

# Multivariate Data Analysis of Molecular Descriptors Estimated by Use of Semiempirical Quantum Chemistry Methods. Principal Properties for Synthetic Screening of 2-Chloromethyloxirane and Analogous Bis-alkylating C<sub>3</sub> Moieties

Hans-René Bjørsvik<sup>†,a</sup> and Hanno Priebe<sup>b</sup>

Nycomed Imaging, <sup>a</sup>Department of Process Chemistry and <sup>b</sup>Department of Development Chemistry, P.O. Box 4220 Torshov, N-0401 Oslo, Norway

Bjørsvik, H.-R. and Priebe, H., 1995. Multivariate Data Analysis of Molecular Descriptors Estimated by Use of Semiempirical Quantum Chemistry Methods. Principal Properties for Synthetic Screening of 2-Chloromethyloxirane and Analogous Bis-alkylating C<sub>3</sub> Moieties. – Acta Chem. Scand. 49: 446–456 © Acta Chemica Scandinavica 1995.

The principal properties of 106 different bis-alkylating C<sub>3</sub> moieties have been calculated and evaluated in the synthesis of 3,3',5',5'-tetrakis(2,3-dihydroxypropylcarbamoyl)-2,2',4,4',6,6'-hexaiodo-*N,N'*-(2-hydroxypropane-1,3-diyl)diacetanilide. Sixteen molecular descriptors, most of them estimated by means of semiempirical quantum chemistry methods were used to describe the bis-alkylating C<sub>3</sub> moieties. Hence, a 106 × 16 data matrix was obtained. This data matrix was subjected to principal component analysis (PCA) in order to estimate the principal properties. Based upon a uniform spread in the principal properties, 13 bis-alkylating C<sub>3</sub> reagents were selected to be used in laboratory experiments in an attempt to describe the variation in yield in the synthesis of 3,3',5',5'-tetrakis(2,3-dihydroxypropylcarbamoyl)-2,2',4,4',6,6'-hexaiodo-*N,N'*-(2-hydroxypropane-1,3-diyl)diacetanilide. The relationship between the yield, two experimental descriptors and the principal properties as well as the two-factor interactions and quadratic terms of these were modelled by the partial least-squares (PLSR) method. The results from this modelling shows that the principal properties, and hence the molecular descriptors, contain useful information for design of screening experimental plans where bis-alkylating C<sub>3</sub> moieties need to be evaluated.

Bis-alkylation using C<sub>3</sub> moieties is a useful reaction leading to the structure elements **D** (Schemes 1 and 2) found in several biological active compounds, such as antifungals,<sup>1</sup> antiinflammatory agents,<sup>2</sup> β-adrenoreceptor antagonists,<sup>3,4</sup> antiamebics,<sup>5</sup> antiarteriosclerotics,<sup>6</sup> and spermicides.<sup>7</sup> This structure element also exists in macrocyclic compounds,<sup>8</sup> in X-ray absorbing agents,<sup>9</sup> and in flame restraining agents.<sup>10</sup> Hence, these C<sub>3</sub> bis-alkylating reagents constitute an important building block in various procedures in synthetic organic chemistry.

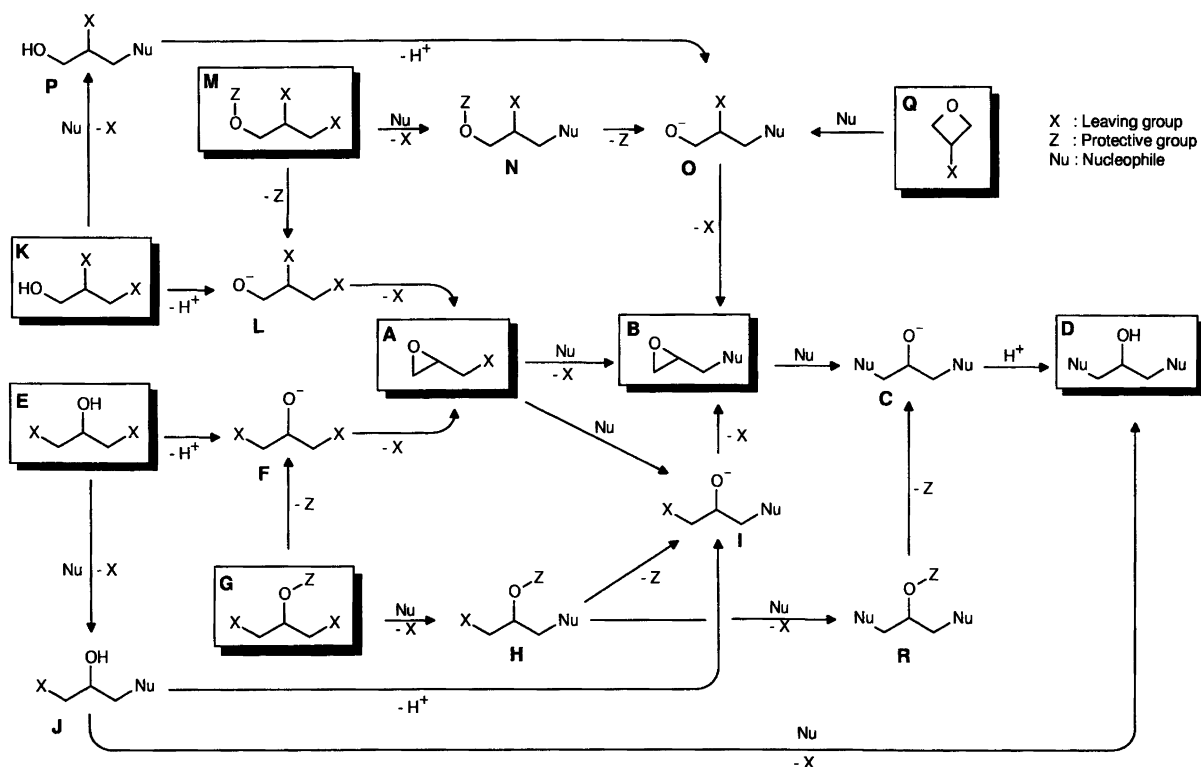
Many of these compounds are prepared by using epichlorohydrin (2-chloromethyloxirane) **A** (X = Cl), a well known and readily available C<sub>3</sub> synthon, useful for the synthesis of 1,2,3-trifunctionalized compounds.<sup>11</sup> Epichlorohydrin can react with equal or different nucleo-

philic compounds at C1 and C3 (sequence: **A** → **B** → **C** → **D**) leading to a hydroxy group at the C2 atom of the C<sub>3</sub> moiety. There are many compatible reagents showing the same type of chemical conversion as epichlorohydrin, e.g., 1,3-dihalohydrins **E**, 1,2-dihalohydrins **K**, or analogues with several other types of leaving group, X, such as ammonium (NH<sub>2</sub>, NHR, and NR<sub>2</sub>), betylates (ammonioalkanesulfonate esters), brosylates (4-bromobenzenesulfonates), alkyl fluorosulfonates, other halogens, nonaflates (nonafluorobutanesulfonates), nosylates (4-nitrobenzenesulfonates), mesylates (methanesulfonates), tosylates (4-methylbenzenesulfonates), tresylates (2,2,2-trifluoroethanesulfonates), and triflates (trifluoromethanesulfonates).<sup>12</sup> All of these leaving groups may be used in any combination.

On an industrial scale, epichlorohydrin is produced from **E** or **K** (X = Cl) and calcium hydroxide.<sup>13</sup> Thus **E** and **K** are 'pro-reagents' since they form **A** *in situ* under basic conditions according to the sequence:

<sup>†</sup> To whom correspondence should be addressed.

Present address: Euro Vanillin, PO Box 330, N-1701 Sarpsborg, Norway.



