Correlation between Yield and Reduced Mass of Raw Materials in Enzymatic Reactions

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

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

Correlation between the yield in the enzymatic reaction and the molecular weight of the substrate as an approximation of the reduced mass of the raw materials was clarified. The correlation was expressed by the same regression equation as in general organic chemical reactions. The coefficient of the regression equation to distinguish between intramolecular and intermolecular reactions were better when the values for intramolecular reactions were used in the plot of literature yields versus predicted yields. It was also found that the adjustment of the reduced mass by the number of rotatable bonds was not necessary and was found to be a good representation of the characteristics of the enzymatic reaction.

Keywords: Yield |Reduced Mass |Correlation |Prediction

Previous reports have been revealed that in many organic chemical reactions, there is a correlation between the reduced mass of the raw material and the yield, as expressed by the following regression equation.[1-4] The accuracy of the

$$yield = -0.186\frac{M'_{AB}}{n_A n_B n_I} + 100 \qquad (1)$$

$$M'_{AB} = \frac{M'_A M'_B}{M'_A + M'_B}$$

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

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

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

M'_{AB} : adjusted reduced mass of molecular A and B
$M'_{A(B)}$: adjusted molecular weight of molecular A(B)
$M_{A(B)}$: molecular weight of molecular A(B)
$n_{A(B)}$: number of reaction sites A(B)
n_I : intermolecular = 1, intramolecular = 2
$R_{A(B)}$: number of rotatable bonds of molecular A(B)
a : coefficient for adjustment by molecular weight

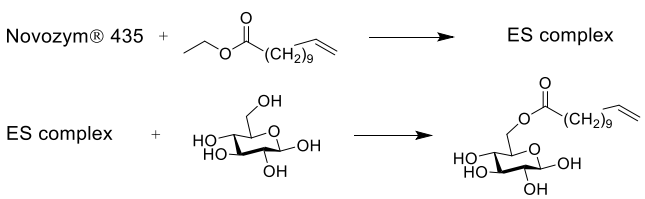
regression equation can be improved by adjusting the reduced mass by the number of rotatable bonds (NORB), which is related to rotational kinetic energy, and by the use of a coefficient to distinguish intramolecular reactions has revealed.[5] This paper reports on the analysis of enzymatic reactions, which are basically chemical reactions, to assess if the characteristics of the enzymatic reaction are reflected.

Molecular weight of the enzyme is at least 10000. If the molecular weight of the substrate is less than 1000, the error between the reduced mass of the enzyme and the substrate and the molecular weight of the substrate is at the most 10%. In the case of enzymes with molecular weights of 16500 to 675000 and substrates with molecular weights of 300 or less, which account for 92% of the examples analyzed in this study, the error between the reduced mass and the molecular weight of the substrate is less than 1.8%. In addition, since the number of reaction sites, n_A and n_B, in the reaction examples used in this analysis is all 1, equation (2) was used for the analysis instead of the equation (1):

$$yield = -0.186\frac{X}{n_I} + 100 \qquad (2)$$

X = M_B or M'_B.

First, esterification and transesterification reactions using various lipases (EC number 3.1.-) were analyzed.[6-18] As an example, transesterification of carboxylic ester with glucose using Novozym®435 is shown in Scheme 1.[7]



Scheme 1. Reaction scheme for an analysis of lipase-catalyzed transesterification of ethyl 11-dodecenoate with glucose.

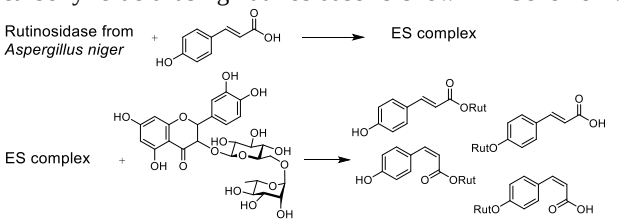
The analysis was divided into two steps: step 1, in which the enzyme and ester form an ES complex, and step 2, in which the ES complex reacts with glucose, as shown in Table 1. Since it was considered to be an intramolecular reaction due to the distance between the reaction sites and the conformation of the substrate molecule was limited, n_I was set to be 2 and the molecular weight was not adjusted by the number of rotatable bonds. As the result, the predicted and literature yields were in good agreement. Other examples of the reaction using lipase were analyzed in the same manner and the results are summarized in Table 6 after removal of outliers by Smirnov-Grubbs test as a significance level of 0.05. Literature versus predicted yield plot is included in Figure 1 as a regression through the origin (RTO) model and expressed as blue rhombus (◆).

