ELSEVIER

Journal of Molecular Structure: THEOCHEM 725 (2005) 31–37

THEO CHEM

www.elsevier.com/locate/theochem

A simple method for reaction rate prediction of ester hydrolysis

Hongzhou Zhang[a,*], Xianggui Qu[b], Howard Ando[a]

[a]Pfizer Global Research and Development, Michigan Laboratories, 2800 Plymouth Road, Ann Arbor, MI 48104, USA
[b]Department of Mathematics and Statistics, Oakland University, Rochester, MI 48309, USA

Received 14 July 2004; revised 17 February 2005; accepted 28 February 2005
Available online 31 May 2005

Abstract

Ester hydrolysis is described by chemical descriptors calculated from quantum mechanic methods. Statistical analysis with good fitting (cross-validated $R^2_{adj} = 92.4\%$) and predicting capability (cross-validated $R^2_{adj} = 91.9\%$ for test molecules) were established through multiple linear regression. Contributions of descriptors in these statistical models were compared and the two-variable models provide sufficient information at high level of calculation.

Keywords: Ester hydrolysis; Linear regression; Variable selection; Reaction rate prediction; Quantum mechanic

1. Introduction

Ester hydrolysis has been studied extensively due to its important role in biological, chemical, environmental and industrial processes [1]. As an important functional group, carboxylic acids and their derivatives such as esters and amides, are widely present in many drug and natural molecules [2]. Ester derivatives of drug molecules are the excellent prodrug candidates to improve drug's physical or pharmaceutical properties [3,4]. As a result, their formation and degradation processes are of particular interest to many pharmaceutical and pharmacokinetics, dynamics and Metabolism scientists.

Reaction mechanisms under neutral or acid/base catalyzed conditions has been widely explored and the addition-elimination mechanism [5] via tetrahedral intermediate is generally accepted. The reaction rate is altered by the electronic and steric effects of the substituent groups and therefore can be related by Hammett equation [6]

$$\log \frac{k}{k^0} = \rho\sigma \tag{1}$$

where k^0 and k are reaction rate constants of the reference and substituted compounds, and ρ and σ are the Hammett reaction constant and substitute parameter. A more general form of Hammett equation or the *linear free energy relationship* (LFER) is

$$\Delta G^{\neq} = m\Delta G_a + b, \tag{2}$$

where $\Delta G^{\neq}$ is free energy of activation, based on transition state theory, ΔG_a represents thermodynamic free energy of reaction; m is slope and b is intercept. LFERs make a connection between thermodynamic data and kinetic values for a series of reactions possessing similar mechanisms, so that it is possible to create structure reactivity relationships for a series of molecules.

Collette [7] predicted base-catalyzed ester hydrolysis rate using the infrared interferograms method. The correlation coefficient between the measured and predicted values was 0.941 (or $R^2 = 0.887$) for 36 out of 41 compounds. With the rapid development of computer hardware and theoretical calculation software, molecular-orbital calculation, have started to compete with Hammett constant obtained experimentally [8]. Chaudry and Popelier [9] applied quantum topological molecular similarity (QTMS) analysis to the Collette's dataset. In their paper, different levels of theories were carried out and total 12 descriptors were used with partial least squares (PLS) methods. Their best model was from Hartree–Fork level with 6-31G(d) basis set with a cross-validated $R^2 = 0.869$. The more expensive density function theory calculation, B3LYP/6-11+G(2d,p) did not offer any gain in terms of

* Corresponding author. Tel.: +1 734 622 5844; fax: +1 734 622 2782.
E-mail address: hongzhou.zhang@pfizer.com (H. Zhang).

0166-1280/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.theochem.2005.02.086

cross-validated R^2 while the model from semi-empirical AM1 calculation had a cross-validated $R^2 = 0.665$.

While PLS [10] technique, like Principle Component Regression (PCR) can be applied to datasets with larger number of descriptors than the number of data records, it makes the model explanation much harder since it introduce latent variables, which is the combination of the original descriptors. On the other hand, a simple MLR model with limited set of descriptors is easier to understand and more natural to prevent over-fitting problem. In this paper, initially eight chemical descriptors were chosen based on analysis of the reaction mechanism and identification of significant descriptors was achieved through multiple regression and best subset analysis. Linear models with just significant descriptors were compared at different levels of calculation.

2. Data generation and statistics methods

All 41 esters from Collette's collection were analyzed (Table 1). The collection covers various esters, such as alkyl alkanoate (RC(=O)OR′), alkyl benzoate (ArC(=O)OR), benzyl alkonoate (RC(=O)OAr) and benzyl benzoate (ArC(=O)OAr).

