Environment Agency

using science to create a better place

Calculation of Molecular Dimensions Related to Indicators for Low Bioaccumulation Potential

Science report

SCHO0109BPGT-E-P

ea/br/e/sci/v1

The Environment Agency is the leading public body protecting and improving the environment in England and Wales.

It's our job to make sure that air, land and water are looked after by everyone in today's society, so that tomorrow's generations inherit a cleaner, healthier world.

Our work includes tackling flooding and pollution incidents, reducing industry's impacts on the environment, cleaning up rivers, coastal waters and contaminated land, and improving wildlife habitats.

This report is the result of research commissioned and funded by the Environment Agency's Science Programme.

Published by:
Environment Agency, Rio House, Waterside Drive, Aztec West, Almondsbury, Bristol, BS32 4UD
Tel: 01454 624400 Fax: 01454 624409
www.environment-agency.gov.uk

ISBN: 978-1-84432-978-6

This report is printed on Cyclus Print, a 100% recycled stock, which is 100% post consumer waste and is totally chlorine free. Water used is treated and in most cases returned to source in better condition than removed.

Further copies of this report are available from:
The Environment Agency's National Customer Contact Centre by emailing enquiries@environment-agency.gov.uk or by telephoning 08708 506506.

Author(s):
Brooke D N and Cronin M T D

Dissemination Status:
Publicly available

Keywords:
Bioaccumulation, molecular dimensions, indicators, modelling, PBT

Research Contractor:
Building Research Establishment Ltd, Bucknalls Lane, Garston, Watford WD25 9XX

Liverpool John Moores University, School of Pharmacy and Chemistry, Byrom Street, Liverpool L3 3AF

Environment Agency's Project Manager:
I Doyle, Science Department, Chemicals Assessment Unit (CAU), Wallingford OX10 8BD.
Tel: +44 (0)1491 828557

Environment Agency's Project Executive:
S Robertson (CAU)

Product Code: SCHO0109BPGT-E-P

Science at the Environment Agency

Science underpins the work of the Environment Agency. It provides an up-to-date understanding of the world about us and helps us to develop monitoring tools and techniques to manage our environment as efficiently and effectively as possible.

The work of the Environment Agency's Science Department is a key ingredient in the partnership between research, policy and operations that enables the Environment Agency to protect and restore our environment.

The science programme focuses on five main areas of activity:

- **Setting the agenda**, by identifying where strategic science can inform our evidence-based policies, advisory and regulatory roles;
- **Funding science**, by supporting programmes, projects and people in response to long-term strategic needs, medium-term policy priorities and shorter-term operational requirements;
- **Managing science**, by ensuring that our programmes and projects are fit for purpose and executed according to international scientific standards;
- **Carrying out science**, by undertaking research – either by contracting it out to research organisations and consultancies or by doing it ourselves;
- **Delivering information, advice, tools and techniques**, by making appropriate products available to our policy and operations staff.

Steve Killeen

Head of Science

Executive Summary

This study assesses various methods for calculating molecular dimensions, which may be suitable as indicators of a low potential for chemical bioaccumulation in wildlife. There is a particular need for clear and unambiguous measures of molecular dimensions to ensure that they can be used consistently under the REACH Regulation.

A number of computational chemistry and molecular modelling tools were assessed for their ability to calculate molecular dimensions and five software packages (OASIS, MOE, TSAR, Mol2Mol and SPARTAN) were chosen for further analysis using a set of 69 compounds that represent a broad range of sizes and shapes. Other potentially useful software packages (e.g., CROSS, Dragon, MOPAC, ChemDraw, etc.) are largely unexplored at this point in time.

The conformation of a molecule used for the calculations greatly affects its molecular dimensions. Conformational analysis (using OASIS) and molecular dynamics (using MOE) showed that there may also be considerable variation in the molecular dimensions calculated by different modelling packages. There are no particular reasons to select any specific molecular conformation for analysis. A possible solution is to consider a range of possible molecular dimensions for all relevant conformers, and choose either the minimum value or an average.

One indicator that is mentioned in the REACH guidance documentation – the maximum molecular length – appears to have been developed manually over twenty years ago, using a very small data set. It is not recommended that this descriptor, in its current form, be used as an indicator. Other molecular dimensions for indicators, e.g., effective diameter, are better described and can be related to the outputs from some of the modelling tools, namely OASIS, Mol2Mol and SPARTAN. It is not possible to say which of these provides the “best” values. A preference for OASIS could be based on the fact that the indicator values in the REACH guidance document were derived using results from this tool. However, screening the tested substances against the 17 Å and 700 Da indicators produced the same results for OASIS and Mol2Mol, albeit on a small data set.

Contents

1 Introduction

1.1 Purpose of this study

The purpose of this study was to evaluate computer programs that can be used to calculate the molecular descriptors that are used as indicators for low bioaccumulation potential under the REACH Regulation. There is a particular need for consistency given the range of methods available.

1.2 Structure of the report

Section 1 provides an overview of the descriptors used for the size and shape of molecules, and describes the indicators currently used in the REACH guidance.

Section 2 briefly presents the models and methods used in this study (more detailed information on the models is included in Appendix A) and the chemical structures selected for the calculations.

Section 3 presents the results of the calculations and analysis of the molecular descriptors.

Section 4 considers the effect of accounting for different conformers on the calculations, for those models which include this possibility.

Section 5 summarises the main findings from the analysis with reference, in particular, to the indicators included in the REACH guidance.

Section 6 presents the conclusions and recommendations.

The final two sections contain acknowledgments (Section 7) and references (Section 8).

Appendix A contains details of the models used and notes on other models that were investigated, but were not of immediate use for this study.

Appendix B contains structure diagrams and SMILES strings for the structures used in the study.

Appendix C contains the full calculation results from the four models used in the study (for the parameters related to molecular dimensions as explained in Section 3).

Appendix D contains the detailed results of the correlation analyses described in Section 3.

Appendix E contains the detailed results of the conformation and molecular dynamics analyses as summarised in Section 4.

1.3 Theoretical Basis of the Measurement of Molecular Dimensions

It is a fundamental principle in chemistry that a molecule has a defined shape and size. This is key to the understanding of phenomena as diverse as chemical reactions, receptor binding, dissolution and the passage of molecules through a membrane. It is convenient to treat a molecule as a solid object where each atom is a “ball” or “sphere” with a fixed diameter based on its size. The size of an atom is usually defined as the “nonbonding atomic” radius or its van der Waals radius.

However, the treatment of a molecule as a series of solid balls is misleading. A molecule consists of a series of filled molecular orbitals, each orbital (in its ground state) is filled by a pair of electrons. Depending on the nature of the atoms within the molecule; electrons may be distributed evenly across the molecule or attracted to a certain area in the molecule and hence produce a polarity effect or dipole moment. Molecular orbital theory implies an outer shell or highest occupied molecular orbital (HOMO) and this is often assumed to be the “surface” of a molecule.

Conventionally it is assumed that two molecules will repel each other when they come within a certain distance of each other. If the molecules have sufficient energy and come close enough together to exchange electrons, then a covalent reaction may take place. Quantification of these distances and surfaces can only be assumed. However, the surface of a molecule may be estimated as the surface of individual atoms (i.e., the van der Waals radius). From this, assuming a reasonable conformation can be obtained (usually by computation) the dimensions of a molecule can be estimated.

1.4 Overview of Descriptors for the Size and Shape of Molecules

There are a number of models for predicting the bioaccumulative effects of pollutants from their chemical structure. Within these models, and the approaches used in their application, there are cut-off criteria. The implication is that “large” molecules do not accumulate in sufficient quantity to cause concern because they are not able to penetrate cell membranes.

There are a number of approaches used to measure and calculate the properties of molecules. Traditionally, quantitative structure-activity relationships (QSARs) have been developed that utilise descriptors for the hydrophobic, electronic and steric effects of molecules. The ability to describe the size and shape of molecules is often considered fundamental to the modelling of receptor binding phenomena (e.g., drug / ligand interactions) and membrane passage.

The modelling and description of molecular size and shape should be considered separately. Size is normally associated with the gross dimensions of a molecule, while shape is a more subtle quality that is difficult to quantify. There are many further complicating factors, such as whether the molecule is considered in two or three dimensions.

There are a small number of reviews of the available steric descriptors of molecules, especially those applied in QSAR (see for instance relevant sections in Dearden 1990 and Todeschini and Consonni 2000). Some of the more frequently used descriptors for the size and shape of molecules are listed in Table 1.1.

Table 1.1 Commonly used descriptors of the size and shape (steric properties) of molecules

Descriptor	Type	2D or 3D dependent	Comments
Molecular weight	Size – gross descriptor of molecular bulk.	2D	Fundamental and comprehensible.
Molecular surface area (including accessible surface area etc.)	Size – generally thought to be a gross descriptor of molecular bulk.	3D	Dependent on conformation.
Molecular volume (including van der Waals volume etc.)	Size – generally thought to be a gross descriptor of molecular bulk.	3D	Dependent on conformation.
Molecular dimensions (e.g., longest distance between two atoms)	Size – probably the most direct assessment of molecular size.	3D	Dependent on conformation.
Topological indices (e.g., molecular connectivities, kappa shape indices, various miscellaneous indices)	Size and to a lesser extent shape. Features from gross molecular bulk, to branching and presence / absence of particular molecular features (e.g., rings).	2D	Not dependent on conformation.

Steric descriptors, particularly topological parameters such as molecular connectivities and the electrotopological state indices, are commonly used in QSAR. Topological parameters are descriptors of a molecule calculated directly from knowledge of the connections between atoms (i.e., the bonds) and how these may be counted. Large molecules have more connections with “paths”, i.e., the number of bonds connected together is higher, and so their indices will be greater. The most commonly applied topological descriptors are the molecular connectivities, which are calculated for the whole molecule from graph theory (i.e., knowledge of the connections between the atoms). Extension of these indices includes the electrotopological state indices for individual atoms and atoms types that encode, in addition to topological information, information about the electronic environment of each atom. The extended use of topological indices is predominantly due to the ease of use and the speed with which they can be calculated, combined with low cost. However, there are many problems with their use, including the lack of interpretability (hence contradicting the requirement of the OECD Principles for the Validation of QSARs) and co-linearity. Co-linearity, i.e., that the descriptors hold the same information, is a common issue with most steric descriptors.

There is widespread use and acceptance of steric descriptors, but those molecular dimensions that are suitable indicators for low bioaccumulation potential are a small subset of those available.

The 3-D modelling of molecules is well established and there is a wide variety of computational chemistry approaches available to the interested user. For small molecules, including those described in this report (the process may be different for macromolecules such as proteins), molecular modelling usually requires the graphical entry of a molecular structure into a software package. Graphical entry (or drawing) of a chemical structure can be performed in a number of ways. After the structure has been entered a process of energy minimisation, or optimisation, is required, which calculates a minimum energy value for the 3-D structure of the molecule. There are several approaches to this; one of the most common for small molecules is the use of molecular orbital (MO) theory. MO theory attempts to model the molecule by treating it as a series of molecular orbitals. There are a number of calculation methods (classically termed "Hamiltonians") and one of the most common of these (which is applied in this report) is the Austin Model 1 (AM1).

An extension of molecular modelling software, with particular application to the development of QSARs, is the development of molecular spreadsheets. A molecular spreadsheet has the appearance of a typical spreadsheet, e.g., Microsoft Excel®, but is able to import and export molecular structures and perform calculations on the structures. Thus, in this study, simple 2-D (SMILES) strings representing each structure were read into a molecular spreadsheet. The spreadsheet calculated a "reasonable" 3-D geometry for each structure – this is a rough, low-energy structure calculated by molecular mechanics that are considered more rapid, but less accurate, than molecular orbital methods. From this conformation the spreadsheet then performed an energy optimisation calculation using the molecular orbital AM1 Hamiltonian. The updated structure, as well as any calculated values and other descriptors (e.g., topological descriptors) may be stored in the spreadsheet. As with a normal spreadsheet, each row represents the information for a single chemical, each column a single property. The spreadsheet may allow for limited statistical analysis or the data can be exported to a bespoke statistical package.

Less well established in the scientific literature is the effect that different conformers of a molecule will have on property (including dimension) calculations. Different conformers of a molecule exist because the rotation of bonds (usually single bonds) can give a molecule different shapes. Thus a straight chain alkane can, in theory, undergo bond rotation to any number of shapes, each of which may have different dimensions. Not every combination of bond angles is energetically feasible and conformation analysis attempts, through calculation, to determine the energetically feasible conformers (or shapes) for a molecule. A molecule with many rotatable (single) bonds is expected to be highly flexible. Measures of molecular flexibility, which are usually a count of rotatable bonds or similar, are available and have been of considerable importance in estimating drug bioavailability.

Molecular dynamics and its associated computational calculations relate to how the atoms of a molecule are given energy and interact with other atoms - according to the laws of physics. Atoms move, or vibrate, along their bonds and the movement of atoms in a molecule can be calculated and hence viewed. The advantage of this approach in modelling small molecules is that it often reveals a range of energetically feasible conformations that cannot be obtained by other means. As noted above, for flexible molecules, these different conformations may have significantly different molecular dimensions.

1.5 Indicators of Compounds with Limited Bioaccumulation Potential

A group of indicators for limited bioaccumulation potential has been developed as part of the guidance for the assessment of substances as PBT or vPvB substances under REACH (EC, 2008). These indicators are set out below:

Used within a weight of evidence approach and with expert judgment, a chemical may be considered as not bioaccumulative (i.e., unlikely to have a BCF > 2,000) using the following types of evidence:

An average maximum diameter ($D_{max\ aver}$) of greater than 17 Å (1.7 nm) plus a molecular weight of greater than 1100 (Da)
A maximum molecular length (MML) of greater than 43 Å (4.3 nm)
Octanol-water partition coefficient as $\log_{10}$ ($\log K_{ow}$) > 10
A measured octanol solubility (mg/l) < 0.002 × molecular weigh (MW) (without observed toxicity or other indicators of bioaccumulation)

In addition to indicators 2, 3 and 4 above, and again within a weight of evidence approach and with expert judgment, an indicator for considering a chemical as possibly not being very bioaccumulative (i.e., unlikely to have a BCF > 5,000) is if it has:

A $D_{max\ aver}$ of greater than 17 Å (1.7 nm) plus a molecular weight of greater than 700 (Da)

In using the indicators above it should be noted that 1 and 2 are generally considered potential barriers to uptake, 3 is considered a general indicator of uptake, distribution and availability (i.e., bioaccumulation in lipid containing parts of the organism) and 4 is an indicator of potential mass storage in lipid tissues.

Currently there appears to be some confusion over the exact definition of these indicators, particularly those relating to molecular diameters.

2 Methodology

2.1 Calculation of molecular dimensions

In order to assess the calculation of molecular descriptors a set of compounds was investigated using a variety of molecular software packages. The individual software packages, along with details such as vendor, version used etc, are described in Appendix A. They were chosen on the basis of availability at no or little cost, or already being available to the main investigator.

Following the calculation of descriptors for all compounds, a number of subjective comments regarding the operation of the software can be made. The comments are intended to reflect the ease of use of the software in the hands of expert users. They also reflect the state-of-art of these software packages. The comments are summarised below.

General Comments:
All software ran on a stand-alone PC running Windows XP. Computation times were rapid (seconds or minutes per compound – depending on the level of conformational analysis or dynamics).
All software requires a degree of expertise for use. However, a competent (biological or chemical) scientist, once trained, should be able to use it easily.

OASIS -
is familiar to many environmental researchers and regulators in the European Union (and beyond).
calculated descriptors efficiently.
was easy to use and automatically performed the conformational analysis.
is the basis of the platform for the first phase of the OECD (Q)SAR Application Toolbox.
The full version of OASIS is a commercial package and requires a license. However, as stated above, a version of OASIS is the platform for the current version of the OECD (Q)SAR Application Toolbox (which is available free of charge). Clarification regarding use of OASIS to develop and use bioaccumulation indicators should be gained from the developer (Laboratory of Mathematical Chemistry, Bourgas, Bulgaria).

MOE -
is primarily a drug design tool and therefore includes a large number of features specifically for this activity.
calculated descriptors efficiently, although dynamics could not be performed (in the first run) on a small number of molecules.
was easy to use and was able to perform a molecular dynamics analysis.
is a commercial package and requires a license.

TSAR -
is a well-established molecular spreadsheet.
calculated descriptors efficiently.
does not perform conformational analysis or molecular dynamics.
was easy to use.
is a commercial package and requires a license.

Mol2Mol -
is a less frequently used molecular modelling and manipulation package.
calculated descriptors efficiently.
does not perform conformational analysis or molecular dynamics.
was easy to use.
is a commercial package and requires a license. It is the cheapest of the commercial packages considered.

Spartan –
is a comprehensive molecular modelling package.
calculated descriptors efficiently, although calculation can be slow.
does not perform conformational analysis or molecular dynamics.
was easy to use.
is a commercial package and requires a license.

A number of other software packages were also considered for use in this study, but these did not provide suitable descriptors. These are noted in Table 2.1.

Table 2.1 Software packages considered but not used in this study to calculate molecular dimensions

Software	Used to Calculate Molecular Dimensions in this Study	Comments	Description in Appendix A.n
ChemOffice (version 6.0)	No	Did not provide useful measurements for molecular dimensions. Performs a molecular dynamics run, but does not store molecular co-ordinates.	A.5
Dragon	No	Did not provide useful measurements for molecular dimensions.	A.6
MOPAC	No	A package for molecular orbital calculations. Not suited to calculate molecular dimensions. The calculations are known to be the same as those performed by the VAMP routine in TSAR.	-
Nemesis	No	Performs basic molecular modelling. Will calculate a surface area that can be visualised, but does not quantify the area.	-

The method described by Cash and Nabholz (2002) has not been considered.

2.2 Molecules included

Descriptors were calculated for a total of 69 molecules. These molecules comprised:

46 molecules considered to be of environmental importance in terms of bioaccumulation (as supplied by the Environment Agency). These included molecules that have unknown bioaccumulative potential but are considered difficult to test. This set included a number of highly flexible molecules. As noted in Section 1.2, a highly flexible molecule is deemed to be one that has many rotatable (single) bonds. This means that rotation of the molecule may occur at many bonds and result in many different conformations (and hence a range of molecular dimensions).

A “reference” set of twenty molecules taken from Patel and Cronin (2001). The original data set published by Patel and Cronin (2001) was devised solely to investigate the parameterisation of molecular linearity. Molecular linearity is a term to describe whether a molecule is “linear” or “non-linear” (i.e., spherical). As an example, hexene and cyclohexane have the same molecular formula (C_6H_{12}) but differing shapes; hexene is likely to adopt a linear conformation and cyclohexane is spherical. The data set was developed with the sole purpose of assessing a series of molecules which ranged from small to large and from spherical to oblong. Molecular structures were written to cover these ranges. It should be understood that these chemicals are not necessarily of commercial, or other, importance; they were simply written to account for a selection of shapes. Therefore, this data set is artificial, although all the structures are chemically feasible. From the original set of two hundred compounds, twenty were chosen to reflect a range of sizes (with an emphasis on larger molecules) and to reflect compounds with large amounts of molecular flexibility. In no way should these twenty compounds be thought of as being of environmental importance.

Three dye structures (from Ciba Specialty Chemicals, Basel, Switzerland).

The same set of compounds was submitted to all of the software packages to determine the usefulness, or otherwise, of the descriptors that were calculated. Relevant descriptors for molecular dimensions were identified and are described in detail below. Less relevant descriptors for molecular dimensions are summarised in the sections on each software package in Appendix A.

The names, SMILES and chiral SMILES of the 69 compounds analysed are available in Appendix B. This file is available electronically if required.

2.2.1 Chirality

The chirality of the compounds was not defined before modelling and was not available for the majority of the compounds analysed (where known it was applied). In order to have a consistent definition of the molecule and the same molecular structure entered into different software, chiral SMILES were developed (see Appendix B) and were used as the starting point for all calculations.

2.2.2 Conformational Analysis and Molecular Dynamics

Molecules are made up of bonds and atoms and certain bonding arrangements allow for considerable flexibility in molecules. Typically single bonds will rotate. If a molecule

has a large number of rotatable bonds then it is likely to form a number of different conformations (i.e., shapes) when energy is added to the system. Thus, molecules under physiological temperature and conditions are not considered to be static, but flexible and dynamic entities. This has significant implications for the calculation of molecular dimensions as these may vary considerably, depending on which conformation is assessed.

Traditionally in the calculation of descriptors for QSAR analysis, molecular conformation is either not relevant (e.g., 2-D SMILES input for the calculation of log K_{ow}) or is assumed to be the lowest energy conformation. Lowest energy structures are calculated most reliably from molecular orbital theory (either semi-empirical or *ab initio*). However, it has long been recognised that there are problems with the use of a single lowest energy conformation. These include the "trapping" of a structure in a local minimum and molecular flexibility under physiological conditions that makes consideration of a single conformation unrealistic.

As described in Section 1.3, conformational analysis is a computational approach to calculate all possible low energy (and hence thermodynamically possible) molecular conformers. Theoretically it is possible to calculate molecular dimensions for these different molecular conformations. A variety of techniques are available to perform these analyses. Some rely on calculation of different conformers from an evaluation of molecular structure and the occurrence of rotatable bonds (e.g., OASIS), others apply molecular dynamics calculations to determine different conformers.

In this study the effect of conformers and flexibility was assessed in two ways:

The OASIS software calculated all stable conformers within 10 kcal/mol of the lowest energy conformation. A flexible molecule is likely to have more conformers that are energetically stable than a less flexible molecule. This type of search is restricted to energetically favourable conformers only.

A molecular dynamics simulation was performed in MOE. Dynamics is intended to identify a range of conformers, but not necessarily the most energetically stable.

3 Calculations of Values Relating to Molecular Dimensions

This section describes the calculation and analysis of molecular descriptors using the software described in Section 2.1. Many hundreds of descriptors were calculated during the study and only a very small proportion is described in detail. Those descriptors relate directly to molecular dimensions of importance to the development, or application, of indicators for low bioaccumulation potential.

The descriptors relevant to molecular dimensions, as calculated by the software described below, are included in Appendix C.

All the descriptors calculated, including those from conformation analysis and dynamics and those deemed to be not relevant to molecular dimensions, are available electronically. Please note that the file size is very large and there are many thousands of values in total.

3.1 OASIS

OASIS calculates a number of descriptors relevant to parameterising molecular dimensions. A number of these descriptors have been used previously for the modelling of bioaccumulation (cf. Dimitrov et al., 2003). The details of the OASIS software and a summary of the computational methodology are included in Appendix A.1.

The descriptors listed in Table 3.1 were found to be relevant to describing molecular dimensions.

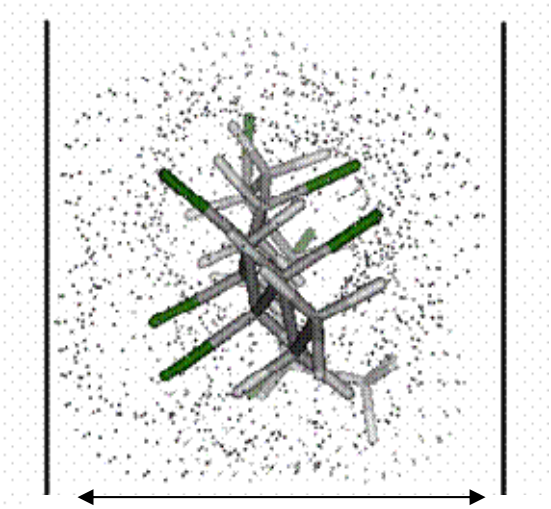
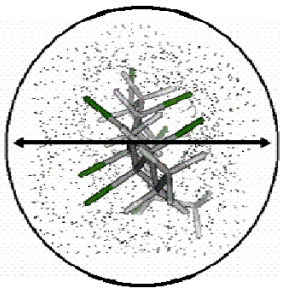
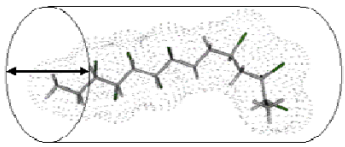
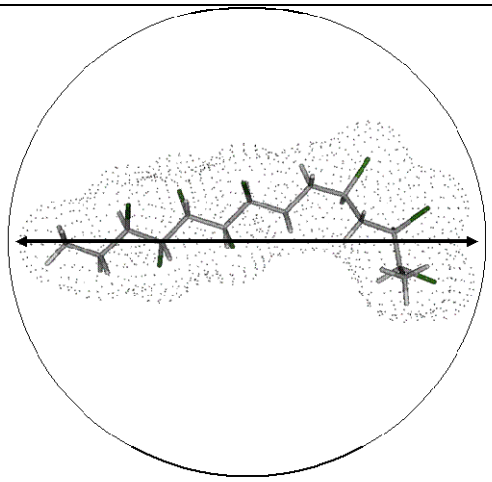
Table 3.1 Descriptors calculated for molecular dimensions calculated by OASIS.

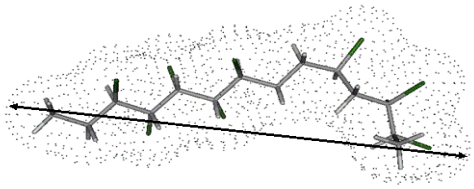
Descriptor	Abbreviation	Unit	Description
Minimum diameter of the molecule.	D_{min}	Å	The minimum distance between two parallel planes circumscribed[a] the molecule.
Effective cross-sectional diameter of the molecule.	D_{eff}	Å	The minimum diameter of infinite cylinders circumscribed[a] the molecule
Maximum diameter of the molecule.	D_{max}	Å	Minimum diameter of spheres circumscribed[a] the molecule.
Maximal distance in the molecule.	Max Distance	Å	The greatest inter-atomic distance between two non-bonded atoms.

[a]As defined in Dimitrov et al (2003). It may be better to describe this as "…. circumscribing the molecule"

These descriptors are shown diagrammatically in Figure 3.1.

Figure 3.1 Diagrams to illustrate the molecular dimensions calculated by OASIS, using the most appropriate view of molecule number 20 ($C_{14}H_{22}Cl_8$ (59.9% wt. Cl))

Minimum diameter of the molecule	 Diammin is the minimum distance between two parallel planes surrounding a molecule
Effective cross-sectional diameter of the molecule	  Diameff is the minimum diameter of a cyclinder that surrounds a molecule
Maximum diameter of the molecule (N.B. the cross section of the sphere is shown here)	 Diammax is the minimum diameter of a sphere that surrounds the molecule

Maximal distance in the molecule	 Max Distance is the maximum distance between two atoms in a molecule

3.1.1 Analysis of the OASIS Descriptors

The descriptors were calculated from OASIS for the energy minimised structure (results of the conformational analysis are below). The correlation matrix of the four descriptors is reported in Table D 1 in Appendix D. Figure D 1 shows the relationships between these descriptors graphically.

The results indicate that for the energy minimised structures considered, the two parameters D_{max} and Max Distance are collinear. However, there is little correlation between the three Diameters (D_{eff}, D_{max} and D_{min}). This is possibly surprising as they are calculated from the same structures and broadly account for the size of the molecule. It suggests that these parameters encode different information about the dimensions of molecules. Therefore, their interchangeable use may require further assessment.

Due to their widespread use, and to their use in developing one of the indicator values, the three Diameters calculated by OASIS are used to assess the meaning (if any) of parameters calculated by other software.

3.2 MOE

The Molecular Operating Environment (MOE) software is a flexible package for molecular modelling. It is used particularly in structure-based development of novel compounds. The details of the MOE software and a summary of the computational methodology are included in Appendix A.2.

As part of its molecular modelling capability, MOE is able to calculate numerous descriptors and perform dynamics calculations. Of the descriptors available in MOE, the descriptors listed in Table 3.2 were found to be relevant to describing molecular dimensions.

Table 3.2 Descriptors calculated for molecular dimensions calculated by MOE

Descriptor	Definition
rgyr	Radius of gyration.
std_dim1	Standard dimension 1: the square root of the largest eigenvalue of the covariance matrix of the atomic coordinates. A standard dimension is equivalent to the standard deviation along a principal component axis.
std_dim2	Standard dimension 2: the square root of the second largest eigenvalue of the covariance matrix of the atomic coordinates. A standard dimension is equivalent to the standard deviation along a principal component axis.
std_dim3	Standard dimension 3: the square root of the third largest eigenvalue of the covariance matrix of the atomic coordinates. A standard dimension is equivalent to the standard deviation along a principal component axis.

The radius of gyration can be defined as the root mean square distance of the objects' parts from either its centre of gravity or an axis. It appears to be most commonly used for polymers and is used to describe the dimensions of a polymer chain. The radius of gyration (R) of a particular molecule at a given time is defined as:

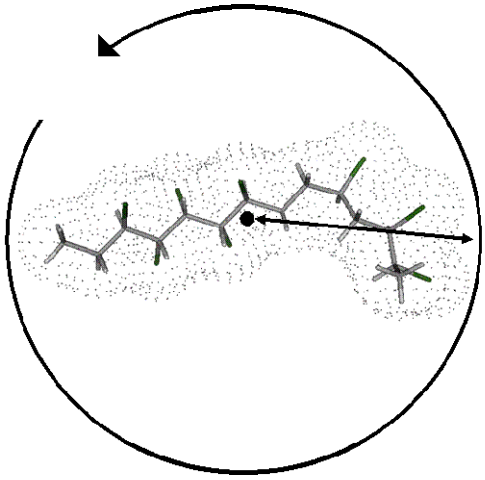
$$R_g^2 = \frac{1}{N}\sum_{k=1}^{N}(r_k - r_{mean})^2$$

Where r_{mean} is the mean position of the monomers.

The standard dimensions are calculated from a covariance matrix (i.e., a principal component analysis of atomic co-ordinates). This is effectively a data reduction technique from atomic co-ordinates and not a direct measurement of molecular dimensions. As such it is more difficult to visualise this descriptor, however it is useful to assess the role of dynamics in providing different conformations.

A diagrammatic representation of the radius of gyration is shown in Figure 3.2. It should be noted that no diagram can be produced for the standard dimensions. These are statistical reductions of atomic co-ordinates and, therefore, cannot be demonstrated graphically.

Figure 3.2 Diagrammatic approximation of the radius of gyration (marked as the double arrow) from the centre of mass for molecule number 20 ($C_{14}H_{22}Cl_8$ (59.9% wt. Cl))



3.2.1 Analysis of the MOE Descriptors

The descriptors were calculated from MOE for the energy minimised structure (results of the molecular dynamics analysis are below). The correlation matrix of the four descriptors is shown in Table D 2 in Appendix D. Figure D 2 shows the relationships between these descriptors graphically.

The analysis indicates that there is a reasonable relationship between rgyr and std_dim1. This is not surprising as they both account for the bulk of the molecule (i.e., the maximal amount of information from the covariance matrix of atomic co-ordinates). There is little relationship between the remaining standard dimensions (std_dim1, std_dim2, std_dim3). This is to be expected as these parameters describe different (and probably less useful) molecular information.

3.2.2 Comparison of MOE Descriptors with OASIS Molecular Dimensions

In order to assess the meaning of the descriptors calculated by MOE, a correlation matrix showing their relationship to the OASIS molecular dimensions is given in Table D 3 and is illustrated graphically in Figure D 3 in Appendix D.

The analysis of the MOE and OASIS descriptors shows that, while some trends are visible, e.g., between D_{max} and both rgyr and std_dim1, few strong relationships exist between the descriptors. The MOE descriptors are not direct measurements of molecular dimensions in the same way as the OASIS descriptors. While they are measures of molecular size (the radius of gyration appears to provide a measure of how far the molecule extends from a central point within it) the standard dimensions are derived from the atomic co-ordinates using statistical analysis. Thus, these descriptors encode molecular dimensions, but are not directly comparable with, for instance, the more fundamental measures of molecular length and width provided by OASIS. This suggests that the MOE descriptors are less likely to account for molecular dimensions directly.

3.3 TSAR

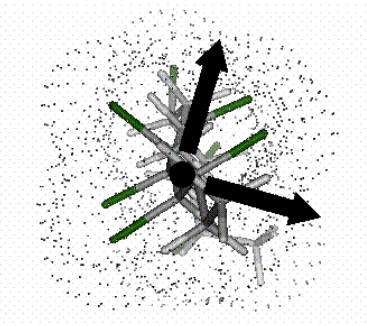
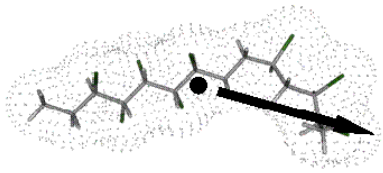
As defined in Section 1.2, a molecular spreadsheet allows the user to include molecular structures and perform calculations on those molecules. TSAR for Windows (TSAR) is a molecular spreadsheet capable of calculating molecular properties. Geometry optimisation can be performed through an interface to the VAMP programme, which performs calculations in a similar manner to MOPAC. Details of the TSAR software and a summary of the computational methodology are included in Appendix A.3.

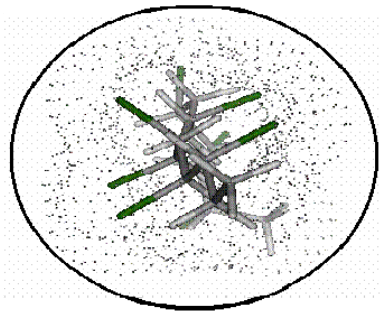
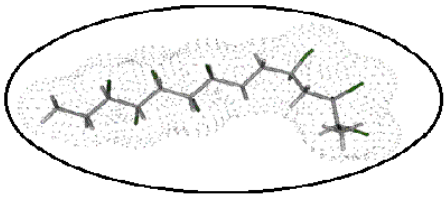
As part of its descriptor calculation capability TSAR provides numerous descriptors and performs molecular orbital calculations. Of the descriptors available in TSAR the following are relevant to describing molecular dimensions. The moment of inertia and principal axes of inertia for a molecule are calculated using the inertia tensor, following standard methods of calculation (Riley, 1983). These results are reported as Moment 1 Size (IMS1), Moment 1 Length (IML1) etc. The volume defined by these values is calculated and reported as the Ellipsoid Volume (Ell Vol).

In addition, the molecule and an ellipsoid of inertia can be viewed. The ellipsoid's principal axes are aligned with the axes of the inertia tensor. The length of each axis is inversely proportional to the moment of inertia around that axis. The resulting ellipsoid is then scaled so that the atom furthest from the centre of gravity of the molecule appears on the ellipsoid surface.

The moment of inertia is the distance from a given axis at which the mass of a body could be concentrated without altering the rotational inertia of the body about that axis. This forms a directional property, which is demonstrated graphically by the length of the arrows. Figure 3.3 demonstrates graphically the moments of inertia and ellipsoidal volume for molecule number 20.

Figure 3.3 Diagrammatic approximation of the moments of inertia and ellipsoidal volume for molecule number 20 ($C_{14}H_{22}Cl_8$ (59.9% wt. Cl)).

Moments of Inertia	  Moments of inertia derived from the centre of mass

Ellipsoidal Volume	  Ellipsoidal volume is calculated as the volume of the ellipsoid defined as shown

3.3.1 Analysis of the TSAR Descriptors

Using TSAR, the descriptors were calculated for the AM1 energy minimised structure where, as noted in Section 1.2, AM1 is the molecular orbital calculation method. The correlation matrix of the seven descriptors considered is shown in Table D 4 (Appendix D).

Analysis of the inter-relationships between the TSAR descriptors indicates that, with the exception of the IMS2 and IMS3 and IML2 and IML3, there is little significant correlation. Thus, we can conclude that the moments of inertia are describing different molecular properties. Ellipsoidal volume is partially correlated to all the descriptors, hence giving some insight to their meaning.

3.3.2 Comparison of TSAR Descriptors with OASIS Molecular Dimensions

In order to assess the meaning of the descriptors calculated by TSAR, a correlation matrix showing their relationship to the OASIS molecular dimensions is given in Table D 5 in Appendix D and is illustrated graphically in Figure D 4.

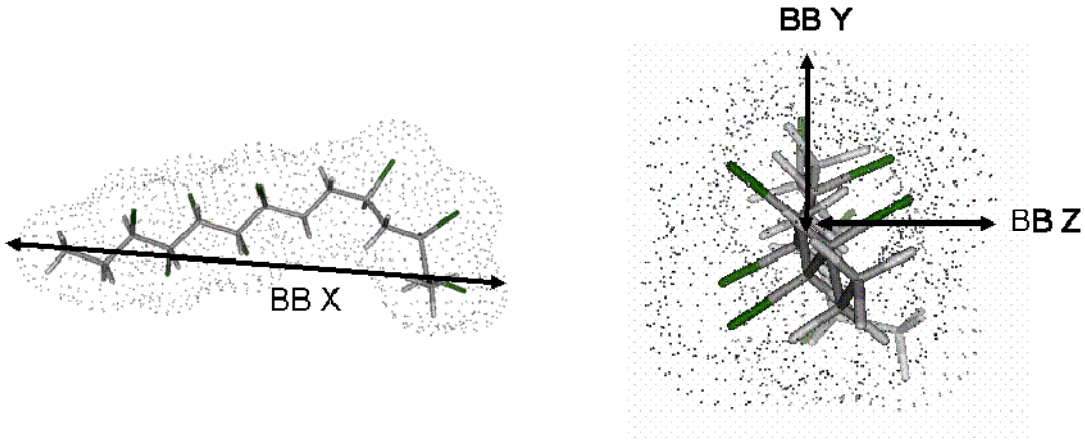
Comparison of the TSAR descriptors with those from OASIS shows few significant trends. There is a moderate relationship between both IML2 and IML3 and D_{eff}. However, overall, there are few definitive relationships between descriptors calculated using TSAR and OASIS. This can be explained by considering the definitions of the descriptors. The moments of inertia are related to molecular dimensions, but are thought to give an indication of the distribution of the mass of a molecule from a central point. In terms of molecules, a larger compound will have a larger distribution of mass around the centre and thus have a larger moment of inertia. The moments are parameterised in different directions (usually starting with the largest) and for the different properties. Thus, while they are describing the relative proportions of a molecule, they are not directly related to molecular dimensions. Similarly, ellipsoidal volume is a measure of the volume of a ellipsoid placed around the molecule. Intuitively, larger molecules will have greater volumes, but this is not a direct measurement of molecular dimensions. This suggests that the TSAR descriptors may not be a suitable surrogate for molecular dimensions.

3.4 Mol2Mol

Mol2Mol is a simple to use and straight-forward software package for the manipulation of chemical structures and calculation of molecular properties. The details of the Mol2Mol software and a summary of the computational methodology are included in Appendix A.4.

