Available online at www.sciencedirect.com
ScienceDirect
Tetrahedron 64 (2008) 4602–4612
Tetrahedron
www.elsevier.com/locate/tet

Molecular centrality for synthetic design of convergent reactions

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

[a] *Organic Synthesis Research Laboratory, Sumitomo Chemical Co., Ltd, 1-98, Kasugade-naka 3-chome, Konohana-ku, Osaka 554-8558, Japan*
[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*

Received 17 December 2007; received in revised form 3 March 2008; accepted 3 March 2008
Available online 7 March 2008

Abstract

With convergent synthesis in mind, we defined a new parameter to quantify the degree of centrality of each atom and bond in a molecule, so-called molecular centrality, which was defined based on squared node distances. The centrality becomes higher as the location of atoms gets closer to the center of a molecule. From the results of validation with 40 organic compounds reported about their total syntheses, it became clear that our molecular centrality was an effective index to evaluate synthetically important bonds. Additionally, it was confirmed that the highest attaching bond centrality of each product moderately correlated with molecular complexity changes of each step. The parameter is quantitative among molecules and suitable for statistical analysis.

Keywords: Synthetic route design; Synthesis design system; Retrosynthesis; AIPHOS; Molecular centrality; Atom centrality; Bond centrality; Molecular complexity

1. Introduction

The first computer-aided synthesis design system OCSS was published in 1969,[1] which was the predecessor of LHASA.[2] After this publication, many systems have been reported and some of them are still being developed.[3] We also have been developing a synthesis design system AIPHOS,[4] which was first published in 1986.[4b] By compiling a database and a knowledgebase from organic reactions reported in journals, the system proposes plausible reactions and precursors for an inputted target. Chemical companies are looking forward to apply such a system for industrial synthetic processes of new electronic materials, and pharmaceutical and agricultural compounds.

In the development of industrial processes, choice of starting materials is one of the most important works. For meeting the request, most of the synthesis design systems have been designed to continuously generate precursors until reaching practical and commercially available compounds. If a target is small enough to be synthesized in a few steps, it is possible that an exhaustive search for possible synthetic routes is finished in practical time. For a larger target, however, the number of proposed routes will increase drastically and sometimes end up with a meaningless result, and at the same time the execution time grows drastically too. As a matter of course, an efficient search strategy for synthetically important bonds is required.

In 1971, about bond selection for efficient retrosynthesis, Corey proposed the first heuristics strategy to perceive strategic bonds in LHASA.[5] In 1981, Bertz demonstrated a sophisticated concept, 'molecular complexity' that was a quantitative approach[6a] and has already been widely recognized.[6] According to the concept of molecular complexity, the retrosynthetically best bond in a molecule is the largest decrease of molecular

* Corresponding authors.
E-mail addresses: tanakaa1@sc.sumitomo-chem.co.jp (A. Tanaka), funatsu@chemsys.t.u-tokyo.ac.jp (K. Funatsu).

0040-4020/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tet.2008.03.007

complexity on disconnection of each bond. Some synthesis design systems, such as SYNGEN[7] and CONAN,[8] load the concept to evaluate efficiencies of bond disconnections in the targets. This evaluation requires disconnection of each bond in order to calculate differences of molecular complexities from the corresponding precursors.

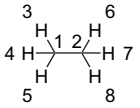
Besides molecular complexity, we have been seeking for a similar quantitative property that is easily calculated without structure transformation such as bond disconnection, at the same time, the property is hopefully comparable to molecular complexity.

We have therefore developed a new topological property with convergent synthesis in mind. The index is able to quantify the degrees of location in a molecule, and defined on each atom and bond. The property becomes higher as the atoms get closer to the center of the molecule. For the property, we named atom and bond centrality, and molecular centrality[9] as the generic name. With the use of reported silphinene syntheses, molecular centralities in the reaction centers of products were compared with the differences of molecular complexities in each step. In addition, using 40 complicate organic compounds whose total syntheses have already been reported, molecular centralities of the last attaching bond for each molecule were studied. In this paper, we focused on topological parameters exploring retrosynthetically important bonds without taking stereochemistry into consideration.

2. Definition of molecular centrality

Convergent synthesis is a standard strategy to improve the efficiency of multi-step synthesis of a complicated organic molecule. A few segments are prepared independently to connect together at almost the last step, hence the last building bond is often located closer to the center of the target.

For purpose of identifying locations of atoms in a molecule, we tried to use the sum of node distances[10] from all of other atoms, Σdist. (The following table summarizes node distances, squared node distances, and atom centralities of ethane, Ref. 10.)



i	dist(i,j)								dist^2(i,j)								$D(i)$	$D'(i)$	Atom centrality(i)
	j	1	2	3	4	5	6	7	8	j	1	2	3	4	5	6	7	8			
1		0	1	1	1	1	2	2	2		0	1	1	1	1	4	4	4	16	2.5	1.818
2		1	0	2	2	2	1	1	1		1	0	4	4	4	1	1	1	16	2.5	1.818
3		1	2	0	2	2	3	3	3		1	4	0	4	4	9	9	9	40	1	0.727
4		1	2	2	0	2	3	3	3		1	4	4	0	4	9	9	9	40	1	0.727
5		1	2	2	2	0	3	3	3		1	4	4	4	0	9	9	9	40	1	0.727
6		2	1	3	3	3	0	2	2		4	1	9	9	9	0	4	4	40	1	0.727
7		2	1	3	3	3	2	0	2		4	1	9	9	9	4	0	4	40	1	0.727
8		2	1	3	3	3	2	2	0		4	1	9	9	9	4	4	0	40	1	0.727

NUM_ATOM=8, D_{max}=40, D'_{av}=1.375.

Hydrogen atoms were included in node distances and in reaction centers. As the values of Σdist of atoms became smaller, the atoms got closer to the center of a molecule. However, it was found that Σdist sometimes shows the same numeric number for nonequivalent atoms. For example, *n*-hexane contains two sorts of atoms with the value 64, where one is on terminal carbons and another is on hydrogen atoms of 3- or 4-position carbons (Fig. 1, **1a**). To reduce the duplication and to emphasize the differences among atoms, we decided to use the sum of squared node distances $D(i)$ as shown in Eq. 1 (**1b**).

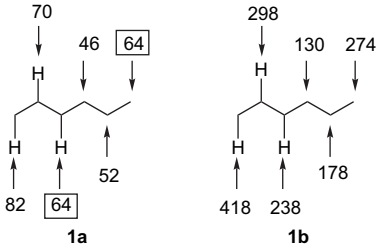


Figure 1. The sum of node distances and squared node distances for each atom. The values of equivalent atoms are omitted.

$$D(i) = \sum_{j=1}^{\text{all}} \text{dist}^2(i,j) \tag{1}$$

In Eq. 1, i and j mean node numbers of atoms in a molecule, and dist^2(i,j) is the squared node distance between atoms i and j.

Atoms on the edge of a molecule had the maximum D_{max}. For each atom, $D(i)$ of a atom i was divided by D_{max} to

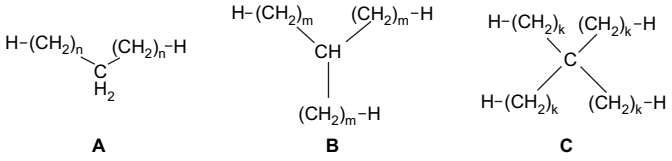


Figure 2. Saturated hydrocarbons **A** (n=0–39), **B** (m=0–29), and **C** (k=0–19) to study distributions of atom centralities.

