Acta Cryst. (1997). D**53**, 56–77

Displacement-Parameter Weighted Coordinate Comparison: I. Detection of Significant Structural Differences Between Oxidation States

CLARE PETERS-LIBEU AND ELINOR T. ADMAN*

Department of Biological Structure, Box 357420, University of Washington, Seattle, WA 98195–7420, USA. E-mail: adman@u.washington.edu

(Received 6 November 1995; accepted 10 June 1996)

Abstract

A method, displacement-parameter weighted coordinate comparison, for comparing closely related structures is developed. A 'probability of similarity' is derived from the overlap between the most probable volumes occupied by two analogous atoms in a comparison. The distribution of distances between atom pairs (difference distances), where each distance is weighted by the probability of similarity, is examined. The subset of atom pairs with normally distributed difference distances is used to estimate errors in the difference distances and to calculate the probability that the difference is significant for each atom-pair. These probabilities of difference are shown to correlate with features observed in difference maps for diffraction data from oxidized and reduced forms of the cupredoxin, pseudoazurin. The pair-wise probability of difference also is shown to be correlated with regions in previously published plastocyanin structures which differ upon copper oxidation state change, and which differ in a similar manner when temperature is changed, or pH is changed.

1. Introduction

On completing the refinement of a new variant of a known structure, one wants to determine which are significant differences among related models. In the absence of full-matrix least-squares refinement that would yield estimates of precision, one usually estimates an overall error or upper limits from the agreement of the model with the data using such methods as Luzzati plots (Luzzati, 1952) or *SIGMAA* plots (Read, 1986, 1990). However, these methods may overestimate the coordinate error for well ordered parts of the models. Moreover, often many small differences associated with significant density in difference maps between related structures are judged not significant by these methods.

Error estimates have been made previously by explicit functional correlation of the distance between positions of atom pairs (difference distances) in two structures to be compared and their displacement parameters (*B* factors) averaged over pairs of atoms with similar scattering factors (Chambers & Stroud, 1979; Perry *et al.*, 1990; Bott & Frane, 1990; Guss, Harrowell, Murata, Norris & Freeman, 1986). The reciprocal of displacement parameters has been used as a weight for calculation of an overall root-mean-square difference distance (Adman *et al.*, 1989). The basis of these approaches is the observation that atoms with larger *B* values have larger coordinate errors associated with them (Cruickshank, 1949). While these methods have been moderately successful for comparisons of proteins in which the displacement parameters are small or the difference in positions relatively large, we found them unsuitable for comparing models of oxidized and reduced pseudoazurin because features seen in difference maps were not detected by these methods.

Pseudoazurin is a small copper-containing electron-transfer protein, a member of the cupredoxin family, sharing a similar fold and copper coordination with other members of the family such as plastocyanin, azurin and amicyanin (Adman, 1991). In all these proteins, the copper center is maintained in a distorted tetrahedral geometry by a network of hydrogen bonds between other residues in the protein and the copper ligands. The unusual geometry of the copper center is midway between the tetrahedral form favored by Cu^{II} and the trigonal geometry favored by Cu^{I}. The characteristically positive redox potential of the copper center in the cupredoxin family may result from the network of hydrogen bonds favoring the reduced form of the protein.

As with the study of the two oxidation states of plastocyanin (Guss *et al.*, 1986), the goal of our work with pseudoazurin has been to determine whether the copper geometry is rigidly maintained upon reduction, and if the surrounding protein changes in response to the reduction of the copper. An $F_{o,ox} - F_{o,red}$ difference map indicates that the copper and one of its ligands, the S^{γ} atom of Cys78, as well as a second pair of residues (Met7 and Pro35) are shifted upon change in oxidation state (Nishiyama *et al.*, 1992). Similar shifts at the copper site were observed in the plastocyanin study. A common difficulty in both these studies is that the magnitude of the observed difference distances at the copper site between the oxidized and reduced forms of both proteins is on the order of the difference expected from random error in the models estimated through traditional methods. Average difference distances do not correlate well with average *B* values between replicate structures of pseudoazurin or between different oxidation states, precluding use of the previously published

ISSN 0907-4449

average-error estimation methods. This same lack of functional correlation with B value precluded use of the more recently described procedure of Stroud & Fauman (1995). Thus, we sought an approach to estimating error that would allow us to look at individual difference distances.

The displacement parameter (B factor) describes the spatial distribution of possible positions of the electrons associated with each atom and is usually assumed to be isotropic. Together the mean position and displacement parameter define the most probable volume for each atom. The overlap between the most probable volumes for each pair of analogous atoms can be used to judge the significance of the difference distance between the atoms. If their overlap is equivalent, pairs of atoms with unusually large difference distances and displacement parameters can be assigned the same level of confidence as pairs of atoms with unusually small displacement parameters and difference distances. In our method, displacement-parameter weighted coordinate comparison (DPWCC), we use the overlap as a weight to obtain a set of weighted difference distances whose distribution is normal, and, therefore, whose properties are related to the random component of the error in the comparison and thus most likely to reflect model error, independent of systematic differences. We will show that DPWCC allows us to report as significant those differences in coordinates that are quite apparent in a difference map but which frequently are relegated to being not significant because the coordinate error estimated from aggregate methods (such as Luzzati plots) is too large. In this paper, we develop DPWCC as a method for comparing closely related coordinate sets, from duplicate determinations, single-site mutants, or different oxidation states of a protein, and in a future paper we will show how the method can be used in structures where differences such as domain motion can be evaluated.

2. Methods

2.1. *Derivation of weighting function*

The collection of difference distances between two coordinate sets (1 and 2) containing N atoms can be considered to be estimates (or measurements) of an overall mean difference distance with an associated variance. If the difference distances are normally distributed, that is, there are no systematic differences between coordinate sets, then calculation of the mean of the distribution is equivalent to calculating the grand mean of a large number of measurements, as if the relative position of a pair of points has been determined independently a large number of times and used in N independent comparisons. The variance could then represent the precision of that 'measurement'. In the comparison of protein models it is more useful to think of the the distribution as composed of two subsets: a large number of difference distances which result from random error and a number of difference distances which arise from either systematic errors or 'true' significant differences. In order to identify pairs of atoms with significant difference distances, we need to first ensure that a set of difference distances exists that is normally distributed.

It seems reasonable to first evaluate some weights, $W_s(d_i)$, that will reflect the confidence in the determination of each difference distance, d_i. These weights can be used to calculate d_W, a weighted average difference distance,

$$d_W = \sum W_s(d_i)d_i / \sum W_s(d_i), \tag{1}$$

between mean positions for all pairs of equivalent atoms. The variance of the weighted distribution is the weighted variance $\sigma^2_W(1,2)$,

$$\sigma^2_W(1,2) = [\sum W_s(d_i)(d_i - d_W)^2] / \sum W_s(d_i). \tag{2}$$

If d_W and $\sigma_W(1,2)$ describe the distribution of a large fraction of the difference distances, then it seems reasonable that d_W and $\sigma_W(1,2)$ reflect the random error in the comparison of the protein models, and that the outliers to this distribution contain significant differences.

We suggest that a suitable weight is the probability that the two atoms are not in different positions, given their individual isotropic displacement parameters. This probability can be calculated from the overlap predicted from the individual B values. If the B values are such that the atom-pair overlap is complete, the atom positions are not significantly different.

The displacement parameter (B) for an atom assumed to be undergoing three-dimensional isotropic simple harmonic motion, is defined as,

$$B = 24\pi^2\langle r^2\rangle, \tag{3}$$

where $\langle r^2\rangle$ is the amplitude of the harmonic motion (and $\langle r^2\rangle = \langle u_x^2\rangle + \langle u_y^2\rangle + \langle u_z^2\rangle$) (James, 1982). Together the position and the displacement parameter specify an isotropic Gaussian probability distribution for the atom. An equivalent description of the atom position is that the probability distribution for the electron density associated with the atom, $D(r)$, is Gaussian or,

$$D(r) = (1/2\pi\sigma^2)^{3/2} \exp(-|r_o - r|^2/2\sigma^2), \tag{4}$$

where σ is related to the width of the Gaussian distribution, and r_o is the mean position of the electrons. Combining (1) and (2), the probability density for an atom with a mean position of $r_o = (x_o, y_o, z_o)$ and a displacement parameter of B can be shown to be,

$$D(r) = (4\pi^2/B)^{3/2} \exp[-(4\pi^2/B)|r_o - r|^2]. \tag{5}$$

$D(r)$ is the probability that the atom can be found in a volume centered on position **r** given simple harmonic motion.

For each pair of atoms, the probability, which we call the probability of similarity, that the difference (d) in position between the mean position of atom 1 and the mean position of atom 2, is not significantly different from zero is equivalent to the probability that the observed probability distribution for atom 1 and the observed probability distribution for atom 2 result from measurements of the same probability distribution.

Given a Gaussian distribution of difference distances one can show that (Bevington, 1969) the probability $P(d)$ that a value $r \leq d$ (between two observations) will be observed is,

$$P(d) = 2(c/\pi^{1/2}) \int_d^\infty dr \exp-(cr)^2$$
$$= 1 - \mathrm{erf}(cd)$$

From this we define the probability of similarity, $P_s(d)$, with the constant $c = 1/[(2)^{1/2}\sigma(d)]$ as,

$$P_s(d) = 1.0 - \mathrm{erf}[d/(2)^{1/2}\sigma(d)]. \quad (6)$$

If one takes the variance of a particular difference distance as the sum of the variances of the probability density for atom 1 and atom 2.

$$\sigma^2(d) = (B_1 + B_2)/8\pi^2, \quad (7)$$

and the difference distance is,

$$d = |\mathbf{r}_1 - \mathbf{r}_2|. \quad (8)$$

then $P_s(d)$ can be written in terms of the mean positions of the atoms and their displacement parameters as,

$$P_s(d|r_1, B_1, r_2, B_2) = 1.0 - \mathrm{erf}\{[4\pi^2/(B_1 + B_2)]^{1/2} \times |\mathbf{r}_1 - \mathbf{r}_2|\}, \quad (9)$$

where the subscripts 1 and 2 refer to the first and second atom in the pair.*

The probability P_s tends towards 1.0 when the mean positions of the atoms are the same and tends towards 0.0 when the atoms are far apart. More importantly for the present analysis, if the average displacement parameter is very large then the probability that the atoms are not significantly different again tends towards 1.

Different scattering factors must also be taken into account in evaluating the contribution of an individual atom pair to the collection of distances. This can be approximated by multiplying the probability of similarity by the ratio of the atomic numbers such that,

$$W_s(d_i) = P_s(d_i)Z_h/Z_l, \quad (10)$$

where Z_l is the atomic number of the lightest atoms in the model, Z_h that of the heavier atoms and $W_s(d_i)$ is the final weight.

2.2. *Identification of a normally distributed subset of difference distances*

Calculation of the probability of similarity assumes that the transformation of the second protein structure places the second atom in the pair within the peak of the probability function of the first atom, for the majority of atoms. If this is not the case, there will be a large number of difference distances apparently not normally distributed, and it could indicate, for example, that one region of the structure has changed significantly, and a transformation matrix has to be chosen using the subset of atoms which will satisfy the above assumption.

Whether or not a subset of the distribution of difference distances follows a normal distribution can be determined by plotting f_o, the fraction of all distances d_i with normal coordinate $(d_i - d_W)/\sigma < z_i$,

$$f_o^i = 1/N \sum_{k=1}^{j} n(z_i), \quad (11)$$

[where N = total number of distances (number of atoms in the comparison) and z is the particular bin of $(d_i - d_W)/\sigma$, n_i is the number of distances in that bin] *versus* the predicted value, f_p, of the standard normal distribution of the normal coordinate z,

$$f_p^i = \mathrm{erf}(z_i/2). \quad (12)$$

If the difference distances are distributed normally with a weighted mean, d_W, and weighted variance $\sigma_W^2(1,2)$, then the plot of f_o *versus* f_p should be linear with slope 1.0 and intercept 0.0. For the distribution of difference distances, the slope and the intercept of f_o *versus* f_p, their errors and correlation coefficient, c,

$$c = (f_o^i - \mathbf{f}_o)(f_p^i - \mathbf{f}_p)/\{[\sum(f_o^i - \mathbf{f}_o)^2][(f_p^i - \mathbf{f}_p)^2]\}^{1/2}, \quad (13)$$

($\mathbf{f}_o$ and $\mathbf{f}_p$ are the respective means of the observed and predicted fractions) can be calculated using least squares. The correlation coefficient is a sensitive test of the linearity of the plot (Pagano & Gauvreau, 1993). If the slope is equal to one within the estimated error in the slope, the intercept equal to 0.0 within the error of the intercept and the correlation coefficient greater than 0.95, then we consider the subset of difference distances to be normally distributed.

2.3. *Calculation of average error of positions*

We suggest that the weighted standard deviation, $\sigma_W(1,2)$ of the normally distributed difference distances can be regarded as an estimate of the error associated with determining the positions of the atoms if the weighted mean d_W is not statistically different from zero. The statistical significance of the weighted mean can be tested by calculating the coefficient of agreement,

* $\mathrm{erf}(x)$ is the error function, which is evaluated as $\mathrm{erf}(x) = (2/\pi^{1/2})\int_0^x \exp(-t^2)dt$. The error function is zero when x equals zero and 1 when x is infinite.

$$C = \exp[-d_W^2/2\sigma_W^2(1,2)]. \quad (14)$$

If the coefficient of agreement is less than 0.25, the weighted mean fails the test at a 95% confidence level and the distribution has a finite probability of having not arisen as a random sampling of a Gaussian distribution with a mean of zero.

If the mean is statistically different from zero, then there may be a systematic displacement of large numbers of residues. Two examples of such an effect are either poor superpositions or a significant difference in a hinge angle between two domains. Distributions in which the d_W is significantly different from zero must be analyzed differently (Libeu & Adman, 1997).

On the other hand, if many of the difference distances are nearly zero (for example in comparison of two sequential steps of refinement), then the mean and the variance may represent only a small number of non-zero difference distances in which case the weighted mean is not equivalent to the grand mean of a large number of independent comparisons, and a comparison is not valid. Very large values of the coefficient of agreement indicate a high degree of correlation between the two models.

