Quantitative Prediction of Rate Constants for Aqueous Racemization To Avoid Pointless Stereoselective Syntheses

Andrew Ballard, Hiwa O. Ahmad, Stefania Narduolo, Lucy Rosa, Nikki Chand, David A. Cosgrove, Peter Varkonyi, Nabil Asaad, Simone Tomasi, Niklaas J. Buurma,* and Andrew G. Leach*

Abstract: *Racemization has a large impact upon the biological properties of molecules but the chemical scope of compounds with known rate constants for racemization in aqueous conditions was hitherto limited. To address this remarkable blind spot, we have measured the kinetics for racemization of 28 compounds using circular dichroism and ^1H NMR spectroscopy. We show that rate constants for racemization (measured by ourselves and others) correlate well with deprotonation energies from quantum mechanical (QM) and group contribution calculations. Such calculations thus provide predictions of the second-order rate constants for general-base-catalyzed racemization that are usefully accurate. When applied to recent publications describing the stereoselective synthesis of compounds of purported biological value, the calculations reveal that racemization would be sufficiently fast to render these expensive syntheses pointless.*

Thalidomide racemizes in a matter of hours and yet it remains a poster child for enantioselective synthesis which would not have saved its victims.[1] The status quo in enantioselective synthesis thus ignores the cruel blind spot that we address in this paper: racemization.

Although necessary in dynamic kinetic resolution protocols,[2,3] racemization and epimerization can cause safe compounds to become toxic or lose efficacy,[4–11] lead to misidentification of chiral compounds extracted from natural sources,[12] etc. Ignoring racemization thus leads to wasted material and human resources.

Racemization is a particular problem because its detection requires chiral analytical methods.[13,14] Hence, few reports disclose rate constants for racemization under aqueous conditions.[1, 15–20] Chiral centers with certain combinations of substituents have been posited to be prone to general-base-catalyzed racemization although with little supporting data.[21–23]

We therefore classified stereogenic carbon atoms according to their attached substituents. Each substituent is identified as one of sixty types,[24] which encompass more than 99.95% of all such substituents in the GOSTAR database.[25] The ten most frequently occurring substituents are listed in Figure 1; the H required for general-base-catalyzed racemization is prominent.[24] Groups labelled * were selected for experimental study.

Based on prevalence, earlier work,[21–23] and chemical intuition, several compounds were selected for detailed kinetic studies.[26–28] The rate constants for general-base-catalyzed racemization were derived for a range of 11 arylglycine derivatives (**1**, **2** and **3**), 12 hydantoins (**4**, **5** and **6**) and 5 thiohydantoins (**7** and **8**). Briefly, at several buffer concentrations, circular dichroism spectroscopy (CD) followed the decrease in ellipticity and/or ^1H NMR spectroscopy the incorporation of D from deuterated buffers. The pseudo-
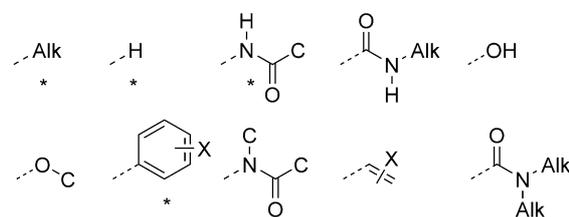
Figure 1. The ten substituents appearing most frequently adjacent to carbon stereogenic centers in the GOSTAR database. Alk=alkyl, C=carbon-linked alkyl or aromatic group and X=any group. *=selected for experimental study.

[*] Dr. A. Ballard, Dr. H. O. Ahmad, Dr. S. Narduolo, Dr. N. J. Buurma
Physical Organic Chemistry Centre
School of Chemistry, Cardiff University
Main Building, Park Place, Cardiff, CF10 3AT (UK)
E-mail: Buurma@Cardiff.ac.uk

Dr. H. O. Ahmad
Pharmaceutical Chemistry Department
College of Pharmacy, Hawler Medical University
Erbil, Kurdistan Region (Iraq)

L. Rosa, N. Chand, Dr. D. A. Cosgrove, Dr. N. Asaad
AstraZeneca Pharmaceuticals
Mereside, Alderley Park, Macclesfield, SK10 4TG (UK)

Dr. P. Varkonyi
AstraZeneca R+D
Pepparedsleden 1, 43183 Mölndal (Sweden)

Dr. S. Tomasi
AstraZeneca, Charter Way, Silk Road Business Park
Macclesfield, SK10 2NA (UK)

Dr. A. G. Leach
School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University
Liverpool, L3 3AF (UK)
E-mail: a.g.leach@ljmu.ac.uk

