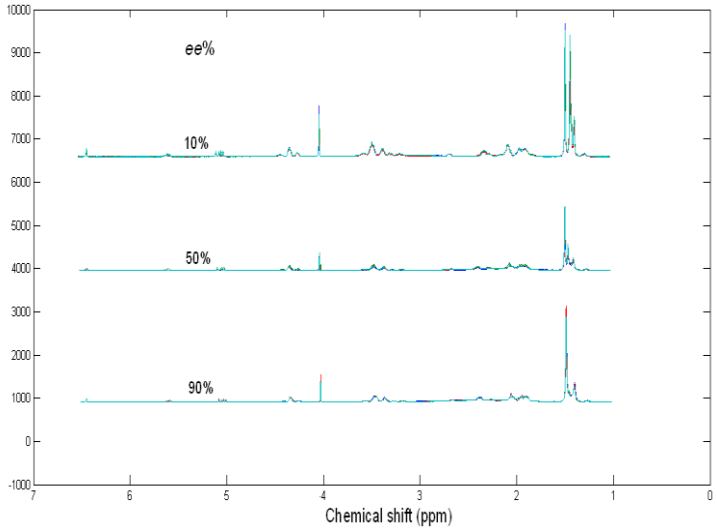

Figure S2. Original NMR spectra of Boc-protected proline at different *ee* values (10%, 50%, and 90%) with quinine.

Table 1 LS-SVM model analysis of unknown solutions (ee% expressed as %R in the data) of chiral Boc-protected piperidine-2-carboxylic with quinine as a chiral auxiliary

analytes	**given**[a]	**LS-SVM**[b]	**error**
1	40	47.91	7.91
2	40	34.08	5.92
3	40	40.39	0.39
4	40	50.68	10.68
5	70	73.20	3.20
6	70	75.36	5.36
7	70	70.92	0.92
8	70	74.19	0.42
	Average	error	4.35

[a]Enantomeric compositions of the solution prepared from L- and racemic amino acids.

[b]Prediction by the LS-SVM model using the spectra of chemical shift range from 1.3471 to 1.5202.

Table 2. LS-SVM model analysis of unknown solutions (*ee*% expressed as %L in the data) of chiral Boc-protected proline with quinine as a chiral auxiliary

Analytes	**Given**[a]	**LS-SVM**[b]	**Error**
1	40	42.36	2.36
2	40	44.44	4.44
3	40	35.99	4.01
4	40	40.27	0.27
5	70	76.33	6.33
6	70	73.43	3.43
7	70	68.12	1.88
8	70	69.77	0.23
	Average error		2.86

[a]Enantomeric compositions of the solution prepared from L- and racemic amino acids.

[b]Prediction by the LS-SVM model using the spectra of chemical shift range from 1.3471 to 1.5202.

Experimental Procedures

Enantiomerically pure quinidine (**1** Optical purity 99%), L-Alanine (Optical purity 99%), L-proline(Optical purity 99%), L-Piperidine-2-carboxylic acid (Optical purity 99%) were obtained from shanghai darui finechemical Co., Ltd. The protection of the amino acid was according with reported methods. Typically, a concentrated stock solution (6.67mM) of the 16:84 mixture of quinine and chiral protected amino acids was prepared by accurately weighing out the dried complex and dissolving it in an accurately measured volume of $CDCl_3$ (0.5 mL). The NMR spectra were recorded after several minutes of thermal equilibration time. NMR spectra were recorded on a Bruker Avance 500 MHz spectrometer. All spectra were recorded using 16 scans at 298K. A full spectrum of each sample was recorded referenced to TMS. An exponential window function with a line-broadening factor of 1Hz was applied to the FID before Fourier transformation. The H^1-NMR spectra were phased and baseline-corrected using Topspin 2.1(Bruker) and were automatically reduced by using the AMIX (Bruker GmbH, Germany) software package to continuous integral segments of equal width of 0.004ppm corresponding to thechemical shift range $_1H$, δ 6.9–8.8 after removing the solvent resonance region (δ 7.2-7.3) prior toprincipal component analysis (PCA) for fingerprinting.