$\sigma_W(1,2)$ may be an overestimate of the error in comparison of positions of atoms such as S or metal atoms since the calculation of $\sigma_W(1,2)$ is dominated by contributions from lighter atoms. However, for well resolved atoms with low displacement parameters determined from complete data sets, the estimated error for heavier atoms can be shown to be approximately proportional to the curvature of the electron density (Cruickshank, 1949). Since the ratio of the curvature for different types of atoms can be approximated as the inverse ratio of their atomic numbers, one can then estimate the error for heavier atoms as,

$$\sigma_h \simeq Z_l\sigma_l/Z_h, \quad (15)$$

Z_l and Z_h are the atomic numbers of the light atom and the heavy atom, respectively, and σ_l is the estimated error of the light atom.

Once a normally distributed subset of difference distances is found, along with d_W and σ_W, the entire list of weighted difference distances can be examined to tabulate which distances fall within the normally distributed subset, and which do not. The pairs of atoms which have a high probability of being different, are pairs of atoms which can be considered to be significantly different given both their displacement parameters and the standard deviation of the distribution of difference distances. These pairs of atoms have both a low $P_s(d_i)$ [(9)] and a high normal coordinate $(d_i - d_W/\sigma)$. The joint probability we call the probability of difference,

$$P_d(d_i) = [1.0 - W_s(d_i)]\mathrm{erf}[(d_iZ_h/Z_l - d_W)/(2)^{1/2}\sigma_W(1,2)] \quad \text{if } (d_iZ_h/Z_l - d_W)/(2)^{1/2}\sigma_W(1,2) > 0$$
$$= 0.0 \quad \text{if } (d_iZ_h/Z_l - d_W)/(2)^{1/2}\sigma_W(1,2) \le 0. \quad (16)$$

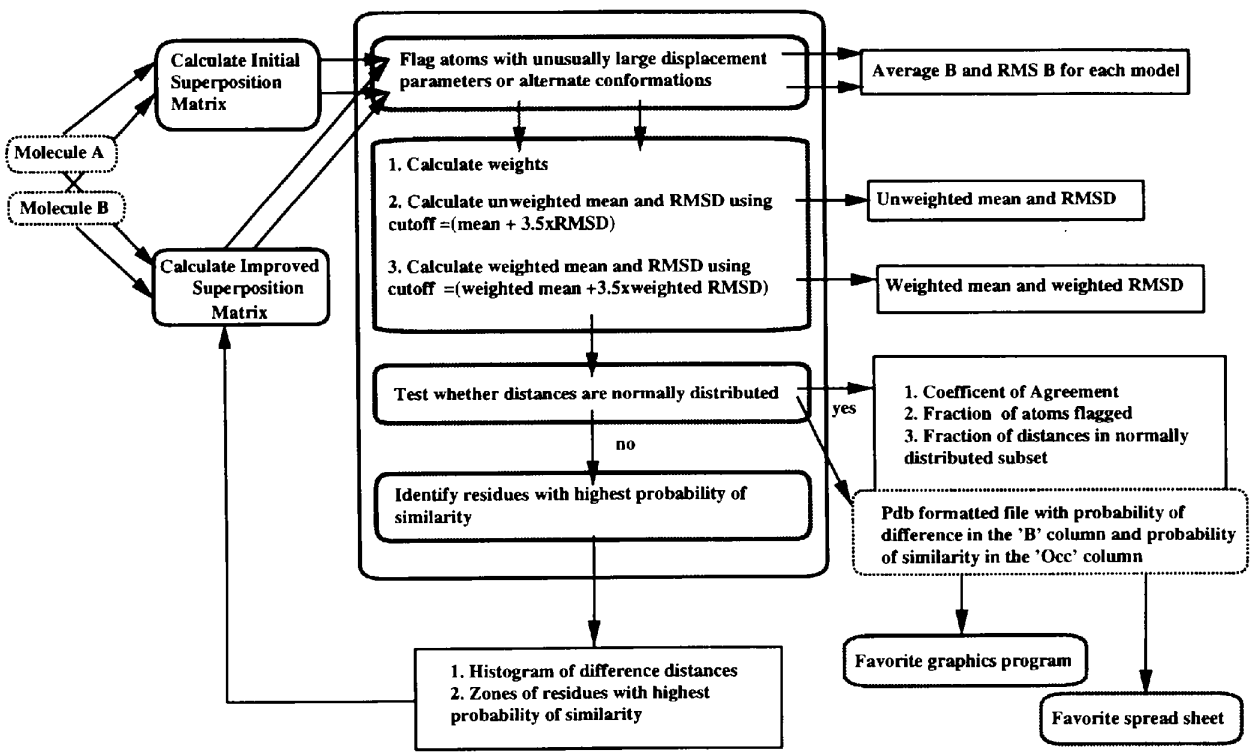


Fig. 1. Flow chart of the displacement-parameter coordinate comparison method.

Table 1. *Refinement statistics for models used in DPWCC comparisons (na = not available)*

(*a*) Pseudoazurin

PDB* code	Model	B_{av}	d_{min} (Å)	d_{max} (Å)	R§	No. of reflections	% Complete	No. of atoms	Deviation bonds (Å)	Deviation angles (°)	Refinement method	Max average error¶	DPI** (Å)
	Seattle oxidized†	20	22.3	1.6	0.17	16225	88	1051	0.03	2.0	*X-PLOR*	0.09	0.06
	Seattle reduced†	25	22.3	1.7	0.16	12839	84	1077	0.02	2.0	*X-PLOR*	0.11	0.07
1paz	Athens oxidized‡	21	9.0	1.55	0.18	18269	91	1026	0.01	2.2	*PROLSQ*	0.15	0.05
1pza	Athens reduced‡	25	10.0	1.8	0.18	12660	99	1016	0.017	3.2	*PROLSQ*	na	0.08

(*b*) Plastocyanin

1pcy	Oxidized††	14	7.0	1.6	0.17	8285	82	784	0.02	na	*PROLSQ*	0.20	0.08
2pcy	Apo‡‡	13	—	1.8	0.16	7891	100	781	na	na	*PROLSQ*	na	0.08
3pcy	Hg-Pc§§	12	5.5	1.9	0.16	6206	100	799	na	na	*PROLSQ*	na	0.11
4pcy	Red pH 7.8¶¶	16	7.0	2.1	0.15	4324	92	774	0.02	8.7	*PROLSQ*	na	0.18
5pcy	Red pH 7.0¶¶	15	7.0	1.8	0.16	6971	88	784	0.02	8.0	*PROLSQ*	na	0.10
6pcy	Red pH 3.8¶¶	16	7.0	1.9	0.15	5810	86	787	0.02	7.6	*PROLSQ*	na	0.11
1pnd	Ox 173 K¶¶	8	6.0	1.6	0.15	7393	74	936	0.01	2.3	*PROLSQ*	0.13	0.08
1pnc	Ox 173 K***	5	8.0	1.6	0.15	7393	73	909	na	na	*EREF*	na	0.09
1plc	Ox 1.33†††	13	8.0	1.33	0.15	14303	76	895	0.02	2.6	*PROLSQ*	0.15	0.04

*Bernstein *et al.* (1977). †Libeu *et al.* (1997). ‡Vakoufari, Wilson & Petratos (1994). §$R = \sum ||F_o| - |F_c|| / \sum F_o$. ¶From Luzzati plots. **Diffraction precision indicator (Cruickshank, 1996). ††Guss & Freeman (1983). ‡‡Garrett, Clingeleffer, Guss, Rogers & Freeman (1984). §§Church, Guss, Potter & Freeman (1986). ¶¶ Guss, Harrowell, Murata, Norris & Freeman (1986). ***Fields *et al.* (1994). †††Guss, Bartunik & Freeman (1992).

We will show below that $P_d(d_i)$ is highly correlated with peaks in a difference Fourier map calculated with observed diffraction data, and thus becomes a useful criterion for screening entire structures for significant differences.

2.4. *Implementation*

The overall procedure used in the comparisons is summarized in Fig. 1. A trial rotation matrix is calculated using the main-chain atoms with equal weights. The least-squares procedure of Kabsch (1978) as implemented in *LSQKAB* in *CCP*4 (Collaborative Computational Project, Number 4, 1994) is used to determine the initial transformation matrix. Then the various statistics, d_W, σ_W, linearity coefficient [c, (13)], agreement [C, (14)], probability of similarity [P_s, (9)] are calculated. If a subset of normally distributed difference distances is not found ($c \leq 0.95$, $C < 0.25$) using the initial transformation matrix, or if there are regions of the main chain that have consistently low probabilities of similarity, then the transformation matrix is recalculated excluding all atoms with low probability of similarity. In all the comparisons discussed in this paper coordinate sets were aligned using the main-chain atoms excluding those from residues with displacement parameters over 60 Å^2. Improvement of the transformation matrices was not required.

The structures of the oxidized and reduced forms of native pseudoazurin have each been determined independently in our laboratory (Seattle) (Libeu, Kukimoto, Nishiyama, Turley & Adman, 1997) and in the laboratory of Dr Petratos (Athens) (Vakoufari, Wilson & Petratos, 1994; Petratos, Banner, Beppu, Wilson & Tsernoglou, 1987). All four models are related in that the original Athens oxidized model was used to complete the 2 Å Seattle model of oxidized pseudoazurin (Adman *et al.*, 1989), and the starting model for each reduced model was its corresponding oxidized model. Similarly, the plastocyanins are all related by the use of an initial model of the oxidized form to solve the other models. Therefore, the reduced plastocyanin pH 7.8, pH 7.0 and pH 3.8 models (Guss *et al.*, 1986) as well as the oxidized 1.33 Å and low-temperature models (Fields *et al.*, 1994; Guss, Bartunik & Freeman 1992) are descendents of the 1.6 Å oxidized plastocyanin model. Table 1 summarizes the relevant statistics for each model used in this paper.

3. Results

3.1. *Estimate of average error*

An overall estimate of the average error was obtained from the weighted mean of difference distances and σ_W for the pseudoazurin and plastocyanin comparisons shown in Table 2 (*a* and *b*). For comparison, we also calculated the unweighted mean difference distance and unweighted root-mean-square deviation (r.m.s.d.), by first calculating an overall mean and r.m.s.d. and then eliminating atoms whose difference distances were greater than 3.4 times the r.m.s.d. plus the mean and recalculating the mean and r.m.s.d. with the remaining pairs. The large variation (from 0.10 to 0.34 Å) in the unweighted r.m.s.d. for pseudoazurin is primarily because of different interpretations of partially disordered loops, particularly the loops containing residues 52–55 and 93–96, as well as residues in the C-terminal helix. The Athens structures were compared to omit maps of loops calculated with the Seattle data and both interpretations appear to fit the data equally well. By taking appropriate

Table 2. *Statistics from displacement-parameter-weighted coordinate comparison*

(*a*) Pseudoazurin

Model 1	Model 2	d* (Å)	R.m.s.d† (Å)	d_w‡ (Å)	σ_w§ (Å)	%atoms ¶	Agreement**	σ_{DPI}†† (Å)
Seattle Oxidized	Athens Oxidized	0.16	0.32	0.11	0.09	98	0.47	0.14
Seattle Reduced	Athens Reduced	0.16	0.13	0.12	0.09	98	0.41	0.19
Athens Oxidized	Athens Reduced	0.15	0.10	0.11	0;08	98	0.35	0.16
Seattle Oxidized	Seattle Reduced	0.22	0.31	0.15	0.11	97	0.39	0.16
Seattle Oxidized	Athens Reduced	0.20	0.18	0.15	0.11	98	0.38	0.17
Seattle Reduced	Athens Oxidized	0.18	0.34	0.12	0.10	97	0.46	0.16

(*b*) Plastocyanin

Model 1	Model 2	d (Å)	R.m.s.d. (Å)	d_w (Å)	σ_w (Å)	% Atoms	Agreement	σ_{DPI} (Å)
Oxidized 1.6 Å	Apo	0.19	0.21	0.13	0.11	97	0.49	0.20
	Hg substituted	0.28	0.29	0.17	0.13	95	0.43	0.23
	Reduced pH 7.8	0.18	0.13	0.13	0.10	98	0.37	0.35
	Reduced pH 7.0	0.17	0.12	0.13	0.09	98	0.36	0.21
	Reduced pH 3.8	0.23	0.21	0.16	0.11	98	0.37	0.24
	Oxidized 173 K‡‡	0.39	0.45	0.22	0.14	95	0.30	0.20
	Oxidized 173 K§§	0.40	0.50	0.21	0.14	94	0.32	0.21
	Oxidized 1.33	0.15	0.20	0.11	0.08	98	0.41	0.15
Apo	Hg substituted	0.34	0.35	0.20	0.15	95	0.39	0.23
	Reduced pH 7.8	0.23	0.21	0.16	0.11	97	0.35	0.35
	Reduced pH 7.0	0.21	0.21	0.14	0.11	97	0.42	0.22
	Reduced pH 3.8	0.27	0.22	0.19	0.14	98	0.40	0.24
	Oxidized 173 K‡‡	0.36	0.37	0.22	0.14	95	0.29	0.20
	Oxidized 173 K§§	0.38	0.47	0.21	0.13	94	0.30	0.22
	Oxidized 1.33	0.21	0.35	0.13	0.09	97	0.33	0.20
Hg substituted	Reduced pH 7.8	0.29	0.26	0.19	0.14	95	0.40	0.37
	Reduced pH 7.0	0.28	0.29	0.17	0.14	95	0.46	0.25
	Reduced pH 3.8	0.26	0.27	0.16	0.13	95	0.45	0.27
	Oxidized 173 K‡‡	0.49	0.56	0.25	0.16	91	0.33	0.24
	Oxidized 173 K§§	0.51	0.62	0.23	0.16	89	0.32	0.25
	Oxidized 1.33	0.28	0.40	0.16	0.12	95	0.41	0.32
Red pH 7.8	Reduced pH 7.0	0.18	0.12	0.14	0.09	98	0.32	0.36
	Reduced pH 3.8	0.24	0.21	0.17	0.11	98	0.33	0.38
	Oxidized 173 K‡‡	0.41	0.38	0.25	0.15	95	0.28	0.35
	Oxidized 173 K§§	0.41	0.39	0.23	0.15	94	0.32	0.36
	Oxidized 1.33	0.21	0.18	0.15	0.10	98	0.29	0.19
Red pH 7.0	Reduced pH 3.8	0.19	0.20	0.13	0.10	98	0.39	0.26
	Oxidized 173 K‡‡	0.40	0.39	0.24	0.14	95	0.25	0.22
	Oxidized 173 K§§	0.40	0.42	0.23	0.14	94	0.28	0.23
	Oxidized 1.33	0.20	0.21	0.14	0.09	98	0.28	0.19
Red pH 3.8	Oxidized 173 K‡‡	0.46	0.44	0.25	0.17	92	0.32	0.24
	Oxidized 173 K§§	0.46	0.45	0.24	0.16	92	0.33	0.26
	Oxidized 1.33	0.26	0.25	0.17	0.12	98	0.34	0.20
Oxidized 173 K‡‡	Oxidized 173 K§§	0.11	0.08	0.08	0.05	98	0.29	0.22
	Oxidized 1.33	0.33	0.29	0.21	0.13	96	0.29	0.15
Oxidized 173 K§§	Oxidized 1.33	0.32	0.30	0.20	0.13	96	0.31	0.17

* Mean difference distance, $\sum d_i/n$. † Root-mean-square deviation of difference distances $[\sum(d_i - d)^2/n]^{1/2}$. ‡ d_w, equation (1). § σ_w, equation (2). ¶ Percentage of atoms that have difference distances that are normally distributed. **Coefficient of agreement, equation (14). †† Diffraction precision indicator, equation (18). ‡‡ *PROLSQ*. §§ *EREF*.

account of loops with high B values DPWCC suggests more consistent errors of 0.09–0.11 Å. The same analysis of plastocyanin yields overall errors of 0.05–0.17 Å, where the unweighted r.m.s.d. yielded 0.08–0.62 Å.

