PRECISION AND RELIABILITY IN MOLECULAR STRUCTURE DETERMINATION

J. Hašek

Institute of Macromolecular Chemistry, Academy of Sciences of CR, Heyrovského nám.2, 162 06 Praha 6, Czech Republic

Protein structure database (PDB) [1] is a primary source of information about the structure of biological macromolecules. It contains almost 20.000 of experimentally determined structures. About 15 % of them are determined by NMR techniques, 85 % by X-ray crystallography (see Tab.1).

In spite of the fact that the final calculations and refinement of the structure with NMR and X-ray data can be performed with the same computer program (e.g. XPLOR, CNS [2]) there are significant differences in meaning and also in presentation of structure.

NMR measurement. Roughly speaking, the most important information from NMR measurement is identification of atoms laying near (3 - 6 Å) each other in spite of the fact that they are far along the chain. The final description structure is obtained by searching for all molecular models satisfying these experimental restrains using the methods of molecular modelling. Thus, generally speaking, the reliability of the structure model is derived from a completeness of the experimental restrains, and the exact atom coordinates are optimized by methods of molecular modelling. The molecular structure in PDB is described as a number of individual structures often interpreted as snapshots of a molecule in movement.

X-ray diffraction experiment. The primary result of X-ray diffraction experiment is a map of electron density averaged over time of measurement and all structure units in crystal. However, it is really never published in this form. The atomic coordinates send to the PDB are determined as centers of electron density of individual atoms. Moving parts of molecule correspond to areas with low or smashed electron density. At this moment, the X-ray scientist starts to look for several alternative conformations which are all refined under the restrain that the sum of occupation factors is 1. Thus the information about molecular movement is hidden in a single file of atom coordinates as alternative conformations for individual side chains and also as temperature factors B [3] describing the mean atomic displacement u(Å) around the mean positions of individual atoms . The dependence of B on the mean atomic displacement u (Å) is ilustrated in Tab. 2.

Another term sometimes misunderstood is resolution. The precision of atomic positions is not a simple function of resolution and depends on more factors. An approximate relation between the expected standard deviation of atomic position and the resolution is illustrated in Tab.3.

The talk will show complementarity of X-ray and NMR techniques and some rules for working with data obtained by X-ray crystallography.

1. Protein DataBank (PDB). Research Collaboratory for Structural Bioinformatics (RCSB) http://rutgers.rcsb.org/pdb/.
2. Giacovazzo C., Monaco H.L., Viterbo D., Scordari F., Gilli G., Zanotti G., Catti M. Fundamentals of Crystallography. Oxford University Press, 2000.
3. International Tables for Crystallography. Crystallography of Biological Mactromolecules. Vol.F. Kluwer Acad.Publ., Dorndrecht 1999.

Table 1. Number of structures of biological macromolecules deposited in the Protein structure database. The theoretical models are not being collected in the PDB since 2002 *(information taken from the PDB Holdings List: 04-Feb-2003)*

	Proteins Viruses	Protein/NA Complexes	Nucleic Acids	Carbohydrates	Details of measurement	total
X-ray	15.507	734	638	14	8.755(i)	16.893
NMR	2.481	89	496	4	1.457(ii)	3.070
Total	17.988	823	1.134	18	10.212	19.963

(i) Deposition of reflection intensities, (ii) Deposition of restrains gained by NMR measurement

Table 2. Exact relation between the temperature factor B (Å^2) and the effective atom width (the mean atomic displacement u (Å)). $B = 8\ ^2\langle u^2\rangle$.

B (Å)	4	8	16	32	64	128
Mean atomic displacement (Å)	0.23	0.32	0.45	0.64	0.90	1.80

Table 3. Typical average coordinate inaccuracy $\langle\ _x\rangle$ (mean expected standard deviation) as a function of the limit for diffraction measurement (resolution). Data collected from randomly selected structures found in literature.

Resolution (Å)	5	3.0	2.4	1.9	1.5	1.3	1.0	0.8	0.6
Expected e.s.d. (Å)	> 3	0.7	0.4	0.2	0.1	0.07	0.05	0.03	0.01