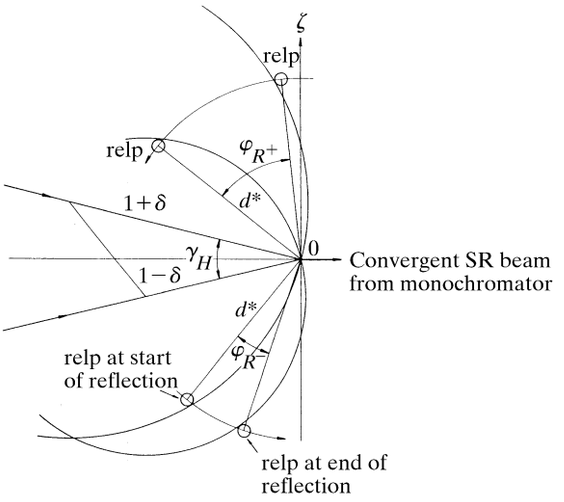


Fig. 8.1.7.3. The rocking width of an individual reflection for the case of Fig. 8.1.7.1(*c*) and a vertical rotation axis. From Greenhough & Helliwell (1982). Copyright (1982) International Union of Crystallography.

perpendicular (vertical) rotation axis, whereas the formula for $\varphi_R$ above is for a horizontal axis. This is purely for didactic reasons since the interrelationship of the components is then much clearer.

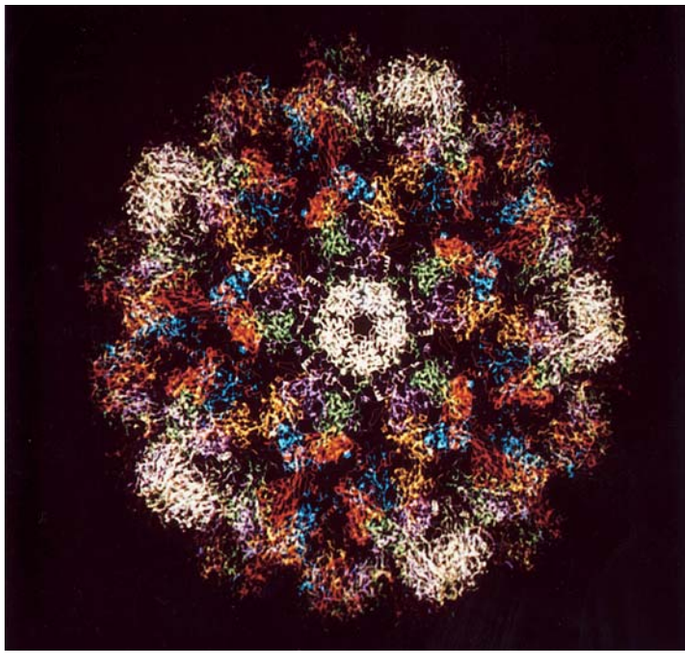

Fig. 8.1.8.2. A view of SV40 virus (based on Liddington *et al.*, 1991) determined using data recorded at the SRS wiggler station 9.6 (Fig. 8.1.4.1*a*).

### 8.1.8. Scientific utilization of SR in protein crystallography

There are a myriad of applications and results of the use of SR in crystallography. Helliwell (1992) has produced an extensive survey and tabulations of SR and macromolecular crystallography applications; Chapter 9 therein concentrates on anomalous scattering and Chapter 10 on high resolution, large unit cells, small crystals, weak scattering efficiency and time-resolved data collection. The field has expanded so dramatically, in fact, that an equivalent survey today would be vast. Table 8.1.4.1 lists the home pages of the facilities, where the specifications and details of the beamlines can be found (*e.g.* all the publications at Daresbury in the protein crystallography area, commencing with NINA in 1976, are to be found at http://www.dl.ac.uk/srs/px/publications.html). The examples below cite extreme cases of the largest unit cell (virus and

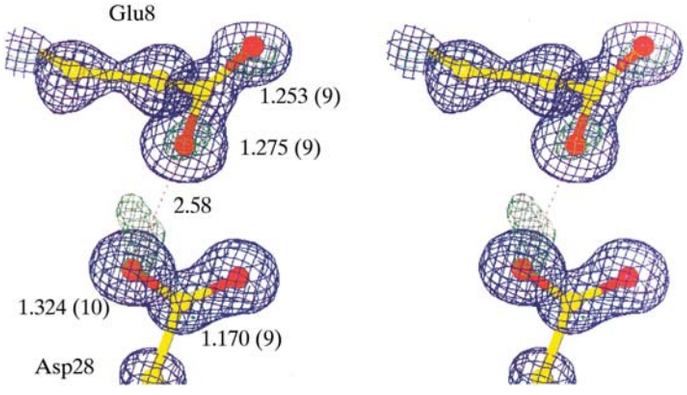


Fig. 8.1.8.1. Determination of the protonation states of carboxylic acid side chains in proteins *via* hydrogen atoms and resolved single and double bond lengths. After Deacon *et al.* (1997) using CHESS. Reproduced by permission of The Royal Society of Chemistry.