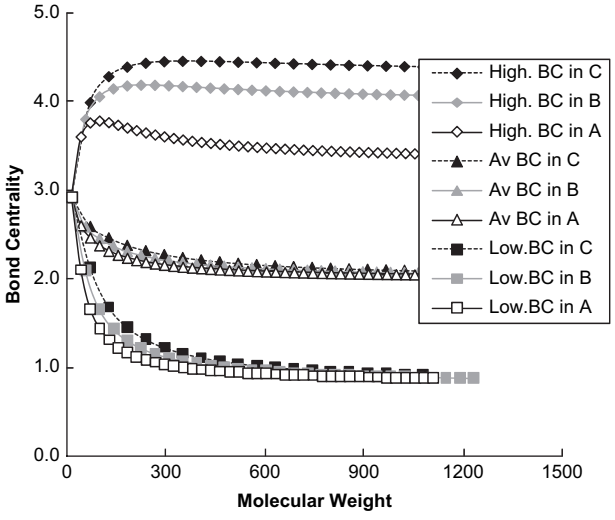


Figure 3. The line and dashed line show dependences of three kinds of bond centralities as for all bonds in **A**, **B**, and **C** on their molecular weights. The three kinds of bond centralities are highest (High.), lowest (Low.), and average (Av) bond centralities (BC).

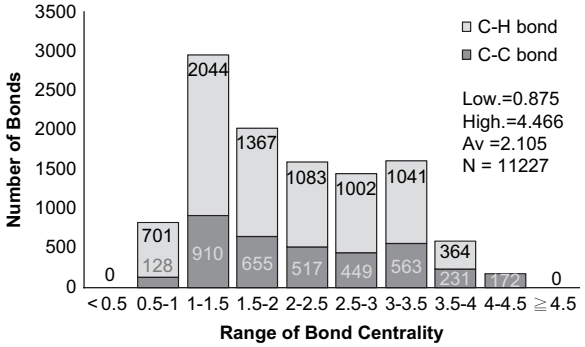


Figure 4. The bar chart shows a frequency distribution of all bond centralities in **A**, **B**, and **C**. The lowest, highest, and average of C−H bonds (N=7602) were 0.875, 3.960, and 2.027, and the lowest, highest, and average of C−C bonds (N=3625) were 0.917, 4.466, and 2.267.

estimate the ratio of $D(i)$ to $D_{\max}$. Additionally the reciprocal number of the ratio $D'(i)$ was prepared to give larger values to more central atoms in a molecule (Eq. 2). Sequentially, $D'(i)$ was normalized by its average value D'_{av} calculated from Eq. 3. The normalized $D'(i)$ was defined as atom centrality (Eq. 4).[10]

$$D'(i) = \frac{1}{\frac{D(i)}{D_{\max}}} = \frac{D_{\max}}{D(i)} \tag{2}$$

$$D'_{\text{av}} = \sum_{i=1}^{\text{all}} \frac{D'(i)}{\text{NUM\_ATOM}} \tag{3}$$

$$\text{Atom centrality}(i) = \frac{D'(i)}{D'_{\text{av}}} \tag{4}$$

In Eq. 3, NUM_ATOM is the number of atoms including hydrogen atoms in a molecule. Bond centrality was defined by summation of two edge atoms' centralities as shown in Eq. 5.

$$\text{Bond centrality}(i,j) = \text{Atom centrality}(i) + \text{Atom centrality}(j) \tag{5}$$

Molecular centrality consists of atom and bond centrality. The parameters are in numerical rating scale to indicate the degree of center in a molecule.

In linear, tertiary, and quaternary saturated hydrocarbons, **A**, **B**, and **C** (Fig. 2), the highest, lowest, and average bond centralities were plotted against molecular weights as shown in Figure 3.

In **A**, **B**, and **C**, bond with the highest bond centrality was four central C−C bonds in **C**, and the next was three central C−C bonds in **B**, followed by two central C−C bonds in **A**. For example, in molecules with molecular weights around 600, the highest centrality of **C** is 4.449 (k=10, MW=577.12), **B** is 4.126 (m=15, MW=605.17), and **A** is 3.473 (n=21, MW=605.17). The lowest bond centralities were on terminal C−H of carbon chains in **A**, **B**, and **C**. As molecular weights increased, the lowest and average came close to the values 2 and 1. In addition, when the number of carbons was around or more than 15, the lowest and average bond centralities became nearly constant and independent on molecular weights.

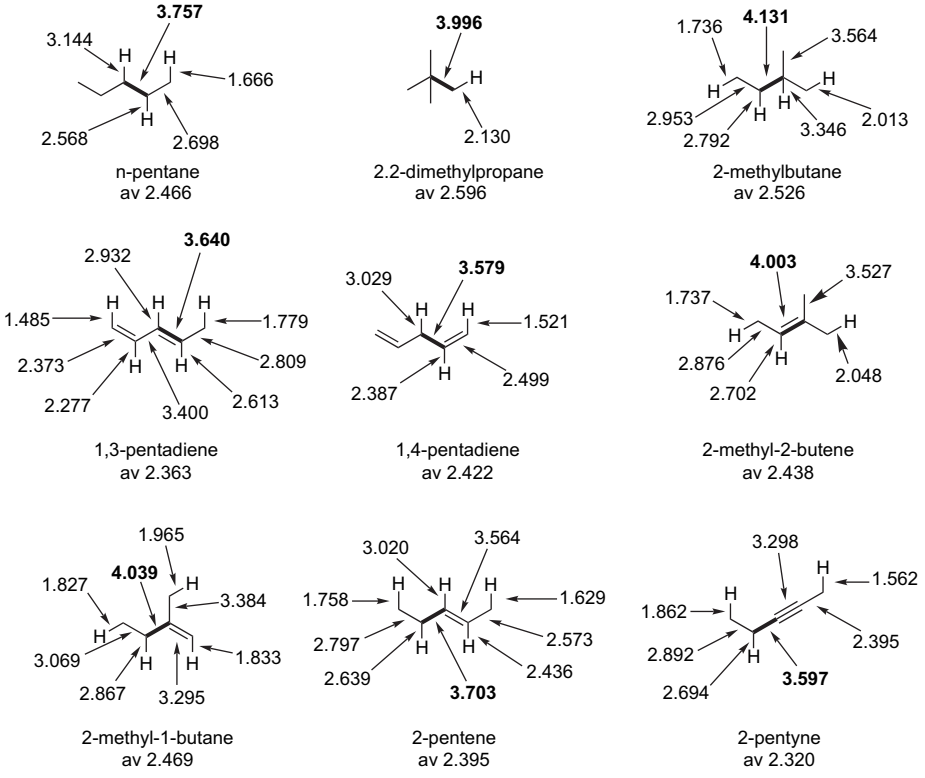
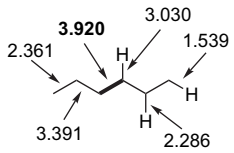
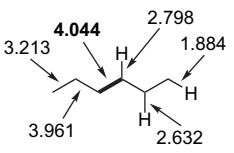
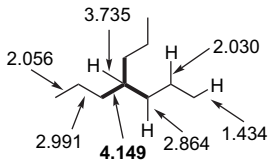
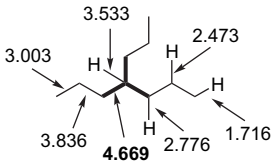
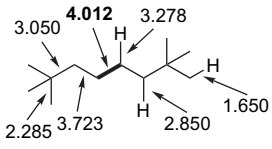
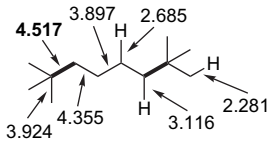
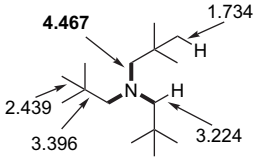
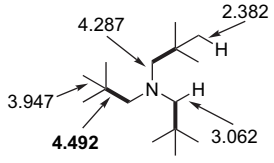


Figure 5. Bond centralities of C5-hydrocarbons. Each bold line indicates the highest centrality bond in each molecule.

Table 1
Bond centrality and ECBC in **2**, **3**, **4**, and **5**

Compound	Bond centrality[a]	ECBC[a]
2		
3		
4		
5		

[a] Bold lines indicate bonds with the highest bond centrality and ECBC.

Therefore, we assumed that bond centrality was a comparable parameter among different molecules.