The distribution of difference distances, average displacement parameters and the probability of similarity for both pseudoazurin and plastocyanin are shown in Fig. 2, along with the probability of similarity as a function of difference distance. In the pseudoazurin comparison, the ten pairs of atoms with an unusually high probability of similarity and difference distances greater than $3.4\sigma_W$ are the terminal atoms of surface residues in which

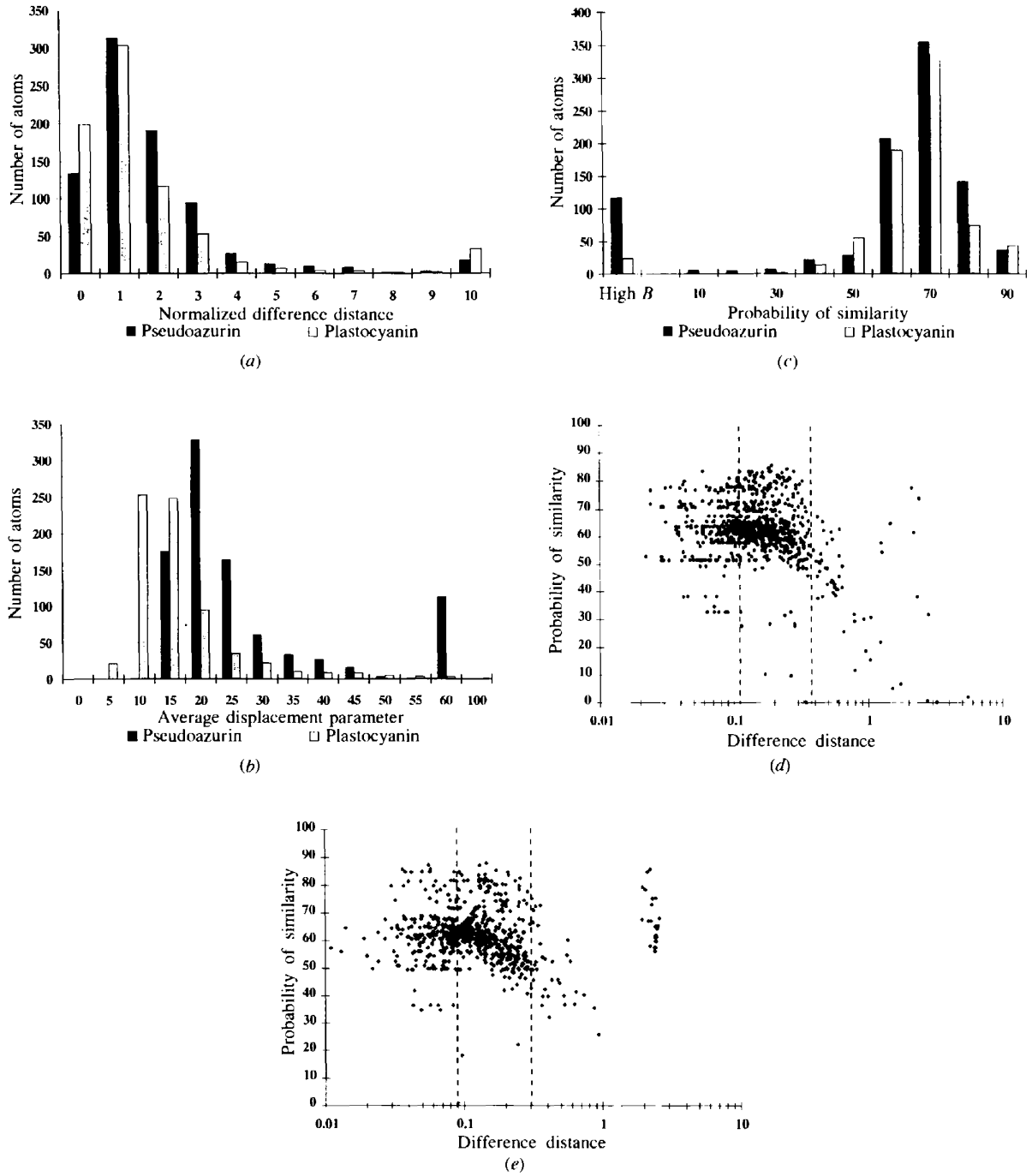


Fig. 2. Relationship between the probability of similarity and difference distance. (*a*) Distribution of difference distances for comparisons of Seattle oxidized and reduced pseudoazurin, oxidized 1.6 Å plastocyanin and reduced pH 7 plastocyanin, (*b*) the distribution of the average displacement parameters for the same comparisons. (*c*) The distribution of the probability of similarity for the same comparisons. The 'high *B*' column represents the 10% of the pairs of atoms that were excluded before the start of the calculations. These are the pairs of atoms in which one atom had a *B* value greater than 60 $Å^2$ or was in side chains that have been modeled in more than one conformation. In the pseudoazurin structures, elimination of the pairs with *B* values greater than 60 $Å^2$ primarily removed the atoms from the first N-terminal residue and the final three C-terminal residues in addition to the terminal atoms of some surface lysines. All of these atoms were not associated with convincing electron density in the final difference maps for the refinement of the oxidized and reduced pseudoazurin models. Removal of these atoms does not affect the calculation of the weighted mean and σ_W, but is necessary to reduced the number of differences which appear significant, but are not, in the identification of the pairs of atoms with significant deviations. The relationships between difference distance and probability of similarity are shown in (*d*) for the pseudoazurin comparison, (*e*) for the plastocyanin comparison. The vertical bars in (*d*) and (*e*) are σ_W and $3.4\sigma_W$ for each comparison.

the displacement of the side chain involves a rotation around a Cβ or Cγ. All of these pairs have convincing density in both final $2F_o - F_c$ maps, but residual density in the final $F_o - F_c$ maps. We believe that these pairs are outliers because the isotropic displacement parameters are not an adequate description of electron density resulting from libration of these residues in the crystals. These residues would most likely be better modeled with anisotropic thermal parameters. None of the pairs in this last category contribute to the weighted mean and σ_W. Although the two proteins are topologically very similar, each has a characteristic distribution of difference distances, average displacement parameter, and probability of similarity.

Fig. 3 shows the dependence of the probability of difference (16) on the difference distance for the comparison of the oxidized and reduced Seattle models and the comparison of the oxidized and reduced pH 7 plastocyanin. Pairs of distances which fall in the upper left (region I) are atoms with high probability of difference, but small difference distance, while those in the lower right (region III) show large differences, but are unlikely to be significant because the probability of difference is small. Regions I and II contain the difference distances likely to be significant. The probability of difference is less than 40% for 80% of the atom pairs so that about 20% of the atom pairs can be considered as possibly significantly different.

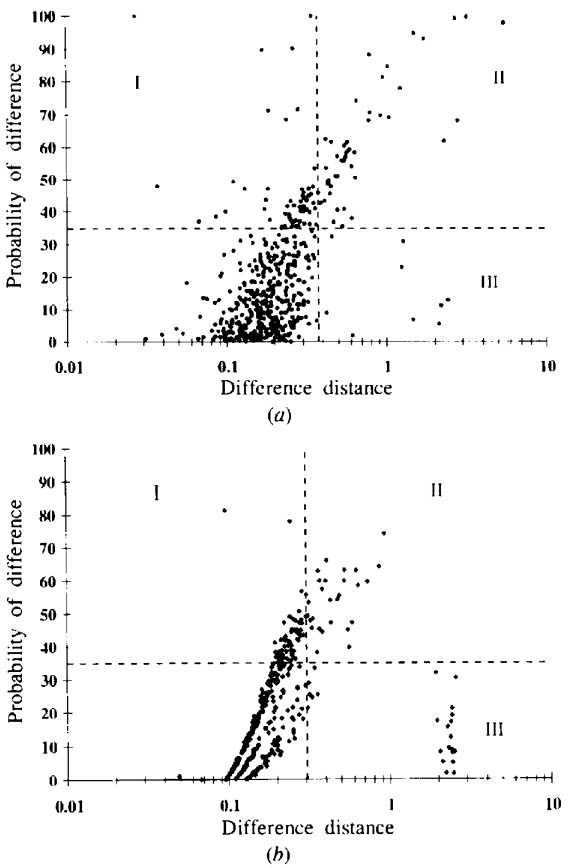


Fig. 3. The dependence of the probability of difference upon difference distance (*a*) the comparison between Seattle oxidized and Seattle reduced pseudoazurin models, (*b*) the comparison between the oxidized 1.6 Å plastocyanin and reduced pH 7.0 plastocyanin models. In order to simplify the plots only the atoms with probability of difference greater than zero have been plotted. The vertical dotted line represents $3.4\sigma_W$ for each, the horizontal dotted line, a probability of difference of 0.35. Regions I, II and III are discussed in the text.

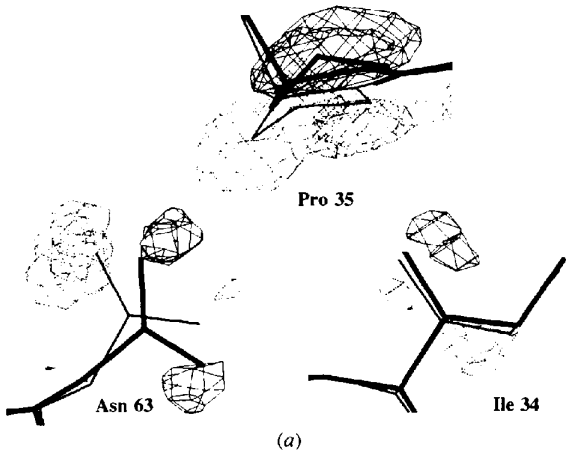
The differences in region I ($3.4\sigma_W = 0.37$ for the pseudoazurin comparison and 0.31 for plastocyanin comparison) include S^γ and the carbonyl O atom of the cysteine ligand in both comparisons as well as the Cu atom in pseudoazurin. Although both the S atom of cysteine and the Cu atom are preferentially weighted because of the dependence on atomic weight of the probability of difference, high atomic number does not guarantee a high probability of difference: neither the S^δ of methionine ligand in both comparisons or the S^δ of Met16 in pseudoazurin have high probability of difference. Like the lighter atoms in region I both S^γ of Cys78 and the Cu atom have average displacement parameters between 10 and 20 $Å^2$, which are unusually small for their difference distances given their atomic number. Thus, these are significant shifts.

