Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 12 February 2003
Accepted 23 April 2003

Anomalous difference signal in protein crystals

Qun Shen,[a]* Jun Wang[b] and Steven E. Ealick[b]

[a]Cornell High Energy Synchrotron Source (CHESS), Wilson Laboratory, Cornell University, Ithaca, NY 14853, USA, and [b]Department of Chemistry and Chemical Biology, Baker Laboratory, Cornell University, Ithaca, NY 14853, USA. Correspondence e-mail: qs11@cornell.edu

A scattering-angle-dependent formula is obtained for estimation of anomalous diffraction signals in Friedel pairs in protein crystals. In addition to the angle dependences of atomic form factor and temperature factor, a correlation form factor is introduced for the first time to take into account anomalous-scattering clusters such as disulfide bonds. Good agreements with experimental observations suggest that such correlated anomalous difference signals may be visible from measurements before a complete structural solution is obtained.

Single-wavelength (SAD) and multiple-wavelength anomalous diffraction (MAD) have become practical methods for phasing macromolecular crystal structures. Both methods make use of the anomalous intensity difference signals in Friedel pairs from the imaginary part f'' of the anomalous dispersion for a given anomalous-scattering element. In a SAD or MAD experiment, it is often desirable to have a good estimate of the anomalous signal to be measured in order to help design the overall experiment for a given protein. This is especially important for weak anomalous phasing signals from *e.g.* intrinsic sulfur atoms in native protein structures.

In the existing literature, estimates of average anomalous difference signal $\langle|\Delta F_\pm|\rangle$ of Friedel pairs are based on the following formula by Hendrickson & Teeter (1981):

$$\frac{\langle|\Delta F_\pm|\rangle}{\langle F\rangle} = \left(\frac{2N_A}{N_P}\right)^{1/2} \frac{f''}{Z_{\rm eff}}, \tag{1}$$

where N_A is the number of anomalous scatterers and N_P is the number of non-hydrogen atoms in the protein, $Z_{\rm eff}$ = 6.7 is the average number of electrons per non-hydrogen atom. This equation is valid for zero scattering angle $2\theta = 0$ only. Published data from SAD phasing experiments (Dauter *et al.*, 1999) show good general agreements with (1) at low resolutions, but consistent deviations exist at higher scattering angles. This is true even when angular dependences of the atomic scattering factor for protein atoms are taken into account (Dauter *et al.*, 2002). The deviations are discernable beyond the larger measurement errors associated with generally weaker reflections at higher resolutions.

In this communication, we show that the experimental discrepancy from (1) can be explained if, in addition to the atomic scattering angular dependence, an effective temperature factor and a pair-correlation form factor are included in the formula. For completeness, an elementary derivation of a modified equation (1) is given below and an empirical expression is introduced for the effective atomic form factor. Our results show good agreement with the experimental data on insulin.

For a given Bragg reflection $\mathbf{H}$, its structure factor is

$$F(\mathbf{H}) = F'(\mathbf{H}) + iF''(\mathbf{H}),$$

where

$$F'(\mathbf{H}) = \sum_j (f_j + f_j') \exp(i2\pi\mathbf{H}\cdot\mathbf{x}_j) \equiv F' \exp(i\varphi'), \tag{2}$$

$$F''(\mathbf{H}) = \sum_j f_j'' \exp(i2\pi\mathbf{H}\cdot\mathbf{x}_j) \equiv F'' \exp(i\varphi''). \tag{3}$$

For a Friedel pair $\pm\mathbf{H}$, we have that

$$\begin{aligned} F(\pm\mathbf{H}) &= F' \exp(\pm i\varphi') + iF'' \exp(\pm i\varphi'') \\ &= (F'\cos\varphi' \mp F''\sin\varphi'') + i(F''\cos\varphi'' \pm F'\sin\varphi') \end{aligned}$$

and its amplitude is

$$\begin{aligned} |F(\pm\mathbf{H})| &= [(F'\cos\varphi' \mp F''\sin\varphi'')^2 + (F''\cos\varphi'' \pm F'\sin\varphi')^2]^{1/2} \\ &= [F'^2 + F''^2 \pm 2F'F''\sin(\varphi' - \varphi'')]^{1/2} \\ &\approx (F'^2 + F''^2)^{1/2}\left[1 \pm \frac{F'F''\sin(\varphi' - \varphi'')}{F'^2 + F''^2}\right]. \end{aligned}$$

The last step is permitted since $F' \gg F''$. Thus the anomalous difference signal is given by

$$\begin{aligned} |\Delta F_\pm| &= \big||F(+\mathbf{H})| - |F(-\mathbf{H})|\big| \\ &\approx \frac{2F'F''|\sin(\varphi' - \varphi'')|}{(F'^2 + F''^2)^{1/2}}. \end{aligned}$$

