



HAL
open science

Predicting thermochemical parameters of oxygen-containing heterocycles using simple QSPR models

Nico Adams, Joachim Clauss, Marc Meunier, Ulrich Schubert

► **To cite this version:**

Nico Adams, Joachim Clauss, Marc Meunier, Ulrich Schubert. Predicting thermochemical parameters of oxygen-containing heterocycles using simple QSPR models. *Molecular Simulation*, 2006, 32 (02), pp.125-134. 10.1080/08927020500474300 . hal-00514974

HAL Id: hal-00514974

<https://hal.science/hal-00514974>

Submitted on 4 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Predicting thermochemical parameters of oxygen-containing heterocycles using simple QSPR models

Journal:	<i>Molecular Simulation/Journal of Experimental Nanoscience</i>
Manuscript ID:	GMOS-2005-0066
Journal:	Molecular Simulation
Date Submitted by the Author:	09-Nov-2005
Complete List of Authors:	Adams, Nico; Eindhoven University of Technology, Department of Chemistry Clauss, Joachim; Ticona, Core Research Meunier, Marc; Accelrys Ltd Schubert, Ulrich; Eindhoven University of Technology, Department of Chemistry
Keywords:	thermochemistry, QSPR, enthalpy of formation, entropy of formation, heat capacity

SCHOLARONE™
Manuscripts

1
2
3
4 **Predicting thermochemical parameters of oxygen-containing**
5
6 **heterocycles using simple QSPR models**
7
8

9
10 NICO ADAMS[†], JOACHIM CLAUSS[‡], MARC MEUNIER⁺, ULRICH S.
11
12 SCHUBERT^{†*}
13
14

15 [†]Laboratory of Macromolecular Chemistry and Nanoscience, Eindhoven University of
16 Technology and Dutch Polymer Institute, P. O. Box 513, 5600 MB Eindhoven, The
17 Netherlands (u.s.schubert@tue.nl)
18
19
20
21

22
23 [‡]Ticona GmbH, Core Technology, Professor-Staudinger-Strasse, 65451 Kelsterbach,
24
25 Germany
26
27

28 ⁺Accelrys Ltd., 334 Cambridge Science Park, Cambridge CB4 0WN, United
29
30 Kingdom
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Quantitative structure-property relationships for the prediction of standard enthalpies and entropies of formation as well as standard molar heat capacities for small oxygen heterocyclic compounds were developed, using 1D, 2D and 3D descriptors and experimental or computed thermochemical data. To develop the models, the data set was split into test and training sets using D-optimal experimental design to generate a diverse training set. Internal ($R^2_{\text{cross-validated}} = 0.898 - 0.998$) and external ($R^2_{\text{cross-validated}} = 0.847 - 0.996$) validation showed the models to be both stable and highly predictive. Enthalpies of formation were best described by electrotopological, atomic composition and molecular refractivity descriptors, while Kier and Hall χ and κ descriptors as well as the number of rotatable bonds appear frequently in models describing the entropy of formation of these compounds. Heat Capacity models often feature the molecular area descriptor as well as well as the Kier and Hall $^0\chi$ descriptor and the number of methyl groups present in the molecule.

Keywords: thermochemistry, QSPR, enthalpy of formation, entropy of formation, heat capacity, heterocycles

1. Introduction

The availability of good-quality thermochemical data for small molecules is of great importance for a number of problems in chemistry and chemical engineering. Oxygen heterocycles are key ingredients in a number of industrial processes[1] and their omnipresence in our environment has led to a significant interest in the way in which they are broken down both in nature as well as in the body.[2] As far as their industrial usage goes, oxygen heterocycles are key ingredients in the manufacture of polyacetals. Polyoxymethylene (POM), for example, is manufactured by polymerizing 1,3,5-trioxane and, if a co-monomer is included, by copolymerizing various 1,3-dioxolanes.[3] Generally, the world market for engineering plastics is growing and, as an example, the demand for POM in China alone was estimated to increase from 140 kt in 2003 to 180 kt in 2005.[4] Tetrahydrofuran is another important monomer, which has attracted a significant amount of industrial attention recently, with the BASF opening the world's largest polyTHF plant in Caojing (China) in the spring of 2005.[5] All of this has led to a need to obtain good thermodynamic data for this class of compounds.

A significant amount of effort has been devoted to the development of methodologies for the estimation of enthalpies and entropies of formation as well as molar heat capacities. As early as the 1950s, Benson *et al.* published a general method for estimating the thermochemical properties of chemical species on the basis of group additive contributions.[6-8] The group additive method makes the assumption that most molecular properties are made up of additive contributions from individual atoms or bonds in the molecule. With the advent of high-performance computing, thermochemical parameters could also be estimated using computational tools, ranging from semiempirical methods[9] through to DFT[10] and other *ab-initio*[11]

1
2
3 calculations. Furthermore, Gibbs-ensemble Monte Carlo simulations can also be used
4
5 to derive thermodynamic properties.[12]
6
7

8
9 In an early study, Lay and co-workers reported thermochemical data for a 34-
10
11 membered dataset of three-to-six membered oxygen-containing heterocyclic
12
13 hydrocarbons calculated using the semiempirical PM3 method[13, 14] and developed
14
15 a set of group additivity ring corrections for use with Benson's group additivity
16
17 parameters.[15] The authors later expanded this work and, using a combination of *ab-*
18
19 *initio* calculations and isodesmic reactions, developed thermochemical and group
20
21 additive parameters for linear[16] and cyclic alkyl peroxides.[17] In a subsequent
22
23 study, Shirel and Pulay investigated the stability of oxo- and chloro-substituted
24
25 trioxanes[18] and Saito and Fuwa conducted an extensive study concerning the
26
27 thermochemical properties of polychlorinated dibenzo-p-dioxins, dibenzofurans and
28
29 polychlorinated biphenyls using the PM3 Hamiltonian.[9] Notario *et al.* studied
30
31 dibenzofurans using *ab initio* calculations at the GAUSSIAN-3 G3(MP2)//B3LYP
32
33 level, albeit with a much smaller compound set.[11] Li *et al.* calculated
34
35 thermochemical parameters for 76 polybrominated dibenzo-p-dioxins using
36
37 B3LYP/6-31G(d) functional and basis set.[19] To the best of our knowledge, no
38
39 quantitative structure-property relationships (QSPRs) for thermochemical parameters
40
41 for small oxygen heterocycles have been developed so far. The present paper aims to
42
43 fill this gap, using 1D, 2D and 3D descriptors. QSPRs for the prediction of enthalpies
44
45 of formation were generated on the basis of available experimental data, while models
46
47 for entropies of formation and heat capacities were, due to the paucity of available
48
49 experimental data, developed on the basis of validated computed values.
50
51
52
53
54
55
56
57
58
59
60

2. Computational Procedure

2.1 DFT and Semi-Empirical Calculations

Molecular energies, geometries and vibrational frequencies were determined using the DMol³. [20, 21] Geometry optimizations were performed using general gradient corrected Perdew, Burke, Ernzerhof (PBE), the revised PBE (RPBE) functional, [22] the Becke, Lee, Yang, Parr (BLYP) correlation functional, [23, 24] or the Hamprecht (HCTH) functional, [25] using a double numerical basis set including a polarization functions (DNP). [20, 21] Optimum structures were confirmed as such by the absence of imaginary vibrations (self-consistent field density convergence: 1×10^{-6} Ha). Semiempirical calculations were carried out using the PM3 method [13, 14] as implemented in VAMP. [26-28] Enthalpies and entropies of formation as well as heat capacities were also estimated using a modified version of Benson's group additive method as implemented in an electronic form by the National Institutes of Standards (NIST). [6, 29-31]

2.2 QSPR studies

Descriptors were calculated using the MS QSAR 3.2 and TSAR 3.3 software packages [28] and experimental thermochemistry data was taken from the Computational Chemistry Benchmark and Comparison Database [32] (CCBCD) or the Chemistry Webbook, [29] both maintained by NIST. QSPR equations were developed on the basis of experimental (enthalpies of formation) or computational data (entropies of formation and heat capacities) if insufficient experimental data was available. To develop the QSPR models, D-optimal design was used to split the dataset into a training and a test set. Regression equations were derived using genetic

1
2
3 algorithms[33] to select key descriptors. Models were validated by predicting
4 thermochemical properties of the test set molecules.
5
6

7 8 9 **2.3 Validation of computational data**

10
11 When comparing the calculated structural data for all combinations of functional and
12 basis set used in this study to experimentally determined values contained in the
13 CCCBD for ethylene oxide (**1**), 1,3,5-trioxane (**2**) and furan (**3**) (Figure 1), it could be
14 shown that the latter are reproduced with good to excellent accuracy (Table 1).
15
16
17
18
19

20
21
22 [Insert Figure 1, Table 1 about here]
23
24

25 As expected, the agreement between experimental data and structures computed the
26 PM3 Hamiltonian is less good. Furthermore, different functionals and basis sets are in
27 good agreement w.r.t. the computed entropies and heat capacities. All further
28 calculations of entropies and heat capacities using DFT methods, were therefore
29 carried out using the PBE functional in connection with a DNP basis set.
30
31
32
33
34
35

