Pergamon

Tetrahedron: *Asymmetry* 13 (2002) 377–382

TETRAHEDRON: *ASYMMETRY*

Stereoselectivity in biocatalytic enantioconvergent reactions and a computer program for its determination

Kurt Faber and Wolfgang Kroutil*

Department of Chemistry, Organic & Bioorganic Chemistry, University of Graz, Heinrichstrasse 28, 8010 Graz, Austria

Received 17 January 2002; accepted 18 February 2002

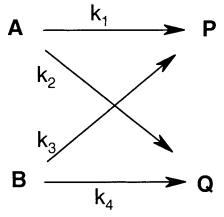
Abstract—For the description of the stereoselectivity of (bio)catalytic asymmetric reactions which may proceed via different regio- or stereo-isomeric pathways (e.g. catalysed by epoxide hydrolases, dehalogenases, sulfatases or glycosidases), a parameter '*RI*' (Retention–Inversion ratio) was introduced in analogy to the Enantiomeric Ratio (*E*), which describes enantioselectivity. A computer program was developed for the treatment of the kinetics of such single-step processes, which offer the potential of deracemization, i.e. a single stereoisomeric product is formed from a racemate in an enantioconvergent fashion. By analysis of experimentally determined progress curves of the enantiomeric excess of substrate and product (e.e.$_S$, e.e.$_P$, respectively) and the conversion (*c*), relative first-order rate constants k_i, the enantioselectivity (*E*) and the *RI* ratio (*RI*) can be determined; in addition, processes can be simulated based on assumed k_i values.

1. Introduction

Due to the diastereomeric nature of the interaction of enantiomers with a chiral catalyst, enantiomers are generally transformed at different reaction rates. If the difference between the rates of reaction of the two enantiomers is large enough, one enantiomer can be entirely transformed to the corresponding product, while its counterpart remains untouched, which allows their separation. This process is generally referred to as kinetic resolution. The generally accepted parameter for the description of the quality (i.e. 'enantioselectivity') of such processes is the so-called Enantiomeric Ratio (*E*),[1–3] which is equal to the ratio of the reaction rates of the enantiomers.

A crucial prerequisite for the applicability of *E* values is the fact that both enantiomers must be transformed through identical stereochemical pathways, which (for the overwhelming majority of cases) involves retention of configuration at the stereogenic centre(s). This precondition is usually fulfilled for the majority of biotransformations, such as hydrolysis/condensation reactions catalysed by proteases, esterases and lipases.

In contrast, there are several types of enzymes which are able to act through more than a single stereochemical pathway: In particular, sulfatases,[4,5,†] dehalogenases,[6,7,‡] glycosidases[8,§] and epoxide hydrolases[9,¶] show the potential to convert both substrate-enantiomers through *two* stereochemical pathways, leading to *opposite* product enantiomers as generalised in Scheme 1.



Scheme 1. General pathways and annotation of first-order rate constants k_i.

† Depending on the enzyme type, sulfatases may act via retention (breaking the O–S bond) or inversion of configuration (breaking the C–O bond).

‡ Depending on the enzyme type, hydrolysis of the C–Hal bond may occur with inversion or retention of configuration.

§ Although the formation/cleavage of glycosidic bonds may proceed via retention or inversion, these reactions have to be considered as a separate case, since diastereomers are involved rather than enantiomers.

¶ Epoxide hydrolases transform an epoxide through retention or inversion of a stereogenic centre, which equals a reaction proceeding through opposite regioselectivity via attack at either oxirane carbon atom.