Table 1. An example of analysis of transesterification using lipase.

step	A	B	M_B[a]	X^b/n_I	$Y_{lit.}$[c]	Y_{int}[d]	$Y_{pred.}$[e]
1	Novozym®435	ethyl 11-dodecenoate	226	113		79	
2	ES complex	glucose	181	90	62	83	66

[a]molecular weight of the substrate B. [b]$X=M_B$. [c]literature yield of the final product. [d]predicted yield of each step. [e]predicted yield of the final product.

Second, glycosylations using glycosidase (EC number 3.2.-) were analyzed.[19-23] As an example, rutinosylation of carboxylic acid using rutinosidase is shown in Scheme 2.[19]



Rut = 6-*O*-α–L-rhamnopyranosyl-β-D-glucopyranosyl- (rutinosyl)

Scheme 2. Reaction scheme for an analysis of glycosidase-catalyzed rutinosylation of *p*-hydroxycinnamic acid.

The analysis was divided into two steps: step 1, in which the enzyme and carboxylic acid form an ES complex, and step 2, in which the ES complex reacts with rutin, as shown in Table 2. As in the analysis of scheme1, n_I was set to be 2 and the molecular weight was not adjusted by the number of rotatable bonds. As the result, the predicted and literature yields were in very good agreement. Other examples of the reaction using

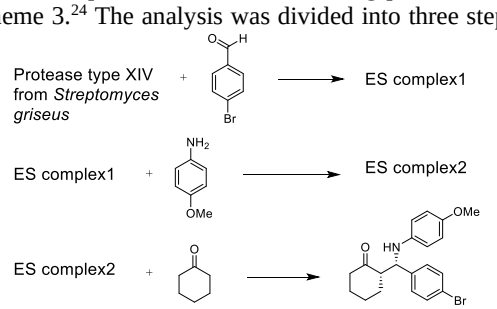
Table 2. An example of analysis of rutinosylation using rutinosidase.

step	A	B	M_B[a]	X^b/n_I	$Y_{lit.}$[c]	Y_{int}[d]	$Y_{pred.}$[e]
1	rutinosidase from *Aspergillus niger*	(*E*)-*p*-coumaric acid	164	82		85	
2	ES complex	rutin	613	306	36	43	36

[a]molecular weight of the substrate B. [b]$X=M_B$. [c]total of the literature yield of four products. [d]predicted yield of each step. [e]predicted yield of the final product.

glycosidase were analyzed in the same manner and the results are summarized in Table 6 after removal of outliers by Smirnov-Grubbs test as a significance level of 0.05. Literature versus predicted yield plot is included in Figure 1 and expressed as pink open circle (○).

Third, Michael addition/cyclization and Mannich reaction using protease (EC number 3.4.-) were analyzed.[24-26] As an example, Mannich reaction using protease is shown in Scheme 3.[24] The analysis was divided into three steps: step 1,



Scheme 3. Reaction scheme for an analysis of protease-catalyzed Mannich reaction.

in which the enzyme and 4-bromobenzaldehyde form an ES complex1, step 2, in which the ES complex1 reacts with 4-methoxyaniline to form an ES complex2, and step 3, in which the ES complex2 reacts with cyclohexanone, as shown in Table 3. As in the analysis of above examples, n_I was set to be 2 and the molecular weight was not adjusted by the number of rotatable bonds. As the result, the predicted and literature yields were in very good agreement. Other examples of the

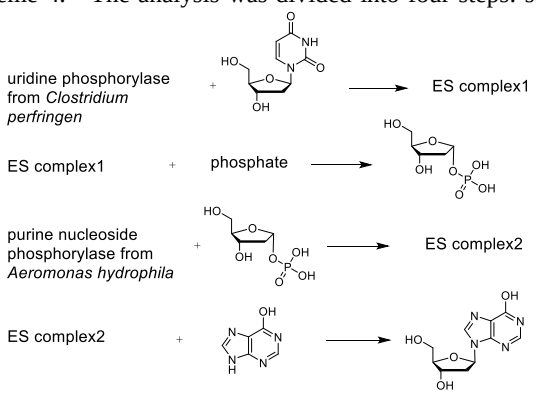
Table 3. An example of analysis of Mannich reaction using protease.