NMR Forced Racemization Experiments. The different level of enantiomeric excess (*ee*) of amino acids (10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%) were prepared from accurately weighing and combining of certain amount of (L)-amino acids and corresponding racemic ones, and sample volume constant (0.5 mL $CDCl_3$), the mol ratio of quinine and amino acid is 16:84. The experiment of each sample with certain *ee* was repeated 4 times to ensue reproducibility.

Principal Component Analysis (PCA):

Principal Component Analysis (PCA) was carried out on MATLAB 2007b (The MathWorks, Inc.Natick, MA). Data were visualized in the form of the PC scores plots. PCA allows multivariate data to be represented in three-dimensional space. The first three principal components are invoked to visualize the objects in two-dimension space. The first principal component (PC1) is detected according to the direction of the maximum variance. The second principal component (PC2) is orthogonal to PC1 and has the maximum possible variance. The

third principal component (PC3) is orthogonal to PC1 and PC2, and has the maximum possible variance.

Support vector machine (SVM):

The foundation for SVMs was proposed in 1992, and has been used extensively for classification problems in many areas of chemistry（**Boser, B.; Guyon, I.; Vapnik, V. Proc. 5th Annu. ACM Workshop Comput. Learn. Theory 1992, 144–152**）. The original theory of SVM was a valuable tool for solving pattern recognition and classification problems, and then SVM has been extended to develop nonlinear regression models capable of quantitative prediction. The basic idea of SVM is to map the original data set X into a higher dimensional feature space via non-linear mapping function and then perform linear regression in the hyperspace, for example, in Fig.S3 (**origin from Visualization and Recovery of the (Bio) chemical Interesting Variables in Data Analysis with Support Vector Machine Classification.** ***Anal. Chem.*** **2010, 82, 7000–7007**), it is obvious that we are not able to separate the classes by a linear model. However, if we project this data in a higher dimensional space by using a mapping function, it is easy to separate the classes by a linear model. So, the SVM method can serve as particularly powerful computational tool for correlating spectral data with known compositional.

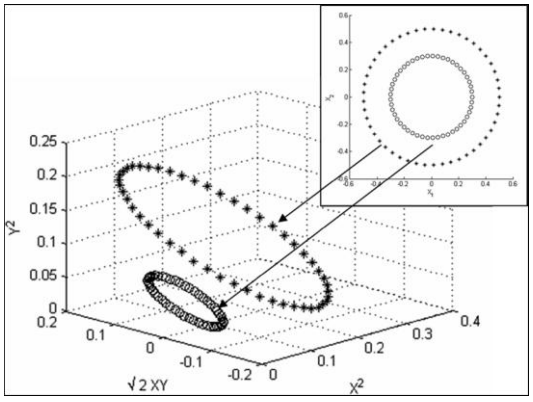

Figure S3. The synthetic benchmark data set. A three-dimensional space (b) makes a two dimensional data set which contains two nonlinearly separable classes (outer and inner circles) linearly separable using the nonlinear mapping function (kernel function).

Least square-support vector machine (LS-SVM):

LS-SVM is a modified version of SVM proposed by Suykens (**J. A. K. Suykens, T. V. Gestel, J. D. Brabanter, B. D. Moor and J. Vandewalle,** ***Least-Squares Support Vector Machines,*** **World Scientific, Singapore, 2002**）and has been applied to the chemometric field including multivariate

calibration and classification problems. Compared to SVM, LS-SVM has an important advantage of only requiring solving a set of linear equations instead of the classical quadratic programming problem, significantly reducing the computational complexity (**Ren S., Gao L., Improvement of the prediction ability of multivariate calibration by a method based on the combination of data fusion and least squares support vector machines, *Analyst,* 2011, 136, 1252**). The techniques implemented in the LS-SVM regression have the goal of analyzing the unknown *ee* values. The LS-SVM regression program was performed with MATLAB 2007b (The MathWorks, Inc.Natick, MA) using LS-SVMlab1.5aw Toolbox. LS-SVM was used as protocol for prediction, and whole spectra was employed as input set for LS-SVM to built regression model. The *ee* training sets at seven different values were used as training data, the *ee* prediction sets at two values, the output values from the LS-SVM mode are compared to the true values.