

Generality of Correlation between Yield and Reduced Mass of Raw Materials in Organic Reactions

Masatoshi Kawashima
Laboratory Section, Azuma Co., Ltd.
1-6-13 Tamasaki-Nishi, Ichihara, Chiba 290-0044

E-mail: kawashima@azuma-g.co.jp

To validate a generality of the correlation between product yield and reduced mass of raw materials, the regression analysis of 129 reaction examples (55 as a sample size) including at least 66 types of reactions used in syntheses of natural products such as peptides and terpenes was conducted. It was possible to predict a yield of a variety of synthetic reactions for a synthesis of natural product with many aliphatic carbon chains by applying a reduced mass, adjusted with a molecular weight and the number of rotatable bond, to the regression equation. Moreover, it was found that the increase in yield due to a use of the adjusted reduced mass correlated with the harmonic mean of the molar heat capacity of raw materials and was expressed as a second-order approximation within the analysis range.

Keywords: Yield |Reduced Mass |Correlation |Prediction

The author has reported the correlation between yield and reduced mass of raw materials in intermolecular/intramolecular coupling reactions, cyclization reactions, and reactions of fullerene.^{1,2,3} The same regression equation (1),

$$\text{yield} = -0.186M_{AB}/n + 100,$$

where M_{AB} is the reduced mass per mole, $n = n_A n_B n_I$, n_A and n_B are the total number of each reaction site in molecular A and molecular B, respectively, and n_I is the number to distinguish whether it is an intramolecular reaction or intermolecular reaction, was held in the reactions. Reactions selected for the analysis were chiefly metal-catalyzed cross coupling reactions, cycloadditions, and Bingel reaction. Therefore, to find out about other types of reactions, papers on the synthesis of several natural products were selected and analyzed in this report.

First, a part of reactions used in a total synthesis of Antillatoxin shown in Scheme 1 was analyzed by using equation (1).⁴ In the case of intramolecular reaction, M_B was set to be the same as M_A , the variable n_A and n_B was set to the same value, and the variable n_I was set to 2 as previous reports.^{2,3} In a previous report, when a molecular weight of a product exceeded about 1500, the mass according to the number of rotatable bonds (NORB) of the raw material was subtracted from the molecular weight and an adjusted reduced mass (M'_{AB}) was obtained from the adjusted molecular weight ($M'_{A(B)}$).^{3,5} In this report, to further improve the accuracy, the mass according to NORB ($R_{A(B)}$) was adjusted according to $M_{A(B)}$, and subtracted from $M_{A(B)}$ to obtain $M'_{A(B)}$ as shown in (3). Then, M'_{AB} was obtained from M'_A and M'_B as presented in (2) and was used in (1). The coefficient of 14.03 in (3) is the mass of the methylene chain, which is often a cause of

$$\text{yield} = -0.186 \frac{M'_{AB}}{n} + 100 \quad (1)$$

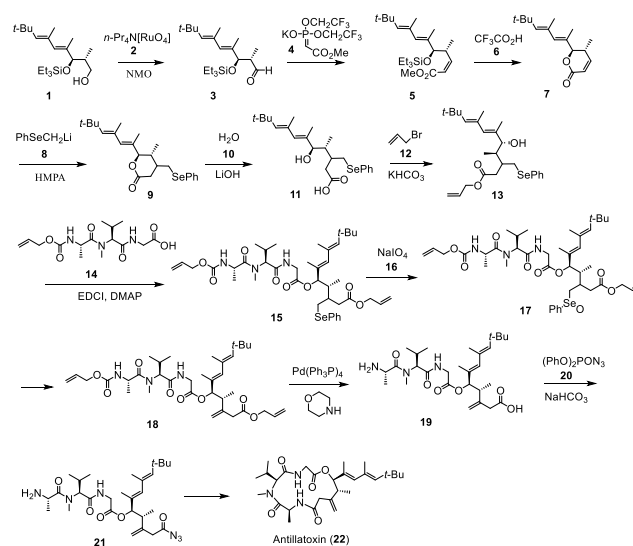
$$M'_{AB} = \frac{M'_A M'_B}{M'_A + M'_B} \quad (2)$$

