The influence of positional errors on the Debye effects

The relation between a Gaussian perturbation of the atomic positional parameters and the average squared structure-factor amplitude is presented. Using an error-dependent radial distance distribution of an atomic protein model, it can be shown that the Debye effects diminish exponentially as a function of increasing positional errors. These relations can be used to estimate the quality of an atomic model and the corresponding phases. The limiting case of equal atoms with an infinitely large coordinate error results in the classical Wilson model.

1. Introduction

The deviations from a straight line in a Wilson plot (Wilson, 1942, 1949) are known as Debye effects (Giacovazzo, 1998). These deviations are mainly caused by the stereochemistry of the molecular structure and can be modelled by the Debye equation (Debye, 1915),

\[ E(|F_h|^2)_o = \sum_j \sum_k f_j(h)f_k(h) \frac{\sin(2\pi d_{jk}h)}{2\pi d_{jk}h}. \]  

(1)

The subscript \( o \) in (1) denotes that the averaging is carried out over all orientations of \( F_h \) for a given reciprocal-lattice spacing \( h \). See Table 1 for notation. Note that in (1) we do not account for lattice periodicity or other packing effects and the expression is thus only strictly valid for a single unit cell or molecule. The effect of the shape of the molecular envelope on the Wilson plot is not directly accounted for. This point is discussed by Morris, Blanc et al. (2004). The need for these assumptions will however be eliminated, as will be discussed below. (1) can be rewritten in terms of the radial distance distribution \( f_{\text{rad}}(d) \). Assuming equal atoms, one arrives at

\[ E(|F_h|^2)_o = Nf^2(h) \left[ 1 + (N - 1) \int f_{\text{rad}}(d) \frac{\sin(2\pi dh)}{2\pi dh} \, dd \right]. \]  

(2)

For the trigonometric part of the structure-factor amplitudes, (2) can be transformed into

\[ E(|E_h|^2)_o = 1 + \gamma(h), \]  

with

\[ \gamma(h) = (N - 1) \int f_{\text{rad}}(d) \frac{\sin(2\pi dh)}{2\pi dh} \, dd. \]  

(4)

In the Wilson approximation, the atoms are independent and uniformly distributed throughout the unit cell, resulting in \( \gamma(h) = 0 \) (Giacovazzo, 1998). An excess of lack of specific interatomic distances results in a non-zero interference and affects the mean-squared structure-factor amplitude. This is demonstrated in Fig. 1, where the radial distance distribution of a typical protein is multiplied by a
Table 1
Notation.

- $\mathbb{E}[\tau(x)j]$ is the expectation value of $\tau(x)$ by integration over $x$
- $d_{jk}$ is the vector between an atom $j$ and $k$
- $d_{ab}$ is the length of $d_{jk}$
- $h$ is the reciprocal-lattice spacing
- $E_{h}$ is the calculated structure factor
- $\{d_{ab}\}_m$ is the average observed intensity in a resolution shell $m$
- $\langle E_{h} \rangle_m$ is the expected average observed intensity in a resolution shell $m$
- $f(j)$ is the form factor of atom $j$
- $\sigma(\pmb{r})$ is the variance of the Gaussian error of the positional parameters
- $f_{rad}(\sigma^2)$ is the radial distance distribution of a model with an error with variance equal to $\sigma^2$ along each direction
- $q_{h}$ is a Gaussian error on atom $j$
- $N$ is the number of atoms
- $NCM(d_{ab}, \sigma^2)$ is the non-central Maxwell distribution
- $\gamma(h)$ is a resolution-dependent term that describes the Debye–Waller factor
- $\gamma(h|\sigma^2)$ is a resolution- and error-dependent term that describes the Debye–Waller factor
- $\gamma(h)_{ob}$ is an empirical $\gamma(h)$ curve obtained from a large number of structures
- $\gamma(h)_{obs}$ is an estimate of $\gamma(h)$ obtained from experimental data
- $e_{B}$ is the Babinet bulk-solvent scale factor
- $B_{s}$ is the Babinet bulk-solvent $B$ value
- $e_{0}$ is the scale factor
- $B_{wil}$ is the Wilson plot $B$ value
- FOM is the Figure of merit defined as $\mathbb{E}[\cos(\Delta\phi)]; \Delta\phi$ is the phase difference

\[
sinc(2\pihd) = \sin(2\pihd)/(2\pihd) \text{ term [see expression (4)]}, \text{ for } 1/h \text{ equal to } 1.1, 2.2 \text{ and } 4.5 \text{ Å. In the same plots, the curves for a structure with a uniform independent distribution of atoms (hereafter denoted as a random structure) are shown. The interatomic distances arising from chemically bonded atoms (1–2 distances) at about 1.4 Å and atoms involved in bond-angle distances (1–3 distances) at about 2.4 Å are the two major contributors to the differences between the radial distance distribution of a protein structure and a random structure. An excess of interatomic distances compared with the random case is also found around 3.8 Å, a typical $C^{\alpha} – C^{\alpha}$ distance. At larger distances, differences between the radial distance distribution of a protein and a random structure are also found owing to secondary structure.}

