lot-to-lot variations in lyophilized protein formulations (Hirakura *et al.*, 2007), and has been explored for use in structure-based generic assays (Allaire *et al.*, 2009).

3.7.7.2. *Calculation of protein powder patterns (with Kenny Ståhl)*

The Powder Diffraction File contains a few experimental powder patterns of proteins. These include silk fibroin protein (00-054-1394), tubulin (00-036-1547 and 00-036-1548), insulin (00-060-1360 through 00-060-1368), tomato bushy stunt virus (00-003-0001) and tobacco mosaic virus (00-003-0003 and 00-003-0004). Patterns have not yet been calculated from the structures in the Protein Data Bank because the calculated intensities generally fit poorly to those in experimental patterns.

Protein structures in the PDB do not generally contain H-atom positions, and the contributions from the disordered solvent in the solvent channels (which is the major source of the discrepancy) is not described (Hartmann *et al.*, 2010). The conventional Lorentz factor tends to infinity when approaching $2\theta = 0^\circ$. Differences in data-collection temperatures and solvent content between powder and single-crystal specimens often mean that the lattice parameters differ. The relatively poor scattering from the protein and the large scattering from the mother liquor and sample holder result in significant background contributions to experimental powder patterns.

Optimization of the lattice parameters is generally straightforward and is important because most protein crystal structures are determined at low temperatures, while powder data are collected under ambient conditions. Protein crystals contain 30–80% disordered solvent. The solvent contribution to the diffraction pattern is most important for the low-angle powder data. In conventional protein crystallography several correction models have been developed (Moews & Kretsinger, 1975; Phillips, 1980; Jiang & Brünger, 1994), but the flat bulk-solvent model is the simplest one which yields a realistic correction (Jiang & Brünger, 1994; Hartmann *et al.*, 2010). This model includes two parameters: $k_{\rm sol}$, which defines the level of electron density in the solvent region, and $B_{\rm sol}$, which defines the steepness of the border between the solvent and macromolecular regions. These parameters are typically refined in contemporary software and cluster around $k_{\rm sol} = 0.35$ e Å^{-3} and $B_{\rm sol} = 46$ Å^2 (Fokine & Urzhumtsev, 2002).

The flat bulk-solvent correction can be applied using *phenix.pdbtools* (Adams *et al.*, 2010), which requires a PDB coordinate file and values of $k_{\rm sol}$ and $B_{\rm sol}$ as input. Average values can be used, but refined values or values from the Electron Density Server (EDS; Kleywegt *et al.*, 2004) can improve the results. The bulk-solvent correction is highly anisotropic, and both parameters affect the anisotropy.

The ideal H-atom positions can be calculated using *phenix.pdbtools*. The solvent and hydrogen contributions to the pattern can be significant (Fig. 3.7.13).

The Lorentz factor L describes the fraction of a reflection that is in the diffracting condition. For Bragg–Brentano and Debye–Scherrer geometries it is given by

$$L = \frac{1}{\sin 2\theta}\frac{1}{\sin\theta}. \qquad (3.7.3)$$

This equation assumes ideal crystals, resulting in infinitesimally small reciprocal-lattice points. The true size of the lattice points depends on the crystallite size and imperfections (strain). This smearing needs to be included in the Lorentz factor at low angles. A revised Lorentz factor for protein powder diffraction has been derived (Hartmann *et al.*, 2010),

$$L_{\rm rev} = \frac{1}{\sin 2\theta}\frac{1}{\sin\theta}\frac{\sin^2\theta}{(\sin^2\theta + \lambda^2\eta^2/12)}, \qquad (3.7.4)$$

in which η reflects the distribution of scattering-vector amplitudes. For Guinier geometry these equations become more complex (Hartmann *et al.*, 2010). Fig. 3.7.14 shows that the Lorentz factor has a smaller effect than the solvent and H atoms, but that it is still significant. By applying these corrections it should be possible for the ICDD editorial staff to calculate useful powder patterns from PDB entries that could be included in the Powder Diffraction File.

Separating the background from the diffraction pattern is not straightforward (Frankaer *et al.*, 2011). Estimation of the background is greatly assisted by a correct calculated pattern. The calculated pattern can be scaled to the experimental data using *PROTPOW* (http://www.kemi.dtu.dk/english/Research/PhysicalChemistry/Protein_og_roentgenkrystallografi/Protpow).

Ståhl *et al.* (2013) have demonstrated that existing search/match procedures can be used to identify proteins using their powder patterns, and that powder patterns calculated from Protein Data Bank coordinates with proper care can be added to a database and included in the search/match procedure. Several problems can be foreseen when including large amounts of protein data into the Powder Diffraction File. It may be worthwhile including powder patterns with several levels of solvent correction, rather than just an average value. Asymmetry from instrumental effects and specimen transparency, which can affect the peak positions, needs to be taken into account. The use of an average thermal expansion coefficient may be sufficient to account for the differences in lattice parameters between low-temperature single-crystal structures and powder patterns measured under ambient conditions.

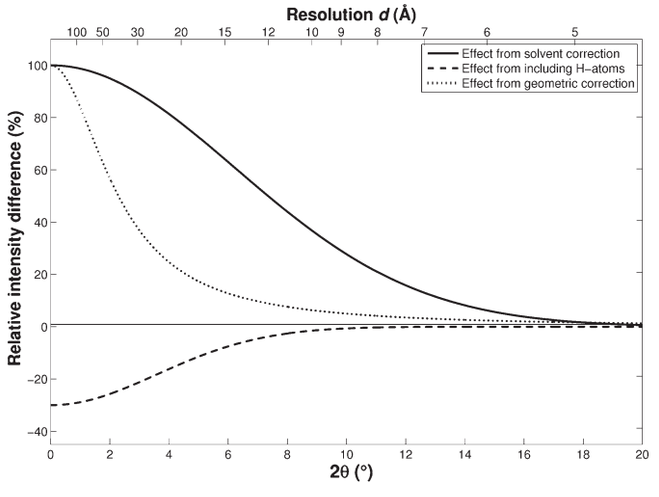

Figure 3.7.13
Overview of the trends from the different corrections. The effects are shown as the relative intensity difference $(I_{\rm non\text{-}corr} - I_{\rm corr})/I_{\rm non\text{-}corr}$ plotted as functions of the scattering angle 2θ (using Cu $K\alpha_1$) and resolution $d = \lambda/(2\sin\theta)$. The curves are based on average corrections of lysozyme and insulin data. $I_{\rm non\text{-}corr}$ is the raw intensity from a calculated pattern which has only been Lorentz corrected. The geometric correction curve was calculated using $\eta = 0.045$ Å^{-1}. From Hartmann *et al.* (2010).