

Chen Jiang · Daliang Li · Jiwu Wen · Tianpa You

## QSAR study of the enantiomeric excess in asymmetric catalytic reactions with topological indices and an artificial neural network

Received: 20 January 2006 / Accepted: 22 March 2006 / Published online: 15 June 2006  
© Springer-Verlag 2006

**Abstract** The relationships between the enantiomer excess of product in catalytic asymmetric reactions and the structures of the catalysts or reagents in several asymmetric reactions were studied using a backpropagation (BP) neural network with topological indices and their chiral expansions. The trained network can be used to screen new asymmetric catalysts, estimate catalytic effects, design reaction environments, and prove or improve the proposed reaction mechanism.

**Keywords** Asymmetric catalytic reaction · Enantiomeric excess · Topological indices · Artificial neural network · Multiple regression

### Introduction

Chirality or asymmetry is a common and important phenomenon in nature. Biological activity and many other important properties are often very different between a pair of enantiomers. Thus, the great need to synthesize enantiomerically pure chiral compounds leads more and more scientists to focus on asymmetric synthesis and its applications in chemistry, biomedicine, and agriculture [1, 2]. Catalytic asymmetric reactions, in which one chiral catalyst molecule may produce millions of chiral product molecules, lie at the center of this stage [3, 4]. Dozens of catalytic asymmetric reactions were developed [5, 6] and varieties of asymmetric catalysts were synthesized, including hundreds of excellent ones that lead to enantiomer excesses (e.e.%) of more than 90%.

However, designing an asymmetric catalyst still remains an experimental job. In most cases, the details of the mechanism of catalysis are unclear and intermediates of the reaction are intangible. Researchers are often unable to pursue explicit instructions to design a new catalyst because the established principles for such tasks are few except for experience and inspiration. In previous research, we have developed a method to inspect the relationships between the structure of the catalysts or the reagents and the e.e.% of products in some types of catalytic asymmetric reactions by topological molecule descriptors and multiple linear regression (MLR) [7]. This approach was shown to be useful for predicting the effect of novel asymmetric catalysts and revealing mechanisms of reactions. We have applied it in our catalyst-screening work [8]. However, the linear models are unable to fit the complex or nonlinear factors of the reactions, such as temperature.

Thus, the calculations were improved with artificial neural networks (ANN) in this paper. As a computational model for simulating the structure of a biological neural system, they can be adapted to both linear and nonlinear models [9, 10]. ANN can also automatically seek out the exceptional cases that may cause normal MLR calculations to fail. Here, some catalytic asymmetric reaction sets were used to train with neural networks, and the knowledge acquired by the networks was applied to new situations. Compared with MLR, ANN may give better precision and broaden the scope of predictions.

### Method of calculation and experimental data

#### Topological molecular descriptor

The topological molecular descriptors  $A_{x_1}$ ,  $A_{x_2}$ , and  $A_{x_3}$  suggested by Yao and Xu [11] for organic molecules with heteroatoms were used in our study to measure the structure of catalyst or reagent. These three descriptors are derived from the path matrix of a hydrogen-suppressed

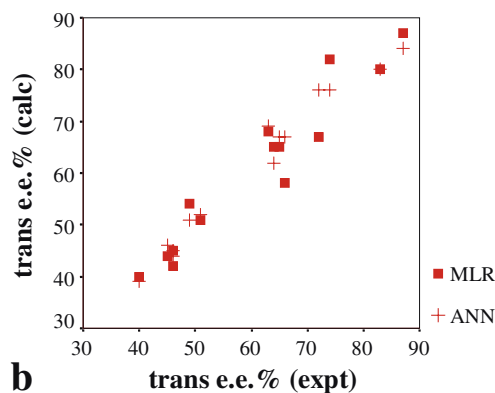
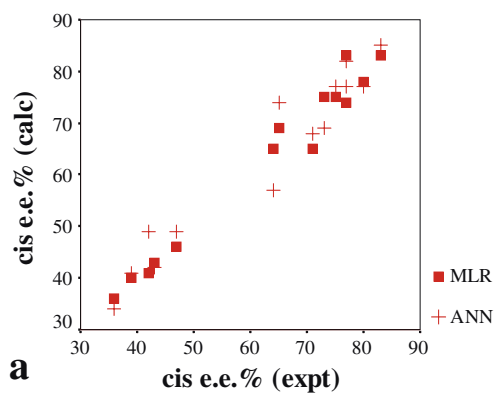
C. Jiang · D. Li · J. Wen · T. You (✉)  
Department of Chemistry,  
University of Science and Technology of China,  
Hefei, 230026, People's Republic of China  
e-mail: ytp@ustc.edu.cn  
Tel.: +86-551-3606944