A frequency distribution of all bond centralities in **A**, **B**, and **C** is shown in Figure 4. The average of all bond centralities was 2.105, and the most frequent distribution was in the range from 1 to 1.5. Although **B** and **C** consist of tertiary and quaternary carbons, there were many linear C–C bonds. As a result, the range of the most frequent distribution was lower than the average.

Figure 5 shows bond centralities of nine C5-hydrocarbons, and each bold line indicates the highest bond centrality, which

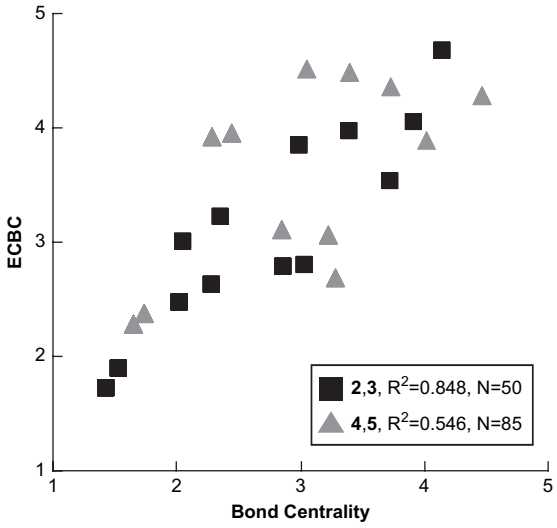


Figure 6. The plots of bond centrality versus ECBC for all bonds in **2**, **3**, **4**, and **5**.

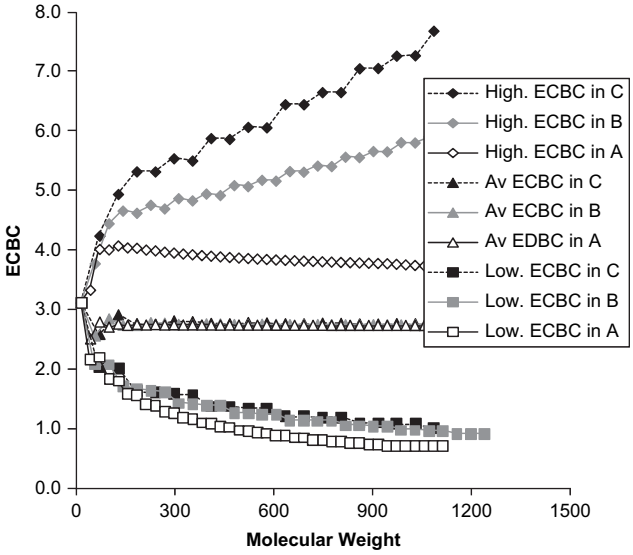


Figure 7. The plots show dependences of three kinds of ECBC for all bonds in **A**, **B**, and **C** on their molecular weights. The three kinds of ECBC are highest (High.), lowest (Low.) and average (Av) ECBC.

was obviously located in the center of each molecule. As compared with unsaturated hydrocarbons, saturated hydrocarbons had slightly higher average due to the contribution of abundant C–H bonds.

3. Result and discussion

3.1. Comparison with extended connectivity

Extended connectivity (EC) is well known as one of the topological parameters for numbering atoms in a molecule, which was proposed by Morgan.[11] EC for each atom is defined as follows: (1) the number of neighbor atoms is set as an initial value EC_1, which are classified by the same values, (2) the sum of neighbor atoms' EC_1 is set as EC_2 and classified, (3) EC_n is continuously calculated and classified, until the number of classes on EC_n becomes less than or equal to the number of previous classes, (4) the next-to-last EC_n value is used as the final EC for each atom in the molecule.

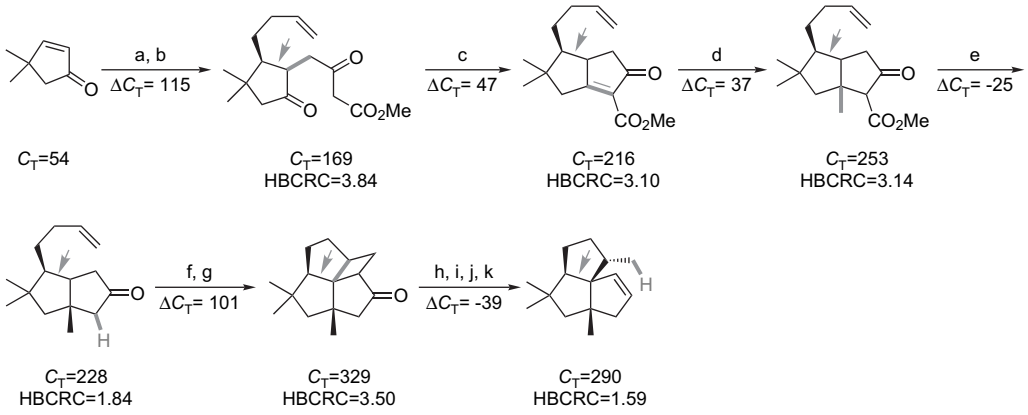
EC was normalized by its average value, called *n*EC (Eq. 6). For estimating bonds, EC bond centrality (ECBC) was newly defined by summation of two edge-atoms' *n*EC (Eq. 7).

$$n\mathrm{EC}(i) = \frac{\mathrm{EC}(i)}{\mathrm{avEC}} \tag{6}$$

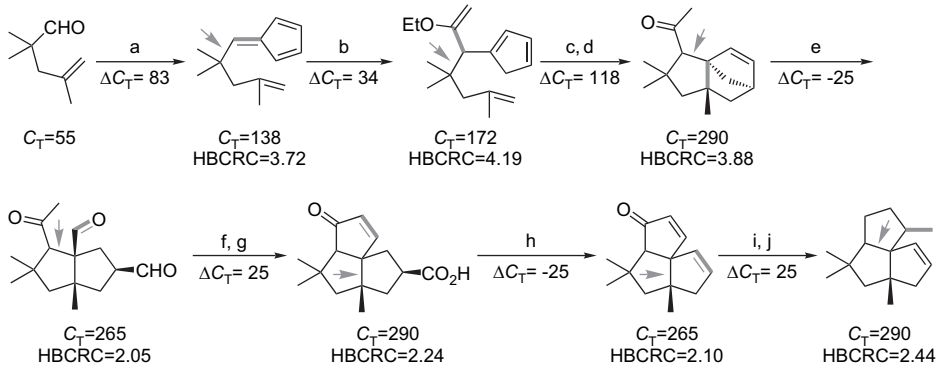
$$\mathrm{ECBC}(k) = n\mathrm{EC}(i) + n\mathrm{EC}(j) \tag{7}$$

In Eq. 6, i is a node number of a atom, and EC(i) and *n*EC(i) are EC and *n*EC of the atom i, and avEC is the average of EC(i). In Eq. 7, k is an ID number of a bond, and i and j are the edge of node numbers of the bond k, and ECBC(k) is ECBC of the bond k.

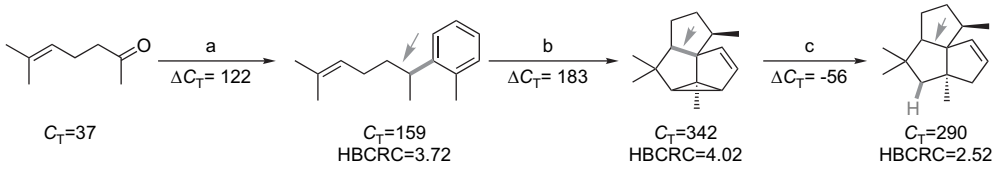
It is possible that ECBC was regarded as one of the bond properties to reflect structures. So, ECBC was compared with bond centrality using four samples, *n*-hexane (**2**), 4-propylheptane (**3**), 2,2,7,7-tetramethyloctane (**4**), and tris-(2,2-dimethylpropyl)-amine (**5**) (Table 1).