step	A	B	M_B[a]	X^b/n_I	$Y_{lit.}$[c]	Y_{int}[d]	$Y_{pred.}$[e]
1	protease type XIV from *Streptomyces griseus*	4-bromo-benzaldehyde	185	93		83	
2	ES complex1	4-methoxy-aniline	123	62		89	
3	ES complex2	cyclohexanone	98	49	66	91	67

[a]molecular weight of the substrate B. [b]$X=M_B$. [c]literature yield of the final product. [d]predicted yield of each step. [e]predicted yield of the final product.

reaction using protease were analyzed in the same manner and the results are summarized in Table 6 after removal of outliers by Smirnov-Grubbs test as a significance level of 0.05. Literature versus predicted yield plot is included in Figure 1 and expressed as red circle (●).

Fourth, phosphorylations using phosphorylase (EC number 2.4.-) were analyzed.[27-29] As an example, transglycosylation reaction using phosphorylase is shown in Scheme 4.[27] The analysis was divided into four steps: step 1,



Scheme 4. Reaction scheme for an analysis of phosphorylase-catalyzed transglycosylation of deoxyuridine

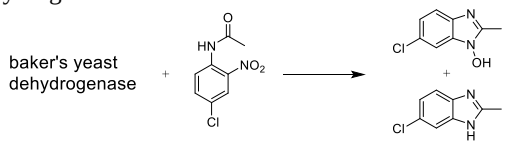
in which the uridine phosphorylase and 2'-deoxyuridine form an ES complex1, step 2, in which the ES complex1 reacts with phosphate ion to form 2'-deoxyribose 1-phosphate, step 3, in which the purine nucleoside phosphorylase reacts with 2'-deoxyribose 1-phosphate to form an ES complex2, and step 4, in which the ES complex reacts with hypoxanthine, as shown in Table 4. As in the analysis of above examples, n_I was set to be 2 and the molecular weight was not adjusted by the number of rotatable bonds. As the result, the predicted and literature yields were in very good agreement. Other examples of the reaction using phosphorylase were analyzed in the same manner and the results are summarized in Table 6 after removal of outliers by Smirnov-Grubbs test as a significance level of 0.05. Literature versus predicted yield plot is included in Figure 1 and expressed as green triangle (▲).

Table 4. An example of analysis of transglycosylation using phosphorylase.

step	A	B	M_B[a]	X^b/n_I	$Y_{lit.}$[c]	Y_{int}[d]	$Y_{pred.}$[e]
1	uridine phosphorylase from *Clostridium perfringens*	2'-deoxyuridine	228	114		79	
2	ES complex1	phosphate	96	48		91	
3	purine nucleoside phosphorylase from *Aeromonas hydrophila*	2'-deoxyribose 1-phosphate	213	107		80	
4	ES complex2	hypoxanthine	136	68	50	87	50

[a]molecular weight of the substrate B. [b]$X=M_B$. [c]literature yield of the final product. [d]predicted yield of each step. [e]predicted yield of the final product.

Fifth, reduction of nitro and azide groups and oxidation of alcohol using dehydrogenase (EC number 1.1.-) were analyzed.[30-33] As an example, reduction of nitro group using dehydrogenase is shown in Scheme 5.[31]



Scheme 5. Reaction scheme for an analysis of reductive cyclization of *p*-chloro *o*-nitroacetanilide by baker's yeast.

The analysis was conducted as one step, in which the baker's yeast dehydrogenase and the substrate form an ES complex to react with NADPH in the dehydrogenase, as shown in Table 5. As in the analysis of scheme1, n_I was set to be 2 and the molecular weight was not adjusted by the number of rotatable bonds. As the result, the predicted and literature yields were in very good agreement. Other examples of the reaction using

Table 5. An example of analysis of reduction using dehydrogenase.

step	A	B	M_B[a]	X^b/n_I	$Y_{lit.}$[c]	Y_{int}[d]	$Y_{pred.}$[e]
1	baker's yeast alcohol dehydrogenase	*p*-chloro *o*-nitro-acetanilide	215	107	79	80	80

[a]molecular weight of the substrate B. [b]$X=M_B$. [c]total of the literature yields of two products. [d]predicted yield of each step. [e]predicted yield of the final product.

dehydrogenase were analyzed in the same manner and the results are summarized in Table 6 after removal of outliers by Smirnov-Grubbs test as a significance level of 0.05. Literature versus predicted yield plot is included in Figure 1 and expressed as sky blue square (■).