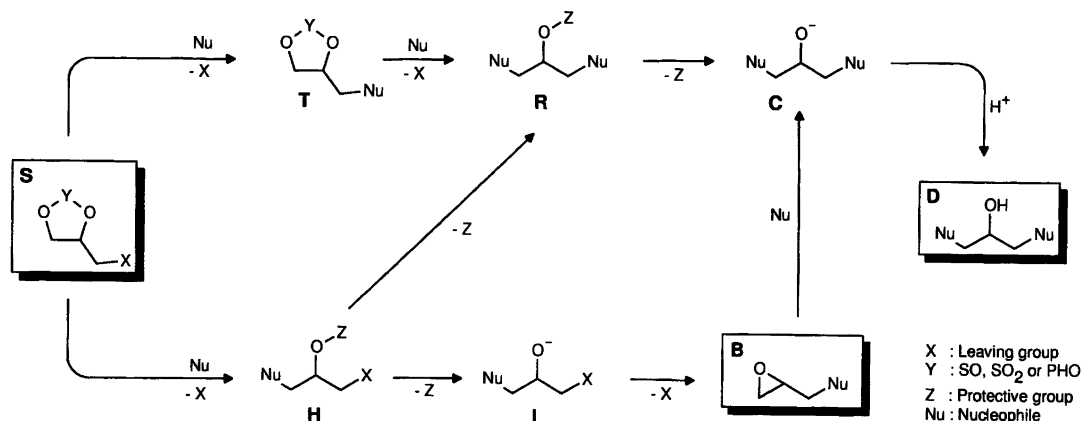
Scheme 1.

$E \rightarrow F \rightarrow A \rightarrow B \rightarrow C \rightarrow D$  or  $K \rightarrow L \rightarrow A \rightarrow B \rightarrow C \rightarrow D$ ,  
 or alternatively according to:  $E \rightarrow J \rightarrow I \rightarrow B \rightarrow C \rightarrow D$ ,  
 $E \rightarrow J \rightarrow D$ , or  $K \rightarrow P \rightarrow O \rightarrow B \rightarrow C \rightarrow D$ .

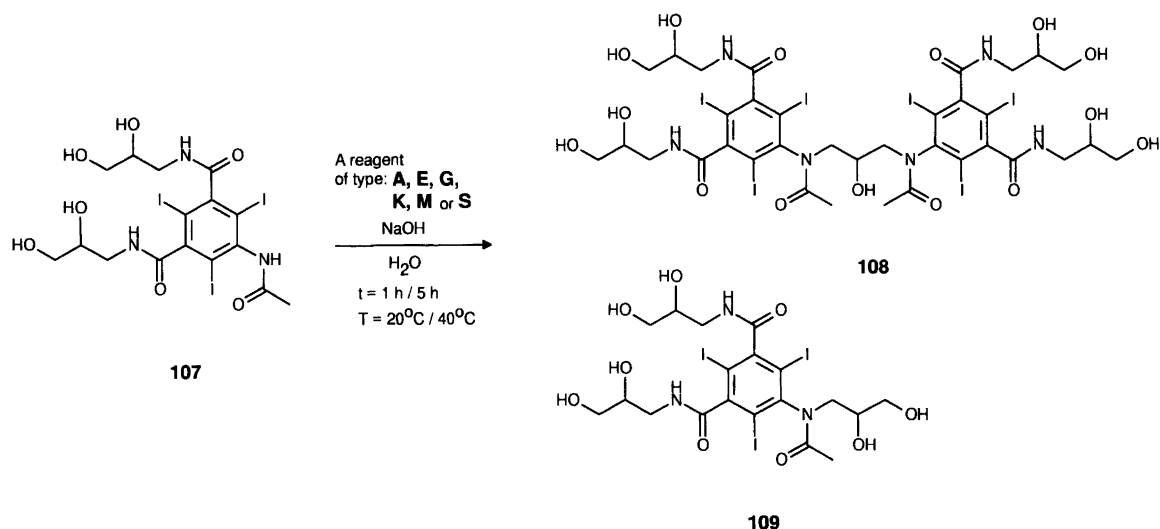
Protection of the hydroxy group in E or K by Z, affords another group of 'pro-reagents', G and M, respectively. Z can be acyl, benzyl, carbamate, phosphate, etc., removable under basic conditions. Hence, G and M offer additional mechanistic paths for reactions with nucleophiles under basic conditions:  $G \rightarrow F \rightarrow A \rightarrow B \rightarrow C \rightarrow D$ ,  $G \rightarrow H \rightarrow I \rightarrow B \rightarrow C \rightarrow D$ ,  $G \rightarrow H \rightarrow R \rightarrow C \rightarrow D$ ,  $M \rightarrow L \rightarrow A \rightarrow B \rightarrow C \rightarrow D$ , and  $M \rightarrow N \rightarrow O \rightarrow B \rightarrow C \rightarrow D$ . The 3-haloxyethanes Q may also form D with nucleophiles under basic conditions:  $Q \rightarrow O \rightarrow B \rightarrow C \rightarrow D$ .

The epoxy function in A can also be replaced by cyclic sulfite,<sup>14,15</sup> cyclic sulfate<sup>14</sup> or cyclic phosphite functions, S (Scheme 2), where Y is SO, SO<sub>2</sub> and PHO, respectively. The structure element S is expected to show corresponding ring-opening reactions as in epoxides according to the following reaction sequences:  $S \rightarrow H \rightarrow I \rightarrow B \rightarrow C \rightarrow D$ ,  $S \rightarrow H \rightarrow R \rightarrow C \rightarrow D$ , and  $S \rightarrow T \rightarrow R \rightarrow C \rightarrow D$ .

On basis of the reagent structures A, E, G, K, M, and S, varying the leaving groups X and the protective groups Z and Y, a multitudinous list of C<sub>3</sub> bis-alkylating reagents, which may all show the same type of chemical conversion as epichlorohydrin, can be compiled.



Scheme 2.



Scheme 3.

This list of bis-alkylating reagents offers a wide range of choices in development of synthetic procedures. This is particularly relevant when the substrate is a poly-nucleophile or when an ambident nucleophile is present and the synthetic procedure must provide optimal conditions both for yield and selectivity. To facilitate selection of a few reagents to be tested during experimental runs, either measured chemical and physical properties, spectroscopic data or computed molecular descriptors may be used to define a data matrix. A data reduction technique is used to obtain a *design-space* based on the *principal properties*.<sup>16,17</sup> Subsequently a systematic reagent selection principle, e.g., the statistical experimental design<sup>18–20</sup> approach is applied to select a reduced set of representative reagents to be tested in the laboratory.

This article illustrates the approach by showing that the variation in yield of a synthetic organic reaction (Scheme 3) may be explained by the calculated molecular descriptors in terms of the *principal properties*. This is shown by selecting a relevant set of bis-alkylating reagents for laboratory experimentation. The yields obtained from these experiments are then multivariately correlated to the corresponding *principal properties* of the reagents used.

## Methods and results

*Description of the bis-alkylating C<sub>3</sub> moieties.* Semiempirical quantum chemistry methods<sup>21</sup> constitute a set of versatile methods for describing molecular properties. These methods are very useful for the organic chemist as molecule properties may be estimated even for substances that never have been prepared. Hence, rare reagents can also be included in a common multivariate classification by estimating some of their molecular properties. In the present work, 106 different bis-alkylating C<sub>3</sub> reagents were studied. For each of these, sixteen different molecule

descriptors were determined and subjected to the multivariate data analysis.  $x_1$ : MW [g mol<sup>-1</sup>] (molecular weight),  $x_2$ : vdWS [Å<sup>2</sup>] (van der Waals surface),  $x_3$ :  $E^{298}$  [cal mol<sup>-1</sup>] (enthalpy at 298 K),  $x_4$ :  $S^{298}$  [cal (mol K)<sup>-1</sup>] (entropy at 298 K),  $x_5$ :  $C_p^{298}$  [cal (mol K)<sup>-1</sup>] (heat capacity at 298 K),  $x_6$ : ZPE [kcal mol<sup>-1</sup>] (zero point energy),  $x_7$ :  $H_f$  [kcal mol<sup>-1</sup>] (heat of formation),  $x_8$ :  $\mu$  [Debye] (dipole moment),  $x_9$ :  $\epsilon_{\text{HOMO}}$  [eV] (energy of highest occupied molecular orbital),  $x_{10}$ :  $\epsilon_{\text{LUMO}}$  [eV] (energy of lowest unoccupied molecular orbital),  $x_{11}$ :  $\eta$  [eV] (absolute electronegativity),  $x_{12}$ :  $\chi$  [eV] (the hardness),  $x_{13}$ :  $\delta_{\text{O}}$  [charge] (atomic point charge on O in the C<sub>3</sub> moiety),  $x_{14}$ :  $\delta_{\text{C}_2}$  [charge] (atomic point charge on C2 in the C<sub>3</sub> moiety),  $x_{15}$ :  $\delta_{\text{C}_t}$  [charge] (the lowest atomic point charge on terminal C in the C<sub>3</sub> moiety), and  $x_{16}$ :  $\delta_{\text{C}_h}$  [charge] (the highest atomic point charge on terminal C in the C<sub>3</sub> moiety). The list of the investigated bis-alkylating C<sub>3</sub> moieties is given in the Experimental section with the corresponding calculated results, the scores and loadings in Tables 1 and 2, respectively.