$$M'_{A(B)} = M_{A(B)} - 14.03aR_{A(B)} \quad (3)$$

$$a = 0.00177M_{A(B)} \quad (M_{A(B)} \leq 768) \quad (4)$$

$$a = 1.36 \quad (M_{A(B)} > 768) \quad (5)$$

a rotatable bond. For a case of a reaction using condensation reagent for esterification or amidation, the intermediate involved in the rate-determining step was also considered for analysis. An analysis was conducted as shown in Table 1 and resulted in a regression coefficient of 1.29 and a coefficient of determination of 0.94. Literature versus predicted yields plot is included in Figure 1(a) as a regression through the origin (RTO) model and expressed as blue circle.



Scheme 1. Reaction scheme for an analysis of reactions used in a total synthesis of Antillatoxin, HMPA = hexamethylphosphoramide, EDCI = 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride, DMAP = 4-dimethylaminopyridine.

Second, reactions used in a synthesis of a fragment for a total synthesis of Microcystin-LF depicted in Scheme 2 was analyzed.⁶ An analysis was conducted as shown in Table 2. The predicted yields are in close agreement with the literature yield to result in a regression coefficient of 1.10 and a coefficient of determination of 0.99. Literature versus predicted yields plot is included in Figure 1(a) and expressed as red rhombus.

Table 1. An analysis of reactions used in a synthesis of a part of a total synthesis of Antillatoxin.

A	B	M_A^a	M_B^b	M_{AB}^c	n_A^d	n_B^e	n_i^f	n_g^g	$Y_{lit.}^h$	$Y_{int.}^i$	$Y_{pred.}^j$
1	2	341	351	132	1	1	1	1		75	
3	4	339	356	138	1	1	1	1		74	
5	6	395	114	77	1	1	1	1	52	86	48
7	8	248	177	91	1	1	1	1	50	83	83
9	10	419	18	17	1	1	1	1		97	
11	12	437	121	85	1	1	1	1		84	
13	14	478	500	146	1	1	1	1	34	73	59
15	16	803	214	129	1	1	1	1	67	76	76
17	17	819	819	163	1	1	2	2		85	
18	18	646	646	146	1	1	2	2		86	
19	20	522	275	134	1	1	1	1		75	
21	21	547	547	158	1	1	2	2	29	85	47

^amolecular weight of molecule A. ^bmolecular weight of molecule B. ^cadjusted reduced mass. ^dnumber of reaction sites of molecule A. ^enumber of reaction sites of molecule B. ^fintermolecular reaction = 1, intramolecular reaction = 2. ^g $n = n_A n_B n_i$. ^hliterature yield. ⁱpredicted yield of intermediate in each step. ^jpredicted yield.

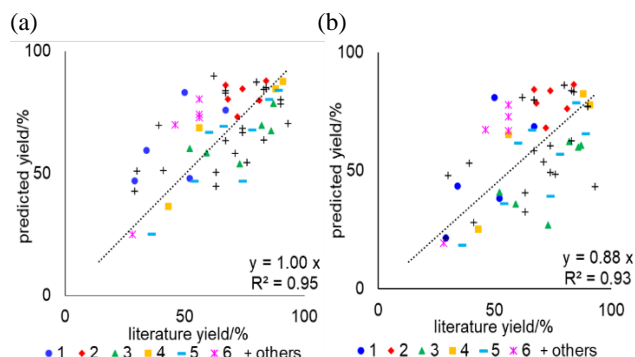
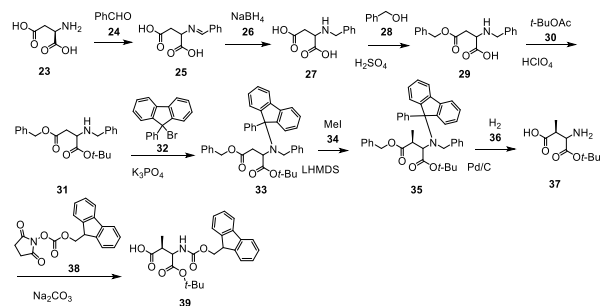


Figure 1. Literature versus predicted yields plot of the reactions used in the synthesis of natural products; ●1: data of Table 1, ◆2: data of Table2, ▲3: data of Table 3, ■4: data of Table4, +5: data of Table5, *6: data of Table 6, +: data of others. (a) calculated by using M'_{AB} . (b) calculated by using M_{AB} .