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
https://doi.org/10.1002/anie.201709163.

first-order rate constants for these processes were corrected for hydrolysis side reactions if required.[24] Plotting the first-order rate constants for racemization or H/D exchange against the concentration of the basic component of the buffer yielded the second-order rate constants for general-base-catalyzed racemization. These were corrected for reaction temperature and substrate protonation state.[24]

For predictive modeling, a mechanistic understanding is beneficial. Racemization of the stereogenic centers studied here could occur by either an S_E1 or an S_E2 mechanism. For hydantoins (e.g. **4–6**) both the S_E1 and S_E2 mechanisms have been proposed previously,[18,29] but we have shown that these reactions occur via the S_E1 mechanism.[30] Further, Hammett plots for **1a-h** show a positive slope and better correlation with σ^- than σ suggesting that a negative charge is formed on the stereogenic center in the rate-determining step of the racemization reaction, in line with an S_E1 mechanism.[24]

The experimental data were correlated with deprotonation energies ($\Delta\Delta G(R_1,R_2,R_3)$, Scheme 1) from B3LYP/6–$31 + G^{**}$ calculations incorporating aqueous solvation using the PCM protocol.[24] Second-order rate constants for general-base-catalyzed racemization, k_{gb}, correlate well with $\Delta\Delta G$-(R_1,R_2,R_3) for **1–3** and **4–8**.[24]

The set of compounds was supplemented with literature data for **9-16**,[24] leading to the relationship with $\Delta\Delta G$-(R_1,R_2,R_3) shown at the top of Figure 2. The line of best fit has equation $\log(k_{gb}) = -0.20 \times \Delta\Delta G(R_1,R_2,R_3) - 14.28$, with an R^2 value of 0.68 and root mean square error of 0.61, that is, reproducing rate constants to within a factor of approximately 4. Clopidogrel is excluded from this analysis due to large experimental uncertainties.[24]

The computational procedure was extended to include a group-contribution approach that is amenable to rapid analysis of chiral compounds and is described with examples in section S5. A simple representative (R) of each substituent type was selected and $\Delta\Delta G(R_1,R_2,R_3)$ computed with $R_1 = R$

1a (■), 1b (♦), 1c (▲), 1d (●), 1e (□), 1f (◇), 1g (△), 1h (○), 2 (■), 3a (■), 3b (♦), 4a (■), 4b (♦), 4c (▲), 4d (●), 4e (□), 4f (◇), 4g (△), 5a (■), 5b (♦), 5c (▲), 5f (◇), 6 (■), 7a (■), 7b (♦), 7c (▲), 7d (●), 8 (■), 9 (■), 10 (♦), 11 (▲), 12 (●), 13 (□), 14 (◇), 15 (△), 16 (○). Bottom: A (■), B (♦), C (▲), D (■), D' (◇), E (♦), F (▲), G (●), H (●), I (□), J (◇), K (□), L (△), M (○) and N (○).

Figure 2. Second-order rate constants for racemization under aqueous general-base-catalyzed conditions plotted against: computed $\Delta\Delta G$-(R_1,R_2,R_3) values (top) and $\Sigma\Delta\Delta G$ values (bottom). Clopidogrel (**16**) is excluded from the line.
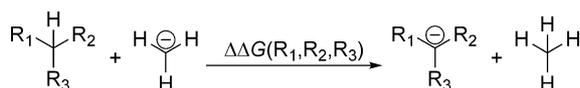
Scheme 1.

1
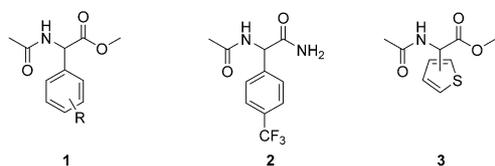
R = H (**a**); 4-OH (**b**); 4-Me (**c**); 4-F (**d**); 4-Cl (**e**); 3-F (**f**); 3-Cl (**g**); 4-CF₃ (**h**)

2

3

thiophen-2-yl (**a**); thiophen-3-yl (**b**)

4   5   6   7   8
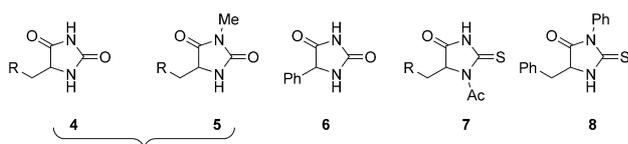
R = phenyl (**a**); 4-fluorophenyl (**b**); 4-methoxyphenyl (**c**); tetrahydrofuran-3-yl (**d**); OH (**e**); pyridin-3-yl (**f**); N-methylpyrrol-2-yl (**g**)