Scheme 1. Crimmin's synthetic route of silphinene.[13] Each bold gray line indicates the bond of HBCRC, and each bond pointed by gray arrow is the most central in each molecule. Reagents and conditions: (a) (1) CH_2=CHC_2H_4MgBr, Bu_3PCuI, THF; (2) $ICH_2C(OMe)$=$CHCO_2Me$, HMPA; (b) $HClO_4$, CH_2Cl_2; (c) NaOMe, MeOH; (d) Me_2CuLi, Et_2O; (e) LiCl, H_2O, DMSO, heat; (f) ETSA, Bu_4NF, THF; (g) *hν*, hexane; (h) Me_3SiI, MeCN, reflux; (i) Bu_3SnH, benzene, reflux; (j) LDA, ^{t}BuOH, THF, $(EtO)_2POCl$; (k) Li, $MeNH_2$.



Scheme 2. Sternbach's synthetic route of silphinene.[14] Each bold gray line indicates the bond of HBCRC, and each bond pointed by gray arrow is the most central in each molecule. Reagent and conditions: (a) Na^+Cp^-, THF; (b) Li(OEt)C=CH_2, THF, 0 °C; (c) benzene, 160 °C; (d)PyH·OTs, acetone, H_2O; (e) O_3, CH_2Cl_2, −78 °C; (f) KOH, MeOH, room temperature; (g) Jones reagent, acetone; (h) $Pd(OAc)_4$, $Cu(OAc)_2\cdot H_2O$, pyridine; (i) Me_2CuLi, ether, −78 °C; (j) N_2H_4, K_2CO_3, triethylene glycol, 180 °C.



Scheme 3. Wender's synthetic route of silphinene.[15] Each bold gray line indicates the bond of HBCRC, and each bond pointed by gray arrow is the most central in each molecule. Reagent and conditions: (a) *o*-bromotoluene, Li, Et_2O; (b) *hν*, pentane; 25 °C; (c) Li, $MeNH_2$, −78 °C.

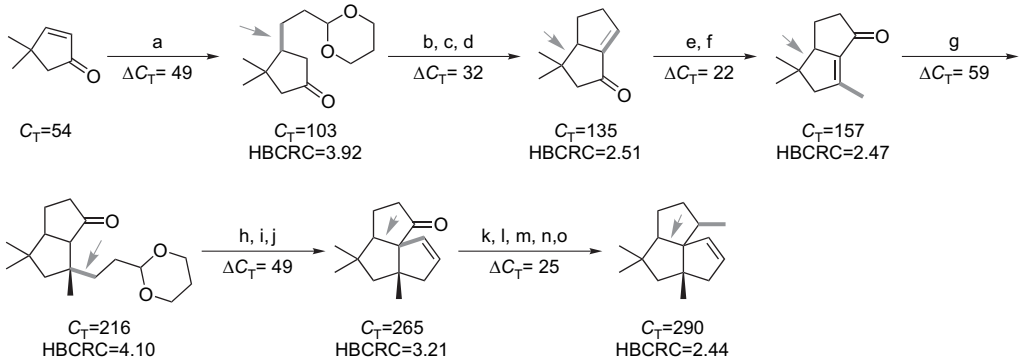
Because EC depends on its neighbor atoms, bonds with divergent skeleton have higher ECBC. As for no branching structure at the central of **2** and **3**, ECBC showed high similarity to bond centrality (Fig. 6). On the other hand, **4** and **5** had branching structures at the edge of them, therefore the highest ECBC in **4** and **5** were not located on the central bonds (Fig. 5), which was different from bond centrality.

ECBCs were calculated for all bonds in **A**, **B**, and **C** to investigate dependence on molecular weights and branching structures (Fig. 7). The result came out that ECBC depended not only on molecular weights but also on branching structures.

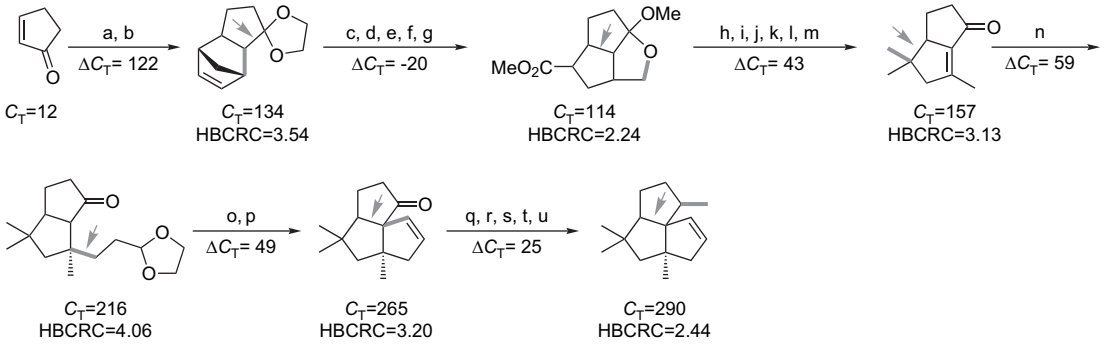
Due to the definition, ECBC did not always reflect the features of the whole molecule. Considering the dependence on molecular weight, bond centrality was more suitable for evaluating centeredness of bonds in a molecule than ECBC.

3.2. Comparison with molecular complexity in silphinene syntheses

Chanon and co-workers have reported the study of molecular complexity (C_T) using published synthetic routes of polyquinane series, silphinene, hirsutene, and corioline.[12] With the use of eight synthetic routes of silphinene, we investigated



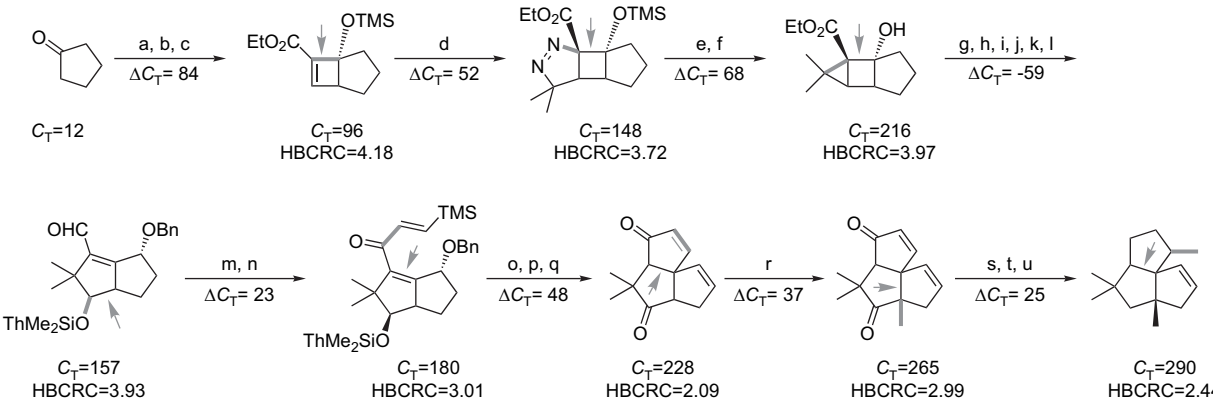
Scheme 4. Paquette's synthetic route of silphinene.[16] Each bold gray line indicates the bond of HBCRC, and each bond pointed by gray arrow is the most central in each molecule. Reagent and conditions: (a) 3-bromopropionaldehyde trimethylene acetal, Mg, THF, −78 °C, $CuBr\cdot Me_2S$; (b) aq HCl; (c) $MeSO_2Cl$, Et_3N, CH_2Cl_2, 0 °C; (d) DBU, CH_2Cl_2; (e) MeLi, ether, −78 °C; (f) PCC, CH_2Cl_2; (g) 3-bromopropionaldehyde trimethylene acetal, Mg, THF, −78 °C, $CuBr\cdot Me_2S$; (h) aq HCl, THF; (i) TsCl, pyridine, THF; (j) 200 °C, 25 Torr; (k) MeLi, hexane, −78 °C; (l) TsOH, benzene; (m) $MeCO_3H$, $NaHCO_3$, $CHCl_3$, 0 °C; (n) $BF_3\cdot Et_2O$, CH_2Cl_2; (o) N_2H_4, K_3CO_3, diethylene glycol, 150 °C.