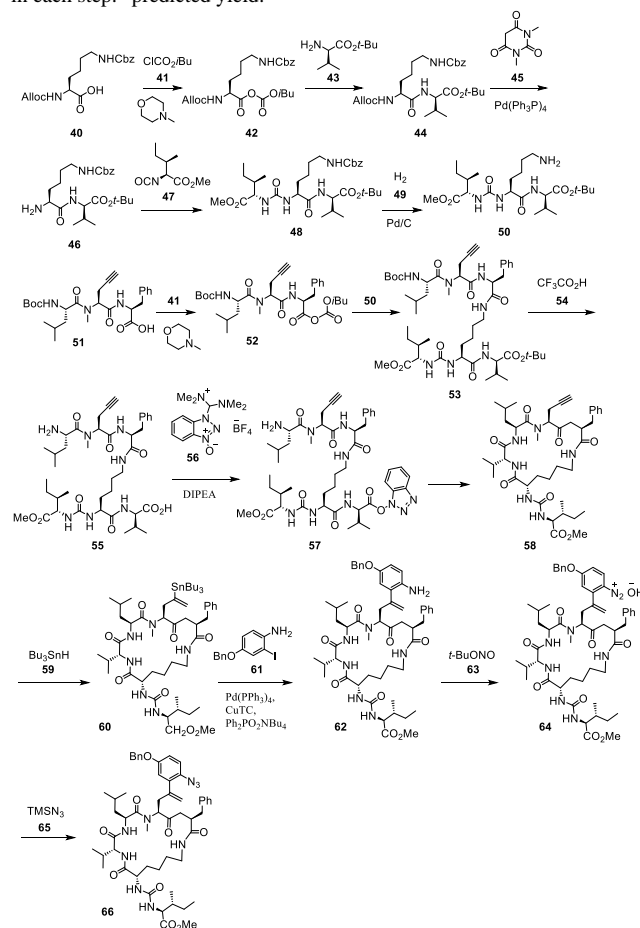
Scheme 2. Reaction scheme for an analysis of reactions used in a synthesis of a fragment for a total synthesis of Microcystin-LF, Ac = acetyl.

Third, reactions used in a synthesis of a part of a total synthesis of mozamide A shown in Scheme 3 was analyzed.⁷ An analysis was carried out as shown in Table 3. The predicted yields are lower than the literature yields, and a regression coefficients (0.87) are somewhat smaller, but a coefficient of determination (0.98) is larger. Literature versus predicted yields plot is included in Figure 1(a) and expressed as green triangle.

Table 2. An analysis of reactions used in a synthesis of a fragment for Microcystin-LF.

A	B	M_A^a	M_B^b	M_{AB}^c	n_A^d	n_B^e	n_i^f	n_g^g	$Y_{lit.}^h$	$Y_{int.}^i$	$Y_{pred.}^j$
23	24	133	106	54	1	1	1	1		90	
25	26	221	38	31	1	1	1	1	74	94	85
27	28	223	108	65	1	1	1	1	84	88	88
29	30	313	116	75	1	1	1	1	67	86	86
31	32	369	321	145	1	1	1	1	72	73	73
33	34	610	142	106	1	1	1	1	68	80	80
35	36	624	2	2	1	1	1	1		100	
37	38	203	337	107	1	1	1	1	81	80	80

^amolecular weight of molecule A. ^bmolecular weight of molecule B. ^cadjusted reduced mass. ^dnumber of reaction sites of molecule A. ^enumber of reaction sites of molecule B. ^fintermolecular reaction = 1, intramolecular reaction = 2. ^g $n = n_A n_B n_i$. ^hliterature yield. ⁱpredicted yield of intermediate in each step. ^jpredicted yield.



