MS22.01.05 A DATABASE STUDY OF THE BONDING AND
CONFORMATION OF BIS-SULFONYLAMIDES AND
IMIDES. By Alajos Kálman* and Petra Bambociz, Central
Research Institute for Chemistry, Hungarian Academy of Sciences,
H-1525 Budapest 114, P.O.Box 17, Hungary.

The bonding and conformational characteristics of bis-
sulfonamides and analogous imides are compared. Structures
(44 altogether) of type R-SO₂-NQ-SO₂-R' (R, R' = Me, Et, or Aryl)
were retrieved from the Cambridge Structural Database (March
1994 release, 120481 entries). They are either neutral (Q = H, alkyl
and aryl group or hetero atom like O and S) or charged (Q = +
or metal, like In, Hg, etc.).

- Analysis of the S-N bonds varying in the range 1.57-1.76 Å
  and their observed influence upon the other S-X and X-S-Y
  parameters enabled us:
  1. to characterize the S-N bonds against the differences in
     the nitrogen environment (charged N, N-metal and N-covalent bonds),
  2. to probe the applicability of the bond order conservation
     principle (Johnston, 1961) and the VSEPR rules (Gillespie, 1963)
     on the S(VI) [O.O', N.C] tetrahedra,
  3. to describe the intra- and intermolecular conditions of the
     folded conformation vs. the energetically favourable open form of
     the C-SO₂-NQ-SO₂-C moieties,
  4. to substantiate the inequalities of the O-S-N angles
     which is attributable to the synclinal position of one of the
     S-O bonds with N-lone pair (Kálman et al., 1981).

The marked difference between the archetypes of the S-N
bonds indicates that the interdependence of the S-X bonds in the
NSO₂ tetrahedra are principally governed by the environment of
the nitrogen atoms. (Sponsor: Hungarian Res. Fund, Grant No.
OTKA T014539).

Cryst., B37, 868-877.

MS22.01.06 AUTOMATIC DETERMINATION OF STRUCTURAL
SUBCLASSES. Shishan Guo, Suzanne Fortier, Janice
I. Glasgow, Chemistry Department and Computing and Information
Science, Queen's University, Kingston, Ontario, Canada K7L 3N6

With the rapid growth of crystallographic databases, fully
automatic methods for mining knowledge from these databases
are needed. Several classification algorithms are already incorpo-
rated into the databases. While these have greatly facilitated
the analysis and classification of datasets, considerable user interven-
tion is still required. For example, extensive examination of the
dataset may be needed for the selection of clustering algorithm,
data parameters, similarity measure, similarity threshold, stopping
point, etc. Furthermore, different choices of algorithms and metrics
often yield different results. It is thus important to evaluate the
robustness of the results and assess their possible dependence on
artifacts of the approach used. Thus, a fully automated classifica-
tion approach requires methods for both pre-classification data
preparation and post-classification result assessment. This contribu-
tion presents a method for the automatic determination of struc-
tural subclasses in datasets retrieved from the CSD. Subclasses/
clusters are obtained by undergoing a comprehensive automated
data preparation which is followed by applying clustering algorithms
and then by undergoing post-clustering evaluation of the results.
The automatic pre-processing component is based on a comprehensive
analysis of histograms and scattergrams generated for potential
classification parameters. This process helps identify informative
parameters and gives a preliminary clustering of the dataset. For
postclassification evaluation, plots of a clustering similarity index
are used to assess how the results are affected by different algo-
rithms and by the introduction of random noise into the dataset.
These plots help understand the nature of the datasets being
analysed by revealing characteristic features associated with the
degree of overlap among the subclasses and by identifying where
maximum similarity occurs. Application of the automatic classifi-
cation approach to four representative datasets - valine, hexas-
opyronose sugars, steroid side-chains and six-membered rings
will be presented.

PS22.01.07 PLUTO: CRYSTAL PACKING ENERGIES IN
THE CAMBRIDGE STRUCTURAL DATABASE. Sam
Motherwell, Cambridge Crystallographic Data Centre, 12 Union
Rd., Cambridge, CB2 1EZ. U.K.

The Pluto program has been enhanced to allow visual
exploration of the packing energy environment of organic
molecules. Facilities allow selection of the N strongest interactions.
This allows comparison of structures on the basis of the packing
energy enclosure shell about a molecule. The energy is calculated
by the empirical atom-pair method, using the tables collected by
Gavezzotti and others. This allows one to easily compare energy
shells for molecules in polymorphs, and in general examine packing
patterns on an energy basis. Selective display is allowed at a given
energy threshold revealing the most important interactions in the
packing. It is also possible to pick any pair of molecules in the
packing diagram for energy display. Molecules may be simplified
to molecular centre points, or spheres, or enclosure surfaces as
ellipsoids or contact probed by a hydrogen atom. Coordinate
information may be presented in a simple free format, or the
standard Cambridge Structural Database FDAT. or SHELX.

Example. Benzene interactions < -1 Kcal/mole

PS22.01.08 PREQUEST: A DATA INPUT PROGRAM FOR
THE CAMBRIDGE STRUCTURAL DATABASE. Paul
Edgington, Clare Macrane, Sam Motherwell, Cambridge Crystal-
lographic Data Centre, 12 Union Rd., Cambridge, CB2 1EZ. UK.

The Prequest program enables users to create new entries for
the Cambridge Structural Database (CSD) from common input
formats such as CIF and SHELX. Important checks are made for
validity and self-consistency of data fields, enabling users to add
entries to a local (private) CSD database fully searchable with
Prequest. This gives the user a method for quickly comparing molec-
ular geometry parameters of current experimental results with the
main CSD.

The program also prepares the way for trouble-free deposit-
ing of data with the Cambridge Crystallographic Data Centre
(CCDC), as private communications to be added to the widely
distributed main CSD file.

The use of the CIF input format is now well established.