graph of a molecule. We have expanded them to asymmetric compounds,  $A_{x_{1-3}}^*$  as shown in the following [7].

$$G_1 = (g_{ij}^1) = \begin{cases} \left\{ \begin{array}{ll} \text{square root of vertex degrees} & (i = 1) \\ \text{square root of van der Waals radii of atoms} & (i = 2) \end{array} \right. & (1 \leq i \leq 2) \\ \left\{ \begin{array}{ll} 1 & \text{if there is a path of length 1 between vertices } i - 2 \text{ and } j, \text{ the atoms } i, j \text{ are nonchiral} \\ 1 \pm \delta & \text{if there is a path of length 1 between vertices } i - 2 \text{ and } j, \text{ the atom is in } S \text{ or } R \text{ conformation} \\ 0 & \text{if there is no path of length 1 between vertices } i - 2 \text{ and } j \end{array} \right. & (i > 2) \end{cases}$$

$$G_2 = (g_{ij}^2) = \begin{cases} \left\{ \begin{array}{ll} \text{square root of vertex degrees} & (i = 1) \\ \text{square root of van der Waals radii of atoms} & (i = 2) \end{array} \right. & (1 \leq i \leq 2) \\ \left\{ \begin{array}{ll} 2 & \text{if there is a path of length 1 between vertices } i - 2 \text{ and } j, \text{ the atoms } i, j \text{ are nonchiral} \\ 2 \pm \delta & \text{if there is a path of length 1 between vertices } i - 2 \text{ and } j, \text{ the atom is in } S \text{ or } R \text{ conformation} \\ 0 & \text{if there is no path of length 1 between vertices } i - 2 \text{ and } j \end{array} \right. & (i > 2) \end{cases}$$

$$G_3 = (g_{ij}^3) = \begin{cases} \left\{ \begin{array}{ll} \text{square root of vertex degrees} & (i = 1) \\ \text{square root of van der Waals radii of atoms} & (i = 2) \end{array} \right. & (1 \leq i \leq 2) \\ \left\{ \begin{array}{ll} 3 & \text{if there is a path of length 1 between vertices } i - 2 \text{ and } j, \text{ the atoms } i, j \text{ are nonchiral} \\ 3 \pm \delta & \text{if there is a path of length 1 between vertices } i - 2 \text{ and } j, \text{ the atom is in } S \text{ or } R \text{ conformation} \\ 0 & \text{if there is no path of length 1 between vertices } i - 2 \text{ and } j \end{array} \right. & (i > 2) \end{cases}$$



**Fig. 1** Plot of experimental vs ANN calculated e.e.%(cis) and e.e.%(trans) of the asymmetric cyclopropanation reaction catalyzed by imine-amine ligands

where  $\delta$  is an adjustable parameter between 0 and 1 to identify a pair of enantiomers. The three topological indices  $A_{x_1}^*$ ,  $A_{x_2}^*$ , and  $A_{x_3}^*$  are defined as:

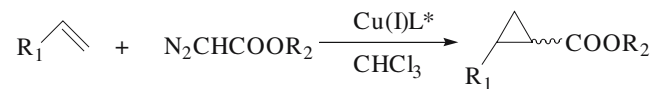
$$Z_1 = G'_1 \times G_1, Z_2 = G'_2 \times G_2, Z_3 = G'_3 \times G_3$$

$$A_{x_1}^* = \lambda_{\max_1}/2, A_{x_2}^* = \lambda_{\max_2}/2, A_{x_3}^* = \lambda_{\max_3}/2$$

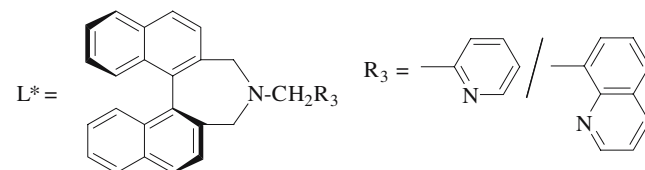
where  $\lambda_{\max_1} \sim \lambda_{\max_3}$  are the largest Eigenvalues of matrices  $Z_1 \sim Z_3$ .

