J Mol Model (2006) 12: 521–527
DOI 10.1007/s00894-005-0087-6

ORIGINAL PAPER

Jinwei Gao · Xueye Wang · Xinliang Yu · Xiaobing Li · Hanlu Wang

Calculation of polyamides melting point by quantum-chemical method and BP artificial neural networks

Received: 23 June 2005 / Accepted: 11 November 2005 / Published online: 14 January 2006

Abstract Quantitative structure–property relationships (QSPR) for the melting point of the polyamides have been determined. All descriptors were calculated from molecular structures at the B3LYP/6–31G(d) level and a QSPR model was generated by multiple linear regression (MLR). The important molecular descriptors for polyamide melting-point temperatures (T_m) are the number of benzene rings in the backbone chain, the proportion of methylene and acylamino in the backbone chain, the total molecular energy and the atomic charge for the oxygen atom in the acylamino group. The MLR determination coefficient (r^2) and the standard error of estimation for the model are 0.865 and 21.34 K, respectively. In addition to the nonlinear regression technique, error back-propagation artificial neural networks (BPANN) was used to study the relationships between molecular structures and melting-point temperatures. It is concluded that melting-point temperatures for polyamides can be described by molecular chain rigidity and interchain attractive interactions. The more accurate predicted results were obtained from BPANN.

Keywords Polyamides · Melting point · QSPR · DFT · BP artificial neural networks

Introduction

The behaviors of melting and crystallization are of fundamental importance for polymer materials. However, they depend on the arrangement of the molecules in the crystal lattice as well as on the strength of the pairwise group interactions. Many available molecular descriptors do not describe the many-body crystal packing effects and intermolecular forces in condensed media satisfactorily [1]. The relationship between equilibrium melting temperatures and the molecular structures is difficult to correlate. Early in 1955, Bunn wrote a paper on this topic and considered the stiffness and the interchain attractive interactions of the chain molecules as the controlling factors for the melting points [2], and this idea was widely accepted [3, 4]. Almost at the same time, Flory considered the stiffness in his lattice statistics [5], but assumed the interchain attraction to be small and neglected its contribution. This neglect was subsequently criticized [6, 7]. On the experimental side, the exact measurement of the equilibrium melting point of polymers is also complex. Kinetic factors such as the small size of the crystallites and their metastability in chain folding, and structural factors such as the defects of the crystallizable chains and the impure environment of crystallites contribute to variations in the melting point. To date, QSPR for melting point temperatures (T_m) of small molecules such as alkanes [8], alkyl benzenes [9] and acrylic carbonyl [10] has been studied, as summarized by Kartritzky and coworkers [11]. However, T_m QSPR for polymers have rarely been published. Polyamides are a kind of old and important material and have been widely used in most fields, so research for polyamides remains interesting.

In this paper, a QSPR model for polyamides T_m is reported, regarding the stiffness and the interchain attractive interactions of the chain molecules to be dominant factors. Based on quantum-chemical method and BPANN, this model can predict the melting points of a set of low and medium molecular weight homo- and copolymers, and the models derived can be used to predict T_m for unknown polyamides.

Procedure

The process of the QSPR model generation includes the experimental data, the molecular descriptors, the QSPR model, and the validation process.

J. Gao · X. Wang (✉) · X. Yu · X. Li · H. Wang
College of Chemistry, Xiangtan University,
Xiangtan 411105, People's Republic of China
e-mail: wxueye@xtu.edu.cn
Tel.: +86-0732-8292206

Experimental data

All the experimental data are quoted from Reference [12] and the web [13]. A sample of 80 polyamides with extensive structural diversity was selected as the working sample shown in Tables 1 and 2. All experimental data were randomly divided into two groups, a training set consisting of 41 polyamides and a test set consisting of 39 polyamides. The T_m for 80 polyamides included in the set ranged between 428 and 701 K with an average value of 273 K.

