DOI: 10.1002/ejoc.200800392

Construction of a Statistical Evaluation Model Based on Molecular Centrality to Find Retrosynthetically Important Bonds in Organic Compounds

Akio Tanaka,*[a] Takashi Kawai,[a] Tsutomu Matsumoto,[a] Mihoko Fujii,[a] Tetsuhiko Takabatake,[a] Hideho Okamoto,[b] and Kimito Funatsu*[c]

Keywords: Synthesis design / Molecular centrality / Bond centrality / Bond dissociation energy / Retrosynthesis / Molecular complexity

For the purpose of finding retrosynthetically important bonds in a molecule, a new evaluation score has been defined through a logistic regression analysis of known reactions stored in reaction databases. We conceived that reaction center bonds in reaction databases describe one of the most retrosynthetically important bonds for each product structure. The derived statistical equation consists of bond centrality and bond dissociation energy terms. The equation shows that synthetically useful bonds tend to be more central in a molecule and to be weaker bonds. Coefficients in two statistical equations derived from two different reaction data sets are quite similar to each other. From a comparison of molecular complexities and validation with 35 complicated organic compounds, the evaluation equation was proved to be useful.

Introduction

Many organic chemists use reaction database systems to explore synthetic routes.[1] By using such systems, researchers envisage reactions to simplify structures of targets and look for the corresponding literature. Synthesis design systems have been developed to find the synthetic routes of targets exhaustively by using computational methods. The development of such systems has a long history. The first computer-aided synthesis design system OCSS was published in 1969[2] and was the predecessor of the more well-known LHASA.[3] Many systems have been published since then and some are still being developed.[4] The systems propose plausible reactions and precursors for input targets, and most of these systems have been designed to continuously generate precursors until reaching practical and commercially available compounds.

In the development of an industrial process it is important to find low-cost starting materials that are stable and reaction conditions that can be scaled up. Thus, it is necessary to explore a wide variety of synthetic routes exhaustively.

[a] Organic Synthesis Research Laboratory, Sumitomo Chemical Co., Ltd.,
1-98, Kasugade-naka 3-chome, Konohana-ku, Osaka, 554-8558, Japan
E-mail: tanakaa1@sc.sumitomo-chem.co.jp
[b] Center for Research and Advancement in Higher Education, Kyushu University,
4-2-1, Ropponmatsu, Chuo-ku, Fukuoka, 810-8560, Japan
[c] Fukan Environmental Engineering, Department of Chemical System Engineering, The University of Tokyo,
Hongo 7-3-1, Bunkyo-ku, Tokyo, 113-8656, Japan
E-mail: funatsu@chemsys.t.u-tokyo.ac.jp

As a target becomes larger, the number of proposed routes increases drastically and users often encounter the difficult task of evaluating the details of massive syntheses trees. To control the 'combinatorial explosion' in synthesis design, only retrosynthetically important bonds should be cut to simplify targets and intermediates effectively.

To find retrosynthetically useful bonds, Corey proposed the first heuristics strategy in 1971.[5] In 1981, Bertz demonstrated a quantitative evaluation method of a 'molecular complex' based on graph theory.[6] According to the concept of molecular complexity, the most important bond in a molecule gives the largest decrease of complexity on bond disconnection.[7]

As a new quantitative index for evaluating bonds, recently we reported on 'molecular centrality', which takes convergent synthesis into consideration.[8] Molecular centrality consists of bond and atom centrality, which are numerical rating scales used to describe the location of atoms and bonds within a molecule. The values increase towards the center of a molecule. It was found that the index was able to estimate retrosynthetically important bonds. However, it was also found that synthetic accessibility should also be considered.

In the synthesis design system WODCA, strategic bonds have so far been perceived by multiple physicochemical parameters.[9] On the other hand, we have investigated evaluation equations composed of not only physicochemical parameters but also topological properties. When more than one parameter is investigated, the weights of the parameters become important. To decide the weights, we attempted to extract the characteristics of reaction center bonds in reac-

tion databases, which is one of the most reliable tools available for chemists to explore synthetic routes.

In consideration of molecular centrality and synthetic accessibility, new evaluation equations have been constructed by statistical methods with two different data sets of reaction databases as training sets. The equations were tested with 35 complicated organic compounds reported in journals by comparison with an equation based on only bond centrality. In addition, the scores from the equations were compared with molecular complexities. In this paper, a new evaluation equation to determine retrosynthetically important bonds in a molecule is described.

Statistical Data and Method

Independent Variables

For statistical analysis, three bond properties, bond centrality (BC), bond dissociation energy (BDE),[10] and the number of chiral centers (NCC),[11] were used as independent variables. BC is a parameter used to describe the location of bonds in a molecule, and the BDE and NCC are indices that represent the degree of synthetic accessibility and intractableness related to the asymmetric syntheses of bonds. The properties of all bonds in the product of each reaction scheme were calculated by statistical analyses.

In organic synthesis, convergent synthesis is a standard strategy for improving the efficiency of a multi-step synthesis of a complicated organic molecule. A few segments are prepared independently and connected together in one of the last steps, hence the last building bond is often located closer to the center of the target.

Molecular centrality is the numerical number that represents the location of atoms and bonds within a molecule and consists of bond and atom centrality. Molecular centrality is defined on the basis of squared node distances, expressed as Equations (1)–(5). The index is a quantitative measure and additionally it is comparable for all bonds in different molecules.

$$D(i) = \sum_{j=1}^{N} dist^2(i,j) \quad (1)$$

$$D'(i) = \frac{1}{\frac{D(i)}{D\,max}} = \frac{D\,max.}{D(i)} \quad (2)$$

$$D'av. = \frac{\sum_{i=1}^{N} D'(i)}{N} \quad (3)$$

$$Atom\ centrality(i) = \frac{D'(i)}{D'av.} \quad (4)$$

$$Bond\ centrality(i,j) = Atom\ centrality(i) + Atom\ centrality(j) \quad (5)$$

In Equations (1)–(5), N is the number of atoms including hydrogen atoms, i and j describe the node numbers of atoms in a molecule, and $dist^2(i,j)$ is the squared node distance between atoms i and j, where the node distance is the smallest number of bonds between the two atoms. $D(i)$ is the value of i, which is the sum of $dist^2(i,j)$ from $j = 1$ to N [Equation (1)]. $Dmax$ is the largest value of $D(i)$. $D'(i)$ is calculated by dividing $Dmax$ by $D(i)$ [Equation (2)], and $D'av$ is the average of $D'(i)$ in a molecule. $Atom\ centrality(i)$ is obtained by dividing $D'(i)$ by $D'av$ [Equation (4)]. For each bond, $bond\ centrality(i,j)$ is defined by the sum of atom centralities of edge atoms i and j [Equation (5)]. The bond centrality was adopted as an independent variable.

With regard to the synthetic accessibility of each bond, the BDE was adopted as a measure of the ease of formation. Compared with carbon–carbon bonds, carbon–heteroatom and heteroatom–heteroatom bonds are generally more facile to make. In other words, the bonds have the potential to easily simplify the structures. In general, the BDEs of these bonds are relatively low. In addition, it is better to make heteroatom-containing bonds in the later steps because they are more reactive, which leads to an increase in the chance of side-reactions in the long synthesis procedure.

In cases of generating bonds with one or two chiral centers, synthetic methods are often limited. The synthetic difficulty stemming from asymmetric bond formation is represented by NCC to study the effects of such bonds on synthetic strategies.

Dependent Variable and Statistical Method

As data sources for statistical analyses, reaction databases storing fact data were used. In general, organic chemists use reaction database systems to find applicable reactions that may realize the synthesis of partial structures of targets and also to find reported synthetic reactions of fixed target structures. We emphasized the latter, 'fixed target structures', the usage of reaction databases, and consequently we conceived that reaction center bonds in product structures in the databases could be regarded as one of the best candidates of retrosynthetically important bonds in the target structures.