Again since $F' \gg F''$, we obtain that (see also Blundell & Johnson, 1976)

$$|\Delta F_\pm| \approx 2F''|\sin(\varphi' - \varphi'')|. \tag{4}$$

If it is assumed that the anomalous scatterers are randomly located in the structure, the root-mean-square (r.m.s.) expectation value of $|\sin(\varphi' - \varphi'')|$ is $\langle\sin^2(\varphi' - \varphi'')\rangle^{1/2} = 1/2^{1/2}$. Thus the average anomalous signal becomes

$$\langle|\Delta F_\pm|\rangle = 2^{1/2}\langle F''\rangle.$$

Its ratio to the Friedel-averaged structure factor $\langle F\rangle$ is thus

$$\frac{\langle|\Delta F_\pm|\rangle}{\langle F\rangle} = \frac{2^{1/2}\langle F''\rangle}{\langle F\rangle}. \tag{5}$$

The average value of the structure factor $\langle F\rangle$ can be obtained in the following way. If it is assumed that N_P atoms with an effective form factor $f_{\rm eff}$ are randomly positioned in the unit cell, the average intensity is simply $N_P f_{\rm eff}^2$. Thus the effective amplitude or structure factor is

$$\langle F\rangle = N_P^{1/2} f_{\rm eff}.$$

Similarly, for N_A anomalous scatterers randomly positioned in the unit cell, the effective $\langle F''\rangle = N_A^{1/2} f''$. However, if the anomalous scatterers are clustered or correlated, we have that

$$\langle F''\rangle = N_A^{1/2} f'' C_A,$$

where C_A is a correlation form factor for the anomalous scatterers which is determined by interatomic distances of the scatterers and is scattering-angle-dependent.

To include the effect of thermal vibrations, we introduce an average difference B factor ΔB, defined as the difference between the average B factors of the protein atoms, B_P, and that of the anomalous atoms, B_A. In general, $B_A < B_P$ and thus $\Delta B > 0$ since the anomalous scatterers are more tightly restrained by the structure (although exceptions do exist). Therefore, (5) becomes

$$\frac{\langle|\Delta F_\pm|\rangle}{\langle F\rangle} = \left(\frac{2N_A}{N_P}\right)^{1/2} \frac{f''}{f_{\rm eff}} C_A \exp\left[\Delta B\left(\frac{\sin\theta}{\lambda}\right)^2\right], \qquad (6)$$

which agrees with (1) in the case of $2\theta = 0$ and $C_A = 1$.

Within the usual resolution range of 10^3–1 Å for protein crystals, the anomalous dispersion corrections f' and f'' are essentially independent of scattering angle or $\sin\theta/\lambda$ (Warren, 1969). The atomic form factor $f_{\rm eff}$, however, does depend on $\sin\theta/\lambda$. To take into account this dependence, we use the nine tabulated parameters for each element given in *International Tables for Crystallography* (2001), and an atom-population ratio of C:N:O:S = 4.83:1.31:1.45:0.03 to calculate $f_{\rm eff}$. The ratio is obtained by counting the number of amino acid residues in nine representative proteins (Richards, 1986). The result, $f_{\rm eff}$ *versus* $1/d = (2\sin\theta/\lambda)$, with d being the d spacing, is shown in Fig. 1 as open circles. We note that $Z_{\rm eff}$ given in (1) is simply the value of $f_{\rm eff}$ at $1/d = 0$. Also shown is a fit to the calculations using a Lorentzian plus a constant, which can be used as a general expression (inset, Fig. 1)for the atomic form factor of non-hydrogen atoms in a protein.

In Fig. 2, we show an example of the calculated anomalous signal, using (6), from sulfur atoms at $\lambda = 1.55$ Å in cubic insulin. There are $N_P = 532$ non-hydrogen atoms (including 81 ordered water molecules) and $N_A = 6$ sulfur atoms in insulin (Badger & Dodson, 1991). For sulfur, $f'' = 0.5567$ at $\lambda = 1.55$ Å (CXRO, 2002). We use a value of $\Delta B = 3$ for the B-factor difference between the protein and the sulfur atoms. With these values, (6) yields a d-spacing dependence (dashed curve, Fig. 2) that is in reasonable agreement with the measured values by Wang & Ealick (2003). Although some of the increase in the experimental values of $\langle|\Delta F_\pm|\rangle/\langle F\rangle$ with scattering angle may result from increased experimental errors as the count rates decrease at higher angles, some consistent deviation from the theoretical dashed curve is apparent even in the lower-resolution region.