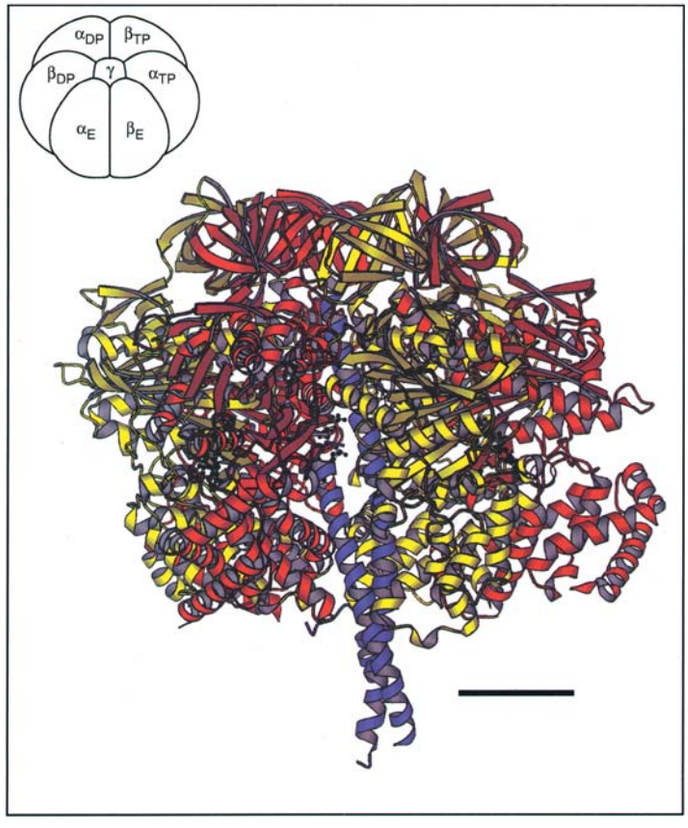


Fig. 8.1.8.3. The protein crystal structure of $F_1$ ATPase, one of the largest non-symmetrical protein structure complexes, solved using SR data recorded at the SRS wiggler 9.6, Daresbury. The scale bar is 20 Å long. Reprinted with permission from *Nature* (Abrahams *et al.*, 1994). Copyright (1994) MacMillan Magazines Limited.

non-virus) cases, the weakest anomalous-scattering signal utilized to date for MAD, the fastest time-resolved Laue study and the highest-resolution structure determinations to date. Another phasing technique involving multiple ('*n*-beam') diffraction is also being applied to proteins [Weckert & Hümmer (1997) at the ESRF and the NSLS]. These examples at least indicate the present bounds of capability of the various sub-fields of SR and macromolecular crystallography.

#### 8.1.8.1. *Atomic and ultra high resolution macromolecular crystallography*

The use of high SR intensity, cryo-freezing of a protein crystal to largely overcome radiation damage and sensitive, automatic area detectors (CCDs and/or image plates) is allowing diffraction data to be recorded at resolutions equivalent to smaller molecule (chemical) crystallography. In a growing number of protein crystal structure studies, atomic resolution (1.2 Å or better) is achievable (Dauter *et al.*, 1997). The 'X-ray data to parameter' ratio can be favourable enough for single and double bonds, *e.g.* in carboxyl side chains, to be resolved [Fig. 8.1.8.1; Deacon *et al.* (1997) for concanavalin A at 0.94 Å resolution]. Along with this bond distance precision, one can see the reactive proton directly. This approach complements H/D exchange neutron diffraction studies. Neutron studies have recently expanded in scope by employing Laue geometry in a synergistic development with SR Laue diffraction (Helliwell & Wilkinson, 1994; Helliwell, 1997*b*; Habash *et al.*, 1997, 2000). The scope and accuracy of protein crystal structures has been transformed.

#### 8.1.8.2. *Small crystals*

Compensating for small crystal sample volume by increasing the intensity at the sample has been of major interest from the outset, and tests have shown that the use of micron-sized samples is feasible (Hedman *et al.*, 1985). Third-generation high-brilliance sources are optimized for this application *via* micron-sized focal spot beams, as described in the *ESRF Foundation Phase Report* (1987). Applications of the ESRF microfocus beamline include the determination of the structure of the bacteriorhodopsin crystal at high resolution from micro-crystals (Pebay-Peyroula *et al.*, 1997). Experiments using extremely thin plates involving only 1000 protein molecular layers are described by Mayans & Wilmanns (1999) on the BW7B wiggler beamline at DESY, Hamburg. A review of small crystals and SR, including tabulated sample scattering efficiencies, can be found in Helliwell (1992), pp. 410–414.