In this paper, a retrosynthetically important bond means the disconnected bond in the retrosynthetic direction, and thus reactions with connecting bond formations were focused upon. A connecting bond is a bond newly created from two separated atoms. In the database, the following reaction schemes were used for statistical studies.

a) Reactions with a single product were used to remove reactions with byproducts.

b) Only reactions with complete mapping information between reactants and a product were used to make clear the reaction center bonds.

c) Reactions consisting of only one or two connection bond formations were considered to highlight and simplify the characteristics of the bond environment. The goal of the

statistical analysis was to make evaluation equations for bonds in molecules, but not for reaction types.

d) Reactions that include either bond order or functional group changes are ignored because the reactions do not contribute to the construction of skeletons, even if they include connecting bond formations.

e) Reactions involving organometallic compounds are omitted because of the difficulty in estimating the BDEs of organometallic bonds by this method.[10]

For reaction schemes satisfying the above conditions, the parameters BC, BDE, and NCC of the connecting and other bonds in the product structures were calculated as independent variables [Equations (6)–(8)].

$$P_{(event)} = \frac{1}{1+\exp(-Z)} \quad (6)$$

$$Z = a_0 + a_1X_1 + a_2X_2 + \ldots + a_iX_i \quad (7)$$

$$P_{(not\ event)} = 1 - P_{(event)} \quad (8)$$

Logistic regression analysis (LoRA)[12] was used for the statistical analyses. The regression analysis is a widely used technique not only in the field of natural science such as biochemistry[13] and ecology,[14] but also in the field of social science.[15] In organic chemistry, prediction studies of bond reactivity[16] and spectra[17] have been reported.

LoRA is generally applied to a phenomenon in which dependent variables are qualitative data and independent variables are quantitative data. The dependent variables are usually represented by binary digits, 0 and 1, and they correspond to probability. Probability $P_{(event)}$ is expressed by regression equations that are sigmoidal functions [Equation (6)] by using independent variables X_i [Equation (7)] and it has a range of 0 to 1.[12c]

BC, BDE, and NCC have been considered as independent variables X_i in Equation (7), and for the dependent variables corresponding to $P_{(event)}$ in Equation (6), connecting bonds of each reaction are given 1 and other bonds are given 0.

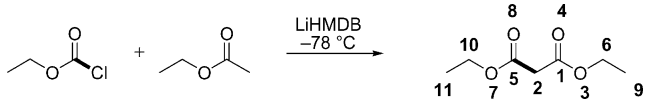
Unfortunately the definition of dependent variables creates awkward data sets without any modification. When a certain reaction scheme consists of one connecting bond in the reaction center, the connecting bond leads to one data set with the dependent variable value 1, and all bonds except for the connecting bond provide the same number of data sets as the bonds with the dependent variable value 0. The unbalanced distribution of data sets may have a bad influence on the statistical analyses, which are eventually incomplete. To solve this problem, the most dissimilar data set from the connecting bonds was extracted on behalf of all bonds except for the connecting bonds. The dissimilarities from the connecting bonds were decided by standardizing variable values for all bonds [Equation (9)]. The standardization involves the average of each variable value being converted into 0, and the deviations being normalized to 1.

$$X_i(j) = \left(x_i(j) - \bar{x}_i \right) \Big/ S_{x_i} \quad (9)$$

$$DIS(j,k) = \sum_{i=1}^{N} (X_i(j) - X_i(k))^2 \quad (10)$$

In Equations (9) and (10), i and j are bond ID numbers in a molecule, with i corresponding to a connecting bond and j to any other bond except a connecting bond. In Equation (9), $x_i(j)$ is the ith dependent variable value of bond j, $X_i(j)$ is a standardized variable value converted from $x_i(j)$, $\bar{x}_i$ is the average of the ith dependent variable x_i of all bonds in a molecule, and S_{x_i} is the standard deviation of x_i. In Equation (10), $DIS(j,k)$ is the dissimilarity between the bonds i and j, and N is the number of atoms in a molecule. The bond with the largest DIS value is recognized as the most dissimilar bond compared with the connecting bond. When more than one bond with the same largest DIS value is found, a bond is randomly extracted and others are discarded. We call the most dissimilar bond a 'nonconnecting bond'. When a product has more than one connecting bond, each corresponding nonconnecting bond is chosen but excludes the other connecting bonds.

The following procedure describes a strategy to extract statistical datasets from a reaction in ORGSYN[18] (Scheme 1 and Table 1). The product, diethyl malonate, was focused upon.



Scheme 1. A sample reaction in ORGSYN.[18]

a) All connecting bonds in the product were recognized from mapping information between reactants and products in the scheme (atoms 2 and 5 in Scheme 1 and bond j = 4 in Table 1).

b) After adding hydrogen atoms, BC, BDE, and NCC were calculated for all bonds [$x_i(j)$ in Table 1].

c) BC, BDE, and NCC were standardized to sBC, sBDE, and sNCC by Equation (9) [$X_i(j)$ in Table 1].

d) The dissimilarities of the connecting bond (j = 4 in Table 1), $DIS(j,4)$, were calculated by Equation (10).

e) Bonds with the largest $DIS(j,4)$ (= 9.341) value were recognized (j = 14, 15, 16, 20, 21, and 22 in Table 1) and one bond (j = 4) was extracted as a nonconnecting bond. Then the connecting (j = 4) and nonconnecting bonds (j = 14) were used for LoRA (gray frame in Table 1).

The statistical parameter P, which corresponds to $P_{(event)}$ in Equation (6), is regarded as the equation for estimating retrosynthetically important bonds. The scores of P are comparable for bonds in other molecules. As the scores of the bonds get closer to 1, the bonds are recognized as more useful for retrosynthesis. Furthermore, it is possible to regard bonds as unworthy of consideration when the scores are less than 0.5.

Table 1. Bond properties of diethyl malonate in Scheme 1.[a]

Bond	Atoms		RC[b]	$x_i(j)$			$X_i(j)$			$DIS(j,4)$
i	atom1	atom2		BDE[c]	BC[d]	NCC[e]	sBDE[f]	sBC[f]	sNCC[f]	
1	1	2	0	90.81	3.627	0	-0.143	1.734	0.000	0.000
2	1	3	0	43.44	3.119	0	-1.601	1.096	0.000	2.531
3	1	4	0	25.66	2.853	0	-2.148	0.762	0.000	4.962
4	*2*	*5*	*1*	*90.81*	*3.627*	*0*	*-0.143*	*1.734*	*0.000*	*0.000*
5	2	12(H)	0	107.87	3.101	0	0.382	1.074	0.000	0.711
6	2	13(H)	0	107.87	3.101	0	0.382	1.074	0.000	0.711
7	3	6	0	92.83	2.441	0	-0.081	0.245	0.000	2.221
8	5	7	0	43.44	3.119	0	-1.601	1.096	0.000	2.531
9	5	8	0	25.66	2.853	0	-2.148	0.762	0.000	4.962
10	6	9	0	96.60	1.836	0	0.035	-0.514	0.000	5.087
11	6	14(H)	0	104.55	1.810	0	0.279	-0.547	0.000	5.382
12	6	15(H)	0	104.55	1.810	0	0.279	-0.547	0.000	5.382
13	7	10	0	92.83	2.441	0	-0.081	0.245	0.000	2.221
14	9	16(H)	0	127.95	1.369	0	0.999	-1.101	0.000	**9.341**
15	9	17(H)	0	127.95	1.369	0	0.999	-1.101	0.000	**9.341**
16	9	18(H)	0	127.95	1.369	0	0.999	-1.101	0.000	**9.341**
17	10	11	0	96.60	1.836	0	0.035	-0.514	0.000	5.087
18	10	19(H)	0	104.55	1.810	0	0.279	-0.547	0.000	5.382
19	10	20(H)	0	104.55	1.810	0	0.279	-0.547	0.000	5.382
20	11	21(H)	0	127.95	1.369	0	0.999	-1.101	0.000	**9.341**
21	11	22(H)	0	127.95	1.369	0	0.999	-1.101	0.000	**9.341**
22	11	23(H)	0	127.95	1.369	0	0.999	-1.101	0.000	**9.341**

[a] Gray rectangular areas represent extracted data sets. [b] 1 refers to a reaction center bond and 0 refers to a nonreaction center bond. [c] Bond dissociation energy [kcal/mol]. [d] Bond centrality. [e] The number of chiral centers in a bond. [f] Standardized variables.

Table 2. Statistics of datasets obtained from ORGSYN and C40–80.