The qualitative effects of these interatomic distances on the average squared structure-factor amplitude are summarized in Table 2 in terms of positive or negative contributions to the integral in (4). As shown by Morris & Bricogne (2003), the 1–2 and 1–3 distances are (in part) responsible for a large peak in the mean $|E_{h}|^2$ value at around $1/h = 1.1$ Å. Both the 1–2 and 1–3 distances have a positive contribution to $\gamma(h)$ for $1/h = 1.1$ Å (Fig. 2). Also, the lack of interatomic distances between the latter two coordination shells where the sinc function is negative results in an effectively larger value of $\gamma(h)$ compared with a random structure. Other pronounced peaks in the average of $|E_{h}|^2$ as a function of resolution lie around $1/h = 2.2$ and $1/h = 4.5$ Å. At $1/h = 2.2$ Å, the 1–2 distances have a negative contribution while the 1–3 distances have a positive contribution to the average $|E_{h}|^2$ value. A similar observation can be made for the peak at $1/h = 4.5$ Å: the 1–2 distances and effects of secondary structure (here loosely defined as distance features between 4.5 and 7 Å) have a positive contribution to the average $|E_{h}|^2$ value, whereas the 1–3 and $C^{\alpha} – C^{\alpha}$ distances at 3.8 Å have a negative contribution.

This interpretation of the well known features in the Wilson plot of a protein structure leads to the statement that changes in the radial distance distribution caused by model errors have an effect on the averaged calculated $|E_{h}|^2$ and $|F_{h}|^2$ values. The

![Figure 1](image.png)

Sinc functions multiplied by the radial distance distribution of a random structure (thin line) and lysozyme (thick line, PDB code 102l) for (a) $h = 1/1.1$, (b) $h = 1/2.2$ and (c) $h = 1/4.5$ Å. See Table 2, Fig. 2 and the main text for further details.
influence of a coordinate error on the Debye effects can therefore be used to assess the quality of an atomic model and the corresponding phases. The function \( \gamma(h) \) is expected to be essentially the same for a wide range of protein structures as judged from the well known features of the Wilson plot (Popov & Bourenkov, 2003; Cowtan, 1998; Blessing et al., 1996), although a dependence on the secondary structure is present (Morris, Blanc et al., 2004). An empirical \( \gamma(h) \) curve obtained from a PDB analysis (Bernstein et al., 1977; Berman et al., 2000) can be used to estimate \( \gamma(h)_{\text{obs}} \) from the observed structure-factor amplitudes of the protein under consideration.

2. Methods

2.1. The influence of the coordinate error on the Debye effects

The error-dependence of the mean-squared structure-factor amplitude was examined by introducing an error-dependent radial distance distribution \( f_{\text{rad}}(d|\sigma^2) \) into (2).

\[
\mathbb{E}(|F_h|_q^2)_{\sigma^2} = N f^2(h) \left[ 1 + (N - 1) \int_0^{\infty} f_{\text{rad}}(d|\sigma^2) \frac{\sin(2\pi h d)}{2\pi h d} \, dd \right]
\]

where \( \gamma(h|\sigma^2) \) is the error-dependent variant of (4). The expectation value of the trigonometric part of the structure-factor amplitude becomes

\[
\mathbb{E}(|F_h|_r^2)_{\sigma^2} = 1 + \gamma(h|\sigma^2).
\]