Indicator variable [12, 13]

An indicator variable was also used to identify simple asymmetric substituents in some molecules. A variable representing chirality was introduced to the network



**Scheme 1** Asymmetric cyclopropanation reaction catalyzed by imine-amine ligands



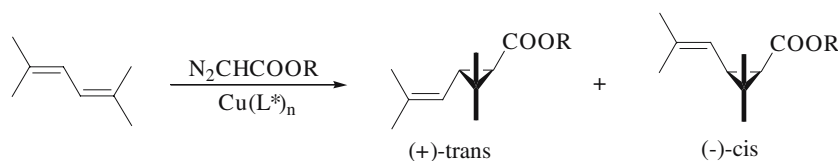
**Scheme 2** Catalyst ligands for the reaction shown in Scheme 1

**Table 1** Substituent pattern and the observed and calculated e.e.% of the asymmetric cyclopropanation reaction catalyzed by imine\_amine ligands

Entry	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	e.e.%(cis)			e.e.%(trans)		
				(expt)	(ANN)	(MLR)	(expt)	(ANN)	(MLR)
1	Ph	Et	pyridinyl	36	36	34	40	40	39
2	Ph	Et	quinolinyl	43	43	42	45	44	46
3	Ph	DCM <sup>a</sup>	pyridinyl	71	65	68	63	68	69
4	Ph	DCM	quinolinyl	80	78	77	83	80	76
5	Ph	L-menthyl	pyridinyl	77	83	77	74	87	76
6	Ph	L-menthyl	quinolinyl	83	83	85	87	87	83
7	Ph	D-menthyl	pyridinyl	39	40	41	46	45	45
8	Ph	D-menthyl	quinolinyl	47	46	49	51	51	52
9 <sup>b</sup>	4-MePh	L-menthyl	pyridinyl	65	69	74	66	58	67
10	4-MePh	L-menthyl	quinolinyl	77	74	82	72	62	76
11	4-MeOPh	L-menthyl	pyridinyl	73	75	69	64	65	59
12	4-MeOPh	L-menthyl	quinolinyl	75	75	77	65	65	67
13	4-CIPh	L-menthyl	pyridinyl	42	41	49	46	42	44
14	4-CIPh	L-menthyl	quinolinyl	64	65	57	49	54	51
Squared correlation coefficient					0.97	0.95		0.93	0.95

<sup>a</sup>Dicyclohexylmethyl diazoacetate<sup>b</sup>Validation case for MLR

**Scheme 3** Asymmetric cyclopropanation reaction catalyzed by Schiff-base ligands



training. If the substituent is chiral, this variable is set to +1 or -1; otherwise, the value is 0.

#### Quantitative structure activity relationship calculation

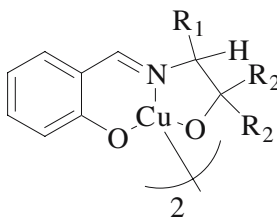
The ANN calculations were carried out in ObjectANN, a generalized neural network program developed by us. Feed-forward networks with three layers were trained with the backpropagation (BP) of errors algorithm [14, 15]. In each calculation, entries were divided into several groups and cross-validation was carried out.

In all multiple regression equations,  $n$  is the number of cases used in analysis,  $r^2$  is the squared correlation coefficient,  $s$  is the standard error of the estimates, and  $F$  is the Fisher significance ratio. In each MLR model, some cases were randomly selected from the experimental data for an internal validation.

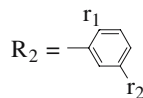
#### Experimental data

All experimental data were obtained from published papers.