Calculation of molecular descriptor

It is impossible to calculate descriptors directly for an entire molecule with high molecular weight. In this paper, models consisting of repeating units end-capped by hydrogen were chosen as a study model to calculate molecular descriptors. The single point energy and thermochemistry data were calculated by density function theory (B3LYP/6–31G(d)) using Gaussian 03 [14]. The necessary quantum-chemical descriptors, which include frontier orbital energies, dipolar

Table 1 Experimental and fit values for training set

Number	Name	T_m(K), exp	T_m(K), train (MLR)	T_m(K), train (BPANN)
1	Poly(undecyliminocarbonyl)	428	486	438
2	Poly(iminohexane-1,6-diyliminoheptanedioyl)	475	511	474
3	Poly(heptano-7-lactam)	498	532	510
4	Poly(imino-1-oxo-3-phenylpropane-1,3-diyl)	522	526	487
5	Poly(2-methyl-2-propyl-3-aminopropionic acid)	527	535	510
6	Poly(2-ethyl-2-methyl-3-aminopropionic acid)	538	541	520
7	Poly(propano-3-lactam)	550	556	552
8	Poly(azelaoyl oxalodihydrazide)	566	554	574
9	Poly(iminobutane-1,4-diyliminoterephthaloyl)	668	607	656
10	Poly[imino(1-oxododecane-1,12-diyl)]	448	502	456
11	Poly(10-aminocapric acid)	453	513	478
12	Poly(hexyliminocarbonyl)	468	516	490
13	Poly[(3-methylphenyl)iminocarbonyl]	473	528	492
14	Poly[imino(1-oxoundecane-1,11-diyl)]	480	508	467
15	Poly(pentamethylene pimelamide)	487	517	484
16	Poly(hexano-6-lactam)	493	537	520
17	Poly(tetramethylene azelamide)	496	523	484
18	Poly(iminoglutarylimino-2,2-dimethylpentamethylene)	504	546	524
19	Poly(imino-2,2-dimethyl-1-oxo-3-phenylpropane-1,3-diyl)	508	547	524
20	Poly(pentamethylene malonamide)	515	546	528
21	Poly[(phenylimino)carbonyl]	520	531	503
22	Poly(2-butyl-2-methyl-3-aminopropionic acid)	524	530	499
23	Poly[imino(1-methyl-3-oxoprop-1-ene-1,3-diyl)]	525	553	544
24	Poly(hydrazine-1,2-diyldecanedioyl)	527	523	502
25	Poly(4-aminobutyric acid)	531	550	542
26	Poly(tetramethylene succinamide)	533	548	537
27	Poly(imino-3-isopropyl-1-oxopropane-1,3-diyl)	538	544	531
28	Poly(imino-1,2-phenyleneiminoadipoyl)	542	569	556
29	Poly(gamma-ethylglutamic acid)	543	546	549
30	Poly(iminomalonyliminobutane-1,4-diyl)	548	551	548
31	Poly(hexamethylene oxamide)	549	548	538
32	Poly(alpha-butyl-L-aspartate)	556	538	545
33	Poly(iminomethyleneiminohexanedioyl)	558	551	548
34	Poly(alpha-ethyl-L-aspartate)	560	551	568
35	Poly(iminooxalyliminotetramethylene)	565	558	560
36	Poly(3-aminoisobutyric acid)	570	549	541
37	Poly(alpha-hexyl-L-aspartate)	580	584	574
38	Poly[3-(4-aminophenyl)propionic acid]	635	613	656
39	Poly[ethylene (*p*-carboxyphenylene)acetamide]	667	610	666
40	Poly(imino-1,4-phenyleneiminoadipoyl)	672	614	662
41	Poly(imino-1,4-phenylenecarbonyl)	682	622	679