Data		ORGSYN		C40-80	
		Connecting bond	Non-connecting bond	Connecting bond	Non-connecting bond
Num. of reactions in DB		5690		3989	
Num. of available reactions		2990		1333	
Num. of conv. datasets		3654	3654	1685	1685
NB.[a] of structures in conv. datasets	Min.	4		75	
	Max.	150		248	
	Av.[b]	27.39		127.44	
	SD.[c]	12.79		26.09	
BC	Min.	0.90	0.87	0.77	0.73
	Max.	4.69	4.61	5.33	4.92
	Av.	2.93	2.03	3.23	2.46
	SD.	0.82	0.97	0.89	1.23
BDE[d]	Min.	2.46	1.42	7.23	3.30
	Max.	257.39	827.92	474.88	2441.20
	Av.	59.75	99.25	63.64	135.20
	SD.	31.77	56.48	48.76	152.31
NCC	Min.	0	0	0	0
	Max.	2	2	2	2
	Av.	0.09	0.11	0.33	0.75
	SD.	0.33	0.39	0.50	0.89

[a] The number of bonds including bonds with hydrogen atoms. [b] Average. [c] Standard deviation. [d] In kcal/mol. Although the BDEs of a few bonds in the structures could not be predicted precisely, they were included in the statistical datasets.

All the reactions in ORGSYN[19] and other reactions with product yields of ≥90%, including products with 40–80 carbon atoms in CCR[20] and REFLIB,[21] designated C40–80, were converted into two datasets. With two datasets, equations for LoRA were generated and these were compared with molecular complexity indices and tested with 35 organic compounds described in Nicolaou's total synthesis textbook.[22]

Results and Discussions

Collection of Statistical Datasets from ORGSYN and C40–80

From 5690 reactions in ORGSYN and 3989 reactions in C40–80, 2990 and 1333 products, respectively, were taken out and converted into 7308 (= 3654 × 2) and 3370 (= 1685 × 2) datasets for LoRA (Table 2). The frequency distributions of the number of bonds in ORGSYN and C40–80 are shown in Figure 1 and Figure 2. The average numbers of bonds in the ORGSYN and C40–80 datasets were 27.4 and 127.4. The distribution of the numbers of bonds is clearly different for the two datasets.

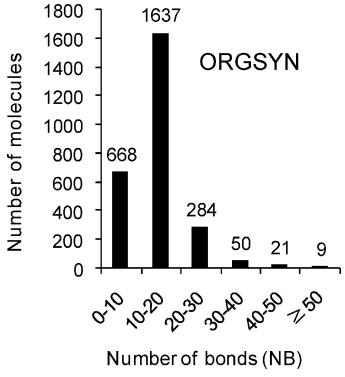


Figure 1. Frequency distribution of the number of bonds in each product in ORGSYN.

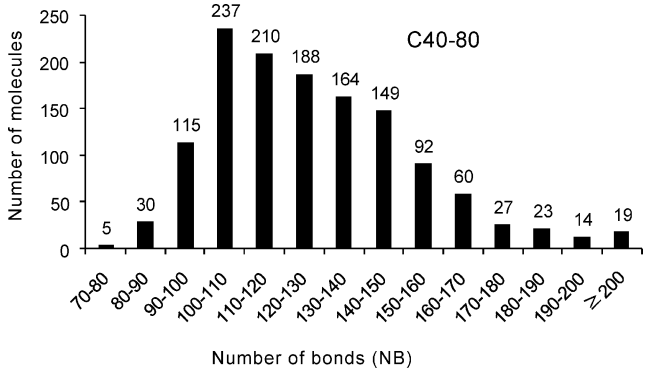


Figure 2. Frequency distribution of the number of bonds in each product in C40–80.

The frequency distributions of three independent variable values in datasets from ORGSYN and C40–80 are shown in Figures 3–8. The distribution shapes for BC are similar for the two databases and are clearly dissimilar be-

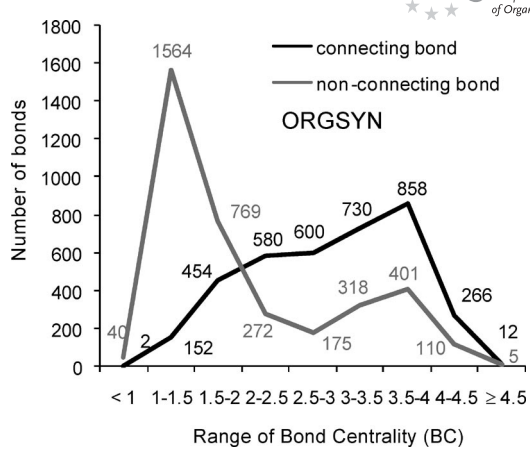


Figure 3. Frequency distributions of connecting (black line) and nonconnecting (gray line) bonds versus ranges of BC in datasets from ORGSYN.

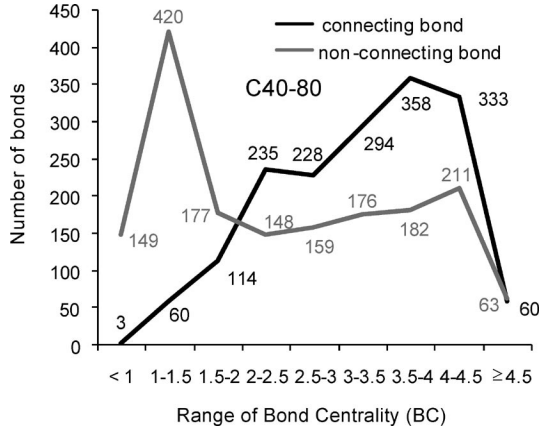


Figure 4. Frequency distributions of connecting (black line) and nonconnecting (gray line) bonds versus ranges of BC in datasets from C40–80.

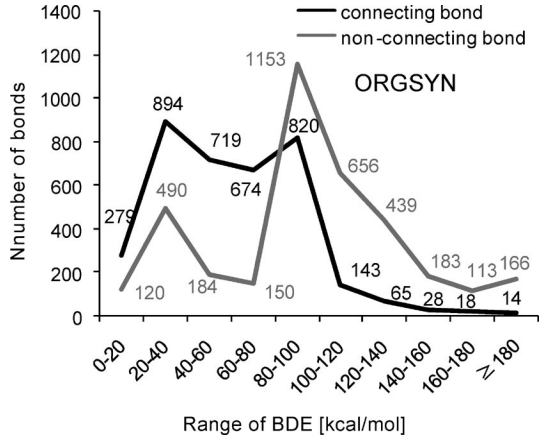


Figure 5. Frequency distributions of connecting (black line) and nonconnecting (gray line) bonds versus ranges of BDE in datasets from ORGSYN.

tween connecting and nonconnecting bonds (Figures 3 and 4). Although the most frequent distribution of connecting bonds was in the range 3.5–4, nonconnecting bonds were in the range 1–1.5. Thus, connecting bonds have a tendency to be located more centrally within molecules. Although BDE distributions were moderately different for two datasets, it was found that connecting bonds have lower BDEs than nonconnecting bonds (Figures 5 and 6). For NCC, both connecting and nonconnecting bonds are similar except for NCC with a value of 2 in the datasets from C40–80 (Figures 7 and 8).

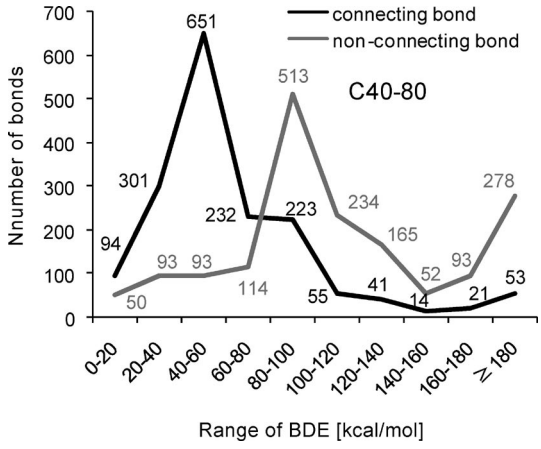


Figure 6. Frequency distributions of connecting (black line) and nonconnecting (gray line) bonds versus ranges of BDE in datasets from C40–80.