The Mol2Mol software calculates a small number of descriptors. The descriptors relevant to molecular dimensions include the "bounding box" for the molecule and from that the calculation of its x, y and z dimensions. Calculation of the bounding box takes into account the van der Waals radii of the atoms following an energy minimisation procedure. The farthest extension is always the x dimension. The calculation is performed in the following way. The largest atom-to-atom distance is looked for and the molecule is re-orientated so that these two atoms li.e., along the x axis. The molecule is rotated around the X axis and of the resulting bounding boxes the one with the minimum cross section (in the YZ plane) is accepted as the definitive one. Searching for the optimal arrangement may take considerable time for molecules with large numbers of atoms (more than 13,000-15,000 atoms). These dimensions are shown graphically in Figure 3.4.

Figure 3.4 Diagrammatic approximation BB X (longest), BB Y and BB Z properties for molecule number 20 ($C_{14}H_{22}Cl_8$ (59.9% wt. Cl)).



3.4.1 Analysis of the Mol2Mol Descriptors

The descriptors were calculated from Mol2Mol for the energy minimised structures. The correlation matrix of the three descriptors considered is shown in Table D 6 in Appendix D.

Analysis of the inter-relationships between the three Mol2Mol descriptors indicates that there is little significant correlation and hence we may conclude that each dimension is describing a different molecular property.

3.4.2 Comparison of Mol2Mol Descriptors with OASIS Molecular Dimensions

In order to assess the meaning of the descriptors calculated by Mol2Mol, a correlation matrix showing their relationship to the OASIS molecular dimensions is given in Table D 7 in Appendix D and this is illustrated graphically in Figure D 5.

The comparison of the Mol2Mol descriptors with those from OASIS indicates few significant trends. There is a moderate relationship between the X and Z dimensions of the Bounding Box (BB X; BB Z) and D_{min} as well as D_{max} and BB X. This latter relationship has a higher correlation ($r = 0.94$) when the $D_{max\ aver}$ value is considered. In some ways the molecular length descriptors from Mol2Mol may be considered the "purest" form of the molecular dimensions as they should be the most fundamental. Thus, it is encouraging that BB X (the longest dimension in a molecule) is related to D_{max} (a similar property). This analysis illustrates the problem caused by the possibility of different conformations. The other properties (BB Y, BB Z) are less likely to be correlated to OASIS descriptors as they are dimensions in different directions, i.e., perpendicular to the longest dimension.

Despite these relationships, overall, there is little correspondence between the Mol2Mol descriptors and those calculated from OASIS. This is possibly surprising as both methods appear, at face value, to be calculating similar molecular dimensions.

3.5 Spartan

Spartan is a molecular modelling and computational chemistry package. The details of the Spartan software and a summary of the computational methodology are included in Appendix A.5.

The Spartan software calculates a number of descriptors. The relevant descriptors for molecular dimensions include the calculation of how "Long" and "Thick" a molecule is.

Spartan was utilised to calculate descriptors for a limited number of molecules as noted in Appendix C.5.

3.5.1 Analysis of the Spartan Descriptors

The descriptors were calculated from Spartan for the equilibrium conformation from AM1 calculations. The correlation matrix for the two descriptors that were considered is shown as part of Table D 8 in Appendix D.

Analysis of the inter-relationships between the two Spartan descriptors indicates that there is little significant correlation and hence we may conclude that Long and Thick each describe a different molecular property.

3.5.2 Comparison of Spartan Descriptors with Mol2Mol BB X and OASIS Molecular Dimensions

In order to assess the meaning of the descriptors calculated by Spartan, a correlation matrix showing their relationship to BB X and the OASIS molecular dimensions is given in Table D 8 in Appendix D and presented graphically in Figure D 6.

The results show that the Spartan Long descriptor is well correlated with BB X and D_{max}. This suggests that these three descriptors account for the longest distance between atoms in molecules.

The Spartan Thick descriptor is moderately well correlated with D_{eff}. This suggests that both describe the “thickness” of the molecule. In OASIS this descriptor is an attempt to parameterise a cylinder placed around the molecule.

4 Effect of Molecular Flexibility – Conformational Analysis and Molecular Dynamics

To investigate the effect of flexibility, conformation analysis and molecular dynamics were performed using the OASIS and MOE packages, respectively (as described above and in Appendices A.1 and A.2).

All of these analyses produced vast amounts of data. The analyses calculate descriptors for each conformation (up to 30 per molecule in OASIS) and for 200 dynamic simulations in MOE. Full data are available if required, but the Excel files take up several megabytes of disk space and so are not included in this report. For clarity and simplicity the following measures have been identified for each of the relevant descriptors:

minimum value for all conformations
maximum value for all conformations
average value (arithmetic mean) for all conformations
value for the lowest energy (minimised) structure (for comparison)

Values for a) to d) were obtained in Microsoft Excel for each compound and summarised in a master table.

A more extensive discussion of this area is included in Appendix E along with the tables and figures referred to in this summary section.

4.1 Conformational Analysis using OASIS

A full set of minimum, maximum, average and energy minimised values for diameter (D_{min}, D_{eff}, D_{max}) and Max Distance for the compounds considered are given in Appendix C.

Analysis of the values for the conformers provides a good illustration of the effect of conformational analysis. Figure E 1 to Figure E 4 (in Appendix E) show the distribution of these descriptors by increasing descriptor value. The plots show the minimum and maximum values as a “range” with the average value shown in the range.

Figure E 1 to Figure E 4 show clearly that there is a considerable effect of conformational analysis. To illustrate this, for each descriptor the ranges of parameters for the three molecules with the greatest average value of that descriptor (and hence which are known to be flexible) are recorded in Table E 1 to Table E4.

The results indicate that conformation flexibility can have a considerable effect on the value of descriptors. A 40% variation in the average value from minimum to maximum is common and even greater ranges occur.

4.1.1 Comparison of OASIS Descriptors following Conformational Analysis

In order to determine the effect of the different conformers, the maximum, minimum and average values for the four OASIS descriptors were analysed. These descriptors were also compared against each other as well as those for the energy minimised structure. The intention of this part of the study was:

To determine if there was any relationship between the descriptors calculated for different conformers (i.e., could they be scaled to each other).
To determine if, for instance, the energy minimised structure could be used as a suitable conformer to encode the information from the other conformers.

The overall aim of this analysis was to assess which, if any, conformation was appropriate to describe the properties of a molecule. Table E 5 to Table E 8 show the correlation matrices for the descriptor values, for the maximum, minimum and average values, and those for the energy minimised structure; these relationships are shown graphically in Figure E 5 to Figure E 8.

The analysis appears to confirm that for the OASIS descriptors there is a significant difference between the minimum, maximum and average values from conformational analysis. This variation appears to be greatest for the largest compounds, which reflects their greater intrinsic flexibility. Therefore, it is not possible to determine which conformation to use as the most appropriate representation of the dimensions of the molecule.

However, for all descriptors there appears to be a good relationship between the values for the energy minimised structures and the maximum value. Therefore, the value for the energy minimised structure could act as a surrogate for the maximum value and negate the need for conformational analysis.

4.2 Molecular Dynamics using MOE

A full set of minimum, maximum, average and energy minimised values for rgyr, std_dim1, std_dim2 and std_dim3, for the compounds considered, are given in Appendix C.

Analysis of the values for the conformers provides a good illustration of the effect of molecular dynamics. Figure E 9 to Figure E 12 show the distribution of these descriptors by increasing descriptor value. The plots show the minimum and maximum values as a “range” with the average value shown in the range.

4.2.1 Comparison of MOE Descriptors following Molecular Dynamics

In order to determine the effect of the molecular dynamics procedure, the maximum, minimum and average values for the four MOE descriptors were analysed. These descriptors were compared against each other and those for the energy minimised structure. Table E 9 to Table E 12 show the correlation matrices for the values of the descriptors for the maximum, minimum and average values and those for the energy minimised structure; these relationships are shown graphically in Figure E 13 to Figure E 16.

The analysis shows that the descriptor rgyr is little affected by the variation in conformers following molecular dynamics. This may be expected as this is not a true measure of molecular dimensions, but a “globularity” measure which may be less susceptible to alteration by changing conformation.

The other descriptors calculated by MOE (std_dim1, std_dim2, std_dim3) do, however, show a significant difference between the minimum, maximum and average values from molecular dynamics. The variation appears to be greatest for the largest compounds, which reflects their greater intrinsic flexibility. Therefore, it is not possible to determine which conformation to use as the most appropriate representation of the dimensions of the molecule.

For one of the MOE descriptors (rgyr), and possibly std_dim1, there appears to be a good relationship between the values for the energy minimised structures and the maximum value. Therefore, the value for the energy minimised structure could act as a surrogate for the maximum value and negate the need for conformational analysis.

5 Relationships Between Descriptors and Indicators

This section summarises the main findings of the analyses with particular regard to the main indicators included in Annex 1 of the PBT Assessment Guidance (see Section 1.2). The indicators considered are the (average) maximum diameter, maximum molecular length and molecular weight.

5.1 Average Maximum Diameter ($D_{max\ aver}$)

The unit for $D_{max\ aver}$ is the Angstrom (Å). 1 Å equals 0.1 nanometre (nm or 10^{-9} metre).

Appendix R.11-1 of the REACH PBT assessment guidance (EC, 2008) makes reference to a number of studies relating to the use of D_{max}. These resulted in the following indicators being included in the Guidance:

Possibly not B: a $D_{max\ aver}$ of > 17 Å plus a molecular weight greater than 1100 (Da)
Possibly not vB: a $D_{max\ aver}$ of > 17 Å plus a molecular weight greater than 700 (Da)

The initial justification for the $D_{max\ aver}$ value was provided by Dimitrov et al (2002, 2003), who demonstrated the relationship between log BCF and D_{max} for 694 chemicals. From their analysis it can be clearly shown that there is a relationship between increasing D_{max} and log BCF. There is also an indication that molecules with D_{max} greater than 1.5 nm have a log BCF less than 3.3. Dimitrov et al (2003) showed a very similar relationship between log BCF and D_{max}, and surmised that a D_{max} of 14.7Å provides an indicator for a BCF of 5500.

Dimitrov et al (2002) justified the existence of such a transition point by suggesting a change in the mechanism of uptake of chemicals: passive diffusion of small molecules through the phospholipid bilayer of the membrane is replaced by the more conservative mechanism of exocytosis and endocytosis of larger molecules. The value of 1.5 nm for the threshold is comparable with cell membrane architecture. The threshold for maximum diameter is comparable with the half thickness of one (of the two) lipid layers that constitute the cell membrane.

Dimitrov et al (2005) revised the indicator by modelling the effect of molecular size using the following equation:

$$F_{MS} = 1 - \frac{1}{1 + e^{-c(D_{max\ aver} - D_{max\_thr})}}$$

Where D_{max_thr} is the threshold (indicator) for molecular size, which was assumed to be a more conservative value of 17Å, based on the analysis in Dimitrov et al (2003).

To illustrate the potential use of the indicators, the $D_{max\ aver}$ values[1] and MW have been calculated, using OASIS, for the compounds described in Section 2.2. These values are listed in Table 5.1. Analysis of these results indicates that 21 of the 69 compounds have a $D_{max\ aver}$ greater than 17 Å. Of these, only 5 compounds have a molecular weight greater than 700 Da, which is the indicator that a compound is possibly not very bioaccumulative (vB). No compound in this study had a molecular weight greater than 1100 Da, so the indicator that a compound is possibly not bioaccumulative (B) would not be triggered.

Table 5.1 D_{max}, BB X, Long and MW for the compounds analysed in this study. Compounds are ordered according to D_{max} (less than and greater 17Å) and compounds with MW greater than 700 Da are identified.

ID	Chemical name (where known)	D_{max} (Å)	BB X (Å)	Long (Å)	MW (Da)
Compounds with D_{max} less than 17 Å					
R13		7.8	7.6	8.3	74.1
R66		10.8	10.6	11.6	134.2
40	2-styrylphenol	11.3	10.0		198.3
R111		11.6	10.8	11.6	184.3
41	4-styrylphenol	12.1	10.4		198.3
12	C10H18Cl4 (51.7% wt. Cl)	12.9	12.7		280.1
R20		13.0	15.1	15.0	158.3
R165		13.5	12.6	13.6	296.5
R91		13.5	13.5	15.6	236.5
R103		13.7	14.9	14.1	240.5
R120		13.8	14.2	13.6	268.6
R198		14.0	11.8	13.6	376.6
42	2,4-Distyryl phenol	14.0	13.8	12.5	302.4
43	2,6-Distyrylphenol	14.0	11.8	13.6	302.4
13	C11H19Cl5 (54.0% wt Cl)	14.0	14.4		328.6
R68		14.1	15.3	15.4	198.4
10	2,2’,3,3’,4,5,5’,6,6’-Nonabromodiphenyl ether (BDE208)	14.1	12.6	14.7	880.2
R124		14.4	15.0	15.1	310.7
4	2,2',4,4',5,5'-Hexabromodiphenyl ether (BDE153)	14.4	12.6		643.6
3	2,2',4,4',6-Pentabromodiphenyl ether (BDE100)	14.5	12.6		564.7
2	2,2',4,4',5-Pentabromodiphenyl ether (BDE99)	14.5	12.5		564.7
1	2,2',4,4'-Tetrabromodiphenyl ether (BDE47)	14.5	12.5		485.8
6	2,2',3,4,4',5',6-Heptabromodiphenyl ether (BDE183)	14.5	12.6		722.5
14	C11H18Cl6 (58.7% wt. Cl)	14.5	12.4	13.6	363.0
R118		14.5	14.7	14.1	282.6
5	2,2’,4,4’,5,6’-Hexabromodiphenyl ether (BDE154)	14.5	12.6		643.6
7	2,2',3,3',4,4',6,6'-Octabromodiphenyl ether (BDE197)	14.5	12.9		801.3
8	2,2’,3,3’,4,4’,5,5’,6-Nonabromodiphenyl ether (BDE206)	14.6	12.9		880.2
15	C12H21Cl5 (51.8% wt. Cl)	14.6	15.8		342.6
9	2,2’,3,3’,4,4’,5,6,6’-Nonabromodiphenyl ether	14.7	12.9		880.2

[1] $D_{max\ aver}$ is the average value for D_{max} over the conformers considered, see Appendix E.

ID	Chemical name (where known)	D_{max} (Å)	BB X (Å)	Long (Å)	MW (Da)
	(BDE207)				
11	Decabromodiphenyl ether	14.7	14.9	13.0	959.1
R26		14.7	17.8	18.8	184.4
16	C13H22Cl6 (54.5% wt. Cl)	14.9	14.7		391.1
44	2,4,6-Tristyrylphenol	15.1	14.9		406.6
46	1,2-Bis(pentabromophenyl) ethane	15.3	15.0	17.0	971.2
R173		15.3	15.7	16.6	310.7
18	C14H24Cl6 (52.6% wt. Cl)	15.4	16.6		405.1
37	Di-(tert-dodecyl) polysulphide v2	15.4	14.3		434.9
17	C14H26Cl4 (42.3% wt. Cl)	15.7	17.8		336.2
R172		15.7	15.1	15.3	422.9
19	C14H24Cl6 (52.6% wt. Cl)	15.9	14.8		405.1
20	C14H22Cl8 (59.9% wt. Cl)	16.2	17.7		474.0
R190		16.3	17.7	17.3	479.0
21	C14H22Cl8 (59.9% wt. Cl)	16.4	17.0		474.0
36	Di-(tert-dodecyl) polysulphide	16.5	18.0	16.0	434.9
22	C14H18Cl12 (69.9% wt.Cl)	16.5	16.8		611.7
35	Di-(tert-nonyl) polysulphide v2	16.5	17.2		414.9
23	C15H26Cl6 (50.8% wt. Cl)	16.7	17.7	17.7	419.1
Compounds with D_{max} greater than 17 Å, but MW less than 700 Da					
34	Di-(tert-nonyl) polysulphide	17.1	17.3	16.7	414.9
24	C16H27Cl7 (53.2% wt. Cl)	17.3	19.5		467.6
39	Di-(tert-dodecyl) pentasulphide v2	17.3	17.9		499.0
R191		17.3	17.0	16.5	366.8
25	C17H29Cl7 (51.6% wt. Cl)	17.9	15.1		481.6
R135		18.6	21.8	16.3	460.7
38	Di-(tert-dodecyl) pentasulphide	18.8	21.5		499.0
R141		18.8	20.8	19.8	476.6
26	C18H31Cl7 (50.2% wt. Cl)	18.9	20.8		495.6
27	C19H32Cl8 (52.2% wt. Cl)	19.2	21.6		544.1
28	C20H34Cl8 (50.9% wt. Cl)	20.3	23.0		558.1
D1	CI Pigment Orange 73	20.7	19.4		400.6
R130		21.4	25.1	14.9	518.7
29	C22H37Cl9 (51.5% wt. Cl)	21.7	23.9		620.6
30	C24H40Cl10 (52.0% wt. Cl)	22.5	24.0		683.1
D3	CI Pigment Yellow 13	25.4	28.8		685.7
Compounds with D_{max} greater than 17 Å and MW greater than 700 Da					
45	Ethylene bistetrabromophthalimide (1H-Isoindole-1,3(2H)-dione, 2,2'-(1,2-ethane)	17.6	18.3	17.0	951.4
31	C26H43Cl11 (52.4% wt. Cl)	24.5	27.5		745.6
D2	CI Pigment Violet 37	25.0	23.7		726.8
32	C28H46Cl12 (52.7% wt. Cl)	26.1	26.9		808.1
33	C30H49Cl13 (53.0% wt. Cl)	27.4	25.7		870.6

Table 5.1 also includes the BB X values calculated with Mol2Mol. An additional seven molecules have a BB X value greater than 17 Å, and one has a value less than this, when compared to the D_{max} results. The same five substances are identified as having a dimension greater than the indicator and a molecular weight above 700 Da. There is a good correlation between the BB X and D_{max} values (see Table D 7 in Appendix D). The

values of BB X tend to be larger than the corresponding D_{max} values (although there are many exceptions); this may be due to the fact that Mol2Mol does not address the conformational flexibility of the molecules.

Of the 30 structures for which calculations were performed with SPARTAN, four have a value for Long greater than 17 Å. For all of these, the molecular weight is less than 700 Da. The BB X values for all four are also above 17 Å, but the D_{max} values are less than this for three of the structures. SPARTAN calculations were only carried out on one of the structures that met the indicators, based on the results of the other two programs; the Long value for this substance is 17 Å, which is exactly the indicator value. There does not appear to be a consistent pattern to the values in comparison to those from the other two programs included in Table 5.1.

Overall it is clear that different conclusions would be drawn on the basis of the output from the three programs, although this is based on a small number of cases, especially for SPARTAN.

The analysis in Section 3 showed that none of the descriptors from TSAR and MOE were strongly related to D_{max}.

5.2 Maximum Molecular Length (MML)

The unit for MML is the Angstrom (Å). 1 Å equals 0.1 nanometre (nm or 10^{-9} metre).

Appendix R.11-1 of the REACH PBT assessment guidance (EC, 2008) refers to the study by Opperhuizen et al (1987) who proposed that a substance of length greater than 43 Å would not pass membranes. This included both the gills and the gut, and was based on a series of bioaccumulation and bioconcentration studies with linear and cyclic polydimethylsiloxanes (silicones) of varying chain length. This chain length corresponds to a molecular dimension (molecular length) of 43 Å, equal to the length of the polydimethylsiloxane (PDMS) congener where reduced uptake was observed. This has been translated into the following statement in the Guidance:

A maximum molecular length of 43 Å indicates no uptake and indicates a chemical is not B or vB.

This is a very general statement and while it may be appropriate for PDMS and fatty acids, it may be too broad to be widely applicable outside of this series of chemicals. Indeed, among the chemicals in this study none would be excluded on this basis. There is no single, easily calculated, parameter termed “maximum molecular length” (MML). The most representative value is probably the greatest inter-atomic distance between two non-bonded atoms (Max Distance) as calculated by OASIS[2]. For the 69 compounds considered in this study, following conformational analysis, the greatest Max Distance is 34.768Å (for $C_{30}H_{49}Cl_{13}$ (53.0% wt. Cl)). Therefore, none of the compounds considered in this study, which includes a wide range of structures, would be identified by the indicator based on MML.

[2] It is worth noting that the Max Distance values for the tested substances are all exceeded by the corresponding D_{max} values (average, minimised etc). Hence the D_{max} values may be more relevant for the comparison, in particular the D_{max} value of the longest conformer. Although the values are higher, none are above 37 Å. The values are in Appendix C.

Concern over the use of the MML parameter also relates to its calculation. It is clear from Opperhuizen et al (1985) that it was not the purpose of their study to develop indicators for low bioaccumulation for widespread application. In addition, this was not a computationally calculated parameter – indeed it is very unlikely that such a calculation would have been performed for such large molecules in the mid-1980s. It appears from Opperhuizen et al (1985) (page 278) that the MML calculation for PDMS was performed by hand. Therefore, while it is valid in the context of the paper, it is unlikely to be comparable to the computational calculations performed in this study.

5.3 Molecular Weight (MW)

The unit for molecular weight is the Dalton (Da) or g/mol (for molar mass).

Molecular weight (MW) is a fundamental and unambiguous parameter. It is simple to calculate and has been used in QSAR studies for many years. MW is considered the simplest measure of molecular size; it is simple to calculate and not affected by conformations. There has been a belief for many years that “large” molecules will not pass through membranes (unless there is an active transport mechanism) and hence will not have biological activity. MW is a good descriptor for this, and upper limits for (pharmacological) activity are commonly assumed to be in the region of 500-700 Da. Indeed, the Lipinski rule of five uses an upper limit of 500 Da.

A compounding factor is the decrease in aqueous solubility associated with increasing molecular size. Regardless of whether large compounds can, or cannot, permeate membranes, they may not be sufficiently soluble to allow for bioavailability.

The limits of 700 and 1100 Da in the PBT Guidance are quite conservative as would be expected with indicators based on existing legislative limits. Few molecules with a MW greater than 700 Da would be expected to show any significant biological activity because of the issues of aqueous solubility, bioavailability and membrane permeability. It may be possible to use aqueous solubility as an additonal indicator for BCF. There are a variety of methods to calculate solubility (of varying precision) should a measured value not be available. The reduction in biological activity would be considerably greater for compounds with MW greater than 1100 Da, and so such compounds would not be expected to be bioavailable. It may be that a lower indicator value for MW could be considered.

The molecular weights for the compounds analysed in this study are given in Table 5.1. The table shows that only 5 of the 69 compounds have a MW greater than 700 Da and none of the compounds have MW greater than 1100 Da. At this time it is not known what the MW distribution of chemicals will be under REACH, but it is likely that there will be only a small proportion with MW greater than 700 Da and even fewer with MW greater than 1100 Da.

6 Conclusions and Recommendations

A number of conclusions and recommendations can be made from the data analysed in this report.

From a technical perspective, all tested software ran efficiently and rapidly. Potentially useful descriptors were obtained from MOE, Mol2Mol, OASIS, SPARTAN and TSAR. MOE and OASIS have the advantage of performing molecular dynamics and conformational analysis, respectively. There were generally few problems in obtaining descriptors for the compounds from all the software packages, although molecular dynamics did not run in MOE for all compounds. With regard to acceptability, the OASIS software is available commercially and forms the basis of the platform underpinning the freely available OECD (Q)SAR Application Toolbox.

OASIS, Mol2Mol and SPARTAN provide results that can be directly compared to some of the indicators. The values from MOE and TSAR take a different form.

The units for dimensions appear to be clear. Most usable descriptors, e.g., D_{max} from OASIS and descriptors from Mol2Mol (e.g., BB X), are in Angstrom and molecular weight is in Dalton. It is recommended that units in Å or Da be considered for the indicators as they are unambiguous and may be related to physiological processes.

Whilst some descriptors are well correlated, e.g., $D_{max\ aver}$ and BB X ($r = 0.94$), there is generally considerable variation between methods used to predict molecular dimensions. Descriptors from energy minimized structures (as opposed to those subjected to conformational analysis) were less well correlated, e.g., for $D_{max\ minimized}$ and BB X, $r = 0.83$. Visual analysis of the data revealed that all methods produce results showing the correct trend, i.e., molecules perceived to be larger have greater values. However, there is no practical method to assess which method is “correct”, because the actual dimensions of the molecules are unknown. Techniques such as X-ray crystallography could potentially be used to derive actual molecular dimensions, although these would be for crystalline structures.

The maximum molecular length (MML) is a simple and fundamental property that could provide a useful indicator if developed further. The current indicator based on this property should be treated with caution and may need to be described more fully on the basis of a broader group of compounds. MML appears not to have been calculated from a computational or molecular modelling approach and the indicator based on it is currently ill-defined. As a result it is not possible to relate any of the outputs from the modelling tools directly to this indicator and this makes its difficult to use. Further work should be carried out to develop indicators for the same property based on the currently available descriptors.

The descriptors maximum diameter and maximum cross-sectional diameter appear to be well defined in the literature and are well described in the REACH PBT assessment guidance. Further clarification of these descriptors is available in this report. These descriptors appear to be calculated well by the OASIS software.

Other software packages (e.g., Dragon, MOPAC, ChemDraw etc) calculate a large number of descriptors that could potentially relate to dimensions. These are largely

unexplored at this point in time, although will suffer from a lack of conformation analysis. Other software (e.g., CROSS) is not readily available and thus at this time would not appear to be a suitable method for recommendation for general use in the calculation of descriptors.

The practice of applying molecular dimensions as indicators for low or negligible uptake appears to be acceptable. Any of the descriptors analysed in detail in this study could, potentially, be used to provide such an indicator. To be used correctly, indicators would need to be scaled appropriately to take account of differences in methodologies used to calculate them. The definition of novel indicators would require further research.

The conformation used for calculations, obtained either through conformational analysis or molecular dynamics, greatly affects molecular dimensions. Deviations in dimensions of approximately 50% of the average were commonly observed.

There are no particular reasons to select any specific molecular conformation for analysis. Pragmatically, the range of values could be considered (or possibly 95% limits), within acceptable limits (i.e., avoiding unrealistically high energy conformers). From a precautionary point of view, the minimum value from any conformer should be utilized, although an average value may prove more realistic.

The descriptors D_{max} from OASIS, BB X from Mol2Mol and Long from SPARTAN appear to be comparable in principle with the 17 Å indicator. However, the values of these three descriptors for individual structures can differ considerably and decisions based on the output of each program could be different. As already noted, it is not possible to select a “correct” value in terms of the actual dimensions. A potential approach could be to use OASIS (D_{max}), as the indicator value was derived using values of this descriptor. Alternatively, similar exercises could be carried out to develop equivalent indicators for each program. The use of several different software tools could lead to variable results for the same substance (exceeding the indicator for one but not for another); however, this does not contradict the weight of evidence approach proposed in the guidance.

7 Acknowledgements

The assistance of various software developers is gratefully acknowledged. In particular the help of Professor Ovanes Mekenyan and his research group, the Laboratory of Mathematical Chemistry, Bourgas, Bulgaria was invaluable.

The authors gratefully acknowledge the contribution of Dr. Hugo Camenzind, Ciba Specialty Chemicals, Basel, Switzerland for the provision of calculations from Spartan and information about that software.

8 References

Anon (2007) Comparative Analysis of Molecular Radii, Volume and Membrane Transport Rates for Alkane Polysulfide Compounds. A report prepared for the Polysulfides Trade Association

Bassan A, Worth AP (2008) The integrated use of models for the properties and effects of chemicals by means of a structured workflow. QSAR Comb. Sci. 27: 6-20.

Cash GC, Nabholz JV (2002) Minimum cross-sectional diameter: calculating when molecules may not fit through a biological membrane. Environmental Toxicology and Chemistry. 21: 2095-2098.

Dearden JC (1990) Physico-chemical descriptors. In: Karcher W, Devillers J (eds) *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*. European Community, Brussels pp. 25-59

Dimitrov S, Dimitrova N, Parkerton T, Comber M, Bonnell M, Mekenyan O (2005) Base-line model for identifying the bioaccumulation potential of chemicals. SAR QSAR Environ. Res. 16: 531-554.

Dimitrov SD, Dimitrova NC, Walker JD, Veith GD, Mekenyan OG (2002) Predicting bioconcentration factors of highly hydrophobic chemicals: effects of molecular size. Pure Appl. Chem. 74: 1823-1830.

Dimitrov SD, Dimitrova NC, Walker JD, Veith GD, Mekenyan OG (2003) Bioconcentration potential predictions based on molecular attributes – an early warning approach for chemicals found in humans, birds, fish and wildlife. QSAR Comb. Sci. 22: 58-68.

EC (2008). Guidance on information requirements and chemical safety assessment. Chapter R.11: PBT Assessment. May 2008. Available from http://reach.jrc.it/docs/guidance_document/information_requirements_r11_en.pdf?vers=20_08_08

Opperhuizen A, Damen HWJ, Asyee GM, van der Steen JMD (1987) Uptake and elimination by fish of polydimethylsiloxanes (Silicones) after dietary and aqueous exposure. Toxicol. Environ. Chem. 13: 265-285.

Patel H, Cronin MTD (2001) A novel index for the description of molecular linearity. J. Chem. Inf. Comp. Sci. 41: 1228-1236.

Riley KF (1983) Mathematical Methods for the Physical Sciences. Cambridge University Press.

Todeschini R, Consonni V (2000) *Handbook of Molecular Descriptors.* Wiley-VCH.

Wermuth CG, Ganellin CR, Lindberg P, Mitscher LA (1998) Glossary of terms used in medicinal chemistry. Pure Appl. Chem. 70: 1129-1143.

9 Glossary

Disclaimer: These definitions have been provided in good faith by the authors and are intended to be used only in the context of this report. Some definitions have been adapted from standard texts (Wermuth et al 1998; Bassan and Worth 2008).

Term	Definition	Reference
2-D molecular descriptors	Descriptors of molecular properties that are calculated from the 2-D structure i.e., without the need to determine a 3-D structure. Examples include calculated log K_{ow}, MW etc.	
3-D molecular descriptors	Descriptors of the molecular properties that are calculated from the 3-D structure of a molecule. Therefore they require some form of molecular modelling and calculation of a minimum energy structure. They are considered to be conformationally dependent. Examples include molecular dimensions, dipole moment, molecular orbital properties etc.	
Ab initio calculation method	A high level molecular orbital calculation method. It is time consuming but considered to be the most accurate calculation method.	
Analogue approach and category approach	The terms category approach and analogue approach describe techniques for grouping chemicals. The term analogue approach is used when the grouping is based on a very limited number of chemicals, where trends in properties are not apparent. A chemical category is a group of chemicals whose physicochemical and human health and/or environmental toxicological properties and/or environmental fate properties are likely to be similar or follow a regular pattern as a result of structural similarity (or other similarity characteristic). In principle, there should be sufficient members in the chemical category to enable the detection of trends across endpoints. As the number of chemicals being grouped into a category increases, the potential for developing hypotheses and making generalisations about the trends will also increase, and hence increase the robustness of the evaluation.	Bassan and Worth (2008)
"Bounding Box" of the molecule	The bounding box of the molecule is calculated taking into account the van der Waals radii of the atoms following an energy minimisation procedure. The dimensions of the box can be calculated and used as molecular dimensions.	
ChemOffice	Molecular Modelling package to provide visualisation and display of molecular surfaces, orbitals, electrostatic potentials, charge densities and spin densities.	
Computational Chemistry	Computational chemistry is a discipline using mathematical methods for the calculation of molecular properties or for the simulation of molecular behaviour.	Wermuth et al (1998)
Dragon	Software for the calculation of 2-D molecular descriptors.	
Effective cross-sectional diameter of the molecule	The minimum diameter of infinite cylinders circumscribing the molecule.	
Electronic properties (of a molecule)	Those descriptors of a molecule that parametrise its electronic properties. The electronic properties include the ionisation constant, hydrogen bonding, dipole moment, nucleophilic and electrophilic reactivity, amongst others. These are described by descriptors such as the pKa, molecular orbital calculations and many others.	

Term	Definition	Reference
Electrotopological state descriptors	These are an extension of molecular connectivities to the atomic level such that they encode information about the unique electronic environment of each atom along with its topology. They may be summed for different atom types across molecules.	
Ellipsoidal Volume	A measure of the volume of a ellipsoid placed around the molecule.	
Expert system	This is a broadly used term for any formal system, generally computer-based, which enables a user to obtain rational predictions about the properties or biological activity of chemicals. Expert systems may be classified as knowledge-based (when the rules are based on expert knowledge), induction rule-based (when statistical methods are used to automatically derive the rules) or hybrid (when both approaches are present). One or more databases may additionally be integrated in the system.	Bassan and Worth (2008)
Hamiltonian	The level of calculation for semi-empirical molecular orbital calculations. Typical Hamiltons include AM1, PM3, PM5 etc.	
Hydrophobic properties (of a molecule)	Those descriptors of a molecule that parametrise its hydrophobic properties. The hydrophobic properties include the ability to distribute between oil and water. These are described by descriptors such as the octanol-water partition coefficient.	
Lowest energy structure	The calculated 3-D structure of a molecule with the lowest energy (in a vacuum at 0 Kelvin). This is commonly used for the calculation of descriptors when conformational analysis is not available.	
Maximal distance in the molecule	The greatest inter-atomic distance between two non-bonded atoms.	
Maximum diameter of the molecule	Minimum diameter of spheres circumscribing the molecule.	
Membrane passage	The passage of a molecule through a membrane such as the skin, gut wall, cell wall etc. This is a vital physiological process. It is assumed that this is mainly achieved by passive diffusion, although active and facilitated transport are very important processes.	
Minimum diameter of the molecule	The minimum distance between two parallel planes circumscribing the molecule.	
MOE	Molecular Operating Environment commercial software from Chemical Computing Group used particularly for drug design. It includes a number of molecular modelling and molecular orbital properties calculations capabilities.	
Mol2Mol	A simple software package to enable molecular file transformations along with limited molecular modelling and calculations.	

Term	Definition	Reference
Molecular chirality	Chiral molecules are isomers that are non-superimposable on their mirror-image. Chiral centres in molecules may, in theory, alter the molecular dimensions but in practice the effect is quite small.	
Molecular conformer	The conformer of a molecule is the 3-D construction of a molecule. Molecules will have a number of conformers, the number of energetically feasible conformers rising with the degree of flexibility of a molecule.	
Molecular connectivities	These are probably the most popular and widely used topological descriptors. They largely encode the size and bulk of a molecule.	
Molecular descriptors	This is a term used in the development of quantitative structure-activity relationships (QSARs). Molecular descriptors parameterise the properties of a molecule, usually in three areas: hydrophobic, electronic and steric. It is the relationship of these molecular descriptors with the biological or other activity (or physico-chemical property) that forms the basis of a (Q)SAR.	
Molecular dimensions	The dimensions of a molecule. There are no standard or formal definitions. However, these are usually thought of as a maximum length within a molecule i.e., the maximum distance between two points on the surface, and properties or dimensions perpendicular to that. Further dimensions described in this report include the enclosure of a molecule in a circle or oval, and attempts to quantify that object.	
Molecular dynamics	Molecular dynamics and its associated computational calculations relate to the process by which the atoms of a molecule are given energy and allowed to interact with other atoms according to the laws of physics. This means that atoms will move or vibrate along their bonds and the motion of atoms in a molecule can be calculated and hence viewed. Molecular dynamics can be used to calculate different molecular conformations.	
Molecular flexibility	The intrinsic flexibility of a molecule. A flexible molecule is deemed to be one that has many rotatable (single) bonds. Measures of molecular flexibility, which are usually a count of rotatable bonds or similar, are available and have had considerable importance in determining drug bioavailability.	
Molecular linearity	A term to describe whether a molecule is “linear” or “non-linear” (i.e., spherical). As an example, hexene and cyclohexane have the same molecular formula (C_6H_{12}) but differing shapes with hexene likely to adopt a linear conformation and cyclohexane being spherical.	
Molecular Modelling	Molecular modeling is a technique for the investigation of molecular structures and properties using computational chemistry and graphical visualization techniques in order to provide a plausible three-dimensional representation under a given set of circumstances.	Wermuth et al 1998

Term	Definition	Reference
Molecular Modelling Software	Software for modelling the properties of molecules. It usually enables the calculation of properties of molecules, including lowest energy structures, different conformers and molecular orbital descriptors. Most molecular modelling software is commercial and will require purchase.	
Molecular orbital theory	Molecular orbital theory allows for the (theoretical) consideration of a molecule through its molecular orbitals. For the purposes of QSAR it allows calculations to be performed on molecules at a variety of levels. These calculations can determine the theoretical 3-D structure of a molecule, or a series of conformers, or allow for the calculation of (predominantly) electronic properties.	
Molecular shape	Each molecule has a nominal shape, dependent on its 3-D conformation. The shape of any molecule is vital for most chemical and biological interactions. There is no standard description of molecular shapes (in contrast to, for instance, crystals) but they may range from being described as flat, globular, spherical, chain-like etc.	
Molecular size	The overall size of a molecule. This is a dimension and in theory it should be possible to calculate, or estimate, it.	
Molecular spreadsheet	This has the appearance of a typical spreadsheet e.g., Microsoft Excel, but is able to import and export molecular structures and perform calculations on the structures. These are usually commercial software packages, the example used in this study was TSAR for Windows.	
Molecular surface area	The surface area of a molecule, usually defined by the van der Waals radii of the individual atoms. It is easily calculated although it requires 3-D molecular modelling and is conformation dependent. It is considered to be a steric descriptor and occasionally used in QSAR analysis.	
Molecular volume	The volume of a molecule, usually defined by the van der Waals radii of the individual atoms. It is easily calculated although it requires 3-D molecular modelling and is conformation dependent. It is considered to be a steric descriptor and occasionally used in QSAR analysis.	
Molecular weight (molecular mass)	The ratio of the mass of a molecule to the unified atomic mass unit. It is simple and fundamental to calculate, and is the easiest steric parameter describing molecular size or bulk for QSAR analysis.	
Moment of Inertia	This the distance from a given axis at which the mass of a body could be concentrated without altering the rotational inertia of the body about that axis. This forms a directional property.	
MOPAC	Freely available molecular orbital calculation package. Whilst MOPAC is an “industry” standard for semi-empirical calculations it does not have a graphical interface.	
Nanometre	10^{-9} metre. 10 Å = 1 nanometre.	