Table 1 Principal properties.

<i>i</i>	PC scores					
	<i>t</i> <sub>1</sub>	<i>t</i> <sub>2</sub>	<i>t</i> <sub>3</sub>	<i>t</i> <sub>4</sub>	<i>t</i> <sub>5</sub>	<i>t</i> <sub>6</sub>
1	-3.842	-1.794	2.504	2.078	-0.157	0.012
2	-2.139	-1.076	-0.097	-0.272	0.249	-0.510
3	-2.201	0.169	0.099	-0.197	0.196	0.049
4	-1.867	1.194	-0.336	-0.494	0.077	0.194
5	-2.572	-0.277	0.945	-0.292	0.647	-0.302
6	-2.185	0.808	0.547	-0.704	0.746	-0.132
7	-1.887	2.081	0.043	-0.864	0.633	0.051
8	-1.955	1.345	0.530	-1.181	0.985	-0.416
9	-1.751	2.291	0.449	-1.355	1.040	-0.417
10	-1.640	2.869	0.548	-1.606	1.205	-0.670
11	-3.745	-1.672	2.505	1.760	-0.027	0.131
12	-2.616	-1.631	-0.135	0.927	-0.193	-0.404
13	-2.089	-1.064	-1.435	0.997	-0.498	-0.462

Table 1 Continued

<i>i</i>	PC scores					
	<i>t</i> <sub>1</sub>	<i>t</i> <sub>2</sub>	<i>t</i> <sub>3</sub>	<i>t</i> <sub>4</sub>	<i>t</i> <sub>5</sub>	<i>t</i> <sub>6</sub>
14	-1.802	-0.466	-2.326	1.327	-0.830	-0.543
15	-2.721	-0.711	0.904	0.114	0.346	-0.128
16	-2.622	-0.801	0.077	0.585	-0.064	-0.478
17	-2.140	-0.276	-1.183	0.672	-0.354	-0.619
18	-1.855	0.333	-2.065	1.026	-0.685	-0.700
19	-2.169	0.225	0.101	-0.553	0.346	-0.206
20	-2.230	0.260	-0.498	0.162	-0.120	-0.416
21	-2.024	0.618	-1.221	0.561	-0.473	-0.590
22	-1.774	1.189	-2.075	0.931	-0.806	-0.715
23	-1.849	1.163	-0.364	-0.886	0.268	-0.226
24	-1.929	1.265	-0.959	-0.143	-0.192	-0.408
25	-1.843	1.654	-1.483	0.426	-0.589	-0.536
26	-1.705	2.126	-2.146	0.887	-0.936	-0.711
27	-4.246	-1.493	1.531	1.808	-0.219	0.378
28	-3.634	-0.660	0.523	0.377	0.677	-0.796
29	-3.090	0.711	-0.382	-0.101	0.611	-0.539
30	-2.742	1.490	-0.789	-0.497	0.510	-0.538
31	-0.040	-1.425	-1.946	-0.659	0.232	0.624
32	2.087	0.069	-1.862	0.319	0.089	2.783
33	-1.018	-0.447	-0.569	-0.642	-0.831	0.665
34	-1.100	0.204	-0.392	-1.016	-0.726	0.644
35	-1.037	0.817	-0.308	-1.200	-0.815	0.471
36	-0.979	1.568	-0.253	-1.253	-0.957	0.446
37	-0.922	-2.403	0.830	-1.117	-1.641	0.422
38	-0.914	-1.553	1.033	-1.481	-1.622	0.611
39	-0.817	-0.239	0.933	-1.414	-1.788	0.975
40	-0.513	0.738	0.502	-1.684	-1.880	1.008
41	-1.694	-1.280	1.092	0.837	0.527	0.467
42	-1.601	-0.540	0.378	1.034	0.081	0.139
43	-1.005	0.687	-0.865	0.949	-0.346	-0.032
44	-0.660	2.132	-1.777	1.240	-0.835	-0.114
45	-1.579	-1.344	0.803	1.140	0.116	1.314
46	-1.222	-0.033	0.727	-0.204	0.891	0.498
47	-0.699	1.014	0.690	-1.188	1.263	-0.025
48	-0.259	3.561	0.294	-1.426	1.341	0.771
49	0.074	-2.993	1.732	-0.908	-1.642	-0.179
50	0.132	-1.246	2.081	-1.723	-1.005	-0.503
51	0.522	0.559	2.037	-2.297	-0.737	-0.553
52	0.854	2.004	1.699	-2.490	-0.916	-0.763
53	4.512	-2.690	1.836	-0.087	-1.107	0.184
54	4.618	-0.692	1.067	0.749	-1.561	0.199
55	4.839	1.219	0.580	1.648	-1.951	-0.132
56	5.029	1.545	0.481	1.834	-2.070	-0.384
57	5.476	4.221	-0.190	2.916	-2.713	-0.971
58	4.370	-2.418	1.881	-0.300	-1.044	0.018
59	4.293	0.421	2.196	-0.632	-0.291	-0.573
60	4.956	3.236	2.383	-0.572	-0.152	-0.820
61	5.522	5.648	2.685	-0.492	-0.120	-1.118
62	-1.426	-1.780	0.645	0.831	0.328	0.614
63	-1.174	-0.501	1.001	-0.275	0.888	0.513
64	-0.587	0.896	0.715	-1.229	1.244	0.197
65	-0.206	2.280	0.703	-1.804	1.451	0.000
66	-0.732	0.298	0.528	-0.775	0.910	0.326
67	-0.496	1.399	0.166	-0.977	0.824	0.413
68	-1.363	-1.886	0.995	0.417	0.756	-0.252
69	-1.455	-0.913	0.524	0.888	0.319	-0.442
70	-0.951	0.694	-0.718	1.033	-0.157	-0.371
71	-0.543	2.323	-1.724	1.359	-0.675	-0.288
72	-1.030	0.328	-0.176	0.596	-0.022	-0.327
73	-4.393	-0.534	2.061	1.471	0.358	0.591
74	-3.571	-0.424	-0.117	1.178	0.045	0.155
75	-3.046	-0.259	-1.395	1.161	-0.245	-0.221
76	-2.726	0.044	-2.259	1.362	-0.575	-0.449

Table 1 Continued

<i>i</i>	PC scores					
	<i>t</i> <sub>1</sub>	<i>t</i> <sub>2</sub>	<i>t</i> <sub>3</sub>	<i>t</i> <sub>4</sub>	<i>t</i> <sub>5</sub>	<i>t</i> <sub>6</sub>
77	-0.075	-0.345	-1.584	-0.975	0.852	0.952
78	2.043	0.201	-1.223	-0.334	0.795	2.535
79	2.213	-1.029	-0.958	-0.165	0.447	1.047
80	5.481	0.703	-0.231	1.953	0.958	2.040
81	2.372	-1.458	-0.783	-0.273	0.640	0.821
82	5.950	0.044	-0.359	1.578	0.665	3.214
83	-0.090	-1.497	1.239	1.899	0.574	0.664
84	0.006	0.019	1.728	0.954	1.104	0.448
85	0.579	1.546	1.316	0.014	1.457	0.304
86	0.951	2.923	1.296	-0.576	1.636	0.190
87	1.425	-2.222	-1.122	-0.602	0.992	-2.131
88	6.773	-2.084	0.136	0.921	1.183	-0.923
89	2.471	-2.546	3.182	2.840	0.396	-1.817
90	4.784	-3.176	-1.989	-1.942	1.181	-2.345
91	6.234	-0.305	-1.466	0.614	1.677	-0.835
92	5.496	0.519	-0.812	1.357	0.976	0.598
93	-0.200	-1.263	1.408	1.398	1.004	0.128
94	-0.198	-0.474	0.974	1.785	0.587	-0.184
95	0.370	0.921	-0.377	1.721	0.134	-0.165
96	0.728	2.455	-1.348	2.015	-0.373	-0.157
97	7.212	-2.564	-1.069	-1.123	1.459	-1.310
98	1.339	-1.858	-2.083	-0.850	0.357	-0.253
99	-0.467	-2.114	-1.803	-0.869	0.112	-0.149
100	1.591	0.279	-2.385	0.274	0.458	0.781
101	4.482	-2.549	-2.236	-1.312	0.273	-0.603
102	1.377	-1.820	-2.233	-0.938	0.495	-0.564
103	0.106	-2.901	-0.418	-2.375	-1.307	0.137
104	0.060	-1.800	-0.341	-2.613	-1.297	0.287
105	0.125	-0.381	-0.478	-2.501	-1.476	0.672
106	0.199	0.599	-0.518	-2.483	-1.586	0.730