#### 8.1.8.3. *Time-resolved macromolecular crystallography*

Time-resolved SR Laue diffraction of light-sensitive proteins, such as CO Mb studied with sub-nanosecond time resolution in pump-probe experiments (see Srajer *et al.*, 1996), are showing direct structural changes as a function of time. Enzymes, likewise, are being studied directly by time-resolved methods *via* a variety of reaction initiation methods, including pH jump, substrate diffusion and light flash of caged compounds pre-equilibrated in the crystal. Flash freezing is used to trap molecular structures at optimal times in a reaction determined either by microspectrophotometry or repeated Laue 'flash photography'. For overviews, see the books edited by Cruickshank *et al.* (1992) and Helliwell & Rentzepis (1997). Enzyme reaction rates can be altered through site-directed mutagenesis (*e.g.* see Niemann *et al.*, 1994; Helliwell *et al.*, 1998) and matched to diffraction-data acquisition times.

#### 8.1.8.4. *Multi-macromolecular complexes*

Multi-macromolecular complexes, such as viruses (Rossmann *et al.*, 1985; Acharya *et al.*, 1989; Liddington *et al.*, 1991) (Fig. 8.1.8.2), the nucleosome (Luger *et al.*, 1997), light-harvesting complex (McDermott *et al.*, 1995) and the 13-subunit membrane-bound protein cytochrome *c* oxidase (Tsukihara *et al.*, 1996), and large-scale molecular assemblies like muscle (Holmes, 1998) are very firmly recognizable as biological entities whose crystal structure determinations rely on SR. These single-crystal structure determinations involve extremely large unit cells and are now tractable despite very weak scattering strength. The crystals often show extreme sensitivity to radiation (hundreds, even a thousand, crystals have been used to constitute a single data set). Cryocrystallography radiation protection is now used extensively in crystallographic data collection on whole ribosome crystals (Hope *et al.*, 1989); SR is essential for this structure determination (Yonath, 1992; Yonath *et al.*, 1998; Ban *et al.*, 1998). These large-scale molecular assemblies often combine electron-microscope and diffraction techniques with SR X-ray crystallography and diffraction for low-to-high resolution detail, respectively. A major surge in results has come from the ESRF, where the X-ray undulator radiation, of incredible intensity and collimation in a number of beamlines (Helliwell, 1987; Miller, 1994; Branden, 1994; Lindley, 1999), has been harnessed to yield atomic level crystal structures of the 780 Å diameter blue tongue virus (Grimes *et al.*, 1997, 1998) and the nucleosome core particle (Luger *et al.*, 1997). A very large multi-protein complex solved using data from the Daresbury SRS wiggler is the $F_1$ ATPase structure (Fig. 8.1.8.3), for which a share in the Nobel Prize for Chemistry in 1997 was awarded to John Walker in Cambridge. The structure (Abrahams *et al.*, 1994; Abrahams & Leslie, 1996) and the amino-acid sequence data, along with fluorescence microscopy, show how biochemical energy is harnessed to drive the proton pump across biological membranes, thus corroborating hypotheses about this process made over many years. This study, made tractable by the SRS wiggler high-intensity protein crystallography station (Fig. 8.1.4.1), illustrates the considerable further scope possible with yet stronger, more brilliant SR undulator and multipole wiggler sources.

#### 8.1.8.5. *Optimized anomalous dispersion (MAD), improved MIR data and 'structural genomics'*

Rapid protein structure determination *via* the MAD method of seleno protein variants (Hendrickson *et al.*, 1990), as well as xenon pressure derivatives (Schiltz *et al.*, 1997), and improved heavy-atom isomorphous replacement data are removing a major bottle-neck in protein crystallography, that of phase determination. One example of a successful MAD study with an especially weak anomalous signal, from one selenium atom per 147 amino acids, undertaken at ESRF BM14, is that of van Montfort *et al.* (1998). At another extreme is the largest number of anomalous scatterer sites; for example, Turner *et al.* (1998), using the NSLS, reported the successful determination of 30 selenium atoms in 96 kDa of protein (one dimer) in the asymmetric unit using one-wavelength anomalous differences (at peak) as *E* values and '*Shake n' Bake*' (Miller *et al.*, 1994), followed by MAD phasing from three-wavelength data and solvent flattening. Overall, as the number of protein structures in the Protein Data Bank doubles every few years (currently the number is 9000), the possibility of considering whole genome-level structure determinations arises (Chayen *et al.*, 1996; Chayen & Helliwell, 1998). The human genome, the determination of the amino-acid sequence of which is currently underway, comprises some 100 000 proteins. Of these, some 40% are membrane bound and somewhat difficult to crystallize. A MAD

protein crystal structure currently requires roughly 1 day of SR BM beamtime. A coordination of 20 SR instruments worldwide, or an SR machine devoted solely to the project, could make major progress in 20 years. This estimate assumes no further speeding up of the technique, such as would acrue with faster detectors like the pixel detector. The smaller yeast genome, comprising amino-acid sequences of 10 000 proteins, has recently been completed. The molecular-weight histogram peaks at 30 kDa. Assuming, on average, that one amino acid out of 56 is a methionine, it is clear that the MAD method, and six Se atoms on average for each protein, is a good match to the task. This approach, along with homology modelling and genetic alignment techniques, opens the immense potential for 'structural genomics' as a basis for understanding and controlling disease (*e.g.* see Bugg *et al.*, 1993). SR and crystallography are now intricately intertwined in their scientific futures and in facilities provision (Helliwell, 1998).



references