**Scheme 4** Catalyst for the asymmetric cyclopropanation reaction catalyzed by Schiff-base ligands



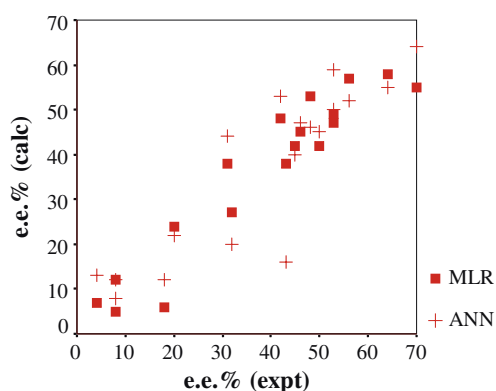
R<sub>1</sub> = methyl, benzyl, isopropyl, isobutyl



(r<sub>1</sub>, r<sub>2</sub>) = (H, H), (OMe, H), (OBu, H), (OBu, Pr), (OBu, tBu), (O-Octyl, OBu)

**Table 2** Substituent pattern and the observed and calculated e.e.% of the asymmetric cyclopropanation reaction catalyzed by Schiff-base ligands

Entry	R <sub>1</sub>	R <sub>2</sub>	(R <sub>2</sub> )		e.e.% (expt)	(ANN)	(MLR)
			r <sub>1</sub>	r <sub>2</sub>			
1	Me		H	H	4	7	13
2	Me		OMe	H	20	24	22
3	Me		OBu	H	50	42	45
4	Me		OBu	Me	48	53	46
5 <sup>a</sup>	Me		OBu	Pr	56	57	52
6	Me		OBu	tBu	64	58	55
7	Me		O-Octyl	tBu	70	55	64
8	CH <sub>2</sub> Ph		H	H	8	12	8
9	CH <sub>2</sub> Ph		OMe	H	43	38	16
10	CH <sub>2</sub> Ph		OBu	H	45	42	40
11	CH <sub>2</sub> Ph		OBu	Pr	46	45	47
12	CH <sub>2</sub> Ph		OBu	tBu	53	47	50
13	CH <sub>2</sub> Ph		O-Octyl	tBu	53	49	59
14	CHMe <sub>2</sub>		H	H	8	5	12
15	CHMe <sub>2</sub>		OBu	H	31	38	44
16	CHMe <sub>2</sub>		OBu	tBu	42	48	53
17 <sup>a</sup>	CHMeEt		H	H	18	6	12
18	CHMeEt		OMe	H	32	27	20
Squared correlation coefficient						0.90	0.85

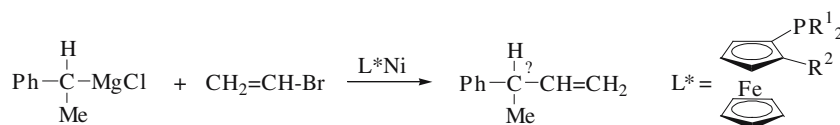
<sup>a</sup>Validation cases for MLR**Fig. 2** Plot of experimental vs ANN calculated e.e.% of the asymmetric cyclopropanation reaction catalyzed by Schiff-base ligands

## Result and discussion

### Asymmetric cyclopropanation catalyzed by imine–amine ligands

An asymmetric cyclopropanation reaction of olefins with diazoacetates catalyzed by imine–amine ligands [16] was investigated as shown in Scheme 1. The catalyst ligands of this reaction ( $L^*$ ) are shown in Scheme 2. Four enantiomers [(1*S*, 2*R*), (1*S*, 2*S*), (1*R*, 2*R*), and (1*R*, 2*S*)] of *trans*- and *cis*-products are observed from the reaction simultaneously.