* Corresponding author. Tel.: +43-(0)316-380-5350; fax: +43-(0)-316-380-9840; e-mail: wolfgang.kroutil@uni-graz.at

0957-4166/02/$ - see front matter
PII: S0957-4166(02)00084-8

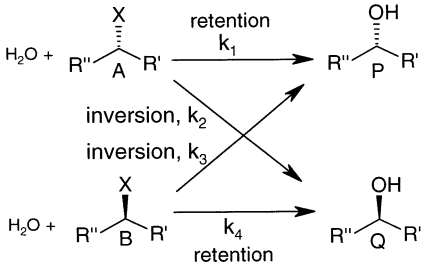
As a consequence, the e.e. of the product is not only a function of the enantioselectivity (E) and the conversion, but also of the ratio of retention to inversion (RI ratio) of each enantiomer—described as the ratios of stereochemical pathways k_1/k_2 and k_4/k_3. Thus, equations for the determination of the E value in kinetic resolutions based on e.e.$_P$—i.e. $E=f(\text{e.e.}_S, \text{e.e.}_P)$ or $f(\text{e.e.}_P, c)$—are not applicable. However, since the stereoselectivity of pathways has no influence on the enantiomeric composition of the *substrate*, $E=f(\text{e.e.}_S, c)$ can still be used. In this case, E equals $(k_1+k_2)/(k_4+k_3)$ and it describes the ability of the chiral catalyst to differentiate between substrate enantiomers.

On first glimpse, the transformation of substrate enantiomers through two stereochemical pathways—leading to opposite product enantiomers—appears rather complex and difficult to control. However, if the stereochemical pathways of enantiomers are *opposite*, e.g. $k_1>k_2$ and $k_3>k_4$ (Scheme 1), both substrate enantiomers yield the *same* stereoisomeric product in 100% theoretical yield and so-called deracemization can be achieved.[10] The overall performance of an enantioconvergent process is determined by the matching interplay of all four rate constants k_i and two main parameters are required for their characterisation: (i) enantioselectivity and (ii) RI ratio.

Whereas the enantioselectivity E denotes the relative rate of enantiomers, i.e. $(k_1+k_2)/(k_3+k_4)$, the RI ratio stands for the relative flux of materials through different stereochemical pathways. In order to describe the latter parameter, we propose the parameter 'RI'—in analogy to the E value—defined as the ratio of relative rates for opposite stereochemical pathways. For enantiomers **A** and **B**, $RI_A=k_1/k_2$ (or k_2/k_1 so that $RI\geq 1$) and RI_B is defined by analogy.

For a full and accurate description of the quality of such processes, the determination of not only the *enantio*, but also the retention–inversion ratio (RI ratio) is required.

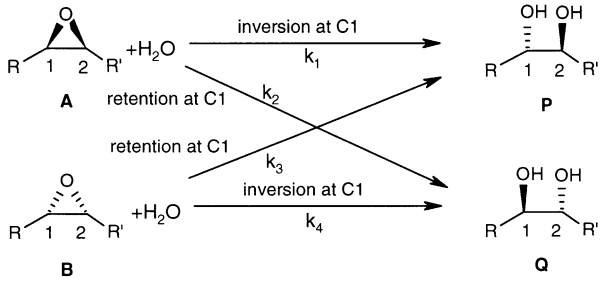
The principle of enantioconvergence may be applied to the following biocatalytic systems:



Scheme 2. Possible pathways for reactions of sulfatase and dehalogenase enzymes.

(i) In the case of dehalogenases and sulfatases—which may act via retention or inversion of configuration (Scheme 2)—the ratios k_1/k_2 and k_4/k_3 denote the number of molecules (issued from **A** or **B**) transformed through retention or inversion.
(ii) The general system described in Scheme 1 may likewise be applied to another type of enzyme: epoxide hydrolases. These enzymes have recently attracted much attention, particularly in view of the facile production of nonracemic *vic*-diols from (±)-epoxides (see Scheme 3).[11] In general, the oxirane carbon atom being attacked undergoes inversion of configuration.‖ In this case, however, the RI_A of pathways denotes the *regioselectivity*, i.e. the ratio of the attack at both oxirane carbon atoms (for instance, C(1) versus C(2), Scheme 3).** The overall type of kinetics and the option to transform both epoxide enantiomers into a single stereoisomeric *vic*-diol remains the same. In the case of 2,3-di- and trisubstituted oxiranes, the existence of all four possible reaction pathways as depicted in Scheme 3 was experimentally proven.[13]