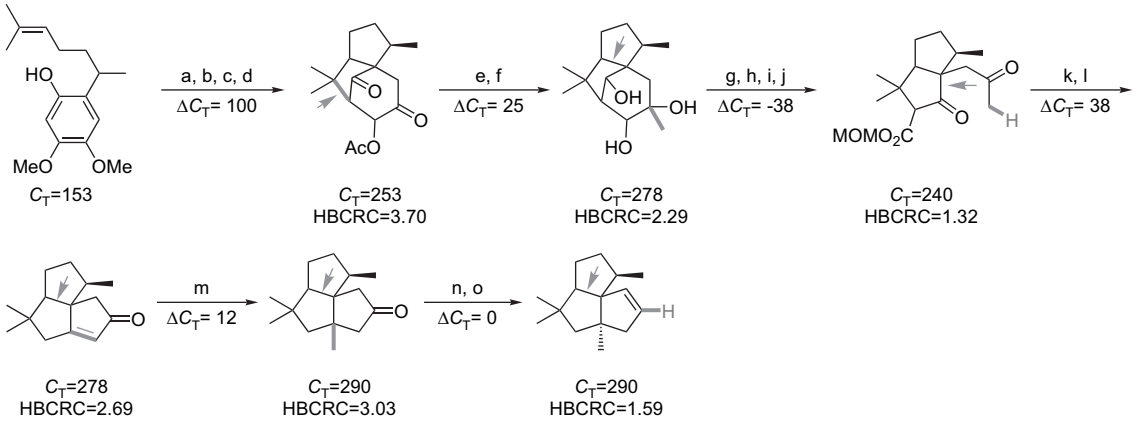
Scheme 5. Ito's synthetic route of silphinene.[17] Each bold gray line indicates the bond of HBCRC, and each bond pointed by gray arrow is the most central in each molecule. Reagent and conditions: (a) cyclopentadiene; (b) $(CH_2OH)_2$, TsOH; (c) $NaIO_4$, OsO_4, $NaHCO_3$; (d) $NaBH_4$; (e) MeOH, HCl; (f) Jones oxidn; (g) CH_2N_2; (h) LDA, MeI; (i) $LiAlH_4$; (j) PCC; (k) N_2H_4, KOH; (l) TMSCl, NaI, MeCN; (m) DBU; (n) 3-bromopropionaldehyde dimethylene acetal, Mg, CuI; (o) HCl, THF, H_2O; (p) $POCl_3$, pyridine; (q) MeLi; (r) $SOCl_2$, pyridine; (s) *m*-CPBA; (t) $BF_3\cdot OEt_2$, 0 °C; (u) N_2H_4, KOH.

correlation between differences of molecular complexity ΔC_T and the highest bond centrality of attaching bond in reaction centers (HBCRC) in each step. The literature's molecular complexity has been developed by Hendrickson and Toczek.[6b]

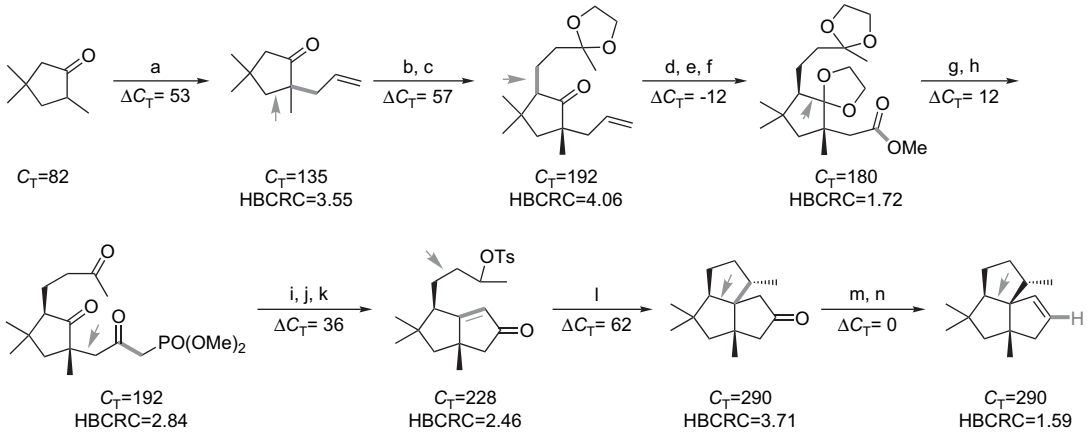
In the Chanon's paper,[12] the reaction schemes consisted of only the steps associated with a change in the carbon skeletons. ΔC_T of each step was calculated by subtracting C_T of each reactant from each product, and HBCRC was used in each



Scheme 6. Franck-Neumann's synthetic route of silphinene.[18] Each bold gray line indicates the bond of HBCRC, and each bond pointed by gray arrow is the most central in each molecule. Reagent and conditions: (a) LDA, TMSCl; (b) ethyl propynoate, $ZrCl_4$, CH_2Cl_2, 20 °C; (c) $CF_3SO_3SiMe_3$, NEt_3, CH_2Cl_2, 20 °C; (d) Me_2CN_2, −50 °C; (e) *hν*, acetone, PhCOMe; (f) TBAF, CH_2Cl_2, 20 °C; (g) H_2SO_4, Et_2O, 20 °C; (h) $CF_3SO_3SiMe_2Thexyl$, NEt_3, 25 °C; (i) SeO_2, dioxane, 110 °C; (j) $BnOCNHCCl_3$, CF_3SO_3H, CH_2Cl_2, 40 °C; (k) DIBAL, benzene, 20 °C; (l) MnO_2, CH_2Cl_2, 20 °C; (m) BrMg(CH═CH)TMS, THF, −30 °C; (n) MnO_2, CH_2Cl_2, 40 °C; (o) $BF_3\cdot Et_2O$, PhEt, 125 °C; (p) TBAF, THF, 25 °C; (q) $(COCl)_2$, DMSO; (r) LDA, MeI; (s) Me_2CuLi; (t) N_2H_4, KOH; (u) N_2H_4, KOH.



Scheme 7. Yamamura's synthetic route of silphinene.[19] Each bold gray line indicates the bond of HBCRC, and each bond pointed by gray arrow is the most central in each molecule. Reagent and conditions: (a) anoic oxidation; (b) DIBAL, THF, −78 °C; (c) Ac_2O, DMAP, pyridine; (d) $(CO_2H)_2$, MeOH, 60 °C; (e) MeMgBr, THF, 0 °C; (f) $LiAlH_4$, THF, −78 °C; (g) $Pd(OAc)_4$, MeOH; (h) $NaClO_2$, acetone; (i) MOMCl, K_2CO_3, DMF; (j) PDC, 4 Å MS, CH_2Cl_2, −30 °C; (k) HCl; (l) NaOEt, EtOH, reflux; (m) $Me_2Cu(CN)Li_2$, Et_2O, −30 °C; (n) LDA, ^{t}BuOH, THF, $(EtO)_2POCl$; (o) Li, $MeNH_2$.



Scheme 8. Nagarajan's synthetic route of silphinene.[20] Each bold gray line indicates the bond of HBCRC, and each bond pointed by gray arrow is the most central in each molecule. Reagent and conditions: (a) NaH, THF, allyl bromide, 65 °C; (b) LDA, THF, 2-methyl-1,3-dioxolan-2-yl-acetaldehyde; (c) Li, NH_3, −78 °C; (d) $RuCl_3$, $NaIO_4$, MeCN; (e) CH_2N_2, ether, 0 °C; (f) $CH(OEt)_3$, $(CH_2OH)_2$, TsOH; (g) $CH_2PO(OMe)_2$, ^{n}BuLi, THF, −78 °C; (h) HCl; (i) $^n Bu_4N^+OH^-$, benzene; (j) $NaBH_4$, 0 °C; (k) *p*-MPTC; (l) $^n Bu_3SnH$, PhMe, 80 °C; (m) LDA, ^{t}BuOH, THF, $(EtO)_2POCl$; (n) Li, $MeNH_2$.