In the ANN method, we chose the topological index  $A_{x_1}$  to represent substituents  $R_1$ ,  $R_2$  of the reactants and  $R_3$  of the catalyst while four descriptors,  $A_{x_1}$  for  $R_2$ ,  $R_3$  and  $A_{x_1}$ ,  $A_{x_3}$  for  $R_1$  were used in the MLR. An indicator variable  $C_{R_2}$  that indicates the chirality of group  $R_2$  was also introduced into both calculations. If  $R_2$  is in the *L* conformation,  $C_{R_2} = -1$ ; if  $R_2$  is in the *D* conformation,  $C_{R_2} = +1$ ; otherwise,  $C_{R_2} = 0$ ; e.e.(*cis*)% and e.e.(*trans*)% were the dependent variables. The BP network was constructed with



**Scheme 5** Reaction and catalyst ligands for the asymmetric Grignard cross-coupling reaction catalyzed by ferrocenylphosphines

four input neurons, two hidden layer neurons, and two output neurons so that it could simultaneously predict *cis* and *trans* e.e.%. The results are shown in Table 1 and Fig. 1 and in Eqs. 1 and 2 (MLR only).

$$e.e.\%(cis) = -526.4 + 50.9A_{x_1}^{R_1} - 14.6A_{x_3}^{R_1} + 1.4A_{x_1}^{R_2} + 0.7A_{x_1}^{R_3} - 17.9C_{R_2}$$

$$r^2 = 0.953, F = 28.308, s = 5.045, n = 13$$

$$e.e.\%(trans) = -533.5 + 53.3A_{x_1}^{R_1} - 15.9A_{x_3}^{R_1} + 1.2A_{x_1}^{R_2} + 0.6A_{x_1}^{R_3} - 15.5C_{R_2}$$

$$r^2 = 0.947, F = 24.811, s = 4.676, n = 13$$

The table shows that most *cis* e.e.% values predicted by the ANN model are superior to the ones given by the MLR model except for entry 5. As for *trans* e.e.%, the MLR seems to be a little better than the ANN. In this reaction, both MLR and ANN fit the experimental data well, and the ANN model provided results of comparable accuracy to MLR with less input variables.

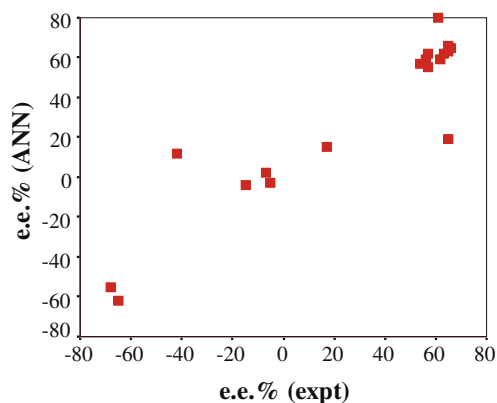
#### Asymmetric cyclopropanation catalyzed by Schiff-base ligands

The well-known work on cyclopropanation by Aratani [17], as shown in Scheme 3, was also investigated. The asymmetric catalysts derived from Schiff-bases are shown

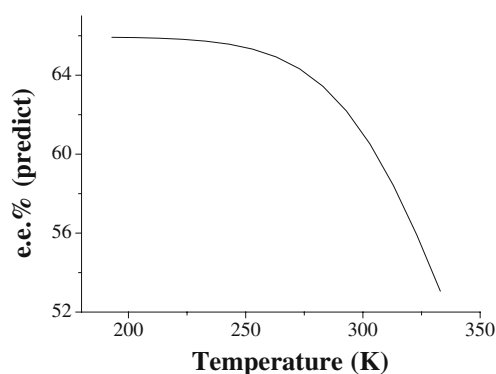
in Scheme 4, which may be considered as a fixed framework with varied substituents  $R_1$  and  $R_2$ , and  $R_2$  varies by two even smaller substituents  $r_1$  and  $r_2$ . Three indices ( $A_{x_1}, A_{x_2}$  for  $R_1$  and  $A_{x_3}$  for  $R_2$ ) as input variables and e.e.% of the (+)-*trans* product as the output variables were prepared for a BP network with two hidden layer neurons. In our previous MLR calculations, three substituents  $R_1, r_1$ , and  $r_2$  were used to represent the Schiff-base catalyst ligand as the e.e.% apparently cannot be fitted well with an entire  $R_2$  in a linear relation. The results are shown in Table 2 and Fig. 2 and in Eq. 3 (MLR only).