36
37 In order to determine how accurately DFT methods predict standard entropies of
38 formation as well as molar heat capacities, the geometries of 84 compounds were
39 optimized using the PBE/DNP functional and basis set combination and
40 thermodynamic data were calculated. Although sufficient experimental data is
41 available, enthalpies of formation were also computed using PM3 and Benson's
42 method. The current commercial implementation of PBE/DNP in DMol³ is not
43 suitable for the calculation of standard enthalpies of formation, as these are calculated
44 using a database of atomic binding energies, which is not currently available for the
45 PBE functional. Tables of experimental and calculated data are given in the
46 supporting information.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 PM3 parameters have been optimized to reproduce experimental enthalpies of
4 formation at 298.15 K. Consequently the method performs well when compared to
5 experimental data ($R = 0.984$, $R^2 = 0.968$, $SD = 9.21 \text{ kcal mol}^{-1}$). The use of DFT
6 optimized structures and subsequent enthalpy prediction using PM3, did not lead to
7 improved data. The results are in good agreement with those previously obtained by
8 Lay *et al.*,[15] with the somewhat higher standard deviation and lower correlation
9 coefficients reflecting the much larger and more diverse set used in the present study.
10 For those cases, for which group additive parameters were available, Benson's
11 method performs well. When comparing experimental and computed data, a
12 correlation coefficient of $R = 0.997$ ($R^2 = 0.993$) and a standard deviation of 3.63 kcal
13 mol^{-1} was determined. *cis*- and *trans*-2,2,4,6-tetramethyl-1,3-dioxin are the only
14 significant outliers in this case and are overestimated by 5.7 and 8.8 kcal mol^{-1} .
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

32 Density functional theory, the PM3 Hamiltonian as well as Benson's group additive
33 method (where appropriate) were used to calculate standard entropies of formation.
34 Unfortunately, there is significantly less entropy than enthalpy data available in the
35 literature and therefore the calculations could only be validated using a significantly
36 smaller data set (8 datapoints). The risk inherent in such a small dataset is that it could
37 lead to either a serious over- (in cases in which there is a good accidental agreement
38 between experimental and computed data) or underestimation (in case the
39 experimental data is very noisy or there are experimental errors) of the accuracy of
40 the computational methods evaluated here. This also means that any comparison
41 between methods may be affected by a certain amount of uncertainty. In the absence
42 of further data, however, this is the best that can currently be achieved. On this basis,
43 all three methods gave satisfactory results, with DFT giving a slightly better
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 correlation between calculated and experimental values than the other two methods
4
5 (Figure 2).
6
7

8
9 [Insert Figure 2 about here]

10
11 DFT calculations give rise to a correlation coefficient of $R = 0.963$ ($R^2 = 0.927$) and
12
13 a standard deviation of $2.72 \text{ cal mol}^{-1}\text{K}^{-1}$, whereas the Benson model gives $R = 0.912$
14
15 ($R^2 = 0.832$) and a standard deviation of $3.44 \text{ cal mol}^{-1}\text{K}^{-1}$. This is reflected in a
16
17 certain amount of disagreement between the two models (Figure 3).
18
19

20
21
22 [Insert Figure 3 about here]

23
24
25 The major outliers (in addition to the *cis*- and *trans*-2,2,4,6-tetramethyl-1,3-dioxines)
26
27 are compounds containing an oxetanone or a carbonate motif, probably indicating that
28
29 the parameterization of the Benson model is not optimal for this type of structures.
30
31 PM3 delivers results close to those of the DFT calculations ($R = 0.961$, $R^2 = 0.923$,
32
33 $SD = 3.21 \text{ cal mol}^{-1}\text{K}^{-1}$).
34
35

36
37 Regarding the prediction of standard heat capacities, all methods give good to
38
39 excellent agreement between experimentally determined and calculated heat
40
41 capacities. Overall, PM3 seems to perform best ($R = 0.979$, $R^2 = 0.958$, $SD = 0.70 \text{ cal}$
42
43 $\text{mol}^{-1}\text{K}^{-1}$), followed by Benson's group additive method ($R = 0.968$, $R^2 = 0.938$, $SD =$
44
45 $1.93 \text{ cal mol}^{-1}\text{K}^{-1}$) and PBE/DNP ($R = 0.935$, $R^2 = 0.875$, $SD = 1.24 \text{ cal mol}^{-1}\text{K}^{-1}$)
46
47 although the differences between the methods are small (Figure 4).
48
49
50

51
52 [Insert Figure 4 about here]

53
54
55 Given the fact, that all three different computational methodologies give very good
56
57 agreement between predicted and experimentally determined values, it is suggested
58
59
60

1
2
3 that these methods provide high-quality data, suitable for the development of
4 quantitative structure-property relationships in the absence of experimental data.
5
6
7
8
9

10 11 **3. Thermochemical data from Quantitative Structure-Property Relationships** 12 13 **(QSPRs)**

14 15 16 17 **3.1 Subset Selection**

18
19
20 The selection of diverse subsets of molecules for model development is a non-trivial
21 problem and a number of different approaches, such as clustering techniques,[34]
22 random selection,[35-38] activity sampling,[38-40] self-organizing maps[41, 42] as
23 well as a number of experimental design approaches[43, 44] have been reported in the
24 literature. In a comparative study, Massart *et al.* demonstrated that both D-Optimal
25 and Kennard stone designs ultimately led to better models than random sampling or
26 self-organizing maps[36] and other authors have also reported favourable
27 experiences.[45-47] D-optimal designs aim to maximize the determinant of the
28 variance-covariance matrix $|\mathbf{X}'\mathbf{X}|$, where X is the information matrix of independent
29 covariables. This determinant will be at a maximum for compound sets, which have a
30 maximum variance (*i.e.* span a large chemical parameter space) and a minimum co-
31 variance (*i.e.* there is minimum similarity between the molecules).[48] In a first step,
32 therefore, 126 different 1D, 2D and 3D descriptors were calculated for each
33 compound in the dataset. Subsequent principal component analysis showed that the
34 first 27 principal components explain 99% of the variance in the dataset. The
35 maximum and minimum values of the first 7 principal component (80 % variance
36 explained) vectors were used as inputs for a D-optimal design, resulting in an
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 ensemble of 46 candidate points in virtual space, representing 55 % of the compounds
4
5
6 in the dataset (Table 2).
7

8
9 [Insert Table 2 approximately here]
10

11 A simple Euclidean distance measure was used to identify those compounds in “real”
12 chemistry space that lie closest to the design points. A visual examination of the
13 scores for the first vs. the second principal component show that the selected training
14 set is diverse and well distributed over the whole dataset (Figure 5). Those
15 compounds not included in the training set were used for external validation of the
16 QSPR model (test set).
17
18
19
20
21
22
23
24
25

26
27 [Insert Figure 5 approximately here]
28

29 30 **3.2 QSPR Model Development.**

31
32 The diverse subset of 46 compounds was used to develop QSPR models for standard
33 enthalpies and entropies of formation and molar heat capacities. Enthalpy models
34 were constructed using experimental data, whereas entropy and heat capacity models
35 were developed using computed data (DFT). Model construction was carried out
36 using genetic algorithm driven linear regression methods.[33] At the beginning of the
37 optimization procedure, 500 equations were randomly selected and evolved until
38 convergence was achieved. To guard against overfitting, the maximum equation
39 length was set to 5 independent variables in accordance with the recommendation that
40 a regression model with k independent variables and n compounds in the training set,
41 should satisfy the $n > 4k$ criterion (in this study, $n = 46$ and $k = 5$).[49] Care was taken
42 to penalize equations with both large sum-of-squares errors and large numbers of
43 independent variables.[50] The top 5 models for standard enthalpies, entropies and
44 molar heat capacities, together with validation data, are given in Table 3.
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 **3.2.1 Standard Enthalpy of Formation.** In the case of standard enthalpies of
5 formation, internal as well as external validation suggests that the models are
6 extremely robust: both adjusted (R^2_{adj}) as well as cross-validated (R^2_{cv}) coefficients of
7 determination are not significantly different, indicating that the models are both
8 robust and predictive. The top-performing model (H1) has a coefficient of
9 determination of 0.921 for the training set and 0.852 for the external test set (Table 3,
10 Figure 6).
11
12
13
14
15
16
17
18
19

20 [Insert Table 3 and Figure 6 approximately here]
21
22

23 The standard deviation for the training set is 16.75 kcal mol⁻¹. However, it can be seen
24 that a number of outliers are present in the training (O2-O5) and test (O1) sets. All of
25 these, with the exception of O3 (*tert*-butylperoxymethyl oxirane) are furan derivatives
26 and O4-O6 all contain acetoxy-substituents and O6 an additional nitro-group.
27 Removing the outliers from the dataset, results in an improved value of R^2 of 0.980
28 and a standard deviation of 8.45 kcal mol⁻¹. It should be noted, that, once outliers have
29 been removed from the dataset, the standard deviation is approximately comparable to
30 results obtained from PM3 calculations. Full tables of computed results are given in
31 the supporting information.
32
33
34
35
36
37
38
39
40
41
42
43

44 The good agreement between experimental and predicted data, shows the value of
45 using diverse subsets, such as those generated *via* D-optimal design for the
46 development of QSPR equations. Examination of the top-performing models shows
47 that a number of descriptors are repeatedly represented. The electrotopological
48 S_{ddsN} descriptor[51] appears in all five models, closely followed by the
49 S_{dssC} [51], molecular refractivity and atomic composition descriptors, all at three
50 counts each. Electro topological descriptors, or E-state indices, were introduced by
51
52
53
54
55
56
57
58
59
60

1
2
3 Kier and Hall.[51] Each atom in a molecular graph is represented by an E-state, which
4 encodes the electronic state of an atom as influenced by the other electronic states of
5
6 all the other atoms in the molecule, within the context of the molecular graph. The E-
7
8 state for a given atom, therefore, varies from molecular structure to molecular
9
10 structure.
11
12