R = H (**a**); phenyl (**b**); 1H-indol-3-yl (**c**); methylthiomethyl (**d**)

amfepramone **9**

ketorolac **11**

pioglitazone **12**

cathinone **10**

thalidomide **13**

14   15

clopidogrel **16**
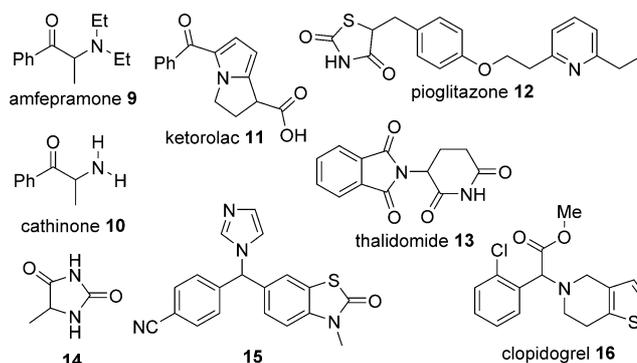
and $R_2 = R_3 = H$. These $\Delta\Delta G(R,H,H)$ values indicate how much R stabilizes an adjacent anion. The sum $\Delta\Delta G$-$(R_1,H,H) + \Delta\Delta G(R_2,H,H) + \Delta\Delta G(R_3,H,H)$, for the three non-H substituents around a chiral carbon atom is referred to here as $\Sigma\Delta\Delta G$. When two or three of the groups provide stabilization through charge delocalization, a cross-conjugation correction is applied to reflect the reduced ability of the second group to stabilize the anion caused by the presence of the first.[24]

Table 1: Stereogenic center types.

Center type	Non-H substituents	Representative compounds [number]
A	phenyl, reversed secondary amide, ester	**1a–h** [8]
B	phenyl, reversed secondary amide, primary amide	**2** [1]
C	5-membered aromatic, reversed secondary amide, ester	**3** [2]
D/D′	alkyl, reversed secondary amide, acidic secondary amide	**D = 4a–g, 5a–c + f** [11] **D′ = 14** [1]
E	phenyl, reversed secondary amide, acidic secondary amide	**6** [1]
F	alkyl, aminothiooxo imide, acidic secondary amide	**7a–d** [4]
G	alkyl, reversed secondary thioamide, acidic secondary amide	**8** [1]
H	ketone, dialkyl tertiary amine, alkyl	**9** [1]
I	ketone, primary amine, alkyl	**10** [1]
J	carboxylic acid, 5-membered aromatic, alkyl	**11** [1]
K	thioether, alkyl, acidic secondary amide	**12** [1]
L	imide, alkyl, acidic secondary amide	**13** [1]
M	phenyl, phenyl, 5-membered aromatic	**15** [1]
N	ester, dialkyl tertiary amine, phenyl	**16** [1]

The 35 compounds studied fall into fourteen chiral carbon atom types (**A–N**, Table 1) that have one hydrogen attached.[1,15–19,31] The second-order rate constants for general-base-catalyzed racemization are plotted against $\Sigma\Delta\Delta G$ at the bottom of Figure 2. When a chiral center type is represented by more than one compound, the mean value of $\log(k_{gb})$ for all representatives was used and the full range of values is shown as a vertical line. This prevents any center type from dominating the linear fit.

For the phenylglycine esters (**A**), substituent effects can cause up to a log unit variation from the line of best fit. This is likely to be representative of general substituent effects.[24] Grouping five-membered aromatic rings together (**C**) masks variation of 1.6 log units, likely reflecting the more direct influence of heteroatoms in aromatic rings. A relatively diverse set of alkyl substituents in the 5-position on a hydantoin ring (group **D**) or thiohydantoin (**F**) causes little variation in rate constants for racemization. In general, variation caused by substitution or structural variation within classes is less than two orders of magnitude and typically less than one order of magnitude.

At the bottom of Figure 2, two subgroups are apparent: those involving a cyclic anion with the potential to be aromatic and those that do not. For the non-aromatic set (shown in red) a line of best fit with equation $\log(k_{gb}) = -0.11 \times \Sigma\Delta\Delta G - 9.81$ was found ($R^2 = 0.78$ and RMSE = 0.40 log units) and for the aromatic anion set (shown in blue) the line of best fit has equation $\log(k_{gb}) = -0.26 \times \Sigma\Delta\Delta G - 16.95$ ($R^2 = 0.92$ and RMSE =

0.39 log units). The RMSE for all compounds computed individually is 0.64 for non-aromatic anions (excluding clopidogrel) and 0.37 for aromatic anions, that is, predictions are typically within 5-fold. Although not a perfect guide, the group contribution approach provides an easily applied, useful and rapid filtering that can even be used for very large databases.