Term	Definition	Reference
OASIS	Software package from the Laboratory of Mathematical Chemistry, Bourgas "Prof. As. Zlatarov" University, Bulgaria. The software integrates molecular modelling with toxicity and property prediction capabilities. Part of the OASIS software will be available in the OECD (Q)SAR Application Toolbox.	
OECD Principles for the Validation of (Q)SARs	A series of rules to assist in the evaluation of a (Q)SAR for use for regulatory purposes. These state that to facilitate the consideration of a (Q)SAR model for regulatory purposes, it should be associated with the following information: i) a defined endpoint ii) an unambiguous algorithm iii) a defined domain of applicability iv) appropriate measures of goodness-of-fit, robustness and predictive power v) a mechanistic interpretation, if possible.	
OECD QSAR Application Toolbox Freeware: developed for OECD by Laboratory of Mathematical Chemistry, University of Bourgas, Bulgaria	Software tool (under development) that allows the user to: a) make (Q)SAR estimations for single chemicals; b) receive summary information on the validation results of the model according to the OECD validation principles; c) receive a list of analogues, together with their (Q)SAR estimates; d) receive estimates for metabolite activation/detoxification information. The Toolbox has the following Workflow: Step 1 – information on biotic and abiotic reactions Step 2 – classification schemes Step 3 – structural alerts Step 5a – checking membership of existing chemical categories Step 5b – chemical similarity assessment, application of the analogue approach Step 6 – application of (Q)SARs	Bassan and Worth (2008)
Quantitative Structure-Activity Relationship (QSAR)	A Quantitative Structure-Activity Relationship (QSAR) is a quantitative relationship between a biological activity (e.g., toxicity) and one or more molecular descriptors that are used to predict the activity. A molecular descriptor is a structural or physicochemical property of a molecule, or part of a molecule, which specifies a particular characteristic of the molecule and is used as an independent variable in a QSAR.	Bassan and Worth (2008)
Radius of Gyration (MOE)	The root mean square distance of the objects' parts from either its centre of gravity or an axis. It appears to be most commonly used for polymers, and is used to describe the dimensions of a polymer chain.	

Term	Definition	Reference
Read-across	A method for filling data gaps in either the analogue or category approaches. Endpoint information for one chemical is used to make a prediction of the endpoint for another chemical, which is considered to be similar in some way. In principle, read-across can be used to assess physicochemical properties, environmental fate and (eco)toxicity effects. It may be performed in a qualitative or quantitative manner. In qualitative read-across, the potential of a chemical to exhibit a property is inferred from the established potential of one or more analogues. In quantitative read-across, the numerical value of a property (or potency of an endpoint) of a chemical is inferred from the quantitative data of one or more analogues.	Bassan and Worth (2008)
Receptor binding	The binding of a molecule (the ligand) into a receptor. This is normally thought to bring about a pharmacological response, enzyme inhibition etc.	
Semi-empirical calculation method	A molecular orbital calculation method which makes several assumptions to increase speed. It is rapid, although the results are considered to be less accurate. There are many calculation methods, or Hamiltonians, the most common of which include AM1, PM3 and PM5.	
SMILES	This is a 2-D string that encodes the molecular structure. It is very commonly used and simple to learn and apply. It ignores the presence of hydrogens on the molecule. As it is 2-D it contains no 3-D or conformational information. Examples include “CCCCC” for pentane or “c1ccccc1” for benzene. The SMILES strings for all molecules considered in this study are given in the Appendices.	
Spartan	Comprehensive molecular modelling package which includes molecular orbital calculations.	
Standard Diameter (MOE)	It is calculated from a covariance matrix (i.e., a principal component analysis) of atomic co-ordinates. It is effectively a data reduction technique from atomic co-ordinates and not a direct measurement of molecular dimensions.	
Steric properties (of a molecule)	Those descriptors of a molecule that parametise its steric properties. The steric properties include the size and shape of a molecule. These are described by descriptors such as molecular weight, surface area, volume, dimensions, topological indices and many others.	

Term	Definition	Reference
Structure Activity Relationship (SAR)	Structure Activity Relationships (SARs) are theoretical models that can be used to predict, in a qualitative manner, the physicochemical, biological (e.g., toxicological) and fate properties of molecules from knowledge of chemical structure. More specifically, a SAR is a qualitative relationship (i.e., association) between a molecular (sub)structure and the presence or absence of a given biological activity, or the capacity to modulate a biological activity imparted by another substructure. The term substructure refers to an atom, or group of adjacently connected atoms, in a molecule. A substructure associated with the presence of a biological activity is sometimes called a structural alert. A SAR can also be based on the ensemble of steric and electronic features considered necessary to ensure the intermolecular interaction with a specific biological target molecule, which results in the manifestation of a specific biological effect. In this case, the SAR is sometimes called a 3D SAR or pharmacophore.	Bassan and Worth (2008)
Topological descriptors (of a molecule)	A sub-set of 2-D molecular descriptors for the steric properties of molecules. They are calculated from the knowledge of the connections between the atoms in the molecule. As such they are considered to encode the size, and occasionally the shape, of a molecule. They are quick and simple to calculate and popular for the development of QSARs by some practitioners.	
TSAR	Tools for Structure-Activity Relationships molecular spreadsheet marketed by Accelrys Inc. It includes molecular orbital theory and properties calculations.	
VAMP	Molecular orbital calculation package, analogous to MOPAC, included in the TSAR for Windows software	

10 Acronyms

2-D	2-Dimensional
3-D	3-Dimensional
Å	Angstrom
AM1	Austin Model 1
B	Bioaccumulative
BB X, BB Y, BB Z	X, Y and Z dimensions of the Bounding Box of the molecule
BCF	Bioconcentration Factor
Da	Dalton
D_{eff}	Effective cross-sectional diameter of the molecule
D_{max}	Maximum diameter of the molecule
D_{min}	Minimum diameter of the molecule
$D_{max\ aver}$	Average Maximum Diameter
D_{max_thr}	Threshold for molecular size in terms of maximum diameter of a molecule
EA	Environment Agency
Ellvol	Ellipsoidal volume
HOMO	Highest occupied molecular orbital
IMS1, IMS2, IMS3,	Moment of Inertia weighted by size along the 1st, 2nd and 3rd principal axes.
IML1, IML2, IML3	Moment of Inertia weighted by length (L) along the 1st, 2nd and 3rd principal axes.
Long	Spartan descriptor for the “length” of the molecule
Max Distance	Maximal distance in the molecule
MML	Maximum Molecular Length
MO	Molecular Orbital
MOE	Molecular Operating Environment (Commercial Software)
MOPAC	Molecular Orbital PACkage (freely available software)
MW	Molecular Weight
nm	nanometre
OECD	Organisation for Economic Co-operation and Development
P (or K_{ow})	Octanol-water partition coefficient
PBT	Persistent, Bioaccumulative and Toxic
PC	Personal Computer
PDMS	PolyDiMethylSiloxane
QSAR	Quantitative Structure-Activity Relationships
REACH	Registration, Evaluation, Authorisation and restriction of Chemicals
Rgyr	Radius of gyration
RIP	REACH Implementation Project
SMILES	Simplified Molecular Input Line Entry System
Std_dim1	Standard dimension 1
Std_dim2	Standard dimension 2
Std_dim3	Standard dimension 3
Thick	Spartan descriptor for the “thickness” of the molecule
TSAR	Tools for Structure-Activity Relationships (Commercial Software)
vPvB	very Persistent very Bioaccumulative

Appendix A – Details of Software Packages Used and Calculation Methods

Appendix A.1. OASIS

Name of Software	OASIS Basic version 1.06
Supplier	Laboratory of Mathematical Chemistry (LMC)
Supplier Address and Contact Details	Prof. Ovanes Mekenyan, Ph. D., D. Sci. Department of Physical Chemistry Laboratory of Mathematical Chemistry, Head Bourgas "Prof. As. Zlatarov" University Yakimov St. #1 8010 Bourgas, Bulgaria phone: ++359 56 880230 (also personal fax) ++359 56 858343 fax: ++359 56 880249
Internet Address	http://www.oasis-lmc.org/
Brief Description	OASIS Basic is a platform for developing 3-D QSAR models. The system provides selection of active conformers as 3D structure representatives of chemicals submitted to QSAR analysis. For that purpose chemicals are 2D-3D migrated, conformationally multiplied and quantum-chemically optimised using a straightforward procedure based on a genetic algorithm. Conformational flexibility in cyclic structures is accounted for. The number of conformers to represent conformational space is defined automatically by the system depending on the flexibility of chemicals. The selection of active conformers is based on user-imposed hypothesis on interaction mechanism. Thus, the conformational representatives of the chemicals under study could be selected according to their stability, electrophilicity, reactivity or any other user defined property. Calculated steric and electronic descriptors of selected conformers and physicochemical parameters of chemicals can be exported as excel (or html) files for a next QSAR (statistical) analysis. The system allows visualization of the conformer distributions of chemicals and assessment of chemical similarity accounting for flexibility of chemical structures.
Platform	Windows on a PC
Approximate Cost	OASIS Basic is commercial software. Details on cost are available from the supplier. A limited version of the OASIS software (e.g., parameter calculation) will be available in the OECD (Q)SAR Application Toolbox.
Summary of Descriptor(s) Available Appropriate to Molecular Dimensions	Steric descriptors Quantum-chemical descriptors Descriptors for molecular volume/surface area Other descriptors are available in the OASIS Basic package, but are considered not relevant to the parameterisation of molecular dimensions so are not included here.
Calculation Method(s)	2-D descriptors are calculated from standard methods. 3-D descriptors are calculated from geometry optimised structures (AM1 method) Conformationally multiplied based on genetic algorithm
Details of calculation method	The process used to create the data was as follows: 1. SMILES strings imported into OASIS Basic.

	2. Chemicals are 2D-3D migrated. 3. Chemicals are conformationally multiplied and quantum-chemically optimized using a straightforward procedure based on genetic algorithm. 4. Calculation of 2D, 3D (steric, electronic) and physicochemical parameters. 5. Calculated values were saved as ODB file. 6. ODB file were imported into an Excel file as a final descriptor list.
Ease of Use. Speed of Calculation.	OASIS Basic is a simple and automated software to use. Multiple input formats are supported including .MOL, .SDF, .SMI, .CMP, .ODB, .RDF, .XYZ, .INCHI, .MOL2, .TXT file types. Descriptors are split into 2 distinct blocks – structure and conformer. There are local and global descriptors. It is important to note that the time taken to perform conformer searching and optimisation is highly dependant upon molecular size and flexibility. This process may take some time depending on the molecule in question. Descriptors are calculated extremely rapidly using an automated procedure. When considering conformers, minimum and maximum descriptor values are given for the population of confomers. Calculated descriptors can then be saved as an ODB file or be exported as an Excel file for subsequent use in Excel etc. OASIS Basic can be used by any computer literate user.
Notes	OASIS Basic can use multiple input formats, however OASIS Basic converts them into non-standard SMILES notations as a working format. Details of these SMILES notations can be found in the software help files.
Descriptors Calculated Descriptors are listed in a separate file.	Deff Dmax Dmin GEOM._WIENER MAX._DISTANCE VAN_D._WAALS_SUR VAN_D._WAALS_VOL VdWSurf_PPSA1 VdWSurf_PPSA2 VdWSurf_PPSA3 VdWSurf_PNSA1 VdWSurf_PNSA2 VdWSurf_PNSA3 VdWSurf_DPSA1 VdWSurf_DPSA2 VdWSurf_DPSA3 VOLUME_POLARIZAB

Appendix A.2. MOE

Name of Software	Molecular Operating Environment 2006.08 (MOE)
Supplier	Chemical Computing Group (CCG)
Supplier Address and Contact Details	Chemical Computing Group St John's Innovation Centre Cowley Road, Cambridge, United Kingdom CB4 0WS tel: +44 1223 422320 fax: +44 1223 422318
Internet Address	http://www.chemcomp.com/
Brief Description	MOE is comprehensive computational chemistry package capable of using a wide range of methods. These include molecular docking, simple binary QSAR, molecular descriptor generation and simulation methods such as molecular dynamics.
Platform	Windows on a PC
Approximate Cost	MOE is commercial software. Details on cost are available from the supplier.
Summary of Descriptor(s) Available Appropriate to Molecular Dimensions	Water accessible surface area, molecular density, molecular globularity, standardised molecular dimensions 1, 2 and 3, molecular volume and van der Waals surface area. Molecular globularity is a measure if how spherical a molecule is, a value of 1 indicates a perfect sphere, whilst a value of 0 indicates a perfectly flat, linear molecule. The standardised molecular dimensions 1, 2 and 3 are calculated from the square route of the first, second and third eigenvalues of the molecular connectivity matrix, respectively. Thus they represent the first, second and third largest atom to atom distances in the molecule. Other descriptors are available in the MOE package, but are considered not relevant to the parameterisation of molecular dimensions so are not included here.
Calculation Method(s)	Default molecular dynamics run, followed by descriptor calculation for each conformer.
Details of calculation method	The process used to create the data was as follows: 1. SMILES strings imported into MOE 2006.08. 2. 3D structures created using the default energy minimisation function, ensuring retain chiralty option was checked. 3. For each chemical in the database a simple molecular dynamics run was performed using the default settings within MOE. Each chemical created a new database containing the retained conformers (200 per molecule). 4. Descriptors calculated for each conformer. 5. Descriptors as shown in work sheet calculated for each conformer of each structure.
Ease of Use. Speed of Calculation.	MOE is a comprehensive piece of software requiring some knowledge of computational chemistry. However well-presented

	manuals, with good 'work through' examples for each technique are available with the software and on the internet. Entry of structures is via a wide range of data formats, including SMILES strings. The software operates using databases. Initially data is imported into a database and these chemicals are then acted upon (via whichever method one chooses). The output is stored in further databases (the exact nature of which depends on the calculation being undertaken). Calculation time is extremely quick as one of MOE's primary functions is high throughput virtual screening. Data can easily be transferred to other software such as Word or Excel using several standardised file formats. MOE could be used by any computer literate scientist who understands SMILES and the basics of chemical structure. However, some time and effort would be required to get to grips with MOE's extensive functionality. This would require use of the extensive training material available.
Notes	Other functionality to assess conformational flexibility in addition to molecular dynamics exist.
Descriptors Calculated Descriptors are listed in a separate file.	Water accessible surface area (asa) Mass density (dens) Globularity (glob) Radius of gyration (rgyr) Standard dimension 1 (std_dim1) Standard dimension 2 (std_dim2) Standard dimension 3 (std_dim3) van der Waals volume (vol) van der Waals surface area (vsa)

Appendix A.3. TSAR

Name of Software	TSAR
Supplier	Accelrys Ltd.
Supplier Address and Contact Details	334 Cambridge Science Park Cambridge CB4 0WN Tel: +44 (0)1223 228500 Fax: +44 (0)1223 228501
Internet Address	http://www.accelrys.com/products/tsar/index.html
Brief Description	TSAR is a molecular spreadsheet. It allows for entry of structures as SMILES strings with conversion to 3-D structures and calculation of a wide variety of descriptors
Platform	Windows on a PC
Approximate Cost	TSAR is commercial software. Details on cost are available from the supplier.
Summary of Descriptor(s) Available Appropriate to Molecular Dimensions	Molecular connectivities Kappa indices Miscellaneous topological indices Number of rotatable bonds Molecular Mass, Surface Area, Volume, Ellipsoidal Volume Moments of Inertia Other descriptors are available in the TSAR package, but are considered not relevant to the parameterisation of molecular dimensions so are not included here.
Calculation Method(s)	2-D descriptors are calculated from standard methods. 3-D descriptors are calculated from geometry optimised structure (AM1 method)
Details of calculation method	The process used to create the data was as follows: 1. SMILES strings imported into TSAR V3.3. 2. Corina (within TSAR) was used to create 3D structures using standard default values. 3. Atomic charges calculated for each molecule using the Charge-2 methodology. 4. COSMIC (within TSAR) used to roughly optimise the structures by molecular mechanics methods. 5. VAMP (within TSAR) used to re-optimise the structures, using the AM1 Hamiltonian, restricted Hartree Fock formalisation. 6. Descriptors as shown in work sheet calculated for each structure.
Ease of Use. Speed of Calculation.	TSAR is a simple and logical piece of software to use. Comprehensive and well-presented manuals are available. Entry of structures is via SMILES strings which may be cut and pasted into TSAR. Conversion of SMILES into 3-D structures is trivial. Molecular mechanics (COSMIC) calculations take a few seconds per structure. Molecular orbital calculations (VAMP) take less than a couple of minutes for a large molecule, considerably less

	for a small molecule with little flexibility. Calculation of other descriptors is trivial. Data can be cut and pasted from TSAR into other software, e.g., Word or Excel. TSAR could be used by any computer literate scientist who understands SMILES and the basics of chemical structure.
Notes	No conformational analysis is available. It is the experience of the authors that in common with many molecular orbital software packages, VAMP will not perform calculations for “large” molecules, i.e., over 600 Da.
Descriptors Calculated (those in [square parentheses] are considered not to be relevant to molecular dimensions) Descriptors are listed in a separate file.	[Total Energy] [Energy of the Lowest Unoccupied Molecular Orbital (LUMO)] [Energy of the Highest Unoccupied Molecular Orbital (HOMO)] Zero order molecular connectivity (Kier Chi0 index) Zero order valence-corrected molecular connectivity (Kier ChiV0 index) First order molecular connectivity (Kier Chi1 index) First order valence-corrected molecular connectivity (Kier ChiV1 index) Second order molecular connectivity (Kier Chi2 index) Second order valence-corrected molecular connectivity (Kier ChiV2 index) Third order cluster molecular connectivity (Kier Chi3 (cluster) index) Third order valence-corrected cluster molecular connectivity (Kier ChiV3 (cluster) index) Fourth order cluster molecular connectivity (Kier Chi4 (cluster) index) Fourth order valence-corrected cluster molecular connectivity (Kier ChiV4 (cluster) index) Fourth order path-cluster molecular connectivity (Kier Chi4 (cluster) index) Fourth order valence-corrected path-cluster molecular connectivity (Kier ChiV4 (cluster) index) Third order molecular connectivity (Kier Chi3 index) Fourth order molecular connectivity (Kier Chi4 index) Fifth order molecular connectivity (Kier Chi5 index) Sixth order molecular connectivity (Kier Chi6 index) Third order valence-corrected molecular connectivity (Kier ChiV3 index) Fourth order valence-corrected molecular connectivity (Kier ChiV4 index) Fifth order valence-corrected molecular connectivity (Kier ChiV5 index) Sixth order valence-corrected molecular connectivity (Kier ChiV6 index) Third order ring molecular connectivity (Kier Chi3 (ring) index) Fourth order ring molecular connectivity (Kier Chi4 (ring) index) Fifth order ring molecular connectivity (Kier Chi5 (ring) index) Sixth order ring molecular connectivity (Kier Chi6 (ring) index) Third order ring valence-corrected molecular connectivity (Kier ChiV3 (ring) index) Fourth order ring valence-corrected molecular connectivity (Kier ChiV4 (ring) index) Fifth order ring valence-corrected molecular connectivity (Kier ChiV5

	(ring) index) Sixth order ring valence-corrected molecular connectivity (Kier ChiV6 (ring) index) First Order Kappa Index (Kappa1 index) Second Order Kappa Index (Kappa2 index) Third Order Kappa Index (Kappa3 index) First Order Kappa alpha Index (KAlpha1 index) Second Order Kappa alpha Index (KAlpha2 index) Third Order Kappa alpha Index (KAlpha3 index) Shape Flexibility index Rotatable Bonds Randic Topological index Balaban Topological index Wiener Topological index Molecular Mass Molecular Surface Area Molecular Volume Inertia Moment 1 Size Inertia Moment 2 Size Inertia Moment 3 Size Inertia Moment 1 Length Inertia Moment 2 Length Inertia Moment 3 Length Ellipsoidal Volume

Appendix A.4. Mol2Mol

Name of Software	Mol2Mol version 5.6.0
Website	http://web.interware.hu/frenzy/mol2mol/index.html
Supplier	Shareit
Supplier Address and Contact Details	CEO: John Strosahl Digital River GMBH Vogelsanger Str. 78 D-50823 Cologne Germany
Internet Address	www.shareit.com
Developer	Tamas E. Gunda
Developer Address and Contact Details	University of Debrecen H-4010 Debrecen Hungary Tel: (+36 52) 512 900 / 22472, 22479 Fax: (+36 52) 512 914 E-mail: tgunda2@puma.unideb.hu
Brief Description	The current version of Mol2Mol™ recognizes, reads and writes about 50 different file formats (including MOL and SDF formats). It contains a simple graphic display module to inspect the currently loaded molecule. Once a structure is loaded into Mol2Mol a number of functions are available to the user. These are summarised below: • Calculation of basic geometrical data: distances, angles, dihedral angles. • Calculation of atom pyramidalities, angle of rings, distances from least square planes, ring puckerings. • Changing of bond types, atom types. • Add or delete hydrogens, structural waters. • Add hydrogens in pH dependent mode. • Slicing of biopolymers or multiple files to individual molecules. • Batch or many-to-many conversion. • Merging of individual simple files to multiple ones. • Browsing multiple files. • Conversion of multiple structural files. • Conversion of proteins to "backbone molecules", Ramachandran plot. • Checking of peptide backbone geometry, adding CH_3, CH_2 centroid pseudo atoms.

	• Manual editing of problematic files... and many others.
Platform	Windows on a PC
Approximate Cost	Mol2Mol is commercial software. Details on cost are available from the supplier.
Summary of Descriptor(s) Available Appropriate to Molecular Dimensions	Mol2Mol is not primarily designed to generate molecular descriptors, but it is able to calculate a number of basic geometrical descriptors. The most useful is the ability to calculate a theoretical 3-dimensional box (termed a "bounding box") that fits a given molecule. This gives an indication as to the length and breadth of a molecule.
Calculation Method(s)	All geometrical descriptors were calculated from the 3D input file (SDF).
Details of calculation method	The process used to create the data was as follows: 1. Geometry optimised structures were imported into Mol2Mol in SDF format. 2. Once the structure is loaded a bounding box was calculated and the results given in the text window. 3. These data were recovered manually and entered into an Excel worksheet.
Ease of Use. Speed of Calculation.	Mol2Mol is a simple and logical piece of software to use. Multiple input formats are supported making it a very flexible tool. Given the simple nature of the geometrical descriptors, all measurements are calculated extremely quickly (less than a second). Mol2Mol can be used by any computer literate user.
Notes	No optimisation function is available. The software is dependent upon the structures given in the input file. Therefore a suitable optimised 3D structure is required prior to use. .
Descriptors Calculated Descriptor values are listed in a separate file.	Bounding Box (X Dimension (Å)) Bounding Box (Y Dimension (Å)) Bounding Box (Z Dimension (Å))

Appendix A.5.SPARTAN

Name of Software	Spartan '04
Supplier	Wavefunction Inc.
Supplier Address and Contact Details	Wavefunction Inc. 18401 Von Karman Avenue, Suite 370 Irvine, CA 92612 USA Tel: + 1 (949) 955 2120 Fax: + 1 (949) 955 2118
Internet Address	http://www.wavefun.com/products/spartan.html
Brief Description	With good visualization and a wide range of well documented computational methods in a single user-friendly software tool, Spartan delivers the full power of molecular modelling to chemists. It is a comprehensive molecular modelling package which provides access to high level molecular orbital and *ab initio* calculations.
Platform	Windows, Mackintosh on a PC, Linux, Unix
Approximate Cost	Spartan is commercial software. Details on cost are available from the supplier.
Summary of Descriptor(s) Available Appropriate to Molecular Dimensions	Molecular length and thickness ('measured' within the 3D model). Other descriptors like 'area' and 'volume' are available in Spartan, but are considered not relevant to the parameterisation of molecular dimensions so are not included here.
Calculation Method(s)	*Molecular Mechanics*: Both the SYBYL and MMFF94 force fields are supported. SYBYL extends throughout the entire Periodic Table while MMFF94 has been specifically parameterized to reproduce geometries and conformations of organic molecules and biopolymers. *Semi-Empirical* /Molecular Orbital: MNDO, AM1 and PM3 methods are supported. MNDO/d extensions for heavy main-group elements have been implemented and PM3 parameters for most transition metals are available. *Hartree-Fock*/Molecular Orbital: Hartree-Fock models useful for predicting structure, energy and property calculations, in particular for organic molecules. A variety of standard basis sets are supported: STO-3G, 3-21G, 6-31G*, 6-311G*, cc-pVDZ, cc-pVTZ and cc-pVQZ, with extensions including (d), (d, p), (2d), (2d, 2p), (2df, 2dp), (3d, 3p), (3df, 3dp) and diffuse functions and/or additional polarization functions…
Details of calculation method used	The process used to create the data was as follows: 1. SMILES strings imported and opened to/in SPARTAN '04. 2. *Equilibrium conformation* calculated with the semi-empiric AM1 optimization (in the gas phase). 3. Distances for the molecule spacing 'measured' within the relaxed 3D structure (from the peripherical atoms, corrected with the addition of 2 Van-der-Waal radii (e.g., 1.0 Å for H).
Ease of Use. Speed of	Spartan is comprehensive software requiring some knowledge of molecular modelling and computational chemistry techniques.

Calculation.	However, well-presented manuals with good 'work through' examples for each technique are available with the software. Entry of structures is via a wide range of data formats (import): Spartan, SYBYL MOL and MOL2, PDB, MacroModel, SMILES, XYZ, SDF, TGF, SKC, CIF, and CDX files. Calculation time is relatively slow. Data can easily be transferred to other software, such as Word or Excel, using several standardised formats for export: Spartan, SYBYL MOL and MOL2, PDB, MacroModel, smiles, and XYZ molecule files, graphics as JPG, PNG, and BMP files, animations as AVI files. Spartan could be used by any computer literate scientist who understands the basics of chemical structures and the transformation of different molecule formats.
Notes	"Standardised molecular dimensions" would have to be calculated separately (?)
Descriptors Calculated Descriptors are listed in a separate file.	Atomic Charges Mulliken and Natural Bond Orbital Charges are available as are charges based on fits to electrostatic potentials. Additional Properties Weight, Area, Volume, Symmetry Group, HOMO and LUMO Energies, Polar Surface Area, LogP, Ovality, Q-Minus, Q-Plus, Electronegativity and Hardness Electrical Dipole, quadrapole and higher moments, polarizabilities (including alpha, beta, and gamma terms). Thermodynamics Enthalpies, entropies and free energies as well as isotope effects, based on calculated geometries and IR vibrational frequencies.

Appendix A.6. ChemOffice

Name of Software	CS Chem3D Ultra
Supplier	CambridgeSoft
Supplier Address and Contact Details	100 Cambridge Park Drive Cambridge MA, 02140-2317, USA. Tel: 800 315-7300 Fax: 617 588-9390
Internet Address	http://www.cambridgesoft.com/software/details/?ds=3
Brief Description	Chem3D provides visualization and display of molecular surfaces, orbitals, electrostatic potentials, charge densities and spin densities. Chem3D utilizes MOPAC, Gaussian, GAMESS and extended Hückel to compute molecular properties. ChemProp computes Connolly surface areas, molecular volumes and properties, including Tinker, ClogP, molar refractivity, critical temperature and pressure.
Platform	Windows on a PC
Approximate Cost	Chem3D is commercial software. Details on cost are available from the supplier.
Summary	Although Chem3D Ultra is able to perform molecular dynamics calculations, it is unable to calculate any descriptors that relate to molecular dimensions. Therefore, it of no use in this application.

Appendix A.7. Dragon

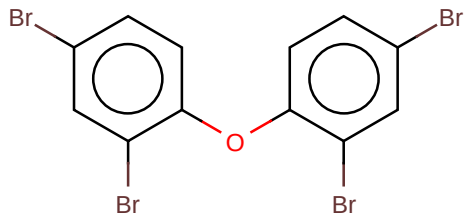
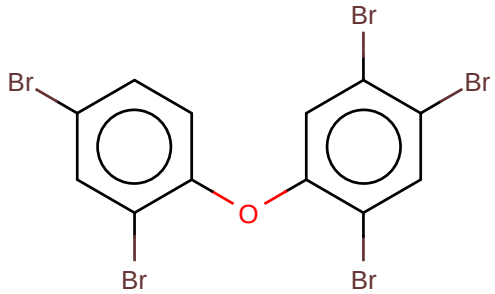
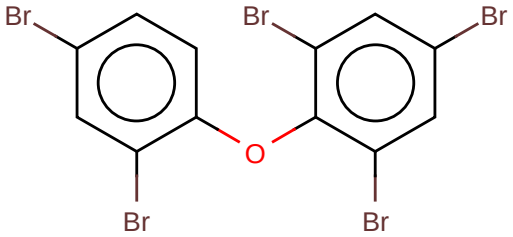
Name of Software	DRAGON Professional version 5.4
Supplier	Talete SRL
Supplier Address and Contact Details	Via V. Pisani, 13-20124 Milano, Italy. Tel/Fax: +39 02 66981300
Internet Address	http://www.talete.mi.it/dragon_exp.htm
Brief Description	DRAGON is an application for the calculation of molecular descriptors, intended for use in developing structure-activity relationships or structure-property relationships, in addition to similarity analysis and HTS of molecular databases. DRAGON has the ability to generate a large number of molecular descriptors (1664).
Platform	Windows on a PC
Approximate Cost	DRAGON is commercial software. Details on cost are available from the supplier.
Summary of Descriptor(s) Available Appropriate to Molecular Dimensions	Topological descriptors Edge adjacency indices Randic molecular descriptors Geometrical descriptors WHIM descriptors Other descriptors are available in the DRAGON package, but are not relevant to the parameterisation of molecular dimensions and are not included here.
Calculation Method(s)	2-D descriptors are calculated from standard methods. 3-D descriptors are calculated from geometry optimised structures (AM1 method)
Details of calculation method	The process used to create the data was as follows: 1. Geometry optimised structures were imported into DRAGON Professional version 5.4 in SDF format. 2. Descriptors groups to be calculated were selected from those available. 3. Descriptors were calculated. 4. Calculated values were saved as text files. 5. Text files were imported into Excel and redundant descriptors, not of use here, were removed. 6. A final descriptor list was saved as an Excel workbook.
Ease of Use. Speed of Calculation.	DRAGON is a simple and logical piece of software to use. Multiple input formats are supported including .Mol, .SDF, .SMI and .HIN file types. Descriptors are split into 20 distinct blocks in order to aid the user and minimise output size.

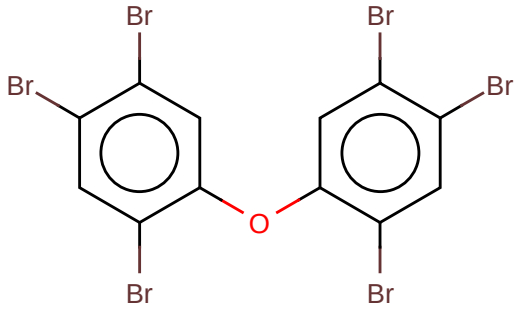
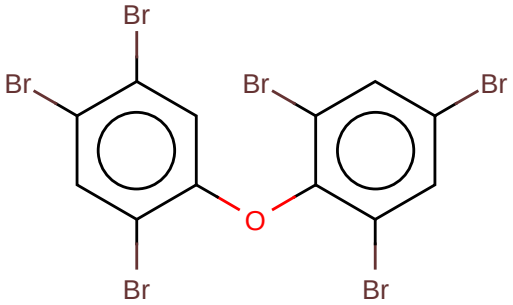
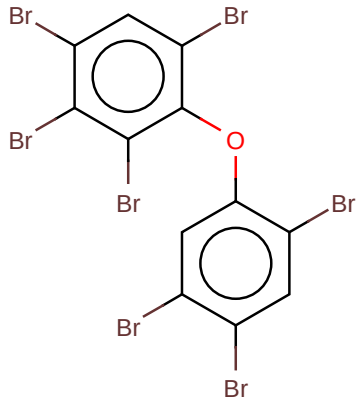
	Once selected, the desired descriptors are calculated extremely rapidly, taking no more than a few seconds. Even with much larger molecules and when calculating the full suite of available descriptors, this time rarely exceeds 1 minute. Calculated descriptors can then be saved as simple text files for subsequent use in Excel etc. The user can also select which blocks of descriptors to save in a desired file. DRAGON Professional can be used by any computer literate user.
Notes	No optimisation function is available. The software is dependent upon the structures given in the input file. Therefore, a suitable optimised 3D structure is required prior to use. There are occasions where DRAGON is unable to calculate certain descriptors for some compounds. These are usually 3D descriptors being calculated for particularly large compounds.
Descriptors Calculated Descriptors are listed in a separate file.	Total structure connectivity index 1-path Kier alpha-modified shape index 2-path Kier alpha-modified shape index 3-path Kier alpha-modified shape index Kier flexibility index Path/walk 2 - Randic shape index Path/walk 3 - Randic shape index Path/walk 4 - Randic shape index Path/walk 5 - Randic shape index Edge connectivity index of order 0 Edge connectivity index of order 1 Shape profile no. 1 Shape profile no. 2 Shape profile no. 3 Shape profile no. 4 Shape profile no. 5 Shape profile no. 6 Shape profile no. 7 Shape profile no. 8 Shape profile no. 9 Shape profile no. 10 Shape profile no. 11 Shape profile no. 12 Shape profile no. 13 Shape profile no. 14 Shape profile no. 15 Shape profile no. 16 Shape profile no. 17 Shape profile no. 18 Shape profile no. 19 Shape profile no. 20 Average shape profile index of order 2 3D-Wiener index 3D-Balaban index 3D-Harary index Radius of gyration (mass weighted) Span R

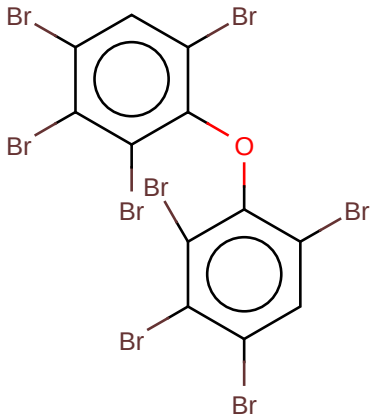
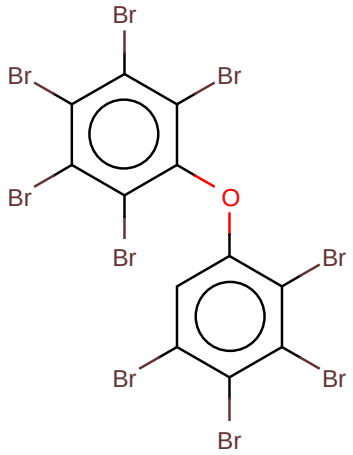
	Average span R Molecular eccentricity Spherosity Asphericity Folding degree index 3D Petitjean shape index Length-to-breadth ratio by WHIM 1st component size directional WHIM index / unweighted 2nd component size directional WHIM index / unweighted 3rd component size directional WHIM index / unweighted 1st component shape directional WHIM index / unweighted 2nd component shape directional WHIM index / unweighted T total size index / unweighted A total size index / unweighted K global shape index / unweighted V total size index / unweighted Further definition of the molecular descriptors calculated by DRAGON is given in the text: R. Todeschini and V. Consonni. Handbook of Molecular Descriptors. *Methods and Principles in Medicinal Chemistry*. Vol. 11.