In a previous study we presented a method to determine the relative first-order rate constants k_i (Scheme 1) of the enzymatic opening of 2,3-dialkyl substituted epoxides performed in ^{18}O-labelled water by using chiral GC–MS analysis of the *vic*-glycol products.[13] These rate constants, k_i, offer the possibility to describe the regio- and enantioselectivity. The method showed excellent reproducibility and a small experimental error. Furthermore, the regioselectivity for both enantiomers could be determined simultaneously based on experiments using racemic starting material. This is a significant advantage over a procedure that requires the separate hydrolysis of enantiomers in pure form.[14] A related method to determine the regioselectivity of epoxide hydrolase-catalysed reactions



Scheme 3. Possible pathways for biocatalytic hydrolysis of epoxides.

‖ Depending on the substitution pattern, this inversion may not be effective, i.e. it remains 'invisible' and apparent retention is observed.

** For bacterial epoxide hydrolases it was shown that for monoalkyl and 2,2-dialkyl oxiranes the attack of the (formal) hydroxyl ion occurs exclusively at the non-substituted carbon.[12]

using ^{18}O-labelled water (or labelled substrates) has also been described for enzymes of fungal,[15] insect,[16] plant[17] or mammalian[18] origin. More recently, a method was published, which calculated the regioselectivity (expressed as 'regioselective coefficients α') by non-linear regression using commercial software[††] based on the Marquardt–Levenberg algorithm.[19,20] α-Coefficients were obtained by non-linear regression of fitting the plots of $e.e._{(S,P)}$ versus conversion. However, non-linear curve-fitting is very sensitive to experimental errors arising from sample manipulation leading to unreliable results.

These considerations prompted us to develop a model, which shows the following advantages:

(i) Easily determined datasets of e.e. values at different conversion ($e.e._S$, $e.e._P$, conversion) are used to calculate the key parameters which allow for the description of the enantioselectivity E and the retention–inversion ratio (RI ratio), respectively.
(ii) By combining the non-linear curve fitting method with our previously reported system, the problem was reduced to a linear fit.
(iii) By using all data sets simultaneously, the error could be minimised.
(iv) A computer program was written which enables not only the calculation of kinetic constants by rapid analysis of experimental data, but also the simulation of the stereochemical outcome of such processes based on assumed k_i values.

2. Proposal of model for analysis

2.1. Definition of the *RI* value

For the description of regioselectivity, the so-called α value has been proposed.[19] It is defined as the *percentage* of molecules, which are transformed via a certain pathway in relation to the sum. For instance, for **A** (according to Scheme 3) α_A equals $100\times k_1/(k_1+k_2)$ [%]. For practical reasons—such as in case of the asymmetric (bio)hydrolysis of 2,3-dialkyl or non-symmetrical tri- and tetra-substituted oxiranes, the definition of α (denoted as a fraction of a sum) may cause problems, since it is not clear which regiospecific pathway α refers to and a detailed pathway analysis is required.