Table 2 Experimental and predicted values for test set

Number	Name	T_m(K), exp	$T_{m\ 0}$(K), pred (MLR)	T_m(K), pred (BPANN)
1	Poly(13-aminotridecanoic acid)	455	446	460
2	Poly(isobutyliminocarbonyl)	488	512	499
3	Poly(iminoadipoyliminoheptane-1,7-diyl)	511	485	524
4	Poly(5-aminovaleric acid)	533	531	543
5	Poly(iminomethyleneimino-1,7-dioxoheptane-1,7-diyl)	548	536	546
6	Poly(hydrazine-1,2-diylsuccinylhydrazine-1,2-diyladipoyl)	558	593	567
7	Poly(imino-1,3-phenyleneiminoglutaryl)	596	620	596
8	Poly(iminoethyleneiminoterephthaloyl)	701	676	678
9	Poly(dec-9-enyliminocarbonyl)	451	447	490
10	Poly(6-methyl-7-aminoenanthic acid)	459	501	528
11	Poly[imino(1-oxooctane-1,8-diyl)]	471	499	525
12	Poly(9-aminononanoic acid)	476	488	519
13	Poly(pentyliminocarbonyl)	482	501	517
14	Poly(imino-2,3-dimethoxybutanedioyliminopentamethylene)	490	474	522
15	Poly[(2-methylbutylimino)carbonyl]	493	502	520
16	Poly(hexamethylene 3-methyladipamide)	503	489	529
17	Poly(iminobutane-1,4-diyliminoheptanedioyl)	506	507	534
18	Poly(iminoglutaryliminohexane-1,6-diyl)	512	506	533
19	Poly(iminomalonyliminohexane-1,6-diyl)	516	527	542
20	Poly(ethyliminocarbonyl)	523	534	535
21	Poly(iminopentane-1,5-diyliminoadipoyl)	524	505	532
22	Poly(iminoglutaryliminopentane-1,5-diyl)	527	517	540
23	Poly(iminobutane-1,4-diyliminooctanedioyl)	528	495	529
24	Poly[D(−)-3-methyl-6-aminocaproic acid]	533	510	540
25	Poly(imino-2,2,3,3-tetramethyl-1-oxopropane-1,3-diyl)	537	509	540
26	Poly(iminomethyleneiminodecanedioyl)	539	504	531
27	Poly(iminoethyleneiminohexanedioyl)	543	537	548
28	Poly(iminobutanedioyliminohexane-1,6-diyl)	548	516	539
29	Poly(iminomethyleneimino-1,8-dioxooctane-1,8-diyl)	549	525	540
30	Poly(imino-1,2-phenyleneiminoglutaryl)	553	559	566
31	Poly(iminomethyleneiminoglutaryl)	557	557	555
32	Poly(alpha-isobutylaspartic acid)	559	549	555
33	Poly(glycine)	561	562	561
34	Poly(3-amino-4-methylbutyric acid)	569	531	549
35	Poly[imino(1,2,2-trimethyl-3-oxopropane-1,3-diyl)	573	523	546
36	Poly(hydrazine-1,2-diylethanedioyl)	598	588	559
37	Poly(*p*-phenylene glutaramide)	638	664	608
38	Poly[trimethylene (*p*-carboxyphenylene)acetamide]	670	655	658
39	Poly(hydrazine-1,2-diylterephthaloyl)	673	676	673

moment, polar moments, atomic charges and total energy were calculated. A few other descriptors are defined, such as NCA, PMA, *L*, *B*, and LB. The detailed descriptors are described in Table 3.

The PMA value can be calculated using the following formula: PMA=$N_{methylene}/N_{acylamino}$, where $N_{methylene}$ is the number of methylenes in the backbone chain, and $N_{acylamino}$ is the number of acylaminos in the backbone chain.

Method and QSPR model

The correlation analysis of the QSPR model was carried out by two methods. MLR generates a form as $Y=a_1 + a_2 X_2 + a_3 X_3 + \ldots + a_n X_n$, where X_2,..., X_n are the descriptors and a_1, a_2,..., a_n are the regression coefficient parameters. The quality of the regression model is measured using primarily four statistical parameters: the determination coefficient (r^2), the standard error of estimation (*SD*), *T*-value (*T*) and the significance level value (*P*). Moreover, variance inflation factors (VIF) are calculated in order to detect multicollinearities in the model using the formula VIF=1/(1−r^2). That a VIF is lower than 10 indicates that the model originates from independent molecular properties.