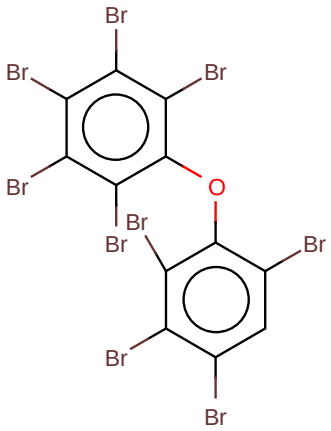
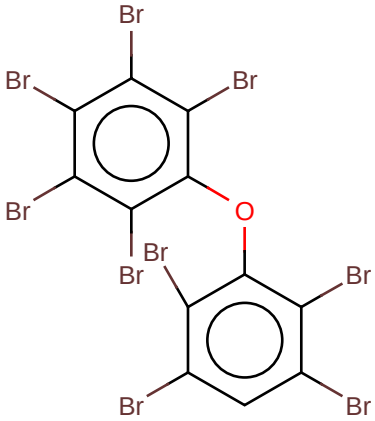
Appendix B – Structures, SMILES and Definitive Chiral SMILES

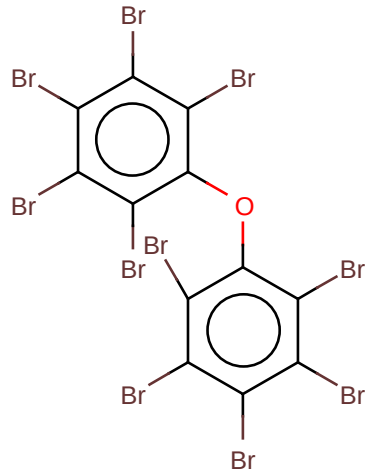
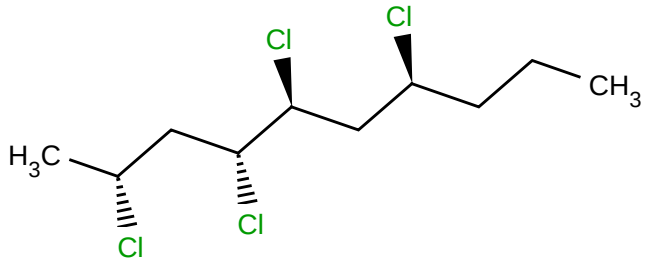
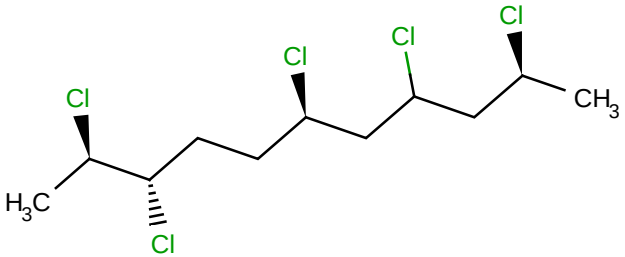
Table B1. Summary of the names and structures of the compounds considered for the calculation of molecular dimensions. (This information is available electronically)

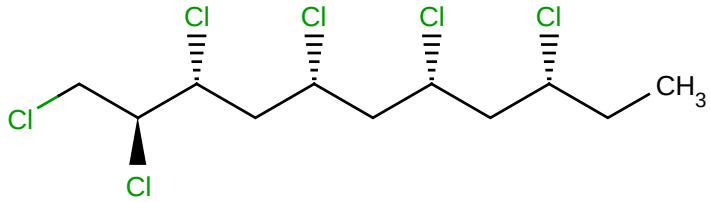
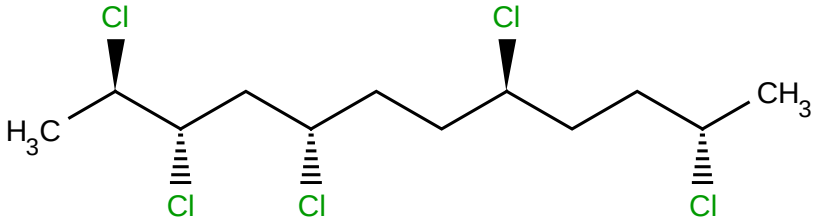
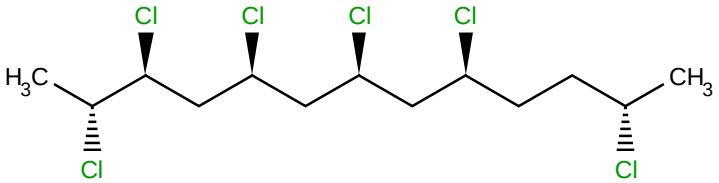
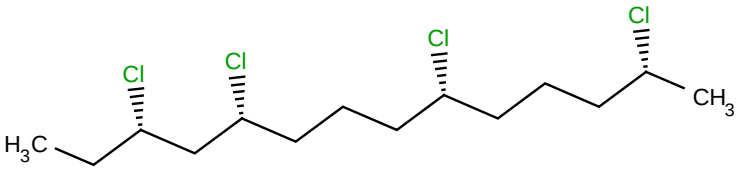
ID	Name (if known)	
Environment Agency Test Set		
1	2,2',4,4'-Tetrabromodiphenyl ether (BDE47)	
2	2,2',4,4',5-Pentabromodiphenyl ether (BDE99)	
3	2,2',4,4',6-Pentabromodiphenyl ether (BDE100)	

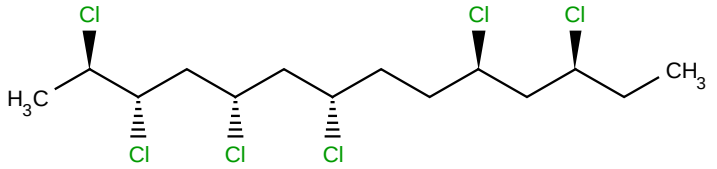
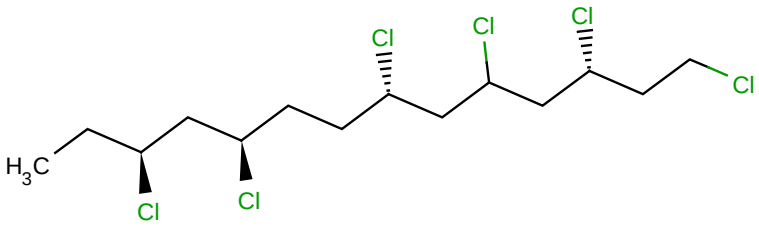
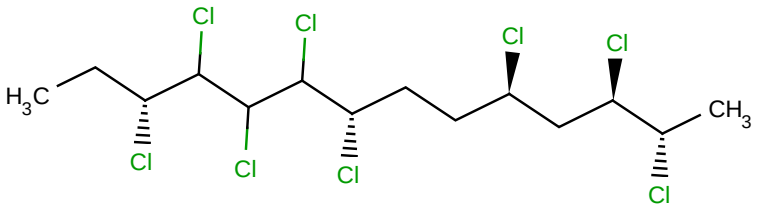
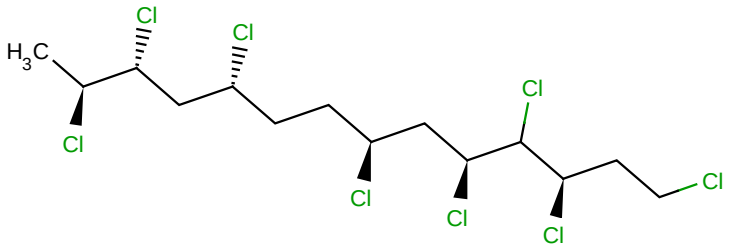
4	2,2',4,4',5,5'-Hexabromodiphenyl ether (BDE153)	
5	2,2',4,4',5,6'-Hexabromodiphenyl ether (BDE154)	
6	2,2',3,4,4',5',6-Heptabromodiphenyl ether (BDE183)	

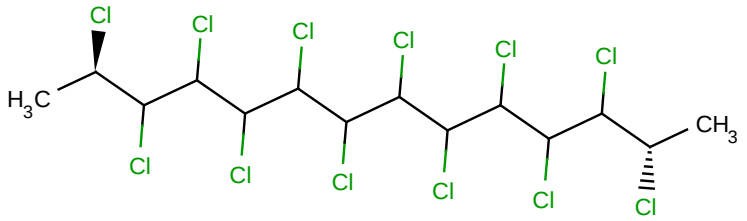
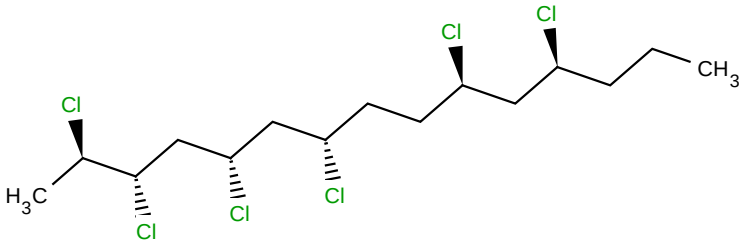
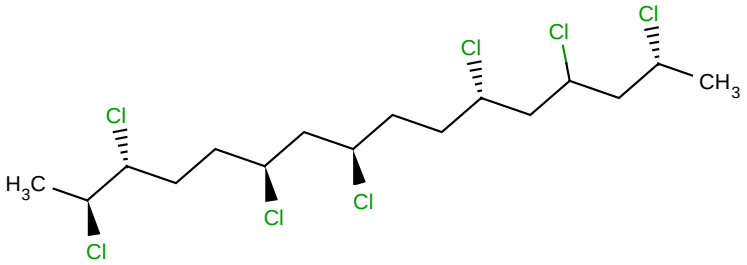
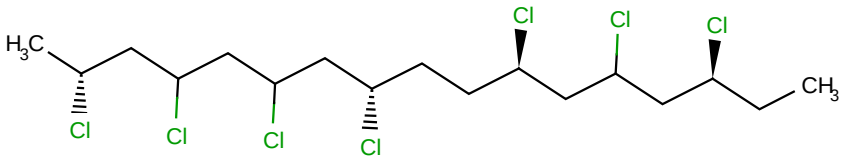
7	2,2',3,3',4,4',6,6'-Octabromodiphenyl ether (BDE197)	
8	2,2’,3,3’,4,4’,5,5’,6-Nonabromodiphenyl ether (BDE206)	

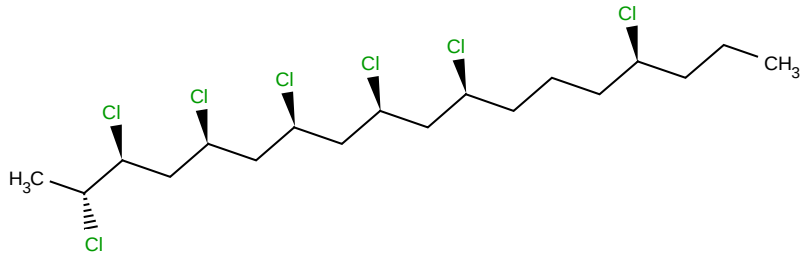
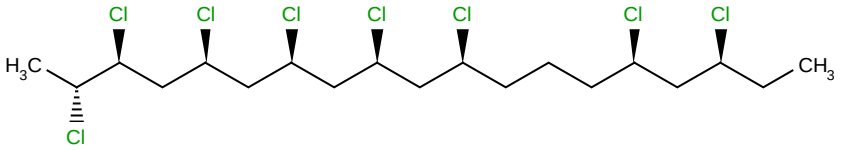
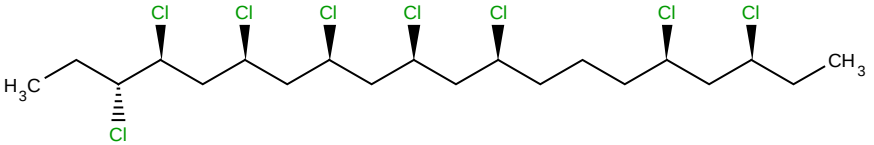
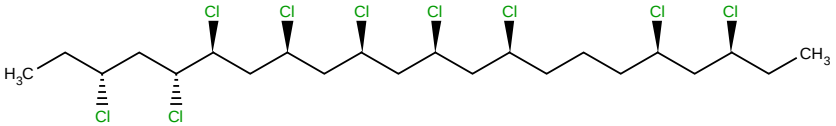
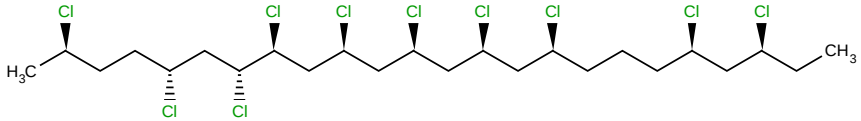
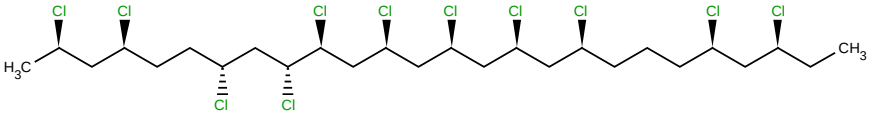
9	2,2’,3,3’,4,4’,5,6,6’-Nonabromodiphenyl ether (BDE207)	
10	BDE208 = 2,2’,3,3’,4,5,5’,6,6’-Nonabromodiphenyl ether (BDE208)	

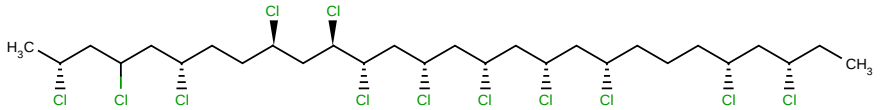
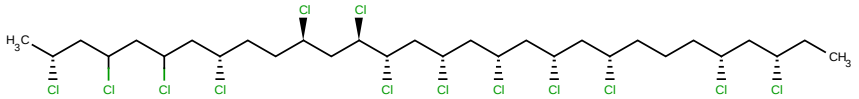
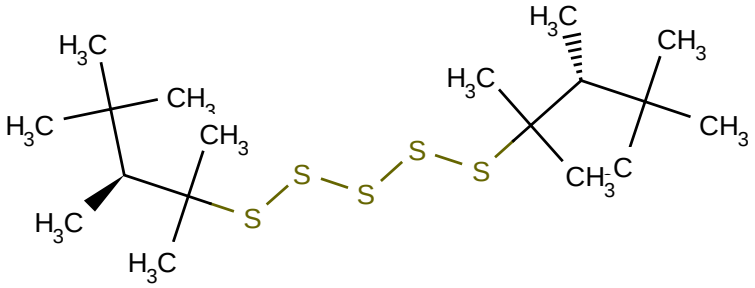
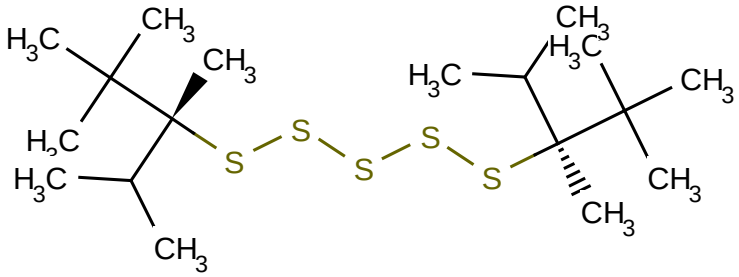
11	Decabromodiphenyl ether	
12	$C_{10}H_{18}Cl_4$ (51.7% wt. Cl) – component of short chain chlorinated paraffin	
13	$C_{11}H_{19}Cl_5$ (54.0% wt Cl)	

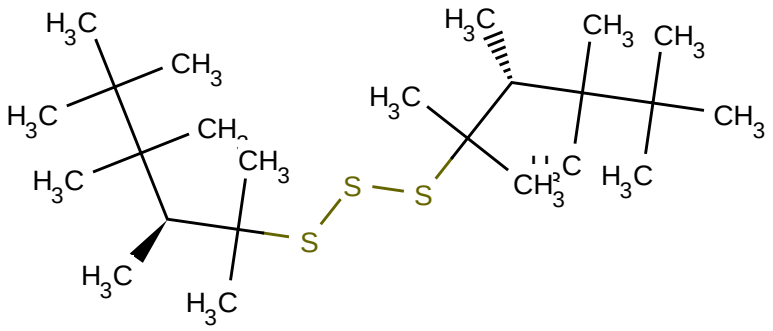
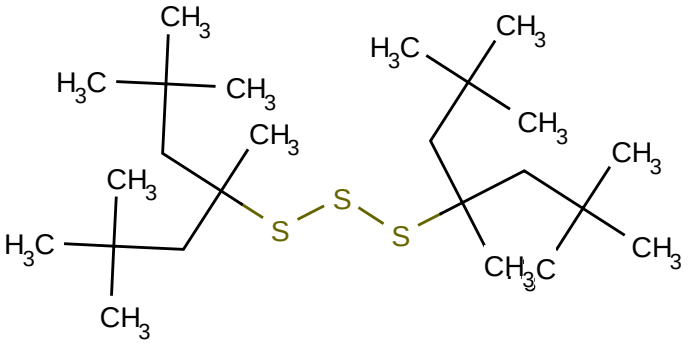
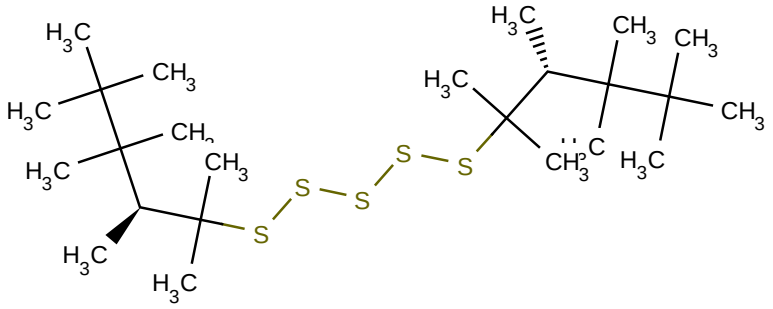
14	$C_{11}H_{18}Cl_6$ (58.7% wt. Cl)	
15	$C_{12}H_{21}Cl_5$ (51.8% wt. Cl)	
16	$C_{13}H_{22}Cl_6$ (54.5% wt. Cl)	
17	$C_{14}H_{26}Cl_4$ (42.3% wt. Cl) – component of medium chain chlorinated paraffin	

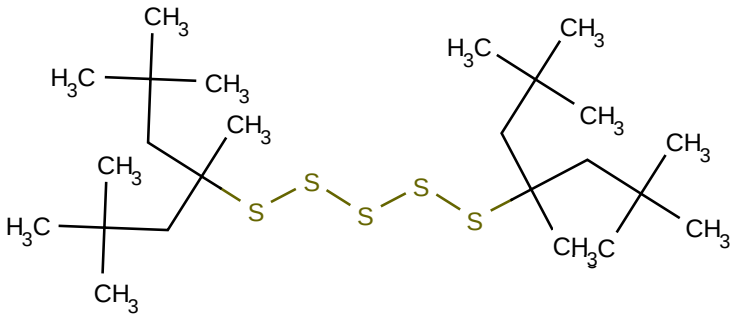
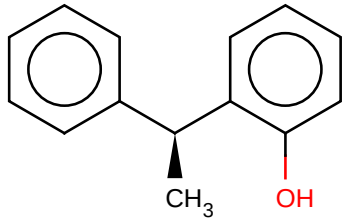
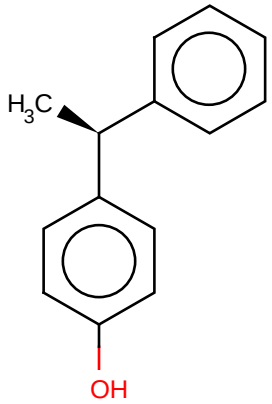
18	$C_{14}H_{24}Cl_6$ (52.6% wt. Cl)	
19	$C_{14}H_{24}Cl_6$ (52.6% wt. Cl)	
20	$C_{14}H_{22}Cl_8$ (59.9% wt. Cl)	
21	$C_{14}H_{22}Cl_8$ (59.9% wt. Cl)	

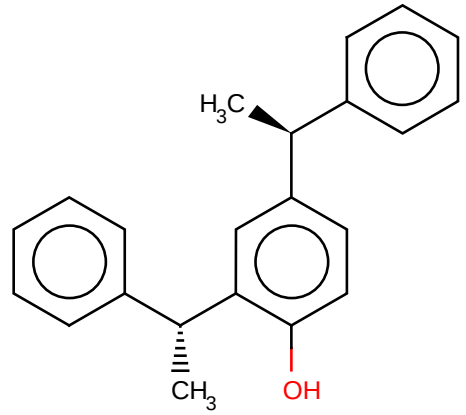
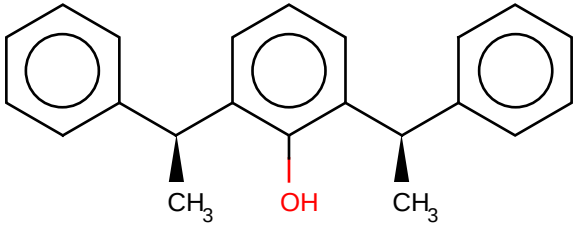
22	$C_{14}H_{18}Cl_{12}$ (69.9% wt.Cl)	
23	$C_{15}H_{26}Cl_6$ (50.8% wt. Cl)	
24	$C_{16}H_{27}Cl_7$ (53.2% wt. Cl)	
25	$C_{17}H_{29}Cl_7$ (51.6% wt. Cl)	

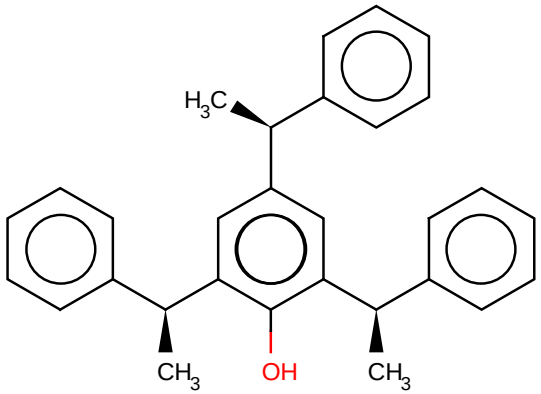
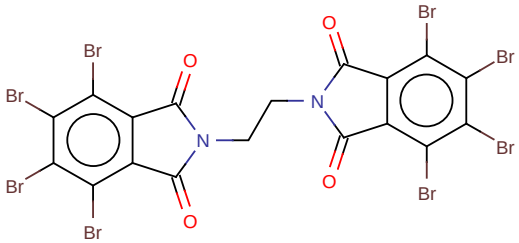
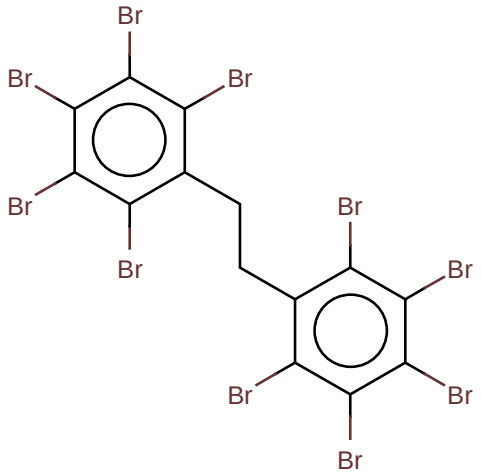
26	$C_{18}H_{31}Cl_7$ (50.2% wt. Cl) – component of long chain chlorinated paraffin	
27	$C_{19}H_{32}Cl_8$ (52.2% wt. Cl)	
28	$C_{20}H_{34}Cl_8$ (50.9% wt. Cl)	
29	$C_{22}H_{37}Cl_9$ (51.5% wt. Cl)	
30	$C_{24}H_{40}Cl_{10}$ (52.0% wt. Cl)	
31	$C_{26}H_{43}Cl_{11}$ (52.4% wt. Cl)	

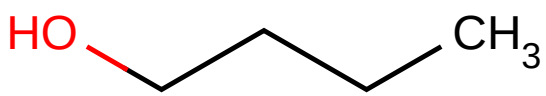
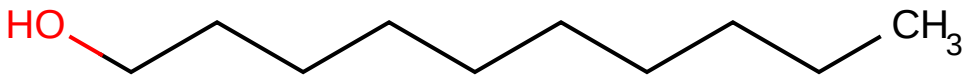
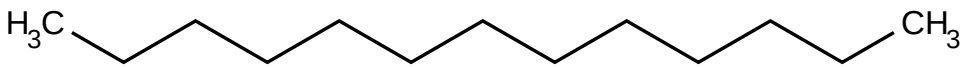
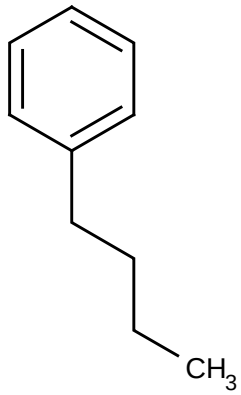
32	$C_{28}H_{46}Cl_{12}$ (52.7% wt. Cl)	
33	$C_{30}H_{49}Cl_{13}$ (53.0% wt. Cl)	
34	Di-(tert-nonyl) polysulphide	
35	Di-(tert-nonyl) polysulphide v2	

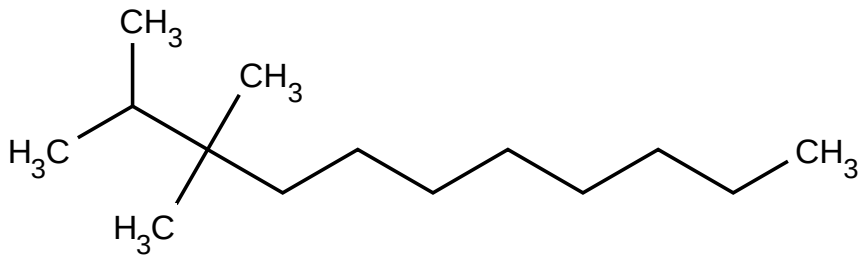
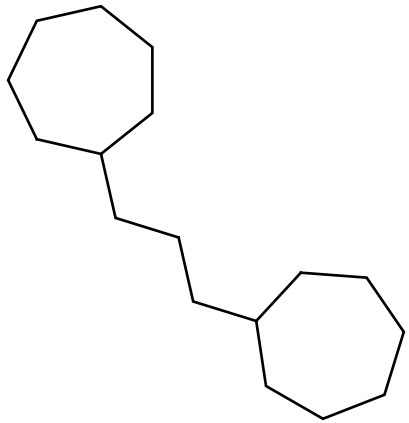
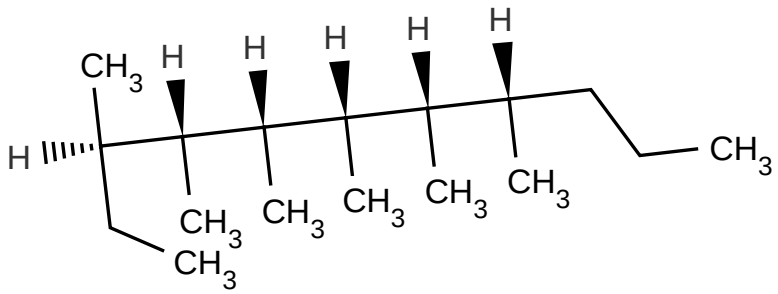
36	Di-(tert-dodecyl) polysulphide	
37	Di-(tert-dodecyl) polysulphide v2	
38	Di-(*tert*-dodecyl) pentasulphide	

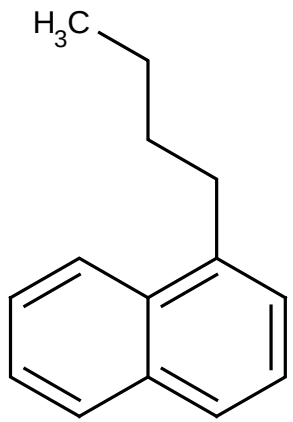
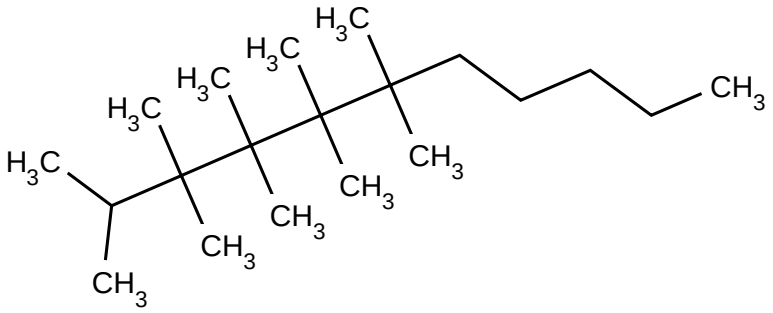
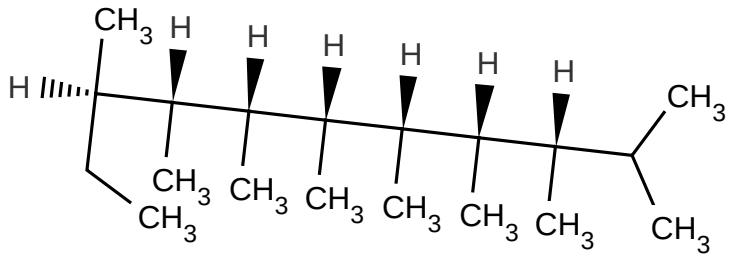
39	Di-(*tert*-dodecyl) pentasulphide v2	
40	2-styrylphenol	
41	4-styrylphenol	

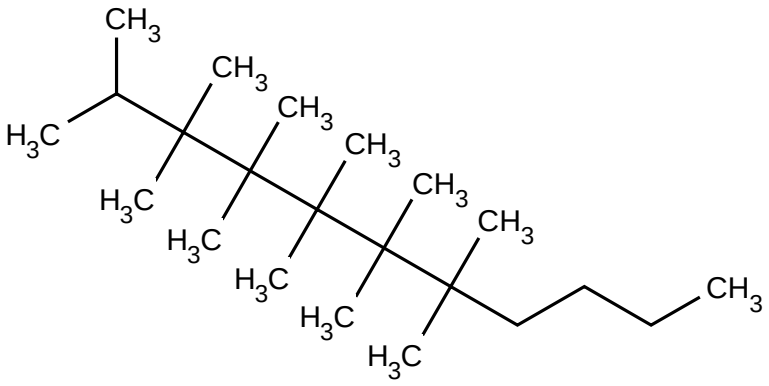
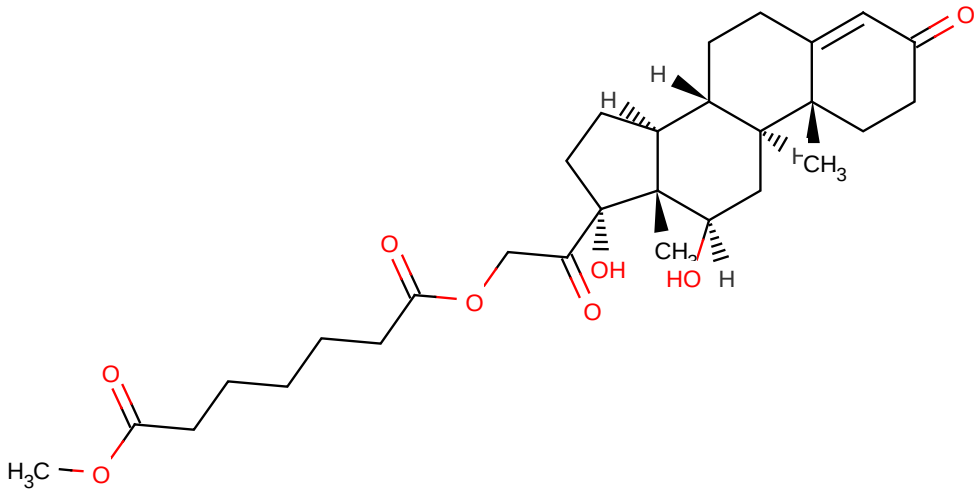
42	2,4-Distyryl phenol	
43	2,6-Distyrylphenol	

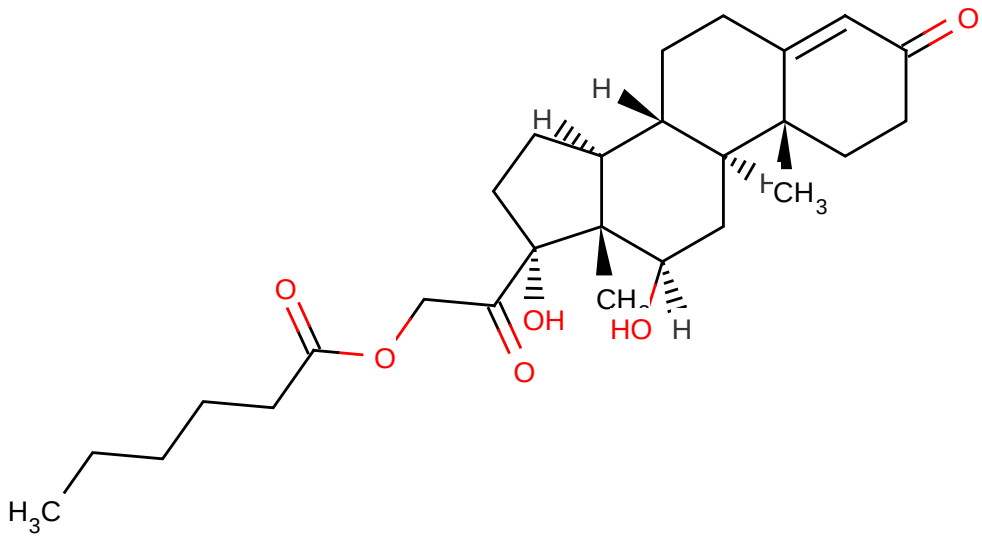
44	2,4,6-Tristyrylphenol	
45	Ethylene bistetrabromophthalimide (1H-Isoindole-1,3(2H)-dione, 2,2'-(1,2-ethanediyl)bis 4,5,6,7-tetrabromo-)	
46	1,2-Bis(pentabromophenyl) ethane	

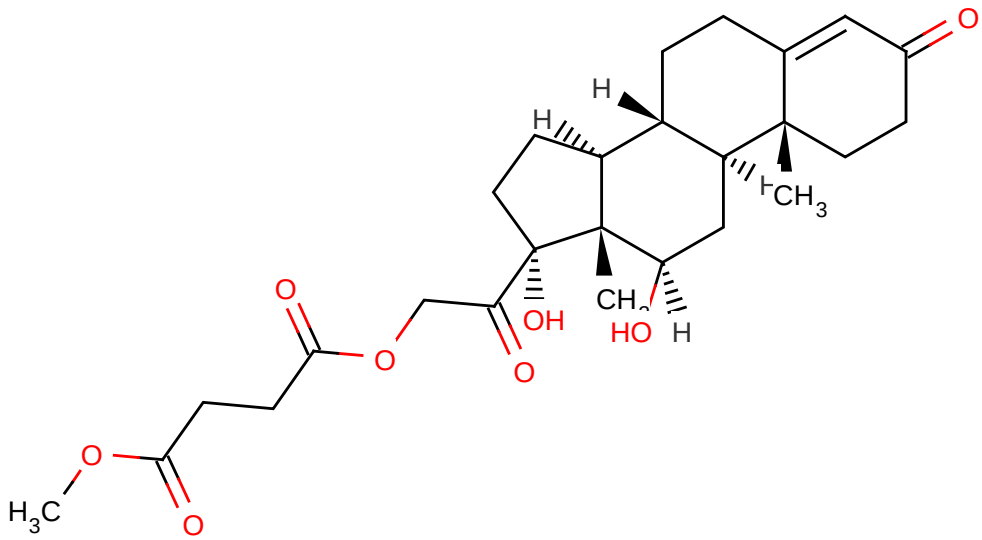
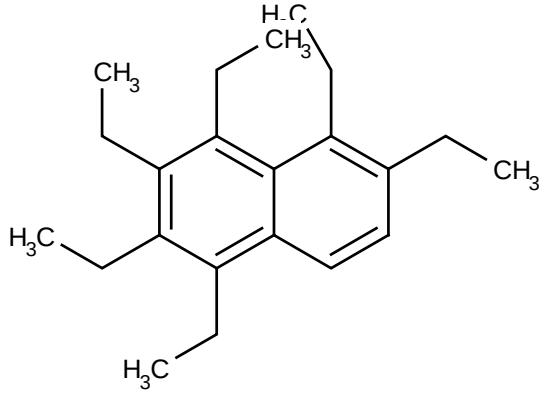
Reference Chemicals		
R13	Butanol	
R20	Decanol	
R26	Tridecane	
R66	Butylbenzene	

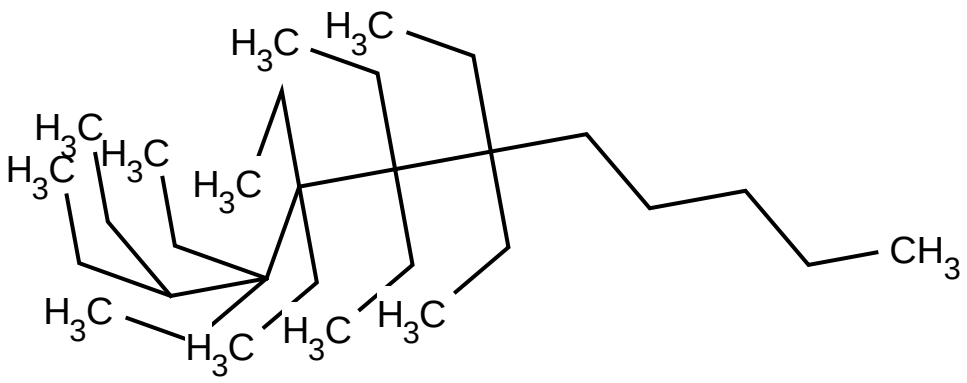
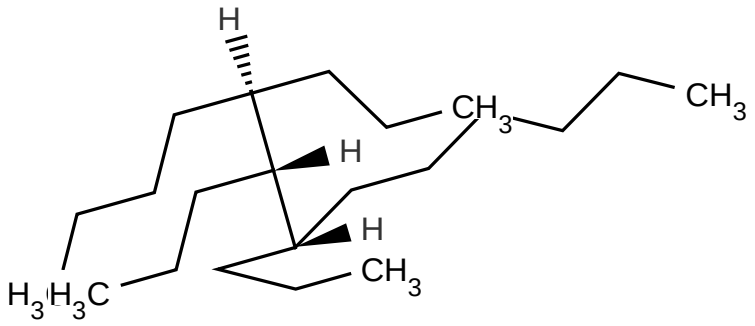
R68	2-methyl-3-dimethyl-undecance	
R91	Reference Compound 91	
R103	Reference Compound 103	

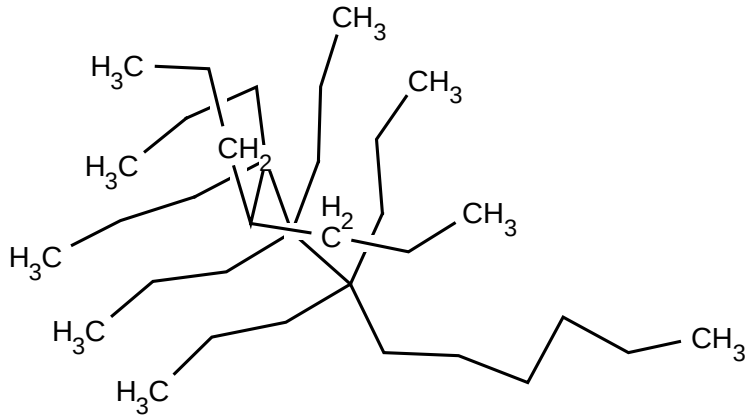
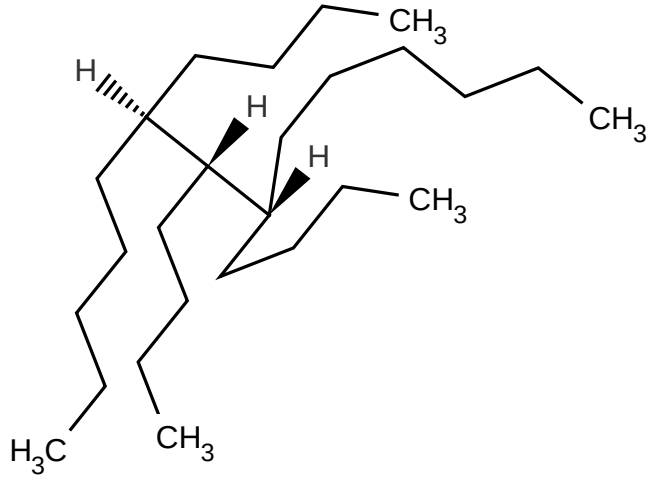
R111	Reference Compound 111	
R118	Reference Compound 118	
R120	Reference Compound 120	

