# Characterization of Hydrophobic Double-Ester Polymers by Molecular Modeling and Multivariate Data Analysis

## Rune Fossheim and Astri Rogstad*

Nycomed Imaging AS, Nycomed Amersham, PO Box 4220 Torshov, N-0401 Oslo, Norway

Fossheim, R. and Rogstad, A., 1999. Characterization of Hydrophobic Double-Ester Polymers by Molecular Modeling and Multivariate Data Analysis. – Acta Chem. Scand. 53: 95–99. © Acta Chemica Scandinavica 1999.

A set of 64 hydrophobic double-ester polymers and 4 benchmark polymers [polyethylene, poly(glycolide), poly($R$-lactide) and polyethylene terephthalate] having two polymer lengths ($N = 15$ and 60) have been characterized using results from molecular modeling in conjunction with multivariate statistical methods. Using a set of 23 extensive and intensive polymer descriptors, a principal component analysis (PCA) model with three principal components (PCs) which explain 72% of the variance in the data set was obtained. The clustering of the polymers in the PCA projections identified effects both of molecular weight (MW) and chemical constitution of the polymers. An MW-dependent (polydispersity) direction can be identified in PCA space, corresponding to the semi-parallel lines (PDs): $PC2 = A \times PC1 + B$, where $A < 0$ and $B$ is a constant that depends on the chemical constitution of the monomers. The distance between the high-MW and low-MW points on the PD lines was roughly proportional to $\Delta$MW. The PC1 scores for one of the double esters and the benchmark compounds correlated with glass transition temperatures ($T_g$) for each of the HMW and LMW series. It was also found that the slopes of the PD lines ($A$) are correlated with the average number MW ($M_n$) correction term to $T_g$ in the Fox and Flory equation ($-0.002\,715 \times T_{g\infty}^3/M_n$). These relationships suggest the usefulness of the PCA description in predicting $T_g$ values for double-ester polymers and for other semi-quantitative comparisons of the polymers.

Double-ester-based polymers, containing the methylene di-ester linkage –CO–O–CHR–O–CO–, comprise one niche in the large field of biodegradable polyesters which may eventually find use within medical and therapeutical applications.[1–3] The double-ester synthetic routes represent a possibility for making biodegradable polymers with a wide range of properties, including hydrophobic polymers and surfactant-type block copolymers.[4,5]
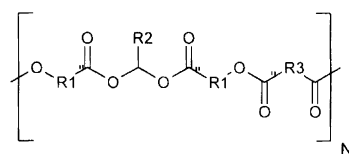
Since even limited systematic variations of structural fragments very soon result in a vast number of polymers that could potentially be made, a prior-to-synthesis characterization of a wider range of such polymers may be useful, since it can eventually direct the synthetic efforts. Combining molecular modeling methods for calculation of molecular descriptors and multivariate statistical methods for data analysis is one possibility for such prior-to-synthesis characterization and generally for identification of structure–property relationships. Such characterizations have been performed for a series of hydrophobic double-ester polymers and four benchmark polymers.

## Methods

*Compounds.* The double-ester polymers that have been included in the study were made as combinations of the four hydroxy acids (R1), four double-esters (R2) and four linkers (R3) shown in Table 1. This spanning of the double-esters gives in all 64 monomers. Two lengths of each polymer were considered for characterization, corresponding to the number of monomer units $N = 15$ (LMW) and $N = 60$ (HMW). The polymers were considered as hydroxyl and carboxylate (protonated) end-capped.

In addition to the double-ester polymers, four synthesized and well known polymers were included as benchmark compounds in the study. These were polyethylene (PE), poly(glycolide) (PGL), poly($R$-lactide) (PLA) and polyethylene terephthalate (PET). Since the monomers in these cases were much smaller than the block polymer double-esters, two descriptions of the LMW benchmark polymers were considered: one based on the chemical monomer unit and one based on a number of monomers giving monomer lengths comparable to the double-ester block monomers. The latter oligomers consisted of 32 (PE), 10 (PGL, PLA) and 2

* To whom correspondence should be addressed.