The MLR process was performed using the analysis software DPS 2.0 [15]. After the linear regression, the relationship between the descriptors selected and the experimental data was analyzed by BPANN. The ANN

Table 3 Quantum-chemical and other descriptors

Number	Descriptors	Definition	Unit
1	ΔE_g	Energy difference of HOMO and LUMO	au
2	Q_{O-}	Atomic charge for oxygen in acylamino	au
3	Q_{H+}	The most positive atomic charge for hydrogen atom	au
4	E_t	Total molecular energy	au
5	μ	Molecular dipole moment	Debye
6	α_{tot}	Molecular polarizability	au
7	L	The length of side chain	–
8	B	The number of side chains	n
9	NCA	The number of total carbon atom in repeating unit	n
10	PMA	The proportion of methylene to acylamino in the backbone chain	–
11	LB	The value of benzene rings in the backbone chain	–

consists of a fully connected three-layer system, and each neuron in a given layer is fully connected to all neurons in the adjacent levels. The number of neurons of the input layer is equal to that of the molecular descriptors taken from the best MLR analysis. The output layer contained one neuron representing the calculated value. The network was trained with the BP algorithm, and the learning rate was 0.0001 and momentum is 0.01. The number of neurons in the hidden layer was optimized by trial and error training assays. The trained network was used to predict the values of the test set. The program BPANN was written in our lab.

Table 4 Four-descriptor MLR model selected for polyamides in data set[a]

Descriptor	Coefficient	Standard error	*t*-value	*P* value	VIF
Intercept	300.15	13.70	21.89	0.0000	–
PMA	−23.58	2.23	−10.55	0.0000	4.5
LB	55.99	2.94	19.03	0.0000	1.1
Q_{O-}	53.10	18.19	2.91	0.0059	1.0
E_t	−0.31	0.04	−8.42	0.0000	4.4

$$T_m(K)=300.1545-23.5783PMA+55.9903LB+53.0969Q_O-0.3119E_t \quad (1)$$

[a]Model statistics: r^2=0.900; *F*-value=104.5009; *P* value<0.00001; *SD*=9.98; r^2_{CV}=0.931; *N*=41

Table 6 Summary of factor analysis for the model

Factor	Eigenvalue	Percentage total	Cumulative
1	1.993	49.8	49.8
2	0.992	24.8	74.6
3	0.896	22.4	97.0
4	0.117	2.9	100.0

Table 7 Varimax rotated factor matrix for the descriptors used in the predicted T_m for polyamides

Descriptor	Factor 1	Factor 2	Factor 3
PMA	0.957	0.018	0.153
LB	0.127	0.006	0.991
Q_{O-}	0.040	0.999	0.006
E_t	−0.968	−0.047	−0.056

Results and discussion

Regression model

A subset of more than 25 descriptors was used in this study. Some of the descriptors are described in Table 3. After the examination by MLR, the best QSPR model was obtained and the correlation between T_m and the dependent descriptors is shown in Table 4. Only four descriptors (LB, PMA, E_t and Q_{O-}) were selected for the best correlation models. The results shown in Table 5 for the MLR analysis of variance show that the model developed is significant, *T*-values of all descriptors are more than two, and the *SD* does not exceed by 25%. The *P* value is as low as 0.005, which indicates that the chosen molecular descriptors can successfully predict T_m of polyamides. The multicollinearity among the descriptors was tested using the VIF method. The VIF values shown in Table 4 are less than ten, which suggest that there is no serious multicollinearity among the descriptors.

Factor analysis was used to test general multivariate relationships between the selected descriptors in the predicted model [16]. These results are shown in Tables 6 and 7. Three factors have Eigenvalues more than one, and consequently data variability can be expressed in terms of three new reduced groups of variables named factors. The first factor explains 50% of the total variance and the second one 25%, yielding a total of 75%. To have an easier data interpretation from the calculated factors,

Table 5 Summary of the results produced by different methods

Number	Method	Training set	Test set	r^2_{train}	SD_{train}	r^2_{pred}	SD_{pred}	Figure	Equation
1	MLR	41	39	0.900	9.98	0.865	21.34	1	1
2	BPANN	41	39	0.931	15.37	0.882	15.19	2	–
3	Cross-validation MLR	80−*i*	*i*	–	–	0.931	12.13	3	–
4	Cross-validation BPANN	80−*i*	*i*	–	–	0.958	8.87	4	–