2.1. Computational method

Calculations were carried out with different modules of Accelrys' MaterialsStudio v2.2. The gas phase geometry optimization was performed using semi-empirical method AM1 from VAMP [11] module and later density functional theory (DFT) from DMol^3 module [12]. DMol^3 uses

Table 1
Ester hydrolysis reaction rate [7] and the related calculated descriptors at AM1 level of calculation

Molecule	log *k*	Bond length			mulliken charge			HOMO	LUMO
		C1=O1	C1–O2	O2–R	C1	O1	O2		
Ethyl acetate	−0.96	1.233	1.368	1.434	0.345	−0.388	−0.328	−11.224	1.152
Ethyl formate	1.41	1.229	1.36	1.435	0.247	−0.388	−0.335	−11.359	1.146
Ethyl chloroacetate	1.56	1.231	1.364	1.437	0.347	−0.362	−0.318	−11.281	0.304
Ethyl bromoacetate	1.7	1.231	1.365	1.437	0.36	−0.362	−0.317	−11.115	−0.123
Ethyl *n*-butyrate	−1.26	1.233	1.367	1.435	0.347	−0.388	−0.326	−11.104	1.191
Ethyl isobutyrate	−1.49	1.233	1.368	1.435	0.349	−0.387	−0.326	−10.995	1.239
Ethyl aminoacetate	−0.19	1.232	1.365	1.436	0.299	−0.385	−0.342	−10.249	1.047
Ethyl benzoate	−1.5	1.235	1.37	1.435	0.401	−0.388	−0.324	−9.996	−0.038
Ethyl *p*-fluorobenzoate	−1.41	1.235	1.37	1.436	0.404	−0.386	−0.325	−9.934	−0.656
Ethyl dibromoacetate	2.31	1.232	1.361	1.439	0.375	−0.355	−0.295	−11.174	−0.529
Ethyl *p*-nitrobenzoate	−0.13	1.234	1.367	1.438	0.397	−0.371	−0.321	−10.802	−1.618
Ethyl *p*-aminobenzoate	−2.59	1.236	1.372	1.434	0.409	−0.398	−0.328	−8.835	−0.153
Ethyl trichloroacetate	3.41	1.228	1.361	1.44	0.361	−0.327	−0.296	−11.6	0.679
Ethyl acrylate	−1.11	1.232	1.368	1.435	0.353	−0.385	−0.327	−10.363	0.923
Ethyl 2-bromopropionate	1	1.232	1.364	1.437	0.359	−0.367	−0.311	−11.025	−0.102
Ethyl pivalate	−2.77	1.233	1.367	1.44	0.354	−0.389	−0.323	−10.986	1.288
Methyl formate	1.56	1.229	1.362	1.428	0.243	−0.385	−0.336	−11.567	1.09
Benzyl acetate	−0.71	1.232	1.37	1.434	0.349	−0.381	−0.327	−9.587	0.332
n-butyl acetate	−1.06	1.233	1.368	1.438	0.345	−0.386	−0.329	−11.173	1.151
n-propyl acetate	−1.06	1.233	1.367	1.438	0.345	−0.387	−0.329	−11.196	1.153
Methyl acetate	−0.74	1.233	1.369	1.427	0.342	−0.385	−0.328	−11.402	1.103
Isopropyl formate	1.04	1.229	1.358	1.446	0.25	−0.391	−0.334	−11.327	1.191
Methyl benzoate	−1.1	1.235	1.372	1.428	0.398	−0.385	−0.325	−10.02	−0.379
Benzyl benzoate	−2.1	1.235	1.37	1.442	0.4	−0.388	−0.319	−9.613	0.377
Isopropyl acetate	−1.52	1.233	1.366	1.445	0.348	−0.391	−0.326	−11.19	1.194
n-butyl formate	1.34	1.229	1.36	1.438	0.246	−0.388	−0.335	−11.264	1.146
n-propyl formate	1.36	1.229	1.36	1.438	0.246	−0.388	−0.336	−11.284	1.147
sec-butyl acetate	−1.76	1.233	1.367	1.438	0.345	−0.388	−0.327	−11.229	1.159
2-chloroethyl acetate	−0.41	1.232	1.37	1.436	0.348	−0.383	−0.328	−11.3	0.983
2-methoxyethyl acetate	−0.69	1.233	1.368	1.43	0.343	−0.393	−0.315	−10.619	1.068
Methyl *p*-fluorobenzoate	−1.15	1.235	1.371	1.429	0.4	−0.383	−0.325	−9.959	−0.689
Methyl *p*-hydroxybenzoate	−1.52	1.235	1.372	1.428	0.403	−0.39	−0.326	−9.536	−0.397
Methyl *p*-aminobenzoate	−2.35	1.236	1.374	1.428	0.406	−0.395	−0.328	−8.852	−0.18
Isopropyl hydroxybenzoate	−2.23	1.236	1.368	1.446	0.409	−0.396	−0.324	−9.494	−0.34
Methyl *m*-aminobenzoate	−1.47	1.235	1.372	1.428	0.396	−0.384	−0.325	−8.834	−0.294
Isopropyl *p*-aminobenzoate	−3.04	1.236	1.37	1.445	0.41	−0.4	−0.326	−8.825	−0.134
Methyl 2,4-D	1.06	1.233	1.361	1.43	0.31	−0.375	−0.299	−9.436	−0.341
2-butoxy 2,4-D	1.48	1.231	1.363	1.435	0.313	−0.367	−0.313	−9.579	−0.467
n-octyl 2,4-D	0.57	1.231	1.362	1.436	0.318	−0.37	−0.311	−9.523	−0.386
Methyl methacrylate	−1.25	1.234	1.372	1.428	0.384	−0.381	−0.328	−10.487	0.077
Ethyl iodoacetate	1.21	1.278	1.367	1.441	0.362	−0.366	−0.32	−10.933	−0.398