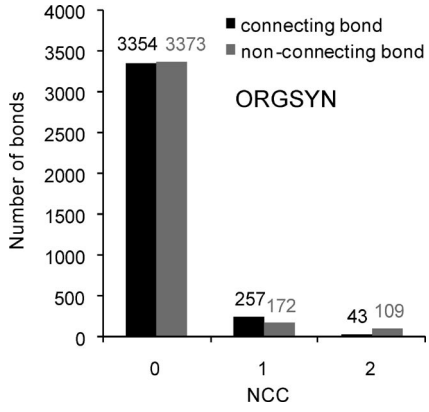


Figure 7. Frequency distributions of the number of connecting (black bars) and nonconnecting (gray bars) bonds versus NCC in datasets from ORGSYN.

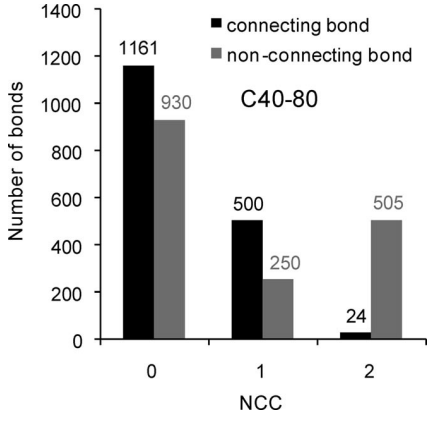


Figure 8. Frequency distributions of the number of connecting (black bars) and nonconnecting (gray bars) bonds versus NCC in datasets from C40–80.

Results of LoRA with Datasets from ORGSYN and C40–80

For the purpose of finding the most suitable evaluation equation, eight equations, P1–P8, were constructed by using a combination of two or three of the independent variables BC, BDE, and NCC from ORGSYN and C40–80 (Table 3).

In Table 3, Cox and Snell's R^2 is an attempt to imitate the interpretation of multiple R^2 based on their likelihood.[23] Nagelkerke's R^2 is a further modification of Cox and Snell's R^2 to assure that it can vary from 0 to 1.[24] The predictive power is a percentage of the number of connecting bonds with scores ≥ 0.5 of all the connecting bonds and the percentage of nonconnecting bonds with scores <0.5 of all the nonconnecting bonds.

All of the coefficients in P1–P8 showed not only similar plus–minus signs but also quantities, except for the coefficient of NCC in P3, and R^2 and the predictive powers of P2 and P3, which are lower. The values of the coefficients of NCC from ORGSYN and C40–80 were different due to dataset dependency.

The coefficient signs of BC, BDE, and NCC were plus, minus, and minus, respectively. Higher evaluation scores, which are synonymous with higher priority bonds, are expected when the bonds are closer to the center of the molecules, have weaker BDE energies, and contain smaller numbers of chiral centers.

According to the R^2 values in Table 3, P1 and P5, which consist of two variables, BC and BDE, are the best evaluation equations in ORGSYN and C40–80, respectively. P4 and P8 also produce good values of R^2 but they consist of three variables, so P1 and P5 are the better equations. On the other hand, NCC was found to have a negative influence on the equations.

Figures 9 and 10 show the score distributions of the connecting and nonconnecting bonds in ORGSYN and C40–80 evaluated by P1 and P5, respectively. For ORGSYN, the evaluation equation P1 is expressed very well in the scores for the nonconnecting bonds, but not so well in the scores for the connecting bonds, for which the distribution is relatively broad in the range of 0.5–1. On the other hand, the distribution of scores for C40–80 by P5 draws a sharp contrast between the connecting and nonconnecting bonds. This tendency is also observed in the comparison of R^2 values for ORGSYN and C40–80 for any combination of variables (Figure 11).

The difference between P1 and P5 is attributed to the difference between reactions in ORGSYN and C40–80. Whereas ORGSYN contains not only reactions to synthe-

Table 3. Coefficients, R^2, predictive powers, and averages of the scores determined by LoRA of ORGSYN and C40–80.

Eq.	DB	Coefficient			Const.	R^2		Predictive power [%]			Average of score	
		BC	BDE	NCC		Cox & Snell	Nagelkerke	Connecting bond	Nonconnecting bond	Average	Connecting bond	Nonconnecting bond
P1	ORGSYN	0.9949	–0.0198	–	–0.9006	0.3118	0.4157	81.99	74.58	78.28	0.6708	0.3299
P2		0.8237	–	–0.7780	–1.9950	0.1483	0.1977	65.19	69.43	67.31	0.5773	0.4227
P3		–	–0.0204	0.0976	1.5646	0.1615	0.2154	68.20	70.96	69.58	0.5832	0.4169
P4		0.9468	–0.0202	–0.6289	–0.7295	0.2963	0.3950	80.27	73.13	76.70	0.6600	0.3389
P5	C40–80	1.2152	–0.0206	–	–1.2287	0.4274	0.5698	84.81	83.15	83.98	0.7456	0.2541
P6		1.5011	–	–2.4865	–2.7058	0.3900	0.5199	83.80	87.95	85.88	0.7279	0.2722
P7		–	–0.0197	–0.8606	2.2968	0.2999	0.3998	83.09	76.14	79.61	0.6685	0.3305
P8		1.3875	–0.0178	–2.0014	–1.2237	0.4255	0.5673	84.63	86.23	85.43	0.7473	0.2538

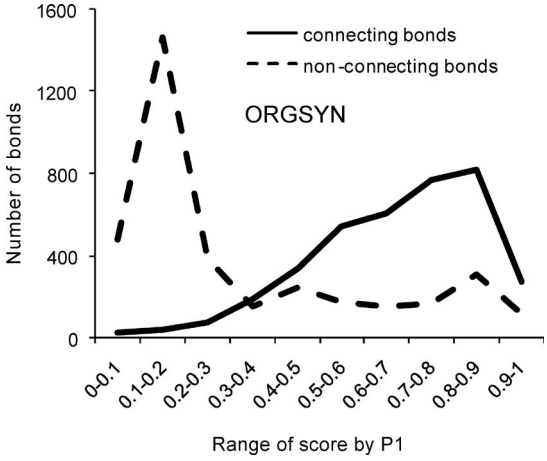


Figure 9. Frequency distributions of the evaluation scores of connecting and nonconnecting bonds in ORGSYN calculated by P1.

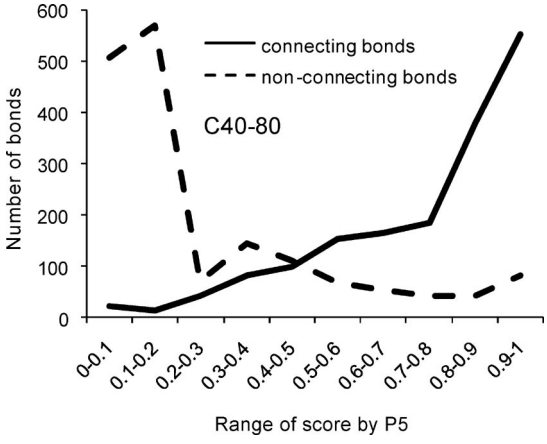


Figure 10. Frequency distributions of the evaluation scores of connecting and nonconnecting bonds in C40–80 calculated by P5.

size certain targets but also sample reactions to demonstrate the scope and limitations of certain catalysts, C40–80 contains only the former reactions because it only contains much larger product structures than ORGSYN. Hence, reactions in C40–80 are more suitable than those in ORGSYN for this statistical study.

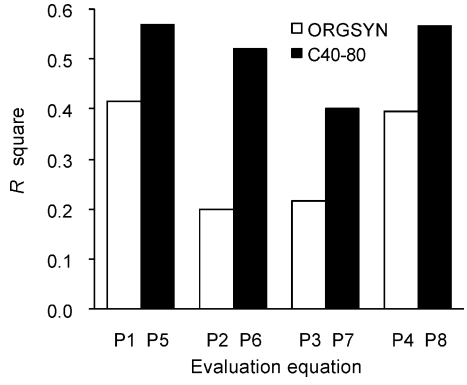


Figure 11. R^2 values from evaluation equations P1–P4 with ORGSYN and P5–P8 with C40–80.

To validate the evaluation equation model P5 statistically derived from the data set C40–80, another data set, ORGSYN, was used as a test set. Figure 12 shows the frequency distributions of connecting and nonconnecting bond scores in ORGSYN evaluated by P5. As is expected from the similar coefficients determined for P1 and P5 (Table 3), the distributions are similar to those of P1 in ORGSYN (Figure 9). The validated distribution shows that P5 is a reasonable and satisfactory equation for estimating retrosynthetically important bonds.