As shown above, the literature yields and predicted yields showed good agreement in all cases when the value of n_I was set to 2. When n_I is set to 1, the coefficient of determination decreases and the regression coefficient decreases significantly, indicating that this value is inappropriate, and that the enzymatic reaction can be explained by the value set for the intramolecular reaction. When n_I was set to 2 and M_B was adjusted by NORB, the regression coefficient approached 1, but there was no improvement in the coefficient of determination. To set the regression coefficient to 1, it is possible to set the n_I value to 2.7, and the coefficient of determination to the maximum value of 0.95, which seems to indicate that the reaction probability is even higher than in the case of intramolecular reactions (n_I=2).[34] These indicate that the structural change of the substrate is limited, i.e., entropy is reduced, which can be understood as a characteristic of enzymatic reactions. In summary, equation (1) is not only applicable to general organic chemical reactions including enzymatic reactions, but also can be applied to enzymatic reactions with the simplified equation (2) when the molecular weight of the substrate is small. It is worth noting that these equations can be applied to the reaction of ES complex formation. In other words, the n_I value in the equations can be said to represent specificity and affinity, so it is likely to be applicable to other biochemical reactions such as antigen-antibody reaction and signal transduction, although the n_I value needs to be set to a larger value.

Table 6. Analysis of enzymatic reaction.

EC number	1.1.-	2.4.-	3.1.-	3.2.-	3.4.-	all			
X[a]	M_B[b]					M_B[b]		M'_B[c]	
n_I[d]	2					2	1	2	1
sample size	13	26	21	12	44	116	116	116	116
coefficient of determination (R^2)	0.97	0.88	0.90	0.82	0.94	0.95	0.86	0.94	0.91
regression coefficient (β)	0.94	0.76	0.87	0.90	0.98	0.90	0.57	0.96	0.67

[a]X: M_B or M'_B. [b]M_B: unadjusted with NORB. [c]M'_B: adjusted with NORB. [d]coefficient to distinguish where intermolecular reaction (n_I=1) or intramolecular reaction (n_I=2).

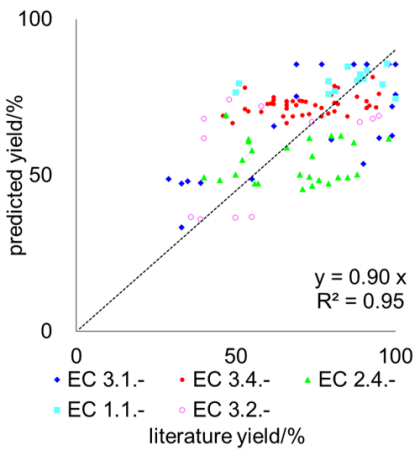


Figure 1. Literature versus predicted yield plot of enzymatic reaction. EC 3.1.-: lipase, EC 3.2.-: glycosidase, E 3.4.-: protease, EC 2.4.-: phosphorylase, EC 1.1.-: dehydrogenase.

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 M. Kawashima, *ChemRxiv*. **2021**, Preprint. https://doi.org/10.26434/chemrxiv.13536080.v1

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 M. Sari, Y. Chung, F. Agatha, H. K. Kim, *Appl. Biochem. Biotechnol.* **2005**, *127*, 17.

7 P.-J. Yao, G.-J. Huang, W. Yan, X. Zhang, Q. Li, Y.-A. Wei, *Adv. Chem. Engineer. Sci*. **2012**, *2*, 204.

8 P. Sun, Y. Chen, H. Wang, J. Li, J. Gao, H. Wang, X. Zheng, S. Zhang, *Eur. Food Res. Technol.* **2011**, *233*, 253.

9 E. Heŕmánková-Vavŕíková, A. Kŕenková, L. Petrásková, C. S. Chambers, J. Zápal, M. Kuzma, K. Valentová, V. Kŕen, *Int. J. Mol. Sci*. **2017**, *18*, 1074.