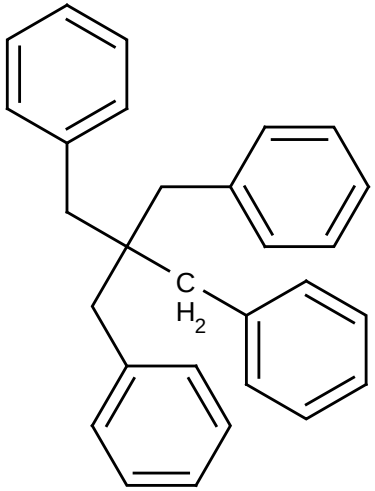
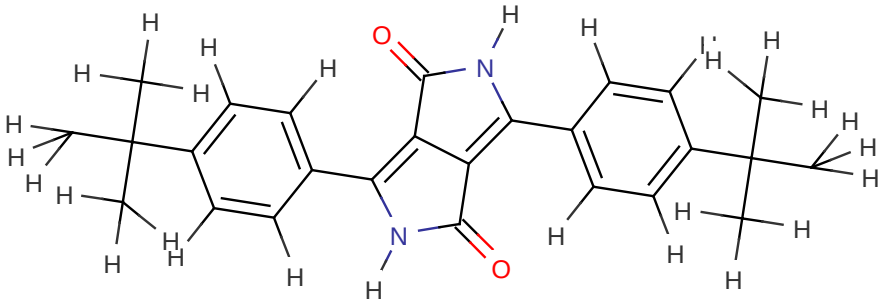
R124	Reference Compound 124	
R130	Reference Compound 130	

R135	Reference Compound 135	

R141	Reference Compound 141	
R165	Reference Compound 165	

R172	Reference Compound 172	
R173	Reference Compound 173	

R190	Reference Compound 190	
R191	Reference Compound 191	

R198	Reference Compound 198	
Ciba Dyes		
D1	CI Pigment Orange 73	

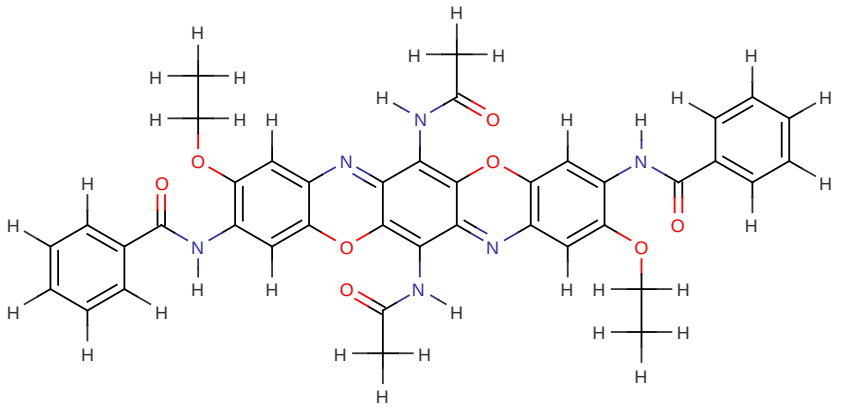
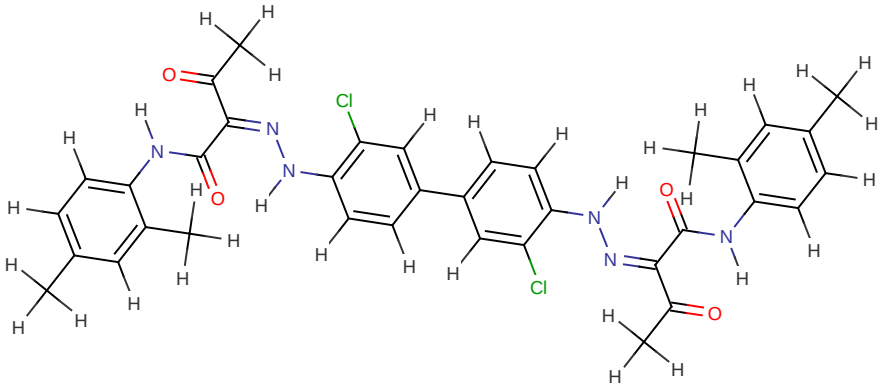
D2	CI Pigment Violet 37	
D3	CI Pigment Yellow 13	

Table B2 Summary of the names, SMILES and chiral SMILES of the compounds considered for the calculation of molecular dimensions. (This information is available electronically)

ID	Name	Original SMILES	Chiral SMILES
Environment Agency Test Set			
1	2,2',4,4'-Tetrabromodiphenyl ether (BDE47)	c1(Br)cc(Br)ccc1Oc2c(Br)cc(Br)cc2	BrC1=CC=C(OC2=C(Br)C=C(Br)C=C2)C(Br)=C1
2	2,2',4,4',5-Pentabromodiphenyl ether (BDE99)	c1(Br)cc(Br)c(Br)cc1Oc2c(Br)cc(Br)cc2	BrC1=CC=C(OC2=C(Br)C=C(Br)C(Br)=C2)C(Br)=C1
3	2,2',4,4',6-Pentabromodiphenyl ether (BDE100)	c1(Br)cc(Br)cc(Br)c1Oc2c(Br)cc(Br)cc2	BrC1=CC=C(OC2=C(Br)C=C(Br)C=C2Br)C(Br)=C1
4	2,2',4,4',5,5'-Hexabromodiphenyl ether (BDE153)	c1(Br)cc(Br)c(Br)cc1Oc2c(Br)cc(Br)c(Br)c2	BrC1=CC(Br)=C(OC2=CC(Br)=C(Br)C=C2Br)C=C1Br
5	2,2',4,4',5,6'-Hexabromodiphenyl ether (BDE154)	c1(Br)cc(Br)cc(Br)c1Oc2c(Br)cc(Br)c(Br)c2	BrC1=CC(Br)=C(OC2=CC(Br)=C(Br)C=C2Br)C(Br)=C1
6	2,2',3,4,4',5',6-Heptabromodiphenyl ether (BDE183)	c1(Br)cc(Br)c(Br)c(Br)c1Oc2c(Br)cc(Br)c(Br)c2	BrC1=CC(Br)=C(Br)C=C1OC2=C(Br)C=C(Br)C(Br)=C2Br
7	2,2',3,3',4,4',6,6'-Octabromodiphenyl ether (BDE197)	c1(Br)cc(Br)c(Br)c(Br)c1Oc2c(Br)cc(Br)c(Br)c2(Br)	BrC1=CC(Br)=C(OC2=C(Br)C(Br)=C(Br)C=C2Br)C(Br)=C1Br
8	2,2',3,3',4,4',5,5',6-Nonabromodiphenyl ether (BDE206)	c1(Br)c(Br)c(Br)c(Br)c(Br)c1Oc2cc(Br)c(Br)c(Br)c2(Br)	BrC1=CC(OC2=C(Br)C(Br)=C(Br)C(Br)=C2Br)=C(Br)C(Br)=C1Br
9	2,2',3,3',4,4',5,6,6'-Nonabromodiphenyl ether (BDE207)	c1(Br)c(Br)c(Br)c(Br)c(Br)c1Oc2c(Br)cc(Br)c(Br)c2(Br)	BrC1=CC(Br)=C(Br)C(Br)=C1OC2=C(Br)C(Br)=C(Br)C(Br)=C2Br
10	BDE208 = 2,2',3,3',4,5,5',6,6'-Nonabromodiphenyl ether (BDE208)	c1(Br)c(Br)c(Br)c(Br)c(Br)c1Oc2c(Br)c(Br)cc(Br)c2(Br)	BrC1=CC(Br)=C(Br)C(OC2=C(Br)C(Br)=C(Br)C(Br)=C2Br)=C1Br
11	Decabromodiphenyl ether	c1(Br)c(Br)c(Br)c(Br)c(Br)c1Oc2c(Br)c(Br)c(Br)c(Br)c2(Br)	BrC1=C(Br)C(Br)=C(OC2=C(Br)C(Br)=C(Br)C(Br)=C2Br)C(Br)=C1Br
12	$C_{10}H_{18}Cl_4$ (51.7% wt. Cl)	CC(Cl)CC(Cl)C(Cl)CC(Cl)CCC	[H][C@](C)(Cl)C[C@@]([H])(Cl)[C@@]([H])(Cl)C[C@@]([H])(Cl)CCC
13	$C_{11}H_{19}Cl_5$ (54.0% wt Cl)	CC(Cl)C(Cl)CCC(Cl)CC(Cl)CC(Cl)C	[H][C@@](C)(Cl)C[C@]([H])(Cl)C[C@]([H])(Cl)CC[C@]([H])(Cl)[C@@]([H])(C)Cl
14	$C_{11}H_{18}Cl_6$ (58.7% wt. Cl)	CCC(Cl)CC(Cl)CC(Cl)CC(Cl)C(Cl)C(Cl)	[H][C@@](Cl)(CC)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)[C@@]([H])(Cl)CCl
15	$C_{12}H_{21}Cl_5$ (51.8% wt. Cl)	CC(Cl)C(Cl)CC(Cl)CCC(Cl)CCC(Cl)C	[H][C@@](C)(Cl)CC[C@]([H])(Cl)CC[C@]([H])(Cl)C[C@]([H])(Cl)[C@@]([H])(C)Cl
16	$C_{13}H_{22}Cl_6$ (54.5% wt. Cl)	CC(Cl)C(Cl)CC(Cl)CC(Cl)CC(Cl)CCC(Cl)C	[H][C@@](C)(Cl)CC[C@]([H])(Cl)C[C@]([H])(Cl)C[C@]([H])(Cl)C[C@]([H])(Cl)[C@@]([H])(C)Cl

17	$C_{14}H_{26}Cl_4$ (42.3% wt. Cl)	CC(Cl)CCCC(Cl)CCCC(Cl)CC(Cl)CC	[H][C@](C)(Cl)CCC[C@@]([H])(Cl)CCC[C@@]([H])(Cl)C[C@@]([H])(Cl)CC
18	$C_{14}H_{24}Cl_6$ (52.6% wt. Cl)	CC(Cl)C(Cl)CC(Cl)CC(Cl)CCC(Cl)CC(Cl)CC	[H][C@](C)(Cl)[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)CC[C@@]([H])(Cl)C[C@@]([H])(Cl)CC
19	$C_{14}H_{24}Cl_6$ (52.6% wt. Cl)	C(Cl)CC(Cl)CC(Cl)CC(Cl)CCC(Cl)CC(Cl)CC	[H][C@](Cl)(CC)C[C@]([H])(Cl)CC[C@]([H])(Cl)C[C@]([H])(Cl)C[C@]([H])(Cl)CCCl
20	$C_{14}H_{22}Cl_8$ (59.9% wt. Cl)	CCC(Cl)C(Cl)C(Cl)C(Cl)C(Cl)CCC(Cl)CC(Cl)C(Cl)C	[H][C@@](C)(Cl)[C@]([H])(Cl)C[C@]([H])(Cl)CC[C@]([H])(Cl)[C@]([H])(Cl)[C@]([H])(Cl)[C@]([H])(Cl)CC
21	$C_{14}H_{22}Cl_8$ (59.9% wt. Cl)	C(Cl)CC(Cl)C(Cl)C(Cl)CC(Cl)CCC(Cl)CC(Cl)C(Cl)C	[H][C@@](C)(Cl)[C@]([H])(Cl)C[C@]([H])(Cl)CC[C@]([H])(Cl)C[C@]([H])(Cl)[C@]([H])(Cl)[C@]([H])(Cl)CCCl
22	$C_{14}H_{18}Cl_{12}$ (69.9% wt.Cl)	CC(Cl)C(Cl)C(Cl)C(Cl)C(Cl)C(Cl)C(Cl)C(Cl)C(Cl)C(Cl)C(Cl)C(Cl)C(Cl)C	[H][C@](C)(Cl)[C@@]([H])(Cl)[C@@]([H])(Cl)[C@@]([H])(Cl)[C@@]([H])(Cl)[C@@]([H])(Cl)[C@@]([H])(Cl)[C@@]([H])(Cl)[C@@]([H])(Cl)[C@@]([H])(Cl)[C@]([H])(C)Cl
23	$C_{15}H_{26}Cl_6$ (50.8% wt. Cl)	CC(Cl)C(Cl)CC(Cl)CC(Cl)CCC(Cl)CC(Cl)CCC	[H][C@](C)(Cl)[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)CC[C@@]([H])(Cl)C[C@@]([H])(Cl)CCC
24	$C_{16}H_{27}Cl_7$ (53.2% wt. Cl)	CC(Cl)CC(Cl)CC(Cl)CCC(Cl)CC(Cl)CCC(Cl)C(Cl)C	[H][C@](C)(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)CC[C@@]([H])(Cl)C[C@@]([H])(Cl)CC[C@@]([H])(Cl)[C@]([H])(C)Cl
25	$C_{17}H_{29}Cl_7$ (51.6% wt. Cl)	CC(Cl)CC(Cl)CC(Cl)CC(Cl)CCC(Cl)CC(Cl)CC(Cl)CC	[H][C@](C)(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)CC[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)CC
26	$C_{18}H_{31}Cl_7$ (50.2% wt. Cl)	CC(Cl)C(Cl)CC(Cl)CC(Cl)CC(Cl)CC(Cl)CCCC(Cl)CCC	[H][C@](C)(Cl)[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)CCC[C@@]([H])(Cl)CCC
27	$C_{19}H_{32}Cl_8$ (52.2% wt. Cl)	CC(Cl)C(Cl)CC(Cl)CC(Cl)CC(Cl)CC(Cl)CCCC(Cl)CC(Cl)CC	[H][C@](C)(Cl)[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)CCC[C@@]([H])(Cl)C[C@@]([H])(Cl)CC
28	$C_{20}H_{34}Cl_8$ (50.9% wt. Cl)	CCC(Cl)C(Cl)CC(Cl)CC(Cl)CC(Cl)CC(Cl)CCCC(Cl)CC(Cl)CC	[H][C@](Cl)(CC)C[C@]([H])(Cl)CCC[C@]([H])(Cl)C[C@]([H])(Cl)C[C@]([H])(Cl)C[C@]([H])(Cl)C[C@]([H])(Cl)C[C@]([H])(Cl)CC
29	$C_{22}H_{37}Cl_9$ (51.5% wt. Cl)	CCC(Cl)CC(Cl)C(Cl)CC(Cl)CC(Cl)CC(Cl)CC(Cl)CCCC(Cl)CC(Cl)CC	[H][C@](Cl)(CC)C[C@]([H])(Cl)CCC[C@]([H])(Cl)C[C@]([H])(Cl)C[C@]([H])(Cl)C[C@]([H])(Cl)C[C@]([H])(Cl)C[C@]([H])(Cl)C[C@]([H])(Cl)CC

30	$C_{24}H_{40}Cl_{10}$ (52.0% wt. Cl)	CC(Cl)CCC(Cl)CC(Cl)C(Cl)CC(Cl)CC(Cl)CC(Cl)CC(Cl)CCCC(Cl)CC(Cl)CC	[H][C@](C)(Cl)CC[C@@]([H])(Cl)C[C@@]([H])(Cl)[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)CCC[C@@]([H])(Cl)C[C@@]([H])(Cl)CC
31	$C_{26}H_{43}Cl_{11}$ (52.4% wt. Cl)	CC(Cl)CC(Cl)CCC(Cl)CC(Cl)C(Cl)CC(Cl)CC(Cl)CC(Cl)CC(Cl)CCCC(Cl)CC(Cl)CC	[H][C@](C)(Cl)C[C@@]([H])(Cl)CC[C@@]([H])(Cl)C[C@@]([H])(Cl)[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)CCC[C@@]([H])(Cl)C[C@@]([H])(Cl)CC
32	$C_{28}H_{46}Cl_{12}$ (52.7% wt. Cl)	CC(Cl)CC(Cl)CC(Cl)CCC(Cl)CC(Cl)C(Cl)CC(Cl)CC(Cl)CC(Cl)CC(Cl)CCCC(Cl)CC(Cl)CC	[H][C@](C)(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)CC[C@@]([H])(Cl)C[C@@]([H])(Cl)[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)CCC[C@@]([H])(Cl)C[C@@]([H])(Cl)CC
33	$C_{30}H_{49}Cl_{13}$ (53.0% wt. Cl)	CC(Cl)CC(Cl)CC(Cl)CC(Cl)CCC(Cl)CC(Cl)C(Cl)CC(Cl)CC(Cl)CC(Cl)CC(Cl)CCCC(Cl)CC(Cl)CC	[H][C@](C)(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)CC[C@@]([H])(Cl)C[C@@]([H])(Cl)[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)C[C@@]([H])(Cl)CCC[C@@]([H])(Cl)C[C@@]([H])(Cl)CC
34	Di-(tert-nonyl) polysulphide	CC(C)(C)C(C)C(C)(C)SSSSSC(C)(C)C(C)C(C)(C)C	[H][C@](C)(C(C)(C)C)C(C)(C)SSSSSC(C)(C)[C@]([H])(C)C(C)(C)C
35	Di-(tert-nonyl) polysulphide v2	CC(C)(C)C(C)(C(C)C)SSSSSC(C)(C(C)C)C(C)(C)C	CC(C)[C@@](C)(SSSSS[C@@](C)(C(C)C)C(C)(C)C)C(C)(C)C
36	Di-(tert-dodecyl) polysulphide	CC(C)(C)C(C)(C)C(C)C(C)(C)SSSC(C)(C)C(C)C(C)(C)C(C)(C)C	[H][C@](C)(C(C)(C)SSSC(C)(C)[C@@]([H])(C)C(C)(C)C(C)(C)C)C(C)(C)C(C)(C)C
37	Di-(tert-dodecyl) polysulphide v2	CC(C)(C)CC(C)(CC(C)(C)(C))SSSC(C)(CC(C)(C)(C))CC(C)(C)(C)	CC(C)(C)CC(C)(CC(C)(C)C)SSSC(C)(CC(C)(C)C)CC(C)(C)C
38	Di-(*tert*-dodecyl) pentasulphide	CC(C)(C)C(C)(C)C(C)C(C)(C)SSSSSC(C)(C)C(C)C(C)(C)C(C)(C)C	[H][C@](C)(C(C)(C)SSSSSC(C)(C)[C@@]([H])(C)C(C)(C)C(C)(C)C)C(C)(C)C(C)(C)C(C)(C)C
39	Di-(*tert*-dodecyl) pentasulphide v2	CC(C)(C)CC(C)(CC(C)(C)(C))SSSSSC(C)(CC(C)(C)(C))CC(C)(C)(C)	CC(C)(C)CC(C)(CC(C)(C)C)SSSSSC(C)(CC(C)(C)C)CC(C)(C)C
40	2-styrylphenol	Oc1ccccc1C(C)c2ccccc2	[H][C@](C)(C1=CC=CC=C1)C2=C(O)C=CC=C2
41	4-styrylphenol	Oc1ccc(cc1)C(C)c2ccccc2	[H][C@](C)(C1=CC=CC=C1)C2=CC=C(O)C=C2
42	2,4-Distyryl phenol	c1ccccc1C(C)c2c(O)ccc(c2)C(C)c3ccccc3	[H][C@](C)(C1=CC=CC=C1)C2=CC=C(O)C(=C2)[C@]([H])(C)C3=CC=CC=C3
43	2,6-Distyrylphenol	c1ccccc1C(C)c2cccc(c2O)C(C)c3ccccc3	[H][C@](C)(C1=CC=CC=C1)C2=CC=CC(=C2O)[C@]([H])(C)C3=CC=CC=C3
44	2,4,6-Tristyrylphenol	c1ccccc1C(C)c2c(O)c(C(C)c3ccccc3)cc(c2)C(C)c4ccccc4	[H][C@](C)(C1=CC=CC=C1)C2=CC(=C(O)C(=C2)[C@]([H])(C)C3=CC=CC=C3)[C@@]([H])(C)C4=CC=CC=C4

45	Ethylene bistetrabromophthalimide (1H-Isoindole-1,3(2H)-dione, 2,2'-(1,2-ethanediyl)bis 4,5,6,7-tetrabromo-)	O=C(N(C(=O)c1c(c(c(c2Br)Br)Br)Br)CCN(C(=O)c(c3c(c(c4Br)Br)Br)c4Br)C3=O)c12	BrC1=C2C(=O)N(CCN3C(=O)C4=C(C3=O)C(Br)=C(Br)C(Br)=C4Br)C(=O)C2=C(Br)C(Br)=C1Br
46	1,2-Bis(pentabromophenyl) ethane	C(c(c(c(c(c1Br)Br)Br)Br)c1Br)Cc(c(c(c(c2Br)Br)Br)Br)c2Br	BrC1=C(Br)C(Br)=C(CCC2=C(Br)C(Br)=C(Br)C(Br)=C2Br)C(Br)=C1Br
Reference Chemicals			
R13		CCCCO	CCCCO
R20		CCCCCCCCCCO	CCCCCCCCCCO
R26		CCCCCCCCCCCCC	CCCCCCCCCCCCC
R66		CCCCc1ccccc1	CCCCC1=CC=CC=C1
R68		CC(C)C(C)(C)CCCCCCCC	CCCCCCCCC(C)(C)C(C)C
R91		C1CCCCCC1CCCC2CCCCCC2	C(CC1CCCCCC1)CC2CCCCCC2
R103		CCC(C)C(C)C(C)C(C)C(C)C(C)CCC	[H][C@@](C)(CC)[C@@]([H])(C)[C@@]([H])(C)[C@@]([H])(C)[C@@]([H])(C)[C@@]([H])(C)CCC
R111		CCCCc1cccc2ccccc12	CCCCC1=C2C=CC=CC2=CC=C1
R118		CC(C)C(C)(C)C(C)(C)C(C)(C)C(C)(C)CCCCC	CCCCCC(C)(C)C(C)(C)C(C)(C)C(C)(C)C(C)C
R120		CCC(C)C(C)C(C)C(C)C(C)C(C)C(C)C(C)C	[H][C@@](C)(CC)[C@@]([H])(C)[C@@]([H])(C)[C@@]([H])(C)[C@@]([H])(C)[C@@]([H])(C)[C@@]([H])(C)C(C)C
R124		CC(C)C(C)(C)C(C)(C)C(C)(C)C(C)(C)C(C)(C)CCCC	CCCCC(C)(C)C(C)(C)C(C)(C)C(C)(C)C(C)(C)C(C)C
R130		COC(=O)CCCCCC(=O)OCC(=O)C3(O)CCC4C2CCC1=CC(=O)CCC1(C)C2CC(O)C34C	[H][C@@]12CC[C@](O)(C(=O)COC(=O)CCCCCC(=O)OC)[C@@]1(C)[C@]([H])(O)C[C@@]3([H])[C@@]2([H])CCC4=CC(=O)CC[C@]34C
R135		CCCCCC(=O)OCC(=O)C3(O)CCC4C2CCC1=CC(=O)CCC1(C)C2CC(O)C34C	[H][C@@]12CC[C@](O)(C(=O)COC(=O)CCCCC)[C@@]1(C)[C@]([H])(O)C[C@@]3([H])[C@@]2([H])CCC4=CC(=O)CC[C@]34C
R141		COC(=O)CCC(=O)OCC(=O)C3(O)CCC4C2CCC1=CC(=O)CCC1(C)C2CC(O)C34C	[H][C@@]12CC[C@](O)(C(=O)COC(=O)CCC(=O)OC)[C@@]1(C)[C@]([H])(O)C[C@@]3([H])[C@@]2([H])CCC4=CC(=O)CC[C@]34C
R165		CCc1c(CC)c(CC)c(CC)c2c(CC)c(CC)ccc12	CCC1=C(CC)C2=C(CC)C(CC)=C(CC)C(CC)=C2C=C1
R172		CCC(CC)C(CC)(CC)C(CC)(CC)C(CC)(CC)C(CC)(CC)CCCCC	CCCCCC(CC)(CC)C(CC)(CC)C(CC)(CC)C(CC)(CC)C(CC)CC
R173		CCCCC(CCC)C(CCC)C(CCC)CCCCCC	[H][C@@](CCC)(CCCC)[C@@]([H])(CCC)[C@@]([H])(CCC)CCCCCC

R190		CCCC(CCC)C(CCC)(CCC)C(CCC)(CCC)C(CCC)(CCC)CCCCCC	CCCCCCC(CCC)(CCC)C(CCC)(CCC)C(CCC)(CCC)C(CCC)CCC
R191		CCCCCC(CCCC)C(CCCC)C(CCCC)CCCCCC	[H][C@@](CCCC)(CCCCC)[C@@]([H])(CCCC)[C@@]([H])(CCCC)CCCCCC
R198		c1ccccc1CC(Cc2ccccc2)(Cc3ccccc3)Cc1ccccc1	C(C1=CC=CC=C1)C(CC2=CC=CC=C2)(CC3=CC=CC=C3)CC4=CC=CC=C4
Ciba Dyes			
D1	CI Pigment Orange 73		[H]N1C(=O)C2=C(N([H])C(=O)C2=C1C3=C([H])C([H])=C(C([H])=C3[H])C(C([H])([H])[H])(C([H])([H])[H])C([H])([H])[H])C4=C([H])C([H])=C(C([H])=C4[H])C(C([H])([H])[H])(C([H])([H])[H])C([H])([H])[H]
D2	CI Pigment Violet 37		[H]N(C(=O)C1=C([H])C([H])=C([H])C([H])=C1[H])C2=C([H])C3=C(N=C4C(O3)=C(N([H])C(=O)C([H])([H])[H])C5=NC6=C([H])C(OC([H])([H])C([H])([H])[H])=C(N([H])C(=O)C7=C([H])C([H])=C([H])C([H])=C7[H])C([H])=C6OC5=C4N([H])C(=O)C([H])([H])[H])C([H])=C2OC([H])([H])C([H])([H])[H]
D3	CI Pigment Yellow 13		[H]N(\N=C(/C(=O)N([H])C1=C([H])C([H])=C(C([H])=C1C([H])([H])[H])C([H])([H])[H])C(=O)C([H])([H])[H])C2=C([H])C([H])=C(C([H])=C2Cl)C3=C([H])C(Cl)=C(N([H])\N=C(/C(=O)N([H])C4=C([H])C([H])=C(C([H])=C4C([H])([H])[H])C([H])([H])[H])C(=O)C([H])([H])[H])C([H])=C3[H]

Appendix C – Calculated Molecular Dimensions

Appendix C.1. OASIS

ID	Diam eff - min	Diam eff - max	Diam eff - ave	Diam eff - minimise	Diam max - min	Diam max - max	Diam max - ave	Diam max - minimise	Diam min - min	Diam min - max	Diam min - ave	Diam min - minimise	Max Distance - min	Max Distance - max	Max Distance - ave	Max Distance - minimise
Unit	Å	Å	Å	Å	Å	Å	Å	Å	Å	Å	Å	Å	Å	Å	Å	Å
1	7.8382	8.6564	8.2473	8.5819	14.1943	14.7490	14.4716	14.2381	5.9139	6.9599	6.4369	7.0783	10.5943	11.1490	10.8716	10.6381
2	7.9195	9.7908	8.8552	9.1509	14.1825	14.7507	14.4666	14.1527	5.9780	7.6497	6.8138	7.3122	10.5825	11.1507	10.8666	10.5527
3	8.6467	8.6636	8.6551	8.6588	14.3189	14.6079	14.4634	14.3695	6.7364	7.3353	7.0359	7.2595	10.7189	11.0079	10.8634	10.7695
4	7.8409	9.9336	8.8872	9.4983	14.0962	14.6869	14.3915	14.2939	6.0588	7.6978	6.8783	7.0670	10.4962	11.0869	10.7915	10.6939
5	8.6475	9.5465	9.0970	8.6503	14.2934	14.7420	14.5177	14.5640	6.8074	7.9457	7.3766	6.7518	10.6934	11.1420	10.9177	10.9640
6	8.6546	9.7692	9.2119	8.7397	14.2487	14.7175	14.4831	14.7242	6.8015	8.2667	7.5341	7.3681	10.6487	11.1175	10.8831	11.1242
7	8.6576	9.5008	9.0792	8.6580	14.4920	14.5967	14.5444	14.5126	7.1978	8.1716	7.6847	7.2482	10.8920	10.9967	10.9444	10.9126
8	9.4160	9.6632	9.5396	9.8005	14.4961	14.6521	14.5741	14.3763	7.8468	8.3510	8.0989	7.9656	10.8961	11.0521	10.9741	10.7763
9	9.3288	9.6857	9.5073	9.4405	14.5283	14.8072	14.6677	14.5998	7.5986	8.2713	7.9350	7.6923	10.9283	11.2072	11.0677	10.9998
10	9.6394	9.7217	9.6806	9.6136	13.8268	14.3402	14.0835	13.8252	8.0846	8.2427	8.1636	8.2908	10.4724	10.7402	10.6063	10.4817
11	9.6494	9.7375	9.6934	9.7364	14.5819	14.7915	14.6867	14.5834	8.0939	8.3393	8.2166	8.0737	10.9819	11.1915	11.0867	10.9834
12	7.1410	8.9834	8.0622	8.6720	10.7813	15.0858	12.9336	13.8459	5.1875	7.2370	6.2122	5.9579	8.0513	12.9258	10.4886	11.6859
13	7.4741	9.0828	8.2784	9.4931	11.9577	16.0882	14.0230	12.8492	5.0245	7.1768	6.1006	5.7109	9.3515	13.9282	11.6399	10.1264
14	7.5595	10.3312	8.9453	8.0442	12.7535	16.2253	14.4894	15.3606	5.9393	7.5032	6.7213	6.2066	9.4535	13.4953	11.4744	12.8227
15	7.2498	9.8130	8.5314	8.8056	11.9527	17.2506	14.6016	14.4265	5.7285	7.9416	6.8351	6.7684	9.2227	15.0906	12.1566	11.5737
16	7.9665	11.0145	9.4905	8.7081	11.7962	17.9588	14.8775	16.9884	5.3662	7.9046	6.6354	6.1543	9.1206	15.7988	12.4597	14.8284
17	7.1717	10.5126	8.8421	7.2790	12.1138	19.3176	15.7157	19.5985	5.3324	7.8581	6.5953	6.0301	8.9177	17.1576	13.0377	16.8685
18	7.8828	11.0722	9.4775	9.2695	12.5501	18.1651	15.3576	17.2080	5.9892	8.3224	7.1558	6.3927	9.3776	16.0051	12.6914	14.6682
19	7.9146	10.4114	9.1630	9.0541	12.1196	19.6308	15.8752	18.9541	6.1021	8.1882	7.1452	6.4144	8.8196	16.9008	12.8602	16.2241
20	7.9848	11.2780	9.6314	8.9617	12.8550	19.6420	16.2485	15.1621	6.2090	8.1537	7.1813	7.1536	10.3610	17.4820	13.9215	12.7963
21	8.1095	10.9207	9.5151	8.0775	13.4465	19.3198	16.3831	20.4547	6.4986	8.5857	7.5421	6.3447	10.1465	16.5898	13.3681	17.7247
22	8.2493	10.9173	9.5833	9.4165	14.1607	18.9113	16.5360	17.1674	6.9450	8.0844	7.5147	7.5316	10.8607	16.7513	13.8060	14.4374
23	7.9771	10.7443	9.3607	7.7272	12.8664	20.6196	16.7430	19.1229	5.8129	8.5017	7.1573	6.7844	9.5664	18.4596	14.0130	16.4934
24	8.2080	11.7189	9.9635	10.9065	13.6865	20.8839	17.2852	15.9322	5.8712	8.4731	7.1722	8.1797	10.9000	18.7239	14.8119	13.2022
25	7.6147	11.5796	9.5971	9.1300	13.6796	22.2000	17.9398	19.6575	6.3604	8.7070	7.5337	7.2993	10.7548	20.0400	15.3974	17.4975
26	8.3164	12.1701	10.2433	11.6279	13.6931	24.0670	18.8800	19.1074	5.8285	8.7663	7.2974	6.6287	10.9631	21.3370	16.1500	16.5951
27	7.9943	12.3449	10.1696	11.4182	13.6870	24.7731	19.2300	21.7657	6.1709	9.1651	7.6680	6.9052	10.9570	22.5635	16.7602	19.0357
28	8.2435	12.8529	10.5482	14.0027	14.6671	25.8596	20.2634	17.1317	6.5362	9.3321	7.9342	6.4235	11.3671	23.6996	17.5334	13.9414
29	7.8743	13.4311	10.6527	10.0688	14.9570	28.3444	21.6507	25.3481	6.7022	9.7918	8.2470	8.1120	12.2270	26.1844	19.2057	23.1881
30	8.8111	14.6289	11.7200	12.9305	14.4942	30.5423	22.5183	16.3424	7.2089	10.788	8.9984	9.6640	11.7642	28.3823	20.0733	13.6124
31	8.2773	14.6548	11.4660	8.8401	15.1754	33.7774	24.4764	33.7610	7.0060	11.111	9.0589	8.0365	12.4454	31.0474	21.7464	31.0310
32	9.6186	15.4751	12.5468	14.2953	17.3917	34.7752	26.0834	17.1441	7.3496	11.860	9.6051	11.5621	14.0917	32.6152	23.3534	14.4141
33	8.6611	16.6901	12.6756	16.6949	17.9488	36.9278	27.4383	27.5005	7.6588	11.955	9.8071	8.0733	15.2188	34.7678	24.9933	24.7705
34	7.2569	10.3428	8.7998	7.2923	13.7409	20.5362	17.1386	20.5166	6.5246	8.0591	7.2919	6.7799	11.5809	18.3762	14.9786	18.3566
35	8.2835	9.0302	8.6568	8.9170	15.3225	17.7576	16.5400	17.9899	6.8602	7.9886	7.4244	6.9434	13.1625	15.5976	14.3800	15.8299
36	7.5887	10.2268	8.9077	8.9763	13.3518	19.6494	16.5006	17.6423	6.8306	8.3692	7.5999	7.0492	11.1918	17.4894	14.3406	15.4823

37	10.1750	11.4394	10.8072	11.4384	14.3122	16.5459	15.4290	15.1938	6.5326	9.1716	7.8521	6.5524	12.1522	14.3859	13.2690	13.0338
38	7.6744	11.3393	9.5069	8.7697	14.6626	22.9689	18.8158	21.2687	6.9383	8.5362	7.7373	7.3929	12.5026	20.8089	16.6558	19.1087
39	8.9109	11.2394	10.0752	11.4910	14.5979	20.0146	17.3062	18.2710	6.5685	9.5189	8.0437	6.6402	12.4379	17.8546	15.1462	16.1110
40	7.3835	7.8508	7.6171	7.4877	11.0697	11.6207	11.3452	11.2530	6.2348	6.8320	6.5334	6.4409	8.9097	9.4607	9.1852	9.0930
41	7.3729	7.9795	7.6762	7.2372	11.7802	12.3252	12.0527	11.7342	5.5448	6.4035	5.9741	6.4912	9.6202	10.1652	9.8927	9.5742
42	7.9523	10.4026	9.1775	9.7588	11.9114	16.0802	13.9958	13.0977	6.3089	7.9070	7.1080	7.0031	9.5650	13.9202	11.7426	10.9377
43	7.7796	10.2942	9.0369	9.0418	11.8793	16.1276	14.0034	15.1464	6.5738	7.7677	7.1708	7.0576	9.7193	13.9676	11.8434	12.9864
44	10.5242	13.3006	11.9124	12.8425	14.0431	16.1692	15.1062	15.6378	7.1511	8.9942	8.0726	8.0960	11.8831	14.0092	12.9462	13.4778
45	10.0080	10.5154	10.2617	10.0600	15.1365	20.1470	17.6417	20.0555	5.1221	9.3178	7.2200	6.4813	11.5365	16.5470	14.0417	16.4555
46	9.3806	9.8288	9.6047	9.3596	13.9182	16.5835	15.2509	16.5848	4.7010	8.7009	6.7010	4.6993	10.3182	12.9835	11.6509	12.9848
R13	5.0170	5.7265	5.3718	5.0098	7.3283	8.2938	7.8111	8.3004	4.0070	4.8791	4.4431	4.0612	5.1683	6.0021	5.5852	5.9609
R20	5.1122	8.2007	6.6564	5.1482	10.3091	15.7497	13.0294	15.7588	4.4488	5.9859	5.2174	4.2009	8.1492	13.4136	10.7814	13.4472
R26	5.0736	9.4373	7.2554	5.1067	10.4528	18.9566	14.7047	18.9475	4.5410	6.2228	5.3819	4.3080	8.2928	16.7966	12.5447	16.7875
R66	6.4841	7.0302	6.7572	6.4996	9.6857	11.8493	10.7675	11.7502	4.0591	6.2003	5.1297	5.0377	7.5257	9.6893	8.6075	9.5902
R68	6.8519	8.8514	7.8516	6.8474	11.6768	16.4531	14.0649	16.4109	5.7958	6.7530	6.2744	5.8285	9.5168	14.2931	11.9049	14.2509
R91	7.1872	8.1498	7.6685	7.2663	11.3261	15.7081	13.5171	15.2612	5.5122	7.2557	6.3840	5.9179	9.1661	13.5482	11.3571	13.1012
R103	7.6410	9.2562	8.4486	8.2646	11.7089	15.6823	13.6956	13.8734	6.2343	7.3890	6.8116	6.8895	9.5489	13.5223	11.5356	11.7134
R111	7.7692	8.9609	8.3650	8.3634	10.3592	12.9010	11.6301	12.9250	4.1283	6.3788	5.2536	4.0298	8.1992	10.7410	9.4701	10.7650
R118	7.2396	9.9427	8.5912	7.2032	12.2827	16.7025	14.4926	16.5523	6.5436	7.3371	6.9404	6.5849	10.1227	14.5425	12.3326	14.3923
R120	7.6403	9.4625	8.5514	8.5192	12.2484	15.3251	13.7867	14.6508	6.4109	7.5094	6.9601	6.8141	10.0884	13.1651	11.6267	12.4908
R124	7.2842	10.0504	8.6673	9.0276	11.9068	16.8348	14.3708	14.1948	6.6551	7.6880	7.1716	7.5853	9.7468	14.6748	12.2108	12.0348
R130	8.5359	13.6983	11.1171	8.8481	16.1663	26.6743	21.4203	24.5590	6.4290	8.9786	7.7038	7.8024	13.4463	24.2343	18.8403	22.1190
R135	8.4562	12.8172	10.6367	11.3890	14.5454	22.5642	18.5548	16.8266	6.4956	8.2728	7.3842	7.1283	12.1054	20.1242	16.1148	14.3866
R141	8.2986	11.6371	9.9678	8.1694	14.9675	22.6872	18.8274	23.4642	6.4615	8.2164	7.3389	6.5695	12.5275	20.2472	16.3874	21.0242
R165	10.1652	10.6566	10.4109	10.1384	13.1088	13.9148	13.5118	13.6726	6.8866	7.8474	7.3670	7.3858	10.9488	11.7548	11.3518	11.5126
R172	9.7773	11.8332	10.8052	9.9850	13.6298	17.8449	15.7374	17.0661	8.2116	9.5293	8.8705	8.3246	11.4698	15.6849	13.5774	14.9061
R173	10.5345	12.1004	11.3174	12.3547	13.0328	17.6404	15.3366	16.5428	8.0268	9.9390	8.9829	8.2923	10.8728	15.4804	13.1766	14.3828
R190	11.4614	13.3800	12.4207	12.1755	14.4999	18.0307	16.2653	17.0352	9.7062	11.077	10.391	10.2223	12.3399	15.8707	14.1053	14.8752
R191	10.9552	13.9195	12.4373	12.1165	14.4600	20.1624	17.3112	18.2109	8.4111	10.945	9.6783	8.9729	12.3000	18.0024	15.1512	16.0509
R198	10.4445	12.2747	11.3596	12.1702	13.0562	14.8864	13.9713	13.1750	6.6526	10.367	8.5099	9.4784	10.8962	12.7264	11.8113	11.0150
D1	7.9576	8.0388	7.9982	8.0369	20.5344	20.8082	20.6713	20.6722	6.0901	6.4424	6.2662	6.2810	18.3744	18.6482	18.5113	18.5122
D2	12.4590	13.8833	13.1712	13.2320	21.0314	28.8947	24.9630	28.8371	6.6552	10.950	8.8031	6.6067	18.5251	26.7347	22.6299	26.6771
D3	11.3913	13.8103	12.6008	11.3639	19.4078	31.2921	25.3500	31.4530	7.1183	11.860	9.4895	7.5379	17.2478	29.1321	23.1900	29.2930