*Table 1.* Spanning and numbering of LMW and HMW double-esters included in PCA analysis.[a]



| R3 span ↓ | R1 span → −(CH₂)− | −1,4-phenyl− | −(CH₂)₉− | −(CH₂)₁₅− |
|---|---|---|---|---|
| −CO−R3−CO− = | H: **1, 5** | H: **33, 37** | H: **65, 69** | H: **97, 101** |
| | Me: **9, 13** | Me: **41, 45** | Me: **73, 77** | Me: **105, 109** |
| −CO−[b] | t-Bu: **17, 21** | t-Bu: **49, 53** | t-Bu: **81, 85** | t-Bu: **113, 117** |
| | Ph: **25, 29** | Ph: **57, 61** | Ph: **89, 93** | Ph: **121, 125** |
| −(CH₂)₂− | H: **2, 6** | H: **34, 38** | H: **66, 70** | H: **98, 102** |
| | Me: **10, 14** | Me: **42, 46** | Me: **74, 78** | Me: **106, 110** |
| | t-Bu: **18, 22** | t-Bu: **50, 54** | t-Bu: **82, 86** | t-Bu: **114, 118** |
| | Ph: **26, 30** | Ph: **58, 62** | Ph: **90, 94** | Ph: **122, 126** |
| −(CH₂)₄− | H: **3, 7** | H: **35, 39** | H: **67, 71** | H: **99, 103** |
| | Me: **11, 15** | Me: **43, 47** | Me: **75, 79** | Me: **107, 111** |
| | t-Bu: **19, 23** | t-Bu: **51, 55** | t-Bu: **83, 87** | t-Bu: **115, 119** |
| | Ph: **27, 31** | Ph: **59, 63** | Ph: **91, 95** | Ph: **123, 127** |
| −1,2-phenyl− | H: **4, 8** | H: **36, 40** | H: **68, 72** | H: **100, 104** |
| | Me: **12, 16** | Me: **44, 48** | Me: **76, 80** | Me: **108, 112** |
| | t-Bu: **20, 24** | t-Bu: **52, 56** | t-Bu: **84, 88** | t-Bu: **116, 120** |
| | Ph: **28, 32** | Ph: **60, 64** | Ph: **92, 96** | Ph: **124, 128** |

[a]Each line in the data cells gives LMW no. and HMW no., respectively, of the various R2s. PE: **129** (LMW monomer description), **130** (LMW oligomer description) and **131** (HMW), PGL: **132–134**, PLA: **135–137** and PET: **139–140**. [b]Carbonate-ester (−O(CO)O−) span.

(PET) monomers. PE was methyl end-capped and the three other polymers were hydroxyl and carboxyl end-capped as the double-ester polymers.

*Parameters.* The 23 polymer parameters used to characterize the compounds are given and described in Table 2. They comprise a list of 10 extensive (dependent of molecular weight) and 13 intensive (independent of molecular weight) parameters which were derived using the molecular modeling programs MOPAC 6.0[8] and Hyperchem 4.5.[9] The complete design matrix for the polymers (Table 3) is available as supplementary material.

*Multivariate data analysis.* A principal component analysis (PCA) of the date matrix was performed with the multivariate statistical program Unscrambler.[10] All variables were scaled to equal weight using $1/s_d$. The calculations were performed in several steps. The initial model was examined to exclude variables insignificant to the model. The refined model was validated by cross-validation on all samples. Only the first three principal components (PC) were examined, as those accounted for most of the variance in the experimental data set. Relationships between PC1 scores and glass transition temperatures available for five of the compounds were investigated.

## Results and discussion

Three PCs described 72% of the variance in the data set when all variables were included. The score and loading plots showed good structure in the data set and the variables and the compounds were well spread in the projections of the model and distributed in clusters. The PC1/PC2 and PC2/PC3 projections of the loadings plots are shown in Figs. 1 and 2, whereas the PC1/PC2 and PC2/PC3 projections of the scores plots are shown in Figs. 3 and 4. Some of the benchmark polymers were identified as outliers in the model. All variables except DIPOL, ELNEG, NOMO and PHEME contributed well to at least one of the first three PCs. These four variables were excluded and the model refined and cross-validated. Three PCs accounted for 82% of the variance in the refined model, and the clustering was more clear. As there is no experimental noise in the data set, three PCs appeared to capture the correlations in the data set and thus reveal the data structure. For this purpose the first three PCs were considered the most interesting.