The subscript \( q \) in (5) and (6) denotes that the expectation value is obtained by integrating over the errors of the positional parameters. The errors of the positional parameters are assumed to be distributed independently for each atom, according to a spherically symmetric Gaussian with variance terms equal to \( \sigma_m^2 \) along the \( x, y, \) and \( z \) directions. The change in the radial distance distribution can then be written in accordance with Zwart & Lamzin (2003),

\[
f_{\text{rad}}^{\text{obs}}(d|\sigma^2) = \int_0^{\infty} f_{\text{rad}}(d|\sigma^2) \, \text{d}d, \quad (7)
\]

where

\[
\text{NCM}(d|\sigma^2, \sigma_m^2) = \frac{1}{[2\pi(\sigma^2)]^{1/2}} \exp \left[ - \frac{(d - d_{\text{mean}})^2}{2\sigma^2} \right] \times \frac{d}{d_{\text{mean}}} \left[ 1 - \exp \left( - \frac{2\sigma_m^2 d}{\sigma^2} \right) \right] \quad (8)
\]

and \( \sigma^2 \) is the sum of the variances of the error terms of the positional parameters of a pair of atoms,

\[
\sigma^2 = \sigma_m^2 + \sigma_m^2. \quad (9)
\]

For errors with a variance equal to \( \sigma_m^2 \) for all atoms, \( \sigma^2 \) becomes equal to \( 2\sigma_m^2 \). When atom \( j \) is not perturbed and an atom \( k \) has a positional error with variance \( \sigma_m^2 \), \( \sigma^2 \) becomes equal to \( \sigma_m^2 \).

It can be shown (Appendix A) that

\[
\gamma(h|\sigma^2) = \exp(-2\pi^2 h^2 \sigma^2) \gamma(h). \quad (10)
\]

The exponential multiplier has the same form as the \( D \) term in the work of Luzzati (1952) and Read (1990). In the limiting case of an infinitely large error and \( h \neq 0 \), \( \gamma(h|\sigma^2) \) in (10) becomes zero, effectively resulting in the Wilson approximation of independent uniformly distributed atoms and thus \( \mathbb{E}(|F_h|^2) = 1 \). The dependence of \( \mathbb{E}(|F_h|^2) \) as a function of the coordinate error and the resolution can be used to estimate \( \sigma_m^2 \) and corresponding figures of merit. Let us model the expected average calculated intensity as a function of resolution and model error as follows:

\[
\mathbb{E}(|F_h|^2)_{\sigma^2} = k_p \exp(-B_{\text{wil}} h^2/2) \times \sum_j f_j^2(h)[1 + \exp(-4\pi^2 \sigma_m^2 h^2) \gamma(h)]. \quad (11)
\]

When \( \gamma(h) \) is known or a good estimate of it is available denoted as \( \gamma(h)_{\text{obs}} \), a least-squares minimization of the difference between the average calculated squared structure factor amplitude as a function of resolution (\( |F_h|^2 \) obtained from a model) and its expected value [\( \mathbb{E}(|F_h|^2) \)] can be carried out. The latter expectation value is calculated using (11), thus allowing estimation of the scale factor \( k_p \), the Wilson plot \( B \) value \( B_{\text{wil}} \) and the variance of the error model \( \sigma_m^2 \). This variance can in turn be used to estimate phase probabilities (Sim, 1958, 1959) and corresponding figures of merit, which are defined as the expected value of the cosine of the phase difference between the available and error-free phases, \( \gamma(h)_{\text{obs}} \) [an estimate of \( \gamma(h) \) of the protein under consideration] can be obtained from the observed X-ray data and from an empirically obtained standard \( \gamma(h) \) curve, denoted as \( \gamma(h)_{\text{PDB}} \), using a procedure outlined in Appendix B.

The standard \( \gamma(h)_{\text{PDB}} \) curve was obtained from the analysis of 100 protein structures from the PDB. The structure-factor
amplitudes were calculated and have subsequently been normalized in resolution bins,

\[ \langle |F_h|^2 \rangle = \frac{\sum_h |F_h|^2}{N_h \sum_h f^2(h)}. \]  (12)

The subscript \( h \) in (12) denotes the summation over the \( N_h \) reflections that fall within the resolution bin \( h \). Although this is more computationally intensive than obtaining the \( \gamma(h) \) profiles via the radial distance distribution and the Debye equation (2), it has the advantage that lattice periodicity and non-equal atom effects are incorporated. The resulting mean \( \gamma(h) \) profile is shown in Fig. 3 together with a \( \gamma(h) \) profile obtained using (4) and a \( \gamma(h) \) profile obtained using (12) for equal-atom structures for comparison.

As shown in Appendix B, the use of the empirically obtained \( \gamma(h)_{\text{PDB}} \) curve for the estimation of \( \gamma(h)_{\text{obs}} \) avoids the need for the single equal-atom molecule approximation.