3.2. *Correlation of the probability of difference with difference Fourier maps*

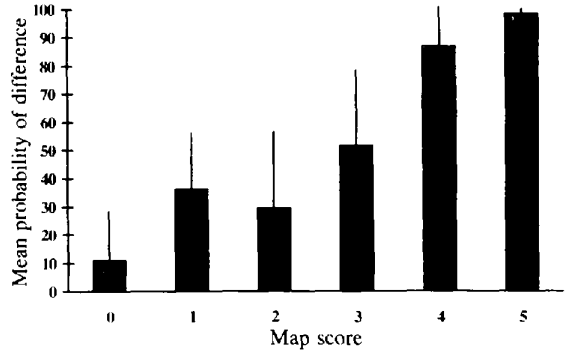
Before the probability of difference can be used reliably to predict significant differences it is necessary to show that it correlates with features in difference Fourier maps.* Before examining the differences predicted by DPWCC we devised a scoring system to quantify the character of difference density found at each atom position. Atoms associated with difference density greater than 5σ and good density for all of the side-

* The relationship between the probability of difference and the magnitudes of difference Fourier map peaks observed for a pair of atoms cannot be derived in a general way since, for real models, the relationship between the height of the difference peak and the distance between corresponding pairs of atoms will be a function of the completeness of the data, resolution, and phase errors as well as any other factors that contribute to the formation of ripples in the electron-density map. However, a minimum requirement for correlation between a pair of atoms having a high probability of difference and large peaks in the difference map is that both functions must have extrema under the same conditions. The probability of difference [(16)] is a monotonically decreasing function that asymptotically approaches zero when the difference distance approaches zero or the displacement parameter becomes very large. That is, when the probability of difference is zero, the volume containing the pair of atoms in the $|F_{o1} - F_{o2}|$ difference map should not contain any significant difference density. At the other extreme, atoms which are at clearly different positions will have maximum probability of difference, and will also have non-overlapping electron distributions and, therefore, large peaks in difference Fourier maps. Therefore, in most cases, the atoms with highest peaks in the difference map should also have a high probability of difference when the models are compared, if the quality of the diffraction data is sufficient to resolve the two atoms.

chain atoms in the residue that are different between the two models were assigned a map score of 5. Pro35 in Fig. 4(*a*) is an example of a residue with a map score of 5. Atoms associated with 5σ difference peaks but attached to atoms in the residue not associated with difference density were assigned a map score of 4. A map score of 3 was assigned to atoms that had clear easily interpretable shift peaks at the 3σ level, while a map score of 2 was assigned to atoms associated with 3σ density that was not clearly interpretable. In Fig. 4(*a*), Asn63 is an example of a residue given a map score of 3, and Ile34, a map score of 2. A final category of map score equal to 1 was assigned to atoms with no 3σ difference density, but bonded to an atom associated with difference density. The latter category includes atoms that are likely to move because of the restraints used in the refinement. Fig. 4(*b*) shows the results of the map examination for oxidized *versus* reduced pseudoazurin. Atoms with difference probabilities greater than 80% account for all of the 5σ difference peaks associated with the protein, except for the very large peak near the Cu. Similarly, all pairs with a map score greater than 4 had a probability of difference greater than 60%. The large peak at the copper site is equivalent to a small shift in the copper position, 0.037 Å, and a relatively large change in the B value of the copper from 15 to 25 Å^2. The resulting difference probability is 48%. The wider range of peaks assigned a map score of 3 is principally caused by an artificial elevation of the height of their associated shift peaks resulting from proximity of atoms in ligand residues to the large shift peak associated with the copper. Nevertheless, the overall map score and the probability of difference are significantly correlated with a correlation coefficient of 0.62 for 835 pairs of atoms.



(*a*)



(*b*)

Fig. 4. Correlation between the $F_{o,ox} - F_{o,red}$ difference map and DPWCC mean probability of difference for the comparison of Seattle oxidized and Seattle reduced pseudoazurin. The height of the bars is the mean probability of difference (16) associated with each map score, while the error bars represent the ± standard deviations of the probability of difference. (*a*) Examples of map scores. The atoms of the Pro35 are an example of atoms with a map score of 5. $N^{\delta 1}$ of Asn63 was assigned a map score of 3 and C^1 of Ile34 a map score of 2, respectively. Map scores of 1 were assigned to atoms connected to atoms with higher maps scores, but not associated with difference peaks, while map scores of zero were assigned to atoms in residues not associated with difference peaks. The dark contours are +3σ difference density and the dotted contours are −3σ difference density. The thick residues are from the Seattle oxidized model, while the thin residues correspond to the Seattle reduced model. (*b*) The mean and root-mean-square deviation (error bars) of probability of difference for each category of map score.

3.3. *Use of individual difference probabilities*

Since difference probabilities are reasonably well correlated with peak height in the $F_{o,ox} - F_{o,red}$ difference maps, it is now possible to screen for significant differences among the coordinate sets on that basis. Six pseudoazurin comparisons are shown in Fig. 5 where the probability of difference is shown for each atom in the sequence. The comparisons between the two oxidized models (Fig. 5*a*) and the two reduced models (Fig. 5*d*) show that less than 1% of the atoms have a probability of difference greater than 40%. These differences result from differences in the interpretation by two different laboratories of a small number of ambiguous regions in the electron-density maps. In contrast, in all four comparisons between oxidized and reduced models (Figs. 5*b*, 5*c*, 5*e* and 5*f*) a cluster of atoms surrounding residues Met7, Pro35 and the loop containing three of the copper ligands (77–87) all have probabilities of difference greater than 40%. All of these residues have been previously reported as being associated with the largest features in the $F_{o,ox} - F_{o,red}$ difference maps (Nishiyama *et al.*, 1992; Vakoufari *et al.*, 1994). Although the graphs for the comparisons between the oxidized and reduced models all have very similar features, the probability of differences between the Athens oxidized and reduced models are systematically smaller, because of systematically higher displacement parameters in the Athens models.

In the interpretation of the difference probability results, the pattern of the peaks is more important than absolute height. The stereo pictures in Fig. 6 show side chains of residues in pseudoazurin containing atoms with probability of difference > 0.35, colored by probability

of difference. In the comparison between the Seattle and Athens oxidized models (Fig. 6*a*), all of the yellow to red atoms (a probability of difference between 50 and 100%) are associated with either the disordered loops or surface lysines. There are no identifiable clusters of residues with a high probability of difference associated with the copper. The differences are primarily localized to surface residues and are well distributed around the protein.

In contrast, comparison of Seattle oxidized and reduced models (Fig. 6*b*) shows yellow to red atoms tightly clustered around either the copper or Met7 and Pro35 with a few in scattered surface lysines and in the disordered loops. Residues in the cluster surrounding Met7 and Pro35 also pack against residues in the loop that contains three copper ligands. These residues have a probability of difference between 30 and 40%, and appear to provide the mechanical connection between Met7

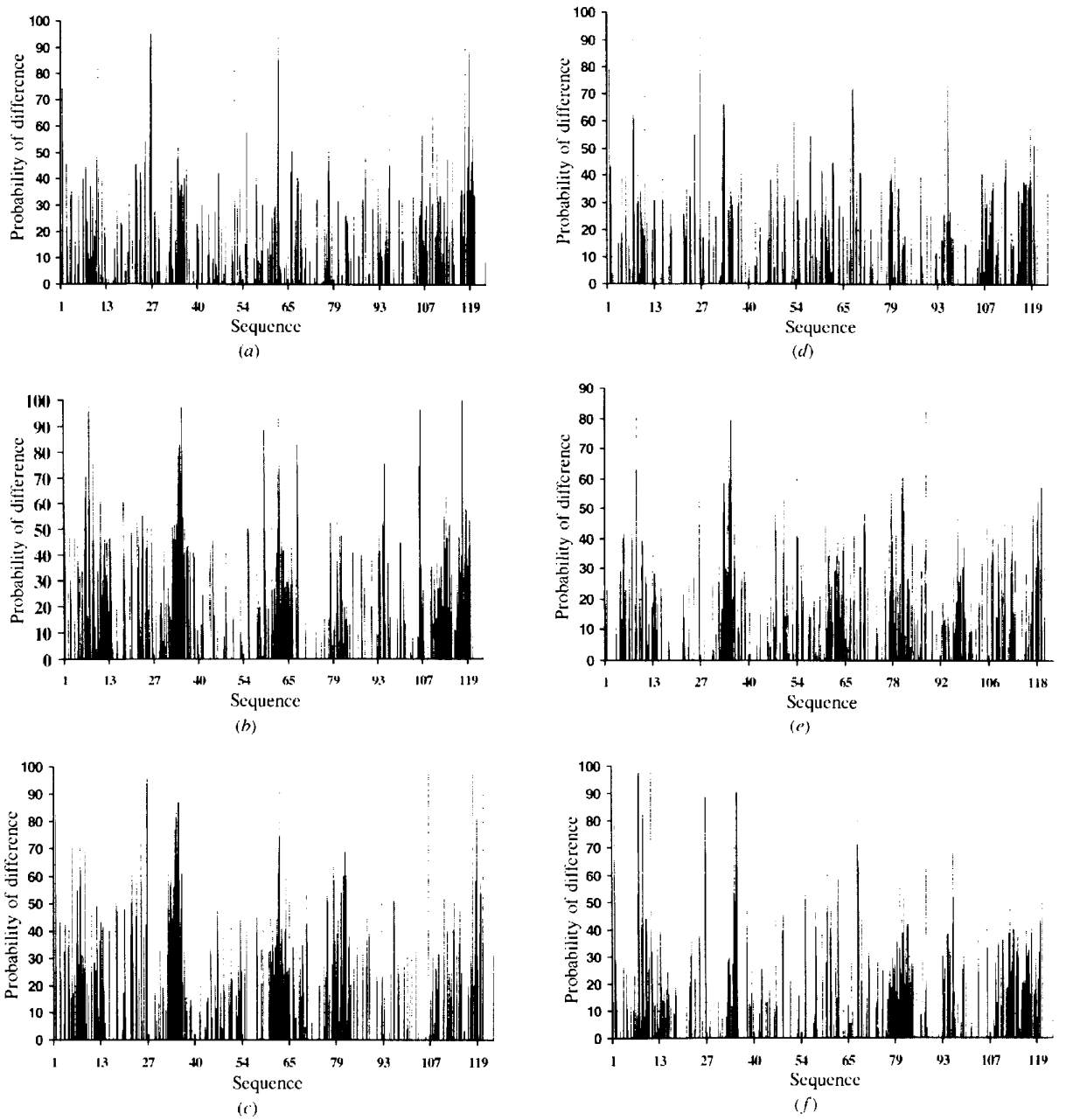


Fig. 5. DPWCC probability of difference for each atom in the pseudoazurin comparisons. (*a*) Seattle oxidized and Athens oxidized, (*b*) Seattle oxidized and Seattle reduced, (*c*) Seattle oxidized and Athens reduced, (*d*) Athens reduced and Seattle reduced, (*e*) Athens oxidized and Athens reduced and (*f*) Athens oxidized and Seattle reduced.

and Pro35 and the copper region. For example (Fig. 6*c*), the phenyl ring of Phe18 is close enough to the S^{δ} of Met86 for a dipole interaction between the aromatic ring and the sulfur. Similarly Tyr33 packs against Met7, Phe18, and Val42 as well as being sterically connected to Pro35. Val 42 has multiple conformations in the oxidized form, and only one in the reduced, consistent with these interactions. The shifts of all four atoms involved in the different tilts of the aromatic rings are so small (0.2–0.3 Å) they would not possibly be considered to be significant in the absence of probability weighting. These shifts are particularly interesting because alternating chains of S atoms and aromatic rings have been suggested to facilitate the passage of electrons through proteins (Morgan, Tatsch, Gushard, McAdon & Warme, 1978).

Also highlighted by the difference probability analysis are residues near Pro80 (red atom in center top of Fig. 6*b*) which undergoes a small change in the pucker of the proline ring upon change in oxidation state accommodating the shift of the copper and the cysteine. Mutation of Pro80 to an alanine has been shown to

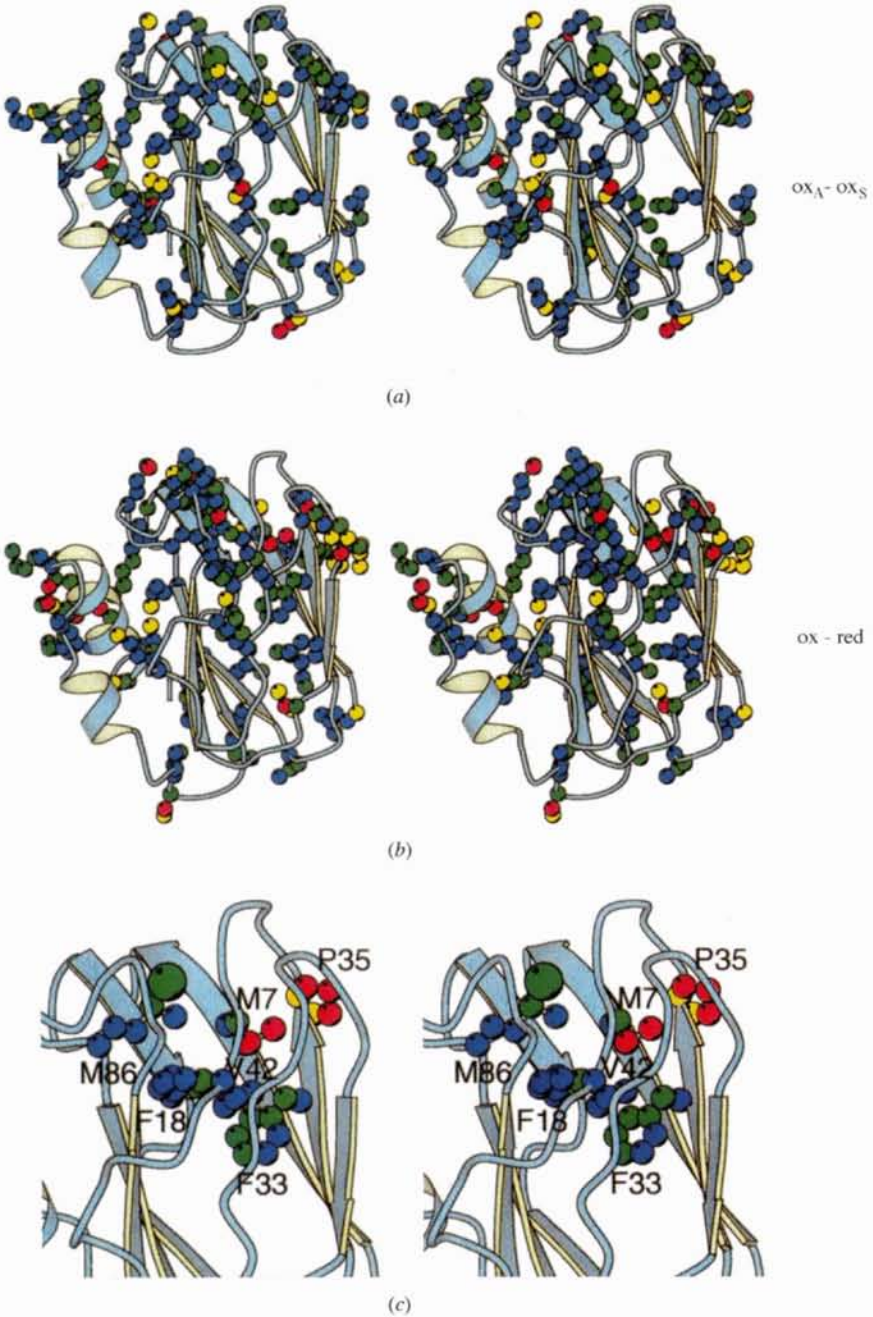


Fig. 6. Stereoviews of DPWCC probability of differences for pseudoazurin. (*a*) Seattle oxidized and Athens oxidized (*b*) Seattle oxidized and Seattle reduced and (*c*) close-up of residues involved in concerted differences (see text). The atoms are colored so that blue is a probability of difference less than 25%, green between 25–50%, yellow 50–70%, and red greater than 70%. Only residues which contain at least one atom with probability of difference >35% are shown. Figs. 6, 7, 8 and 9 were prepared using *MOLSCRIPT* (Kraulis, 1991)

Table 3. *Analysis of copper-site geometry, differences and errors*

(*a*) Bond lengths and angles

The estimated error in the difference in bond lengths, $\Delta l(1,2)$, were calculated with the formula $\sigma 1(1,2) = 6\sigma_w[(1/N_1^2) + (1/N_2^2)]^{1/2}$, where N_1 and N_2 are the atomic weights of the atoms in the bond. The estimated error in the difference in bond angles, $\Delta\varphi(1,2)$, was approximated as $\sigma\varphi(1,2) \leq \sigma_w[A_1(1/l_{11}^2 + 1/l_{12}^2) - 0.5A_2(1/l_{21}^2 + l/l_{22}^2)]$ where the coefficients A_1 and A_2 are functions of the angle φ in models 1 and 2 such that $A_i = (1/\sin^2\varphi_i)(\cos^2\varphi_i + \frac{1}{3})$, and the notation l_{ij} denotes the jth bond defining angle φ in the ith model. If i and j are assigned such that $l_{11} > l_{12}$ then $\Delta\varphi(1,2)$ is an upper bound on the estimated error of the angle determined by those bonds. This computation assumes that the positional errors are isotropic. Δ's in bold are considered significant differences.