Principal component analysis<sup>22,23</sup> (PCA) and principal properties.<sup>24</sup> The descriptor variables  $x_1$ – $x_{16}$  estimated by use of semiempirical methods, can be considered as point measurements of different molecular properties for each of the 106 different bis-alkylating reagents. The results from these calculations are available as *supplementary material* deposited at The Royal Danish Library.<sup>25</sup> From several viewpoints, these descriptor variables portray different energies and electronic behaviour for the characterised molecule. It is reasonable to assume that molecular descriptors which describe the *same* intrinsic property of the molecule will be correlated over the set of compounds, while other descriptors which depend on *different* intrinsic properties will be almost uncorrelated over the whole set of compounds. By using the PCA method, these features will be taken into account, since the PCA method is able to extract the uncorrelated principal information from the data matrix  $X$ .

The PCA method corresponds to factorisation of the data matrix  $X$ , which in this case consists of 106 reagents each described by 16 molecular descriptors, into means ( $\bar{x}_k$ ), the principal component scores ( $t_{ia}$ ) (which displays the between-compound variation), the principal component loadings ( $p_{ak}$ ) (which describes the within-compound variation of the descriptors), and residuals ( $\epsilon_{ik}$ ).

Table 2. PC loadings

	P <sub>a1</sub>	P <sub>a2</sub>	P <sub>a3</sub>	P <sub>a4</sub>	P <sub>a5</sub>	P <sub>a6</sub>	P <sub>a7</sub>	P <sub>a8</sub>	P <sub>a9</sub>	P <sub>a10</sub>	P <sub>a11</sub>	P <sub>a12</sub>	P <sub>a13</sub>	P <sub>a14</sub>	P <sub>a15</sub>	P <sub>a16</sub>
a	MW	vdWS	E <sup>298</sup>	S <sup>298</sup>	C <sub>p</sub> <sup>298</sup>	ZPE	H <sub>f</sub>	μ	ε <sub>HOMO</sub>	ε <sub>HOMO</sub>	η	χ	δ <sub>0</sub>	δ <sub>C<sub>2</sub></sub>	δ <sub>C<sub>L</sub></sub>	δ <sub>C<sub>H</sub></sub>
1	0.283	0.341	0.354	0.352	0.355	0.319	-0.261	0.234	0.013	-0.255	-0.231	-0.232	-0.137	0.085	0.063	0.034
2	0.312	0.122	0.003	0.042	-0.009	0.014	0.328	-0.149	0.524	-0.018	-0.212	-0.234	-0.023	-0.047	-0.496	-0.364
3	0.069	0.103	0.091	0.097	0.076	0.120	-0.150	0.026	-0.147	0.435	-0.339	0.446	-0.295	0.435	-0.038	-0.348
4	0.097	0.179	0.145	0.145	0.144	0.260	0.063	-0.169	0.183	0.325	-0.380	0.212	0.287	-0.461	0.245	0.336
5	-0.157	0.012	0.081	0.060	0.097	0.170	0.068	-0.203	-0.049	-0.065	0.081	-0.037	0.772	0.385	0.121	-0.328
6	-0.256	-0.054	-0.082	-0.122	-0.034	0.222	0.237	0.731	0.352	0.052	-0.202	-0.100	0.045	0.230	0.153	0.116

This is described mathematically as in eqn. (1).

$$x_{ik} = \bar{x}_k + \sum_{a=1}^A t_{ia} p_{ak} + \varepsilon_{ik} \quad (1)$$

The first principal component describes the largest systematic variation in the data matrix  $X$ , the second described the next largest and so on. In the final model, the number of principal components,  $A$ , is much smaller than the number of variables,  $K$ . Hence, the principal component model gives a considerable simplification. In order to determine the optimal number of principal components, the cross-validation method<sup>26</sup> with 106 cross-validation elements randomly selected was used. A six principal component model ( $A = 6$ ) was determined. This model explained 90% ( $42 + 14 + 11 + 11 + 6 + 6$ ) of the variance in the data matrix  $X$ .

The principal component scores  $t_{ia}$ , listed in Table 1, show the location of the information for the alkylating reagent  $i = 1-106$  along the  $a$ th principal component and describe the between-reagent variation. Hence, sets of molecular descriptors which are similar will thus have similar score value. Fig. 1 shows three two-way plots of the estimated scores for principal component (PC) #1 versus PC #2, PC #1 versus PC #3, and PC #2 versus PC #3.

The loadings  $p_{ak}$  describe the within-agent variation of the set of molecular descriptors. The absolute value of the loadings,  $p_{ak}$ , tells how much a certain molecule descriptor,  $k = 1-16$ , contributes to the  $a$ th principal component, whereas the signs indicate whether a particular descriptor is negatively or positively correlated with the principal

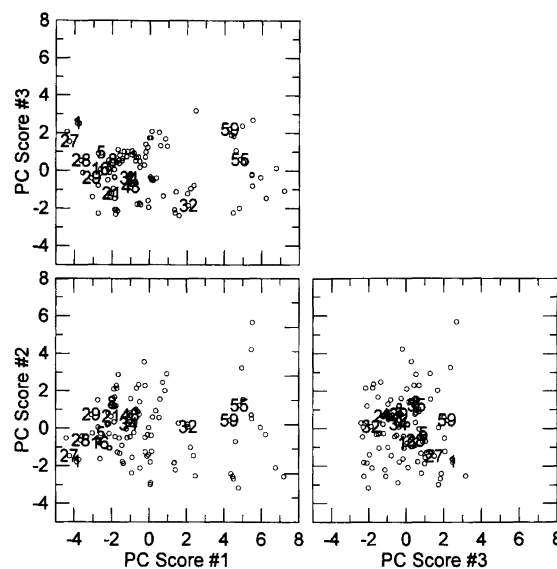


Fig. 1. Two-way score plots. Principal component (PC) score #1 versus PC #2, PC #1 versus PC #3, and PC #2 versus PC #3. The following reagents were selected for laboratory experimentation: **1, 5, 8, 16, 21, 27, 28, 29, 32, 34, 43, 55** and **59**. These reagents are indicated by their object numbers in the plots.

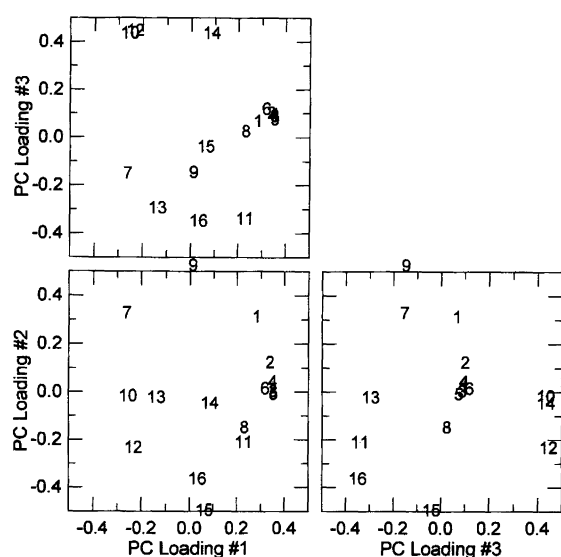


Fig. 2. Two-way loading plots. Principal component (PC) loading # 1 versus PC # 2, PC # 1 versus PC # 3, and PC # 2 versus PC # 3.

component. The loadings  $p_{ak}$  are listed in Table 2. Fig. 2 shows three two-way plots of the loadings for pc # 1 versus pc # 2 versus pc # 3, and pc # 2 versus pc # 3. From these plots, it is readily seen how the molecular descriptors are correlated.

Hence, molecular descriptors which depend on the same intrinsic properties will be described by the same principal component, while molecular descriptors which

depend on different underlying intrinsic properties will be described by *different* principal components. As principal components are mutually orthogonal, such an analysis will reveal which different underlying intrinsic properties may influence the experimental results. Detailed descriptions of the principal component analysis method are given in Refs. 22 and 23.