13
14
15 The presence of the S_{ddsN} descriptor and its negative contribution indicates that the
16
17 presence and number of nitro-groups in the molecule has a significant bearing on the
18
19 standard enthalpy of formation. Interestingly, three of the 6 outliers are nitro
20
21 compounds (O1, O4, O6). The S_{dssC} descriptor makes a positive contribution to the
22
23 equations. Again, it is probably not surprising that the descriptor should be present, as
24
25 presence and number of double bonds in the system can be expected to have a
26
27 significant bearing on the enthalpy of formation. Molecular refractivity is defined as
28
29
30

$$MR = \frac{n^2 - 1}{n^2 + 2} \times \frac{M_w}{d} \quad (1)$$

31
32
33 where n is the refractive index, M_w the molecular weight and d the density. As n
34
35 usually does not change significantly, the molecular refractivity is effectively a
36
37 measure of volume and therefore the size of the molecule, albeit coupled to
38
39 polarizability information.[52] The atomic composition index, finally, is an
40
41 information content descriptor and the name is programmatic in this context – the
42
43 descriptor encodes the elemental composition of a molecule. It is intuitively
44
45 comprehensible, why such a descriptor should encode information about enthalpies of
46
47 formation.
48
49
50
51
52
53
54
55

56
57 **3.2.2 Standard Entropy of Formation.** Little experimental entropy data is available
58
59 for the compounds contained in the dataset (8 out of 84 compounds). Furthermore the
60

1
2
3 available data is tightly clustered and not suitable for the development of high quality
4 models. QSPRs for the standard entropy of formation of small oxygen-containing
5 heterocycles were therefore developed using calculated data. As discussed above,
6 density functional methods perform marginally better than PM3. Therefore, data from
7 the PBE/DNP calculations was used to develop the equations. The models appear to
8 be very stable and predictive with $R^2 = 0.988$, $R^2_{\text{adjusted}} = 0.987$ and $R^2_{\text{cross-validated}} =$
9 0.984 in the training and $R^2 = 0.958$ and $R^2_{\text{adjusted}} = 0.957$ in the test set, using model
10 S1. The standard deviation is $1.99 \text{ cal mol}^{-1}\text{K}^{-1}$ in the training set (Table 3, Figure 7).
11 The descriptors which appear most frequently in the top 5 models (Table 3) are the $^1\kappa$
12 and $^3\chi$ descriptors[53] as well as the rotatable bond count.
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27

28 [Insert Figure 7 approximately here]
29
30

31 **3.3.3 Standard Heat Capacity.** Again, very little experimental heat capacity data is
32 available (19 out of 84 compounds) for model development and the data is tightly
33 clustered. As there is no real difference between the heat capacity data computed
34 using DFT and semiempirical methods, data generated using density functional theory
35 was used to derive the QSPR equations. Again, the models show extremely good
36 performance, both in terms of training and validation sets (Table 3, Figure 8). The
37 highest performing model C1 gave $R^2 = 0.994$, $R^2_{\text{adjusted}} = 0.993$ and $R^2_{\text{cross-validated}} =$
38 0.993 in the training set and $R^2 = 0.964$ and $R^2_{\text{adjusted}} = 0.963$ for the validation set.
39 The most frequently observed descriptor here, is the molecular area (vdW area)
40 descriptor. Its presence is probably not surprising as it describes the van der Waals
41 area of the molecule and therefore also its size. As the heat capacity is defined as the
42 amount of heat required to change the temperature of a substance by one degree,
43 larger molecules will need more heat than smaller ones, which, in turn explains the
44 correlation with the molecular area descriptor. The only other descriptors appearing
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 multiple times are ${}^0\chi$ and the methyl group count, both at two counts each. ${}^0\chi$
4
5 describes the immediate bonding environment of atoms in a molecule, while
6
7 containing relatively information about the connectivity of the molecular skeleton.
8
9

10
11 While the QSPRs for both the entropy of formation as well as the heat capacities were
12
13 developed using computed data, one would have to expect that similar robust models
14
15 could be developed for experimental data on the basis of the fact that all three
16
17 different computational methods (see above) are in close agreement with each other
18
19 and with the available experimental data; *i.e.* the computed results must be close the
20
21 experimental values, were these available.
22
23
24
25
26
27
28

29 **4. Summary and Conclusions.**

30
31
32 Several computational ways of obtaining thermochemical parameters for small
33
34 oxygen-containing heterocycles were investigated and compared and QSPR models
35
36 for the prediction of standard enthalpies and entropies of formation as well as
37
38 standard heat capacities were developed. Robust and predictive quantitative structure-
39
40 property relationships were developed for all three thermodynamic parameters on the
41
42 basis of experimental or validated computed data. It could be shown that QSPR
43
44 models can be a fast and powerful tool for the prediction of thermodynamic
45
46 parameters of small oxygen-containing heterocycles.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **5. Acknowledgements.**
4

5
6 The authors wish to acknowledge the Dutch Polymer Institute (DPI), Project #500, for
7
8 financial support of this work. Furthermore, Brian Pauw is gratefully acknowledged
9
10 for help with MatLab programming.
11
12

13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review Only

References

- [1] S. Penczek. Cationic Ring-Opening Polymerization (CROP) Major Mechanistic phenomena. *J. Polym. Sci, Part A*, **38**, 1919 (2000).
- [2] A. Baccarelli, P. Mocarelli, D. G. Patterson Jr., M. Bonzini, A. C. Pesatori, N. Caporaso, M. T. Landi. Immunologic effects of dioxin: new results from Seveso and comparison with other studies. *Env. Health Perspect.*, **110**, 1169 (2002).
- [3] H. Cramail, A. Deffieux. Cationic Polymerization. In *Synthesis of Polymers*, A. D. Schlueter (Ed.), p. 231, Wiley-VCH, Weinheim (1999).
- [4] E. Brock. Asia - future market for engineering plastics. Available online at: http://www.ticona.com/ticona/ed_brock_16_6_uk_pdf.pdf (accessed November 2005).
- [5] BASF. Press Release. Available online at: <http://media.basf.com/en/presse/mitteilungen/pm.htm?pmid=1761&id=c0P1t7gw7bcp0oS> (accessed November 2005).
- [6] S. W. Benson. *Thermochemical Kinetics*. John Wiley & Sons, New York (1976).
- [7] S. W. Benson, J. H. Buss. Additivity rules for the estimation of molecular properties. Thermodynamic properties. *J. Chem. Phys.*, **29**, 546 (1958).
- [8] S. W. Benson, F. R. Cruickshank, D. M. Golden, G. R. Haugen, H. E. O'Neal, A. S. Rodgers, R. Shaw, R. Walsh. Additivity rules for the estimation of thermochemical parameters. *Chem. Rev.*, **69**, 279 (1969).
- [9] N. Saito, A. Fuwa. Prediction for thermodynamic function of dioxins for gas phase using semiempirical molecular orbital method with PM3 Hamiltonian. *Chemosphere*, **40**, 131 (2000).
- [10] C. F. Wilcox, Y.-X. Zhang, S. H. Bauer. The thermochemistry of TNAZ (1,3,3-trinitroazetidine) and related species: G3(MP2)//B3LYP heats of formation. *J. Molec. Struct.: THEOCHEM*, **538**, 67 (2001).
- [11] R. Notario, M. V. Roux, O. Castano. The enthalpy of formation of dibenzofuran and some related oxygen containing heterocycles in the gas phase. *Phys. Chem. Chem. Phys.*, **3**, 3717 (2001).
- [12] M. G. Ahunbay, S. Kranias, V. Lachet, P. Ungerer. Prediction of thermodynamic properties of heavy hydrocarbons by Monte Carlo simulations. *Fluid Phase Equilibria*, **224**, 73 (2004).
- [13] J. J. P. Stewart. Optimization of parameters for semiempirical methods I. Method. *J. Comp. Chem.*, **10**, 209 (1989).
- [14] J. J. P. Stewart. Optimization of parameters for semiempirical methods II. Applications. *J. Comp. Chem.*, **10**, 221 (1989).
- [15] T. S. Lay, T. Yamada, P.-L. Tsai, J. W. Bozelli. Thermodynamic parameters and group additivity ring corrections for three-to-six-membered oxygen heterocyclic hydrocarbons. *J. Phys. Chem. B*, **101**, 2471 (1997).
- [16] T. S. Lay, J. W. Bozelli. Enthalpies of Formation and Group Additivity of Alkyl Peroxides and trioxides. *J. Phys. Chem. A*, **101**, 9505 (1997).
- [17] T. S. Lay, J. W. Bozelli. Enthalpies of formation of cyclic alkyl peroxides: dioxirane, 1,2-dioxetane, 1,2-dioxolane and 1,2-dioxane. *Chem. Phys. Lett.*, **268**, 175 (1997).
- [18] M. L. Shirel, P. Pulay. Stability of novel oxo- and chloro-substituted trioxanes. *J. Am. Chem. Soc.*, **121**, (1999).