Pseudoazurin

	$N_{40}^{\delta 1}$ (Å)	S_{78}^{γ} (Å)	$N_{81}^{\delta 1}$ (Å)	S_{86}^{δ} (Å)	N_1-S^{γ} (°)	N_1-N_2 (°)	N_1-S^{δ} (°)	$S^{\gamma}-N_2$ (°)	$S^{\gamma}-S^{\delta}$ (°)	N_2-S^{δ} (°)
A-Ox	2.16	2.16	2.12	2.76	136	100	87	112	108	112
S-Ox	2.03	2.07	2.06	2.68	136	98	87	113	109	112
S-red	2.11	2.17	2.24	2.75	140	99	90	105	109	112
A-red	2.16	2.17	2.29	2.90	140	102	85	108	106	110

Plastocyanin

	$N_{37}^{\delta 1}$ (Å)	S_{84}^{γ} (Å)	$N_{87}^{\delta 1}$ (Å)	S_{92}^{δ} (Å)	N_1-S^{γ} (°)	N_1-N_2 (°)	N_1-S^{δ} (°)	$S^{\gamma}-N_2$ (°)	$S^{\gamma}-S^{\delta}$ (°)	N_2-S^{δ} (°)
Ox 1.6	2.04	2.13	2.10	2.90	132	97	85	123	108	103
Ox 1.33	1.91	2.07	2.06	2.82	132	97	89	121	110	101
Red pH 7.8	2.12	2.11	2.25	2.90	141	92	90	112	114	102
Red pH 7.0	2.13	2.17	2.39	2.87	136	99	88	110	113	106
Red pH 3.8	2.12	21.3	5.17	2.51	141	96	95	86	124	98

(*b*) Differences (Δ) and errors (σ) estimated from DPWCC

Pseudoazurin

		$-N_{40}^{\delta 1}$ (Å)	$-S_{78}^{\gamma}$ (Å)	$-N_{81}^{\delta 1}$ (Å)	$-S_{86}^{\delta}$ (Å)	N_1-S^{γ} (°)	N_1-N_2 (°)	N_1-S^{δ} (°)	$S^{\gamma}-N_2$ (°)	$S^{\gamma}-S^{\delta}$ (°)	N_2-S^{δ} (°)
Sox–Aox	Δ	−0.13	−0.08	−0.07	−0.08	0.4	1.6	0.2	1.1	0.4	0.3
	σ	0.08	0.04	0.08	0.04	3.3	1.7	1.4	1.5	0.9	1.8
Sred–Ared	Δ	−0.05	0.00	−0.05	**−0.15**	0.1	−2.7	**4.9**	**−3.1**	**3.4**	2.0
	σ	0.08	0.04	0.08	**0.04**	4.1	2.0	**1.6**	**1.4**	**0.9**	1.9
Aox–Ared	Δ	0.00	−0.01	**−0.17**	**−0.14**	−3.9	−2.4	**1.7**	**3.6**	**2.2**	1.7
	σ	0.07	0.03	**0.07**	**0.03**	3.2	1.7	**1.4**	**1.4**	**0.9**	1.7
Sox–Sred	Δ	−0.08	−0.10	−0.18	−0.07	−4.2	−1.4	−2.9	**7.8**	0.9	0.1
	σ	0.09	0.05	0.09	0.05	4.4	2.3	1.9	**1.8**	1.2	2.4
Sox–Ared	Δ	−0.13	−0.10	**−0.23**	**−0.22**	−4.3	−4.0	1.9	**4.7**	**2.6**	2.0
	σ	0.09	0.05	**0.09**	**0.05**	3.9	2.0	1.7	**1.7**	**1.2**	2.0
Aox–Sred	Δ	0.05	−0.01	−0.12	0.01	−3.8	0.3	−3.1	**6.7**	−1.3	0.3
	σ	0.08	0.04	0.08	0.04	3.7	2.0	1.6	**1.5**	0.9	2.0

Plastocyanin

		$N_{37}^{\delta 1}$ (Å)	S_{84}^{γ} (Å)	$N_{87}^{\delta 1}$ (Å)	S_{92}^{δ} (Å)	N_1-S^{γ} (°)	N_1-N_2 (°)	N_1-S^{δ} (°)	$S^{\gamma}-N_2$ (°)	$S^{\gamma}-S^{\delta}$ (°)	N_2-S^{δ} (°)
Ox 1.6–	Δ	0.13	0.06	0.04	**0.08**	0.40	0.6	−3.1	1.9	1.8	2.1
Ox 1.3	σ	0.07	0.03	0.07	**0.03**	3.10	1.8	1.5	1.9	0.9	1.5
Ox 1.6–	Δ	−0.08	0.02	−0.14	0.00	**−8.8**	**4.8**	**−4.5**	**11.3**	**−5.5**	0.6
Red pH 7.8	σ	0.08	0.04	0.08	0.04	**3.7**	**1.9**	**1.7**	**2.6**	**1.2**	1.7
Ox 1.6–	Δ	−0.09	−0.04	**−0.29**	0.03	−4.2	−2.5	−2.5	**13.0**	**−5.3**	−3.3
Red pH 7.0	σ	0.09	0.04	**0.09**	0.04	3.4	2.0	1.6	**1.7**	**0.9**	1.7
Ox 1.6–	Δ	−0.08	0.00	**−3.04**	**0.39**	−8.5	0.6	−9.6	**36.9**	**−15.6**	4.6
Red pH 3.8	σ	0.10	0.05	**0.10**	**0.05**	4.5	2.4	2.1	**1.8**	**1.3**	1.7

(*c*) Errors from full-matrix least squares for pseudoazurin

	$-N_{40}^{\delta 1}$ (Å)	$-S_{78}^{\gamma}$ (Å)	$-N_{81}^{\delta 1}$ (Å)	$-S_{86}^{\delta}$ (Å)	N_1-S^{γ} (°)	N_1-N_2 (°)	N_1-S^{δ} (°)	$S^{\gamma}-N_2$ (°)	$S^{\gamma}-S^{\delta}$ (°)	N_2-S^{δ} (°)
Oxidized native	0.03	0.02	0.03	0.02	1	1	1	1	0.5	1
Reduced pH 7.0	0.04	0.02	0.05	0.03	1	2	1	1	0.8	1
Ox, red	0.05	0.03	0.06	0.04	1	2	1	1	0.9	1

increase the redox potential of pseudoazurin by 139 mV (Nishiyama *et al.*, 1992) and to affect the direction of the shift in position of the copper and the cysteine (Libeu *et al.*, 1997).

Although plots such as Fig. 5 were examined for plastocyanin, the most useful way to look at the results of the comparison is the color-coded figure, Fig. 7. In this one can see that the pattern of differences between oxidized and reduced plastocyanin is quite different from the pattern for pseudoazurin. For the comparison between oxidized and reduced plastocyanin at pH 7.0, the only significant shifts near the copper site are the copper itself, the S^{γ} of Cys84, a small rotation of the ring of His87, the movement of the carbonyl O atom of Cys84 and a change in the pucker of the ring of Pro86 (not shown because P_d is < 0.50) (Fig. 7*a*). Just as in pseudoazurin, the change in the pucker of the proline ring accommodates the movement of the cysteine side chain and the copper. No large shifts are seen in the region analogous to Met7 and Pro35 in pseudoazurin

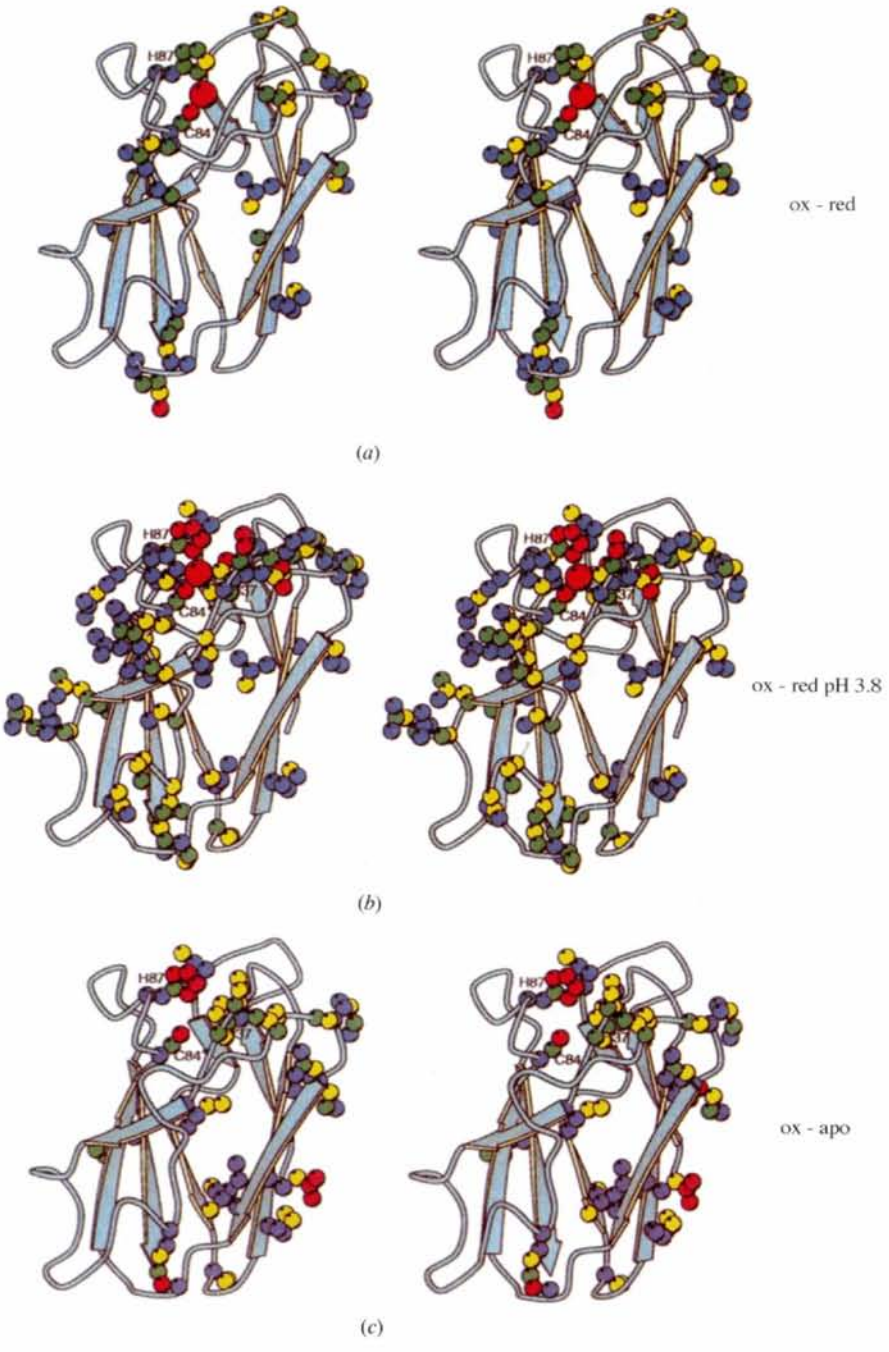
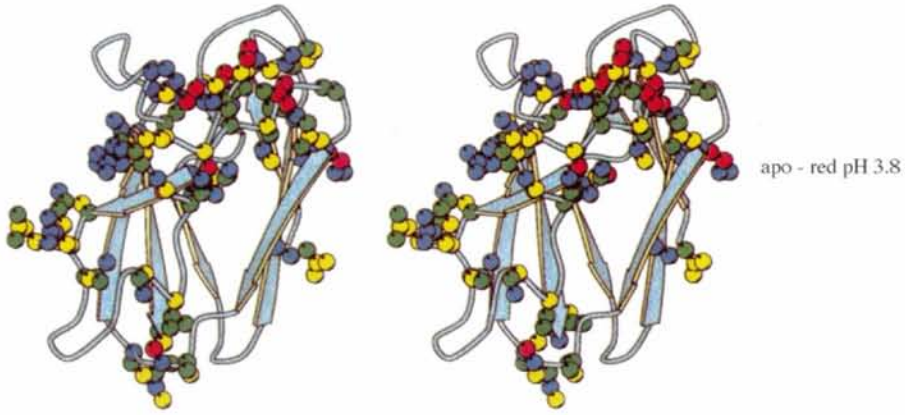


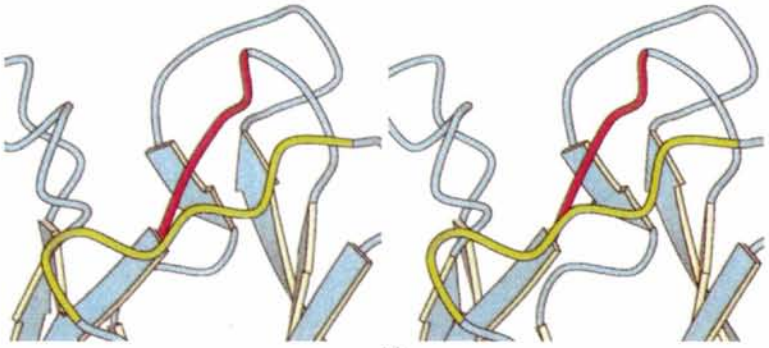
Fig. 7. Stereoviews of DPWCC probability of differences for plastocyanin. (*a*) Oxidized 1.6 Å and reduced pH 7.0, (*b*) oxidized 1.6 Å and reduced pH 3.8, and (*c*) oxidized 1.6 Å and apo.

possibly because the packing of this region is quite different in plastocyanin. Although two internal aromatic side chains are conserved and a similar pattern of small rotations of the aromatic rings is observed, residues analogous to Met7 and Pro35 are leucine and asparagine respectively, and may be more flexible in response to oxidation-state changes.