Scheme 3. Reaction scheme for an analysis of reactions used in a synthesis of a part of a total synthesis of mozamide A, Alloc = allyloxycarbonyl, Cbz = benzyloxycarbonyl, DIPEA = diisopropylethylamine, CuTC = copper(I) thiophene-2-carboxylate, TMS = trimethylsilyl.

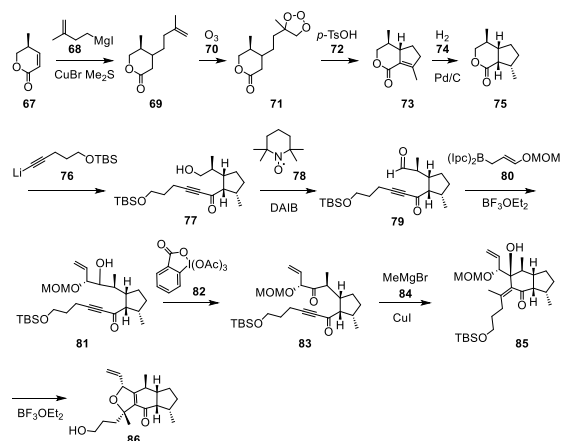
Fourth, reactions used in a synthesis of a part of total synthesis of (+)-leucosceptroid B shown in Scheme 4 was analyzed.⁸ An analysis was conducted as shown in Table 4. The predicted and literature yields were in good agreement, resulting in a regression coefficient of 0.99 and a coefficient of determination of 0.99. Literature versus predicted yields plot is included in Figure 1(a) and expressed as orange square.

Fifth, the reactions used in total synthesis of (+)-leucosceptroid B and (+)-leucosceptroid A shown in

Table 3. An analysis of reactions used in a synthesis of a part of a total synthesis of mozamide A.

A	B	M_A^a	M_B^b	M'_{AB}^c	n_A^d	n_B^e	n_i^f	n^g	Y_{lit}^h	Y_{int}^i	Y_{pred}^j
40	41	364	137	83	1	1	1	1		85	
42	43	465	173	95	1	1	1	1	82	82	70
44	45	520	156	98	1	1	1	1		82	
46	47	436	171	95	1	1	1	1	86	82	67
48	49	607	2	2	1	1	1	1		100	
51	41	488	137	90	1	1	1	1		83	
52	50	588	473	148	1	1	1	1	52	73	60
53	54	942	114	83	1	1	1	1		85	
55	56	786	321	130	1	1	1	1		76	
57	57	903	903	168	1	1	2	2	73	84	54
58	59	767	291	115	1	1	1	1	87	79	79
60	61	1060	325	160	1	1	1	1		70	
62	63	966	103	75	1	1	1	1		86	
64	65	978	115	17	1	1	1	1	59	97	59

^amolecular weight of molecule A. ^bmolecular weight of molecule B. ^cadjusted reduced mass. ^dnumber of reaction sites of molecule A. ^enumber of reaction sites of molecule B. ^fintermolecular reaction = 1, intramolecular reaction = 2. ^g $n = n_A n_B n_i$. ^hliterature yield. ⁱpredicted yield of intermediate in each step. ^jpredicted yield.



Scheme 4. Reaction scheme for an analysis of reactions used in a synthesis of a part of a total synthesis of (+)-leucosceptroid B, *p*-TsOH = *p*-toluenesulfonic acid, TBS = *tert*-butyldimethylsilyl, DAIB = (diacetoxyiodo)benzene, MOM = methoxymethyl, ipc = isopinocampheyl.

Table 4. An analysis of reactions used in a part of a total synthesis of (+)-leucosceptroid B.