3. Results

3.1. Coordinate error-dependent \( \gamma(h) \) profiles

A Monte Carlo simulation has been carried out to compute the average value of \( \langle \sin(2\pi d h) / 2\pi d h \rangle \) with \( d \) distributed according to the non-central Maxwell distribution in order to test the validity of (10). This has been carried out for a number of errors and various values of \( h \). The numerical results have been subsequently compared with the results from (10). A plot of the average values of \( \sin(2\pi d h) / 2\pi d h \) against the expected values is shown in Fig. 4. To visualize the effect of the reduction of the Debye effects with increasing positional error, the atomic model of lysozyme (PDB code 102l) has been used to compute a number of error-dependent radial distance distributions and corresponding \( \gamma(h|\sigma^2) \) profiles via (5) and (7). The resulting profiles are shown in Fig. 5. The contribution of H atoms has been omitted.

3.2. Model and phase quality estimates

Estimates of \( \sigma_m^2 \) and corresponding figures of merit have been obtained using the described least-squares procedure, with a number of different errors on the model. The first model used was the ARP/wARP (Perrakis et al., 1999) distributed example (with the X-ray data) of leish-

Figure 3
Empirical \( \gamma(h) \) curves determined from a selection of good-quality atomic protein models using (12) for deposited protein models (Crystal structure), the same protein models but with all atoms as O and B values set to 20 Å² (Crystal structure, equal atoms) and using the radial distance distribution (5) (Single molecule, equal atoms). The differences between the curves are ascribed to packing effects and the assumption of equal atoms in (5).

Figure 4
The expected value of \( \gamma(h|\sigma^2) \) given by (10) is plotted against the average value of \( \sin(2\pi d h) / 2\pi d h \) for a distance distributed according to the non-central Maxwell distribution with a target distance equal to 2.5 Å at various values of \( \sigma^2 \) and \( h \) (black diamonds). The least-squares line fitted through the points has a slope of 1 and a correlation coefficient of 1.

Figure 5
The effect of a coordinate error on the Debye effects calculated for lysozyme (102l) using (6) and (7).
research papers

manolysin (PSP; courtesy of P. Metcalf). The final model has been randomized by adding a Gaussian error to the positional parameters with an r.m.s.d. of 1.5 Å. Structure-factor amplitudes have been calculated from this model using REFMACS (Murshudov et al., 1997) and have been used to estimate $\sigma_{\text{m}}^2$. The overall scale factor, Wilson plot $B$ value, bulk-solvent parameters and $\gamma(h)_{\text{obs}}$ have been estimated from the measured experimental data (a zero coordinate error was assumed) as outlined in Appendix B. $\gamma(h)_{\text{obs}}$ was subsequently used to estimate the scale factor, the Wilson plot $B$ value and the variance of the Gaussian error model, $\sigma_{\text{m}}^2$. Given the structure-factor amplitudes calculated from the randomized model. In Fig. 6, a bulk-solvent-corrected Wilson plot from the experimental data is shown (Observed $I$) as well as a fitted curve on the basis of $\gamma(h)_{\text{obs}}$ (Theoretical $I$; 0 Å r.m.s.d.). The least-squares residual estimated the coordinate error to be 1.7 Å. The corresponding Wilson plots using the calculated structure-factor amplitudes of the randomized model are also shown in Fig. 6.

The same final model of PSP was scrambled by introduction of a Gaussian error to the atomic parameters and underwent five unrestrained refinement cycles with REFMACS using the full resolution range of the observed structure-factor amplitudes, with the use of cross-validation. The resulting coordinates have been used to calculate model structure-factor amplitudes and figures of merit were estimated. This has been carried out for a number of different errors as well as for a phase set calculated from a model with an r.m.s.d. of 3 Å, which was used for free-atom modelling with ARP/wARP without the use of cross-validation. A similar set of tests has been carried out on the 1hf8 model and X-ray data set from the PDB. For 1hf8 the data was truncated at the low-resolution side because of poor data quality for $1/h > 7$ Å. The results of these test are summarized in Figs. 7 and 8. Another test has been carried out on a number of intermediate free-atom modelling structures. Solvent-flattened experimental phases of PSP were used to carry out a free-atom modelling experiment without the use of cross-validation. The estimated figures of merit and r.m.s.d. values are shown in Fig. 9.