From the loadings plots it can be seen that MW and MR are strongly coupled, which is not unexpected since MR is related to molar volume and all the polymers contain only O, C and H. LOGP, NROT and NMET are correlated, signifying that the most hydrophobic polymers have many (rotatable) methylene groups. The

*Table 2.* Extensive and intensive polymer parameters used in PCA description.

| Extensive parameters | | | Intensive parameters | | |
| --- | --- | --- | --- | --- | --- |
| No. | Name | Parameter | No. | Name | Parameter |
| 1 | NOMO | No. of monomers[a] | 5 | PERC | Percent carbon |
| 2 | MW | Molecular weight | 6 | PERO | Percent oxygen |
| 3 | MR | Molar refractivity[b] | 7 | PERH | Percent hydrogen |
| 4 | LOGP | Water/octanol partition coefficient[c] | 10 | FLEX | Ratio NROT/no. of bonds[f] |
| 8 | NROT | No. of rotatable bonds[d] | 11 | COIL | Bond population average for *gauche* and *syn* torsion angles[g] |
| 9 | NRIG | No. of rigid bonds[e] | 12 | MLEN | Monomer length[h] |
| 13 | NMET | No. of -CHR- units[i] | 16 | MEPO | Ratio NMET/NPOL |
| 14 | NPOL | No. of carbonyl units[j] | 17 | PHEME | Ratio NPHE/NMET |
| 15 | NPHE | No. of phenyl groups[k] | 19 | FRBU | Ratio SIBU/MR |
| 18 | SIBU | Side-chain steric bulk[l] | 20 | DIPOL | Dipole moment[m] |
| | | | 21 | ELNEG | Electronegativity[n] |
| | | | 22 | HARD | Hardness[n] |
| | | | 23 | DELN | Ratio ELNEG/HARD[o] |

[a]The number of monomers in the polymer; i.e. 15 or 60 for the double-ester polymers and the benchmark polymers based on the oligomer description. For the true monomer description of the latter compounds the number of monomers is 480 (PE), 150 (PGL and PLA) and 60 (PET). [b]The molar refractivity for each polymer was obtained by calculating the polarizability (in $Å^3$) for the monomers, $P(M)$ using the semi-empirical quantum mechanical method AM1, incorporated in the MOPAC 6.0 program package. [Keywords: AM1 GNORM=0.5 POLAR (XYZ GEO-OK)]. The molar refractivity for the polymers was calculated from eqn. (1);

$$MR = [P(M) - P(Me/OH)] \times N + P(OH/H) \qquad (1)$$

using the fact that molar refractivity is proportional to polarizability. $P(Me/OH)$ is the polarizability for the monomer end-capping groups (Me and OH) and $P(OH/H)$ is the polarizability for the polymer end-capping groups (OH and H). $N$ is the number of monomers. [c]The water–1-octanol partition coefficient was calculated from atomic hydrophobicity constants according to Ghose and Crippon.[6] The calculated log $P$ is extensive. However, the measured log $P$ may because of shielding, coiling and other effects in polymers be quite different from the calculated log $P$. [d]The number of rotatable bonds in the polymer was enumerated by summing the number of $C(sp^3)$–$C/O(sp^3)$ single bonds, the number of Ph–C/O bonds and the O–CX–O double-ester bonds excluding the t-Bu substituted double-ester. These are the bonds where rotation expectedly can occur at room temperature. [e]The number of rigid bonds in the polymers was enumerated by summing all configurationally locked bonds, the O–CH(t-Bu)–O double-ester bonds, the R(O)–O and the OR(O)–O ester bonds. NROT+NRIG equals the number of bonds in the polymer. [f]The polymer flexibility index is taken as the number of (rotatable bonds)/(number of bonds). This index is 1.0 for a completely flexible polymer like PE, and 0.0 for a conformationally and/or configurationally locked polymer. [g]This parameter describes the likelihood for coiled conformations. Model molecules for the different bond types in the polymer were made using the Hyperchem 4.5 program. The energies for the *anti*, *gauche* and eventual *syn* conformations were determined using MM+ molecular mechanics calculations. The *gauche*+*syn* populations $(X_i)$ for each bond type were determined using the Boltzman distribution, and the COIL parameter was calculated according to eqn. (2):

$$COIL = \sum X_i/NB \qquad (2)$$

where the sum is over all the bonds in the monomer and NB is the number of bonds in the monomer. [h]The monomer length (in Å) obtained from MM+ geometry optimized geometries of extended conformations. [i]The number of methylene and double-ester units in the polymer. [j]The number of carbonyl groups in the polymer. [k]The number of phenyl groups in the main chain and the side-chains. [l]The side-chain steric bulk is taken as the molar refractivity (MR) of the side-chains, in excess of that of a PE chain according to eqn. (3):

$$SIBU = MR = P(S-R) - P(R) \qquad (3)$$

where $P(S-R)$ is ethane, $n$-pentane and toluene and $P(R)$ is methane for calculation of the double-ester steric bulk. $P(S-R)$ is $p$-xylene and $o$-xylene and $P(S)$ is hexane and butane for polymers containing phenyl groups in the main chain. [m]Dipole moment (in debyes) for the AM1 geometry optimized monomers. [n]Electronegativity and hardness for the monomers calculated according to Pearson.[7] [o]Reactivity calculated according to Pearson.