$$e.e.\% = -13.7 - 0.3A_{x_1}^{R_1} + 11.1A_{x_2}^{r_1} - 7.9A_{x_3}^{r_1} + 1.2A_{x_1}^{r_2}$$

$$r^2 = 0.870, F = 18.409, s = 8.225, n = 16$$

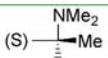
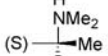
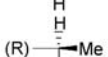
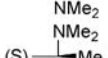
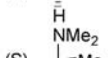
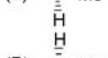
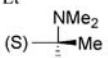
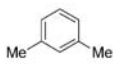
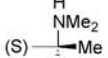
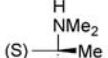
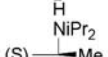
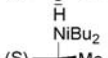

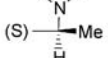
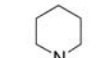
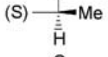
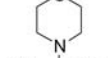


**Fig. 3** Plot of experimental vs ANN calculated e.e.% of the asymmetric Grignard cross-coupling reaction catalyzed by ferrocenylphosphines



**Fig. 4** Relations of temperature vs ANN-predicted e.e.% of the asymmetric Grignard cross-coupling reaction catalyzed by ferrocenylphosphines

**Table 3** Substituent pattern and the observed and calculated e.e.% of asymmetric Grignard cross-coupling reaction catalyzed by ferrocenylphosphines

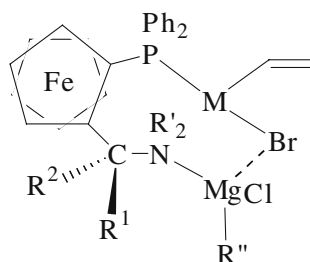
Entry	R <sup>1</sup>	R <sup>2</sup>	Planar Chirality	Temperature (K)	e.e.(%) (expt)	e.e.(%) (ANN)
1	Ph		R	273	63	62
2	Ph		R	253	66	65
3	Ph		S	253	-68 <sup>a</sup>	-55
4	Ph		R	318	56	59
5	Ph		R	298	61	80
6	Ph		R	273	54	57
7	Ph	CH <sub>2</sub> NMe <sub>2</sub>	S	273	-65 <sup>a</sup>	-62
8	Ph	Et	R	273	-5 <sup>a</sup>	-3
9	PhMe		R	273	65	66
10			R	273	65	63
11	PhOMe		R	273	57	62
12	Ph		R	273	-7 <sup>a</sup>	2
13	Ph		R	273	-15 <sup>a</sup>	-4
14	Ph		R	273	62	59
15	Ph		R	273	-42 <sup>a</sup>	12
16	Ph		R	273	17	15
17	Ph		R	273	65	19
18	Ph		R	273	57	55
Squared correlation coefficient					0.86	

Negative e.e.% means enantiomeric excess of the S-confirmation product

In this case, the predicted values given by the ANN model are superior to the MLR ones for entries 1, 5, 6, 9, 10, 11, and 13–16. In entries 2, 3, 4, 7, 8, 12, and 17, the

MLR seems to be better fitted. The ANN model still has room to be improved, especially for the entries when R<sub>1</sub>=Me. However, the ANN provides a superior global

**Scheme 6** Intermediate of the Grignard cross-coupling reaction



model with a larger squared correlation coefficient (0.90 to 0.85 of MLR) and lower standard error (6.38 to 8.25 of MLR). Moreover, the trained network may be used to predict more flexible patterns of substituent  $R_2$ , whereas the MLR is limited to disubstituted phenyl substituents.

#### Asymmetric Grignard cross-coupling catalyzed by ferrocenylphosphines

Finally, we treated the classic work on asymmetric Grignard cross-coupling reactions following Hayashi and Konishi [18] as shown in Scheme 5. As usual, the e.e.% of the product of  $R$  conformation was used as the output variable for a BP network with three hidden layer neurons.