4. Discussion and conclusions

As seen in Fig. 4, the exponential multiplier describes the changes in the Debye effects as a function of coordinate error

Figure 6
Wilson plots for experimental bulk-solvent-corrected PSP structure-factor amplitudes (Observed $I$) and the fit using the estimated $\gamma(h)_{\text{obs}}$ (Theoretical $I$; 0 Å r.m.s.d.). Similar curves are shown for the structure-factor amplitudes calculated from the model with an r.m.s.d. of 1.5 Å on the positional parameters (Calculated $I$). A fit of the Wilson plot of the calculated structure-factor amplitudes using $\gamma(h)_{\text{obs}}$ and assuming a coordinate error of 1.7 Å is also shown (Calculated $I$; 1.7 Å r.m.s.d.). In the upper right corner, the exponentiated negative least-squares criteria is shown versus the r.m.s.d.

Figure 7
Estimated figures of merit for the PSP data set (20–2.0 Å resolution; $B_{\text{ext}} = 18$ Å$^2$) for various scrambled models refined without restraints (u.r.), as well as from a free-atom model (f.a.m.) obtained from phases generated from the final model randomized by 3 Å r.m.s.d. In the latter case, no cross-validation has been used. (Estimated FOM) denotes the average figure of merit estimated via the described method. For comparison, the REFMACS estimate is also given. (True FOM) is defined as the average value of the cosine of the phase differences between the final and scrambled model.

Figure 8
Estimated figures of merit for the 1hf8 data set (7–2.0 Å resolution; $B_{\text{ext}} = 32$ Å$^2$) for various unrestrained (u.r.) refined scrambled models. The true figure of merit is taken to be equal to cosine of the phase difference of the final and current model. (Estimated FOM) denotes the average figure of merit estimated via the described method. For comparison, the REFMACS estimate is also given.
rather well and offers an easy way of modelling these effects. The differences in the \( \gamma(h)_{\text{PD}} \) profiles computed using expression (4) and via binning of structure-factor amplitudes are ascribed to the underlying assumptions. The effect of packing only affects the low-resolution part of the \( \gamma(h) \) curve, whereas the effect of an equal-atom assumption introduces substantial differences over the whole resolution range. However, the curves can be scaled together using an exponential model similar to the Babinet terms used to model the effects of the bulk solvent. The estimates of the figures of merit shown in Figs. 7, 8 and 9 are close to the \textsc{refmac5} estimates, indicating that the effects studied contain enough information to predict, within certain limits, the accuracy of the model and corresponding phases. In the case of phases originating from a model with an r.m.s.d. of 3 Å and extreme model bias owing to the subsequent free-atom modelling without cross-validation (see Fig. 7; r.m.s.d. = 3 Å + f.a.m.) the average estimated figure of merit is still larger than the true average cosine of the phase error but is a better estimate than that obtained by \textsc{refmac5}.

A key point is that the presented error-estimation method is rather sensitive to the quality of the low-resolution part of the data set used. This is ascribed to the fact that the Debye effects at high resolution diminish faster than those at lower resolution. Most information on the value of \( \sigma^2 \) when the error is (moderately) large is thus obtained from the low-resolution part of the data. More appropriate weighting schemes in the least-squares procedure or a maximum-likelihood approach can possibly account for this sensitivity and might be useful in reducing the observed bias in the estimates. Linked to this is the need for a model describing the behaviour of the bulk solvent and its effect on the average structure-factor amplitude. The exponential model (Tronrud, 1997) used here is known for its limitations (Fokine & Urzhumtsev, 2002), but has been widely used because of its simplicity. The presented method is sensitive to non-randomly incomplete data, such as missing strong low-angle reflections. These effects can in principle be modelled by using the characteristics of the truncated Wilson distribution (Parthasarathy & Sekar, 1993a), rather than ignoring a subset of valuable structure-factor amplitudes as performed for the 1hf8 data set. Furthermore, the method is based on an assumption of independent Gaussian errors on each atom. Violation of this assumption undermines the basic principles of the method, which is largely designed for usage during free-atom modelling experiments.

Ideally, the dependence of the average squared structure-factor amplitude as a function of resolution and model error should be used in conjunction with other error-estimation methods, such as \( \sigma_A \) (Read, 1986), in the hope of enhancing the overall quality of the error estimates. Further incorporation of prior knowledge in the error-estimation method outlined in this paper might enhance the results. If secondary-structural information is available or if the classification procedure outlined by Morris, Blanc et al. (2004) proves to be reliable and robust, then protein-specific \( \gamma(h)_{\text{PD}} \) curves can be utilized to obtain more accurate \( \gamma(h)_{\text{obs}} \) estimates.