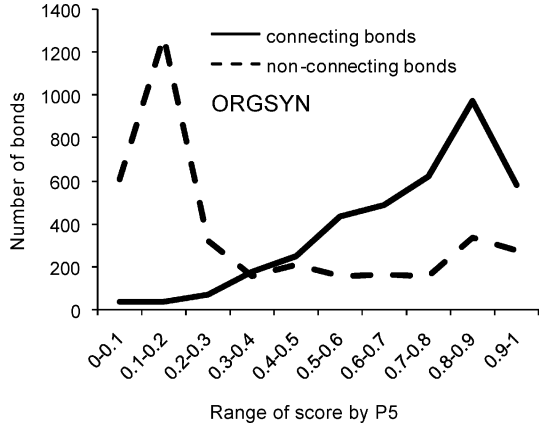


Figure 12. Frequency distributions of scores of connecting and nonconnecting bonds in ORGSYN evaluated by P5 for validation of the model.

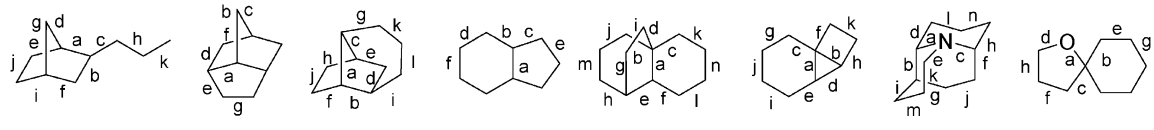


Figure 13. Structures **S1**–**S8** for comparison of P5 with Δtwc and ΔN_T.

Table 4. BC, BDE, P5, Δtwc, and ΔN_T values for each bond in **S1**–**S8**.

		BC	BDE[a]	P5	Δtwc	ΔN_T		BC	BDE	P5	Δtwc	ΔN_T
S1	a	3.868	69.96	0.884	24574	280	g	2.592	64.78	0.643	19412	218
	b	3.672	70.69	0.855	20170	244	h	2.767	82.96	0.605	6444	178
	c	3.728	79.39	0.841	14922	264	i	2.270	72.35	0.510	16698	214
	d	3.049	64.25	0.760	20870	227	j	2.159	73.78	0.469	13724	187
	e	2.998	71.63	0.719	18442	226	k	1.879	85.16	0.332	2292	90
	f	2.914	71.51	0.698	19300	232				R^2 vs. P5	0.5622	0.8134
S2	a	3.280	49.32	0.851	12178	456	e	2.774	65.06	0.690	8008	398
	b	3.005	50.61	0.799	9994	405	f	2.737	64.79	0.682	8470	396
	c	2.680	51.36	0.725	8774	385	g	2.448	66.99	0.590	5928	337
	d	2.883	64.30	0.721	9016	407				R^2 vs. P5	0.9604	0.8231
S3	a	3.483	22.78	0.927	36066	684	g	3.072	65.13	0.762	22858	634
	b	3.344	30.54	0.901	33444	692	h	2.680	45.59	0.748	22746	574
	c	3.344	36.14	0.890	3606	678	i	2.829	65.69	0.702	20418	616
	d	2.979	44.62	0.813	25792	602	j	2.233	47.50	0.624	17602	488
	e	2.979	44.74	0.813	26166	598	k	2.379	68.12	0.564	13104	498
	f	2.898	45.08	0.796	24834	594	l	2.275	68.25	0.533	12552	492
										R^2 vs. P5	0.9658	0.8625
S4	a	3.732	77.49	0.847	7930	176	d	2.585	81.93	0.556	3470	125
	b	3.322	79.29	0.764	5618	158	e	2.390	81.5	0.498	3870	117
	c	3.158	79.08	0.727	5728	149	f	2.257	82.55	0.453	2892	114
										R^2 vs. P5	0.9123	0.984
S5	a	3.936	82.71	0.864	343442	1696	h	2.658	81.15	0.582	158038	1368
	b	3.468	78.49	0.797	232292	1512	i	2.648	84.38	0.562	160310	1155
	c	3.468	78.62	0.797	225752	1558	j	2.490	83.07	0.522	121542	1154
	d	3.436	80.43	0.784	249516	1442	k	2.468	83.43	0.513	107578	1176
	e	3.447	81.88	0.781	271238	1624	l	2.313	83.92	0.463	95356	1144
	f	3.243	80.04	0.743	189472	1494	m	2.169	83.74	0.421	103784	1106
	g	2.791	82.84	0.612	180548	1314	n	2.006	84.73	0.369	76054	1038
										R^2 vs. P5	0.894	0.9254
S6	a	3.550	71.61	0.833	22834	400	g	2.736	80.46	0.608	6438	321
	b	3.352	71.42	0.798	23032	378	h	2.518	77.79	0.557	13664	333
	c	3.352	76.75	0.780	14902	424	i	2.518	80.97	0.541	5568	307
	d	3.225	71.79	0.771	20236	373	j	2.346	81.50	0.486	4320	282
	e	3.106	78.24	0.718	11852	396	k	2.198	79.41	0.452	10074	279
	f	3.032	76.05	0.709	16164	365				R^2 vs. P5	0.6922	0.8772
S7	a	3.566	72.28	0.834	2599114	1552	h	2.731	74.30	0.636	778316	1312
	b	3.407	70.83	0.810	1073232	1558	i	2.783	79.20	0.628	424422	1312
	c	3.355	73.23	0.793	2467144	1558	j	2.598	80.11	0.569	372246	1137
	d	3.080	73.95	0.729	824696	1369	k	2.285	75.43	0.498	643626	1072
	e	3.039	75.10	0.714	2064058	1369	l	2.321	81.76	0.477	330234	1072
	f	2.869	73.14	0.679	805362	1311	m	2.188	81.95	0.436	322346	1053
	g	2.926	78.03	0.672	444014	1311	n	2.183	81.95	0.434	322346	1053
										R^2 vs. P5	0.5618	0.9601
S8	a	3.518	47.22	0.888	45359	242	e	2.898	78.89	0.661	14142	196
	b	3.792	74.54	0.863	30785	256	f	2.460	79.56	0.530	17562	179
	c	3.613	83.84	0.808	31962	242	g	2.219	79.78	0.456	8733	166
	d	2.336	43.57	0.671	28067	179	h	1.960	87.21	0.345	16826	158
										R^2 vs. P5	0.6831	0.8766

[a] kcal/mol.

Comparison of P5 with Molecular Complexity Indices, *twc* and N_T

The evaluation score of P5 combining BC and BDE has been compared with molecular complexity indices, the total number of structures, N_T,[7c] and the total walk count, *twc*.[7d] It has already been reported that the indices have been proven to be successful for finding the simplest and thus most useful precursors.

Because P5 provides evaluation scores for each bond, the corresponding bond values related to the complexity indices have been defined by differences in the complexities between targets and the precursors produced by missing bonds, which are called Δtwc and ΔN_T respectively.

With the aim of a numerical comparison between P5 and the molecular complexities indices, Δtwc, and ΔN_T, eight structures (Figure 13) were selected from an article.[25] **S1**–**S8** are shown with their bonds labeled alphabetically, with the descending scores of P5 in each compound displayed in alphabetic order. Thus, bond 'a' has the highest score of P5 and 'b' is the second highest.

Table 4 lists bond centralities, BDE, scores of P5, Δtwc, and ΔN_T for all bonds in **S1**–**S8** and the square of the correlation coefficients R^2 between P5 and Δtwc, and between P5 and ΔN_T.

As shown in Figure 14 and Figure 15, P5 is linearly correlated with two complexity indices, which increase in proportion to the increase of P5. ΔN_T shows a good correlation with P5 that is independent of the structures, but Δtwc shows a relatively poor correlation with P5 for **S1**, **S7**, and **S8** (Table 4). The slopes of the lines are dependent on the structures because P5 is expressed as a sigmoidal function, which results in slopes in the range of 0 and 1. Meanwhile, Δtwc and ΔN_T were determined based on the number of bonds and branching in a structure. Comparison of the molecular complexities clearly demonstrates the usefulness of P5.