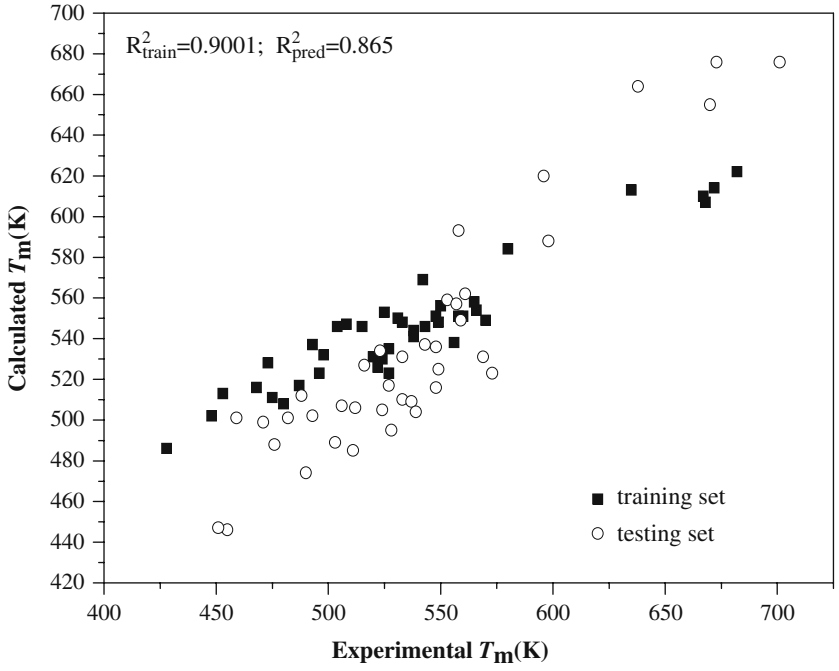

Fig. 1 Experimental vs calculated T_m using MLR

the Varimax Factor Matrix was rotated, and the results are show in Table 7. The results suggest that the first factor is loaded by both E_t and PMA, and two descriptors associate with molecular stability and polarity. The second factor is linked to atomic charge. The third factor is dominated by LB, a structure descriptor that associates with molecular flexibility. Therefore, both geometrical and electronic descriptors determine T_m for polyamides.

The BPANN model was generated using the descriptors selected by the MLR with a four-two-one network model. The calculated results are shown in Table 5. More details can be found in Tables 1 and 2, where the fitting values and the predicted values by the two model methods for training set and test set are shown. The results of regression and prediction for polyamides T_m are shown in Figs. 1 and 2. The test set was not used for the training process, so the predicted values can successfully test validity and the credibility of the model.

To explore the reliability of the proposed methods further, the method of leave-one-out cross-validation was used. A data set consisting of 20 polyamides randomly