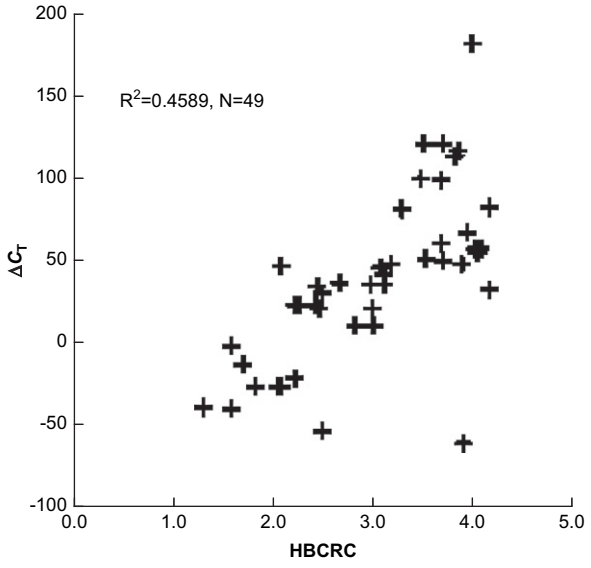


Figure 8. The plots of ΔC_T versus HBCRC of each step in silphinene syntheses.

product. Schemes 1–8 describe the values of ΔC_T and HBCRC. At the same time, the corresponding bonds of HBCRC and of the highest bonds centrality in molecules were shown by bold lines and pointing with gray arrows, respectively.

A moderately linear correlation was recognized between HBCRC of product and ΔC_T of each step (Fig. 8). As HBCRC became higher, ΔC_T was larger. This relation represents that bond disconnections of larger HBCRC's bonds in products generate synthetic precursors with larger decrease in molecular complexity. This is reasonable and bond centrality proved to be useful for finding retrosynthetically important bond. Some separate plots in Figure 8 were due to rearrangement steps that sometimes result in smaller ΔC_T.

3.3. Study of bond centrality in several natural occurring compounds

Using 40 organic compounds described in Nicolaou's total synthesis textbook,[21] validation study of bond centrality was

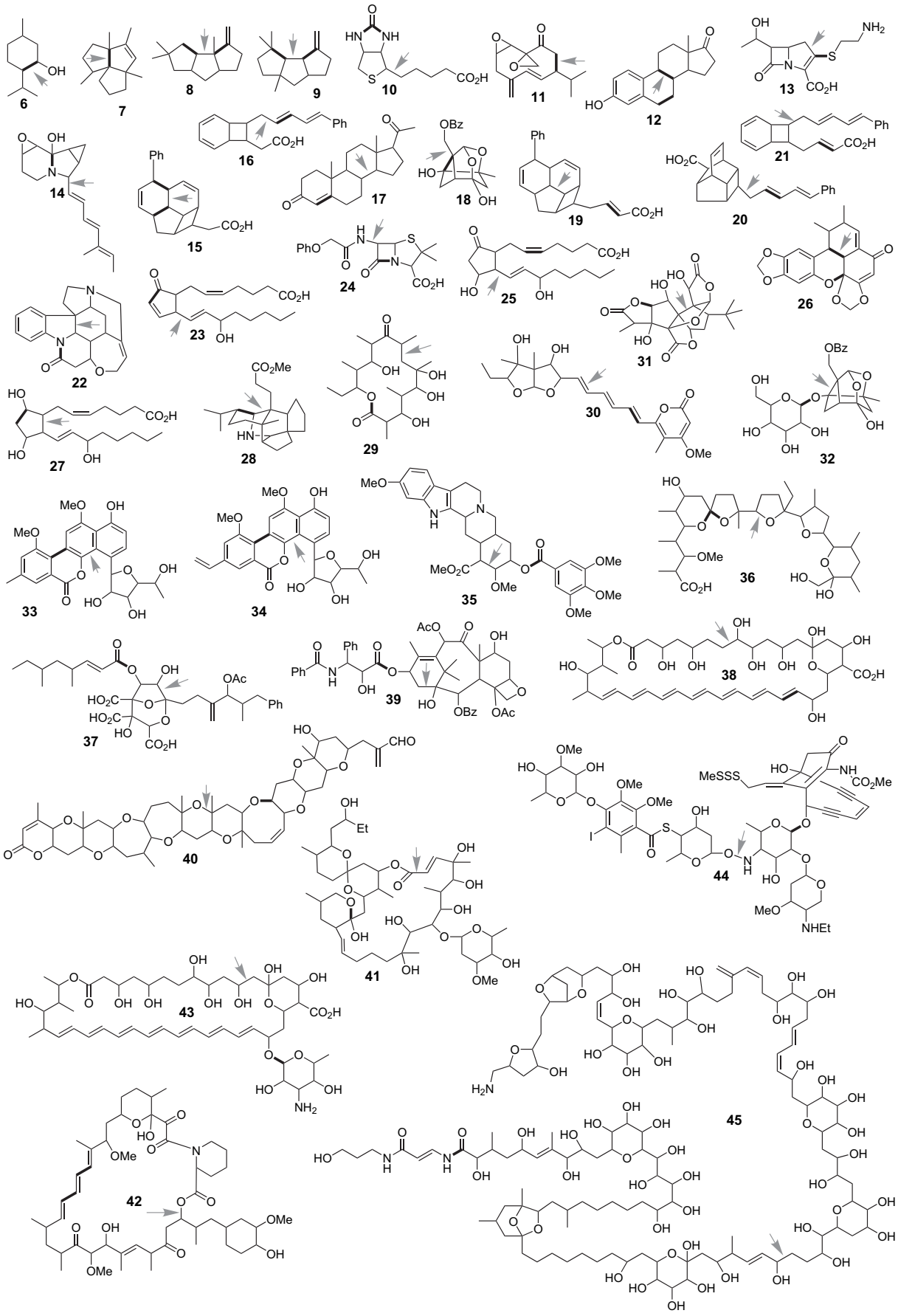


Figure 9. Targets. Bold lines indicate the last attaching bonds, and each bond pointed by gray arrows has the highest centrality in each molecule calculated by Eq. 5.