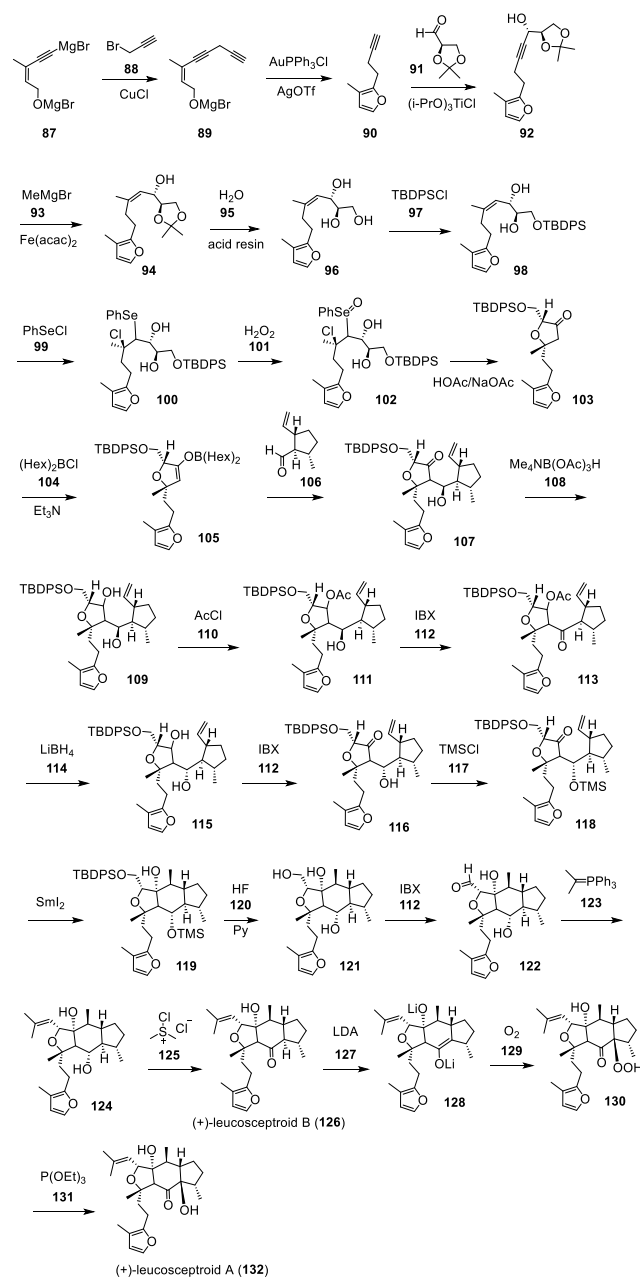
A	B	M_A^a	M_B^b	M'_{AB}^c	n_A^d	n_B^e	n_i^f	n^g	Y_{lit}^h	Y_{int}^i	Y_{pred}^j
67	68	112	220	66	1	1	1	1		88	
69	70	182	48	36	1	1	1	1		93	
71	72	230	172	85	1	1	1	1		84	
73	74	166	2	2	1	1	1	1	56	100	69
75	76	168	204	81	1	1	1	1		85	
77	78	367	156	89	1	1	1	1		84	
79	80	365	386	147	1	1	1	1		73	
81	82	467	424	158	1	1	1	1	43	71	36
83	84	465	119	83	1	1	1	1	88	85	85
85	85	481	481	133	1	1	2	2	91	88	88

^amolecular weight of molecule A. ^bmolecular weight of molecule B. ^cadjusted reduced mass. ^dnumber of reaction sites of molecule A. ^enumber of reaction sites of molecule B. ^fintermolecular reaction = 1, intramolecular reaction = 2. ^g $n = n_A n_B n_i$. ^hliterature yield. ⁱpredicted yield of intermediate in each step. ^jpredicted yield.

Scheme 5 was analyzed.⁹ An analysis was conducted as expressed in Table 5. The predicted and literature yields were

in relatively good agreement, resulting in a regression coefficient of 0.90 and a coefficient of determination of 0.98. Literature versus predicted yields plot is included in Figure 1(a) and expressed as light blue minus.