MEPO, FLEX, MLEN and COIL parameters were weakly correlated, and weakly negatively correlated to the parameters NPHE, DELN, FRBU and SIBU. Thus compounds with a large methylene to carbonyl content are usually long and flexible and easy to coil, whereas the phenyl-rich compounds are more rigid and have more steric bulk in the side-chains. They are also generally more reactive (Pearson DELN). The Pearson hardness parameter (HARD) was negatively correlated to these parameters, corresponding to phenyl groups being 'soft' and reactive and methylene groups being hard and unreactive. Furthermore, PERC and PERO were negatively correlated as expected.

The leverage of the parameters was such that oxygen-rich and/or phenyl-containing compounds with bulk in the side-chains were placed on positive PC1, whereas hydrophobic, flexible and methylene containing polymers with high Pearson hardness were placed on negative
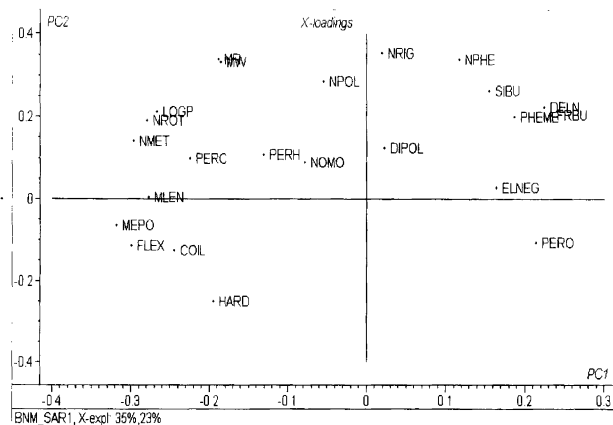
Fig. 1. Loadings plot of all variables in the principal regression model; projection PC1/PC2.
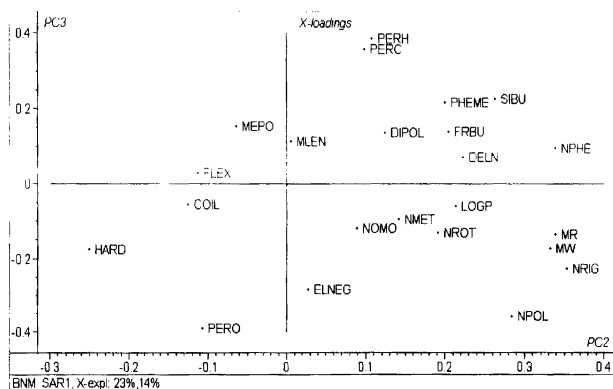


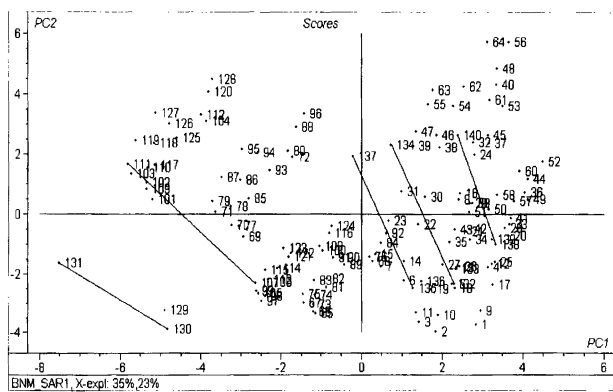Fig. 2. Loadings plot of all variables in the principal regression model; projection PC2/PC3



Fig. 3. Principal component classification of the complete data set in the PC1/PC2 projection. PD lines have been included for the benchmark compounds and 107, 111 (P73).

PC1. High-MW, phenyl-containing and rigid polymers with high Pearson reactivity (DELN) will be placed on positive PC2, whereas elemental composition is most important for placement along PC3.