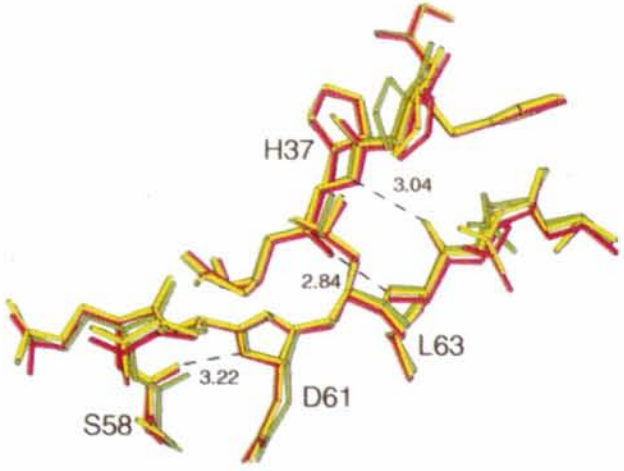
For the comparison between oxidized and reduced plastocyanin at pH 3.8 (Fig. 7*b*), DPWCC highlights changes in the positions of all four ligands, the copper as well as shifts in the loop that contains the ligand His37, and some more remote shifts in the loop containing residues 59–63 which form hydrogen bonds with residues in the His37 loop. The dramatic change in conformation of the copper site results from protonation and rotation of His87 so that the copper moves from standard distorted tetrahedral geometry to nearly trigonal geometry with the remaining three ligands (Guss *et al.*, 1986). Relaxation into trigonal geometry is accompanied by a significant increase in the distance between the



(*d*)



(*e*)



(*f*)

Fig. 7 (*cont.*) (*d*) apo and pH 3.8 (*e*) schematic showing loop 34–38 (red) and loop 58–65 (green). (*f*) Close up of loops 34–38 and 58–65 for which there are significant differences between apo (red), oxidized (yellow) and reduced pH 3.8 (green) plastocyanin. The hydrogen-bond lengths shown are from the oxidized 1.6 Å model. (*a*) to (*d*) are colored as in Fig. 6, but only side chains in which at least one atom with probability of difference > 50% are shown.

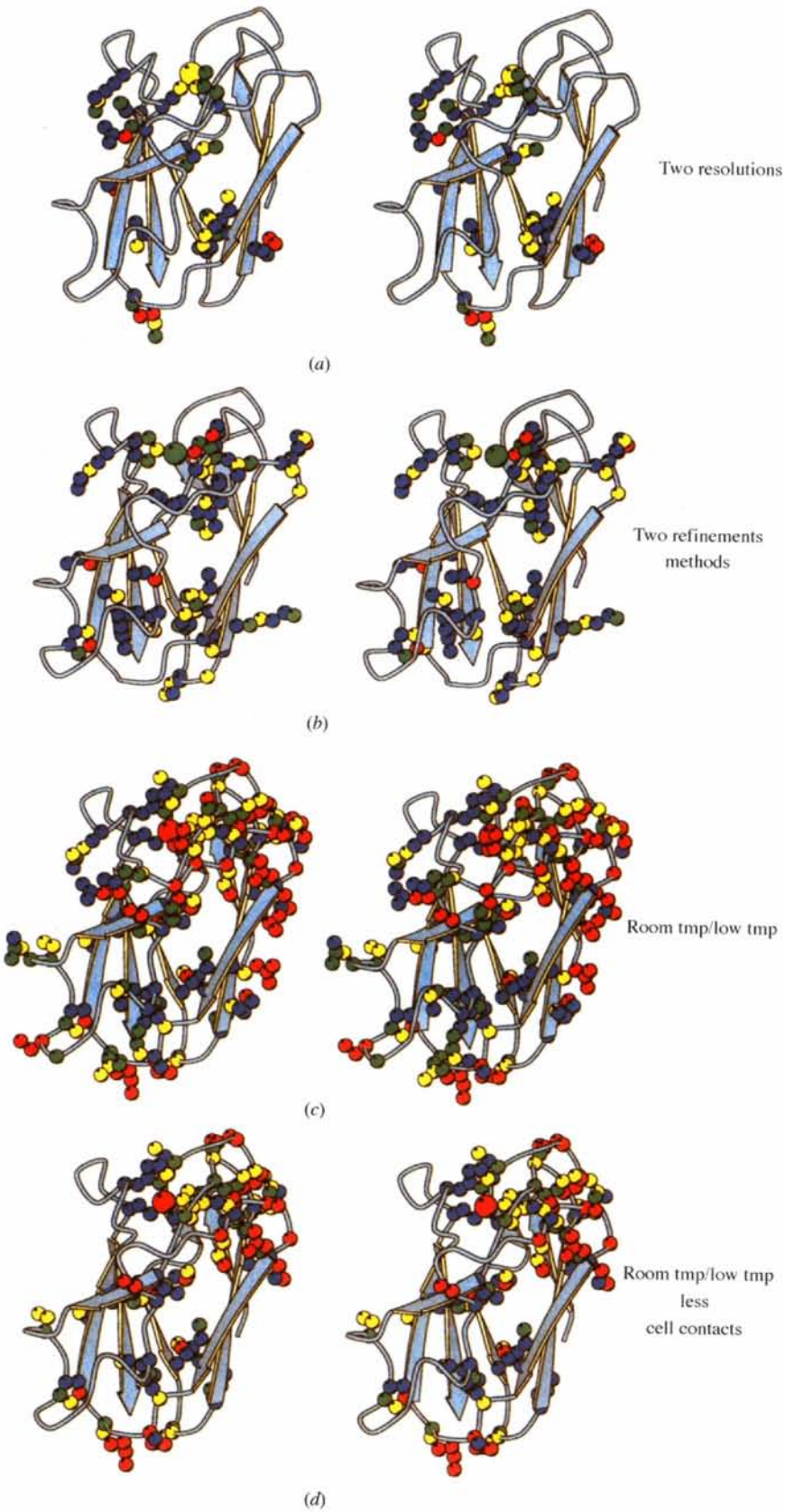


Fig. 8. Stereoviews of DPWCC probability of differences for plastocyanin. (*a*) At two resolutions: oxidized 1.6 Å and oxidized 1.3 Å models, (*b*) with two different refinement methods Ox-173 K (*PROLSQ*) and Ox-173 K (*EREF*) models and (*c*) at two different temperatures oxidized 1.6 Å and Ox-173 K (*PROLSQ*) model, all colored as in Fig. 6. (*d*) same as (*c*) excluding residues involved in crystal packing interactions. Residues involved in crystal packing interactions were defined as residues with at least one atom within 3.7 Å of a residue in a symmetry-related molecule.

carbonyl O atom of His87 and the amide N atom of Met92 as well as disruption of the hydrogen bond between the amide of His87 and the carbonyl O atom of Cys84.

A similar but less extensive cluster of changes is observed when comparing the apo and oxidized forms (Fig. 7*c*). However, the pattern of significant differences shows that the apo and pH 3.8 forms are distinct (Fig. 7*d*). As shown in Fig. 7(*f*), the oxidized form in this region (yellow) appears to be intermediate to the reduced, pH 3.8 (green) and the apo (red). Residues 59–62 are all involved in crystal contacts, so that the changes here may reflect subtle differences resulting from packing of molecules in the crystal. These differences may also be evidence that this loop adopts a wider range of conformations in solution and that the crystal form selects particular conformations. Alternatively, the difference in conformation of these loops between the reduced pH 3.8 form and the apo form may indicate that the presence of the copper influences the apparent rigidity of the copper site in plastocyanin.

Residues 59–63 also exhibit changes between room temperature and 173 K. Figs. 8(*a*) and 8(*b*) compare models of oxidized plastocyanin determined at the same temperature. In both comparisons, the structures are identical. However, significant differences are found between the oxidized models determined at different temperatures (Fig. 8*c*). In this comparison, the loop containing 59–63 as well as the loop containing His37 show significant differences. Residues which have difference probability > 0.5, and are not involved in cell contacts in either high- or low-temperature structures are shown in Fig. 8(*d*).

Many residues with high probability of difference in the high *versus* low temperature comparison are either in protein cell contacts or adjacent to them. Finding significant differences in the positions of the residues at cell contacts is not unexpected. Upon cooling, the cell volume of the plastocyanin crystals decreases by 4.2% (Fields *et al.*, 1994). There is independent evidence that the dynamical properties of plastocyanin change upon lowering the temperature: a discontinuity of the linewidth and background intensity of the resonance Raman signal associated with Cu—S^{γ} of the cysteine ligand in french bean plastocyanin has been observed between 220 and 260 K (Woodruff, Norton, Swanson & Fry, 1984). Our observation that the regions not involved in cell contacts that have significant differences upon cooling, reduction at pH 3.8, or upon removal of the copper, are the same, suggests that the origin of all the differences is most likely dynamical and that the flexibility of the loops is affected both by crystal packing and the presence and oxidation state of the copper.

Finally, for each pseudoazurin and plastocyanin comparison, σ_W was used to estimate the error in the difference in bond lengths and angles for the Cu atom and ligands. Our estimate of 0.04 Å for the comparison of oxidized and reduced plastocyanin agrees well with the error of 0.05 Å for the copper–ligand bonds previously estimated using the correlation of random errors with *B* value (Guss *et al.*, 1986). Table 3 summarizes the error estimates for bond lengths and angles of the copper centers for both pseudoazurin and plastocyanin. For each oxidized and reduced comparison, the estimated errors are larger than the combined e.s.d.'s (Table 3*c*) derived from the restrained least-squares full-matrix refinement of the oxidized and reduced pseudoazurin (Libeu *et al.*, 1997).

4. Discussion

4.1. *Analysis of replicate structures and of oxidation-state changes with DPWCC*

In the comparisons of both the pseudoazurin and plastocyanin replicate structures, a disturbing tendency for either the copper or one of the ligands to have a high probability of difference was observed even when none was expected. For example, in the comparison of Seattle oxidized and Athens oxidized pseudoazurin (Fig. 6*c*), the probability of difference for the copper was 48% for a shift of 0.03 Å and displacement parameters of 15 $Å^2$ for the Seattle model and 14 $Å^2$ for the Athens model. This relatively large change in the copper position is associated with the apparently significant difference in the bond distance for Met86 (Table 3). In our final $F_o - F_c$ difference maps for both oxidized and reduced pseudoazurin, small peaks of density that could be attributed to anisotropic movement of the copper are found near the copper site. Thus, this apparently significant difference between the Athens and Seattle models may result from the inappropriate modeling of the copper site with an isotropic temperature factor.

The tilt of the histidine rings also differs between replicate plastocyanin models, suggested to result from use of different restraints during refinement (Fields *et al.*, 1994). The Seattle pseudoazurin models refined with *X-PLOR* (Brünger, 1992) and the Athens models with *PROLSQ* (Hendrickson & Konnert, 1980) exhibit similar effects. In all the comparisons of oxidized pseudoazurin models using the same refinement program we find that the probability of difference for the copper is zero. All the well ordered Pro, Phe, Met and His side chains with moderate probabilities of difference between models refined with *X-PLOR* and models refined with *PROLSQ* are not significantly different in comparisons between oxidized pseudoazurin models refined only with *PROLSQ*. The most consistent results from DPWCC will be obtained only for comparisons in which both models were refined with the same refinement program.

In the plastocyanin comparisons, several changes at the copper site were shown to be significant according to our analysis. The Cu—N_{87}^{δ} bond significantly lengthens with lower pH and the Cu—S^{δ} bond significantly shortens. There are also significant differences between the

oxidized and reduced S^{γ}—Cu—$N_{87}{}^{\delta}$ and S^{γ}—Cu—S^{δ} angles. Comparisons between the higher resolution 1.3 Å oxidized plastocyanin model and the reduced plastocyanin models, yielded identical conclusions (data not shown).