Appendix C.2. MOE

ID	rgyr max	rgyr min	rgyr ave	rgyr - minimise	std_dim1 max	std_dim1 min	std_dim1 ave	std_dim1 - minimise	std_dim2 max	std_dim2 min	std_dim2 ave	std_dim2 - minimise	std_dim3 max	std_dim3 min	std_dim3 ave	std_dim3 - minimise
Unit	Å	Å	Å	Å	Å²	Å²	Å²	Å²	Å²	Å²	Å²	Å²	Å²	Å²	Å²	Å²
1	4.23456	3.89939	4.08348	4.13202	3.08502	2.71123	2.92498	2.90759	1.49834	1.18154	1.29662	1.20371	1.21769	0.78370	1.05391	1.17315
2	4.28347	3.85164	4.11026	4.13454	3.11343	2.63582	2.92194	2.89935	1.55902	1.19254	1.35256	1.28386	1.23594	0.73484	1.08546	1.18715
3	4.09369	3.75072	3.97534	3.83612	3.10625	2.61278	2.93534	2.97060	1.48296	1.17225	1.33353	1.24155	1.26032	0.87537	1.08008	1.14654
4	4.35829	4.05443	4.20508	4.18046	3.09978	2.73579	2.93290	2.90113	1.62529	1.25663	1.39701	1.31722	1.31847	0.82408	1.11298	1.23767
5	4.18056	3.92653	4.06417	4.05032	3.12279	2.77864	3.01520	3.00566	1.53943	1.19672	1.29314	1.25914	1.22496	0.96449	1.13830	1.17299
6	4.28908	3.93163	4.12202	4.13351	3.12909	2.73804	2.95327	2.96263	1.59145	1.24633	1.40442	1.33929	1.27637	0.85800	1.16201	1.22189
7	4.25483	3.91523	4.07752	4.00524	3.16667	2.90478	3.04005	3.01727	1.52539	1.24092	1.37329	1.31627	1.31893	1.00914	1.21242	1.29969
8	4.44235	4.22053	4.33896	4.34068	3.15475	2.86313	3.01292	3.01951	1.66215	1.33414	1.44255	1.38772	1.34771	1.01575	1.23650	1.28031
9	4.32783	4.09808	4.23271	4.21752	3.20802	2.88008	3.07849	3.06464	1.54822	1.30365	1.39732	1.34187	1.34574	1.03478	1.22894	1.28750
10	4.16710	3.96632	4.09769	4.08925	3.15725	2.82394	3.03936	3.03205	1.55844	1.32557	1.39936	1.34598	1.36280	1.13365	1.27997	1.33269
11	4.35837	4.17262	4.26556	4.25533	3.20353	2.96106	3.08764	3.07627	1.54606	1.33919	1.42182	1.36531	1.36652	1.10004	1.28423	1.34569
12	3.70535	3.39591	3.54281	3.65331	3.85049	3.28155	3.58383	3.73281	1.35422	0.98064	1.11529	1.08131	1.03304	0.84387	0.93133	0.88345
13	4.48056	3.73256	3.98135	4.45574	4.11115	2.80154	3.72079	4.04354	1.75866	1.03247	1.22309	1.09643	1.07747	0.86657	0.98701	0.99406
14	4.46863	3.53224	3.87946	4.46868	3.96355	2.52472	3.54053	3.96353	1.97298	1.13228	1.25444	1.30476	1.23121	0.92792	1.02987	0.94160
15	4.58105	3.72044	4.24621	4.37384	4.29899	2.71777	3.75638	4.03954	1.98274	1.12810	1.58770	1.37685	1.12416	0.83619	0.95008	0.93730
16	4.94422	4.46877	4.70139	4.82138	4.83624	4.29510	4.58973	4.73384	1.47340	1.01622	1.13345	1.03459	1.06563	0.83924	0.94262	0.92327
17	5.42612	4.20465	4.58595	5.42620	5.12355	3.26166	3.96241	5.12360	2.17779	1.13232	1.69367	1.13332	1.29142	0.83987	1.16119	0.98290
18	4.98095	3.68850	4.34328	4.96589	4.86175	2.87119	3.85247	4.47698	2.24168	1.13790	1.78418	1.66454	1.37109	0.90157	1.06649	0.93645
19	5.01374	3.77700	4.32992	4.99470	4.93931	2.50671	3.93655	4.86950	2.42710	1.06427	1.66951	1.11415	1.56311	0.88083	1.08199	1.02046
20	5.16535	4.19937	4.93973	5.04377	5.05434	3.83983	4.70030	4.81170	1.81153	1.18550	1.42157	1.49688	1.17227	0.89992	1.01296	0.97778
21	5.61688	4.55292	5.10681	5.60459	4.99391	3.33864	4.23064	4.99391	2.13366	1.10787	1.68485	1.33747	1.28637	0.91018	1.05422	0.93937
22	5.05196	4.84910	4.96768	4.95265	5.21099	4.71604	5.08610	5.10827	1.41813	1.12441	1.19276	1.16030	1.19308	0.94436	1.00897	0.95853
23	5.36898	4.30642	4.89597	5.01211	5.30794	4.18645	4.92126	4.61642	1.84501	1.07348	1.36112	1.71018	1.17531	0.86147	1.01351	1.02154
24	6.03388	4.32782	5.47115	5.82863	5.80129	3.21332	5.00887	5.27334	2.40776	1.05911	1.71040	1.36548	1.31624	0.88074	1.03212	1.08859
25	5.82890	4.11710	5.24334	5.62701	5.64455	2.86073	4.73013	5.35199	2.59911	1.38902	2.00328	1.68261	1.43387	0.89693	1.07724	0.98614
26	6.10852	4.45293	5.30289	6.02052	5.97049	3.74880	5.13856	5.63129	2.81059	1.35325	2.03589	2.15873	1.42540	0.90464	1.12497	0.91712
27	6.20433	4.90307	5.44774	6.15699	5.95542	3.97665	4.87391	5.95563	2.49846	1.53677	1.95949	1.89407	1.75510	0.87953	1.25372	0.88095
28	6.52906	4.47856	5.03555	6.36191	6.34423	2.85551	3.75457	6.34441	3.18849	1.82480	2.72400	1.82518	1.76314	1.00063	1.23043	1.18307
29	7.47107	4.69936	5.28536	7.37076	7.60564	3.45505	4.33240	7.40976	3.19865	1.39410	2.69194	1.70001	2.09742	0.95746	1.32944	0.99634
30	7.43919	5.34916	6.41557	7.43922	7.58772	4.45783	5.95233	7.48900	3.31642	1.44466	2.45968	1.77971	2.18289	0.92112	1.42591	1.12446
31	8.20678	5.35513	6.30502	8.20683	8.42463	3.83836	5.21088	8.42485	3.66862	1.80798	3.15379	1.80829	1.92874	1.06586	1.39139	1.13144
32	9.16167	6.69247	7.99789	9.11624	9.32361	5.68344	7.65869	9.26259	3.45296	1.26835	2.45248	1.37411	2.00753	0.97230	1.44374	1.29143
33	8.54471	5.29023	6.16027	8.54473	8.25482	3.78705	4.90076	8.25487	4.75603	2.47932	3.33752	2.78427	2.57945	1.06946	1.76775	1.26646
34	4.77227	3.63082	3.86066	4.29753	5.45340	3.28481	3.68727	4.66774	2.10018	1.51002	2.00681	1.79515	1.63296	1.19761	1.32505	1.35288
35	4.34701	3.67363	3.89893	3.84495	4.74997	3.63199	4.03630	3.88065	1.90552	1.62723	1.76360	1.67963	1.64150	1.28232	1.48109	1.49329
36	4.95002	3.87446	4.19588	4.76425	5.30526	3.46481	4.03995	5.01968	2.19861	1.51399	1.97574	1.61295	1.71526	1.29212	1.53432	1.29285
37	4.37699	3.93881	4.18022	4.16946	4.26986	3.46278	3.98193	3.96725	2.58355	1.90416	2.18494	2.38852	1.93880	1.21816	1.68007	1.21919

38	5.21753	4.21379	4.75806	5.21759	5.69343	3.84813	5.02490	5.65225	2.12807	1.74151	1.89248	1.82778	1.80708	1.39076	1.49405	1.44746
39	5.20361	4.31757	4.73874	4.50244	5.49877	4.17706	4.83177	4.45952	2.58330	1.93641	2.35141	2.24918	1.97042	1.28270	1.53663	1.63269
40	2.91752	2.69945	2.81135	2.81320	2.74325	2.32679	2.55266	2.52248	1.61371	1.33647	1.48073	1.49867	1.44482	1.02182	1.23334	1.29646
41	3.14905	2.91334	3.04295	3.02986	2.85822	2.41806	2.69845	2.66251	1.67313	1.33808	1.50073	1.49838	1.34941	0.85288	1.16485	1.22702
42	4.10140	3.29947	3.74436	3.84741	3.94756	2.39637	3.26255	3.45167	2.39507	1.46366	1.89178	1.72519	1.66354	1.20162	1.44557	1.50310
43	4.12840	3.34660	3.63523	3.97888	3.97849	2.75550	3.21594	3.76108	2.26394	1.60756	2.05739	1.80465	1.48779	1.03801	1.30264	1.16282
44	4.32898	3.76864	4.06830	4.24325	3.64913	2.70068	3.13410	3.43335	2.87891	2.08394	2.51851	2.51962	2.05464	1.36023	1.83720	1.63745
45	6.45257	6.07470	6.30205	6.42992	4.63091	4.30106	4.49194	4.61452	1.74297	1.33768	1.60821	1.67527	1.11055	0.29419	0.70458	0.29518
46	5.07195	3.72425	4.48478	5.07089	3.69594	1.93784	3.03467	3.69594	1.94477	1.53988	1.73828	1.73497	1.70270	0.34109	1.06060	0.34204
R13	1.98171	1.84655	1.92397	1.91752	1.97188	1.69689	1.86764	1.82636	0.91314	0.76041	0.83199	0.85890	0.75857	0.62142	0.65811	0.66136
R20	4.20910	3.64912	3.99298	4.10992	4.13485	3.44552	3.92254	4.01308	1.26784	0.84008	0.93305	0.89177	0.87073	0.68039	0.75394	0.69922
R26	4.83642	3.16801	4.00072	4.82703	2.11952	0.87511	1.51816	4.97820	1.22976	0.70378	0.93735	0.89428	*	*	*	*
R66	2.81921	2.31749	2.68977	2.72260	2.94800	2.12061	2.73660	2.78111	1.35128	1.11349	1.19074	1.12890	1.15266	0.50822	0.73945	0.83109
R68	4.23231	3.25080	3.76520	4.19553	4.26947	2.95124	3.63182	4.23262	1.77841	1.15322	1.45742	1.17926	1.20947	0.92311	1.05407	0.92642
R91	4.29146	3.53247	4.05006	4.09804	4.22790	3.21660	3.92530	3.97814	1.64624	1.18159	1.35553	1.38567	1.32598	0.84683	1.08645	1.05073
R103	3.73596	3.34005	3.51581	3.62786	3.62618	2.98583	3.27742	3.45225	1.72985	1.36428	1.54434	1.54811	1.47920	1.11060	1.32587	1.17904
R111	3.17263	2.70484	2.91122	2.94432	3.26252	2.45462	2.85075	2.86551	1.72916	1.56986	1.63560	1.69717	1.17788	0.45327	0.75493	0.77883
R118	4.06220	3.75327	3.95269	3.88768	3.99645	3.53383	3.85167	3.72445	1.64745	1.31246	1.38796	1.52605	1.34215	1.20255	1.28828	1.31372
R120	3.93363	3.65379	3.81937	3.78571	3.82967	3.40618	3.66113	3.62881	1.53437	1.31780	1.41042	1.36492	1.39442	1.25072	1.31887	1.36282
R124	3.87860	3.56088	3.75081	3.92426	3.69849	3.17720	3.50861	3.77683	1.79039	1.45829	1.58117	1.49749	1.50396	1.36819	1.43221	1.32955
R130	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
R135	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
R141	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
R165	3.49392	3.32525	3.41419	3.38903	3.15562	2.92054	3.05179	3.03084	2.38692	2.10692	2.24719	2.24310	1.45596	1.09912	1.25049	1.22572
R172	4.43917	3.94720	4.25802	4.11477	3.96170	2.82993	3.52662	3.63553	2.80896	1.91526	2.36746	2.02356	1.94522	1.21606	1.72451	1.65346
R173	4.38707	3.75873	3.98551	4.21434	3.93263	2.85364	3.18616	3.43404	2.58742	2.04214	2.34767	2.52138	1.94018	1.39908	1.68383	1.61834
R190	4.36361	4.03972	4.20274	4.48372	3.49979	2.82015	3.08071	3.84994	2.70127	2.31727	2.55140	2.24192	2.30063	2.01057	2.19472	1.91272
R191	4.82022	4.18277	4.46686	4.62996	4.07444	2.81113	3.25539	3.50171	3.04638	2.31739	2.67609	2.93702	2.57876	1.76003	2.27814	1.94337
R198	4.00534	3.79282	3.93022	3.89624	3.26141	2.46690	2.82154	2.39719	2.60168	1.96810	2.33186	2.39327	2.31903	1.31776	1.98822	2.38444
D1	5.37706	4.90159	5.26174	5.30854	6.12944	5.31142	5.96057	6.05403	1.69765	1.29913	1.42837	1.41010	1.42377	0.89123	1.10323	1.01539
D2	6.71125	6.34832	6.59798	6.63224	6.60559	6.05962	6.38310	6.43006	3.27269	2.83769	3.08055	3.14220	1.74888	0.69445	1.10933	0.99554
D3	8.32127	6.75086	7.79490	8.13000	8.93691	6.40815	8.09927	8.65015	2.97226	1.86786	2.41657	2.53001	2.35888	0.78731	1.48401	0.85921

[¥] units are stated as Å in the MOE documentation, however as these variables are square of eigenvectors, it seems unlikely this is correct.

* indicates the calculation was not performed

Appendix C.3. TSAR

ID	Name	IMS1	IMS2	IMS3	IML1	IML2	IML3	Ell Vol
Unit		kg m^2	kg m^2	kg m^2	kg m^2	kg m^2	kg m^2	$Å^3$
1	2,2',4,4'-Tetrabromodiphenyl ether (BDE47)	355.81	1135.63	1302.96	10.3340	3.23781	2.82200	395.51
2	2,2',4,4',5-Pentabromodiphenyl ether (BDE99)	500.25	1251.58	1494.42	9.2562	3.69968	3.09847	444.46
3	2,2',4,4',6-Pentabromodiphenyl ether (BDE100)	452.12	1258.12	1274.09	8.7563	3.14669	3.10724	358.62
4	2,2',4,4',5,5'-Hexabromodiphenyl ether (BDE153)	546.93	1518.29	1715.77	10.2023	3.67511	3.25212	510.77
5	2,2',4,4',5,6'-Hexabromodiphenyl ether (BDE154)	604.71	1336.87	1474.45	7.9615	3.60124	3.26521	392.15
6	2,2',3,4,4',5',6-Heptabromodiphenyl ether (BDE183)	681.11	1640.92	1715.46	8.7579	3.63520	3.47724	463.72
7	2,2',3,3',4,4',6,6'-Octabromodiphenyl ether (BDE197)	662.43	1793.66	1918.24	10.0915	3.72694	3.48489	549.02
8	2,2',3,3',4,4',5,5',6-Nonabromodiphenyl ether (BDE206)	848.16	2254.49	2337.85	9.7055	3.65130	3.52111	522.67
9	2,2',3,3',4,4',5,6,6'-Nonabromodiphenyl ether (BDE207)	733.05	2097.59	2288.02	10.0330	3.50629	3.21448	473.67
10	BDE208 = 2,2',3,3',4,5,5',6,6'-Nonabromodiphenyl ether (BDE208)	867.23	1888.36	2046.49	8.9934	4.13021	3.81106	592.97
11	Decabromodiphenyl ether	900.69	2325.19	2460.97	9.2547	3.58493	3.38713	470.72
12	C10H18Cl4 (51.7% wt. Cl)	118.96	477.89	530.80	12.8659	3.20259	2.88335	497.66
13	C11H19Cl5 (54.0% wt Cl)	136.99	853.17	953.17	21.7645	3.49474	3.12810	996.62
14	C11H18Cl6 (58.7% wt. Cl)	247.51	741.99	877.40	13.9968	4.66901	3.94844	1080.85
15	C12H21Cl5 (51.8% wt. Cl)	149.30	1026.82	1094.26	16.9254	2.46095	2.30929	402.91
16	C13H22Cl6 (54.5% wt. Cl)	283.63	1036.01	1113.37	15.8827	4.34827	4.04615	1170.50
17	C14H26Cl4 (42.3% wt. Cl)	107.31	1445.48	1521.60	44.8489	3.32952	3.16295	1978.40
18	C14H24Cl6 (52.6% wt. Cl)	234.40	1383.60	1463.84	23.9538	4.05813	3.83568	1561.82
19	C14H24Cl6 (52.6% wt. Cl)	313.92	1336.01	1433.76	18.5749	4.36450	4.06695	1381.08
20	C14H22Cl8 (59.9% wt. Cl)	238.91	1677.96	1738.35	19.5631	2.78542	2.68866	613.70
21	C14H22Cl8 (59.9% wt. Cl)	297.18	1912.72	1950.48	21.2953	3.30862	3.24457	957.58
22	C14H18Cl12 (69.9% wt.Cl)	393.58	1925.21	2015.76	18.4027	3.76221	3.59320	1042.07
23	C15H26Cl6 (50.8% wt. Cl)	253.01	1566.54	1629.13	24.6661	3.98373	3.83069	1576.73
24	C16H27Cl7 (53.2% wt. Cl)	275.39	2476.53	2563.84	30.6514	3.40838	3.29231	1440.75
25	C17H29Cl7 (51.6% wt. Cl)	504.81	1670.13	1916.78	18.8971	5.71177	4.97679	2250.10
26	C18H31Cl7 (50.2% wt. Cl)	377.01	2250.80	2347.73	20.7014	3.46753	3.32437	999.58
27	C19H32Cl8 (52.2% wt. Cl)	419.48	3049.66	3141.51	26.9959	3.71328	3.60472	1513.61
28	C20H34Cl8 (50.9% wt. Cl)	430.14	3248.40	3332.92	26.9536	3.56911	3.47861	1401.75
29	C22H37Cl9 (51.5% wt. Cl)	440.50	4321.66	4346.41	39.1204	3.98748	3.96477	2590.65
30	C24H40Cl10 (52.0% wt. Cl)	1003.43	4619.36	5083.30	31.7864	6.90471	6.27453	5768.41
31	C26H43Cl11 (52.4% wt. Cl)	954.22	6506.47	6913.03	34.2209	5.01876	4.72361	3398.21
32	C28H46Cl12 (52.7% wt. Cl)	1757.09	7082.80	8146.84	30.6674	7.60791	6.61426	6464.17
33	C30H49Cl13 (53.0% wt. Cl)	2454.78	6457.88	8047.40	22.1733	8.42853	6.76372	5294.87

34	Di-(tert-nonyl) polysulphide	169.68	1664.06	1737.91	34.6981	3.53802	3.38767	1742.03
35	Di-(tert-nonyl) polysulphide v2	163.49	1499.94	1518.48	33.7969	3.68386	3.63889	1897.74
36	Di-(tert-dodecyl) polysulphide	180.24	1599.36	1628.27	26.1187	2.94347	2.89120	931.06
37	Di-(tert-dodecyl) polysulphide v2	371.85	875.47	1081.78	11.5031	4.88590	3.95411	930.89
38	Di-(tert-dodecyl) pentasulphide	198.29	2969.67	3003.15	55.1852	3.68481	3.64373	3103.64
39	Di-(tert-dodecyl) pentasulphide v2	337.21	2206.23	2384.15	29.3163	4.48087	4.14648	2281.60
40	2-styrylphenol	76.32	217.28	228.55	9.2247	3.24020	3.08045	385.68
41	4-styrylphenol	67.62	257.84	269.54	12.7665	3.34827	3.20303	573.51
42	2,4-Distyryl phenol	177.85	617.91	672.07	14.6362	4.21264	3.87316	1000.31
43	2,6-Distyrylphenol	220.94	470.22	579.25	9.5561	4.49017	3.64500	655.13
44	2,4,6-Tristyrylphenol	474.37	799.28	1068.70	10.7402	6.37429	4.76731	1367.12
45	Ethylene bistetrabromophthalimide [1H-Isoindole-1,3(2H)-dione, 2,2'-(1,2-ethanediyl)bis 4,5,6,7-tetrabromo-]	820.31	5794.62	6361.82	24.4970	3.46791	3.15872	1124.03
46	1,2-Bis(pentabromophenyl) ethane	897.69	2829.25	3695.74	10.1618	3.22424	2.46830	338.75
R13	CCCCO	4.58	41.34	43.72	13.8240	1.53175	1.44838	128.47
R20	CCCCCCCCCCO	10.85	427.59	432.96	57.8702	1.46805	1.44985	515.95
R26	CCCCCCCCCCCCC	13.44	689.61	695.94	74.6897	1.45566	1.44242	656.91
R66	CCCCc1ccccc1	22.88	147.51	156.33	14.5782	2.26153	2.13391	294.69
R68	CC(C)C(C)(C)CCCCCCCC	37.86	539.01	547.48	35.5060	2.49391	2.45533	910.71
R91	C1CCCCCC1CCCC2CCCCCC2	74.71	586.99	605.22	23.3792	2.97565	2.88600	841.00
R103	CCC(C)C(C)C(C)C(C)C(C)C(C)CCC	79.85	471.66	479.72	18.2129	3.08341	3.03161	713.14
R111	CCCCc1cccc2ccccc12	76.20	191.02	251.64	9.3679	3.73673	2.83663	415.93
R118	CC(C)C(C)(C)C(C)(C)C(C)(C)C(C)(C)CCCCC	116.76	582.85	608.39	17.0740	3.42045	3.27689	801.62
R120	CCC(C)C(C)C(C)C(C)C(C)C(C)C(C)C(C)C	107.27	538.21	548.73	16.2716	3.24305	3.18092	703.11
R124	CC(C)C(C)(C)C(C)(C)C(C)(C)C(C)(C)C(C)(C)CCCC	122.91	648.86	664.51	16.8136	3.18496	3.10993	697.60
R130	COC(=O)CCCCCC(=O)OCC(=O)C3(O)CCC4C2CCC1=CC(=O)CCC1(C)C2CC(O)C34C	243.06	3635.18	3727.75	56.7856	3.79683	3.70254	3343.86
R135	CCCCCC(=O)OCC(=O)C3(O)CCC4C2CCC1=CC(=O)CCC1(C)C2CC(O)C34C	226.62	2079.52	2164.73	34.0646	3.71225	3.56613	1888.97
R141	COC(=O)CCC(=O)OCC(=O)C3(O)CCC4C2CCC1=CC(=O)CCC1(C)C2CC(O)C34C	247.04	2202.07	2311.00	34.0438	3.81925	3.63924	1982.05
R165	CCc1c(CC)c(CC)c(CC)c2c(CC)c(CC)ccc12	210.74	369.17	498.73	8.4941	4.84886	3.58915	619.21
R172	CCC(CC)C(CC)(CC)C(CC)(CC)C(CC)(CC)C(CC)(CC)CCCCC	331.07	875.57	960.73	12.5893	4.76026	4.33835	1089.04
R173	CCCCC(CCC)C(CCC)C(CCC)CCCCCC	361.33	622.13	788.99	11.0227	6.40196	5.04803	1492.15
R190	CCCC(CCC)C(CCC)(CCC)C(CCC)(CCC)C(CCC)(CCC)CCCCCC	498.81	1195.53	1286.15	12.3939	5.17116	4.80681	1290.45
R191	CCCCCC(CCCC)C(CCCC)C(CCCC)CCCCCC	623.39	819.03	1080.68	10.4638	7.96432	6.03601	2107.06
R198	c1ccccc1CC(Cc2ccccc2)(Cc3ccccc3)Cc1ccccc1	606.04	610.33	610.82	6.3038	6.25946	6.25450	1033.76

D1	CI Pigment Orange 73.mol	141.96	1750.15	1849.99	34.1755	2.77199	2.62240	1040.62
D2	CI Pigment Violet 37.mol	851.76	4426.12	5066.49	34.3085	6.60235	5.76785	5472.72
D3	CI Pigment Yellow 13.mol	538.31	7150.59	7508.52	67.5846	5.08793	4.84539	6979.22

Appendix C.4. Mol2Mol

ID	Chemical name	Mol3mol BB X	Mol3mol BB Y	Mol3mol BB Z
Unit		Å	Å	Å
1	2,2',4,4'-Tetrabromodiphenyl ether (BDE47)	12.540	5.760	6.230
2	2,2',4,4',5-Pentabromodiphenyl ether (BDE99)	12.539	7.206	5.900
3	2,2',4,4',6-Pentabromodiphenyl ether (BDE100)	12.632	5.889	6.157
4	2,2',4,4',5,5'-Hexabromodiphenyl ether (BDE153)	12.627	7.487	5.442
5	2,2',4,4',5,6'-Hexabromodiphenyl ether (BDE154)	12.643	6.741	6.786
6	2,2',3,4,4',5',6-Heptabromodiphenyl ether (BDE183)	12.642	6.726	6.869
7	2,2',3,3',4,4',6,6'-Octabromodiphenyl ether (BDE197)	12.887	6.569	6.474
8	2,2',3,3',4,4',5,5',6-Nonabromodiphenyl ether (BDE206)	12.861	7.093	7.385
9	2,2',3,3',4,4',5,6,6'-Nonabromodiphenyl ether (BDE207)	12.886	6.491	6.522
10	BDE208 = 2,2',3,3',4,5,5',6,6'-Nonabromodiphenyl ether (BDE208)	12.618	7.172	7.726
11	Decabromodiphenyl ether	12.895	6.539	6.488
12	C10H18Cl4 (51.7% wt. Cl)	12.686	4.705	5.739
13	C11H19Cl5 (54.0% wt Cl)	14.411	6.919	4.249
14	C11H18Cl6 (58.7% wt. Cl)	12.382	5.684	7.185
15	C12H21Cl5 (51.8% wt. Cl)	15.843	4.630	5.975
16	C13H22Cl6 (54.5% wt. Cl)	14.698	5.995	7.773
17	C14H26Cl4 (42.3% wt. Cl)	17.827	4.113	6.571
18	C14H24Cl6 (52.6% wt. Cl)	16.623	6.077	7.220
19	C14H24Cl6 (52.6% wt. Cl)	14.784	8.903	7.397
20	C14H22Cl8 (59.9% wt. Cl)	17.735	5.969	6.179
21	C14H22Cl8 (59.9% wt. Cl)	16.966	6.530	6.658
22	C14H18Cl12 (69.9% wt.Cl)	16.803	6.327	7.150
23	C15H26Cl6 (50.8% wt. Cl)	17.657	6.755	6.354
24	C16H27Cl7 (53.2% wt. Cl)	19.524	6.771	6.250
25	C17H29Cl7 (51.6% wt. Cl)	15.080	9.189	7.333
26	C18H31Cl7 (50.2% wt. Cl)	20.825	7.190	7.100
27	C19H32Cl8 (52.2% wt. Cl)	21.591	6.850	7.590
28	C20H34Cl8 (50.9% wt. Cl)	22.965	7.549	6.976
29	C22H37Cl9 (51.5% wt. Cl)	23.859	7.802	7.232
30	C24H40Cl10 (52.0% wt. Cl)	23.977	7.606	12.229
31	C26H43Cl11 (52.4% wt. Cl)	27.548	8.668	8.605
32	C28H46Cl12 (52.7% wt. Cl)	26.949	8.688	13.896
33	C30H49Cl13 (53.0% wt. Cl)	25.748	15.209	8.209
34	Di-(tert-nonyl) polysulphide	17.318	6.758	7.370

35	Di-(tert-nonyl) polysulphide v2	17.217	6.937	7.578
36	Di-(tert-dodecyl) polysulphide	17.950	5.648	6.130
37	Di-(tert-dodecyl) polysulphide v2	14.306	9.734	5.869
38	Di-(tert-dodecyl) pentasulphide	21.519	5.760	7.265
39	Di-(tert-dodecyl) pentasulphide v2	17.862	11.338	6.807
40	2-styrylphenol	9.950	6.135	5.707
41	4-styrylphenol	10.397	6.264	5.459
42	2,4-Distyryl phenol	13.824	8.413	6.462
43	2,6-Distyrylphenol	11.797	6.199	10.458
44	2,4,6-Tristyrylphenol	14.948	10.176	7.800
45	Ethylene bistetrabromophthalimide [1H-Isoindole-1,3(2H)-dione, 2,2'-(1,2-ethanediyl)bis 4,5,6,7-tetrabromo-]	18.276	9.889	4.840
46	1,2-Bis(pentabromophenyl) ethane	14.952	7.690	3.565
R13	CCCCO	7.553	3.259	2.856
R20	CCCCCCCCCCO	15.050	3.285	2.856
R26	CCCCCCCCCCCCC	17.812	3.224	2.845
R66	CCCCc1ccccc1	10.552	4.651	4.605
R68	CC(C)C(C)(C)CCCCCCCC	15.295	4.881	5.462
R91	C1CCCCCC1CCCC2CCCCCC2	13.536	6.576	5.794
R103	CCC(C)C(C)C(C)C(C)C(C)C(C)CCC	14.713	6.336	5.793
R111	CCCCc1cccc2ccccc12	10.822	7.430	4.403
R118	CC(C)C(C)(C)C(C)(C)C(C)(C)C(C)(C)CCCCC	14.680	6.138	6.429
R120	CCC(C)C(C)C(C)C(C)C(C)C(C)C(C)C(C)C	14.162	6.243	6.376
R124	CC(C)C(C)(C)C(C)(C)C(C)(C)C(C)(C)C(C)(C)CCCC	15.041	6.041	6.056
R130	COC(=O)CCCCCC(=O)OCC(=O)C3(O)CCC4C2CCC1=CC(=O)CCC1(C)C2CC(O)C34C	25.121	6.178	8.020
R135	CCCCCC(=O)OCC(=O)C3(O)CCC4C2CCC1=CC(=O)CCC1(C)C2CC(O)C34C	21.774	6.684	7.810
R141	COC(=O)CCC(=O)OCC(=O)C3(O)CCC4C2CCC1=CC(=O)CCC1(C)C2CC(O)C34C	20.804	8.072	6.226
R165	CCc1c(CC)c(CC)c(CC)c2c(CC)c(CC)ccc12	12.615	7.130	8.295
R172	CCC(CC)C(CC)(CC)C(CC)(CC)C(CC)(CC)C(CC)(CC)CCCCC	15.117	9.170	7.712
R173	CCCCC(CCC)C(CCC)C(CCC)CCCCCC	15.713	12.080	8.426
R190	CCCC(CCC)C(CCC)(CCC)C(CCC)(CCC)C(CCC)(CCC)CCCCCC	17.721	10.426	9.806
R191	CCCCCC(CCCC)C(CCCC)C(CCCC)CCCCCC	17.035	10.421	13.764
R198	c1ccccc1CC(Cc2ccccc2)(Cc3ccccc3)Cc1ccccc1	11.870	10.887	10.421
D1	CI Pigment Orange 73	19.352	5.102	7.056
D2	CI Pigment Violet 37	23.691	14.293	7.419
D3	CI Pigment Yellow 13	28.841	9.813	7.299

Appendix C.5. SPARTAN

ID	Chemical name	Long (Å)	Thick(Å)
10	BDE208 = 2,2',3,3',4,5,5',6,6'-Nonabromodiphenyl ether (BDE208)	**14.7**	**10.5**
11	Decabromodiphenyl ether	**14.9**	**10.7**
14	$C_{11}H_{18}Cl_6$ (58.7% wt. Cl)	**13.6**	**6.8**
23	$C_{15}H_{26}Cl_6$ (50.8% wt. Cl)	**17.7**	**8.4**
34	Di-(tert-nonyl) polysulphide	**16.7**	**8.4**
36	Di-(tert-dodecyl) polysulphide	**16.0**	**9.2**
42	2,4-Distyryl phenol	**12.5**	**8.8**
43	2,6-Distyrylphenol	**13.6**	**9.2**
45	Ethylene bistetrabromophthalimide (1H-Isoindole-1,3(2H)-dione, 2,2'-(1,2-ethanediyl)bis 4,5,6,7-tetrabromo-)	**17.0**	**11.8**
46	1,2-Bis(pentabromophenyl) ethane	**17.0**	**5.5**
R13	CCCCO	**8.3**	**4.2**
R20	CCCCCCCCCCO	**15.0**	**7.7**
R26	CCCCCCCCCCCCC	**18.8**	**4.2**
R66	CCCCc1ccccc1	**11.6**	**6.3**
R68	CC(C)C(C)(C)CCCCCCCC	**15.4**	**7.6**
R91	C1CCCCCC1CCCC2CCCCCC2	**15.6**	**7.1**
R103	CCC(C)C(C)C(C)C(C)C(C)C(C)CCC	**14.1**	**7.7**
R111	CCCCc1cccc2ccccc12	**11.6**	**8.8**
R118	CC(C)C(C)(C)C(C)(C)C(C)(C)C(C)(C)CCCCC	**14.1**	**10.0**
R120	CCC(C)C(C)C(C)C(C)C(C)C(C)C(C)C(C)C	**13.6**	**7.9**
R124	CC(C)C(C)(C)C(C)(C)C(C)(C)C(C)(C)C(C)(C)CCCC	**15.1**	**7.6**
R130	COC(=O)CCCCCC(=O)OCC(=O)C3(O)CCC4C2CCC1=CC(=O)CCC1(C)C2CC(O)C34C	**14.9**	**11.4**
R135	CCCCCC(=O)OCC(=O)C3(O)CCC4C2CCC1=CC(=O)CCC1(C)C2CC(O)C34C	**16.3**	**12.8**
R141	COC(=O)CCC(=O)OCC(=O)C3(O)CCC4C2CCC1=CC(=O)CCC1(C)C2CC(O)C34C	**19.8**	**7.6**
R165	CCc1c(CC)c(CC)c(CC)c2c(CC)c(CC)ccc12	**13.6**	**11.6**
R172	CCC(CC)C(CC)(CC)C(CC)(CC)C(CC)(CC)C(CC)(CC)CCCCC	**15.3**	**10.3**
R173	CCCCC(CCC)C(CCC)C(CCC)CCCCCC	**16.6**	**11.6**
R190	CCCC(CCC)C(CCC)(CCC)C(CCC)(CCC)C(CCC)(CCC)CCCCCC	**17.3**	**12.6**
R191	CCCCCC(CCCC)C(CCCC)C(CCCC)CCCCCC	**16.5**	**13.3**
R198	c1ccccc1CC(Cc2ccccc2)(Cc3ccccc3)Cc1ccccc1	**13.6**	**11.7**

Appendix D – Interrelationships Between Calculated Descriptors

OASIS

Table D 1 Correlation matrix for OASIS descriptors from energy minimised structures.

	D_{eff} minimum	D_{max} minimum	D_{min} minimum
D_{max} minimised	0.303		
D_{min} minimised	0.658	0.179	
Max Distance minimised	0.289	0.993	0.165

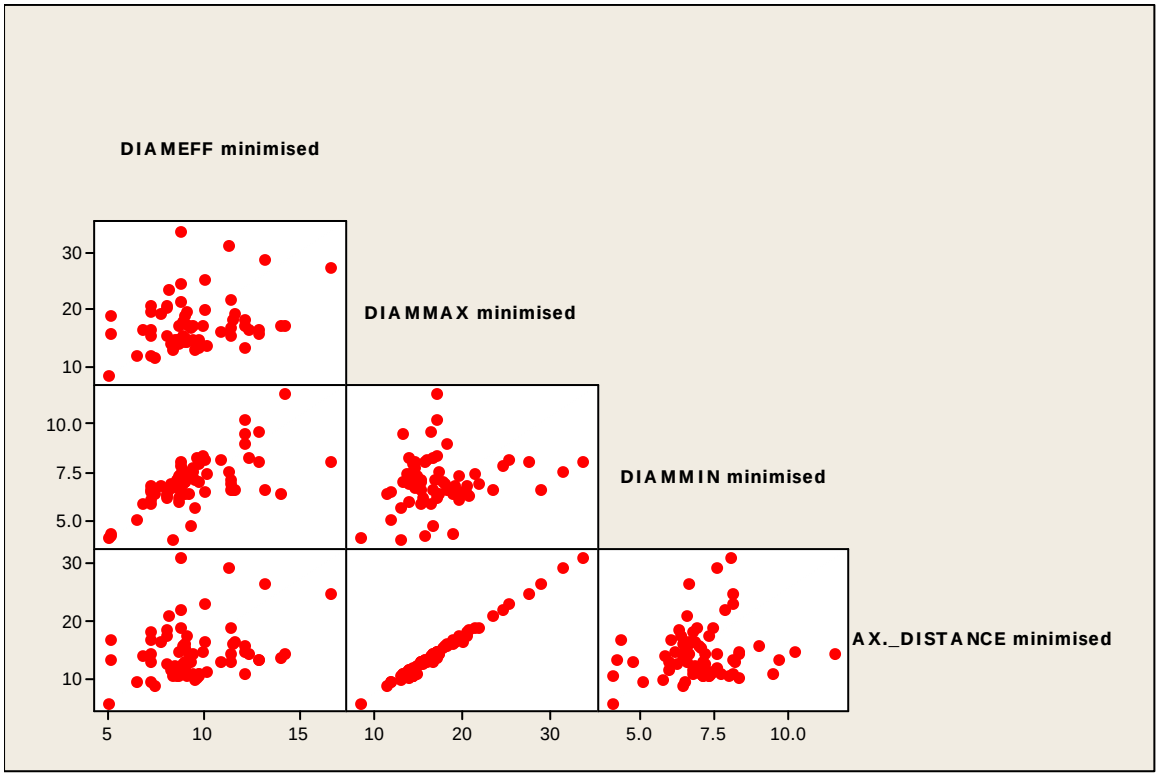


Figure D 1 Graphical representations of the relationships between OASIS descriptors from energy minimised structures.