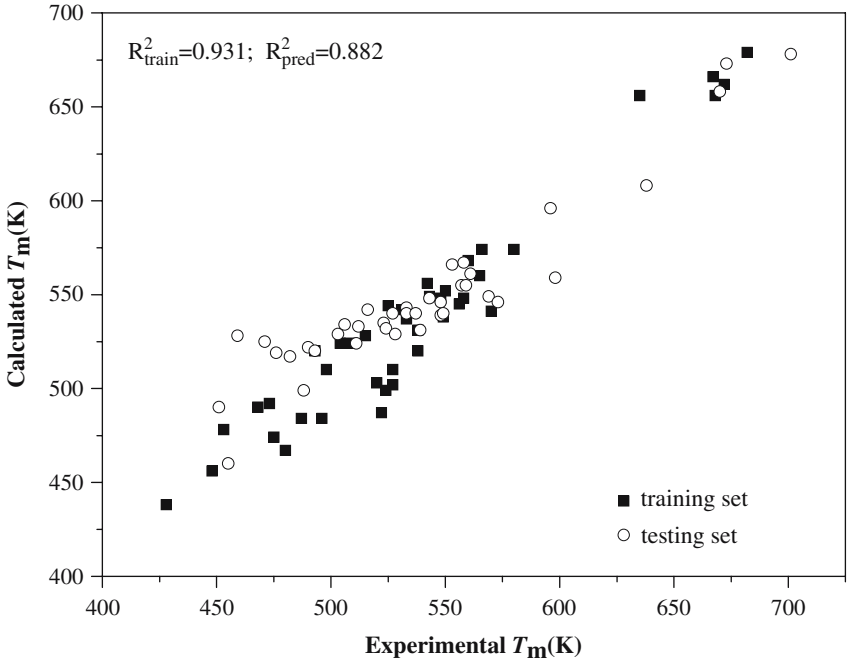

Fig. 2 Experimental vs calculated T_m using BPANN

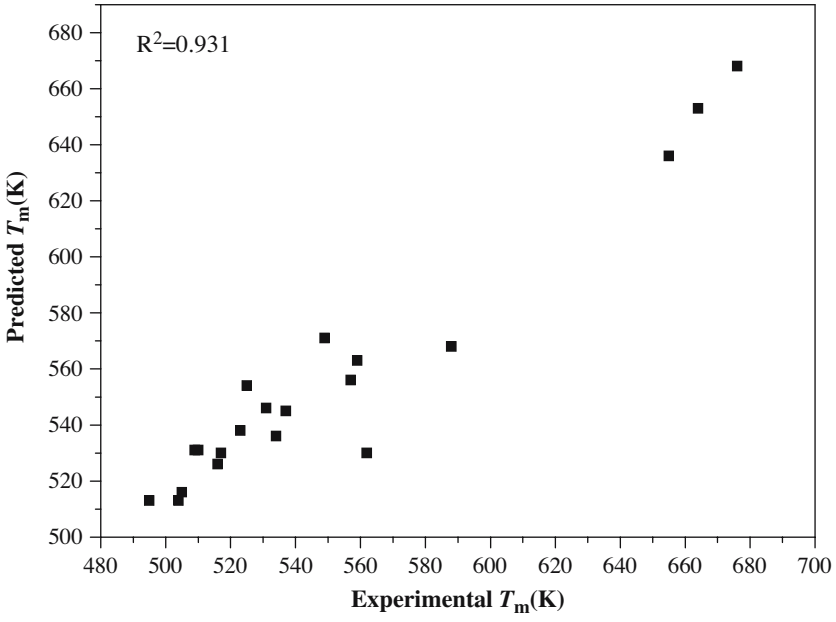

Fig. 3 Experimental vs predicted T_m with cross-validation method (MLR)

selected from full data set was predicted. Table 5, Figs. 3 and 4 summarize these predicted results, and the BPANN is clearly superior to the MLR.

Descriptor interpretation

T_m can be expressed by the equation $T_m=\Delta H_m/\Delta S_m$. Where ΔH_m is heat of fusion, and ΔS_m is the entropy of fusion. According to this equation, T_m can be measured by the proportion ΔH_m to ΔS_m. An increase of ΔH_m and a decrease of ΔS_m lead to an increase of T_m. Four descriptors were chosen in T_m model: PMA, LB,E_t and Q_{O-}. According to the *t*-test value, the most significant descriptor is LB, which is a geometrical descriptor and relates to the chain rigidity. It can influence ΔS_m and lead to a significant positive correlation with T_m. The second most important descriptor in the model is PMA, which accounts for the proportion of the polar groups (acylamino) to the nonpolar groups (methylene), which influences ΔH_m directly, if the value of PMA increases, the polarity of the main chain increases, which can form more hydrogen bonds among main chains and heighten intermolecular forces. Consequently, T_m will be higher.

The total energy of a molecule, E_t, is a thermochemistry descriptor including translational, rotational, and electronic energy, and influences molecular stability directly. It has a negative correlation with T_m.

