s8b.m4.o1 **Structural systematics in molecular inorganic chemistry**. A. G. Orpen, *School of Chemistry, University of Bristol, BRISTOL BS8 1TS, U.K*.
Keywords: methods crystallography, data mining, data bases.

There are now over 200,000 structures in the Cambridge Structural Database (CSD)[1] of which *ca*. 60% are of molecular inorganic species. As such the CSD is an unrivalled source on information on intramolecular and intermolecular geometry in this aspect of chemistry.

Areas of molecular inorganic chemistry to which the CSD have been usefully applied include the following:

- Typical molecular dimensions[2]
- Molecular flexibility of metal complexes[3]
- Reaction pathways in metal cluster complexes[4]
- Bonding in metal phosphine complexes[4]
- Conformational analysis in metal phosphines[5]
- Intermolecular bonding:
  Secondary bonding in main group element chemistry[6]
  Metal-assisted hydrogen bonding[7]

Examples of these applications and the new science resulting will be discussed.

[1] F.H. Allen *et al.*, *Acc. Chem. Res.*, 1983, **16**, 146; F.H. Allen *et al.*, *J. Chem. Inf. Comput. Sci.,* 1987, **31**, 187.
[2] A.G. Orpen, *Acta Crystallogr., Sect. D*, 1998, **54**, 1194-1198.
[3] A. Martín and A.G. Orpen, *J. Am. Chem. Soc.*, 1996, **118**, 1464-1470.
[4] A.G. Orpen *Chem. Soc. Revs*, 1993, 191-197.
[5] J.J Barker and A.G. Orpen. *Acta Crystallogr., Sect. B*, 1999, **55**, 203-208.
[6] J. Starbuck, N.C. Norman, A.G. Orpen. *New J. Chem*, 1999, **23**, 969-972.
[7] G. Aullón, D. Bellamy, L. Brammer, E. A. Bruton and A.G. Orpen. *Chem. Commun.,* 1998, 653-4.

s8b.m4.o2 **Recognition of spatial motifs in protein structures**. D. Madsen, G. J. Kleywegt, *Dept. of Cell and Molecular Biology, Uppsala University, Biomedical Centre, Box 596, SE-75421 Uppsala, Sweden*.
Keywords: methods crystallography, data mining, data bases.

The number of biomacromolecular structures in the Protein Data Bank grows near-exponentially. In order to exploit the enormous amount of raw structural information, new methods to organize and access it need to be developed. One example of such new methods are those embodied in the program package SPASM. It uses a compact representation of various subsets of the structures available in the PDB, allowing users to query the database to determine if a small "motif" (the spatial arrangement of a small set of residues, including their main chain or side chain atoms or both) occurs in any other know structures. Hence, the program facilitates a 3D counterpart to the detection of ProSite-style patterns in protein sequences. Also, SPASM can query its database with "fuzzy patterns" (which have relaxed requirements on the nature of the matched residues, for instance an Asp might be matched to an Asp or Glu). Recently, a new program has been developed that does full atom-by-atom superpositioning of the user's motif and each of the hits found in the database. This makes that SPASM now provides a fast, reliable and convenient method to answer the question "has this protein structural motif been observed before ?".

Recently, a web-interface to the program has been developed (a public demo-version is available at a relatively slow server:

http://xray.bmc.uu.se/cgi-bin/dennis/spasm.pl). With the interface the user is guided from the input of a small PDB file (that contains the spatial motif of interest) to displaying the user's selection of the hits with Rasmol (or similar programs).

The algorithm and data structure of the programs will be described and a few applications will be discussed.