A possible explanation of the deviation is the effect of correlated sulfur atom clusters, such as disulfide bridges. Indeed, when a spherically averaged correlation form factor known as the radial (pair) distribution function (Warren, 1969)

$$C_A = 1 + \frac{\sin(2\pi a/d)}{2\pi a/d} \qquad (7)$$

is included, with an S—S bond length $a = 2.0$ Å, the agreement with the data appears to be much better, as shown by the solid curve in Fig. 2. This suggests that it may be possible to deduce the existence of heavy-atom clusters from accurately measured anomalous difference ratios $\langle|\Delta F_\pm|\rangle/\langle F\rangle$ as a function of diffraction resolution, even before a structural solution is attempted. It is also interesting to note that, at extremely low resolutions, the correlated anomalous signal is always greater, by as much as a factor of two in the case of disulfide bonds, than the average value for random positions.

In conclusion, we have presented a general scattering-angle- or d-spacing-dependent formula, equation (6), for the calculation of anomalous difference signals in Friedel pairs of protein crystals. We have shown that the angle dependence arises from three factors: the atomic scattering factor of the protein atoms, the B-factor difference between the protein and the anomalous scatterers, and the correlation from possible anomalous atom clusters such as disulfide bridges. The results from the formula appear to agree with the experimental data very well in the case of the intrinsic sulfur signal from a cubic insulin crystal. The simple formula may be helpful in optimal designs of a SAD or MAD experiment.

This work is supported by NSF and NIGMS Grant DMR 97-13424 through CHESS, NIH Grant GM-46733 through Hauptman–Woodward Institute in Buffalo, and NIH Grant RR-15301 through the NE-CAT at Argonne.

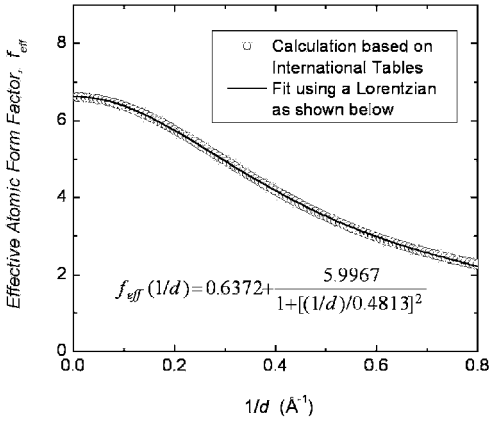

Figure 1
Effective atomic scattering factor $f_{\rm eff}$ for non-hydrogen protein atoms, shown as open circles, as calculated using *International Tables for Crystallography* (2001). The solid curve is a fit to the calculations using a Lorentzian plus a constant.

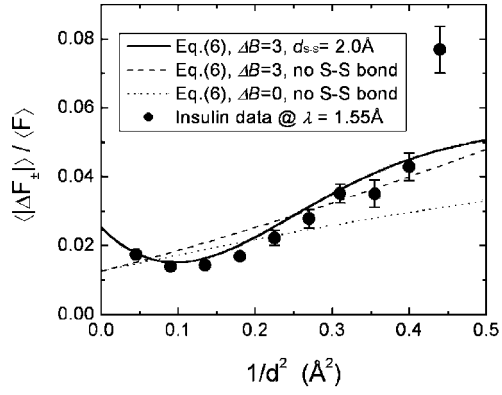

Figure 2
Calculated S anomalous difference signal $\langle|\Delta F_\pm|\rangle/\langle F\rangle$ using equation (6), compared with experimental data on insulin (Wang & Ealick, 2003). No S—S correlations are included for the dotted ($\Delta B = 0$) and the dashed curves ($\Delta B = 3$), while the solid curve includes a correlation form factor, equation (7), owing to disulfide bonds of $a = 2.0$ Å.

References

Badger, J. & Dodson, G. G. (1991). Deposited with the Protein Data Bank, PDB ID = 9INS.

Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. New York: Academic Press.

CXRO (2002). Center for X-ray Optics, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, http://www.cxro.lbl.gov/.

Dauter, Z., Dauter, M. & Dodson, E. (2002). *Acta Cryst.* D**58**, 494–506.

Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.

Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.

International Tables for Crystallography (2001). Vol. B. Dordrecht: Kluwer Academic Publishers.

Richards, F. M. (1986). *Protein Structure, Folding and Design*, pp. 171–196. New York: Alan R. Liss, Inc.

Wang, J. & Ealick, S. E. (2003). Proc. Am. Crystallogr. Assoc. Meet. Abstr. W0371.

Warren, B. E. (1969). *X-ray Diffraction.* Reading, MA: Addison-Wesley.