10 R. Awang, M. Basri, S. Ahmad, A. B. Salleh, *J. Oleo Sci*. **2005**, *54*, 305.

11 V. Osuna1, I. A. Rivero, *J. Mex. Chem. Soc*. **2012**, *56*, 176.

12 C. H. Kwon, J. H. Lee, S. W. Kim, J. W. Kang, *J. Microbiol. Biotechnol.* **2009**, *19*, 1596.

13 V. M Almeida, C. R. Branco, S. A Assis, I. J. Vieira, R. Braz-Filho, A. Branco, *Chem. Cent. J.* **2012**, *6*, 41.

14 F. Monot, F. Borzeix, M. Bardin, J.-P. Vandecasteele, *Appl Microbiol. Biotechnol.* **1991**, *35*, 759.

15 M. H. Sipponen, M. Farooq, J. Koivisto, A. Pellis, J. Seitsonen**,** M. Österberg, *Nat. Commun.* **2018**, *9*, 2300.

16 A. Schär, L. Nyström, *Eur. J. Lipd Sci. Technol.* **2016**, *118*, 1557.

17 C. Hidayat, K. Fitria, P. H. Supriyanto, *Agric. Agric. Sci. Procedia* **2016**, *9*, 353.

18 P. Youa, E. Sua, X. Yangb, D. Maob, D. Weia, *J. Mol. Catal. B Enzym.* **2011**, *71*, 152.

19 I. Bassanini, J. Kapešová, L. Petrásková, H. Pelantová, K. Markošová, M. Rebroš, K. Valentová, M. Kotik, K. Káňová, P. Bojarová, J. Cvačka, L. Turková, E. E. Ferrandi, I. Bayout, S. Riva, V. Křena, *Adv. Synth. Catal*. **2019**, *361*, 2627.

20 D. Bilaničova1, V. Mastihuba, M. Mastihubova, J. Balešova**,** Š. Schmidt, *Czech J. Food Sci*. **2010**, *28*, 69.

21 M. Y. Rather, S. Mishra, V. Verma, S. Chand, *Bioresour. Technol*. **2012**, *107*, 287.

22 K. Lirdprapamongkol, J. Svasti, *Biotechnol. Lett*. **2000**, *22*, 1889.

23 Z. Tao, J. Liu, Y. Jiang, L. Gong, B. Yang, *Sci. Rep.* **2017**, *7*, 12445.

24 Y. Xue, L.-P. Li, Y.-H. He, Z. Guan, *Nature* **2012,** *2*, 761.

25 X. Ding, X.-D. Zhang, C.-L. Dong, Z. Guan, Y.-H. He, *Catal. Lett.* **2018**, *148*, 757.

26 C. Li, Y.-J. Zhou, N. Wang, X.-W. Feng, K. Li, X.-Q. Yu, *J. Biotechnol.* **2010**, *150*, 539.

27 L. Tamborini, C. Previtali, F. Annunziata, T. Bavaro, M. Terreni, E. Calleri, F. Rinaldi, A. Pinto, G. Speranza, D. Ubiali, P. Conti, *Molecules* **2020**, *25*, 1223.

28 T. Nihira, Y. Saito, K. Ohtsubo, H. Nakai, M. Kitaoka, *PLoS One* **2014**, *9*, e86548.

29 A. Hatano, H. Wakana, N. Terado, A. Kojima, C. Nishioka, Y. Iizuka, T. Imaizumi, S. Uehara, *Catal. Sci. Technol.* **2019**, *9*, 5122.

30 M. M. Cajnko, U. Novak, M. Grilc, B. Likozar, *Biotechnol. Biofuels* **2020**, *13*, 66.

31 A. Navarro-Ocaña, L. F. Olguín, H. Luna, M. Jiménez-Estrada, E. Bárzana, *J. Chem. Soc., Perkin Trans*. 1, **2001**, 2754.

32 W. Baik, D. I. Kim, H. J. Lee, W.-J. Chung, B. H. Kim, S. W. Lee, *Tetrahedron Lett.* **1997**, *38*, 4579.

33 A. Kamal, Y. Damayanthi, B. S. N. Reddy, B. Lakminarayana, B. S. P. Reddy, *Chem. Commun*. **1997**, 1015.

34 To calculate the predicted yield from M'_{AB}, the NORB (R_A) of the enzyme was first obtained by multiplying the number of amino acid residues of the enzyme by the average NORB of 20 amino acids (2.9). From the R_A of the enzyme and the R_B of the substrate, M_A and M_B were adjusted to obtain M'_A and M'_B, respectively, and M'_{AB} was calculated from these adjusted molecular weights and applied to equation (1) to calculate the predicted yield. The error was very small compared to the case when equation (2) was used (M_B or M'_B was used as variable X in equation (2)) because of the reason that the ratio of M'_A to M_B was 57 at least.