Table 2
Bond centralities in targets sorted by molecular weights

ID	Corresponding organic compounds	MW	MF	NB[a]	Bond centralities in reaction center			Bond centralities in target		
					HBCRC	Rank order[b]	Relative ranking[c] [%]	Max.	Min.	Av
6	Menthol	156.27	$C_{10}H_{20}O$	31	3.840	1	3.2	3.840	1.387	2.387
7	Isocomene	204.36	$C_{15}H_{24}$	41	4.309	1	2.4	4.309	1.527	2.422
8	Hirsutene	204.36	$C_{15}H_{24}$	41	3.752	3	7.3	4.015	1.476	2.373
9	$\Delta^{9(12)}$-Capnellene	204.36	$C_{15}H_{24}$	41	3.912	3	7.3	4.135	1.415	2.392
10	Biotin	244.31	$C_{10}H_{16}N_2O_3S$	33	1.849	21	63.6	3.855	0.956	2.280
11	Periplanone B	248.32	$C_{15}H_{20}O_3$	40	3.583	1	2.5	3.583	1.533	2.264
12	Estrone	270.37	$C_{18}H_{22}O_2$	45	3.825	1	2.2	3.825	0.865	2.295
13	Thienamycin	272.33	$C_{11}H_{16}N_2O_4S$	35	3.332	5	14.3	3.711	1.108	2.290
14	Indolizomycin	273.38	$C_{17}H_{23}NO_2$	46	2.876	9	19.6	3.876	1.106	2.213
15	Endiandric acid A	306.40	$C_{21}H_{22}O_2$	49	3.354	3	6.1	3.625	0.917	2.261
16	Endiandric acid D, E	306.40	$C_{21}H_{22}O_2$	47	3.605	3	6.4	3.826	0.948	2.186
17	Progesterone	314.47	$C_{21}H_{30}O_2$	56	1.952	34	60.7	3.887	1.136	2.304
18	Paeoniflorigenin	318.33	$C_{17}H_{18}O_6$	45	3.526	3	6.7	3.895	0.920	2.242
19	Endiandric acid B	332.44	$C_{23}H_{24}O_2$	53	1.521	38	71.7	3.514	0.790	2.247
20	Endiandric acid C	332.44	$C_{23}H_{24}O_2$	53	3.364	4	7.5	3.730	0.879	2.188
21	Endiandric acid F, G	332.44	$C_{23}H_{24}O_2$	51	1.907	29	56.9	3.860	0.929	2.189
22	Strychnine	334.42	$C_{21}H_{22}N_2O_2$	53	2.553	17	32.1	3.675	1.125	2.276
23	Prostaglandin A_2	334.46	$C_{20}H_{30}O_4$	54	2.638	17	31.5	3.721	0.917	2.191
24	Penicillin V	350.40	$C_{16}H_{18}N_2O_5S$	44	2.955	10	22.7	3.722	0.970	2.193
25	Prostaglandin E_2	352.47	$C_{20}H_{32}O_5$	57	2.592	20	35.1	3.792	0.898	2.200
26	Carpanone	354.36	$C_{20}H_{18}O_6$	49	3.916	1	2.0	3.916	1.030	2.242
27	Prostaglandin $F_{2\alpha}$	354.49	$C_{20}H_{34}O_5$	59	2.768	16	27.1	3.843	0.892	2.205
28	Methyl homoseco-daphniphyllate	359.55	$C_{23}H_{37}NO_2$	67	3.438	6	9.0	4.320	0.930	2.325
29	Erythronolide B	402.53	$C_{21}H_{38}O_7$	66	2.829	9	13.6	2.960	1.424	2.243
30	Asteltoxin	418.49	$C_{23}H_{30}O_7$	62	2.652	18	29.0	3.707	1.019	2.152
31	Ginkgolide B	424.40	$C_{20}H_{24}O_{10}$	59	2.134	32	54.2	4.207	1.231	2.328
32	Paeoniflorin	480.47	$C_{23}H_{28}O_{11}$	67	3.535	6	9.0	4.131	0.873	2.233
33	Gilvocarcin M	482.49	$C_{26}H_{26}O_9$	65	3.249	8	12.3	4.048	1.142	2.210
34	Gilvocarcin V	494.50	$C_{27}H_{26}O_9$	66	3.329	7	10.6	4.032	1.000	2.208
35	Reserpine	608.69	$C_{33}H_{40}N_2O_9$	89	3.023	20	22.5	3.646	0.888	2.167
36	Monensin	670.88	$C_{36}H_{62}O_{11}$	113	3.377	12	10.6	3.887	1.074	2.177
37	Zaragozic acid A	690.74	$C_{35}H_{46}O_{14}$	97	3.215	15	15.5	3.927	0.994	2.169
38	Amphoteronolide B	778.93	$C_{41}H_{62}O_{14}$	118	2.142	46	39.0	2.758	1.402	2.105
39	Taxol	853.92	$C_{47}H_{51}NO_{14}$	119	3.409	15	12.6	4.063	0.828	2.169
40	Brevetoxin B	895.10	$C_{50}H_{70}O_{14}$	144	2.977	27	18.8	3.604	0.870	2.123
41	Cytovaricin	901.14	$C_{47}H_{80}O_{16}$	147	2.671	34	23.1	3.184	1.054	2.135
42	Rapamycin	914.19	$C_{51}H_{79}NO_{13}$	147	1.827	102	69.4	3.496	1.191	2.126
43	Amphotericin B	924.09	$C_{47}H_{73}NO_{17}$	140	2.305	52	37.1	2.930	1.251	2.103
44	Calicheamicin γ_1^I	1366.35	$C_{55}H_{72}IN_3O_{21}S_4$	162	3.067	31	19.1	3.877	0.966	2.107
45	Palytoxin	2680.18	$C_{129}H_{223}N_3O_{54}$	418	1.064	361	86.4	3.416	0.821	2.033

[a] The number of bonds in a molecule including hydrogens.
[b] Order of the HBCRC in all bonds.
[c] (Rank order/NB)×100 [%].

performed on the last attaching bond of 40 targets (Fig. 9). Here, the last attaching bonds in total synthesis did not include simple functional group conversions, bond order changes, and one-carbon elongation reactions. The compounds were composed of various numbers of bonds from 31 to 418 including hydrogen atoms.

Table 2 summarizes the highest, lowest, and average bond centralities, and the highest bond centrality in the last attaching reaction center (HBCRC), rank order of HBCRC, and relative ranking of HBCRC against all bonds. The relative ranking is a percentage of the rank order in all bonds that is expressed as the current bond order divided by the number of bonds.

The frequency distributions of all bond centralities and HBCRC are shown in Figure 10. The most frequent distribution of HBCRC was in the range 3–3.5, which was much higher than the most frequent distribution of all bond centralities that ranges from 1.5 to 2. Additionally, the average of HBCRC was 2.95 that was higher than 2.18 for all bond centralities. The distribution obviously shows that the last attaching bonds are more centrally located in the targets.

The correlation between the HBCRCs and their relative rankings was analyzed (Fig. 11). In spite of that only atom centrality was normalized in Eq. 4, it was confirmed that

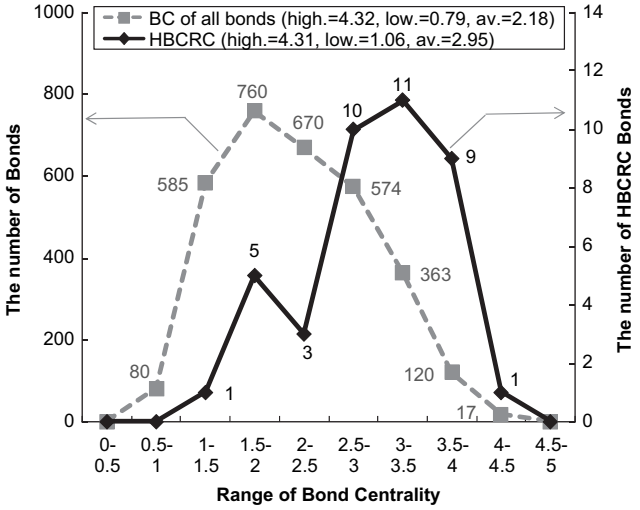


Figure 10. The black line chart shows a frequency distribution of HBCRC in **6** to **45**, and the gray dashed line chart shows a frequency distribution of all bond centralities (BC). The highest, lowest, and average of BC were 4.32, 0.79, and 2.18, and HBCRC were 4.31, 1.06, and 2.95.

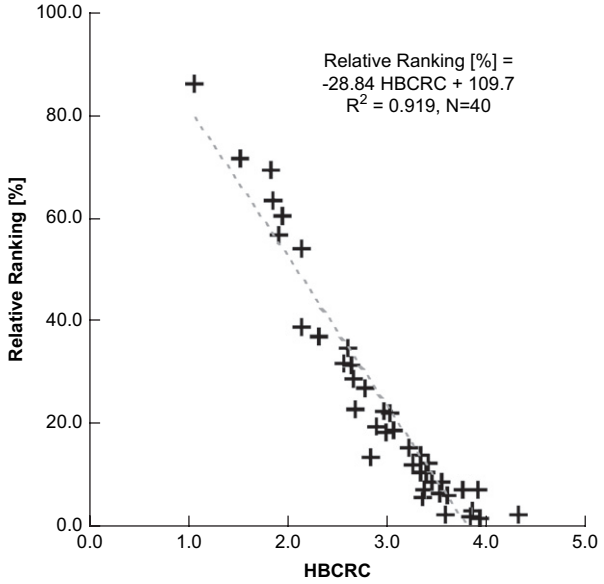


Figure 11. Plots of HBCRCs versus their relative rankings. HBCRC stands for the highest bond centrality in a reaction center of a target.