The results for the pseudoazurin comparisons are less clear cut. Although the comparisons of each reduced model and the two oxidized models are consistent (Figs. 5 and 6), the comparisons involving the reduced Athens and the reduced Seattle models disagree as to the significant changes observed at the copper site (Table 3). The disagreement in fact may be because of different pH values of the crystals used for data collection for the reduced models. The Athens reduced model was determined at pH 7.8 and the Seattle reduced model at pH 7.0. The disagreement between the two sets of coordinates arises from a small differences in the position of the Cu and the Met S^{δ} in the Seattle reduced and Athens reduced forms. The longer Cu—S^{δ} distance in the Athens reduced form is consistent with the apparently longer distance in reduced pH 7.8 plastocyanin compared with reduced pH 7.0. In the case of plastocyanin these distances result from different admixtures of pH 7.8 (long Cu—S^{δ} bond, four-coordinate Cu, unprotonated His87) and pH 3.8 (short Cu—S^{δ} bond, three-coordinate Cu, protonated and flipped out His87) forms, where the pK of His87 is 5.1 (Sykes, 1990). In the pseudoazurin comparison, because of the high NH_4SO_4 in the stabilizing solution, the actual pH difference between the crystals from two laboratories could be negligible or as large as 1.2 pH units depending on how and when the pH was measured. Since similarity to the plastocyanin results suggested that the differences may truly represent pH-dependent behavior of the His ligand, we collected diffraction data on reduced pseudoazurin crystals at pH 7.8. Our preliminary difference maps show that there is only one pair of significant shift peaks in the $(F_{o,\ \text{red pH 7.8}} - F_{o,\ \text{red pH 7.0}})$ difference map. These peaks are located at the copper position and suggest that indeed the Cu—S^{δ} bond is longer at higher pH, as suggested by the comparison between the Athens and Seattle reduced pseudoazurin models. The reason for the pH-dependent effect is not clear. The pK_a of His81 is 4.84 in a closely related pseudoazurin (Dennison, Kohzuma, McFarlane, Suzuki, & Sykes, 1994), so that a possible pH effect is unlikely to originate from direct protonation of a copper ligand. The possible origin of the pH effect in pseudoazurin may be result from pH-dependent behavior of the region surrounding Met7, where the pK_a of an adjacent His6 has been measured in *A. cycloclastes* pseudoazurin to be 7.21 (Dennison *et al.*, 1994).

In order to find out if the significant changes of the copper ligand angles were accompanied by changes in the hydrogen-bonding network surrounding the copper site, the estimated error in the hydrogen-bond lengths was used to screen all the hydrogen bonds within a 5 Å radius of the copper. Fig. 9 shows relevant hydrogen bonds for the comparison between oxidized and reduced pseudoazurin. In comparisons of duplicate structures, no significant changes in the hydrogen bonds were found; however between the oxidized and reduced pairs an interesting pattern was observed. In both plastocyanin and pseudoazurin, the distance between the carbonyl O atom of the cysteine and the amide N atom of the histidine that detaches from the copper in the low-pH reduced form significantly lengthens, effectively severing one of the three most important hydrogen bonds for maintaining the geometry of the copper center. The loss of this constraining hydrogen bond is consistent with the observed differences in S—Cu—N^{δ} angle in both pseudoazurin and plastocyanin.

Many of the atoms highlighted by the probability of difference analysis are in non-polar residues and appear to participate in small shifts in the packing of the loop that contains the ligands Cys78, His81 and Met86. The probability of difference analysis indicates that our focus in the comparison of these structures should shift from considering only the hydrogen bonds to the copper ligands, to also considering the packing interactions made by the ligands, as seen for example in Fig. 6(*c*). The implication of all of these changes for the mechanism of reduction of pseudoazurin will be discussed in our paper that reports the structural results for the P80A and P80I mutants of pseudoazurin (Libeu *et al.*, 1997).

4.2. *Evaluating the validity of a comparison*

σ_W is only an error estimate if the parameters describing the models in the comparison are determined independently and have relatively few systematic differences. Models that have only undergone a few cycles of refinement from the starting molecular replacement solution will have very low σ_W when compared to the starting model because the models are still highly

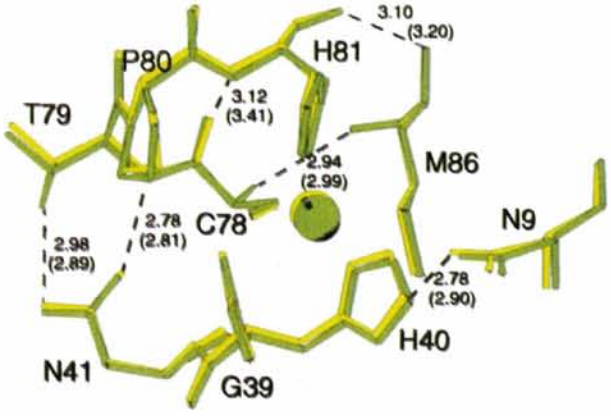


Fig. 9. Selected hydrogen bonds between protein atoms within 5 Å of the copper site Seattle oxidized (yellow) and Seattle reduced (green) models. The bond lengths in parentheses are for the Seattle reduced model.

correlated. When choosing test models from the Protein Data Bank, we did find some mutant/native model pairs that had σ_W as low as 0.02 Å and a coefficient of agreement over 0.70. In tests with models from different stages of refinement of the Seattle oxidized pseudoazurin model the agreement between the model at the end of each cycle of refinement and the original 2 Å model fell from 100% at the 0th cycle to 0.47 after 12 cycles of conjugate-gradient refinement as implemented in *PROLSQ*. Similarly, the agreement between the reduced model and the original 2 Å oxidized model was 0.49 after nine cycles of conjugate-gradient refinement. The rapid decrease in agreement between models is partly a result of the loose bond-length restraints, 0.035 Å, used in the refinement. The agreement (14) between all the final pseudoazurin structures is between 0.35 and 0.47 (Table 2), meaning that between 37 and 42% of the pairs of atoms are expected to have negligible difference distances. This proportion is consistent with visual inspection of the superimposed models, as well as with the histogram of difference distances in Fig. 2. All the backbone atoms in the strands and most of the side-chain atoms that contribute to the core of the barrel overlap completely in the four structures. We believe that for σ_W to be a valid estimate of the average error in a comparison, the agreement should not be much greater than the fraction of atoms that forms the core of the protein. For most small compact proteins, comparisons with agreements over 60% should be treated cautiously.

Too low an agreement can be a sign of a poor transformation matrix. For example, a relative displacement of 0.1 Å between the molecular centers of the two oxidized pseudoazurin models reduces the coefficient of agreement from 0.46 to 0.05. Transformation matrices calculated from the main-chain atoms are usually sufficient to align almost all the models so that a normally distributed subset of difference distances contains more than 90% of the difference distances and the agreement is over 0.25. Agreements over 0.25 imply that the weighted mean is not significantly different from zero at a 95% confidence level and that our approximation of the distribution of difference distances as a normal one is valid. An example described in Libeu & Adman (1997) on lysozyme and ribonuclease shows that when domains of the protein have moved relative to each other, it is more difficult to obtain an appropriate transformation matrix and the agreement is much less than 0.25.

4.3. *Correlation between σ_W and other estimates of the quality of models used in the comparison*

In order to evaluate the predictiveness of σ_W as an error estimate the relative quality of each model must be established independently. Comparison of the pseudoazurin entries (Table 2) highlights the difficulty comparing models that result from different refinement strategies. The Athens models have higher final R values, but are based on more complete data sets and have lower deviations from ideal bond lengths. Upper estimates of the average error in models from Luzzati plots (Luzzati, 1952) are available for three of the models and slightly favor the Seattle models. Since the Luzzati estimates are derived from the agreement between the observed and calculated structure factors in different shells of $\sin\theta/\lambda$, the disparity in the Luzzati estimates between oxidized models most likely arises from lower reflection-to-parameter ratio for the refinement of the Seattle models rather than true differences in precision between the models. The corresponding DPWCC estimate of precision, σ_W, predicts that the comparisons between the Athens models and the two oxidized models have the least error.

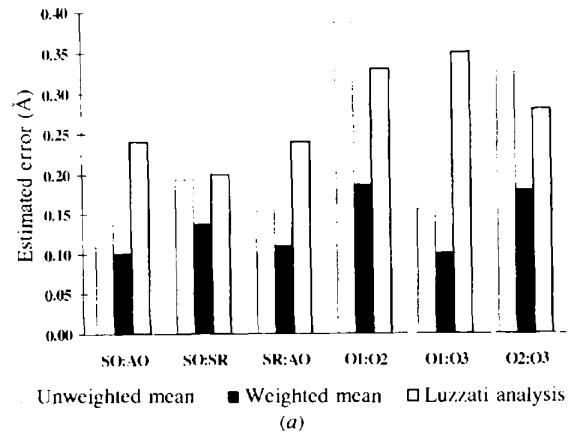
Cruickshank (1995, 1996) has proposed an alternative statistic for comparing the reliability of protein models, the diffraction precision indicator (DPI). He defines DPI as a function of the completeness of the data (C), the ratio of the number of atoms to degrees of freedom, (N/P), the resolution ($d_{\min}$) and the crystallographic R factor (R).

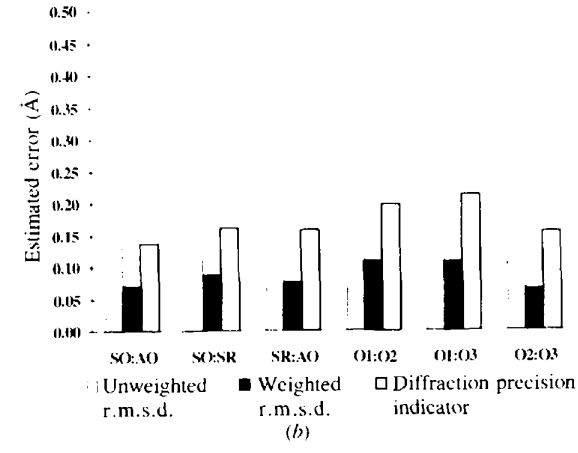
$$\sigma_d = 0.7(N/P)^{1/2}C^{-1/3}d_{\min}R. \qquad (17)$$

A statistic similar to the DPI has been shown to correlate reasonably well to e.s.d.'s of C atoms in small molecules estimated from full-matrix refinement (Allen, Cole & Howard, 1995). The DPI has the advantage that it can be calculated from commonly published parameters in the absence of the diffraction data and used to estimate a relative precision for each of the comparisons $\sigma_d(1,2)$ from,

$$\sigma_d(1,2) = \{3[\sigma_d(1)^2 + \sigma_d(2)^2]\}^{1/2}. \qquad (18)$$

The weighted and unweighted r.m.s.d. difference distances are compared with the estimated error from the DPI estimates, and the weighted and unweighted mean difference distances are compared with the Luzzati error estimates for the pseudoazurin and plastocyanin comparisons in Figs. 10(*a*) and 10(*b*). The weighted mean and σ_W are consistently smaller than the estimated error in the comparison from the appropriate source. Since both the Luzzati estimates and the DPI are derived from assumptions regarding overall agreement of structure factors, this pattern is consistent with the nature of the σ_W and the weighted mean. The weighted mean and σ_W represents the actual mean and variance in the distribution of difference distances from the refined models. This variance has contributions from both the diffraction data and the restraints, while both the DPI and the Luzzati estimate are estimates of error in the model resulting from the diffraction data alone. The inclusion of restraints during refinement acts to increase the precision in the models to better than the expected precision from the quality of the diffraction data (Cruickshank, 1996). Thus, the difference between the estimated errors from the DPI and the DPWCC results may reflect the contribution of the restraints to increasing the precision





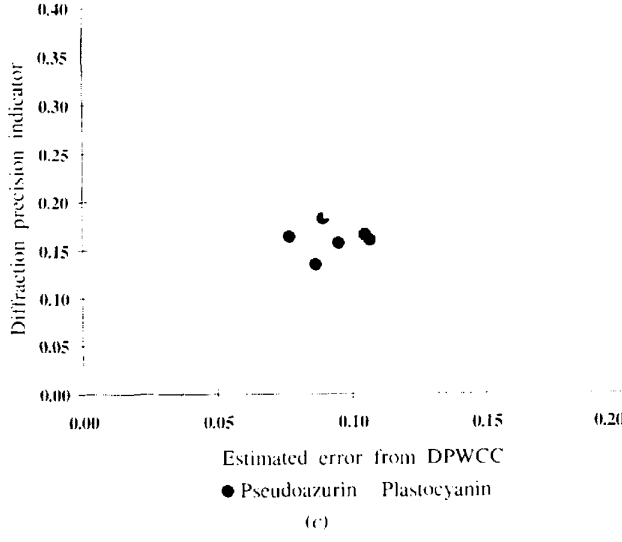


Fig. 10. Agreement between different methods of error estimation: (*a*) d_w and Luzzati estimate σ_w and (*b*) DPI for selected pseudoazurin and plastocyanin comparisons [The models are Seattle oxidized pseudoazurin (SO), Athens oxidized pseudoazurin (AO), Seattle reduced pseudoazurin (SR), oxidized 1.6 Å plastocyanin (O1), oxidized 1.3 Å plastocyanin (O2) and *PROLSQ* 173 K oxidized plastocyanin (O3)] and (*c*) σ_w and DPI for all comparisons: for pseudoazurin and for plastocyanin.

of the models. In a sense, σ_W is a lower bound to the error in the comparison while the DPI represents an upper bound (Fig. 10*c*), although it is still larger than the e.s.d.'s estimated from full-matrix restrained least-squares refinement (Table 3*c*).

4.4. *Comparison with other methods of error analysis*

Incorporation of the information in the displacement parameters into the analysis of similar structures is not a new idea. DPWCC has direct connections with the $1/B$ method (Adman *et al.*, 1989). The leading term of the Taylor expansion of the erf(x/a) about zero is x/a, so that the $1/B$ weighting method is an approximation of the full relationship for small distances. Other studies have assumed an empirical relationship between the B values and the error in the position (Chambers & Stroud, 1979; Perry *et al.*, 1990; Bott & Frane, 1990; Guss *et al.*, 1986; Stroud & Fauman, 1995). *Although superficially similar, the displacement-weighting method is a very different approach, because it does not depend on a strict correlation between the displacement parameter and the error in the position.* Defining the particular relationship between difference distances and displacement parameters for each set of models is not necessary in order to identify significantly different pairs of atoms, and in some cases may prove to be a handicap. Assumption of a *particular* relationship between the distribution of difference distance and displacement parameter can lead to poor prediction of significant differences.