*Multivariate calibration using partial least squares regression*<sup>27,28</sup> (PLSR). In order to validate the relevance of the principal properties described above, a synthetic procedure using the bis-alkylating C<sub>3</sub> reagents was studied. On the basis of the two-way score plots in Fig. 1, the reagents in Table 3, listed with their corresponding measured HPLC results, were tested in the synthetic procedure (Scheme 3) for 3,3',5',5'-tetrakis(2,3-dihydroxypropyl-carbamoyl)-2,2',4,4',6,6'-hexaiodo-*N,N'*-(2-hydroxypropane-1,3-diyl)diacetanilide **108** from 5-acetamido-*N,N'*-bis(2,3-dihydroxypropyl)-2,4,6-triiodoisophthalamide **107** and one of the bis-alkylating reagents of the type A, E, G, K, M, Q or S (Schemes 1 and 2). The HPLC results are measurements of the reaction mixtures at reaction time 1 h and 5 h, at reaction temperatures of 20 °C and 40 °C for each of the reagents. The selection of the bis-alkylating reagents spans the variation in the different reagent structures described in Schemes 1 and 2, and the variation in the intrinsic properties among the *principal properties*. This selection gives a uniform spread of properties of the bis-alkylating reagents for the experimental work.

Table 3 Experimental conditions and HPLC results.

Exp. #	Reagent	Experimental variables <sup>a</sup>		Responses <sup>b</sup>				
		$\pi_1$	$\pi_2$	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\gamma_4$	$\gamma_r$
1	<b>1</b>	1	20	0.30	97.40	0.00	0.30	2.00
2	1,3-Difluoropropan-2-ol	5	20	2.80	81.30	1.20	0.40	14.30
3		1	40	9.50	67.00	3.00	0.80	19.70
4		5	40	39.00	36.60	8.10	4.50	11.80
5	<b>5</b>	1	20	9.70	73.20	16.30	0.30	0.50
6	1,3-Dichloropropan-2-ol	5	20	44.20	36.00	17.40	1.90	0.50
7		1	40	43.40	36.29	17.67	1.59	1.05
8		5	40	61.46	18.91	14.89	3.92	0.82
9	<b>8</b>	1	20	9.30	74.00	15.80	0.40	0.50
10	1,3-Dibromopropan-2-ol	5	20	42.80	37.50	17.50	1.70	0.50
11		1	40	44.20	34.90	19.10	1.60	0.20
12		5	40	60.50	19.10	16.80	3.30	0.30
13	<b>16</b>	1	20	4.20	89.30	5.80	0.30	0.40
14	2,3-Dichloropropan-1-ol	5	20	36.00	51.70	9.70	1.80	0.80
15		1	40	40.40	45.10	9.10	2.40	3.00
16		5	40	56.40	28.00	9.50	4.70	1.40
17	<b>21</b>	1	20	0.40	97.10	2.00	0.20	0.30
18	2,3-Dibromopropan-1-ol	5	20	1.00	92.80	5.40	0.30	0.50

Table 3 Continued

Exp. #	Reagent	Experimental variables <sup>a</sup>		Responses <sup>b</sup>				
		$\pi_1$	$\pi_2$	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\gamma_4$	$\gamma_r$
19		1	40	2.33	90.60	6.29	0.39	0.39
20		5	40	8.47	79.61	9.75	0.46	1.71
21	<b>27</b>	1	20	1.70	73.00	1.40	0.50	23.40
22	2-Fluoromethyloxirane	5	20	20.70	45.50	5.00	6.30	22.50
23		1	40	29.80	42.60	7.90	7.50	12.20
24		5	40	41.90	31.00	11.10	13.80	2.20
25	<b>28</b>	1	20	17.70	60.60	17.90	1.50	2.30
26	2-Chloromethyloxirane	5	20	45.90	29.00	10.70	12.10	2.30
27		1	40	46.47	28.73	10.86	11.54	2.40
28		5	40	47.39	26.44	10.53	13.37	2.27
29	<b>29</b>	1	20	20.00	58.90	17.30	1.90	1.90
30	2-Bromomethyloxirane	5	20	46.90	27.90	10.10	12.60	2.50
31		1	40	45.30	30.80	12.20	10.70	1.00
32		5	40	47.70	26.40	10.80	13.60	1.50
33	<b>32</b>	1	20	0.90	97.90	0.70	0.30	0.20
34	4-Trifluoromethylbenzene-	5	20	5.90	92.40	0.90	0.30	0.50
35	sulfonic acid oxiran-2-ylmethyl	1	40	17.87	76.41	3.46	1.15	1.11
36	ester	5	40	44.47	35.74	8.18	9.37	2.24
37	<b>34</b>	1	20	0.60	96.70	2.00	0.40	0.30
38	4-Chloromethyl-	5	20	0.60	93.60	5.20	0.40	0.20
39	1,3,2-dioxathiolane	1	40	0.00	58.70	38.60	0.00	2.70
40	2-oxide	5	40	0.00	40.50	54.50	0.00	5.00
41	<b>43</b>	1	20	0.40	97.10	2.00	0.20	0.30
42	2,3-Dibromopropan-1-yl	5	20	1.00	92.80	5.40	0.30	0.50
43	carbamate	1	40	2.33	90.06	6.29	0.39	0.93
44		5	40	8.47	79.61	9.75	0.46	1.71
45	<b>55</b>	1	20	—	—	—	—	—
46	Phosphoric acid	5	20	—	—	—	—	—
47	bis(2,3-dibromopropyl) ester	1	40	0.80	97.70	0.73	0.30	0.47
48	(2,3-dichloropropyl) ester	5	40	1.14	97.10	0.70	0.40	0.66
49	<b>59</b>	1	20	—	—	—	—	—
50	Phosphoric acid	5	20	—	—	—	—	—
51	tris(2-chloro-1-chloromethylethyl)	1	40	8.50	89.20	1.60	0.40	0.30
52	ester	5	40	15.00	80.60	2.20	0.90	1.30

<sup>a</sup> Experimental variables:  $\pi_1$ , reaction time [h];  $\pi_2$ , reaction temperature [ $^{\circ}$ C]. <sup>b</sup> Responses:  $\gamma_1$ , HPLC area% of **22**;  $\gamma_2$ , HPLC area% of **21**;  $\gamma_3$ , HPLC area% of **23**;  $\gamma_4$ , HPLC sum area% of *O*- and *N*-over-alkylated products of **22**;  $\gamma_r = 100 - \sum_{i=1}^4 \gamma_i$ .

The validation of the *principal properties* was based upon the hypothesis that the reaction **107**  $\rightarrow$  **108** proceeded according to the same or a similar reaction mechanism for each of the reagents.

The relationship between the yield of **108**, Scheme 3, the principal properties  $t_1$ – $t_6$  for the explored bis-alkylating reagents and the experimental conditions, reaction time and temperature, was modelled by the PLSR method. The PLSR method relates a matrix *X*, here the experimental conditions and principal properties, to a

vector *y*, here the yield of **108** for each of the explored reagents. This PLSR model can predict the yield for reagents not explored in laboratory experiments, although only for the studied reaction **107**  $\rightarrow$  **108**. However, the PCA model can be applied for any bis-alkylating  $C_3$  reaction. A good correlation between obtained versus predicted yield indicates that the descriptor variables contain relevant information to describe the response *y*, the yield of **108**.

The PLSR modelling is accomplished through a set of

partial least-squares (PLS) components. A PLS-component represents systematic regression found in the data. Each reagent is represented in the PLS components by its scalar values called *PLS scores*. Each variable in the data matrices is represented by *PLS loadings*, similar to PC-loadings. The PLSR analysis also includes determination of some statistical parameters and the optimal number of PLS components to be used in the model. All these parameters are useful for investigating the model, interpretation of the data and for detecting outliers in the data set.