Sixth, reactions used in a total synthesis of (–)-6-*epi*-ophiobolin N shown in Scheme 6 was analyzed.¹⁰ An analysis was carried out as shown in Table 6. Although the predicted yields had higher than the literature yield (a regression coefficient = 1.35), a coefficient of determination was very large (0.99). Literature versus predicted yields plot is included in Figure 1(a) and expressed as pink asterisk.



Scheme 5. Reaction scheme for an analysis of reactions used in a total synthesis of (+)-leucosceptroid B and (+)-leucosceptroid A, TBDPS = *tert*-butyldiphenylsilyl, Ac = acetyl, IBX = 2-iodoxybenzoic acid, TMS = trimethylsilyl, LDA = lithium diisopropylamide.

Finally, analysis of reactions used in a synthesis of (–)-alotaketol C, (–)-alotaketol D, (+)-luffarin I, (+)-luffarin A, (–)-nitidasin, and four fragments that differ from the fragment of Microcyctin-LF described above were conducted as the same manner as described above and were included in Figure 1(a), and expressed as black plus.^{6,11,12,13} The data shown in Figure 1(a) are after outliers by Smirnov-Grubbs test as a significance level of 0.05. Looking at the color-coded plots in Figure 1(a), we can see that the relationship between literature and predicted yields is characteristic of each literature. In other words, there are three types of plots: those (●, ◆, *) that are clustered on the upper side of the regression line, those (▲) that clustered on the lower side of the regression line, and those (■, —) that are relatively well aligned with the regression line. There is no significant difference in the type of reaction used in each synthesis, suggesting that one of the reasons for this is that the purification efficiency or purities differ for each literature. The regression line summarizing the all results of the above analyses is also shown in Figure 1(a) as a dotted line. A regression coefficient and a determination coefficient were 1.00 and 0.95, respectively. Figure 1(b) shows a result obtained by using M_{AB} calculated without adjusting $M_{A(B)}$ by NORB. The effect of an adjustment by NORB is evident from difference in these regression coefficients and difference in coefficients of determination. A significant relationship between harmonic mean of molar heat capacity (C_p) of raw materials and an increase in a yield calculated using M'_{AB} compared to a yield calculated using M_{AB} was obtained using data of Tables 4 and 5 and is presented in Figure 2.¹⁴ The correlation was expressed as a quadratic approximation within the analysis range, and a high coefficient of determination was observed ($R^2 = 0.94$);

$$y = 11.2C_{pAB}^2 + 11.9C_{pAB} - 0.5,$$

where y is an increase in yield due to an adjustment by NORB, C_{pAB} is the harmonic mean of molar heat capacity of raw materials. Although a sample size for each reaction type used in the analysis was small, high correlation was observed throughout despite a large number of reaction types. Moreover, the fact that the regression equation for the correlation between yield and reduced mass, which held in C-C, C-N, C-O couplings, cycloadditions, and Bingel reactions, is valid for the present analysis indicates high generality as a regression equation for yield prediction in thermal organic reactions.^{1,2,3} This equation can be applied when the reaction conditions are appropriate. If the reaction conditions are optimized but deviate significantly from the regression equation, it may be because the molecular weight of the chemical species involved in the reaction or the reaction mechanism is different from what is assumed.