Using the expected averaged squared structure-factor amplitude as a function of resolution and model error as a source of information during refinement seems to be an interesting option. This is in effect a reciprocal-space variant of the addition of restraints to atoms on the basis of known radial distance distributions in proteins (Sheldrick, 1998; Scheres & Gros, 2001). A more radical and possibly better approach would be to improve the description of structure-factor probability distributions by taking into account stereochemistry \textit{a priori}, as suggested by Bricogne (1997a,b).

**APPENDIX A**

The expectation value of \( \gamma(h \mid \sigma^2) \)

Consider the expression for the squared trigonometric part of the structure-factor amplitude of an atomic model with an error

\[
|E_h|^2 = 1 + \sum_{j,k \neq k} \exp[-2\pi i h(d_{jk} + q_{jk})].
\]

with \( d_{jk} \) as an interatomic vector and \( q_{jk} \) a vector drawn from a three-dimensional Gaussian centred on the origin with a variances in all directions equal to \( \sigma^2 \). The latter expression is equal to

\[
|E_h|^2 = 1 + \sum_{j,k \neq k} \exp(-2\pi i h d_{jk}) \exp(-2\pi i h q_{jk}).
\]

Averaging over the vectors \( q_{jk} \) results in (Luzzati, 1952)

\[
\mathbb{E}(|E_h|^2)_q = 1 + \exp(-2\pi^2 h^2 \sigma^2) \sum_{j,k \neq k} \exp(-2\pi i h d_{jk}).
\]

Subsequent averaging over all orientations results in

\[
\mathbb{E}(|E_h|^2)_{o,q} = 1 + \exp(-2\pi^2 h^2 \sigma^2) \sum_{j,k \neq k} \frac{\sin(2\pi h d_{jk})}{2\pi h d_{jk}}
\]

and thus

**Figure 9**

Figures of merit and corresponding r.m.s.d. estimates of the PSP data set during free-atom modelling. (True FOM) is defined as the average cosine of the phase difference between the final and current model. (Estimated FOM) denotes the average figure of merit estimated \textit{via} the described method. For comparison, the \textsc{refmac5} estimate is also given.
\[ \gamma(h|\sigma^2) = \exp(-2\pi^2 h^2 \sigma^2)\gamma(h). \]  

(17)

First averaging over the orientations and subsequently over the distances is more complicated, but results in the same expression. In this study, the transition from (15) to (17) was carried out numerically rather than using the Debye formula. This eliminated the need for the approximations discussed in §1.

**APPENDIX B**

**Determination of \(\gamma(h)_{\text{obs}}\)**

The \(\gamma(h)\) profile corresponding to the protein under investigation can be obtained using the \(\gamma(h)_{\text{PDB}}\) profile obtained from the PDB analysis by minimizing the least-squares target

\[ Q = \sum m w_m (I_{\text{obs},m} - I_{0,m})^2, \]  

(18)

where \(I_{\text{obs}}\) is the average observed squared structure-factor amplitude in resolution bin \(m\). \(I_{0,m}\) is the expected average observed squared structure-factor amplitude on the basis of the following model:

\[ I_{0,m} = k_p \exp(-B_{\text{Wil}}h_m^2/2)[1 - k_s \exp(-B_s h_m^2/4)]^2 \]

\[ \times \sum k f_k^2(h_m)[1 + \gamma(h_m)_{\text{PDB}}]. \]  

(19)

\(k_p\) is a scale factor and \(B_{\text{Wil}}\) the Wilson plot \(B\) value. \(k_s\) and \(B_s\) are Babinet bulk-solvent correction factors (Tronrud, 1997).

The weights \(w_m\) are the sum of estimated variances of the mean intensities per resolution shell \(h_m\). Minimizing \(Q\) as a function of \(k_p, B_{\text{Wil}}, k_s\) and \(B_s\), results in a set of parameters that can be used to obtain a \(\gamma(h)\) profile from the observed data. This procedure is similar to the Wilson scaling routine in ARP/wARP (Morris, Zwart et al., 2004), with the main difference that the reference profile [1 + \(\gamma(h)_{\text{PDB}}\)] used here is replaced with an experimentally obtained Wilson plot (Popov & Bourenkov, 2003). Since the average experimental Wilson plot already contains bulk-solvent contributions, there is no need to include Babinet terms. The presented scaling procedure is similar to that described by Parthasarathy & Sekar (1993b).

The authors would like to thank G. Bourenkov, A. Popov, H. Schenk, A. Perrakis, R. J. Morris and G. Murshudov for stimulating discussions. PHZ thanks K. Cowtan for his patient help on the use of the Clipper Libraries (Cowtan, 2002).

**References**