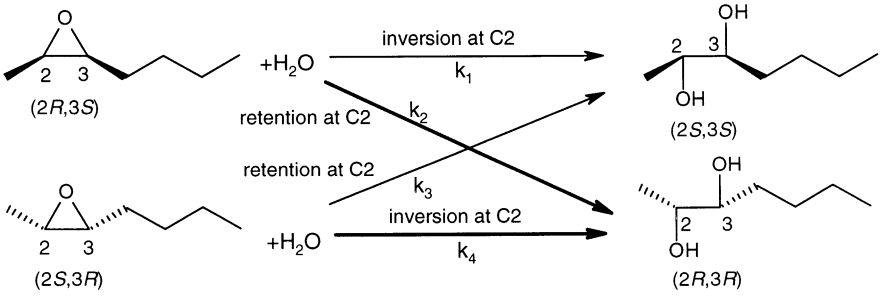
Since the enantiomeric ratio (E) for the description of enantioselectivity is defined as the *ratio* of the reaction rates of enantiomers [i.e. $(k_1+k_2)/(k_4+k_3)$, Scheme 3] rather than the *fraction* of its sum, we found it more appropriate, to define a parameter for regioselectivity (RI ratio) for epoxide hydrolases accordingly for the sake of clarity. Thus, the parameter RI (retention–inversion ratio) is used throughout this study: $RI_A=k_1/k_2$ and $RI_B=k_4/k_3$ or its reciprocal value, so that $S\geq 1$. For full characterisation, the absolute configuration of the stereogenic centre being preferentially attacked should be added in brackets. In order to fully describe the stereoselective pathways for processes following Scheme 1, three parameters are needed, and we propose that these should be RI_A, RI_B and E.

As an example, the complete kinetic characterisation of the enzymatic hydrolysis of (±)-*cis*-2-heptene oxide employing *Rhodococcus ruber* DSM 43338 is given (Scheme 4).[13]

In this case, '$RI_{(2S,3R)}$' describes the regioselectivity for the 2*S*,3*R* isomer, and (2*S*) denotes that the (2*S*,3*R*) enantiomer is preferentially attacked at its (2*S*) centre.

In this particular case, attack at the (2*S*) centre occurs 498 times faster than at the (3*R*) counterpart. The absolute configuration noted in conjunction with the enantioselectivity (E) indicates that the (2*S*,3*R*)-configured substrate enantiomer reacts faster.

It is a puzzling fact, that for an ideal enantioconvergent system, *low* enantioselectivity E—which ensures that both substrate enantiomers (**A**+**B**) are transformed at comparable rates—is desirable to create a *fast* overall process. To obtain a stereoisomeric product in high enantiomeric excess, the stereoselectivity should be *high* and *opposite*, meaning on the level of rate constants that e.g. k_2 and k_4 should be much greater than k_1 and k_3. In order to illustrate these phenomena, a comparison of processes is given below.



Scheme 4. Enzymatic hydrolysis of (±)-*cis*-2-heptenoxide as an example. $RI_{(2S,3R)}=498$ (2*S*); $RI_{(2R,3S)}=4.5$ (3*S*); $E=6.5$ (2*S*,3*R*).

[††] Jandel Sigmaplot®.

The key data of the enzymatic hydrolysis of (±)-*cis*-2-heptene oxide[13] can be given as shown in Table 1.

The hydrolysis of (±)-*cis*-2-heptene oxide (Table 1) by *R. ruber* DSM 43338 and DSM 44540 are examples for a process with moderate E value but (at least one) excellent regioselectivity, which still resulted in the formation of the corresponding (2*R*,3*R*)-heptanediol in 91% e.e. and 79% isolated yield for *R. ruber* DSM 44540.

The system can of course also be applied to tri- or tetra-substituted oxiranes, such as 2-methyl-2,3-epoxynonane, using the first-order relative rate constants published before.[21]

The hydrolysis of (±)-2-methyl-2,3-epoxynonane using *Streptomyces lavendulae* ATCC 55209 (Table 2) represents an ideal case of an enantioconvergent process. Low enantioselectivity ($E=4.2$) and excellent opposite regioselectivity for each enantiomer ensures the formation of (3*R*)-2-methyl-nonane-2,3-diol in 97% e.e. and 60% conversion. Although the regioselectivity $RI_{(S)}$ of *M. paraffinicum* NCIMB 10420 for (*S*)-2-methyl-2,3-epoxynonane is excellent, the regioselectivity for the (*R*)-enantiomer is low, resulting in the formation of (3*R*)-2-methyl-nonane-2,3-diol in reduced e.e. (83%) at 90% conversion. Without all three values, enantioselectivity and two values for regioselectivity, a critical assessment of these processes is impossible.