HBCRC linked to bond centrality was highly correlated with relative rankings of HBCRC (R^2=0.919). The correlation dashed line indicates that a bond with more than 2.76 HBCRC has the relative ranking within top 30%. The correlation suggests that HBCRC is a quantitative variable among different molecules.

According to the relative rankings in Table 2, HBCRC of 28 targets were ranked within top 30% in 40 targets, and 12 targets were ranked out of top 30% (**10**, **17**, **19**, **21**, **22**, **23**, **25**, **31**, **38**, **42**, **43**, and **45**) including three macrocyclic compounds, **38**, **42**, and **43**. The bond centrality properly estimated chain and branch structures, but it was difficult to deal with macrocycles. Targets **19** and **21** were derivatives of other targets **15** and **16**,[22,23] so it was not necessary to find other possible synthetic routes. In fact, the precursors **15** and **16** were synthesized by building high centrality bonds whose relative rankings were 6.1 and 6.4%. Removing the three macrocycles and the two derivatives from all the 40 targets, 35 targets included 28 targets (80%) that were synthesized at the bonds in top 30% relative ranking of bond centrality.

Targets **10**, **17**, **23**, and **25** have simply saturated C–C bonds in the center and easily synthesizable bonds around the periphery of the structures. In addition, they are relatively small. In these cases, instead of topological strategy, synthetic accessibilities were given higher priority.

In this paper, stereochemistry was not considered, although all of 40 targets contained stereocenters. As for 33 targets, bonds of the highest bond centralities consist of chiral atoms. On the other hand, only 19 targets have the last attaching bonds with chiral atoms in reported syntheses (Fig. 9). The difference is thought to be due to synthetic intractableness in asymmetric synthesis. In order to offset the gap, at least the bonds consisting of two chirals would be assigned to lower priority into molecular centrality.

4. Conclusion

For the purpose of quick search of the retrosynthetically best bond in targets, we defined a new parameter, molecular centrality, based on squared node distances, which was much simpler than the concept of molecular complexity. The result of comparison with molecular complexity in silphinene syntheses and examination with 40 organic compounds made it clear that bond centrality was useful for estimation of important bonds in retrosynthesis. We also learned that not only the molecular centrality but also synthetic accessibilities were important. Considering both factors with statistical analysis would probably realize better evaluation to find the best bond for retrosynthesis. The bond centrality is suitable for statistical analysis because the parameter of molecular centrality is fairly quantitative and comparable among molecules.

References and notes

1. Corey, E. J.; Wipke, W. T. *Science* **1969**, *166*, 178.
2. Corey, E. J.; Petersson, A. *J. Am. Chem. Soc.* **1972**, *94*, 460.
3. (a) Barone, R.; Chanon. *Encyclopedia of Computational Chemistry*; Paul von Rague Schleyer, Ed.; Wiley: Chichester, UK, 1998; p 2931; (b) Baron, R.; Chanon. *Handbook of Chemoinformatics*; Gasteiger, J., Ed.; VCH: Weinheim, 2003; Vol. 4, p 1428; (c) Pfoerter, M.; Sitzmann, M. *Handbook of Chemoinformatics*; Gasteiger, J., Ed.; VCH: Weinheim, 2003; Vol. 4, p 1457; (d) Todd, M. H. *Chem. Soc. Rev.* **2005**, *34*, 247.
4. (a) Funatsu, K.; Sasaki, S. *Tetrahedron Comput. Methodol.* **1988**, *1*, 27; (b) Funatsu, K.; Sasaki, S. *Chemistry* **1986**, *41*, 632; (c) Satoh, K.; Funatsu, K. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 316.
5. (a) Corey, E. J. Q. *Rev. Chem. Soc.* **1971**, *25*, 455; (b) Corey, E. J.; Howe, W. J.; Orf, H. W.; Pensak, D. A.; Petersson, G. *J. Am. Chem. Soc.* **1975**, *97*, 6116; (c) Corey, E. J.; Cheng, X.-M. *The Logic of Chemical Synthesis*; Wiley: New York, NY, Chichester, Brisbane, Toronto, Singapore, 1989.
6. (a) Bertz, S. H. *J. Am. Chem. Soc.* **1981**, *103*, 3599; (b) Hendrickson, J. B.; Huang, P.; Toczko, A. G. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 63; (c) Wender, P. A.; Miller, B. L. *Organic Synthesis Theory and Applications*; Hudlicky, T., Ed.; JAI: 1993; p 27; (d) Randić, M.; Plavšić, D. *Croat.*

Chem. Acta **2002**, *75*, 107; (e) Barone, R.; Petitjean, M.; Baralotto, C.; Chanon, M. *J. Phys. Org. Chem.* **2003**, *16*, 9; (f) Ruecher, C.; Ruecker, G.; Bertz, S. H. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 378; (g) Whitlock, H. W. *J. Org. Chem.* **1998**, *63*, 7982.

7. (a) Hendrickson, J. B. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 1286; (b) Hendrickson, J. B. *CHEMTECH* **1998**, *28*, 35.
8. Barone, R.; Chanon, M. *Tetrahedron* **2005**, *61*, 8916.
9. A Perl script was prepared for calculating molecular centrality. The script is available only for academic usage via email at funatsu@chemsys.t.u-tokyo.ac.jp.
10. Node distance is the number of bonds between two atoms in a molecule. If two atoms, *i* and *j*, are neighbors to each other, dist(*i*,*j*) is equal to 1.
11. (a) Morgan, H. L. *J. Chem. Doc.* **1965**, *5*, 107; (b) Gasteiger, J.; Engel, T. *Chemoinformatics*; VCH: Weinheim, 2003; p 59.
12. Chanon, M.; Barone, R.; Baralotto, C.; Julliard, M.; Hendrickson, J. B. *Synthesis* **1998**, 1559.
13. Crimmins, M. T.; Mascarella, S. W. *J. Am. Chem. Soc.* **1986**, *108*, 3435.
14. Sternbach, D. D.; Hughes, J. W.; Burdi, D. F.; Banks, B. A. *J. Am. Chem. Soc.* **1985**, *107*, 2149.
15. Wender, P. A.; Ternansky, R. J. *Tetrahedron Lett.* **1985**, *26*, 2625.
16. Paquette, L. A.; Leone-Bay, A. *J. Am. Chem. Soc.* **1983**, *105*, 7352.
17. Tsunoda, T.; Kodama, M.; Ito, S. *Tetrahedron Lett.* **1983**, *24*, 83.
18. (a) Franck-Neumann, M.; Miesch, M.; Gross, L. *Tetrahedron Lett.* **1991**, *32*, 2135; (b) Miesch, M.; Miesch-Gross, L.; Franck-Neumann, M. *Tetrahedron* **1997**, *53*, 2103.
19. Shizuri, Y.; Ohkubo, M.; Yamamura, S. *Tetrahedron Lett.* **1989**, *30*, 3797.
20. Rao, Y. K.; Nagarajan, M. *Tetrahedron Lett.* **1988**, *29*, 107.
21. Nicolaou, K. C.; Sorensen, E. J. *Classics in Total Synthesis*; VCH: Weinheim, New York, NY, Basel, Cambridge, Tokyo, 1996. All targets were investigated except for sugars and an organometallic compound, vitamin B-12.
22. Nicolaou, K. C.; Petasis, N. A.; Zipkin, R. E.; Uenishi, J. *J. Am. Chem. Soc.* **1982**, *104*, 5555.
23. Nicolaou, K. C.; Petasis, N. A.; Uenishi, J.; Zipkin, R. E. *J. Am. Chem. Soc.* **1982**, *104*, 5557.