Considering the scores plots in Figs. 3 and 4, the C15 and C9 hydroxy acid compounds are on negative PC1, whereas the glycolic acid and p-hydroxybenzoic acid compounds are mainly on positive PC1. The separation
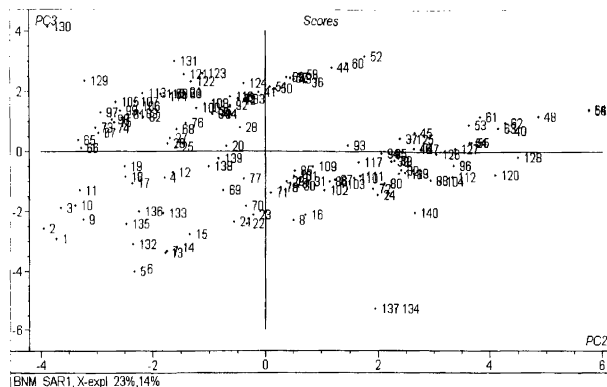


Fig. 4. Principal component classification of the complete data set in the PC1/PC2 projection.

of the latter two polymer classes is mainly along PC2, and it can be seen that for all the polymers, phenyl substitution raises the compounds on the PC2 axis in accordance with the loadings plots.

In addition to the distribution caused by different chemical constitution of the polymers discussed above, there was an evident classification caused by molecular weight seen in the PC1/PC2 projection. All LMW/HMW points for the polymers were close to pseudo-parallel lines (PDs) with negative slope: $PC2 = A \times PC1 + B$, where $A < 0$ and $B$ is a constant that depends on the chemical constitution of the monomers. The distance between the LMW and HMW points was proportional to $\Delta MW$, and the PD lines could thus be identified as an $M_n$ (average number MW) dependent direction in PC space. Accordingly, the PCA description incorporated a property description of the polymers that took into account both differences caused by chemical constitution and differences caused by differences in molecular weight. The separation of MW and chemical constitution effects fell naturally out of the PC analysis, which is based on a wide range of both extensive and intensive polymer parameters. Since the MW effect was expressed along the PD lines, it would be expected that physico-chemical differences would be manifested mostly in the direction normal to this. This was also in accordance with the loadings plots which showed that the majority of the intensive polymer descriptors were located close to the line $PC2 \approx PC1$ (COIL, FLEX, MEPO, DELN, FRBU).

Thus we can identify double esters that are similar to the benchmark compounds independently of MW effects by examining PD lines that are close to the benchmark lines. PE is found closest to the C15/C9 hydroxy acid compounds, although none of the double esters resembles PE to a large extent. This is reasonable, since PE is the only polymer lacking both carbonyl and phenyl groups. PGL/PLA and PET are found in the glycolic acid and p-hydroxybenzoic acid cluster. The polymer description based on the true monomer and oligomer descriptions for these compounds turn out very similar since the LMW points are almost overlapping. The polymers that are calculated similar to PLA are glycolic acid esters
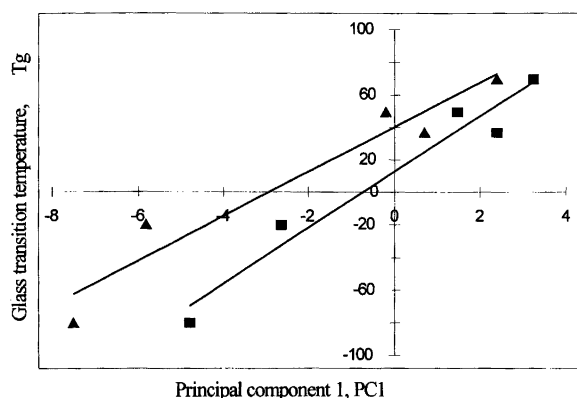
Fig. 5. $T_g$-values plotted against PC1 scores for the HMW series (filled triangles) and the LMW series (filled squares).
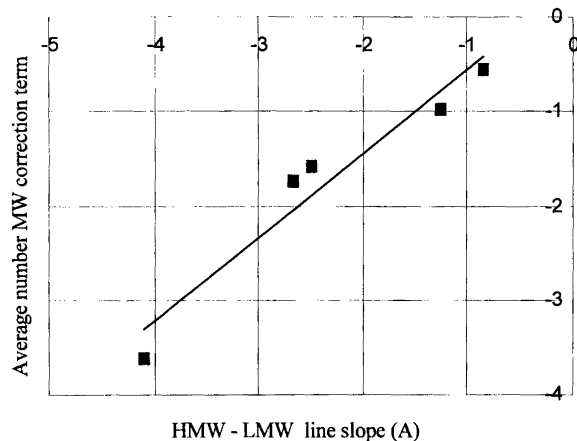


Fig. 6. Number average MW $(M_n)$ correction term $(K_g/M_n)$ in the equation of Fox and Flory plotted against the slopes of the PD lines.