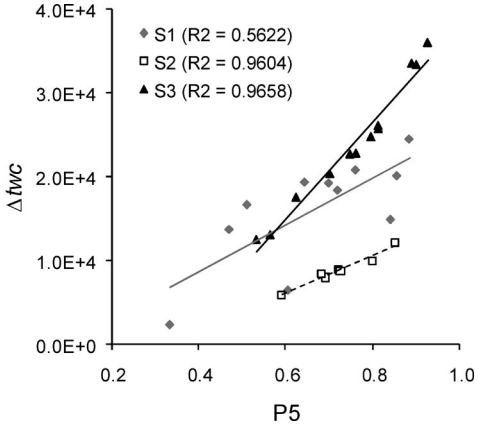


Figure 14. Plots of Δtwc vs. the P5 score for each bond in **S1**, **S2**, and **S3**.

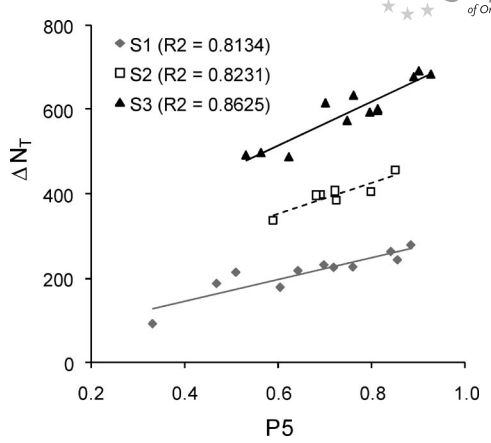


Figure 15. Plots of ΔN_T vs. the P5 score for each bond in **S1**, **S2**, and **S3**.

Examination of the Evaluation Score with Several Naturally Occurring Compounds

The concrete bond positions of the highest scores of P5, simple BC, and the last connecting bonds in the reported syntheses of the 35 organic compounds described in Nicolaou's total synthesis textbook[22] are shown in Figure 16. Based on the last connecting bond in each target, the performances of the evaluation equations P5–P8 (Table 3) were examined by comparison with simple BC.[8] Here, the last connecting bonds in the total synthesis did not include simple functional group conversions, bond order changes, or one-carbon elongation reactions because these connecting bonds do not contribute much to the construction of target skeletons, but they do contribute to the adjustment of target structures.

Table 5 reports the molecular weights, molecular formulae, the number of bonds (NB), and the three bond properties, BC, BDE, and NCC, of the 35 targets. Table 6 shows the ranking order sorted by the highest bond centralities in connecting reaction centers (HBCRC) and the scores of P5, and the corresponding relative rankings, which are the ranking orders divided by the number of bonds expressed in percentages. In addition, ΔRR is defined as the subtraction of the relative ranking of P5 from that of the HBCRC for each target. When ΔRR for the connecting bond becomes positive, P5 can evaluate the reported connecting bond as a retrosynthetically important bond more precisely than BC. As for the bonds with hydrogen atoms, BC, BDE, and NCC were calculated, but the bonds were not included as last connecting bonds.

The average of the relative rankings of P5 (16.9) is smaller than that of HBCRC (26.3) and the values of ΔRR are consistently positive except in five cases (Figure 17). For several structures, the estimation capability of P5 is considerably improved compared with BC.

The ΔRR values of the targets **5**, **15**, **16**, **26**, **33**, and **35** are more than 20.0 because the BDEs of the connecting bonds are relatively low, like carbon–heteroatom bonds. The coefficients of the BDEs in P5 indicate that a lower

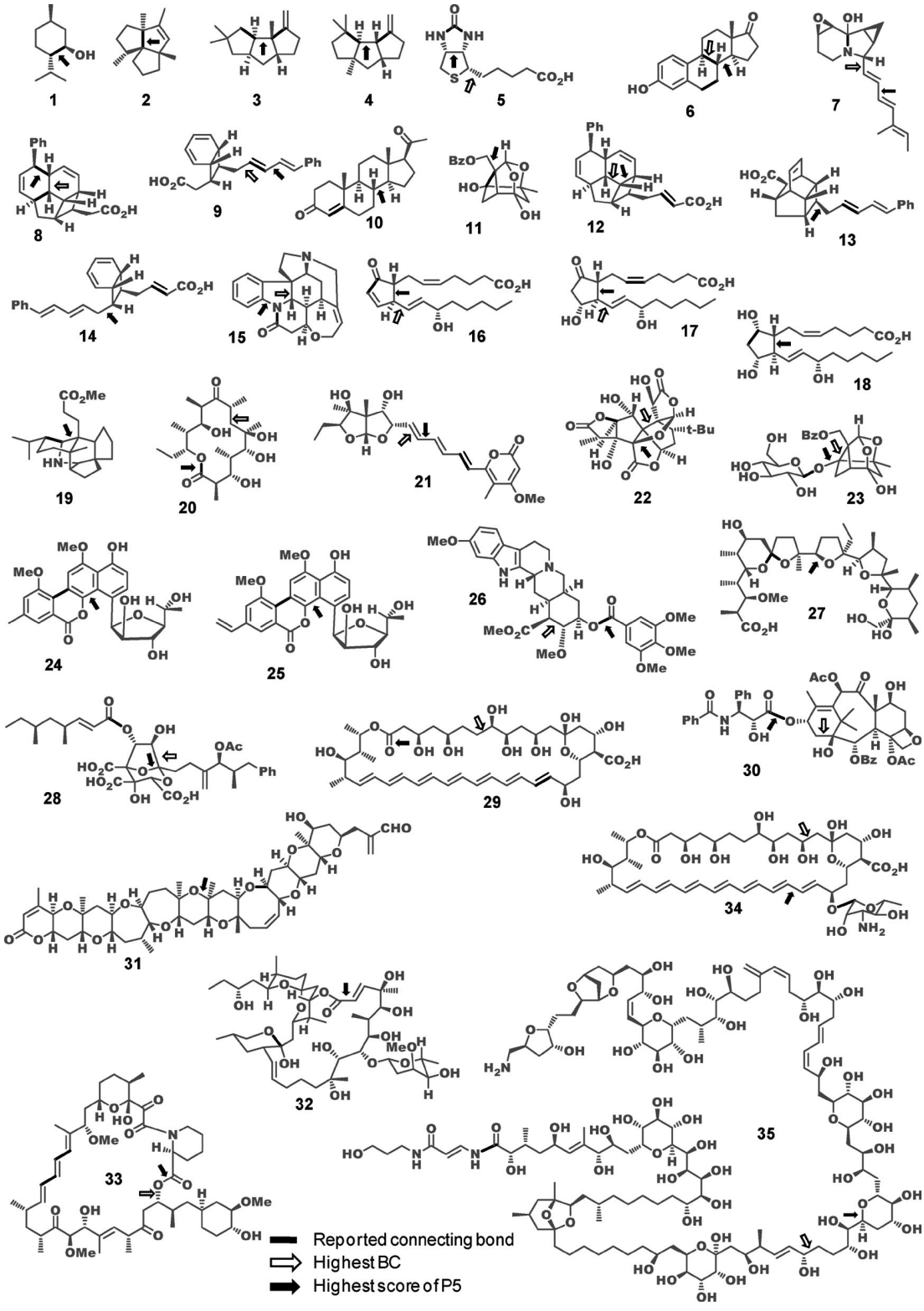


Figure 16. Targets from Nicolaou's total synthesis textbook.[22] The black bold bonds represent the reported connecting bonds, the black full arrows point to the highest score bonds of P5, and the open arrows point to bonds with the highest bond centralities. When bonds with the highest bond centralities are the same as the highest score bonds of P5, the open arrows are omitted.

BDE raises the score of P5 (Table 3). P5 recognizes the ester bonds in **20**, **26**, and **30** as the best bonds, in agreement with the reported connecting bonds, but BC does not recognize them. This suggests that BDE is a fundamental parameter for evaluating the importance of retrosynthetically important bonds.

Table 5. The list of targets and the properties of their connecting bonds in reaction centers.