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- [19] X.-W. Li, E. Shibata, T. Nakamura. Theoretical calculation of thermodynamic properties of polybrominated dibenzo-p-dioxins. *J. Chem. Eng. Data*, **48**, 727 (2003).
- [20] J. Delley. An all-electron numerical method for solving the local density functional for polyatomic molecules. *J. Chem. Phys.*, **92**, 508 (1990).
- [21] J. Delley. From molecules to solids with the DMol3 approach. *J. Chem. Phys.*, **113**, 7756 (2000).
- [22] B. Hammer, L. B. Hansen, J. K. Norskov. Improved adsorption energetics within density-functional theory using revised Perdew-Burke Ernzerhof functionals. *Phys. Rev. B.*, **59**, 7413 (1999).
- [23] A. D. J. Becke. A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phys.*, **88**, 2547 (1988).
- [24] C. Lee, W. Yang, R. G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B.*, **37**, 785 (1988).
- [25] A. D. Boese, N. C. Handy. A new parametrization of exchange-correlation generalized gradient approximation functionals. *J. Chem. Phys.*, **114**, 5497 (2001).
- [26] G. Rauhut, T. Clark. Multicenter point charge model for high-quality molecular electrostatic potentials from AM1 calculations. *J. Comp. Chem.*, **15**, 503 (1993).
- [27] B. Beck, G. Rauhut, T. Clark. The natural atomic orbital point charge model for PM3: multipole moments and molecular electrostatic potentials. *J. Comp. Chem.*, **15**, 1064 (1994).
- [28] Accelrys. homepage. Available online at: <http://www.accelrys.com> (accessed November 2005).
- [29] National Institute of Standards. Chemistry Webbook. Available online at: <http://webbook.nist.gov/chemistry> (accessed November 2005).
- [30] R. C. Reid, J. M. Prausnitz, B. E. Poling. *The Properties of Gases and Liquids*. MacGraw-Hill Book Company, London (1987).
- [31] W. J. Lyman, W. F. Reehl, D. H. Rosenblatt. *Handbook of Chemical Property Estimation Methods. Environmental Behaviour of Environmental Compounds*. McGraw-Hill Book Company, London (1982).
- [32] R. D. Johnson III. NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101. Available online at: <http://srdata.nist.gov/ccbdb> (accessed November 2005).
- [33] D. Rogers, A. J. Hopfinger. Application of genetic function approximation in quantitative structure-activity relationships and quantitative structure-property relationships. *J. Chem. Inf. Comput. Sci.*, **34**, 854 (1994).
- [34] C. L. Senese, A. J. Hopfinger. A simple clustering technique to improve QSAR model selection and predictivity: application to a receptor independent 4D-QSAR analysis of cyclic urea derived inhibitors of HIV-1 protease. *J. Chem. Inf. Comput. Sci.*, **43**, 2180 (2003).
- [35] A. Yasri, D. Hartsough. Toward an optimal procedure for variable selection and QSAR Model Building. *J. Chem. Inf. Comput. Sci.*, **41**, 1218 (2001).
- [36] W. Wu, B. Walczak, D. L. Massart, S. Heuerding, F. Erni, I. R. Last, K. A. Prebble. Artificial neural networks in classification of NIR spectral data: design of the training set. *Chemometr. Intell. Lab. Syst.*, **33**, 35 (1996).

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- [37] T. Poetter, H. Matter. Random or rational design? Evaluation of diverse compound subsets from chemical structure databases. *J. Med Chem.*, **41**, 478 (1990).
- [38] A. Golbraikh, A. Tropsha. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *J. Comp.-Aided Molec. Design*, **16**, 357 (2002).
- [39] G. V. Kauffman, P. C. Jurs. QSAR and k-nearest neighbour classification analysis of selective cyclooxygenase-2 inhibitors using topologically based numerical descriptors. *J. Chem. Inf. Comput. Sci.*, **41**, 1553 (2001).
- [40] B. E. Mattioni, P. C. Jurs. Development of quantitative structure-activity relationship and classification models for a set of carbonic anhydrase inhibitors. *J. Chem. Inf. Comput. Sci.*, **42**, 94 (2002).
- [41] R. Guha, J. R. Serra, P. C. Jurs. Generation of QSAR sets with a self-organizing map. *J. Molec. Graph. Model.*, **23**, 1 (2004).
- [42] E. Bayram, P. Santago, R. Harris, Y.-D. Xiao, A. J. Clauset, J. D. Schmitt. Genetic algorithms and self-organizing maps: a powerful combination for modeling complex QSAR and QSPR problems. *J. Comp.-Aid. Molec. Des.*, **18**, 483 (2004).
- [43] M. Sjostrom, L. Eriksson. Applications of Statistical Experimental Design. In *Chemometrics Methods in Molecular Design*, H. van de Waterbeed (Ed.), p. 63, VCH, Weinheim (1995).
- [44] L. Eriksson, E. Johansson. Multivariate Design and Modeling in QSAR. *Chemometr. Intell. Lab. Syst.*, **34**, 1 (1996).
- [45] P. Gramatica, E. Papa. QSAR Modeling of bioconcentration factor by theoretical molecular descriptors. *QSAR Comb. Sci.*, **22**, 374 (2003).
- [46] P. Gramatica, P. Pilutti, E. Papa. QSAR Prediction of ozone tropospheric degradation. *QSAR Comb. Sci.*, **22**, 364 (2003).
- [47] E. Marengo, R. Todeschini. A new algorithm for optimal, distance-based experimental design. *Chemom. Intell. Lab. Sys.*, **16**, 37 (1992).
- [48] D. C. Montgomery. *Design and Analysis of Experiments*. John Wiley, New York (2001).
- [49] A. Tropsha, P. Grammatica, V. K. Gombar. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.*, **22**, 69 (2003).
- [50] Accelrys. *Accelrys MS Modeling 3.2 Documentation*. (2005).
- [51] L. H. Hall, B. Mohny, L. B. Kier. The electrotopological state: structure information at the atomic level for molecular graphs. *J. Chem. Inf. Comput. Sci.*, **31**, 76 (1991).
- [52] C. Hansch, R. Garg, A. Kurup. Searching for Allosteric Effects via QSARs. *Bioorganic & Medicinal Chemistry*, **9**, 283 (2001).
- [53] L. H. Hall, L. B. Kier. The molecular connectivity chi indexes and kappa shape indexes in structure property modeling. In *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd (Ed.), pp. (1992).

Figure Captions

Figure 1: Ethylene oxide (1), 1,3,5-trioxane (2) and furan (3).

Figure 2: Experimentally determined Entropies of formation *vs.* results derived using (a) DFT (PBE/DNP), (b) PM3 and (c) Benson's group additive method.

Figure 3: Comparison of entropies of formation calculated using DFT (PBE/DNP) and Benson's group additive method.

Figure 4: Experimentally determined molar heat capacities *vs.* results derived from (a) DFT (PBE/DNP), (b) PM3 and (c) Benson's group additive method.

Figure 5: Principal Component Analysis of all descriptors (PC1 – PC2: EV = 55 %) for the data set chemicals split into training (●) and test (●) set.

Figure 6: Predicted *vs.* experimental standard enthalpies of formation for both training (●) and test sets (●), using model H1.

Figure 7: Predicted *vs.* computational standard entropies of formation for both training (●) and test sets (●), using model S1.

Figure 8: Predicted *vs.* computational standard heat capacities for both training (●) and test sets (●), using model C1.

Table Headings

Table 1: Experimentally determined and calculated geometries for ethylene oxide (**1**), 1,3,5-trioxane (**2**) and furan (**3**) (Bond lengths are given in Ångströms (Å), bond angles in degrees (°), $\Delta S_f^{298.15}$ and $C_p^{298.15}$ in $\text{cal mol}^{-1}\text{K}^{-1}$).

Table 2: Results from principal component analysis.

Table 3: QSPR Models for standard enthalpies and entropies of formation and molar heat capacities. (R^2 = coefficient of determination; R^2_{adjusted} = adjusted coefficient of determination, R^2_{cv} = cross-validated coefficient of determination; cRB = number of rotatable bonds; HA = number of hydrogen bond acceptors; S_ddsN = E state keys (sums) = S_ddsN; MD = molecular density; $^3\kappa$ = 3-Kappa (Kier and Hall); MR = molecular refractivity; SC_c = Subgraph counts (3): chain; AC = atomic composition; S_dssC = E-state keys (sums): S_dssC; $^3\chi$ = 3-Chi (chain) (Kier and Hall); $^1\kappa$ = 1-Kappa (atom modified) (Kier and Hall); VDM = Vertex distance/magnitude; S_ssO = E-state keys (sums): S_ssO; AlogP = AlogP; S_sCH₃ = E-state keys: S_sCH₃; MA = molecular area; MF = number of methyl groups; SC_p = subgraph counts (0): path)

Table 1

	Expt.	BLYP			RPBE			PBE			HCTH			PM3	
		DNP	DND	DN	DNP	DND	DN	DNP	DND	DN	DNP	DND	DN		
(1)	d O1-C2	1.452	1.453	1.452	1.529	1.445	1.445	1.518	1.438	1.437	1.510	1.424	1.424	1.493	1.432
	d C2-C3	1.459	1.475	1.474	1.490	1.475	1.475	1.490	1.438	1.469	1.485	1.461	1.460	1.477	1.484
	d C2-H4	1.084	1.093	1.096	1.091	1.096	1.098	1.093	1.095	1.097	1.092	1.093	1.094	1.090	1.096
	∠ O1-C2-C3	59.203	59.504	59.499	60.846	59.326	59.327	60.615	59.261	59.261	60.542	59.142	59.150	60.357	58.806
	∠ C2-O1-C3	61.594	60.997	61.007	58.318	61.350	61.349	58.786	61.481	61.481	58.923	61.715	61.70	59.283	62.388
	∠ H4-C2-H5	116.750	115.659	115.529	115.898	115.635	115.485	115.919	115.713	115.583	115.901	115.503	115.366	115.551	111.654
	∠ O1-C2-H4	114.704	114.987	115.032	114.042	115.136	115.126	114.183	115.228	115.164	114.334	115.063	115.088	114.355	116.313
	Δ S _r (298.15)	58.08	59.55	59.55	59.93	59.63	59.60	59.93	59.54	59.52	59.87	59.36	59.31	59.61	58.03
	C _p (298.15)	11.44	11.61	11.61	12.03	11.61	11.83	12.14	11.78	11.71	12.07	11.18	11.08	11.36	11.51
(2)	d O-C	1.422	1.431	1.431	1.477	1.428	1.427	1.471	1.418	1.419	1.463	1.410	1.409	1.451	1.410
	d C-H _{eq}		1.095	1.095	1.089	1.096	1.097	1.091	1.091	1.096	1.090	1.092	1.093	1.087	1.097
	d C-H _{axial}		1.111	1.111	1.107	1.110	1.112	1.108	1.112	1.112	1.108	1.108	1.109	1.104	1.107
	∠ O-C-O	112.2	111.708	111.708	111.127	111.577	111.868	111.268	111.784	111.800	111.211	111.861	111.925	111.297	107.523
	∠ C-O-C	110.3	109.542	109.524	110.124	109.061	109.046	109.773	108.525	108.943	109.620	109.569	109.495	110.585	112.979
	Δ S _r (298.15)	68.09	71.88	71.88	72.79	71.49	71.51	72.31	71.71	71.36	71.91	74.89	73.75	71.75	70.62
	C _p (298.15)	19.57	20.67	20.67	21.79	20.46	20.41	21.29	20.52	20.23	20.97	20.97	20.68	20.53	21.53
(3)	d O1-C2	1.362	1.382	1.382	1.382	1.378	1.378	1.416	1.371	1.371	1.409	1.362	1.361	1.398	1.378
	d C2-C4	1.361	1.365	1.365	1.382	1.368	1.368	1.374	1.364	1.364	1.370	1.364	1.363	1.368	1.373
	d C4-C5	1.4338	1.438	1.438	1.438	1.437	1.437	1.451	1.431	1.431	1.444	1.426	1.426	1.438	1.441
	d C3-H7	1.0760	1.082	1.082	1.082	1.085	1.085	1.083	1.084	1.084	1.082	1.080	1.082	1.078	1.085
	d C4-H8	1.0760	1.084	1.084	1.084	1.087	1.087	1.086	1.086	1.086	1.085	1.082	1.084	1.081	1.086
	∠ O1-C2-C4	110.700	110.348	110.348	110.348	110.497	110.497	109.677	110.449	110.449	109.603	110.391	110.427	109.683	110.238
	∠ C2-O1-C3	106.60	106.322	106.322	106.322	106.347	106.347	106.171	106.477	106.477	106.309	106.827	106.796	106.850	106.857
	∠ C2-C4-C5	106.00	106.491	106.491	106.491	106.329	106.329	107.237	106.313	106.313	107.242	106.195	106.175	107.042	106.334