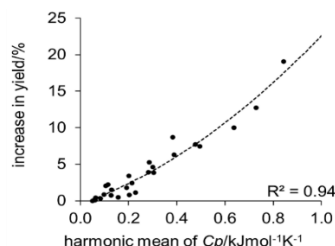
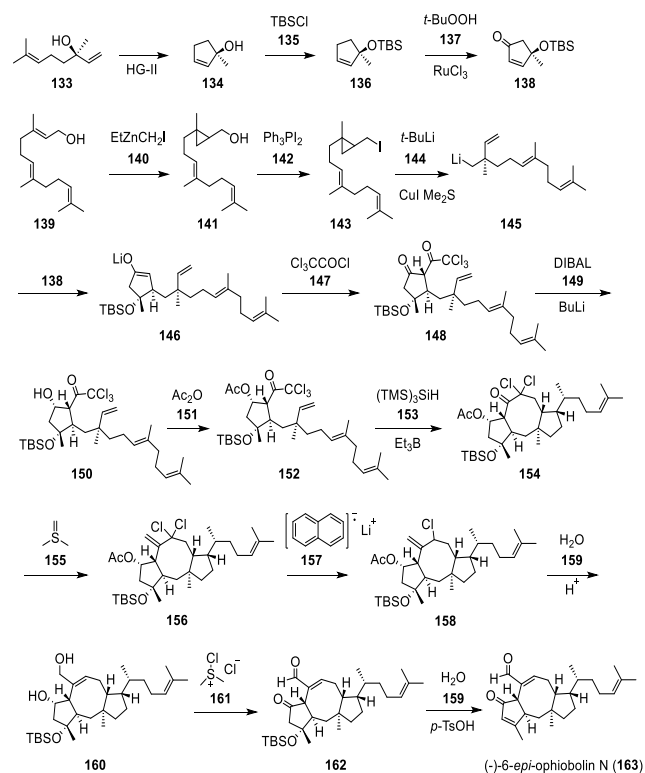


Figure 2. Relationship between a harmonic mean of C_p of raw materials and an increase in yield due to an adjustment by NORB.

Table 5. An analysis of reactions used in a total synthesis of (+)-leucosceptroid B and (+)-leucosceptroid A.

A	B	M_A^a	M_B^b	M'_{AB}^c	n_A^d	n_B^e	n_f^f	n_g^g	Y_{lit}^h	Y_{int}^i	Y_{pred}^j
87	88	303	119	81	1	1	1	1	85		
89	89	134	134	59	1	1	2	2	85	95	80
90	91	134	130	64	1	1	1	1		88	
92	93	264	238	114	1	1	1	1	66	79	69
94	95	280	18	16	1	1	1	1		97	
96	97	240	275	108	1	1	1	1		80	
98	99	479	192	118	1	1	1	1		78	
100	101	634	34	31	1	1	1	1		94	
102	102	650	650	188	1	1	2	2	54	83	47
103	104	477	217	102	1	1	1	1		81	
105	106	657	138	86	1	1	1	1	78	84	68
107	108	615	263	143	1	1	1	1		73	
109	110	617	79	66	1	1	1	1		88	
111	112	659	280	164	1	1	1	1		69	
113	114	657	22	21	1	1	1	1		96	
115	112	617	280	163	1	1	1	1		70	
116	117	615	109	86	1	1	1	1	36	84	25
118	118	687	687	139	1	1	2	2		87	
119	120	689	20	19	1	1	1	1	89	96	84
121	112	379	280	132	1	1	1	1		75	
122	123	376	304	135	1	1	1	1		75	
124	125	403	133	90	1	1	1	1	74	83	47
126	127	401	107	75	1	1	1	1		86	
128	129	412	32	29	1	1	1	1		95	
130	131	433	166	95	1	1	1	1	60	82	67

^amolecular weight of molecule A. ^bmolecular weight of molecule B. ^cadjusted reduced mass. ^dnumber of reaction sites of molecule A. ^enumber of reaction sites of molecule B. ^fintermolecular reaction = 1, intramolecular reaction = 2. ^g $n = n_A n_B n_i$. ^hliterature yield. ⁱpredicted yield of intermediate in each step. ^jpredicted yield.



Scheme 6. Reaction scheme for analysis of reactions used in a total synthesis of (–)-6-*epi*-ophiobolin N, HG-II = Hoveyda-Grubbs' 2nd generation catalyst, TBS = *tert*-butyldimethylsilyl, DIBAL = diisobutylaluminum hydride, Ac=acetyl, TMS = trimethylsilyl.

Table 6. An analysis of reactions used in a total synthesis of (-)-6-*epi*-ophiobolin N.