For the particular example of chiral pharmaceuticals, our analysis can be applied to predict half-lives of racemization in physiological conditions. The rate constants for racemization of thalidomide at different phosphate buffer concentrations compared to that in blood suggest that, in terms of availability of catalytically active general bases, blood is approximately equivalent to a 0.15 M phosphate buffer at pH 7.2. Therefore, with k_{gb} predicted by the QM or group contribution method, the required half-lives can be predicted.[1]

A comprehensive workflow has now arisen: rapid analysis with a group contribution based method can trigger quantum mechanical calculations, which in turn can trigger an experimental protocol (Figure 3). Compounds at high risk of racemization can be avoided and racemization risk can be suppressed by design.

To illustrate the degree to which racemization in aqueous conditions is an overlooked issue, we have surveyed recent editions of leading chemistry journals, using our knowledge of the group contributions, to identify several articles envisaging biological applications. This was not an exhaustive search. Compounds described were subject to group contribution calculations. It is disappointing to reveal (Table 2) that liability to racemize under physiological conditions is more commonplace than would be possible if it were properly understood and controlled, as many chemists seem to believe.

In summary, we describe an approach to quantitatively predict the racemization risk that is generally applicable and allows synthetic chemists to avoid racemization-prone targets or understand erosion of the enantiomeric excess (*ee*). The approach allows quantitative assessment of the risk of chiral compounds of turning into racemic mixtures when used as pharmaceuticals or for other purposes.[36]

Acknowledgements
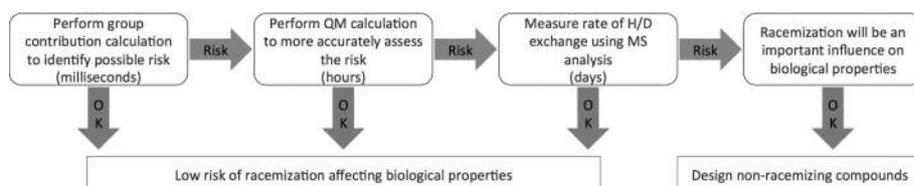
Figure 3. Workflow for identifying compounds at risk of racemization, in parentheses is the typical time taken to process one compound at each step.

Table 2: Examples of potentially pointless stereoselective syntheses from recent literature.

Reference	Representative compound	$\Sigma\Delta\Delta G$ [kcal mol^{-1}]	Predicted percentage[a]
[32]		−48.8	28%
[33]		−50.7	40%
[34]		−54.1	70%
[35]		−46.9	19%
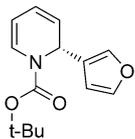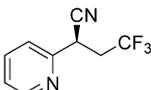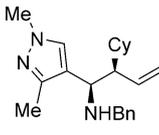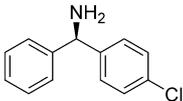
[a] Racemized within 24 hours.

table-of-contents image is based on a photo taken at the Groningen Museum of "The Cathedral" by Rodin (collection Musée Rodin).

Conflict of interest

A.G.L. is a shareholder in AstraZeneca.

[1] M. Reist, P. A. Carrupt, E. Francotte, B. Testa, *Chem. Res. Toxicol.* **1998**, *11*, 1521–1528.
[2] K. Nakano, M. Kitamura, in *Separation of Enantiomers*, Wiley-VCH Verlag GmbH & Co. KGaA, **2014**, pp. 161–216.
[3] M. T. El Gihani, J. M. Williams, *Curr. Opin. Chem. Biol.* **1999**, *3*, 11–15.
[4] E. J. Ariens, W. Soudijn, P. B. M. W. M. Timmermans, *Stereochemistry and Biological Activity of Drugs*, Blackwell Scientific Publications, Oxford, England, **1983**.
[5] E. J. Ariens, *Stereochem. Biol. Act. Drugs Rev. Int. Symp.* **1983**, 11–32.
[6] E. J. Ariëns, *Trends Pharmacol. Sci.* **1993**, *14*, 68–75.
[7] S. W. Smith, *Toxicol. Sci.* **2009**, *110*, 4–30.
[8] A. K. Peepliwal, S. B. Bagade, C. G. Bonde, *J. Biomed. Sci. Res.* **2010**, *2*, 29–45.
[9] A. G. Leach, E. A. Pilling, A. A. Rabow, S. Tomasi, N. Asaad, N. J. Buurma, A. Ballard, S. Narduolo, *MedChemComm* **2012**, *3*, 528–540.