MOE

Table D 2 Correlation matrix for MOE descriptors from energy minimised structures.

	Rgyr - minimised	Std_dim1 – minimised	Std_dim2 - minimised
Std_dim1 - minimised	0.944		
Std_dim2 – minimised	0.327	0.273	
Std_dim3 - minimised	-0.158	-0.189	0.435

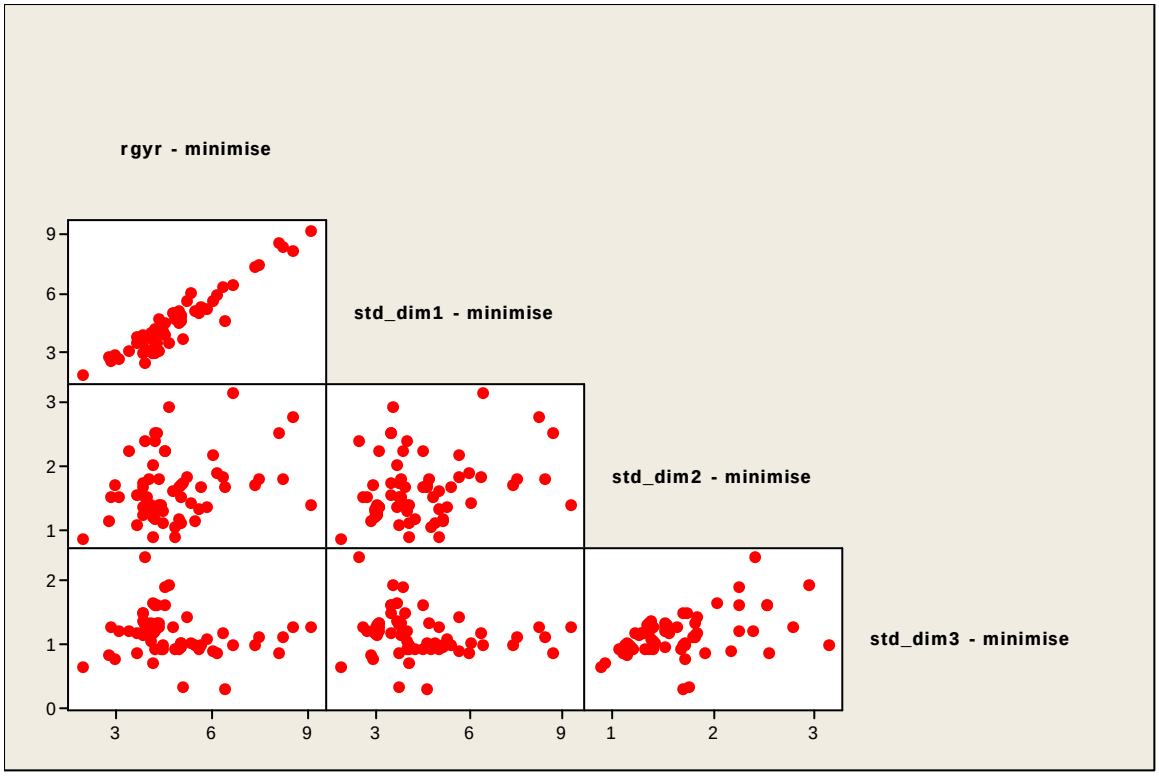


Figure D 2 Graphical representations of the relationships between MOE descriptors from energy minimised structures

Table D 3 Correlation matrix for MOE and OASIS descriptors from energy minimised structures.

	Rgyr - minimised	Std_dim1 - minimised	Std_dim2 - minimised	Std_dim3 – minimised	D_{eff} minimum	D_{max} minimum
Std_dim1 - minimised	0.944					
Std_dim2 – minimised	0.327	0.273				
Std_dim3 - minimised	-0.158	-0.189	0.435			
D_{eff} minimum	0.584	0.459	0.726	0.372		
D_{max} minimum	0.796	0.807	0.459	-0.108	0.303	
D_{min} minimum	0.425	0.304	0.413	0.646	0.658	0.179

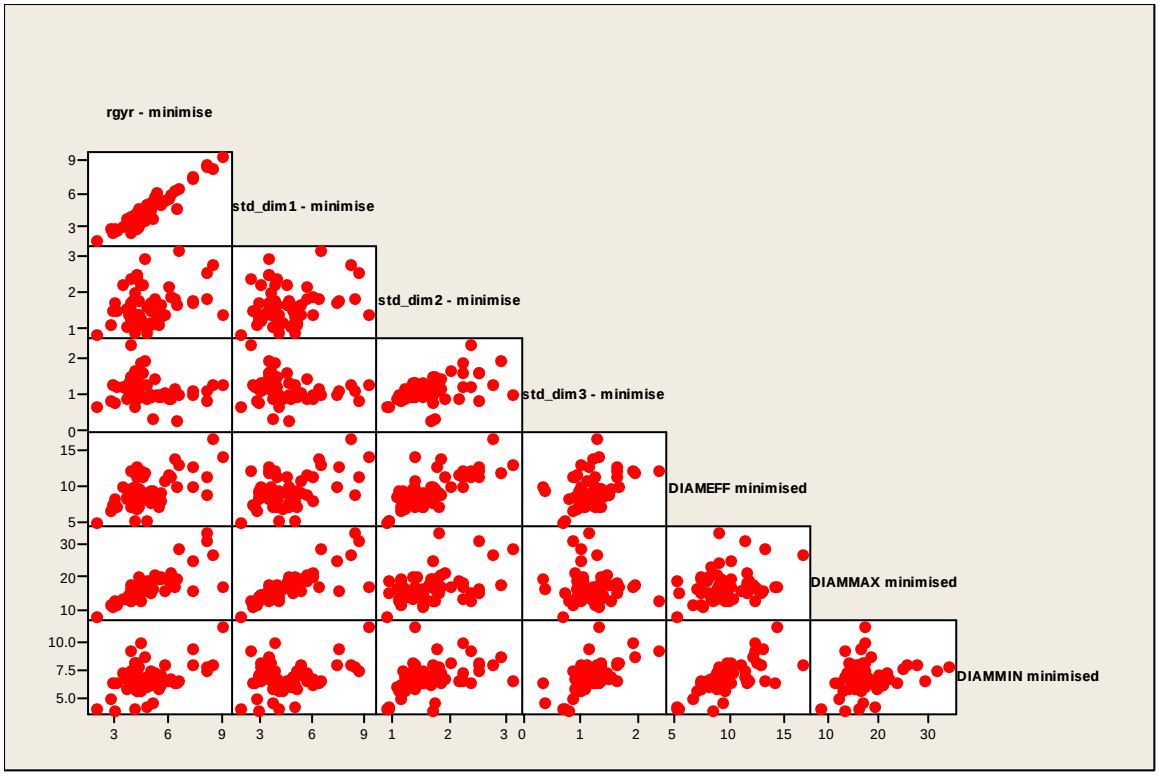


Figure D 3 Graphical representations of the relationships between MOE and OASIS descriptors from energy minimised structures.

TSAR

Table D 4 Correlation matrix for TSAR descriptors from energy minimised structures.

	IMS1	IMS2	IMS3	IML1	IML2	IML3
IMS2	0.715					
IMS3	0.764	0.995				
IML1	-0.117	0.420	0.381			
IML2	0.671	0.439	0.481	-0.149		
IML3	0.665	0.509	0.535	-0.074	0.969	
Ell Vol	0.549	0.804	0.808	0.519	0.652	0.711

Table D 5 Correlation matrix for TSAR and OASIS descriptors from energy minimised structures.

	IMS1	IMS2	IMS3	IML1	IML2	IML3	Ell Vol	D_{eff} minimum	D_{max} minimum
IMS2	0.715								
IMS3	0.764	0.995							
IML1	-0.117	0.420	0.381						
IML2	0.671	0.439	0.481	-0.149					
IML3	0.665	0.509	0.535	-0.074	0.969				
Ell Vol	0.549	0.804	0.808	0.519	0.652	0.711			
D_{eff} minimum	0.688	0.543	0.573	-0.128	0.803	0.798	0.567		
D_{max} minimum	0.352	0.744	0.725	0.608	0.349	0.411	0.704	0.303	
D_{min} minimum	0.553	0.417	0.417	-0.149	0.686	0.755	0.459	0.658	0.179

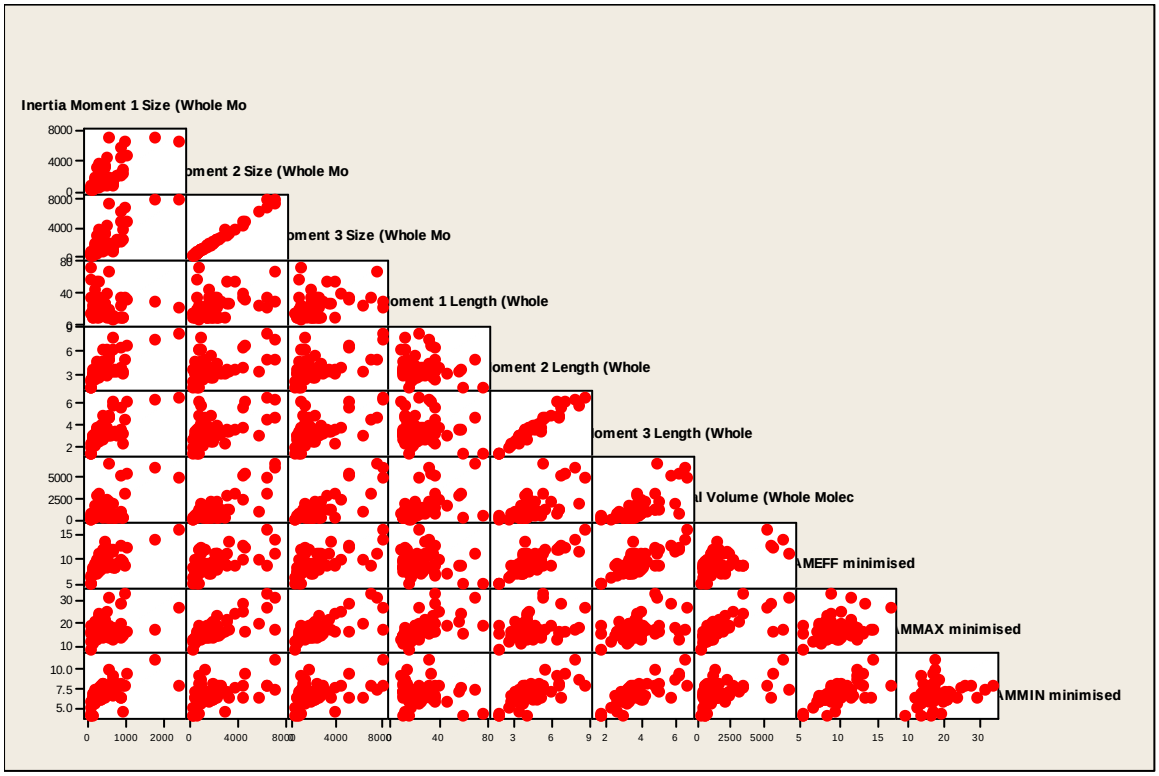


Figure D 4 Graphical representations of the relationships between TSAR and OASIS descriptors from energy minimised structures.

MOL2MOL

Table D 6 Correlation matrix for Mol2Mol descriptors from energy minimised structures.

	Mol2mol BB X	Mol2mol BB Y
Mol2mol BB Y	0.365	
Mol2mol BB Z	0.400	0.459

Table D 7 Correlation matrix for Mol2Mol and OASIS descriptors from energy minimised structures.

	Mol2mol BB X	Mol2mol BB Y	Mol2mol BB Z	D_{eff} minimum	D_{max} minimum
Mol2mol BB Y	0.365				
Mol2mol BB Z	0.400	0.459			
D_{eff} minimum	0.513	0.795	0.633		
D_{max} minimum	0.841	0.431	0.246	0.303	
D_{min} minimum	0.374	0.483	0.830	0.658	0.179

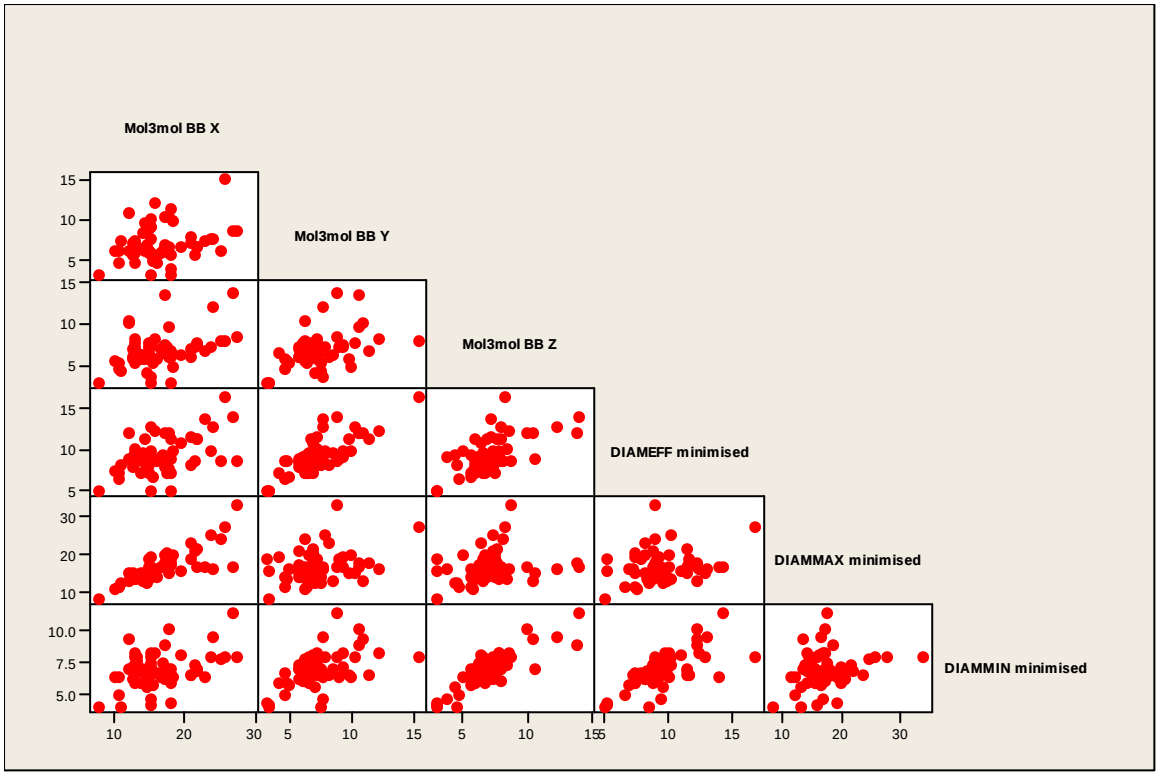


Figure D 5 Graphical representations of the relationships between Mol2Mol and OASIS descriptors from energy minimised structures.

SPARTAN

Table D 8 Correlation matrix for Spartan, Mol2Mol BB X and OASIS descriptors from energy minimised structures.

	Spartan Long	Spartan Thick	Mol2Mol BB X	D_{eff} minimised	D_{max} minimised
Spartan Thick	0.203				
Mol2Mol BB X	0.871	0.314			
D_{eff} minimised	0.220	0.840	0.513		
D_{max} minimised	0.911	0.193	0.841	0.303	
D_{min} minimised	0.258	0.782	0.374	0.658	0.179

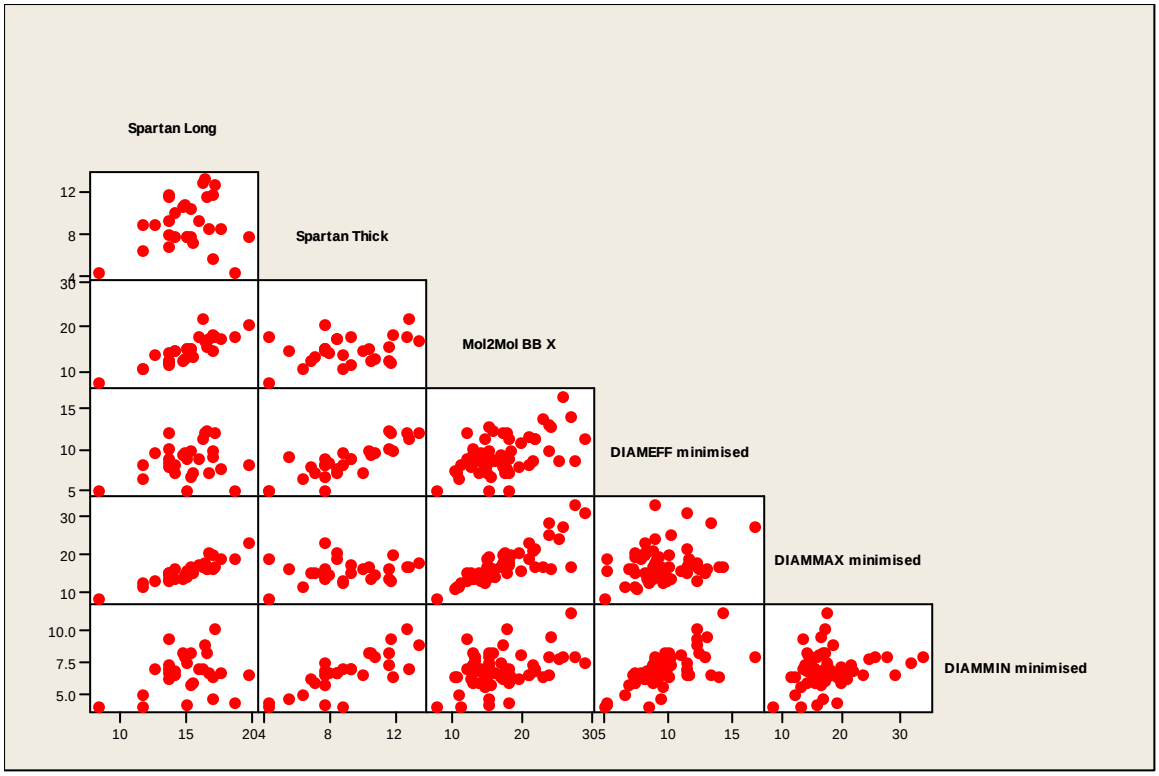


Figure D 6 Graphical representations of the relationships between Spartan, Mol2Mol BB X and OASIS descriptors from energy minimised structures.

Appendix E – Detailed Results for Conformational Analysis and Molecular Dynamics

Conformational Analysis and Molecular Dynamics

In order to investigate the effect of flexibility, conformation analysis was performed using the OASIS software and molecular dynamics with the MOE (as described in Appendices A.1 and A.2).

These analyses produce a vast amount of data. They calculate descriptors for each conformation (up to 30 per molecule in OASIS) and for 200 dynamic simulations for MOE. Full data are available if required, but the Excel files take up several megabytes of disk space and so are not included in this report. For clarity and simplicity the following measures have been identified for each of the relevant descriptors:

minimum value for all conformations
maximum value for all conformations
average value (arithmetic mean) for all conformations
value for the lowest energy (minimised) structure (for comparison)

Values for a) to d) were obtained in Microsoft Excel for each compound and summarised in a master table.

Conformational Analysis using OASIS

A full set of minimum, maximum, average and energy minimised values for D_{min}, D_{eff}, D_{max} and Max Distance for the compounds considered are given in Appendix C.

Analysis of the values for the conformers provides a good illustration of the effect of conformational analysis. Figure E 1 to Figure E 4 show the distribution of these descriptors by increasing descriptor value. The plots show the minimum and maximum values as a “range” with the average value in the range.

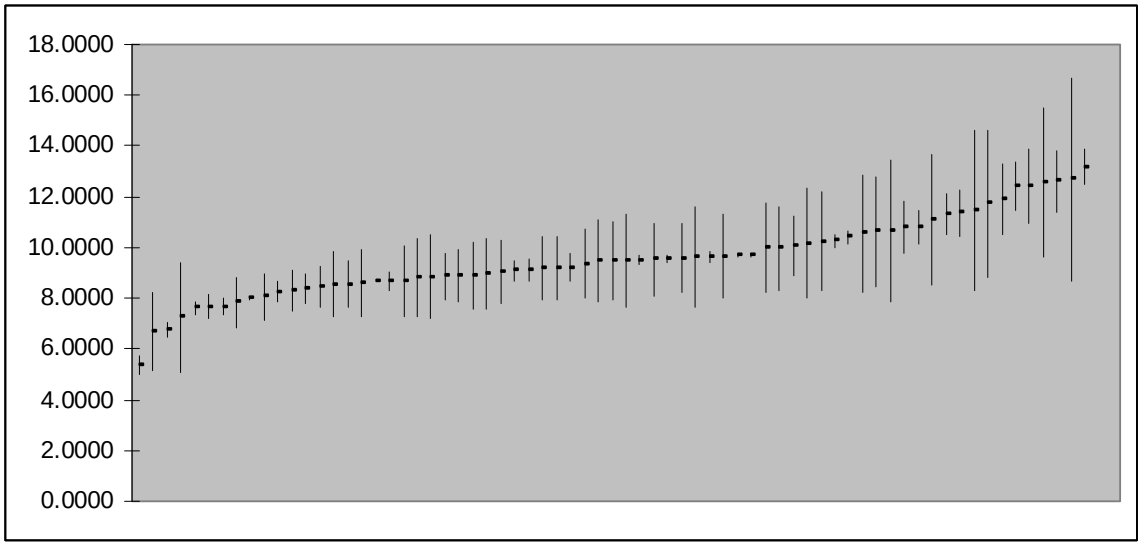


Figure E 1 Plot to illustrate the variability of the D_{eff} values (Å) for the compounds considered, with the maximum, minimum and average values for all conformers.

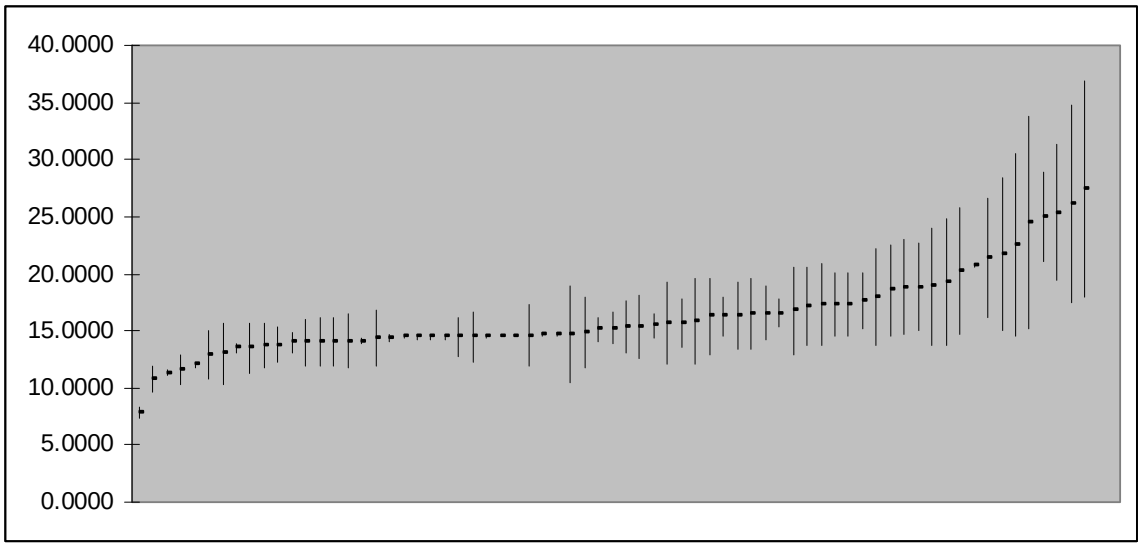


Figure E 2 Plot to illustrate the variability of the D_{max} values (Å) for the compounds considered, with the maximum, minimum and average values for all conformers.

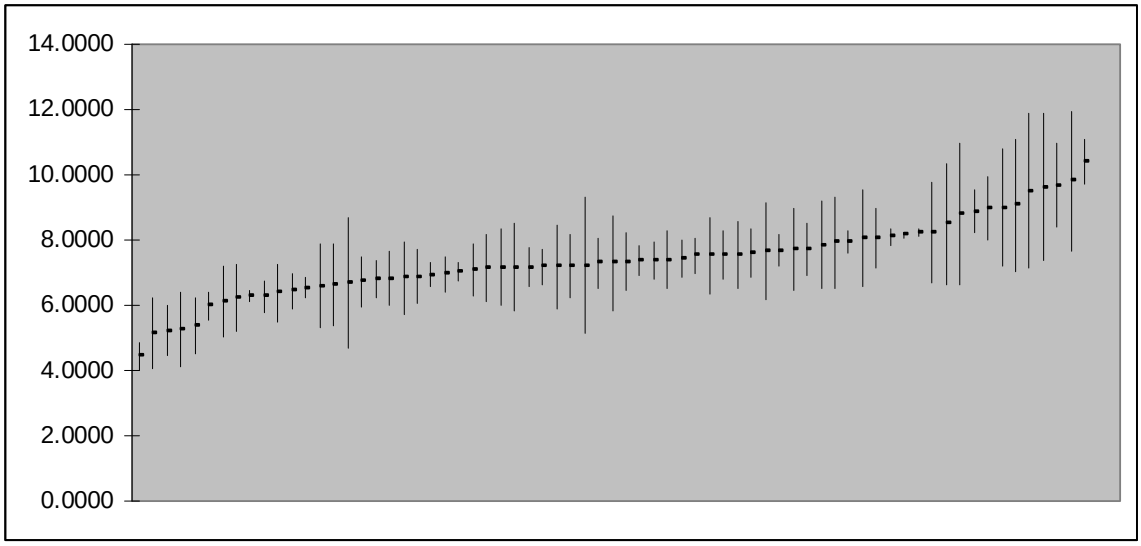


Figure E 3 Plot to illustrate the variability of the D_{min} values (Å) for the compounds considered, with the maximum, minimum and average values for all conformers.

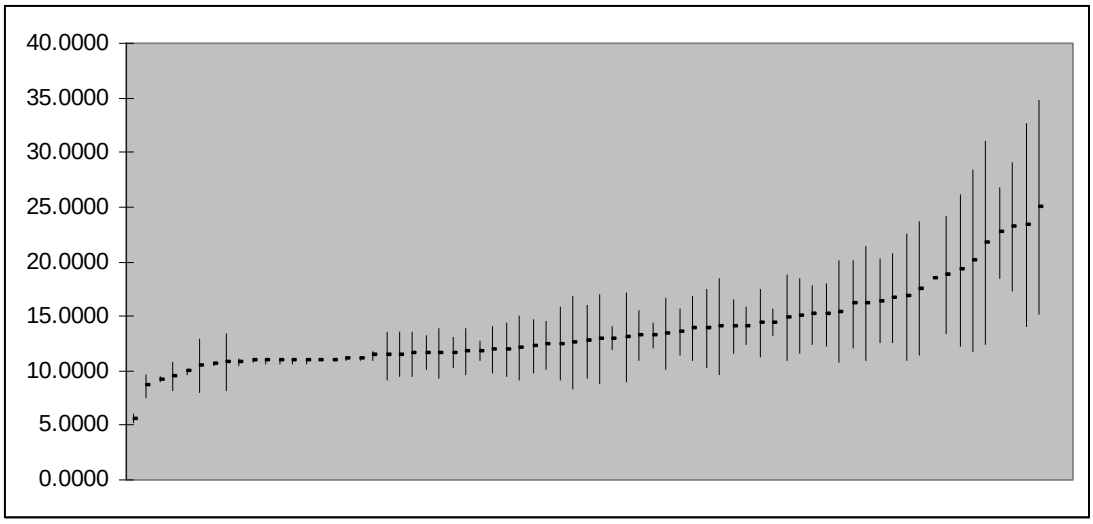
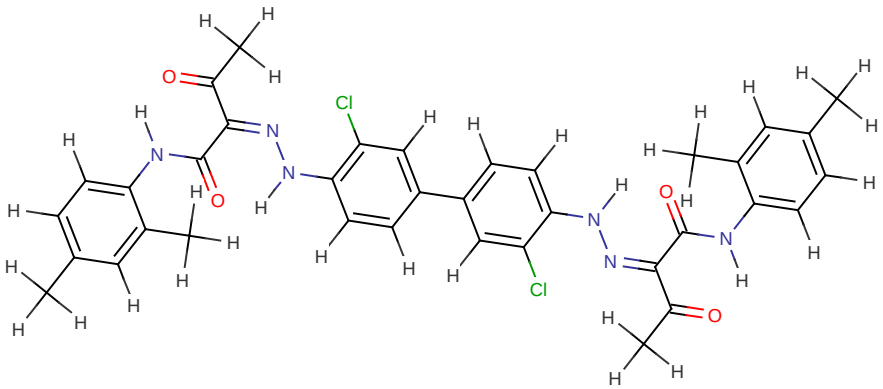
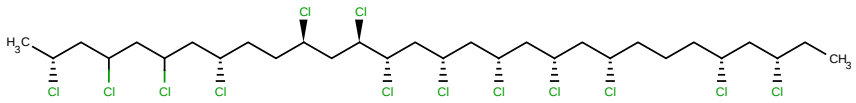
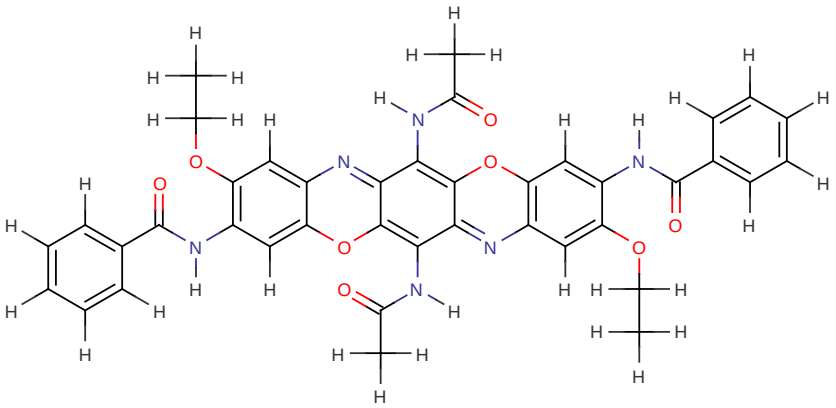


Figure E 4 Plot to illustrate the variability of the Max Distance values (Å) for the compounds considered, with the maximum, minimum and average values for all conformers.

Figure E 1 to Figure E 4 show clearly that there is a considerable effect of conformational analysis. To illustrate this, for each descriptor the ranges of parameters for the three molecules with the greatest average value of that descriptor (and hence which are known to be flexible) are recorded in Table E 1 to Table E4.

Table E 1 Following conformational analysis, the average, minimum and maximum values (together with % variation) for D_{eff} for the three compounds with the greatest average value.

ID[a]	Structure	Conf[b]	D_{eff} (Å)				
			average	Minimum	maximum	Max - min	% variation of average
D3		30	12.6008	11.3913	13.8103	2.4190	19.2
33		30	12.6756	8.6611	16.6901	8.0291	63.3
D2		30	13.1712	12.4590	13.8833	1.4242	10.8

[a]Refer to Table B in Appendix B
[b]Conf is the number of conformations considered (30 appears to be the maximum)

Table E 2 Following conformational analysis, the average, minimum and maximum values (together with % variation) for D_{max} for the three compounds with the greatest average value.

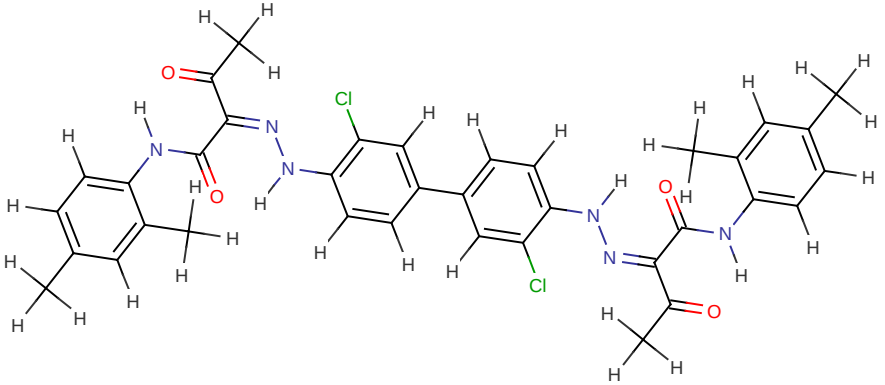
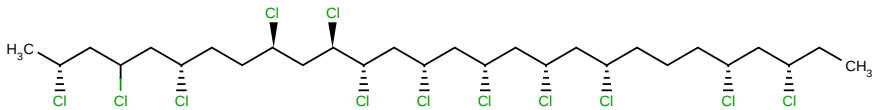
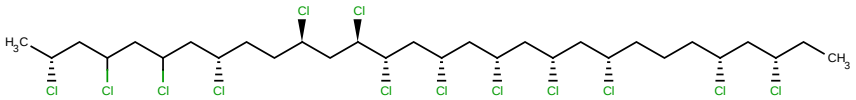
ID	Structure	Conf	D_{max} (Å)				
			average	Minimum	maximum	Max - min	% variation of average
D3		30	25.3500	19.4078	31.2921	11.8842	46.9
32		30	26.0834	17.3917	34.7752	17.3835	66.69
33		30	27.4383	17.9488	36.9278	18.9790	69.2

Table E 3 Following conformational analysis, the average, minimum and maximum values (together with % variation) for D_{min} for the three compounds with the greatest average value.

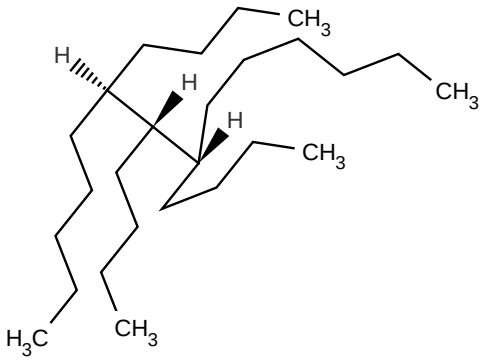
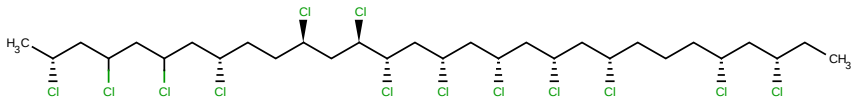
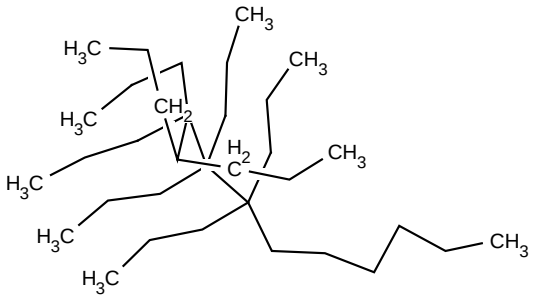
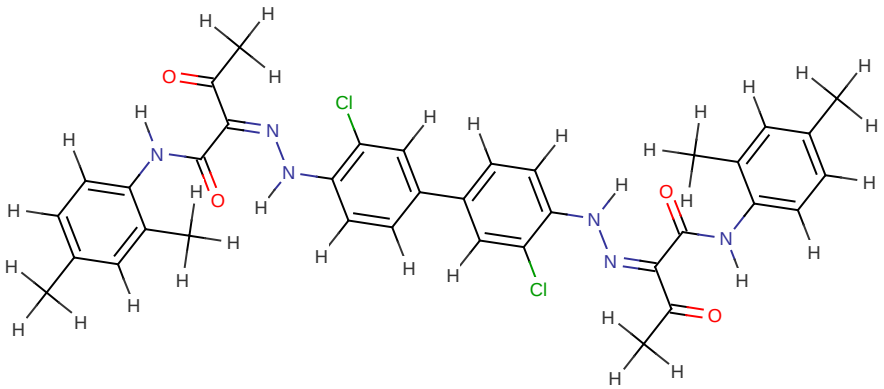
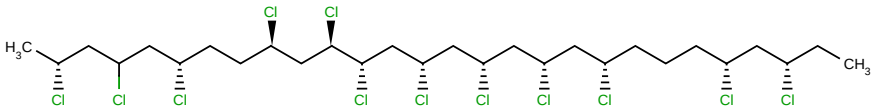
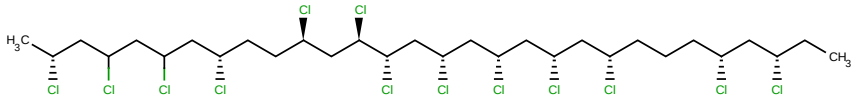
ID	Structure	Conf	D_{min} (Å)				
			average	Minimum	maximum	Max - min	% variation of average
R191		30	9.6783	8.4111	10.9454	2.5343	26.2
33		30	9.8071	7.6588	11.9554	4.2966	43.8
R190		15	10.3917	9.7062	11.0771	1.3709	13.2

Table E 4 Following conformational analysis, the average, minimum and maximum values (together with % variation) for Max Distance for the three compounds with the greatest average value.

ID	Structure	Conf[a]	Max Distance (Å)				
			average	Minimum	maximum	Max - min	% variation of average
D3		30	23.1900	17.2478	29.1321	11.8842	51.2
32		30	23.3534	14.0917	32.6152	18.5235	79.3
33		30	24.9933	15.2188	34.7678	19.549	78.2

The results in Figure E 1 to Figure E 4 and Table E 1 to Table E4 indicate that conformation flexibility can have a considerable effect of the value of descriptors. A 40% variation between the minimum and maximum average value is common and greater ranges occur.