3. Features of the computer program '*Stereo*'©

The computer program was written in Visual Basic 6.0 and is available as shareware from the ftp server <ftp://borgc185.kfunigraz.ac.at>.‡‡

3.1. Abbreviations and annotations

In the following formulas, **A**, **B**, **P** and **Q** stand for the concentration of the respective species for substrate (**A**, **B**) and product enantiomers, (**P**, **Q**), respectively.

Constants k_1 through k_4 (k_i) are calculated as relative first-order rate constants. The conversion of the reaction (c) is defined as $100\times(\mathbf{P}+\mathbf{Q})/(\mathbf{A}_0+\mathbf{B}_0)$ [%], where $\mathbf{A}_0$ and $\mathbf{B}_0$ are the concentration of **A** and **B** at the onset of the reaction. Racemic starting material is assumed, thus $\mathbf{A}_0=\mathbf{B}_0$.

The enantioselectivity is denoted as the enantiomeric ratio (E), defined as the ability of the (bio)catalyst to discriminate between the two substrate enantiomers, thus $E=(k_1+k_2)/(k_3+k_4)$. E.e.$_S$ and e.e.$_P$ are the enantiomeric excess of the substrate and product, respectively.

3.2. Analysis

The relative first-order rate constants k_i can be calculated by using datasets consisting of **A**, **B**, **P**, **Q** and c; at least three sets are required and data input is performed using a spreadsheet. The data for **A**, **B**, **P**, **Q** consist of the respective concentrations or equivalents thereof (such as percentage, area counts from GC/HPLC analysis). These data are used to calculate the e.e. In order to avoid mistakes caused by a wrong algebraic sign, the 'e.e.' is not the direct input for the spreadsheet, since during further calculations, negative e.e. values also have to be used. The latter indicate the opposite enantiomer being in excess. The datasets should be chosen in such a way, that the corresponding conversion is within a range of 25–75% to minimise errors. By using the button 'Calculate', the k_i values are

Table 1. Stereoselectivities for the enzymatic hydrolysis of *rac*-*cis*-2-heptene oxide

Biocatalyst	c (%)	E.e. (%)	E	$RI_{(2S,3R)}$	$RI_{(2R,3S)}$
R. ruber DSM 43338	90	83	6.5 (2*S*,3*R*)	498 (2*S*)	4.5 (3*S*)
R. ruber DSM 44540	79	91	14.2 (2*S*,3*R*)	326 (2*S*)	2.8 (3*S*)
R. equi IFO 3730	80	65	4 (2*S*,3*R*)	39 (2*S*)	1.5 (3*S*)

Table 2. Stereoselectivities for the enzymatic hydrolysis of 2-methyl-2,3-epoxynonane

Biocatalyst	c (%)	E.e. (%)	E	$RI_{(S)}$	$RI_{(R)}$
S. lavendulae ATCC 55209	60	97	1.3 (*S*)	126 (*S*)	99 (C-2)
M. paraffinicum NCIMB 10420	90	83	4.2 (*S*)	723 (*S*)	6.3 (C-2)

‡‡ We have tried our best to ensure maximum program stability. However, we strongly encourage all users to report any kinds of incompatibilities, malfunctions and/or errors to wolfgang.kroutil@uni-graz.at.

calculated as well as the corresponding values for the enantioselectivity (E) and the RI ratio.

3.3. Simulation

Starting from calculated or assumed first-order rate constants, plots of e.e. versus conversion or e.e. and concentration versus time can simulated.

The validity of the computer program was tested for the enantioconvergent hydrolysis of trisubstituted epoxides.[21] Comparison of the calculated parameters with data obtained from independent studies performed in ^{18}O-labelled water showed excellent compatibility.