The last descriptor is atomic charge for the oxygen in the acylamino group. The charge distribution facilitates the

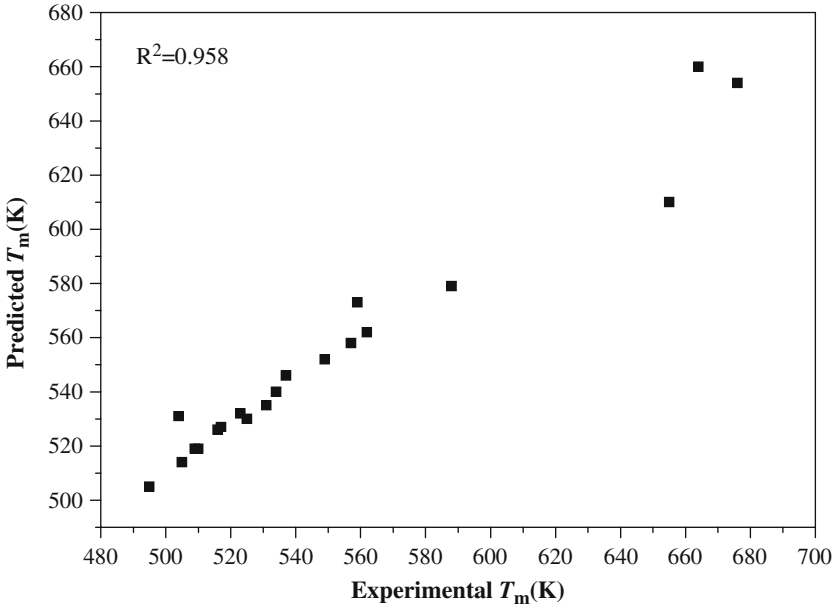

Fig. 4 Experimental vs predicted T_m with cross-validation method (BPANN)

formation of dipole–dipole interactions among main chains and can be a function of polarity as well as the molecular packing ability in the crystalline solid. The atomic charge for oxygen in acylamino can also reflect hydrogen-bonding properties, which influence ΔH_m. The change has a positive correlation with T_m in this model.

Conclusion

The calculations demonstrate that the four-descriptor model can predict T_m successfully for polyamides, and MLR and BPANN are practical methods for building the QSPR models. In comparison with MLR, BPANN proves to be more accurate with the following statistics: $r^2_{pred} = 0.882$ for the test set, $r^2_{CV} = 0.958$ for cross-validation method. The descriptors selected for the T_m model are representative, and the value of T_m is governed mainly by the molecular rigidity and molecular polarity. Following work will use geometrical, electronic and topology descriptors.

Acknowledgement The authors like to thank the financial support from the Scientific Research Fund of Hunan Provincial Education Department (05A002) for the research work.

References

1. Katritzky AR, Lobanov VS, Karelson M (1995) Chem Soc Rev 24:279–287
2. Bunn CW (1955) J Polym Sci 16:323–343
3. Wunderlich B (1973) Crystal structure, morphology, defects, macromolecular physics, vol 1. Academic, New York, pp 68
4. Tadokoro H (1979) Structure of crystalline polymers. Wiley, New York, pp 15
5. Flory PJ (1956) Proc R Soc London A234:60–73
6. Gujirati PD, Goldstein MJ (1981) J Chem Phys 74:2596–2603
7. Nagle JF, Gujirati PD, Goldstein MJ (1984) J Phys Chem 88:4599–4608
8. Toropov A, Toropova A, Ismailov T, Bonchev D (1998) J Mol Struct: (THEOCHEM) 424:237–247
9. Firpo M, Gavernet L, Castro EA, Toropov AA (2000) J Mol Struct: (THEOCHEM) 501–502:419–425
10. Toropov AA, Toropova AP (2002) J Mol Struct: (THEOCHEM) 581:11–15
11. Katritzky AR, Jain R, Lomak A, Petrukhin R (2001) Perspective 1:261–265
12. Brandrup J (1999) Polymer handbook. 4th edn. Wiley, New York
13. Polymer Database: http://polymer.nims.go.jp/PoLyInfo/
14. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, MonTgomery JA, Jr, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson BG, Chen W, Wong MW, Andres JL, Head-Gordon M, Replogle ES, Pople JA (2003) Gaussian 03, Revision B.05. Gaussian Inc, Pittsburgh, Pennsylvania
15. Tang Q, Feng M (2002) Practical Statistics and DPS Data Processing System. Science, Beijing
16. Wold S, Johanson E, Cocchi M (1993) In: Kubinyi H (ed) 3D QSAR in drug design: theory, method and applications. ESCOM, Leiden, pp 523–550