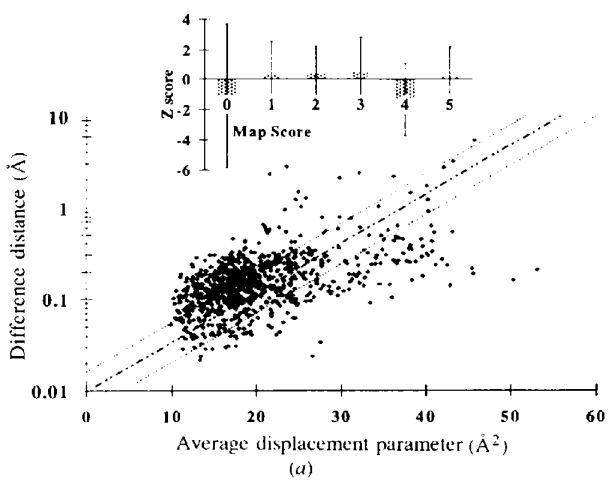
To illustrate that assuming a particular functional form for the dependence of difference distances on displacement parameter is not predictive, we compared the oxidized and reduced form of pseudoazurin using the form $\log(d) = a + bB$ (Bott & Frane, 1990), and r.m.s.(d) $= a + bB + cB^2$ (Chambers & Stroud, 1979; Guss *et al.*, 1986; Perry *et al.*, 1990) where d is the difference distance, r.m.s.(d) is the root-mean-square difference distance for a range of displacement parameters (B), B is the displacement parameter, and a, b and c are coefficients determined from the data. Fig. 11(*a*) is $\log(d)$ *versus* B, and Fig. 11(*b*), r.m.s.(d) *versus* B. The best fits to the logarithmic function and to the linear function were obtained using only average B values less than 40 Å^2. However, neither function allowed us to isolate significant differences that correlated with differences seen in difference Fourier maps, as seen in Fig. 11(*a*) (inset) and Fig. 11(*b*) (inset). Both methods resulted in scores that are not significantly correlated with the map score. The correlation coefficients are 0.07 and 0.17 respectively for 835 pairs of atoms, while the correlation coefficient between the map score and the probability of difference is 0.62. The logarithmic function predicted errors that were much larger than the observed difference distances for Met7 and Pro35, excluding them from being considered significantly different. The average displacement parameters for our pseudoazurin models are significantly higher, and were

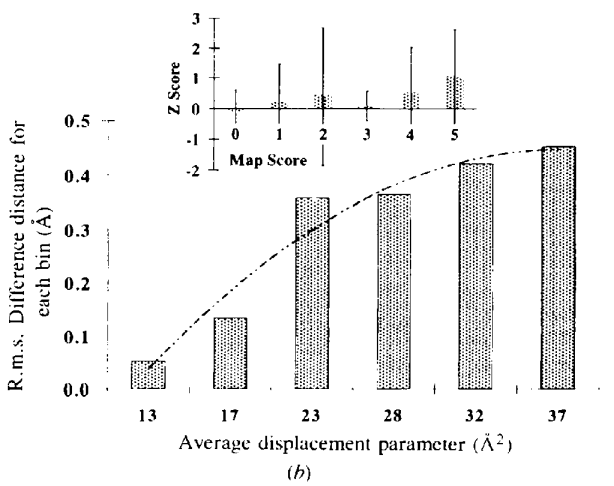
possibly less tightly restrained than those used by Bott and Frane. Larger temperature factors in pseudoazurin may contribute the failure of the logarithmic function to predict which atoms are significantly displaced between oxidized and reduced pseudoazurin.

The lack of correlation with credible features in the difference maps was largely the reason we feel that the more elaborate development by Stroud & Fauman (1995) is not appropriate, either.

The quadratic (or linear if $c = 0$) relationship of r.m.s.(d) with B to our surprise turned out to have negative curvature (Fig. 11c) unlike the fit for the plastocyanin models (Guss *et al.*, 1986) or the thymidylate synthase models (Perry *et al.*, 1990). Again this may be because of overall higher B values, and/or looser B-value restraints. In the case of thymidylate synthase, the difference distances between the models are also much larger because the two models were only 60% identical. The fit predicts difference distance deviations that are too small for pairs with average displacement parameters greater than 15 Å^2 and difference distance deviations that are too large for small B values in both cases not predicting significant map features. We noticed however that the mean difference distance *versus* the B for displacement parameters less than 40 Å^2 is linear with a correlation coefficient of 0.99 (Fig. 11c). As shown in Fig. 11(c) (inset), the scores from the linear fit to the average difference distance are more predictive of differences seen in difference maps, than the logarithmic fit. Nevertheless, even though Met7 and Pro35 are predicted to be significantly different in this case, many other atoms with low map scores are also predicted to be significantly different, reflected in the large standard deviation of difference distances for atom pairs with map scores of four and five. Thus, the linear function of mean d is also inappropriate for predicting significant peaks in difference maps.

All the methods described above that assume a particular functional form for correlating difference distance with B value also assume that difference distances represent only random error. Our method *identifies* the randomly distributed difference distances and, therefore, the systematic differences, without assuming a functional relationship between difference distance and





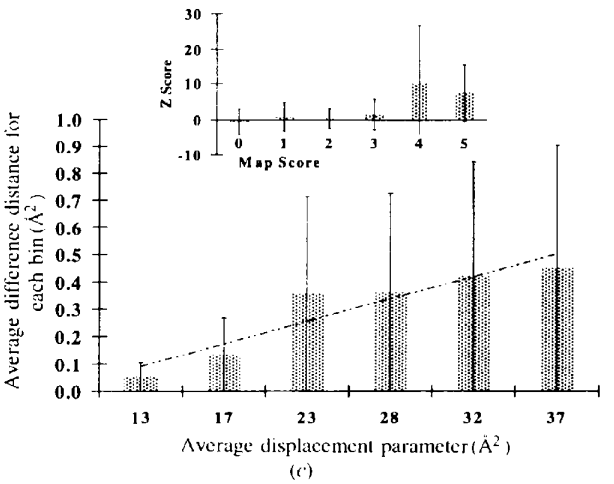


Fig. 11. Fit of difference distances to displacement parameters and correlation with map score. (*a*) Plot of log(d) *versus* B. The best fit is given by the heavy dotted line. The light dotted lines mark the fit ± RMSE (root-mean-square error). Inset: correlation of Z score with map score. The Z score for each pair of atoms is assigned by dividing the deviation from the line by the RMSE from the regression $Z = [\log(d) - a - bB]$/RMSE (Bott & Frane, 1990), where $a = 3.5$, $b = 37$ and RMSE = 0.53. The map score is defined in text. (*b*) R.m.s.(d) *versus* B value, where r.m.s.(d) is obtained from atom pairs with B values in intervals; inset: correlation of Z score with map score The Z score is $Z = [d - (a + bB + cB^2)]$/RMSE where $a = -0.0003$, $b = 0.0466$, $c = -0.5445$ and RMSE = 0.46.; (*c*) mean difference distance *versus* B value, again over 5 Å^2 intervals; inset: correlation of Z score with map score. The Z score is $Z = [d - (aB + b)]$/RMSE, where $a = 0.015$, $b = -0.096$ and RMSE = 0.99.

displacement parameter. In our study of temperature dependent structural shifts in lysozyme and ribonuclease (Libeu & Adman, 1997), we have successfully applied DPWCC analysis to protein models with very different distributions of displacement parameters. Thus, DPWCC is generally applicable and is more suitable for yielding reasonable estimates of error and also reflecting actual features seen in difference maps than previous methods.

5. Conclusions

Displacement-parameter weighting can be successfully applied to find small changes in nearly identical proteins, given that (i) the pair of proteins is similar enough that a transformation matrix can be found that superimposes more than 50% of the atoms within the radius defined by their average displacement parameter, and (ii) the pair of proteins is independently determined. Without an understanding of which residues are most likely to contain errors in the comparison (*e.g.* residues for which there is little density), one can still accept as significant, differences which are not. Although clusters of residues with high probability of difference are most likely significantly different, difficult to interpret, multiresidue disordered regions may also have high probabilities of difference depending on the restraints on the displacement parameters used in refinement.

Bearing these requirements in mind, one can evaluate the suitability of different transformation matrices, the agreement between models (15), and ultimately obtain overall estimates of error (σ_W) and thus error for bonds of interest, and quickly identify residues that are most likely to be significantly different between the models (using 16); since the differences with the highest probability of difference correlate well with what can be seen directly in difference Fourier maps. A strict cutoff for a value of the probability of difference need not be applied for all comparisons. Instead, the probability of difference for each pair assigns a relative order of differences for each comparison which can be used as a guide for interpretation of the structural results and design of future experiments.

In the case of comparing the oxidation states of pseudoazurin and plastocyanin, DPWCC analysis has caused us to shift our focus from the hydrogen-bonding network that maintains the geometry of the copper center, to considering how secondary packing and flexibility of the protein impact the stability of the copper center. While DPWCC confirms our interpretation of our difference maps for pseudoazurin, it can also be used to confidently compare structures, when combined with chemical knowledge, even if diffraction data are not available. Because of the requirement that the two models are very similar, DPWCC should be most useful for comparisons where the differences are expected to be small, such as ligand-binding studies and analyses of the effects of mutations.

Errors derived from DPWCC may be too large if comparisons contain large systematic errors and too small if models are not truly independently determined, but in general the error estimates correlate well with other independent error estimates. Errors for the copper and its ligands estimated from restrained full-matrix least-squares refinement of the pseudoazurin structures are very comparable.

This work was supported by NIH Grant GM 31770 and Molecular Biophysics Training Grant T32-GM08268. We are grateful to D. Cruickshank and Lyle Jensen for critical input to an early version of this manuscript, to Myrna Jewett for help preparing figures, and to E. Merritt for advice and encouragement in the early stages of the project.

References

Adman, E. T. (1991) *Adv. Protein Chem.* **42**, 145–197.
Adman, E. T., Turley, S., Bramson, R., Petratos, K., Banner, D., Tsernoglou, D., Beppu, T. & Watanabe, H. (1989). *J. Biol. Chem.* **264**, 87–99.
Allen, F. H., Cole, J. C. & Howard, J. A. K. (1995). *Acta Cryst.* A**51**, 95–111.
Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
Bevington, P. R. (1969). *Data Reduction and Error Analysis for the Physical Sciences*, pp. 66–91. New York: McGraw-Hill.
Bott, R. & Frane, J. (1990). *Protein Eng.* **3**, 649–657.
Brünger, A. T. (1992). *X-PLOR Version 3.1 A System for X-ray Crystallography and NMR.* New Haven: Yale University Press.
Chambers, J. L. & Stroud, R. M. (1979). *Acta Cryst.* B**35**, 1861–1874.
Church, W. B., Guss, J. M., Potter, J. J. & Freeman, H. C. (1986). *J. Biol. Chem* **261**, 234–237.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cruickshank, D. W. J. (1949). *Acta Cryst.* **2**, 65–82.
Cruickshank, D. W. J. (1995). Personal communication.
Cruickshank, D. W. J. (1996). *Proceedings of the CCP4 Study Weekend, January 1996*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 11–22. CLRC COnference Proceedings DL-Conf96–001.
Dennison, C., Kohzuma, T., McFarlane, W., Suzuki, S. & Sykes, A. G. (1994). *J. Chem. Soc. Dalton Trans.* pp. 437–443.
Fields, B. A., Bartsch, H. H., Bartunik, H. D., Cordes, F., Guss, J. M. & Freeman, H. C. (1994). *Acta Cryst.* D**50**, 709–730.
Garrett, T. P. J, Clingeleffer, D. J., Guss, J. M., Rogers, S. J. & Freeman, H. C. (1984). *J. Biol. Chem.* **259**, 2822–2825.
Guss, J. M., Bartunik, H. D. & Freeman, H. C. (1992). *Acta Cryst.* B**48**, 790–811.
Guss, J. M. & Freeman, H. C. (1983). *J. Mol. Biol.* **169**, 521–563.
Guss, J. M., Harrowell, P. K., Murata, M., Norris, V. A. & Freeman, H. C. (1986). *J. Mol. Biol.* **192**, 361–387.

Hendrickson, W. A. & Konnert, J. H. (1980). *Biomolecular Structure, Function, Conformation and Evolution*, edited by R. Srinivasan, pp. 43–57. Oxford: Pergamon Press.
James, R. W. (1982). *The Optical Principles of the Diffraction of X-rays*, pp. 20–25. Woodbridge, CT: Ox Bow Press.
Kabsch, W (1978). *Acta Cryst.* A**32**, 922–923.
Kraulis, P. (1991). *J. Appl. Cryst.* **24**, 946–950.
Libeu, C. A. P & Adman, E. T. (1997). In preparation.
Libeu, C. A. P., Kukimoto, M., Nishiyama, M., Turley, S. & Adman, E. T. (1997) In preparation.
Luzzatti, P. V. (1952). *Acta Cryst.* **5**, 802–810.
Morgan, R. S., Tatsch, C. E., Gushard, R. H., McAdon, J. M. & Warme, P. K. (1978). *Int. J. Pept. Protein Res.* **11**, 209–217.
Nishiyama, M., Suzuki, J., Ohnuki, T., Chang, H. C., Horinuchi, S., Turley, S., Adman, E. T. & Beppu, T. (1992). *Protein Eng.* **5**, 177–184.
Pagano, M. & Gauvreau, K. (1993). *Principles of Biostatistics*, pp 379–408. Belmont, CA: Wadsworth, Inc.
Perry, K. M., Fauman, E. B., Finer-Moore, J. S., Montfort, W. R., Maley, G. F., Maley, F. & Stroud, R. M. (1990). *Proteins*, **8**, 315–333.
Petratos, K., Banner, D. W., Beppu, T., Wilson, T. S. & Tsernoglou, D. (1987). *FEBS Lett.* **218**, 209–214.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Read, R. J. (1990). *Acta Cryst.* A**46**, 900–912.
Stroud, R. M. & Fauman, E. B. (1995). *Protein Sci.* **4**, 2392–2404.
Sykes, A. (1990) *Struct. Bond.* **75**, 175–224.
Vakoufari, E., Wilson, K. S. & Petratos, K. (1994). *FEBS Lett.* **347**, 203–206.
Woodruff, W. H., Norton, K. A., Swanson, B. I. & Fry, H. A. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 1263–1267.