As the series of substituents of  $R^2$  in the catalyst has two steric configurations, the expansion of topological descriptors  $A_{x1-3}^*$  must be used. Here, the adjustable parameters used were  $\delta=+0.5$  for ( $S$ ) conformation and  $\delta=-0.5$  for the ( $R$ ) conformation.  $A_{x1}$  for  $R^1$ ,  $A_{x1}^*$  for  $R^2$  were chosen for the input layer of the network. The planar chirality of the ferrocene was also considered as an indicator variable  $C_{\text{planar}}$ .  $C_{\text{planar}}=+1$  for the  $R$  conformation and  $-1$  for the  $S$  conformation. Temperature, which normally acts as a nonlinear factor affecting the enantiomer excess, was nearly impossible to include in MLR calculations. However, we tried to input it into a BP network. The results are shown in Table 3 and Fig. 3.

Figure 3 shows that entries 15 and 17 are not well predicted by the network. This may be caused by deficiencies of the model, but another explanation exists. The mechanism suggested by Hayashi proposed that the intermediate shown in Scheme 6 must be formed first. The chair conformation of the hexatomic ring of entry 15 and another  $N$  atom as an additional donor to the metal atom of entry 17 can disturb the normal formation process of the intermediate. This may be the reason why these two cases cannot be fitted well. These exceptions may suggest a reexamination of the mechanism of the reaction.

With  $R^1$  as Ph and  $R^2$  as , a series of temperatures from 193 to 333 K ( $-80$  to  $50$  °C) were input into the trained network to investigate the relationship between e.e.% and

temperature. The results are shown Fig. 4. Although the extrapolation ability of the ANN is not comparable to interpolation in many situations, this case leads to a result that tally with the known fact that a decrease of temperature benefits the enantiomer excess. From the curve, it may be observed that the improvements of e.e.% below 250 K are not as significant and may even be unable to match for the decrease of yield of reaction. Above 300 K, the e.e.% falls dramatically. Thus, 250–300 K may be considered a suitable temperature range for the reaction.

## Conclusions

Models may be developed for varieties of asymmetric catalytic reactions using molecule topological indices and feed-forward neural network calculation. Our method can be helpful for screening new asymmetric catalysts, instructing the design of reaction environments, and to test or improve the proposed mechanism.

**Acknowledgement** Financial support from the National Natural Science Foundation of China (no. 20472077) is gratefully acknowledged.

## References

1. Eliel EL (1962) Stereochemistry of carbon compounds. McGraw-Hill, New York, pp 81–87
2. Mislow K (1966) Introduction to stereochemistry. W. A. Benjamin, New York, pp 122–123
3. Ojima I (1993) Catalytic asymmetric synthesis. VCH, New York, pp 1–6
4. Morrison JD, Mosher HS (1986) Asymmetric organic reactions. Prentice-Hall, Englewood Cliffs, NJ, pp 1–28
5. Trost BM, Crawley ML (2003) Chem Rev 103:2921–2944
6. Soai K, Niwa S (1992) Chem Rev 92:833–856
7. Chen J, Yougui L, Qingshan T, Tianpa Y (2003) J Chem Inf Comput Sci 43:1876–1881
8. Qingshan T, Chen J, Yougui L, Changsheng J, Tianpa Y (2004) J Mol Catal A Chem 219:315–317
9. Zupan J, Gasteiger J (1991) Anal Chim Acta 248:1–30
10. Kateman G (1993) Chemom Intell Lab Syst Lab Inf Manag 19:135–142
11. Yao Y-Y, Xu L (1993) J Chem Inf Comput Sci 33:590–594
12. Daniels C, Wood FS (1986) Fitting equations to data, 2nd edn. Wiley, New York, pp 56–57
13. Draper NR, Smith H (1981) Applied regression analysis, 2nd edn. Wiley, New York, pp 241–243
14. Zupan J, Gasteiger J (1993) Neural networks for chemists: an introduction. VCH, Weinheim, Germany, pp 119–148
15. Svozil D, Kvasnicka V, Pospichal J (1997) Chemom Intell Lab Syst Lab Inf Manag 39:43–62
16. Ma JA, Wang LX (2001) Tetrahedron: Asymmetry 12:2801–2804
17. Aratani T (1985) Pure Appl Chem 57:1839–1844
18. Hayashi T, Konishi M (1982) J Am Chem Soc 104:180–186