Finally, these parameters can be rearranged to obtain the predictive model in eqn. (2)

$$\eta = \beta_0 + \alpha_1\pi_1 + \alpha_2\pi_2 + \beta_1t_1 + \beta_2t_2 + \dots + \beta_6t_6 + \beta_{12}t_1t_2 + \beta_{13}t_1t_3 + \dots + \beta_{46}t_4t_6 + \beta_{56}t_5t_6 + \beta_{11}t_1^2 + \beta_{22}t_2^2 + \dots + \beta_{22}t_6^2 \quad (2)$$

where  $\beta_0$  is the mean response value,  $\alpha_1$  and  $\alpha_2$  are the regression coefficients for the experimental conditions and  $\beta_1$ – $\beta_6$  are the regression coefficients for the principal properties (from Table 1). The coefficients  $\beta_{12}, \beta_{13}, \dots, \beta_{56}$  and  $\beta_{11}, \beta_{22}, \dots, \beta_{66}$  are the two-factor interaction terms between the principal properties and the quadratic terms, respectively. These two-factor interaction terms and quadratic terms are included in order to explain non-linear effects.

However, in this case, to obtain a satisfactory adaptation of the descriptors  $\pi_1$  and  $\pi_2$ , (the reaction time and temperature) and  $t_1$ – $t_6$  (the principal properties) to the measured response  $y$ , a logarithmic transformation, eqn. (3) of the response  $y$  was necessary. Hence, the PLSR modelling is done using the log transformed  $y$ , and the predictions of the yield of the **108** are carried out by eqns. (2) and (4).

$$\eta = \log\left(\frac{y+1}{100-y}\right) \quad (3)$$

$$y = \frac{10^{(\eta+2)} - 1}{10^\eta + 1} \quad (4)$$

The method and mathematics for multivariate calibration, prediction and model validation using the PLSR and cross-validation methods are well described in Refs. 26 and 27.

In the introductory PLSR analysis, entries #28, #37, #40, #45, and #46 of Table 3 were detected as outliers. These were removed from the data set, and a final PLSR model with six PLS components which accounted for more than 90% of the variance in  $\eta$  [ $=f(y)$ ] was determined to be significant according to cross-validation with the maximum number, i.e., 43, of cross-validation segments.<sup>26</sup> Fig. 3(a) shows the measured yield of **108** versus the predicted yield of **108**, where the predicted value is retrieved from the *cross-validation prediction*. Hence, this plot may be considered as a plot of a test set.

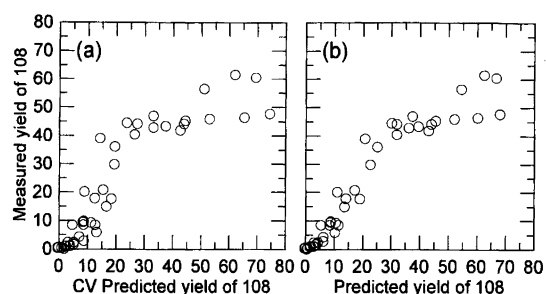


Fig. 3. Plot of observed versus predicted yield of **108**. The determined values are from HPLC analysis (area%) of the reaction mixture when different alkylating reagents were used with different reaction temperatures and reaction times. (a) Predictions from the cross-validation calculations. RMSEP=9.30 and  $R^2=0.95$ . (b) Predictions using the final model RMSEP=6.73 and  $R^2=0.97$ .

The statistics show an RMSEP of 9.30 and an  $R^2$  of 0.95. Fig. 3(b) shows the measured yield of **108** versus the predicted yield of **108** using the final parameters in eqns. (2) and (4). The statistics show an RMSEP of 6.73 and an  $R^2$  of 0.97. This shows the fairly good predictive ability of the PLSR model.

## Conclusion and discussion

The predictive power of the PLSR model is acceptable and fulfils the requirements for a model to be used in exploratory organic synthesis. The final synthesis must in the most cases be tuned by studying the experimental variables in a statistically designed experimental plan. The predictive ability of the model can be seen in the plot of observed versus predicted yield of **108**, Fig. 3. However, some deviation is seen for some of the reagents. This indicates that these reagents are in fact, or may, not be members of the general class of compounds as presupposed. Although, other conditions such as, e.g., the selected solvent, in this case water, may have a detrimental effect on the present reaction with these particular reagents. Such conditions can lead to other reaction mechanisms yielding undesired product(s).

If the selection of the test reagents were based on a larger variation among the whole set with respect to the PC score #2, an increased predictive ability of the PLSR model may be achieved. Although, on the basis of the obtained results, it can be concluded that the selected molecular descriptors  $x_1$ – $x_{16}$ , and thus the calculated principal properties from the PCA modelling, contain relevant information to be used in the design and modelling of bis-alkylating reactions.

An experimental plan based on the principal properties will allow selection of bis-alkylating reagents in such a way that effective variation in molecular properties is achieved. This will provide the possibility for increased freedom and efficiency in planning and performing synthetic procedure design, as well in the laboratory experi-



mentation and may lead to improved understanding of the chemistry of the reaction.

*Proposals of experimental designs.* Based upon the estimated principal properties the following experimental designs are proposed when new synthetic procedures of reactions with bis-alkylating  $C_3$  reagents are explored. Design #1 'uniformly spread' in PC score #1 versus PC score #2: **25, 53, 56, 69, 76, 77, 96, 100, 104**. Design #2 'uniformly spread' in PC score #1 versus PC score #2: **10, 28, 29, 37, 66, 82, 86, 89**. Design #3 'cube design and a centre' in PC score #1 versus PC score #2: **25, 41, 77, 81, 85**. Design #4 'star design' in PC score #1 versus PC score #2: **35, 49, 59, 69, 71, 76, 84, 87, 100**.

If for any reason one or more of the reagents in the designs above cannot be used in a certain reaction, this (these) can be substituted by another reagent with similar principal properties. For this purpose a plot of PC score #1 versus PC score #2 (values given in Table 1) should be used.

Carlson *et al.*<sup>16,17,29-32</sup> have previously shown that principal properties to be used in synthetic screening can be derived from experimentally measured physical and chemical properties. In the present work, it has been shown that molecular descriptors calculated through semiempirical quantum chemistry methods can be used to derive principal properties and successfully predict yields for a reaction. The results in Table 1 may, hopefully, be applied to the development of any reaction with a bis-alkylating  $C_3$  reagent step. This is particularly useful for improving the yield and/or the selectivity in reactions where the substrate is an ambident or a poly-nucleophile.

## Experimental

*Chemicals.* The substrate 5-acetamido-*N,N'*-bis(2,3-dihydroxypropyl)-2,4,6-triiodoisophthalamide **107** was obtained from Nycomed. Sodium hydroxide was from EKA Nobel and hydrochloric acid from Merck. The bis-alkylating reagents were supplied by Janssen Chimica (2-chloromethyloxirane, 2-bromomethyloxirane, and 2,3-dibromopropan-1-ol), Aldrich (1,3-difluoropropan-2-ol), Fluka Chemika (1,3-dichloropropan-2-ol), TCI-Tokyo Kasei [1,3-dibromopropan-2-ol, 2,3-dichloropropan-1-ol, phosphoric acid tris(2-chloro-1-chloromethylethyl) ester, phosphoric acid bis(2,3-dibromopropyl) ester (2,3-dichloropropyl) ester, and 2,3-dibromopropyl carbamate], and Sigma (2-fluoromethyloxirane, 4-trifluoromethylbenzenesulfonic acid oxiran-2-ylmethyl ester). 4-Chloromethyl-1,3,2-dioxathiolane 2-oxide was prepared in house.<sup>15</sup>

*Procedure for the bis-alkylating reaction.* 5-Acetamido-*N,N'*-bis(2,3-dihydroxypropyl)-2,4,6-triiodoisophthalamide **107** (100 g, 0.134 mol) was dissolved in a solution of sodium hydroxide (6.43 g, 0.161 mol, 1.2 equiv.) in purified water (200 ml). When the substrate **107** was completely dissolved, the bis-alkylating reagent (0.067 mol,

0.5 equiv.), see Table 3, was added. Samples from the reaction mixture were withdrawn after 1 h and 5 h and quenched by addition of concentrated hydrochloric acid. The composition of the mixture was determined by HPLC, and the results are given in Table 3.

*HPLC analysis.* Two eluents, A and B, were used. Eluent A was pure acetonitrile (Merck) and eluent B was purified water. The eluents were degassed with helium prior to use.

The solvent delivery system was gradient pumping 3–10% of eluent A, for 30 min (linear gradient profile), then 10–50% of eluent A for 40 min (concave gradient profile), with a flow rate of 1.0 ml min<sup>-1</sup>. The columns used were RP-18, Brownlee Columns (L 250 mm, ID 4.6 mm) packed with octadecylsilyl silica gel with particle size 5  $\mu$ m. A UV detector operating at wavelength  $\lambda = 254$  nm with a cell compartment having an optical path length of 10 mm was used.