Comparison of OASIS Descriptors following Conformational Analysis

In order to determine the effect of the different conformers, the maximum, minimum and average values for the four OASIS descriptors were analysed. These descriptors were compared against each other as well as those for the energy minimised structure. The intention of this part of the study was to determine if:

there was any relationship between the descriptors calculated for different conformers (i.e., could they be scaled to each other).
the energy minimised structure, for instance, could be used as a suitable conformer to encode the information from the other conformers.

The overall aim of this analysis was to assess which, if any, conformation was appropriate to describe the conformational flexibility of a molecule. Table E 5 to Table E 8 show the correlation matrices for the values of the descriptors for the maximum, minimum and average values and those for the energy minimised structure. These relationships are shown graphically in Figure E5 to Figure E8.

Table E 5 Correlation matrix for the minimum, maximum and average values for D_{eff} (following conformational analysis) in addition to the energy minimised structure.

	D_{eff} - minimum	D_{eff} - maximum	D_{eff} - average
D_{eff} - maximum	0.580		
D_{eff} - average	0.840	0.930	
D_{eff} - minimised	0.724	0.804	0.864

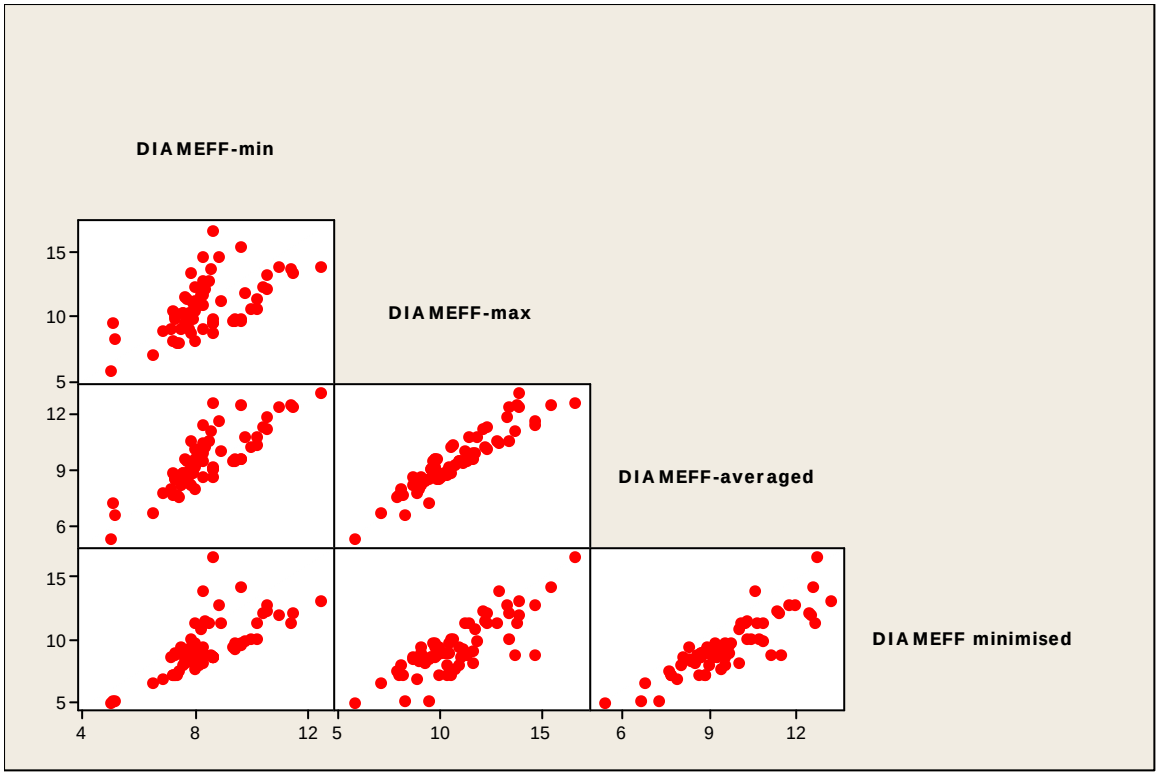


Figure E 5 Plots of the relationships between the minimum, maximum and average values for D_{eff} (following conformational analysis) in addition to the energy minimised structure.

Table E 6 Correlation matrix for the minimum, maximum and average values for D_{max} (following conformational analysis) in addition to the energy minimised structure.

	D_{max} - minimum	D_{max} - maximum	D_{max} - average
D_{max} - maximum	0.654		
D_{max} - average	0.812	0.973	
D_{max} - minimised	0.657	0.832	0.844

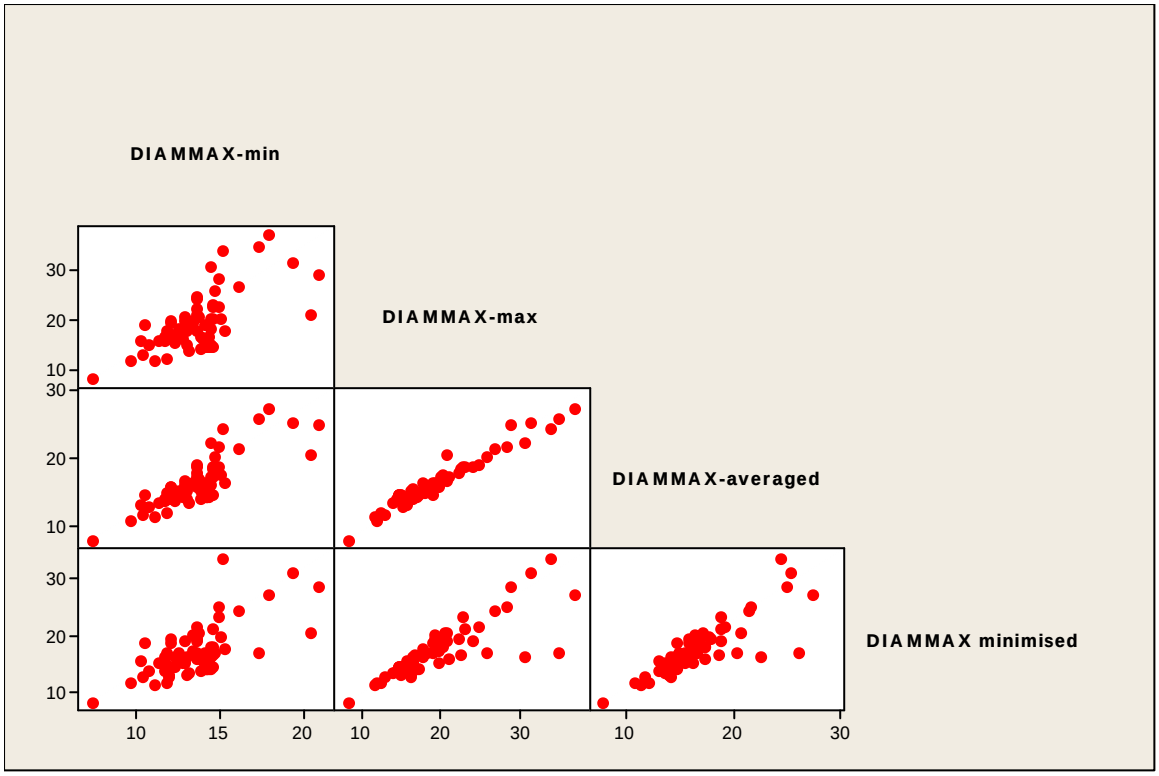


Figure E 6 Plots of the relationships between the minimum, maximum and average values for D_{max} (following conformational analysis) in addition to the energy minimised structure.

Table E 7 Correlation matrix for the minimum, maximum and average values for D_{min} (following conformational analysis) in addition to the energy minimised structure.

	D_{min} - minimum	D_{min} - maximum	D_{min} - average
D_{min} - maximum	0.652		
D_{min} - average	0.876	0.937	
D_{min} - minimised	0.826	0.729	0.845

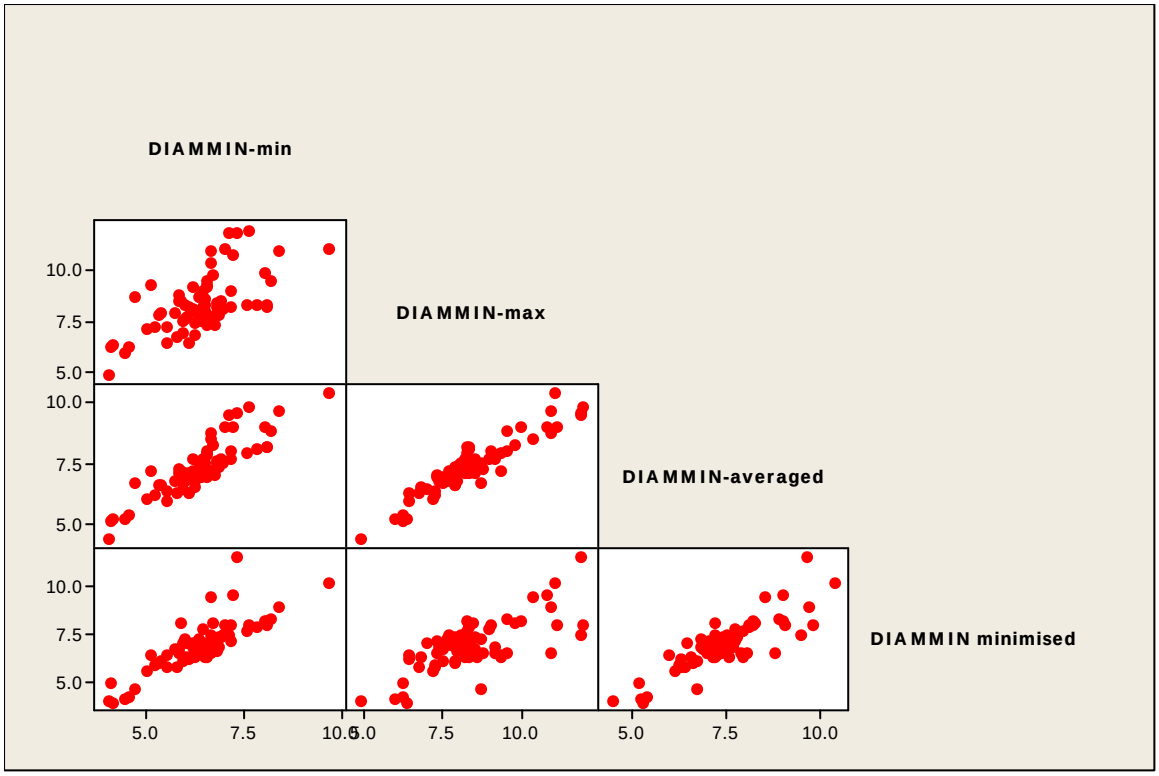


Figure E 7 Plots of the relationships between the minimum, maximum and average values for D_{min} (following conformational analysis) in addition to the energy minimised structure.

Table E 8 Correlation matrix for the minimum, maximum and average values for Max Distance (following conformational analysis) in addition to the energy minimised structure.

	Max Distance - minimum	Max Distance - maximum	Max Distance - average
Max Distance - maximum	0.650		
Max Distance - average	0.805	0.974	
Max Distance - minimised	0.669	0.826	0.844

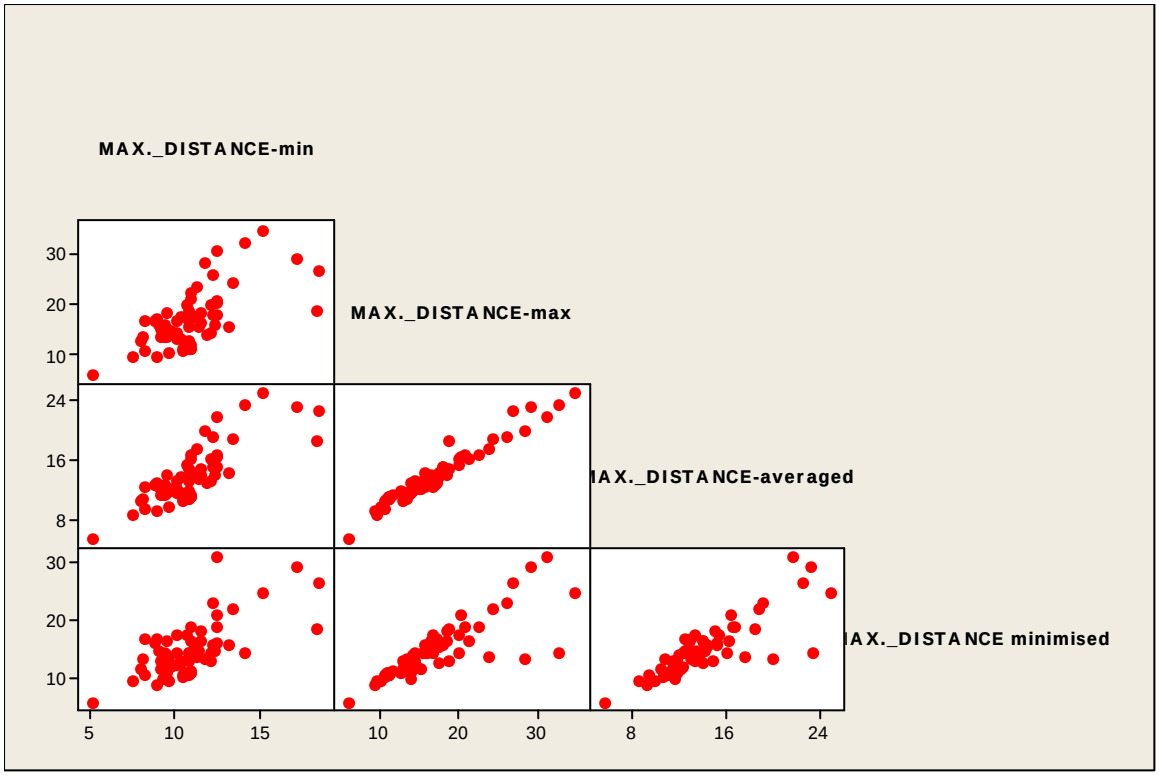


Figure E 8 Plots of the relationships between the minimum, maximum and average values for Max Distance (following conformational analysis) in addition to the energy minimised structure.

The analysis appears to confirm that for the OASIS descriptors there is a significant difference between the minimum, maximum and average values from conformational analysis. The variation appears to be greatest for the largest compounds, which reflects their greater intrinsic flexibility. Therefore, it is not possible to determine which is the correct conformation to use to represent the dimensions of the molecule.

For all descriptors there appears to be a good relationship between the values for the energy minimised structures and the maximum value. Therefore, the value for the energy minimised structure could act as a surrogate for the maximum value and negate the requirement for conformational analysis.

Molecular Dynamics using MOE

A full set of minimum, maximum, average and energy minimised values for rgyr, std_dim1, std_dim2 and std_dim3 for the compounds considered are given in Appendix C.

Analysis of the values for the conformers provides a good illustration of the effect of molecular dynamics. Figure E 9 to Figure E 12 show the distribution of these descriptors by increasing descriptor value. The plots show the minimum and maximum values as a “range” with the average value in the range.

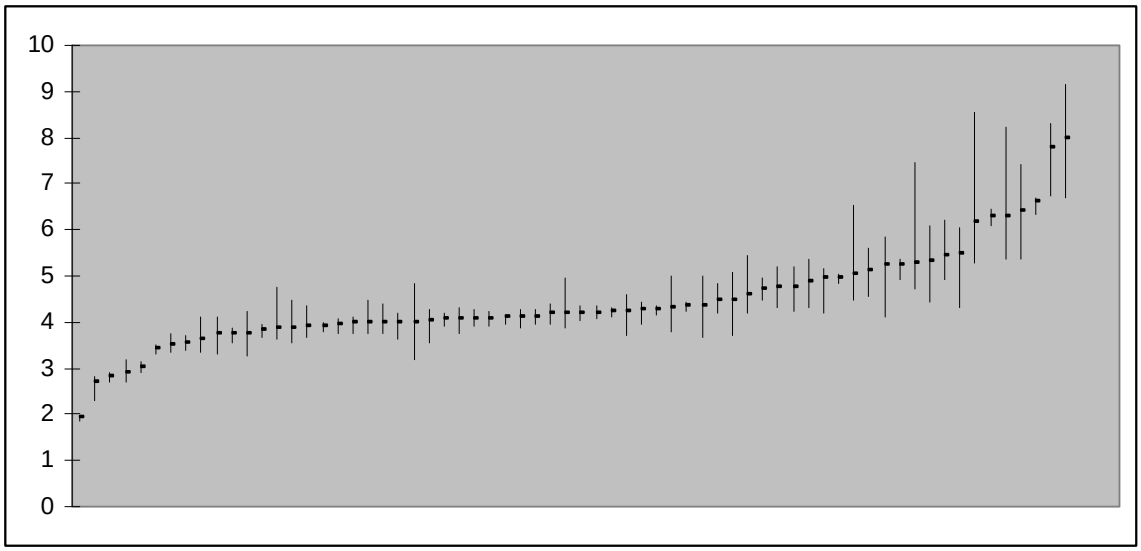


Figure E 9 Plot to illustrate the variability of rgyr values (Å) for the compounds, with the maximum, minimum and average values for all conformers.

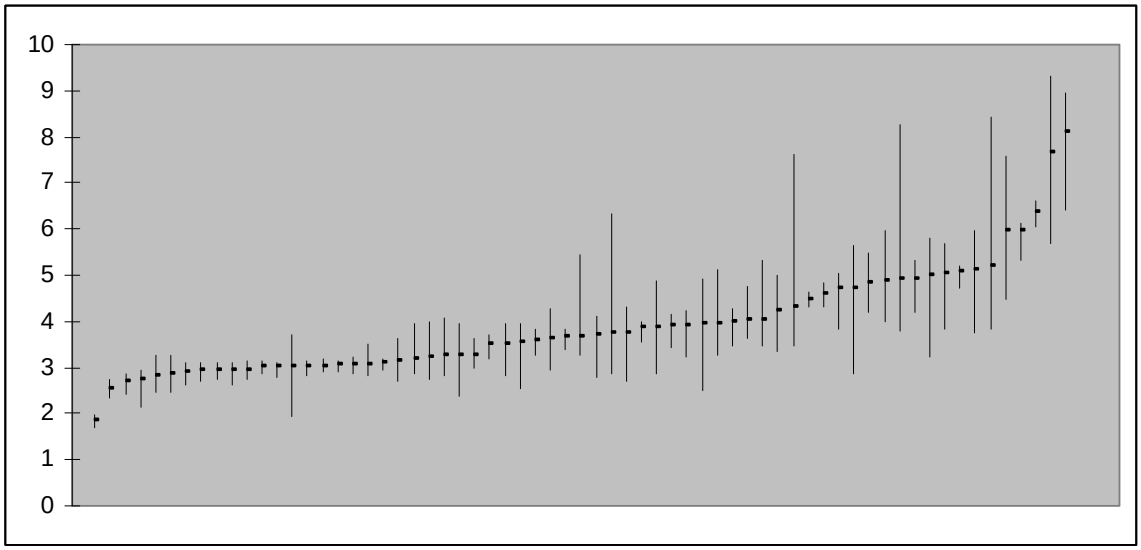


Figure E 10 Plot to illustrate the variability of std_dim1 values (Å) for the compounds, with the maximum, minimum and average values for all conformers.

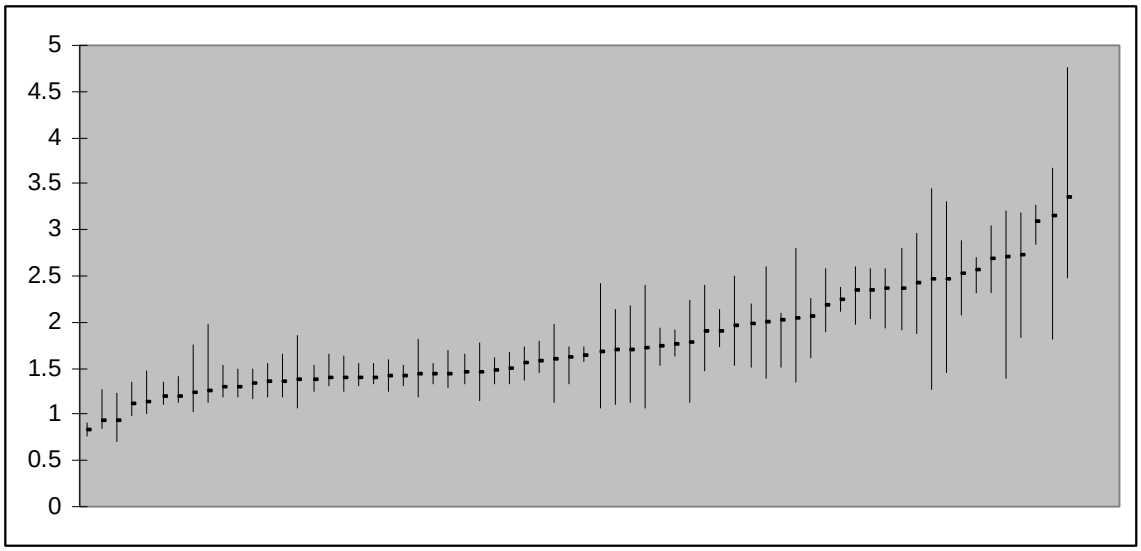


Figure E 11 Plot to illustrate the variability of std_dim2 values (Å) for the compounds, with the maximum, minimum and average values for all conformers.

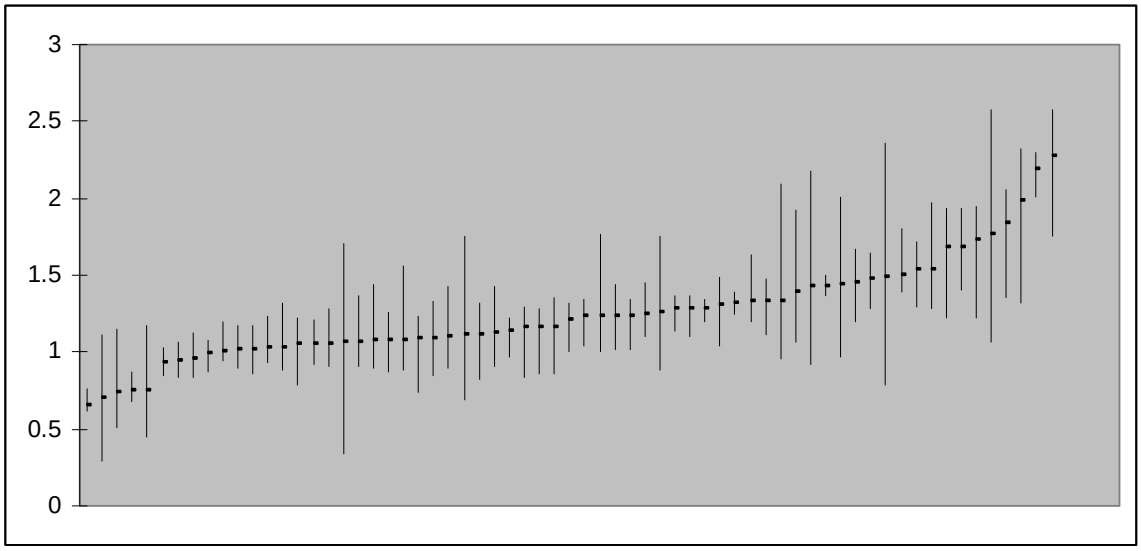


Figure E 12 Plot to illustrate the variability of std_dim3 values (Å) for the compounds, with the maximum, minimum and average values for all conformers.

Comparison of MOE Descriptors following Molecular Dynamics

In order to determine the effect of the molecular dynamics procedure, the maximum, minimum and average values for the four MOE descriptors were analysed. These descriptors were compared against each other and those for the energy minimised structure. Table E 9 to Table E 12 show the correlation matrices for the values of the descriptors for the maximum, minimum and average values as well as those for the energy minimised structure; these relationship are also shown graphically in Figure E 13 to Figure E 16.

Table E 9 Correlation matrix for the minimum, maximum and average values for rgyr (following conformational analysis) in addition to the energy minimised structure.

	rgyr max	rgyr min	rgyr ave
rgyr min	0.888		
rgyr ave	0.952	0.969	
rgyr - minimised	0.996	0.883	0.947

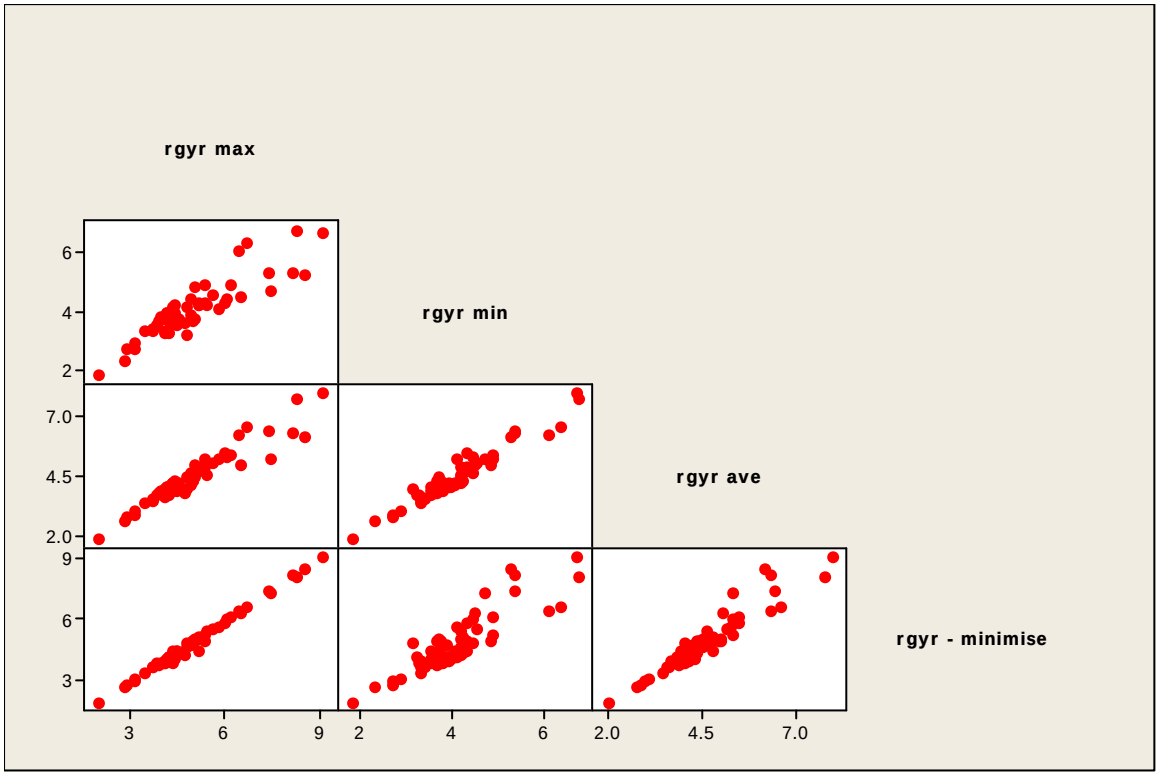


Figure E 13 Plots of the relationships between the minimum, maximum and average values for rgyr (following molecular dynamics) in addition to the energy minimised structure.

Table E 10 Correlation matrix for the minimum, maximum and average values for std_dim1 (following conformational analysis) in addition to the energy minimised structure.

	std_dim1 maximum	std_dim1 minimum	std_dim1 average
std_dim1 minimum	0.765		
std_dim1 average	0.902	0.926	
std_dim1 - minimised	0.961	0.694	0.846

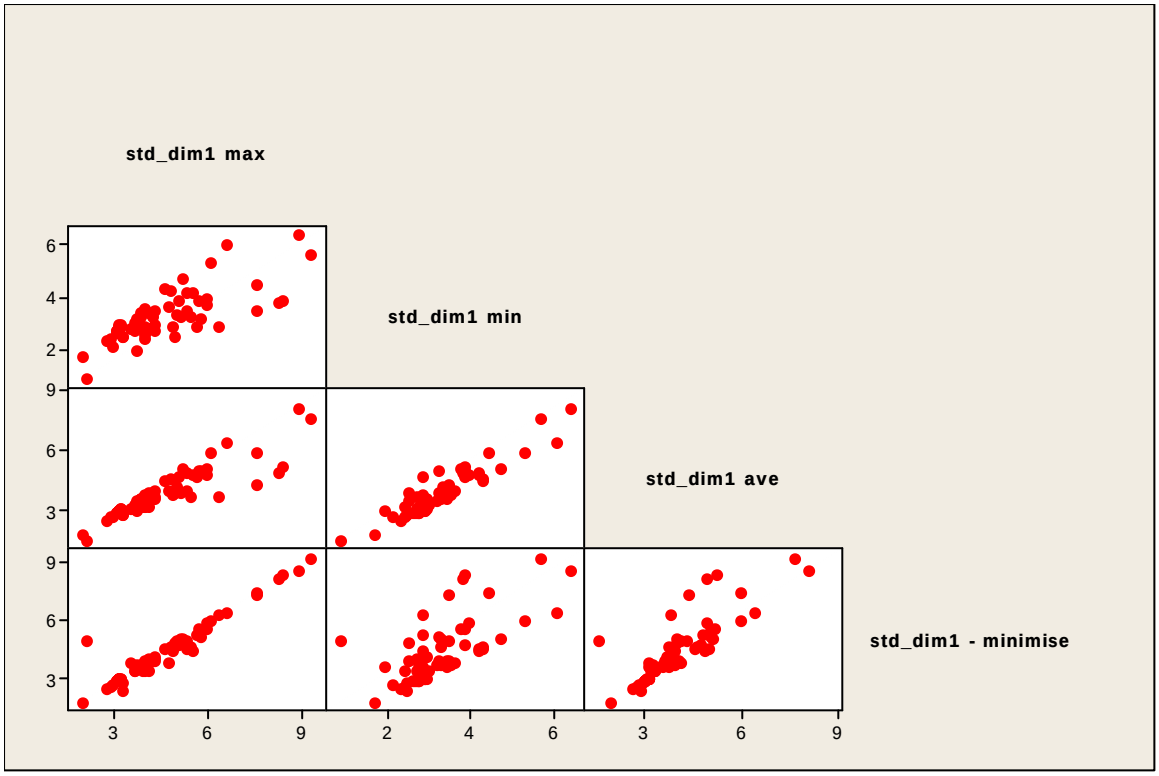


Figure E 14 Plots of the relationships between the minimum, maximum and average values for std_dim1 (following molecular dynamics) in addition to the energy minimised structure.

Table E 11 Correlation matrix for the minimum, maximum and average values for std_dim2 (following conformational analysis) in addition to the energy minimised structure.

	std_dim2 maximum	std_dim2 minimum	std_dim2 average
std_dim2 minimum	0.712		
std_dim2 average	0.951	0.851	
std_dim2 - minimised	0.741	0.935	0.840

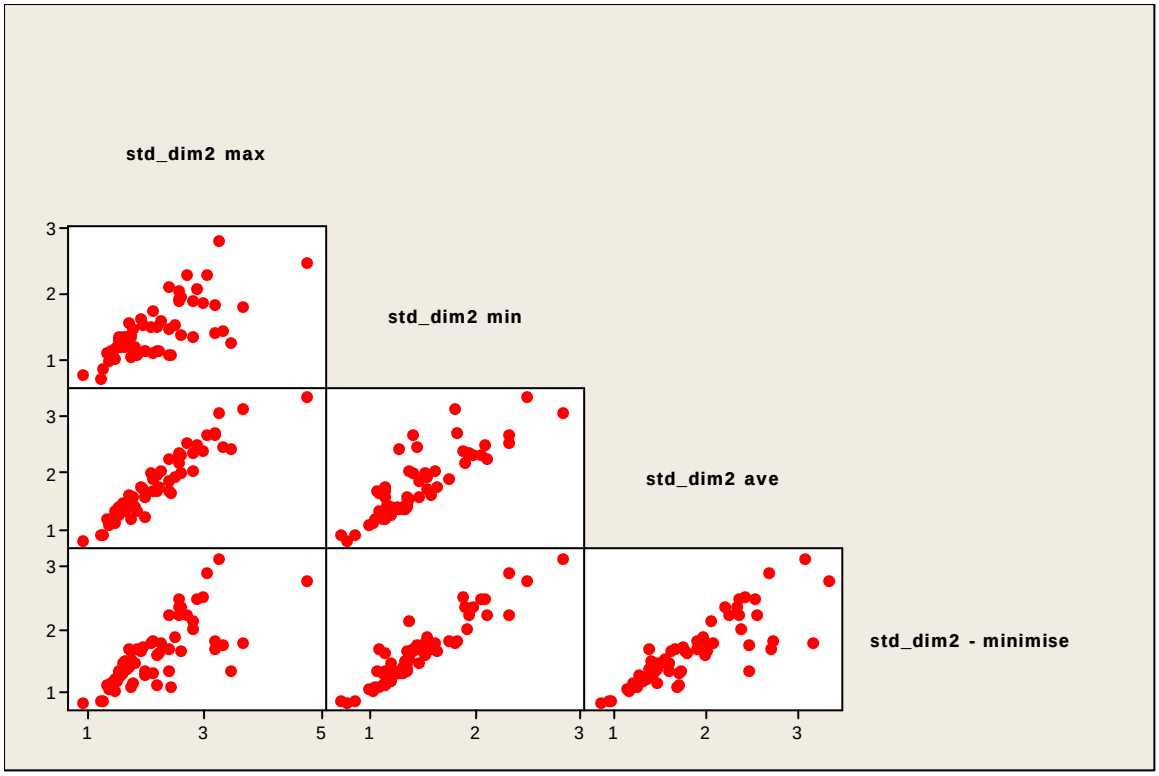


Figure E 15 Plots of the relationships between the minimum, maximum and average values for std_dim2 (following molecular dynamics) in addition to the energy minimised structure.

Table E 12 Correlation matrix for the minimum, maximum and average values for std_dim3 (following conformational analysis) in addition to the energy minimised structure.

	std_dim3 maximum	std_dim3 minimum	std_dim3 average
std_dim3 minimum	0.578		
std_dim3 average	0.873	0.861	
std_dim3 - minimised	0.576	0.858	0.836

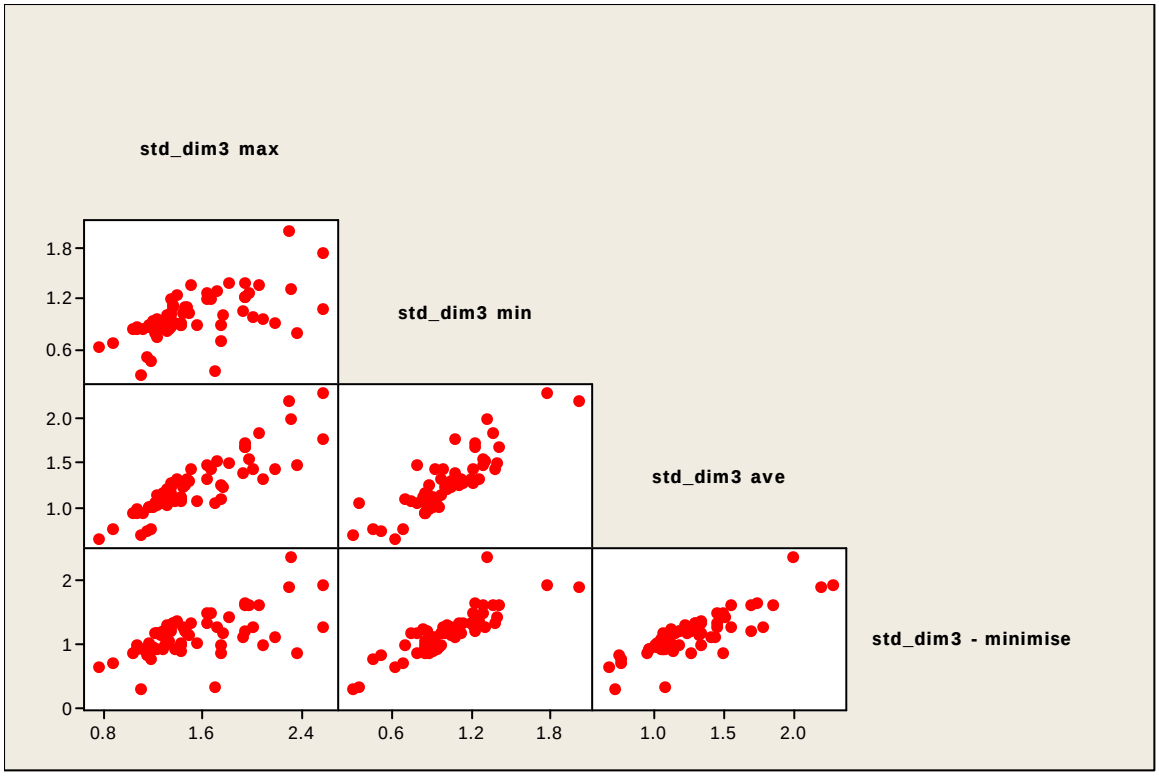


Figure E 16 Plots of the relationships between the minimum, maximum and average values for std_dim3 (following molecular dynamics) in addition to the energy minimised structure.

The analysis shows (Figure E 13) that the descriptor rgyr is little affected by the variation in conformers following molecular dynamics. This may be expected as this is not a true measure of molecular dimensions, but a “globularity” measure which may be less susceptible to conformers.

The other descriptors calculated by MOE (std_dim1, std_dim2, std_dim3) do, however, show a significant difference between the minimum, maximum and average values from molecular dynamics. The variation appears to be greatest for the largest compounds, which reflects their greater intrinsic flexibility. Therefore, it is not possible to determine which conformation to use to correctly represent the dimensions of the molecule.

We welcome views from our users, stakeholders and the public, including comments about the content and presentation of this report. If you are happy with our service, please tell us about it. It helps us to identify good practice and rewards our staff. If you are unhappy with our service, please let us know how we can improve it.

We are The Environment Agency. It's our job to look after your environment and make it **a better place** – for you, and for future generations.

Your environment is the air you breathe, the water you drink and the ground you walk on. Working with business, Government and society as a whole, we are making your environment cleaner and healthier.

The Environment Agency. Out there, making your environment a better place.

Published by:

Environment Agency
Rio House
Waterside Drive, Aztec West
Almondsbury, Bristol BS32 4UD
Tel: 0870 8506506
Email: enquiries@environment-agency.gov.uk
www.environment-agency.gov.uk