[10] F. Lovering, J. Bikker, C. Humblet, *J. Med. Chem.* **2009**, *52*, 6752–6756.
[11] J. T. Bagdanoff, Y. Xu, G. Dollinger, E. Martin, *J. Med. Chem.* **2015**, *58*, 5781–5788.
[12] F. Pellati, G. Cannazza, S. Benvenuti, *J. Chromatogr. A* **2010**, *1217*, 3503–3510.
[13] L. Calabrese, A. Fleischer, *Am. J. Med.* **2000**, *108*, 487–495.
[14] I. Ali, V. K. Gupta, H. Y. Aboul-Enein, P. Singh, B. Sharma, *Chirality* **2007**, *19*, 453–463.
[15] M. Reist, L. H. Christiansen, P. Christoffersen, P.-A. Carrupt, B. Testa, *Chirality* **1995**, *7*, 469–473.
[16] M. Reist, M. Roy-de Vos, J. P. Montseny, J. M. Mayer, P. A. Carrupt, Y. Berger, B. Testa, *Drug Metab. Dispos. Biol. Fate Chem.* **2000**, *28*, 1405–1410.
[17] C. Danel, C. Foulon, J.-F. Goossens, J.-P. Bonte, C. Vaccher, *Tetrahedron: Asymmetry* **2006**, *17*, 2317–2321.
[18] M. Reist, P.-A. Carrupt, B. Testa, S. Lehmann, J. J. Hansen, *Helv. Chim. Acta* **1996**, *79*, 767–778.
[19] M. Brandl, D. Conley, D. Johnson, D. Johnson, *J. Pharm. Sci.* **1995**, *84*, 1045–1048.
[20] B. Jamali, I. Bjornsdottir, O. Nordfang, S. H. Hansen, *J. Pharm. Biomed. Anal.* **2008**, *46*, 82–87.
[21] B. Testa, P.-A. Carrupt, J. Gal, *Chirality* **1993**, *5*, 105–111.
[22] P. Jenner, B. Testa, *Drug Metab. Rev.* **1973**, *2*, 117–184.
[23] E. J. Ebbers, G. J. A. Ariaans, J. P. M. Houbiers, A. Bruggink, B. Zwanenburg, *Tetrahedron* **1997**, *53*, 9417–9476.
[24] For further details, see supporting information.
[25] http://www.gostardb.com/gostar/loginEntry.do.
[26] A. Ballard, Kinetics and Mechanism of H/D Exchange Reactions and Racemization in Aqueous Solutions: Configurational Stability of Ester and Amide Arylglycine Derivatives, Kinetics and mechanism of H/D exchange reactions and racemization in aqueous solutions: configurational stability of ester and amide arylglycine derivatives, Cardiff University, **2011**.
[27] S. Narduolo, The Mechanism of Racemization of 5-Substituted Hydantoins in Aqueous Solution, The mechanism of racemization of 5-substituted hydantoins in aqueous solution, Cardiff University, **2011**.
[28] H. O. Ahmad, Kinetics and Mechanism of Racemization Reactions of Configurationally Labile Stereogenic Centres in Drug-like Molecules in Aqueous Solutions; Thiohydantoins and Related Compounds, Cardiff University, **2015**.
[29] R. A. Lazarus, *J. Org. Chem.* **1990**, *55*, 4755–4757.
[30] Manuscript in preparation.
[31] A. K. Mills, A. E. W. Smith, *Helv. Chim. Acta* **1960**, *43*, 1915–1927.
[32] J. P. Lutz, S. T. Chau, A. G. Doyle, *Chem. Sci.* **2016**, *7*, 4105–4109.
[33] F. Wang, D. Wang, X. Wan, L. Wu, P. Chen, G. Liu, *J. Am. Chem. Soc.* **2016**, *138*, 15547–15550.
[34] R. Y. Liu, Y. Yang, S. L. Buchwald, *Angew. Chem.* **2016**, *128*, 14283–14286.
[35] I. V. Pavlidis, M. S. Weiß, M. Genz, P. Spurr, S. P. Hanlon, B. Wirz, H. Iding, U. T. Bornscheuer, *Nat. Chem.* **2016**, *8*, 1076–1082.
[36] Information about data underpinning the results presented here, including how to access them, can be found in the Cardiff University data catalogue at http://doi.org/10.17035/d.2017.0043106638.