A	B	M_A^a	M_B^b	M_{AB}^c	n_A^d	n_B^e	n_t^f	n^g	$Y_{lit.}^h$	$Y_{int.}^i$	$Y_{pred.}^j$
133	133	154	154	68	1	1	2	2		94	
134	135	98	151	58	1	1	1	1		89	
136	137	212	90	60	1	1	1	1	56	89	74
139	140	222	235	99	1	1	1	1		82	
141	142	236	516	136	1	1	1	1		75	
143	144	346	64	51	1	1	1	1		91	
145	138	226	226	97	1	1	1	1		82	
146	147	453	182	114	1	1	1	1		79	
148	149	592	142	97	1	1	1	1		82	
150	151	594	102	78	1	1	1	1	28	86	25
152	153	636	249	146	1	1	1	1	56	73	73
154	155	602	76	63	1	1	1	1		88	
156	157	600	135	98	1	1	1	1		82	
158	159	565	18	17	1	1	1	1	46	97	70
160	161	505	133	91	1	1	1	1		83	
162	159	501	18	17	1	1	1	1	56	97	80

^amolecular weight of molecule A. ^bmolecular weight of molecule B. ^cadjusted reduced mass. ^dnumber of reaction sites of molecule A. ^enumber of reaction sites of molecule B. ^fintermolecular reaction = 1, intramolecular reaction = 2. ^g $n = n_A n_B n_t$. ^hliterature yield. ⁱpredicted yield of intermediate in each step. ^jpredicted yield.

References and Notes

- 1 M. Kawashima, *ChemRxiv*. **2020**, Preprint. <https://doi.org/10.26434/chemrxiv.12993218.v1>
- 2 M. Kawashima, *ChemRxiv*. **2020**, Preprint. <https://doi.org/10.26434/chemrxiv.13350974.v1>
- 3 M. Kawashima, *ChemRxiv*. **2020**, Preprint. <https://doi.org/10.26434/chemrxiv.13465718.v1>
- 4 F. Yokokawa, H. Fujiwara, T. Shioiri, *Tetrahedron Lett.* **1999**, *40*, 1915.
- 5 D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple, *J. Med. Chem.* **2002**, *45*, 2615.
- 6 I. Zemskov, S. Altaner, D. R. Dietrich, V. Wittmann, *Org. Chem.* **2017**, *82*, 3680.
- 7 L. Junk, U. Kazmaier, *J. Org. Chem.* **2019**, *84*, 2489.
- 8 X. Huang, L. Song, J. Xu, G. Zhu, B. Liu, *Angew. Chem., Int. Ed.* **2013**, *52*, 952.
- 9 S. Guo, J. Liu, D. Ma, *Angew. Chem., Int. Ed.* **2015**, *54*, 1298.
- 10 Z. G. Brill, H. K. Grover, T. J. Maimone, *Science* **2016**, *352*, 1078.
- 11 Y. Chen, J. Zhao, S. Li, J. Xu, *Nat. Prod. Rep.* **2019**, *36*, 263.
- 12 A. Urosa, I. S. Marcos, D. Díez, G. B. Plata, J. M. Padrón, P. Basabe, *Mol. Divers.* **2016**, *20*, 369.
- 13 a) D. T. Hog, F. M. E. Huber, P. Mayer, D. Trauner, *Angew. Chem., Int. Ed.* **2014**, *53*, 8513. b) D. T. Hog, F. M. E. Huber, G. Jiménez-Osés, P. Mayer, K. N. Houk, D. Trauner, *Chem. Eur. J.* **2015**, *21*, 13646.
- 14 Molar heat capacity was calculated by MOPAC 6.0 with PM3 using Winmostar (V4, X-Ability Co. Ltd., Tokyo, Japan, 2012) as Gui interface. Reactions using Dess-Martin periodinane or 2-iodoxybenzoic acid and reactions of compounds containing lithium elements out of the data in Tables 4 and 5 were excluded from the calculation.