\angle O1-C3-H7	115.90	115.586	115.586	115.586	115.613	115.613	115.461	115.722	115.722	115.572	115.652	115.556	115.570	115.492
Δ S _r (298.15)	63.82	65.36	65.36	65.36	65.39	65.39	65.46	65.28	65.28	65.37	64.99	64.95	65.03	64.58
C _p (298.15)	15.63	15.87	15.87	15.87	15.89	15.89	15.94	15.74	15.74	15.81	15.09	15.03	14.99	15.56

For Peer Review Only

Table 2

Principal Component	Variance explained	Cumulative Variance	Min	Max
1	0.449	0.449	-1.634	3.841
2	0.100	0.549	-2.147	2.058
3	0.074	0.624	-3.526	5.235
4	0.065	0.689	-2.437	2.985
5	0.048	0.737	-1.771	2.768
6	0.031	0.768	-4.603	2.148
7	0.028	0.796	-2.554	2.722

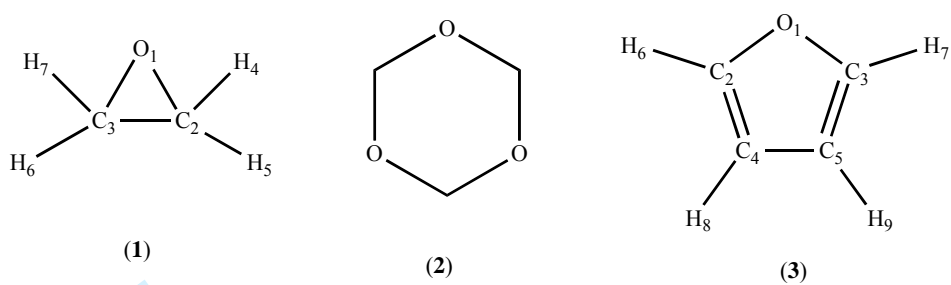
For Peer Review Only

Table 3

	Equation	R (test set)	R ² (test set)	R ² _{adj} (test set)	R ² _{cv}	F	SD (test set)
H1	$\Delta H_f^{298.15} = (10.741 \times \text{RB}) - (73.681 \times \text{HA}) - (176.609 \times \text{S_ddsN}) + (340.444 \times \text{MD}) - 284.461$	0.959 (0.923)	0.921 (0.852)	0.913 (0.847)	0.898	119	16.75 (13.87)
H2	$\Delta H_f^{298.15} = (74.772 \times \text{HA}) + (14.942 \times {}^3\kappa) - (165.827 \times \text{S_ddsN}) + (392.964 \times \text{MD}) - 352.086$	0.958 (0.835)	0.918 (0.697)	0.909 (0.688)	0.889	115	17.01 (18.16)
H3	$\Delta H_f^{298.15} = (7.911 \times \text{MR}) + (32.762 \times \text{SC_c}) - (13.531 \times \text{AC}) + (43.197 \times \text{S_dssC}) - (109.266 \times \text{S_ddsN}) + 7.575$	0.972 (0.839)	0.946 (0.704)	0.938 (0.695)	0.931	138	14.06 (19.66)
H4	$\Delta H_f^{298.15} = (7.935 \times \text{MR}) + (200.597 \times {}^3\chi) - (13.576 \times \text{AC}) + (43.116 \times \text{S_dssC}) - (109.465316780 \times \text{S_ddsN}) + 7.992$	0.972 (0.834)	0.945 (0.696)	0.938 (0.687)	0.930	137	14.17 (19.86)
H5	$\Delta H_f^{298.15} = (7.935 \times \text{MR}) + (115.815 \times {}^3\chi) - (13.575 \times \text{AC}) + (43.116 \times \text{S_dssC}) - (109.465 \times \text{S_ddsN}) + 7.992$	0.972 (0.834)	0.945 (0.696)	0.938 (0.687)	0.930	137	14.17 (19.86)
S1	$\Delta S_f^{298.15} = (7.773 \times {}^1\kappa) + (23.597 \times {}^3\chi) - (0.004 \times \text{VDM}) - (0.400 \times \text{S_ssO}) + 45.165$	0.994 (0.979)	0.988 (0.958)	0.987 (0.957)	0.984	826	1.98 (2.19)
S2	$\Delta S_f^{298.15} = (7.773 \times {}^1\kappa) + (40.8716 \times {}^3\chi) - (0.004 \times \text{VDM}) - (0.400 \times \text{S_ssO}) + 45.166$	0.994 (0.979)	0.987 (0.958)	0.987 (0.957)	0.984	826	1.98 (2.19)
S3	$\Delta S_f^{298.15} = (1.883 \times \text{RB}) + (5.346 \times {}^1\kappa) + 53.712$	0.988 (0.968)	0.977 (0.937)	0.975 (0.936)	0.971	876	2.76 (2.23)
S4	$\Delta S_f^{298.15} = (1.201 \times \text{RB}) + (5.753 \times {}^1\kappa) + (24.050 \times {}^3\chi) + 51.105$	0.991 (0.976)	0.983 (0.953)	0.981 (0.952)	0.978	767	2.39 (2.04)
S5	$\Delta S_f^{298.15} = (1.201 \times \text{RB}) + (5.753 \times {}^1\kappa) + (13.885 \times$	0.991 (0.976)	0.983 (0.953)	0.981 (0.952)	0.978	767	2.39 (2.04)

1							
2							
3							
4		$^3\chi) + 51.105$					
5	C1	$\Delta C_p^{298.15} = (1.008 \times ^0\chi) +$	0.997	0.994	0.993	0.993	2095
6		$(0.951 \times \text{MF}) + (0.198 \times$	(0.982)	(0.964)	(0.963)		1.08
7		$\text{MA}) - 5.529$					(1.46)
8							
9	C2	$\Delta C_p^{298.15} = (1.179 \times ^0\chi) +$	0.997	0.993	0.993	0.992	2005
10		$(0.471 \times \text{S_sCH}_3) +$	(0.982)	(0.964)	(0.963)		1.09
11		$(0.191 \times \text{MA})$					(1.45)
12							
13	C3	$\Delta C_p^{298.15} = (2.039 \times ^3\chi) +$	0.995	0.991	0.990	0.989	2239
14		$(0.243 \times \text{MA}) - 6.787$	(0.984)	(0.967)	(0.966)		1.29
15							(1.39)
16	C4	$\Delta C_p^{298.15} = (0.732 \times ^1\kappa) +$	0.997	0.993	0.993	0.992	1944
17		$(0.921 \times \text{MF}) + (0.204 \times$	(0.981)	(0.962)	(0.961)		1.12
18		$\text{MA}) - 4.842$					(1.51)
19							
20	C5	$\Delta C_p^{298.15} = (4.087 \times \text{SC_p})$	0.996	0.993	0.992	0.992	1925
21		$- (7.661 \times ^1\chi) + (0.226 \times$	(0.984)	(0.969)	(0.968)		1.12
22		$\text{MA}) - 6.229$					(1.35)
23							
24							
25							
26							
27							
28							
29							
30							
31							
32							
33							
34							
35							
36							
37							
38							
39							
40							
41							
42							
43							
44							
45							
46							
47							
48							
49							
50							
51							
52							
53							
54							
55							
56							
57							
58							
59							
60							