4. Theory

In order to allow a mathematical analysis, the following assumptions and preconditions were made:

1. The reaction pathways can be correlated to Scheme 1.
2. The specific activity of the enzyme remains constant during the whole period of the reaction, implying that enzyme deactivation—caused by pH, temperature, chemical or mechanical stress—can be omitted.
3. Absence of inhibition.
4. Spontaneous (non-biocatalysed) reactions can be neglected.
5. All reactions are irreversible.

From Scheme 1 it can be deduced that the enantioselectivity E equals $(k_1+k_2)/(k_3+k_4)$ or its reciprocal value, depending on which one is greater than 1.

$$E=\frac{k_1+k_2}{k_3+k_4} \text{ or } E=\frac{k_3+k_4}{k_1+k_2} \tag{1}$$

Following the approach of Moussou et al.,[19] 'regioselectivity coefficients' (α) can be deduced [Eqs. (2) and (3)]:

$$\alpha_A=\frac{k_1}{k_1+k_2} \tag{2}$$

$$\text{and } \alpha_B=\frac{k_4}{k_3+k_4} \tag{3}$$

After a simple transformation using Eq. (7) from the above-mentioned study, one obtains the following equation (using $\alpha_{(S)}=\alpha_A$ and $\alpha_{(R)}=\alpha_B$):

$$\underbrace{ee_P c - ee_S + ee_S c}_{a} = \underbrace{\alpha_A(c - ee_S + ee_S c)}_{b} + \underbrace{\alpha_B(ee_S c - c - ee_S)}_{c} \tag{4}$$

In order to simplify Eq. (4) for further calculations, the parameters a, b and c are introduced. Since several datasets of e.e.$_S$, e.e.$_P$, and c are used for the calculations, several equations are subsequently obtained and thus the parameters a, b and c are indexed in general a_i, b_i and c_i. By taking two subsequent equations:

$$a_i=b_i\alpha_A+c_i\alpha_B \tag{5}$$

$$a_{i+1}=b_{i+1}\alpha_A+c_{i+1}\alpha_B \quad i=1 \text{ to } (n-1) \tag{6}$$

one can easily eliminate α_A to obtain Eq. (7).

$$\underbrace{(c_{i+1}b_i - c_i b_{i+1})}_{x_i}\alpha_B = \underbrace{a_{i+1}b_i - a_i b_{i+1}}_{y_i} \quad \text{for i = 1 to (n-1)} \tag{7}$$

By writing the system in the above manner, one would obtain $(n-1)$ equations. Of course, it is possible to get many more equations by combining all datasets with each other—e.g. by combining dataset i with datasets $(i+2)$ or $(i+3)$—but experiments have shown, that this does not improve the accuracy of the results.

Since Eq. (7) is a linear equation system, α_B can be calculated by using a simple linear regression (Eq. (8)). By applying the same method, α_A is obtained.

$$\alpha_B=\frac{\sum_{i=1}^{n-1} y_i x_i}{\sum_{i=1}^{n-1} x_i^2} \tag{8}$$

Since we are only interested in relative k_i values rather than absolute ones, k_3 is arbitrarily set to 1. Knowing α_B (Eq. (8)) we can use Eq. (3) to obtain k_4.

In the next step, the enantioselectivity E—corresponding to $(k_1+k_2)/(k_3+k_4)$ or $(k_3+k_4)/(k_1+k_2)$—is calculated from the known equations as a function of e.e.$_S$ and c, again by applying linear regression using all data sets. Combining this calculation with k_3 (set to 1), k_4 and Eq. (1) we obtain (k_1+k_2). Knowing (k_1+k_2) we calculate k_1 from Eq. (2). Now k_2 can be obtained. The last step consists of the determination which of the k_i is the smallest and to divide all k_i by k_{smallest}, so that the smallest k_i value equals 1.