ID	Compound	MW	Molecular formula	NB[a]	BDE[b]	BC[c]	NCC[d]
1	menthol	156.27	$C_{10}H_{20}O$	31	75.88	3.840	2
2	isocomene	204.36	$C_{15}H_{24}$	41	49.19	4.309	2
3	hirsutene	204.36	$C_{15}H_{24}$	41	73.50	3.752	1
					77.01	3.244	1
4	$\Delta^{9(12)}$-capnellene	204.36	$C_{15}H_{24}$	41	69.47	3.912	1
					78.59	3.183	1
5	biotin	244.31	$C_{10}H_{16}N_2O_3S$	33	27.76	1.849	0
					27.76	1.616	0
6	estrone	270.37	$C_{18}H_{22}O_2$	45	115.51	3.825	2
					115.88	2.813	0
7	indolizomycin	273.38	$C_{17}H_{23}NO_2$	46	80.57	2.876	0
8	endiandric acid A	306.40	$C_{21}H_{22}O_2$	49	50.60	3.354	2
					40.27	3.283	2
9	endiandric acid D	306.40	$C_{21}H_{22}O_2$	47	63.01	3.605	0
10	progesterone	314.47	$C_{21}H_{30}O_2$	56	59.27	1.952	0
11	paeoniflorigenin	318.33	$C_{17}H_{18}O_6$	45	64.86	3.526	0
12	endiandric acid B	332.44	$C_{23}H_{24}O_2$	53	68.34	1.521	0
13	endiandric acid C	332.44	$C_{23}H_{24}O_2$	53	63.00	3.364	0
14	endiandric acid F	332.44	$C_{23}H_{24}O_2$	51	68.35	1.907	0
15	strychnine	334.42	$C_{21}H_{22}N_2O_2$	53	34.19	2.553	0
16	prostaglandin A_2	334.46	$C_{20}H_{30}O_4$	54	18.58	2.638	0
17	prostaglandin E_2	352.47	$C_{20}H_{32}O_5$	57	81.59	2.592	0
18	prostaglandin $F_{2\alpha}$	354.49	$C_{20}H_{34}O_5$	59	79.98	2.768	1
19	methyl homoseco-daphniphyllate	359.55	$C_{23}H_{37}NO_2$	67	66.46	3.438	2
20	erythronolide B	402.53	$C_{21}H_{38}O_7$	66	44.25	2.829	0
21	asteltoxin	418.49	$C_{23}H_{30}O_7$	62	50.00	2.652	0
22	ginkgolide B	424.40	$C_{20}H_{24}O_{10}$	59	99.54	2.134	0
23	paeoniflorin	480.47	$C_{23}H_{28}O_{11}$	67	40.94	3.535	1
24	gilvocarcin M	482.49	$C_{26}H_{26}O_9$	65	23.49	3.249	0
25	gilvocarcin V	494.50	$C_{27}H_{26}O_9$	66	23.73	3.329	0
26	reserpine	608.69	$C_{33}H_{40}N_2O_9$	89	34.19	3.023	0
27	monensin	670.88	$C_{36}H_{62}O_{11}$	113	38.43	3.377	1
					39.18	3.008	1
28	zaragozic acid A	690.74	$C_{35}H_{46}O_{14}$	97	40.24	3.215	0
29	amphoteronolide B	778.93	$C_{41}H_{62}O_{14}$	118	53.76	2.142	0
30	taxol	853.92	$C_{47}H_{51}NO_{14}$	119	43.83	3.409	0
31	brevetoxin B	895.10	$C_{50}H_{70}O_{14}$	144	47.40	2.977	1
32	cytovaricin	901.14	$C_{47}H_{80}O_{16}$	147	48.35	2.671	1
33	rapamycin	914.19	$C_{51}H_{79}NO_{13}$	147	15.38	1.827	0
					15.38	1.802	0
34	amphotericin B	924.09	$C_{47}H_{73}NO_{17}$	140	53.05	2.305	1
35	palytoxin	2680.18	$C_{129}H_{223}N_3O_{54}$	418	29.18	1.064	0

[a] The number of bonds, including bonds with hydrogen atoms. [b] Bond dissociation energies of connecting bonds. [c] Bond centralities of connecting bonds. [d] The number of chiral centers of connecting bonds.

The reported connecting bond in **6** presents a large negative ΔRR value (–11.1). The bond in the B ring is a CH_2–CH_2 bond and so the BDE of the bond is higher than others. As a result, the score of P5 is lower than the BC.

The relative rankings of **12**, **14**, **22**, and **35** by P5 and BC are more than 50%, which means the reported connecting bonds are regarded as below-average bonds for retrosynthesis. The bonds are located on relatively terminal structures, and so the values of BC are lower. For **12** and **14**, depending on the perspective, the bonds correspond to functional group conversions and do not contribute to structure constructions. In fact, literature reports have shown that **12** and **14** have been synthesized from **8** and a stereoisomer of **9**, respectively.[26,27] Targets **22** and **35** are evidently too complicated to evaluate the bonds by P5 and BC.

Thirty of the targets (85.71%) have scores ≥0.5 by P5. This means that the last connecting bond in more than 85% of the targets are recognized as retrosynthetically important bonds.

As a result of the validation, P5 gave much better evaluation scores than BC. The equation consists of BC and BDE, with BDE used to indicate bond reactivity. Even though BDE is limited to indicating reaction reactivity, it improved the evaluation equation. Instead of BDE, more precise reactivity parameters, such as PETRA param-

Table 6. HBCRC, P5, and ΔRR of the connecting bond in each target.

ID	HBCRC[a] BC	HBCRC[a] Ranking order	Relative ranking [%][b]	P5 Score	P5 Ranking order	Relative ranking [%][b]	ΔRR[c]
1	3.840	1	3.2	0.867	1	3.2	0.0
2	4.309	1	2.4	0.952	1	2.4	0.0
3	3.752	3	7.3	0.860	3	7.3	0.0
4	3.912	3	7.3	0.890	3	7.3	0.0
5	1.849	21	63.6	0.610	7	21.2	42.4
6	3.825	1	2.2	0.739	6	13.3	–11.1
7	2.876	9	19.6	0.647	12	26.1	–6.5
8	3.354	3	6.1	0.873	1	2.0	4.1
9	3.605	3	6.4	0.865	3	6.4	0.0
10	1.952	34	60.7	0.481	25	44.6	16.1
11	3.526	3	6.7	0.848	2	4.4	2.3
12	1.521	38	71.7	0.313	42	79.2	–7.5
13	3.364	4	7.5	0.827	3	5.7	1.8
14	1.907	29	56.9	0.421	29	56.9	0.0
15	2.553	17	32.1	0.763	4	7.5	24.6
16	2.638	17	31.5	0.831	3	5.6	25.9
17	2.592	20	35.1	0.560	19	33.3	1.8
18	2.768	16	27.1	0.620	17	28.8	–1.7
19	3.438	6	9.0	0.829	6	9.0	0.0
20	2.829	9	13.6	0.786	1	1.5	12.1
21	2.652	18	29.0	0.724	8	12.9	16.1
22	2.134	32	54.2	0.335	39	66.1	–11.9
23	3.535	6	9.0	0.902	3	4.5	4.5
24	3.249	8	12.3	0.903	6	9.2	3.1
25	3.329	7	10.6	0.911	6	9.1	1.5
26	3.023	20	22.5	0.851	1	1.1	21.4
27	3.377	12	10.6	0.889	4	3.5	7.1
28	3.215	15	15.5	0.864	5	5.2	10.3
29	2.142	46	39.0	0.566	29	24.6	14.4
30	3.409	15	12.6	0.882	1	0.8	11.8
31	2.977	27	18.8	0.804	9	6.3	12.5
32	2.671	34	23.1	0.735	6	4.1	19.0
33	1.827	102	69.4	0.663	16	10.9	58.5
34	2.305	52	37.1	0.618	26	18.6	18.5
35	1.064	361	86.4	0.369	209	50.0	36.4
Average			26.3			16.9	9.4

[a] Highest bond centrality in connecting reaction centers. [b] (Ranking order/NB) × 100. [c] Subtraction of relative ranking of P5 from that of HBCRC.