Figure 1



For Peer Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 2

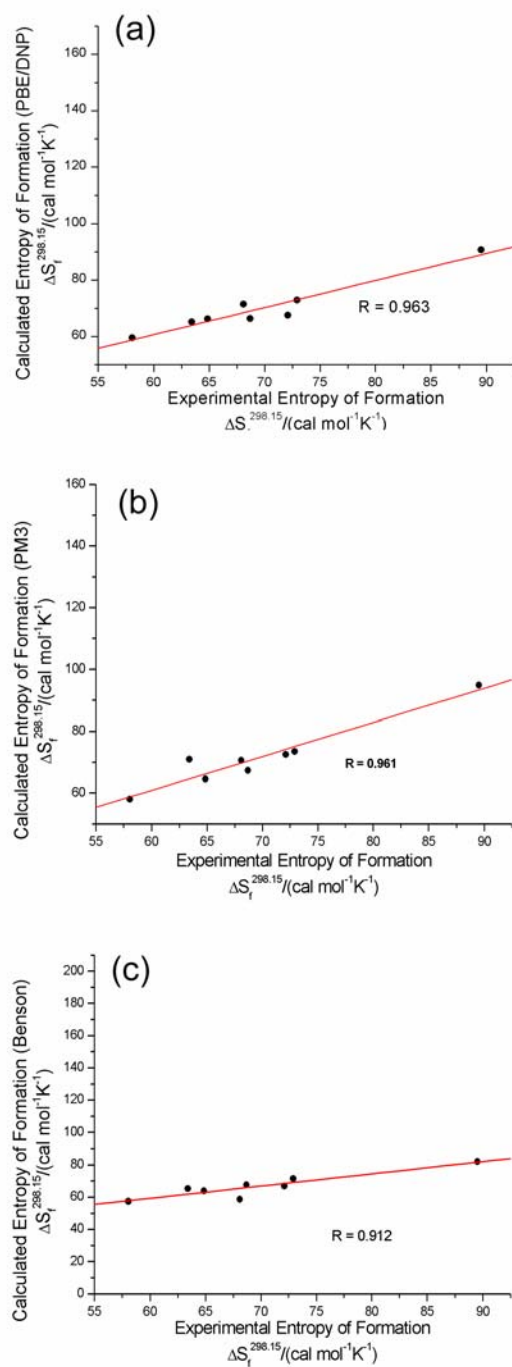
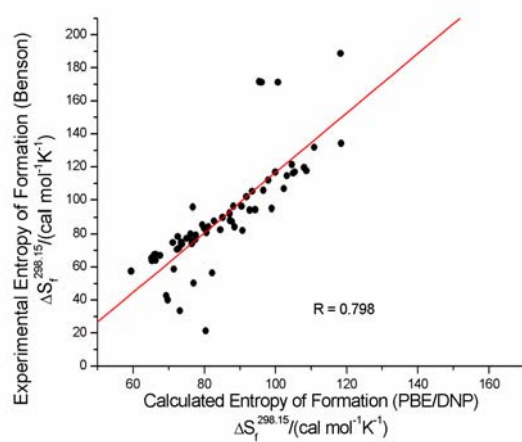


Figure 3



FCF Peer Review Only

Figure 4

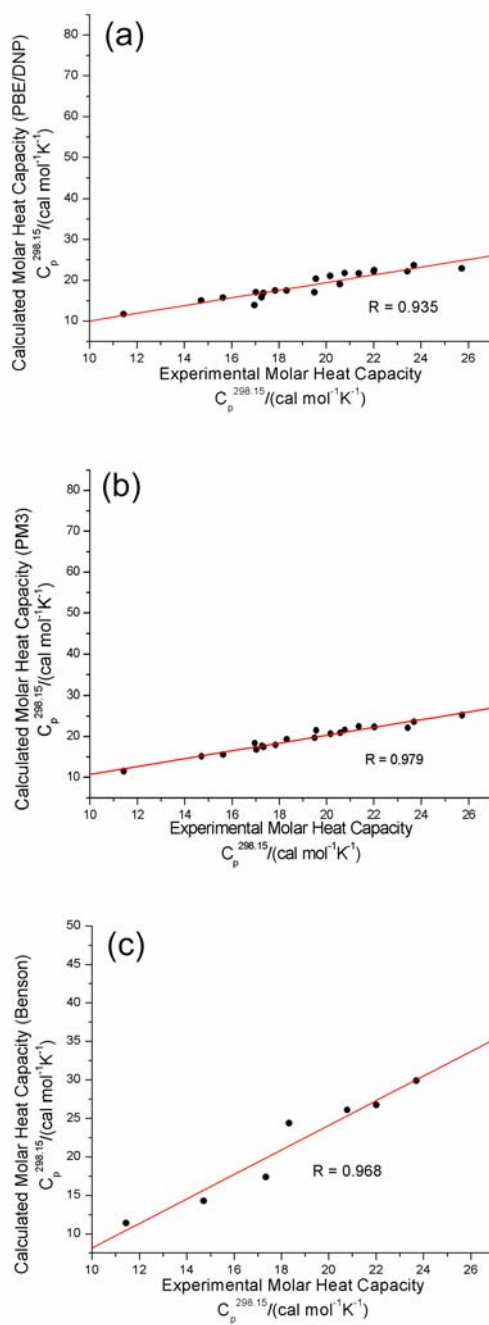


Figure 5

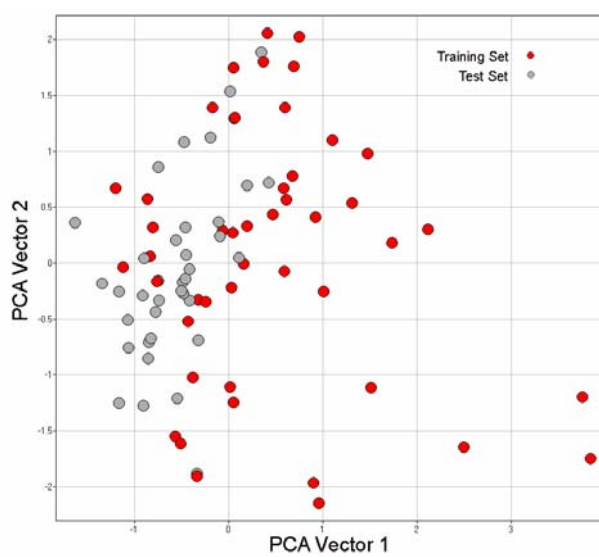
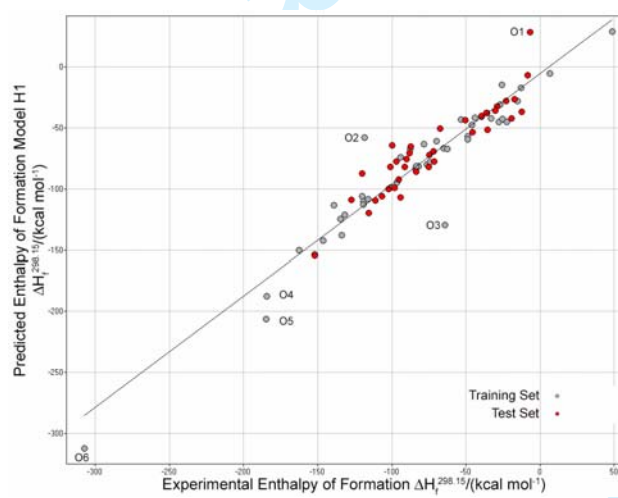


Figure 6



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 7

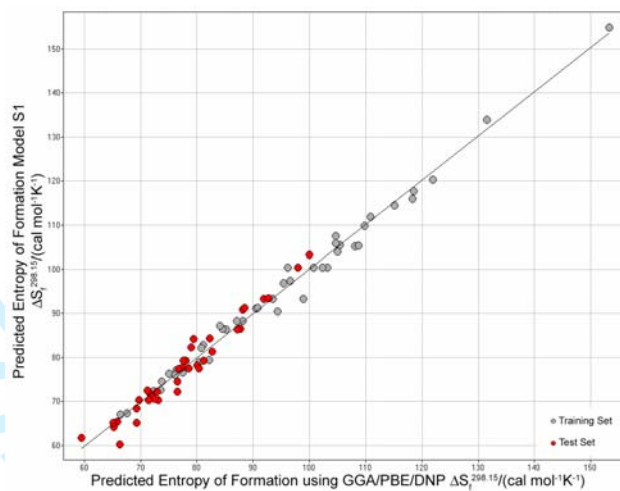
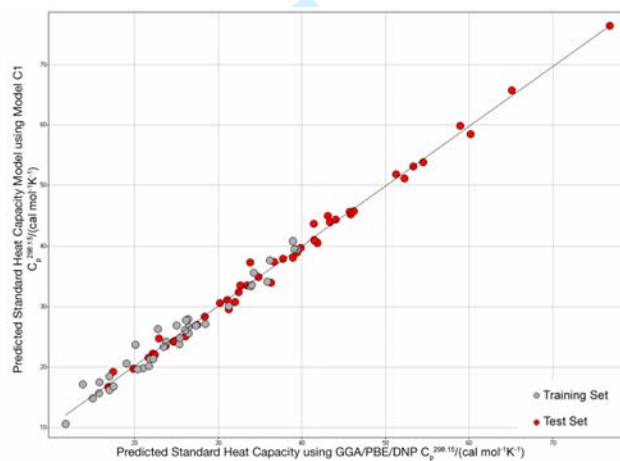


Figure 8



1
2
3
4 **Predicting thermochemical parameters of oxygen-containing heterocycles using**
5
6
7 **simple QSPR models**
8

9
10 NICO ADAMS[†], JOACHIM CLAUSS[‡], MARC MEUNIER⁺, ULRICH S. SCHUBERT^{†*}
11

12
13 [†]Laboratory of Macromolecular Chemistry and Nanoscience, Eindhoven University of Technology
14 and Dutch Polymer Institute, P. O. Box 513, 5600 MB Eindhoven, The Netherlands
15
16 (u.s.schubert@tue.nl)
17

18
19
20
21 [‡]Ticona GmbH, Core Technology, Professor-Staudinger-Strasse, 65451 Kelsterbach, Germany
22

23
24 ⁺Accelrys Ltd., 334 Cambridge Science Park, Cambridge CB4 0WN, United Kingdom
25
26
27
28
29
30

31
32
33 **Supporting Information**
34
35
36
37

-
- 38
39
40 1. Experimental (as available) and calculated standard enthalpies of formation for oxygen containing
41 heterocycles
42
43
44 2. Experimental (as available) and calculated standard entropies of formation for oxygen containing
45 heterocycles
46
47
48 3. Experimental (as available) and calculated standard molar heat capacities for oxygen containing
49 heterocycles
50
51
52
53 4. References
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1. Experimental (as available) and calculated standard enthalpies of formation for oxygen containing heterocycles

All values are given in kcal mol⁻¹. Experimental data were taken from ref [1] or ref [2].