In order to minimise errors and to obtain maximal accuracy, it is recommended to use only datasets at conversions within a range of 25–75% for the calculation.

Acknowledgements

This work was performed within the Spezialforschungsbereich 'Biokatalyse' and financial support from the Fonds zur Förderung der Wissenschaftlichen Forschung (Project no. F-104) is gratefully acknowledged.

References

1. Chen, C. S.; Fujimoto, Y.; Girdaukas, G.; Sih, C. J. *J. Am. Chem. Soc.* **1982**, *104*, 7294–7299.
2. Straathof, A. J. J.; Jongejan, J. A. *Enzyme Microb. Technol.* **1997**, *21*, 559–571.
3. Enantioselectivity E is also called s factor or k_{rel} in non-enzymatic kinetic resolutions: Kagan, H. B.; Fiaud, J.-C. *Top. Stereochem.* **1988**, *18*, 249–330.
4. Spencer, B. *Biochem. J.* **1958**, *69*, 155–159.
5. Bartholomew, B.; Dodgson, K. S.; Matcham, G. W. J.; Shaw, D. J.; White, G. F. *Biochem. J.* **1977**, *165*, 575–580.
6. Soda, K.; Kurihara, T.; Liu, J.-Q.; Nardi-Dei, V.; Park, C.; Miyagi, M.; Tsunasawa, S.; Esaki, N. *Pure Appl. Chem.* **1996**, *68*, 2097–2103.
7. Weightman, A. J.; Weightman, A. L.; Slater, J. H. J. *Gen. Microbiol.* **1982**, *128*, 1755–1762.
8. Sinnott, M. L. *Chem. Rev.* **1990**, *90*, 1171–1202.
9. Nardini, M.; Ridder, I. S.; Rozeboom, H. J.; Kalk, K. H.; Rink, R.; Janssen, D. B.; Dijkstra, B. W. *J. Biol. Chem.* **1999**, *274*, 14579–14586.
10. Faber, K. *Chem. Eur. J.* **2001**, *7*, 5004–5010.
11. Orru, R. V. A.; Archelas, A.; Furstoss, R.; Faber, K. *Adv. Biochem. Eng. Biotechnol.* **1999**, *63*, 145–167.
12. Mischitz, M.; Mirtl, C.; Saf, R.; Faber, K. *Tetrahedron: Asymmetry* **1996**, *7*, 2041–2046.
13. Kroutil, W.; Mischitz, M.; Faber, K. *J. Chem. Soc., Perkin Trans. 1* **1997**, 3629–3636.
14. Zeldin, D. C.; Wei, S.; Falck, J. R.; Hammock, B. D.; Snapper, J. R.; Capdevila, J. H. *Arch. Biochem. Biophys.* **1995**, *316*, 443–451.
15. Pedragosa-Moreau, S.; Archelas, A.; Furstoss, R. *Bioorg. Med. Chem.* **1994**, *2*, 609–616.
16. Linderman, R. J.; Walker, E. A.; Haney, C.; Roe, R. M. *Tetrahedron* **1995**, *51*, 10845–10856.
17. Blée, E.; Schuber, F. *J. Biol. Chem.* **1992**, *267*, 11881–11887.
18. Dietze, E. C.; Kuwano, E.; Casas, J.; Hammock, B. D. *Biochem. Pharmacol.* **1991**, *42*, 1163–1175.
19. Moussou, P.; Archelas, A.; Baratti, J.; Furstoss, R. *Tetrahedron: Asymmetry* **1998**, *9*, 1539–1547.
20. Moussou, P.; Archelas, A.; Baratti, J.; Furstoss, R. *J. Mol. Catal. B* **1998**, *5*, 213–217.
21. Steinreiber, A.; Mayer, S. F.; Saf, R.; Faber, K. *Tetrahedron: Asymmetry* **2001**, *12*, 1519–1528. To avoid confusion, we replaced the α value in this paper by *RI*.