*Calculations of the molecular descriptors.* The energy optimisation of the structures was achieved by use of the semiempirical method AM1 (Austin Model 1)<sup>21,33-35</sup> implemented in the MOPAC software<sup>36</sup> version 5.0 and version 6.0 or AMSOL version 3.01 (which is built upon the AMPAC version 2.1). The calculations were done in two steps, with the following keywords: (1) AM1 BONDS PRECISE GNORM = 0.2 and (2) AM1 BONDS PRECISE GNORM = 0.2 NOINTER NOXYZ THERMO(298,298) ROT = 1.

The AM1 routine was used to calculate molecular weight, enthalpy at 298 K, entropy at 298 K, heat capacity at 298 K, zero point energy, heat of formation, dipole moment, energy of highest occupied molecular orbital, energy of lowest unoccupied molecular orbital, atomic point charge on O in the  $C_3$  moiety, atomic point charge on C2 in the  $C_3$  moiety, the lowest atomic point charge on the terminal C in the  $C_3$  moiety and the highest atomic point charge on the terminal C in the  $C_3$  moiety. The AM1 calculations were done on a VAX computer under VMS.

The absolute electronegativity and the hardness were calculated based upon the calculated energy of highest occupied molecular orbital ( $\epsilon_{\text{HOMO}}$ ) and the energy of lowest unoccupied molecular orbital ( $\epsilon_{\text{LUMO}}$ ) (from the AM1 calculations) according to Pearson.<sup>37</sup>

The final energy-optimised structures from the AM1 computations were submitted for calculations of the van der Waals surfaces [ $\text{\AA}^2$ ] using INSIGHTII software under IRIX 4.05 on a Silicon Graphics 4D/35 computer.

*List of the 106  $C_3$  bis-alkylating moieties investigated using the semiempirical method AM1.* The compound number given here is also the entry number in Table 1, which gives the results of the multivariate computations. **1**, 1,3-Difluoropropan-2-ol; **2**, 1-chloro-3-fluoropropan-2-ol; **3**, 1-bromo-3-fluoropropan-2-ol; **4**, 1-fluoro-3-iodopropan-2-ol; **5**, 1,3-dichloropropan-2-ol; **6**, 1-bromo-3-chloro-

propan-2-ol; **7**, 1-chloro-3-iodopropan-2-ol; **8**, 1,3-dibromopropan-2-ol; **9**, 1-bromo-3-iodopropan-2-ol; **10**, 1,3-diiodopropan-2-ol; **11**, 2,3-difluoropropan-1-ol; **12**, 2-chloro-3-fluoropropan-1-ol; **13**, 2-bromo-3-fluoropropan-1-ol; **14**, 3-fluoro-2-iodopropan-1-ol; **15**, 3-chloro-2-fluoropropan-1-ol; **16**, 2,3-dichloropropan-1-ol; **17**, 2-bromo-3-chloropropan-1-ol; **18**, 3-chloro-2-iodopropan-1-ol; **19**, 3-bromo-2-fluoropropan-1-ol; **20**, 3-bromo-2-chloropropan-1-ol; **21**, 2,3-dibromopropan-1-ol; **22**, 3-bromo-2-iodopropan-1-ol; **23**, 2-fluoro-3-iodopropan-1-ol; **24**, 2-chloro-3-iodopropan-1-ol; **25**, 2-bromo-3-iodopropan-1-ol; **26**, 2,3-diiodopropan-1-ol; **27**, 2-fluoromethyloxirane; **28**, 2-chloromethyloxirane; **29**, 2-bromomethyloxirane; **30**, 2-iodomethyloxirane; **31**, trifluoromethanesulfonic acid oxiran-2-ylmethyl ester; **32**, 4-trifluoromethylbenzenesulfonic acid oxiran-2-ylmethyl ester; **33**, 4-fluoromethyl-1,3,2-dioxathiolane 2-oxide; **34**, 4-chloromethyl-1,3,2-dioxathiolane 2-oxide; **35**, 4-bromomethyl-1,3,2-dioxathiolane 2-oxide; **36**, 4-iodomethyl-1,3,2-dioxathiolane 2-oxide; **37**, 4-fluoromethyl-2-oxo-[1,3,2λ<sup>5</sup>]dioxaphospholane; **38**, 4-chloromethyl-2-oxo-[1,3,2λ<sup>5</sup>]dioxaphospholane; **39**, 4-bromomethyl-2-oxo-[1,3,2λ<sup>5</sup>]dioxaphospholane; **40**, 4-iodomethyl-2-oxo-[1,3,2λ<sup>5</sup>]dioxaphospholane; **41**, 2,3-difluoropropyl carbamate; **42**, 2,3-dichloropropyl carbamate; **43**, 2,3-dibromopropyl carbamate; **44**, 2,3-diiodopropyl carbamate; **45**, 1,3-difluoro-2-propyl carbamate; **46**, 1,3-dichloro-2-propyl carbamate; **47**, 1,3-dibromo-2-propyl carbamate; **48**, 1,3-diiodo-2-propyl carbamate; **49**, phosphoric acid 2-fluoro-1-fluoromethylethyl ester; **50**, phosphoric acid 2-chloro-1-chloromethylethyl ester; **51**, phosphoric acid 2-bromo-1-bromomethylethyl ester; **52**, phosphoric acid 2-iodo-1-iodomethylethyl ester; **53**, phosphoric acid tris(2,3-difluoropropyl) ester; **54**, phosphoric acid tris(2,3-dichloropropyl) ester; **55**, phosphoric acid bis(2,3-dibromopropyl) ester (2,3-dichloropropyl) ester; **56**, phosphoric acid tris(2,3-dibromopropyl) ester; **57**, phosphoric acid tris(2,3-diiodopropyl) ester; **58**, phosphoric acid tris(2-fluoro-1-fluoromethylethyl) ester; **59**, phosphoric acid tris(2-chloro-1-chloromethylethyl) ester; **60**, phosphoric acid tris(2-bromo-1-bromomethylethyl) ester; **61**, phosphoric acid tris(2-iodo-1-iodomethylethyl) ester; **62**, acetic acid 2-fluoro-1-fluoromethylethyl ester; **63**, acetic acid 2-chloro-1-chloromethylethyl ester; **64**, acetic acid 2-bromo-1-bromomethylethyl ester; **65**, acetic acid 2-iodo-1-iodomethylethyl ester; **66**, acetic acid 2-bromo-1-chloromethylethyl ester; **67**, acetic acid 1-chloromethyl-2-iodoethyl ester; **68**, acetic acid 2,3-difluoropropyl ester; **69**, acetic acid 2,3-dichloropropyl ester; **70**, acetic acid 2,3-dibromopropyl ester; **71**, acetic acid 2,3-diiodopropyl ester; **72**, acetic acid 3-bromo-2-chloropropyl ester; **73**, 3-fluorooxetane; **74**, 3-chlorooxetane; **75**, 3-bromooxetane; **76**, 3-iodooxetane; **77**, methanesulfonic acid oxetan-3-yl ester; **78**, toluene-4-sulfonic acid oxetan-3-yl ester; **79**, methanesulfonic acid 3-hydroxy-2-methanesulfonyloxypropyl ester; **80**, toluene-4-sulfonic acid 3-hydroxy-2-(toluene-4-sulfonyloxy)propyl ester; **81**, methanesulfonic acid 2-hydroxy-3-methanesulfonyloxypropyl ester;