Name (Whole Molecule)	$\Delta H_f^{298.15}$ (experimental)	$\Delta H_f^{298.15}$ (PM3)	$\Delta H_f^{298.15}$ (Benson)	$\Delta H_f^{298.15}$ Model H1
Oxirane	-12.58	-8.12	-12.50	-37.04
Propylene oxide	-22.63	-16.56	-21.80	-44.98
Chloromethyl oxirane	-25.8	-18.98	-27.20	-14.58
<i>cis</i> -1,2-Epoxy cyclopentane	-23.2	-17.05		-28.00
Cyclohexene oxide	-29.21	-24.89		-32.38
Propoxymethyl oxirane	-65.15	-60.03	-66.00	-66.78
1-Methylethoxymethyl oxirane	-71.56	-60.40	-70.40	-77.76
<i>cis</i> -1,2-Epoxy cycloheptane	-36.4	-28.69		-37.81
<i>cis</i> -2,3-Epoxybicyclo[2.2.1]heptane	-12.9	0.97		-17.09
Butoxymethyl oxirane	-69.77	-65.25	-71.00	-61.09
2-Methylpropoxymethyl oxirane	-74.69	-66.04	-73.20	-72.22
<i>tert</i> -Butoxymethyl oxirane	-76.41	-67.13	-79.50	-79.96
<i>cis</i> -1,2-Epoxy cyclooctane	-39.46	-29.11		-40.15
<i>cis</i> -9-Oxabicyclo[6.1.0]nonane	-39.46	-32.19		-40.64
<i>tert</i> -Butylperoxymethyl oxirane	-64.415	-51.28	-65.30	-129.64
Glycidyl butyrate	-119.9	-106.00	-118.10	-106.11
3-Methylbutoxymethyl oxirane	-78.44	-70.66	-77.30	-63.22
2-Propenoic acid	-94.22	-77.74	-91.20	-107.19
Phenoxymethyl oxirane	-27.82	-17.58	-27.00	-45.42
Phenylmethoxymethyl oxirane	-32.98	-20.23	-28.90	-42.15
Oxetane	-19.25	-26.69	-18.60	-42.04
β -Propiolactone	-67.61	-62.23	-67.60	-50.49
3,3-Dimethyl oxetane	-35.43	-41.12	-33.50	-51.48
4-Methylene oxetanone	-45.47	-37.08	-45.50	-53.72
3,3-Dimethyl oxetanone	-84.29	-75.79	-81.30	-83.93
3,3-Bis-chloromethyl oxetane	-48.75	-45.74	-44.40	-57.02
3-Ethyl-3-chloromethyl oxetane	-46.2	-47.26	-42.30	-47.55
Tetrahydrofuran	-44.03	-51.28	-43.30	-41.63
γ -Butyrolactone	-87.33	-91.44		-65.49
Tetrahydrofuran methanol	-88.20	-97.12	-88.40	-70.62
Dihydro-5-methyl furanone	-97.16	-97.74		-77.34
5-Hexyldihydro-2 furanone	-118.60	-124.28		-57.70
2,3-Dihydrofuran	-17.27	-24.79	-21.40	-26.44
2,3-Dihydro-5-methyl-furan	-30.21	-33.58	-29.90	-35.96
5-Nitro-2-acetoxy-2,5-dihydrofurfural diacetate	-307.20	-273.94		-312.21
Furan	-8.29	-4.02	-8.00	-6.93
2-Furanmethanol	-50.62	-51.77	-52.30	-43.52
2-Furancarboxaldehyde	-36.10	-37.25	-34.10	-37.89
Furylethylene	6.60	11.923	5.30	-5.49
2-Nitrofuran	-6.90	-10.36		28.43

1					
2					
3	Furancarboxylic acid	-94.40	-91.80	-95.10	-74.19
4	Methyl furoate	-96.80	-83.65	-90.30	-94.69
5	3-2-Furanyl-2-propenal	-25.30	-23.05	-25.00	-42.84
6	5-Nitro-2-furancarboxylic acid				
7	methyl ester	-87.70	-86.77		-67.79
8	3-5-Nitro-2-furyl-2-propenal	-15.50	-27.88		-28.06
9	2-Diacetoxymethyl furan	-184.70	-172.73	-199.40	-206.30
10	2-Diacetoxymethyl-5-nitrofuran	-184.40	-180.65		-187.61
11	1,3-Diphenylisobenzofuran	48.40	70.07		28.98
12	1,3-Dioxolane	-72.10	-82.46	-72.70	-69.06
13	Ethylene carbonate	-120.10	-125.012	-127.20	-87.18
14	2-Methyl-1,3-dioxolane	-83.70	-89.47	-83.00	-85.99
15	2-Methoxy-1,3-dioxolane	-115.50	-129.65	-118.60	-119.91
16	Propylene carbonate	-139.22	-131.69	-136.50	-113.23
17	2-Methyl-4-methylene-1,3-				
18	dioxolane	-91.50	-68.01		-81.92
19	2-Phenyl dioxolane	-49.07	-53.65	-50.80	-59.18
20	2-Ethoxy-4,4,5,5-tetramethyl-1,3-				
21	dioxolane	-162.40	-154.69	-162.90	-149.97
22	2-Methyl-2-phenyl-1,3-dioxolane	-62.60	-58.75	-63.30	-67.19
23	1,3-Dioxol-2-one	-100.05	-94.76		-64.11
24	2,4-Dimethyl-1,3-dioxole	-101.10	-69.43		-81.79
25	Tetrahydropyran	-53.16	-57.39	-53.60	-43.02
26	Tetrahydropyran-2-one	-90.30	-95.91		-75.84
27	Dihydropyran-2,6-dione	-127.20	-134.10	-131.20	-109.20
28	2-2-Methoxyethoxy-				
29	tetrahydropyran	-134.56	-136.10	-134.60	-124.68
30	3,4-Dihydropyran	-26.96	-33.28	-29.90	-30.85
31	1,3-Dioxane	-81.99	-87.76	-83.40	-81.59
32	2-Methyl-1,3-dioxane	-95.4	-92.22	-93.80	-92.33
33	4-Methyl-1,3-dioxane	-90.45	-91.23	-92.70	-92.87
34	2-Hydroxymethyl-1,3-dioxane	-132.00	-132.54	-129.60	-121.21
35	5,5-Dimethyl-1,3-dioxane	-100.67	-99.11	-98.30	-98.84
36	<i>trans</i> -4,5-Dimethyl-1,3-dioxane	-98.20	-94.46	-99.00	-99.23
37	<i>cis</i> -2,4-Dimethyl-1,3-dioxane	-102.30	-95.26	-102.10	-100.19
38	2,4,6-Trimethyl-1,3-dioxane	-106.80	-100.23	-112.30	-106.12
39	<i>cis</i> -2,2,4,6-Tetramethyl-1,3-				
40	dioxane	-119.10	-105.84	-124.80	-110.48
41	<i>trans</i> -2,2,4,6-Tetramethyl-1,3-				
42	dioxane	-116.00	-108.19	124.80	-108.47
43	1,4-Dimethyl-2,6,7-				
44	Trioxabicyclo[2.2.2]octane	-133.88	-132.37		-137.82
45	2,4,10-				
46	Trioxatricyclo[3.3.1.1(3,7)]decane	-119.32	-128.55	-119.30	-112.86
47	5,5-Dimethyl-2-phenyl-1,3-				
48	dioxane	-74.40	-64.80	-76.50	-77.86
49	2,5,5-Trimethyl-2-phenyl-1,3-				
50	dioxane	-83.81	-71.51	-89.00	-81.37
51	1,4-Dioxane	-75.36	-83.11	-75.40	-81.82
52	1,4-Dioxane-2,5-dione	-146.30	-156.37		-142.22

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

3,6-Dihydro-1,2-dioxin		-9.26	-7.50	-69.06
1,3,5-Trioxane	-111.32	-122.41	-111.20	-109.23
Paraldehyde	-152.10	-136.91	-142.40	-153.74
<i>cis</i> -2,4,6-Trimethyl-1,3,5-trioxane	-152.06	-136.35	-142.40	-154.77

For Peer Review Only

2. Experimental (as available) and calculated standard entropies of formation for oxygen-containing heterocycles

All values are given in $\text{cal mol}^{-1}\text{K}^{-1}$. Experimental data were taken from ref [1] or ref [2].