The relationship between the PD lines and $T_g$-values may be investigated further by plotting the calculated slopes $(A)$ from the PC12 projections against the number-average MW $(M_n)$ correction term $(K_g/M_n)$ in the equation of Fox and Flory, eqn. (4):

$$T_g = T_{g\infty} - K_g/M_n = T_{g\infty} - 0.002\,715 \times T_{g\infty}^3/M_n \qquad (4)$$

The correction term was calculated from the experimental $T_g$-values and $M_n$, which was calculated with $n = (60 + 15)/2$. This correlation is shown in Fig. 6. For the limited series of compounds it is indicated that the slopes express the number-average MW $(M_n)$ correction term for $T_g$ in this equation. Thus, the PCA description naturally gives a method for prediction of $T_{g\infty}$-values from the slopes of the PD lines.

The indicated relationships between the PD lines and experimental glass transition temperatures suggest the potential usefulness of the PCA description in evaluating similarities between the polymers. In particular the PCA description may be useful for identifying polymers with $T_g$-values in ranges above or below those of the polymers PLA and PGL. It is anticipated that comparisons can also be performed for other polymer properties than $T_g$-values, on the basis of the natural clustering of the polymers according to molecular weight and chemical constitution.

*Supplementary material.* The design matrix for the PCA analysis (Table 3) may be obtained from the authors on request.

(nos. 2/6, 3/7, 10/14, 11/15 and 19/23), with the exception of two compounds which are C9 hydroxy acids with phenyl groups (nos. 84/88 and 92/96). The PGL similar compounds are all glycolic acid or *p*-hydroxybenzoic esters, (nos. 1/5, 9/13, 18/22, 26/30, 27/31 and 35/39). The PET-resembling double-esters all contain at least one phenyl group in the monomers, either in the double-ester (25/29), but most often in the main chain (nos. 4/8, 12/16, 20/24, 28/32, 50/54 and 51/55), as in the benchmark compound.

Glass transition temperatures taken from the literature[11,12] and reports[13] are plotted against the PC1 scores for the HMW and LMW series of polymers for one synthesized double-ester (nos. 107 and 111; P73) and for the benchmark polymers in Fig. 5. For the limited number of data points it was found that within each MW group (LMW range 5800–10800, HMW range 23000–43300) the $T_g$-values correlate with the PC1 scores, indicating that semi-quantitative interpolations of $T_g$-values between those of PET and PE can be made for the double esters.

## References

1. Vert, M. *Eur. Mater. Res. Soc. Monogr. 3 (Biomaterials)* (1992) 223.
2. Piskin, E. *J. Biomater. Sci. Polym. 6* (1995) 775.
3. Hayashi, T. *Progr. Polym. Sci. 19* (1994) 663.
4. Dugstad, H., Foss, P. A., Klaveness, J., Rongved, P., Skurtveit, R., Solberg, J. and Aukrust, I. R. F. *Pat. WO* 9521631 A1 950817 and Dugstad, H., Foss, P. A., Klaveness, J., Redford, K., Rongved, P., Skurtveit, R. and Strande, P. *Pat. WO* 9506518 A1 950309.
5. Klaveness, J., Strande, P., Rongved, P., Aksnes, E., Holm, K. H., Redford, K., Solberg, J., Stori, Aa. and Vik, I. B. *Det 8. Organisk Kjemiske Vintermøte*, Beito 1993.
6. Ghose, A. K. and Crippen, G. M. *J. Chem. Inf. Comput. Sci. 27* (1987) 21.
7. Pearson, R. G. *J. Org. Chem. 54* (1989) 1423.
8. *MOPAC 6.0 A*, General Molecular Orbital Package, QCPE Program no. 455, Department of Chemistry, Indiana University, Bloomington, IN.
9. *Hyperchem 4.5*, Hypercube, Inc. Waterloo, Ontario, Canada.
10. *Unscrambler 6.1*, CAMO ASA, Trondheim, Norway.
11. Kitao, T., Sugiyama, Y., Konishi, T., Kimura, Y., Ohya, S. and Ariki, M. *Kobunshi Ronbunshu 41* (1984) 717.
12. Bicerano, J. *Prediction of Polymer Properties*, 2nd edn., Marcel Decker Inc., New York 1996.
13. $T_g$ for 107, 111 is ca. $-45$ to $-2\,°C$; Redford, K. *SINTEF Report* 21.05.1995 and Thatcher, J. and Stori, Aa. *SINTEF Report* 06.09.1996.