**82**, toluene-4-sulfonic acid 2-hydroxy-3-(toluene-4-sulfonyloxy)propyl ester; **83**, 2,2-dimethylpropionic acid 2-fluoro-1-fluoromethylethyl ester; **84**, 2,2-dimethylpropionic acid 2-chloro-1-chloromethylethyl ester; **85**, 2,2-dimethylpropionic acid 2-bromo-1-bromomethylethyl ester; **86**, 2,2-dimethylpropionic acid 2-iodo-1-iodomethylethyl ester; **87**, fluorosulfonic acid 2-hydroxy-3-(fluorosulfonyloxy)propyl ester; **88**, 1,1,2,2,3,3,4,4,4-nonafluorobutane-1-sulfonic acid 2-hydroxy-3-(1,1,2,2,3,3,4,4,4-nonafluorobutane-1-sulfonyloxy)propyl ester; **89**, 2,2,2-trifluoroethanesulfonic acid 2-hydroxy-3-(2,2,2-trifluoroethanesulfonyloxy)propyl ester; **90**, trifluoromethanesulfonic acid 2-hydroxy-3-(trifluoromethanesulfonyloxy)propyl ester; **91**, 4-nitrobenzenesulfonic acid 2-hydroxy-3-(4-nitrobenzenesulfonyloxy)propyl ester; **92**, 4-bromobenzenesulfonic acid 2-hydroxy-3-(4-bromobenzenesulfonyloxy)propyl ester; **93**, 2,2-dimethylpropionic acid 2-fluoro-1-fluoromethylethyl ester; **94**, 2,2-dimethylpropionic acid 2-chloro-1-chloromethylethyl ester; **95**, 2,2-dimethylpropionic acid 2-bromo-1-bromomethylethyl ester; **96**, 2,2-dimethylpropionic acid 2-iodo-1-iodomethylethyl ester; **97**, 2,2-dimethylpropionic acid 2,3-bis(trifluoromethanesulfonyloxy)propyl ester; **98**, trifluoromethanesulfonic acid oxiran-2-ylmethyl ester; **99**, fluorosulfonic acid oxiran-2-ylmethyl ester; **100**, 4-bromobenzenesulfonic acid oxiran-2-ylmethyl ester; **101**, 1,1,2,2,3,3,4,4,4-nonafluorobutane-1-sulfonic acid oxiran-2-ylmethyl ester; **102**, 2,2,2-trifluoroethanesulfonic acid oxiran-2-ylmethyl ester; **103**, 4-fluoromethyl-1,3,2-dioxathiolane 2,2-dioxide; **104**, 4-chloromethyl-1,3,2-dioxathiolane 2,2-dioxide; **105**, 4-bromomethyl-1,3,2-dioxathiolane 2,2-dioxide; **106**, 4-iodomethyl-1,3,2-dioxathiolane 2,2-dioxide.

*Multivariate computation.* Prior to computing the principal component models, both for the PCA and the PLSR methods, each descriptor variable was scaled to unit variance. This was done to avoid the possibility that different units of measurement of the descriptors distort the variance structure and thereby bias the projections. The Unscrambler<sup>®</sup> version 5.03 software<sup>38</sup> under DOS 5.0 on a COMPAQ 486/50DX microcomputer, was used for the multivariate data analysis on the scaled data. To avoid overfitting in deriving the principal component models, the cross-validation method<sup>26</sup> with maximum number of segments, i.e., equal to number of rows in the matrix, was used for determining the significant number of principal components.

*Acknowledgements.* The authors would like to express their gratitude to the technical staff at Nycomed Imaging, Departments of Chemical Analysis, Development Chemistry and Process Chemistry for excellent technical assistance. Special gratitude is directed to Dr. Rune Fosheim at the Department of Chemical Synthesis for helping us with the calculations of the van der Waals surfaces. Nycomed Imaging AS is gratefully acknowledged for permission to publish the present work.

## References

1. Sangwan, N. K., Verma, B. S. and Dhindsa, K. S. *Indian J. Chem., Sect. B* 31 (1992) 590.
2. Bhalla, M., Naithani, P. K., Kumar, A., Bhalla, T. N. and Shanker, K. *Indian J. Chem., Sect. B* 31 (1992) 183.
3. Kettmann, V., Csoellei, J., Racanska, E. and Svec, P. *Eur. J. Med. Chem.* 26 (1991) 843.
4. Nadolol, C07C 217/38, WO 91/04244, PCT/CA90/00316.
5. Sinha, A. K., Rastogi, S. N. and Das, S. R. *Indian J. Chem., Sect. B* 30 (1991) 1041.
6. Eguchi, Y., Sasaki, F., Takashima, Y., Nakajima, M. and Ishikawa, M. *Chem. Pharm. Bull.* 39 (1991) 795.
7. Dwivedi, A. K., Shukla, V. K., Bhandari, K., Setty, B. S., Hamboj, V. P. and Khanna, N. M. *Indian J. Chem., Sect. B* 30 (1991) 281.
8. Stibor, I., Holy, P., Zavada, J., Koudelka, J., Novak, J., Zajicek, J. and Belohradsky, M. *J. Chem. Soc., Chem. Commun.* (1990) 1581.
9. Nyegaard & Co., EP 0 108 638 A1, Int. Cl. C 07 C 103/78, A 61 K 49/04.
10. Liu, Y. and Chen, H. *Chem. Abstr.* 110 (23436), p N/A, (1987).
11. Warren, S. *Organic Synthesis – The Disconnection Approach*, Wiley, New York 1992, p. 357.
12. March, J. *Advanced Organic Chemistry, 4th ed.*, Wiley, New York 1992, pp. 352–357.
13. Weissermel, K. and Arpe, H. J. *Industrielle Organische Chemie, 2. Aufl.*, Verlag Chemie Weinheim 1978, p. 277.
14. Lohray, B. B. *Synthesis* (1992) 1035.
15. De la Mare, P. B. D., Klyne, W., Millen, D. J., Pritchard, J. G. and Watson, D. *J. Chem. Soc.* (1956) 1813.
16. Hellberg, S., Sjöström, M. and Wold, S. *Acta Chem. Scand. B* 40 (1986) 135.
17. Hellberg, S., Sjöström, M., Skagerberg, B. and Wold, S. *J. Med. Chem.* 30 (1987) 1126.
18. Box, G. E. P., Hunter, W. G. and Hunter, J. S. *Statistics for Experimenters, An Introduction to Design, Data Analysis, and Model Building*, Wiley, New York 1978, pp. 306–351, pp. 374–418.
19. Box, G. E. P. and Hunter, J. S. *Ann. Math. Statist.* 28 (1957) 195.
20. Box, G. E. P. and Draper, N. R. *Empirical Model-Building and Response Surfaces*, Wiley, New York 1987, pp. 305–312, 440–441.
21. Clark, T. *A Handbook of Computational Chemistry, A Practical Guide to Chemical Structure and Energy Calculations*, Wiley, New York 1985.
22. Jolliffe, I. T. *Principal Component Analysis*, Springer-Verlag, New York 1986, pp. 1–114.
23. Jackson, J. E. *A User's Guide to Principal Components*, Wiley, New York 1991, pp. 1–62.
24. Carlson, R. *Design and Optimization in Organic Synthesis*, Vol. 8 in *Data Handling in Science and Technology*, Elsevier, Amsterdam, 1992, pp. 337–387.
25. Bjørsvik, H.-R. and Priebe, H. *Semi-empirical Quantum Chemistry Calculations of Bis-alkylating C<sub>3</sub> Moieties*, Deposited material at The Royal Danish Library, The Danish Office, P.O. Box 2149, DK Copenhagen, Denmark.
26. Wold, S. *Technometrics* 20 (1978) 397.
27. Martens, H. and Næs, T. *Multivariate Calibration*, Wiley, New York 1991, pp. 116–165.
28. Jackson, J. E. *A User's Guide to Principal Components*, Wiley, New York 1991, pp. 282–290.
29. Carlson, R., Lundsted, T. and Albano, C. *Acta Chem. Scand., Ser. B* 39 (1985) 79.
30. Carlson, R., Prochazka, M. P. and Lundsted, T. *Acta Chem. Scand., Ser. B* 42 (1988) 145.
31. Carlson, R., Prochazka, M. P. and Lundsted, T. *Acta Chem. Scand., Ser. B* 42 (1988) 157.
32. Carlson, R., Lundsted, T., Nordahl, Å. and Prochazka, M. P. *Acta Chem. Scand., Ser. B* 40 (1986) 522.
33. Dewar, M. J. S., Zoebisch, E. G., Healy, E. F. and Stewart, J. J. P. *J. Am. Chem. Soc.* 107 (1985) 3902.
34. Dewar, M. J. S. and Dieter, K. M. *J. Am. Chem. Soc.* 108 (1986) 8075.
35. Stewart, J. J. P. *J. Comput.-Aided Mol. Design* 4 (1990) 1.
36. Stewart, J. J. P. *MOPAC 93.00 Manual*, Fujitsu Limited, Tokyo, 1993.
37. Pearson, R. G. *J. Org. Chem.* 54 (1989) 1423.
38. The Unscrambler®, Extended Memory Version 5.03. User's Guide, October 1993. CAMO A/S, Olav Tryggvasons gt 24, N-7011 Trondheim, Norway.

Received December 5, 1994.