Name (Whole Molecule)	$\Delta S_f^{298.15}$ (experimental)	$\Delta S_f^{298.15}$ (DFT)	$\Delta S_f^{298.15}$ (PM3)	$\Delta S_f^{298.15}$ (Benson)	$\Delta S_f^{298.15}$ (Model S1)
Oxirane	58.08	59.54	58.06	57.40	61.69
Propylene oxide	68.69	66.38	67.37	67.59	67.05
Chloromethyl oxirane		75.01	75.60	77.23	76.38
<i>cis</i> -1,2-Epoxy cyclopentane		71.89	72.83		71.45
Cyclohexene oxide		77.45	76.86		77.83
Propoxymethyl oxirane		93.49	98.17	105.29	93.39
1-Methylethoxymethyl oxirane		91.85	95.73	101.93	93.36
<i>cis</i> -1,2-Epoxy cycloheptane		82.27	81.25		84.36
<i>cis</i> -2,3-Epoxy bicyclo[2.2.1]heptane		76.31	75.26		77.19
Butoxymethyl oxirane		103.19	106.23	114.71	100.48
2-Methylpropoxymethyl oxirane		97.98	102.74	112.04	100.46
<i>tert</i> -Butoxymethyl oxirane		102.33	97.98	106.79	100.39
<i>cis</i> -1,2-Epoxy cyclooctane		88.12	86.71		90.94
<i>cis</i> -9-Oxabicyclo[6.1.0]nonane		88.15	85.48		90.94
<i>tert</i> -Butylperoxymethyl oxirane		105.40	108.86	116.91	105.39
Glycidyl butyrate		108.05	108.75	119.61	105.21
3-Methylbutoxymethyl oxirane		104.63	113.32	121.46	107.50
2-Propenoic acid		99.97	107.43	116.81	103.24
Phenoxymethyl oxirane		96.62	100.22	105.77	97.46
Phenylmethoxymethyl oxirane		105.02	109.95	116.24	104.00
Oxetane	64.87	66.29	64.59	64.02	60.23
β -Propiolactone		69.26	68.08	42.60	65.12
3,3-Dimethyl oxetane		76.51	77.37	75.97	74.53
4-Methylene oxetanone		73.13	72.92	33.52	70.35
3,3-Dimethyl oxetanone		82.22	82.43	56.46	79.41
3,3-Bis-chloromethyl oxetane		98.93	92.93	95.23	93.36
3-Ethyl-3-chloromethyl oxetane		90.46	91.63	96.40	91.11
Tetrahydrofuran	72.11	67.57	72.57	66.94	67.31
γ -Butyrolactone		71.40	74.21		72.21
Tetrahydrofuran methanol		82.75	85.15	87.78	81.40
Dihydro-5-methyl furanone		81.19	81.36		79.33
5-Hexyldihydro-2 furanone		115.12	118.59		114.44
2,3-Dihydrofuran		65.91	68.07	67.27	65.44
2,3-Dihydro-5-methyl-furan		71.16	76.68	74.81	72.54
5-Nitro-2-acetoxy-2,5-dihydrofurfural diacetate		153.23	150.23		154.97
Furan		65.28	64.06	63.78	64.25
2-Furanmethanol		80.03	81.11	83.34	78.26
2-Furancarboxaldehyde		76.12	76.26	79.97	76.07
Furylethylene		77.47	77.96	76.56	76.59
2-Nitrofuran		78.98	80.47		82.36
Furancarboxylic acid		81.10	82.84	84.27	82.89
Methyl furoate		88.17	91.32	96.40	88.25

1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					
43					
44					
45					
46					
47					
48					
49					
50					
51					
52					
53					
54					
55					
56					
57					
58					
59					
60					

3. Experimental (as available) and calculated standard molar heat capacities for oxygen-containing heterocycles.

All values are given in $\text{cal mol}^{-1}\text{K}^{-1}$. Experimental data were taken from ref [1] or ref [2].

Name (Whole Molecule)	$C_p^{298.15}$ (experimental)	$C_p^{298.15}$ (DFT)	$C_p^{298.15}$ (PM3)	$C_p^{298.15}$ (Benson)	$C_p^{298.15}$ (Model C1)
Oxirane	11.44	11.75	11.514	11.38	10.56
Propylene oxide	17.34	16.86	17.454	17.38	16.71
Chloromethyl oxirane		19.88	20.139	19.93	19.76
<i>cis</i> -1,2-Epoxy cyclopentane		21.83	21.222		21.28
Cyclohexene oxide		26.49	25.783		25.64
Propoxymethyl oxirane		34.79	35.031	36.26	34.97
1-Methylethoxymethyl oxirane		34.23	35.374	36.76	35.59
<i>cis</i> -1,2-Epoxy cycloheptane		31.26	30.365		30.08
<i>cis</i> -2,3-Epoxybicyclo[2.2.1]heptane		27.47	26.025		26.98
Butoxymethyl oxirane		39.87	39.925	41.76	39.75
2-Methylpropoxymethyl oxirane		38.94	39.902	41.49	40.84
<i>tert</i> -Butoxymethyl oxirane		41.88	40.545	42.48	40.54
<i>cis</i> -1,2-Epoxy cyclooctane		35.87	35.138		34.16
<i>cis</i> -9-Oxabicyclo[6.1.0]nonane		36.29	34.851		34.01
<i>tert</i> -Butylperoxymethyl oxirane		45.86	45.427	46.88	45.22
Glycidyl butyrate		41.48	41.33		40.99
3-Methylbutoxymethyl oxirane		43.10	44.993	46.99	44.91
2-Propenoic acid		39.42	40.069		39.52
Phenoxymethyl oxirane		38.95	38.203		38.11
Phenylmethoxymethyl oxirane		43.36	43.497		43.92
Oxetane	14.71	15.01	15.095	14.28	14.82
β -Propiolactone	17.03	17.01	16.816		16.19
3,3-Dimethyl oxetane		25.01	26.612	25.53	26.93
4-Methylene oxetanone	20.17	21.03	20.675		19.86
3,3-Dimethyl oxetanone		28.41	28.442		28.34
3,3-Bis-chloromethyl oxetane		32.47	32.038	30.63	32.47
3-Ethyl-3-chloromethyl oxetane		32.64	33.845	33.58	33.54
Tetrahydrofuran	18.32	17.46	19.361	24.38	19.25
γ -Butyrolactone	20.58	19.06	20.875		20.60
Tetrahydrofuran methanol		28.46	28.418	33.48	27.15
Dihydro-5-methyl furanone		26.44	26.347		26.68
5-Hexyldihydro-2 furanone		51.27	50.647		51.84
2,3-Dihydrofuran	17.27	15.79	17.685		17.52
2,3-Dihydro-5-methyl-furan		20.13	23.508		23.69
5-Nitro-2-acetoxy-2,5-dihydrofurfural diacetate		76.81	76.805		76.47
Furan	15.63	15.74	15.562		15.72
2-Furanmethanol		25.35	24.873		23.83
2-Furancarboxaldehyde	23.44	22.15	22.132		22.21
Furylethylene		24.68	24.219		24.21
2-Nitrofuran		23.81	23.849		24.20

1					
2					
3					
4	Furancarboxylic acid		26.07	25.952	25.12
5	Methyl furoate		31.11	31.112	31.11
6	3-2-Furanyl-2-propenal		30.23	30.399	30.62
7	5-Nitro-2-furancarboxylic acid				
8	methyl ester			39.584	39.40
9	3-5-Nitro-2-furyl-2-propenal		39.39	38.798	38.99
10	2-Diacetoxymethyl furan		52.22	51.316	51.16
11	2-Diacetoxymethyl-5-nitrofuran		58.93	59.599	59.89
12	1,3-Diphenylisobenzofuran		65.07	64.896	65.77
13	1,3-Dioxolane	16.97	13.86	18.345	17.18
14	Ethylene carbonate	19.5	17.03	19.681	18.46
15	2-Methyl-1,3-dioxolane		23.51	23.799	23.33
16	2-Methoxy-1,3-dioxolane		22.79	27.522	26.32
17	Propylene carbonate	25.72	22.91	25.232	24.75
18	2-Methyl-4-methylene-1,3-				
19	dioxolane		27.36	27.541	26.92
20	2-Phenyl dioxolane		36.69	37.117	37.43
21	2-Ethoxy-4,4,5,5-tetramethyl-1,3-				
22	dioxolane		54.49	54.263	53.85
23	2-Methyl-2-phenyl-1,3-dioxolane		41.44	42.731	43.63
24	1,3-Dioxol-2-one	17.84	17.49	17.915	16.83
25	2,4-Dimethyl-1,3-dioxole		26.44	28.095	27.94
26	Tetrahydropyran	23.70	23.70	23.599	29.88
27	Tetrahydropyran-2-one		25.40	25.38	24.84
28	Dihydropyran-2,6-dione		26.01	27.019	26.12
29	2-2-Methoxyethoxy-				
30	tetrahydropyran		45.69	46.087	45.67
31	3,4-Dihydropyran	22.03	22.44	22.291	22.04
32	1,3-Dioxane	21.37	21.68	22.471	21.57
33	2-Methyl-1,3-dioxane		26.13	28.056	27.73
34	4-Methyl-1,3-dioxane		27.77	28.271	27.68
35	2-Hydroxymethyl-1,3-dioxane		31.24	31.612	29.57
36	5,5-Dimethyl-1,3-dioxane		33.46	33.759	33.58
37	<i>trans</i> -4,5-Dimethyl-1,3-dioxane		33.87	33.775	33.34
38	<i>cis</i> -2,4-Dimethyl-1,3-dioxane		33.98	33.743	33.65
39	2,4,6-Trimethyl-1,3-dioxane		39.14	39.356	39.52
40	<i>cis</i> -2,2,4,6-Tetramethyl-1,3-				
41	dioxane		46.18	45.245	45.73
42	<i>trans</i> -2,2,4,6-Tetramethyl-1,3-				
43	dioxane		44.00	44.531	44.36
44	1,4-Dimethyl-2,6,7-				
45	Trioxabicyclo[2.2.2]octane		33.79	38.958	37.33
46	2,4,10-				
47	Trioxatricyclo[3.3.1.1(3,7)]decane		31.99	31.176	30.76
48	5,5-Dimethyl-2-phenyl-1,3-				
49	dioxane		53.32	52.699	53.16
50	2,5,5-Trimethyl-2-phenyl-1,3-				
51	dioxane		60.15	58.064	58.46
52	1,4-Dioxane	22.02	22.23	22.384	26.76
53	1,4-Dioxane-2,5-dione		24.79	25.943	24.31

3,6-Dihydro-1,2-dioxin	20.78	21.76	21.608	26.10	20.22
1,3,5-Trioxane	19.57	20.34	21.526		19.63
Paraldehyde		37.75	38.31		37.93
<i>cis</i> -2,4,6-Trimethyl-1,3,5-trioxane		36.21	38.382		37.73

4. References

- [1] P.J. Linstrom, W.G. Mallard (Eds.). *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, National Institute of Standards and Technology (NIST). Available online at <http://webbook.nist.gov> (accessed November 2005).
- [2] R. D. Johnson III (Ed.). *NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 10*, National Institute of Standards and Technology (NIST). Available online at <http://srdata.nist.gov/cccbdb> (accessed November 2005).