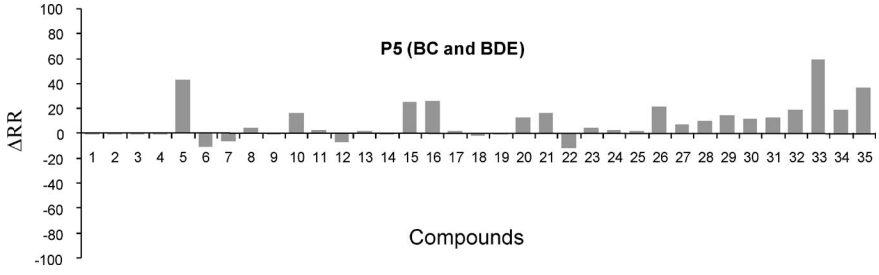


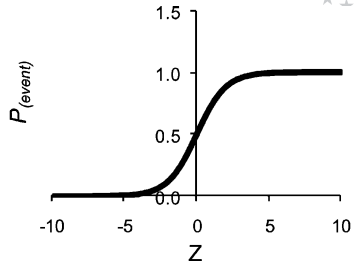
Figure 17. ΔRR for the last connecting bonds in the 35 targets.

eters,[28] which are physicochemical values such as polarizability[29] developed by Gasteiger and co-workers, may have the potential to improve evaluation equations.

Conclusion

We have developed a new evaluation equation for retrosynthetically important bonds, which has been successfully built from reaction databases by means of logistic regression analysis (LoRA). Connecting bonds of products in the reaction schemes were assumed to be one of the best candidate bonds for retrosynthesis. With three bond properties, BC, BDE, and NCC, of the connecting and nonconnecting bonds, LoRA has been performed to survey important factors and their contributions. The best statistical equation, P5, which consists of BC and BDE, predicted the

reported connecting bonds in reaction centers as retrosynthetically important bonds more precisely than simple BC. The coefficients and constants in the equations were quite similar in spite of different source reaction databases. They showed that synthetically useful bonds tended to be more central in a molecule and to be weaker bonds. In addition, the scores are quantitative values with important bonds being distinguished from others when the score is ≥0.5.



[1] E. Zass, *Encyclopedia of Computational Chemistry* (Ed.: P. v. R. Schleyer), Wiley, Chichester, **1998**, pp. 2402–2420.

[2] E. J. Corey, W. T. Wipke, *Science* **1969**, *166*, 178–192.

[3] E. J. Corey, A. Petersson, *J. Am. Chem. Soc.* **1972**, *94*, 460–465.

[4] a) R. Barone, M. Chanon, *Encyclopedia of Computational Chemistry* (Ed.: P. v. R. Schleyer), Wiley, Chichester, **1998**, pp. 2931–2948; b) R. Baron, M. Chanon, *Handbook of Chemoinformatics* (Ed.: J. Gasteiger), Wiley-VCH, Weinheim, **2003**, vol. 4, pp. 1428–1456; c) M. Pfoerter, M. Sitzmann, *Handbook of Chemoinformatics* (Ed.: J. Gasteiger), Wiley-VCH, Weinheim, **2003**, vol. 4, pp. 1457–1507; d) M. H. Todd, *Chem. Soc. Rev.* **2005**, *34*, 247–266.

[5] a) E. J. Corey, *Q. Rev. Chem. Soc.* **1971**, *25*, 455–482; b) E. J. Corey, W. J. Howe, H. W. Orf, D. A. Pensak, G. Petersson, *J. Am. Chem. Soc.* **1975**, *97*, 6116–6124; c) E. J. Corey, X. M. Cheng, *The Logic of Chemical Synthesis* Wiley, New York, **1989**.

[6] S. H. Bertz, *J. Am. Chem. Soc.* **1981**, *103*, 3599–3601.

[7] a) J. B. Hendrickson, P. Huang, A. G. Toczko, *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 63–67; b) P. A.Wender, B. L. Miller, *Organic Synthesis Theory and Applications* (Ed.: T. Hudlicky), JAI Press, **1993**, p. 27–66; c) S. H. Bertz, T. J. Sommer, *Chem. Commun.* **1997**, 2409–2410; d) G. Rücker, G. Rücker, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 99–106; e) M. Randić, D. Plavšić, *Croat. Chem. Acta* **2002**, *75*, 107–116; f) R. Barone, M. Petitjean, C. Baralotto, M. Chanon, *J. Phys. Org. Chem.* **2003**, *16*, 9–15; g) H. W. Whitlock, *J. Org. Chem.* **1998**, *63*, 7982–7989.

[8] A. Tanaka, T. Kawai, M. Fujii, T. Matsumoto, T. Takabatake, H. Okamoto, K. Funatsu, *Tetrahedron* **2008**, *64*, 4602–4612.

[9] a) J. Gasteiger, W. D. Ihlenfeldt, R. Fick, J. R. Rose, *J. Chem. Inf. Sci.* **1992**, *32*, 700–712; b) J. Gasteiger, W. D. Ihlenfeldt, P. Röse, *Recl. Trav. Chim. Pays-Bas.* **1992**, *111*, 270–290; c) W. D. Ihlenfeldt, J. Gasteiger, *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 2613–2633.

[10] J. Gasteiger, M. Marsili, M. G. Hutchings, H. Saller, P. Löw, P. Röse, K. Rafeiner, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 467–476.

[11] NCC is the number of chiral center atoms in a bond. The chiral information was obtained from chiral flags, 1 or 2, in the atom block formatted by the SDFile converted from ISIS reaction databases.

[12] a) SPSS was used for analyses: http://www.spss.com; b) A. Agresti, *Categorical Data Analysis*, Wiley-Interscience, New York, **2002**; c) the sigmoidal curve of Equation (6) is as follows:

[13] E. Vittinghoff; *Regression Methods In Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*, Springer, New York, **2004**.

[14] S. Manel, J.-M. Diasb, S. J. Ormerdoc, *Ecol. Model.* **1999**, *120*, 337–347.

[15] S. W. Menard, *Applied Logistic Regression Analysis (Quantitative Applications in the Social Sciences)*, 2nd ed., Sage, USA, **2001**.

[16] J. Gasteiger, K.-P. Schulz, C. Kredler, *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 385–394.

[17] J. Gasteiger, W. Hanebeck, K.-P. Schulz, *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 264–271.

[18] M. Lang, S. Lang-Fugmann, W. Steglich, *Org. Synth.* **2000**, *78*, 113–122.

[19] ORGSYN contains new general synthetic methods and proven compound preparations: http://www.orgsyn.org/. The source is *Organic Syntheses*, 1980–present, published by Wiley.

[20] Current Chemical Reactions covers the latest synthetic methods reported in the world's leading organic chemistry journals, providing access to over 400000 reactions: http://scientific.thomson.com/products/ccr/.

[21] The Reference Library of Synthetic Methodology database contains a broad collection of novel methods for organic synthesis, abstracted in the chemical literature from 1900 to 1991: http://www.mdl.com/products/knowledge/reflib_synthetic_meth/.

[22] K. C. Nicolaou, E. J. Sorensen, *Classics in Total Synthesis*, VCH, Weinheim, **1996**. All naturally occurring compounds in the textbook have been investigated except for sugars, and unpredictable compounds about BDE, periplanone B, thienamycin, penicillin V, carpanone, and calicheamicin γ^{I}_{1}.

[23] J. E. Snell, *Applied statistics, principles and examples*, Chapman & Hall, CRC Press, Florida, **1981**.

[24] N. J. D. Nagelkerke, *Biometrika* **1991**, *78*, 691–692.

[25] C. Rücher, G. Rücker, S. H. Bertz, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 378–386.

[26] K. C. Nicolaou, N. A. Petasis, R. E. Zipkin, J. Uenishi, *J. Am. Chem. Soc.* **1982**, *104*, 5555–5557.

[27] C. K. Nicolaou, N. A. Petasis, J. Uenishi, R. E. Zipkin, *J. Am. Chem. Soc.* **1982**, *104*, 5557–5558.

[28] a) J. Gasteiger, M. Marsili, *Tetrahedron* **1980**, *36*, 3219–3228; b) J. Gasteiger, M. G. Hutchings, *Tetrahedron Lett.* **1983**, *24*, 2537–2540; c) The PETRA program package is opened through the Internet: http://www2.ccc.uni-erlangen.de/software/petra/.

[29] J. Gasteiger, M. G. Hutchings, *J. Am. Chem. Soc.* **1984**, *106*, 6489–6495.

Received: April 19, 2008
Published Online: October 31, 2008

